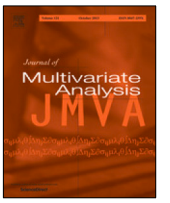
Journal of Multivariate Analysis 187 (2022) 104900

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva



Covariance function versus covariance matrix estimation in efficient semi-parametric regression for longitudinal data analysis

Shengji Jia [a,*], Chunming Zhang [b], Haoran Lu [c]

[a] School of Statistics and Mathematics; Interdisciplinary Research Institute of Data Science, Shanghai Lixin University of Accounting and Finance, Shanghai 201209, China
[b] Department of Statistics, University of Wisconsin, Madison, WI 53706, USA
[c] Department of Statistics, University of Georgia, Athens, GA 30602, USA

ARTICLE INFO

Article history:
Received 3 October 2020
Received in revised form 24 October 2021
Accepted 25 October 2021
Available online 5 November 2021

AMS 2020 subject classifications:
primary 62H12
secondary 62G05

Keywords:
Covariance function
High-dimensional covariance matrix
Local linear regression
Method of regularization
Profile weighted least squares
Varying-coefficient partially linear model

ABSTRACT

Improving estimation efficiency for regression coefficients is an important issue in the analysis of longitudinal data, which involves estimating the covariance matrix of the within-subject errors. In the balanced or nearly balanced setting, we can also regard the covariance matrix of the dependent errors as the bivariate covariance function evaluated at specific time points. In this paper, we compare the performance of the proposed regularized-covariance-function-based estimator and the conventional high-dimensional covariance matrix estimator of the within-subject errors. It shows that when the number p of the time points in each subject is large enough compared to the number n of the subjects, i.e., $p \gg n^{1/4}\log n$, the estimation errors of the high-dimensional covariance matrix will be accumulated, therefore the error bound of the proposed regularized-covariance-function-based estimator will be smaller than that of the high-dimensional covariance matrix estimator in Frobenius norm. We also assess the performance of these two estimators for the incomplete longitudinal data. All the comparisons and theoretical results are illustrated using both simulated and real data.

1. Introduction

There has been substantial recent interest in nonparametric and semi-parametric methods for longitudinal or clustered data with dependence within subjects (or clusters), see [2,7,9,18], among others. Improving estimation efficiency is an important issue in the analysis of longitudinal data. In the nonparametric setting, Lin and Carroll [9] have recommended an approach which ignores the within-subject correlation completely and treats the data as if they are independent. However, Wang et al. [15] have shown that, in the semi-parametric setting, the estimator for parametric component will achieve the semi-parametric efficiency bound if the within-subject correlation structure is specified correctly. Thus estimating the covariance matrix of the within-subject errors is an important issue in the semi-parametric model for longitudinal data.

Many authors have investigated the problem of within-subject correlation structure in longitudinal data. Challenges arise in estimating the covariance matrix if the data are collected at irregular or subject specific time points. In this case, instead of estimating covariance matrix, people usually estimate the covariance function of the error process. There

* Corresponding author.
E-mail address: 20200026@lixin.edu.cn (S. Jia).

https://doi.org/10.1016/j.jmva.2021.104900
0047-259X/

also has been a vast volume of work on modeling covariance functions in longitudinal data in literature. For example, Fan et al. [3,4] have proposed a quasi-maximum likelihood method to model covariance functions. The quasi-maximum likelihood method relies on correctly assuming the form of the covariance functions and can be misspecified easily. Li [8] has applied bivariate kernel smoothing techniques to estimate covariance functions. The bivariate kernel smoothing techniques do not utilize the structures of the covariance functions, thus the proposed estimator is not guaranteed to be positive definite and has slower rates of convergence. Jia et al. [7] have proposed the regularized method to estimate covariance functions, which has faster rates of convergence and is guaranteed to be positive semi-definite. Besides, other authors have tried to estimate the error processes directly, for example, Wu and Zhang [17,19] have proposed local linear mixed-effects (LLME) method, and Yao et al. [6,20] have proposed functional principal component analysis (FPCA) method to analyze longitudinal data.

In this paper, we will focus on the balanced or nearly balanced longitudinal data where the number p of the time points in each subject is allowed to increase together with the number n of the subjects. In this case, we can either estimate the high-dimensional covariance matrix of the with-subject errors by some existing methods directly [12–14], or we can first estimate the covariance function of the error process by the methods mentioned previously, and then derive the estimated covariance matrix by focusing on the elements of covariance function evaluated at given time points. We may ask which method of estimating the covariance matrix is more precise, or has faster rates of convergence and thus will result in a more efficient estimator for the parametric component in the semi-parametric model.

Let us consider a semi-parametric varying-coefficient partially linear model

$$Y(t) = \boldsymbol{X}(t)^\top \boldsymbol{\alpha}(t) + \boldsymbol{Z}(t)^\top \boldsymbol{\beta} + \eta(t) + \zeta(t), \tag{1}$$

where $\boldsymbol{\alpha}(t)$ comprises m unknown smooth functions, $\boldsymbol{\beta}$ is a q-dimensional unknown parameter vector, and $\eta(t)$ captures the within-subject dependence with smooth covariance function $R_\eta(t_1, t_2)$, $\zeta(t)$ is just the measurement error with covariance function $\sigma_\zeta^2(t_1)\mathrm{I}(t_1 = t_2)$. All the temporal correlations of the errors are relegated to $\eta(t)$, so this decomposition is unique. Nonparametric models for longitudinal data can be viewed as special cases of model (1). Moreover, model (1) is also a useful extension of the partially linear model, the time-varying coefficient model and the partially linear additive model.

Intuitively, the dimension of the covariance matrix of within-subject errors will increase as the number p of time points in each subject increases, and thus the total estimation error will be accumulated, whereas the covariance function will be exactly the same and not dependent on p. So when p and n become larger, the estimator of the covariance function will be more precise.

The balanced or nearly balanced longitudinal data usually come from the experiments in agriculture, biology and medicine. In these cases the observations are supposed to be collected at the designed time points in advance. But occasionally we may still lose the observations at part of these given time points for some subjects, which results in the incomplete longitudinal data. This is a special case of the unbalanced data, but most conventional high-dimensional covariance matrix techniques based on sample covariance matrix still work: we can derive each element of the sample covariance matrix by calculating the pairwise sample covariance, although the resulting sample covariance matrix is no longer positive definite. We are interested in the performance of covariance matrix estimators using different methods: the extreme case is that the sample covariance matrix does not exist and thus the high-dimensional covariance matrix techniques fail while the estimated covariance function still exists when the observations at time t_1 are missed for half of the subjects and the observations at time t_2 are missed for the other half. In this paper:

(i) We propose a new estimator for the covariance matrix of the within-subject errors, which is called the regularized-covariance-function-based estimator.
(ii) We derive the error bounds of the covariance matrices estimators (both the proposed regularized-covariance-function-based estimator and the high-dimensional covariance matrix estimator [12]) in Frobenius norm.
(iii) We compare the upper bounds of the convergence rates for these two estimators in Frobenius norm and show two cases when the regularized-covariance-function-based estimator outperforms the high-dimensional covariance matrix estimator in Section 4: the "high-rank sparse" case and "low-rank dense" case when $p \gg n^{1/4} \log n$.
(iv) We compare the performance of these two estimators for the incomplete longitudinal data and study the influence of the incompleteness on these two different approaches.

The rest of the article is organized as follows. Section 2 describes the general framework (profile weighted least squares estimation procedure) of estimating the parametric components in the varying-coefficient partially lineal model. Two different estimators for the covariance matrix of within-subject errors are described in Section 3. Theoretical properties and comparisons of the two estimators are presented in Section 4. In Sections 5 and 6 the proposed method is illustrated via simulation studies and real data examples, respectively. All technical proofs are relegated to Appendix.

2. Framework of the efficient estimation procedure

Suppose all longitudinal observations from different subjects (or clusters) are made on a fixed time interval $\mathcal{T} \subset \mathbb{R}$, e.g., $\mathcal{T} = [0, 1]$. The data consist of n independent subjects. For the ith subject, $i \in \{1, \ldots, n\}$, the response variable $Y_i(t)$ and the covariates $\{\boldsymbol{X}_i(t), \boldsymbol{Z}_i(t)\}$ are collected at time points $t = t_j$, $j \in \{1, \ldots, p\}$, where p is the total number of

observations for each subject. One special case is the equally spaced time points scheme, i.e., $t_j = j/p, j \in \{1, \ldots, p\}$. In this article, we consider the semi-parametric varying-coefficient partially linear model:

$$Y(t) = \boldsymbol{X}(t)^\top \boldsymbol{\alpha}(t) + \boldsymbol{Z}(t)^\top \boldsymbol{\beta} + \varepsilon(t),$$

i.e.,

$$Y_i(t_j) = \boldsymbol{X}_i(t_j)^\top \boldsymbol{\alpha}(t_j) + \boldsymbol{Z}_i(t_j)^\top \boldsymbol{\beta} + \varepsilon_i(t_j), \quad i \in \{1, \ldots, n\},\ j \in \{1, \ldots, p\}, \tag{2}$$

where $\boldsymbol{\alpha}(t)$ comprises m unknown smooth functions, $\boldsymbol{\beta}$ is a q-dimensional unknown parameter vector, and $\varepsilon_i(t),\ i \in \{1, \ldots, n\}$ are the i.i.d. error processes with $\mathrm{E}\{\varepsilon_i(t) \mid \boldsymbol{X}_i(t), \boldsymbol{Z}_i(t)\} = 0$. To consider the within-subject dependence, we assume that $\varepsilon_i(t)$ can be decomposed into two independent error processes:

$$\varepsilon_i(t) = \eta_i(t) + \zeta_i(t),$$

where $\eta_i(t),\ i \in \{1, \ldots, n\}$ are the i.i.d. mean zero error processes capturing the within-subject dependence or temporal correlation, and $\zeta_i(t),\ i \in \{1, \ldots, n\}$ are the i.i.d. measurement errors, see [5,20]. For $t_1, t_2 \in \mathcal{T}$, suppose

$$\mathrm{Cov}\{\eta_i(t_1), \eta_i(t_2)\} = R_\eta(t_1, t_2), \quad \mathrm{Cov}\{\zeta_i(t_1), \zeta_i(t_2)\} = \sigma_\zeta^2(t_1)\mathrm{I}(t_1 = t_2), \quad i \in \{1, \ldots, n\},$$

where $\mathrm{I}(\cdot)$ is an indicator function, $R_\eta(\cdot, \cdot)$ and $\sigma_\zeta^2(\cdot)$ are smooth functions. Then the covariance function $R(t_1, t_2)$ of $\varepsilon_i(t) = \eta_i(t) + \zeta_i(t)$ is given by

$$R(t_1, t_2) \equiv \mathrm{Cov}\{\varepsilon_i(t_1), \varepsilon_i(t_2)\} = R_\eta(t_1, t_2) + \sigma_\zeta^2(t_1)\mathrm{I}(t_1 = t_2), \quad i \in \{1, \ldots, n\}, \tag{3}$$

which is a smooth surface except on the diagonal points where $t_1 = t_2$.

