SCIENCE CHINA
Mathematics

• ARTICLES •

October 2020 Vol. 63 No. 10: 2117–2136
https://doi.org/10.1007/s11425-018-9466-8

Statistical inference for multivariate longitudinal data with irregular auto-correlated error process

Youquan Pei[1,*], Yiming Tang[2] & Tao Huang[2]

[1]*School of Economics, Shandong University, Jinan* 250100, *China;*
[2]*School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai* 200433, *China*

Email: ypeistat@163.com, jstangyiming@163.com, huang.tao@mail.shufe.edu.cn

Received March 21, 2018; accepted December 4, 2018; published online July 6, 2020

Abstract Multivariate longitudinal data arise frequently in a variety of applications, where multiple outcomes are measured repeatedly from the same subject. In this paper, we first propose a two-stage weighted least square estimation procedure for the regression coefficients when the random error follows an irregular autoregressive (AR) process, and establish asymptotic normality properties for the resulting estimators. We then apply the smoothly clipped absolute deviation (SCAD) variable selection approach to determine the order of the AR error process. We further propose a test statistic to check whether multiple responses are correlated at the same observation time, and derive the asymptotic distribution of the proposed test statistic. Several simulated examples and real data analysis are presented to illustrate the finite-sample performance of the proposed method.

Keywords multivariate longitudinal data, autoregressive error, two-stage weighted least square, hypothesis testing

MSC(2010) 62F03, 62F05, 62F86

Citation: Pei Y Q, Tang Y M, Huang T. Statistical inference for multivariate longitudinal data with irregular auto-correlated error process. Sci China Math, 2020, 63: 2117–2136, https://doi.org/10.1007/s11425-018-9466-8

1 Introduction

Longitudinal data arise frequently in several areas of scientific research, and a variety of statistical models have been proposed in the last few decades for analyzing such data. However, most of these models are confined to the analysis of univariate longitudinal data (see [3, 4, 7, 9]).

In practice, multivariate longitudinal data can arise when a set of different outcomes of the same unit is measured repeatedly over time. For example, in a data set about the quality of paper making [12], several physical characteristics of the paper, including the tensile index (ng/g), burst index (kPa m^2/g), tear index (nN m^2/g), and drainability of pulp (Schopper-Riegler (SR) number) were repeatedly measured at beating times of 5, 15, 30, 45 and 60 minutes for 48 batches of pine sulfate pulp. Instead of modeling each longitudinal response variable separately, it is natural and important to model multivariate longitudinal responses simultaneously. Practically, it provides a unique opportunity for one to study the joint evolution

* Corresponding author

of various responses over time. Numerically, it may improve the estimation efficiency by incorporating the correlation information between various responses.

The analysis of multivariate longitudinal data can be challenging compared with that of traditional univariate longitudinal data, because (a) the error variances are likely to be different for different responses, (b) the errors are likely to be correlated for the same response measured at different times, and (c) the errors are also likely to be correlated among responses measured at the same time (see [8, 17] for a detailed description).

There are some previous studies on the statistical analysis of multivariate longitudinal data. However, nearly all of these have assumed that the error term follows normal distribution (see, e.g., [2, 6, 10]). When the model assumptions are valid, these methods should be effective. However, in practice, it is difficult to obtain sufficient prior information to properly specify parametric models for the error term. There have also been studies on parametric or nonparametric modeling of multivariate longitudinal data (see [12, 15]). However, these methods require estimations of the covariance matrix of the error term, which can be challenging when the number of observation times is large. To avoid the need to estimate the covariance matrix, Bai et al. [1] proposed an irregular-time autoregressive (AR) model with the aim of directly modeling the error process itself, where the informative correlation structure is fitted by the time distance adaptive autoregressive process that can automatically accommodate irregular and possibly subject-specific time points of longitudinal data.

In this paper, following the approach adopted by [1], we introduce a model that includes multivariate longitudinal data with an auto-correlated error process. The model is given by

$$Y_{i,j,k} = \mathbf{X}_{i,j}^{\tau}\boldsymbol{\beta}_k + \varepsilon_{i,j,k}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, m_i, \quad k = 1, \ldots, q, \tag{1.1}$$

where $Y_{i,j,k}$ denotes the k-th response observed at time t_j of the i-th subject, $\mathbf{X}_{ij} = (X_{i,j,1}, \ldots, X_{i,j,p})^{\tau}$ is a p-dimensional vector of covariates, $\boldsymbol{\beta}_k = (\beta_{k,1}, \ldots, \beta_{k,p})^{\tau}$ is a p-dimensional vector of regression coefficients for the k-th response, $\varepsilon_{i,j,k}$ represents a zero-mean stochastic process, and τ indicates the matrix transpose. In this paper, we assume that the dimensions of p and q are smaller than n.

In practice, it is known that longitudinal data are collected over a period of time and are auto-correlated. Motivated by [1], we can naturally generalize their idea from univariate longitudinal data to multivariate longitudinal data by assuming that there are correlations between various responses at the same observation time and that these correlations do not change with time.

Specifically, let $d_{i,j,s} = t_{i,j} - t_{i,j-s}$ be the time distance between the j-th and $(j-s)$-th observations for the i-th subject. We model $\varepsilon_{i,j,k}$ as

$$\varepsilon_{i,j,k} = \sum_{s=1}^{d_k}(a_{k,s} + b_{k,s}d_{i,j,s})\varepsilon_{i,j-s,k} + e_{i,j,k}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, m_i, \quad k = 1, \ldots, q, \tag{1.2}$$

where

$$\{(a_{k,s}, b_{k,s}) : k = 1, \ldots, q, s = 1, \ldots, d_k\}$$

are unknown autoregressive parameters, and $e_{i,j,k}$'s are random errors that satisfy $\mathrm{E}(\mathbf{e}_{i,j}) = \mathbf{0}$ and $\mathrm{cov}(\mathbf{e}_{i,j}) = \mathbf{V}$, where $\mathbf{e}_{i,j} = (e_{i,j,1}, \ldots, e_{i,j,q})^{\tau}$.

Note that the model (1.2) is a natural generalization of the standard AR model. It comprises a stationary part, $\sum_{s=1}^{d_k} a_{k,s}\varepsilon_{i,j-s,k}$ and a non-stationary part, $\sum_{s=1}^{d_k} b_{k,s}d_{i,j,s}\varepsilon_{i,j-s,k}$ that can capture the irregular and subject-specific characteristics of longitudinal data.

The remainder of this paper is organized as follows. In Section 2, we propose a two-stage weighted least square estimation procedure for the regression coefficients, $\boldsymbol{\beta}_k$, and autoregressive coefficient vectors, $\mathbf{a}_k = (a_{k,1}, \ldots, a_{k,d})^{\tau}$ and $\mathbf{b}_k = (b_{k,1}, \ldots, b_{k,d})^{\tau}$, and then apply the SCAD variable selection approach [5] to determine the order of the auto-correlated error process. In Section 3, we further propose a test statistic to examine whether multiple responses are correlated at the same observation time. In Section 4, we study the asymptotic properties of the proposed estimator and test statistic. Numerical simulations and an analysis of real data are presented in Section 5 to illustrate the finite-sample performances of the proposed approach. We give the conclusion in Section 6. The technical proofs are presented in Appendix A.

2 Model estimation

2.1 Initial least square estimation

For convenience, we use the same autoregressive order, d for different $\varepsilon_{i,j,k}$ $(k = 1, \ldots, q)$. There is, however, no problem in allowing the errors to depend on different orders, d_k.

To formulate the models (1.1) and (1.2) in the matrix forms, we introduce some notations:

$$\mathbf{y}_{i,j} = (y_{i,j,1}, \ldots, y_{i,j,q})^\tau, \quad \mathbf{y}_i = (\mathbf{y}_{i,1}^\tau, \ldots, \mathbf{y}_{i,m_i}^\tau)^\tau, \quad \mathbf{y} = (\mathbf{y}_1^\tau, \ldots, \mathbf{y}_n^\tau)^\tau,$$
$$\boldsymbol{\varepsilon}_{i,j} = (\epsilon_{i,j,1}, \ldots, \epsilon_{i,j,q})^\tau, \quad \boldsymbol{\varepsilon}_i = (\boldsymbol{\varepsilon}_{i,1}^\tau, \ldots, \boldsymbol{\varepsilon}_{i,m_i}^\tau)^\tau, \quad \boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1^\tau, \ldots, \boldsymbol{\varepsilon}_n^\tau)^\tau,$$
$$\mathbf{e}_{i,j} = (e_{i,j,1}, \ldots, e_{i,j,q})^\tau, \quad \mathbf{e}_i = (\mathbf{e}_{i,d+1}^\tau, \ldots, \mathbf{e}_{i,m_i}^\tau)^\tau, \quad \mathbf{e} = (\mathbf{e}_1^\tau, \ldots, \mathbf{e}_n^\tau)^\tau,$$
$$X_i = \begin{pmatrix} 1 & x_{i,1,1} & \ldots & x_{i,1,p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{i,m_i,1} & \ldots & x_{i,m_i,p} \end{pmatrix}_{m_i \times (p+1)}, \quad \beta_0 = \begin{pmatrix} \beta_{1,0} & \beta_{1,1} & \ldots & \beta_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{q,0} & \beta_{q,1} & \ldots & \beta_{q,p} \end{pmatrix}_{q \times (p+1)}.$$

Furthermore, we write $\mathbf{X} = (X_1^\tau, \ldots, X_n^\tau)^\tau$ and $\boldsymbol{\beta} = \mathrm{vec}(\beta_0)$, where $\boldsymbol{\beta}$ is created by connecting all columns of β_0 one after another. Thus, we can rewrite the model (1.1) in the matrix form as

$$\mathbf{y} = (\mathbf{X} \otimes I_q)\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2.1}$$

where $\otimes$ denotes the Kronecker product, and I_q is a $q \times q$ identity matrix. Now, we aim to consistently estimate the regression parameters $\boldsymbol{\beta}$ and the autoregressive coefficients $\{(\mathbf{a}_k, \mathbf{b}_k) : k = 1, \ldots, q\}$. The first-stage estimation procedure is shown as follows:

Assuming the errors $\varepsilon_{i,j,k}$ are independent, we obtain an initial estimate of $\boldsymbol{\beta}$ using the ordinary least square method

$$\hat{\boldsymbol{\beta}} = [(\mathbf{X} \otimes I_q)^\tau (\mathbf{X} \otimes I_q)]^{-1} (\mathbf{X} \otimes I_q)^\tau \mathbf{y}. \tag{2.2}$$

Then, given the estimated error, $\hat{\varepsilon}_{i,j,k} = y_{i,j,k} - \mathbf{X}_{i,j}^\tau \hat{\boldsymbol{\beta}}_k$, we can estimate $\mathbf{a}_k$ and $\mathbf{b}_k$ by minimizing

$$Q(\mathbf{a}_k, \mathbf{b}_k) = \sum_{i=1}^n \sum_{j=d+1}^{m_i} \left[\hat{\varepsilon}_{i,j,k} - \sum_{s=1}^d (a_{k,s} + b_{k,s} d_{i,j,s}) \hat{\varepsilon}_{i,j-s,k} \right]^2, \quad k = 1, \ldots, q.$$

Note that the estimates of $\hat{\boldsymbol{\beta}}$, $\hat{\mathbf{a}}_k$ and $\hat{\mathbf{b}}_k$ are consistent but may not be efficient, and we therefore propose a two-stage estimation approach to improve the efficiency.

2.2 Two-stage weighted least square estimation

In the second stage, we improve the estimation efficiency for the regression parameters and autoregressive coefficients by taking the correlation between responses into account. Denote

$$\boldsymbol{\eta}_{i,j} = (\hat{\varepsilon}_{i,j,1}, \ldots, \hat{\varepsilon}_{i,j,q})^\tau, \quad \boldsymbol{\eta}_i = (\boldsymbol{\eta}_{i,d+1}^\tau, \ldots, \boldsymbol{\eta}_{i,m_i}^\tau)^\tau, \quad \boldsymbol{\eta} = (\boldsymbol{\eta}_1^\tau, \ldots, \boldsymbol{\eta}_n^\tau)^\tau,$$
$$\boldsymbol{\delta}_{i,j} = \boldsymbol{\delta}_{i,j,1}^\tau \oplus \cdots \oplus \boldsymbol{\delta}_{i,j,q}^\tau, \quad \boldsymbol{\delta}_i = (\boldsymbol{\delta}_{i,d+1}, \ldots, \boldsymbol{\delta}_{i,m_i})^\tau, \quad \boldsymbol{\delta} = (\boldsymbol{\delta}_1^\tau, \ldots, \boldsymbol{\delta}_n^\tau)^\tau,$$

where $\oplus$ denotes the direct sum of vectors and $\boldsymbol{\delta}_{i,j,k} = (\hat{\varepsilon}_{i,j-1,k}, d_{i,j,1}\hat{\varepsilon}_{i,j-1,k}, \ldots, \hat{\varepsilon}_{i,j-d,k}, d_{i,j,d}\hat{\varepsilon}_{i,j-d,k})^\tau$.

