



Efficient Estimation for Varying-Coefficient Mixed Effects Models with Functional Response Data

Xiong Cai¹ · Liugen Xue¹ · Xiaolong Pu² · Xingyu Yan²

Received: 20 January 2020 / Published online: 9 June 2020
© The Author(s) 2020

Abstract

In this article, we focus on the estimation of varying-coefficient mixed effects models for longitudinal and sparse functional response data, by using the generalized least squares method coupling a modified local kernel smoothing technique. This approach provides a useful framework that simultaneously takes into account the within-subject covariance and all observation information in the estimation to improve efficiency. We establish both uniform consistency and pointwise asymptotic normality for the proposed estimators of varying-coefficient functions. Numerical studies are carried out to illustrate the finite sample performance of the proposed procedure. An application to the white matter tract dataset obtained from Alzheimer’s Disease Neuroimaging Initiative study is also provided.

Keywords Functional varying coefficient models · Within-subject correlation · Local kernel smoothing · Efficient estimation · Functional responses

1 Introduction

Longitudinal and functional data are frequently encountered in biomedicine, epidemiology, and other fields of natural and social sciences. See, Hand and Crowder (1996), Ramsay and Silverman (2005) and Ferraty (2011) for abundant interesting examples. In recent years, various functional regression models are proposed to conduct a better description and analysis for this complicated data, since it allows the responses or the regressors, or both, are curves. According to the difference in the concrete forms of the responses and regressors, functional regression models are usually subdivided into

✉ Liugen Xue
lgxue@bjut.edu.cn

¹ College of Applied Sciences, Beijing University of Technology, Beijing 100124, People’s Republic of China

² Key Laboratory of Advanced Theory and Application in Statistics and Data Science-MOE, School of Statistics, East China Normal University, Shanghai 200062, People’s Republic of China

three categories, including scalar-on-function regression (see, e.g., Cai and Hall 2006; Chen, Hall and Müller 2011; Sang et al. 2018), function-on-function regression (Yao et al. 2005a; He et al. 2010; Wang et al. 2018) and function-on-scalar regression (Wu et al. 2013; Zhang and Wang 2015). For more details about the functional regression models and its applications, one can see Morris (2015), Wang et al. (2016) and Reiss et al. (2017) for an overview.

Motivated by the regression analysis of imaging (or functional) data, such as diffusion tensor imaging (DTI) and positron emission tomography (PET) (Zhu et al. 2007; Friston 2009), many nonparametric and semi-parametric regressions for functional responses have been proposed, see, for example, Luo et al. (2016) and Li et al. (2017). A functional varying-coefficient mixed effects model, which can be used for modeling the dynamic impacts of the covariates of interest on the response, is one of the most powerful tools in addressing these scientific problems for imaging data. Let $Y_i(t)$ be the functional response variable for subject i , $i = 1, \dots, n$, and $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ be the associated p -dimensional covariate vector, then the functional varying-coefficient mixed effects model takes the form

$$Y_i(t) = \mathbf{x}_i^T \boldsymbol{\alpha}(t) + \eta_i(t) + \varepsilon_i(t), \quad (1)$$

where $\boldsymbol{\alpha}(t) = (\alpha_1(t), \dots, \alpha_p(t))^T$ is the p -dimensional unknown but smooth coefficient function vector, $\varepsilon_i(t)$ are measurement errors, and $\eta_i(t)$ is the random effect function that characterizes individual curve variation (subject-effect) from $\mathbf{x}_i^T \boldsymbol{\alpha}(t)$. Without loss of generality, throughout this paper we assume that $t \in \mathcal{T} = [0, 1]$. Moreover, $\{\eta_i(t) : t \in \mathcal{T}\}$ is a stochastic process indexed by $t \in \mathcal{T}$ that captures the within-curve dependence. We assume that $\eta_i(t)$ are independent copies of $\eta(t) \sim \text{SP}(0, \Sigma)$, where $\text{SP}(\mu, \Sigma)$ denotes a stochastic process with mean function $\mu(t)$ and covariance function $\Sigma(s, t)$.

Model (1), as a classical function-on-scalar regression has been widely applied for functional data analysis and extensively studied in the existing literature. For instance, for model (1) without involving the random effect item, Wu and Chiang (2000) proposed two kernel estimators based on componentwise local least-squares criteria to estimate the varying-coefficient functions; Chiang et al. (2001) employed a componentwise smoothing spline method to estimate the time-varying coefficients and studied the corresponding asymptotic properties. Zhang and Chen (2007) adopted the idea of ‘‘smoothing first, then estimation’’ to construct the local polynomial kernel reconstruction-based estimator for mean, covariance and error variance functions for model (1) and derived their asymptotics. Zhu et al. (2012) extended model (1) to a multivariate case and systematically studied the theoretical properties of the resulting estimators obtained by local linear smoother. Zhu et al. (2014) further proposed a spatially varying coefficient model (SVCM) for neuroimaging study and investigated the asymptotic properties of the parameter estimators. Other more studies on varying-coefficient models and mixed effects models for longitudinal and functional data, readers may refer to Fan and Zhang (2008) and Chen and Wang (2011) for comprehensive discussions.

However, these aforementioned estimation approaches for model (1) are mainly based on kernel smoothing and smoothing spline, both of which ignore the within-

subject correlation information in the estimation procedures. For longitudinal and functional data, it is well known that there exists a high correlation between the observations on the grid of each datum. As a result, these existing methods might lead to a loss of efficiency. Additionally, the kernel smoothing method utilizes only observations in the neighborhood of the target point for estimation. It is noticeable that these neighborhoods usually contain only a few observations in longitudinal studies, and the within-subject correlation may cause observations out of the neighborhoods carrying significant information. In view of this, the local kernel method might result in inefficient estimators. Therefore, some new techniques improving the efficiency of the conventional local kernel method are highly desirable in longitudinal and functional data analysis. Chen et al. (2010) proposed a global partial likelihood method to study the nonparametric proportional hazards model. The main idea of their global approach is to use all observations instead of only local observations near the target point for estimation to improve the efficiency. Zhou et al. (2018) employed that global smoothing idea combining with the full quasi-likelihood method to conduct the estimation of mean and covariance functions for sparse functional data. Motivated by the idea of global smoothing, we present an efficient estimation procedure for estimating the varying-coefficient functions in model (1) for longitudinal and sparse functional response data. This procedure combines the generalized least squares method and a modified local kernel smoothing criterion which uses the observations on the whole grid of each subject in estimation. An iterative but simple algorithm is proposed to compute the estimator.

The proposed method overcomes the potential drawbacks of existing local kernel methods and improves the efficiency of the resulting estimators. Specifically, there are several advantages for the new method: (i) the proposed estimation procedure involves the global smoothing idea that not only absorbs the advantages of the local linear smoother by using the observations near the target point for estimation, but also utilizes the information available at data outside the neighborhood, which is not used in the local kernel smoothing principle; (ii) the within-subject covariance structure is also incorporated into the estimation procedure to improve the estimators; and (iii) the iterative procedure is easy to implement so that it can be applied to solve various application problems. We derive both uniform consistency and asymptotic normality for the estimated varying coefficient functions. Simulation studies are conducted to compare the performances of our method and the local linear method in terms of root mean squared error, and the results reveal that the proposed method has a smaller root mean squared error than its competitor. This discovery further indicates that the results obtained by our method can more accurately characterize the dynamic relationship between a set of covariates and the functional response variable in application problems. Finally, a real data example is presented to illustrate the newly proposed approach.

The rest of this article is organized as follows. Section 2 introduces the estimation procedure. The theoretical properties including consistency and asymptotic normality of the estimators are established in Sect. 3. Section 4 presents the simulation studies which are performed to evaluate the performance of the method. Section 5 is devoted to applying this method to analyze the white matter tract dataset obtained

from Alzheimer's Disease Neuroimaging Initiative (ADNI). Section 6 concludes the paper and discusses future work. Technical proofs are given in the "Appendix".

2 Methods

In this paper, we focus on sparse data situations. Suppose that we observe the random functions Y_i at discrete points t_{ij} , that is,

$$Y_{ij} = \mathbf{x}_i^T \boldsymbol{\alpha}(t_{ij}) + \eta_i(t_{ij}) + \varepsilon_{ij}, \quad j = 1, \dots, m_i, \quad i = 1, \dots, n, \quad (2)$$

where the observation locations t_{ij} are independent and identically distributed random variables with density function $f(\cdot)$. Throughout this paper, we shall assume that \mathbf{x}_i , t_{ij} , $\eta(\cdot)$ and $\varepsilon(\cdot)$ are all independent and the noise terms, ε_{ij} , are independent and identically distributed random errors with mean zero and variance σ^2 . Suppose that $2 \leq m_i \leq M$, for some finite constant M , for all $i = 1, \dots, n$. Our proposed estimation procedure contains two key steps: (I) calculate the initial estimate $\tilde{\boldsymbol{\alpha}}(t)$ of $\boldsymbol{\alpha}(t)$ and construct the nonparametric covariance estimation for $\Sigma(s, t)$; (II) use the estimated covariance function and the initial estimate $\tilde{\boldsymbol{\alpha}}(t)$ to develop a refined estimate for $\boldsymbol{\alpha}(t)$, denoted by $\hat{\boldsymbol{\alpha}}(t)$.

2.1 Initial estimation

We adopt the local linear scatter-plot smoother (Fan and Gijbels 1996) to estimate the varying coefficient function $\boldsymbol{\alpha}(t)$ and the covariance function $\Sigma(s, t)$. For any T in a small neighborhood of t , applying the Taylor's expansion, we have $\alpha_k(T) \approx \alpha_k(t) + \dot{\alpha}_k(t)(T - t) \equiv a_k + b_k(T - t)$ for $k = 1, \dots, p$, where $\dot{\alpha}_k(t) = d\alpha_k(t)/dt$. Denote $\mathcal{A}_0 = (a_1, \dots, a_p)^T$ and $\mathcal{A} = (a_1, \dots, a_p, b_1, \dots, b_p)^T$. Let $K(\cdot)$ be a kernel function and $K_h(\cdot) = h^{-1}K(\cdot/h)$ be a scaled kernel with bandwidth h . Then the local linear estimate of the regression coefficients, $\boldsymbol{\alpha}(t)$, is simply $\tilde{\boldsymbol{\alpha}}(t) = \tilde{\mathcal{A}}_0$ where

$$\tilde{\mathcal{A}} = \arg \min_{\mathcal{A}} \left\{ \sum_{i=1}^n \sum_{j=1}^{m_i} \left[Y_{ij} - \sum_{k=1}^p [a_k + b_k(t_{ij} - t)] x_{ik} \right]^2 K_{h_\alpha}(t_{ij} - t) \right\}, \quad (3)$$

Let $C_{ijl} = (Y_{ij} - \mathbf{x}_i^T \tilde{\boldsymbol{\alpha}}(t_{ij}))(Y_{il} - \mathbf{x}_i^T \tilde{\boldsymbol{\alpha}}(t_{il}))$ be the "raw" covariances, where $\tilde{\boldsymbol{\alpha}}(\cdot)$ is the estimated varying coefficient function obtained from (3). Note that $E(C_{ijl} | \mathbf{x}_i, t_{ij}, t_{il}) \approx \Sigma(t_{ij}, t_{il}) + \sigma^2 I(t_{ij} = t_{il})$, where $I(\cdot)$ is an indicator function. Therefore the local linear surface smoother for covariance function $\Sigma(s, t)$ can be achieved by the raw covariances with their diagonal elements being removed, and this yields that $\hat{\Sigma}(s, t) = \hat{\boldsymbol{\beta}}_0$, where $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2)^T$ and

$$\widehat{\beta} = \arg \min_{\beta} \left\{ \sum_{i=1}^n \sum_{1 \leq j \neq l \leq m_i} K_{h_C}(t_{ij} - s) K_{h_C}(t_{il} - t) \times [C_{ijl} - \beta_0 - \beta_1(t_{ij} - s) - \beta_2(t_{il} - t)]^2 \right\}. \tag{4}$$

In order to estimate σ^2 , we first estimate the variance $R(t) = \Sigma(t, t) + \sigma^2$ of $Y(t)$. Let $\widehat{R}(t)$ be the local linear smoother of $R(t)$, then $\widehat{R}(t) = \widehat{\beta}_0$, where $\widehat{\beta} = (\widehat{\beta}_0, \widehat{\beta}_1)^T$ and

$$\widehat{\beta} = \arg \min_{\beta} \left\{ \sum_{i=1}^n \sum_{j=1}^{m_i} [C_{ijj} - \beta_0 - \beta_1(t_{ij} - t)]^2 K_{h_V}(t_{ij} - t) \right\}. \tag{5}$$

The estimate for the variance σ^2 is then,

$$\widehat{\sigma}^2 = \frac{1}{|\mathcal{T}|} \int_{\mathcal{T}} \{\widehat{R}(t) - \widehat{\Sigma}(t, t)\} dt, \tag{6}$$

where the notation $|\mathcal{T}|$ denotes the length of the interval \mathcal{T} . The more detailed process for the estimation of mean and covariance functions for longitudinal and sparse functional data, one can refer to Yao et al. (2005b) and Li and Hsing (2010).

