Article:

Chen, Jia orcid.org/0000-0002-2791-2486, Li, Degui orcid.org/0000-0001-6802-308X and Xia, Yingcun (2019) Estimation of a rank-reduced functional-coefficient panel data model with serial correlation. Journal of Multivariate Analysis. pp. 456-479.

https://doi.org/10.1016/j.jmva.2019.04.005

Estimation of a rank-reduced functional-coefficient panel data model with serial correlation

Jia Chen[a,*], Degui Li[b], Yingcun Xia[c]

[a]Department of Economics and Related Studies, University of York, YO10 5DD, United Kingdom
[b]Department of Mathematics, University of York, YO10 5DD, United Kingdom
[c]Department of Statistics and Applied Probability, National University of Singapore, 117546, Singapore

Abstract

We consider estimation of a functional-coefficient panel data model. This model is useful for modeling time varying and cross-sectionally heterogeneous relationships between economic variables. We allow for serial correlation and heteroscedasticity in the model. When the number of explanatory variables is large, we impose a rank-reduced structure on the model's functional coefficients to reduce the number of functions to be estimated and thus improve estimation efficiency. To adjust for serial correlation and further improve estimation efficiency, we use a Cholesky decomposition on the serial covariance matrices to produce a transformation of the original panel data model. By applying the standard semiparametric profile least squares method to the transformed model, more efficient estimates of the coefficient functions can be obtained. Under some regularity conditions, we derive the asymptotic distribution for the developed semiparametric estimators and show their efficiency improvement under correct specification of the serial covariance matrices. To attain this efficiency gain when the serial covariance structure is unknown, we propose approaches to consistently estimate the lower triangular matrix in the Cholesky decomposition for balanced panel data, and the serial covariance matrices for unbalanced panel data. Numerical studies, including Monte Carlo experiments and an empirical application to economic growth data, show that the developed semiparametric method works reasonably well in finite samples.

Keywords: Cholesky decomposition, Functional coefficients, Local linear smoothing, Panel data, Principal component analysis, Profile least squares, Within-subject covariance.
2010 MSC: 62G05, 62E20

1. Introduction

Nonparametric panel data models have received increasing attention in the past two decades [16, 17, 25, 34, 35]. When the dimension of explanatory variables in nonparametric panel regression is large, to circumvent the "curse of dimensionality", some nonparametric and semiparametric modeling techniques, such as functional-coefficient models, additive models, partially linear models and single-index models, have been extensively studied in the literature [2, 5, 12, 23, 30**?**]. In this paper, we consider functional-coefficient panel data models as they are a natural generalization of classical linear regression models and provide a flexible framework for depicting the relationship between the response and explanatory variables. A detailed introduction on estimation and inference of functional-coefficient models with independent or weakly dependent data can be found in [3, 6, 13, 14, 31] and the references therein. In this paper, we do not impose any restriction about the serial correlatedness of the model error terms and allow for arbitrary serial correlation and heteroscedasticity to give our model and method wider applicability.

Consider a set of panel data $(Y_{ij}, \mathbf{X}_{ij}, u_{ij})$ with $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m_i\}$, where Y_{ij} is the response variable of interest, $\mathbf{X}_{ij}$ is a d-dimensional vector of explanatory variables whose first element is 1 and u_{ij} is a univariate random

*Corresponding author. Email address: jia.chen@york.ac.uk

variable. The variable u_{ij} can be chosen as calendar time or some other index variable in practical applications. The functional-coefficient panel model is defined, for $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m_i\}$, by

$$Y_{ij} = \mathbf{X}_{ij}^{\top}\boldsymbol{\beta}_0(u_{ij}) + \varepsilon_{ij}, \tag{1}$$

where $\boldsymbol{\beta}_0$ is a d-dimensional vector of functional coefficients, ε_{ij} is the random error that is cross-sectionally independent but serially correlated and satisfies

$$\mathrm{E}(\varepsilon_{ij}|\mathbf{X}_{ij}, u_{ij}) = 0 \quad \text{and} \quad \mathrm{cov}\,(\boldsymbol{\varepsilon}_i|\mathbf{X}_i, \boldsymbol{u}_i) = \boldsymbol{\Sigma}_{i,0} \tag{2}$$

almost surely (a.s.), where $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{im_i})^{\top}$, $\mathbf{X}_i = (\mathbf{X}_{i1}, \ldots, \mathbf{X}_{im_i})^{\top}$ and $\mathbf{u}_i = (u_{i1}, \ldots, u_{im_i})^{\top}$. Throughout the paper, we annotate the true parameters and functions of the data generating process with 0 in their subscripts.

In this paper, we consider short panels, where the number of cross-sectional units n is large, but the number of time series observations m_i is fixed and relatively small and may vary across units. Our main interest is to estimate the functional coefficients $\boldsymbol{\beta}_0$ in model (1). However, when the dimension d is large or even moderately large, the nonparametric estimation of the functional coefficients is unstable. Therefore, an appropriate dimension reduction approach needs to be employed to reduce the number of nonparametric coefficient functions to be estimated. Motivated by [19], we achieve this dimension reduction by extracting the principal components of the functional coefficients. More specifically, we assume that there exist a vector of functions $\boldsymbol{\gamma}_0 = (\gamma_{1,0}, \ldots, \gamma_{d_0,0})^{\top}$, a d-dimensional vector of parameters $\boldsymbol{\theta}_0$ and a $d \times d_0$ (with $d_0 \leq d$) matrix of parameters $\boldsymbol{\Theta}_0 = (\boldsymbol{\Theta}_0(1), \ldots, \boldsymbol{\Theta}_0(d_0))$ in which $\boldsymbol{\Theta}_0(k)$, for each $k \in \{1, \ldots, d_0\}$, is a d-dimensional column vector, such that

$$\boldsymbol{\beta}_0(u) = \boldsymbol{\theta}_0 + \boldsymbol{\Theta}_0\boldsymbol{\gamma}_0(u). \tag{3}$$

We call $\gamma_{1,0}$, ..., $\gamma_{d_0,0}$ the principal functional coefficients since they can be seen as the principal functions of $\boldsymbol{\beta}_0$. The positive integer d_0 is usually unknown in practice but typically much smaller than d. This integer can be estimated by a simple ratio criterion [21] given in Section 5.1. We call model (1) with the imposition (3) a rank-reduced functional-coefficient panel data model. By imposing the structure (3) on the functional coefficients $\boldsymbol{\beta}_0$, we reduce the number of nonparametric functions from d to d_0.

In practice, economic panel data are often found to be serially correlated. There are two possible ways to deal with serial correlation in econometric analysis. The first is to treat it as a result of model misspecification and then adjust the model accordingly to eliminate serial correlation, e.g., by using a dynamic model instead of a static model. The second is to correct for serial correlation in estimation and statistical inference directly without changing the model specification. In the present paper, we will take the second approach.

In the presence of serial correlation, a direct application of the estimation procedure proposed by Jiang et al. [19], which ignores the serial correlation, would certainly affect the efficiency of functional coefficients estimation. To account for serial correlation, some modified nonparametric and semiparametric methods have been introduced in the statistics literature, making certain functional transformation or nonparametric/semiparametric estimation of the serial covariance matrices $\boldsymbol{\Sigma}_{i,0}$ [11, 24, 26, 36]. In this paper, we adjust for serial correlation by using a Cholesky decomposition on the serial covariance matrices $\boldsymbol{\Sigma}_{i,0}$ to obtain a transform of the original panel data model so that the errors of the transformed model are free from serial correlation. We then apply a semiparametric estimation method to the transformed model to obtain estimates of $\boldsymbol{\theta}_0$, $\boldsymbol{\Theta}_0$ and $\boldsymbol{\gamma}_0$. To the best of our knowledge, this paper is among the first to combine the rank-reduced structure on the model functional coefficients and the Cholesky decomposition on the serial covariance matrices in the estimation methodology and systematically study the relevant asymptotic properties. Under some regularity conditions, we establish the asymptotic distribution theory for the proposed semiparametric estimators. In particular, we show that using the rank-reduced structure on the functional coefficients and the Cholesky decomposition on the serial covariance matrices can improve the estimation efficiency of the principal functional coefficients and thus that of the functional coefficients when the serial covariance structure is correctly specified up to a constant multiple.

However, the true serial covariance structure is usually unknown, and its misspecification could lead to efficiency loss. Hence, we introduce two different approaches for consistent estimation of the lower triangular matrix in the Cholesky decomposition for balanced and unbalanced panel data. By using these consistent estimates, we attain the same efficiency gain as when the serial covariance matrices are correctly specified. The simulation studies show that the developed semiparametric approach works reasonably well in finite samples.

The functional-coefficient panel modelling framework (1) can cover a special case where the univariate index variable is time-invariant. In this case, for each $i \in \{1, \ldots, n\}$, $j \in \{1, \ldots, m_i\}$, the functional-coefficient panel data model (1) becomes

$$Y_{ij} = \mathbf{X}_{ij}^\top \boldsymbol{\beta}_0(u_i) + \varepsilon_{ij} \tag{4}$$

and the moment conditions in (2) become

$$\mathrm{E}(\varepsilon_{ij}|\mathbf{X}_{ij}, u_i) = 0 \quad \text{and} \quad \mathrm{cov}(\boldsymbol{\varepsilon}_i|\mathbf{X}_i, u_i) = \boldsymbol{\Sigma}_{i,0} \ \text{ a.s.},$$

where u_i is the time-invariant index variable. The estimation methodology developed in our paper is applicable to the above model setting. In fact, model (4) is of interest for economic panel studies where the relationships under study are cross-sectionally heterogeneous but do not vary significantly over time, as in the econometric growth study we provide in Section 6. In this empirical study, the index variable is chosen as the income level of a country at the beginning of the sampling period (i.e., 1986), as previous research (e.g., [4]) has found that differences in initial conditions can account for a large amount of the variation in the effects of various economic factors on growth. By using the proposed approach, two principal functional coefficients are identified out of the 10 functional coefficients. Furthermore, our method is shown to have a better out-of-sample prediction performance than the one that does not employ the rank-reduced structure or the one that ignores the serial correlation in the data.

The rest of the paper is organized as follows. In Section 2, we introduce identification conditions and two semiparametric estimation procedures. In Section 3, we present the asymptotic theory for the proposed estimators. In Section 4, we consider estimation of the lower triangular matrix in the Cholesky decomposition for balanced panel data, and the serial covariance matrices for unbalanced panel data. In Section 5, we discuss methods to obtain initial parameter estimates and determine the value of d_0 and conduct Monte Carlo simulation studies. Section 6 gives an empirical application to panel data on economic growth. Section 7 concludes the paper. Proofs of the asymptotic theorems are given in the Appendix.

2. Model identification and estimation

Before estimating the rank-reduced functional-coefficient model, we first introduce some identification conditions that enable $\boldsymbol{\Theta}_0$ and $\boldsymbol{\gamma}_0$ to be uniquely determined. We then proceed to provide a semiparametric estimation procedure that ignores serial correlation. The parameter estimates from this first procedure are consistent and can be used in the second semiparametric procedure in Section 2.4 as initial parameter values to produce more efficient estimates of the functional coefficients. The second estimation procedure utilises a Cholesky decomposition of the serial covariance matrices and adjusts for serial correlation in its estimation of the functional coefficients. To focus on the idea, we assume in this section that the number of principal functional coefficients d_0 is known. We will discuss the estimation of d_0 later in Section 5.

2.1. Model identification

Assume that the principal functional coefficients $\boldsymbol{\gamma}_0$ satisfy

$$\mathrm{E}\{\boldsymbol{\gamma}_0(u_{ij})\} = \mathbf{0}_{d_0} \quad \text{and} \quad \mathrm{cov}\{\boldsymbol{\gamma}_0(u_{ij})\} = \mathrm{diag}(\lambda_1, \ldots, \lambda_{d_0}), \tag{5}$$

where $\mathbf{0}_{d_0}$ is a d_0-dimensional null vector, $0 \le d_0 \le d$, and the diagonal numbers $\lambda_1, \ldots, \lambda_{d_0}$ can be seen as the eigenvalues of the covariance matrix $\mathrm{cov}\{\boldsymbol{\beta}_0(u_{ij})\}$. If we further assume that

$$\lambda_1 > \cdots > \lambda_{d_0} > 0 \quad \text{and} \quad \boldsymbol{\Theta}_0^\top \boldsymbol{\Theta}_0 = \mathbf{I}_{d_0},$$

where $\mathbf{I}_{d_0}$ is a $d_0 \times d_0$ identity matrix, then $\boldsymbol{\Theta}_0$ and $\boldsymbol{\gamma}_0(u)$ are identifiable up to a possible sign change. Similar identification conditions can be found in [1] and [19]. The paper by Boneva et al. [1] considers identification and estimation of heterogeneous nonparametric panel data models with a univariate regressor.

2.2. *A semiparametric method ignoring serial correlation*

We now introduce a semiparametric profile least squares method that ignores serial correlation in its estimation of parameters and principal functional coefficients. Combining (1) and (3), we may write the rank-reduced functional-coefficient panel data model as

$$Y_{ij} = \mathbf{X}_{ij}^{\top}\boldsymbol{\theta}_0 + \sum_{k=1}^{d_0} \gamma_{k,0}(u_{ij})\mathbf{X}_{ij}^{\top}\boldsymbol{\Theta}_0(k) + \varepsilon_{ij} \tag{6}$$

for each $i \in \{1, \ldots, n\}$, $j \in \{1, \ldots, m_i\}$.

For given $\boldsymbol{\theta}$ and $\boldsymbol{\Theta}$, we estimate the principal functional coefficients $\boldsymbol{\gamma}_0$ at a given point u by using the local linear smoothing method [10]. Define the kernel-weighted loss function

$$\mathcal{L}_n\{\mathbf{a}(u), \mathbf{b}(u)\big|\boldsymbol{\theta}, \boldsymbol{\Theta}\} = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left\{Y_{ij} - \mathbf{X}_{ij}^{\top}\boldsymbol{\theta} - \sum_{k=1}^{d_0} a_k(u)\mathbf{X}_{ij}^{\top}\boldsymbol{\Theta}(k) - \sum_{k=1}^{d_0} b_k(u)(u_{ij}-u)\mathbf{X}_{ij}^{\top}\boldsymbol{\Theta}(k)\right\}^2 K\left(\frac{u_{ij}-u}{h}\right), \tag{7}$$

where $\mathbf{a}(u) = (a_1(u), \ldots, a_{d_0}(u))^{\top}$ and $\mathbf{b}(u) = (b_1(u), \ldots, b_{d_0}(u))^{\top}$, K is a kernel function and h is a bandwidth. Let $\widehat{\mathbf{a}}(u) = (\widehat{a}_1(u), \ldots, \widehat{a}_{d_0}(u))^{\top}$ and $\widehat{\mathbf{b}}(u) = (\widehat{b}_1(u), \ldots, \widehat{b}_{d_0}(u))^{\top}$ be the solution to the minimization of the loss function $\mathcal{L}_n\{\mathbf{a}(u), \mathbf{b}(u)|\boldsymbol{\theta}, \boldsymbol{\Theta}\}$ with respect to $\mathbf{a}(u)$ and $\mathbf{b}(u)$. Then, the local linear estimates of the principal functional coefficients for given $\boldsymbol{\theta}$ and $\boldsymbol{\Theta}$ can be obtained

$$\widehat{\boldsymbol{\gamma}}(u|\boldsymbol{\theta}, \boldsymbol{\Theta}) = \widehat{\mathbf{a}}(u) \quad \text{with} \quad \forall_{k\in\{1,\ldots,d_0\}} \quad \widehat{\gamma}_k(u|\boldsymbol{\theta}, \boldsymbol{\Theta}) = \widehat{a}_k(u). \tag{8}$$

Given the condition $\mathrm{E}\{\boldsymbol{\gamma}_0(u_{ij})\} = \mathbf{0}_{d_0}$ in (5), we centralize $\widehat{\gamma}_k(u|\boldsymbol{\theta}, \boldsymbol{\Theta})$ for each $k \in \{1, \ldots, d_0\}$ to obtain

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

where $N(n) = m_1 + \cdots + m_n$. Replacing $\gamma_{k,0}(u_{ij})$ with $\widehat{\gamma}_k^{*}(u_{ij}|\boldsymbol{\theta}, \boldsymbol{\Theta})$ in (6), we can estimate $\boldsymbol{\theta}_0$ and $\boldsymbol{\Theta}_0$ by minimizing the least squares loss function

$$Q_n(\boldsymbol{\theta}, \boldsymbol{\Theta}) = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left\{Y_{ij} - \mathbf{X}_{ij}^{\top}\boldsymbol{\theta} - \sum_{k=1}^{d_0}\widehat{\gamma}_k^{*}(u_{ij}|\boldsymbol{\theta}, \boldsymbol{\Theta})\mathbf{X}_{ij}^{\top}\boldsymbol{\Theta}(k)\right\}^2. \tag{9}$$

In general, the solution to the minimization of the loss function $Q_n(\boldsymbol{\theta}, \boldsymbol{\Theta})$ in (9) can be obtained via an iterative algorithm. A proper choice of the initial estimates of $\boldsymbol{\theta}_0$ and $\boldsymbol{\Theta}_0$ may help save computational time and improve estimation accuracy in finite samples. Section 5.1 will discuss how the consistent initial estimates can be obtained.

Let $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\Theta}}$ be the minimizers to $Q_n(\boldsymbol{\theta}, \boldsymbol{\Theta})$ in (9). The final local linear estimate of the principal functional coefficients $\boldsymbol{\gamma}_0(u)$ is

$$\widehat{\boldsymbol{\gamma}}(u) = \widehat{\boldsymbol{\gamma}}(u|\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Theta}}). \tag{10}$$

Subsequently, an estimate for the model functional coefficients is

$$\widehat{\boldsymbol{\beta}}(u) = \widehat{\boldsymbol{\theta}} + \widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\gamma}}(u). \tag{11}$$

The above semiparametric profile least squares estimation can be seen as a generalization of the method in [19] to the panel data setting. As can be seen, this method does not take into account the serial correlation in the data, and hence may entail an efficiency loss. To address this problem, we will modify this semiparametric estimation method for serial correlation by utilizing a Cholesky decomposition on the serial covariance matrices.

