Ann Inst Stat Math (2016) 68:181–208
DOI 10.1007/s10463-014-0488-3

Estimation and inference in functional single-index models

Shujie Ma

Received: 10 February 2014 / Revised: 31 July 2014 / Published online: 1 October 2014

Abstract We propose a functional single-index model (FSiM) to study the link between a scalar response variable and multiple functional predictors, in which the mean of the response is related to the linear predictors via an unknown link function. The FSiM serves as a good tool for dimension reduction in regression with multiple predictors and it is more flexible than functional linear models. Assuming that the functional predictors are observed at discrete points, we use B-spline basis functions to estimate the slope functions and the link function based on the least-squares criterion, and propose an iterative estimating procedure. Moreover, we provide uniform convergence rates of the proposed spline estimators in the FSiM, and construct asymptotic simultaneous confidence bands for the slope functions for inference. Our proposed method is illustrated by simulation studies and by an analysis of a diffusion tensor imaging data application.

Keywords Single-index models · Functional data analysis · Functional linear models · B splines · Confidence bands · Simultaneous inference

1 Introduction

In the literature of functional data analysis, functional linear models have been applied frequently to study the relationship between a scalar response and a functional predictor as given in Cardot et al. (2003), Ramsay and Silverman (2005), Yao et al. (2005), Cai and Hall (2006), Hall and Horowitz (2007), Li and Hsing (2007), among others. Let

Ma's research was partially supported by NSF grant DMS 1306972.

S. Ma (✉)
Department of Statistics, University of California-Riverside, Riverside, CA 92521, USA
e-mail: shujie.ma@ucr.edu

Y be a real random response and $X(t)$ be a square-integrable random function defined on some compact set $\mathcal{T}$. A functional linear model (FLM) is given as

$$Y = \int_{\mathcal{T}} \beta(t)X(t)\mathrm{d}t + \varepsilon, \tag{1}$$

where $\beta(\cdot)$ is a square-integrable unknown function defined on R and ε is the error term satisfying $E(\varepsilon|X(t)t \in \mathcal{T}) = 0$. Two commonly used approaches for estimating the slope function $\beta(\cdot)$ include the principal component regression (PCR) method based on spectral decompositions of both the covariance of X and its estimator (see Yao et al. 2005; Hall and Horowitz 2007) and the basis function expansions method with both $\beta(\cdot)$ and $X(\cdot)$ approximated by basis functions such as B-spline functions in Cardot et al. (2003) and Fourier basis functions in Li and Hsing (2007). Consequently, convergence rates of the resulting estimators by both methods have been studied in these papers.

To gain flexibility, Müller and Stadtmüller (2005) proposed a generalized functional linear model (GFLM) in which the expected value of the response is related to the linear predictor via a link function $g(\cdot)$ given as

$$Y = g\left(\int_{\mathcal{T}} \beta(t)X(t)\mathrm{d}t\right) + \varepsilon. \tag{2}$$

The proposed GFLM is useful when the linearity assumption is violated, and moreover, it can be naturally applied to the logistic, binomial or Poisson regression with a functional predictor. With multiple functional predictors, Gertheiss et al. (2013) proposed a penalized likelihood method to achieve variable selection and model estimation simultaneously. In a recent work of McLean et al. (2014), they proposed a functional additive model to allow a nonlinear relationship between the response and each functional predictor when multiple predictors present.

The generalized models studied in the above papers are apparently more flexible than the linear model given in (1). However, with a pre-selected parametric form for the link function $g(\cdot)$, the model may be misspecified. As extension, Chen et al. (2011) re-visited the GFLM (2) with an unknown link function which is estimated by nonparametric kernel smoothing. A global polynomial convergence rate for the estimator of $g(\cdot)$ has been studied in Chen et al. (2011).

The model with an unknown link function studied in Chen et al. (2011) contains only one functional predictor $X(\cdot)$. In real data applications, multiple predictors $X_1(t), \ldots, X_p(t)$ may be involved. A motivating example is a diffusion tensor imaging (DTI) study (Goldsmith et al. 2011, 2012) in which DTI brain scans are recorded for many multiple-sclerosis (MS) patients to assess the effect of neurodegeneration on disability. This data application contains a scalar outcome which is the paced auditory serial addition test (PASAT) score and two functional predictors which are the mean diffusivity profile of the corpus callosum tract (CCA) and the parallel diffusivity profile of the right corticospinal tract (RCST). Let $\mathbf{X}(t) = \{X_1(t), \ldots, X_p(t)\}^{\mathrm{T}}$ with $X_k(t)$, $k = 1, \ldots, p$, defined on a compact interval $\mathcal{T}$. To study the relationship between the scalar response (PASAT score) and the functional predictors (CCA and RCST), we consider the following functional single-index model (FSiM):

$$
\begin{aligned}
Y &= g\left(\int_{\mathcal{T}} \sum_{k=1}^{p} \beta_k(t) X_k(t) \mathrm{d}t\right) + \varepsilon \\
&= g\left(\int_{\mathcal{T}} \beta(t)^{\mathrm{T}} \mathbf{X}(t) \mathrm{d}t\right) + \varepsilon,
\end{aligned} \tag{3}
$$

where $\beta(t) = \{(\beta_1(t), \ldots, \beta_p(t))\}^{\mathrm{T}}$ are unknown smooth coefficient functions, and $g(\cdot)$ is an unknown nonconstant smooth link function. With the functional predictors observed at discrete points, we approximate the coefficient functions $\beta_k(\cdot)$ and the link function $g(\cdot)$ by two different sets of B-spline basis expansions (Boor 2001) with increasing numbers of basis functions, and propose an iterative method based on the least-squares criterion to estimate those nonparametric functions. With multiple functions $\beta_k(\cdot)$ involved, spline approximation is a more convenient way than kernel smoothing used in Chen et al. (2011), since those functions can be approximated by one-step linear combinations of spline basis functions. In contrast, it needs some iterative methods such as backfitting with kernel smoothing, which may be neither computationally efficient nor stable.

Moreover, we provide uniform convergence rates of the proposed spline estimators of both $\beta_k(\cdot)$ and $g(\cdot)$, which is technically challenging due to the involvement of different spline basis expansions. We conclude that the rates are slightly slower than the classical uniform nonparametric rate with a univariate nonparametric function. We also derive pointwise asymptotic normality for the estimator of $\beta_k(t)$. The asymptotic properties established provide a theoretical guidance for estimation in the FSiM (3) with spline basis functions.

Another challenge is to conduct statistical inference for the slope functions $\beta_k(t)$ in the FSiM (3). For the single-index model (SiM): $Y = g(\sum_{k=1}^{p} \beta_k X_k) + \varepsilon$, where $\beta = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$ are unknown parameters, estimation of it has been studied substantially in the literature. Given the asymptotic normality of the estimators (see Ichimura 1993 for the profile estimators; Carroll et al. 1997 for the backfitting estimators; Xia et al. 2002 and Jiang and Wang 2011 for the minimum average variance estimation), developing statistical inferential tools such as confidence intervals, Wald test, and likelihood ratio test for the index parameters β_k becomes straightforward. In the literature related to FLMs, many existing works focus on estimation of the slope functions $\beta_k(t)$, and some of them, for example Müller and Stadtmüller (2005) and Yao et al. (2005), provide pointwise asymptotic distributions for the functional estimators so that pointwise confidence intervals can be constructed. However, for functions $\beta_k(t)$ defined over a domain $\mathcal{T}$ instead of a single point, developing global inferential tools may be more needed. It is, however, of greater technical difficulty in devising such tools and establishing their theoretical properties. It is noteworthy that Goldsmith et al. (2011) proposed a mixed model approach by penalized splines for GFLM (2), so that confidence intervals and likelihood ratio tests can be obtained using standard mixed effects software.

In this paper, we construct asymptotic simultaneous confidence band (SCB) for $\beta_k(t)$ in the FSiM (3). SCBs provide a powerful tool for global inference of functions. Existing works on SCBs in the functional data setting are very limited due to the complex data nature, see Bunea et al. (2011) and Ma et al. (2012) for SCBs for the mean function and Crainiceanu et al. (2012) for a bootstrap method on the mean difference between two correlated functional processes. In this paper, we derive asymptotic SCBs for slope functions $\beta_k(t)$ of the functional predictors, which have not been investigated in this context, based on the absolute maxima distribution of a Gaussian process and strong approximation lemma.

The rest of this paper is organized as follows. Section 2 gives conditions for the model identification. Section 3 introduces the estimation procedure and presents asymptotic properties of the proposed estimators. In Sect. 4, we evaluate finite sample properties of the proposed estimation and inference procedures via simulation studies. Section 5 illustrates the proposed method through the analysis of the diffusion tensor imaging study. All technical details including detailed proofs are provided in the "Appendix".

2 Model identifiability

Suppose, we have a scalar response Y and functional predictors $X_1(t), \ldots, X_p(t)$, defined on a compact interval $\mathcal{T}$. Let $(Y_1, \mathbf{X}_1(t)), \ldots, (Y_n, \mathbf{X}_n(t))$, where $\mathbf{X}_i(t) = \{X_{i,1}(t), \ldots, X_{i,p}(t)\}^{\mathrm{T}}$, be independent and identically distributed observations from $(Y, \mathbf{X}(t))$, where $\mathbf{X}(t) = \{X_1(t), \ldots, X_p(t)\}^{\mathrm{T}}$. We assume that

(C1) $\beta_k(t)$ and $X_{i,k}(t)$ are square-integrable functions and random functions, which can be written as $\beta_k(t) = \sum_{l=1}^{\infty} \delta_{l,k}\varphi_l(t)$ and $X_{i,k}(t) = \sum_{l=1}^{\infty} \xi_{il,k}\varphi_l(t)$, for each $k = 1, \ldots, p$ and $t \in \mathcal{T}$, and suitable coefficients $\delta_k = \{\delta_{l,k}\}$ and $\xi_{i,k} = \{\xi_{il,k}\}$, where $\varphi_l(t)$, $l = 1, 2, \ldots$, are an orthonormal basis of the function space $L_2(\mathcal{T})$, $\delta_{l,k}$ are deterministic coefficients, and $\xi_{il,k}$ are random coefficients from iid copies of $\xi_{l,k}$ with variance 1. For all $1 \le i \le n$ and $1 \le k \le p$, $\sup_{t\in\mathcal{T}} |X_{i,k}(t)| < \infty$.

To ensure identification of model (3), we assume the following condition:

(C2) $\sum_{k=1}^{p} \int_{\mathcal{T}} \beta_k^2(t)\mathrm{d}t = 1$, $\beta_1(t)$ is nonconstant and monotone nondecreasing over $t \in \mathcal{T}$, or $g(u)$ is monotone nondecreasing over $u \in \mathcal{I}$.