2.2 Refined estimation for varying-coefficient functions

We know that the local kernel smoothing uses only local observations near the target point for estimation. However, the within-subject correlation, presented in the longitudinal and functional data, may result in observations outside the neighborhood of the target point carrying relevant information. Therefore, the local linear method used in Sect. 2.1 for estimating the varying-coefficient functions may reduce to loss of efficiency. In this subsection, we employ the generalized least squares method incorporating a modification of local linear smoothing to conduct a refined estimator for $\alpha(\cdot)$. This new method not only adopts the idea that uses all observations instead of the local observations near the target point for estimation but also incorporates the estimated covariance structure which can be used to account for the within-subject correlation into the estimation process to improve the resulting estimators.

Write $Y_i = (Y_{i1}, \dots, Y_{i,m_i})^T, t_i = (t_{i1}, \dots, t_{i,m_i})^T, \alpha(t_i) = (\alpha_1(t_i), \dots, \alpha_p(t_i))^T$ and $\alpha_j(t_i) = (\alpha_j(t_{i1}), \dots, \alpha_j(t_{i,m_i}))^T$ for $j = 1, \dots, p$. Note that $E(Y_i | x_i, t_i) = \alpha^T(t_i) x_i$ and $\text{cov}(Y_i | x_i, t_i) = \Sigma_i + \sigma^2 I_{m_i}$, where $\Sigma_i = \{\Sigma(t_{ij}, t_{il})\}_{j,l=1}^{m_i}$ is the covariance matrix of $\eta_i(t_i) = (\eta_i(t_{i1}), \dots, \eta_i(t_{i,m_i}))^T$ and I_{m_i} is the identity matrix of order m_i for $i = 1, \dots, n$. For convenience, let $V_i = \Sigma_i + \sigma^2 I_{m_i}$. Then the generalized least squares estimator of $\alpha(t)$ is obtained by minimizing the follow objective function

$$Q(\alpha) = \sum_{i=1}^n \{Y_i - \alpha^T(t_i) x_i\}^T V_i^{-1} \{Y_i - \alpha^T(t_i) x_i\}. \tag{7}$$

Firstly, we replace V_i with its estimates. Since we do not assume any structure on the covariance function of $Y(t)$ other than that it is a smooth function, the local linear estimates presented in Sect. 2.1 is recommended. Let $\widehat{\Sigma}_i$ be the estimated covariance matrix of Σ_i and $\widehat{\sigma}^2$ be the estimated variance of measurement errors obtained from (4), (5) and (6). Then V_i is substituted by $\widehat{V}_i = \widehat{\Sigma}_i + \widehat{\sigma}^2 I_{m_i}$.

Secondly, we adopt a similar strategy in Chen et al. (2010) and Zhou et al. (2018) to approximate $\alpha(t)$. Specifically, let $B_n(t)$ be a neighborhood of t , it is easy to see that

$$\alpha(t_{ij}) = \alpha(t_{ij})I(t_{ij} \in B_n(t)) + \alpha(t_{ij})\{1 - I(t_{ij} \in B_n(t))\}. \tag{8}$$

Based on the idea of local linear smoother, we use linear function $\zeta + \dot{\zeta}(t_{ij} - t)$ to approximate $\alpha(t_{ij})$ in the first term of (8) and replace $I(t_{ij} \in B_n(t))$ with a kernel function $K\{(t_{ij} - t)/h\}$. Then for t_{ij} in a neighborhood of t , we have

$$\alpha(t_{ij}) \approx \{\zeta + \dot{\zeta}(t_{ij} - t)\}hK_h(t_{ij} - t) + \alpha(t_{ij})\{1 - hK_h(t_{ij} - t)\},$$

where $\zeta = (\zeta_1, \dots, \zeta_p)^T$, $\dot{\zeta} = (\dot{\zeta}_1, \dots, \dot{\zeta}_p)^T$ and $K_h(\cdot) = h^{-1}K(\cdot/h)$ is a scaled kernel function with bandwidth h . This expansion is of great use in longitudinal and functional data analysis by the reason of it not only absorbs the advantage of local linear smoother but also utilizes the information available at data outside the neighborhood region $B_n(t)$. Substituting this approximation and the estimated covariance matrix into (7), we can obtain the refined estimator for $\vartheta = (\zeta^T, \dot{\zeta}^T)^T$ by solving the following equation

$$\frac{\partial Q(\vartheta; \alpha)}{\partial \vartheta} = \sum_{i=1}^n \Theta_i^T K_{i,h} \widehat{V}_i^{-1} \left\{ Y_i - hK_{i,h} \Theta_i \vartheta - (I_{m_i} - hK_{i,h}) \alpha^T(t_i) \mathbf{x}_i \right\} = 0, \tag{9}$$

where $K_{i,h} = \text{diag}(K_h(t_{i1} - t), \dots, K_h(t_{i,m_i} - t))$ and

$$\Theta_i = \begin{pmatrix} x_{i1} \cdots x_{ip} & x_{i1}(t_{i1} - t) & \cdots & x_{ip}(t_{i1} - t) \\ \vdots & \ddots & \ddots & \vdots \\ x_{i1} \cdots x_{ip} & x_{i1}(t_{i,m_i} - t) & \cdots & x_{ip}(t_{i,m_i} - t) \end{pmatrix}.$$

Note that the refined estimator can not be directly obtained by solving the the Eq. (9) as there still exists an unknown function vector $\alpha(t_i)$. Therefore, we present an iterative procedure by substituting $\alpha(t_i)$ with its previous estimator to address this problem. Let $t_1 < t_2 < \dots < t_M$ be the distinct points of t_{ij} for $i = 1, \dots, n$ and $j = 1, \dots, m_i$. The iterative algorithm is described as follows.

- *Step 0* Taking the estimator fulfilled by (3) as the initial values of $\alpha(t)$, denoted by $\alpha_{[0]}(t) = \tilde{\alpha}(t)$ for $t = t_1, \dots, t_M$.
- *Step 1* For each $t = t_1, \dots, t_M$, solve the Eq. (9) in which $\alpha(t)$ is replaced by its previous estimator, that is,

$$\sum_{i=1}^n \Theta_i^T K_{i,h} \widehat{V}_i^{-1} \{ Y_i - hK_{i,h} \Theta_i \vartheta - (I_{m_i} - hK_{i,h}) \alpha_{[s-1]}^T(t_i) \mathbf{x}_i \} = 0, \tag{10}$$

where $\alpha_{[s-1]}^T(t_i)$ is the $(s - 1)$ th iterative estimator of $\alpha^T(t_i)$. (10) yields the estimator $\hat{\vartheta}(t)$, given by

$$\hat{\vartheta}(t) = \begin{pmatrix} \hat{\zeta}(t) \\ \hat{\xi}(t) \end{pmatrix} = \left\{ \sum_{i=1}^n \Theta_i^T \mathbf{K}_{i,h} \hat{\mathbf{V}}_i^{-1} h \mathbf{K}_{i,h} \Theta_i \right\}^{-1} \times \sum_{i=1}^n \Theta_i^T \mathbf{K}_{i,h} \hat{\mathbf{V}}_i^{-1} \{Y_i - (I_{m_i} - h \mathbf{K}_{i,h}) \alpha_{[s-1]}^T(t_i) \mathbf{x}_i\}. \quad (11)$$

Hence, $\alpha_{[s]}(t) = \hat{\zeta}(t)$.

- Step 2 Repeat Step 1 until convergence.

The selection of the bandwidth h for the estimation procedure presented above is achieved by the leave-one-curve-out cross-validation (Rice and Silverman 1991). Specifically, let $\hat{\alpha}^{(-i)}(\cdot)$ be the estimated varying-coefficient function of $\alpha(\cdot)$ after removing the i th subject from the data. Then the optimal bandwidth h can be chosen by minimizing the cross-validation score based on the squared prediction error given by

$$PE(h) = \frac{1}{n} \sum_{i=1}^n \frac{1}{m_i} \sum_{j=1}^{m_i} \{Y_i(t_{ij}) - \mathbf{x}_i^T \hat{\alpha}(t_{ij})^{(-i)}\}^2.$$

Remark 1 The proposed iterative estimation procedure can be readily implemented as there is a closed-form at each step. Compared to the local linear method used in Zhu et al. (2012), the newly presented estimation approach may be more efficient since it takes the within-subject correlation structure and the observations on the whole grid of each subject into consideration, two of which are significantly important in longitudinal and functional data analysis. It is not hard to see that the estimation in (3) is, in essence, a local marginal estimation, assuming that the within-subject covariance matrices are identities. This leads to a working independence estimator for $\alpha(t)$ which ignores the correlation among the observations on the grid of each subject, and further causes a loss of efficiency.

Remark 2 Replacing $\alpha(t)$ with its estimate to model (1), we also can conduct the estimation of each random effect function $\eta_i(t)$. Many existing non-parametric methods such as the local linear regression have been developed to directly estimate all individual functions $\eta_i(t)$, $i = 1, \dots, n$, (see, e.g., Zhu et al. 2012; Li et al. 2017). However, these non-parametric methods may be not suitable for our sparse data situations when the number of observation points for each subject is small. Functional principle component analysis (FPCA) as a key technique in characterizing the features of unknown functions in functional data analysis, is another feasible method for the estimation of $\eta_i(t)$. We recommend using the version of FPCA developed by Yao et al. (2005b) which they refer to as principal components analysis through conditional expectation (PACE) for longitudinal data to estimate each random effect function $\eta_i(t)$. The PACE approach is appealing since it is implemented mainly based on the estimated population covariance function of $\eta_i(t)$, which utilizes the whole sample information and

can be easily obtained by using the method presented in Sect. 2.1. Comprehensive surveys on FPCA for longitudinal or sparse functional data were developed in Yao et al. (2005b) and Hall et al. (2006).

3 Asymptotic properties

In this section, we study the asymptotic properties of our estimators. Some regularity conditions needed for the theoretical developments are stated in “Appendix A.1”. All these conditions are very mild and most of these are also adopted by Chen et al. (2010) and Zhou et al. (2018). The detailed proofs for the following theorems are given in “Appendix A.2–A.3”.

To start with, we give some notations which will be used in the theorems. Let $v_r = \int u^r K(u)du$ and $v_r = \int u^r K^2(u)du$ for notational simplicity. For a vector of functions $\alpha(t)$ of t , denote $\dot{\alpha}(t) = d\alpha(t)/dt$ and $\ddot{\alpha}(t) = d^2\alpha(t)/dt^2$, which are the componentwise derivatives. Let $\mathcal{V}_{i,jk}$ be the (j, k) th element of $V_i^{-1} = (\Sigma_i + \sigma^2 I_{m_i})^{-1}$. Denote $\gamma_1(s) = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \sum_{j=1}^{m_i} E[\mathcal{V}_{i,jk} | t_{ij} = s]$, $\gamma_2(s_1, s_2) = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \sum_{j \neq k}^{m_i} E[\mathcal{V}_{i,jk} | t_{ij} = s_1, t_{ik} = s_2]$ and $\Omega_X = E[\mathbf{xx}^T]$. Let \mathcal{I} be the identity operator, and \mathcal{A} be the linear operator satisfying $\mathcal{A}(H)(t) = \int_z \gamma_2(t, z) / \gamma_1(t) H(z) f(z) dz$ for any function H .

Theorem 1 *Under the regularity Conditions (C1)–(C8) in “Appendix A.1”, when $n \rightarrow \infty$, we have*

$$\sup_{t \in [0,1]} |\widehat{\alpha}_j(t) - \alpha_j(t)| \rightarrow 0, \quad j = 1, \dots, p,$$

in probability.

Theorem 2 *Suppose that conditions of Theorem 1 hold. Then, for any $t \in (0, 1)$, we have*

$$(\mathcal{I} + \mathcal{A})(\widehat{\alpha} - \alpha)(t) = (nh)^{-1/2} \mathbf{\Gamma}(t)^{1/2} \boldsymbol{\varphi} + \frac{1}{2} v_2 h^2 \ddot{\alpha}(t) + o_p\{h^2 + (nh)^{-1/2}\},$$

where $\mathbf{\Gamma}(t) = v_0[\Omega_X f(t) \gamma_1(t)]^{-1}$ and $\boldsymbol{\varphi}$ is a random vector following the standard normal distribution.

Theorem 2 implies the following Corollary 3.