We first improve the estimation efficiency for the autoregressive coefficients $\mathbf{a}_k$ and $\mathbf{b}_k$ by minimizing

$$(\boldsymbol{\eta} - \boldsymbol{\delta}\, \boldsymbol{\theta})^\tau \widehat{\boldsymbol{\Phi}}^{-1} (\boldsymbol{\eta} - \boldsymbol{\delta}\, \boldsymbol{\theta}),$$

where $\boldsymbol{\theta} = (a_{1,1}, b_{1,1}, \ldots, a_{1,d}, b_{1,d}, \ldots, a_{q,1}, b_{q,1}, \ldots, a_{q,d}, b_{q,d})^\tau$, and $\widehat{\boldsymbol{\Phi}} = \mathrm{diag}(\widehat{\boldsymbol{\Phi}}_1, \ldots, \widehat{\boldsymbol{\Phi}}_n)$ is a weight matrix. Furthermore,

$$\widehat{\boldsymbol{\Phi}}_i = \mathrm{I}_{m_i - d} \otimes \widehat{\mathbf{V}}, \quad \widehat{\mathbf{V}} = \frac{1}{n} \sum_{i=1}^n \frac{1}{m_i - d} \sum_{j=d+1}^{m_i} \hat{\mathbf{e}}_{ij} \hat{\mathbf{e}}_{ij}^\tau.$$

Note that $\mathbf{\Phi}_i$ is an $(m_i-d)q \times (m_i-d)q$ block-diagonal matrix, where each submatrix describes the correlation between the responses at the same observation time.

We then improve the estimation efficiency for the regression coefficients $\boldsymbol{\beta}$ by applying the difference-based method to the models (1.1) and (1.2),

$$y_{i,j,k}-\sum_{s=1}^{d}(a_{k,s}+b_{k,s}d_{i,j,s})(y_{i,j-s,k}-X_{i,j-s}^{\tau}\boldsymbol{\beta}_k)=X_{i,j}^{\tau}\boldsymbol{\beta}_k+e_{i,j,k},\quad j=d+1,\dots,m_i.$$

By the simple calculations, we have

$$y_{i,j,k}-\sum_{s=1}^{d}(a_{k,s}+b_{k,s}d_{i,j,s})y_{i,j-s,k}=\left[X_{i,j}^{\tau}-\sum_{s=1}^{d}(a_{k,s}+b_{k,s}d_{i,j,s})X_{i,j-s}^{\tau}\right]\boldsymbol{\beta}_k+e_{i,j,k}. \tag{2.3}$$

Let us denote

$$y_{i,j,k}^{*}=y_{i,j,k}-\sum_{s=1}^{d}(a_{k,s}+b_{k,s}d_{i,j,s})y_{i,j-s,k}\quad\text{and}\quad X_{i,j,k}^{*}=X_{i,j}-\sum_{s=1}^{d}(a_{k,s}+b_{k,s}d_{i,j,s})X_{i,j-s}.$$

Since $y_{i,j,k}^{*}$ and $X_{i,j,k}^{*}$ contain unknown parameters, $a_{k,s}$ and $b_{k,s}$, we replace them with $\tilde{y}_{i,j,k}$ and $\tilde{X}_{i,j,k}$, respectively, where

$$\tilde{y}_{i,j,k}=y_{i,j,k}-\sum_{s=1}^{d}(\hat{a}_{k,s}+\hat{b}_{k,s}d_{i,j,s})y_{i,j-s,k},\quad j=d+1,\dots,m_i,$$
$$\tilde{X}_{i,j,k}=X_{i,j}-\sum_{s=1}^{d}(\hat{a}_{k,s}+\hat{b}_{k,s}d_{i,j,s})X_{i,j-s},\quad j=d+1,\dots,m_i.$$

Because the correlation structure of the first d observations cannot be estimated based on the model (2.2) with a lag order of d, we define $\tilde{y}_{i,j,k}=y_{i,j,k}$. Denote

$$\begin{aligned}
&\tilde{\mathbf{y}}_{i,j}=(\tilde{y}_{i,j,1},\dots,\tilde{y}_{i,j,q})^{\tau},\quad \tilde{\mathbf{y}}_i=(\tilde{\mathbf{y}}_{i,1}^{\tau},\dots,\tilde{\mathbf{y}}_{i,m_i}^{\tau})^{\tau},\quad \tilde{\mathbf{y}}=(\tilde{\mathbf{y}}_1^{\tau},\dots,\tilde{\mathbf{y}}_n^{\tau})^{\tau},\\
&\tilde{\mathbf{e}}_i=(\boldsymbol{\varepsilon}_{i,1}^{\tau},\dots,\boldsymbol{\varepsilon}_{i,d}^{\tau},\mathbf{e}_{i,d+1}^{\tau},\dots,\mathbf{e}_{i,m_i}^{\tau})^{\tau},\quad \tilde{\mathbf{e}}=(\tilde{\mathbf{e}}_1^{\tau},\dots,\tilde{\mathbf{e}}_n^{\tau})^{\tau},\\
&\tilde{X}_{i,j}=\begin{cases}X_{i,j}\otimes I_q, & \text{if } j=1,\dots,d,\\ \tilde{X}_{i,j,1}\oplus\cdots\oplus\tilde{X}_{i,j,q}, & \text{if } j=d+1,\dots,m_i,\end{cases}\\
&\tilde{X}_i=(\tilde{X}_{i,1},\dots,\tilde{X}_{i,m_i})^{\tau},\quad \tilde{X}=(\tilde{X}_1^{\tau},\dots,\tilde{X}_n^{\tau})^{\tau}.
\end{aligned}$$

Hence, the multivariate longitudinal data model (2.3) can then be rewritten as

$$\tilde{\mathbf{y}}=\tilde{\mathbf{X}}\boldsymbol{\beta}+\tilde{\mathbf{e}}. \tag{2.4}$$

Using the weighted least square method, we can estimate $\boldsymbol{\beta}$ more efficiently using

$$\widetilde{\boldsymbol{\beta}}=[\widetilde{\mathbf{X}}^{\tau}\widehat{\mathbf{\Sigma}}^{-1}\widetilde{\mathbf{X}}]^{-1}\widetilde{\mathbf{X}}^{\tau}\widehat{\mathbf{\Sigma}}^{-1}\widetilde{\mathbf{y}}, \tag{2.5}$$

where $\widehat{\mathbf{\Sigma}}=\mathrm{diag}(\widehat{\mathbf{\Sigma}}_1,\dots,\widehat{\mathbf{\Sigma}}_n)$, and $\widehat{\mathbf{\Sigma}}_i=\widehat{\mathbf{W}}\oplus\widehat{\mathbf{\Phi}}_i, i=1,\dots,n$. For the first dq elements of $\tilde{\mathbf{e}}_i$, the covariance matrix can be estimated as $\widehat{\mathbf{W}}=\frac{1}{n}\sum_{i=1}^{n}\hat{\boldsymbol{\varepsilon}}_i\hat{\boldsymbol{\varepsilon}}_i^{\tau}$, where $\hat{\boldsymbol{\varepsilon}}_i=(\hat{\varepsilon}_{i,1,1},\dots,\hat{\varepsilon}_{i,1,q},\dots,\hat{\varepsilon}_{i,d,1},\dots,\hat{\varepsilon}_{i,d,q})^{\tau}$.

2.3 Determination of AR order

In practice, the true AR order in the errors is not known, and misspecification of the lagged order will result in less-efficient estimation and reduced precision of prediction. Therefore, the correct determination of the lagged order for the AR error structure is a matter that demands our attention.

With reference to the model (1.2), we can start from a large-order AR model and establish an algorithm to reduce the complexity of the model. We thus propose to simultaneously estimate the autoregressive coefficients $\mathbf{a}_k$ and $\mathbf{b}_k$, and to determine the order of the k-th AR error process by minimizing the following penalized least square function:

$$\begin{aligned} &\mathrm{P}(\mathbf{a}_k, \mathbf{b}_k) \\ &= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=d_{k0}+1}^{m_i}\left(\hat{\varepsilon}_{i,j,k} - \sum_{s=1}^{d_{k0}}(a_{k,s} + b_{k,s}d_{i,j,s})\hat{\varepsilon}_{i,j-s,k}\right)^2 + N\sum_{s=1}^{d_{k0}} p_\lambda(\|\boldsymbol{\theta}_{k,s}\|), \quad k = 1,\ldots,q, \end{aligned} \tag{2.6}$$

where d_{k0} is a pre-specified constant such that $d_{k0} > d$, λ is the tuning parameter, and $p_\lambda(\cdot)$ is the SCAD penalty function [5]. By minimizing the above objective function, we can specify the significant $a_{k,s}$ and $b_{k,s}$, and the corresponding autoregressive order.

Local quadratic approximation. We assume, without loss of generality, that $d_{k0} \equiv d_0$, and then the equivalent matrix form of (2.6) is

$$Q(\boldsymbol{\theta}) = \frac{1}{2}(\boldsymbol{\eta} - \boldsymbol{\delta}\ \boldsymbol{\theta})^\tau(\boldsymbol{\eta} - \boldsymbol{\delta}\ \boldsymbol{\theta}) + Nq\sum_{j=1}^{2d_0q} p_\lambda(|\theta_j|). \tag{2.7}$$

It is difficult to minimize (2.7) directly because the objective function is irregular at the origin and does not have the second derivative at some points. By local quadratic approximation to the SCAD penalty function [5], given an estimate $\theta_j^{(k)}$ in the k-th iteration, we set $\hat{\theta}_j = 0$ if $|\theta_j^{(k)}|$ is close to 0; otherwise, the SCAD penalty is locally approximated by a quadratic function as

$$[p_\lambda(|\theta_j|)]' = p'_\lambda(|\theta_j|)\,\mathrm{sgn}(\theta_j) \approx p'_\lambda(|\theta_j^{(k)}|)/|\theta_j^{(k)}|\theta_j.$$

We can apply an iterative ridge regression to find the minimizer of (2.7):

$$\boldsymbol{\theta}^{(k+1)} = [\boldsymbol{\delta}^\tau\boldsymbol{\delta} + Nq\ \Sigma_\lambda(\boldsymbol{\theta}^{(k)})]^{-1}\boldsymbol{\delta}^\tau\boldsymbol{\eta}, \tag{2.8}$$

where $\Sigma_\lambda(\boldsymbol{\theta}^{(k)}) = \mathrm{diag}\{p'_\lambda(|\theta_1^{(k)}|)/|\theta_1^{(k)}|, \ldots, p'_\lambda(|\theta_{2d_0q}^{(k)}|)/|\theta_{2d_0q}^{(k)}|\}$.

Tuning parameter selection. We also need to select a proper tuning parameter λ for the SCAD penalty function. Wang et al. [13] advocated using the Bayesian information criterion (BIC) tuning parameter selector for linear regression, and showed that it yields an oracle estimator in an asymptotic sense. Therefore, we propose a BIC tuning parameter selector by minimizing

$$\mathrm{BIC}(\lambda) = \log(\mathrm{RSS}/Nq) + \hat{s}\frac{\log(Nq)}{Nq},$$

where $\mathrm{RSS} = \|\boldsymbol{\eta} - \boldsymbol{\delta}\ \hat{\boldsymbol{\theta}}(\lambda)\|^2$ is the residual sum of squares (RSS) and $\hat{s}$ is the number of the estimated nonzero coefficients for a given λ.

