UNIVERSITY *of York*

Discussion Papers in Economics

No. 15/17

New Semiparametric Estimation Procedure for Functional Coefficient Longitudinal Data Models

Jia Chen, Degui Li and Yingcun Xia

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

New Semiparametric Estimation Procedure for Functional Coefficient Longitudinal Data Models

Jia Chen* The University of York

Degui Li† The University of York

Yingcun Xia‡ National University of Singapore

July 30, 2015

Abstract

In order to achieve dimension reduction for the nonparametric functional coefficients and improve the estimation efficiency, in this paper we introduce a novel semiparametric estimation procedure which combines a principal component analysis of the functional coefficients and a Cholesky decomposition of the within-subject covariance matrices. Under some regularity conditions, we derive the asymptotic distribution for the proposed semiparametric estimators and show that the efficiency of the estimation of the (principal) functional coefficients can be improved when the within-subject covariance structure is correctly specified. Furthermore, we apply two approaches to consistently estimate the Cholesky decomposition, which avoid a possible misspecification of the within-subject covariance structure and ensure the efficiency improvement for the estimation of the (principal) functional coefficients. Some numerical studies including Monte Carlo experiments and an empirical application show that the developed semiparametric method works reasonably well in finite samples.

Keywords: Cholesky decomposition, functional coefficients, local linear smoothing, principal component analysis, profile least squares, within-subject covariance.

JEL Classifications: C14, C23, C51.

*Department of Economics and Related Studies, University of York, Heslington, YO10 5DD, UK.

†Department of Mathematics, University of York, Heslington, YO10 5DD, UK.

‡Department of Statistics and Applied Probability, National University of Singapore, 117546, Singapore.

1 Introduction

Nonparametric estimation of longitudinal data models has received increasing attention in the past decade, see, for example, Lin and Carroll (2000), Lin and Ying (2001), Wang (2003), Wu and Zhang (2006) and Yao and Li (2013). When the dimension of the covariates in the nonparametric component is larger than three, to circumvent the well-known "curse of dimensionality", some nonparametric and semiparametric modelling techniques, such as functional coefficient models, additive models, partially linear models and single-index models, have been extensively studied in the literature (Hoover *et al*, 1998; Fan and Li, 2004; Fan *et al*, 2007; Zhang *et al*, 2009; Jiang and Wang, 2011; Chen *et al*, 2013; Wang *et al*, 2014). In this paper, we study the functional coefficient longitudinal data models as they are a natural generalisation of the classical linear regression models and provide a flexible framework to describe the relationship between response and covariates. A detailed introduction on estimation and inference of functional coefficient models in the context of independent or weakly dependent data can be found in Chen and Tsay (1993), Fan and Zhang (1999), Xia *et al* (2004), Li and Liang (2008), Wang and Xia (2009), Zhou and Liang (2009), Kai *et al* (2011), Jiang *et al* (2013) and the references therein. In this paper, we propose a novel semiparametric method for longitudinal data that could result in a more efficient estimation of the functional coefficient models.

Consider a set of longitudinal data $\big(Y_{ij}, \mathbf{X}_{ij}, U_{ij}\big)$, $i = 1, \ldots, n$, $j = 1, \ldots, m_i$, where Y_{ij} is the response variable of interest, $\mathbf{X}_{ij}$ is a d-dimensional vector of random covariates and U_{ij} is a univariate random covariate. The variable U_{ij} can be chosen as a calendar time or some other *index* variable in practical applications. The functional coefficient longitudinal data model takes the following form

$$Y_{ij} = \mathbf{X}_{ij}^{\tau}\boldsymbol{\beta}(U_{ij}) + \varepsilon_{ij}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, m_i, \tag{1.1}$$

where $\boldsymbol{\beta}(\cdot)$ is a d-dimensional vector of functional coefficients, ε_{ij} is the random error term satisfying

$$\mathsf{E}\left[\varepsilon_{ij}|\mathbf{X}_{ij}, U_{ij}\right] = 0 \quad \text{and} \quad \mathsf{Cov}\left[\varepsilon_i|\mathbf{X}_i, U_i\right] = \boldsymbol{\Sigma}_i \tag{1.2}$$

almost surely ($a.s.$) in which $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{im_i})^{\tau}$, $\mathbf{X}_i = (\mathbf{X}_{i1}, \ldots, \mathbf{X}_{im_i})^{\tau}$ and $U_i = (U_{i1}, \ldots, U_{im_i})^{\tau}$. Throughout the paper, we assume that the number of the subjects n is large, but the number of the observations for each subject m_i is fixed.

The main interest of the paper is to estimate the functional coefficients $\boldsymbol{\beta}(\cdot)$. However, when the dimension of the functional coefficients, d, is large or even moderately large, the resulting nonparametric estimation of the coefficient functions would be very unstable. In this paper, we aim to address this challenge through the principal components of the functional coefficients. This approach is partly motivated by a recent paper by Jiang *et al* (2013). Suppose that there exist a vector of *principal functional coefficients* $\boldsymbol{\gamma}(u) = \left[\gamma_1(u), \ldots, \gamma_{d_0}(u)\right]^{\tau}$, a d-dimensional vector of parameters $\boldsymbol{\theta}$ and a $d \times d_0$ matrix of parameters $\boldsymbol{\Theta}$ such that

$$\boldsymbol{\beta}(u) = \boldsymbol{\theta} + \boldsymbol{\Theta}\boldsymbol{\gamma}(u) \tag{1.3}$$

in which $\boldsymbol{\Theta} = \left[\boldsymbol{\Theta}(1), \boldsymbol{\Theta}(2), \ldots, \boldsymbol{\Theta}(d_0)\right]$ with $\boldsymbol{\Theta}(k)$ being a d-dimensional column vector of constants. The positive integer d_0 is usually unknown in practice but typically much smaller than d. This integer will be determined later in the paper by using the ratio criterion (Lam and Yao, 2012). To avoid confusion, throughout the paper, we let $\boldsymbol{\theta}_0$, $\boldsymbol{\Theta}_0 = \left[\boldsymbol{\Theta}_0(1), \boldsymbol{\Theta}_0(2), \ldots, \boldsymbol{\Theta}_0(d_0)\right]$ and $\boldsymbol{\gamma}_0(\cdot) = \left[\gamma_{10}(\cdot), \ldots, \gamma_{d_0 0}(\cdot)\right]^{\tau}$ and $\boldsymbol{\beta}_0(\cdot)$ be the true values of $\boldsymbol{\theta}$, $\boldsymbol{\Theta}$, $\boldsymbol{\gamma}(\cdot)$ and $\boldsymbol{\beta}(\cdot)$, respectively.

The conditions in (1.2) allow the longitudinal data to be correlated within each subject. To account for the within-subject correlation, some modified nonparametric and semiparametric methods have been introduced in the literature through certain functional transformation or nonparametric (or semiparametric) estimation of the within-subject covariance matrices $\boldsymbol{\Sigma}_i$, see, for example, Linton *et al* (2003), Wang (2003), Fan *et al* (2007), Li (2011), Yao and Li (2013), Honda *et al* (2014). A direct application of the estimation procedure proposed in Jiang *et al* (2013) by ignoring the within-subject correlation would certainly affect the efficiency of the functional coefficients estimation. To address this problem, we first use, as in Yao and Li (2013), a Cholesky decomposition on the within-subject covariance matrices $\boldsymbol{\Sigma}_i$, and then propose a semiparametric estimation method to estimate $\boldsymbol{\theta}_0$, $\boldsymbol{\Theta}_0$ and $\boldsymbol{\gamma}_0(\cdot)$. Under some regularity conditions, we establish the asymptotic distribution theory for the proposed semiparametric estimators. In particular, we show that the combination of a principal component analysis of the functional coefficients and a Cholesky decomposition of the within-subject covariance matrices could improve the efficiency of the estimation of the principal functional coefficients and thus the model functional coefficients when the within-subject covariance structure is correctly specified up to a constant multiple. However, the true within-subject covariance structure is usually unknown in practice and misspecification of the covariance matrix could lead to loss of estimation efficiency. Hence we further introduce two different approaches to consistently estimate the parameters in the Cholesky decomposition for cases of balanced and unbalanced longitudinal data, respectively. By using these consistent estimation methods, we can avoid a possible misspecification of the within-subject covariance structure and thus ensure the efficiency improvement for the estimation of $\boldsymbol{\gamma}_0(\cdot)$ and $\boldsymbol{\beta}_0(\cdot)$. We also provide some numerical studies in Section 5, which show that the developed semiparametric approach works reasonably well in finite samples.

The rest of the paper is organised as follows. In Section 2, we introduce an identification condition and the semiparametric estimation procedure for model (1.1). In Section 3, we give the asymptotic theorems for the proposed estimators. In Section 4, we consider the estimation of the autoregressive coefficients in the Cholesky decomposition. In Section 5, we conduct some numerical studies to illustrate the finite sample performance of the proposed methods. Section 6 concludes the paper. The technical assumptions and proofs of the asymptotic theorems are given in the appendix.

2 Model identification and estimation

In this section, we first give some identification conditions to ensure that $\mathbf{\Theta}_0$ and $\boldsymbol{\gamma}_0(\cdot)$ are uniquely determined, followed by a semiparametric estimation procedure which ignores the within-subject correlation, then introduce a Cholesky decomposition to analyse the within-subject covariance structure, and finally propose a semiparametric method to estimate the unknown components which accounts for the within-subject correlation. To focus on the idea behind the semiparametric method, we assume in this section that the number of the principal functional coefficients (d_0) is known. We will discuss a method to estimate d_0 in Section 5.

2.1. Model identification

It is easy to see that, after reparameterization and rotation to $\boldsymbol{\gamma}_0(.)$ in (1.3), the principal functional coefficients $\boldsymbol{\gamma}_0(.)$ can satisfy

$$\mathsf{E}[\boldsymbol{\gamma}_0(U_{ij})] = \mathbf{0}_{d_0}, \text{ and } \mathsf{Cov}[\boldsymbol{\gamma}_0(U_{ij})] = \mathsf{diag}(\lambda_1, ..., \lambda_{d_0}), \tag{2.1}$$

where $\mathbf{0}_k$ is a k-dimensional null vector, and $0 \leq d_0 \leq d$. If we further assume that

$$\lambda_1 > \ldots > \lambda_{d_0} > 0,$$

then $\mathbf{\Theta}$ and $\boldsymbol{\gamma}_0(u)$ are identifiable up to possible sign difference. Details can be found in Jiang *et al* (2013). Körber *et al* (2015) also use a condition similar to (2.1) to identify their heterogeneous nonparametric panel data models with a univariate regressor.

2.2. Semiparametric estimation procedure which ignores the within-subject correlation

We next introduce a semiparametric profile least squares method that does not account for the within-subject correlation in longitudinal data to estimate $\boldsymbol{\theta}_0$, $\mathbf{\Theta}_0$ and the principal functional coefficients $\boldsymbol{\gamma}_0(\cdot)$. Combining (1.1) and (1.3), we may write the principal functional coefficient longitudinal data model as

$$Y_{ij} = \mathbf{X}_{ij}^{\tau}\boldsymbol{\theta}_0 + \sum_{k=1}^{d_0} \gamma_{k0}(U_{ij})\mathbf{X}_{ij}^{\tau}\boldsymbol{\Theta}_0(k) + \varepsilon_{ij}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, m_i. \tag{2.2}$$

For given $\boldsymbol{\theta}$ and $\boldsymbol{\Theta}$, we estimate the principal functional coefficients at a given point u by using the local linear smoothing method (Fan and Gijbels, 1996). Define the kernel-weighted loss function

$$\begin{aligned} L_n\big(\mathbf{a}(u), \mathbf{b}(u) \big| \boldsymbol{\theta}, \boldsymbol{\Theta}\big) &= \sum_{i=1}^{n}\sum_{j=1}^{m_i}\Bigg[Y_{ij} - \mathbf{X}_{ij}^{\tau}\boldsymbol{\theta} - \sum_{k=1}^{d_0} a_k(u)\mathbf{X}_{ij}^{\tau}\boldsymbol{\Theta}(k) \\ &\quad - \sum_{k=1}^{d_0} b_k(u)(U_{ij} - u)\mathbf{X}_{ij}^{\tau}\boldsymbol{\Theta}(k)\Bigg]^2 K\left(\frac{U_{ij} - u}{h}\right), \end{aligned} \tag{2.3}$$

where $\mathbf{a}(u) = \big[a_1(u), \ldots, a_{d_0}(u)\big]^{\tau}$ and $\mathbf{b}(u) = \big[b_1(u), \ldots, b_{d_0}(u)\big]^{\tau}$, $K(\cdot)$ is a kernel function and h is a bandwidth. Let $\widehat{\mathbf{a}}(u) = \big[\widehat{a}_1(u), \ldots, \widehat{a}_{d_0}(u)\big]^{\tau}$ and $\widehat{\mathbf{b}}(u) = \big[\widehat{b}_1(u), \ldots, \widehat{b}_{d_0}(u)\big]^{\tau}$ be the solution which minimises the loss function defined in (2.3). Then, the local linear estimate of the principal functional coefficients for given $\boldsymbol{\theta}$ and $\boldsymbol{\Theta}$ can be obtained by

$$\widehat{\boldsymbol{\gamma}}(u|\boldsymbol{\theta}, \boldsymbol{\Theta}) = \big[\widehat{\gamma}_1(u|\boldsymbol{\theta}, \boldsymbol{\Theta}), \ldots, \widehat{\gamma}_{d_0}(u|\boldsymbol{\theta}, \boldsymbol{\Theta})\big]^{\tau} \quad \text{with} \quad \widehat{\gamma}_k(u|\boldsymbol{\theta}, \boldsymbol{\Theta}) = \widehat{a}_k(u). \tag{2.4}$$

Replacing $\gamma_{k0}(U_{ij})$ in model (2.2) by

$$\widetilde{\gamma}_k(U_{ij}|\boldsymbol{\theta}, \boldsymbol{\Theta}) = \widehat{\gamma}_k(U_{ij}|\boldsymbol{\theta}, \boldsymbol{\Theta}) - \frac{1}{N(n)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\widehat{\gamma}_k(U_{ij}|\boldsymbol{\theta}, \boldsymbol{\Theta})$$

where $N(n) = \sum_{i=1}^{n} m_i$, we can estimate $\boldsymbol{\theta}_0$ and $\boldsymbol{\Theta}_0$ through minimising the loss function

$$Q_n(\boldsymbol{\theta}, \boldsymbol{\Theta}) = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\Bigg[Y_{ij} - \mathbf{X}_{ij}^{\tau}\boldsymbol{\theta} - \sum_{k=1}^{d_0}\widetilde{\gamma}_k(U_{ij}|\boldsymbol{\theta}, \boldsymbol{\Theta})\mathbf{X}_{ij}^{\tau}\boldsymbol{\Theta}(k)\Bigg]^2, \tag{2.5}$$

and the resulting estimates are denoted as $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\Theta}}$, respectively. Then we obtain a local linear estimate of the principal functional coefficients $\boldsymbol{\gamma}_0(u)$ by

$$\widehat{\boldsymbol{\gamma}}(u) \equiv \widehat{\boldsymbol{\gamma}}(u|\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Theta}}). \tag{2.6}$$

Subsequently, an estimate for the model functional coefficients is

$$\widehat{\boldsymbol{\beta}}(u) = \widehat{\boldsymbol{\theta}} + \widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\gamma}}(u). \tag{2.7}$$

The above semiparametric profile least squares estimation method can be seen as a generalisation of Jiang *et al* (2013) to the setting of longitudinal data. However, this estimation method for the model functional coefficients $\boldsymbol{\beta}(u)$ does not consider possible within-subject correlation structure in the longitudinal data, which may lead to an efficiency loss. To address this problem, we will modify the semiparametric estimation method in Section 2.4 below by using the Cholesky decomposition of the within-subject covariance matrices.

2.3. Cholesky decomposition

The Cholesky decomposition has been widely used in the literature to analyse the within-subject covariance matrices, see, for example, Pourahmadi (1999), Leng *et al* (2010), Yao and Li (2013), Liu and Li (2015), and Zhang *et al* (2015). In particular, Yao and Li (2013) apply such a decomposition to improve the nonparametric estimation efficiency in longitudinal data models with a univariate covariate. Next, we consider a more general model setting with multivariate covariates and a dimension reduction on the functional coefficients via a principal component analysis in (1.3).

For each within-subject covariance matrix $\mathbf{\Sigma}_i$, by the Cholesky decomposition, there exists a lower triangular matrix $\mathbf{C}_i$ with diagonal elements being one such that

$$\mathbf{C}_i\mathbf{\Sigma}_i\mathbf{C}_i^\tau = \mathbf{\Delta}_i = \mathsf{diag}\left\{\rho_{i1}, \cdots, \rho_{im_i}\right\}, \tag{2.8}$$

where $\rho_{ij} > 0$ for $i = 1, \ldots, n$ and $j = 1, \ldots, m_i$. Let $\eta_i = (\eta_{i1}, \ldots, \eta_{im_i})^\tau = \mathbf{C}_i\varepsilon_i$ and $c_{i,jk}$ be the (j, k)-th entry of the minus of $\mathbf{C}_i$, i.e., $-\mathbf{C}_i$. For each $i = 1, \ldots, n$, it is easy to see that $\varepsilon_{i1} = \eta_{i1}$ and

$$\varepsilon_{ij} = \eta_{ij} + \sum_{k=1}^{j-1} c_{i,jk}\varepsilon_{ik}, \quad j = 2, \ldots, m_i. \tag{2.9}$$

Throughout this paper, we call $c_{i,jk}$ *the autoregressive coefficients in the Cholesky decomposition*, as (2.9) can be seen as an autoregressive model. Using (2.2) and (2.9), we can re-write model (1.1) as

$$Y_{i1} = \mathbf{X}_{i1}^\tau\boldsymbol{\theta}_0 + \sum_{k=1}^{d_0} \gamma_{k0}(U_{i1})\mathbf{X}_{i1}^\tau\mathbf{\Theta}_0(k) + \eta_{i1} \tag{2.10}$$

and

$$Y_{ij} = \mathbf{X}_{ij}^\tau\boldsymbol{\theta}_0 + \sum_{k=1}^{d_0} \gamma_{k0}(U_{ij})\mathbf{X}_{ij}^\tau\mathbf{\Theta}_0(k) + \sum_{k=1}^{j-1} c_{i,jk}\varepsilon_{ik} + \eta_{ij}, \quad j = 2, \ldots, m_i, \tag{2.11}$$

for $i = 1, \ldots, n$. By (2.8), the transformed errors η_{ij} in (2.10) and (2.11) are independent over both i and j and have variance $\mathsf{Var}(\eta_{ij}) = \rho_{ij}$. However, in (2.11) both the parameter matrices $\mathbf{C}_i$ and the random error vectors ε_i are unobserved. Hence, in practical applications, we need to replace them by their estimated values.

As suggested in Fan *et al* (2007) and Yao and Li (2013), we may replace $\mathbf{\Sigma}_i$ by a working covariance matrix $\mathbf{\Sigma}_i^\diamond$. Such a replacement in the estimation procedure would not affect the consistency of the resulting estimator even if $\mathbf{\Sigma}_i^\diamond \neq \mathbf{\Sigma}_i$. Then, we apply the Cholesky decomposition to the working covariance matrix $\mathbf{\Sigma}_i^\diamond$ and find a lower triangular matrix $\mathbf{C}_i^\diamond$ with main diagonal elements being one and positive constants $\rho_{i1}^\diamond, \ldots, \rho_{im_i}^\diamond$ such that

$$\mathbf{C}_i^\diamond \mathbf{\Sigma}_i^\diamond (\mathbf{C}_i^\diamond)^\tau = \mathbf{\Delta}_i^\diamond = \mathsf{diag}\left\{\rho_{i1}^\diamond, \ldots, \rho_{im_i}^\diamond\right\}, \quad i = 1, \ldots, n. \tag{2.12}$$

Let $\widetilde{\varepsilon}_i = (\widetilde{\varepsilon}_{i1}, \ldots, \widetilde{\varepsilon}_{im_i})^\tau$ with $\widetilde{\varepsilon}_{ij} = Y_{ij} - \mathbf{X}_{ij}^\tau \widetilde{\boldsymbol{\beta}}(U_{ij})$, where $\widetilde{\boldsymbol{\beta}}(\cdot)$ is a local linear estimation of $\boldsymbol{\beta}_0(\cdot)$ by using a kernel function $K(\cdot)$ and a bandwidth b. Denote

$$\widetilde{Y}_{i1} = Y_{i1}, \quad \widetilde{Y}_{ij} = Y_{ij} - \sum_{k=1}^{j-1} c_{i,jk}^\diamond \widetilde{\varepsilon}_{ik}, \quad j = 2, \ldots, m_i, \tag{2.13}$$

where $c_{i,jk}^\diamond$ is the (j,k) component of $-\mathbf{C}_i^\diamond$. Therefore, we can further approximate (2.10) and (2.11) by

$$\widetilde{Y}_{ij} \approx \mathbf{X}_{ij}^\tau \boldsymbol{\theta}_0 + \sum_{k=1}^{d_0} \gamma_{k0}(U_{ij}) \mathbf{X}_{ij}^\tau \mathbf{\Theta}_0(k) + \eta_{ij}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, m_i. \tag{2.14}$$

The estimation methods introduced in Section 2.4 are then applied to the model approximation (2.14). Furthermore, Theorem 3.2 in Section 3 will show that the efficiency improvement of the developed nonparametric estimation in (2.15) and (2.16) below relies on a correct specification of the lower triangular matrix $\mathbf{C}_i$. To avoid the misspecification of $\mathbf{C}_i$, we will discuss in Section 4 how to consistently estimate the elements in $\mathbf{C}_i$ for both balanced and unbalanced longitudinal data.