2.3. *Cholesky decomposition*

The Cholesky decomposition has been widely used in recent literature to analyze covariance matrices [22, 27, 32, 36, 38]. In particular, Yao and Li [36] apply this technique to improve the nonparametric estimation efficiency in panel data models with a univariate regressor. In this paper, we consider a more general model setting, where there are multivariate regressors and a rank-reduced functional-coefficient structure.

For the serial covariance matrix $\mathbf{\Sigma}_{i,0}$ of each cross section, by a Cholesky decomposition, there exist a lower triangular matrix $\mathbf{C}_{i,0}$ with diagonal elements being 1s and a diagonal matrix $\mathbf{\Lambda}_{i,0}$ with diagonal elements being ρ_{i1}, ..., $\rho_{im_i,0}$, such that

$$\mathbf{C}_{i,0}\mathbf{\Sigma}_{i,0}\mathbf{C}_{i,0}^\top = \mathbf{\Lambda}_{i,0} = \mathrm{diag}(\rho_{i1,0}, \ldots, \rho_{im_i,0}), \tag{12}$$

where $\rho_{ij,0} > 0$ for $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m_i\}$. Define $\boldsymbol{\eta}_i = (\eta_{i1}, \ldots, \eta_{im_i})^\top = \mathbf{C}_{i,0}\boldsymbol{\varepsilon}_i$. Then, $\mathrm{cov}(\boldsymbol{\eta}_i) = \mathbf{C}_{i,0}\mathbf{\Sigma}_{i,0}\mathbf{C}_{i,0}^\top = \mathbf{\Lambda}_{i,0}$, which is a diagonal matrix. Hence, η_{ij} are serially uncorrelated. Denote by $c_{i,jk,0}$ the (j,k)th entry of the minus of $\mathbf{C}_{i,0}$, i.e., $-\mathbf{C}_{i,0}$. For each $i \in \{1, \ldots, n\}$, it is easy to see that $\varepsilon_{i1} = \eta_{i1}$ and, for all $j \in \{2, \ldots, m_i\}$,

$$\varepsilon_{ij} = \eta_{ij} + \sum_{k=1}^{j-1} c_{i,jk,0}\varepsilon_{ik}. \tag{13}$$

Given the autoregressive representation for ε_{ij} in (13), we call $c_{i,jk,0}$ the autoregressive coefficients in the Cholesky decomposition (12). Using (6) and (13), we can re-write (6) as, for each $i \in \{1, \ldots, n\}$,

$$Y_{i1} = \mathbf{X}_{i1}^\top\boldsymbol{\theta}_0 + \sum_{k=1}^{d_0} \gamma_{k,0}(U_{i1})\mathbf{X}_{i1}^\top\mathbf{\Theta}_0(k) + \eta_{i1}, \tag{14}$$

and, for all $j \in \{2, \ldots, m_i\}$,

$$Y_{ij} - \sum_{k=1}^{j-1} c_{i,jk,0}\varepsilon_{ik} = \mathbf{X}_{ij}^\top\boldsymbol{\theta}_0 + \sum_{k=1}^{d_0} \gamma_{k,0}(u_{ij})\mathbf{X}_{ij}^\top\mathbf{\Theta}_0(k) + \eta_{ij}. \tag{15}$$

Since η_{ij} are both serially and cross-sectionally uncorrelated, if $\sum_{k=1}^{j-1} c_{i,jk,0}\varepsilon_{ik}$ were known, by treating $Y_{ij}-\sum_{k=1}^{j-1} c_{i,jk,0}\varepsilon_{ik}$ as the response variable and η_{ij} as the error term, we would have a panel data model in (14) and (15) that has no serial and cross-sectional correlation on the errors. However, the parameters $c_{i,jk,0}$ and the random errors ε_{ij} are unobservable in practical applications. Hence, we need to replace them with their estimates.

As suggested by [11] and [36], we may replace $\mathbf{\Sigma}_{i,0}$ 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 $\mathbf{\Sigma}_i^\diamond \neq \mathbf{\Sigma}_{i,0}$. Hence, we apply the Cholesky decomposition to a working covariance matrix $\mathbf{\Sigma}_i^\diamond$ and find a lower triangular matrix $\mathbf{C}_i^\diamond$ whose main diagonal elements are 1s and a diagonal matrix $\mathbf{\Lambda}_i^\diamond$ whose diagonal elements are positive constants $\rho_{i1}^\diamond, \ldots, \rho_{im_i}^\diamond$, such that, for all $i \in \{1, \ldots, n\}$,

$$\mathbf{C}_i^\diamond\mathbf{\Sigma}_i^\diamond(\mathbf{C}_i^\diamond)^\top = \mathbf{\Lambda}_i^\diamond = \mathrm{diag}(\rho_{i1}^\diamond, \ldots, \rho_{im_i}^\diamond). \tag{16}$$

Given an initial local linear estimate $\widetilde{\boldsymbol{\beta}}$ computed directly from the functional-coefficient model (1) using a kernel function K and a bandwidth b, let $\widetilde{\boldsymbol{\varepsilon}}_i = (\widetilde{\varepsilon}_{i1}, \ldots, \widetilde{\varepsilon}_{im_i})^\top$, where $\widetilde{\varepsilon}_{ij} = Y_{ij} - \mathbf{X}_{ij}^\top\widetilde{\boldsymbol{\beta}}(u_{ij})$. Define

$$\widetilde{Y}_{i1} = Y_{i1}, \quad \forall_{j\in\{2,\ldots,m_i\}}\ \widetilde{Y}_{ij} = Y_{ij} - \sum_{k=1}^{j-1} c_{i,jk}^\diamond\widetilde{\varepsilon}_{ik}, \tag{17}$$

where $c_{i,jk}^\diamond$ is the (j,k)th entry of $-\mathbf{C}_i^\diamond$. Approximate (14) and (15) by

$$\widetilde{Y}_{ij} \approx \mathbf{X}_{ij}^\top\boldsymbol{\theta}_0 + \sum_{k=1}^{d_0} \gamma_{k,0}(u_{ij})\mathbf{X}_{ij}^\top\mathbf{\Theta}_0(k) + \eta_{ij}, \tag{18}$$

for all $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m_i\}$. Section 2.4 below provides an estimation method for model (18). Theorem 1 in Section 3 will show that the efficiency improvement of this estimator relies on correct specification of the lower triangular matrices $\mathbf{C}_{i,0}$ and the serial covariance matrices $\mathbf{\Sigma}_{i,0}$. To avoid misspecification and ensure efficiency improvement, we will provide two methods in Section 4 for consistently estimating $\mathbf{C}_{i,0}$ or $\mathbf{\Sigma}_{i,0}$ for balanced and unbalanced panel data, respectively.

2.4. A semiparametric method adjusting for serial correlation

After the Cholesky decomposition (16) on the working covariance matrices $\mathbf{\Sigma}_i^\diamond$ and the subsequent transformation in (17), the error terms η_{ij} in (18) are uncorrelated over both i and j and have variances $\mathrm{var}(\eta_{ij}) = \rho_{ij}^\diamond$. Hence, for any given $\sqrt{n}$-consistent parameter estimates $\overline{\boldsymbol{\theta}}$ and $\overline{\mathbf{\Theta}}$, we define the loss function

$$\overline{\mathcal{L}}_n\{\mathbf{a}(u), \mathbf{b}(u)\} = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left\{\widetilde{Y}_{ij} - \mathbf{X}_{ij}^\top\overline{\boldsymbol{\theta}} - \sum_{k=1}^{d_0} a_k(u)\mathbf{X}_{ij}^\top\overline{\mathbf{\Theta}}(k) - \sum_{k=1}^{d_0} b_k(u)(u_{ij}-u)\mathbf{X}_{ij}^\top\overline{\mathbf{\Theta}}(k)\right\}^2 (\rho_{ij}^\diamond)^{-1} K\left(\frac{u_{ij}-u}{h}\right), \qquad (19)$$

where $\overline{\mathbf{\Theta}}(k)$ is the kth column of $\overline{\mathbf{\Theta}}$. Let $\overline{\mathbf{a}}(u) = (\overline{a}_1(u), \ldots, \overline{a}_{d_0}(u))^\top$ and $\overline{\mathbf{b}}(u) = (\overline{b}_1(u), \ldots, \overline{b}_{d_0}(u))^\top$ be the solution to the minimization of $\overline{\mathcal{L}}_n\{\mathbf{a}(u), \mathbf{b}(u)\}$ in (19), which has a closed form. A modified local linear estimate of the principal functional coefficients is

$$\overline{\boldsymbol{\gamma}}(u) \equiv \overline{\boldsymbol{\gamma}}(u|\overline{\boldsymbol{\theta}}, \overline{\mathbf{\Theta}}) = (\overline{\gamma}_1(u), \ldots, \overline{\gamma}_{d_0}(u))^\top = \overline{\mathbf{a}}(u), \qquad (20)$$

and a subsequent estimate of the functional coefficients is

$$\overline{\boldsymbol{\beta}}(u) = \overline{\boldsymbol{\theta}} + \overline{\mathbf{\Theta}}\,\overline{\boldsymbol{\gamma}}(u). \qquad (21)$$

Proposition 1 in Section 3 shows that $\widehat{\boldsymbol{\theta}}$ and $\widehat{\mathbf{\Theta}}$ constructed in Section 2.2 are consistent with a $\sqrt{n}$-convergence rate. Hence, we may choose $\overline{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}$ and $\overline{\mathbf{\Theta}} = \widehat{\mathbf{\Theta}}$ in the above estimation procedure (19)–(21). Although the parametric estimators $\widehat{\boldsymbol{\theta}}$ and $\widehat{\mathbf{\Theta}}$ do not account for serial correlation, using them in the above procedure would not affect the asymptotic efficiency of the functional-coefficient estimation $\overline{\boldsymbol{\beta}}$, as the convergence rates for parametric estimators are much faster than the point-wise convergence rates of nonparametric estimators, e.g., the asymptotic theorems in Section 3. With properly chosen working serial covariance matrices, the semiparametric estimation procedure proposed in this section provides a feasible approach to more efficiently estimate the coefficient functions.

3. Asymptotic theorems

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

$$\begin{aligned}
&\mathbf{X}_{ij,k}(\mathbf{\Theta}) = \mathbf{X}_{ij}^\top\mathbf{\Theta}(k), \quad \mathbf{X}_{ij}(\mathbf{\Theta}) = (\mathbf{X}_{ij,1}(\mathbf{\Theta}), \ldots, \mathbf{X}_{ij,d_0}(\mathbf{\Theta}))^\top,\\
&\mathbf{X}_{ij}^*(\mathbf{\Theta}) = \mathbf{X}_{ij} - \mathbf{\Delta}_{\mathbf{X}}^*(u_{ij}|\mathbf{\Theta})\mathbf{X}_{ij}(\mathbf{\Theta}), \quad \mathbf{X}_{ij}^* \equiv \mathbf{X}_{ij}^*(\mathbf{\Theta}_0),\\
&\mathbf{\Delta}_{\mathbf{X}}^*(u_{ij}|\mathbf{\Theta}) = \mathbf{\Delta}_2(u_{ij}|\mathbf{\Theta})\mathbf{\Delta}_1^+(u_{ij}|\mathbf{\Theta}) - \mathrm{E}\{\mathbf{\Delta}_2(u_{ij}|\mathbf{\Theta})\mathbf{\Delta}_1^+(u_{ij}|\mathbf{\Theta})\},
\end{aligned}$$

where $\mathbf{\Delta}_1(u|\mathbf{\Theta}) = \mathrm{E}\{\mathbf{X}_{ij}(\mathbf{\Theta})\mathbf{X}_{ij}^\top(\mathbf{\Theta})|u_{ij} = u\}$, $\mathbf{\Delta}_2(u|\mathbf{\Theta}) = \mathrm{E}\{\mathbf{X}_{ij}\mathbf{X}_{ij}^\top(\mathbf{\Theta})|u_{ij} = u\}$, and for a square matrix $\mathbf{A}$, $\mathbf{A}^+$ denotes its generalised inverse. Let

$$\mathbf{W}_1 = \mathrm{E}\left\{\begin{pmatrix} 1 & \boldsymbol{\gamma}_0^\top(u_{ij}) \\ \boldsymbol{\gamma}_0(u_{ij}) & \boldsymbol{\gamma}_0(u_{ij})\boldsymbol{\gamma}_0^\top(u_{ij})\end{pmatrix} \otimes \left(\mathbf{X}_{ij}^*\mathbf{X}_{ij}^{*\top}\right)\right\}$$

and

$$\mathbf{V}_n = (\mathbf{V}_n^\top(0), \mathbf{V}_n^\top(1), \ldots, \mathbf{V}_n^\top(d_0))^\top,$$

where $\otimes$ denotes the Kronecker product,

$$\mathbf{V}_n(k) = \frac{1}{\sqrt{N(n)}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\{\gamma_{k,0}(u_{ij})(\mathbf{X}_{ij}^* - \mathbf{X}_{ij}^\S)\varepsilon_{ij} + \mathbf{X}_{ij}^\ddagger(k)\varepsilon_{ij} + \mathbf{\Delta}_k^\ddagger\boldsymbol{\gamma}_0(u_{ij})\}$$

for $k \in \{0, \ldots, d_0\}$, $\gamma_{0,0} \equiv 1$, $N(n) = m_1 + \cdots + m_n$,

$$\begin{aligned}
\mathbf{X}_{ij}^\S &= \mathrm{E}\{\mathbf{\Delta}_2(u_{ij}|\mathbf{\Theta}_0)\mathbf{\Delta}_1^+(u_{ij}|\mathbf{\Theta}_0)\}\mathbf{X}_{ij}(\mathbf{\Theta}_0),\\
\mathbf{X}_{ij}^\ddagger(k) &= \mathrm{E}\{\mathbf{\Delta}_2(u_{ij}|\mathbf{\Theta}_0)\mathbf{\Delta}_1^+(u_{ij}|\mathbf{\Theta}_0)\}\mathrm{E}\{\gamma_{k,0}(u_{ij})\mathbf{\Delta}_1(u_{ij}|\mathbf{\Theta}_0)\}\mathbf{\Delta}_1^+(u_{ij}|\mathbf{\Theta}_0)\mathbf{X}_{ij}(\mathbf{\Theta}_0),\\
\mathbf{\Delta}_k^\ddagger &= \mathrm{E}\{\mathbf{\Delta}_2(u_{ij}|\mathbf{\Theta}_0)\mathbf{\Delta}_1^+(u_{ij}|\mathbf{\Theta}_0)\}\mathrm{E}\{\gamma_{k,0}(u_{ij})\mathbf{\Delta}_1(u_{ij}|\mathbf{\Theta}_0)\}.
\end{aligned}$$

Assume that there exists a $d(d_0+1)\times d(d_0+1)$ matrix $\mathbf{W}_2$ such that

$$\lim_{n\to\infty} \mathrm{E}(\mathbf{V}_n\mathbf{V}_n^\top) = \mathbf{W}_2.$$

We next list the assumptions used to prove the asymptotic theory. Some of these assumptions might be relaxed at the cost of lengthier proofs.

Assumption 1. (i) The kernel function K 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 $\ln(h^{-1})/(nh)\to 0$, where $\delta<1-1/\zeta$ with ζ being defined in Assumption 2(iii) below.

Assumption 2. (i) The random elements in $\{(u_{ij},\mathbf{X}_{ij},\varepsilon_{ij}):\ j\in\{1,\dots,m_i\}\}$ are independent over i. Furthermore, u_{ij} and $\mathbf{X}_{ij}$ are identically distributed over both i and j, and $\mathrm{E}\{\varepsilon_{i_1,k}\varepsilon_{i_2,\ell}\mid(u_{ij},\mathbf{X}_{ij}),\ i\in\{1,\dots,n\},\ j\in\{1,\dots,m_i\}\}=0$ 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 is positive and bounded away from zero on $\mathcal{U}$. The joint density function of (u_{ij},u_{ik}), f_{jk}, exists and is continuous for any $j\neq k$.

(iii) There exists a positive number $\zeta>2$ such that $\mathbf{E}(\|\mathbf{X}_{ij}\|^\zeta+|\varepsilon_{ij}|^\zeta)<\infty$. Furthermore, when $\mathbf{\Theta}$ is in a small neighborhood of $\mathbf{\Theta}_0$, the matrix $\mathbf{\Delta}_1(u|\mathbf{\Theta})$ is continuous, positive definite and twice differentiable for any $u\in\mathcal{U}$.

Assumption 3. The functional coefficients $\boldsymbol{\beta}_0$ and principal functional coefficients $\boldsymbol{\gamma}_0$ have continuous second-order derivatives in $\mathcal{U}$.

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

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

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

where $\xi_n^\star=\{\ln(b^{-1})/(nb)\}^{1/2}$.

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 as in [29]. The bandwidth condition $nh^4\to 0$ in Assumption 1(ii) indicates that under-smoothing is needed to derive the $\sqrt{n}$-convergence rates for the parameter estimation. Assumption 2(i) requires the underlying panel data to be cross-sectionally independent, which is not uncommon in the literature; see, e.g., [18]. However, the restriction of identical distribution on u_{ij} and $\mathbf{X}_{ij}$ can be relaxed at the cost of more complicated expressions for $\mathbf{W}_1$ and $\mathbf{V}_n$. The smoothness conditions on f, $\boldsymbol{\beta}_0$ and $\boldsymbol{\gamma}_0$ in Assumptions 2(ii) and 3 are needed as local linear smoothing of the nonparametric functional coefficients is used; see, e.g., [10]. Assumption 4 is mainly used in Theorem 1 for proving that the term $\varepsilon_{ik}-\widetilde{\varepsilon}_{ik}$ is asymptotically negligible in the estimation of the principal functional coefficients.