Model (3) is identified under Condition (C2). To see this, let

$$g\left(\int_{\mathcal{T}} \sum_{k=1}^{p} \beta_k(t) X_{i,k}(t)\mathrm{d}t\right) = h\left(\int_{\mathcal{T}} \sum_{k=1}^{p} \alpha_k(t) X_{i,k}(t)\mathrm{d}t\right),$$

where $\beta_k(t)$ and $\alpha_k(t)$ are coefficient functions satisfying Conditions (C1) and (C2), and $g(\cdot)$ and $h(\cdot)$ are nonconstant link functions. By Condition (C1), we have $\beta_k(t) = \sum_{l=1}^{\infty} \delta_{l,k}\varphi_l(t)$ and $\alpha_k(t) = \sum_{l=1}^{\infty} \gamma_{l,k}\varphi_l(t)$ for some coefficients $\delta_k = \{\delta_{l,k}\}$ and $\gamma_k = \{\gamma_{l,k}\}$, and thus $g(\sum_{k=1}^{p} \delta_k^{\mathrm{T}} \xi_{i,k}) = h(\sum_{k=1}^{p} \gamma_k^{\mathrm{T}} \xi_{i,k})$. Let $U_i = \sum_{k=1}^{p} \delta_k^{\mathrm{T}} \xi_{i,k}$. Assume that $g'(U_i) \neq 0$ for some i. Thus,

$$(\partial g/\partial \xi_{il,k})/(\partial g/\partial \xi_{i1,1}) = \delta_{l,k}/\delta_{1,1} = \gamma_{l,k}/\gamma_{1,1} = (\partial h/\partial \xi_{il,k})/(\partial h/\partial \xi_{i1,1}).$$

Therefore, we have $\delta_{1,1} = \gamma_{1,1}$ or $\delta_{1,1} = -\gamma_{1,1}$ if $\sum_{k=1}^{p} \sum_{l=1}^{\infty} \delta_{l,k}^2 = 1$, and thus $\delta_{l,k} = \gamma_{l,k}$ or $\delta_{l,k} = -\gamma_{l,k}$, for all l and k, which leads to $\beta_k(t) = \alpha_k(t)$ or $\beta_k(t) =$

$-\alpha_k(t)$ for all $t \in \mathcal{T}$ and $k = 1, \dots, p$. To this end, by assuming that $\beta_1(t)$ is nonconstant and monotone nondecreasing over $t \in \mathcal{T}$ or the link function $g(u)$ is monotone nondecreasing over $u \in \mathcal{I}$, model (3) is identified.

3 Estimation

In real data applications, the curve $X_{i,k}(t)$ is discretized at dense time points $t_{i1}, \dots, t_{i,m_i+1}$ with $m_i \to \infty$ as $n \to \infty$. We first represent the integral $\int_{\mathcal{T}} \beta_k(t) X_{i,k}(t) \mathrm{d}t$ by $\sum_{j=1}^{m_i} (t_{i,j+1} - t_{ij}) \beta_k(t_{ij}) X_{i,k}(t_{ij})$ using Riemann integration, and then approximate the coefficient functions $\beta_k(t)$ and the link function $g(u)$ by B-spline functions. The estimation procedure is described as follows. Suppose that we have the qth order normalized B-spline basis functions $\mathbf{B}_1(u) = \{B_{s,1}(u), 1 \le s \le J_{n,1}\}^{\mathrm{T}}$ with knot sequence $\{\tau_s\}$ satisfying $\tau_1 = \cdots = \tau_q < \tau_{q+1} < \cdots < \tau_{J_{n,1}} < \tau_{J_{n,1}+1} = \cdots = \tau_{J_{n,1}+q}$, and the qth order normalized B-spline basis functions $\mathbf{B}_2(t) = \{B_{r,2}(t), 1 \le r \le J_{n,2}\}^{\mathrm{T}}$ with knot sequence $\{\upsilon_r\}$ satisfying $\upsilon_1 = \cdots = \upsilon_q < \upsilon_{q+1} < \cdots < \upsilon_{J_{n,2}} < \upsilon_{J_{n,2}+1} = \cdots = \upsilon_{J_{n,2}+q}$. The knot sequences also satisfy

$$\max_{q \le s \le J_{n,1}} |\tau_{s+1} - \tau_s| / \min_{q \le s \le J_{n,1}} |\tau_{s+1} - \tau_s| \le M_1,$$
$$\max_{q \le r \le J_{n,2}} |\upsilon_{r+1} - \upsilon_r| / \min_{q \le r \le J_{n,2}} |\upsilon_{r+1} - \upsilon_r| \le M_2,$$

uniformly in n for some constants $0 < M_1 < \infty$ and $0 < M_2 < \infty$. The above assumption on the distances between neighboring knots is typical in the polynomial spline regression literature to study asymptotic properties of the spline estimators, see Huang (2003). Then, $g(\cdot)$ and $\beta_k(\cdot)$ can be well approximated by B-spline functions such that

$$g(u) \approx g^0(u) = \sum_{r=1}^{J_{n,1}} \lambda_r B_{r,1}(u) = \mathbf{B}_1(u)^{\mathrm{T}} \lambda,$$
$$\beta_k(t) \approx \beta_k^0(t) = \sum_{s=1}^{J_{n,2}} \delta_{s,k} B_{s,2}(t) = \mathbf{B}_2(t)^{\mathrm{T}} \delta_k,$$

where $\lambda = (\lambda_1, \dots, \lambda_{J_{n,1}})^{\mathrm{T}}$ and $\delta_k = (\delta_{1,k}, \dots, \delta_{J_{n,2},k})^{\mathrm{T}}$. For simplicity, we use the splines with the same order to estimate $g(\cdot)$ and $\beta_k(\cdot)$. Our results can be easily extended to the case with splines of different orders. By the B-spline property, the derivative of $\beta_1^0(t)$ is

$$\dot{\beta}_1^0(t) = \partial \beta_1^0(t) / \partial t = \sum_{s=1}^{J_{n,2}} \delta_{s,1} \dot{B}_{s,2}(t) = \sum_{s=2}^{J_{n,2}} \frac{(q-1)(\delta_{s,1} - \delta_{s-1,1})}{\upsilon_{s+q-1} - \upsilon_s} B_{s,2}^{q-1}(t),$$

where $\{B_{s,2}^{q-1}(t) : 2 \le s \le J_{n,2}\}$ are B-spline functions with order $q-1$. Thus, when $\delta_{1,1} \le \cdots \le \delta_{J_{n,2},1}$, the B-spline function $\beta_1^0(t)$ is nondecreasing in $t \in \mathcal{T}$.

Let $\delta = (\delta_1^{\mathrm{T}}, \ldots, \delta_p^{\mathrm{T}})^{\mathrm{T}}$ and $\mathbf{W}_n = \int_{\mathcal{T}} \mathbf{B}_2(t)\mathbf{B}_2(t)^{\mathrm{T}} \mathrm{d}t$. Denote

$$\Phi_{i,sk} = \sum_{j=1}^{m_i} (t_{i,j+1} - t_{ij}) B_{s,2}(t_{ij}) X_{i,k}(t_{ij}),\ \Phi_{i,k} = (\Phi_{i,1k}, \ldots, \Phi_{i,J_{n,2}k})^{\mathrm{T}}, \quad (4)$$

and $\Phi_i = (\Phi_{i,1}^{\mathrm{T}}, \ldots, \Phi_{i,p}^{\mathrm{T}})^{\mathrm{T}}$. For given $\beta_k^0(\cdot)$, i.e., given δ_k, for $1 \le k \le p$, we obtain the least-squares estimator $\widehat{\lambda} = (\widehat{\lambda}_1, \ldots, \widehat{\lambda}_{J_{n,1}})^{\mathrm{T}}$ of $\lambda = (\lambda_1, \ldots, \lambda_{J_{n,1}})^{\mathrm{T}}$ and $\widehat{\delta}_k = (\widehat{\delta}_{1,k}, \ldots, \widehat{\delta}_{J_n,k})^{\mathrm{T}}$ of $\delta_k = (\delta_{1,k}, \ldots, \delta_{J_n,k})^{\mathrm{T}}$ by minimizing

$$L_n(\lambda, \delta) = \sum_{i=1}^{n} \left\{ Y_i - \sum_{s=1}^{J_{n,1}} \lambda_s B_{s,1}(\Phi_i^{\mathrm{T}} \delta) \right\}^2. \quad (5)$$

In practice, simultaneously estimating λ and δ from minimizing (5) cannot be achieved. Thus, we apply an iterative estimation method as frequently used in the single-index model literature given as below.

Step 0. (Initialization step). Obtain an initial value $\widehat{\delta}^0 = (\widehat{\delta}_1^{0\mathrm{T}}, \ldots, \widehat{\delta}_p^{0\mathrm{T}})^{\mathrm{T}}$ of δ by assuming a parametric form for the link function $g(\cdot)$, and set $\widehat{\delta}^0 = \widehat{\delta}^0 / (\sum_{k=1}^{p} \widehat{\delta}_k^{0\mathrm{T}} \mathbf{W}_n \widehat{\delta}_k^0)^{1/2}$.

Step I. Let $\widehat{\delta}$ be the estimate from the previous iteration. Find $\widehat{\lambda}(\widehat{\delta})$ by minimizing

$$\sum_{i=1}^{n} \left\{ Y_i - \sum_{s=1}^{J_{n,1}} \lambda_s B_{s,1}(\Phi_i^{\mathrm{T}} \widehat{\delta}) \right\}^2,$$

and thus

$$\widehat{\lambda}(\widehat{\delta}) = \{\widehat{\lambda}_1(\widehat{\delta}), \ldots, \widehat{\lambda}_{J_{n,1}}(\widehat{\delta})\}^{\mathrm{T}} = \{\mathcal{B}(\widehat{\delta})^{\mathrm{T}} \mathcal{B}(\widehat{\delta})\}^{-1} \mathcal{B}(\widehat{\delta})^{\mathrm{T}} \mathbf{Y}, \quad (6)$$

where $\mathcal{B}(\widehat{\delta}) = [\{\mathbf{B}_1(\Phi_1^{\mathrm{T}} \widehat{\delta}), \ldots, \mathbf{B}_1(\Phi_n^{\mathrm{T}} \widehat{\delta})\}^{\mathrm{T}}]_{n \times J_{n,1}}$ and $\mathbf{Y} = (Y_1, \ldots, Y_n)^{\mathrm{T}}$.