Corollary 3 *Under the Conditions (C1)–(C8) in the “Appendix A.1”, we have*

$$\sqrt{nh} \left\{ (\widehat{\alpha} - \alpha)(t) - \frac{1}{2} (\mathcal{I} + \mathcal{A})^{-1} (\ddot{\alpha})(t) v_2 h^2 \right\} \xrightarrow{\mathcal{L}} N(0, \mathbf{\Xi}(t))$$

for any $t \in (0, 1)$, where $\mathbf{\Xi}(t) = \{(\mathcal{I} + \mathcal{A})^{-1} (\mathbf{\Gamma}^{1/2})(t)\} \{(\mathcal{I} + \mathcal{A})^{-1} (\mathbf{\Gamma}^{1/2})(t)\}^T$.

Remark 3 Theorem 1 shows the proposed estimator $\widehat{\alpha}_j(\cdot), j = 1, \dots, p$, is uniformly convergent. Theorem 2 shows the asymptotic bias of $(\mathcal{I} + \mathcal{A})(\widehat{\alpha} - \alpha)(t)$ is of order $O(h^2)$, and the asymptotic variance is of order $O((nh)^{-1})$. As a consequence, the theoretical optimal bandwidth is of order $n^{-1/5}$. These results are similar to that of Zhou et al. (2018). The asymptotic properties presented above can be viewed as a generalization of the results in Zhou et al. (2018) from mean function estimation to varying-coefficient function estimation. These results will be more widely used in solving various practical problems than that of Zhou et al. (2018). Especially, it can provide theoretical guidance for analyzing the dynamic relationship between a set of covariates and the functional response variable, rather than only exploring the feature of functional variables.

4 Simulation studies

In this section, we are going to perform some simulation studies to illustrate the finite sample properties of the proposed estimation method. In consideration of the estimation procedure we proposed is aimed at estimating the varying-coefficient functions, we will use the accuracy of the varying-coefficient function estimators to demonstrate how well the procedure works. The performances are evaluated via the bias, standard deviation (SD), and root mean squared error (RMSE), at given grid points, which also are used by Zhou et al. (2018). Specifically, the bias, SD and RMSE of an estimator $\widehat{\alpha}(\cdot)$ are respectively defined as

$$\text{Bias} = \left[\frac{1}{n_g} \sum_{j=1}^{n_g} \{E\widehat{\alpha}(t_j) - \alpha(t_j)\}^2 \right]^{1/2}, \quad \text{SD} = \left[\frac{1}{n_g} \sum_{j=1}^{n_g} E\{\widehat{\alpha}(t_j) - E\widehat{\alpha}(t_j)\}^2 \right]^{1/2}$$

and $\text{RMSE} = (\text{bias}^2 + \text{SD}^2)^{1/2}$, where $\widehat{\alpha}(\cdot)$ is an estimator of $\alpha(\cdot)$, $E\widehat{\alpha}(t_j)$ is estimated by the sample mean based on the N simulated data sets, and $t_j(j = 1, \dots, n_g)$ are the grid points at which the function $\widehat{\alpha}(\cdot)$ is evaluated. In addition, comparisons are also made between our method and the local linear method employed in Zhu et al. (2012) (Local hereinafter). In the following simulated examples, the Epanechnikov kernel is used for nonparametric regression. We set grid points $n_g = 100$ and compute the results based on $N = 500$ replications.

Example 4 Generate data from the following model

$$Y_{ij} = \mathbf{x}_i^T \boldsymbol{\alpha}(t_{ij}) + \eta_i(t_{ij}) + \varepsilon_{ij},$$

where $t_{ij} \sim U[0, 1], \varepsilon_{ij} \sim N(0, \sigma^2)$ and $\mathbf{x}_i = (1, x_{i1}, x_{i2})^T$ for $i = 1, \dots, n, j = 1, \dots, m_i$. Moreover, $(x_{i1}, x_{i2})^T \sim N((0, 0)^T, \text{diag}(0.9, 0.9) + 0.1(1, 1)^{\otimes 2})$, and $\eta_i(t_{ij}) = \xi_{i1}\psi_1(t_{ij}) + \xi_{i2}\psi_2(t_{ij})$, where $\xi_{ij} \sim N(0, \lambda_j^2)$ for $j = 1, 2$. Furthermore, $(x_{i1}, x_{i2}), t_{ij}, \varepsilon_{ij}, \xi_{i1}$ and ξ_{i2} are independent random variables. We set $(\sigma, \lambda_1, \lambda_2) = (0.3, 1, 0.5)$, and the varying-coefficient functions, and eigenfunctions as follows:

$$\alpha_1(t) = t^2, \quad \alpha_2(t) = -3(t - 0.5)^2 + 1, \quad \alpha_3 = 4t(t - 1),$$

$$\psi_1(t) = \sqrt{2} \sin(2\pi t), \quad \psi_2(t) = \sqrt{2} \cos(2\pi t).$$

We first conducted simulation studies under the following two regular time points settings.

- Setting 1: $m_i = 5$ with various sample size: $n = 200$ and $n = 400$.
- Setting 2: $n = 150$ with various time points: $m_i = 7$ and $m_i = 13$.

Setting 1, two different sample size $n = 200, 400$, and a fixed time points $m_i = 5$ are designed. Setting 2, sample size $n = 150$ is fixed with various time points settings $m_i = 7$ and $m_i = 13$, which guarantees the total number of observations is comparable to that in Setting 1, by decreasing the number of subjects and increasing the observation number from the same individual. Tables 1 and 2 report the numerical results for the proposed method and Local method under Setting 1 and Setting 2, respectively.

Table 1 The simulation results for Setting 1 in Example 4. Bias, SD and RMSE of $\hat{\alpha}_1(\cdot)$, $\hat{\alpha}_2(\cdot)$ and $\hat{\alpha}_3(\cdot)$ using the proposed method and the Local method

	$m_i = 5, n = 200$			$m_i = 5, n = 400$		
	$\hat{\alpha}_1(\cdot)$	$\hat{\alpha}_2(\cdot)$	$\hat{\alpha}_3(\cdot)$	$\hat{\alpha}_1(\cdot)$	$\hat{\alpha}_2(\cdot)$	$\hat{\alpha}_3(\cdot)$
<i>Prop.</i>						
Bias	0.0120	0.0375	0.0569	0.0109	0.0279	0.0382
SD	0.0798	0.0856	0.0823	0.0613	0.0623	0.0611
RMSE	0.0807	0.0935	0.1001	0.0623	0.0683	0.0721
<i>Local</i>						
Bias	0.0135	0.0394	0.0575	0.0112	0.0296	0.0396
SD	0.0840	0.0899	0.0881	0.0636	0.0646	0.0635
RMSE	0.0851	0.0982	0.1052	0.0646	0.0711	0.0748

Table 2 The simulation results for Setting 2 in Example 4. Bias, SD and RMSE of $\hat{\alpha}_1(\cdot)$, $\hat{\alpha}_2(\cdot)$ and $\hat{\alpha}_3(\cdot)$ using the proposed method and the Local method

	$m_i = 7, n = 150$			$m_i = 13, n = 150$		
	$\hat{\alpha}_1(\cdot)$	$\hat{\alpha}_2(\cdot)$	$\hat{\alpha}_3(\cdot)$	$\hat{\alpha}_1(\cdot)$	$\hat{\alpha}_2(\cdot)$	$\hat{\alpha}_3(\cdot)$
<i>Prop.</i>						
Bias	0.0145	0.0424	0.0557	0.0155	0.0432	0.0554
SD	0.0858	0.0878	0.0877	0.0782	0.0771	0.0792
RMSE	0.0870	0.0975	0.1039	0.0797	0.0884	0.0967
<i>Local</i>						
Bias	0.0152	0.0440	0.0561	0.0168	0.0442	0.0572
SD	0.0911	0.0930	0.0931	0.0833	0.0820	0.0845
RMSE	0.0924	0.1029	0.1087	0.0850	0.0932	0.1020

Example 5 To illustrate that the proposed procedure is applicable even to the case that the observation times for each trajectory are randomly missed, we replicated Example 4 with an irregular time points design used in Fan, Huang and Li (2007). Specifically, we will generate the observation times randomly, assuming that each individual has a set of “scheduled” observation points, $\{0, 1, 2, \dots, 11\}$, and each scheduled time, except time 0, has a 20% probability of being skipped, then a uniform $[0, 1]$ random variable is added to a nonskipped scheduled time so that the actual observation time is in fact a random perturbation of a scheduled time. Table 3 presents the numerical results for this case with sample size $n = 200$ and 400.

Tables 1, 2 and 3 summarize the numerical performance of the proposed method and the Local method through the Bias, SD and RMSE. It is shown from Tables 1, 2 and 3 that the RMSE of the proposed method is consistently smaller than that of the Local method for all varying-coefficient function estimators $\hat{\alpha}_1(t)$, $\hat{\alpha}_2(t)$ and $\hat{\alpha}_3(t)$, which indicates that the proposed method performs better than the Local method. Moreover, the proposed method has less bias and variance than the Local method, mainly because of the incorporation of the correlation information and the modified local linear technique. Tables 1, 2 and 3 also show that the Bias, SD and RMSE values decrease as the sample size increases, which corroborates with our consistency and asymptotic normality established in Sect. 3. From Tables 1 and 2, we find that the resulting estimators obtained in Example 4 with Setting 1 have smaller RMSE than those obtained in Example 4 with Setting 2 when the total number of observations for these two settings is close. This finding implies that the sparsity is beneficial to the performance of the proposed method and the estimators of the varying-coefficient functions can be improved with increasing the number of subjects. For all simulation cases, we chose $h_0 = 0.12$ as the bandwidth for the initial estimates of the varying-coefficient functions and the estimates of the covariance functions. We discover that the proposed method is not sensitive to the initial values. Furthermore, the optimal bandwidth for the proposed method is slightly larger than that for the Local method. This is likely because the correlation matrix is incorporated into the estimation process

Table 3 The simulation results for Example 5. Bias, SD and RMSE of $\hat{\alpha}_1(\cdot)$, $\hat{\alpha}_2(\cdot)$ and $\hat{\alpha}_3(\cdot)$ using the proposed method and the Local method

	$n = 200$			$n = 400$		
	$\hat{\alpha}_1(\cdot)$	$\hat{\alpha}_2(\cdot)$	$\hat{\alpha}_3(\cdot)$	$\hat{\alpha}_1(\cdot)$	$\hat{\alpha}_2(\cdot)$	$\hat{\alpha}_3(\cdot)$
<i>Prop.</i>						
Bias	0.0112	0.0337	0.0447	0.0096	0.0243	0.0325
SD	0.0724	0.0723	0.0773	0.0562	0.0597	0.0568
RMSE	0.0733	0.0798	0.0893	0.0570	0.0645	0.0654
<i>Local</i>						
Bias	0.0125	0.0371	0.0469	0.0100	0.0268	0.0365
SD	0.0752	0.0745	0.0802	0.0574	0.0606	0.0576
RMSE	0.0762	0.0832	0.0929	0.0583	0.0663	0.0682

Table 4 The mean and SD of RASE for estimators $\widehat{\eta}_i(\cdot)$ using the PACE method and the local linear method with data generated from Example 5

n	PACE		Local linear	
	Mean	SD	Mean	SD
100	0.3009	0.0393	0.6692	0.0599
200	0.2461	0.0278	0.6599	0.0486
400	0.2073	0.0165	0.6521	0.0322

of the proposed method, which can be treated as the weight as well as the kernel matrix, and further lower data usage. Therefore, the larger bandwidth for the proposed method is needed to enlarge the included range of data.

We also apply the PACE method to Example 5 to conduct the estimation of $\eta_i(t)$. A comparison is made between the PACE method and the local linear approach. The PACE method is implemented with the FPCA function available in the package `fcpa` (Dai et al. 2019). We choose the first four functional principal components to construct the estimators of $\eta_i(t)$ ($i = 1, \dots, n$) for the PACE method. The bandwidth for the local linear method is selected by generalized cross-validation (GCV). We evaluate the performances through the root average square errors (RASE) given by

$$\text{RASE} = \left[\frac{1}{n} \sum_{i=1}^n \frac{1}{n_g^*} \sum_{j=1}^{n_g^*} \{ \widehat{\eta}_i(t_{ij}^*) - \eta_i(t_{ij}^*) \}^2 \right]^{1/2},$$

where t_{ij}^* ($j = 1, \dots, n_g^*$) are the grid points at which the function $\widehat{\eta}_i(\cdot)$ is evaluated. Since the observation points for each subject are sparse, we set $n_g^* = 50$ to calculate the RASE. Table 4 presents the mean and standard deviation (SD) of RASE, both of which are computed based on 200 simulation replications. We can see from Table 4 that all the mean and SD of RASE for the PACE method are significantly smaller than that of the local linear method. These results imply that the PACE method is superior to the local linear method for the estimation of $\eta_i(t)$. Moreover, we observe that the mean value of RASE for the PACE method decreases as the sample size increases, while the value for the local linear method has hardly changed. This is likely due to the fact that the PACE method for constructing the estimators $\widehat{\eta}_i(t)$ ($i = 1, \dots, n$) is mainly based on the estimated population covariance function of $\eta_i(t)$, which applies the whole sample information in estimation. Nevertheless, the local linear method for directly estimating each $\eta_i(t)$ only utilizes the observations on the grid of its corresponding curve, therefore, the resulting estimators could not be improved by increasing the size of samples.