We now give the asymptotic distribution theory for the parameter estimators $\widehat{\boldsymbol{\theta}}$ and $\widehat{\mathbf{\Theta}}$ and the nonparametric estimator $\widehat{\boldsymbol{\gamma}}$ defined in Section 2.2.

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

$$\lim_{n\to\infty}\frac{1}{N(n)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sigma_{ij}^2=c_\sigma,$$

where $\sigma_{ij}^2=\mathrm{E}(\varepsilon_{ij}^2)$. Then we have

$$\sqrt{N(n)}\begin{bmatrix}\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\\ \mathrm{vec}(\widehat{\mathbf{\Theta}})-\mathrm{vec}(\mathbf{\Theta}_0)\end{bmatrix}\rightsquigarrow\mathcal{N}\left(\mathbf{0}_{d(d_0+1)},\ \mathbf{W}_1^+\mathbf{W}_2\mathbf{W}_1^+\right)\tag{22}$$

and

$$\sqrt{N(n)h}\left[\widehat{\boldsymbol{\gamma}}(u) - \boldsymbol{\gamma}_0(u) - \frac{1}{2}\mu_2\boldsymbol{\gamma}_0''(u)h^2\right] \rightsquigarrow \mathcal{N}[\mathbf{0}_{d_0},\ \omega(u)\boldsymbol{\Delta}_1^+(u|\boldsymbol{\Theta}_0)], \tag{23}$$

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 is the density function of u_{ij}.

Remark 1. As the number of observations for each individual, m_i, is assumed to be fixed, we have $N(n) \propto n$, where $a_n \propto b_n$ denotes that $0 < \underline{c} \leq a_n/b_n \leq \bar{c} < \infty$ when n is sufficiently large. The above proposition shows that the estimation of the parameters $\boldsymbol{\theta}_0$ and $\boldsymbol{\Theta}_0$ from Section 2.2 has the well-known $\sqrt{n}$-convergence rate, and the nonparametric estimation of $\boldsymbol{\gamma}_0$ 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 [19] to the panel data setting. By (11), (22) and (23), we can show that the asymptotic variance of the nonparametric estimation $\widehat{\boldsymbol{\beta}}(u)$ defined in (11) is $\omega(u)\boldsymbol{\Theta}_0\boldsymbol{\Delta}_1^+(u|\boldsymbol{\Theta}_0)\boldsymbol{\Theta}_0^\top/\{N(n)h\}$. In contrast, a direct local linear estimation of the functional coefficients, with the same kernel function and bandwidth but ignoring the rank-reduced structure, has the asymptotic variance of $\omega(u)\boldsymbol{\Delta}^+(u)/\{N(n)h\}$, where $\boldsymbol{\Delta}(u) = \mathrm{E}(\mathbf{X}_{ij}\mathbf{X}_{ij}^\top|u_{ij} = u)$. Following the argument in [19], the estimator $\widehat{\boldsymbol{\beta}}(u)$ is asymptotically more efficient when d_0 is smaller than d.

Remark 2. The asymptotic covariance matrices in (22) and (23) can be estimated by replacing the unknown elements involved by the relevant estimated values. For example, $\omega(u)$ in (23) can be estimated by

$$\widetilde{\omega}(u) = \nu_0\widetilde{c}_\sigma/\widetilde{f}(u),\ \ \widetilde{c}_\sigma = \frac{1}{N(n)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\widetilde{\varepsilon}_{ij}^2,$$

where $\widetilde{f}(u)$ is the conventional kernel estimate of the density function $f(u)$ and $\widetilde{\varepsilon}_{ij}$ is defined as in Section 2.3. The matrix $\boldsymbol{\Delta}_1(u|\boldsymbol{\Theta}_0)$ can be consistently estimated by $\widetilde{\boldsymbol{\Delta}}_1(u|\widehat{\boldsymbol{\Theta}})$ which is obtained via kernel-based regression by treating $\mathbf{X}_{ij}(\widehat{\boldsymbol{\Theta}})\mathbf{X}_{ij}^\top(\widehat{\boldsymbol{\Theta}})$ and u_{ij} as the response and regressor, respectively. Then we may use $\widetilde{\omega}(u)\widetilde{\boldsymbol{\Delta}}_1^+(u|\widehat{\boldsymbol{\Theta}})$ as the estimated covariance matrix for the asymptotic normal distribution in (23). For the estimation of asymptotic covariance matrix in (22), we need to estimate $\mathbf{W}_1$ and $\mathbf{W}_2$. We next only discuss how to estimate $\mathbf{W}_1$ as the estimation of $\mathbf{W}_2$ can be constructed analogously. Let $\widetilde{\boldsymbol{\Delta}}_2(u|\widehat{\boldsymbol{\Theta}})$ be the estimate of $\boldsymbol{\Delta}_2(u|\boldsymbol{\Theta}_0)$, which is constructed via kernel-based regression by treating $\mathbf{X}_{ij}\mathbf{X}_{ij}^\top(\widehat{\boldsymbol{\Theta}})$ and u_{ij} as the response and regressor, respectively. Define

$$\widetilde{\mathbf{X}}_{ij}^* = \mathbf{X}_{ij} - \widetilde{\boldsymbol{\Delta}}_{\mathbf{X}}^*(u_{ij}|\widehat{\boldsymbol{\Theta}})\mathbf{X}_{ij}(\widehat{\boldsymbol{\Theta}})$$

with

$$\widetilde{\boldsymbol{\Delta}}_{\mathbf{X}}^*(u_{ij}|\widehat{\boldsymbol{\Theta}}) = \widetilde{\boldsymbol{\Delta}}_2(u_{ij}|\widehat{\boldsymbol{\Theta}})\widetilde{\boldsymbol{\Delta}}_1^+(u_{ij}|\widehat{\boldsymbol{\Theta}}) - \frac{1}{N(n)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\widetilde{\boldsymbol{\Delta}}_2(u_{ij}|\widehat{\boldsymbol{\Theta}})\widetilde{\boldsymbol{\Delta}}_1^+(u_{ij}|\widehat{\boldsymbol{\Theta}}).$$

We finally estimate $\mathbf{W}_1$ by

$$\widetilde{\mathbf{W}}_1 = \frac{1}{N(n)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\begin{pmatrix} 1 & \widehat{\boldsymbol{\gamma}}^\top(u_{ij}) \\ \widehat{\boldsymbol{\gamma}}(u_{ij}) & \widehat{\boldsymbol{\gamma}}(u_{ij})\widehat{\boldsymbol{\gamma}}^\top(u_{ij}) \end{pmatrix} \otimes \left(\widetilde{\mathbf{X}}_{ij}^*\widetilde{\mathbf{X}}_{ij}^{*\top}\right).$$

Estimation of the asymptotic variance plays an important role when studying inference on the functional coefficients and constructing relevant test statistics. Examination of the above asymptotic variance estimation in finite samples is not the main focus of the present paper and will be addressed in our future research.

We next give the asymptotic theory for $\overline{\boldsymbol{\gamma}}(u)$, the local linear estimator of the principal functional coefficients defined in Section 2.4. Denote $e_{ij} = \eta_{ij} + \sum_{k=1}^{j-1}(c_{i,jk,0} - c_{i,jk}^\diamond)\varepsilon_{ik}$ and $\tau_{ij} = \mathrm{E}(e_{ij}^2)$. Suppose that there exist two positive constants: c_τ and c_ρ such that

$$\lim_{n\to\infty}\frac{1}{N(n)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{\tau_{ij}}{(\rho_{ij}^\diamond)^2} = c_\tau \quad \text{and} \quad \lim_{n\to\infty}\frac{1}{N(n)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}(\rho_{ij}^\diamond)^{-1} = c_\rho. \tag{24}$$

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

Theorem 1. *Suppose that Assumptions 1–4 and* (24) *are satisfied. Then we have*

$$\sqrt{N(n)h}\left\{\overline{\boldsymbol{\gamma}}(u)-\boldsymbol{\gamma}_0(u)-\frac{1}{2}\mu_2\boldsymbol{\gamma}_0''(u)h^2\right\}\rightsquigarrow \mathcal{N}[\mathbf{0}_{d_0},\ \omega_\diamond(u)\boldsymbol{\Delta}_1^+(u|\boldsymbol{\Theta}_0)], \tag{25}$$

where $\omega_\diamond(u)=\nu_0 c_\tau/\{c_\rho^2 f(u)\}$.

Remark 3. The above theorem can be seen as a generalization of Theorem 2 in [36] to the principal functional-coefficient models for unbalanced panel data. The estimation of the asymptotic covariance matrix in (25) can be constructed in the same way as discussed in Remark 2. Theorem 1 shows that $\overline{\boldsymbol{\gamma}}$ retains the point-wise convergence rate of $O_P(h^2+1/\sqrt{nh})$ even if the working serial covariance matrix $\boldsymbol{\Sigma}_i^\diamond$ is misspecified. However, a misspecified $\boldsymbol{\Sigma}_i^\diamond$ would lead to a larger asymptotic variance than when it is correctly specified. This can be seen by noting that for each i and j, η_{ij} and $\{\varepsilon_{i1},\dots,\varepsilon_{i,j-1}\}$ are independent. Hence, we have

$$\tau_{ij}=\mathrm{E}(\eta_{ij}^2)+\mathrm{E}\left[\left\{\sum_{k=1}^{j-1}(c_{i,jk,0}-c_{i,jk}^\diamond)\varepsilon_{ik}\right\}^2\right]=\rho_{ij,0}+\tau_{ij}^\diamond,$$

where $\tau_{ij}^\diamond=\mathrm{E}[\{\sum_{k=1}^{j-1}(c_{i,jk,0}-c_{i,jk}^\diamond)\varepsilon_{ik}\}^2]$. By (24), we have

$$\begin{aligned}
\omega_\diamond(u) &= \nu_0 c_\tau/\{c_\rho^2 f(u)\}\\
&= \frac{\nu_0}{f(u)}\left\{\lim_{n\to\infty}\frac{1}{N(n)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{\tau_{ij}}{(\rho_{ij}^\diamond)^2}\right\}\left\{\lim_{n\to\infty}\frac{1}{N(n)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}(\rho_{ij}^\diamond)^{-1}\right\}^{-2}\\
&= \frac{\nu_0}{f(u)}\left\{\lim_{n\to\infty}\frac{1}{N(n)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{\rho_{ij,0}}{(\rho_{ij}^\diamond)^2}\right\}\left\{\lim_{n\to\infty}\frac{1}{N(n)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}(\rho_{ij}^\diamond)^{-1}\right\}^{-2}\\
&+ \frac{\nu_0}{f(u)}\left\{\lim_{n\to\infty}\frac{1}{N(n)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{\tau_{ij}^\diamond}{(\rho_{ij}^\diamond)^2}\right\}\left\{\lim_{n\to\infty}\frac{1}{N(n)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}(\rho_{ij}^\diamond)^{-1}\right\}^{-2}\\
&\equiv \omega_1^\diamond(u)+\omega_2^\diamond(u).
\end{aligned} \tag{26}$$

When $\boldsymbol{\Sigma}_i^\diamond$ is correctly specified, $\tau_{ij}^\diamond=0$ for all $i\in\{1,\dots,n\}$ and $j\in\{1,\dots,m_i\}$. However, $\tau_{ij}^\diamond>0$ for some i and j when $\boldsymbol{\Sigma}_i^\diamond$ is misspecified. Hence, $\omega_2^\diamond(u)=0$ when $\boldsymbol{\Sigma}_i^\diamond$ is correctly specified and $\omega_2^\diamond(u)>0$ when $\boldsymbol{\Sigma}_i^\diamond$ is misspecified, showing that misspecification in the working serial covariance matrices would lead to efficiency loss in the principal functional-coefficient estimator $\overline{\boldsymbol{\gamma}}$ and thus efficiency loss in the functional-coefficient estimator $\overline{\boldsymbol{\beta}}$. To avoid this, we will introduce, in the next section, two methods for consistently estimating $\mathbf{C}_{i,0}$ or $\boldsymbol{\Sigma}_{i,0}$ for both balanced and unbalanced panel data.

We next compare the asymptotic covariances between the two local linear estimators of the principal functional coefficients: $\widehat{\boldsymbol{\gamma}}(u)$ and $\overline{\boldsymbol{\gamma}}(u)$. The former ignores serial correlation and the latter has adjusted for serial correlation. For the latter, we consider the case where the serial covariance matrices are correctly specified up to a constant multiple, i.e., $\boldsymbol{\Sigma}_i^\diamond=c_0\boldsymbol{\Sigma}_{i,0}$, where $0<c_0<\infty$. In this case $\omega_2^\diamond(u)\equiv 0$, and hence

$$\omega_\diamond(u)=\omega_1^\diamond(u)=\frac{\nu_0}{f(u)}\left\{\lim_{n\to\infty}\frac{1}{N(n)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\rho_{ij,0}^{-1}\right\}^{-1}.$$

Assume that there exists a positive constant $c_{\rho,0}$ such that

$$0<c_{\rho,0}=\lim_{n\to\infty}\frac{1}{N(n)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\rho_{ij,0}^{-1}. \tag{27}$$

By using the harmonic mean value inequality, we have

$$\left\{\frac{1}{N(n)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\rho_{ij,0}^{-1}\right\}^{-1} \le \frac{1}{N(n)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\rho_{ij,0}. \tag{28}$$

The equality holds only when all the $\rho_{ij,0}$ are the same. Furthermore, by the relation $\varepsilon_{i1} = \eta_{i1}$, Eq. (13) and the fact that η_{ij} and $\{\varepsilon_{i1}, \ldots, \varepsilon_{i,j-1}\}$ are independent for $j \in \{2, \ldots, m_i\}$, we have

$$\rho_{i1,0} = \sigma_{i1}^2, \quad \rho_{ij,0} \le \sigma_{ij}^2,$$

indicating that

$$\frac{1}{N(n)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\rho_{ij,0} \le \frac{1}{N(n)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sigma_{ij}^2. \tag{29}$$

By (28), (29), Proposition 1 and Theorem 1, we immediately obtain the following asymptotic result.

Corollary 1. *Suppose that the conditions in Proposition 1 and Theorem 1 and Eq.* (27) *hold, and* $\mathbf{\Sigma}_i^\diamond = c_0\mathbf{\Sigma}_{i,0}$ *for some* $0 < c_0 < \infty$. *The local linear estimator* $\overline{\boldsymbol{\gamma}}(u)$ *defined in Section 2.4 is asymptotically more efficient than* $\widehat{\boldsymbol{\gamma}}(u)$ *in Section 2.2 that ignores serial correlation.*

Remark 4. By (11), (21) and Corollary 1, we can show that the asymptotic variance of the local linear estimator $\widehat{\boldsymbol{\beta}}$ minus that of $\overline{\boldsymbol{\beta}}(u)$ is non-negative definite, and thus $\overline{\boldsymbol{\beta}}$ is asymptotically more efficient than $\widehat{\boldsymbol{\beta}}$. Furthermore, noting that the asymptotic bias of $\overline{\boldsymbol{\gamma}}$ and $\widehat{\boldsymbol{\gamma}}$ are the same, the mean squared error of $\overline{\boldsymbol{\gamma}}$ (or $\overline{\boldsymbol{\beta}}$) is asymptotically smaller than that of $\widehat{\boldsymbol{\gamma}}$ (or $\widehat{\boldsymbol{\beta}}$). This will be later verified for finite samples in simulation.

Remark 5. For the functional-coefficient panel data model (4) with time-invariant index variable u_i, we may show that, with some minor modifications, the asymptotic properties given in this section still hold if the index variables, $u_1, \ldots u_n$, are assumed to be independent and identically distributed and satisfy the smoothness conditions in Assumption 2(ii). Details are omitted here to save the space.

4. Estimation of the Cholesky decomposition

The asymptotic theorems in Section 3 show that the efficiency gain in the local linear estimator of the (principal) functional coefficients which adjusts for serial correlation over the ordinary local linear estimator given in Section 2.2 depends on correct specification of the true serial covariance matrices $\mathbf{\Sigma}_{i,0}$, or equivalently, the lower triangular matrices $\mathbf{C}_{i,0}$ and the diagonal matrices $\mathbf{\Lambda}_{i,0}$ in the Cholesky decomposition. In applications, we may use working covariance matrices constructed from a semiparametric method proposed in [11], which relies on a parametric specification of the serial correlation matrix. In this section, we introduce two different methods for estimating $\mathbf{C}_{i,0}$ or $\mathbf{\Sigma}_{i,0}$. For balanced panel data, we will estimate $\mathbf{C}_{i,0}$ directly together with model parameters $\boldsymbol{\theta}_0$ and $\mathbf{\Theta}_0$ via a profile least squares method. For unbalanced panel data, we first use a local linear method to estimate the serial covariance and variance functions consistently and then obtain an estimate of $\mathbf{C}_{i,0}$ via (12) or (16). The estimation of $\mathbf{\Lambda}_{i,0}$ will be discussed as well for the above two cases.