In practice, estimation of $\{\boldsymbol{\alpha}(t), \boldsymbol{\beta}\}$ must be done in multiple steps. Their initial estimators are constructed by ignoring within-subject correlation. With the initial estimators of $\{\boldsymbol{\alpha}(t), \boldsymbol{\beta}\}$, we can estimate the covariance matrix of within-subject errors. Finally, we can estimate $\{\boldsymbol{\alpha}(t), \boldsymbol{\beta}\}$ more efficiently by using the estimator of covariance matrix. We propose the efficient estimators for $\{\boldsymbol{\alpha}(t), \boldsymbol{\beta}\}$ using the profile weighted least squares techniques.

2.1. Step 1: Initial estimator

For a given $\boldsymbol{\beta}$, model (2) reduces to a varying-coefficient model:

$$Y_i(t_j) - \boldsymbol{Z}_i(t_j)^\top \boldsymbol{\beta} = \boldsymbol{X}_i(t_j)^\top \boldsymbol{\alpha}(t_j) + \eta_i(t_j) + \zeta_i(t_j). \tag{4}$$

Ignoring the within-subject correlation or $\eta_i(t)$, we first apply the profile local linear regression to get the initial estimators of $\{\boldsymbol{\alpha}(t), \boldsymbol{\beta}\}$, see [3]. The profile least squares estimators of $\{\boldsymbol{\alpha}(t), \boldsymbol{\beta}\}$ will enjoy the closed form if we use the following matrix notations, let

$$\begin{aligned}
&\boldsymbol{Y} = (\boldsymbol{Y}_1^\top, \ldots, \boldsymbol{Y}_n^\top)^\top, \qquad \boldsymbol{Z} = (\boldsymbol{Z}_1^\top, \ldots, \boldsymbol{Z}_n^\top)^\top, \qquad \boldsymbol{m} = (\boldsymbol{m}_1^\top, \ldots, \boldsymbol{m}_n^\top)^\top,\\
&\boldsymbol{Y}_i = (Y_i(t_1), \ldots, Y_i(t_p))^\top,\ \boldsymbol{Z}_i = (\boldsymbol{Z}_i(t_1), \ldots, \boldsymbol{Z}_i(t_p))^\top,\ \boldsymbol{m}_i = (\boldsymbol{X}_i(t_1)^\top \boldsymbol{\alpha}(t_1), \ldots, \boldsymbol{X}_i(t_p)^\top \boldsymbol{\alpha}(t_p))^\top.
\end{aligned}$$

Now model (4) can be written as

$$\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta} = \boldsymbol{m} + \boldsymbol{\eta} + \boldsymbol{\zeta}, \tag{5}$$

where $\boldsymbol{\eta} = (\eta_1(t_1), \ldots, \eta_1(t_p), \ldots, \eta_n(t_1), \ldots, \eta_n(t_p))^\top$ and $\boldsymbol{\zeta} = (\zeta_1(t_1), \ldots, \zeta_1(t_p), \ldots, \zeta_n(t_1), \ldots, \zeta_n(t_p))^\top$. Let $K(\cdot)$ be a kernel function, $K_h(\cdot) = h^{-1}K(\cdot/h)$ where $h > 0$ is the bandwidth, and $\boldsymbol{X}_i(t) = (X_{i1}(t), \ldots, X_{im}(t))^\top$. Then the local linear estimator $\widehat{\boldsymbol{\alpha}}_{\boldsymbol{\beta}}(t_0) \equiv (\widehat{a}_1, \ldots, \widehat{a}_m)^\top$ for the nonparametric component at time point t_0 can be derived by:

$$\min_{\{a_l, b_l\}_{l=1}^m} \sum_{i=1}^n \sum_{j=1}^p \Big[Y_i(t_j) - \boldsymbol{Z}_i(t_j)^\top \boldsymbol{\beta} - \sum_{l=1}^m \{a_l + b_l(t_j - t_0)\} X_{il}(t_j) \Big]^2 K_h(t_j - t_0).$$

Substituting $\widehat{\boldsymbol{\alpha}}_{\boldsymbol{\beta}}(\cdot)$ into model (5) results in the ordinary least squares estimator of $\boldsymbol{\beta}$:

$$\widehat{\boldsymbol{\beta}}^{\mathrm{ini}} = \{\boldsymbol{Z}^\top (\mathbf{I} - \mathbf{S}_h)^\top (\mathbf{I} - \mathbf{S}_h)\boldsymbol{Z}\}^{-1} \boldsymbol{Z}^\top (\mathbf{I} - \mathbf{S}_h)^\top (\mathbf{I} - \mathbf{S}_h)\boldsymbol{Y}, \tag{6}$$

where $\mathbf{S}_h$ is a smoothing matrix of the local linear smoother and $\mathbf{I}$ is an identity matrix, see [3] for details. The initial estimator of the nonparametric component $\boldsymbol{\alpha}(\cdot)$ is just $\widehat{\boldsymbol{\alpha}}_{\widehat{\boldsymbol{\beta}}^{\mathrm{ini}}}(\cdot)$.

2.2. Step 2: Refined estimator

First, let $\boldsymbol{\Sigma}_0$ and $\widehat{\boldsymbol{\Sigma}}$ be the true and estimated covariance matrix of $(\varepsilon_i(t_1), \ldots, \varepsilon_i(t_p))^\top$ respectively, i.e.,

$$\boldsymbol{\Sigma}_0 = \big[R(t_j, t_k)\big]_{j,k=1}^p, \quad \widehat{\boldsymbol{\Sigma}} = \big[\widehat{R}(t_j, t_k)\big]_{j,k=1}^p, \tag{7}$$

where $R(\cdot,\cdot)$ is defined in (3) and the different choices of the covariance function estimator $\widehat{R}(\cdot,\cdot)$ will be shown in Section 3. Since we ignore the within-subject correlation or $\eta_i(t)$ in Step 1, the initial least squares estimator $\widehat{\boldsymbol{\beta}}^{\text{ini}}$ in (6) is not efficient. To improve the efficiency for estimating $\boldsymbol{\beta}$, we use the profile weighted least squares estimator as follows:

$$\{\mathbf{Z}^\top(\mathbf{I}-\mathbf{S}_h)^\top\mathbf{W}(\mathbf{I}-\mathbf{S}_h)\mathbf{Z}\}^{-1}\mathbf{Z}^\top(\mathbf{I}-\mathbf{S}_h)^\top\mathbf{W}(\mathbf{I}-\mathbf{S}_h)\boldsymbol{Y}, \tag{8}$$

where $\mathbf{W}$ is a weight matrix, called a working covariance matrix. Then the initial estimator $\widehat{\boldsymbol{\beta}}^{\text{ini}}$ in (6) is just a special case of (8) with $\mathbf{W}$ being an identity matrix. As usual, misspecification of the working covariance matrix does not affect the consistency of the resulting estimators, but does affect the efficiency. Fan et al. [3] have shown that the most efficient estimator for $\boldsymbol{\beta}$ among the profile weighted least squares estimators given by (8) is the one that uses the inverse of the true variance–covariance matrix of errors as the weight matrix, that is, $\mathbf{W}=\text{diag}(\boldsymbol{\Sigma}_0^{-1},\ldots,\boldsymbol{\Sigma}_0^{-1})$ is a n-blocked diagonal matrix, with each block being $\boldsymbol{\Sigma}_0^{-1}$. Because $\boldsymbol{\Sigma}_0$ is unknown, we can use the estimator $\widehat{\boldsymbol{\Sigma}}$. Different choices of the covariance matrix estimator $\widehat{\boldsymbol{\Sigma}}$ will be shown and compared in Sections 3 and 4, so the final refined estimator of $\boldsymbol{\beta}$ is

$$\widehat{\boldsymbol{\beta}}=\{\mathbf{Z}^\top(\mathbf{I}-\mathbf{S}_h)^\top\widehat{\mathbf{W}}(\mathbf{I}-\mathbf{S}_h)\mathbf{Z}\}^{-1}\mathbf{Z}^\top(\mathbf{I}-\mathbf{S}_h)^\top\widehat{\mathbf{W}}(\mathbf{I}-\mathbf{S}_h)\boldsymbol{Y}, \tag{9}$$

where $\widehat{\mathbf{W}}=\text{diag}(\widehat{\boldsymbol{\Sigma}}^{-1},\ldots,\widehat{\boldsymbol{\Sigma}}^{-1})$ is a $np\times np$ matrix. The profile least squares estimator for the nonparametric component is simply $\widehat{\boldsymbol{\alpha}}_{\widehat{\boldsymbol{\beta}}}(\cdot)$.