Step II. Update $\widehat{\delta}$ by minimizing

$$\widetilde{L}_n(\delta) = \sum_{i=1}^{n} \left\{ Y_i - \sum_{s=1}^{J_{n,1}} \widehat{\lambda}_s(\widehat{\delta}) B_{s,1}(\Phi_i^{\mathrm{T}} \delta) \right\}^2, \quad (7)$$

with respect to δ, subject to $\delta_{1,1} \le \cdots \le \delta_{J_{n,2},1}$ and let $\widehat{\delta} = \widehat{\delta} / (\sum_{k=1}^{p} \widehat{\delta}_k^{\mathrm{T}} \mathbf{W}_n \widehat{\delta}_k)^{1/2}$. This step is realized by nonlinear optimization with constraint $\delta_{1,1} \le \cdots \le \delta_{J_{n,2},1}$. We use the "constrOptim" package in R software to achieve this optimization. To use the "constrOptim" algorithm, we set $\widehat{\delta}$ from the previous iteration as the starting value, and we also need to specify the gradient of the objective function $\widetilde{L}_n(\delta)$ in (7) given as,

$$\partial \widetilde{L}_n(\delta)/\partial \delta = -2 \sum_{i=1}^{n} \left\{ Y_i - \sum_{s=1}^{J_{n,1}} \widehat{\lambda}_s(\widehat{\delta}) B_{s,1}(\Phi_i^{\mathrm{T}} \delta) \right\} \widehat{\dot{g}}(\Phi_i^{\mathrm{T}} \delta; \widehat{\delta}) \Phi_i,$$

where $\widehat{\dot{g}}(u; \widehat{\delta}) = \dot{\mathbf{B}}_1(u)^{\mathrm{T}} \widehat{\lambda}(\widehat{\delta})$, which is a spline estimator of $\dot{g}(u) = \partial g(u)/\partial u$.

Step III. Repeat steps I and II until convergence.

The final estimators of $\beta_k(t)$ and $g(u)$ are $\widehat{\beta}_k(t) = \mathbf{B}_2(t)^{\mathrm{T}}\widehat{\delta}_k$ and $\widehat{g}(u;\widehat{\delta}) = \mathbf{B}_1(u)^{\mathrm{T}}\widehat{\lambda}(\widehat{\delta})$, respectively. Let $C^{(\alpha)}(\mathcal{X}) = \{\phi|\phi^{(\alpha)} \in C(\mathcal{X})\}$ be the space of the αth smooth functions, and $C^{0,1}(\mathcal{X})$ be the space of Lipschitz continuous function on $\mathcal{X}$, i.e.,

$$C^{0,1}(\mathcal{X}) = \left\{\phi : \|\phi\|_{0,1} = \sup_{w \neq w', w, w' \in \mathcal{X}} \frac{|\phi(w) - \phi(w')|}{|w - w'|} < \infty\right\}.$$

Let $m_{\min} = \min_{1\le i\le n}(m_i)$ and $m_{\max} = \max_{1\le i\le n}(m_i)$. For establishing asymptotic consistency of the estimators, we need the following assumptions.

(C3) For every $1 \le k \le p$, $\beta_k(\cdot) \in C^{(\alpha)}(\mathcal{T})$ and $g(\cdot) \in C^{(\alpha)}(\mathcal{I})$, for some integer $\alpha \ge 2$. For every $1 \le k \le p$, $X_k(\cdot) \in C^{0,1}(\mathcal{T})$, and $E(X_k(\cdot)|U) \in C^{0,1}(\mathcal{T})$, where $U = \int_{\mathcal{T}} \beta(t)^{\mathrm{T}}\mathbf{X}(t)\mathrm{d}t$. The spline order satisfies $q \ge \alpha$.

(C4) For all $1 \le i \le n$,

$$\max_{1\le j\le m_i} |t_{i,j+1} - t_{ij}| / \min_{1\le j\le m_i} |t_{i,j+1} - t_{ij}| \le M_3,$$

for some constant $0 < M_3 < \infty$.

(C5) $\mathrm{Var}\{Y|\mathbf{X}(t), t \in \mathcal{T}\} \le M_4$, for some constant $0 < M_4 < \infty$.

Condition (C3) gives smoothness conditions of the nonparametric functions which are commonly assumed in the spline smoothing literature such as in Zhou et al. (1998). Condition (C4) is the condition for distances between neighboring time points. Examples with equally spaced time points are a special case. Condition (C5) states boundedness of the conditional variance of the response, see Assumption (H.2) in Cardot et al. (2003) for the same condition. To present the asymptotic results, we introduce some notations as follows. For any matrix $\mathbf{A}$, denote $\mathbf{A}^{\otimes 2} = \mathbf{A}\mathbf{A}^{\mathrm{T}}$. For any vector $\zeta = (\zeta_1, \ldots, \zeta_s)^{\mathrm{T}} \in R^s$, denote $\|\zeta\|_\infty = \max_{1\le l\le s}|\zeta_l|$. For any positive numbers a_n and b_n, let $a_n \ll b_n$ denote that $a_n/b_n = o(1)$ and let $a_n \asymp b_n$ denote that $\lim_{n\to\infty} a_n/b_n = c$, for a positive constant c.

According to the result on page 149 of Boor (2001), for $g(\cdot)$ and $\beta_k(\cdot)$ satisfying Condition (C3), there are spline functions

$$g_n(u) = \sum_{s=1}^{J_{n,1}} \lambda_{s,n}^0 B_{s,1}(u) = \mathbf{B}_1(u)^{\mathrm{T}}\lambda_n^0$$

with $\lambda_n^0 \in R^{J_{n,1}}$ and $\beta_{k,n}(t) = \sum_{s=1}^{J_{n,2}} \delta_{s,k,n} B_{s,2}(t) = \mathbf{B}_2(t)^{\mathrm{T}}\delta_{k,n}^0$ with $\delta_{k,n}^0 \in R^{J_{n,2}}$ such that

$$\sup_{u\in\mathcal{I}} |g_n(u) - g(u)| = O(J_{n,1}^{-\alpha}),$$
$$\sup_{t\in\mathcal{T}} |\beta_{k,n}(t) - \beta_k(t)| = O(J_{n,2}^{-\alpha}). \tag{8}$$

Let $\delta_n^0 = (\delta_{1,n}^{0\mathrm{T}}, \ldots, \delta_{p,n}^{0\mathrm{T}})^{\mathrm{T}}$. The proposition below gives the uniform convergence rates of the spline estimators $\widehat{g}(u;\delta_n)$ and $\widehat{\dot{g}}(u;\delta_n)$ for δ_n in a neighborhood of δ_n^0.

Proposition 1 *Under Conditions (C1)–(C5) in the "Appendix", for* $J_{n,1}^3/n = o(1)$, $J_{n,1}J_{n,2}^{-\alpha} = o(1)$, $J_{n,1}m_{\min}^{-1} = o(1)$, *and for* $a_n = o(J_{n,1}^{-1})$, *we have* (i)

$$\sup_{\|\delta_n-\delta_n^0\|_\infty \le a_n} \sup_{u\in\mathcal{I}} |\widehat{g}(u;\delta_n) - g(u)| = O_p\left(a_n + J_{n,1}^{-\alpha} + J_{n,2}^{-\alpha} + m_{\min}^{-1} + \sqrt{\log(n)J_{n,1}n^{-1}}\right);$$

and (ii)

$$\sup_{\|\delta_n-\delta_n^0\|_\infty \le a_n} \sup_{u\in\mathcal{I}} |\widehat{\dot{g}}(u;\delta_n) - \dot{g}(u)| = O_p\left(J_{n,1}a_n + J_{n,1}^{1-\alpha} + J_{n,1}J_{n,2}^{-\alpha} + J_{n,1}m_{\min}^{-1} + \sqrt{\log(n)J_{n,1}^3n^{-1}}\right).$$

To derive the uniform convergence rates of the final estimators $\widehat{\beta}_k(t)$ and $\widehat{g}(u;\widehat{\delta})$, we need the initial value of δ denoted by $\widehat{\delta}_n^0$ to satisfy $\|\widehat{\delta}_n^0 - \delta_n^0\|_\infty = O\{(\log n)^{1/2}J_{n,2}^{1/2}n^{-1/2}\}$. In data analysis, the starting value $\widehat{\delta}_n^0$ is obtained by assuming that $g(\cdot)$ is a linear function, which is commonly used in the single-index model literature, see Carroll et al. (1997) and Xia et al. (2002). We next introduce some notations. Denote $U_i = \int_{\mathcal{T}} \beta(t)^{\mathrm{T}}\mathbf{X}_i(t)\mathrm{d}t$, $\Psi_i = \dot{g}(U_i)\{\Phi_i - E(\Phi_i|U_i)\}$ and $\mathrm{Var}(\varepsilon_i|U_i) = \sigma^2(U_i)$, for $i = 1,\ldots,n$. Let $\mathbf{\Lambda}_k = [(\mathbf{0}_{J_{n,2}\times J_{n,2}},\ldots,\mathbf{I}_{J_{n,2}\times J_{n,2}},\ldots,\mathbf{0}_{J_{n,2}\times J_{n,2}})]_{J_{n,2}\times pJ_{n,2}}$ be the $J_{n,2}\times pJ_{n,2}$ matrix consisting of p matrices of $J_{n,2}\times J_{n,2}$ dimension with the identity matrix $\mathbf{I}_{J_{n,2}\times J_{n,2}}$ as the kth matrix and $\mathbf{0}_{J_{n,2}\times J_{n,2}}$ as other matrices. Then, $\widehat{\beta}_k(t) = \mathbf{B}_2(t)^{\mathrm{T}}\mathbf{\Lambda}_k\widehat{\delta}$. For $1\le k\le p$, define

$$\sigma_{n,k}^2(t) = \mathbf{B}_2(t)^{\mathrm{T}}\mathbf{\Lambda}_k\left\{\sum_{i=1}^n E(\Psi_i^{\otimes 2})\right\}^{-1}\left[\sum_{i=1}^n E\{\sigma^2(U_i)\Psi_i^{\otimes 2}\}\right] \times \left\{\sum_{i=1}^n E(\Psi_i^{\otimes 2})\right\}^{-1}\mathbf{\Lambda}_k^{\mathrm{T}}\mathbf{B}_2(t).$$

Let $\widehat{\delta}$ be the minimizer of $\widetilde{L}_n(\delta)$ given in (7) subject to $\delta_{1,1}\le\cdots\le\delta_{J_{n,2},1}$ satisfying $\|\widehat{\delta}-\delta_n^0\|_\infty\le a_n$ with probability approaching 1, so that we assume the minimizer happens in a consistent neighborhood of δ_n^0. The following theorems present the uniform convergence rates of $\widehat{\beta}_k(t) = \mathbf{B}_2(t)^{\mathrm{T}}\widehat{\delta}_k$ and $\widehat{g}(\cdot;\widehat{\delta}) = \mathbf{B}_1(u)^{\mathrm{T}}\widehat{\lambda}(\widehat{\delta})$, and asymptotic normality for the coefficient estimator $\widehat{\beta}_k(t)$.