4.1. The case of balanced panel data

Consider the case of balanced panel data, i.e., $m_1 = \cdots = m_n \equiv m$. As in [36], we assume that $\boldsymbol{\varepsilon}_i \equiv (\varepsilon_{i1}, \ldots, \varepsilon_{im})^\top$ are independent and identically distributed over i and are independent of the regressors $\mathbf{X}_i$. This implies that the serial covariance matrices $\mathbf{\Sigma}_{i,0}$ are invariant over i, i.e., $\mathbf{\Sigma}_{i,0} = \mathbf{\Sigma}_0$ for all $i \in \{1, \ldots, n\}$. Then, the Cholesky decomposition of $\mathbf{\Sigma}_0$ gives

$$\mathbf{C}_0\mathbf{\Sigma}_0\mathbf{C}_0^\top = \mathbf{\Lambda}_0 = \mathrm{diag}(\rho_{1,0}, \ldots, \rho_{m,0}),$$

where $\mathbf{C}_0$ is a lower triangular matrix with diagonal elements being 1s and $\rho_{1,0}$, ..., $\rho_{m,0}$ are positive constants. As in Section 2.3, we denote $c_{jk,0}$ the (j,k)th entry of the matrix $-\mathbf{C}_0$ and $\boldsymbol{\eta}_i = \mathbf{C}_0\boldsymbol{\varepsilon}_i$ and replace ε_{ik} by $\widetilde{\varepsilon}_{ik}$ in (15), where

$\widetilde{\varepsilon}_{ik} = Y_{ik} - \mathbf{X}_{ik}^\top \widetilde{\boldsymbol{\beta}}(u_{ik})$ is the residual from an initial estimate $\widetilde{\boldsymbol{\beta}}(u_{ik})$ of the functional coefficients. Then, for $i \in \{1, \ldots, n\}$ and $j \in \{2, \ldots, m\}$, the approximating model (18) becomes

$$Y_{ij} \approx \mathbf{X}_{ij}^\top \boldsymbol{\theta}_0 + \sum_{k=1}^{d_0} \gamma_{k,0}(u_{ij}) \mathbf{X}_{ij}^\top \boldsymbol{\Theta}_0(k) + \sum_{k=1}^{j-1} c_{jk,0} \widetilde{\varepsilon}_{ik} + \eta_{ij}. \tag{30}$$

Define $\mathbf{c}_0 = (c_{21,0}, c_{31,0}, c_{32,0}, \ldots, c_{m,m-1,0})^\top$, a vector consisting of the elements of $\mathbf{C}_0$ below the main diagonal.

Based on (30), we next provide a feasible procedure for estimating $\boldsymbol{\theta}_0$, $\boldsymbol{\Theta}_0$ and $\mathbf{c}_0$. For given initial values $\boldsymbol{\theta}$, $\boldsymbol{\Theta}$ and $\mathbf{c} = (c_{21}, c_{31}, c_{32}, \ldots, c_{m,m-1})^\top$, we first obtain the local linear estimate of the principal functional coefficients by minimizing the kernel-weighted loss function

$$\mathcal{L}_n^\sharp(\mathbf{a}(u), \mathbf{b}(u) \big| \boldsymbol{\theta}, \boldsymbol{\Theta}, \mathbf{c}) = \sum_{i=1}^n \sum_{j=2}^m \left\{ Y_{ij} - \mathbf{X}_{ij}^\top \boldsymbol{\theta} - \sum_{k=1}^{j-1} c_{jk} \widetilde{\varepsilon}_{ik} - \sum_{k=1}^{d_0} a_k(u) \mathbf{X}_{ij}^\top \boldsymbol{\Theta}(k) \right.$$
$$\left. - \sum_{k=1}^{d_0} b_k(u)(u_{ij} - u) \mathbf{X}_{ij}^\top \boldsymbol{\Theta}(k) \right\}^2 K\left(\frac{u_{ij} - u}{h_1}\right) \tag{31}$$

with respect to $\mathbf{a}(u)$ and $\mathbf{b}(u)$, where h_1 is a bandwidth. Denote the resulting estimator of the principal functional coefficients by $\check{\boldsymbol{\gamma}}(u|\boldsymbol{\theta}, \boldsymbol{\Theta}, \mathbf{c})$ which has a closed form given $\boldsymbol{\theta}$, $\boldsymbol{\Theta}$ and $\mathbf{c}$. Let $\check{\boldsymbol{\gamma}}^*(u|\boldsymbol{\theta}, \boldsymbol{\Theta}, \mathbf{c})$ be the sample centralization of $\check{\boldsymbol{\gamma}}(u|\boldsymbol{\theta}, \boldsymbol{\Theta}, \mathbf{c})$ and $\check{\gamma}_k^*(u|\boldsymbol{\theta}, \boldsymbol{\Theta}, \mathbf{c})$ the kth element of $\check{\boldsymbol{\gamma}}^*(u|\boldsymbol{\theta}, \boldsymbol{\Theta}, \mathbf{c})$. Then we update the values of $\boldsymbol{\theta}$, $\boldsymbol{\Theta}$ and $\mathbf{c}$ by minimizing the loss function

$$Q_n^\sharp(\boldsymbol{\theta}, \boldsymbol{\Theta}, \mathbf{c}) = \sum_{i=1}^n \sum_{j=2}^m \left\{ Y_{ij} - \mathbf{X}_{ij}^\top \boldsymbol{\theta} - \sum_{k=1}^{d_0} \check{\gamma}_k^*(u_{ij}|\boldsymbol{\theta}, \boldsymbol{\Theta}, \mathbf{c}) \mathbf{X}_{ij}^\top \boldsymbol{\Theta}(k) - \sum_{k=1}^{j-1} c_{jk} \widetilde{\varepsilon}_{ik} \right\}^2. \tag{32}$$

Let $(\overline{\boldsymbol{\theta}}, \overline{\boldsymbol{\Theta}}, \overline{\mathbf{c}})$, in which

$$\overline{\mathbf{c}} = (\overline{c}_{21}, \overline{c}_{31}, \overline{c}_{32}, \ldots, \overline{c}_{m,m-1})^\top, \tag{33}$$

be a minimizer of (32), i.e.,

$$Q_n^\sharp(\overline{\boldsymbol{\theta}}, \overline{\boldsymbol{\Theta}}, \overline{\mathbf{c}}) = \min_{\boldsymbol{\theta}, \boldsymbol{\Theta}, \mathbf{c}} Q_n^\sharp(\boldsymbol{\theta}, \boldsymbol{\Theta}, \mathbf{c}). \tag{34}$$

The above estimation method can be seen as a generalization of the profile likelihood method proposed by [36] from univariate nonparametric panel data models to principal functional-coefficient panel data models. It provides a feasible iterative procedure to obtain the estimates $\overline{\boldsymbol{\theta}}$, $\overline{\boldsymbol{\Theta}}$ and $\overline{\mathbf{c}}$. The initial values $\boldsymbol{\theta}$ and $\boldsymbol{\Theta}$ can be chosen as the estimates $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\Theta}}$ from Section 2.2 and the initial value of $\mathbf{c}$ can be chosen as the corresponding values from a Cholesky decomposition of the sample serial covariance matrix of the residuals $\widetilde{\varepsilon}_{ij}$. In the following proposition we give the convergence rates for $\overline{\boldsymbol{\theta}}$, $\overline{\boldsymbol{\Theta}}$ and $\overline{\mathbf{c}}$.

Proposition 2. *Suppose that Assumptions 1(i) and 2–4 are satisfied and Assumption 1(ii) holds with h_1 replacing h. Furthermore, assume $m_i \equiv m$ and that $\boldsymbol{\varepsilon}_i$ are independent and identically distributed over i, and are independent of the regressors $\mathbf{X}_i$. Then we have*

$$\begin{bmatrix} \overline{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \\ \mathrm{vec}(\overline{\boldsymbol{\Theta}}) - \mathrm{vec}(\boldsymbol{\Theta}_0) \\ \overline{\mathbf{c}} - \mathbf{c}_0 \end{bmatrix} = O_P(1/\sqrt{n}). \tag{35}$$

Given a consistent initial estimate $\widetilde{\boldsymbol{\beta}}(u_{ij})$ of the functional coefficients, the diagonal elements, $\rho_{1,0}, \ldots, \rho_{m,0}$, of $\boldsymbol{\Lambda}_0$ can be consistently estimated by applying a Cholesky decomposition on the sample serial covariance matrix of the residuals. One may then add weights, chosen as the inverse of estimates of $\rho_{2,0}, \ldots, \rho_{m,0}$, in the loss functions $\mathcal{L}_n^\sharp$ and $Q_n^\sharp$. This would not affect the above consistency and convergence rates of the parameter estimators but may potentially make them more efficient.

With the estimates of $c_{ij,0}$ from (33) and (34), we may transform Y_{ij} into $\widetilde{Y}_{ij}$ via (17). By substituting $\rho_{ij}^{\diamond}$ with their estimates, plugging in $\overline{\boldsymbol{\theta}}$ and $\overline{\boldsymbol{\Theta}}$, and then minimizing the resulting loss function in (19), we can obtain the estimator $\overline{\boldsymbol{\gamma}}$ of the principal functional coefficients. Combining Theorem 1 and Proposition 2, we obtain the following asymptotic distribution theory for $\overline{\boldsymbol{\gamma}}$.

Corollary 2. *Suppose that the conditions in Theorem 1 and Proposition 2 are satisfied, and let $c_{i,jk}^{\diamond}$ in* (17) *be replaced by $\overline{c}_{jk}$ defined in* (33) *and* (34)*. Then we have*

$$\sqrt{nmh}\left\{\overline{\boldsymbol{\gamma}}(u)-\boldsymbol{\gamma}_0(u)-\mu_2\boldsymbol{\gamma}_0''(u)h^2/2\right\}\rightsquigarrow \mathcal{N}[\mathbf{0}_{d_0},\ \omega_1^{\sharp}(u)\boldsymbol{\Delta}_1^{+}(u|\boldsymbol{\Theta}_0)], \tag{36}$$

where $\omega_1^{\sharp}(u)=\nu_0/\{c_{\rho,0}f(u)\}$ and $c_{\rho,0}$ is defined in (27).

Remark 6. By using consistent estimates of the autoregressive coefficients $c_{ij,0}$ and estimates of $\rho_{j,0}$ in (19), we can avoid the misspecification of $\boldsymbol{\Sigma}_0$ or $\mathbf{C}_0$ and hence obtain a modified local linear estimator of $\boldsymbol{\gamma}_0(u)$ that is asymptotically more efficient than the local linear estimator $\widetilde{\boldsymbol{\gamma}}(u)$ that ignores the serial correlation present in the panel data.

4.2. The case of unbalanced panel data

The profile least squares estimation method discussed in Section 4.1 strongly relies on the balanced structure of the panel data and cannot be directly extended to unbalanced panel data. Hence, we next introduce a different approach to estimate the matrices $\mathbf{C}_{i,0}$ in the Cholesky decomposition. Motivated by [20] and [24], we assume that the observations on the ith cross-section are taken at time points $t_{i1},\ldots,t_{im_i}$, which fall within a bounded interval $\mathcal{T}$. In the rest of this section, we sometimes use $Y_i(t_{ij})$, $\mathbf{X}_i(t_{ij})$, $u_i(t_{ij})$, and $\varepsilon_i(t_{ij})$ to denote Y_{ij}, $\mathbf{X}_{ij}$, u_{ij} and ε_{ij} in order to emphasize the time at which their observations are taken.

Denote the covariance function of ε by σ_0, i.e., for all $t_{ij},t_{ik}\in\mathcal{T}$,

$$\sigma_0(t_{ij},t_{ik})\equiv \operatorname{cov}\{\varepsilon_i(t_{ij}),\varepsilon_i(t_{ik})\}. \tag{37}$$

Note that σ_0 is a bivariate positive semi-definite function, and $\sigma_0(t_{ij},t_{ik})$ is the (j,k)th element of the serial covariance matrix $\boldsymbol{\Sigma}_{i,0}$. Hence, to estimate $\boldsymbol{\Sigma}_{i,0}$ consistently, we only need to estimate σ_0 consistently. As in some previous studies such as [24] and [37], we assume that $\sigma_0(s,t)$ is continuous everywhere except on the plane $s=t$. This implies that the covariance function can have jumps at the main diagonal of the covariance matrix $\boldsymbol{\Sigma}_{i,0}$, i.e., the so-called nugget effect. Because of the existence of the nugget effect, we estimate $\sigma_0(s,t)$, $s\neq t$, and $\sigma_0^2(t)\equiv\sigma(t,t)$, separately.

The estimation procedure is based on local linear smoothing and uses residuals, $\widetilde{\varepsilon}_{ij}=\widetilde{\varepsilon}_i(t_{ij})=Y_{ij}-\mathbf{X}_{ij}^{\top}\widetilde{\boldsymbol{\beta}}(u_{ij})$, from an initial estimate, $\widetilde{\boldsymbol{\beta}}$, of the functional coefficients: (i) for $s\neq t$, we estimate $\sigma_0(s,t)$ by $\widetilde{\sigma}(s,t)=\widetilde{\sigma}_{10}$, where $\widetilde{\sigma}_{10}$ is defined such that the triplet $(\widetilde{\sigma}_{10},\widetilde{\sigma}_{11},\widetilde{\sigma}_{12})$ minimizes

$$\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1,\neq j}^{m_i}\left\{\widetilde{\varepsilon}_{ij}\widetilde{\varepsilon}_{ik}-\sigma_{10}-\sigma_{11}(t_{ij}-s)-\sigma_{12}(t_{ik}-t)\right\}^2 K\left(\frac{t_{ij}-s}{b_1}\right)K\left(\frac{t_{ik}-t}{b_2}\right) \tag{38}$$

with respect to σ_{10}, σ_{11} and σ_{12}, where K is a kernel function and b_1 and b_2 are two bandwidths; (ii) for $s=t$, we estimate $\sigma_0^2(t)$ by $\widetilde{\sigma}^2(t)=\widetilde{\sigma}_{20}$, where $\widetilde{\sigma}_{20}$ is defined such that the pair $(\widetilde{\sigma}_{20},\widetilde{\sigma}_{21})$ minimizes

$$\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left\{\widetilde{\varepsilon}_{ij}^2-\sigma_{20}-\sigma_{21}(t_{ij}-t)\right\}^2 K\left(\frac{t_{ij}-t}{b_3}\right) \tag{39}$$

with respect to σ_{20} and σ_{21}, where b_3 is a bandwidth. Proposition 1 in [24] shows that thus defined $\widetilde{\sigma}(s,t)$ and $\widetilde{\sigma}^2(t)$ are uniformly consistent over $s,t\in\mathcal{T}$. With these consistent estimates, we can readily construct $\widehat{\boldsymbol{\Sigma}}_i$, a consistent estimator of $\boldsymbol{\Sigma}_{i,0}$, whose (j,k)th element is

$$\widehat{\sigma}(t_{ij},t_{ik})=\widetilde{\sigma}(t_{ij},t_{ik})\mathbf{1}(t_{ij}\neq t_{ik})+\widetilde{\sigma}^2(t_{ij})\mathbf{1}(t_{ij}=t_{ik}), \tag{40}$$

where $\mathbf{1}$ is an indicator function. After applying the Cholesky decomposition to $\widehat{\mathbf{\Sigma}}_i$, we can obtain consistent estimates of $\mathbf{C}_{i,0}$ and $\mathbf{\Lambda}_{i,0}$.

Note that the serial covariance matrix estimate $\widehat{\mathbf{\Sigma}}_i$ constructed above is not necessarily positive definite in finite samples. To ensure its positive definiteness, we need to make modifications on $\widetilde{\sigma}(s,t)$ and $\widetilde{\sigma}^2(t)$. As in [15], let $\widetilde{\iota}_k$ and $\widetilde{\phi}_k$ be the eigenvalues and their corresponding eigenfunctions of $\widetilde{\sigma}$ with $\widetilde{\iota}_1 \geq \widetilde{\iota}_2 \geq \cdots$, and re-define the estimator of $\sigma_0(s,t)$ as

$$\widetilde{\sigma}_\star(s,t) = \sum_{k=1}^{k_0} \widetilde{\iota}_k \widetilde{\phi}_k(s) \widetilde{\phi}_k(t), \quad k_0 = \max\{k : \widetilde{\iota}_k > 0\}. \tag{41}$$

On the other hand, to ensure the non-negativity of $\widetilde{\sigma}^2$, we truncate it by a pre-determined positive parameter ϖ_n, which is very close to zero, and re-define the estimator of $\sigma_0^2(t)$ as $\widetilde{\sigma}^2_\star(t) = \widetilde{\sigma}^2(t) \mathsf{I}(\widetilde{\sigma}^2(t) > \varpi_n)$. The simulation studies in Section 5 below show that the nonparametric estimation method for the serial covariance matrices works reasonably well in finite samples.

5. Monte Carlo simulation studies

In this section, we conduct Monte Carlo experiments to examine the finite sample performance of the estimation methods introduced in Sections 2 and 4. To facilitate practical implementation of the methods, we start this section with the proposal of initial parameter estimates for the iterative estimation procedures in Sections 2.2 and 2.4 and then a method for estimating the number of principal functional coefficients, i.e., d_0. It is then followed by a flowchart showing the flow of implementation process for the estimation methods considered in our numerical studies.

5.1. Choice of initial parameter estimates and estimation of d_0

In order to save the computational time of the iterative semiparametric estimation procedures given in Section 2, we introduce a consistent initial estimate for the model parameters. As in previous sections, let $\widetilde{\boldsymbol{\beta}}$ be the local linear estimate of $\boldsymbol{\beta}_0$ obtained from local linear smoothing on the functional-coefficient model (1) directly, ignoring the principal component structure for $\boldsymbol{\beta}_0$. Given the assumption $\mathrm{E}\{\boldsymbol{\gamma}_0(u_{ij})\} = \mathbf{0}$, we have $\boldsymbol{\theta}_0 = \mathrm{E}\{\boldsymbol{\beta}_0(u_{ij})\}$. Hence, an initial estimate of $\boldsymbol{\theta}_0$ can naturally be chosen as

$$\widetilde{\boldsymbol{\theta}}^{(0)} = \frac{1}{N(n)} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \widetilde{\boldsymbol{\beta}}(u_{ij}). \tag{42}$$

To construct an initial estimate for $\mathbf{\Theta}_0$, we define the $d \times d$ sample covariance matrix for $\widetilde{\boldsymbol{\beta}}(u_{ij})$ as

$$\widetilde{\mathbf{\Sigma}}_{\boldsymbol{\beta}} = \frac{1}{N(n)} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left\{\widetilde{\boldsymbol{\beta}}(u_{ij}) - \widetilde{\boldsymbol{\theta}}^{(0)}\right\} \left\{\widetilde{\boldsymbol{\beta}}(u_{ij}) - \widetilde{\boldsymbol{\theta}}^{(0)}\right\}^\top$$

and let $\widetilde{\mathbf{\Theta}}^{(0)}(k)$ be the eigenvector associated with the kth largest eigenvalue $\widetilde{\lambda}_k$ of $\widetilde{\mathbf{\Sigma}}_{\boldsymbol{\beta}}$. In view of the identification conditions in Section 2.1, when d_0 is known, a natural initial estimate of $\mathbf{\Theta}_0$ is

$$\widetilde{\mathbf{\Theta}}^{(0)} = \left[\widetilde{\mathbf{\Theta}}^{(0)}(1), \ldots, \widetilde{\mathbf{\Theta}}^{(0)}(d_0)\right]. \tag{43}$$

Following the proof of Lemma B.1 in Appendix B, the uniform convergence rate for $\widetilde{\boldsymbol{\beta}}$ is $O_P[b^2 + \{nb/\ln(b^{-1})\}^{-1/2}]$, where b is the bandwidth used in computing $\widetilde{\boldsymbol{\beta}}$. By Theorem 1 in [19], we can also show that the initial estimators in (42) and (43) are consistent with a uniform convergence rate $O_P[b^2 + \{nb/\ln(b^{-1})\}^{-1/2}]$.