3. Proposed estimators of the covariance matrix

3.1. Regularized-covariance-function-based estimator

After we get the initial estimators $\widehat{\boldsymbol{\beta}}^{\text{ini}}$ and $\widehat{\boldsymbol{\alpha}}_{\widehat{\boldsymbol{\beta}}^{\text{ini}}}(t)$ in Section 2, the residuals are

$$\widehat{\varepsilon}_i(t_j)=Y_i(t_j)-\boldsymbol{X}_i(t_j)^\top\widehat{\boldsymbol{\alpha}}_{\widehat{\boldsymbol{\beta}}^{\text{ini}}}(t_j)-\boldsymbol{Z}_i(t_j)^\top\widehat{\boldsymbol{\beta}}^{\text{ini}},\quad i\in\{1,\ldots,n\},\ j\in\{1,\ldots,p\}.$$

Then we can derive the nonparametric regularized estimator of $R_\eta(\cdot,\cdot)$ based on $\widehat{\varepsilon}_i(t_j)$. Since there are too many parameters in $R_\eta(\cdot,\cdot)$, a penalty for over-parametrization is imposed to regularize the covariance function. Jia et al. [7] assumed that $R_\eta(\cdot,\cdot)$ resides in a tensor product of Hilbert space $\mathcal{W}_2^2(\mathcal{T})\otimes\mathcal{W}_2^2(\mathcal{T})$, which is the closure of the following linear space

$$\text{span}\{f(s)g(t):\ f(\cdot),g(\cdot)\in\mathcal{W}_2^2(\mathcal{T})\}, \tag{10}$$

where $\mathcal{W}_2^2(\mathcal{T})=\{f:f',f''$ are absolutely continuous, $f''\in L_2(\mathcal{T})\}$ is a Sobolev space endowed with the squared norm $\int_{\mathcal{T}}(f'')^2$. Because $\mathcal{W}_2^2(\mathcal{T})\otimes\mathcal{W}_2^2(\mathcal{T})$ is dense in the continuous bivariate function space, we can find an element in $\mathcal{W}_2^2(\mathcal{T})\otimes\mathcal{W}_2^2(\mathcal{T})$ that approximates any continuous bivariate function very well. Then let $\widehat{R}_\eta^{\text{ini}}(s,t)\in\mathcal{W}_2^2(\mathcal{T})\otimes\mathcal{W}_2^2(\mathcal{T})$ be the minimizer of

$$\frac{1}{np(p-1)}\sum_{i=1}^{n}\sum_{1\le j_1\ne j_2\le p}\left\{\widehat{\varepsilon}_i(t_{j_1})\widehat{\varepsilon}_i(t_{j_2})-R_\eta(t_{j_1},t_{j_2})\right\}^2+\lambda_{1,n}P(R_\eta), \tag{11}$$

where $\lambda_{1,n}\ge 0$ is a tuning parameter, and $P(R_\eta)$ is a penalty function for $R_\eta(s,t)=\sum_{j\ge1}a_jf_j(s)g_j(t)$ with the following form:

$$\begin{aligned}P(R_\eta)&=\iint_{t,s\in\mathcal{T}}\left\{\frac{\partial^4R_\eta(s,t)}{\partial^2s\partial^2t}\right\}^2dsdt=\iint_{t,s\in\mathcal{T}}\left\{\sum_{j\ge1}a_jf_j''(s)g_j''(t)\right\}^2dsdt\\&=\sum_{i,j\ge1}\sum a_ia_j\int_{\mathcal{T}}\{f_i''(s)f_j''(s)\}ds\int_{\mathcal{T}}\{g_i''(t)g_j''(t)\}dt.\end{aligned}$$

Cai and Yuan [1] have shown that the minimizer $\widehat{R}_\eta^{\text{ini}}$ of (11) must have the following closed form

$$\widehat{R}_\eta^{\text{ini}}(s,t)=\mathbf{H}(s)^\top\widehat{\mathbf{A}}\mathbf{H}(t)=\sum_{j,k=1}^{p}\widehat{\mathbf{A}}_{j,k}H(s,t_j)H(t,t_k), \tag{12}$$

where $\widehat{\mathbf{A}}$ is a $p\times p$ symmetric matrix which can be derived by the package "`ssfcov`", $\widehat{\mathbf{A}}_{j,k}$ is the element in the jth row and kth column of $\widehat{\mathbf{A}}$, and

$$\mathbf{H}(s)=\big(H(s,t_1),\ldots,H(s,t_p)\big)^\top\in\mathbb{R}^p,$$

where $H(s,t)=B_2(s)B_2(t)/4-B_4(|s-t|)/24$ and B_r is the rth Bernoulli polynomial, see [10]. Thanks to the representation (12), Jia et al. [7] showed that the eigenvalues and eigenfunctions of $\widehat{R}_\eta^{\text{ini}}(s,t)=\mathbf{H}(s)^\top\widehat{\mathbf{A}}\mathbf{H}(t)$ can actually be computed explicitly:

$$\widehat{R}_\eta^{\text{ini}}(s,t)=\mathbf{H}(s)^T\mathbf{Q}^{-1/2}\big(\mathbf{Q}^{1/2}\widehat{\mathbf{A}}\mathbf{Q}^{1/2}\big)\mathbf{Q}^{-1/2}\mathbf{H}(t)=\mathbf{H}(s)^T\mathbf{Q}^{-1/2}\left\{\sum_{k=1}^{p}\widehat{\lambda}_{(k)}\widehat{\boldsymbol{u}}_k\widehat{\boldsymbol{u}}_k^T\right\}\mathbf{Q}^{-1/2}\mathbf{H}(t)\equiv\sum_{k=1}^{p}\widehat{\lambda}_{(k)}\widehat{\psi}_k(s)\widehat{\psi}_k(t),$$

where $\widehat{\lambda}_{(1)} \geq \widehat{\lambda}_{(2)} \geq \cdots \geq \widehat{\lambda}_{(p)}$ are the decreasing eigenvalues and $\{\widehat{\boldsymbol{u}}_1, \widehat{\boldsymbol{u}}_2, \ldots, \widehat{\boldsymbol{u}}_p\}$ are the corresponding eigenvectors of $\mathbf{Q}^{1/2}\widehat{\mathbf{A}}\mathbf{Q}^{1/2}$, where $\mathbf{Q} = \left[\int_{\mathcal{T}} H(s,t_j)H(s,t_k)ds\right]_{1\leq j,k\leq p}$, and $\widehat{\psi}_k(\cdot) = \widehat{\boldsymbol{u}}_k^T\mathbf{Q}^{-1/2}\mathbf{H}(\cdot)$. Since $\widehat{\lambda}_{(k)}$'s are not necessarily positive, we need to truncate the negative eigenvalues. Then the adjusted estimator for R_η is defined as

$$\widehat{R}_\eta(s,t) = \sum_{k=1}^{p} \widehat{\lambda}_{(k)} \mathrm{I}(\widehat{\lambda}_{(k)} > \tau)\widehat{\psi}_k(s)\widehat{\psi}_k(t), \quad s,t \in \mathcal{T}, \tag{13}$$

where $\tau \geq 0$ is a predetermined threshold for the eigenvalues (e.g., 0.01) or a percentage (e.g., 1 percent) of the sum of all the positive eigenvalues.

The diagonal element of $R(\cdot,\cdot)$ requires a special treatment since it involves both $R_\eta(t,t)$ and $\sigma_\zeta^2(t)$. Define $\sigma^2(t) \equiv R(t,t) = R_\eta(t,t) + \sigma_\zeta^2(t)$, then $R(s,t)$ in (3) can also be written as

$$R(s,t) = R_\eta(s,t)\mathrm{I}(s \neq t) + \sigma^2(t)\mathrm{I}(s = t), \tag{14}$$

that means, $R(s,t)$ is $R_\eta(s,t)$ when $s \neq t$ and $\sigma^2(t)$ when $s = t$. For the irregular or subject specific setting, Jia et al. [7] used the standard one-dimensional local linear estimator of $\sigma^2(t)$ based on the squared residuals $\widehat{\varepsilon}_i^2(t_j),\ i \in \{1,\ldots,n\}$ and $j \in \{1,\ldots,p\}$. For the balanced and nearly balanced case, we define the following regularized-covariance-function-based estimator $\widehat{\mathbf{\Sigma}}_{\mathrm{recf}}$ of $\mathbf{\Sigma}_0$ as:

$$\widehat{\mathbf{\Sigma}}_{\mathrm{recf}}(j,k) = \begin{cases} \widehat{R}_\eta(t_j,t_k), & \text{if } j \neq k, \\ s_j^2, & \text{if } j = k, \end{cases} \tag{15}$$

where $s_j^2 = (n-1)^{-1}\sum_{i=1}^n \{\widehat{\varepsilon}_i(t_j) - \bar{\varepsilon}(t_j)\}^2$ is the sample variance of $\widehat{\varepsilon}_i(t_j),\ i \in \{1,\ldots,n\}$, with $\bar{\varepsilon}(t_j) = n^{-1}\sum_{i=1}^n \widehat{\varepsilon}_i(t_j)$. It is easy to show that with probability tending to 1, $\widehat{\mathbf{\Sigma}}_{\mathrm{recf}}$ defined in (15) is positive definite. In practice, the diagonal elements of the estimator $\widehat{\mathbf{\Sigma}}_{\mathrm{recf}}$ can be defined as $\widehat{\mathbf{\Sigma}}_{\mathrm{recf}}(j,j) = \max\{s_j^2, \widehat{R}_\eta(t_j,t_j)\}$, $j \in \{1,\ldots,p\}$, to guarantee the positive definiteness when sample size n is small.