Theorem 1 *Under Conditions (C1)–(C5),* $n^{1/(2\alpha+1)} \ll J_{n,2} \ll n^{1/3}(\log n)^{-1}$, $n^{1/(2\alpha+3)} \ll J_{n,1} \ll J_{n,2} \ll J_{n,1}^2$, *and* $n^{1/2}m_{\min}^{-1}J_{n,2}^{-1/2} = o(1)$, *we have for given* $1\le k\le p$, (i) $\sup_{t\in\mathcal{T}}|\widehat{\beta}_k(t)-\beta_k(t)| = O_p(\sqrt{(\log n)J_{n,2}n^{-1}})$; *and* (ii) *for any* $t\in\mathcal{T}$, *as* $n\to\infty$, $\sigma_{n,k}^{-1}(t)\{\widehat{\beta}_k(t)-\beta_k(t)\}\to\mathcal{N}(0,1)$.

Theorem 2 *Under Conditions (C1)–(C5),* $n^{1/(2\alpha+1)} \ll J_{n,2} \ll n^{1/3}(\log n)^{-1}$, $n^{1/(2\alpha+3)} \ll J_{n,1} \ll J_{n,2} \ll J_{n,1}^2$, *and* $n^{1/2}m_{\min}^{-1}J_{n,2}^{-1/2} = o(1)$, *we have* $\sup_{u\in\mathcal{I}} |\widehat{g}(u;\widehat{\delta}) - g(u)| = O_p(J_{n,1}^{-\alpha} + (\log n)^{1/2}J_{n,2}^{1/2}n^{-1/2})$.

Remark let $J_{n,1} \asymp n^{1/(2\alpha+1)}$ and $J_{n,2} \asymp n^{(2v+1)/(2\alpha+1)}$ for a small $v > 0$. Then, we have

$$\sup_{t\in\mathcal{T}} |\widehat{\beta}_k(t) - \beta_k(t)| = O_p\left(\sqrt{(\log n)}n^{(v-\alpha)/(2\alpha+1)}\right)$$

and $\sup_{u\in\mathcal{I}} |\widehat{g}(u;\widehat{\delta}) - g(u)| = O_p(\sqrt{(\log n)}n^{(v-\alpha)/(2\alpha+1)})$. By assuming $\alpha = 2$ such that $\beta_k(t)$ and $g(u)$ are second smooth functions, the uniform convergence rate would be $O_p(\sqrt{(\log n)}n^{(v-2)/5})$ for both estimators, which is slighter slower than the classical uniform nonparametric rate $O_p(\sqrt{(\log n)}n^{-2/5})$.

For $1 \le k \le p$, let $\xi_k(t)$ be a Gaussian process with $E\{\xi_k(t)\} \equiv 0$, $\mathrm{Var}\{\xi_k(t)\} \equiv 1$, and covariance matrix

$$\mathrm{Cov}(\eta_k^0(t), \eta_k^0(t')) = \sigma_{n,k}^{-1}(t)\sigma_{n,k}^{-1}(t')\mathbf{B}_2(t)^{\mathrm{T}}\mathbf{\Lambda}_k\left\{\sum_{i=1}^n E(\Psi_i^{\otimes 2})\right\}^{-1} \times\left[\sum_{i=1}^n E\{\sigma^2(U_i)\Psi_i^{\otimes 2}\}\right]\left\{\sum_{i=1}^n E(\Psi_i^{\otimes 2})\right\}^{-1}\mathbf{\Lambda}_k^{\mathrm{T}}\mathbf{B}_2(t'). \quad (9)$$

Define the $100(1-\alpha)$th percentile of the absolute maxima distribution of $\xi_k(t)$ as $Q_k(\alpha)$ which satisfies

$$P\{\sup_{t\in T} |\xi_k(t)| \le Q_k(\alpha)\} = 1-\alpha.$$

We now state the following theorem used to construct the asymptotic simultaneous confidence bands (ASCB) for coefficient functions $\beta_k(t)$.

Theorem 3 *Under Conditions (C1)–(C5),* $n^{1/(2\alpha+1)} \ll J_{n,2} \ll n^{1/3}(\log n)^{-1}$, $n^{1/(2\alpha+3)} \ll J_{n,1} \ll J_{n,2}$, *and* $n^{1/2}m_{\min}^{-1}J_{n,2}^{-1/2} = o(1)$, *we have*

$$\lim_{n\to\infty} P\left\{\sup_{t\in T} |\sigma_{n,k}^{-1}(t)\{\widehat{\beta}_k(t) - \beta_k(t)\}| \le Q_k(\alpha)\right\} = 1-\alpha.$$

Remark Based on the results in Theorem (3), we can construct $(1-\alpha)100\,\%$ ASCBs for $\beta_k(t)$ given as

$$\widehat{\beta}_k(t) \pm Q_k(\alpha)\sigma_{n,k}(t). \quad (10)$$

4 Simulation

In this section, we study the finite sample performance of the proposed method. The predictor functions are generated as $X_{i,k}(t) = t + \sum_{j=1}^4 \xi_{ij,k}\phi_{j,k}(t)$, $k = 1, 2$, $i = 1, \ldots, n$, where $\xi_{ij,k}$ are i.i.d $N(0, \sigma_j^2)$ with $\sigma_j^2 = 1/2^{(j-1)}$, and $\phi_{1,k}(t) = (1/\sqrt{2})\sin(2\pi t)$, $\phi_{2,k}(t) = (1/\sqrt{2})\cos(2\pi t)$, $\phi_{3,k}(t) = (1/\sqrt{2})\sin(4\pi t)$, $\phi_{4,k}(t) = (1/\sqrt{2})\cos(4\pi t)$. Let $n = 100, 200, 500$, and each predictor function is sampled through $m = 100$ equally spaced measurements in [0, 1]. We let $\beta_1(t) = \cos(\pi t + \pi)$ and $\beta_2(t) = \sin(\pi t) - 2/\pi$, so $\beta_1(t)$ is monotone increasing in t. Let $U_i = \int_{\mathcal{T}} \sum_{k=1}^{2} \beta_k(t) X_{i,k}(t) \mathrm{d}t$. We generate continuous responses from the FSiM $Y_i = g(U_i) + \varepsilon_i$ with three different link functions which lead to three models given as:

M1: $Y_i = 5U_i + \varepsilon_i$ (linear link function);
M2: $Y_i = 5\sin(\pi(U_i - U_{\min})/U_{\mathrm{diff}} - \pi/2) + \varepsilon_i$, where $U_{\mathrm{diff}} = U_{\max} - U_{\min}$ (monotone increasing link function);
M3: $Y_i = 5\sin(\pi(U_i - U_{\min})/U_{\mathrm{diff}}) + \varepsilon_i$ (non-monotone link function).

In the above models, the error term ε_i is simulated from i.i.d. normal distribution with mean 0 and standard deviation $\mathrm{SD}(\varepsilon_i) = 0.5\mathrm{SD}(g(U_i))$. We use cubic splines with order $q = 4$ to approximate the nonparametric functions $g(\cdot)$ and $\beta_k(\cdot)$. The numbers of knots N_1 and N_2 for the B-spline bases $\mathbf{B}_1(u)$ and $\mathbf{B}_2(t)$ are selected by minimizing the BIC criterion given as

$$n \log \left[n^{-1} \sum_{i=1}^{n} \{Y_i - \mathbf{B}_1(\Phi_i^{\mathrm{T}} \widehat{\delta})^{\mathrm{T}} \widehat{\lambda}\}^2 \right] + \log(n)\{N_1 + q + p(N_2 + q)\}.$$

See Xue and Yang (2006) for the AIC and BIC for the number of knots selection in spline regression. To evaluate the performance of the proposed estimation method, we quantify the prediction error by average squared errors (ASE) given as

$$\mathrm{ASE} = n^{-1} \sum_{i=1}^{n} \{\widehat{Y}_i - g(U_i)\}^2,$$

where $\widehat{Y}_i = \widehat{g}(\int_{\mathcal{T}} \sum_{k=1}^{2} \widehat{\beta}_k(t) X_{i,k}(t) \mathrm{d}t)$. We also quantify the errors of the estimated coefficient functions by squared error given as

$$\mathrm{SE}(\widehat{\beta}_k) = \int_{\mathcal{T}} \{\widehat{\beta}_k(t) - \beta_k(t)\}^2 \mathrm{d}t.$$

When we treat the link function $g(\cdot)$ as a linear function, estimation in models M2 and M3 may encounter the misspecification problem. For illustration, we compare the estimation results by the proposed method in the FSiM and results in the functional linear model (FLM) with a linear link function. In the FLM, we estimate the spline parameters $\delta = (\delta_1^{\mathrm{T}}, \ldots, \delta_p^{\mathrm{T}})^{\mathrm{T}}$ in the coefficient functions $\beta_k(t) = \mathbf{B}_2(t)^{\mathrm{T}} \delta_k$ by minimizing

$$L_n(\delta) = \sum_{i=1}^{n} (Y_i - \Phi_i^{\mathrm{T}} \delta)^2, \tag{11}$$

and then let $\widehat{\delta} = \widehat{\delta} / (\sum_{k=1}^{p} \widehat{\delta}_k^{\mathrm{T}} \mathbf{W}_n \widehat{\delta}_k)^{1/2}$. The number of knots is also selected by the BIC criterion.

Table 1 reports the ASEs and SEs for the proposed estimator in the FSiM and estimator obtained from minimizing (11) in the FLM with data generated from models M1-M3, respectively. We also compare estimation results in the FSiM using different starting values in the proposed iterative algorithm in Section 3. Column FSiM-C shows the results, when the starting values $\widehat{\delta}_k^0$ are obtained by minimizing $\int_{\mathcal{T}} \{\beta_k(t) - \mathbf{B}_2(t)^{\mathrm{T}} \delta_k\}^2 \mathrm{d}t$, so that $\widehat{\delta}_k^0$ is a consistent estimator. Clearly, this initial estimator is not

Table 1 The ASEs and SEs for the proposed estimators in the FSiM and estimators from minimizing (11) in the FLM with data generated from models M1, M2 and M3

Model	n	FSiM-C			FSiM			FLM		
		ASE	SE($\widehat{\beta}_1$)	SE($\widehat{\beta}_2$)	ASE	SE($\widehat{\beta}_1$)	SE($\widehat{\beta}_2$)	ASE	SE($\widehat{\beta}_1$)	SE($\widehat{\beta}_2$)
M1	100	0.0666	0.0564	0.0226	0.0732	0.0464	0.0393	0.0613	0.0619	0.0593
	200	0.0306	0.0402	0.0140	0.0376	0.0423	0.0223	0.0302	0.0496	0.0300
	500	0.0116	0.0325	0.0097	0.0143	0.0300	0.0118	0.0116	0.0422	0.0167
M2	100	0.1716	0.0587	0.0251	0.1911	0.0499	0.0381	0.2754	0.1541	0.0811
	200	0.0693	0.0502	0.0159	0.0860	0.0325	0.0196	0.1590	0.1366	0.0599
	500	0.0233	0.0336	0.0103	0.0279	0.0253	0.0108	0.0841	0.1167	0.0455
M3	100	0.0353	0.0479	0.0169	0.0507	0.0642	0.0332	1.2257	0.5839	0.1491
	200	0.0123	0.0388	0.0109	0.0181	0.0539	0.0192	0.9508	0.5247	0.1105
	500	0.0035	0.0367	0.0083	0.0062	0.0505	0.0143	0.6882	0.5215	0.0992

obtainable in real data analysis since the true slope functions $\beta_k(t)$ are unknown. We use it for the purpose of comparing with the results in Column FSiM with the initial estimate $\widehat{\delta}_k^0$ obtained by assuming a linear link function as used in real data applications. We can see that with data generated from all of the three models M1–M3, the values of ASEs and SEs shown in FSiM-C and FSiM are comparable. This result indicates that we obtain reasonable estimates using different initial values. Moreover, we observe that the ASE and SE values decrease as the sample size increases, which corroborates with our consistency results established in Theorems 1 and 2 for $\widehat{\beta}_k(t)$ and $\widehat{g}(u; \widehat{\delta})$. When data are generated from model M3 with a nonlinear and non-monotone link function, using the FLM which is misspecified, it leads to large values of both ASEs and SEs for the resulting estimators comparing to those values in FSiM-C and FSiM. With data generated from model M2 with a nonlinear, but monotone link function, the ASE and SE values in FLM decrease, but they are still larger than those values in FSiM-C and FSiM. For M1 that the FLM is the true model, the results in FSiM-C, FSiM and FLM are comparable. This indicates that although the link function $g(\cdot)$ is linear, the proposed method in the FSiM by estimating $g(\cdot)$ nonparametrically still performs well.