However, the number of principal functional coefficients d_0 is usually unknown in practice, and has to be estimated. A Bayesian Information Criterion (BIC) is proposed by [19] to estimate d_0, which performs well in numerical studies. In this paper, we use a simple ratio method introduced in [21] to estimate d_0. This estimator is defined as

$$\widehat{d_0} = \arg\min_{1 \leq k \leq d} \widetilde{\lambda}_{k+1} / \widetilde{\lambda}_k, \tag{44}$$

where $\widetilde{\lambda}_k$ is the kth largest eigenvalue of $\widetilde{\mathbf{\Sigma}}_{\boldsymbol{\beta}}$, and we define $\widetilde{\lambda}_{d+1} = 0$ 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. In the simulation studies below, we choose ϵ_0 as 0.05, which works well in all the cases considered.

The flow of estimation process for the proposed methods is shown in the flowchart. To facilitate exposition, we use FCM (and a tilde for estimators) to denote the method that estimates $\boldsymbol{\beta}_0(u)$ by applying local linear smoothing directly on the functional-coefficient model $Y_{ij} = \mathbf{X}_{ij}^\top \boldsymbol{\beta}_0(u_{ij}) + \varepsilon_{ij}$ in (1), ignoring both the principal functional-coefficient structure and any serial correlation that may exist, PFCM (and a hat for estimators) to denote the estimation method defined in (7)–(11) of Section 2.2, which ignores serial correlation but takes into account the principal functional-coefficient structure, and PFCM + CD (and an overline for estimators) to denote the method defined in (19)–(21), which takes into account the principal functional-coefficient structure and adjusts for serial correlation via a Cholesky decomposition on error serial covariance matrices and a subsequent transformation on the model.

5.2. Simulation study

For the choice of bandwidth in the context of panel data with serial correlation, one could employ the leave-one-unit-out (i.e., leave out all observations from one cross-sectional unit at a time) cross-validation method; see [33] for details. In the simulation studies below, however, we use fixed bandwidths to save computational time, i.e., $h = 0.4, 0.3, 0.2$ for cross-sectional sizes of $n = 50, 100, 200$, respectively. Throughout the numerical studies, we use the standard normal probability density function as the kernel function.

We now investigate the finite-sample performance of the estimation methods proposed in the foregoing sections. For each $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m_i\}$, we generate data from the following principal functional-coefficient panel data model

$$Y_{ij} = \mathbf{X}_{ij}^\top \boldsymbol{\beta}_0(u_{ij}) + \varepsilon_{ij}, \quad \boldsymbol{\beta}_0(u_{ij}) = \boldsymbol{\theta}_0 + \mathbf{\Theta}_0 \boldsymbol{\gamma}_0(u_{ij}), \tag{45}$$

where $d = 7$, $d_0 = 1$, $\boldsymbol{\theta}_0 = (1, 1, 1, 1, 1, 1, 1)^\top$, $\mathbf{\Theta}_0 = (-1/\sqrt{2}, 1/\sqrt{2}, 0, 0, 0, 0, 0)^\top$, $\boldsymbol{\gamma}_0(u) = 10u(1-u) - 5/3$, the index variables u_{ij} are independently drawn from the uniform distribution on $[0, 1]$, and $\mathbf{X}_{ij} = (1, \mathbf{X}_{ij}^*)^\top$, in which $\mathbf{X}_{ij}^*$ are independently drawn from a 6-dimensional Gaussian distribution with mean $\mathbf{0}_6$ and covariance matrix $\mathbf{I}_6$. The disturbances ε_{ij} will be specified later. The above data generating process is similar to that in [19]. We compare the performance of three estimation methods for the functional coefficients $\boldsymbol{\beta}_0(u)$: the FCM, PFCM and PFCM + CD. For the PFCM + CD method, the autoregressive coefficients $c_{i,jk,0}$ in the Cholesky decomposition are estimated using the method introduced either in Section 4.1 or Section 4.2, depending on whether the panels are balanced or unbalanced.

The accuracy of FCM, PFCM, and PFCM + CD estimates for the functional coefficients $\boldsymbol{\beta}_0(u)$ are evaluated by the average bias, mean squared error (MSE), and mean absolute deviation (MAD) over grid points between 0.05 and 0.95 with an increment of 0.1. Letting $u_1^0, \ldots, u_{N_0}^0$ denote these grid points, the average bias, MSE, and MAD are defined, respectively, as

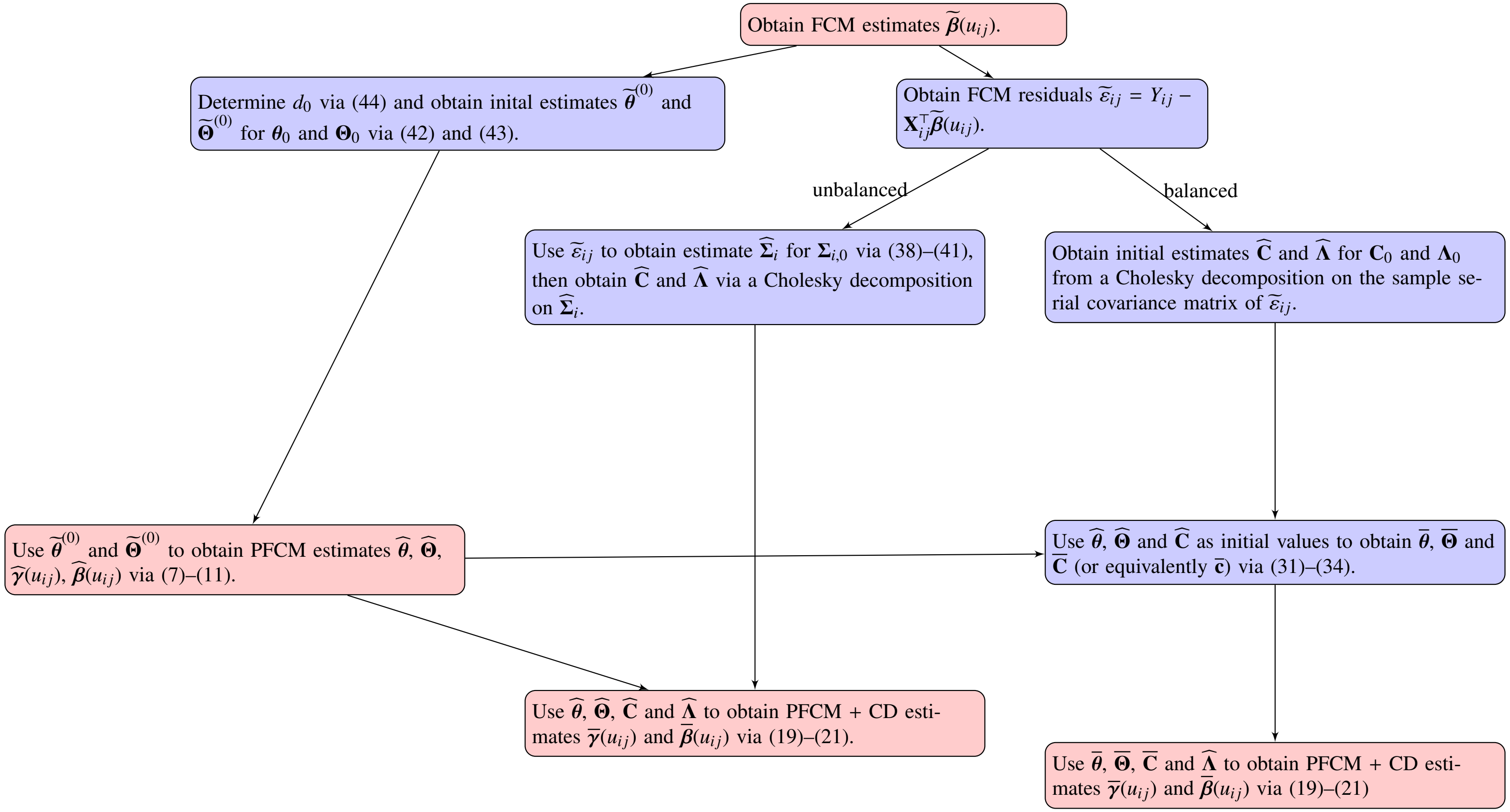
$$\mathrm{Bias}(\breve{\boldsymbol{\beta}}) = \frac{1}{N_0 d} \sum_{k=1}^{d} \sum_{\ell=1}^{N_0} \left\{ \breve{\beta}_k(u_\ell^0) - \beta_{k0}(u_\ell^0) \right\}$$

$$\mathrm{MSE}(\breve{\boldsymbol{\beta}}) = \frac{1}{N_0 d} \sum_{k=1}^{d} \sum_{\ell=1}^{N_0} \left\{ \breve{\beta}_k(u_\ell^0) - \beta_{k0}(u_\ell^0) \right\}^2, \quad \mathrm{MAD}(\breve{\boldsymbol{\beta}}) = \frac{1}{N_0 d} \sum_{k=1}^{d} \sum_{\ell=1}^{N_0} \left| \breve{\beta}_k(u_\ell^0) - \beta_{k0}(u_\ell^0) \right|,$$

where $\breve{\boldsymbol{\beta}}$ denotes an estimate of $\boldsymbol{\beta}_0$. The smaller the values of bias, MSE and MAD, the more accurate the underlying estimator is.

As this paper concerns more efficient nonparametric estimation of functional-coefficient models, we also look at the average standard deviations of the FCM, PFCM and PFCM + CD estimators of $\boldsymbol{\beta}_0$ over the grid points. The PFCM + CD estimates are expected to have the smallest standard deviations when there is medium or strong serial correlation.

We consider both balanced and unbalanced panels for the data generating process in simulation.



Flowchart for the estimation process

Case I: balanced panel data. Set $m_i \equiv 4$ and generate $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{i4})^\top$, $i \in \{1, \ldots, n\}$, independently from a 4-dimensional Gaussian distribution with mean $\mathbf{0}_4$ and covariance matrix $\boldsymbol{\Sigma}_0 = (\sigma_{jk})_{4\times 4}$, where $\sigma_{jk} = \iota^{|j-k|}$ for all $j, k \in \{1, \ldots, 4\}$. Such ε_{ij} are cross-sectionally independent but serially correlated with an AR(1) serial correlation structure. In order to investigate the performance of the three estimation methods under different levels of serial correlation, we choose $\iota = 0.1, 0.5, 0.9$. We also set $n = 50, 100, 200$.

With balanced data, we use the method in Section 4.1 to estimate the autoregressive coefficients in the Cholesky decomposition. To obtain initial estimates of these autoregressive coefficients, we first extract the residuals from the FCM method, i.e., $\widetilde{\varepsilon}_{ij} = Y_{ij} - \mathbf{X}_{ij}^\top \widetilde{\boldsymbol{\beta}}(u_{ij})$, then calculate the sample covariance matrix of $\widetilde{\boldsymbol{\varepsilon}}_i = (\widetilde{\varepsilon}_{i1}, \ldots, \widetilde{\varepsilon}_{i4})^\top$ for all $i \in \{1, \ldots, n\}$, and subsequently apply the Cholesky decomposition to this sample covariance matrix and extract the below-diagonal elements of the resulting lower triangular matrix. Table 1 summaries the average bias, MSEs, and MADs of the three estimation methods and their standard deviations (in parentheses) over 1000 replications for balanced panel under different combinations of n and ι. Table 3 gives the average standard deviations over the grid points between 0 and 1.

From Table 1, we find that the FCM method always has the lowest estimation accuracy (measured in MAD), as it ignores both the rank-reduced structure of $\boldsymbol{\beta}_0(u_{ij}) = \boldsymbol{\theta}_0 + \boldsymbol{\Theta}_0\boldsymbol{\gamma}_0(u_{ij})$ and the serial correlation present in the data. This, on the other hand, shows the benefit of making use of the principal functional-coefficient structure for reducing the dimension of functional coefficients to be estimated. When $\iota = 0.1$, the MADs of PFCM and PFCM + CD are close, which is not surprising because when the serial correlation is weak, the consequence of ignoring it in the estimation is not severe. However, as ι becomes larger (in particular when $\iota = 0.9$), the PFCM + CD outperforms the PFCM, even in small samples of size 50. The average standard deviations in Table 3 indicate that when there is medium to high serial correlation, the PFCM + CD is more efficient than the other two estimators. The above findings are consistent with the asymptotic theorems in Section 3.

Case II: Unbalanced panel data. With unbalanced data, the number of observations, m_i, for each cross-section varies across i. We follow [11] and generate unbalanced observation times in the following way: we first pre-schedule the observation times for each cross-section at 1, 2, 3, 4, 5; each observation point has a probability of 0.2 of being skipped (i.e., having no observation taken at that time); then for each non-skipped observation point a random perturbation (uniformly distributed over $[0, 1]$) is added. This results in unbalanced numbers of observations and different observation times for different cross-sections. All the other specifications of the data generating process are the same as those in Case I.

With unbalanced data, we use the nonparametric estimation method introduced in Section 4.2 to estimate the serial covariance matrices and then apply the Cholesky decomposition to the estimated matrices. As in Case I, we compare the performance of the three estimation methods: the FCM, PFCM and PFCM + CD. The simulation results are summarized in Tables 2 and 4, and the same findings can be observed as those for Case I with balanced data.

In the above simulation studies, the number of principal functional coefficients, d_0, is estimated using the simple ratio method introduced in (44). In order to evaluate the performance of this method, we report, in Tables 5 and 6, the frequencies at which d_0 is correctly estimated (i.e., the estimated value equals its true value) for the cases considered above. The results in these tables show that when the number of observations $N(n) = m_1 + \cdots + m_n$ is about $50 \times 4 = 200$, the percentage of replications in which d_0 is correctly estimated is around 70% for unbalanced data and 85% for balanced data. This percentage rises to around 98% when $N(n)$ increases to around 400 and further to almost 100% when $N(n)$ increases to around 800.

Table 1
Average biases, MADs and MSEs for Case I, with their standard deviations over 1000 replications given within the parentheses next to them.

ι \ n			n = 50		n = 100		n = 200	
$\iota = 0.1$	FCM	Bias	−0.0004	(0.0324)	−0.0010	(0.0213)	−0.0010	(0.0152)
		MAD	0.1644	(0.0183)	0.1213	(0.0119)	0.0818	(0.0083)
		MSE	0.0525	(0.0087)	0.0297	(0.0047)	0.0129	(0.0024)
	PFCM	Bias	0.0002	(0.0303)	−0.0014	(0.0193)	−0.0011	(0.0149)
		MAD	0.1462	(0.0197)	0.1062	(0.0115)	0.0723	(0.0080)
		MSE	0.0474	(0.0095)	0.0264	(0.0045)	0.0113	(0.0023)
	PFCM + CD	Bias	0.0005	(0.0310)	−0.0007	(0.0194)	−0.0001	(0.0148)
		MAD	0.1489	(0.0201)	0.1076	(0.0119)	0.0733	(0.0080)
		MSE	0.0490	(0.0097)	0.0272	(0.0047)	0.0117	(0.0023)
$\iota = 0.5$	FCM	Bias	−0.0005	(0.0346)	−0.0006	(0.0235)	−0.0008	(0.0152)
		MAD	0.1651	(0.0181)	0.1226	(0.0123)	0.0813	(0.0087)
		MSE	0.0532	(0.0091)	0.0306	(0.0051)	0.0128	(0.0027)
	PFCM	Bias	2.7×10^{-5}	(0.0326)	−0.0009	(0.0221)	−0.0010	(0.0145)
		MAD	0.1472	(0.0199)	0.1074	(0.0125)	0.0718	(0.0084)
		MSE	0.0481	(0.0099)	0.0272	(0.0050)	0.0112	(0.0025)
	PFCM + CD	Bias	0.0034	(0.0320)	0.0048	(0.0208)	0.0045	(0.0137)
		MAD	0.1463	(0.0191)	0.1074	(0.0120)	0.0733	(0.0085)
		MSE	0.0489	(0.0095)	0.0287	(0.0054)	0.0126	(0.0030)
$\iota = 0.9$	FCM	Bias	−0.0004	(0.0365)	0.0002	(0.0232)	−0.0002	(0.0163)
		MAD	0.1652	(0.0185)	0.1216	(0.0127)	0.0815	(0.0095)
		MSE	0.0547	(0.0095)	0.0305	(0.0059)	0.0131	(0.0032)
	PFCM	Bias	−0.0006	(0.0341)	-8.7×10^{-5}	(0.0229)	−0.0002	(0.0158)
		MAD	0.1481	(0.0216)	0.1071	(0.0122)	0.0722	(0.0093)
		MSE	0.0503	(0.0122)	0.0273	(0.0057)	0.0115	(0.0032)
	PFCM + CD	Bias	0.0071	(0.0287)	0.0122	(0.0186)	0.0132	(0.0124)
		MAD	0.1351	(0.0170)	0.0993	(0.0107)	0.0690	(0.0090)
		MSE	0.0480	(0.0106)	0.0295	(0.0071)	0.0146	(0.0047)

Table 2
Average biases, MADs and MSEs for Case II, with their standard deviations over 1000 replications given within the parentheses next to them.