To have a more visual display about how well the proposed estimation procedure works, we choose Example 5 with sample size $n = 200$ as an example. Figure 1 exhibits the averaged estimated varying-coefficient functions based on 50 replications and the boxplot of RASE of the varying-coefficient functions over 500 simulations. It can be seen from Fig. 1a–c that the estimated varying-coefficient curves (broken lines), obtained by the proposed method, closely resemble the corresponding true curves (solid lines), reflecting the good performance of the proposed procedure. Figure 1d shows the proposed method performs better than the Local method in terms of RASE.

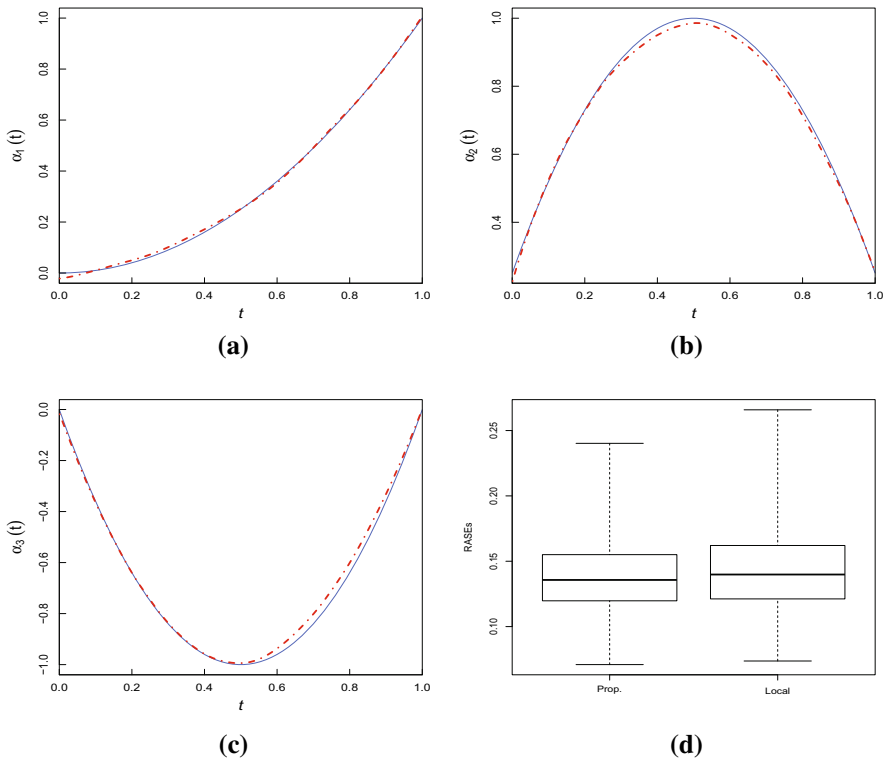


Fig. 1 The averaged estimated varying-coefficient functions in **a)–(c)** over 50 simulations and the boxplot for the three varying-coefficient functions in **d** over 500 simulations under Example 5 with sample size $n = 200$. The solid lines are the true functions and the broken lines are the estimated one in **(a)–(c)**

5 Real data analysis

We apply model (2) and the newly proposed procedure to analyze a real diffusion tensor imaging (DTI) dataset with $n = 213$ subjects that was collected from NIH Alzheimer’s Disease Neuroimaging Initiative (ADNI) study. One goal of the NIH ADNI is to test whether genetic, structural and functional neuroimaging, and clinical data can be integrated to evaluate the progression of mild cognitive impairment (MCI) and early Alzheimer’s disease (AD). For more details on this dataset, see the ADNI publicly available database (<http://adni.loni.usc.edu/>). In diffusion data analysis, fractional anisotropy (FA), which quantifies the directional strength of white matter tract structure, at a particular location, is one of the most used measures, and has been widely applied to statistical analyses in many imaging studies. During the stage of data processing, we calculate the FA curve for each subject using a Tract-Based Spatial Statistics (TBSS) pipeline analysis (Smith et al. 2006) and the FMRIB software library tool. The more details for this process, one can refer to Li et al. (2017).

In this study, 213 subjects with age of 48.4–73.2 years old are sampled to assess the impact of some indicators which are commonly occurred in diffusion data, such

as gender (123 male and 90 female), handedness (193 righthand and 20 left-hand), education level (years, ranges from 9 to 20 (years), mean 15.91), Alzheimer's disease (AD) status (19.6%), mild cognitive impairment (MCI) status (55.1%) and Mini-Mental State Exam (MMSE) of ADNI on the FA. We aim to evaluate the effects of these indicators on the FA using model (2). Specifically, the response $Y_i(t)$ represents the FA curve and the covariate vector \mathbf{x}_i includes an intercept term, the gender variable coded by a dummy variable indicating for male, the age of the subject, a handedness indicator coded by a dummy variable indicating for left-hand, the education level, an indicator for AD status, an indicator for MCI status and MMSE. We randomly selected 15 observations from the 83 grid points of each FA curve, so that we can compare the results obtained from the sparse data and the complete data and further validate the new approach. We standardize all variables to be mean zero and variance one, and apply the proposed procedure established in Sect. 2 to estimate the corresponding coefficients.

Figure 2 shows the estimated varying-coefficient functions corresponding to intercept, gender, age, education, AD status, MCI status and MMSE indicator and the associated 95% pointwise confidence intervals for both sparse and complete data. The pointwise confidence intervals were obtained with the standard error based on 200 bootstrap resampling. The dashed lines are the estimates fulfilled by the proposed method for sparse data and the solid lines are estimated by the local linear method for complete data. Figure 2 indicates that the proposed estimation method works well even though data are sparse; it also reveals that age, handedness, education, AD and MCI have more effects on FA than that of gender and MMSE. The intercept function characterizes the nonlinear overall trend of FA. It can be seen from Fig. 2 that the coefficient curve for gender is positive and that for MMSE is near to the zero line, while the coefficients for the other five covariates of interest are negative, at most of the grid points. These findings may indicate that males tend to have higher FA values than females, MMSE has no remarkable effects on FA, AD and MCI patients tend to have lower FA values and being older, left-hand, or more educated may lead to smaller FA values. These discoveries are also consistent with the previous analysis in Li et al. (2017). Finally, we calculated the squared prediction error (i.e., $SPE = \sum_{i=1}^n \sum_{j=1}^{m_i} \{Y_i(t_{ij}) - \hat{Y}_i^*(t_{ij})\}^2$, where \hat{Y}_i^* is the predicted trajectory for the i th subject by using the data with the i th subject being pulled out.) both for the proposed method and the local linear method. The corresponding SPE values for these two approaches are respectively 18.34 and 18.42, which implies that the proposed method has a smaller prediction error than that of the local linear method for analyzing this data.

Figure 3 presents the estimated eigenvalues and eigenfunctions for the covariance function $\Sigma(s, t)$ with sparse data. The top four nonzero eigenvalues are 0.0028, 0.00047, 0.0003 and 0.00016 respectively. The first four eigenvalues explain about 94% of the total variability, while, it can be seen from Fig. 3 that the remaining eigenvalues rapidly drop to zero. The first eigenfunction, with a dominant eigenvalue accounting for 70% of the total variation, varies simply in structure. While the other eigenfunctions also have quite simple structures and most of them are roughly sinusoidal.

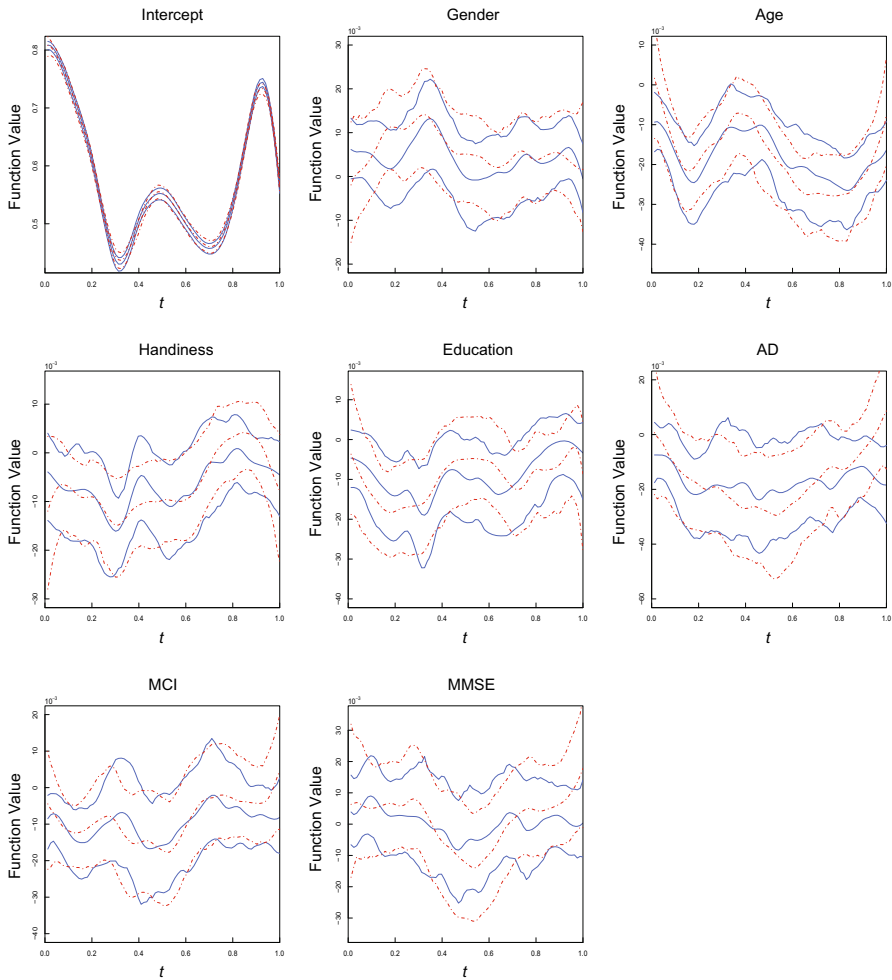


Fig. 2 The estimated varying coefficients for intercept, sex, age, handedness, education, AD, MCI and MMSE and their 95% pointwise confidence intervals for the ADNI data. The dashed lines are the estimates fulfilled by our proposed estimation procedure with sparse data and the solid lines are estimated by the local linear method with complete data

6 Conclusions

In this paper, we used a functional varying-coefficient mixed effects model to investigate the dynamic association between a set of covariates of interest and the functional response. An iterative estimation procedure was developed to estimate the varying-coefficient functions in this model for longitudinal or sparse functional data. The proposed estimation procedure not only employs the idea that uses all observations instead of the local observations near the target point for estimation but also incorporates the estimated covariance structure into the estimation process to improve the resulting estimators. The asymptotic properties including uniform consistency and

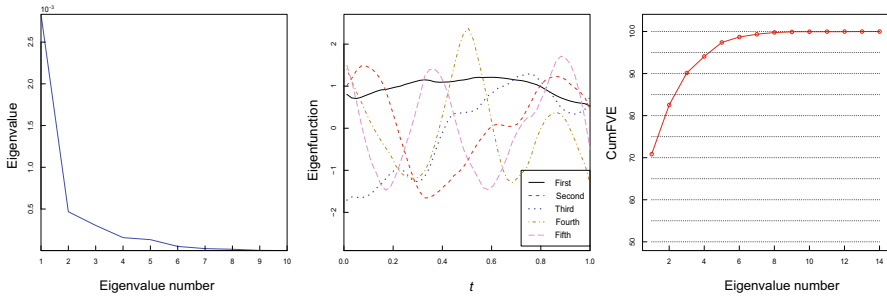


Fig. 3 The estimated eigenvalues in decreasing order (left), the estimated eigenfunctions (middle) and the corresponding cumulative FVE (right) for ADNI data analysis

asymptotic normality for the proposed estimators were also established. Finally, we compared the proposed procedure with the local linear method numerically through several simulation studies. The simulation results reveal that our method performed better than the local linear method in terms of the root mean squared error. A real data example is also provided to illustrate our approach.

Our future research will be focused on addressing several important issues. First, it would be interesting to gain a more efficient estimator for the covariance function by using the newly proposed estimation procedure. Second, applying the proposed estimation procedure to investigate other more complex regression models, such as functional index models (Jiang and Wang 2011), single-index varying coefficient (SIVC) models (Luo et al. 2016) and functional additive models (Zhang, Park and Wang 2013), is also a meaningful issue. Third, it is also worthwhile to conduct statistical inferences such as hypothesis tests for longitudinal and functional data.