We next compare the performance of the proposed FSiM, the FLM method mentioned above and the PFR method in the "refund" R package for fitting the GFLM. For the PFR method, we use the default value for the dimension of the B-spline basis, and set the family as Gaussian. Table 2 presents the ASEs for the three methods with data generated from the models M1, M2 and M3. We can clearly see that the FLM and the PFR methods have very similar ASE values for all cases, and moreover for all other cases except $n = 500$ and M3, the FLM has slightly smaller ASEs than the PFR. This result demonstrates that the regression spline with the number of knots selected by the BIC and the penalized spline with a penalty term controlling the roughness of the fit have similar performance in this context. It also corroborates with the finding in Claeskens et al. (2009) that the asymptotic properties of the penalized spline are similar to those of the regression spline with less knots than data points. Again, when data are generated from M2 and M3 with a nonlinear link function, both FLM and

Table 2 The ASEs for the proposed FSiM, the FLM method and the PFR method in the "refund" R package with data generated from models M1, M2 and M3

Model	$n = 100$			$n = 200$			$n = 500$		
	FSiM	FLM	PFR	FSiM	FLM	PFR	FSiM	FLM	PFR
M1	0.0732	0.0613	0.0641	0.0376	0.0302	0.0346	0.0143	0.0116	0.0125
M2	0.1911	0.2754	0.2873	0.0860	0.1590	0.1684	0.0279	0.0841	0.0846
M3	0.0507	1.2257	1.3094	0.0181	0.9508	0.9747	0.0062	0.6882	0.6780

Table 3 The empirical coverage rate that the true curve $\beta_k(\cdot)$ is covered by the simultaneous confidence bands (10) and the average bandwidth with data generated from model M3

n	Coverage rate		Bandwidths	
	β_1	β_2	β_1	β_2
100	0.926	0.934	0.963	0.942
200	0.944	0.954	0.762	0.745
500	0.952	0.950	0.510	0.504

PFR have larger ASE values than the FSiM. The difference is more obvious for M3 which has a nonlinear and non-monotone link function. The results further illustrates that the proposed FSiM is useful when there exists an unknown and nonlinear link between the response and the functional predictors.

To evaluate the asymptotic standard deviation and confidence bandwidths established in Theorem 3, we construct 95 % asymptotic simultaneous confidence bands for coefficient functions $\beta_k(t)$. Table 3 shows the empirical coverage rate and the average bandwidth among 500 replications with data generated from model M3. We observe that empirical coverage rates are close to the nominal confidence level for all cases and the average bandwidth decreases as the sample size increases.

5 Application

In this section, we apply our proposed method to the diffusion tensor imaging (DTI) study (Goldsmith et al. 2011, 2012). The data can be found from the "refund" package in R software. In this data application, DTI brain scans are recorded for many multiple-sclerosis (MS) patients at several visits with the goal of assessing the effect of neurodegeneration on disability. The scalar outcome is the paced auditory serial addition test (PASAT) score, which is commonly used examination of cognitive function affected by MS with scores ranging between 0 and 60, and two functional predictors are the mean diffusivity profile of the corpus callosum tract (CCA) and the parallel diffusivity profile of the right corticospinal tract (RCST). We refer to Goldsmith et al. (2011) and Goldsmith et al. (2012) for the detailed descriptions of this data application. The dataset has 142 patients with several visits. After deleting missing data, we use the remaining 229 observations in our analysis. Then, a functional single-index model is fitted as

$$Y_i = g\left(\int_{\mathcal{T}} \beta_1(t) X_{i,1}(t)\mathrm{d}t + \int_{\mathcal{T}} \beta_2(t) X_{i,2}(t)\mathrm{d}t\right) + \varepsilon_i, \quad i = 1, \ldots, 229, \qquad (12)$$

where Y_i is the PASAT score, and $X_{i,1}(\cdot)$ and $X_{i,2}(\cdot)$ are the two functional predictors, CCA and RCST, respectively. Measurements of CCA and RCST are taken at 93 and 54 tract locations, respectively. For illustration, Fig. 1 shows 100 trajectories for the two functional predictors CCA and RCST.

For data exploration, we first fit a functional linear model (FLM) for each predictor such that $Y_i = a_k \int_0^1 \beta_k(t) X_{i,k}(t)\mathrm{d}t + \varepsilon_i$, $k = 1, 2$ with $\int_{\mathcal{T}} \beta_k^2(t)\mathrm{d}t = 1$. Figure 2 shows the plots of the estimated coefficient functions $\widehat{\beta_1}(t)$ and $\widehat{\beta_2}(t)$ by fitting the FLM for the two predictors, CCA and RCST, respectively. We observe that for CCA, $\widehat{\beta_1}(t)$ shows an increasing trend followed by a sudden drop in the end, and for RCST,

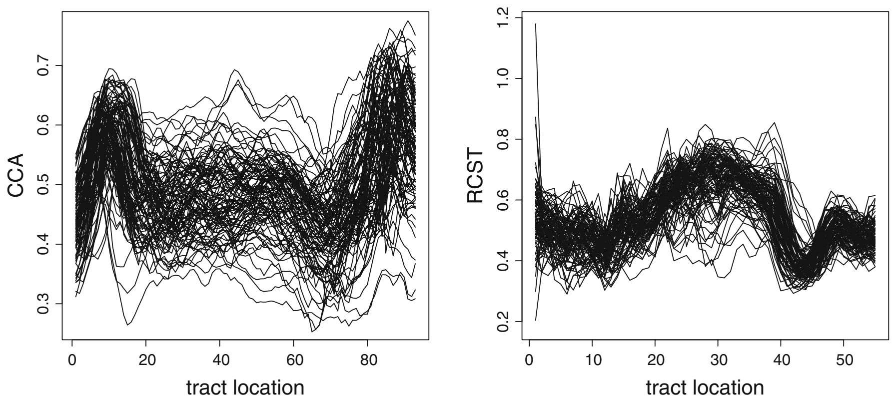

Fig. 1 Plots of the trajectories for the two functional predictors CCA and RCST

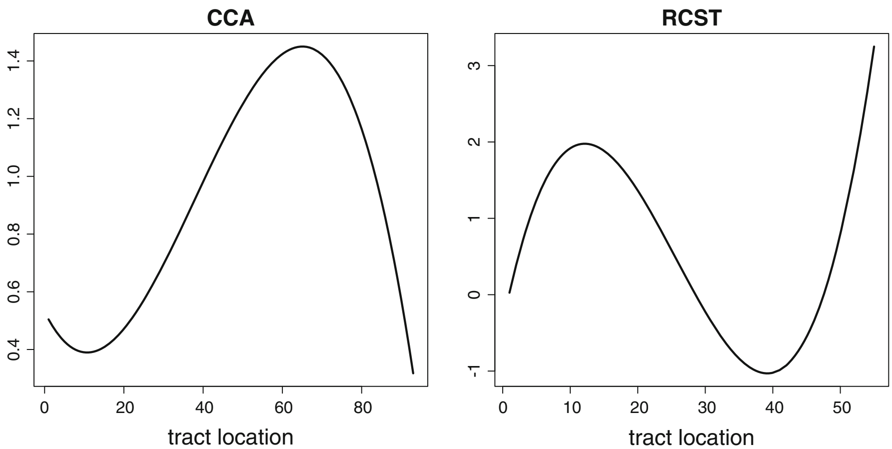

Fig. 2 Plots of the estimated functions $\widehat{\beta_1}(t)$ and $\widehat{\beta_2}(t)$ by fitting the FLM for the two predictors, CCA and RCST, respectively

$\widehat{\beta}_2(t)$ shows a wave pattern. The two plots corroborate with Figure 4 in Goldsmith et al. (2012).