3.2. High-dimensional covariance matrix estimator

From the procedure of estimating the parametric component $\boldsymbol{\beta}$ in Section 2, the estimator of $\mathbf{\Sigma}_0 = \left[R(t_j,t_k)\right]_{j,k=1}^p = \mathrm{Cov}(\boldsymbol{\varepsilon}_i, \boldsymbol{\varepsilon}_i)$, $i \in \{1,\ldots,n\}$, is required to be positive definite, where $\boldsymbol{\varepsilon}_i = (\varepsilon_i(t_1),\ldots,\varepsilon_i(t_p))^\top$. When the dimension p is large, it is well known that the sample covariance matrix $\mathbf{S}$ does not work well. And sometimes it is reasonable to assume that the dependence between $Y(t)$ and $Y(s)$ is weak when $|t-s|$ is relatively large. Thus to estimate the covariance matrix $\mathbf{\Sigma}_0$, we adopt the Positive Definite Sparse Covariance-matrix Estimator (PDSCE) proposed by Rothman [12] when the dimension p of the covariance matrix is large. Other high-dimensional covariance matrix estimators with different constraints and structures may also be considered.

Let $\mathbf{M}^+$ be a diagonal matrix with the same diagonal as $\mathbf{M}$, and define $\mathbf{M}^- = \mathbf{M} - \mathbf{M}^+$. For a matrix $\mathbf{M}$, let $|\mathbf{M}|_q = \|\mathrm{vec}(\mathbf{M})\|_q$ denote the q-norm of the vector formed by stacking the columns of $\mathbf{M}$. Let $\|\mathbf{M}\|_F \equiv |\mathbf{M}|_2$ and $\|\mathbf{M}\|$ denote the Frobenius and spectral norms of $\mathbf{M}$ respectively. When $\mathbf{M}$ is a square matrix, $|\mathbf{M}|$ denotes its determinant, and $\mathbf{M} \succ 0$ indicates that $\mathbf{M}$ is symmetric and positive definite. Let $\mathbf{\Theta}_0 = \mathrm{Corr}(\boldsymbol{\varepsilon}_i, \boldsymbol{\varepsilon}_i)$ denote the true population correlation matrix of $\boldsymbol{\varepsilon}_i = (\varepsilon_i(t_1),\ldots,\varepsilon_i(t_p))^\top$. Rothman [12] proposed the correlation matrix estimator of $\mathbf{\Theta}_0$ by

$$\widehat{\mathbf{\Theta}}_\lambda = \underset{\mathbf{\Theta}\succ 0}{\arg\min}\big(\|\ \mathbf{\Theta} - \mathbf{\Gamma}\ \|_F^2\ /2 - \tau_n \log|\mathbf{\Theta}| + \lambda_{2,n}|\mathbf{\Theta}^-|_1\big), \tag{16}$$

where $\mathbf{\Gamma}$ is the sample correlation matrix of $\widehat{\boldsymbol{\varepsilon}}_i = (\widehat{\varepsilon}_i(t_1),\ldots,\widehat{\varepsilon}_i(t_p))^\top,\ i \in \{1,\ldots,n\}$, and $\lambda_{2,n} \geq 0$ is a tuning parameter, $\tau_n > 0$ is a small value. The logarithmic barrier term ensures the existence of a positive definite solution, and the lasso-type penalty is used to encourage sparse solutions. Then the proposed high-dimensional covariance matrix estimator of $\mathbf{\Sigma}_0$ is defined by

$$\widehat{\mathbf{\Sigma}}_\lambda = (\mathbf{S}^+)^{1/2}\widehat{\mathbf{\Theta}}_\lambda(\mathbf{S}^+)^{1/2}, \tag{17}$$

where $\mathbf{S}$ is the sample covariance matrix for the vectors $\widehat{\boldsymbol{\varepsilon}}_i = (\widehat{\varepsilon}_i(t_1),\ldots,\widehat{\varepsilon}_i(t_p))^\top$, $i \in \{1,\ldots,n\}$.

In the next section, we will derive the upper bounds of the convergence rates for both $\widehat{\mathbf{\Sigma}}_{\mathrm{recf}}$ in (15) and $\widehat{\mathbf{\Sigma}}_\lambda$ in (17) in Frobenius norm, and then compare their performance.

4. Theoretical results

In this section we investigate sampling properties of the covariance matrices estimators as $n \to \infty$ and $p \to \infty$. The proposed estimation procedures are applicable for various formulations for collecting longitudinal data. For example, we could assume that $t_j,\ j \in \{1,\ldots,p\}$ are either independent and identically distributed on $\mathcal{T}$ with a density $f_T(t)$, or the fixed designs $t_j = j/p, j \in \{1,\ldots,p\}$, see [3].

Without loss of generality, in the following theorems we derive the upper bounds of the convergence rates for the two estimated covariance matrices in Frobenius norm based on the true errors $\boldsymbol{\varepsilon}_i,\ i \in \{1,\ldots,n\}$ instead of the residuals

$\widehat{\boldsymbol{\varepsilon}}_i,\ i \in \{1, \ldots, n\}$. Let the pseudo-estimators $\overline{\boldsymbol{\Sigma}}_{\text{recf}}$ and $\overline{\boldsymbol{\Sigma}}_{\lambda}$ be derived by the same procedure in Section 3, with the residuals being replaced by $\boldsymbol{\varepsilon}_i,\ i \in \{1, \ldots, n\}$. In order to present the asymptotic properties of the proposed estimators in Section 3, we need the following technical assumptions.

Assumption 1. The observed time points $t_j,\ j \in \{1, \ldots, p\}$ are independent and identically distributed with a density $f_T(\cdot)$ bounded away from zero on $\mathcal{T}$.

Assumption 2. $\mathrm{E}\{\zeta_i^4(t_j)\} < \infty$ for $j \in \{1, \ldots, p\}$.

Assumption 3. $\lambda_{1,n} = K_1\{\log(n)/(pn)\}^{4/5}$, where K_1 is a sufficiently large constant.

Assumption 4. $\tau_n = O\big(\{n^{-1}s\log(p)\}^{1/2}\|\boldsymbol{\Theta}_0^{-1}\|_F^{-1}\big)$, and $\lambda_{2,n} = K_2\{n^{-1}\log(p)\}^{1/2}$, where K_2 is a sufficiently large constant.

Assumption 5. $(s+1)\log(p) = o(n)$, where s denotes the number of nonzero off-diagonal entries in $\boldsymbol{\Sigma}_0$.

Assumption 6. $\mathrm{E}\big[\exp\{u\varepsilon_i^2(t_j)\}\big] \le C_1 < \infty$ for $0 < |u| < C_2$, $j \in \{1, \ldots, p\}$, and C_1, C_2 are positive constants.

Assumption 7. $0 < \kappa_1 \le \lambda_{\min}(\boldsymbol{\Sigma}_0) \le \lambda_{\max}(\boldsymbol{\Sigma}_0) \le \kappa_2 < \infty$, where κ_1 and κ_2 are constants, $\lambda_{\min}(\boldsymbol{\Sigma}_0)$ and $\lambda_{\max}(\boldsymbol{\Sigma}_0)$ denote the smallest and largest eigenvalues of $\boldsymbol{\Sigma}_0$.

Theorem 1. *Under Assumptions 1–3, we have*

$$\|\overline{\boldsymbol{\Sigma}}_{\text{recf}} - \boldsymbol{\Sigma}_0\|_F^2 = O_p\Big(\Big\{\frac{p^{3/2}\log(n)}{n}\Big\}^{4/5} + \frac{p^2}{n}\Big). \tag{18}$$

The next corollary shows that when the dimension p of the covariance matrix is quite large or small compared to the sample size n, the error bounds of $\overline{\boldsymbol{\Sigma}}_{\text{recf}}$ in (18) are different.

Corollary 1. *Under Assumptions 1–3, according to (18),*

(i) *If $p \gg n^{1/4}\log(n)$, then*

$$\|\overline{\boldsymbol{\Sigma}}_{\text{recf}} - \boldsymbol{\Sigma}_0\|_F^2 = O_p\Big(\frac{p^2}{n}\Big). \tag{19}$$

(ii) *If $p \ll n^{1/4}\log(n)$, then*

$$\|\overline{\boldsymbol{\Sigma}}_{\text{recf}} - \boldsymbol{\Sigma}_0\|_F^2 = O_p\Big(\Big\{\frac{p^{3/2}\log(n)}{n}\Big\}^{4/5}\Big).$$

The following theorem shows the upper bound of the convergence rate for the PDSCE estimator $\overline{\boldsymbol{\Sigma}}_{\lambda}$ in Frobenius norm.

Theorem 2. *Under Assumptions 4–7, we have*

$$\|\overline{\boldsymbol{\Sigma}}_{\lambda} - \boldsymbol{\Sigma}_0\|_F^2 = O_p\Big\{\frac{r(s+1)\log(p)}{n}\Big\}, \tag{20}$$

where r is the rank of $(\overline{\boldsymbol{\Sigma}}_{\lambda} - \boldsymbol{\Sigma}_0)^{\top}(\overline{\boldsymbol{\Sigma}}_{\lambda} - \boldsymbol{\Sigma}_0)$.

Now let us compare the upper bounds of convergence rates (18) and (20). According to Corollary 1, it is obvious that when $p \gg n^{1/4}\log(n)$, the upper bound of convergence rate (19) is smaller than (20) when

$$r(s+1)\log(p)/p^2 \to \infty. \tag{21}$$

The following situations are the sufficient conditions for (21), that means the regularized-covariance-function-based estimator outperforms the Positive Definite Sparse Covariance-matrix Estimator (PDSCE) in terms of the error bounds in Frobenius norm when:

(i) "high-rank sparse" case: $r \asymp p$, $s \asymp p$;
(ii) "low-rank dense" case: $r = O(1)$, $s \asymp p^2$.

To further investigate the relationship between the quantity $r(s+1)$, the dimension p and the sample size n, suppose $p = n^{\theta}$, where $\theta \ge 0$ and $r(s+1) = p^{\gamma} = n^{\theta\gamma}$, it is obvious that $0 \le \gamma \le 3$.