Acknowledgements This work was supported by the National Natural Science Foundation of China (11971001), the Beijing Natural Science Foundation (1182002), the grants from the National Natural Science Foundation of China (11771145, 11801210). Data used in this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (<http://adni.loni.usc.edu>).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix

A.1 Regularity Conditions and Notations

Our proofs use a strategy that is similar to that in Zhou et al. (2018). The following regularity conditions, which will be used to establish the asymptotic properties of the

proposed estimators, are imposed to facilitate the proof and are similar to Zhou et al. (2018). They are not the weakest possible conditions, but mainly for mathematical simplicity and may be modified if necessary.

- (C1) The kernel $K(\cdot)$ is bounded symmetric probability density function with the compact support $[-1, 1]$ and satisfy $v_r = \int u^r K(u)du < \infty$, $v_r = \int u^r K^2(u)du < \infty$.
- (C2) $\max_{1 < i < n} m_i \leq M$, where M is a constant, and the dimension of covariate vector is bounded.
- (C3) $t_{ij}, i = 1, \dots, n, j = 1, \dots, m_i$ are independent and identically distributed copies of a random variable T defined on $[0,1]$. Its density function $f(\cdot)$ is positive and has continuous second derivatives.
- (C4) The covariate vectors \mathbf{x}_i 's are independent and identically distributed with $E(\mathbf{x}_i) = \mu_x$ and $\|\mathbf{x}_i\|_\infty < \infty$. Assume that $E[\mathbf{x}_i \mathbf{x}_i^T] = \Omega_X$ is positive definite.
- (C5) $\{\alpha_l(t), l = 1, \dots, p\}$ have continuous third derivatives on their compact support $[0, 1]$.
- (C6) There exist estimators $\tilde{\alpha}(t), \widehat{\Sigma}(s, t)$ and $\hat{\sigma}$ such that $\sup_{t \in [0,1]} |\tilde{\alpha}_i(t) - \alpha_i(t)| = o_p(1)$ for $i = 1, \dots, p$, $\sup_{s,t \in [0,1]} |\widehat{\Sigma}(s, t) - \Sigma(s, t)| = o_p(1)$ and $|\hat{\sigma}^2 - \sigma^2| = o_p(1)$.
- (C7) $nh^5 \rightarrow c$ and $nh/(\log n)^2 \rightarrow \infty$ as $n \rightarrow \infty$, where c is a constant.
- (C8) If $\mathbf{h}(\cdot) \in \mathcal{F}_p$ satisfies $E[\{\gamma_1(t) + \int_0^1 \gamma_2(t, z) f(z) dz\} \Omega_X \mathbf{h}(t)] = 0$, then $\mathbf{h}(t) \equiv 0$, where $\mathcal{F}_p, \gamma_1(t)$ and $\gamma_2(t, z)$ are defined in the following.

Conditions (C1)-(C5) are mild assumptions for the kernel function, covariates and the function of interest, which are commonly used in many longitudinal literature. Condition (C6) guarantees the convergence of the initial estimators which can be easily achieved by the exiting kernel and spline estimation method. Condition (C7) is the regular condition for the bandwidth, which also has been used in Chen et al. (2010) and Zhou et al. (2018). Condition (C8) ensures identifiable.

For notational simplicity, denote that $\mathcal{F}_p = \{\zeta(\cdot) : \zeta(\cdot) \text{ is continuous on } [0, 1]^p\}$ and $\mathcal{F}_1 = \{\zeta(\cdot) : \zeta(\cdot) \text{ is continuous on } [0, 1]\}$. Let $\mathcal{V}_{i,jk}$ be the (j, k) th element of $V_i^{-1} = (\Sigma_i + \sigma^2 I_{m_i})^{-1}$ and $\widehat{\mathcal{V}}_{i,jk}$ be the corresponding estimates. Denote

$$\gamma_1(s) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} E[\mathcal{V}_{i,jk} | t_{ij} = s]$$

and

$$\gamma_2(s_1, s_2) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sum_{j \neq k}^{m_i} E[\mathcal{V}_{i,jk} | t_{ij} = s_1, t_{ik} = s_2].$$

Define $\mathcal{B}_n^p = \{\zeta : \|\zeta\|_\infty \leq C, \|\zeta(x) - \zeta(y)\| \leq c[|x - y| + b_n], x, y \in [0, 1], \zeta \in \mathcal{F}_p\}$ for some constants $C > 0$ and $c > 0$, where $b_n = h + (nh)^{-1/2}(\log n)^{1/2}$. Before

beginning to proof the main results, we firstly present a lemma, which will be needed to prove the theorems.

Lemma 1 *Suppose that Conditions (C1)-(C7) hold, $\zeta_l(\cdot) \in \mathcal{F}_1$ is any bounded function having continuous third derivatives on its compact support, where $l = 1, \dots, p$. Then, for $J = 1, \dots, p$,*

$$\sup_{t \in [0,1]} |a_{nJ}(t) - Ea_{nJ}(t)| = O_p\{(nh)^{-1/2}(\log n)^{1/2}\},$$

where $a_{nJ}(t) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} \sum_{l=1}^p K_h(t_{ij} - t) \mathcal{V}_{i,jk} \{\alpha_l(t_{ik}) - \zeta_l(t_{ik})\} x_{il} x_{ij}$, and

$$\sup_{t \in [0,1]} |c_{nJ}(t)| = O_p\{h^2 + (nh)^{-1/2}h(\log n)^{1/2}\},$$

where $c_{nJ}(t) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} \sum_{l=1}^p h K_h(t_{ij} - t) \mathcal{V}_{i,jk} K_h(t_{ik} - t) \{\zeta_l(t) + \dot{\zeta}_l(t)(t_{ik} - t) - \zeta_l(t_{ik})\} x_{il} x_{ij}$.

The proof of this lemma is similar to that of Theorem 1 in Zhou et al. (2018) and follows from Theorem 37 and Example 38 in Chapter 2 of Pollard (1984).

A.2 Proof of Theorem 1

Let $\zeta(t) = (\zeta_1(t), \dots, \zeta_p(t))^T$, $\dot{\zeta}(t) = (\dot{\zeta}_1(t), \dots, \dot{\zeta}_p(t))^T$ and $\vartheta(t) = (\zeta^T(t), \dot{\zeta}^T(t))^T$. Denote the score function

$$S_m(\vartheta; t) = \frac{1}{n} \sum_{i=1}^n \Theta_i^T \mathbf{K}_{i,h}(t) \widehat{\mathbf{V}}_i^{-1} \times [Y_i - h \mathbf{K}_{i,h}(t) \Theta_i \vartheta(t) - \{I_{m_i} - h \mathbf{K}_{i,h}(t)\} \zeta^T(t) \mathbf{x}_i]. \tag{A.1}$$

For the sake of convenience, we define for $J = 1, \dots, p$,

$$S_J(\vartheta; t) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} \sum_{l=1}^p K_{ij}(t) \mathcal{V}_{i,jk} \{\alpha_l(t_{ik}) - \zeta_l(t_{ik})\} x_{il} x_{ij} - \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} \sum_{l=1}^p h K_{ij}(t) \mathcal{V}_{i,jk} K_{ik}(t) \{\zeta_l(t) + \dot{\zeta}_l(t)(t_{ik} - t) - \zeta_l(t_{ik})\} x_{il} x_{ij}, \tag{A.2}$$

$$S_J^*(\vartheta; t) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} \sum_{l=1}^p K_{ij}(t)(t_{ij} - t) \mathcal{V}_{i,jk} \{\alpha_l(t_{ik}) - \zeta_l(t_{ik})\} x_{il} x_{ij}$$

$$-\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} \sum_{l=1}^p h K_{ij}(t)(t_{ij} - t) \mathcal{V}_{i,jk} K_{ik}(t) \{ \zeta_l(t) + \dot{\zeta}_l(t)(t_{ik} - t) - \zeta_l(t_{ik}) \} x_{il} x_{ij}$$

and

$$s_j(\boldsymbol{\zeta}; t) = \sum_{l=1}^p \tau_{jl} f(t) \left[\gamma_1(t) \{ \alpha_l(t) - \zeta_l(t) \} + \int_0^1 \gamma_2(t, z) \{ \alpha_l(z) - \zeta_l(z) \} f(z) dz \right] \tag{A.3}$$

where $K_{ij}(t) = K_h(t_{ij} - t)$, $\tau_{jl} = E(x_{ij} x_{il})$, $\gamma_1(s)$ and $\gamma_2(s_1, s_2)$ are defined in ‘‘Appendix A.1’’. Let $S_j^{(1)}(\boldsymbol{\vartheta}; t)$ be the first term in the right of (A.2) and $S_j^{(2)}(\boldsymbol{\vartheta}; t)$ be the second term. Under the Conditions (C1)-(C4), it is easy to get that

$$\begin{aligned} S_j^{(1)}(\boldsymbol{\vartheta}; t) &= E\{S_j^{(1)}(\boldsymbol{\vartheta}; t)\} + O_p \left[\sqrt{\text{Var}\{S_j^{(1)}(\boldsymbol{\vartheta}; t)\}} \right] \\ &= E\{S_j^{(1)}(\boldsymbol{\vartheta}; t)\} + O_p\{(nh)^{-1/2}\} \\ &\rightarrow s_j(\boldsymbol{\zeta}; t). \end{aligned}$$

Similarly, we can check that $S_j^{(2)}(\boldsymbol{\vartheta}; t) = O_p\{(nh)^{-1/2}\}$ and $S_j^*(\boldsymbol{\vartheta}; t) = O_p\{(nh)^{-1/2}\}$.

Let $\mathbf{D}_1(\boldsymbol{\vartheta}; t) = (S_1(\boldsymbol{\vartheta}; t), \dots, S_p(\boldsymbol{\vartheta}; t))^T$, $\mathbf{D}_2(\boldsymbol{\vartheta}; t) = (S_1^*(\boldsymbol{\vartheta}; t), \dots, S_p^*(\boldsymbol{\vartheta}; t))^T$ and $\mathbf{D}(\boldsymbol{\vartheta}; t) = (\mathbf{D}_1(\boldsymbol{\vartheta}; t)^T, \mathbf{D}_2(\boldsymbol{\vartheta}; t)^T)^T$. Note that $\mathbf{S}_m(\boldsymbol{\vartheta}; t) = \mathbf{D}(\boldsymbol{\vartheta}; t) + o_p(1)$, then by Conditions (C1)-(C5), we have

$$\begin{aligned} \mathbf{S}_m(\boldsymbol{\vartheta}; t) &= E[\mathbf{D}(\boldsymbol{\vartheta}; t)] + O_p\{(nh)^{-1/2}\} + o_p(1) \\ &\rightarrow (s_1(\boldsymbol{\zeta}; t), \dots, s_p(\boldsymbol{\zeta}; t))^T \otimes (1, 0)^T \equiv \mathbf{s}_m(\boldsymbol{\zeta}; t) \otimes (1, 0)^T, \end{aligned}$$

where \otimes is the Kronecker product. The proposed estimation procedure shows that $\mathbf{S}_m(\widehat{\boldsymbol{\alpha}}; \widehat{\boldsymbol{\alpha}}; t) = 0$ uniformly over $t \in [0, 1]$, and it is easy to see $s_j(\boldsymbol{\alpha}; t) \equiv 0$ for $j = 1, \dots, p$. Then, it follows from Condition (C8) that $\boldsymbol{\alpha}(\cdot)$ is the unique root of the equation $s_m(\boldsymbol{\zeta}; t) \equiv 0$ over $\boldsymbol{\zeta} \in \mathcal{F}_p$. Therefore, to show the uniform consistency of $\widehat{\boldsymbol{\alpha}}(t)$, it suffices to prove the following three conclusions.

- (i) For any continuous function vector $\boldsymbol{\zeta}$ and bounded function vector $\dot{\boldsymbol{\zeta}}$, $\sup_{t \in [0,1]} \|\mathbf{S}_m(\boldsymbol{\vartheta}; t) - \mathbf{s}_m(\boldsymbol{\zeta}; t) \otimes (1, 0)^T\| = o_p(1)$.
- (ii) $\sup_{t \in [0,1]} \|\mathbf{S}_m(\boldsymbol{\vartheta}; t) - \mathbf{s}_m(\boldsymbol{\zeta}; t) \otimes (1, 0)^T\| = o_p(1)$ uniformly holds over $\boldsymbol{\zeta} \in \mathcal{B}_n^p$ and bounded function vector $\dot{\boldsymbol{\zeta}}$.
- (iii) $P(\widehat{\boldsymbol{\alpha}}(t) \in \mathcal{B}_n^p) \rightarrow 1$.