ι \ n			n = 50		n = 100		n = 200	
$\iota = 0.1$	FCM	Bias	−0.0013	(0.0307)	−0.0013	(0.0198)	0.0002	(0.0144)
		MAD	0.1241	(0.0182)	0.0872	(0.0128)	0.0625	(0.0077)
		MSE	0.0259	(0.0075)	0.0126	(0.0035)	0.0064	(0.0015)
	PFCM	Bias	−0.0019	(0.0302)	−0.0011	(0.0194)	0.0003	(0.0142)
		MAD	0.1118	(0.0221)	0.0708	(0.0140)	0.0465	(0.0085)
		MSE	0.0217	(0.0081)	0.0086	(0.0033)	0.0036	(0.0012)
	PFCM + CD	Bias	−0.0012	(0.0306)	−0.0007	(0.0196)	0.0006	(0.0141)
		MAD	0.1143	(0.0222)	0.0719	(0.0142)	0.0467	(0.0084)
		MSE	0.0228	(0.0084)	0.0089	(0.0034)	0.0037	(0.0012)
$\iota = 0.5$	FCM	Bias	−0.0007	(0.0337)	−0.0010	(0.0206)	0.0010	(0.0152)
		MAD	0.1288	(0.0204)	0.0868	(0.0122)	0.0626	(0.0084)
		MSE	0.0283	(0.0091)	0.0125	(0.0034)	0.0065	(0.0018)
	PFCM	Bias	−0.0010	(0.0334)	−0.0013	(0.0200)	0.0009	(0.0148)
		MAD	0.1162	(0.0245)	0.0708	(0.0133)	0.0469	(0.0089)
		MSE	0.0237	(0.0096)	0.0087	(0.0031)	0.0037	(0.0014)
	PFCM + CD	Bias	0.0025	(0.0319)	0.0011	(0.0193)	0.0021	(0.0140)
		MAD	0.1165	(0.0237)	0.0703	(0.0133)	0.0459	(0.0089)
		MSE	0.0240	(0.0092)	0.0087	(0.0032)	0.0036	(0.0015)
$\iota = 0.9$	FCM	Bias	0.0029	(0.0340)	−0.0008	(0.0222)	−0.0011	(0.0156)
		MAD	0.1241	(0.0219)	0.0871	(0.0140)	0.0633	(0.0090)
		MSE	0.0268	(0.0098)	0.0129	(0.0041)	0.0066	(0.0019)
	PFCM	Bias	0.0022	(0.0329)	−0.0009	(0.0216)	−0.0011	(0.0154)
		MAD	0.1105	(0.0254)	0.0717	(0.0144)	0.0476	(0.0095)
		MSE	0.0222	(0.0103)	0.0091	(0.0038)	0.0039	(0.0016)
	PFCM + CD	Bias	0.0088	(0.0284)	0.0044	(0.0180)	0.0017	(0.0130)
		MAD	0.1000	(0.0224)	0.0633	(0.0130)	0.0394	(0.0087)
		MSE	0.0199	(0.0096)	0.0079	(0.0039)	0.0029	(0.0015)

Table 3
Standard deviations of estimators for Case I.

ι \ n		$n = 50$	$n = 100$	$n = 200$
	FCM	0.1268	0.0855	0.0605
$\iota = 0.1$	PFCM	0.1043	0.0653	0.0476
	PFCM + CD	0.1063	0.0658	0.0478
	FCM	0.1295	0.0876	0.0614
$\iota = 0.5$	PFCM	0.1076	0.0674	0.0485
	PFCM + CD	0.1033	0.0636	0.0447
	FCM	0.1322	0.0891	0.0630
$\iota = 0.9$	PFCM	0.1115	0.0697	0.0503
	PFCM + CD	0.0877	0.0506	0.0326

Table 4
Standard deviations of estimators for Case II.

ι \ n		$n = 50$	$n = 100$	$n = 200$
	FCM	0.1295	0.0941	0.0732
$\iota = 0.1$	PFCM	0.1139	0.0729	0.0524
	PFCM + CD	0.1139	0.0730	0.0523
	FCM	0.1374	0.0942	0.0737
$\iota = 0.5$	PFCM	0.1220	0.0732	0.0532
	PFCM + CD	0.1164	0.0695	0.0507
	FCM	0.1323	0.0953	0.0749
$\iota = 0.9$	PFCM	0.1154	0.0752	0.0545
	PFCM + CD	0.0934	0.0584	0.0417

Table 5
Frequency at which d_0 is corrected estimated for Case I.

ι \ n	$n = 50$	$n = 100$	$n = 200$
$\iota = 0.1$	87.00%	98.40%	100.00%
$\iota = 0.5$	85.80%	98.00%	100.00%
$\iota = 0.9$	85.80%	99.80%	100.00%

Table 6
Frequency at which d_0 is corrected estimated for Case II.

ι \ n	$n = 50$	$n = 100$	$n = 200$
$\iota = 0.1$	68.80%	97.40%	100.00%
$\iota = 0.5$	61.20%	98.40%	100.00%
$\iota = 0.9$	73.60%	98.40%	100.00%

6. An empirical application to economic growth data

As introduced in Section 1, the modeling framework and estimation methodology developed in this paper cover a special case where the index variable is time-invariant. In this section, we apply model (4) and our estimation methods to a cross-country economic growth study where the index variable is chosen as the initial income level of a country over the study period, which is invariant over time.

There has been extensive literature on the econometric analysis of economic growth; see [7] for a review of the literature. A wide range of statistical methods have been employed to study and identify structural patterns in growth data. With the increasing availability of panel data in this area, much of the research now uses panel data methods in their study. In cross-country growth studies, while a single model assuming parameter constancy is easier to implement and more straightforward to interpret, it ignores the heterogeneity that is well documented to exist across countries. One of the most commonly-used econometric models for accounting for heterogeneity is the linear fixed-effect model. Although the fixed-effect model does allow a certain degree of heterogeneity (for the model intercept term), it is limited in its capacity to capture the variation in growth structure across countries. Moreover, the fixed-effect model treats the heterogeneous individual effects as nuisance parameters and, as argued in [9], this treatment of individual effects is inappropriate due to them being of fundamental interest to studying growth differences. One approach that has been used by researchers to address this issue is to include interactions or nonlinear terms of the explanatory variables in the model so that the marginal effects of these explanatory variables can vary across countries and/or over time. In this paper, we use a different approach and employ the (principal) functional-coefficient model to the cross-country growth study. The dependency of the parameters on the index variable, which varies cross-sectionally, allows the model to capture cross-country heterogeneity in the effects of explanatory variables on growth.

The use of functional-coefficient models in growth study is not new. For example, Durlauf et al. [8] estimate a functional-coefficient version of the augmented Solow model and find considerable parameter heterogeneity, especially among the poorer countries. However, to the best of our knowledge, no studies have used a rank-reduced functional-coefficient model. Since there is no consent among economists on the determinants of growth and economic theories do not provide much guidance on this, many different explanatory variables have been used in growth regressions and found to be significant (different studies may produce a different set of significant explanatory variables). However, the inclusion of a large number of explanatory variables in a functional-coefficient model is clearly not advisable due to the large number of functional coefficients having to be estimated. The rank-reduced functional-coefficient model alleviates this problem by extracting and using only the first few principal functional coefficients.

Table 7
Description of variables.

Variable	Description
Y	Annual growth rate of real GDP per capita at 2005 constant prices (%)
u	Log of real GDP per capita at 2005 constant prices (log of 2005 $/person)
X_1	1 for the model intercept
X_2	Annual growth rate of population (%)
X_3	Price level of investment (PPP over investment/XRAT)
X_4	Consumption share of real GDP per capita at 2005 constant prices (%)
X_5	Government consumption share of real GDP per capita at 2005 constant prices(%)
X_6	Investment share of real GDP per capita at 2005 constant prices (%)
X_7	Openness at 2005 constant prices (%)
X_8	Age dependency ratio, young % of working-age population (%)
X_9	Fertility rate, total births per woman
X_{10}	Life expectancy at birth (years)

As we will see in our study below, there are common patterns in some of the functional coefficients and the first two principal functionals account for 99.9% of the total variation among all the 10 functional coefficients.

The data we use are from the Penn World Table and the World Bank World Development Indicators (WDI). They include annual growth rate of real GDP per capita (at 2005 constant prices), real GDP per capita and 9 other variables for 148 countries over the years 1986–2010. Since many variables exhibit little year-to-year variation, we take five-year averages of the data as is done in most panel growth studies. This gives a balanced panel data set with a time series length of 5. Since previous research (e.g., [4]) has found that initial economic conditions can explain a big proportion of cross-sectional variation, we use the log of real GDP per capita at the beginning of the sampling period (i.e., in the year of 1986) as the time-invariant index variable in model (4). A list of the variables used in this study is given in Table 7.

Given that the growth variables are likely to be correlated over time with the correlation structure unknown a priori, we use the methods proposed in Sections 2.4 and 4.1 to estimate the serial correlation and adjust for it in the estimation procedure. The number of principal functional coefficients, d_0, is selected as 2. This cuts the dimension of nonparametric functional coefficient vector from 10 (1 for intercept coefficient and 9 for slope coefficients) to 2. The selection of d_0 can be clearly seen from Figure 1, which depicts the eigenvalues $\widehat{\lambda}_k$ of the sample covariance matrix of the FCM estimates of $\boldsymbol{\beta}_0$ and their consecutive ratios $\widehat{\lambda}_{k+1}/\widehat{\lambda}_k$. The first two principal functional coefficients together account for 99.97% of the total variation. The preliminary (or initial) estimates of $\boldsymbol{\beta}_0$ from the FCM method are shown in Figure 2. One can observe from Figure 2 that there are some common patterns in the 10 estimated coefficient functions: (i) the first pattern, typically seen in $\widetilde{\beta}_1(u)$, $\widetilde{\beta}_2(u)$, $\widetilde{\beta}_9(u)$ and $\widetilde{\beta}_{10}(u)$ (pattern reversed for $\widetilde{\beta}_9(u)$ and $\widetilde{\beta}_{10}(u)$), has two saddle points and goes sharply upwards for smaller values of u ($u \leq 6.4$ approximately) then downwards before going upwards again for larger values of u ($u \geq 8$ approximately), or the reverse of the above; (ii) the second pattern, typically seen in $\widetilde{\beta}_5(u)$, $\widetilde{\beta}_6(u)$ and $\widetilde{\beta}_7(u)$, has a general upward or downward trend with small variations. These typical patterns can be more clearly observed in Figure 3, which plots the PFCM + CD estimates of $\boldsymbol{\beta}_0(u_{ij})$, and are generally captured by the principal coefficient functions shown in Figure 4. Linear combinations of the two principal functions can provide a good description of the 10 functional coefficients.

All the functional coefficients in Figures 2 and 3 show a large amount of variation, especially for smaller values of u. This implies that there is higher heterogeneity in the effects of the explanatory variables on growth across poorer countries (with lower income levels). This is consistent with findings from previous growth studies such as [8]. The estimated intercept coefficient, $\widetilde{\beta}_1(u_{ij})$ or $\widetilde{\beta}_1(u_{ij})$, shows a positive relationship between the initial income level and economic growth for smaller values of u, negative for medium values and positive again for larger values. This is again consistent with findings by Durlauf et al. [8] and Liu and Stengos [28], who find a positive relationship for smaller values of u and a negative relationship for medium values ($6.4 \leq u \leq 9$ approximately). Note that we use data from a much larger set of countries (148 countries) and covering a more recent time period (1986–2010) than [8] (98 countries over 1960–85) and [28] (86 countries over 1960s, 1970s and 1980s).

The PFCM + CD estimates of $\boldsymbol{\theta}_0$ and $\boldsymbol{\Theta}_0$ are listed in Table 8 and their standard errors are shown in parentheses. To again compare the three methods: FCM, PFCM and PFCM + CD, we present in Table 9 the in-sample mean squared

Table 8
PFCM + CD Estimates of $\boldsymbol{\theta}_0$ and $\boldsymbol{\Theta}_0$, with their standard errors given within the parentheses next to them.

$\boldsymbol{\theta}_0$		$\boldsymbol{\Theta}_0(1)$		$\boldsymbol{\Theta}_0(2)$	
4.0013	(2.6550)	−0.9462	(0.9093)	−0.1582	(5.5816)
0.0511	(0.1469)	−0.1951	(0.0541)	0.9573	(0.2762)
0.0015	(0.0040)	−0.0002	(0.0015)	−0.0050	(0.0054)
−0.0341	(0.0090)	0.0016	(0.0028)	0.0110	(0.0177)
−0.0800	(0.0244)	−0.0026	(0.0076)	0.1686	(0.0533)
0.0473	(0.0162)	−0.0049	(0.0060)	0.0806	(0.0347)
0.0077	(0.0031)	0.0001	(0.0009)	0.0124	(0.0065)
0.0276	(0.0240)	−0.0054	(0.0083)	−0.0189	(0.0500)
−1.1393	(0.3227)	0.2578	(0.1118)	0.1482	(0.6393)
0.0206	(0.0310)	0.0119	(0.0092)	−0.0317	(0.0608)

Table 9
In-sample MSEs and average out-of-sample MSPEs

Measure \ Method	FCM	PFCM	PFCM+CD
In-sample MSE	10.3159	10.5190	10.6262
Average out-of-sample MSPE	14.7575	13.7189	13.4521

errors (MSE) and the average out-of-sample mean squared prediction errors (MSPE) (calculated from 10 repeats of 10-fold cross validation). For the in-sample MSE, the full sample is used for estimation and then the mean of squared residuals is calculated. The 10-fold cross validation randomly splits the 148 countries into 10 subsets of roughly equal size (each subset has 14-15 countries) and uses data from 9 subsets of countries for estimation and those from the remaining subset of countries for testing, from which the MSPE is computed. Since data splitting is randomly taken, the out-of-sample MSPE varies for each repeat of the 10-fold cross-validation. From Table 9, we find that the FCM method has the best fit in terms of in-sample MSE, whereas the PFCM-based methods show their advantage in terms of out-of-sample MSPE.

7. Conclusion

This paper introduces a novel rank-reduced functional-coefficient modeling approach for analyzing panel data. By imposing a rank-reduced structure on the functional coefficients, we achieve dimension reduction on the nonparametric coefficient functions. Both the parameters and principal functional coefficients in the model are estimated through an iterative semiparametric procedure. To account for serial correlation in the estimation, we first apply Cholesky decomposition to the serial covariance matrices and then use the elements in the lower triangular matrix from the Cholesky decomposition to construct a model transformation whose errors are free from serial correlation.

Based on this transformed model, we obtain an asymptotically more efficient estimator for the (principal) functional coefficients. Under some regularity conditions, we establish the asymptotic distribution theory for the proposed parametric and nonparametric estimators. The theory shows that the serial covariance matrix needs to be correctly specified up to a constant multiple for the Cholesky decomposition based method to achieve efficiency improvement. Since the serial covariance matrix is usually unknown, we further propose two different approaches to its consistent estimation for balanced and unbalanced panel data.

Simulation studies as well as an empirical application show that the developed semiparametric modeling and estimation methods work reasonably well in finite samples. In particular, from the empirical application, we find that the proposed principal functional-coefficient panel models with reduced number of unknown nonparametric components have more accurate out-sample prediction of the economic growth than the conventional functional-coefficient panel models, and are thus recommended in practical applications when the main interest lies in the out-sample panel prediction.

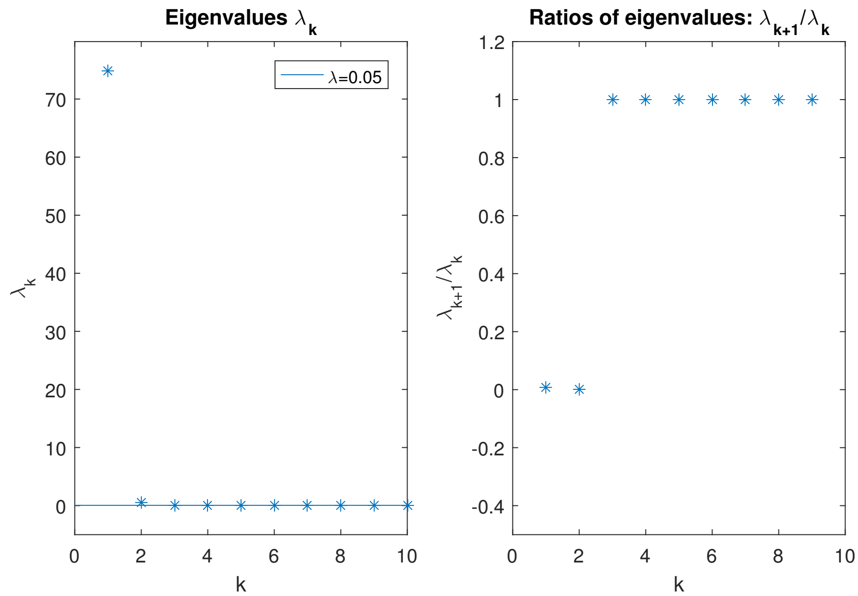


Figure 1: Eigenvalues of the sample covariance matrix of the FCM estimates of $\beta_{1,0}, \ldots, \beta_{10,0}$ and their consecutive ratios.