Theorem 3. *Under Assumptions 1–7, if (θ, γ) satisfy*

(i) $\theta > 1/4$, $2 \le \gamma \le 3$;
(ii) $0 < \theta \le 1/4$, $(6\theta+1)/(5\theta) \le \gamma \le 3$,

Table 1
(Simulation 1) Biases and standard deviations of $\widehat{\beta}_2$ over 100 simulations using different Methods I–IV (with $n = 50$ and $p = 6, 12, 18$).

Method	$p = 6$		$p = 12$		$p = 18$	
	Bias	SD	Bias	SD	Bias	SD
Method I	.0028	.0411	−.0011	.0297	−.0055	.0267
Method II	.0046	.0416	−.0052	.0319	−.0049	.0311
Method III	.0017	.0409	−.0053	.0307	−.0032	.0291
Our Method IV	.0030	.0417	−.0024	.0286	−.0030	.0254

then the upper bound of the convergence rate of the regularized-covariance-function-based estimator $\overline{\Sigma}_{\text{recf}}$ is smaller than that of the Positive Definite Sparse Covariance-matrix Estimator (PDSCE) $\overline{\Sigma}_{\lambda}$ in Frobenius norm.

5. Simulation study

5.1. Simulation 1

In this section, we investigate the finite sample properties of the estimators proposed in Section 3 through Monte Carlo simulations. Suppose the data are generated from the following model:

$$Y_i(t_j) = \boldsymbol{X}_i(t_j)^\top \boldsymbol{\alpha}(t_j) + \boldsymbol{Z}_i(t_j)^\top \boldsymbol{\beta} + \eta_i(t_j) + \zeta_i(t_j), \quad i \in \{1, \ldots, n\},\ j \in \{1, \ldots, p\}.$$

Similar to the real data example ($n = 60$, $p = 11$) in Section 6, we assume that the sample size n is 50, and $p = 6, 12, 18$. Let $\mathcal{T} = [0, 1]$, and the observation time points $t_j = j/(p+1)$, $j \in \{1, \ldots, p\}$. We let the coefficients $\boldsymbol{\alpha}(t)$ and $\boldsymbol{\beta}$ be two dimensional in our simulation, and further set $X_1(t) \equiv 1$ to include an intercept term. We generate the covariates in the following way: For a given t, $(X_2(t), Z_1(t))^\top$ follows a bivariate normal distribution with mean 0, variance 1, and correlation 0.5, and $Z_2(t)$ is a Bernoulli distributed random variable with success probability 0.5 and independent of $X_2(t)$ and $Z_1(t)$. In this simulation we set $\boldsymbol{\beta} = (1, 2)^\top$, $\alpha_1(t) = \sqrt{t}$, and $\alpha_2(t) = \sin(2\pi t)$. The within-subject errors $\eta_i(t)$'s are generated from a temporal mixed effects model (see [7]):

$$\eta_i(t) = \sum_{k=1}^{L} \xi_{i,k} \psi_k(t), \quad i \in \{1, \ldots, n\},$$

where $\xi_{i,k}$'s are independent normal random variables with $N(0, \sigma^2)$. Thus the covariance function of $\eta_i(t)$ is

$$R_\eta(s, t) = \sigma^2 \sum_{k=1}^{L} \psi_k(s)\psi_k(t). \tag{22}$$

We set $L = 1$, $\sigma^2 = 0.1$, and $\psi_1(t) = \cos(\pi t)$, which corresponds to the "low-rank dense" case. Finally we assume the measurement errors $\zeta_i(t_j)$ follow $N(0, \left(\sqrt{0.1}\right)^2)$ and are independent of $\eta_i(t_j)$. For comparison, in each simulated dataset, we fit the model using four different estimators of the covariance matrix $\boldsymbol{\Sigma}_0$ in (7):

Method I. True covariance matrix $\boldsymbol{\Sigma}_0$.
Method II. Sample covariance matrix $\mathbf{S}$.
Method III. Positive Definite Sparse Covariance-matrix Estimator (PDSCE) $\widehat{\boldsymbol{\Sigma}}_{\lambda}$ in (17).
Method IV. Regularized-covariance-function-based estimator $\widehat{\boldsymbol{\Sigma}}_{\text{recf}}$ in (15).

For a fair comparison, we use the same bandwidth $h = 0.15$ when estimating $\boldsymbol{\alpha}(\cdot)$ for all the estimators. Throughout the simulations and the real data examples in the next section, we use the Epanechnikov kernel, and the tuning parameter $\lambda_{1,n}$ in (11) is selected automatically by the package "`ssfcov`", the tuning parameters $\lambda_{2,n}$ and τ_n in (16) are selected by the package "`PDSEC`".

Table 1 and Fig. 1 summarize the results over 100 simulations. We assess the performance of different covariance matrix estimators by calculating the biases and standard errors of the 100 estimators of coefficients. In the table, "Bias" represents the median of the 100 estimators subtracting the true value, "SD" represents the median absolute deviation of the 100 estimators divided by a factor of 0.6745. We also assess the performance of different covariance matrix estimators $\widehat{\boldsymbol{\Sigma}}$ in Method II, III, IV by calculating the average of $\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0\|_F^2/p^2$ over 100 simulations. Since $\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0\|_F^2$ is the sum of p^2 squared terms for a $p \times p$ matrix estimator $\widehat{\boldsymbol{\Sigma}}$, we divide it by a factor p^2 to reduce the influence of dimension p.

From Table 1, all parameter estimators are asymptotically unbiased, which is confirmed from the numerical results: the biases are much smaller than the standard errors in all cases. This is also shown by the box-plots in Fig. 1. In terms of the efficiency, theoretically, the oracle estimator (Method I) using the true covariance matrix $\boldsymbol{\Sigma}_0$ should be the best. For Method II using the sample covariance matrix $\mathbf{S}$, it works well when the dimension p is quite small (e.g., $p = 6$), but the "SD" of the coefficients are much larger than that of the other methods when p is large. For the comparison of our proposed estimator $\widehat{\boldsymbol{\Sigma}}_{\text{recf}}$ (Method IV) and the PDSCE estimator $\widehat{\boldsymbol{\Sigma}}_{\lambda}$ (Method III), when p is small, these two estimators are comparable, and as p becomes large, our proposed estimator $\widehat{\boldsymbol{\Sigma}}_{\text{recf}}$ (Method IV) works better than the other methods.

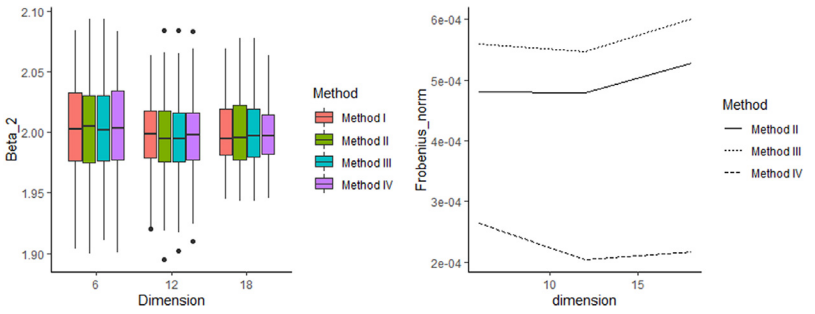


Fig. 1. (Simulation 1) Left panel: boxplots for $\widehat{\beta}_2$ with different methods and different $p = 6, 12, 18$; right panel: average of $\|\widehat{\Sigma} - \Sigma_0\|_F^2/p^2$ over 100 simulations with different $\widehat{\Sigma}$ and different $p = 6, 12, 18$.

Table 2
(Simulation 2) Biases and standard deviations of $\widehat{\beta}_2$ over 100 simulations using different Methods I–IV (with $n = 50$ and $p = 6, 12, 18$).

Method	$p = 6$		$p = 12$		$p = 18$	
	Bias	SD	Bias	SD	Bias	SD
Method I	.0081	.0575	−.0034	.0353	.0066	.0366
Method II	.0062	.0541	−.0036	.0344	.0035	.0351
Method III	.0077	.0566	−.0038	.0378	.0055	.0374
Our Method IV	.0082	.0519	−.0010	.0398	.0059	.0362

Fig. 1 shows the average of $\|\widehat{\Sigma} - \Sigma_0\|_F^2/p^2$ over 100 simulations with different estimators $\widehat{\Sigma}$ and different dimension p. When we compare the error bounds in Frobenius norm, our proposed estimator $\widehat{\Sigma}_{\text{recf}}$ (Method IV) outperforms the other two estimators, and the performance of the PDSCE estimator $\widehat{\Sigma}_\lambda$ (Method III) is even worse than the sample covariance matrix **S** (Method II). This is because the true covariance matrix Σ_0 defined by (22) is not sparse. In the next simulation where the true covariance matrix Σ_0 is sparse or nearly sparse, it is shown that the PDSCE estimator $\widehat{\Sigma}_\lambda$ also works quite well.

Fig. 2 also shows the advantage of our proposed method when the dimension p is large ($p = 18$). From the plot, the true covariance matrix Σ_0 in Fig. 2(a) is very complicated. Fig. 2(b) shows the sample covariance matrix estimator **S** (Method II) and Fig. 2(c) shows the Positive Definite Sparse Covariance-matrix Estimator $\widehat{\Sigma}_\lambda$ (Method III). Our proposed estimator $\widehat{\Sigma}_{\text{recf}}$ in (15) is displayed in Fig. 2(d). It is obvious that our proposed estimator $\widehat{\Sigma}_{\text{recf}}$ (Method IV) performs better than the PDSCE estimator $\widehat{\Sigma}_\lambda$ (Method III) and captures the structure of the true covariance matrix very well when p is large.