2.4. Semiparametric estimation procedure which accounts for the within-subject correlation

To improve the efficiency in the estimation of the principal functional coefficients and the model functional coefficients proposed in Section 2.2, we next make use of the Cholesky decomposition on the working covariance matrices $\mathbf{\Sigma}_i^\diamond$ and the subsequent transformation in (2.13). Note that after the transformation the error terms η_{ij} in (2.14) are independent across both i and j and have variances $\mathsf{Var}(\eta_{ij}) = \rho_{ij}^\diamond$. Hence, for given root-$n$ consistent estimates $\overline{\boldsymbol{\theta}}$ and $\overline{\mathbf{\Theta}}$ of $\boldsymbol{\theta}_0$ and $\mathbf{\Theta}_0$, define the loss function

$$\begin{aligned} \overline{L}_n\big(\mathbf{a}(u), \mathbf{b}(u)\big) \quad = \quad & \sum_{i=1}^n \sum_{j=1}^{m_i} \Bigg[\widetilde{Y}_{ij} - \mathbf{X}_{ij}^\tau \overline{\boldsymbol{\theta}} - \sum_{k=1}^{d_0} a_k(u) \mathbf{X}_{ij}^\tau \overline{\mathbf{\Theta}}(k) \\ & - \sum_{k=1}^{d_0} b_k(u)(U_{ij} - u) \mathbf{X}_{ij}^\tau \overline{\mathbf{\Theta}}(k) \Bigg]^2 \left(\rho_{ij}^\diamond\right)^{-1} K\left(\frac{U_{ij} - u}{h}\right), \end{aligned} \tag{2.15}$$

where $\overline{\mathbf{\Theta}}(k)$ is the k-th column of $\overline{\mathbf{\Theta}}$. Let $\overline{\mathbf{a}}(u) = \left[\overline{a}_1(u), \ldots, \overline{a}_{d_0}(u)\right]^\tau$ and $\overline{\mathbf{b}}(u) = \left[\overline{b}_1(u), \ldots, \overline{b}_{d_0}(u)\right]^\tau$ be the solution to the minimisation of $\overline{L}_n\big(\mathbf{a}(u), \mathbf{b}(u)\big)$ in (2.15). Then, a modified local linear estimate of the principal functional coefficients is obtained by

$$\overline{\boldsymbol{\gamma}}(u) \equiv \overline{\boldsymbol{\gamma}}(u|\overline{\boldsymbol{\theta}}, \overline{\mathbf{\Theta}}) = \left[\overline{\gamma}_1(u), \ldots, \overline{\gamma}_{d_0}(u)\right]^\tau \quad \text{with} \quad \overline{\gamma}_k(u) = \overline{a}_k(u), \tag{2.16}$$

and a subsequent estimate of the functional coefficients is given by

$$\overline{\boldsymbol{\beta}}(u) = \overline{\boldsymbol{\theta}} + \overline{\mathbf{\Theta}}\overline{\boldsymbol{\gamma}}(u). \tag{2.17}$$

Proposition 3.1 indicates that $\widehat{\boldsymbol{\theta}}$ and $\widehat{\mathbf{\Theta}}$ constructed in Section 2.2 are consistent with a root-n convergence rate. Hence, we may choose $\overline{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}$ and $\overline{\mathbf{\Theta}} = \widehat{\mathbf{\Theta}}$ in the above estimation procedure. Although the parametric estimators $\widehat{\boldsymbol{\theta}}$ and $\widehat{\mathbf{\Theta}}$ ignore the possible within-subject correlation in the longitudinal data, it would not affect the asymptotic efficiency of the functional coefficients estimation $\overline{\boldsymbol{\beta}}(\cdot)$ as the convergence rates for the parameter estimators are much faster than the point-wise convergence rates of the nonparametric estimators, see, for example, the asymptotic theorems in Section 3. The semiparametric estimation procedure proposed in this section and that in Section 2.2 provide a feasible approach to estimating the parameters and the principal coefficient functions. An appropriate choice of the initial estimates of $\boldsymbol{\theta}_0$ and $\mathbf{\Theta}_0$ may help save computational time and improve estimation accuracy in finite samples. Section 5.1 below will discuss how to obtain a consistent initial semiparametric estimation.

3 Asymptotic theorems

In this section, we establish the asymptotic properties for the semiparametric estimators defined in Section 2. Define

$$\begin{aligned}
&\mathbf{X}_{ij,k}(\mathbf{\Theta}) = \mathbf{X}_{ij}^\tau \mathbf{\Theta}(k), \quad \mathbf{X}_{ij}(\mathbf{\Theta}) = \left[\mathbf{X}_{ij,1}(\mathbf{\Theta}), \ldots, \mathbf{X}_{ij,d_0}(\mathbf{\Theta})\right]^\tau, \\
&\widetilde{\mathbf{X}}_{ij}(\mathbf{\Theta}) = \mathbf{X}_{ij} - \widetilde{\mathbf{\Delta}}_{\mathbf{X}}(U_{ij}|\mathbf{\Theta})\mathbf{X}_{ij}(\mathbf{\Theta}), \quad \widetilde{\mathbf{X}}_{ij} \equiv \widetilde{\mathbf{X}}_{ij}(\mathbf{\Theta}_0), \\
&\widetilde{\mathbf{\Delta}}_{\mathbf{X}}(U_{ij}|\mathbf{\Theta}) = \mathbf{\Delta}_{\mathbf{X}}(U_{ij}|\mathbf{\Theta})\mathbf{\Delta}^+(U_{ij}|\mathbf{\Theta}) - \mathsf{E}\left[\mathbf{\Delta}_{\mathbf{X}}(U_{ij}|\mathbf{\Theta})\mathbf{\Delta}^+(U_{ij}|\mathbf{\Theta})\right],
\end{aligned}$$

where $\mathbf{\Delta}(u|\mathbf{\Theta}) = \mathsf{E}\big[\mathbf{X}_{ij}(\mathbf{\Theta})\mathbf{X}_{ij}^\tau(\mathbf{\Theta})|U_{ij} = u\big]$, $\mathbf{\Delta}_{\mathbf{X}}(u|\mathbf{\Theta}) = \mathsf{E}\big[\mathbf{X}_{ij}\mathbf{X}_{ij}^\tau(\mathbf{\Theta})|U_{ij} = u\big]$, and $\mathbf{\Delta}^+$ denotes the generalised inverse of the matrix $\mathbf{\Delta}$ which becomes the conventional inverse if $\mathbf{\Delta}$ is positive definite. Let

$$\mathbf{W} = \begin{pmatrix} \mathsf{E}\big[\widetilde{\mathbf{X}}_{ij}\widetilde{\mathbf{X}}_{ij}^\tau\big] & \mathsf{E}\big[\gamma_{10}(U_{ij})\widetilde{\mathbf{X}}_{ij}\widetilde{\mathbf{X}}_{ij}^\tau\big] & \cdots & \mathsf{E}\big[\gamma_{d_00}(U_{ij})\widetilde{\mathbf{X}}_{ij}\widetilde{\mathbf{X}}_{ij}^\tau\big] \\ \mathsf{E}\big[\gamma_{10}(U_{ij})\widetilde{\mathbf{X}}_{ij}\widetilde{\mathbf{X}}_{ij}^\tau\big] & \mathsf{E}\big[\gamma_{10}^2(U_{ij})\widetilde{\mathbf{X}}_{ij}\widetilde{\mathbf{X}}_{ij}^\tau\big] & \cdots & \mathsf{E}\big[\gamma_{10}(U_{ij})\gamma_{d_00}(U_{ij})\widetilde{\mathbf{X}}_{ij}\widetilde{\mathbf{X}}_{ij}^\tau\big] \\ \vdots & \vdots & \vdots & \vdots \\ \mathsf{E}\big[\gamma_{d_00}(U_{ij})\widetilde{\mathbf{X}}_{ij}\widetilde{\mathbf{X}}_{ij}^\tau\big] & \mathsf{E}\big[\gamma_{d_00}(U_{ij})\gamma_{10(U_{ij})}\widetilde{\mathbf{X}}_{ij}\widetilde{\mathbf{X}}_{ij}^\tau\big] & \cdots & \mathsf{E}\big[\gamma_{d_00}^2(U_{ij})\widetilde{\mathbf{X}}_{ij}\widetilde{\mathbf{X}}_{ij}^\tau\big] \end{pmatrix}$$

and

$$\mathbf{V}_n = \Big[\mathbf{V}_n^\tau(0), \mathbf{V}_n^\tau(1), \ldots, \mathbf{V}_n^\tau(d_0)\Big]^\tau,$$

where

$$\mathbf{V}_n(k) = \frac{1}{\sqrt{N(n)}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\Big[\gamma_{k0}(U_{ij})\big(\widetilde{\mathbf{X}}_{ij} - \overline{\mathbf{X}}_{ij}\big)\varepsilon_{ij} + \widehat{\mathbf{X}}_{ij}(k)\varepsilon_{ij} + \widehat{\mathbf{\Delta}}_k\boldsymbol{\gamma}_0(U_{ij})\Big]$$

for $k = 0, 1, \ldots, d_0$, $\gamma_{00}(\cdot) \equiv 1$, $N(n) = \sum_{i=1}^n m_i$, $\overline{\mathbf{X}}_{ij} = \mathsf{E}\big[\mathbf{\Delta}_{\mathbf{X}}(U_{ij}|\mathbf{\Theta}_0)\mathbf{\Delta}^+(U_{ij}|\mathbf{\Theta}_0)\big]\mathbf{X}_{ij}(\mathbf{\Theta}_0)$,

$$\widehat{\mathbf{X}}_{ij}(k) = \mathsf{E}\big[\mathbf{\Delta}_{\mathbf{X}}(U_{ij}|\mathbf{\Theta}_0)\mathbf{\Delta}^+(U_{ij}|\mathbf{\Theta}_0)\big]\mathsf{E}\big[\gamma_{k0}(U_{ij})\mathbf{\Delta}(U_{ij}|\mathbf{\Theta}_0)\big]\mathbf{\Delta}^+(U_{ij}|\mathbf{\Theta}_0)\mathbf{X}_{ij}(\mathbf{\Theta}_0),$$
$$\widehat{\mathbf{\Delta}}_k = \mathsf{E}\big[\mathbf{\Delta}_{\mathbf{X}}(U_{ij}|\mathbf{\Theta}_0)\mathbf{\Delta}^+(U_{ij}|\mathbf{\Theta}_0)\big]\mathsf{E}\big[\gamma_{k0}(U_{ij})\mathbf{\Delta}(U_{ij}|\mathbf{\Theta}_0)\big].$$

We assume that there exists a matrix $\mathbf{W}_1$ such that $\lim_{n\to\infty}\mathsf{E}\big[\mathbf{V}_n\mathbf{V}_n^\tau\big] = \mathbf{W}_1$.

We next give the asymptotic distribution theory for the parameter estimators $\widehat{\boldsymbol{\theta}}$ and $\widehat{\mathbf{\Theta}}$ and the nonparametric estimator $\widehat{\boldsymbol{\gamma}}(\cdot)$ defined in Section 2.1 which ignores the within-subject correlation.

Proposition 3.1. *Suppose that Assumptions 1–3 in Appendix A are satisfied and there exists a positive constant c_σ such that*

$$\lim_{n\to\infty}\frac{1}{N(n)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sigma_{ij}^2 = c_\sigma, \quad \textit{where } \sigma_{ij}^2 = \mathsf{E}[\varepsilon_{ij}^2].$$

Then we have

$$\sqrt{N(n)}\begin{bmatrix} \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \\ \mathsf{vec}(\widehat{\mathbf{\Theta}}) - \mathsf{vec}(\mathbf{\Theta}_0) \end{bmatrix} \xrightarrow{d} \mathsf{N}\left(\mathbf{0}_{d(d_0+1)},\ \mathbf{W}^+\mathbf{W}_1\mathbf{W}^+\right) \tag{3.1}$$

and

$$\sqrt{N(n)h}\left[\widehat{\boldsymbol{\gamma}}(u) - \boldsymbol{\gamma}_0(u) - \frac{1}{2}\mu_2\boldsymbol{\gamma}_0''(u)h^2\right] \xrightarrow{d} \mathsf{N}\left(\mathbf{0}_{d_0},\ \omega(u)\mathbf{\Delta}^+(u|\mathbf{\Theta}_0)\right), \tag{3.2}$$

where $\omega(u) = \nu_0 c_\sigma/f(u)$, $\boldsymbol{\gamma}_0''(u)$ is the second-order derivative of $\boldsymbol{\gamma}_0(u)$, $\mu_j = \int u^j K(u)du$, $\nu_0 = \int K^2(u)du$ and $f(\cdot)$ is the density function of U_{ij}.

Remark 3.1. As the number of the observations for each subject, m_i, is assumed to be fixed, the above proposition shows that the estimation of the parameters $\boldsymbol{\theta}_0$ and $\mathbf{\Theta}_0$ has the well known root-n convergence rate, and the nonparametric estimation of $\boldsymbol{\gamma}_0(\cdot)$ has a point-wise convergence rate of $O_P(h^2 + 1/\sqrt{nh})$. This result can be seen as an extension of Theorems 2 and 3 in Jiang *et al* (2013) to the longitudinal data setting. If we further assume that ε_{ij} is independent and identically distributed over both i and j with $\sigma^2 = \mathsf{Var}[\varepsilon_{ij}^2]$, then the asymptotic variance in (3.1) can be simplified to $\sigma^2\mathbf{W}^+$. By (2.7), (3.1) and (3.2), we can show that, through using the principal component structure on functional coefficients, the asymptotic variance of the nonparametric estimation $\widehat{\boldsymbol{\beta}}(u)$ defined in (2.7) is $\omega(u)\boldsymbol{\Theta}_0\boldsymbol{\Delta}^+(u|\boldsymbol{\Theta}_0)\boldsymbol{\Theta}_0^\tau/[N(n)h]$. In contrast, the direct local linear estimation of the functional coefficients with the same kernel function and bandwidth has the asymptotic variance of $\omega(u)\boldsymbol{\Delta}^+(u)/[N(n)h]$ with $\boldsymbol{\Delta}(u) = \mathsf{E}\left[\mathbf{X}_{ij}\mathbf{X}_{ij}^\tau|U_{ij} = u\right]$. Following the argument in Jiang *et al* (2013), the estimation $\widehat{\boldsymbol{\beta}}(u)$ would be asymptotically more efficient when d_0 is smaller than d.

We next give the asymptotic theory for the local linear estimation of the principal functional coefficients which accounts for the within-subject correlation by utilising the Cholesky decomposition on the within-subject covariance matrices. Define $e_{ij} = \eta_{ij} + \sum_{k=1}^{j-1}(c_{i,jk} - c_{i,jk}^\diamond)\varepsilon_{ik}$. Suppose that there exist two positive constants: c_τ and $c_\rho^\diamond$, such that

$$\lim_{n\to\infty}\frac{1}{N(n)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{\tau_{ij}}{(\rho_{ij}^\diamond)^2} = c_\tau \quad \text{and} \quad \lim_{n\to\infty}\frac{1}{N(n)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left(\rho_{ij}^\diamond\right)^{-1} = c_\rho^\diamond, \tag{3.3}$$

where $\tau_{ij} = \mathsf{E}\left[e_{ij}^2\right]$. Note that for each i and j, η_{ij} and $\{\varepsilon_{i1}, \ldots, \varepsilon_{i,j-1}\}$ are independent. Hence, $\tau_{ij} = \mathsf{E}\left[\eta_{ij}^2\right] + \mathsf{E}\left[\left(\sum_{k=1}^{j-1}(c_{i,jk} - c_{i,jk}^\diamond)\varepsilon_{ik}\right)^2\right] = \rho_{ij} + \tau_{ij}^\diamond$ with $\tau_{ij}^\diamond = \mathsf{E}\left[\left(\sum_{k=1}^{j-1}(c_{i,jk} - c_{i,jk}^\diamond)\varepsilon_{ik}\right)^2\right]$, and

$$\begin{aligned}
\omega^\diamond(u) &\equiv \frac{\nu_0}{f(u)}\left[\lim_{n\to\infty}\frac{1}{N(n)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{\tau_{ij}}{(\rho_{ij}^\diamond)^2}\right]\left[\lim_{n\to\infty}\frac{1}{N(n)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left(\rho_{ij}^\diamond\right)^{-1}\right]^{-2} \\
&= \frac{\nu_0}{f(u)}\left[\lim_{n\to\infty}\frac{1}{N(n)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{\rho_{ij}}{(\rho_{ij}^\diamond)^2}\right]\left[\lim_{n\to\infty}\frac{1}{N(n)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left(\rho_{ij}^\diamond\right)^{-1}\right]^{-2} \\
&\quad +\frac{\nu_0}{f(u)}\left[\lim_{n\to\infty}\frac{1}{N(n)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{\tau_{ij}^\diamond}{(\rho_{ij}^\diamond)^2}\right]\left[\lim_{n\to\infty}\frac{1}{N(n)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left(\rho_{ij}^\diamond\right)^{-1}\right]^{-2} \\
&\equiv \omega_1^\diamond(u) + \omega_2^\diamond(u).
\end{aligned} \tag{3.4}$$

It is easy to find that $\omega^\diamond(u) = \nu_0 c_\tau/[(c_\rho^\diamond)^2 f(u)]$.

The asymptotic distribution theory for $\overline{\boldsymbol{\gamma}}(\cdot)$ is given in the following theorem.

THEOREM 3.2. *Suppose that Assumptions 1–4 in Appendix A and (3.3) are satisfied. Then we have*

$$\sqrt{N(n)h}\left[\overline{\boldsymbol{\gamma}}(u) - \boldsymbol{\gamma}_0(u) - \frac{1}{2}\mu_2\boldsymbol{\gamma}_0''(u)h^2\right] \xrightarrow{d} \mathsf{N}\left(\mathbf{0}_{d_0},\ \omega^\diamond(u)\boldsymbol{\Delta}^+(u|\boldsymbol{\Theta}_0)\right), \tag{3.5}$$

where $\omega^{\diamond}(u)$ is defined in (3.4).

REMARK 3.2. The above theorem can be seen as a generalisation of Theorem 2.2 in Yao and Li (2013) to the principal functional coefficient models with unbalanced longitudinal data. From (3.5), we can find that the point-wise convergence rate of the local linear estimation would remain the same even if there is a misspecification in the working covariance matrix $\mathbf{\Sigma}_i^{\diamond}$. However, the misspecification in $\mathbf{\Sigma}_i^{\diamond}$ would lead to a larger asymptotic variance as represented by a positive $\omega_2^{\diamond}(u)$ in the decomposition of $\omega^{\diamond}(u)$ in (3.4) when $\mathbf{\Sigma}_i^{\diamond}$ is misspecified.

We next compare the asymptotic variances between the two local linear estimators of the principal functional coefficients $\widehat{\gamma}(u)$ and $\overline{\gamma}(u)$ in the case where the within-subject covariance is correctly specified up to a constant multiple, i.e., $\mathbf{\Sigma}_i^{\diamond} = c_0\mathbf{\Sigma}_i$, where $0 < c_0 < \infty$. We can show that $\omega_2^{\diamond}(u) \equiv 0$ and

$$\omega^{\diamond}(u) = \omega_1^{\diamond}(u) = \frac{\nu_0}{f(u)} \left[\lim_{n\to\infty} \frac{1}{N(n)} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \rho_{ij}^{-1} \right]^{-1}.$$

Assume that there exists a positive constant c_ρ such that

$$c_\rho = \lim_{n\to\infty} \frac{1}{N(n)} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \rho_{ij}^{-1}. \tag{3.6}$$

By using the harmonic mean value inequality, we have

$$\left[\frac{1}{N(n)} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \rho_{ij}^{-1} \right]^{-1} \leq \frac{1}{N(n)} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \rho_{ij}, \tag{3.7}$$

and the equality holds only when all the ρ_{ij} are the same. Furthermore, by the Cholesky decomposition, we have

$$\rho_{i1} = \sigma_{i1}^2, \quad \rho_{ij} \leq \sigma_{ij}^2, \quad j = 2, \ldots, m_i,$$

which indicates that

$$\frac{1}{N(n)} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \rho_{ij} \leq \frac{1}{N(n)} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sigma_{ij}^2. \tag{3.8}$$

By (3.7), (3.8), Proposition 3.1 and Theorem 3.2, we immediately obtain the following asymptotic result.