3 Hypothesis testing

As discussed in the introduction, a number of approaches have been proposed for joint modeling of multivariate longitudinal data. However, as [11] noted, the availability of multivariate longitudinal data does not necessarily require the simultaneous construction of a joint model for all outcomes. If there are no correlations among different responses, we can directly apply the commonly used univariate longitudinal data models separately for each outcome. In this section, we are interested in examining whether multiple responses are correlated at a given observation time, say the j-th time, or equivalently, where $\Sigma_j = \mathrm{Cov}(\mathbf{e}_{i,j})$ is a diagonal matrix, i.e.,

$$H_{0j} : \Sigma_j = \mathrm{Cov}(\mathbf{e}_{i,j}) = \mathrm{diag}(\sigma_{11}^2, \ldots, \sigma_{qq}^2). \tag{3.1}$$

Define $\Delta_{i,j,k_1k_2} = e_{i,j,k_1} \cdot e_{i,j,k_2}$, $k_1 < k_2$, and there are $q(q-1)/2$ combinations of Δ_{i,j,k_1k_2}. Under the null hypothesis that the elements in $\mathbf{e}_{i,j}$ are independent, it is easy to know that Δ_{i,j,k_1k_2} is independent of Δ_{i,j,k_3k_4} given $k_1 \neq k_3$ or $k_2 \neq k_4$. In addition, as $n \to \infty$, $n^{-1/2}\sum_{i=1}^n \Delta_{i,j,k_1k_2} \xrightarrow{\mathcal{D}} N(0, \sigma^2_{k_1k_1}\sigma^2_{k_2k_2})$. Then, it holds that

$$\Lambda_j = n^{-1/2}\sum_{i=1}^n \mathbf{\Delta}_{i,j} \xrightarrow{\mathcal{D}} N(0, \Gamma_j) \quad \text{as} \quad n \to \infty,$$

where $\mathbf{\Delta}_{i,j} = (\Delta_{i,j,12}, \ldots, \Delta_{i,j,(q-1)q})^\tau$, $\Gamma_j = \mathrm{diag}(\Gamma_{j,12}, \ldots, \Gamma_{j,(q-1)q})$, and $\Gamma_{j,k_1k_2} = \sigma^2_{k_1k_1}\sigma^2_{k_2k_2}$. In practice, $e_{i,j,k}$ is unobservable and can be replaced with $\hat{e}_{i,j,k} = \hat{\varepsilon}_{i,j,k} - \sum_{s=1}^d (\hat{a}_{k,s} + \hat{b}_{ks} d_{i,j,s})\hat{\varepsilon}_{i,j-s,k}$. We thus propose a test statistic, $\hat{M}_j$ for the null hypothesis, H_{0j}, $\hat{M}_j = \hat{\Lambda}_j^\tau \ \hat{\Gamma}_j^{-1} \ \hat{\Lambda}_j$, where

$$\hat{\Lambda}_j = n^{-1/2}\sum_{i=1}^n \hat{\mathbf{\Delta}}_{i,j}, \quad \hat{\mathbf{\Delta}}_{i,j} = (\hat{\Delta}_{i,j,12}, \ldots, \hat{\Delta}_{i,j,(q-1)q})^\tau, \quad \hat{\Delta}_{i,j,k_1k_2} = \hat{e}_{i,j,k_1} \cdot \hat{e}_{i,j,k_2},$$

and $\hat{\Gamma}_j = \mathrm{diag}(\hat{\sigma}^2_{11}\hat{\sigma}^2_{22}, \ldots, \hat{\sigma}^2_{(q-1)(q-1)}\hat{\sigma}^2_{qq})$.

4 Asymptotic properties

We introduce the following regularity conditions to establish the asymptotic properties of the proposed estimators. These assumptions are imposed for the brevity of our proofs and can be weakened.

(A1) The observation times $t_{i,j}$ are independent and identically distributed (i.i.d.) from an unknown density function $f(t)$ that is defined on the support of $[0, T]$ and is uniformly bounded away from infinity and 0.

(A2) The number of measurements, m_i $(1 \leqslant i \leqslant n)$ are i.i.d. with $0 < \mathrm{E}(m_i) < \infty$.

(A3) The time series processes, $\mathrm{Z}_i = \{(X_{i,j}, \varepsilon_{i,j}) : 1 \leqslant j \leqslant m_i\}$ are independent across i and independent of $e_{i,j}$. Moreover, we assume that the second moment of X exists, that is $\mathrm{E}[\mathbf{X}^\tau_{i,j}\mathbf{X}_{i,j}] \leqslant C < \infty$ for some finite constant C.

Assumption (A1) is a standard assumption for modeling longitudinal data (see [4]). Under (A2), the total sample size, $N = \sum_{i=1}^n m_i$ is of the same order as the number of subjects n. (A3) is a technical assumption needed to establish the asymptotic properties of $\hat{\boldsymbol{\beta}}$. Note that the number of observation times for every subject, m_i is not required to be bounded and our simulation results in Section 5 can verify this assumption.

Throughout this paper, we will use $\hat{\boldsymbol{\beta}}_{\mathrm{ols}}$ and $\hat{\boldsymbol{\theta}}_{\mathrm{ols}}$ to denote the first-stage estimators, and $\hat{\boldsymbol{\beta}}_{\mathrm{wls}}$ and $\hat{\boldsymbol{\theta}}_{\mathrm{wls}}$ to denote the two-stage estimators.

4.1 Asymptotic properties when correlations among responses are ignored

We first introduce some notations. Let $\eta(t) = (\eta_1(t), \ldots, \eta_p(t))^\tau$ be defined by the equation

$$X_{i,j} = \eta(t_{i,j}) + \zeta_{i,j}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, m_i,$$

where $\zeta_{i,j} = (\zeta_{i,j,1}, \ldots, \zeta_{i,j,p})^\tau$ satisfies $\mathrm{E}(\zeta_{i,j} \,|\, t_{i,j}) = 0$. Here, the conditional expectation is considered componentwise. Denote

$$\begin{aligned}
\tilde{\zeta}^{(k)}_{i,j} &= \zeta_{i,j} - \sum_{s=1}^{d_k}(a_{k,s} + b_{k,s}d_{i,j,s})\zeta_{i,j-s}, \\
\tilde{\zeta}_{i,j} &= \tilde{\zeta}^{(1)}_{i,j} \oplus \cdots \oplus \tilde{\zeta}^{(q)}_{i,j}, \quad \delta_{i,j} = \delta_{i,j,1} \oplus \cdots \oplus \delta_{i,j,q}, \\
\delta_{i,j,k} &= (\varepsilon_{i,j-1,k}, d_{i,j,1}\varepsilon_{i,j-1,k}, \ldots, \varepsilon_{i,j-d_k,k}, d_{i,j,d_k}\varepsilon_{i,j-d_k,k})^\tau.
\end{aligned}$$

We assume

$$\frac{1}{Nq}\sum_{i=1}^n \left[\sum_{j=1}^d (X_{i,j} \otimes I_q)(X_{i,j} \otimes I_q)^\tau + \sum_{j=d+1}^{m_i} \tilde{X}_{i,j}\tilde{X}^\tau_{i,j}\right] \xrightarrow{\mathcal{P}} D, \tag{4.1}$$

$$\frac{1}{Nq-ndq}\sum_{i=1}^{n}\sum_{j=d+1}^{m_i}\delta_{i,j}\delta_{i,j}^{\tau}\xrightarrow{\mathcal{P}}\Lambda, \tag{4.2}$$

$$\frac{1}{Nq}\sum_{i=1}^{n}\left\{\left[\sum_{j=1}^{d}(X_{i,j}\otimes I_q)\boldsymbol{\varepsilon}_{i,j}\right]\left[\sum_{j=1}^{d}(X_{i,j}\otimes I_q)\boldsymbol{\varepsilon}_{i,j}\right]^{\tau}+\sigma_e^2\sum_{j=d+1}^{m_i}\tilde{X}_{i,j}\tilde{X}_{i,j}^{\tau}\right\}\xrightarrow{\mathcal{P}}\Delta. \tag{4.3}$$

We then have the following asymptotic results for the ordinary least square estimators, $\hat{\boldsymbol{\beta}}_{\rm ols}$ and $\hat{\boldsymbol{\theta}}_{\rm ols}$ proposed in Subsection 2.1.

Theorem 4.1. *Under regularity conditions* (A1)–(A3) *and* (4.1)–(4.3), *as* $n\to\infty$, *we have*

(i) $\sqrt{Nq}\,(\hat{\boldsymbol{\beta}}_{\rm ols}-\boldsymbol{\beta})\xrightarrow{\mathcal{D}}N(0,D^{-1}\Delta D^{-1})$, *where* D *and* Δ *are defined in* (4.1) *and* (4.3), *respectively*;

(ii) $\sqrt{Nq-ndq}\,(\hat{\boldsymbol{\theta}}_{\rm ols}-\boldsymbol{\theta})\xrightarrow{\mathcal{D}}N(0,\sigma_e^2\Lambda^{-1})$, *where* Λ *is defined in* (4.2) *and* σ_e^2 *is the variance of* $e_{i,j,k}$.

The proofs of Theorem 4.1 and the results below can be found in Appendix A. We now consider the estimations of the covariance matrices for the estimators, which involve estimations of σ_e^2, D, Δ and Λ. First, we estimate σ_e^2 using

$$\hat{\sigma}_e^2=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i-d}\sum_{j=d+1}^{m_i}\frac{1}{q}\sum_{k=1}^{q}\left[\hat{\varepsilon}_{i,j,k}-\sum_{s=1}^{d_k}(\hat{a}_{k,s}+\hat{b}_{k,s}d_{i,j,s})\hat{\varepsilon}_{i,j-s,k}\right]^2,$$

where $\hat{\varepsilon}_{i,j,k}=y_{i,j,k}-X_{i,j}^{\tau}\hat{\boldsymbol{\beta}}_k, i=1,\dots,n, j=1,\dots,m_i, k=1,\dots,q$. Recall that

$$\tilde{X}_{i,j,k}=X_{i,j}-\sum_{s=1}^{d}(\hat{a}_{k,s}+\hat{b}_{k,s}d_{i,j,s})X_{i,j-s},\quad \tilde{X}_{i,j}=\tilde{X}_{i,j,1}\oplus\cdots\oplus\tilde{X}_{i,j,q},$$
$$\hat{\delta}_{i,j,k}=(\hat{\varepsilon}_{i,j-1,k},d_{i,j,1}\hat{\varepsilon}_{i,j-1,k},\dots,\hat{\varepsilon}_{i,j-d,k},d_{i,j,d}\hat{\varepsilon}_{i,j-d_k,k})^{\tau},\quad \hat{\delta}_{i,j}=\hat{\delta}_{i,j,1}\oplus\cdots\oplus\hat{\delta}_{i,j,q}.$$

We then estimate D, Δ and Λ respectively using

$$\hat{D}=\frac{1}{Nq}\sum_{i=1}^{n}\left[\sum_{j=1}^{d}(X_{i,j}\otimes I_q)(X_{i,j}\otimes I_q)^{\tau}+\sum_{j=d+1}^{m_i}\tilde{X}_{i,j}\tilde{X}_{i,j}^{\tau}\right],$$
$$\hat{\Lambda}=\frac{1}{Nq-ndq}\sum_{i=1}^{n}\sum_{j=d+1}^{m_i}\hat{\delta}_{i,j}\hat{\delta}_{i,j}^{\tau},$$
$$\hat{\Delta}=\frac{1}{Nq}\sum_{i=1}^{n}\left\{\left[\sum_{j=1}^{d}(X_{i,j}\otimes I_q)\hat{\boldsymbol{\varepsilon}}_{i,j}\right]\left[\sum_{j=1}^{d}(X_{i,j}\otimes I_q)\hat{\boldsymbol{\varepsilon}}_{i,j}\right]^{\tau}+\hat{\sigma}_e^2\sum_{j=d+1}^{m_i}\tilde{X}_{i,j}\tilde{X}_{i,j}^{\tau}\right\}.$$

Theorem 4.2. *Under regularity conditions* (A1)–(A3) *and* (4.1)–(4.3), *as* $n\to\infty$, *we have*

$$\sqrt{Nq-ndq}\,(\hat{\sigma}_e^2-\sigma_e^2)\xrightarrow{\mathcal{D}}N(0,\mathrm{var}(e_{i,j,k}^2))$$

and

$$\hat{D}\xrightarrow{\mathcal{P}}D,\quad \hat{\Lambda}\xrightarrow{\mathcal{P}}\Lambda,\quad \hat{\Delta}\xrightarrow{\mathcal{P}}\Delta.$$