5.2. Simulation 2

In simulation 2, we study the performance of different methods in the "high-rank sparse" case when the true covariance matrix Σ_0 is sparse or nearly sparse. The data are generated with the same setup as in the previous simulation, except that $\eta_i(t)$ is a Gaussian process with the covariance function:

$$R_\eta(s,t) = \sigma(s)\sigma(t)\rho^{\kappa|s-t|}, \quad s,t \in [0,1]. \tag{23}$$

We set marginal variance $\sigma^2(t) = 0.1$, and set $\rho = 0.4$ and $\kappa = 10$ in (23). When $|s-t| > 0.5$, $R_\eta(s,t) \le 10^{-3}$, which is regarded to be close to 0. We apply the semi-parametric regression methods using the true covariance matrix Σ_0 (Method I), the sample covariance matrix estimator **S** (Method II), the PDSCE estimator $\widehat{\Sigma}_\lambda$ (Method III) and our proposed estimator $\widehat{\Sigma}_{\text{recf}}$ (Method IV) to the simulated data and repeat the simulation 100 times. The selection methods of h, $\lambda_{1,n}$, $\lambda_{2,n}$ and τ_n are the same as those in the previous simulation. The results for the estimated coefficients and covariance matrices are summarized in Table 2, Figs. 3 and 4.

From Table 2 and Fig. 3, again all parameter estimators are asymptotically unbiased since the biases are much smaller than the standard errors. From the right panel of Fig. 3, the PDSCE estimator $\widehat{\Sigma}_\lambda$ (Method III) performs better than the sample covariance matrix estimator **S** (Method II) in terms of the error bounds in Frobenius norm, which is quite different from Simulation 1. In addition, our proposed estimator $\widehat{\Sigma}_{\text{recf}}$ (Method IV) and the PDSCE estimator $\widehat{\Sigma}_\lambda$ (Method III) are comparable when sample size is small ($p = 6$), and the average squared Frobenius norm of our proposed estimator $\widehat{\Sigma}_{\text{recf}}$ (Method IV) is smaller than that of the PDSCE estimator $\widehat{\Sigma}_\lambda$ (Method III) as p becomes larger, which verifies the theorems in Section 4. From Fig. 4, when p is large ($p = 18$), both our proposed estimator $\widehat{\Sigma}_{\text{recf}}$ and the PDSCE estimator $\widehat{\Sigma}_\lambda$ work quite well and capture the structure of the true covariance matrix Σ_0 in Fig. 4(*a*).

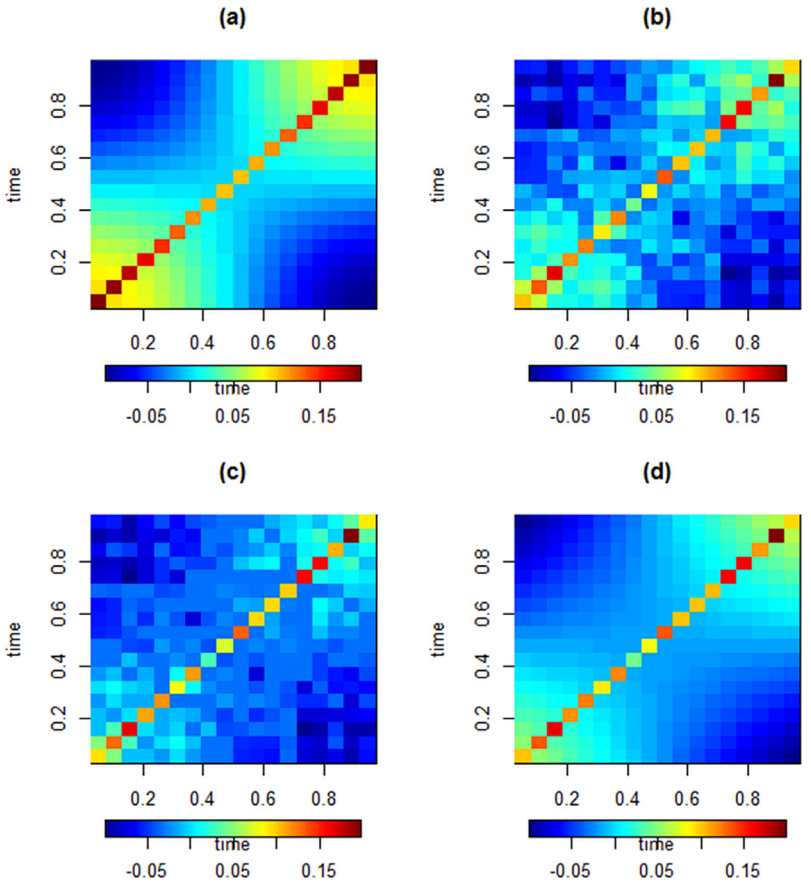


Fig. 2. (Simulation 1) Panel (a): true covariance matrix $\mathbf{\Sigma}_0$ in Method I; panel (b): sample covariance matrix $\mathbf{S}$ in Method II; panel (c): PDSCE estimator $\widehat{\mathbf{\Sigma}}_\lambda$ in Method III; panel (d): our proposed covariance matrix estimator $\widehat{\mathbf{\Sigma}}_{\text{recf}}$ in Method IV for one simulated data (with $p = 18$ and $n = 50$).

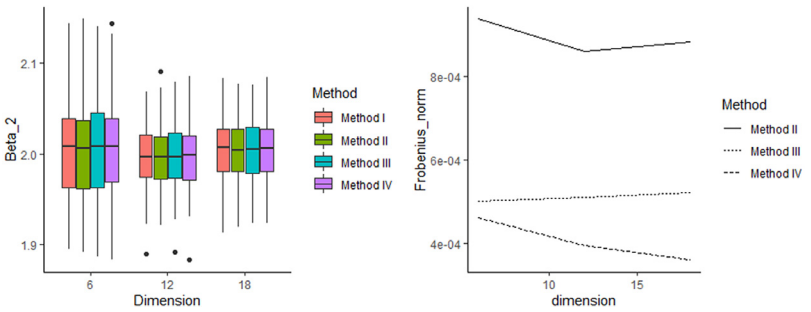


Fig. 3. (Simulation 2) Left panel: boxplots for $\widehat{\beta}_2$ with different methods and different $p = 6, 12, 18$; right panel: average of $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}_0\|_F^2/p^2$ over 100 simulations with different $\widehat{\mathbf{\Sigma}}$ and different $p = 6, 12, 18$.

5.3. Simulation 3

In this simulation we compare the performance of different covariance matrices estimators when data are incomplete and study the influence of the incompleteness on these different approaches. We consider both the "low-rank dense" case and the "high-rank sparse" case in Simulation 1 and 2 when $p = 18$ and $n = 50$. The complete data are generated with the same setup as that in Simulation 1 and 2, with the true covariance function $R_\eta(s, t)$ in (22) and (23) respectively.

We get the incomplete data by the following way: all the 50 subjects are divided into 5 groups. The observation at time point t_k ($k \in \{1, \ldots, 5\}$) for the ith subject is deleted, if the subject i belongs to the kth group, (i.e., $i = 10(k-1) + j,\ j \in \{1, \ldots, 10\}$). Thus we have 17 observations for each subject and a total number of 850 observations

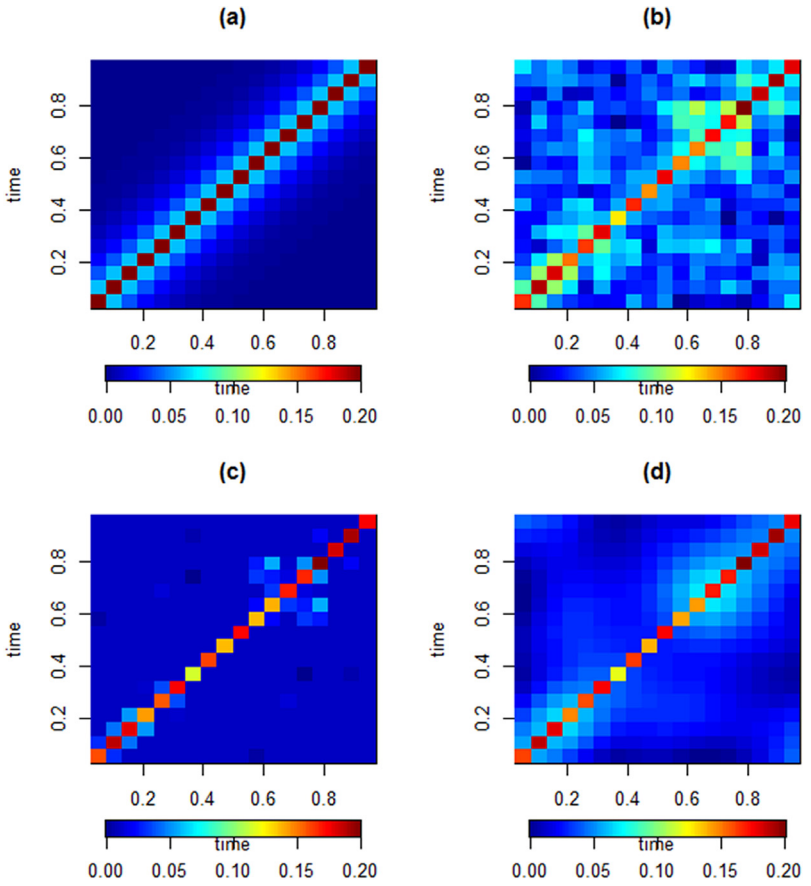


Fig. 4. (Simulation 2) Panel (a): true covariance matrix $\mathbf{\Sigma}_0$ in Method I; panel (b): sample covariance matrix $\mathbf{S}$ in Method II; panel (c): PDSCE estimator $\widehat{\mathbf{\Sigma}}_\lambda$ in Method III; panel (d): our proposed covariance matrix estimator $\widehat{\mathbf{\Sigma}}_{\text{recf}}$ in Method IV for one simulated data (with $p = 18$ and $n = 50$).

Table 3
(Simulation 3) Average squared Frobenius norms over 100 simulations for complete and incomplete data and Relative Loss of Incompleteness in (24) using different Methods II–IV (with $p = 18$ and $n = 50$).