COROLLARY 3.3. *Suppose that the conditions in Proposition 3.1 and Theorem 3.2 and (3.6) are satisfied, and $\mathbf{\Sigma}_i^{\diamond} = c_0\mathbf{\Sigma}_i$ with $0 < c_0 < \infty$. The local linear estimator $\overline{\gamma}(u)$ which accounts for the within-subject correlation is asymptotically more efficient than $\widehat{\gamma}(u)$ which ignores the within-subject correlation.*

Remark 3.3. By (2.7), (2.17) and Corollary 3.3, we can show that the asymptotic variance of the local linear estimation of the model functional coefficients $\widehat{\boldsymbol{\beta}}(u)$ minus that of $\overline{\boldsymbol{\beta}}(u)$ is non-negative definite, and thus $\overline{\boldsymbol{\beta}}(u)$ is asymptotically more efficient than $\widehat{\boldsymbol{\beta}}(u)$. Furthermore, noting that the asymptotic bias terms for $\overline{\boldsymbol{\gamma}}(u)$ and $\widehat{\boldsymbol{\gamma}}(u)$ are the same, the mean squared errors of $\overline{\boldsymbol{\gamma}}(u)$ (or $\overline{\boldsymbol{\beta}}(u)$) are asymptotically smaller than those of $\widehat{\boldsymbol{\gamma}}(u)$ (or $\widehat{\boldsymbol{\beta}}(u)$). This will be justified in our simulation study for the finite sample case.

4 Estimation of the Cholesky decomposition

The asymptotic theorems in Section 3 show that whether the efficiency for the local linear estimation of the (principal) functional coefficients can be improved relies on a correct specification of the within-subject covariance matrix $\boldsymbol{\Sigma}_i$ and the lower triangular matrix $\mathbf{C}_i$ in the Cholesky decomposition. However, in practical applications, the true within-subject covariance matrix is usually unknown. We may construct a working covariance matrix by using a semiparametric method as in Fan *et al* (2007) which relies on a parametric specification of the within-subject correlation matrix. In this section, we will introduce two different estimation methods. In the case of balanced longitudinal data, we will consider estimating $\mathbf{C}_i$ directly together with other parameters $\boldsymbol{\theta}_0$ and $\boldsymbol{\Theta}_0$ via a profile least squares method. In the case of unbalanced longitudinal data, we first use a local linear method to estimate the within-subject covariance function and variance function consistently and then obtain the estimate of the autoregressive coefficients via (2.8) or (2.12).

4.1. The case of balanced longitudinal data

As in Yao and Li (2013), we consider in this section the case of $m_i \equiv m$ which falls into the setting of balanced longitudinal data. We further assume that $\varepsilon_i \equiv (\varepsilon_{i1}, \ldots, \varepsilon_{im})^\tau$ are independent and identically distributed over i and are independent of the covariates, which indicates that the within-subject covariance matrix $\boldsymbol{\Sigma}_i$ is independent of i, i.e., $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}$. Then, the Cholesky decomposition of $\boldsymbol{\Sigma}$ gives

$$\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^\tau = \boldsymbol{\Delta} = \mathsf{diag}\left\{\rho_1, \cdots, \rho_m\right\}, \tag{4.1}$$

where $\mathbf{C}$ is a lower triangular matrix with diagonal elements being one. Similar to the notation used in Section 2.3, we let $c_{jk,0}$ be the (j,k)-th entry of the matrix $-\mathbf{C}$ and $\eta_i = \mathbf{C}\varepsilon_i$ and replace ε_{ik} by $\widetilde{\varepsilon}_{ik}$ in the following estimation procedure, where $\widetilde{\varepsilon}_{ik} = Y_{ik} - \mathbf{X}_{ik}^\tau \widetilde{\boldsymbol{\beta}}(U_{ik})$ is defined as in Section

2.3. Then, for $i = 1, \ldots, n$ and $j = 2, \ldots, m$, the approximating model (2.14) becomes

$$Y_{ij} \approx \mathbf{X}_{ij}^{\tau}\boldsymbol{\theta}_0 + \sum_{k=1}^{d_0} \gamma_{k0}(U_{ij})\mathbf{X}_{ij}^{\tau}\boldsymbol{\Theta}_0(k) + \sum_{k=1}^{j-1} c_{jk,0}\widetilde{\varepsilon}_{ik} + \eta_{ij}. \tag{4.2}$$

Throughout this section, we let $\mathbf{C}_0 = (c_{21,0}, c_{31,0}, c_{32,0}, \ldots, c_{m,m-1,0})^{\tau}$, which consists of the elements of $\mathbf{C}$ below the main diagonal.

Based on (4.2), we develop a feasible procedure for estimating $\boldsymbol{\theta}_0$, $\boldsymbol{\Theta}_0$ and $\mathbf{C}_0$. As in Section 2.2, for given $\boldsymbol{\theta}$, $\boldsymbol{\Theta}$ and $\mathbf{C} = (c_{21}, c_{31}, c_{32}, \ldots, c_{m,m-1})^{\tau}$, we first obtain the local linear estimation of the principal functional coefficients by minimising the kernel-weighted loss function

$$\begin{aligned} L_n^*\big(\mathbf{a}(u), \mathbf{b}(u)\big|\boldsymbol{\theta}, \boldsymbol{\Theta}, \mathbf{C}\big) &= \sum_{i=1}^{n}\sum_{j=2}^{m}\Bigg[Y_{ij} - \mathbf{X}_{ij}^{\tau}\boldsymbol{\theta} - \sum_{k=1}^{j-1} c_{jk}\widetilde{\varepsilon}_{ik} - \sum_{k=1}^{d_0} a_k(u)\mathbf{X}_{ij}^{\tau}\boldsymbol{\Theta}(k) \\ &\qquad - \sum_{k=1}^{d_0} b_k(u)(U_{ij} - u)\mathbf{X}_{ij}^{\tau}\boldsymbol{\Theta}(k)\Bigg]^2 K\left(\frac{U_{ij} - u}{h_*}\right) \end{aligned}$$

where h_* is a bandwidth. The resulting estimator of the principal functional coefficients is denoted by $\widehat{\boldsymbol{\gamma}}(u|\boldsymbol{\theta}, \boldsymbol{\Theta}, \mathbf{C})$. As in Section 2.2, we let $\widetilde{\boldsymbol{\gamma}}(u|\boldsymbol{\theta}, \boldsymbol{\Theta}, \mathbf{C})$ be the sample centralisation of the principal functional-coefficient estimates $\widehat{\boldsymbol{\gamma}}(u|\boldsymbol{\theta}, \boldsymbol{\Theta}, \mathbf{C})$ and let $\widetilde{\gamma}_k(\cdot|\boldsymbol{\theta}, \boldsymbol{\Theta}, \mathbf{C})$ be the k-th element of $\widetilde{\boldsymbol{\gamma}}(u|\boldsymbol{\theta}, \boldsymbol{\Theta}, \mathbf{C})$. We estimate $\boldsymbol{\theta}_0$, $\boldsymbol{\Theta}_0$ and $\mathbf{C}_0$ by minimising the following loss function:

$$Q_n^*(\boldsymbol{\theta}, \boldsymbol{\Theta}, \mathbf{C}) = \sum_{i=1}^{n}\sum_{j=2}^{m}\left[Y_{ij} - \mathbf{X}_{ij}^{\tau}\boldsymbol{\theta} - \sum_{k=1}^{d_0}\widetilde{\gamma}_k(U_{ij}|\boldsymbol{\theta}, \boldsymbol{\Theta}, \mathbf{C})\mathbf{X}_{ij}^{\tau}\boldsymbol{\Theta}(k) - \sum_{k=1}^{j-1} c_{jk}\widetilde{\varepsilon}_{ik}\right]^2, \tag{4.3}$$

Let $\widetilde{\boldsymbol{\theta}}$, $\widetilde{\boldsymbol{\Theta}}$ and

$$\widetilde{\mathbf{C}} = (\widetilde{c}_{21}, \widetilde{c}_{31}, \widetilde{c}_{32}, \ldots, \widetilde{c}_{m,m-1})^{\tau} \tag{4.4}$$

be a minimiser of (4.3), i.e.,

$$Q_n^*(\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\Theta}}, \widetilde{\mathbf{C}}) = \min_{\boldsymbol{\theta}, \boldsymbol{\Theta}, \mathbf{C}} Q_n^*(\boldsymbol{\theta}, \boldsymbol{\Theta}, \mathbf{C}). \tag{4.5}$$

The above estimation method can be seen as a generalisation of the profile likelihood method proposed by Yao and Li (2013) from univariate nonparametric longitudinal data models to principal functional coefficients longitudinal data models. Furthermore, we may add weights (chosen as the inverse of consistent estimators of $\rho_2, \cdots, \rho_m$) in the loss functions $L_n^*(\cdot, \cdot|\boldsymbol{\theta}, \boldsymbol{\Theta}, \mathbf{C})$ and $Q_n^*(\cdot, \cdot, \cdot)$. This would not affect the consistency and convergence rates of the parameter estimators but may potentially make their asymptotic variances smaller. We next give the convergence rates for the estimators $\widetilde{\boldsymbol{\theta}}$, $\widetilde{\boldsymbol{\Theta}}$ and $\widetilde{\mathbf{C}}$.

Proposition 4.1. *Suppose that Assumptions 1(i) and 2–4 in Appendix A are satisfied and Assumption 1(ii) holds when h is replaced by h_*. Furthermore, assume $m_i \equiv m$ and ε_i are independent and identically distributed over i, and are independent of the covariates. Then we have*

$$\begin{bmatrix} \widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \\ \mathsf{vec}(\widetilde{\boldsymbol{\Theta}}) - \mathsf{vec}(\boldsymbol{\Theta}_0) \\ \widetilde{\mathbf{C}} - \mathbf{C}_0 \end{bmatrix} = O_P\left(1/\sqrt{n}\right). \tag{4.6}$$

With the estimates of $c_{ij,0}$ obtained from above, we may transform Y_{ij} into $\widetilde{Y}_{ij}$ via (2.13). The variance of η_{ij}, ρ_j, can be estimated by using the estimated residuals $\widetilde{\varepsilon}_{ij}$ and applying the Cholesky decomposition on the covariance matrix of $\widetilde{\varepsilon}_i = (\widetilde{\varepsilon}_{i1}, \cdots, \widetilde{\varepsilon}_{im})^\tau$. Then, by replacing $\overline{\boldsymbol{\theta}}$ and $\overline{\boldsymbol{\Theta}}$ in (2.15) by $\widetilde{\boldsymbol{\theta}}$ and $\widetilde{\boldsymbol{\Theta}}$, respectively, and minimising the resulting loss function, we can obtain the estimator $\overline{\boldsymbol{\gamma}}(\cdot)$ of the principal functional coefficients. Combining Theorem 3.2 and Proposition 4.1, we obtain the following asymptotic distribution theory for the local linear estimator $\overline{\boldsymbol{\gamma}}(\cdot)$.

Corollary 4.2. *Suppose that the conditions in Theorem 3.2 and Proposition 4.1 are satisfied, and let $c^\diamond_{i,jk}$ in (2.13) be replaced by $\widetilde{c}_{jk}$ defined in (4.4) and (4.5). Then we have*

$$\sqrt{nmh}\left[\overline{\boldsymbol{\gamma}}(u) - \boldsymbol{\gamma}_0(u) - \frac{1}{2}\mu_2\boldsymbol{\gamma}_0''(u)h^2\right] \xrightarrow{d} \mathsf{N}\left(\mathbf{0}_{d_0},\ \omega_1^\diamond(u)\boldsymbol{\Delta}^+(u|\boldsymbol{\Theta}_0)\right), \tag{4.7}$$

where $\omega_1^\diamond(u)$ is defined in (3.4).

Remark 4.1. By estimating the autoregressive coefficients $c_{ij,0}$ in the Cholesky decomposition and the transformation in (2.13) and choosing a consistent estimate of ρ_j in (2.15), we have avoided the misspecification of $\boldsymbol{\Sigma}$ or $\mathbf{C}$ and obtained a more efficient estimator of $\boldsymbol{\gamma}_0(u)$ than $\widehat{\boldsymbol{\gamma}}(u)$ that ignores the within-subject correlation in the longitudinal data.

4.2. The case of unbalanced longitudinal data

The profile least squares estimation method discussed in Section 4.1 strongly relies on the balanced structure of the longitudinal data and cannot be directly extended to a more general setting of unbalanced longitudinal data, which is commonly encountered in practical applications. Hence, we next introduce a different approach to estimate the autoregressive coefficients in the Cholesky decomposition. Motivated by Jiang and Wang (2011) and Li (2011), we assume that the observations on the i-th subject are taken at time points $t_{i1}, \cdots, t_{im_i}$, which fall in a bounded time interval $\mathcal{T}$. Let $Y_{ij} = Y_i(t_{ij})$, $\mathbf{X}_{ij} = \mathbf{X}_i(t_{ij})$, $U_{ij} = U_i(t_{ij})$ and $\varepsilon_{ij} = \varepsilon_i(t_{ij})$. Furthermore, assume that the (j,k)-element of the within-subject covariance matrix $\boldsymbol{\Sigma}_i$ is defined by

$$\mathsf{Cov}(\varepsilon_{ij}, \varepsilon_{ik}) = \mathsf{Cov}\left(\varepsilon_i(t_{ij}), \varepsilon_i(t_{ik})\right) = \sigma(t_{ij}, t_{ik}), \quad t_{ij}, t_{ik} \in \mathcal{T}, \tag{4.8}$$

where $\sigma(\cdot,\cdot)$ is a bivariate positive semi-definite function. As in the literature such as Yao *et al* (2005) and Li (2011), we assume that the bivariate function $\sigma(\cdot,\cdot)$ is continuous everywhere except at the points on the diagonal line, which implies the existence of the so-called *nugget effect*.

It is easy to see that in order to estimate $\mathbf{\Sigma}_i$ consistently, we only need to estimate $\sigma(\cdot,\cdot)$ consistently. As in Section 4.1, we use the estimated residuals $\widetilde{\varepsilon}_{ij} = \widetilde{\varepsilon}_i(t_{ij}) = Y_{ij} - \mathbf{X}_{ij}^{\tau}\widetilde{\boldsymbol{\beta}}(U_{ij})$ with $\widetilde{\boldsymbol{\beta}}(\cdot)$ being the initial local linear estimator of the model functional coefficients. Because of the existence of the nugget effect, we consider the two cases: $\sigma(s,t)$ with $s \neq t$, and $\sigma_0^2(t) = \sigma(t,t)$, separately. The estimation procedure is based on the local linear smoothing. When $s \neq t$, we estimate $\sigma(s,t)$ by $\widetilde{\sigma}(s,t) = \widetilde{\sigma}_{10}$, where $(\widetilde{\sigma}_{10}, \widetilde{\sigma}_{11}, \widetilde{\sigma}_{12})$ is the minimiser to

$$\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1,\neq j}^{m_i}\left[\widetilde{\varepsilon}_{ij}\widetilde{\varepsilon}_{ik} - \sigma_{10} - \sigma_{11}(t_{ij}-s) - \sigma_{12}(t_{ik}-t)\right]^2 K\big(\frac{t_{ij}-s}{b_1}\big)K\big(\frac{t_{ik}-t}{b_2}\big)$$

with respect to σ_{10}, σ_{11} and σ_{12}, where $K(\cdot)$ is a kernel function and b_1 and b_2 are two bandwidths. When $s = t$, we estimate $\sigma_0^2(t)$ by $\widetilde{\sigma}^2(t) = \widetilde{\sigma}_{20}$, where $(\widetilde{\sigma}_{20}, \widetilde{\sigma}_{21})$ is the minimiser to

$$\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left[\widetilde{\varepsilon}_{ij}^2 - \sigma_{20} - \sigma_{21}(t_{ij}-t)\right]^2 K\big(\frac{t_{ij}-t}{b_3}\big)$$

with respect to σ_{20} and σ_{21}, where b_3 is a bandwidth. Proposition 1 in Li (2011) shows that both $\widetilde{\sigma}(s,t)$ and $\widetilde{\sigma}^2(t)$ defined above are uniformly consistent over $s,t \in \mathcal{T}$. Then, we readily construct a consistent estimate of $\mathbf{\Sigma}_i$ by $\widehat{\mathbf{\Sigma}}_i$, where the (j,k)-element of $\widehat{\mathbf{\Sigma}}_i$ is

$$\widehat{\sigma}(t_{ij}, t_{ik}) = \widetilde{\sigma}(t_{ij}, t_{ik})\mathsf{I}(t_{ij} \neq t_{ik}) + \widetilde{\sigma}^2(t_{ij})\mathsf{I}(t_{ij} = t_{ik}) \tag{4.9}$$

where $\mathsf{I}(\cdot)$ is an indicator function. After applying the Cholesky decomposition on $\widehat{\mathbf{\Sigma}}_i$, we can obtain the consistent estimates of the autoregressive coefficients.

To ensure that $\widehat{\mathbf{\Sigma}}_i$ is positive semi-definite in the finite sample studies, we need to make some modification on $\widetilde{\sigma}(s,t)$ and $\widetilde{\sigma}^2(t)$. As in Hall *et al* (2008), we let $\widetilde{\lambda}_k$ and $\widetilde{\phi}_k(\cdot)$ be the eigenvalues and eigenfunctions of $\widetilde{\sigma}(\cdot,\cdot)$ and re-define the estimate of $\sigma(s,t)$ by

$$\widetilde{\sigma}_*(s,t) = \sum_{k=1}^{k_0}\widetilde{\lambda}_k\widetilde{\phi}_k(s)\widetilde{\phi}_k(t), \quad k_0 = \max\{k: \ \widetilde{\lambda}_k > 0\}.$$

On the other hand, to ensure the non-negativity of $\widetilde{\sigma}^2(\cdot)$, we may re-define it through a truncation, i.e., $\widetilde{\sigma}_*^2(t) = \widetilde{\sigma}^2(t)\mathsf{I}(\widetilde{\sigma}^2(t) > \tau_n)$, where τ_n is a pre-determined truncation parameter which is positive and could be very close to zero. The numerical studies in Section 5 below show that this nonparametric estimation method for the covariance matrix works reasonably well in finite samples.

5 Numerical studies

In this section, we give both simulated and empirical examples to examine the finite sample performance of the developed methods. We start with an approach to obtaining some initial parameter estimates for the iterative estimation procedure in Sections 2.2 and 2.4 and a discussion of determining the value of d_0.

5.1. Choice of the initial estimation and determination of d_0

In order to save the computational time of the iterative semiparametric estimation procedure introduced in Section 2, we next discuss the choice of a consistent initial estimation for the parameters. As in the previous sections, we let $\widetilde{\boldsymbol{\beta}}(\cdot)$ be the local linear estimate of the functional coefficients $\boldsymbol{\beta}_0(\cdot)$ with the kernel function $K(\cdot)$ and the bandwidth b. Given the assumption that $\boldsymbol{\theta}_0 = \mathsf{E}[\boldsymbol{\beta}_0(U_{ij})]$, an initial estimate of $\boldsymbol{\theta}_0$ can be chosen as

$$\widehat{\boldsymbol{\theta}}^{(0)} = \frac{1}{N(n)} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \widetilde{\boldsymbol{\beta}}(U_{ij}). \tag{5.1}$$

To construct an initial estimate for $\boldsymbol{\Theta}_0$, we define the $d \times d$ covariance matrix:

$$\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}} = \frac{1}{N(n)} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left[\widetilde{\boldsymbol{\beta}}(U_{ij}) - \widehat{\boldsymbol{\theta}}^{(0)}\right] \left[\widetilde{\boldsymbol{\beta}}(U_{ij}) - \widehat{\boldsymbol{\theta}}^{(0)}\right]^{\tau}$$

and let $\widehat{\boldsymbol{\Theta}}^{(0)}(l)$ be the eigenvector associated with the l-th largest eigenvalue $\widetilde{\lambda}_l$ of $\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}$. From the identification conditions in Section 2.1, a natural initial estimate of $\boldsymbol{\Theta}_0$ is

$$\widehat{\boldsymbol{\Theta}}_k^{(0)} = \left[\widehat{\boldsymbol{\Theta}}^{(0)}(1), \ldots, \widehat{\boldsymbol{\Theta}}^{(0)}(k)\right]. \tag{5.2}$$

if the number of principal functional coefficients is chosen as k. Following the proof of Lemma B.1 in Appendix B, the uniform convergence rates for $\widetilde{\boldsymbol{\beta}}(\cdot)$ is $O_P\left(b^2 + (nb/\log b^{-1})^{-1/2}\right)$. From Theorem 1 in Jiang *et al* (2013), we can also show that the initial estimators of $\boldsymbol{\theta}_0$ and $\boldsymbol{\Theta}_0$ are consistent with the convergence rate $O_P\left(b^2 + (nb/\log b^{-1})^{-1/2}\right)$ when $k = d_0$.

However, as the number of principal functional coefficients is usually unknown in practical applications, we next discuss how to determine this number. Jiang *et al* (2013) propose a Bayesian Information Criterion (BIC) to estimate d_0 and show that it has a good performance in numerical studies. In this paper, we use a simple ratio method introduced in Lam and Yao (2012) to estimate d_0:

$$\widetilde{d}_0 = \arg\min_{1 \le k \le d} \widetilde{\lambda}_{k+1}/\widetilde{\lambda}_k, \tag{5.3}$$

where $\widetilde{\lambda}_k$ is the k-th largest eigenvalue of $\widetilde{\mathbf{\Sigma}}_{\boldsymbol{\beta}}$, $\widetilde{\lambda}_{d+1} = 0$, and $0/0 \equiv 1$. In order to reduce estimation error, we set $\widetilde{\lambda}_k = 0$ if $|\widetilde{\lambda}_k| < \epsilon_0$, where ϵ_0 is a pre-specified small number. The simulation studies below show that this ratio method works well when using an ϵ_0 value of 0.05.