4.2 Asymptotic properties when correlations among responses are taken into account

We define

$$\mathrm{cov}\{(e_{i,j,1},\dots,e_{i,j,q})\,|\,t_{i,d+1},\dots,t_{i,m_i}\}=V,\quad \mathrm{cov}\{(\varepsilon_{i,j,1},\dots,\varepsilon_{i,j,q})\,|\,t_{i,d+1},\dots,t_{i,m_i}\}=W,$$

and assume that

$$\frac{1}{Nq}\sum_{i=1}^{n}\left[\sum_{j=1}^{d}(X_{i,j}\otimes I_q)W(X_{i,j}\otimes I_q)^{\tau}+\sum_{j=d+1}^{m_i}\tilde{X}_{i,j}V\tilde{X}_{i,j}^{\tau}\right]\xrightarrow{\mathcal{P}}C, \tag{4.4}$$

$$\frac{1}{Nq}\sum_{i=1}^{n}\bigg\{\bigg[\sum_{j=1}^{d}(X_{i,j}\otimes I_q)\boldsymbol{\varepsilon}_{i,j}\bigg]\bigg[\sum_{j=1}^{d}(X_{i,j}\otimes I_q)\boldsymbol{\varepsilon}_{i,j}\bigg]^{\tau} + \bigg(\sum_{j=d+1}^{m_i}\tilde{X}_{i,j}\mathbf{e}_{i,j}\bigg)\bigg(\sum_{j=d+1}^{m_i}\tilde{X}_{i,j}\mathbf{e}_{i,j}\bigg)^{\tau}\bigg\}\xrightarrow{\mathcal{P}}\Theta, \tag{4.5}$$

$$\frac{1}{Nq-ndq}\sum_{i=1}^{n}\sum_{j=d+1}^{m_i}\delta_{i,j}V\delta_{i,j}^{\tau}\xrightarrow{\mathcal{P}}A, \tag{4.6}$$

$$\frac{1}{Nq-ndq}\sum_{i=1}^{n}\sum_{j=d+1}^{m_i}\delta_{i,j}V\mathbf{e}_{i,j}\xrightarrow{\mathcal{P}}B. \tag{4.7}$$

Subsequently, we have the following asymptotic results for the two-stage weighted least square estimators, $\hat{\boldsymbol{\beta}}_{\text{wls}}$ and $\hat{\boldsymbol{\theta}}_{\text{wls}}$ proposed in Subsection 2.2.

Theorem 4.3. *Under regularity conditions* (A1)–(A3) *and* (4.4)–(4.7)*, as* $n\to\infty$*, we have*

(i) $\sqrt{Nq}\,(\hat{\boldsymbol{\beta}}_{\text{wls}}-\boldsymbol{\beta})\xrightarrow{\mathcal{D}}N(0,C^{-1}\Theta C^{-1})$*, where* C *and* Θ *are defined in* (4.4) *and* (4.5)*, respectively;*

(ii) $\sqrt{Nq-ndq}\,(\hat{\boldsymbol{\theta}}_{\text{wls}}-\boldsymbol{\theta})\xrightarrow{\mathcal{D}}N(0,A^{-1}BA^{-1})$*, where* A *and* B *are defined in* (4.6) *and* (4.7)*, respectively.*

The proofs of Theorem 4.3 and the results below can be found in Appendix A. We now consider the estimations of the covariance matrices for the estimators, which involve estimations of V, W, C, Θ, A and B. First, we estimate V and W using

$$\hat{V}=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i-d}\sum_{j=d+1}^{m_i}\hat{\mathbf{e}}_{ij}\hat{\mathbf{e}}_{ij}^{\tau},\quad \hat{W}=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{d}\sum_{j=1}^{d}\hat{\boldsymbol{\varepsilon}}_{ij}\hat{\boldsymbol{\varepsilon}}_{ij}^{\tau},$$

where

$$\hat{\boldsymbol{\varepsilon}}_{ij}=(\hat{\varepsilon}_{i,j,1},\ldots,\hat{\varepsilon}_{i,j,q})^{\tau},\quad \hat{\mathbf{e}}_{ij}=(\hat{e}_{i,j,1},\ldots,\hat{e}_{i,j,q})^{\tau},\quad \hat{e}_{i,j,k}=\hat{\varepsilon}_{i,j,k}-\sum_{s=1}^{d_k}(\hat{a}_{k,s}+\hat{b}_{k,s}d_{i,j,s})\hat{\varepsilon}_{i,j-s,k}.$$

We then estimate C, Θ, A and B using

$$\hat{C}=\frac{1}{Nq}\sum_{i=1}^{n}\bigg[\sum_{j=1}^{d}(X_{i,j}\otimes I_q)\hat{W}(X_{i,j}\otimes I_q)^{\tau}+\sum_{j=d+1}^{m_i}\tilde{X}_{i,j}\hat{V}\tilde{X}_{i,j}^{\tau}\bigg],$$

$$\hat{\Theta}=\frac{1}{Nq}\sum_{i=1}^{n}\bigg\{\bigg[\sum_{j=1}^{d}(X_{i,j}\otimes I_q)\hat{\boldsymbol{\varepsilon}}_{i,j}\bigg]\bigg[\sum_{j=1}^{d}(X_{i,j}\otimes I_q)\hat{\boldsymbol{\varepsilon}}_{i,j}\bigg]^{\tau}+\bigg(\sum_{j=d+1}^{m_i}\tilde{X}_{i,j}\hat{\mathbf{e}}_{i,j}\bigg)\bigg(\sum_{j=d+1}^{m_i}\tilde{X}_{i,j}\hat{\mathbf{e}}_{i,j}\bigg)^{\tau}\bigg\},$$

$$\hat{A}=\frac{1}{Nq-ndq}\sum_{i=1}^{n}\sum_{j=d+1}^{m_i}\hat{\delta}_{i,j}\hat{V}\hat{\delta}_{i,j}^{\tau},\quad \hat{B}=\frac{1}{Nq-ndq}\sum_{i=1}^{n}\sum_{j=d+1}^{m_i}\hat{\delta}_{i,j}\hat{V}\hat{\mathbf{e}}_{i,j}.$$

Theorem 4.4. *Under regularity conditions* (A1)–(A3) *and* (4.4)–(4.7)*, as* $n\to\infty$*, we have*

$$\sqrt{N-nd}\,[\text{vech}(\hat{V})-\text{vech}(V)]\xrightarrow{\mathcal{D}}N(0,L_q\,\text{cov}(\mathbf{e}_{i,j}\otimes\mathbf{e}_{i,j})L_q^{\tau}),$$

where vech is a column stacking operator that stacks only the elements on or below the main diagonal of the matrix, L_q is the $\frac{1}{2}q(q+1)\times q^2$ elimination matrix, and $\mathbf{e}_{i,j}=(e_{i,j,1},\ldots,e_{i,j,q})^{\tau}$.

Theorem 4.5. *Under regularity conditions* (A1)–(A3) *and the null hypothesis,* H_{0j}*, as* $n\to\infty$*, we have* $\hat{M}_j\xrightarrow{D}\chi^2_{q(q-1)/2}$*,* $j=1,\ldots,m_i$*.*

5 Numerical studies

5.1 Simulations

We conduct several simulation studies to examine the finite-sample performance of our proposed methods. We first consider a dataset-generating procedure in the case where the true AR order is known.

Example 5.1. The simulated data are generated from the linear model with AR(1) error processes

$$y_{i,j,k} = X_{i,j}^{\tau}\boldsymbol{\beta}_k + \varepsilon_{i,j,k},$$
$$\varepsilon_{i,j,k} = [a_k + b_k(t_{i,j} - t_{i,j-1})]\varepsilon_{i,j-1,k} + e_{i,j,k},$$

where $X_{ij} = \sin(t_{ij}) + \varrho_{ij}$, and ϱ_{ij} follows a standard normal distribution. The observation time t_{ij} follows a uniform distribution on $(0,1)$, and the random error $e_{i,j,k}$ follows a multivariate normal distribution with a zero mean vector and a compound symmetric covariance matrix with ones on its main diagonal and ρ on its off-diagonal entries.

We consider three choices of $\rho = 0.9$, $\rho = 0.6$ and $\rho = 0.3$, corresponding to strongly, moderately and weakly correlated errors, respectively. The regression coefficients $\boldsymbol{\beta}_1 = (0.5, 0.8)$, $\boldsymbol{\beta}_2 = (1.0, 2.0)$ and $\boldsymbol{\beta}_3 = (1.2, 1.5)$, and the autoregressive coefficients $(a_1, b_1) = (0.6, -0.3)$, $(a_2, b_2) = (0.3, -0.4)$ and $(a_3, b_3) = (-0.5, 0.3)$. The sample size, $n = 50$, $n = 100$ and $n = 200$, and the number of within-subject observations, $m_i = m = 5$, $m_i = m = 10$ and $m_i = m = 15$ for each subject. We run the simulation 100 times for each setting.

We compare our proposed method with the univariate method of [1] that ignores correlations between the multiple responses and model them separately. Let us denote the proposed weighted least square estimator by $\hat{\boldsymbol{\beta}}$ and the ordinary least square estimator proposed by [1] $\check{\boldsymbol{\beta}}$, respectively. The average bias (Bias), and average empirical standard deviation (SD) of both estimates calculated on the basis of 100 simulations, are summarized in Table 1.

Table 1 Average bias (Bias) and average empirical standard deviation (SD) of the estimates for regression coefficients calculated from 100 simulations

$\rho = 0.9$	$\hat{\beta}$	m	5			10			15		
		n	50	100	200	50	100	200	50	100	200
First response	$\hat{\beta}_{1,0}$	Bias	0.0063	−0.0020	0.0023	−0.0139	0.0094	−0.0036	0.0059	−0.0068	0.0015
		SD	0.0810	0.0585	0.0336	0.0677	0.0463	0.0368	0.0618	0.0464	0.0294
	$\check{\beta}_{1,0}$	Bias	0.0093	−0.0020	0.0042	−0.0112	0.0091	−0.0044	0.0018	−0.0065	0.0014
		SD	0.0956	0.0717	0.0405	0.0752	0.0544	0.0389	0.0742	0.0521	0.0326
	$\hat{\beta}_{1,1}$	Bias	−0.0036	0.0003	−0.0012	0.0003	−0.0009	−0.0007	0.0023	−0.0005	−0.0004
		SD	0.0351	0.0189	0.0148	0.0221	0.0140	0.0105	0.0168	0.0110	0.0086
	$\check{\beta}_{1,1}$	Bias	−0.0032	−0.0018	−0.0016	0.0014	−0.0016	−0.0002	0.0038	−0.0032	0.0026
		SD	0.0522	0.0372	0.0279	0.0355	0.0265	0.0184	0.0311	0.0232	0.0150
Second response	$\hat{\beta}_{2,0}$	Bias	0.0053	−0.0027	0.0028	−0.0039	0.0067	−0.0004	−0.0009	0.0007	0.0002
		SD	0.0600	0.0396	0.0234	0.0429	0.0287	0.0212	0.0397	0.0304	0.0188
	$\check{\beta}_{2,0}$	Bias	0.0050	−0.0076	0.0049	−0.0016	0.0083	−0.0009	−0.0059	0.0008	−0.0011
		SD	0.0795	0.0500	0.0320	0.0588	0.0349	0.0271	0.0483	0.0375	0.0207
	$\hat{\beta}_{2,1}$	Bias	−0.0041	0.0010	−0.0005	−0.0002	−0.0028	−0.0013	0.0033	−0.0014	0.0001
		SD	0.0318	0.0219	0.0172	0.0245	0.0162	0.0130	0.0183	0.0131	0.0089
	$\check{\beta}_{2,1}$	Bias	−0.0021	0.0019	−0.0010	−0.0001	−0.0048	−0.0010	0.0056	−0.0032	0.0034
		SD	0.0540	0.0374	0.0293	0.0429	0.0261	0.0218	0.0369	0.0268	0.0157
Third response	$\hat{\beta}_{3,0}$	Bias	−0.0015	0.0009	0.0002	−0.0056	0.0020	−0.0001	0.0026	0.0002	0.0016
		SD	0.0439	0.0305	0.0178	0.0303	0.0181	0.0137	0.0228	0.0163	0.0116
	$\check{\beta}_{3,0}$	Bias	−0.0069	0.0010	0.0045	−0.0028	0.0030	0.0004	−0.0035	−0.0006	0.0021
		SD	0.0571	0.0411	0.0262	0.0437	0.0279	0.0178	0.0322	0.0243	0.0148
	$\hat{\beta}_{3,1}$	Bias	0.0004	0.0019	−0.0010	0.0021	−0.0002	−0.0005	0.0011	−0.0015	−0.0007
		SD	0.0346	0.0207	0.0169	0.0197	0.0138	0.0112	0.0158	0.0111	0.0083
	$\check{\beta}_{3,1}$	Bias	0.0083	−0.0014	−0.0045	0.0004	0.0005	−0.0019	0.0052	−0.0009	−0.0003
		SD	0.0560	0.0359	0.0276	0.0392	0.0252	0.0192	0.0294	0.0252	0.0166

Note. $\hat{\beta}$ represents the proposed weighted least square estimator and $\check{\beta}$ represents the naive ordinary least square estimator.