Method	Simulation 1			Simulation 2		
	Complete	Incomplete	Loss	Complete	Incomplete	Loss
Method II	.0164	.0192	17.1%	.0288	.0329	14.2%
Method III	.0191	.0214	12.0%	.0166	.0175	5.4%
Method IV	.0062	.0065	4.8%	.0118	.0122	3.4%

are made. In Method II and Method III, we use the adjusted sample covariance/correlation matrix which based on the pairwise sample covariance/correlation instead of $\mathbf{S}$. It is obvious that when we utilize the pairwise sample covariance to estimate $\mathbf{\Sigma}_0(j,k)$, $1 \le j \ne k \le 5$, only 30 pairs are observed completely, and thus the effective sample size is just 30 for estimating $\mathbf{\Sigma}_0(j,k)$, $1 \le j \ne k \le 5$.

We assess the performance of the different covariance matrices estimators $\widehat{\mathbf{\Sigma}}$ by calculating the average of the squared Frobenius norms $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}_0\|_F^2$ over 100 simulations. When we use the incomplete data, the estimator of covariance matrix will be less precise and the squared Frobenius norm $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}_0\|_F^2$ will be larger than that derived from the complete data. Thus we define the Relative Loss of Incompleteness by

$$\frac{\|\widehat{\mathbf{\Sigma}}^{\text{incom}} - \mathbf{\Sigma}_0\|_F^2 - \|\widehat{\mathbf{\Sigma}}^{\text{com}} - \mathbf{\Sigma}_0\|_F^2}{\|\widehat{\mathbf{\Sigma}}^{\text{com}} - \mathbf{\Sigma}_0\|_F^2}, \tag{24}$$

where $\widehat{\mathbf{\Sigma}}^{\text{com}}$ and $\widehat{\mathbf{\Sigma}}^{\text{incom}}$ are the estimators of covariance matrix based on complete and incomplete data respectively.

Table 3 summarizes the average squared Frobenius norms and the Relative Loss of Incompleteness in (24) for different methods. For the sample covariance matrix (Method II), the relative losses are 17.1% and 14.2% for these two cases respectively, which are much larger than that of other two methods. The PDSCE estimator (Method III) performs quite

Table 4
(Real data case) Comparisons of the estimators and standard errors of β in (25) using different Methods II–IV.

	Method II	Method III	Method IV
$\widehat{\beta}$	1.209	1.735	−0.224
$\widehat{\text{var}}(\widehat{\beta})$	6.166	6.361	7.767

well when the true covariance function is nearly sparse, and the relative loss 5.4% is close to that of our proposed Method IV. For both cases, the relative losses of our proposed Method IV are quite small (4.8% and 3.4% respectively), which shows the advantage of the regularized-covariance-function-based estimator when the data are incomplete.

6. Real data example

We now present an application to the Cattle data study. This dataset has been studied extensively in the literature of longitudinal data analysis, see [11,16]. This dataset comprises the information of an experiment in which cattle were assigned randomly to two treatment groups *A* and *B*, and their body weights were recorded in kilogram. 30 animals received treatment A and another 30 received treatment *B*. The animals were weighted 11 times over a 133-day period: the first 10 measurements for each animal were made at two-week intervals and the last measurement was made one week later. A total of 660 observations were made in this study, with $n = 60$ and $p = 11$. Since no observation is missing, it is considered to be a balanced longitudinal data.

We take the response Y to be the body weights of cattle, Z to be the treatment level (1 for treatment A and 0 for treatment B). The observation time is divided by 140 so that the rescaled observation time t is in between 0 and 1. Now consider a partially linear model

$$Y(t) = \alpha(t) + \beta Z(t) + \eta(t) + \zeta(t). \tag{25}$$

We apply a multifold cross-validation method to select a bandwidth h for $\alpha(t)$. After partitioning the data into 15 groups, we fit model (25) for the data excluding the kth-group, $k \in \{1, \ldots, 15\}$. For the computational issue, we minimize the cross-validation (CV) score on a rough grid $h = 0.5 \times 0.9^b$, $b \in \{1, \ldots, 10\}$. The resulting optimal bandwidth is $h^{\text{opt}} = 0.174$.

Table 4 shows the estimators of β using three different covariance matrix estimators: the sample covariance matrix estimator $\mathbf{S}$ (Method II), the PDSCE estimator $\widehat{\mathbf{\Sigma}}_\lambda$ (Method III) and our proposed estimator $\widehat{\mathbf{\Sigma}}_{\text{recf}}$ (Method IV). The estimator $\widehat{\beta}$ using Method IV is quite different with that using either Method II or Method III. According to the sandwich formula by Fan et al. [3], in Table 4 we also provide the estimated variances $\widehat{\text{var}}(\widehat{\beta})$ of the profile weighted least squares estimators $\widehat{\beta}$ in (9) for different methods. And we conclude that the treatment A is not significantly different from the treatment B for all the three methods.

Since the scales of the estimated covariance matrices are large, and it is hard to distinguish these three covariance matrices estimators from the plots directly, in Fig. 5 we display the estimated correlation matrices for different methods. Fig. 5(a) shows the estimator of correlation matrix using Method II and Fig. 5(b) shows the correlation matrix estimator using Method III. Our proposed estimator of the correlation matrix is displayed in Fig. 5(c). For comparison, we also show the nonparametric estimator of large correlation matrix using a two step estimation procedure proposed by Wu and Pourahmadi [16] in Fig. 5(d). They only studied the 30 cattle with treatment A. From Fig. 5, the estimators of the covariance matrices using Method II and Method III are similar and they are quite different from our proposed Method IV and Wu and Pourahmadi's method.

7. Conclusion

In this article we proposed a nonparametric method to estimate the covariance matrix of the within-subject errors in longitudinal data collected at balanced or nearly balanced time points, which is called the regularized-covariance-function-based estimator. We further derived and compared the error bounds for both the proposed regularized-covariance-function-based estimator and the Positive Definite Sparse Covariance-matrix Estimator (PDSCE) in Frobenius norm. We also investigated the relationship between the rank r, the dimension p, the sparsity s of covariance matrix and the sample size n, so that the upper bound of the convergence rate of the proposed regularized-covariance-function-based estimator is smaller than that of PDSCE estimator in Frobenius norm, which includes the "high-rank sparse" case and "low-rank dense" case when $p \gg n^{1/4} \log n$.

We also assess the performance of these two covariance matrices estimators when the data are incomplete. Both our proposed regularized-covariance-function-based estimator and PDSCE estimator perform quite well in the "high-rank sparse" case, while the Relative Losses of Incompleteness for the PDSCE estimator and the sample covariance matrix are much large than that of our proposed method for the "low-rank dense" case. So the regularized-covariance-function-based estimator is not sensitive to the incompleteness of data.

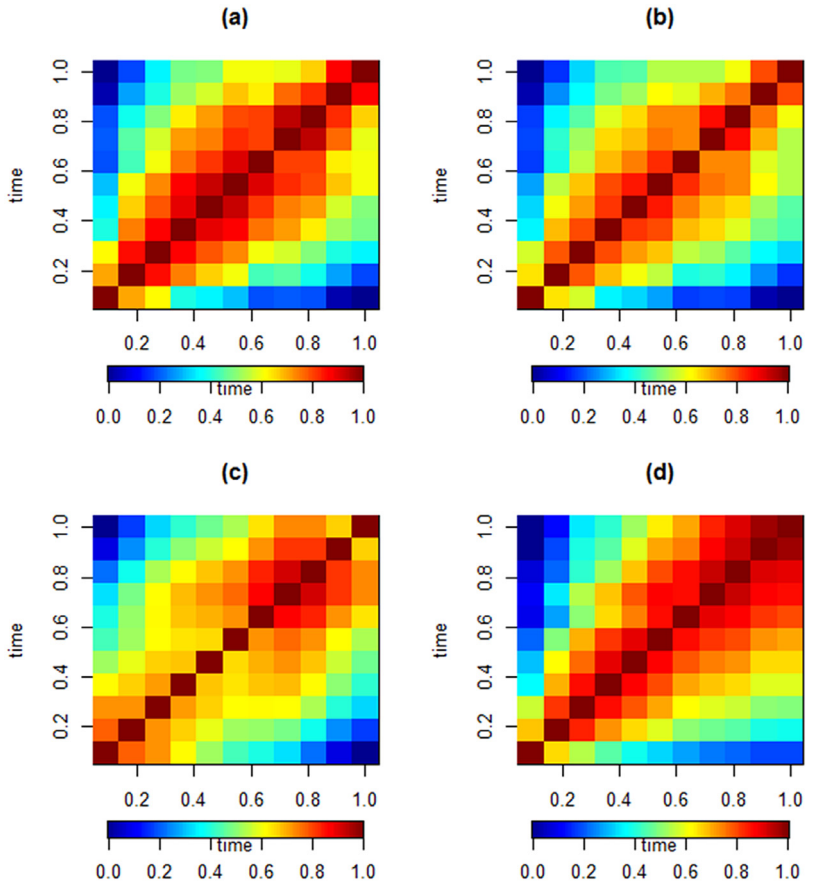


Fig. 5. (Real data case) Panel (a): estimated correlation matrix for Method II; panel (b): estimated correlation matrix for Method III; panel (c): estimated correlation matrix for our proposed Method IV; panel (d): estimated correlation matrix by Wu and Pourahmadi [16].