Once (i)-(iii) are established, applying the Arzela-Ascoli theorem and (iii), it is easy to show that, for any subsequence of $\widehat{\boldsymbol{\alpha}}(t)$, there exist a convergent subsequence $\widehat{\boldsymbol{\alpha}}_{n_k}(t)$ such that, uniformly over $t \in [0, 1]$, $\widehat{\boldsymbol{\alpha}}_{n_k}(t)$ converges to $\boldsymbol{\alpha}^*(t)$ in probability with $\boldsymbol{\alpha}^*(t) \in \mathcal{F}_p$. Observing that

$$s_m(\boldsymbol{\alpha}^*; t) \otimes (1, 0)^T = \mathbf{S}_m(\widehat{\boldsymbol{\alpha}}_{n_k}, \widehat{\boldsymbol{\alpha}}_{n_k}; t) - \{ \mathbf{S}_m(\widehat{\boldsymbol{\alpha}}_{n_k}, \widehat{\boldsymbol{\alpha}}_{n_k}; t) - \mathbf{s}_m(\widehat{\boldsymbol{\alpha}}_{n_k}; t) \otimes (1, 0)^T \}$$

$$- \{s_m(\widehat{\alpha}_{n_k}; t) \otimes (1, 0)^T - s_m(\alpha^*; t) \otimes (1, 0)^T\}$$

Note that $S_m(\widehat{\alpha}_{n_k}, \widehat{\alpha}_{n_k}; t) \equiv 0$, the continuity of $s_m(\zeta; t)$ and (ii) implies that $s_m(\alpha^*; t)$ converges to 0 uniformly over $t \in [0, 1]$. As a result, $s_m(\alpha^*; t) \equiv 0$. Since $s_m(\zeta; t) \equiv 0$ has a unique root α by Condition (C8), we then have $\alpha^*(t) = \alpha(t)$ for all $t \in [0, 1]$, which ensures the uniform consistency of $\widehat{\alpha}$.

Proof of (i). Note that

$$\sup_{t \in [0, 1]} \|S_m(\vartheta; t) - s_m(\zeta; t) \otimes (1, 0)^T\| \leq \sup_{t \in [0, 1]} \|D(\vartheta; t) - s_m(\zeta; t) \otimes (1, 0)^T\| + o_p(1).$$

For convenience, we only give the proof of $\sup_{t \in [0, 1]} \|S_J(\vartheta; t) - s_J(\zeta; t)\|$. The similar arguments lead to the conclusion about $\sup_{t \in [0, 1]} \|D(\vartheta; t) - s_m(\zeta; t) \otimes (1, 0)^T\|$. It follows from (A.2) and (A.3) that

$$\sup_{t \in [0, 1]} \|S_J(\vartheta; t) - s_J(\zeta; t)\| \leq I + II,$$

where

$$I \equiv \sup_{t \in [0, 1]} \left| \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} \sum_{l=1}^p K_{ij}(t) \mathcal{V}_{i,jk} \{ \alpha_l(t_{ik}) - \zeta_l(t_{ik}) \} x_{il} x_{ij} - \sum_{l=1}^p \tau_{jl} f(t) \left[\gamma_1(t) \{ \alpha_l(t) - \zeta_l(t) \} + \int_0^1 \gamma_2(t, z) \{ \alpha_l(z) - \zeta_l(z) \} f(z) dz \right] \right|$$

and

$$II \equiv \sup_{t \in [0, 1]} \left| \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} \sum_{l=1}^p h K_{ij}(t) \mathcal{V}_{i,jk} K_{ik}(t) \{ \zeta_l(t) + \dot{\zeta}_l(t)(t_{ik} - t) - \zeta_l(t_{ik}) \} x_{il} x_{ij} \right|.$$

By Conditions (C1)-(C7) and Lemma 1, it is easily seen that $I \leq O_p\{(nh)^{-1/2}(\log n)^{1/2}\}$ and $II \leq O_p\{h^2 + (nh)^{-1/2}h(\log n)^{1/2}\}$, which implies $\sup_{t \in [0, 1]} \|S_J(\vartheta; t) - s_J(\zeta; t)\| = o_p(1)$, and hence statement (i) follows.

Proof of (ii). By (A.1), we obtain that, for any continuous function vectors $\phi = (\phi_1, \dots, \phi_p)^T$ and any bounded function vector $\dot{\zeta} = (\dot{\zeta}_1, \dots, \dot{\zeta}_p)^T$,

$$\begin{aligned} S_m(\vartheta; t) - S_m(\phi, \dot{\zeta}; t) &= D(\zeta, \dot{\zeta}; t) - D(\phi, \dot{\zeta}; t) + o_p(1) \\ &\rightarrow \{s_m(\zeta; t) - s_m(\phi; t)\} \otimes (1, 0)^T. \end{aligned}$$

According to (A.2), (A.3) and Lemma 1, it can be shown that, uniformly over $t \in [0, 1]$,

$$S_J(\zeta, \dot{\zeta}; t) - S_J(\phi, \dot{\zeta}; t)$$

$$\begin{aligned}
 &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} \sum_{l=1}^p K_{ij}(t) \mathcal{V}_{i,jk} \{ \phi_l(t_{ik}) - \zeta_l(t_{ik}) \} x_{il} x_{ij} \\
 &\quad - \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} \sum_{l=1}^p h K_{ij}(t) \mathcal{V}_{i,jk} K_{ik}(t) \{ \zeta_l(t) + \dot{\zeta}_l(t)(t_{ik} - t) - \zeta_l(t_{ik}) \} x_{il} x_{ij} \\
 &\quad + \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} \sum_{l=1}^p h K_{ij}(t) \mathcal{V}_{i,jk} K_{ik}(t) \{ \phi_l(t) + \dot{\phi}_l(t)(t_{ik} - t) - \phi_l(t_{ik}) \} x_{il} x_{ij} \\
 &= \sum_{l=1}^p \tau_{jl} f(t) \left[\gamma_1(t) \{ \phi_l(t) - \zeta_l(t) \} + \int_0^1 \gamma_2(t, z) \{ \phi_l(z) - \zeta_l(z) \} f(z) dz \right] \\
 &\quad + O_p \{ h^2 + (nh)^{-1/2} (\log n)^{1/2} \} \\
 &\equiv s_j(\boldsymbol{\zeta}; t) - s_j(\boldsymbol{\phi}; t) + O_p \{ h^2 + (nh)^{-1/2} (\log n)^{1/2} \}.
 \end{aligned}$$

Then, for any $\boldsymbol{\zeta} \in \mathcal{B}_n^p$ and bounded function vector $\dot{\boldsymbol{\zeta}}$, we have

$$\sup_{t \in [0,1]} |S_j(\boldsymbol{\zeta}, \dot{\boldsymbol{\zeta}}; t) - S_j(\boldsymbol{\phi}, \dot{\boldsymbol{\zeta}}; t)| = \sum_{l=1}^p \sup_{t \in [0,1]} | \phi_l(t) - \zeta_l(t) | O_p(1) + o_p(1) \tag{A.4}$$

and

$$\sup_{t \in [0,1]} |s_j(\boldsymbol{\zeta}; t) - s_j(\boldsymbol{\phi}; t)| = \sum_{l=1}^p \sup_{t \in [0,1]} | \phi_l(t) - \zeta_l(t) | O_p(1). \tag{A.5}$$

Using the similar arguments in Zhou et al. (2018), (ii) immediately holds by (A.4) and (A.5).

Proof of (iii). Note that

$$\begin{aligned}
 \begin{pmatrix} \widehat{\boldsymbol{\alpha}}(t) \\ \widehat{\boldsymbol{\alpha}}(t) \end{pmatrix} &= \left\{ \frac{1}{n} \sum_{i=1}^n \boldsymbol{\Theta}_i^T \mathbf{K}_{i,h} \widehat{\mathbf{V}}_i^{-1} h \mathbf{K}_{i,h} \boldsymbol{\Theta}_i \right\}^{-1} \\
 &\quad \times \frac{1}{n} \sum_{i=1}^n \boldsymbol{\Theta}_i^T \mathbf{K}_{i,h} \widehat{\mathbf{V}}_i^{-1} \{ \mathbf{Y}_i - \widehat{\boldsymbol{\alpha}}^T(t_i) \mathbf{x}_i + h \mathbf{K}_{i,h} \widehat{\boldsymbol{\alpha}}^T(t_i) \mathbf{x}_i \} \\
 &= \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B} & \mathbf{C} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{U}_0 \\ \mathbf{U}_1 \end{pmatrix} \equiv \mathbf{G}_n^{-1}(t) \mathbf{H}_n(\widehat{\boldsymbol{\alpha}}; t),
 \end{aligned} \tag{A.6}$$

where \mathbf{A} , \mathbf{B} and \mathbf{C} are $p \times p$ symmetric matrices with the (r, s) th elements are, respectively, given by

$$A_{r,s} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} h K_{ij}(t) \widehat{\mathcal{V}}_{i,jk} K_{ik}(t) x_{ir} x_{is},$$

$$B_{r,s} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} h K_{ij}(t)(t_{ij} - t) \widehat{\mathcal{V}}_{i,jk} K_{ik}(t) x_{ir} x_{is},$$

$$C_{r,s} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} h K_{ij}(t)(t_{ij} - t) \widehat{\mathcal{V}}_{i,jk} K_{ik}(t)(t_{ik} - t) x_{ir} x_{is},$$

and $\mathbf{U}_0, \mathbf{U}_1$ are two p -dimensional vectors with the r th elements respectively are

$$U_{0r} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} K_{ij}(t) \widehat{\mathcal{V}}_{i,jk} \left\{ Y_{ik} - \sum_{l=1}^p \widehat{\alpha}_l(t_{ik}) x_{il} \right\} x_{ir}$$

$$+ \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} \sum_{l=1}^p h K_{ij}(t) \widehat{\mathcal{V}}_{i,jk} K_{ik}(t) \widehat{\alpha}_l(t_{ik}) x_{il} x_{ir},$$

$$U_{1r} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} K_{ij}(t)(t_{ij} - t) \widehat{\mathcal{V}}_{i,jk} \left\{ Y_{ik} - \sum_{l=1}^p \widehat{\alpha}_l(t_{ik}) x_{il} \right\} x_{ir}$$

$$+ \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} \sum_{l=1}^p h K_{ij}(t)(t_{ij} - t) \widehat{\mathcal{V}}_{i,jk} K_{ik}(t) \widehat{\alpha}_l(t_{ik}) x_{il} x_{ir},$$

$r, s = 1, \dots, p$. Denote

$$A_{r,s}^* = n^{-1} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} h K_{ij}(t) \mathcal{V}_{i,jk} K_{ik}(t) x_{ir} x_{is},$$

then by Condition (C6), it can be shown that

$$A_{r,s} = A_{r,s}^* + o_p(1) = EA_{r,s}^* + O_p \left\{ \sqrt{\text{Var}(A_{r,s}^*)} \right\} + o_p(1)$$

$$= \tau_{rs} f(t) \{ \gamma_1(t) v_0 + h f(t) \gamma_2(t, t) \} + O_p \{ (nh)^{-1/2} \}.$$

Similarly, we can get that

$$B_{r,s} = O_p \{ h(nh)^{-1/2} \},$$

$$C_{r,s} = \tau_{rs} h^2 f(t) \gamma_1(t) v_2 + O_p \{ h^2(nh)^{-1/2} \},$$

$$U_{0r} = \sum_{l=1}^p f(t) \tau_{rl} \left[v_0 \gamma_1(t) \zeta_l(t) + h f(t) \gamma_2(t, t) \zeta_l(t) + \gamma_1(t) \{ \alpha_l(t) - \zeta_l(t) \} \right]$$

$$+ \int_0^1 \gamma_2(t, z) \{ \alpha_l(z) - \zeta_l(z) \} f(z) dz + O_p \{ (nh)^{-1/2} \},$$

$$U_{1r} = O_p \{ h(nh)^{-1/2} \}.$$

Therefore, we have

$$\begin{aligned} \mathbf{G}_n(t) &= f(t) \begin{pmatrix} \gamma_1(t)v_0 + hf(t)\gamma_2(t, t) & 0 \\ 0 & h^2\gamma_1(t)v_2 \end{pmatrix} \otimes \begin{pmatrix} \tau_{11} & \cdots & \tau_{1p} \\ \vdots & \ddots & \vdots \\ \tau_{p1} & \cdots & \tau_{pp} \end{pmatrix} + o_p(1) \\ &= f(t) \begin{pmatrix} \{\gamma_1(t)v_0 + hf(t)\gamma_2(t, t)\}\Omega_X & 0 \\ 0 & h^2\gamma_1(t)v_2\Omega_X \end{pmatrix} + o_p(1), \end{aligned} \tag{A.7}$$

and $\mathbf{H}_n(\boldsymbol{\zeta}; t) = (H_1(\boldsymbol{\zeta}; t), \dots, H_p(\boldsymbol{\zeta}; t), \mathbf{0}_p)^T + o_p(1)$, where $\mathbf{0}_p$ is a p -dimensional row vector with elements 0, and for $r = 1, \dots, p$,

$$\begin{aligned} H_r(\boldsymbol{\zeta}; t) &= \sum_{l=1}^p f(t)\tau_{rl} \left[v_0\gamma_1(t)\zeta_l(t) + hf(t)\gamma_2(t, t)\zeta_l(t) \right. \\ &\quad \left. + \gamma_1(t)\{\alpha_l(t) - \zeta_l(t)\} + \int_0^1 \gamma_2(t, z)\{\alpha_l(z) - \zeta_l(z)\}f(z)dz \right]. \end{aligned} \tag{A.8}$$