Table 2 Average bias (Bias) and average empirical standard deviation (SD) of the estimates for relatively large observation time m calculated from 100 simulations

$\rho = 0.9$	$\hat{\beta}$	m	15			30			60		
		n	30	40	50	30	40	50	30	40	50
First response	$\hat{\beta}_{1,0}$	Bias	−0.0088	0.0048	0.0059	−0.0093	0.0027	0.0021	−0.0036	0.0043	0.0009
		SD	0.0713	0.0664	0.0618	0.0570	0.0521	0.0454	0.0493	0.0383	0.0335
	$\hat{\beta}_{1,1}$	Bias	0.0019	−0.0019	0.0023	−0.0006	−0.0001	−0.0001	0.0007	−0.0006	0.0006
		SD	0.0242	0.0190	0.0168	0.0157	0.0131	0.0117	0.0102	0.0091	0.0079
Second response	$\hat{\beta}_{2,0}$	Bias	−0.0009	0.0027	−0.0009	−0.0020	0.0005	0.0003	−0.0007	0.0038	0.0017
		SD	0.0481	0.0397	0.0365	0.0383	0.0325	0.0281	0.0280	0.0253	0.0201
	$\hat{\beta}_{2,1}$	Bias	−0.0015	−0.0007	0.0033	−0.0007	−0.0009	−0.0005	0.0008	−0.0006	−0.0001
		SD	0.0281	0.0183	0.0180	0.0161	0.0156	0.0145	0.0111	0.0103	0.0092
Third response	$\hat{\beta}_{3,0}$	Bias	−0.0002	0.0009	0.0026	-0.0007	0.0018	0.0005	−0.0003	0.0012	0.0010
		SD	0.0283	0.0260	0.0228	0.0207	0.0183	0.0166	0.0143	0.0135	0.0117
	$\hat{\beta}_{3,1}$	Bias	−0.0007	−0.0020	0.0011	−0.0009	−0.0027	−0.0008	−0.0004	−0.0001	−0.0003
		SD	0.0234	0.0193	0.0158	0.0140	0.0120	0.0120	0.0098	0.0081	0.0080

Note. n is the number of subjects and m is the number of observation times for each subject. $\beta_{j,0}$ $(k = 1, 2, 3)$ represents the estimated regression coefficients of the intercept for the k-th response and $\beta_{k,1}$ $(k = 1, 2, 3)$ denotes the estimated regression coefficients of the slope for the k-th response.

Table 3 Average bias (Bias) and average empirical standard deviation (SD) of the estimates for autoregressive coefficients calculated from 100 simulations

$\rho = 0.9$	m	5			10			15		
	n	50	100	200	50	100	200	50	100	200
$\hat{a}_1$	Bias	−0.0103	−0.0031	−0.0002	0.0002	0.0012	−0.0015	−0.0012	−0.0052	−0.0023
	SD	0.0572	0.0364	0.0248	0.0296	0.0207	0.0140	0.0241	0.0160	0.0115
$\check{a}_1$	Bias	−0.0196	0.0064	0.0111	−0.0011	−0.0063	−0.0012	−0.0080	−0.0074	−0.0011
	SD	0.1258	0.0809	0.0548	0.0703	0.0479	0.0290	0.0460	0.0363	0.0240
$\hat{b}_1$	Bias	0.0024	−0.0016	−0.0023	−0.0111	−0.0109	−0.0009	−0.0052	0.0097	0.0031
	SD	0.0879	0.0577	0.0363	0.0571	0.0372	0.0286	0.0527	0.0351	0.0271
$\check{b}_1$	Bias	0.0060	−0.0142	−0.0257	−0.0236	−0.0048	−0.0045	−0.0011	0.0161	0.0019
	SD	0.1869	0.1412	0.0886	0.1384	0.0939	0.0573	0.1110	0.0792	0.0543
$\hat{a}_2$	Bias	−0.0031	−0.0020	−0.0023	−0.0019	0.0009	0.0006	−0.0046	−0.0031	−0.0011
	SD	0.0595	0.0402	0.0250	0.0368	0.0221	0.0178	0.0267	0.0191	0.0122
$\check{a}_2$	Bias	−0.0012	0.0057	0.0049	−0.0076	−0.0041	0.0030	−0.0097	−0.0055	−0.0002
	SD	0.1482	0.0925	0.0567	0.0770	0.0535	0.0389	0.0516	0.0384	0.0277
$\hat{b}_2$	Bias	−0.0090	−0.0020	0.0005	−0.0013	−0.0060	−0.0023	−0.0031	−0.0004	0.0017
	SD	0.0903	0.0580	0.0354	0.0705	0.0424	0.0356	0.0608	0.0438	0.0269
$\check{b}_2$	Bias	−0.0187	−0.0052	−0.0154	−0.0017	−0.0001	−0.0047	0.0014	0.0068	0.0006
	SD	0.2237	0.1416	0.0901	0.1578	0.0958	0.0401	0.1192	0.0928	0.0516
$\hat{a}_3$	Bias	−0.0063	−0.0025	−0.0040	−0.0049	0.0005	−0.0004	0.0012	−0.0043	−0.0015
	SD	0.0650	0.0395	0.0264	0.0330	0.0213	0.0401	0.0237	0.0186	0.0126
$\check{a}_3$	Bias	−0.0057	0.0035	0.0044	−0.0044	−0.0068	−0.0015	−0.0046	−0.0060	−0.0026
	SD	0.1325	0.0876	0.0489	0.0757	0.0451	0.0338	0.0495	0.0360	0.0274
$\hat{b}_3$	Bias	0.0027	−0.0029	0.0006	0.0036	0.0005	−0.0045	−0.0091	0.0071	0.0033
	SD	0.0907	0.0619	0.0415	0.0658	0.0368	0.0285	0.0537	0.0393	0.0294
$\check{b}_3$	Bias	−0.0071	−0.0096	−0.0088	−0.0062	0.0204	−0.0024	-0.0012	0.0194	0.0041
	SD	0.2025	0.1308	0.0760	0.1572	0.0940	0.0663	0.1142	0.0899	0.0612

Note. $\hat{a}$ and $\hat{b}$ represent the proposed weighted least square estimator while $\check{a}$ and $\check{b}$ represent the naive ordinary least square estimator.

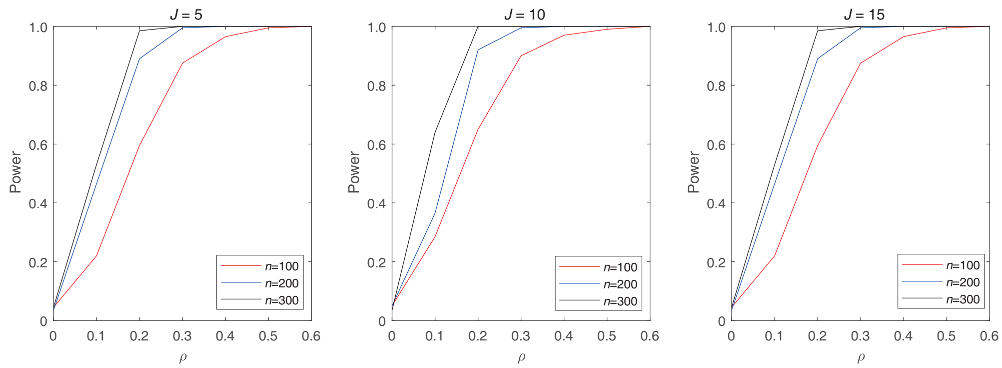

Figure 1 (Color online) Estimated power curves with $\alpha = 0.05$ for different observation times

It indicates that our proposed estimation method for the regression coefficients performs well with finite samples. The SD of our proposed estimator decreases as either the sample size n or the number of replicated observations m increases. Moreover, when correlations between different responses are taken into account, the estimated regression coefficients have smaller empirical standard deviations than in the case where correlations among responses are ignored. We omit the results for $\rho = 0.6$ and $\rho = 0.3$, which are similar.

Moreover, we also conducted additional simulation studies to illustrate that m_i can be large. The simulation settings are exactly the same as in Example 5.1 and the results are presented in Table 2. The results indicate that our proposed estimation method for regression coefficients performs well for relatively large n and m. The standard deviation of our proposed estimator decreases as either the sample size n or the number of replicated observations m increases.

Similarly, the average bias and average empirical standard deviation of estimates for the autoregressive coefficients are presented in Table 3. They show simulation results similar to those in Table 1. We again omit the results for $\rho = 0.6$ and $\rho = 0.3$, which are similar.

Example 5.2. In this example, we aim to show the performance of our proposed testing method. The simulation settings are as presented in Example 5.1, except that the covariance matrix of $e_{i,j,k}$ has $1+\rho$ on the diagonal and ρ on the off-diagonal entries.

We consider different values of ρ being equal to 0, 0.1, 0.2, 0.3, 0.4, 0.5 and 0.6. The larger the ρ value, the stronger the correlation between the responses. For each setting, we run the simulation 200 times. The nominal significance level is $\alpha = 0.05$, and the estimated power curves are shown in Figure 1. We can see that under the null hypothesis, the actual size is close to the nominal size, 0.05, and the power approaches 1 rapidly as ρ or the sample size increases.

Example 5.3. In this example, the simulation settings are again as presented in Example 5.1, except that the orders of autoregressive processes are unknown. Specifically, the error processes are generated as follows:

$$\varepsilon_{i,j,k} = \sum_{s=1}^{6}(a_{k,s} + b_{k,s}d_{i,j,s})\varepsilon_{i,j-s,k} + e_{i,j,k},$$

where $(a_1, b_1) = (0.6, -0.3, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)^\tau$, $(a_2, b_2) = (0.3, -0.4, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)^\tau$ and $(a_3, b_3) = (-0.5, 0.3, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)^\tau$. We run the simulation 200 times, and apply the proposed SCAD variable selection method to simultaneously determine the orders of the AR processes and estimate the autoregressive coefficients. We evaluate the performance of the SCAD penalization procedure using the following three criteria:

(1) **Model size:** the number of $\hat{\theta}_j$ estimated as nonzero. The true model size is 6.

(2) **False positives:** the number of $\hat{\theta}_j$ falsely estimated as nonzero, i.e., $|\{\hat{\theta}_j \neq 0, \theta_j = 0\}|$. The maximum number of false positives is 30.

(3) **False negatives:** the number of $\hat{\theta}_j$ falsely estimated as zero, i.e., $|\{\hat{\theta}_j = 0, \theta_j \neq 0\}|$. The maximum number of false positives is 6.

Table 4 Results of the order determination for the AR error process

m	10			15		
n	50	100	150	50	100	150
Model size	4.36	4.52	5.31	4.87	5.38	5.96
False positives	1.49	1.12	1.40	1.14	0.97	0.96
False negatives	3.13	2.60	2.09	2.28	1.59	1.06

Table 5 Standard deviations of the predicted values for the multivariate longitudinal data model versus three univariate longitudinal data models using 100 simulations

$\rho = 0.9$ m	5			10			15		
n	50	100	200	50	100	200	50	100	200
Univariate$_1$	0.1208	0.0631	0.0534	0.1006	0.0602	0.0423	0.0708	0.0527	0.0374
Multivariate$_1$	0.1006	0.0537	0.0475	0.0893	0.0525	0.0370	0.0624	0.0467	0.0314
Univariate$_2$	0.1028	0.0805	0.0502	0.0790	0.0496	0.0322	0.0586	0.0432	0.0311
Multivariate$_2$	0.0895	0.0642	0.0435	0.0733	0.0447	0.0301	0.0559	0.0383	0.0250
Univariate$_3$	0.0826	0.0747	0.0377	0.0567	0.0323	0.0237	0.0399	0.0284	0.0200
Multivariate$_3$	0.0673	0.0645	0.0292	0.0526	0.0292	0.0210	0.0363	0.0250	0.0164

Note. Univariate$_k$ ($k = 1, 2, 3$) denotes the standard deviation of predicted values based on the k-th univariate longitudinal data modeling, while Multivariate$_k$ denotes the standard deviation of predicted values for the k-th response based on multivariate modeling.