5.2. Simulation study

In this section, we compare the performance of three estimation methods for the varying coefficient functions $\boldsymbol{\beta}_0(u)$ in the following principal functional coefficient longitudinal data model:

$$Y_{ij} = \mathbf{X}_{ij}^{\tau}\boldsymbol{\beta}_0(U_{ij}) + \varepsilon_{ij}, \quad \boldsymbol{\beta}_0(U_{ij}) = \boldsymbol{\theta}_0 + \mathbf{\Theta}_0\boldsymbol{\gamma}_0(U_{ij}), \quad i = 1, \dots, n, \quad j = 1, \dots, m_i, \tag{5.4}$$

where $d = 7$, $d_0 = 1$, $\boldsymbol{\theta}_0 = (1, \dots, 1)^{\tau}$, $\mathbf{\Theta}_0 = (-1/\sqrt{2}, 1/\sqrt{2}, 0, \dots, 0)^{\tau}$, $\boldsymbol{\gamma}_0(u) = 10u(1-u) - 5/3$, the index variables U_{ij} are independently drawn from the uniform distribution in $[0, 1]$, $\mathbf{X}_{ij} = (1, \mathbf{X}_{ij}^*)^{\tau}$ with $\mathbf{X}_{ij}^*$ independently drawn from a 6-dimensional Gaussian distribution with mean $\mathbf{0}$ and covariance matrix I_6 and ε_{ij} will be specified later. The simulation setting is similar to that in Jiang *et al* (2013). The first estimator of $\boldsymbol{\beta}_0(u)$, denoted by *FCM*, is the one obtained by applying local linear smoothing directly on the functional coefficient model equation (5.4). The second estimator, denoted by *PFCM*, is the one detailed in (2.3)–(2.7) of Section 2.2 which ignores the possible within-subject correlation structure. The third, denoted by *PFCM+CD*, is the one detailed in (2.15)–(2.17), where the Cholesky decomposition is employed and the autoregressive coefficients $c_{i,jk}$ are estimated using either the method proposed in Section 4.1 or that in Section 4.2 depending on whether the data is balanced or not.

Given the identification condition $\boldsymbol{\theta}_0 = \mathsf{E}[\boldsymbol{\beta}_0(U_{ij})]$, we use the sample mean of the FCM estimates $\widetilde{\boldsymbol{\beta}}(\cdot)$ defined in (5.1), as the initial estimate of $\boldsymbol{\theta}_0$ in the PFCM and PFCM+CD estimation methods. Based on $\widetilde{\boldsymbol{\beta}}(U_{ij})$, we also perform an eigenanalysis on their sample covariance matrix to determine d_0 and construct the initial estimate of $\mathbf{\Theta}_0$ for PFCM and PFCM+CD as described in Section 5.1. We compare the estimation errors of the three methods in estimating the functional coefficients $\boldsymbol{\beta}(u)$. The estimation error for an estimator is measured as the mean absolute deviation:

$$\frac{1}{nd}\sum_{k=1}^{d}\sum_{i=1}^{n}\frac{1}{m_i}\sum_{j=1}^{m_i}|\beta_k^*(U_{ij}) - \beta_k(U_{ij})|,$$

where $\beta_k^*(\cdot)$ is the estimate of $\beta_k(\cdot)$ by using one of the three estimation methods described above.

We consider the following two cases in the simulation study.

Case I. Consider a data generating process for balanced longitudinal data with $m_i \equiv 4$ and ε_{ij} are Gaussian error terms that are independent over subjects i and have an AR (1) within-subject covariance structure with coefficient ι, i.e, $\text{cov}(\varepsilon_{ij}, \varepsilon_{ik}) = \iota^{|j-k|}$ for $1 \leq i \leq n$, $1 \leq j, k \leq m$, and

$\text{cov}(\varepsilon_{i_1 j}, \varepsilon_{i_2 k}) = 0$ for $1 \le i_1 \ne i_2 \le n$ and $1 \le j, k \le m$. In order to investigate the performance of the three methods under different levels of within-subject correlation, we set $\iota = 0.1, 0.5, 0.9$. The number of subject is chosen to be $n = 50, 100, 200$.

In order to obtain an initial estimate of the autoregressive coefficients in the Cholesky decomposition, we first obtain the residuals of the FCM, $\widetilde{\varepsilon}_{ij} = Y_{ij} - \mathbf{X}_{ij}^{\tau} \widetilde{\boldsymbol{\beta}}(U_{ij})$, and then estimate the within-subject covariance matrix of $\widetilde{\varepsilon}_i = (\widetilde{\varepsilon}_{i1}, \cdots, \widetilde{\varepsilon}_{i4})^{\tau}$ and obtain the initial estimates of the autoregressive coefficients through implementing the Choleskey decomposition on the estimated within-subject covariance matrix. With these initial estimates of the autoregressive coefficients in the Cholesky decomposition as well as the initial estimates of $\boldsymbol{\theta}_0$ and $\boldsymbol{\Theta}_0$, we use the estimation procedure in Section 4.1 to obtain the root-n consistent parametric estimates. Table 5.1 summaries the estimation errors of the three methods and their standard deviations (in parentheses) over 1000 realisations of the setting (5.4) for different combinations of n and ι.

From Table 5.1, we find that the performance of FCM estimation is alway the worst as it ignores both the principal component structure of the functional coefficients and the within-subject correlation of the model errors. This shows that we need to take the principal component structure into account when estimating the functional coefficients, as in PFCM and PFCM+CD estimations. When $\iota = 0.1$, the estimation errors of PFCM and PFCM+CD estimations are close, which is not surprising because within-subject correlation is weak. However, as ι becomes larger ($\iota = 0.5$ and 0.9), the PFCM+CD estimation outperforms the PFCM estimation even when the number of the subjects is as small as 50, which supports the asymptotic theorems in Section 3.

Table 5.1. Estimation errors of FCM, PFCM, and PFCM+CD for Case I

ι \ n		$n = 50$	$n = 100$	$n = 200$
$\iota = 0.1$	FCM	0.1242(0.0182)	0.0879(0.0122)	0.0631(0.0077)
	PFCM	0.0971(0.0207)	0.0620(0.0121)	0.0430(0.0080)
	PFCM+CD	0.0999(0.0209)	0.0629(0.0123)	0.0432(0.0080)
$\iota = 0.5$	FCM	0.1266(0.0185)	0.0888(0.0132)	0.0637(0.0084)
	PFCM	0.0998(0.0218)	0.0629(0.0122)	0.0433(0.0081)
	PFCM+CD	0.0950(0.0198)	0.0588(0.0112)	0.0396(0.0074)
$\iota = 0.9$	FCM	0.1271(0.0223)	0.0897(0.0144)	0.0651(0.0089)
	PFCM	0.1021(0.0239)	0.0651(0.0138)	0.0449(0.0087)
	PFCM+CD	0.0726(0.0158)	0.0429(0.0097)	0.0273(0.0064)

Case II. We then consider the same principal functional coefficient longitudinal data model as in (5.4) but with unbalanced data, i.e., the number of observations for each subject m_i varies across i. The observations for each subject are pre-scheduled to be taken at times 1, 2, 3, 4, 5. However, each observation time has a probability of 0.2 of being skipped (i.e., having no observations taken at that time), which results in unbalanced numbers of observations for different subjects. All the other variables and parameters are the same as those in Case I. We use the nonparametic estimation method introduced in Section 4.2 to estimate the within-subject covariance matrix and then implement the Cholesky decomposition. As in Case I, we compare the three estimation methods: FCM, PFCM and PFCM+CD. The simulation results for this unbalanced longitudinal data case are summarised in Table 5.2 from which we have the similar finding to that in Case I.

The number of principal functional coefficients d_0 is estimated using the simple ratio method introduced in (5.3). In order to evaluate the performance of this method, we report, in Tables 5.3 and 5.4, the percentages of simulation replications in which d_0 is correctly estimated. The results in these two tables show, for both balanced and unbalanced longitudinal data, that when the number of observations $N(n)$ is about 200, the percentage of replications in which d_0 is correctly estimated is around 70%. This percentage rises to around 98% when $N(n)$ increases to around 400 and further to 100% when $N(n)$ increases to around 800.

Table 5.2. Estimation errors of FCM, PFCM, and PFCM+CD for Case II

ι \ n		$n = 50$	$n = 100$	$n = 200$
$\iota = 0.1$	FCM	0.1242(0.0183)	0.0879(0.0117)	0.0634(0.0079)
	PFCM	0.0977(0.0238)	0.0618(0.0119)	0.0429(0.0079)
	PFCM+CD	0.0987(0.0234)	0.0623(0.0120)	0.0430(0.0079)
$\iota = 0.5$	FCM	0.1223(0.0186)	0.0897(0.0130)	0.0644(0.0084)
	PFCM	0.0961(0.0213)	0.0637(0.0130)	0.0439(0.0081)
	PFCM+CD	0.0923(0.0188)	0.0618(0.0125)	0.0422(0.0077)
$\iota = 0.9$	FCM	0.1269(0.0209)	0.0896(0.0142)	0.0658(0.0093)
	PFCM	0.1009(0.0230)	0.0649(0.0137)	0.0458(0.0089)
	PFCM+CD	0.0837(0.0182)	0.0522(0.0115)	0.0358(0.0076)

Table 5.3. Frequency at which d_0 is corrected estimated for Case I

ι \ n	$n = 50$	$n = 100$	$n = 200$
$\iota = 0.1$	72.30%	98.20%	100%
$\iota = 0.5$	69.80%	98.30%	100%
$\iota = 0.9$	70.40%	98.50%	100%

Table 5.4. Frequency at which d_0 is corrected estimated for Case II

ι \ n	$n = 50$	$n = 100$	$n = 200$
$\iota = 0.1$	71.00%	98.20%	100%
$\iota = 0.5$	70.00%	97.20%	100%
$\iota = 0.9$	68.80%	97.20%	100%

5.3. Empirical application

The research on the effects of foreign direct investment on the economic growth of a host country has attracted much interest in the past two decades. There has been a large literature devoted to this research. Various modelling approaches have been used to evaluate the effects of foreign direct investment. While many studies have employed a linear modelling approach, some research (Li and Liu, 2004, Kottaridi and Stengos, 2010) has found a nonlinear relationship between foreign direct investment and economic growth and amended the conventional linear model by adding interaction terms of the foreign direct investment with some control variables. Cai *et al* (2013) use a partially varying coefficient quantile regression model to investigate this problem in order to allow for the effects of foreign direct investment to vary with the initial economic condition of the host country. Chen *et al* (2013) employ a partially linear single index model to enhance the modelling flexibility of conventional linear models while at the same time retain the interpretability of linear models.

In this section, we will use a balanced longitudinal data set collected from 22 Organisation for Economic Co-operation and Development (OECD) countries over the years 1970–2000. Some OECD countries are excluded from the study due to unavailability of the full data. As a common practice in the economic growth literature, we use 5-year averages of involving variables in order to reduce the impact of year-to-year fluctuations. This leaves us a data set with $n = 22$ countries and $m = 7$ time periods. Some previous studies have shown that the effects of foreign direct investment on economic growth depend on the initial economic conditions of the host country. Hence, we first use the following varying coefficient model with the GDP per capita at the beginning of each 5-year period as the index variable:

$$Y_{it} = \beta_1(U_{it})\log(\mathrm{FDI}_{it}) + \beta_2(U_{it})\mathrm{SCH}_{it} + \beta_3(U_{it})\mathrm{POP}_{it} + \beta_4(U_{it})\log(\mathrm{DI}_{it}) + e_{it}, \tag{5.5}$$

where Y_{it} is the GDP per capita growth in country i for time period t, U_{it} is the log of GDP per capita at the beginning of the t-th 5-year period, FDI_{it} is the foreign direct investment in percentage of GDP, SCH_{it} is the average years of schooling, POP_{it} is the population growth rate, and DI_{it} is the domestic investment in percentage of GDP, wherein the domestic investment is measured by gross fixed capital formation. All variables involved are normalised to have zero mean and unit variance.

By employing local linear smoothing to model (5.5), we obtain the estimates of the four varying coefficients, which are plotted in Figure 5.1. From Figure 5.1, we can observe some common pattern in the four curves, especially on the right half. Hence, we may assume that the variation in the coefficient function vector $\boldsymbol{\beta}(U_{it}) = [\beta_1(U_{it}), \beta_2(U_{it}), \beta_3(U_{it}), \beta_4(U_{it})]^\tau$ is generated by the principal coefficient function vector $\boldsymbol{\gamma}(U_{it}) = [\gamma_1(U_{it}), \cdots, \gamma_{d_0}(U_{it})]^{\tau}$:

$$\boldsymbol{\beta}(U_{it}) = \boldsymbol{\theta} + \boldsymbol{\Theta}\boldsymbol{\gamma}(U_{it})$$

where $1 \leq d_0 \leq 4$. By the ratio method introduced in (5.3), the integer d_0 is estimated as $\widetilde{d}_0 = 1$. By using the proposed semiparametric estimation method in combination with the Cholesky decomposition for balanced data to take into account of the within-subject correlation, we obtain the estimate of the principal coefficient function, which is depicted in Figure 5.2. The estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\Theta}$ and their standard errors (in parentheses) are summarised in Table 5.5. Through the principal component analysis on the functional coefficients in model (5.5), we can find that the dimension of the nonparametric components is reduced from 4 to 1, and the estimated principal functional coefficient in Figure 5.2 can, in some sense, reflect the general pattern of the four estimated coefficient functions in Figure 5.1.

TABLE 5.5. Estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\Theta}$ in the empirical example

$\boldsymbol{\theta}$	$\boldsymbol{\Theta}$
0.1205(0.0579)	0.5755(0.1637)
-0.1866(0.0641)	0.2256(0.1513)
0.0271(0.0599)	0.7812(0.1424)
0.4625(0.0614)	0.0874(0.0907)

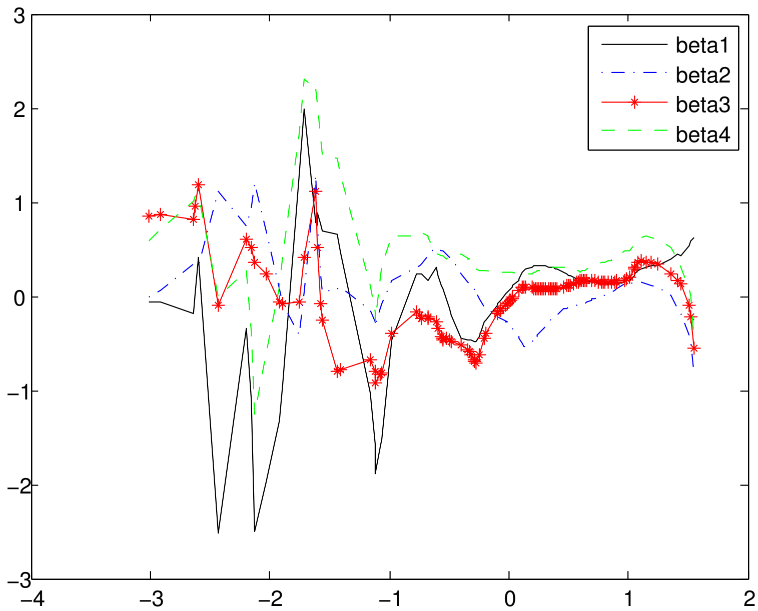

Figure 5.1. Local linear estimates of the varying coefficient functions in model (5.5).

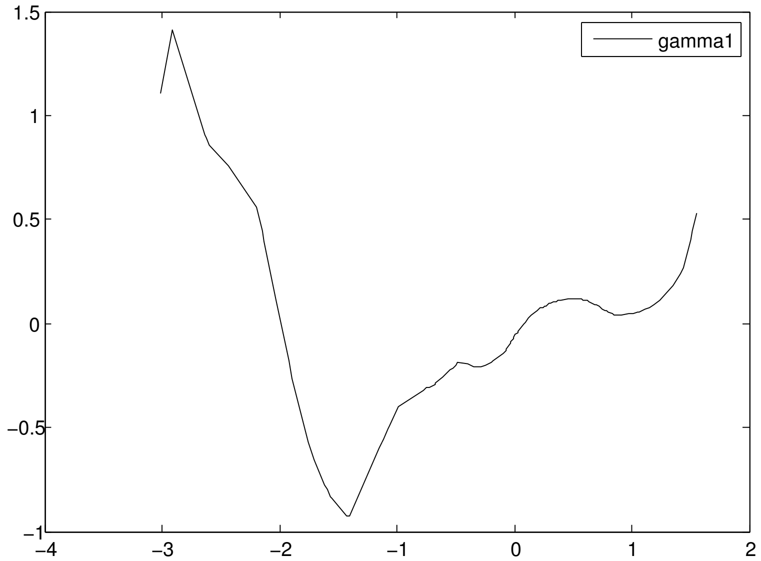

Figure 5.2. Nonparametric estimate of the principal coefficient function in model (5.5).

6 Conclusion

This paper introduces a novel functional-coefficient modelling approach for analysing longitudinal data. By imposing a principal component structure on the functional coefficients, the dimension reduction on the nonparametric components can be achieved. Both the parameters and principal functional coefficients in the model are estimated through an iterative semiparametric estimation procedure. To account for the within-subject correlation in longitudinal data, we apply the Cholesky decomposition to the within-subject covariance matrix and further propose an asymptotically more efficient nonparametric estimation for the (principal) functional coefficients. Under some regularity conditions, we establish the asymptotic distribution theory for the proposed parametric and nonparametric estimation. To ensure the efficiency improvement, the within-subject covariance in longitudinal data needs to be correctly specified up to a constant multiple. Therefore we also propose two approaches to consistently estimate the Cholesky decomposition on the within-subject covariance matrix for balanced and unbalanced longitudinal data respectively. Some simulation studies as well an empirical application on estimating the GDP per capita growth show that the developed semiparametric model and estimation methodology work reasonably well in finite samples.

Appendix A: Assumptions

We next give some assumptions which are used to prove the asymptotic theory. Some of the following technical conditions might be relaxed at the cost of more lengthy proofs.

ASSUMPTION 1. (i) The kernel function $K(\cdot)$ is a continuous and symmetric probability density function with a compact support.

(ii) The bandwidth h satisfies $n^{2\delta-1}h \to \infty$, $nh^4 \to 0$ and $\frac{\log h^{-1}}{nh} \to 0$, where $\delta < 1-1/\zeta$ with ζ defined in Assumption 2(iii) below.

ASSUMPTION 2. (i) The random elements in $\{(U_{ij}, \mathbf{X}_{ij}, \varepsilon_{ij}) : \ j = 1, 2, \ldots, m_i\}$ are independent over i. Furthermore, U_{ij} and $\mathbf{X}_{ij}$ are identically distributed over both i and j, and

$$\mathsf{E}\left[\varepsilon_{i_1 k}\varepsilon_{i_2 l}\big|(U_{ij}, \mathbf{X}_{ij}),\ i = 1, \cdots, n,\ j = 1, \cdots, m_i\right] = 0 \ \ a.s.$$

for $i_1 \neq i_2$.

(ii) The index variable U_{ij} has a continuous density function $f(u)$ and a compact support $\mathcal{U}$. Furthermore, $f(\cdot)$ is positive and bounded away from zero on $\mathcal{U}$. The joint density function of (U_{ij}, U_{ik}), $f_{jk}(\cdot, \cdot)$, exists and is continuous for $j \neq k$.

(iii) There exists a positive number $\zeta > 2$ such that $\mathsf{E}\big[\|\mathbf{X}_{ij}\|^{\zeta} + |\varepsilon_{ij}|^{\zeta}\big] < \infty$. Furthermore, when $\mathbf{\Theta}$ is in a small neighbourhood of $\mathbf{\Theta}_0$, the matrix $\mathbf{\Delta}(u|\mathbf{\Theta})$ defined in Section 3 is continuous, positive definite and twice differentiable for any $u \in \mathcal{U}$.

ASSUMPTION 3. The functional coefficients $\boldsymbol{\beta}(u)$ have continuous second-order derivatives for $u \in \mathcal{U}$.

ASSUMPTION 4. (i) The sixth moment of $\mathbf{X}_{ij}$ exists, i.e., $\mathsf{E}\big[\|\mathbf{X}_{ij}\|^6\big] < \infty$. The matrix $\mathbf{\Delta}(u) \equiv \mathsf{E}[\mathbf{X}_{ij}\mathbf{X}_{ij}^{\tau}|U_{ij} = u]$ is continuous and positive definite for any $u \in \mathcal{U}$.

(ii) The bandwidth b in the initial local linear estimation $\widetilde{\boldsymbol{\beta}}(\cdot)$ satisfies

$$b = o(h), \quad nb \to \infty, \quad (b + \xi_n^*)\xi_n^* = o\big((nh)^{-1/2}\big),$$

where $\xi_n^* = \left(\frac{\log b^{-1}}{nb}\right)^{1/2}$.