We next apply the proposed estimating procedure in Sect. 3. Same as the simulation studies in Sect. 4, we use cubic splines with order $q = 4$ to approximate the nonparametric functions, and use the BIC criterion to select the number of knots. As a result, one interior knot is selected for both of the spline bases. A similar strategy can be found in Huang et al. (2004), Xue and Yang (2006) and Xue et al. (2010). As indicated in Xue et al. (2010), only a small number of knots are often needed. We then obtain the estimated coefficient functions $\widehat{\beta}_1(t)$ and $\widehat{\beta}_2(t)$ and the estimated link function $\widehat{g}(u)$ in model (12) by assuming that $\beta_1(t)$ is a monotone function. Figure 3 shows the estimated coefficient functions $\widehat{\beta}_1(t)$ and $\widehat{\beta}_2(t)$ and their 95 % simultaneous confidence bands, and the estimated link function $\widehat{g}(u)$. The last plot in Fig. 3 indicates that $\widehat{g}(u)$ is a monotone nonlinear function. Thus, the functional linear model (FLM) by assuming linearity of $g(u)$ for this data application may encounter the problem of misspecification. Based on this result, we refit model (12) by assuming that the link function $g(\cdot)$ is monotone increasing. Figure 4 shows the estimated coefficient functions $\widehat{\beta}_1(t)$ and $\widehat{\beta}_2(t)$ and their 95 % simultaneous confidence bands, and the estimated

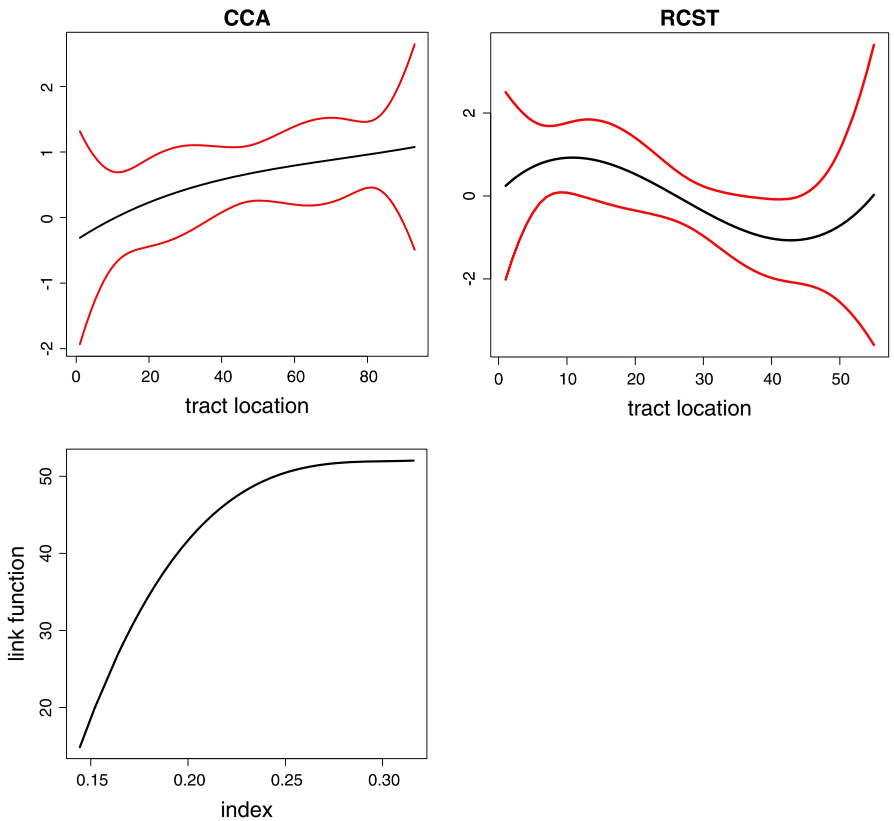

Fig. 3 Plots of the estimated coefficient functions $\widehat{\beta}_1(t)$ and $\widehat{\beta}_2(t)$ and their 95 % simultaneous confidence bands, and the estimated link function $\widehat{g}(u)$ by assuming monotonicity of $\beta_1(t)$

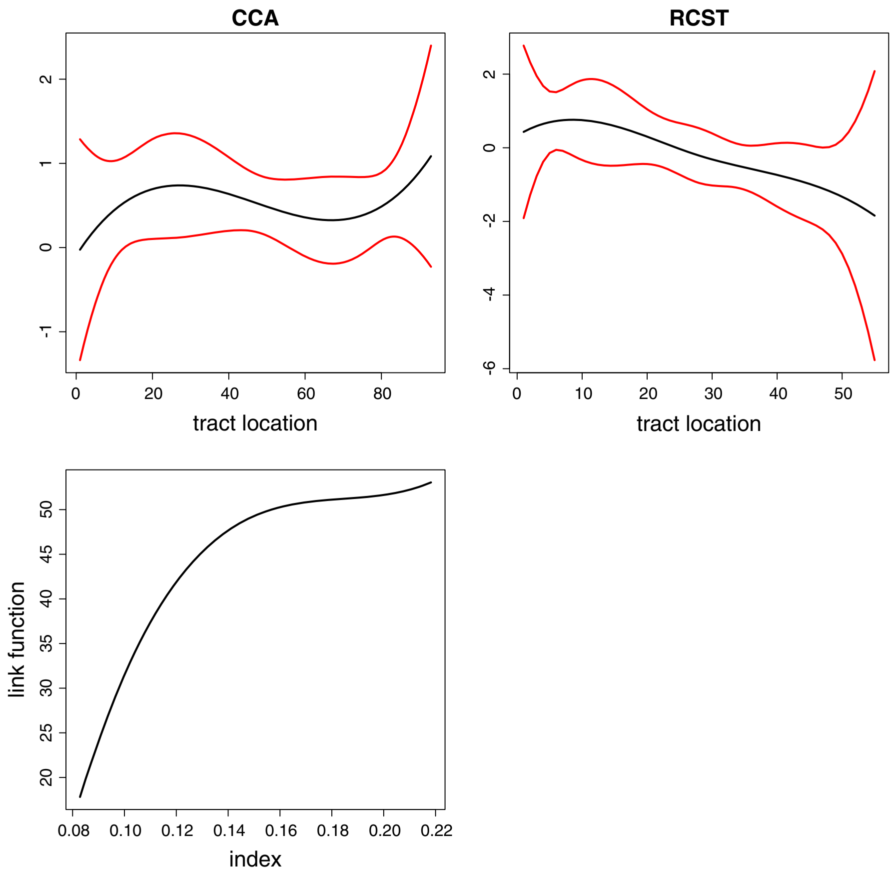

Fig. 4 Plots of the estimated coefficient functions $\widehat{\beta}_1(t)$ and $\widehat{\beta}_2(t)$ and their 95 % simultaneous confidence bands, and the estimated link function $\widehat{g}(u)$ by assuming monotonicity of $g(u)$

link function $\widehat{g}(u)$ by assuming monotonicity of $g(u)$. We can observe a decreasing pattern for $\widehat{\beta}_2(t)$. For model comparison, we obtain the coefficient of determination $R^2 = 0.24, 0.26, 0.19$ for the FSiM (12) by assuming monotonicity of $\beta_1(t)$ and $g(u)$, respectively, and the FLM by assuming that $g(u)$ is linear. Apparently, the FSiM leads to a better fit than the FLM, and the FSiM improves the model fitting by 26.3 and 36.8 %, respectively, for the two methods, comparing to the FLM.

6 Discussion

In this paper, we propose a FSiM to study the link between a scalar response variable and multiple functional predictors. We then use B-spline basis functions to estimate the slope functions and the link function based on the least-squares criterion, and propose an iterative estimating procedure. We select the number of the knots for the B-spline basis by the BIC criterion. Moreover, we provide uniform convergence rates of the proposed spline estimators in the FSiM, and construct asymptotic simultaneous confidence bands for the slope functions for inference. The proposed method is demonstrated to provide a flexible tool to explore possible nonlinear relationships between the response and the predictors. As pointed out by the referees, the penalized splines can also be used to estimate the nonparametric functions with a penalty term in the objective function to control the roughness of the fit. We will consider the penalized spline method as a future work. The asymptotic properties of the resulting estimators, however, still need us to further explore according to the techniques in Cardot et al. (2003), Claeskens et al. (2009) and this paper. Moreover, we will extend the proposed method to the case with a functional response. We will also consider variable selection by penalization when the number of predictors is large.

Appendix

For any positive numbers a_n and b_n, let $a_n \sim b_n$ denote that $\lim_{n\to\infty} a_n/b_n = 1$. For any vector $\zeta = (\zeta_1, \ldots, \zeta_s)^{\mathrm{T}} \in R^s$, denote its L_r norm as $\|\zeta\|_r = (|\zeta_1|^r + \cdots |\zeta_s|^r)^{1/r}$. For any symmetric matrix $\mathbf{A}_{s\times s}$, denote its L_r norm as $\|\mathbf{A}\|_r = \max_{\zeta\in s, \zeta\neq 0} \|\mathbf{A}\zeta\|_r \|\zeta\|_r^{-1}$. For any matrix $\mathbf{A} = (A_{ij})_{i=1,j=1}^{s,t}$, denote $\|\mathbf{A}\|_\infty = \max_{1\le i\le s} \sum_{j=1}^{t} |A_{ij}|$. The estimator $\widehat{g}(u;\delta_n)$ can be rewritten as $\widehat{\widehat{g}}(u;\delta_n) = \mathbf{B}_1^{q-1}(u)^{\mathrm{T}} \mathbf{D}_1 \widehat{\lambda}(\delta_n)$, where

$$\mathbf{B}_1^{q-1}(u) = (B_{r,1}^{q-1}(u) : 2 \le r \le J_{n,1})^{\mathrm{T}}$$

are B-spline functions with order $q-1$, and

$$\mathbf{D}_1 = (q-1)\begin{pmatrix} \frac{-1}{\tau_{q+1}-\tau_2} & \frac{1}{\tau_{q+1}-\tau_2} & 0 & \cdots & 0 \\ 0 & \frac{-1}{\tau_{q+2}-\tau_3} & \frac{1}{\tau_{q+2}-\tau_3} & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{-1}{\tau_{J_{n,1}+q-1}-\tau_{J_{n,1}}} & \frac{1}{\tau_{J_{n,1}+q-1}-\tau_{J_{n,1}}} \end{pmatrix}_{(J_{n,1}-1)\times J_{n,1}}. \tag{13}$$

Proof of Proposition 1 By (6), $\widehat{\lambda}(\delta_n)$ can be decomposed as $\widehat{\lambda}(\delta_n) = \widehat{\lambda}_\varepsilon(\delta_n) + \widehat{\lambda}_g(\delta_n)$, where

$$\widehat{\lambda}_\varepsilon(\delta_n) = \{\mathcal{B}(\delta_n)^{\mathrm{T}}\mathcal{B}(\delta_n)\}^{-1}\mathcal{B}(\delta_n)^{\mathrm{T}}\varepsilon_n,$$
$$\widehat{\lambda}_g(\delta_n) = \{\mathcal{B}(\delta_n)^{\mathrm{T}}\mathcal{B}(\delta_n)\}^{-1}\mathcal{B}(\delta_n)^{\mathrm{T}}\mathbf{g}_n,$$

in which $\varepsilon_n = (\varepsilon_1, \ldots, \varepsilon_n)^{\mathrm{T}}$ and $\mathbf{g}_n = \{g(\int_{\mathcal{T}} \beta(t)^{\mathrm{T}}\mathbf{X}_i(t)\mathrm{d}t), 1 \le i \le n\}^{\mathrm{T}}$. Correspondingly, $\widehat{g}(u;\delta_n)$ is decomposed into $\widehat{g}(u;\delta_n) = \widehat{g}_\varepsilon(u;\delta_n) + \widehat{g}_g(u;\delta_n)$, where