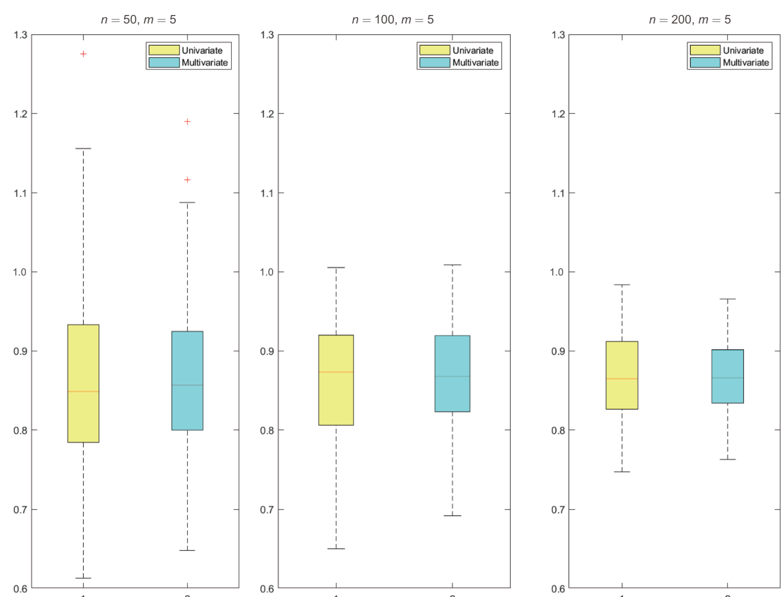

Figure 2 (Color online) Boxplots of the standard deviations for the first response's predicted values using different sample sizes

We summarize the simulation results in Table 4. The table shows that as either the sample size n or the number of replicated observations m increases, the selected model size becomes closer to the true model size, and the false positives and false negatives decrease.

Example 5.4. In this example, the simulation settings are the same as presented in Example 5.1. We want to illustrate the superiority of the multivariate longitudinal data model over several univariate longitudinal data models in terms of prediction accuracy.

In particular, we randomly split the data into a training set and a testing set at a proportion of 80% and 20%. We use the training data set to estimate the regression parameters, and then use the testing data set to evaluate the prediction accuracy. For different sample sizes, $N = n \times m$, we run the simulation 100 times. For each simulation, we can obtain N_1 predicted values of $\hat{y}_i$ ($i = 1, \ldots, N_1$), where N_1 is the sample size of the testing data, and then take the average of these N_1 predicted values. Finally, we calculate the standard deviation of the predicted values using 100 simulations. Note that the smaller the standard deviation, the better the prediction performance. Table 5 shows the standard deviation of the predicted values for the multivariate longitudinal data model and three univariate longitudinal data models under different scenarios.

From Table 5, we can see that if correlation exists among different responses, the standard deviation of the predicted values based on the multivariate longitudinal data model is always smaller than several univariate longitudinal data models. It is easy to see that an increase in sample size, n, or the number of observation time, m, results in a decrease in the standard deviation for both models.

Furthermore, Figure 2 depicts the standard deviations for the predicted values of the first response with different sample sizes, showing that the multivariate longitudinal data model always performs better than the univariate longitudinal data model. This is intuitive, because we can borrow information from other univariate models. We do not present the results here for other similar cases.

5.2 Real data analysis

We apply our proposed method to the paper making data described in the introduction. These data have also been analyzed by [12]. We are interested in four quality response variables, i.e., the tensile index (ng/g), the burst index (kPa m^2/g), the tear index (nN m^2/g), and the drainability of pulp (SR number). We denote them as $\{Y_{\cdot\cdot k}, k = 1, 2, 3, 4\}$, respectively. These $q = 4$ response variables are repeatedly measured at $m = 5$ beating times of 5, 15, 30, 45 and 60 minutes for $n = 48$ batches of pine sulfate pulp. The final quality of the paper may also be affected by characteristics of the pulp, such as the International Standards Organization (ISO) brightness (%), the electrical conductivity (mS/m), and its pH. We consider these three characteristics as model covariates, and respectively denote them as $\{X_{1\cdot\cdot}, X_{2\cdot\cdot}, X_{3\cdot\cdot}\}$. Moreover, we plot the logarithm of the four response variables against beating times in Figure 3, showing a nonlinear time trend for each response. Viroli [12] characterized the nonlinear trends using a function of the form $\lambda_0 + \lambda_1 t + \lambda_2 \log(t)$, where t is the beating time. Thus, we model the multivariate longitudinal data as follows:

$$\log(y_{i,j,k}) = X_{i,j}^{\tau}\boldsymbol{\beta}_k + \lambda_0 + \lambda_1 t + \lambda_2 \log t + \varepsilon_{i,j,k}, \quad i = 1, \ldots, 46, \quad j = 1, \ldots, 5, \quad k = 1, \ldots, 4. \tag{5.1}$$

We first assume that the error process for each response follows an AR(4) model,

$$\varepsilon_{i,j,k} = \sum_{s=1}^{4} (a_{k,s} + b_{k,s} d_{i,j,s}) \varepsilon_{i,j-s,k} + e_{i,j,k}, \quad k = 1, 2, 3, 4. \tag{5.2}$$

The estimated regression coefficients with corresponding p-values in parentheses are listed in Table 5, where “Independency” denotes the results obtained by the initial least square estimation method, “Proposed” denotes the results obtained by the proposed weighted least square estimation method and “Viroli (2012)” denotes the results obtained by [12]. From Table 6, we can conclude that, the effects of covariates brightness and electric conductivity are not significant to the drainability index using our proposed method and [12]; the effect of the covariate brightness is not significant to the tensile index, and the covariate pH is also not significant to the drainability index following the result obtained in [12].

We also apply the SCAD penalized method proposed in Subsection 2.3 to simultaneously determine the orders of the error processes and to estimate the autoregressive coefficients. AR models with lag orders of 1, 1, 1 and 2 were selected for the four responses. The results of fitting based on the selected model are shown in Table 7. The estimated $\hat{a}_k$'s are positive and significant while $\hat{b}_k$'s are zero, which may be due to the fact that the data were balanced.

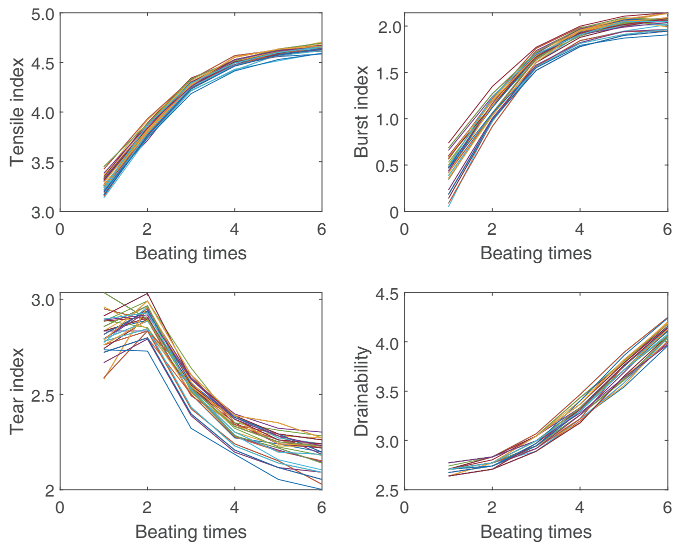

Figure 3 (Color online) Scatter plots for four paper quality variables versus the beating time

Table 6 Estimated regression coefficients and associated p-values (in brackets)

Method	Parameter	Tensile index	Burst index	Tear index	Drainability
Independence	β_1	−0.0182 (0.000)	−0.0501 (0.000)	−0.0497 (0.000)	−0.0055 (**0.373**)
	β_2	−0.0095 (**0.313**)	0.0331 (0.001)	−0.0429 (0.000)	0.0117 (**0.359**)
	β_3	0.0465 (0.004)	0.1223 (0.000)	0.0544 (0.006)	0.0651 (0.003)
	λ_0	4.5579 (0.000)	4.2048 (0.000)	7.8328 (0.000)	2.9489 (0.000)
	λ_1	−0.0061 (0.000)	−0.0086 (0.000)	0.0045 (0.000)	0.0279 (0.000)
	λ_2	0.4670 (0.000)	0.5652 (0.000)	−0.3864 (0.000)	−0.0807 (0.000)
Proposed	β_1	−0.0149 (0.000)	−0.0452 (0.000)	−0.0534 (0.000)	−0.0034 (**0.4900**)
	β_2	−0.0207 (0.001)	−0.0020 (**0.862**)	−0.0474 (0.000)	0.0068 (**0.0508**)
	β_3	0.0238 (0.021)	0.0560 (0.002)	0.0413 (0.031)	0.0571 (0.0010)
	λ_0	4.3746 (0.000)	4.1048 (0.000)	8.2325 (0.000)	2.8083 (0.0000)
	λ_1	−0.0061 (0.000)	−0.0086 (0.000)	0.0045 (0.000)	0.0279 (0.0000)
	λ_2	0.4672 (0.000)	0.5650 (0.000)	−0.3864 (0.000)	−0.0814 (0.0000)
Viroli (2012)	β_1	−0.0169 (0.000)	−0.0470 (0.000)	−0.0529 (0.000)	−0.0022 (**0.971**)
	β_2	−0.0074 (**0.425**)	0.0249 (0.030)	−0.0342 (0.030)	0.0192 (**0.266**)
	β_3	0.0374 (0.000)	0.0937 (0.000)	0.0443 (0.000)	0.0492 (**0.078**)
	λ_0	4.1620 (0.000)	3.6922 (0.000)	8.1458 (0.000)	2.1086 (0.000)
	λ_1	−0.0061 (0.000)	−0.0086 (0.000)	0.0044 (0.000)	0.0283 (0.000)
	λ_2	0.4669 (0.000)	0.5668 (0.000)	−0.3817 (0.000)	−0.0897 (0.000)

Table 7 Estimated autoregressive coefficients and the corresponding standard errors (SE) and associated 95% confidence intervals (CI)

Coefficients	Estimate	SE	CI
$\hat{a}_{1,1}$	0.5094	0.0481	$[0.4116, 0.6043]$
$\hat{a}_{2,1}$	0.5626	0.0388	$[0.4860, 0.6392]$
$\hat{a}_{3,1}$	0.7237	0.0508	$[0.6236, 0.8238]$
$\hat{a}_{4,1}$	0.6571	0.0801	$[0.4986, 0.8155]$
$\hat{a}_{4,2}$	0.3657	0.1027	$[0.1627, 0.5687]$

6 Conclusion

In this article, we developed a two-stage weighted least square method to analyze multivariate longitudinal data with auto-correlated error processes. Our method not only captures the potential correlations among different responses, which may improve the efficiency of our proposed estimators, but also avoids estimating large covariance matrices so as to reduce computational complexity. Moreover, the test statistic proposed in this paper can detect the potential correlations among different responses. Both Monte Carlo simulations and real data analysis demonstrate favorable empirical performances compared with existing methods for modeling univariate longitudinal data.

For future study, we may consider nonparametric or semiparametric models in the cases where the linear model is invalid, e.g., the partial linear model,

$$Y_{i,j,k} = \mathbf{X}_{i,j}^{\tau}\ \boldsymbol{\beta}_k + g(t_{i,j}) + \varepsilon_{i,j,k}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, m_i, \quad k = 1, \ldots, q.$$

In addition, we may also investigate models involving high-dimensional regressors. Wei et al. [14] have studied variable selection and estimation of the varying coefficient model for univariate longitudinal data under the high-dimensional setting. However, they ignored the with-subject correlation. It is important to investigate how to incorporate such correlation into inference for high-dimensional multivariate longitudinal data.

Acknowledgements The first author was supported by the Fundamental Research Funds of Shandong University (Grant No. 2018GN050), the Academic Prosperity Program provided by School of Economics, Shandong University and the Taishan Scholar Program of Shandong Province. The third author was supported by National Natural Science Foundation of China (Grant No. 11871323), the State Key Program in the Major Research Plan of National Natural Science Foundation of China (Grant No. 91546202) and Program for Innovative Research Team of Shanghai University of Finance and Economics. The authors thank the associate editor and two anonymous referees for many helpful comments and suggestions, which greatly improved the paper.