The above assumptions are mild. Assumption 1(i) imposes some commonly-used restrictions on the kernel function. Assumption 1(ii) and the moment conditions in Assumption 2(iii) ensure the applicability of the uniform consistency results for the kernel-based estimation derived in Mack and Silverman (1982). The bandwidth condition $nh^4 \to 0$ in Assumption 1(ii) indicates that under-smoothing is needed to derive the root-n convergence rates for the parameter estimation. Assumption 2(i) shows that the longitudinal data are independent across subjects which is not uncommon in the literature (Diggle *et al*, 2002; Wu and Zhang, 2006). However, the restriction of identical distribution on U_{ij} and $\mathbf{X}_{ij}$ can be relaxed at the cost of more complicated forms for $\mathbf{W}$ and $\mathbf{V}_n$ defined in Section 3. The smoothness conditions on $f(\cdot)$ and $\boldsymbol{\beta}(\cdot)$ in Assumptions 2(ii) and 3 are needed as the local linear smoothing of the nonparametric functional coefficients is used in the present paper (Fan and Gijbels, 1996). Assumption 4 is mainly used to prove that the influence of $\varepsilon_{ik} - \widetilde{\varepsilon}_{ik}$ in the estimation of the principal functional coefficients is asymptotically negligible in the proof of Theorem 3.2.

Appendix B: Proofs of the asymptotic theorems

In this appendix, we give the detailed proofs of the asymptotic theorems stated in Sections 3 and 4. To simplify the proofs, we introduce some notation. For $l = 0, 1, 2, \ldots$, we define

$$\mathbf{S}_{nl}\big(u|\boldsymbol{\Theta}\big) = \frac{1}{N(n)h}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\Big(\frac{U_{ij}-u}{h}\Big)^l \mathbf{X}_{ij}(\boldsymbol{\Theta})\mathbf{X}_{ij}^{\tau}(\boldsymbol{\Theta})K\Big(\frac{U_{ij}-u}{h}\Big),$$

$$\mathbf{T}_{nl}\big(u|\boldsymbol{\theta},\boldsymbol{\Theta}\big) = \frac{1}{N(n)h}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\Big(\frac{U_{ij}-u}{h}\Big)^l \mathbf{X}_{ij}(\boldsymbol{\Theta})\left(Y_{ij} - \mathbf{X}_{ij}^{\tau}\boldsymbol{\theta}\right)K\Big(\frac{U_{ij}-u}{h}\Big),$$

and

$$\mathbf{S}_n(u|\boldsymbol{\Theta}) = \begin{bmatrix} \mathbf{S}_{n0}(u|\boldsymbol{\Theta}) & \mathbf{S}_{n1}(u|\boldsymbol{\Theta}) \\ \mathbf{S}_{n1}(u|\boldsymbol{\Theta}) & \mathbf{S}_{n2}(u|\boldsymbol{\Theta}) \end{bmatrix}, \quad \mathbf{T}_n(u|\boldsymbol{\theta},\boldsymbol{\Theta}) = \begin{bmatrix} \mathbf{T}_{n0}(u|\boldsymbol{\theta},\boldsymbol{\Theta}) \\ \mathbf{T}_{n1}(u|\boldsymbol{\theta},\boldsymbol{\Theta}) \end{bmatrix}.$$

Define $\boldsymbol{\Delta}_{1\boldsymbol{\Theta}}(u|\boldsymbol{\theta}_0-\boldsymbol{\theta}) = \mathsf{E}\big[\mathbf{X}_{ij}(\boldsymbol{\Theta})\mathbf{X}_{ij}^{\tau}(\boldsymbol{\theta}_0-\boldsymbol{\theta})|U_{ij}=u\big]$, $\boldsymbol{\Delta}_{2\boldsymbol{\Theta}}(u|\boldsymbol{\Theta}_0-\boldsymbol{\Theta}) = \mathsf{E}\big[\mathbf{X}_{ij}(\boldsymbol{\Theta})\mathbf{X}_{ij}^{\tau}(\boldsymbol{\Theta}_0 - \boldsymbol{\Theta})|U_{ij}=u\big]$, and

$$\mathbf{T}_{n,\varepsilon}(u|\boldsymbol{\Theta}) = \frac{1}{N(n)hf(u)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\mathbf{X}_{ij}(\boldsymbol{\Theta})\varepsilon_{ij}K\Big(\frac{U_{ij}-u}{h}\Big).$$

We start with a technical lemma on the asymptotic expansion of the local linear estimate $\widehat{\boldsymbol{\gamma}}(u|\boldsymbol{\theta},\boldsymbol{\Theta})$. This lemma is a generalisation of Lemma 1 in Jiang *et al* (2013) to the context of longitudinal data.

LEMMA B.1. *Suppose that Assumptions 1–3 in Appendix A are satisfied. Then for $(\boldsymbol{\theta},\boldsymbol{\Theta})$ in a neighbourhood of $(\boldsymbol{\theta}_0,\boldsymbol{\Theta}_0)$ we have*

$$\begin{aligned}\widehat{\boldsymbol{\gamma}}(u|\boldsymbol{\theta},\boldsymbol{\Theta}) - \boldsymbol{\gamma}_0(u) &= \frac{1}{2}\mu_2 h^2\boldsymbol{\gamma}_0''(u) + \boldsymbol{\Delta}^{+}(u|\boldsymbol{\Theta})\big[\boldsymbol{\Delta}_{1\boldsymbol{\Theta}}(u|\boldsymbol{\theta}_0-\boldsymbol{\theta}) + \boldsymbol{\Delta}_{2\boldsymbol{\Theta}}(u|\boldsymbol{\Theta}_0-\boldsymbol{\Theta})\boldsymbol{\gamma}_0(u)\big] \\ &\quad + \boldsymbol{\Delta}^{+}(u|\boldsymbol{\Theta})\mathbf{T}_{n,\varepsilon}(u|\boldsymbol{\Theta}) + o_P\big(\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| + \|\boldsymbol{\Theta}-\boldsymbol{\Theta}_0\|\big) \\ &\quad + O_P\big(h^3 + h\xi_n + \xi_n^2\big) \qquad \text{(B.1)}\end{aligned}$$

uniformly for $u \in \mathcal{U}$*, where* $\mathcal{U}$ *is the compact support of* U_{ij} *defined in Assumption 2(ii), and* $\xi_n = \left(\frac{\log h^{-1}}{nh}\right)^{1/2}$.

PROOF. By (2.2), we can show that

$$\begin{aligned} Y_{ij} - \mathbf{X}_{ij}^{\tau}\boldsymbol{\theta} &= \mathbf{X}_{ij}^{\tau}\big(\boldsymbol{\theta}_0 - \boldsymbol{\theta}\big) + \sum_{k=1}^{d_0} \gamma_{k0}(U_{ij})\mathbf{X}_{ij}^{\tau}\boldsymbol{\Theta}_0(k) + \varepsilon_{ij} \\ &= \mathbf{X}_{ij}^{\tau}\big(\boldsymbol{\theta}_0 - \boldsymbol{\theta}\big) + \sum_{k=1}^{d_0} \gamma_{k0}(U_{ij})\mathbf{X}_{ij}^{\tau}\big[\boldsymbol{\Theta}_0(k) - \boldsymbol{\Theta}(k)\big] \\ &\quad + \sum_{k=1}^{d_0} \gamma_{k0}(U_{ij})\mathbf{X}_{ij}^{\tau}\boldsymbol{\Theta}(k) + \varepsilon_{ij}. \end{aligned} \tag{B.2}$$

Meanwhile, as $\widehat{\gamma}(u|\boldsymbol{\theta}, \boldsymbol{\Theta})$ is the local linear estimation of $\boldsymbol{\gamma}_0(u)$ for given $\boldsymbol{\theta}$ and $\boldsymbol{\Theta}$, which minimises the kernel-weighted loss function $L\big(\mathbf{a}(u), \mathbf{b}(u)|\boldsymbol{\theta}, \boldsymbol{\Theta}\big)$ defined in (2.3), following the standard argument of local linear smoothing (Fan and Gijbels, 1996), we can obtain

$$\widehat{\gamma}(u|\boldsymbol{\theta}, \boldsymbol{\Theta}) = \big(\mathbf{I}_{d_0},\ \mathbf{N}_{d_0}\big)\mathbf{S}_n^{+}(u|\boldsymbol{\Theta})\mathbf{T}_n(u|\boldsymbol{\theta}, \boldsymbol{\Theta}), \tag{B.3}$$

where $\mathbf{I}_{d_0}$ and $\mathbf{N}_{d_0}$ are the $d_0 \times d_0$ identity matrix and null matrix, respectively. A combination of (B.2) and (B.3) leads to

$$\begin{aligned} \widehat{\gamma}(u|\boldsymbol{\theta}, \boldsymbol{\Theta}) &= \big(\mathbf{I}_{d_0},\ \mathbf{N}_{d_0}\big)\mathbf{S}_n^{+}(u|\boldsymbol{\Theta})\mathbf{T}_n^{(1)}(u|\boldsymbol{\theta}_0 - \boldsymbol{\theta},\ \boldsymbol{\Theta}_0 - \boldsymbol{\Theta}) \\ &\quad + \big(\mathbf{I}_{d_0},\ \mathbf{N}_{d_0}\big)\mathbf{S}_n^{+}(u|\boldsymbol{\Theta})\mathbf{T}_n^{(2)}(u|\boldsymbol{\Theta}) + \big(\mathbf{I}_{d_0},\ \mathbf{N}_{d_0}\big)\mathbf{S}_n^{+}(u|\boldsymbol{\Theta})\mathbf{T}_n^{(3)}(u|\boldsymbol{\Theta}), \end{aligned} \tag{B.4}$$

where $\mathbf{T}_n^{(1)}(u|\boldsymbol{\theta}_0 - \boldsymbol{\theta},\ \boldsymbol{\Theta}_0 - \boldsymbol{\Theta})$, $\mathbf{T}_n^{(2)}(u|\boldsymbol{\Theta})$ and $\mathbf{T}_n^{(3)}(u|\boldsymbol{\Theta})$ are defined as $\mathbf{T}_n(u|\boldsymbol{\theta}, \boldsymbol{\Theta})$ with $Y_{ij} - \mathbf{X}_{ij}^{\tau}\boldsymbol{\theta}$ replaced by $\mathbf{X}_{ij}^{\tau}\big(\boldsymbol{\theta}_0 - \boldsymbol{\theta}\big) + \sum_{k=1}^{d_0} \gamma_{k0}(U_{ij})\mathbf{X}_{ij}^{\tau}\big[\boldsymbol{\Theta}(k) - \boldsymbol{\Theta}(k)\big]$, $\sum_{k=1}^{d_0} \gamma_{k0}(U_{ij})\mathbf{X}_{ij}^{\tau}\boldsymbol{\Theta}(k)$ and ε_{ij}, respectively.

Recalling $\boldsymbol{\Delta}(u|\boldsymbol{\Theta}) = \mathsf{E}\big[\mathbf{X}_{ij}(\boldsymbol{\Theta})\mathbf{X}_{ij}^{\tau}(\boldsymbol{\Theta})|U_{ij} = u\big]$ and using Assumptions 1 and 2 as well as the uniform consistency results for nonparametric kernel-based estimation (Mack and Silverman, 1982), we may prove that

$$\sup_{u\in\mathcal{U}} \big\|\mathbf{S}_{nl}(u|\boldsymbol{\Theta}) - \mu_l f(u)\boldsymbol{\Delta}(u|\boldsymbol{\Theta})\big\| = O_P\big(h^2 + \xi_n\big) \tag{B.5}$$

when l is even; and

$$\sup_{u\in\mathcal{U}} \big\|\mathbf{S}_{nl}(u|\boldsymbol{\Theta}) - \mathbf{N}_{d_0}\big\| = O_P\big(h + \xi_n\big) \tag{B.6}$$

when l is odd. Using (B.5) and (B.6), we can show that

$$\sup_{u\in\mathcal{U}} \big\|\mathbf{S}_n(u|\boldsymbol{\Theta}) - \mathbf{S}(u|\boldsymbol{\Theta})\big\| = O_P\big(h + \xi_n\big), \tag{B.7}$$

where $\mathbf{S}(u|\boldsymbol{\Theta}) = \mathsf{diag}(1, \mu_2) \otimes f(u)\boldsymbol{\Delta}(u|\boldsymbol{\Theta})$.

Using Assumptions 1–3 and the uniform consistency results again for nonparametric kernel-based estimation, we have

$$\begin{aligned}
& \sup_{u\in\mathcal{U}} \left\|\mathbf{T}_n^{(1)}(u|\boldsymbol{\theta}_0-\boldsymbol{\theta}, \boldsymbol{\Theta}_0-\boldsymbol{\Theta}) - \mathbf{T}_{\boldsymbol{\Theta}}^{(1,1)}(u|\boldsymbol{\theta}_0-\boldsymbol{\theta}) - \mathbf{T}_{\boldsymbol{\Theta}}^{(1,2)}(u|\boldsymbol{\Theta}_0-\boldsymbol{\Theta})\right\| \\
= \ & O_P\Big((h+\xi_n)\big(\|\boldsymbol{\theta}_0-\boldsymbol{\theta}\| + \|\boldsymbol{\Theta}_0-\boldsymbol{\Theta}\|\big)\Big) = o_P\Big(\|\boldsymbol{\theta}_0-\boldsymbol{\theta}\| + \|\boldsymbol{\Theta}_0-\boldsymbol{\Theta}\|\Big), \qquad \text{(B.8)}
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{T}_{\boldsymbol{\Theta}}^{(1,1)}(u|\boldsymbol{\theta}_0-\boldsymbol{\theta}) &= (1,0)^\tau \otimes \big[f(u)\boldsymbol{\Delta}_{1\boldsymbol{\Theta}}(u|\boldsymbol{\theta}_0-\boldsymbol{\theta})\big], \\
\mathbf{T}_{\boldsymbol{\Theta}}^{(1,2)}(u|\boldsymbol{\Theta}_0-\boldsymbol{\Theta}) &= (1,0)^\tau \otimes \big[f(u)\boldsymbol{\Delta}_{2\boldsymbol{\Theta}}(u|\boldsymbol{\Theta}_0-\boldsymbol{\Theta})\boldsymbol{\gamma}_0(u)\big].
\end{aligned}$$

By (1.3) and Assumption 3, $\boldsymbol{\gamma}_0(\cdot)$ has continuous second-order derivatives. Applying the Taylor expansion to the principal coefficient functions in $\mathbf{T}_n^{(2)}(u|\boldsymbol{\Theta})$, we obtain, uniformly for $u \in \mathcal{U}$,

$$\big(\mathbf{I}_{d_0},\ \mathbf{N}_{d_0}\big)\mathbf{S}_n^+(u|\boldsymbol{\Theta})\mathbf{T}_n^{(2)}(u|\boldsymbol{\Theta}) = \boldsymbol{\gamma}_0(u) + \frac{1}{2}\mu_2 h^2 \boldsymbol{\gamma}_0''(u) + O_P\big(h^3 + h^2\xi_n\big). \qquad \text{(B.9)}$$

Using (B.7) and the uniform consistency result for $\mathbf{T}_n^{(3)}(u|\boldsymbol{\Theta})$, we can prove that

$$\big(\mathbf{I}_{d_0},\ \mathbf{N}_{d_0}\big)\mathbf{S}_n^+(u|\boldsymbol{\Theta})\mathbf{T}_n^{(3)}(u|\boldsymbol{\Theta}) = \boldsymbol{\Delta}^+(u|\boldsymbol{\Theta})\mathbf{T}_{n,\varepsilon}(u|\boldsymbol{\Theta}) + O_P\big(h\xi_n + \xi_n^2\big), \qquad \text{(B.10)}$$

uniformly for $u \in \mathcal{U}$, where $\mathbf{T}_{n,\varepsilon}(u|\boldsymbol{\Theta})$ is defined above Lemma B.1.

Then, we can complete the proof of (B.1) using (B.4) and (B.7)–(B.10). ∎

We next give another important lemma on the asymptotic representation of the parameter estimators $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\Theta}}$. In the following proofs, we use the consistent estimates of $\boldsymbol{\theta}_0$ and $\boldsymbol{\Theta}_0$ constructed in Section 5.1 as the initial values for the semiparametric estimation procedure proposed in Section 2.

LEMMA B.2. *Suppose that Assumptions 1–3 in Appendix A are satisfied. Then we have*

$$\sqrt{N(n)}\begin{bmatrix} \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \\ \mathsf{vec}(\widehat{\boldsymbol{\Theta}}) - \mathsf{vec}(\boldsymbol{\Theta}_0) \end{bmatrix} = \mathbf{W}^+\mathbf{V}_n(1 + o_P(1)), \qquad \text{(B.11)}$$

where $\mathbf{V}_n$ and $\mathbf{W}$ are defined in Section 3.

PROOF. Define

$$E_{ij}(\boldsymbol{\theta}, \boldsymbol{\Theta}) = Y_{ij} - \mathbf{X}_{ij}^\tau\boldsymbol{\theta} - \sum_{k=1}^{d_0} \widetilde{\gamma}_k(U_{ij}|\boldsymbol{\theta}, \boldsymbol{\Theta})\mathbf{X}_{ij}^\tau\boldsymbol{\Theta}(k).$$

To simplify our notation, we let $\mathsf{vec}(\boldsymbol{\theta},\boldsymbol{\Theta})=\big[\boldsymbol{\theta}^\tau,\mathsf{vec}^\tau(\boldsymbol{\Theta})\big]^\tau$ and $\mathsf{vec}(\boldsymbol{\theta}_0,\boldsymbol{\Theta}_0)=\big[\boldsymbol{\theta}_0^\tau,\mathsf{vec}^\tau(\boldsymbol{\Theta}_0)\big]^\tau$. By the Taylor expansion of the loss function $Q_n(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\Theta}})$ defined in (2.5), we have

$$
\begin{aligned}
Q_n\big(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\Theta}}\big) &= Q_n\big(\boldsymbol{\theta}_0,\boldsymbol{\Theta}_0\big)+\Big[\mathbf{Q}_n^{(1)}\big(\boldsymbol{\theta}_0,\boldsymbol{\Theta}_0\big)\Big]^\tau\Big[\mathsf{vec}(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\Theta}})-\mathsf{vec}(\boldsymbol{\theta}_0,\boldsymbol{\Theta}_0)\Big]\\
&\quad+\frac{1}{2}\Big[\mathsf{vec}(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\Theta}})-\mathsf{vec}(\boldsymbol{\theta}_0,\boldsymbol{\Theta}_0)\Big]^\tau\mathbf{Q}_n^{(2)}\big(\boldsymbol{\theta}_*,\boldsymbol{\Theta}_*\big)\Big[\mathsf{vec}(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\Theta}})-\mathsf{vec}(\boldsymbol{\theta}_0,\boldsymbol{\Theta}_0)\Big], \qquad \text{(B.12)}
\end{aligned}
$$

where $\big(\boldsymbol{\theta}_*,\boldsymbol{\Theta}_*\big)$ lies on the line segment between $\big(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\Theta}}\big)$ and $\big(\boldsymbol{\theta}_0,\boldsymbol{\Theta}_0\big)$,

$$
\mathbf{Q}_n^{(1)}(\boldsymbol{\theta}_0,\boldsymbol{\Theta}_0)=2\sum_{i=1}^n\sum_{j=1}^{m_i}E_{ij}(\boldsymbol{\theta}_0,\boldsymbol{\Theta}_0)\cdot\Big[\frac{\partial E_{ij}(\boldsymbol{\theta},\boldsymbol{\Theta})}{\partial\mathsf{vec}(\boldsymbol{\theta},\boldsymbol{\Theta})}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0,\boldsymbol{\Theta}=\boldsymbol{\Theta}_0}\Big]
$$

and

$$
\begin{aligned}
\mathbf{Q}_n^{(2)}(\boldsymbol{\theta}_*,\boldsymbol{\Theta}_*) &= 2\sum_{i=1}^n\sum_{j=1}^{m_i}\Big[\frac{\partial E_{ij}(\boldsymbol{\theta},\boldsymbol{\Theta})}{\partial\mathsf{vec}(\boldsymbol{\theta},\boldsymbol{\Theta})}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_*,\boldsymbol{\Theta}=\boldsymbol{\Theta}_*}\Big]\Big[\frac{\partial E_{ij}(\boldsymbol{\theta},\boldsymbol{\Theta})}{\partial\mathsf{vec}(\boldsymbol{\theta},\boldsymbol{\Theta})}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_*,\boldsymbol{\Theta}=\boldsymbol{\Theta}_*}\Big]^\tau\\
&\quad+2\sum_{i=1}^n\sum_{j=1}^{m_i}E_{ij}(\boldsymbol{\theta}_*,\boldsymbol{\Theta}_*)\Big[\frac{\partial^2E_{ij}(\boldsymbol{\theta},\boldsymbol{\Theta})}{\partial\mathsf{vec}(\boldsymbol{\theta},\boldsymbol{\Theta})\partial\mathsf{vec}(\boldsymbol{\theta},\boldsymbol{\Theta})^\tau}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_*,\boldsymbol{\Theta}=\boldsymbol{\Theta}_*}\Big].
\end{aligned}
$$