$\widehat{g}_\varepsilon(u;\delta_n) = \mathbf{B}_1(u)^{\mathrm{T}}\widehat{\lambda}_\varepsilon(\delta_n)$ and $\widehat{g}_g(u;\delta_n) = \mathbf{B}_1(u)^{\mathrm{T}}\widehat{\lambda}_g(\delta_n)$. Thus,

$$\begin{aligned}\widehat{g}_g(u;\delta_n) - g_n(u) &= \mathbf{B}_1(u)^{\mathrm{T}}(\widehat{\lambda}_g(\delta_n) - \lambda_n)\\ &= \mathbf{B}_1(u)^{\mathrm{T}}\{\mathcal{B}(\delta_n)^{\mathrm{T}}\mathcal{B}(\delta_n)\}^{-1}\mathcal{B}(\delta_n)^{\mathrm{T}}\{\mathbf{g}_n - \mathcal{B}(\delta_n)\lambda_n\}\\ &= \Psi_1(u) + \Psi_2(u) + \Psi_3(u),\end{aligned}$$

where

$$\begin{aligned}\Psi_1(u) &= \mathbf{B}_1(u)^{\mathrm{T}}\{\mathcal{B}(\delta_n)^{\mathrm{T}}\mathcal{B}(\delta_n)\}^{-1}\mathcal{B}(\delta_n)^{\mathrm{T}}\Bigg[\Bigg\{g\left(\int_{\mathcal{T}}\beta(t)^{\mathrm{T}}\mathbf{X}_i(t)\mathrm{d}t\right)\\ &\quad - g\left(\sum_{j=1}^{m_i}(t_{i,j+1}-t_{ij})\beta(t_{ij})^{\mathrm{T}}\mathbf{X}_i(t_{ij})\right), 1\le i\le n\Bigg\}^{\mathrm{T}}\Bigg],\\ \Psi_2(u) &= \mathbf{B}_1(u)^{\mathrm{T}}\{\mathcal{B}(\delta_n)^{\mathrm{T}}\mathcal{B}(\delta_n)\}^{-1}\mathcal{B}(\delta_n)^{\mathrm{T}}\\ &\quad\times\left\{g\left(\sum_{j=1}^{m_i}(t_{i,j+1}-t_{ij})\beta(t_{ij})^{\mathrm{T}}\mathbf{X}_i(t_{ij})\right) - g(\Phi_i^{\mathrm{T}}\delta_n), 1\le i\le n\right\}^{\mathrm{T}},\\ \Psi_3(u) &= \mathbf{B}_1(u)^{\mathrm{T}}\{\mathcal{B}(\delta_n)^{\mathrm{T}}\mathcal{B}(\delta_n)\}^{-1}\mathcal{B}(\delta_n)^{\mathrm{T}}\\ &\quad\times[\{g(\Phi_i^{\mathrm{T}}\delta_n), 1\le i\le n\}^{\mathrm{T}} - \mathcal{B}(\delta_n)\lambda_n].\end{aligned} \tag{14}$$

By Conditions (C3) and (C4), we have that for all $1\le i\le n$, there exists a constant $0 < C < \infty$ such that

$$\left|g\left(\int_{\mathcal{T}}\beta(t)^{\mathrm{T}}\mathbf{X}_i(t)\mathrm{d}t\right) - g\left(\sum_{j=1}^{m_i}(t_{i,j+1}-t_{ij})\beta(t_{ij})^{\mathrm{T}}\mathbf{X}_i(t_{ij})\right)\right| \le C m_{\min}^{-1}, \tag{15}$$

and by (8) there exist constants $0 < C', C'' < \infty$ such that

$$\begin{aligned}&\left|g\left(\sum_{j=1}^{m_i}(t_{i,j+1}-t_{ij})\beta(t_{ij})^{\mathrm{T}}\mathbf{X}_i(t_{ij})\right) - g(\Phi_i^{\mathrm{T}}\delta_n)\right|\\ &\quad\le C'\left|\sum_{j=1}^{m_i}(t_{i,j+1}-t_{ij})\sum_{k=1}^{p}\{\beta_k(t_{ij}) - \mathbf{B}_2(t_{ij})^{\mathrm{T}}\widetilde{\delta}_{k,n}\}X_{ik}(t_{ij})\right|\\ &\quad\le C''(a_n + J_{n,2}^{-\alpha}).\end{aligned} \tag{16}$$

Moreover, by (8) for all $1\le i\le n$, there exists a constant $0 < C''' < \infty$ such that

$$|g(\Phi_i^{\mathrm{T}}\delta_n) - \mathbf{B}_1(\Phi_i^{\mathrm{T}}\delta_n)\lambda_n| \le C''' J_{n,1}^{-\alpha}. \tag{17}$$

By Theorem 5.4.2 of DeVore and Lorentz (1993) and Berstein's inequality in Boor (2001), one has for large enough n, there are constants $0 < c_1 < C_1 < \infty$, such that

$$c_1 J_{n,1}^{-1} \le \|E\{n^{-1}\mathcal{B}(\delta_n)^{\mathrm{T}}\mathcal{B}(\delta_n)\}\|_2 \le C_1 J_{n,1}^{-1},$$

with probability approaching 1, for $J_{n,1}\log(n)/n = o(1)$,

$$c_1 J_{n,1}^{-1} \le \|\{n^{-1}\mathcal{B}(\delta_n)^{\mathrm{T}}\mathcal{B}(\delta_n)\}\|_2 \le C_1 J_{n,1}^{-1},$$

and thus

$$C_1^{-1} J_{n,1} \le \|\{n^{-1}\mathcal{B}(\delta_n)^{\mathrm{T}}\mathcal{B}(\delta_n)\}^{-1}\|_2 \le c_1^{-1} J_{n,1}. \tag{18}$$

By the above result and Demko (1986), it can be proved that with probability approaching 1 and for large enough n,

$$\|\{n^{-1}\mathcal{B}(\delta_n)^{\mathrm{T}}\mathcal{B}(\delta_n)\}^{-1}\|_\infty \le C_2 J_{n,1}, \tag{19}$$

for some constant $0 < C_2 < \infty$. Therefore, by (14), (15), (16), (17) and (19), we have

$$\begin{aligned}
&\sup_{u\in\mathcal{I}} |\Psi_1(u)| \\
&\le \sup_{u\in\mathcal{I}} \|\mathbf{B}_1(u)\|_\infty \|\{n^{-1}\mathcal{B}(\delta_n)^{\mathrm{T}}\mathcal{B}(\delta_n)\}^{-1}\|_\infty \|n^{-1}\mathcal{B}(\delta_n)^{\mathrm{T}}\mathbf{1}_n\|_\infty O(m_{\min}^{-1}) \\
&= O_p(J_{n,1})O_p(J_{n,1}^{-1})O(m_{\min}^{-1}) = O_p(m_{\min}^{-1}),\ \sup_{u\in\mathcal{I}} |\Psi_2(u)| \\
&\le \sup_{u\in\mathcal{I}} \|\mathbf{B}_1(u)\|_\infty \|\{n^{-1}\mathcal{B}(\delta_n)^{\mathrm{T}}\mathcal{B}(\delta_n)\}^{-1}\|_\infty \|n^{-1}\mathcal{B}(\delta_n)^{\mathrm{T}}\mathbf{1}_n\|_\infty O(a_n + J_{n,2}^{-\alpha}) \\
&= O_p(J_{n,1})O_p(J_{n,1}^{-1})O(a_n + J_{n,2}^{-\alpha}) = O_p(a_n + J_{n,2}^{-\alpha}) \\
&\quad \sup_{u\in\mathcal{I}} |\Psi_3(u)| \\
&\le \sup_{u\in\mathcal{I}} \|\mathbf{B}_1(u)\|_\infty \|\{n^{-1}\mathcal{B}(\delta_n)^{\mathrm{T}}\mathcal{B}(\delta_n)\}^{-1}\|_\infty \|n^{-1}\mathcal{B}(\delta_n)^{\mathrm{T}}\mathbf{1}_n\|_\infty O(J_{n,1}^{-\alpha}) \\
&= O_p(J_{n,1})O_p(J_{n,1}^{-1})O(J_{n,1}^{-\alpha}) = O_p(J_{n,1}^{-\alpha}).
\end{aligned}$$

Thus, we have

$$\sup_{u\in\mathcal{I}} |\widehat{g}_g(u;\delta_n) - g_n(u)| = O_p(a_n + J_{n,1}^{-\alpha} + J_{n,2}^{-\alpha} + m_{\min}^{-1}).$$

Let $\mathbb{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$. By Condition (C5) and (18) for every $u \in \mathcal{I}$, $E\{\widehat{g}_\varepsilon(u;\delta_n)|\mathbb{X}\} = 0$, and

$$E\{\widehat{g}_\varepsilon(u;\delta_n)|\mathbb{X}\}^2 \asymp \mathbf{B}_1(u)^{\mathrm{T}}\{\mathcal{B}(\delta_n)^{\mathrm{T}}\mathcal{B}(\delta_n)\}^{-1}\mathbf{B}_1(u) \asymp J_{n,1}n^{-1}.$$

Thus, it can be proved by Berstein's inequality in Boor (2001) that $\sup_{u\in\mathcal{I}} |\widehat{g}_\varepsilon(u;\delta_n)| = O_p(\sqrt{\log(n)J_{n,1}n^{-1}})$. Therefore, we have

$$\sup_{u\in\mathcal{I}} |\widehat{g}(u;\delta_n) - g_n(u)| = O_p\left(a_n + J_{n,1}^{-\alpha} + J_{n,2}^{-\alpha} + m_{\min}^{-1} + \sqrt{\log(n)J_{n,1}n^{-1}}\right).$$

Result (i) is proved by the above result and (8). It is easy to prove that $\|\mathbf{D}_1\|_\infty = O(J_{n,1})$, where $\mathbf{D}_1$ is defined in (13). Following the similar reasoning as the proof for $\widehat{g}(u;\delta_n)$, the result in (ii) can be proved. □

Lemma 1 *Under Condition* (C3)*, we have that there exists* $\widetilde{\delta}_{1,n}^0 = (\widetilde{\delta}_{r1,n}^0 : 1 \le r \le J_{n,2})^{\mathrm{T}} \in R^{J_{n,2}}$ *with* $\widetilde{\delta}_{11,n}^0 \le \cdots \le \widetilde{\delta}_{J_{n,2}1,n}^0$ *such that* $\sup_{t\in\mathcal{T}} |\beta_1(t) - \widetilde{\beta}_{1,n}(t)| = O(J_{n,2}^{-\alpha})$*, where* $\widetilde{\beta}_{1,n}(t) = \mathbf{B}_2(t)^{\mathrm{T}}\widetilde{\delta}_{1,n}^0$.