References

1 Bai Y, Huang J, Li R, et al. Semiparametric longitudinal model with irregular time autoregressive error process. Statist Sinica, 2015, 25: 507–527

2 Coull B A, Staudenmayer J. Self-modeling regression for multivariate curve data. Statist Sinica, 2004, 14: 695–711

3 Diggle P, Heagerty P, Liang K, et al. Analysis of Longitudinal Data. Oxford: Oxford University Press, 2013

4 Fan J, Huang T, Li R. Analysis of longitudinal data with semiparametric estimation of covariance function. J Amer Statist Assoc, 2007, 102: 632–641

5 Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. J Amer Statist Assoc, 2001, 96: 1348–1360

6 Fieuws S, Verbeke G. Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. Biometrics, 2006, 62: 424–431

7 Fitzmaurice G, Laird N, Ware J. Applied Longitudinal Analysis. Hoboken: Wiley, 2011

8 Gao F, Thompson P, Xiong C, et al. Analyzing multivariate longitudinal data using SAS. In: Proceedings of the Thirty-First Annual SAS Users Group International Conference. Http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.88.8829&rep=rep1&type=pdf, 2006, 1–31

9 Pei Y, Huang T, You J. Nonparametric fixed effects model for panel data with locally stationary regressors. J Econometrics, 2018: 202: 286–305

10 Roy J, Lin X. Latent variable models for longitudinal data with multiple continuous outcomes. Biometrics, 2000, 56: 1047–1054

11 Verbeke G, Fieuws S, Molenberghs G, et al. The analysis of multivariate longitudinal data: A review. Stat Methods Med Res, 2014, 23: 42–59

12 Viroli C. On matrix-variate regression analysis. J Multivariate Anal, 2012, 111: 296–309

13 Wang H, Li R, Tsai C-L. Tuning parameter selectors for the smoothly clipped absolute deviation method. Biometrika, 2007, 94: 553–568

14 Wei F, Huang J, Li H. Variable selection and estimation in high-dimensional varying-coefficient models. Statist Sinica, 2011, 21: 1515–1540

15 Xiang D, Qiu P, Pu X. Nonparametric regression analysis of multivariate longitudinal data. Statist Sinica, 2013, 23: 769–789

16 Zellner A. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. J Amer Statist Assoc, 1962, 57: 348–368

17 Zhu H, Li R, Kong L. Multivariate varying coefficient model for functional responses. Ann statist, 2012, 40: 2634–2666

Appendix A

Proof of Theorem 4.1. We first give a detailed proof for the first part of Theorem 4.1.

- Denote

$$D = \mathrm{E}[(X_i \otimes I_q)^\tau (X_i \otimes I_q)], \quad \Delta = \mathrm{E}[(X_i \otimes I_q)^\tau \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^\tau (X_i \otimes I_q)].$$

We have

$$\begin{aligned}\sqrt{Nq}(\hat{\boldsymbol{\beta}}_{\rm ols} - \boldsymbol{\beta}) &= \sqrt{Nq}\{[(\mathbf{X} \otimes I_q)^\tau (\mathbf{X} \otimes I_q)]^{-1} (\mathbf{X} \otimes I_q)^\tau [(\mathbf{X} \otimes I_q)\boldsymbol{\beta} + \boldsymbol{\varepsilon}] - \boldsymbol{\beta}\} \\ &= \sqrt{Nq}\{[(\mathbf{X} \otimes I_q)^\tau (\mathbf{X} \otimes I_q)]^{-1} (\mathbf{X} \otimes I_q)^\tau \boldsymbol{\varepsilon}\} \\ &= \left\{\frac{1}{Nq}[(\mathbf{X} \otimes I_q)^\tau (\mathbf{X} \otimes I_q)]\right\}^{-1} \left\{\frac{1}{\sqrt{Nq}}(\mathbf{X} \otimes I_q)^\tau \boldsymbol{\varepsilon}\right\} \\ &= \left\{\frac{1}{Nq}\sum_{i=1}^n [(X_i \otimes I_q)^\tau (X_i \otimes I_q)]\right\}^{-1} \left\{\frac{1}{\sqrt{Nq}}\sum_{i=1}^n (X_i \otimes I_q)^\tau \boldsymbol{\varepsilon}_i\right\}.\end{aligned}$$

By the law of large numbers,

$$\frac{1}{Nq}\sum_{i=1}^n [(X_i \otimes I_q)^\tau (X_i \otimes I_q)] \xrightarrow{\mathcal{P}} \mathrm{E}[(X_i \otimes I_q)^\tau (X_i \otimes I_q)].$$

By the central limit theorem,

$$\frac{1}{\sqrt{Nq}}\sum_{i=1}^n (X_i \otimes I_q)^\tau \boldsymbol{\varepsilon}_i \xrightarrow{\mathcal{D}} N(0, \mathrm{E}[(X_i \otimes I_q)^\tau \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^\tau (X_i \otimes I_q)]).$$

Therefore, by Slutsky's lemma,

$$\sqrt{Nq}(\hat{\boldsymbol{\beta}}_{\rm ols} - \boldsymbol{\beta}) \xrightarrow{\mathcal{D}} N(0, D^{-1}\Delta D^{-1}).$$

Next, we give a detailed proof for the second part of Theorem 4.1.

- Recall that

$$\begin{aligned}\sqrt{Nq}(\hat{\boldsymbol{\theta}}_{\rm ols} - \boldsymbol{\theta}) &= \sqrt{Nq - ndq}[(\delta^\tau \delta)^{-1}\delta^\tau(\delta\boldsymbol{\theta} + \tilde{\mathbf{e}}) - \boldsymbol{\theta}] \\ &= \sqrt{Nq - ndq}[(\delta^\tau \delta)^{-1}\delta^\tau \tilde{\mathbf{e}}] \\ &= \left(\frac{1}{Nq - ndq}\delta^\tau \delta\right)^{-1} \left(\frac{1}{\sqrt{Nq - ndq}}\delta^\tau \tilde{\mathbf{e}}\right) \\ &= \left(\frac{1}{Nq - ndq}\sum_{i=1}^n \delta_i^\tau \delta_i\right)^{-1} \left(\frac{1}{\sqrt{Nq - ndq}}\sum_{i=1}^n \delta_i^\tau \tilde{\mathbf{e}}_i\right).\end{aligned}$$

By the law of large numbers,

$$\frac{1}{Nq - ndq}\sum_{i=1}^n \delta_i^\tau \delta_i \xrightarrow{\mathcal{P}} \mathrm{E}(\delta_i^\tau \delta_i).$$

By the central limit theorem,

$$\frac{1}{\sqrt{Nq - ndq}}\sum_{i=1}^n \delta_i^\tau \tilde{\mathbf{e}}_i \xrightarrow{\mathcal{D}} N(0, \sigma_e^2 \mathrm{E}(\delta_i^\tau \delta_i)).$$

Therefore, by Slutsky's lemma,

$$\sqrt{Nq-ndq}(\hat{\boldsymbol{\theta}}_{\mathrm{ols}}-\boldsymbol{\theta}) \xrightarrow{\mathcal{D}} N(0,\sigma_e^2\Lambda^{-1}).$$

This completes the proof. □

Proof of Theorem 4.2. It is obvious that

$$\hat{a}_{k,s}+\hat{b}_{k,s}d_{i,j,s}=a_{k,s}+b_{k,s}d_{i,j,s}-(a_{k,s}-\hat{a}_{k,s})-(b_{k,s}-\hat{b}_{k,s})d_{i,j,s}.$$

By the definition of $\hat{\sigma}_e^2$ in Section 4, we have

$$\begin{aligned}\hat{\sigma}_e^2&=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i-d}\sum_{j=d+1}^{m_i}\frac{1}{q}\sum_{k=1}^{q}\left\{\varepsilon_{i,j,k}-\sum_{s=1}^{d_k}(\hat{a}_{k,s}+\hat{b}_{k,s}d_{i,j,s})\varepsilon_{i,j-s,k}\right\}^2+o_p\left(\frac{1}{\sqrt{N}}\right)\\&=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i-d}\sum_{j=d+1}^{m_i}\frac{1}{q}\sum_{k=1}^{q}\left\{\varepsilon_{i,j,k}-\sum_{s=1}^{d_k}(a_{k,s}+b_{k,s}d_{i,j,s})\varepsilon_{i,j-s,k}\right\}^2+o_p\left(\frac{1}{\sqrt{N}}\right)\\&=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i-d}\sum_{j=d+1}^{m_i}\frac{1}{q}\sum_{k=1}^{q}e_{i,j,k}^2+o_p\left(\frac{1}{\sqrt{N}}\right).\end{aligned}$$

Since $e_{i,j,k}^2$ are i.i.d. random variables with mean σ_e^2 and variance $\mathrm{E}e_{i,j,k}^4-\sigma_e^4$, the first claim of Theorem 4.2 follows from the central limit theorem and Slutsky's lemma. The remaining claims of Theorem 4.2 follow from Theorem 4.1, the law of large numbers and Slutsky's lemma. □

Proof of Theorem 4.3. We first give a detailed proof for the first part of Theorem 4.3.

- Assume the true covariance matrix of $\tilde{\mathbf{e}}_i$ is Σ_i, and denote

$$C=\mathrm{E}(\tilde{X}_i^\tau\Sigma_i^{-1}\tilde{X}_i),\quad \Theta=\mathrm{E}(\tilde{X}_i^\tau\Sigma_i^{-1}\tilde{\mathbf{e}}_i\tilde{\mathbf{e}}_i^\tau\Sigma_i^{-1}\tilde{X}_i).$$

We have

$$\begin{aligned}\sqrt{Nq}(\hat{\boldsymbol{\beta}}_{\mathrm{wls}}-\boldsymbol{\beta})&=\sqrt{Nq}[(\tilde{X}^\tau\Sigma^{-1}\tilde{X})^{-1}\tilde{X}^\tau\Sigma^{-1}(\tilde{X}\boldsymbol{\beta}+\tilde{\mathrm{e}})-\boldsymbol{\beta}]\\&=\sqrt{Nq}[(\tilde{X}^\tau\Sigma^{-1}\tilde{X})^{-1}\tilde{X}^\tau\Sigma^{-1}\tilde{\mathbf{e}}]\\&=\left(\frac{1}{Nq}\tilde{X}^\tau\Sigma^{-1}\tilde{X}\right)^{-1}\left(\frac{1}{\sqrt{Nq}}\tilde{X}^\tau\Sigma^{-1}\tilde{\mathbf{e}}\right)\\&=\left(\frac{1}{Nq}\sum_{i=1}^{n}\tilde{X}_i^\tau\Sigma_i^{-1}\tilde{X}_i\right)^{-1}\left(\frac{1}{\sqrt{Nq}}\sum_{i=1}^{n}\tilde{X}_i^\tau\Sigma_i^{-1}\tilde{\mathbf{e}}_i\right).\end{aligned}$$

By the law of large numbers,

$$\frac{1}{Nq}\sum_{i=1}^{n}\tilde{X}_i^\tau\Sigma_i^{-1}\tilde{X}_i\xrightarrow{\mathcal{P}}\mathrm{E}(\tilde{X}_i^\tau\Sigma_i^{-1}\tilde{X}_i).$$

By the central limit theorem,

$$\frac{1}{\sqrt{Nq}}\sum_{i=1}^{n}\tilde{X}_i^\tau\Sigma_i^{-1}\tilde{\mathbf{e}}_i\xrightarrow{\mathcal{D}}N(0,\mathrm{E}(\tilde{X}_i^\tau\Sigma_i^{-1}\tilde{\mathbf{e}}_i\tilde{\mathbf{e}}_i^\tau\Sigma_i^{-1}\tilde{X}_i)).$$

Therefore, by Slutsky's lemma,

$$\sqrt{Nq}(\hat{\boldsymbol{\beta}}_{\mathrm{wls}}-\boldsymbol{\beta})\xrightarrow{\mathcal{D}}N(0,C^{-1}\Theta C^{-1}).$$

- The proof of the second part is similar to the first part. Denote

$$A=\mathrm{E}(\delta_i^\tau\Sigma_i^{-1}\delta_i),\quad B=\mathrm{E}(\delta_i^\tau\Sigma_i^{-1}\tilde{\mathbf{e}}_i\tilde{\mathbf{e}}_i^\tau\Sigma_i^{-1}\delta_i).$$