Combine (A.7), (A.8) and the first equation of (A.6), we have

$$\widehat{\boldsymbol{\alpha}}(t) = \Omega_X^{-1}(H_1^*(\widehat{\boldsymbol{\alpha}}; t), \dots, H_p^*(\widehat{\boldsymbol{\alpha}}; t))^T + o_p(1),$$

where

$$\begin{aligned} H_r^*(\widehat{\boldsymbol{\alpha}}; t) &= \{\gamma_1(t)v_0 + hf(t)\gamma_2(t, t)\}^{-1} \sum_{l=1}^p \tau_{rl}[v_0\gamma_1(t)\widehat{\alpha}_l(t) \\ &\quad + hf(t)\gamma_2(t, t)\widehat{\alpha}_l(t) + \gamma_1(t)\{\alpha_l(t) - \widehat{\alpha}_l(t)\} + \kappa(\alpha_l, \widehat{\alpha}_l; t)] \end{aligned}$$

and

$$\kappa(\alpha_l, \widehat{\alpha}_l; t) = \int_0^1 \gamma_2(t, z)\{\alpha_l(z) - \widehat{\alpha}_l(z)\}f(z)dz$$

for $r = 1, \dots, p$. Following some straightforward derivations, we can obtain that

$$\sum_{l=1}^p \widehat{\alpha}_l(t)\tau_{rl} = \sum_{l=1}^p \alpha_l(t)\tau_{rl} + \sum_{l=1}^p \kappa(\alpha_l, \widehat{\alpha}_l; t)\tau_{rl}/\gamma_1(t) + o_p(1),$$

where $r = 1, \dots, p$. Let $\boldsymbol{\tau}_r = (\tau_{r1}, \dots, \tau_{rp})^T$ and $\kappa(\widehat{\boldsymbol{\alpha}}; t) = (\kappa(\alpha_1, \widehat{\alpha}_1; t), \dots, \kappa(\alpha_p, \widehat{\alpha}_p; t))^T$. It can be seen that $\boldsymbol{\tau}_r^T \{\widehat{\boldsymbol{\alpha}}(t) - \boldsymbol{\alpha}(t)\} = \boldsymbol{\tau}_r^T \kappa(\widehat{\boldsymbol{\alpha}}; t)/\gamma_1(t) + o_p(1)$. By Conditions (C3)-(C5), (iii) is held immediately. \square

A.3 Proof of Theorem 2

For convenience of notation, denote

$$\beta(t) = (\widehat{\alpha}_1(t) - \alpha_1(t), \dots, \widehat{\alpha}_p(t) - \alpha_p(t), h\{\widehat{\alpha}_1(t) - \dot{\alpha}_1(t)\}, \dots, h\{\widehat{\alpha}_p(t) - \dot{\alpha}_p(t)\})^T,$$

$$e_{n,1} = \sup_{t \in [0,1]} \|\widehat{\alpha}(t) - \alpha(t)\|, \quad e_{n,2} = \sup_{t \in [0,1]} \|h\widehat{\alpha}(t) - h\dot{\alpha}(t)\|, \quad e_n = e_{n,1} + e_{n,2}.$$

Note that $S_J(\widehat{\alpha}, \widehat{\alpha}; t)$ can be rewritten as

$$S_J(\widehat{\alpha}, \widehat{\alpha}; t) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} K_{ij}(t) \mathcal{V}_{i,jk} x_{ij} \{\alpha^T(t_{ik}) - \widehat{\alpha}^T(t_{ik})\} \mathbf{x}_i$$

$$- \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} h K_{ij}(t) \mathcal{V}_{i,jk} K_{ik}(t) x_{ij} \{\widehat{\alpha}^T(t) + \widehat{\alpha}^T(t)(t_{ik} - t) - \widehat{\alpha}^T(t_{ik})\} \mathbf{x}_i.$$

Then, it follows that

$$\frac{\partial S_J(\widehat{\alpha}, \widehat{\alpha}; t)}{\partial \widehat{\alpha}^T(t)} = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} h K_{ij}(t) \mathcal{V}_{i,jk} K_{ik}(t) x_{ij} \mathbf{x}_i^T,$$

$$\frac{\partial S_J(\widehat{\alpha}, \widehat{\alpha}; t)}{\partial \widehat{\alpha}^T(t_{ik})} = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} K_{ij}(t) \mathcal{V}_{i,jk} x_{ij} \{1 - h K_{ik}(t)\} \mathbf{x}_i^T,$$

$$\frac{\partial S_J(\widehat{\alpha}, \widehat{\alpha}; t)}{\partial \{h\widehat{\alpha}^T(t)\}} = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} h K_{ij}(t) \mathcal{V}_{i,jk} K_{ik}(t) x_{ij} \{(t_{ik} - t)/h\} \mathbf{x}_i^T.$$

It is easy to check that $E \{ \partial S_J(\widehat{\alpha}, \widehat{\alpha}; t) / \partial \widehat{\alpha}^T(t) \} = O(1)$, $E \{ \partial S_J(\widehat{\alpha}, \widehat{\alpha}; t) / \partial \widehat{\alpha}^T(t_{ik}) \} = O(1)$ and $E \{ \partial S_J(\widehat{\alpha}, \widehat{\alpha}; t) / \partial \{h\widehat{\alpha}^T(t)\} \} = O(h)$. Using a Taylor expansion and the consistency of $(\widehat{\alpha}, h\widehat{\alpha})$, we have, for $j = 1, \dots, p$,

$$S_J(\widehat{\alpha}, \widehat{\alpha}; t) - S_J(\alpha, \dot{\alpha}; t)$$

$$= -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} h K_{ij}(t) \mathcal{V}_{i,jk} K_{ik}(t) x_{ij} \mathbf{x}_i^T \{\widehat{\alpha}(t) - \alpha(t)\}$$

$$- \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} K_{ij}(t) \mathcal{V}_{i,jk} x_{ij} \{1 - h K_{ik}(t)\} \mathbf{x}_i^T \{\widehat{\alpha}(t_{ik}) - \alpha(t_{ik})\}$$

$$- \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} h K_{ij}(t) \mathcal{V}_{i,jk} K_{ik}(t) x_{ij} \{(t_{ik} - t)/h\} \mathbf{x}_i^T \{h\widehat{\alpha}(t) - h\dot{\alpha}(t)\}$$

$$+ O_p\{\|\widehat{\alpha}(t) - \alpha(t)\|\} + O_p\{\|\widehat{\alpha}(t_{ik}) - \alpha(t_{ik})\|\} + O_p\{h\|\widehat{h\alpha}(t) - h\dot{\alpha}(t)\|\}$$

$$\begin{aligned}
 &= -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} h K_{ij}(t) \mathcal{V}_{i,jk} K_{ik}(t) x_{ij} \times (\mathbf{x}_i^T \{(t_{ik} - t)/h\} \mathbf{x}_i^T) \boldsymbol{\beta}(t) \\
 &\quad - \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} K_{ij}(t) \mathcal{V}_{i,jk} x_{ij} \{1 - h K_{ik}(t)\} \mathbf{x}_i^T \{\widehat{\boldsymbol{\alpha}}(t_{ik}) - \boldsymbol{\alpha}(t_{ik})\} \\
 &\quad + O_p(e_{n,1} + h e_{n,2}) \\
 &= W_1 + W_2 + O_p(e_{n,1} + h e_{n,2}),
 \end{aligned}$$

where

$$\begin{aligned}
 W_1 &= -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} h K_{ij}(t) \mathcal{V}_{i,jk} K_{ik}(t) x_{ij} \times (\mathbf{x}_i^T \{(t_{ik} - t)/h\} \mathbf{x}_i^T) \boldsymbol{\beta}(t), \\
 W_2 &= -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} K_{ij}(t) \mathcal{V}_{i,jk} x_{ij} \{1 - h K_{ik}(t)\} \mathbf{x}_i^T \{\widehat{\boldsymbol{\alpha}}(t_{ik}) - \boldsymbol{\alpha}(t_{ik})\}.
 \end{aligned}$$

Using the similar technique in the proof of Theorem 1, it can be shown that

$$\begin{aligned}
 W_1 &= E W_1 + O_p\{\sqrt{\text{Var}(W_1)}\} \\
 &= - (v_0 \boldsymbol{\tau}_j^T \{\dot{f}(t)/f(t) + \dot{\gamma}_1(t)/\gamma_1(t)\} v_2 h \boldsymbol{\tau}_j^T) \times \gamma_1(t) f(t) \boldsymbol{\beta}(t) \\
 &\quad - h f^2(t) \gamma_2(t, t) \boldsymbol{\tau}_j^T \{\widehat{\boldsymbol{\alpha}}(t) - \boldsymbol{\alpha}(t)\} + o_p(h^2) + O_p\{(nh)^{-1/2} e_n\}
 \end{aligned}$$

and

$$\begin{aligned}
 W_2 &= (v_0 - v_0) f(t) \gamma_1(t) \boldsymbol{\tau}_j^T \{\widehat{\boldsymbol{\alpha}}(t) - \boldsymbol{\alpha}(t)\} + v_0 f(t) \boldsymbol{\tau}_j^T \kappa(\widehat{\boldsymbol{\alpha}}; t) \\
 &\quad + h f^2(t) \gamma_2(t, t) \boldsymbol{\tau}_j^T \{\widehat{\boldsymbol{\alpha}}(t) - \boldsymbol{\alpha}(t)\} + O_p\{(nh)^{-1/2} e_{n,1}\},
 \end{aligned}$$

where $\boldsymbol{\tau}_j = (\tau_{1j}, \dots, \tau_{pj})^T$ and $\kappa(\widehat{\boldsymbol{\alpha}}; t) = (\kappa(\alpha_1, \widehat{\alpha}_1; t), \dots, \kappa(\alpha_p, \widehat{\alpha}_p; t))^T$. Therefore, we can obtain that, for $j = 1, \dots, p$,

$$\begin{aligned}
 &S_j(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\alpha}}; t) - S_j(\boldsymbol{\alpha}, \boldsymbol{\alpha}; t) \\
 &= - (v_0 \boldsymbol{\tau}_j^T \{\dot{f}(t)/f(t) + \dot{\gamma}_1(t)/\gamma_1(t)\} v_2 h \boldsymbol{\tau}_j^T) \times \gamma_1(t) f(t) \boldsymbol{\beta}(t) \\
 &\quad + v_0 f(t) \boldsymbol{\tau}_j^T \kappa(\widehat{\boldsymbol{\alpha}}; t) + o_p(h^2) + O_p\{e_{n,1} + h e_{n,2} + (nh)^{-1/2} e_n\}.
 \end{aligned}$$

Similarly,

$$\begin{aligned}
 &S_j^*(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\alpha}}; t) - S_j^*(\boldsymbol{\alpha}, \boldsymbol{\alpha}; t) \\
 &= -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} h^2 K_{ij}(t) \{(t_{ij} - t)/h\} \mathcal{V}_{i,jk} K_{ik}(t) x_{ij}
 \end{aligned}$$

$$\begin{aligned} & \times (\mathbf{x}_i^T \{(t_{ik} - t)/h\} \mathbf{x}_i^T) \boldsymbol{\beta}(t) - \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} h K_{ij}(t) \{(t_{ij} - t)/h\} \\ & \times \mathcal{V}_{i,jk} x_{ij} \{1 - h K_{ik}(t)\} \mathbf{x}_i^T \{\widehat{\boldsymbol{\alpha}}(t_{ik}) - \boldsymbol{\alpha}(t_{ik})\} + O_p(h^2 e_{n,1} + h e_{n,2}) \\ = & - \left\{ \dot{f}(t)/f(t) + \dot{\gamma}_1(t)/\gamma_1(t) \right\} v_2 h^2 \boldsymbol{\tau}_j^T v_2 h \boldsymbol{\tau}_j^T \gamma_1(t) f(t) \boldsymbol{\beta}(t) \\ & + v_2 h^2 \boldsymbol{\tau}_j^T \{ \kappa_1(\widehat{\boldsymbol{\alpha}}; t) f(t) + \kappa(\widehat{\boldsymbol{\alpha}}; t) \dot{f}(t) \} \\ & + o_p(h^2) + O_p\{h^2 e_{n,1} + h e_{n,2} + (nh)^{-1/2} h e_n\}, \end{aligned}$$

where $\kappa_1(\widehat{\boldsymbol{\alpha}}; t) = (\kappa_1(\alpha_1, \widehat{\alpha}_1; t), \dots, \kappa_1(\alpha_p, \widehat{\alpha}_p; t))^T$ and $\kappa_l(\alpha_l, \widehat{\alpha}_l; t) = \int_0^1 \partial \gamma_2(t, z) / \partial t \{ \alpha_l(z) - \widehat{\alpha}_l(z) \} f(z) dz$ for $l = 1, \dots, p$. Thus, we can conclude that