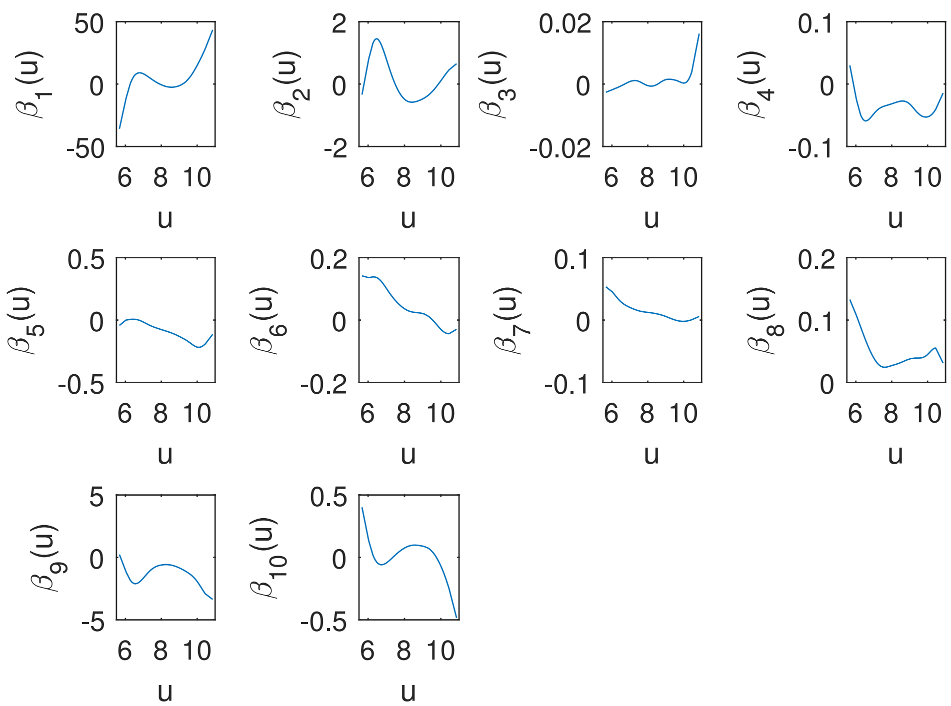


Figure 2: FCM estimates of the functional coefficients.

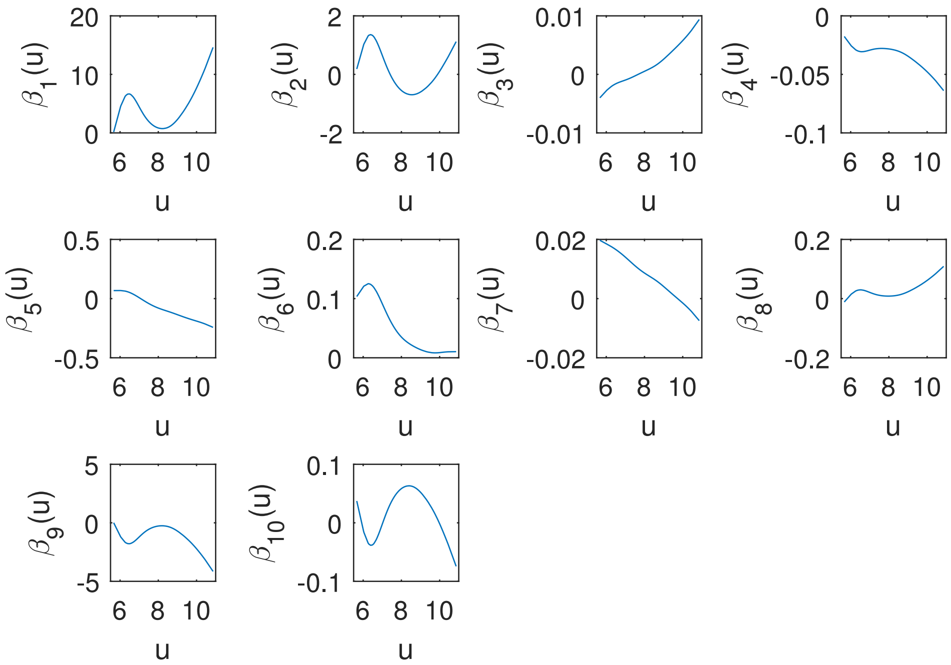


Figure 3: PFCM + CD estimates of the functional coefficients.

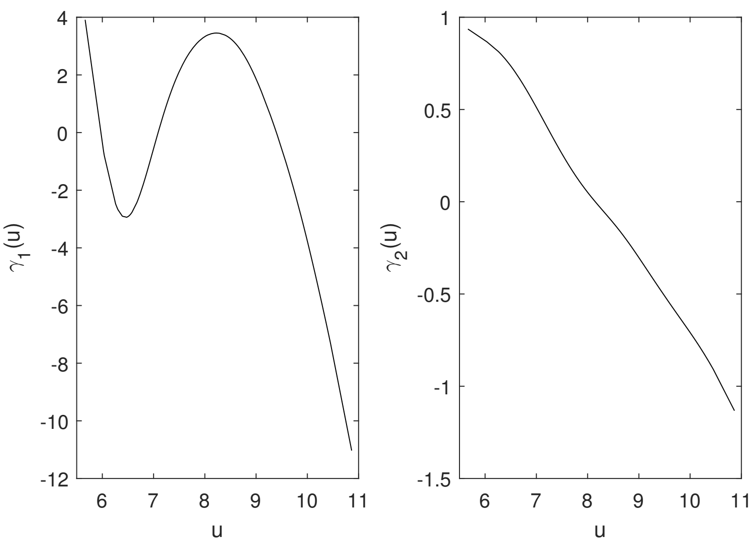


Figure 4: PFCM + CD estimates of the two principal functional coefficients.

Acknowledgments

The authors would like to thank the Editor-in-Chief, Prof. Christian Genest, an Associate Editor and two anonymous referees for their comments and suggestions, which substantially improve an earlier version of the paper.

Appendix: 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 each $\ell \in \{0,1,2\}$, we define

$$\mathbf{S}_{n\ell}(u|\mathbf{\Theta}) = \frac{1}{N(n)h}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\Big(\frac{u_{ij}-u}{h}\Big)^{\ell}\mathbf{X}_{ij}(\mathbf{\Theta})\mathbf{X}_{ij}^{\top}(\mathbf{\Theta})K\Big(\frac{u_{ij}-u}{h}\Big),$$

$$\mathbf{T}_{n\ell}(u|\boldsymbol{\theta},\mathbf{\Theta}) = \frac{1}{N(n)h}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\Big(\frac{u_{ij}-u}{h}\Big)^{\ell}\mathbf{X}_{ij}(\mathbf{\Theta})(Y_{ij}-\mathbf{X}_{ij}^{\top}\boldsymbol{\theta})K\Big(\frac{u_{ij}-u}{h}\Big),$$

and

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

Define $\mathbf{\Delta}_{1\mathbf{\Theta}}(u|\boldsymbol{\theta}_0-\boldsymbol{\theta}) = \mathbf{E}\{\mathbf{X}_{ij}(\mathbf{\Theta})\mathbf{X}_{ij}^{\top}(\boldsymbol{\theta}_0-\boldsymbol{\theta})|u_{ij}=u\}$, $\mathbf{\Delta}_{2\mathbf{\Theta}}(u|\mathbf{\Theta}_0-\mathbf{\Theta}) = \mathbf{E}\{\mathbf{X}_{ij}(\mathbf{\Theta})\mathbf{X}_{ij}^{\top}(\mathbf{\Theta}_0-\mathbf{\Theta})|u_{ij}=u\}$, and

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

We start with two technical lemmas. The first lemma gives the asymptotic expansion of the local linear estimate $\widehat{\boldsymbol{\gamma}}(u|\boldsymbol{\theta},\mathbf{\Theta})$ and is a generalization of Lemma 1 in [19] to the panel data context. The proofs of the lemmas can be found in the Online Supplement.

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

$$\begin{aligned}\widehat{\boldsymbol{\gamma}}(u|\boldsymbol{\theta},\mathbf{\Theta}) - \boldsymbol{\gamma}_0(u) = {}& \mu_2 h^2\boldsymbol{\gamma}_0''(u)/2 + \mathbf{\Delta}_1^{+}(u|\mathbf{\Theta})\{\mathbf{\Delta}_{1\mathbf{\Theta}}(u|\boldsymbol{\theta}_0-\boldsymbol{\theta}) + \mathbf{\Delta}_{2\mathbf{\Theta}}(u|\mathbf{\Theta}_0-\mathbf{\Theta})\boldsymbol{\gamma}_0(u)\} \\ & + \mathbf{\Delta}_1^{+}(u|\mathbf{\Theta})\mathbf{T}_{n,\varepsilon}(u|\mathbf{\Theta}) + o_P(\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| + \|\mathbf{\Theta}-\mathbf{\Theta}_0\|) \\ & + O_P(h^3 + h\xi_n + \xi_n^2) \end{aligned} \tag{46}$$

uniformly in $u \in \mathcal{U}$, where $\mathcal{U}$ is the compact support of u_{ij} defined in Assumption 2(ii), and $\xi_n = \{\ln(h^{-1})/(nh)\}^{1/2}$.

Lemma A.2. *Suppose that Assumptions 1–3 are satisfied. Then we have*

$$\sqrt{N(n)}\begin{bmatrix} \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \\ \mathrm{vec}(\widehat{\mathbf{\Theta}}) - \mathrm{vec}(\mathbf{\Theta}_0) \end{bmatrix} = \mathbf{W}_1^{+}\mathbf{V}_n\{1 + o_P(1)\}, \tag{47}$$

where $\mathbf{V}_n$ and $\mathbf{W}_1$ are defined in Section 3.

Proof of Proposition 1. By Lemma A.2 and applying the Central Limit Theorem on $\mathbf{V}_n$, we readily prove (22). By Lemma A.1 and Eq. (22), we have

$$\begin{aligned}\widehat{\boldsymbol{\gamma}}(u) - \boldsymbol{\gamma}_0(u) &= \mu_2 h^2\boldsymbol{\gamma}_0''(u)/2 + \mathbf{\Delta}_1^{+}(u|\mathbf{\Theta}_0)\mathbf{T}_{n,\varepsilon}(u|\mathbf{\Theta}_0) + O_P(n^{-1/2} + h^3 + h\xi_n + \xi_n^2), \\ &= \mu_2 h^2\boldsymbol{\gamma}_0''(u)/2 + \mathbf{\Delta}_1^{+}(u|\mathbf{\Theta}_0)\mathbf{T}_{n,\varepsilon}(u|\mathbf{\Theta}_0) + o_P\{(nh)^{-1/2}\}. \end{aligned} \tag{48}$$

Using the Central Limit Theorem again, we may show that

$$\sqrt{N(n)h}\,\mathbf{T}_{n,\varepsilon}(u|\mathbf{\Theta}_0) = \frac{1}{\sqrt{N(n)h}f(u)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\mathbf{X}_{ij}(\mathbf{\Theta}_0)\varepsilon_{ij}K\Big(\frac{u_{ij}-u}{h}\Big) \rightsquigarrow \mathcal{N}[\mathbf{0}_{d_0},\ \omega(u)\mathbf{\Delta}_1(u|\mathbf{\Theta}_0)]. \tag{49}$$

We then complete the proof of (23) by using (48) and (49).

The proof of Proposition 1 has thus been completed. □

Proof of Theorem 1. Define

$$\mathbf{S}_{n\ell}^{\diamond}(u|\mathbf{\Theta}) = \frac{1}{N(n)h}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\Big(\frac{u_{ij}-u}{h}\Big)^{\ell}\mathbf{X}_{ij}(\mathbf{\Theta})\mathbf{X}_{ij}^{\top}(\mathbf{\Theta})(\rho_{ij}^{\diamond})^{-1}K\Big(\frac{u_{ij}-u}{h}\Big),$$

$$\mathbf{T}_{n\ell}^{\diamond}(u|\boldsymbol{\theta},\mathbf{\Theta}) = \frac{1}{N(n)h}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\Big(\frac{u_{ij}-u}{h}\Big)^{\ell}\mathbf{X}_{ij}(\mathbf{\Theta})(\widetilde{Y}_{ij}-\mathbf{X}_{ij}^{\top}\boldsymbol{\theta})(\rho_{ij}^{\diamond})^{-1}K\Big(\frac{u_{ij}-u}{h}\Big)$$

and

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

Using the same argument as in the proof of Lemma A.1, we have

$$\overline{\boldsymbol{\gamma}}(u) = (\mathbf{I}_{d_0},\ \mathbf{N}_{d_0})\{\mathbf{S}_{n}^{\diamond}(u|\overline{\mathbf{\Theta}})\}^{+}\mathbf{T}_{n}^{\diamond}(u|\overline{\boldsymbol{\theta}},\overline{\mathbf{\Theta}}), \tag{50}$$

where $\mathbf{N}_{d_0}$ is a $d_0\times d_0$ null matrix. As $\overline{\mathbf{\Theta}}$ is assumed to be $\sqrt{n}$-consistent, we may show that

$$\mathbf{S}_{n}^{\diamond}(u|\overline{\mathbf{\Theta}}) = \mathbf{S}_{n}^{\diamond}(u|\mathbf{\Theta}_0) + O_P(n^{-1/2}) = \mathbf{\Delta}_{n}^{\diamond}(u|\mathbf{\Theta}_0) + O_P\{(nh)^{-1/2}+h\}, \tag{51}$$

where

$$\begin{aligned}\mathbf{\Delta}_{n}^{\diamond}(u|\mathbf{\Theta}) &= \mathrm{diag}(1,\mu_2)\otimes\frac{f(u)}{N(n)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\mathrm{E}\{\mathbf{X}_{ij}(\mathbf{\Theta})\mathbf{X}_{ij}^{\top}(\mathbf{\Theta})(\rho_{ij}^{\diamond})^{-1}|u_{ij}=u\} \\ &= f(u)\left\{\frac{1}{N(n)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}(\rho_{ij}^{\diamond})^{-1}\right\}\times\mathrm{diag}(1,\mu_2)\otimes\mathbf{\Delta}_1(u|\mathbf{\Theta}).\end{aligned}$$

Note that

$$\begin{aligned}Y_{ij}-\mathbf{X}_{ij}^{\top}\boldsymbol{\theta} &= \mathbf{X}_{ij}^{\top}(\boldsymbol{\theta}_0-\boldsymbol{\theta}) + \sum_{k=1}^{d_0}\gamma_{k0}(u_{ij})\mathbf{X}_{ij}^{\top}\mathbf{\Theta}_0(k)+\varepsilon_{ij} \\ &= \mathbf{X}_{ij}^{\top}(\boldsymbol{\theta}_0-\boldsymbol{\theta}) + \sum_{k=1}^{d_0}\gamma_{k0}(u_{ij})\mathbf{X}_{ij}^{\top}\{\mathbf{\Theta}_0(k)-\mathbf{\Theta}(k)\} + \sum_{k=1}^{d_0}\gamma_{k0}(u_{ij})\mathbf{X}_{ij}^{\top}\mathbf{\Theta}(k)+\varepsilon_{ij}.\end{aligned} \tag{52}$$

By (18) and (52), we have

$$\widetilde{Y}_{ij}-\mathbf{X}_{ij}^{\top}\overline{\boldsymbol{\theta}} = \eta_{ij} + \mathbf{X}_{ij}^{\top}(\boldsymbol{\theta}_0-\overline{\boldsymbol{\theta}}) + \sum_{k=1}^{d_0}\gamma_{k,0}(u_{ij})\mathbf{X}_{ij}^{\top}\mathbf{\Theta}_0(k) + \sum_{k=1}^{j-1}(c_{i,jk,0}-c_{i,jk}^{\diamond})\varepsilon_{ik} + \sum_{k=1}^{j-1}c_{i,jk}^{\diamond}(\varepsilon_{ik}-\widetilde{\varepsilon}_{ik}), \tag{53}$$

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 $\widetilde{Y}_{ij}-\mathbf{X}_{ij}^{\top}\overline{\boldsymbol{\theta}}$ replaced by η_{ij}, $\mathbf{X}_{ij}^{\top}(\boldsymbol{\theta}_0-\overline{\boldsymbol{\theta}})$, $\sum_{k=1}^{d_0}\gamma_{k,0}(u_{ij})\mathbf{X}_{ij}^{\top}\mathbf{\Theta}_0(k)$, $\sum_{k=1}^{j-1}(c_{i,jk,0}-c_{i,jk}^{\diamond})\varepsilon_{ik}$ and $\sum_{k=1}^{j-1}c_{i,jk}^{\diamond}(\varepsilon_{ik}-\widetilde{\varepsilon}_{ik})$, respectively. Then, by (50)–(53), we can show that

$$\overline{\boldsymbol{\gamma}}(u) = (\mathbf{I}_{d_0},\ \mathbf{N}_{d_0})\{\mathbf{S}_{n}^{\diamond}(u|\overline{\mathbf{\Theta}})\}^{+}\left\{\sum_{k=1}^{5}\mathbf{T}_{n}^{\diamond}(u,k)\right\} = (\mathbf{I}_{d_0},\ \mathbf{N}_{d_0})\{\mathbf{S}_{n}^{\diamond}(u|\mathbf{\Theta}_0)\}^{+}\left\{\sum_{k=1}^{5}\mathbf{T}_{n}^{\diamond}(u,k)\right\} + O_P(n^{-1/2}). \tag{54}$$

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

$$(\mathbf{I}_{d_0},\ \mathbf{N}_{d_0})\,\{\mathbf{S}_n^{\diamond}(u|\boldsymbol{\Theta}_0)\}^{+}\,\mathbf{T}_n^{\diamond}(u,2) = O_P(n^{-1/2}) = o_P\{(nh)^{-1/2}\}. \tag{55}$$

By some standard calculations, we can also prove that

$$(\mathbf{I}_{d_0},\ \mathbf{N}_{d_0})\,\{\mathbf{S}_n^{\diamond}(u|\boldsymbol{\Theta}_0)\}^{+}\,\mathbf{T}_n^{\diamond}(u,5) = o_P\{(nh)^{-1/2}\}. \tag{56}$$

The proof of (56) will be given later in this Appendix. Applying the Taylor expansion to the principal coefficient functions in $\mathbf{T}_n^{\diamond}(u,3)$, we obtain, uniformly in $u \in \mathcal{U}$,

$$(\mathbf{I}_{d_0},\ \mathbf{N}_{d_0})\{\mathbf{S}_n^{\diamond}(u|\boldsymbol{\Theta}_0)\}^{+}\mathbf{T}_n^{\diamond}(u,3) = \boldsymbol{\gamma}_0(u) + \mu_2 h^2 \boldsymbol{\gamma}_0''(u)/2 + o_P\{h^2 + (nh)^{-1/2}\}. \tag{57}$$

Let $e_{ij} = \eta_{ij} + \sum_{k=1}^{j-1}(c_{i,jk,0} - c_{i,jk}^{\diamond})\varepsilon_{ik}$. By the Central Limit Theorem, we can show that

$$\frac{1}{\sqrt{N(n)h}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\mathbf{X}_{ij}(\boldsymbol{\Theta}_0)e_{ij}(\rho_{ij}^{\diamond})^{-1}K\Big(\frac{u_{ij}-u}{h}\Big) \rightsquigarrow \mathcal{N}[\mathbf{0},\ \boldsymbol{\Omega}_{\diamond}(u|\boldsymbol{\Theta}_0)],$$

where

$$\boldsymbol{\Omega}_{\diamond}(u|\boldsymbol{\Theta}_0) = f(u)\nu_0\boldsymbol{\Delta}_1(u|\boldsymbol{\Theta}_0)\lim_{n\to\infty}\frac{1}{N(n)}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{\tau_{ij}}{(\rho_{ij}^{\diamond})^2} = f(u)\nu_0 c_\tau \boldsymbol{\Delta}_1(u|\boldsymbol{\Theta}_0),$$

$\tau_{ij} = \mathrm{E}(e_{ij}^2)$. This indicates that

$$\sqrt{N(n)h}\,(\mathbf{I}_{d_0},\ \mathbf{N}_{d_0})\,\{\mathbf{S}_n^{\diamond}(u|\boldsymbol{\Theta}_0)\}^{+}\,\{\mathbf{T}_n^{\diamond}(u,1) + \mathbf{T}_n^{\diamond}(u,4)\} \rightsquigarrow \mathcal{N}[\mathbf{0}_{d_0},\ \omega_{\diamond}(u)\boldsymbol{\Delta}_1^{+}(u|\boldsymbol{\Theta}_0)], \tag{58}$$

where $\omega_{\diamond}(u)$ is defined in (26). By (54)–(58), we complete the proof of (25). □

We next give the proof of (56), which shows that the influence of replacing ε_{ik} by $\widetilde{\varepsilon}_{ik}$ can be ignored asymptotically.