Define

$$
\begin{aligned}
\overline{E}_{ij}(\boldsymbol{\theta},\boldsymbol{\Theta}) &= Y_{ij}-\mathbf{X}_{ij}^\tau\boldsymbol{\theta}-\sum_{k=1}^{d_0}\gamma_{k0}(U_{ij})\mathbf{X}_{ij}^\tau\boldsymbol{\Theta}(k),\\
\widetilde{E}_{ij}(\boldsymbol{\theta},\boldsymbol{\Theta}) &= \sum_{k=1}^{d_0}\big[\gamma_{k0}(U_{ij})-\widetilde{\gamma}_k(U_{ij}|\boldsymbol{\theta},\boldsymbol{\Theta})\big]\mathbf{X}_{ij}^\tau\boldsymbol{\Theta}(k).
\end{aligned}
$$

As the initial values in the semiparametric estimation procedure are the consistent estimation of $\big(\boldsymbol{\theta}_0,\boldsymbol{\Theta}_0\big)$, $\big(\boldsymbol{\theta}_*,\boldsymbol{\Theta}_*\big)$ can be made sufficiently close to $\big(\boldsymbol{\theta}_0,\boldsymbol{\Theta}_0\big)$. As a consequence, by Lemma B.1, we may show that

$$
\begin{aligned}
\overline{\mathbf{Q}}_n^{(2)}(\boldsymbol{\theta}_*,\boldsymbol{\Theta}_*) &= 2\sum_{i=1}^n\sum_{j=1}^{m_i}\Big[\frac{\partial E_{ij}(\boldsymbol{\theta},\boldsymbol{\Theta})}{\partial\mathsf{vec}(\boldsymbol{\theta},\boldsymbol{\Theta})}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_*,\boldsymbol{\Theta}=\boldsymbol{\Theta}_*}\Big]\Big[\frac{\partial E_{ij}(\boldsymbol{\theta},\boldsymbol{\Theta})}{\partial\mathsf{vec}(\boldsymbol{\theta},\boldsymbol{\Theta})}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_*,\boldsymbol{\Theta}=\boldsymbol{\Theta}_*}\Big]^\tau\\
&\quad+2\sum_{i=1}^n\sum_{j=1}^{m_i}\overline{E}_{ij}(\boldsymbol{\theta}_*,\boldsymbol{\Theta}_*)\Big[\frac{\partial^2E_{ij}(\boldsymbol{\theta},\boldsymbol{\Theta})}{\partial\mathsf{vec}(\boldsymbol{\theta},\boldsymbol{\Theta})\partial\mathsf{vec}(\boldsymbol{\Theta},\boldsymbol{\Theta})^\tau}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_*,\boldsymbol{\Theta}=\boldsymbol{\Theta}_*}\Big]
\end{aligned}
$$

is the asymptotic leading term of $\mathbf{Q}_n^{(2)}(\boldsymbol{\theta}_*,\boldsymbol{\Theta}_*)$. Let

$$
\begin{aligned}
\overline{\mathbf{Q}}_n^{(2,1)}(\boldsymbol{\theta}_*,\boldsymbol{\Theta}_*) &= 2\sum_{i=1}^n\sum_{j=1}^{m_i}\Big[\frac{\partial E_{ij}(\boldsymbol{\theta},\boldsymbol{\Theta})}{\partial\mathsf{vec}(\boldsymbol{\theta},\boldsymbol{\Theta})}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_*,\boldsymbol{\Theta}=\boldsymbol{\Theta}_*}\Big]\Big[\frac{\partial E_{ij}(\boldsymbol{\theta},\boldsymbol{\Theta})}{\partial\mathsf{vec}(\boldsymbol{\theta},\boldsymbol{\Theta})}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_*,\boldsymbol{\Theta}=\boldsymbol{\Theta}_*}\Big]^\tau,\\
\overline{\mathbf{Q}}_n^{(2,2)}(\boldsymbol{\theta}_*,\boldsymbol{\Theta}_*) &= 2\sum_{i=1}^n\sum_{j=1}^{m_i}\overline{E}_{ij}(\boldsymbol{\theta}_*,\boldsymbol{\Theta}_*)\Big[\frac{\partial^2E_{ij}(\boldsymbol{\theta},\boldsymbol{\Theta})}{\partial\mathsf{vec}(\boldsymbol{\theta},\boldsymbol{\Theta})\partial\mathsf{vec}(\boldsymbol{\Theta},\boldsymbol{\Theta})^\tau}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_*,\boldsymbol{\Theta}=\boldsymbol{\Theta}_*}\Big].
\end{aligned}
$$

Furthermore, we may prove that $\overline{\mathbf{Q}}_n^{(2,2)}(\boldsymbol{\theta}_*, \boldsymbol{\Theta}_*)$ is asymptotically dominated by $\overline{\mathbf{Q}}_n^{(2,1)}(\boldsymbol{\theta}_*, \boldsymbol{\Theta}_*)$ when $\big(\boldsymbol{\theta}_*, \boldsymbol{\Theta}_*\big)$ is close enough to $\big(\boldsymbol{\theta}_0, \boldsymbol{\Theta}_0\big)$. Note that

$$\frac{\partial \overline{E}_{ij}(\boldsymbol{\theta}, \boldsymbol{\Theta})}{\partial \mathsf{vec}(\boldsymbol{\theta}, \boldsymbol{\Theta})} = \left[\mathbf{X}_{ij}^{\tau}, \gamma_{10}(U_{ij})\mathbf{X}_{ij}^{\tau}, \cdots, \gamma_{d_0 0}(U_{ij})\mathbf{X}_{ij}^{\tau}\right]^{\tau},$$

and using Lemma B.1, when $\big(\boldsymbol{\theta}_*, \boldsymbol{\Theta}_*\big)$ is close enough to $\big(\boldsymbol{\theta}_0, \boldsymbol{\Theta}_0\big)$, we may show that

$$\frac{\partial \widetilde{E}_{ij}(\boldsymbol{\theta}, \boldsymbol{\Theta})}{\partial \mathsf{vec}(\boldsymbol{\theta}, \boldsymbol{\Theta})}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_*, \boldsymbol{\Theta}=\boldsymbol{\Theta}_*} = -\begin{bmatrix} \widetilde{\boldsymbol{\Delta}}_{\mathbf{X}}(U_{ij}|\boldsymbol{\Theta}_0)\mathbf{X}_{ij}(\boldsymbol{\Theta}_0) \\ \gamma_{10}(U_{ij})\widetilde{\boldsymbol{\Delta}}_{\mathbf{X}}(U_{ij}|\boldsymbol{\Theta}_0)\mathbf{X}_{ij}(\boldsymbol{\Theta}_0) \\ \vdots \\ \gamma_{d_0 0}(U_{ij})\widetilde{\boldsymbol{\Delta}}_{\mathbf{X}}(U_{ij}|\boldsymbol{\Theta}_0)\mathbf{X}_{ij}(\boldsymbol{\Theta}_0) \end{bmatrix} + o_P(1),$$

where

$$\widetilde{\boldsymbol{\Delta}}_{\mathbf{X}}(U_{ij}|\boldsymbol{\Theta}_0) = \boldsymbol{\Delta}_{\mathbf{X}}(U_{ij}|\boldsymbol{\Theta}_0)\boldsymbol{\Delta}^{+}(U_{ij}|\boldsymbol{\Theta}_0) - \mathsf{E}\left[\boldsymbol{\Delta}_{\mathbf{X}}(U_{ij}|\boldsymbol{\Theta}_0)\boldsymbol{\Delta}^{+}(U_{ij}|\boldsymbol{\Theta}_0)\right].$$

Hence, we have

$$\frac{1}{N(n)} \cdot \frac{\overline{\mathbf{Q}}_n^{(2,1)}(\boldsymbol{\theta}_*, \boldsymbol{\Theta}_*)}{2} = \mathbf{W} + o_P(1) \tag{B.13}$$

when $\big(\boldsymbol{\theta}_*, \boldsymbol{\Theta}_*\big)$ is close enough to $\big(\boldsymbol{\theta}_0, \boldsymbol{\Theta}_0\big)$, where $\mathbf{W}$ is defined in Section 3. Hence, by (B.13) and the previous arguments, we can prove

$$\frac{1}{N(n)} \cdot \frac{\mathbf{Q}_n^{(2)}(\boldsymbol{\theta}_*, \boldsymbol{\Theta}_*)}{2} = \mathbf{W} + o_P(1). \tag{B.14}$$

By (B.12), (B.14) and the definition of $\widehat{\boldsymbol{\Theta}}$, we have

$$\sqrt{N(n)}\big[\mathsf{vec}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Theta}}) - \mathsf{vec}(\boldsymbol{\theta}_0, \boldsymbol{\Theta}_0)\big] = \frac{1}{2\sqrt{N(n)}} \cdot \mathbf{W}^{+}\mathbf{Q}_n^{(1)}(\boldsymbol{\theta}_0, \boldsymbol{\Theta}_0)(1 + o_P(1)). \tag{B.15}$$

Hence, to prove (B.11) we only need to consider the term $\mathbf{Q}_n^{(1)}(\boldsymbol{\theta}_0, \boldsymbol{\Theta}_0)$. Note that

$$\begin{aligned} \frac{1}{2}\mathbf{Q}_n^{(1)}(\boldsymbol{\theta}_0, \boldsymbol{\Theta}_0) &= \sum_{i=1}^{n}\sum_{j=1}^{m_i} \overline{E}_{ij}(\boldsymbol{\theta}_0, \boldsymbol{\Theta}_0) \cdot \Big[\frac{\partial E_{ij}(\boldsymbol{\theta}, \boldsymbol{\Theta})}{\partial \mathsf{vec}(\boldsymbol{\theta}, \boldsymbol{\Theta})}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0, \boldsymbol{\Theta}=\boldsymbol{\Theta}_0}\Big] \\ &\quad + \sum_{i=1}^{n}\sum_{j=1}^{m_i} \widetilde{E}_{ij}(\boldsymbol{\theta}_0, \boldsymbol{\Theta}_0) \cdot \Big[\frac{\partial E_{ij}(\boldsymbol{\theta}, \boldsymbol{\Theta})}{\partial \mathsf{vec}(\boldsymbol{\theta}, \boldsymbol{\Theta})}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0, \boldsymbol{\Theta}=\boldsymbol{\Theta}_0}\Big] \\ &\equiv \overline{\mathbf{Q}}_n^{(1)}(\boldsymbol{\theta}_0, \boldsymbol{\Theta}_0) + \widetilde{\mathbf{Q}}_n^{(1)}(\boldsymbol{\theta}_0, \boldsymbol{\Theta}_0). \end{aligned}$$

By Lemma B.1 and some elementary calculations, we may show that

$$\begin{aligned} \frac{1}{\sqrt{N(n)}}\overline{\mathbf{Q}}_n^{(1)}(\boldsymbol{\theta}_0, \boldsymbol{\Theta}_0) &= \frac{1}{\sqrt{N(n)}}\sum_{i=1}^{n}\sum_{j=1}^{m_i} \overline{E}_{ij}(\boldsymbol{\theta}_0, \boldsymbol{\Theta}_0) \cdot \Big[\frac{\partial E_{ij}(\boldsymbol{\theta}, \boldsymbol{\Theta})}{\partial \mathsf{vec}(\boldsymbol{\theta}, \boldsymbol{\Theta})}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0, \boldsymbol{\Theta}=\boldsymbol{\Theta}_0}\Big] \\ &= \mathbf{V}_n + o_P(1), \end{aligned} \tag{B.16}$$

where $\mathbf{V}_n$ is defined in Section 3.

We next consider $\widetilde{\mathbf{Q}}_n^{(1)}(\boldsymbol{\theta}_0, \boldsymbol{\Theta}_0)$, which is more complicated than $\overline{\mathbf{Q}}_n^{(1)}(\boldsymbol{\theta}_0, \boldsymbol{\Theta}_0)$. By the definition of $\widetilde{\gamma}_k(\cdot|\boldsymbol{\theta}, \boldsymbol{\Theta})$, we have

$$\begin{aligned}
\widetilde{E}_{ij}(\boldsymbol{\theta}_0, \boldsymbol{\Theta}_0) &= \sum_{k=1}^{d_0} \big[\gamma_{k0}(U_{ij}) - \widehat{\gamma}_k(U_{ij}|\boldsymbol{\theta}_0, \boldsymbol{\Theta}_0)\big]\mathbf{X}_{ij}^{\tau}\boldsymbol{\Theta}_0(k) \\
&\quad + \sum_{k=1}^{d_0} \Big[\frac{1}{N(n)} \sum_{p=1}^{n}\sum_{q=1}^{m_p} \widehat{\gamma}_k(U_{pq}|\boldsymbol{\theta}_0, \boldsymbol{\Theta}_0)\Big]\mathbf{X}_{ij}^{\tau}\boldsymbol{\Theta}_0(k) \\
&\equiv \widetilde{E}_{ij,1}(\boldsymbol{\theta}_0, \boldsymbol{\Theta}_0) + \widetilde{E}_{ij,2}(\boldsymbol{\theta}_0, \boldsymbol{\Theta}_0). \qquad \text{(B.17)}
\end{aligned}$$

Following the proof of Lemma B.1 above, we have

$$\widetilde{E}_{ij,1}(\boldsymbol{\theta}_0, \boldsymbol{\Theta}_0) = -\frac{1}{2}\mu_2 h^2 \sum_{k=1}^{d_0} \gamma_{k0}''(U_{ij})\mathbf{X}_{ij}^{\tau}\boldsymbol{\Theta}_0(k) - \mathbf{T}_{n\varepsilon}^{\tau}(U_{ij}|\boldsymbol{\Theta}_0)\boldsymbol{\Delta}^{+}(U_{ij}|\boldsymbol{\Theta}_0)\mathbf{X}_{ij}(\boldsymbol{\Theta}_0).$$

This indicates that

$$\begin{aligned}
&\frac{1}{\sqrt{N(n)}} \sum_{i=1}^{n}\sum_{j=1}^{m_i} \widetilde{E}_{ij,1}(\boldsymbol{\theta}_0, \boldsymbol{\Theta}_0) \cdot \Big[\frac{\partial E_{ij}(\boldsymbol{\theta}, \boldsymbol{\Theta})}{\partial \mathsf{vec}(\boldsymbol{\theta}, \boldsymbol{\Theta})}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0, \boldsymbol{\Theta}=\boldsymbol{\Theta}_0}\Big] \\
= &-\frac{1}{\sqrt{N(n)}} \sum_{i=1}^{n}\sum_{j=1}^{m_i} \mathbf{T}_{n\varepsilon}^{\tau}(U_{ij}|\boldsymbol{\Theta}_0)\boldsymbol{\Delta}^{+}(U_{ij}|\boldsymbol{\Theta}_0)\mathbf{X}_{ij}(\boldsymbol{\Theta}_0) \cdot \Big[\frac{\partial E_{ij}(\boldsymbol{\theta}, \boldsymbol{\Theta})}{\partial \mathsf{vec}(\boldsymbol{\theta}, \boldsymbol{\Theta})}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0, \boldsymbol{\Theta}=\boldsymbol{\Theta}_0}\Big] + O_P(\sqrt{n}h^2) \\
\equiv &-\widetilde{\mathbf{Q}}_n^{(1,1)}(\boldsymbol{\theta}_0, \boldsymbol{\Theta}_0) + o_P(1) \qquad \text{(B.18)}
\end{aligned}$$

as $nh^4 = o(1)$ by using Assumption 1(ii). Noting that the dimension of the random vector $\widetilde{\mathbf{Q}}_n^{(1,1)}(\boldsymbol{\theta}_0, \boldsymbol{\Theta}_0)$ is $d(d_0+1)$, we define the d-dimensional sub-vector:

$$\widetilde{\mathbf{Q}}_n^{(1,1)}(k) = \frac{1}{\sqrt{N(n)}} \sum_{i=1}^{n}\sum_{j=1}^{m_i} \gamma_{k0}(U_{ij})\mathbf{T}_{n\varepsilon}^{\tau}(U_{ij}|\boldsymbol{\Theta}_0)\boldsymbol{\Delta}^{+}(U_{ij}|\boldsymbol{\Theta}_0)\mathbf{X}_{ij}(\boldsymbol{\Theta}_0)\widetilde{\mathbf{X}}_{ij}, \quad k = 0, 1, \ldots, d_0,$$

with $\gamma_{00}(\cdot) \equiv 1$, such that

$$\widetilde{\mathbf{Q}}_n^{(1,1)}(\boldsymbol{\theta}_0, \boldsymbol{\Theta}_0) = \Big\{\big[\widetilde{\mathbf{Q}}_n^{(1,1)}(0)\big]^{\tau}, \big[\widetilde{\mathbf{Q}}_n^{(1,1)}(1)\big]^{\tau}, \ldots, \big[\widetilde{\mathbf{Q}}_n^{(1,1)}(d_0)\big]^{\tau}\Big\}^{\tau} + o_P(1).$$

By the definition of $\mathbf{T}_{n\varepsilon}(U_{ij}|\boldsymbol{\Theta}_0)$, we may show that, for $k = 0, 1, \ldots, d_0$,

$$\widetilde{\mathbf{Q}}_n^{(1,1)}(k) = \frac{1}{[N(n)]^{3/2}h} \sum_{p=1}^{n}\sum_{q=1}^{m_p}\sum_{i=1}^{n}\sum_{j=1}^{m_i} \widetilde{v}_{k,ij}\overline{v}_{pq} K\big(\frac{U_{pq} - U_{ij}}{h}\big),$$

where

$$\widetilde{v}_{k,ij} = \gamma_{k0}(U_{ij})\widetilde{\mathbf{X}}_{ij}\mathbf{X}_{ij}^{\tau}(\boldsymbol{\Theta}_0)\big[f(U_{ij})\boldsymbol{\Delta}(U_{ij}|\boldsymbol{\Theta}_0)\big]^{+}, \quad \overline{v}_{pq} = \mathbf{X}_{pq}(\boldsymbol{\Theta}_0)\varepsilon_{pq}.$$

By the definition of $\widetilde{\mathbf{X}}_{ij}$ as well as Assumptions 1 and 2, it is easy to see that

$$\mathsf{E}\left[\widetilde{v}_{k,ij}K(\frac{U_{pq}-U_{ij}}{h})|U_{ij}=U_{pq}\right]=h\gamma_{k0}(U_{pq})\mathsf{E}\big[\mathbf{\Delta}_{\mathbf{X}}(U_{pq}|\mathbf{\Theta}_0)\mathbf{\Delta}^{+}(U_{pq}|\mathbf{\Theta}_0)\big](1+o_P(1)).$$

Then, letting $\overline{\mathbf{X}}_{pq}=\mathsf{E}\big[\mathbf{\Delta}_{\mathbf{X}}(U_{pq}|\mathbf{\Theta}_0)\mathbf{\Delta}^{+}(U_{pq}|\mathbf{\Theta}_0)\big]\mathbf{X}_{pq}(\mathbf{\Theta}_0)$, by some standard calculations, we may show that

$$\widetilde{\mathbf{Q}}_n^{(1,1)}(\boldsymbol{\theta}_0,\mathbf{\Theta}_0)=\frac{1}{\sqrt{N(n)}}\cdot\sum_{p=1}^{n}\sum_{q=1}^{m_p}\begin{bmatrix}\overline{\mathbf{X}}_{pq}\varepsilon_{pq}\\ \gamma_{10}(U_{pq})\overline{\mathbf{X}}_{pq}\varepsilon_{pq}\\ \vdots\\ \gamma_{d_00}(U_{pq})\overline{\mathbf{X}}_{pq}\varepsilon_{pq}\end{bmatrix}+o_P(1), \tag{B.19}$$

For $\widetilde{E}_{ij,2}(\boldsymbol{\theta}_0,\mathbf{\Theta}_0)$, note that

$$\widetilde{E}_{ij,2}(\boldsymbol{\theta}_0,\mathbf{\Theta}_0)=\widetilde{E}_{ij,21}(\boldsymbol{\theta}_0,\mathbf{\Theta}_0)+\widetilde{E}_{ij,22}(\boldsymbol{\theta}_0,\mathbf{\Theta}_0),$$

where

$$\begin{aligned}\widetilde{E}_{ij,21}(\boldsymbol{\theta}_0,\mathbf{\Theta}_0) &= \sum_{k=1}^{d_0}\Big\{\frac{1}{N(n)}\sum_{p=1}^{n}\sum_{q=1}^{m_p}\big[\widehat{\gamma}_k(U_{pq}|\boldsymbol{\theta}_0,\mathbf{\Theta}_0)-\gamma_{k0}(U_{pq})\big]\Big\}\mathbf{X}_{ij}^{\tau}\mathbf{\Theta}_0(k),\\ \widetilde{E}_{ij,22}(\boldsymbol{\theta}_0,\mathbf{\Theta}_0) &= \sum_{k=1}^{d_0}\Big[\frac{1}{N(n)}\sum_{p=1}^{n}\sum_{q=1}^{m_p}\gamma_{k0}(U_{pq})\Big]\mathbf{X}_{ij}^{\tau}\mathbf{\Theta}_0(k).\end{aligned}$$