$$\begin{aligned} & S_m(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\alpha}}; t) - S_m(\boldsymbol{\alpha}, \boldsymbol{\alpha}; t) \\ = & - \begin{pmatrix} v_0 \Omega_X & \{ \dot{f}(t)/f(t) + \dot{\gamma}_1(t)/\gamma_1(t) \} v_2 h \Omega_X \\ \{ \dot{f}(t)/f(t) + \dot{\gamma}_1(t)/\gamma_1(t) \} v_2 h^2 \Omega_X & v_2 h \Omega_X \end{pmatrix} \\ & \times \gamma_1(t) f(t) \boldsymbol{\beta}(t) + \begin{pmatrix} v_0 f(t) \Omega_X \kappa(\widehat{\boldsymbol{\alpha}}; t) \\ v_2 h^2 \Omega_X \{ \kappa_1(\widehat{\boldsymbol{\alpha}}; t) f(t) + \kappa(\widehat{\boldsymbol{\alpha}}; t) \dot{f}(t) \} \end{pmatrix} + o_p(h^2) \\ & + O_p \left\{ \begin{pmatrix} (e_{n,1} + h e_{n,2}) \mathbf{1}_p \\ (h^2 e_{n,1} + h e_{n,2}) \mathbf{1}_p \end{pmatrix} + (nh)^{-1/2} e_n \begin{pmatrix} \mathbf{1}_p \\ h \mathbf{1}_p \end{pmatrix} \right\}, \end{aligned} \tag{A.9}$$

where $\mathbf{1}_p$ is a p -dimensional column vector with elements 1. Furthermore, by Conditions (C5) and (C6), we have

$$\begin{aligned} S_m(\boldsymbol{\alpha}, \boldsymbol{\alpha}; t) = & \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} K_{ij}(t) \mathcal{V}_{i,jk} (\eta_{ik} + \varepsilon_{ik}) \boldsymbol{\xi}_{ij} \\ & + \frac{1}{2n} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} h K_{ij}(t) \mathcal{V}_{i,jk} K_{ik}(t) (t_{ik} - t)^2 \mathbf{x}_i^T \ddot{\boldsymbol{\alpha}}(t) \boldsymbol{\xi}_{ij} \\ & + \begin{pmatrix} o_p(h^2) \mathbf{1}_p + O_p\{(nh)^{-1/2}\} \mathbf{1}_p \\ o_p(h^4) \mathbf{1}_p + O_p\{h(nh)^{-1/2}\} \mathbf{1}_p \end{pmatrix}, \end{aligned} \tag{A.10}$$

where $\boldsymbol{\xi}_{ij} = (\mathbf{x}_i^T, (t_{ij} - t) \mathbf{x}_i^T)^T$. Note that $S_m(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\alpha}}; t) = 0$, then substitute (A.10) into (A.9), we get

$$\begin{aligned} \boldsymbol{\beta}(t) = & 1/\gamma_1(t) \left(\begin{matrix} \kappa(\widehat{\boldsymbol{\alpha}}; t) \\ \{ \kappa_1(\widehat{\boldsymbol{\alpha}}; t) - \dot{\gamma}_1(t)/\gamma_1(t) \kappa(\widehat{\boldsymbol{\alpha}}; t) \} h \end{matrix} \right) + \frac{1}{nf(t) \gamma_1(t)} \\ & \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} \left\{ K_{ij}(t) \mathcal{V}_{i,jk} (\eta_{ik} + \varepsilon_{ik}) + h K_{ij}(t) \mathcal{V}_{i,jk} K_{ik}(t) (t_{ik} - t)^2 \mathbf{x}_i^T \frac{\ddot{\boldsymbol{\alpha}}(t)}{2} \right\} \\ & \times \begin{pmatrix} \Omega_X^{-1} \mathbf{x}_i [1 - \{ \dot{f}(t)/f(t) + \dot{\gamma}_1(t)/\gamma_1(t) \} (t_{ij} - t) v_2 / v_2] \\ \Omega_X^{-1} \mathbf{x}_i [-h \{ \dot{f}(t)/f(t) + \dot{\gamma}_1(t)/\gamma_1(t) \} + (t_{ij} - t) / (v_2 h)] \end{pmatrix} \end{aligned}$$

$$+ O_p \left\{ \left((e_{n,1} + h e_{n,2}) \mathbf{1}_p \right) + (nh)^{-1/2} e_n \right\} + o_p \left\{ \begin{pmatrix} h^2 \mathbf{1}_p \\ h \mathbf{1}_p \end{pmatrix} \right\}. \tag{A.11}$$

Denote

$$\begin{aligned} \Delta_1(t) &= \frac{1}{nf(t)\gamma_1(t)} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} K_{ij}(t) \mathcal{V}_{i,jk} (\eta_{ik} + \varepsilon_{ik}) \Omega_X^{-1} \mathbf{x}_i, \\ \Delta_2(t) &= \frac{1}{nf(t)\gamma_1(t)} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} h K_{ij}(t) \mathcal{V}_{i,jk} K_{ik}(t) (t_{ik} - t)^2 \mathbf{x}_i^T \frac{\ddot{\alpha}(t)}{2} \Omega_X^{-1} \mathbf{x}_i. \end{aligned}$$

Following some straightforward calculations, we can easily obtain that $E\{\Delta_1(t)\} = 0$, $\text{Var}\{\Delta_1(t)\} = \frac{v_0 \Omega_X^{-1}}{f(t)\gamma_1(t)nh} (1 + o(1))$ and $E\{\Delta_2(t)\} = \frac{1}{2} v_2 h^2 \ddot{\alpha}(t) + o(h^2)$. Hence, it follows that

$$\begin{aligned} \widehat{\alpha}(t) - \alpha(t) &= - \int_z \frac{\gamma_2(t, z)}{\gamma_1(t)} \{\widehat{\alpha}(z) - \alpha(z)\} f(z) dz + \frac{1}{2} v_2 h^2 \ddot{\alpha}(t) + (nh)^{-1/2} \xi_n(t) \\ &+ O_p\{(e_{n,1} + h e_{n,2}) + (nh)^{-1/2} e_n\} + o_p(h^2) \end{aligned} \tag{A.12}$$

and

$$h\{\widehat{\alpha}(t) - \alpha(t)\} = O_p\{e_{n,2} + (nh)^{-1/2}\} + o_p(h), \tag{A.13}$$

where $\xi_n(t)$ converges to the multivariate normal distribution with mean 0 and covariance matrix $\Gamma(t) = v_0[\Omega_X f(t)\gamma_1(t)]^{-1}$. Using Lemma 1 and taking the supremum norm on both sides of (A.12) and (A.13), we have $e_{n,1} = O_p\{h^2 + (nh)^{-1/2}(\log n)^{1/2}\}$ and $e_{n,2} = O_p\{(nh)^{-1/2}(\log n)^{1/2}\} + o_p(h)$. Therefore, it is seen from (A.12) that

$$\begin{aligned} \widehat{\alpha}(t) - \alpha(t) &+ \int_z \frac{\gamma_2(t, z)}{\gamma_1(t)} \{\widehat{\alpha}(z) - \alpha(z)\} f(z) dz \\ &= \frac{1}{2} v_2 h^2 \ddot{\alpha}(t) + (nh)^{-1/2} \Gamma(t)^{1/2} \boldsymbol{\varphi} + o_p\{h^2 + (nh)^{-1/2}\}, \end{aligned}$$

where $\boldsymbol{\varphi}$ is a random vector following the standard normal distribution. Then the proof of Theorem 2 is completed. □

References

Cai TT, Hall P (2006) Prediction in functional linear regression. *Ann Stat* 34(5):2159–2179
 Chen D, Hall P, Müller HG (2011) Single and multiple index functional regression models with nonparametric link. *Ann Stat* 39(3):1720–1747
 Chen H, Wang Y (2011) A penalized spline approach to functional mixed effects model analysis. *Biometrics* 67(3):861–870
 Chen K, Guo S, Sun L, Wang JL (2010) Global partial likelihood for nonparametric proportional hazards models. *J Am Stat Assoc* 105(490):750–760

- Chiang CT, Rice JA, Wu CO (2001) Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables. *J Am Stat Assoc* 96(454):605–619
- Dai X, Hadjipantelis PZ, Han K, Ji H (2019) fdapace: functional data analysis and empirical dynamics. R package version (4):1 <https://CRAN.R-project.org/package=fdapace>
- Fan J, Gijbels I (1996) Local polynomial modelling and its applications. Chapman & Hall, London
- Fan J, Huang T, Li R (2007) Analysis of longitudinal data with semiparametric estimation of covariance function. *J Am Stat Assoc* 102(478):632–641
- Fan J, Zhang W (2008) Statistical methods with varying coefficient models. *Stat Interface* 1(1):179–195
- Ferraty F (2011) Recent advances in functional data analysis and related topics. Springer, New York
- Friston KJ (2009) Modalities, modes, and models in functional neuroimaging. *Science* 326(5951):399–403
- Hall P, Müller HG, Wang JL (2006) Properties of principal component methods for functional and longitudinal data analysis. *Ann Stat* 34(3):1493–1517
- Hand D, Crowder M (1996) Practical longitudinal data analysis. Chapman & Hall, London
- He G, Müller HG, Wang JL, Yang W (2010) Functional linear regression via canonical analysis. *Bernoulli* 16(3):705–729
- Jiang CR, Wang JL (2011) Functional single index models for longitudinal data. *Ann Stat* 39(1):362–388
- Li J, Huang C, Zhu H (2017) A functional varying-coefficient single-index model for functional response data. *J Am Stat Assoc* 112(519):1169–1181
- Li Y, Hsing T (2010) Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Ann Stat* 38(6):3321–3351
- Luo X, Zhu L, Zhu H (2016) Singlervarying coefficient model for functional responses. *Biometrics* 72(4):1275–1284
- Morris JS (2015) Functional regression. *Ann Rev Stat Appl* 2:321–359
- Pollard D (1984) Convergence of stochastic processes. Springer, New York
- Ramsay JO, Silverman BW (2005) Functional data analysis, 2nd edn. Springer, New York
- Reiss PT, Goldsmith J, Shang HL, Ogden RT (2017) Methods for scalar-on-function regression. *Int Stat Rev* 85(2):228–249
- Rice JA, Silverman BW (1991) Estimating the mean and covariance structure nonparametrically when the data are curves. *J R Stat Soc B* 53(1):233–243
- Sang P, Lockhart RA, Cao J (2018) Sparse estimation for functional semiparametric additive models. *J Multivar Anal* 168:105–118
- Smith SM, Jenkinson M, Johansen-Berg H, Rueckert D, Nichols TE, Mackay CE, Watkins KE, Ciccarelli O, Cader MZ, Matthews PM, Behrens TEJ (2006) Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *NeuroImage* 31(4):1487–1505
- Wang H, Zhong PS, Cui Y, Li Y (2018) Unified empirical likelihood ratio tests for functional concurrent linear models and the phase transition from sparse to dense functional data. *J R Stat Soc B* 80(2):343–364
- Wang JL, Chiou JM, Müller HG (2016) Functional data analysis. *Ann Rev Stat Appl* 3:257–295
- Wu CO, Chiang CT (2000) Kernel smoothing on varying coefficient models with longitudinal dependent variable. *Stat Sin* 10(2):433–456
- Wu S, Müller HG, Zhang Z (2013) Functional data analysis for point processes with rare events. *Stat Sin* 23(1):1–23
- Yao F, Müller HG, Wang JL (2005a) Functional linear regression analysis for longitudinal data. *Ann Stat* 33(6):2873–2903
- Yao F, Müller HG, Wang JL (2005b) Functional data analysis for sparse longitudinal data. *J Am Stat Assoc* 100(470):577–590
- Zhang JT, Chen J (2007) Statistical inferences for functional data. *Ann Stat* 35(3):1052–1079
- Zhang X, Park BU, Wang JL (2013) Time-varying additive models for longitudinal data. *J Am Stat Assoc* 108(503):983–998
- Zhang X, Wang JL (2015) Varying-coefficient additive models for functional data. *Biometrika* 102(1):15–32
- Zhou L, Lin H, Liang H (2018) Efficient estimation of the nonparametric mean and covariance functions for longitudinal and sparse functional data. *J Am Stat Assoc* 113(524):1550–1564
- Zhu H, Fan J, Kong L (2014) Spatially varying coefficient model for neuroimaging data with jump discontinuities. *J Am Stat Assoc* 109(507):1084–1098
- Zhu H, Li R, Kong L (2012) Multivariate varying coefficient model for functional responses. *Ann Stat* 40(5):2634

Zhu H, Zhang H, Ibrahim JG, Peterson BS (2007) Statistical analysis of diffusion tensors in diffusion-weighted magnetic resonance imaging data. *J Am Stat Assoc* 102(480):1085–1102

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.