We have

$$\begin{aligned}\sqrt{Nq-ndq}(\hat{\boldsymbol{\theta}}_{\rm wls}-\boldsymbol{\theta}) &= \sqrt{Nq-ndq}[(\delta^\tau\Sigma^{-1}\delta)^{-1}\delta^\tau\Sigma^{-1}(\delta\boldsymbol{\theta}+\tilde{\mathbf{e}})-\boldsymbol{\theta}]\\ &= \sqrt{Nq-ndq}[(\delta^\tau\Sigma^{-1}\delta)^{-1}\delta^\tau\Sigma^{-1}\tilde{\mathbf{e}}]\\ &= \left(\frac{1}{Nq-ndq}\delta^\tau\Sigma^{-1}\delta\right)^{-1}\left(\frac{1}{\sqrt{Nq-ndq}}\delta^\tau\Sigma^{-1}\tilde{\mathbf{e}}\right)\\ &= \left(\frac{1}{Nq-ndq}\sum_{i=1}^n\delta_i^\tau\Sigma_i^{-1}\delta_i\right)^{-1}\left(\frac{1}{\sqrt{Nq-ndq}}\sum_{i=1}^n\delta_i^\tau\Sigma_i^{-1}\tilde{\mathbf{e}}_i\right).\end{aligned}$$

By the law of large numbers,

$$\frac{1}{Nq-ndq}\sum_{i=1}^n\delta_i^\tau\Sigma_i^{-1}\delta_i\xrightarrow{\mathcal{P}}{\rm E}(\delta_i^\tau\Sigma_i^{-1}\delta_i).$$

By the central limit theorem,

$$\frac{1}{\sqrt{Nq-ndq}}\sum_{i=1}^n\delta_i^\tau\Sigma_i^{-1}\tilde{\mathbf{e}}_i\xrightarrow{\mathcal{D}}N(0,{\rm E}(\delta_i^\tau\Sigma_i^{-1}\tilde{\mathbf{e}}_i\tilde{\mathbf{e}}_i^\tau\Sigma_i^{-1}\delta_i)).$$

Therefore, by Slutsky's lemma,

$$\sqrt{Nq-ndq}(\hat{\boldsymbol{\theta}}_{\rm wls}-\boldsymbol{\theta})\xrightarrow{\mathcal{D}}N(0,A^{-1}BA^{-1}).$$

This completes the proof. □

Proof of Theorem 4.4. According to the definition of $\hat{V}=(\hat{\sigma}^2_{k_1,k_2})$ and Theorem 4.3, we can show that

$$\begin{aligned}\hat{\sigma}^2_{k_1,k_2} &= \frac{1}{n}\sum_{i=1}^n\frac{1}{m_i-d}\sum_{j=d+1}^{m_i}\hat{e}_{i,j,k_1}\hat{e}_{i,j,k_2}\\ &= \frac{1}{n}\sum_{i=1}^n\frac{1}{m_i-d}\sum_{j=d+1}^{m_i}\left\{\varepsilon_{i,j,k_1}-\sum_{s=1}^d(\hat{a}_{k,s}+\hat{b}_{k,s}d_{i,j,s})\varepsilon_{i,j-s,k_1}\right\}\\ &\quad\times\left\{\varepsilon_{i,j,k_1}-\sum_{s=1}^d(\hat{a}_{k,s}+\hat{b}_{k,s}d_{i,j,s})\varepsilon_{i,j-s,k_1}\right\}+o_p\left(\frac{1}{\sqrt{N}}\right)\\ &= \frac{1}{n}\sum_{i=1}^n\frac{1}{m_i-d}\sum_{j=d+1}^{m_i}\left\{\varepsilon_{i,j,k_1}-\sum_{s=1}^d(a_{k,s}+b_{k,s}d_{i,j,s})\varepsilon_{i,j-s,k_1}\right\}\\ &\quad\times\left\{\varepsilon_{i,j,k_1}-\sum_{s=1}^d(a_{k,s}+b_{k,s}d_{i,j,s})\varepsilon_{i,j-s,k_1}\right\}+o_p\left(\frac{1}{\sqrt{N}}\right)\\ &= \frac{1}{n}\sum_{i=1}^n\frac{1}{m_i-d}\sum_{j=d+1}^{m_i}e_{i,j,k_1}e_{i,j,k_2}+o_p\left(\frac{1}{\sqrt{N}}\right).\end{aligned}$$

The theorem can be proven by applying the central limit theorem and Slutsky's lemma. □

Proof of Theorem 4.5. By the definition of $\hat{M}_j$,

$$\begin{aligned}\hat{M}_j &= \hat{\Lambda}_j^\tau\hat{\Gamma}_j^{-1}\hat{\Lambda}_j=\left(n^{-1/2}\sum_{i=1}^n\hat{\boldsymbol{\Delta}}_{i,j}^\tau\right)\hat{\Gamma}_j^{-1}\left(n^{-1/2}\sum_{i=1}^n\hat{\boldsymbol{\Delta}}_{i,j}\right)\\ &= \left(\frac{\sum_{i=1}^n\hat{e}_{i,j,1}\hat{e}_{i,j,2}}{\sqrt{n}\sigma_{11}\sigma_{22}}\right)^2+\left(\frac{\sum_{i=1}^n\hat{e}_{i,j,1}\hat{e}_{i,j,3}}{\sqrt{n}\sigma_{11}\sigma_{33}}\right)^2+\cdots+\left(\frac{\sum_{i=1}^n\hat{e}_{i,j,q-1}\hat{e}_{i,j,q}}{\sqrt{n}\sigma_{q-1,q-1}\sigma_{qq}}\right)^2\\ &= \left(\frac{\sum_{i=1}^n e_{i,j,1}e_{i,j,2}}{\sqrt{n}\sigma_{11}\sigma_{22}}\right)^2+\left(\frac{\sum_{i=1}^n e_{i,j,1}e_{i,j,3}}{\sqrt{n}\sigma_{11}\sigma_{33}}\right)^2+\cdots+\left(\frac{\sum_{i=1}^n e_{i,j,q-1}e_{i,j,q}}{\sqrt{n}\sigma_{q-1,q-1}\sigma_{qq}}\right)^2+o_P\left(\frac{1}{\sqrt{N}}\right).\end{aligned}$$

Under the null hypothesis using the central limit theorem and Slutsky's lemma, we can easily show that the limit distribution of $\hat{M}_j$ follows $\chi^2_{q(q-1)/2}$ distribution, where $q(q-1)/2$ is the degree of freedom. □

Discussions about the efficiency of multivariate and univariate longitudinal data models. Keep in mind that if there is no correlation among different responses, the multivariate longitudinal data model is equivalent to several univariate longitudinal data models. However, if this is not the case, the multivariate longitudinal data model will perform better in estimation efficiency. We will discuss this problem in the following steps.

Note that our proposed estimation method for the multivariate longitudinal data model,

$$Y_{i,j,k} = \mathbf{X}_{i,j}^{\tau}\boldsymbol{\beta}_k + \varepsilon_{i,j,k}, \quad i=1,\ldots,n, \quad j=1,\ldots,m_i, \quad k=1,\ldots,q, \tag{A.1}$$

$$\varepsilon_{i,j,k} = \sum_{s=1}^{d_k}(a_{k,s} + b_{k,s}d_{i,j,s})\varepsilon_{i,j-s,k} + e_{i,j,k} \tag{A.2}$$

can be transformed into the following model (A.3) using some difference-based approaches:

$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\mathbf{e}}, \tag{A.3}$$

where the definitions of $\tilde{\mathbf{y}}, \tilde{\mathbf{X}}$ and $\tilde{\mathbf{e}}$ are given in Subsection 2.2, and the estimator $\boldsymbol{\beta}$ is given by

$$\widetilde{\boldsymbol{\beta}} = [\widetilde{\mathbf{X}}^{\tau}\boldsymbol{\Sigma}^{-1}\widetilde{\mathbf{X}}]^{-1}\widetilde{\mathbf{X}}^{\tau}\boldsymbol{\Sigma}^{-1}\widetilde{\mathbf{y}},$$

where $\boldsymbol{\Sigma} = V \otimes I_N$, and V is a $q \times q$ covariance matrix of different responses, e.g.,

$$V = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \ldots & \sigma_{1q} \\ \sigma_{21} & \sigma_{22} & \ldots & \sigma_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{q1} & \sigma_{q2} & \ldots & \sigma_{qq} \end{pmatrix}$$

and I_N is an identity matrix of dimension $N = n \times m$. The variance-covariance matrix of the estimator $\widetilde{\boldsymbol{\beta}}$ is easily shown to be $(\widetilde{\mathbf{X}}^{\tau}\boldsymbol{\Sigma}^{-1}\widetilde{\mathbf{X}})^{-1}$. For convenience, we consider a simple case with a disturbance covariance matrix such that $\sigma_{kk} = \sigma^2$ and $\sigma_{k,l} = \rho\sigma^2$ for $k \neq l$, and then by some elementary calculation, we can arrive at

$$V = \sigma^2[(1-\rho)I_q + \rho\mathbf{1}_q\mathbf{1}_q^{\top}],$$

where I_q is an identity matrix of size $q \times q$ and $\mathbf{1}_q$ is a $q \times 1$ vector of 1's. $V^{-1} = \alpha I_q - \gamma\mathbf{1}_q\mathbf{1}_q^{\top}$ with $\alpha = 1/\sigma^2(1-\rho)$ and $\gamma = \alpha\rho/[1+(q-1)\rho]$. Then, for the covariance matrix of the estimator $\widetilde{\boldsymbol{\beta}}$, we have

$$\begin{aligned}\mathrm{Var}(\widetilde{\boldsymbol{\beta}}) &= [\widetilde{\mathbf{X}}^{\tau}(V^{-1}\otimes I_N)\widetilde{\mathbf{X}}]^{-1} \\ &= \begin{pmatrix} (\alpha-\gamma)\tilde{X}_1^{\tau}\tilde{X}_1 & -\gamma\tilde{X}_1^{\tau}\tilde{X}_2 & \ldots & -\gamma\tilde{X}_1^{\tau}\tilde{X}_q \\ -\gamma\tilde{X}_2^{\tau}\tilde{X}_1 & (\alpha-\gamma)\tilde{X}_2^{\tau}\tilde{X}_2 & \ldots & -\gamma\tilde{X}_2^{\tau}\tilde{X}_q \\ \vdots & \vdots & \ddots & \vdots \\ -\gamma\tilde{X}_q^{\tau}\tilde{X}_1 & -\gamma\tilde{X}_q^{\tau}\tilde{X}_2 & \ldots & (\alpha-\gamma)\tilde{X}_q^{\tau}\tilde{X}_q \end{pmatrix}^{-1},\end{aligned}$$

where $\tilde{X}_k$ $(k=1,\ldots,q)$ is an $N \times p$ matrix with rank p. For the two-dimensional case $(q=2)$, the covariance matrix of the coefficient vector estimator for the first response is given by

$$\mathrm{Var}(\widetilde{\boldsymbol{\beta}}_1) = \left[(\alpha-\gamma)\tilde{X}_1^{\tau}\tilde{X}_1 - \frac{\gamma^2}{\alpha-\gamma}\tilde{X}_1^{\tau}\tilde{X}_2(\tilde{X}_2^{\tau}\tilde{X}_2)^{-1}\tilde{X}_2^{\tau}\tilde{X}_1\right]^{-1}. \tag{A.4}$$

From [16], we can immediately arrive at

$$\mathrm{Var}(\widetilde{\boldsymbol{\beta}}_1) = \frac{(1-\rho^2)^p}{\prod_{j=1}^{p}(1-\rho^2 r_j^2)}\sigma^2(\tilde{X}_1^{\tau}\tilde{X}_1)^{-1}, \tag{A.5}$$

where p is the rank of the matrix $\tilde{X}_k$ and r_j is the j-th canonical correlation coefficient associated with the sets of variables in $\tilde{X}_1$ and $\tilde{X}_2$. From (A.5), we can draw the following conclusions:

- Since $0 \leqslant r_j^2 \leqslant 1$, it is clear that the variance of $\widetilde{\boldsymbol{\beta}}_1$ will be smaller than or equal to $|\sigma^2(\tilde{X}_1^\tau \tilde{X}_1)^{-1}|$, which is the variance of the first univariate longitudinal data model.
- If $\rho = 0$, the multivariate longitudinal data model is equivalent to several univariate longitudinal data models.
- If $\rho \neq 0$, the larger the ρ, the more efficiency can be gained from the multivariate longitudinal data model compared with several univariate longitudinal data models.