Several issues are desirable for future research. First, in the presence of outliers, one should consider a robust method to estimate $\boldsymbol{\alpha}(t)$ and $\boldsymbol{\beta}$ instead of using profile weighted least squares. Second, in simulations different covariance structures (22) and (23) are considered, which are classical examples for the "low-rank dense" and "high-rank sparse" cases. The performance of the regularized-covariance-function-based estimator and the PDSCE estimator under these two cases are quite different, thus in practice it is desirable to develop testing procedures for model assumptions (22) and (23) to decide which method should be applied. Finally, in this paper we only compare the error bounds of the regularized-covariance-function-based estimator and the PDSCE estimator in Frobenius norm. Other norms such as the L_2 norm may also be considered. In the proof of Theorems 1 and 2, we have shown the error bound of PDSCE estimator $\widehat{\boldsymbol{\Sigma}}_{\lambda}$ in L_2 norm, but we only derived the error bound of the covariance function estimator $\widehat{R}_{\eta}(\cdot,\cdot)$ in integrated squared norm. Thus it is also of interest to study the convergence rate of $\widehat{\boldsymbol{\Sigma}}_{\text{recf}}$ in L_2 norm. Besides, other high-dimensional covariance matrix estimators with different constraints and structures instead of PDSCE may also be compared. For example, in Simulation 1, instead of using the PDSCE estimator, one may consider the high-dimensional positive definite low-rank covariance matrix estimator; in Simulation 2, one may use the high-dimensional sparse Toeplitz covariance matrix estimator. This topic is beyond the scope of this article and further research is needed.

Acknowledgments

The authors would thank the Editor, the Associate Editor and two referees for their insightful comments and suggestions. The work of C. Zhang was partially supported by the U.S. National Science Foundation grants DMS-2013486 and DMS-1712418, and the University of Wisconsin-Madison Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation, United States.

Appendix

Proof of Theorem 1. Without loss of generality, we assume $t_j = j/p, j \in \{1, \ldots, p\}$. By the definition of $\overline{\Sigma}_{\text{recf}}$ and $R(\cdot,\cdot)$ in (14), we have

$$\begin{aligned}\|\overline{\Sigma}_{\text{recf}} - \Sigma_0\|_F^2 &= \sum_{1\le j\ne k\le p} \{\overline{R}_\eta(t_j,t_k) - R(t_j,t_k)\}^2 + \sum_{j=1}^p \{s_j^2 - R(t_j,t_j)\}^2 \\ &= \sum_{1\le j\ne k\le p} \{\overline{R}_\eta(t_j,t_k) - R_\eta(t_j,t_k)\}^2 + \sum_{j=1}^p \{s_j^2 - \sigma^2(t_j)\}^2 \\ &\le \sum_{j,k=1}^p \{\overline{R}_\eta(t_j,t_k) - R_\eta(t_j,t_k)\}^2 + \sum_{j=1}^p \{s_j^2 - \sigma^2(t_j)\}^2, \end{aligned} \tag{26}$$

where $\overline{R}_\eta(\cdot,\cdot)$ is defined in [7]. According the definition of integral, and an application of [1] (Theorem 4), as $p \to \infty$,

$$\frac{1}{p^2}\sum_{j,k=1}^p \{\overline{R}_\eta(t_j,t_k) - R_\eta(t_j,t_k)\}^2 \to \|\overline{R}_\eta - R_\eta\|_{L_2}^2 = O_P\big(\{\log(n)/(np)\}^{4/5} + n^{-1}\big), \tag{27}$$

where $\|f(\cdot,\cdot)\|_{L_2} = \left\{\iint_{s,t\in\mathcal{T}} f^2(s,t)dsdt\right\}^{1/2}$ denotes the integrated squared norm of a bivariate function $f(\cdot,\cdot)$. And an application of CLT shows that

$$\sum_{j=1}^p \{s_j^2 - \sigma^2(t_j)\}^2 = O_P\Big(\frac{p}{n}\Big). \tag{28}$$

(26) combined with (27), (28) shows that

$$\|\overline{\Sigma}_{\text{recf}} - \Sigma_0\|_F^2 = O_P\Big(p^2\Big[\Big\{\frac{\log(n)}{np}\Big\}^{4/5} + \frac{1}{n}\Big] + \frac{p}{n}\Big) = O_P\Big(\Big\{\frac{p^{3/2}\log(n)}{n}\Big\}^{4/5} + \frac{p^2}{n}\Big),$$

which completes the proof. □

Proof of Theorem 2. By the definition of Frobenius norm and spectral norm for a general $p \times p$ matrix $\mathbf{M}$,

$$\|\mathbf{M}\|_F = \sqrt{\text{trace}(\mathbf{M}^\top\mathbf{M})}, \quad \|\mathbf{M}\| = \sqrt{\lambda_{\max}(\mathbf{M}^\top\mathbf{M})}.$$

Let $\lambda_{\max}(\mathbf{M}^\top\mathbf{M}) = \lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_r > \lambda_{r+1} = \cdots = \lambda_p = \lambda_{\min}(\mathbf{M}^\top\mathbf{M}) = 0$ be the decreasing eigenvalues of $\mathbf{M}^\top\mathbf{M}$, where $r = \text{rank}(\mathbf{M}^\top\mathbf{M})$, then

$$\|\mathbf{M}\|_F^2 = \text{trace}(\mathbf{M}^\top\mathbf{M}) = \sum_{i=1}^p \lambda_i = \lambda_1 + \cdots + \lambda_r \le r\lambda_{\max}(\mathbf{M}^\top\mathbf{M}) = r\|\mathbf{M}\|^2. \tag{29}$$

According to Theorem 1 of [12], we have

$$\|\overline{\Sigma}_\lambda - \Sigma_0\|^2 = O_p\Big\{\frac{(s+1)\log(p)}{n}\Big\}. \tag{30}$$

Take $\mathbf{M} = \overline{\Sigma}_\lambda - \Sigma_0$, (30) together with (29) completes the proof. □

Proof of Theorem 3. According to Corollary 1, there are two different situations:

(i) If $\theta > 1/4$, then $p \gg n^{1/4}\log(n)$ and the upper bound of the convergence rate for the regularized-covariance-function-based estimator will be smaller than that for the Positive Definite Sparse Covariance-matrix Estimator (PDSCE) when

$$\frac{p^2}{n} = o\Big\{\frac{r(s+1)\log(p)}{n}\Big\}, \quad \text{or} \quad n^{2\theta-1} = o\big(n^{\theta\gamma-1}\log(n)\big),$$

that means $2\theta - 1 \le \theta\gamma - 1$ or $2 \le \gamma$.

(ii) If $\theta \le 1/4$, then $p \ll n^{1/4}\log(n)$ and the upper bound of the convergence rate for the regularized-covariance-function-based estimator will be smaller than that for the Positive Definite Sparse Covariance-matrix Estimator (PDSCE) when

$$\left\{\frac{p^{3/2}\log(n)}{n}\right\}^{4/5} = o\left\{\frac{r(s+1)\log(p)}{n}\right\}, \quad \text{or} \quad n^{(6\theta-4)/5}\{\log(n)\}^{4/5} = o\big(n^{\theta\gamma-1}\log(n)\big),$$

than means $(6\theta-4)/5 \le \theta\gamma - 1$ or $\gamma \ge (6\theta+1)/(5\theta)$. □

References

[1] T. Cai, M. Yuan, Nonparametric Covariance Function Estimation for Functional and Longitudinal Data, Technical Report, 2010.
[2] P.J. Diggle, P. Heagerty, K.Y. Liang, S.L. Zeger, Analysis of Longitudinal Data, second ed., Oxford University Press, New York, 2002.
[3] J. Fan, T. Huang, R. Li, Analysis of longitudinal data with semiparametric estimation of covariance function, J. Amer. Statist. Assoc. 102 (2007) 632–641.
[4] J. Fan, Y. Wu, Semiparametric estimation of covariance matrixes for longitudinal data, J. Amer. Statist. Assoc. 103 (2008) 1520–1533.
[5] P. Hall, H.G. Müller, J.L. Wang, Properties of principal component methods for functional and longitudinal data analysis, Ann. Statist. 34 (2006) 1493–1517.
[6] P. Hall, H.G. Müller, F. Yao, Modelling sparse generalized longitudinal observations with latent Gaussian processes, J. R. Stat. Soc. Ser. B Stat. Methodol. 70 (2008) 703–723.
[7] S. Jia, C. Zhang, H. Wu, Efficient semiparametric regression for longitudinal data with regularised estimation of error covariance function, J. Nonparametr. Stat. 31 (2019) 867–886.
[8] Y. Li, Efficient semiparametric regression for longitudinal data with nonparametric covariance estimation, Biometrika 98 (2011) 355–370.
[9] X. Lin, R.J. Carroll, Semiparametric regression for clustered data using generalized estimating equations, J. Amer. Statist. Assoc. 96 (2001) 1045–1056.
[10] A. Milton, A.S. Irene, Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, New York, 1972.
[11] J. Pan, G. MacKenzie, On modelling mean-covariance structures in longitidinal studies, Biometrika 90 (2003) 239–244.
[12] A.J. Rothman, Positive definite estimators of large covariance matrices, Biometrika 99 (2012) 733–740.
[13] A.J. Rothman, P.J. Bickel, E. Levina, J. Zhu, Sparse permutation invariant covariance estimation, Electron. J. Stat. 2 (2008) 494–515.
[14] A.J. Rothman, E. Levina, J. Zhu, Generalized thresholding of large covariance matrices, J. Amer. Statist. Assoc. 104 (2009) 177–186.
[15] N. Wang, R.J. Carroll, X. Lin, Efficient semiparametric marginal estimation for longitudinal/clustered data, J. Am. Statist. Assoc. 100 (2005) 147–157.
[16] W.B. Wu, M. Pourahmadi, Nonparametric estimation of large covariance matrices of longitudinal data, Biometrika 90 (2003) 831–844.
[17] H.L. Wu, J.T. Zhang, Local polynomial mixed-effects models for longitudinal data, J. Am. Statist. Assoc. 97 (2002) 883–897.
[18] H.L. Wu, J.T. Zhang, The study of long-term HIV dynamics using semiparametric nonlinear mixed-effects models, Stat. Med. 21 (2002) 3655–3675.
[19] H.L. Wu, J.T. Zhang, Nonparametric Regression Methods for Longitudinal Data: Mixed-Effects Modeling Approaches, Wiley, Hoboken, 2006.
[20] F. Yao, H.G. Müller, J.L. Wang, Functional data analysis for sparse longitudinal data, J. Amer. Statist. Assoc. 100 (2005) 577–590.