Let

$$\widetilde{\mathbf{Q}}_n^{(1,2)}(\boldsymbol{\theta}_0,\mathbf{\Theta}_0)=\frac{1}{\sqrt{N(n)}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\widetilde{E}_{ij,21}(\boldsymbol{\theta}_0,\mathbf{\Theta}_0)\cdot\Big[\frac{\partial E_{ij}(\boldsymbol{\theta},\mathbf{\Theta})}{\partial\mathsf{vec}(\boldsymbol{\theta},\mathbf{\Theta})}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0,\mathbf{\Theta}=\mathbf{\Theta}_0}\Big].$$

Similar to the argument on $\widetilde{\mathbf{Q}}_n^{(1,1)}(\mathbf{\Theta}_0)$, we define the d-dimensional sub-vector:

$$\widetilde{\mathbf{Q}}_n^{(1,2)}(k)=\frac{1}{[N(n)]^{3/2}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\gamma_{k0}(U_{ij})\Big\{\sum_{l=1}^{d_0}\sum_{p=1}^{n}\sum_{q=1}^{m_p}\big[\widehat{\gamma}_l(U_{pq}|\boldsymbol{\theta}_0,\mathbf{\Theta}_0)-\gamma_{l0}(U_{pq})\big]\mathbf{X}_{ij}^{\tau}\mathbf{\Theta}_0(l)\Big\}\widetilde{\mathbf{X}}_{ij}$$

for $k=0,1,\ldots,d_0$ such that

$$\widetilde{\mathbf{Q}}_n^{(1,2)}(\boldsymbol{\theta}_0,\mathbf{\Theta}_0)=\Big\{\big[\widetilde{\mathbf{Q}}_n^{(1,2)}(0)\big]^{\tau},\big[\widetilde{\mathbf{Q}}_n^{(1,2)}(1)\big]^{\tau},\ldots,\big[\widetilde{\mathbf{Q}}_n^{(1,2)}(d_0)\big]^{\tau}\Big\}^{\tau}.$$

Then, we may show that

$$\begin{aligned}
\widetilde{\mathbf{Q}}_n^{(1,2)}(k) &\overset{P}{\sim} \frac{1}{[N(n)]^{3/2}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\gamma_{k0}(U_{ij})\widetilde{\mathbf{X}}_{ij}\mathbf{X}_{ij}^{\tau}(\mathbf{\Theta}_0)\Big\{\sum_{p=1}^{n}\sum_{q=1}^{m_p}\mathbf{\Delta}^{+}(U_{pq}|\mathbf{\Theta}_0)\mathbf{T}_{n\varepsilon}(U_{pq}|\mathbf{\Theta}_0)\Big\} \\
&= \frac{1}{[N(n)]^{5/2}h}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\gamma_{k0}(U_{ij})\widetilde{\mathbf{X}}_{ij}\mathbf{X}_{ij}^{\tau}(\mathbf{\Theta}_0)\Big\{\sum_{p=1}^{n}\sum_{q=1}^{m_p}\sum_{s=1}^{n}\sum_{t=1}^{m_s}\big[f(U_{pq})\mathbf{\Delta}(U_{pq}|\mathbf{\Theta}_0)\big]^{+} \\
&\quad \times\mathbf{X}_{st}(\mathbf{\Theta}_0)\varepsilon_{st}K\Big(\frac{U_{st}-U_{pq}}{h}\Big)\Big\} \\
&\overset{P}{\sim} \frac{1}{[N(n)]^{3/2}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\gamma_{k0}(U_{ij})\widetilde{\mathbf{X}}_{ij}\mathbf{X}_{ij}^{\tau}(\mathbf{\Theta}_0)\Big\{\sum_{s=1}^{n}\sum_{t=1}^{m_s}\mathbf{\Delta}^{+}(U_{st}|\mathbf{\Theta}_0)\mathbf{X}_{st}(\mathbf{\Theta}_0)\varepsilon_{st}\Big\} \\
&= \frac{1}{[N(n)]^{3/2}}\sum_{s=1}^{n}\sum_{t=1}^{m_s}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\widetilde{u}_{k,ij}\overline{u}_{st},
\end{aligned}$$

where

$$\widetilde{u}_{k,ij} = \gamma_{k0}(U_{ij})\widetilde{\mathbf{X}}_{ij}\mathbf{X}_{ij}^{\tau}(\mathbf{\Theta}_0), \quad \overline{u}_{st} = \mathbf{\Delta}^{+}(U_{st}|\mathbf{\Theta}_0)\mathbf{X}_{st}(\mathbf{\Theta}_0)\varepsilon_{st}$$

By the Law of Large Numbers, we have

$$\frac{1}{N(n)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\widetilde{u}_{k,ij} = \mathsf{E}\big[\mathbf{\Delta}_{\mathbf{X}}(U_{st}|\mathbf{\Theta}_0)\mathbf{\Delta}^{+}(U_{st}|\mathbf{\Theta}_0)\big]\mathsf{E}\big[\gamma_{k0}(U_{st})\mathbf{\Delta}(U_{st}|\mathbf{\Theta}_0)\big] + o_P(1),$$

which indicates that

$$\widetilde{\mathbf{Q}}_n^{(1,2)}(\boldsymbol{\theta}_0,\mathbf{\Theta}_0) = \frac{1}{\sqrt{N(n)}}\cdot\sum_{s=1}^{n}\sum_{t=1}^{m_s}\begin{bmatrix}\widehat{\mathbf{X}}_{st}(0)\varepsilon_{st}\\ \widehat{\mathbf{X}}_{st}(1)\varepsilon_{st}\\ \vdots \\ \widehat{\mathbf{X}}_{st}(d_0)\varepsilon_{st}\end{bmatrix} + o_P(1), \tag{B.20}$$

where $\widehat{\mathbf{X}}_{st}(k) = \mathsf{E}\big[\mathbf{\Delta}_{\mathbf{X}}(U_{st}|\mathbf{\Theta}_0)\mathbf{\Delta}^{+}(U_{st}|\mathbf{\Theta}_0)\big]\mathsf{E}\big[\gamma_{k0}(U_{st})\mathbf{\Delta}(U_{st}|\mathbf{\Theta}_0)\big]\mathbf{\Delta}^{+}(U_{st}|\mathbf{\Theta}_0)\mathbf{X}_{st}(\mathbf{\Theta}_0)$.

Let

$$\widetilde{\mathbf{Q}}_n^{(1,3)}(\boldsymbol{\theta}_0,\mathbf{\Theta}_0) = \frac{1}{\sqrt{N(n)}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\widetilde{E}_{ij,22}(\boldsymbol{\theta}_0,\mathbf{\Theta}_0)\cdot\Big[\frac{\partial E_{ij}(\boldsymbol{\theta},\mathbf{\Theta})}{\partial\mathsf{vec}(\boldsymbol{\theta},\mathbf{\Theta})}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0,\mathbf{\Theta}=\mathbf{\Theta}_0}\Big].$$

Similar to the above arguments, we also define the d-dimensional sub-vector

$$\widetilde{\mathbf{Q}}_n^{(1,3)}(k) = \frac{1}{\sqrt{N(n)}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\gamma_{k0}(U_{ij})\Big\{\sum_{l=1}^{d_0}\Big[\frac{1}{N(n)}\sum_{p=1}^{n}\sum_{q=1}^{m_p}\gamma_{l0}(U_{pq})\Big]\mathbf{X}_{ij}^{\tau}\mathbf{\Theta}_0(l)\Big\}\widetilde{\mathbf{X}}_{ij}$$

such that

$$\widetilde{\mathbf{Q}}_n^{(1,3)}(\boldsymbol{\theta}_0,\mathbf{\Theta}_0) = \Big\{\big[\widetilde{\mathbf{Q}}_n^{(1,3)}(0)\big]^{\tau},\big[\widetilde{\mathbf{Q}}_n^{(1,3)}(1)\big]^{\tau},\ldots,\big[\widetilde{\mathbf{Q}}_n^{(1,3)}(d_0)\big]^{\tau}\Big\}^{\tau}.$$

Following the argument in the proof of (B.20), we have

$$\begin{aligned}
\widetilde{\mathbf{Q}}_n^{(1,3)}(k) &= \frac{1}{[N(n)]^{3/2}}\sum_{i=1}^n\sum_{j=1}^{m_i}\gamma_{k0}(U_{ij})\widetilde{\mathbf{X}}_{ij}\Big\{\sum_{l=1}^{d_0}\Big[\sum_{p=1}^n\sum_{q=1}^{m_p}\gamma_{k0}(U_{pq})\Big]\mathbf{X}_{ij}^{\tau}\mathbf{\Theta}_0(l)\Big\}\mathbf{X}_{ij} \\
&= \frac{1}{[N(n)]^{3/2}}\sum_{p=1}^n\sum_{q=1}^{m_p}\sum_{i=1}^n\sum_{j=1}^{m_i}\gamma_{k0}(U_{ij})\widetilde{\mathbf{X}}_{ij}\mathbf{X}_{ij}^{\tau}(\mathbf{\Theta}_0)\boldsymbol{\gamma}_0(U_{pq}), \\
&\overset{P}{\sim} \frac{1}{\sqrt{N(n)}}\sum_{p=1}^n\sum_{q=1}^{m_p}\mathsf{E}\big[\mathbf{\Delta}_{\mathbf{X}}(U_{pq}|\mathbf{\Theta}_0)\mathbf{\Delta}^+(U_{pq}|\mathbf{\Theta}_0)\big]\mathsf{E}\big[\gamma_{k0}(U_{pq})\mathbf{\Delta}(U_{pq}|\mathbf{\Theta}_0)\big]\boldsymbol{\gamma}_0(U_{pq}), \\
&\equiv \frac{1}{\sqrt{N(n)}}\sum_{p=1}^n\sum_{q=1}^{m_p}\widehat{\mathbf{\Delta}}_k\boldsymbol{\gamma}_0(U_{pq}),
\end{aligned}$$

which indicates that

$$\widetilde{\mathbf{Q}}_n^{(1,3)}(\boldsymbol{\theta}_0,\mathbf{\Theta}_0) = \frac{1}{\sqrt{N(n)}}\cdot\sum_{p=1}^n\sum_{q=1}^{m_p}\begin{bmatrix}\widehat{\mathbf{\Delta}}_0\boldsymbol{\gamma}_0(U_{pq})\\ \widehat{\mathbf{\Delta}}_1\boldsymbol{\gamma}_0(U_{pq})\\ \vdots \\ \widehat{\mathbf{\Delta}}_{d_0}\boldsymbol{\gamma}_0(U_{pq})\end{bmatrix} + o_P(1). \tag{B.21}$$

By (B.15), (B.16), (B.19)–(B.21), we can complete the proof of (B.11). ■

Proof of Proposition 3.1. By Lemma B.2 and then applying the classical central limit theorem on $\mathbf{V}_n$, we readily prove (3.1). By Lemma B.1 and (3.1), we have

$$\begin{aligned}
\widehat{\boldsymbol{\gamma}}(u)-\boldsymbol{\gamma}_0(u) &= \frac{1}{2}\mu_2h^2\boldsymbol{\gamma}_0''(u)+\mathbf{\Delta}^+(u|\mathbf{\Theta}_0)\mathbf{T}_{n\varepsilon}(u|\mathbf{\Theta}_0)+O_P\big(n^{-1/2}+h^3+h\xi_n+\xi_n^2\big), \\
&= \frac{1}{2}\mu_2h^2\boldsymbol{\gamma}_0''(u)+\mathbf{\Delta}^+(u|\mathbf{\Theta}_0)\mathbf{T}_{n\varepsilon}(u|\mathbf{\Theta}_0)+o_P\big((nh)^{-1/2}\big).
\end{aligned}\tag{B.22}$$

Using the central limit theorem, we may show that

$$\begin{aligned}
\sqrt{N(n)h}\mathbf{T}_{n\varepsilon}(u|\mathbf{\Theta}_0) &= \frac{1}{\sqrt{N(n)}}\sum_{i=1}^n\Big[\frac{1}{f(u)\sqrt{h}}\sum_{j=1}^{m_i}\mathbf{X}_{ij}(\mathbf{\Theta}_0)\varepsilon_{ij}K\Big(\frac{U_{ij}-u}{h}\Big)\Big] \\
&\xrightarrow{d} \mathsf{N}\big(\mathbf{0}_{d_0},\ \omega(u)\mathbf{\Delta}(u|\mathbf{\Theta}_0)\big).
\end{aligned}\tag{B.23}$$

We then complete the proof of (3.2) by using (B.22) and (B.23).

The proof of Proposition 3.1 has thus been completed. ■

Proof of Theorem 3.2. Define

$$\mathbf{S}_{nl}^{\diamond}(u|\mathbf{\Theta}) = \frac{1}{N(n)h}\sum_{i=1}^n\sum_{j=1}^{m_i}\Big(\frac{U_{ij}-u}{h}\Big)^l\mathbf{X}_{ij}(\mathbf{\Theta})\mathbf{X}_{ij}^{\tau}(\mathbf{\Theta})\big(\rho_{ij}^{\diamond}\big)^{-1}K\Big(\frac{U_{ij}-u}{h}\Big),$$

$$\mathbf{T}_{nl}^{\diamond}(u|\boldsymbol{\theta},\mathbf{\Theta}) = \frac{1}{N(n)h}\sum_{i=1}^n\sum_{j=1}^{m_i}\Big(\frac{U_{ij}-u}{h}\Big)^l\mathbf{X}_{ij}(\mathbf{\Theta})\big[\widetilde{Y}_{ij}-\mathbf{X}_{ij}^{\tau}\boldsymbol{\theta}\big]\big(\rho_{ij}^{\diamond}\big)^{-1}K\Big(\frac{U_{ij}-u}{h}\Big)$$

and

$$\mathbf{S}_n^\diamond(u|\mathbf{\Theta}) = \begin{bmatrix} \mathbf{S}_{n0}^\diamond(u|\mathbf{\Theta}) & \mathbf{S}_{n1}^\diamond(u|\mathbf{\Theta}) \\ \mathbf{S}_{n1}^\diamond(u|\mathbf{\Theta}) & \mathbf{S}_{n2}^\diamond(u|\mathbf{\Theta}) \end{bmatrix}, \quad \mathbf{T}_n^\diamond(u|\boldsymbol{\theta},\mathbf{\Theta}) = \begin{bmatrix} \mathbf{T}_{n0}^\diamond(u|\boldsymbol{\theta},\mathbf{\Theta}) \\ \mathbf{T}_{n1}^\diamond(u|\boldsymbol{\theta},\mathbf{\Theta}) \end{bmatrix}.$$

As in the proof of Lemma B.1, we readily have

$$\overline{\gamma}(u) = \big(\mathbf{I}_{d_0},\ \mathbf{N}_{d_0}\big)\big[\mathbf{S}_n^\diamond(u|\overline{\mathbf{\Theta}})\big]^+\mathbf{T}_n^\diamond(u|\overline{\boldsymbol{\theta}},\overline{\mathbf{\Theta}}). \tag{B.24}$$

As $\overline{\mathbf{\Theta}}$ is assumed to be root-n consistent, we may show that

$$\mathbf{S}_n^\diamond(u|\overline{\mathbf{\Theta}}) = \mathbf{S}_n^\diamond(u|\mathbf{\Theta}_0) + O_P(n^{-1/2}) = \mathbf{\Delta}_n^\diamond(u|\mathbf{\Theta}_0) + O_P(n^{-1/2}+h), \tag{B.25}$$

where

$$\begin{aligned} \mathbf{\Delta}_n^\diamond(u|\mathbf{\Theta}) &= \mathsf{diag}(1,\mu_2)\otimes\frac{f(u)}{N(n)}\sum_{i=1}^n\sum_{j=1}^{m_i}\mathsf{E}\big[\mathbf{X}_{ij}(\mathbf{\Theta})\mathbf{X}_{ij}^\tau(\mathbf{\Theta})\big(\rho_{ij}^\diamond\big)^{-1}|U_{ij}=u\big] \\ &= f(u)\Big[\frac{1}{N(n)}\sum_{i=1}^n\sum_{j=1}^{m_i}\big(\rho_{ij}^\diamond\big)^{-1}\Big]\cdot\mathsf{diag}(1,\mu_2)\otimes\mathbf{\Delta}(u|\mathbf{\Theta}_0). \end{aligned}$$

By (2.14) and (B.2), we have

$$\begin{aligned} \widetilde{Y}_{ij}-\mathbf{X}_{ij}^\tau\overline{\boldsymbol{\theta}} &= \eta_{ij}+\mathbf{X}_{ij}^\tau\big(\boldsymbol{\theta}_0-\overline{\boldsymbol{\theta}}\big)+\sum_{k=1}^{d_0}\gamma_{k0}(U_{ij})\mathbf{X}_{ij}^\tau\mathbf{\Theta}_0(k) \\ &\quad+\sum_{k=1}^{j-1}(c_{i,jk}-c_{i,jk}^\diamond)\varepsilon_{ik}+\sum_{k=1}^{j-1}c_{i,jk}^\diamond(\varepsilon_{ik}-\widetilde{\varepsilon}_{ik}), \end{aligned} \tag{B.26}$$

where $\sum_{k=1}^0\cdot\equiv 0$. Let $\mathbf{T}_n^\diamond(u,1)$, $\mathbf{T}_n^\diamond(u,2)$, $\mathbf{T}_n^\diamond(u,3)$, $\mathbf{T}_n^\diamond(u,4)$ and $\mathbf{T}_n^\diamond(u,5)$ be defined as $\mathbf{T}_n^\diamond(u|\overline{\boldsymbol{\theta}},\overline{\mathbf{\Theta}})$ with $\widetilde{Y}_{ij}-\mathbf{X}_{ij}^\tau\overline{\boldsymbol{\theta}}$ replaced by η_{ij}, $\mathbf{X}_{ij}^\tau\big(\boldsymbol{\theta}_0-\overline{\boldsymbol{\theta}}\big)$, $\sum_{k=1}^{d_0}\gamma_{k0}(U_{ij})\mathbf{X}_{ij}^\tau\mathbf{\Theta}_0(k)$, $\sum_{k=1}^{j-1}(c_{i,jk}-c_{i,jk}^\diamond)\varepsilon_{ik}$ and $\sum_{k=1}^{j-1}c_{i,jk}^\diamond(\varepsilon_{ik}-\widetilde{\varepsilon}_{ik})$, respectively. Then, by (B.24)–(B.26), we can show that

$$\begin{aligned} \overline{\gamma}(u) &= \big(\mathbf{I}_{d_0},\ \mathbf{N}_{d_0}\big)\big[\mathbf{S}_n^\diamond(u|\overline{\mathbf{\Theta}})\big]^+\Big[\sum_{k=1}^5\mathbf{T}_n^\diamond(u,k)\Big] \\ &= \big(\mathbf{I}_{d_0},\ \mathbf{N}_{d_0}\big)\big[\mathbf{S}_n^\diamond(u|\mathbf{\Theta}_0)\big]^+\Big[\sum_{k=1}^5\mathbf{T}_n^\diamond(u,k)\Big]+O_P(n^{-1/2}). \end{aligned} \tag{B.27}$$

As $\overline{\boldsymbol{\theta}}$ and $\overline{\mathbf{\Theta}}$ are root-n consistent, we can prove that

$$\big(\mathbf{I}_{d_0},\ \mathbf{N}_{d_0}\big)\big[\mathbf{S}_n^\diamond(u|\mathbf{\Theta}_0)\big]^+\mathbf{T}_n^\diamond(u,2) = O_P(n^{-1/2}) = o_P\big((nh)^{-1/2}\big). \tag{B.28}$$

By some standard calculations, we can also prove that

$$\big(\mathbf{I}_{d_0},\ \mathbf{N}_{d_0}\big)\big[\mathbf{S}_n^\diamond(u|\mathbf{\Theta}_0)\big]^+\mathbf{T}_n^\diamond(u,5) = o_P\big((nh)^{-1/2}\big). \tag{B.29}$$

The proof of (B.29) will be given later in this appendix. Similar to the proof of (B.9), we have

$$\big(\mathbf{I}_{d_0},\ \mathbf{N}_{d_0}\big)\big[\mathbf{S}_n^{\diamond}(u|\boldsymbol{\Theta}_0)\big]^{+}\mathbf{T}_n^{\diamond}(u,3)=\boldsymbol{\gamma}_0(u)+\frac{1}{2}\mu_2h^2\boldsymbol{\gamma}_0''(u)+o_P\big(h^2+(nh)^{-1/2}\big). \tag{B.30}$$

Let $e_{ij}=\eta_{ij}+\sum_{k=1}^{j-1}(c_{i,jk}-c_{i,jk}^{\diamond})\varepsilon_{ik}$. By the central limit theorem, we can show that

$$\frac{1}{\sqrt{N(n)h}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\mathbf{X}_{ij}(\boldsymbol{\Theta}_0)e_{ij}\big(\rho_{ij}^{\diamond}\big)^{-1}K\Big(\frac{U_{ij}-u}{h}\Big)\xrightarrow{d}\mathsf{N}\big(\mathbf{0},\ \boldsymbol{\Omega}^{\diamond}(u|\boldsymbol{\Theta}_0)\big),$$

where

$$\boldsymbol{\Omega}^{\diamond}(u|\boldsymbol{\Theta}_0)=f(u)\nu_0\boldsymbol{\Delta}(u|\boldsymbol{\Theta}_0)\lim_{n\to\infty}\frac{1}{N(n)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{\tau_{ij}}{\big(\rho_{ij}^{\diamond}\big)^2},\qquad \tau_{ij}=\mathsf{E}[e_{ij}^2].$$

This indicates

$$\sqrt{N(n)h}\big(\mathbf{I}_{d_0},\ \mathbf{N}_{d_0}\big)\big[\mathbf{S}_n^{\diamond}(u|\boldsymbol{\Theta}_0)\big]^{+}\big[\mathbf{T}_n^{\diamond}(u,1)+\mathbf{T}_n^{\diamond}(u,4)\big]\xrightarrow{d}\mathsf{N}\big(\mathbf{0}_{d_0},\ \omega^{\diamond}(u)\boldsymbol{\Delta}^{-1}(u|\boldsymbol{\Theta}_0)\big), \tag{B.31}$$

where $\omega^{\diamond}(u)$ is defined in (3.4). By (B.27)–(B.31), we complete the proof of (3.5). ■

We next give the proof of (B.29), which shows that the influence of replacing ε_{ik} by $\widetilde{\varepsilon}_{ik}$ can be ignored asymptotically.