Proof of (56). Recall that $\widetilde{\boldsymbol{\beta}}$ is a local linear estimation of $\boldsymbol{\beta}_0$ with the kernel function K and bandwidth b, i.e.,

$$\widetilde{\boldsymbol{\beta}}(u) = (\mathbf{I}_d,\ \mathbf{N}_d)\mathbf{S}_n^{+}(u)\mathbf{T}_n(u), \tag{59}$$

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

$$\varepsilon_{ik} - \widetilde{\varepsilon}_{ik} = \mathbf{X}_{ik}^{\top}\{\widetilde{\boldsymbol{\beta}}(u_{ik}) - \boldsymbol{\beta}_0(u_{ik})\},$$

which indicates that

$$\mathbf{T}_{n\ell}^{\diamond}(u,5) \quad = \quad \frac{1}{N(n)h}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\Big(\frac{u_{ij}-u}{h}\Big)^{\ell}\mathbf{X}_{ij}(\overline{\boldsymbol{\Theta}})(\rho_{ij}^{\diamond})^{-1}K\Big(\frac{u_{ij}-u}{h}\Big)\sum_{k=1}^{j-1}c_{i,jk}^{\diamond}\mathbf{X}_{ik}^{\top}\Big\{\widetilde{\boldsymbol{\beta}}(u_{ik}) - \boldsymbol{\beta}_0(u_{ik})\Big\} \tag{60}$$

for $\ell \in \{0,1\}$, 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 (56), we only need to show that

$$\mathbf{T}_{n\ell}^{\diamond}(u,5) = o_P\{(nh)^{-1/2}\} \tag{61}$$

for $\ell \in \{0,1\}$. We next only prove (61) for the case of $\ell = 0$ as the case of $\ell = 1$ can be proved in the same way. As $\overline{\boldsymbol{\Theta}}$ is $\sqrt{n}$-consistent, we may replace $\overline{\boldsymbol{\Theta}}$ by $\boldsymbol{\Theta}_0$ in this proof. Furthermore, noting that

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

where $\mathbf{\Delta}(u)$ is defined in Assumption 4(i) and $\xi_n^\star = \{\ln(b^{-1})/(nb)\}^{1/2}$, we may show that

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

Then, by (60) and Assumption 4(ii), we have

$$\begin{aligned}
\mathbf{T}_{n0}^\diamond(u,5) &= \frac{1}{N^2(n)hb} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \mathbf{X}_{ij}(\mathbf{\Theta}_0) K\left(\frac{u_{ij} - u}{h}\right) \sum_{k=1}^{j-1} c_{i,jk}^\diamond \sum_{p=1}^{n} \sum_{q=1}^{m_p} \varepsilon_{pq} \mathbf{X}_{ik}^\top \\
&\qquad \times f^{-1}(u_{ik}) \mathbf{\Delta}^+(u_{ik}) \mathbf{X}_{pq} K\left(\frac{u_{pq} - u_{ik}}{b}\right) + O_P\{b^2 + (b + \xi_n^\star)\xi_n^\star\} \\
&= \frac{1}{N^2(n)hb} \sum_{p=1}^{n} \sum_{q=1}^{m_p} \varepsilon_{pq} \nu_{pq}(u) + o_P\{h^2 + (nh)^{-1/2}\},
\end{aligned} \tag{62}$$

where

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

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

$$\begin{aligned}
\mathrm{E}\Big[\Big\{\sum_{p=1}^{n} \sum_{q=1}^{m_p} \varepsilon_{pq} \nu_{pq}(u)\Big\}^2\Big] &= \sum_{p=1}^{n} \sum_{q_1=1}^{m_p} \sum_{q_2=1}^{m_p} \mathrm{E}\{\varepsilon_{pq_1} \varepsilon_{pq_2} \nu_{pq_1}(u) \nu_{pq_2}(u)\} \\
&= O\Bigg[\sum_{p=1}^{n} \sum_{q_1=1}^{m_p} \sum_{q_2=1}^{m_p} \sum_{i_1=1}^{n} \sum_{i_2=1,\neq i_1}^{n} \sum_{j_1=1}^{m_{i_1}} \sum_{j_2=1}^{m_{i_2}} \sum_{k_1=1}^{j_1-1} \sum_{k_2=1}^{j_2-1} \mathrm{E}\Big\{K\Big(\frac{u_{i_1 j_1} - u}{h}\Big) \\
&\qquad \times K\Big(\frac{u_{pq_1} - u_{i_1 k_1}}{b}\Big) K\Big(\frac{u_{i_2 j_2} - u}{h}\Big) K\Big(\frac{u_{pq_2} - u_{i_2 k_2}}{b}\Big)\Big\}\Bigg] \\
&\quad + O\Bigg[\sum_{p=1}^{n} \sum_{q_1=1}^{m_p} \sum_{q_2=1}^{m_p} \sum_{i_1=1}^{n} \sum_{j_1=1}^{m_{i_1}} \sum_{j_2=1}^{m_{i_1}} \sum_{k_1=1}^{j_1-1} \sum_{k_2=1}^{j_2-1} \mathrm{E}\Big\{K\Big(\frac{u_{i_1 j_1} - u}{h}\Big) \\
&\qquad \times K\Big(\frac{u_{pq_1} - u_{i_1 k_1}}{b}\Big) K\Big(\frac{u_{i_1 j_2} - u}{h}\Big) K\Big(\frac{u_{pq_2} - u_{i_1 k_2}}{b}\Big)\Big\}\Bigg] \\
&= O(n^3 h^2 b^2 + n^2 hb),
\end{aligned}$$

which indicates that

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

Therefore, we complete the proofs of (61) and (56). $\square$

In order to prove Proposition 2, we need to use the following two technical lemmas, which are similar to Lemmas A.1 and A.2. Their proofs can again be found in the Online Supplement. Define

$$\mathbf{T}_{n,\eta}(u|\mathbf{\Theta}) = \frac{1}{n(m-1)h_1 f(u)} \sum_{i=1}^{n} \sum_{j=2}^{m} \mathbf{X}_{ij}(\mathbf{\Theta}) \eta_{ij} K\Big(\frac{u_{ij} - u}{h_1}\Big)$$

and

$$\mathbf{T}_{n,\varepsilon}(u|\mathbf{\Theta}, \mathbf{c}) = \frac{1}{n(m-1)h_1 f(u)} \sum_{i=1}^{n} \sum_{j=2}^{m} \mathbf{X}_{ij}(\mathbf{\Theta}) K\Big(\frac{u_{ij} - u}{h_1}\Big) \sum_{k=1}^{j-1} c_{jk,0}(\varepsilon_{ik} - \widetilde{\varepsilon}_{ik}).$$

Lemma A.3. *Suppose that the conditions in Proposition 2 are satisfied. Then we have*

$$\check{\boldsymbol{\gamma}}(u|\boldsymbol{\theta},\boldsymbol{\Theta},\mathbf{c})-\boldsymbol{\gamma}_0(u)=\mu_2h_1^2\boldsymbol{\gamma}_0''(u)/2+\boldsymbol{\Delta}_1^+(u|\boldsymbol{\Theta})\{\boldsymbol{\Delta}_{1\boldsymbol{\Theta}}(u|\boldsymbol{\theta}_0-\boldsymbol{\theta})+\boldsymbol{\Delta}_{2\boldsymbol{\Theta}}(u|\boldsymbol{\Theta}_0-\boldsymbol{\Theta})\boldsymbol{\gamma}_0(u)\}$$
$$+\boldsymbol{\Delta}_1^+(u|\boldsymbol{\Theta})\{\mathbf{T}_{n,\eta}(u|\boldsymbol{\Theta})+\mathbf{T}_{n,\varepsilon}(u|\boldsymbol{\Theta},\mathbf{c})\}\quad+o_P(\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|+\|\boldsymbol{\Theta}-\boldsymbol{\Theta}_0\|$$
$$+\|\mathbf{c}-\mathbf{c}_0\|)+O_P(h_1^3+h_1\xi_{n1}+\xi_{n1}^2)\quad(63)$$

uniformly in $u\in\mathcal{U}$, where $\xi_{n1}=\{\ln(h_1^{-1})/(nh_1)\}^{1/2}$ and the remaining notation is the same as that in Lemma A.1.

Let

$$\mathbf{F}_{ij}=\left[\mathbf{0}^\top_{\frac{(j-2)(j-1)}{2}},\varepsilon_{i1},\ldots,\varepsilon_{ij-1},\mathbf{0}^\top_{\frac{m(m-1)}{2}-\frac{j(j-1)}{2}}\right]^\top,\quad \mathbf{W}_\mathbf{F}=\mathrm{diag}\left\{\mathbf{W}_1,\ \frac{1}{m-1}\sum_{j=2}^m\mathrm{E}(\mathbf{F}_{1j}\mathbf{F}_{1j}^\top)\right\},$$

where $\mathbf{W}_1$ is defined as in Section 3.

Lemma A.4. *Suppose that the conditions in Proposition 2 are satisfied. Then we have*

$$\sqrt{n(m-1)}\begin{bmatrix}\overline{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\\ \mathrm{vec}(\overline{\boldsymbol{\Theta}})-\mathrm{vec}(\boldsymbol{\Theta}_0)\\ \overline{\mathbf{c}}-\mathbf{c}_0\end{bmatrix}=\mathbf{W}_\mathbf{F}^+\mathbf{V}_n^\sharp\{1+o_P(1)\},\quad(64)$$

where $\mathbf{V}_n^\sharp$ is a $\{d(d_0+1)+m(m-1)/2\}$-dimensional random vector satisfying $\mathbf{V}_n^\sharp=O_P(1)$.

Proof of Proposition 2. By Lemma A.4, we readily have (35) in Proposition 2. □

References

[1] L. Boneva, O. Linton, M. Vogt, A semiparametric model for heterogeneous panel data with fixed effects, J. Econometrics 188 (2015) 327–345.
[2] Z. Cai, Q. Li, Nonparametric estimation of varying coefficient dynamic panel data models, Econom. Theory 24 (2008) 1321–1342.
[3] Z. Cai, J. Fan, R. Li, Efficient estimation and inferences for varying-coefficient models, J. Amer. Statist. Assoc. 95 (2000) 888–902.
[4] F. Canova, A. Marcet, The poor stay poor: Non-convergence across countries and regions, Centre for Economic Policy Research Discussion Paper 1265, 1995.
[5] J. Chen, J. Gao, D. Li, Estimation in partially linear single-index panel data models with fixed effects, J. Bus. Econ. Statist. 31 (2013) 315–330.
[6] R. Chen, R.S. Tsay, Functional coefficient autoregressive models, J. Amer. Statist. Assoc. 88 (1993) 298–308.
[7] S. Durlauf, P. Johnson, J. Temple, Growth econometrics, in: P. Aghion, S. Durlauf (Eds), Handbook of Economic Growth, Vol. 1A, North-Holland, Amsterdam, 2005.
[8] S. Durlauf, A. Kourtellos, A. Minkin, The local Solow growth model, Eur. Econ. Rev. 45 (2001) 928–940.
[9] S. Durlauf, D. Quah, The new empirics of economic growth, in: J. Taylor, M. Woodford (Eds), Handbook of Macroeconomics, North-Holland, Amsterdam, 1999.
[10] J. Fan, I. Gijbels, Local Polynomial Modelling and Its Applications, Chapman & Hall, London, 1996.
[11] J. Fan, T. Huang, R. Li, Analysis of longitudinal data with semiparametric estimation of covariance function, J. Amer. Statist. Assoc. 102 (2007) 632–641.
[12] J. Fan, R. Li, New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis, J. Amer. Statist. Assoc. 99 (2004) 710–723.
[13] J. Fan, W. Zhang, Statistical estimation in varying coefficient models, Ann. Statist. 27 (1999) 1491–1518.
[14] J. Fan, W. Zhang, Statistical methods with varying coefficient models, Statist. Interface 1 (2008) 179–195.
[15] P. Hall, H.-G. Müller, F. Yao, Modelling sparse generalized longitudinal observations with latent Gaussian processes, J. R. Stat. Soc. Ser. B (Stat. Methodol.) 70 (2008) 703–723.
[16] D. Henderson, R. Carroll, Q. Li, Nonparametric estimation and testing of fixed effects panel data models, J. Econometrics 144 (2008) 257–275.
[17] V. Hjellvik, R. Chen, D. Tjøstheim, Nonparametric estimation and testing in panels of intercorrelated time series, J. Time Series Anal. 25 (2004) 831–872.
[18] C. Hsiao, Analysis of Panel Data, Cambridge University Press, 2003.
[19] Q. Jiang, H. Wang, Y. Xia, G. Jiang, On a principal varying coefficient model, J. Amer. Statist. Assoc. 108 (2013) 228–236.
[20] C. Jiang, J. Wang, Functional single-index models for longitudinal data, Ann. Statist. 39 (2011) 362–388.
[21] C. Lam, Q. Yao, Factor modelling for high-dimensional time series: inference for the number of factors, Ann. Statist. 40 (2012) 694–726.
[22] C. Leng, W. Zhang, J. Pan, Semiparametric mean-covariance regression analysis for longitudinal data, J. Amer. Statist. Assoc. 105 (2010) 181–193.

[23] D. Li, J. Chen, J. Gao, Nonparametric time-varying coefficient panel data models with fixed effects, Econometrics J. 14 (2011) 387–408.
[24] Y. Li, Efficient semiparametric regression for longitudinal data with nonparametric covariance estimation, Biometrika 98 (2011) 355–370.
[25] X. Lin, R. Carroll, Nonparametric function estimation for clustered data when the predictor is measured without/with error, J. Amer. Statist. Assoc. 95 (2000) 520–534.
[26] O. Linton, E. Mammen, X. Lin, R. Carroll, Accounting for correlation in marginal longitudinal nonparametric regression, Second Seattle Symposium on Biostatistics, 2003.
[27] S. Liu, G. Li, Varying-coefficient mean-covariance regression analysis for longitudinal data, J. Statist. Plann. Infer. 160 (2015) 89–106.
[28] Z. Liu, T. Stengos, Non-linearities in cross country growth regressions: a semiparametric approach, J. Appl. Econom. 14 (1999) 527–538.
[29] Y.P. Mack, B.W. Silverman, Weak and strong uniform consistency of kernel regression estimates, Z. Wahrscheinlichkeit. verwandte Gebiete 61 (1982) 405–415.
[30] E. Mammen, B. Støve, D. Tjøstheim, Nonparametric additive models for panels of time series, Econom. Theory 25 (2009) 442–481.
[31] B.U. Park, E. Mammen, Y.K. Lee, E.R. Lee, Varying coefficient regression models: A review and new developments, Int. Stat. Rev. 83 (2015) 36–64.
[32] M. Pourahmadi, Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation, Biometrika 86 (1999) 677–690.
[33] Y. Sun, R. Carroll, D. Li, Semiparametric estimation of fixed effects panel data varying coefficient models, Adv. Econom. 25 (2009) 101–129.
[34] A. Ullah, N. Roy, Nonparametric and semiparametric econometrics of panel data, in: A. Ullah, D.E.A. Giles (Eds), Handbook of Applied Economics Statistics, Marcel Dekker, New York, pp. 579–604.
[35] H. Wu, J. Zhang, Nonparametric Regression Methods for Longitudinal Data Analysis: Mixed-Effects Modeling Approaches, Wiley, New York, 2006.
[36] W. Yao, R. Li, New local estimation procedure for nonparametric regression function of longitudinal data, J. R. Stat. Soc. Ser. B (Stat. Methodol.) 75 (2013) 123–138.
[37] F. Yao, H.-G. Müller, J.-L. Wang, Functional data analysis for sparse longitudinal data, J. Amer. Statist. Assoc. 100 (2005) 577–590.
[38] W. Zhang, C. Leng, C. Tang, A joint modelling approach for longitudinal studies, J. R. Stat. Soc. Series B (Stat. Methodol.) 77 (2015) 219–238.