Proof of (B.29). Recall that $\widetilde{\boldsymbol{\beta}}(\cdot)$ is a local linear estimation of $\boldsymbol{\beta}_0(\cdot)$ with the kernel function $K(\cdot)$ and bandwidth b, i.e.,

$$\widetilde{\boldsymbol{\beta}}(u)=\big(\mathbf{I}_d,\ \mathbf{N}_d\big)\mathbf{S}_n^{+}(u)\mathbf{T}_n(u), \tag{B.32}$$

where $\mathbf{S}_n(\cdot)$ and $\mathbf{T}_n(\cdot)$ are defined similar to $\mathbf{S}_n(\cdot|\boldsymbol{\Theta})$and $\mathbf{T}_n(u|\boldsymbol{\theta},\boldsymbol{\Theta})$ with $\mathbf{X}_{ij}(\boldsymbol{\Theta})$, $Y_{ij}-\mathbf{X}_{ij}^{\tau}\boldsymbol{\theta}$ and h replaced by $\mathbf{X}_{ij}$, Y_{ij} and b, respectively. Hence,

$$\varepsilon_{ik}-\widetilde{\varepsilon}_{ik}=\mathbf{X}_{ik}^{\tau}\big[\widetilde{\boldsymbol{\beta}}(U_{ik})-\boldsymbol{\beta}_0(U_{ik})\big]$$

which indicates that

$$\mathbf{T}_{nl}^{\diamond}(u,5)=\frac{1}{N(n)h}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left(\frac{U_{ij}-u}{h}\right)^{l}\mathbf{X}_{ij}(\overline{\boldsymbol{\Theta}})\big(\rho_{ij}^{\diamond}\big)^{-1}K\left(\frac{U_{ij}-u}{h}\right)\sum_{k=1}^{j-1}c_{i,jk}^{\diamond}\mathbf{X}_{ik}^{\tau}\left[\widetilde{\boldsymbol{\beta}}(U_{ik})-\boldsymbol{\beta}_0(U_{ik})\right] \tag{B.33}$$

for $l=0$ and $l=1$, where $\mathbf{T}_{n0}^{\diamond}(u,5)$ is the first d_0 elements in $\mathbf{T}_n^{\diamond}(u,5)$, and $\mathbf{T}_{n1}^{\diamond}(u,5)$ is the last d_0 elements in $\mathbf{T}_n^{\diamond}(u,5)$.

In order to complete the proof of (B.29), we only need to show that

$$\mathbf{T}_{nl}^{\diamond}(u,5)=o_P\big((nh)^{-1/2}\big),\quad l=0,1. \tag{B.34}$$

We next only prove (B.34) for the case of $l=0$ as the case of $l=1$ can be proved in the same way. As $\overline{\boldsymbol{\Theta}}$ is root-n consistent, we may replace $\overline{\boldsymbol{\Theta}}$ by $\boldsymbol{\Theta}_0$ in this proof. Furthermore, noting that

$$\sup_{u\in\mathcal{U}} \|\mathbf{S}_n(u) - \mathsf{diag}(1,\mu_2)\otimes f(u)\boldsymbol{\Delta}(u)\| = O_P(b+\xi_n^*)$$

where $\boldsymbol{\Delta}(u)$ is defined in Assumption 4(i) and $\xi_n^* = \left(\frac{\log b^{-1}}{nb}\right)^{1/2}$, we may show that

$$\widetilde{\boldsymbol{\beta}}(U_{ik}) - \boldsymbol{\beta}_0(U_{ik}) = \frac{1+O_P(b+\xi_n^*)}{N(n)bf(U_{ik})}\sum_{p=1}^{n}\sum_{q=1}^{m_p}\varepsilon_{pq}\boldsymbol{\Delta}^+(U_{ik})\mathbf{X}_{pq}K\left(\frac{U_{pq}-U_{ik}}{b}\right) + O_P(b^2).$$

Then, by (B.33) and Assumption 4(ii), we have

$$\begin{aligned}
\mathbf{T}_{n0}^{\diamond}(u,5) &= \frac{1}{N^2(n)hb}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\mathbf{X}_{ij}(\boldsymbol{\Theta}_0)K\left(\frac{U_{ij}-u}{h}\right)\sum_{k=1}^{j-1}c_{i,jk}^{\diamond}\sum_{p=1}^{n}\sum_{q=1}^{m_p}\varepsilon_{pq}\mathbf{X}_{ik}^{\tau} \\
&\quad \times f^{-1}(U_{ik})\boldsymbol{\Delta}^+(U_{ik})\mathbf{X}_{pq}K\left(\frac{U_{pq}-U_{ik}}{b}\right) + O_P\big(b^2+(b+\xi_n^*)\xi_n^*\big) \\
&= \frac{1}{N^2(n)hb}\sum_{p=1}^{n}\sum_{q=1}^{m_p}\varepsilon_{pq}\nu_{pq}(u) + o_P\big(h^2+(nh)^{-1/2}\big), \qquad \text{(B.35)}
\end{aligned}$$

where

$$\nu_{pq}(u) = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{j-1}c_{i,jk}^{\diamond}\mathbf{X}_{ij}(\boldsymbol{\Theta}_0)\mathbf{X}_{ik}^{\tau}f^{-1}(U_{ik})\boldsymbol{\Delta}^+(U_{ik})\mathbf{X}_{pq}K\left(\frac{U_{ij}-u}{h}\right)K\left(\frac{U_{pq}-U_{ik}}{b}\right).$$

To evaluate the asymptotic order of $\mathbf{T}_{n0}^{\diamond}(u,5)$, we next calculate the order for the variance of $\sum_{p=1}^{n}\sum_{q=1}^{m_p}\varepsilon_{pq}\nu_{pq}(u)$. Note that

$$\begin{aligned}
\mathsf{E}\Big\{\Big[\sum_{p=1}^{n}\sum_{q=1}^{m_p}\varepsilon_{pq}\nu_{pq}(u)\Big]^2\Big\} &= \sum_{p=1}^{n}\sum_{q_1=1}^{m_p}\sum_{q_2=1}^{m_p}\mathsf{E}\Big[\varepsilon_{pq_1}\varepsilon_{pq_2}\nu_{pq_1}(u)\nu_{pq_2}(u)\Big] \\
&= O\left(\sum_{p=1}^{n}\sum_{q_1=1}^{m_p}\sum_{q_2=1}^{m_p}\sum_{i_1=1}^{n}\sum_{i_2=1,\neq i_1}^{n}\sum_{j_1=1}^{m_{i_1}}\sum_{j_2=1}^{m_{i_2}}\sum_{k_1=1}^{j_1-1}\sum_{k_2=1}^{j_2-1}\mathsf{E}\left[K\left(\frac{U_{i_1j_1}-u}{h}\right)\right.\right. \\
&\qquad \left.\left. K\left(\frac{U_{pq_1}-U_{i_1k_1}}{b}\right)K\left(\frac{U_{i_2j_2}-u}{h}\right)K\left(\frac{U_{pq_2}-U_{i_2k_2}}{b}\right)\right]\right) \\
&\quad +O\left(\sum_{p=1}^{n}\sum_{q_1=1}^{m_p}\sum_{q_2=1}^{m_p}\sum_{i_1=1}^{n}\sum_{j_1=1}^{m_{i_1}}\sum_{j_2=1}^{m_{i_1}}\sum_{k_1=1}^{j_1-1}\sum_{k_2=1}^{j_2-1}\mathsf{E}\left[K\left(\frac{U_{i_1j_1}-u}{h}\right)\right.\right. \\
&\qquad \left.\left. K\left(\frac{U_{pq_1}-U_{i_1k_1}}{b}\right)K\left(\frac{U_{i_1j_2}-u}{h}\right)K\left(\frac{U_{pq_2}-U_{i_1k_2}}{b}\right)\right]\right) \\
&= O(n^3h^2b^2+n^2hb),
\end{aligned}$$

which indicates that

$$\mathbf{T}_{n0}^{\diamond}(u,5) = O_P\big(n^{-1/2} + (n^2hb)^{-1/2}\big) = o_P\big((nh)^{-1/2}\big).$$

Therefore, we complete the proofs of (B.34) and (B.29). ■

In order to prove Proposition 4.1, we need to use the following two technical lemmas, which are similar to Lemmas B.1 and B.2 above. Define

$$\mathbf{T}_{n,\eta}(u|\mathbf{\Theta}) = \frac{1}{n(m-1)hf(u)}\sum_{i=1}^{n}\sum_{j=2}^{m}\mathbf{X}_{ij}(\mathbf{\Theta})\eta_{ij}K\Big(\frac{U_{ij}-u}{h}\Big)$$

and

$$\widetilde{\mathbf{T}}_{n,\varepsilon}(u|\mathbf{\Theta},\mathbf{C}) = \frac{1}{n(m-1)hf(u)}\sum_{i=1}^{n}\sum_{j=2}^{m}\mathbf{X}_{ij}(\mathbf{\Theta})K\Big(\frac{U_{ij}-u}{h}\Big)\sum_{k=1}^{j-1}c_{jk}(\varepsilon_{ik}-\widetilde{\varepsilon}_{ik}).$$

Lemma B.3. *Suppose that the conditions in Proposition 4.1 are satisfied. Then we have*

$$\begin{aligned}\widetilde{\boldsymbol{\gamma}}(u|\boldsymbol{\theta},\mathbf{\Theta},\mathbf{C}) - \boldsymbol{\gamma}_0(u) &= \frac{1}{2}\mu_2 h_*^2\boldsymbol{\gamma}_0''(u) + \mathbf{\Delta}^{+}(u|\mathbf{\Theta})\big[\mathbf{\Delta}_{1\mathbf{\Theta}}(u|\boldsymbol{\theta}_0-\boldsymbol{\theta}) + \mathbf{\Delta}_{2\mathbf{\Theta}}(u|\mathbf{\Theta}_0-\mathbf{\Theta})\boldsymbol{\gamma}_0(u)\big]\\ &\quad + \mathbf{\Delta}^{+}(u|\mathbf{\Theta})\big[\mathbf{T}_{n,\eta}(u|\mathbf{\Theta}) + \widetilde{\mathbf{T}}_{n,\varepsilon}(u|\mathbf{\Theta},\mathbf{C})\big] + o_P\big(\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| + \|\mathbf{\Theta}-\mathbf{\Theta}_0\|\\ &\quad + \|\mathbf{C}-\mathbf{C}_0\|\big) + O_P\big(h_*^3 + h_*\xi_{n*} + \xi_{n*}^2\big) \qquad \text{(B.36)}\end{aligned}$$

uniformly for $u \in \mathcal{U}$, where $\xi_{n*} = \big(\frac{\log h_*^{-1}}{nh_*}\big)^{1/2}$ and the remaining notation is the same as that in Lemma B.1.

Proof. Note that for $i = 1,\ldots,n$ and $j = 2,\ldots,m$,

$$\begin{aligned}Y_{ij} - \mathbf{X}_{ij}^{\tau}\boldsymbol{\theta} - \sum_{k=1}^{j-1}c_{jk}\widetilde{\varepsilon}_{ik} &= \eta_{ij} + \mathbf{X}_{ij}^{\tau}\big(\boldsymbol{\theta}_0-\boldsymbol{\theta}\big) + \sum_{k=1}^{d_0}\gamma_{k0}(U_{ij})\mathbf{X}_{ij}^{\tau}\big[\mathbf{\Theta}_0(k)-\mathbf{\Theta}(k)\big]\\ &\quad + \sum_{k=1}^{d_0}\gamma_{k0}(U_{ij})\mathbf{X}_{ij}^{\tau}\mathbf{\Theta}(k) + \Big(\sum_{k=1}^{j-1}c_{jk,0}\varepsilon_{ik} - \sum_{k=1}^{j-1}c_{jk}\widetilde{\varepsilon}_{ik}\Big)\end{aligned}$$

and

$$c_{jk,0}\varepsilon_{ik} - c_{jk}\widetilde{\varepsilon}_{ik} = \big(c_{jk,0}-c_{jk}\big)\varepsilon_{ik} + c_{jk}(\varepsilon_{ik}-\widetilde{\varepsilon}_{ik}).$$

Then, following the proof of Lemma B.1 and using the fact that

$$\frac{1}{n(m-1)hf(u)}\sum_{i=1}^{n}\sum_{j=2}^{m}\mathbf{X}_{ij}(\mathbf{\Theta})K\Big(\frac{U_{ij}-u}{h}\Big)\sum_{k=1}^{j-1}(c_{jk,0}-c_{jk})\varepsilon_{ik} = O_P\big(\xi_{n*}\cdot\|\mathbf{C}-\mathbf{C}_0\|\big) = o_P(\|\mathbf{C}-\mathbf{C}_0\|).$$

Then we may prove (B.36) in a manner similar to that in the proof of Lemma B.1. This completes the proof of Lemma B.3. ■

Let

$$\mathbf{F}_{ij} = \left[\mathbf{0}^{\tau}_{\frac{(j-2)(j-1)}{2}}, \varepsilon_{i1}, \ldots, \varepsilon_{ij-1}, \mathbf{0}^{\tau}_{\frac{m(m-1)}{2}-\frac{j(j-1)}{2}}\right]^{\tau},$$

$$\mathbf{W}_{\mathbf{F}} = \mathsf{diag}\left\{\mathbf{W}, \ \frac{1}{m-1}\sum_{j=2}^{m}\mathsf{E}\left[\mathbf{F}_{1j}\mathbf{F}_{1j}^{\tau}\right]\right\},$$

where $\mathbf{W}$ is defined as in Section 3.

Lemma B.4. *Suppose that the conditions in Proposition 4.1 are satisfied. Then we have*

$$\sqrt{n(m-1)}\begin{bmatrix} \widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \\ \mathsf{vec}(\widetilde{\boldsymbol{\Theta}}) - \mathsf{vec}(\boldsymbol{\Theta}_0) \\ \widetilde{\mathbf{C}} - \mathbf{C}_0 \end{bmatrix} = \mathbf{W}_{\mathbf{F}}^{+}\mathbf{V}_{n*}(1 + o_P(1)), \tag{B.37}$$

where $\mathbf{V}_{n*}$ is a $[d(d_0+1) + m(m-1)/2]$-dimensional random variable satisfying $\mathbf{V}_{n*} = O_P(1)$.

Proof Lemma B.4. By using Lemma B.3 and the arguments in the proof of (B.29), the proof of Lemma B.4 is similar to the proof of Lemma B.2 above. Hence, details are omitted here. ■

Proof of Proposition 4.1. By Lemma B.4 above, we readily have (4.6) in Proposition 4.1. ■

References

[1] Cai, Z., Chen, L. and Fang, Y. (2013). Quantile panel data models with partially varying coefficients: estimating the growth effect of FDI. Working paper.

[2] Chen, J., Gao, J. and Li, D. (2013). Estimation in partially linear single-index panel data models with fixed effects. *Journal of Business and Economic Statistics* **31**, 315–330.

[3] Chen, R. and Tsay, R. S. (1993). Functional coefficient autoregressive models. *Journal of American Statistical Association* **88**, 298-308.

[4] Diggle, P. J., Heagerty, P., Liang, K. Y. and Zeger, S. L. (2002). *Analysis of Longitudinal Data.* Oxford University Press, London.

[5] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications.* Chapman and Hall, London.

[6] Fan, J. and Li, R. (2004). New estimation and model selection procedures for semiparametric modelling in longitudinal data analysis. *Journal of the American Statistical Association* **99**, 710–723.

[7] Fan, J., Huang, T. and Li, R. (2007). Analysis of longitudinal data with semiparametric estimation of covariance function. *Journal of the American Statistical Association* **102**, 632-641.

[8] Fan, J. and Zhang, W. (1999). Statistical estimation in varying coefficient models. *Annals of Statistics* **27**, 1491-1518.

[9] Hall, P., Müller, H. G. and Yao, F. (2008). Modelling sparse generalized longitudinal observations with latent Gaussian processes. *Journal of the Royal Statistical Society Series B* **70**, 703–723.

[10] Honda, T., Cheng, M. and Li, J. (2014). Efficient estimation in semivarying coefficient models for longitudinal/clustered data. Working paper.

[11] Hoover, D. R., Rice, J. A., Wu, C. O. and Yang, L. P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* **85**, 809-822.

[12] Jiang, Q., Wang, H., Xia, Y. and Jiang, G. (2013). On a principal varying coefficient model. *Journal of the American Statistical Association* **108**, 228-236.

[13] Jiang, C. and Wang, J. (2011). Functional singl–index models for longitudinal data. *Annals of Statistics* **39**, 362–388.

[14] Kai, B., Li, R. and Zou, H. (2011). New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. *Annals of Statistics* **39**, 305–332.

[15] Körber, L., Linton, O. and Vogt, M. (2015). A semiparametric model for heterogeneous panel data with fixed effects. To appear in *Journal of Econometrics*.

[16] Kottaridi, C. and Stengos, T. (2010). Foreign direct investment, human capital and nonlinearities in economic growth. *Journal of Macroeconomics* **32**, 858–871.

[17] Lam, C. and Yao, Q. (2012). Factor modelling for high-dimensional time series: inference for the number of factors. *Annals of Statistics* **40**, 694–726.

[18] Leng, C., Zhang, W. and Pan, J. (2010). Semiparametric mean-covariance regression analysis for longitudinal data. *Journal of the American Statistical Association* **105**, 181–193.

[19] Li, R. and Liang, H. (2008). Variable selection in semiparametric regression modeling. *Annals of Statistics* **36**, 261–286.

[20] Li, X. and Liu, X. (2004). Foreign direct investment and economic growth: an increasingly endogenous relationship. *World Development* **33**, 393–406.

[21] Li, Y. (2011). Efficient semiparametric regression for longitudinal data with nonparametric covariance estimation. *Biometrika* **98**, 355–370.

[22] Lin, X. and Carroll, R. (2000). Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American Statistical Association* **95**, 520–534.

[23] Lin, D. and Ying, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data (with discussions). *Journal of the American Statistical Association* **96**, 103–126.

[24] Linton, O., Mammen, E., Lin, X. and Carroll, R. (2003). Accounting for correlation in marginal longitudinal nonparametric regression. *Second Seattle Symposium on Biostatistics.*

[25] Liu, S. and Li, G. (2015). Varying-coefficient mean-covariance regression analysis for longitudinal data. *Journal of Statistical Planning and Inference* **160**, 89–106.

[26] Mack, Y. P. and Silverman, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Zeitschrift fur Wahrscheinlichskeittheorie und verwandte Gebiete* **61**, 405–415.

[27] Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation. *Biometrika* **86**, 677–690.

[28] Wang, L., Xue, L., Qu, A. and Liang, H. (2014). Estimation and model selection in generalized additive partial linear models for high-dimensional correlated data. *Annals of Statistics* **42**, 592–624.

[29] Wang, N. (2003). Marginal nonparametric kernel regression accounting for within-subject correlation. *Biometrika* **90**, 43–52.

[30] Wang, H. and Xia, Y. (2009). Shrinkage estimation of the varying-coefficient model. *Journal of the American Statistical Association* **104**, 747–757.

[31] Wu, H. and Zhang, J. (2006). *Nonparametric Regression Methods for Longitudinal Data Analysis: Mixed-Effects Modelling Approaches.* Wiley Series in Probability and Statistics.

[32] Xia, Y., Zhang, W. and Tong, H. (2004). Efficient estimation for semivarying-coefficient models. *Biometrika* **91**, 661–681.

[33] Yao, W. and Li, R. (2013). New local estimation procedure for nonparametric regression function of longitudinal data. *Journal of the Royal Statistical Society, Series B* **75**, 123–138.

[34] Yao, F., Müller, H. G. and Wang, J. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* **100**, 577–590.

[35] Zhang, W., Fan, J. and Sun,Y. (2009). A semiparametric model for cluster data. *Annals of Statistics* **37**, 2377-2408.

[36] Zhang, W., Leng, C. and Tang, C. (2015). A joint modeling approach for longitudinal studies. *Journal of the Royal Statistical Society Series B* **77**, 219–238.

[37] Zhou, Y. and Liang, H. (2009). Statistical inference for semiparametric varying-coefficient partially linear models with error-prone linear covariates. *Annals of Statistics* **37**, 427–458.