Proof By choosing $\epsilon_1 < \cdots < \epsilon_{J_{n,2}}$, we define

$$\widetilde{\beta}^0_{1,n}(t) = \sum_{r=1}^{J_{n,2}} \beta_1(\epsilon_r) B_{r,2}(t),$$

which is monotone nondecreasing function in t. By the fact that for t in $[\upsilon_{r_1}, \upsilon_{r_1+1})$, $\sum_{r=\upsilon_{r_1}+1-q}^{\upsilon_{r_1}} B_{r,2}(t) = 1$, we have

$$\beta_1(\widetilde{t}) = \sum_{r=\upsilon_{r_1}+1-q}^{\upsilon_{r_1}} \beta_1(\widetilde{t}) B_{r,2}(t)$$

for $\widetilde{t} \in [\upsilon_{r_1}, \upsilon_{r_1+1})$, and thus

$$\begin{aligned} |\beta_1(\widetilde{t}) - \widetilde{\beta}^0_{1,n}(\widetilde{t})| &\le \sum_{r=\upsilon_{r_1}+1-q}^{\upsilon_{r_1}} |\beta_1(\widetilde{t}) - \beta_1(\epsilon_r)| B_{r,2}(t) \\ &\le q \max_{\upsilon_{r_1}+1-q \le r \le \upsilon_{r_1}} |\beta_1(\widetilde{t}) - \beta_1(\epsilon_r)|. \end{aligned}$$

Let $h = \max_{q \le l \le J_{n,2}} (\upsilon_{r+1} - \upsilon_r)$. Define

$$\omega(\beta_1; h) = \max\{|\beta_1(t_1) - \beta_1(t_2)| : |t_1 - t_2| \le h\}.$$

Then, $\omega(\beta_1; h)$ is a monotone and subadditivity function of h , that is, $\omega(\beta_1; h) \le \omega(\beta_1; h_1 + h_2) \le \omega(\beta_1; h_1) + \omega(\beta_1; h_2)$ for $h_1 > 0$ and $h_2 > 0$. See Lemma 2.19 of Wu (2010) for the detailed proof. We choose $\epsilon_r = \upsilon_1 + (r-1)(\upsilon_{q+1} - \upsilon_q)/q$ for $r = 1, \ldots, q$ and $\epsilon_r = \upsilon_r$ for $r = q+1, \ldots, J_{n,2}$ to guarantee that $\epsilon_{r+1} - \epsilon_r > 0$. Then, we have $|\epsilon_r - \upsilon_r| \le h$ for $r = 1, \ldots, J_{n,2}$. Moreover, for $\widetilde{t} \in [\upsilon_{r_1}, \upsilon_{r_1+1})$ and $r_1 - q \le r \le r_1$, $|\widetilde{t} - \epsilon_r| \le (q+1)h$. Therefore, we have

$$\sup_t |\beta_1(t) - \widetilde{\beta}^0_{1,n}(t)| \le q\omega(\beta_1; (q+1)h) \le (q+1)q\omega(\beta_1; h).$$

The last step follows from the subadditivity of $\omega(\beta_1; h)$. Let

$$G_q = \{\mathbf{B}_2(t)^{\mathrm{T}}\delta_1, \delta_1 \in \mathbf{R}^{J_{n,2}}, \delta_{11} \le \cdots \le \delta_{J_{n,2}1}\}.$$

Denote $d(\beta_1, G_q)$ as the distance of β_1 from G_q. Following the reasoning as given in Lemma 2.19 of Wu (2010), it can be shown that for any $g \in G_q$, we have

$$d(\beta_1, G_q) \le ch\|\partial(\beta_1 - g)/\partial t\|_\infty,$$

for some constant $0 < c < \infty$, and thus

$$d(\beta_1, G_q) \le chd(\partial\beta_1/\partial t, G_{q-1}),$$

where $G_{q-1,q} = \{\partial g/\partial t, g \in G_q\}$. Proceeding in this way, we can derive

$$d(\beta_1, G_q) \leq ch^{\alpha} \|\partial^{\alpha} g/\partial t^{\alpha}\|_{\infty}.$$

Thus, the result in Lemma 1 follows from the above result and Condition (C3). □

Lemma 2 *Let $\widehat{\delta}$ be the minimizer of $\widetilde{L}_n(\delta)$ given in (7) subject to $\delta_{1,1} \leq \cdots \leq \delta_{J_{n,2},1}$ satisfying $\|\widehat{\delta} - \delta_n^0\|_{\infty} \leq a_n$ with probability approaching 1, where $\widetilde{\delta}_n^0 = (\widetilde{\delta}_{1,n}^{0\mathrm{T}}, \ldots, \delta_{p,n}^{0\mathrm{T}})^{\mathrm{T}}$, under the assumptions in Theorem 1, we have*

$$\|\widehat{\delta}_n - \widetilde{\delta}_n^0\|_{\infty} = O_p\{(\log n)^{1/2} J_{n,2}^{1/2} n^{-1/2}\}. \tag{20}$$

Proof Let $\widehat{\delta}_n$ be the minimizer of $\widetilde{L}_n(\delta)$ and $\|\widehat{\delta}_n - \delta_n^0\|_{\infty} \leq a_n$. By Taylor's expansion, we have

$$\widehat{\delta}_n - \delta_n^0 = \{L_n(\widetilde{\delta}_n^0)/\partial\delta\partial\delta^{\mathrm{T}}\}^{-1}\{-L_n(\widetilde{\delta}_n^0)/\partial\delta\}\{1 + o_p(1)\}.$$

Moreover,

$$\begin{aligned}
&-\widetilde{L}_n(\widetilde{\delta}_n^0)/\partial\delta \\
&= \sum_{i=1}^{n}\{Y_i - \widehat{g}(\Phi_i^{\mathrm{T}}\widetilde{\delta}_n^0, \widehat{\delta})\}[\widehat{\dot{g}}(\Phi_i^{\mathrm{T}}\widetilde{\delta}_n^0, \widehat{\delta})\Phi_i + \{\widehat{\lambda}(\widehat{\delta})^{\mathrm{T}}/\partial\delta\}\mathbf{B}_1(\Phi_i^{\mathrm{T}}\delta_n^0)] \\
&= \Theta_1 + \Theta_2,
\end{aligned}$$

where

$$\begin{aligned}
\Theta_1 &= \sum_{i=1}^{n}\{Y_i - g(U_i)\}\Psi_i, \\
\Theta_2 &= \sum_{i=1}^{n}\{g(U_i) - \widehat{g}(\Phi_i^{\mathrm{T}}\widetilde{\delta}_n^0, \widehat{\delta})\}\Psi_i,
\end{aligned}$$

and

$$\Psi_i = \widehat{\dot{g}}(\Phi_i^{\mathrm{T}}\widetilde{\delta}_n^0, \widehat{\delta})\Phi_i + \{\widehat{\lambda}(\widehat{\delta})^{\mathrm{T}}/\partial\delta\}\mathbf{B}_1(\Phi_i^{\mathrm{T}}\delta_n^0).$$

By Berstein's inequality Boor (2001), it can be proved that $\|\Theta_1\|_{\infty} = O_p((\log n)^{1/2} n^{1/2} J_{n,2}^{-1/2})$. Next, we will show that $\|\Theta_2\|_{\infty} = o_p(n^{1/2} J_{n,2}^{-1/2})$. By Proposition 1 and the assumption in Theorem 1, we have

$$|g(U_i) - \widehat{g}(\Phi_i^{\mathrm{T}}\widetilde{\delta}_n^0, \widehat{\delta})| = O_p\left(a_n + J_{n,1}^{-\alpha} + m_{\min}^{-1} + \sqrt{\log(n) J_{n,1} n^{-1}}\right) = o_p(1).$$

By the law of large numbers, we have $\sum_{i=1}^{n} \|\Psi_i\|_{\infty} = O_p(n^{1/2} J_{n,2}^{-1/2})$. Therefore, $\|\Theta_2\|_{\infty} = o_p(n^{1/2} J_{n,2}^{-1/2})$. Thus, we have $\| - L_n(\widetilde{\delta}_n^0)/\partial\delta\|_{\infty} = O_p((\log n)^{1/2} n^{1/2} J_{n,2}^{-1/2})$. Moreover,

$$L_n(\delta_n^0)/\partial\delta\partial\delta^{\mathrm{T}} = \left(\sum_{i=1}^{n} \Psi_i \Psi_i^{\mathrm{T}}\right)(1 + o_p(1)) \asymp n J_{n,2}^{-1}.$$

Therefore, we have $\|\widehat{\delta}_n - \widetilde{\delta}_n^0\|_\infty = O_p((\log n)^{1/2} n^{-1/2} J_{n,2}^{1/2})$. Since $\widetilde{\delta}_{11,n}^0 \leq \cdots \leq \widetilde{\delta}_{J_{n,2}1,n}^0$, then with probability approaching 1, $\widehat{\delta}_n = \widehat{\delta}$. □

Lemma 3 *Under the assumptions in Theorem 1,*

$$\left\| -\widetilde{L}_n(\widetilde{\delta}_n^0)/\partial\delta - 2\sum_{i=1}^{n} \{Y_i - g(U_i)\}\dot{g}(U_i)\{\Phi_i - E(\Phi_i|U_i)\} \right\|_\infty = o_p(n^{1/2} J_{n,2}^{-1/2}).$$

Proof By (6), we have $\widehat{\lambda}(\delta) = \{\mathcal{B}(\delta)^{\mathrm{T}}\mathcal{B}(\delta)\}^{-1}\mathcal{B}(\delta)^{\mathrm{T}}\mathbf{Y}_n$, where $\mathbf{Y}_n = (Y_1, \ldots, Y_n)^{\mathrm{T}}$. Thus,

$$\begin{aligned}
&\mathbf{B}_1(\Phi_i^{\mathrm{T}}\delta)^{\mathrm{T}}\{\partial\widehat{\lambda}(\delta)/\partial\delta^{\mathrm{T}}\} \\
&\quad = \mathbf{B}_1(\Phi_i^{\mathrm{T}}\delta)^{\mathrm{T}}\{\partial(\widehat{\lambda}(\delta) - \lambda_n^0)/\partial\delta^{\mathrm{T}}\} \\
&\quad = \mathbf{B}_1(\Phi_i^{\mathrm{T}}\delta)^{\mathrm{T}}\partial[\{\mathcal{B}(\delta)^{\mathrm{T}}\mathcal{B}(\delta)\}^{-1}\mathcal{B}(\delta)^{\mathrm{T}}(\mathbf{Y}_n - \mathcal{B}(\delta)\lambda_n^0)]/\partial\delta^{\mathrm{T}} \\
&\quad = \Omega_1(\delta) + \Omega_2(\delta),
\end{aligned} \tag{21}$$

where

$$\begin{aligned}
\Omega_1(\delta) &= -\mathbf{B}_1(\Phi_i^{\mathrm{T}}\delta)^{\mathrm{T}}\{\mathcal{B}(\delta)^{\mathrm{T}}\mathcal{B}(\delta)\}^{-1}\mathcal{B}(\delta)^{\mathrm{T}}\{\dot{g}_n(\Phi_i^{\mathrm{T}}\delta)\Phi_i, 1 \leq i \leq n\}^{\mathrm{T}}, \\
\Omega_2(\delta) &= \mathbf{B}_1(\Phi_i^{\mathrm{T}}\delta)^{\mathrm{T}}[\partial[\{\mathcal{B}(\delta)^{\mathrm{T}}\mathcal{B}(\delta)\}^{-1}\mathcal{B}(\delta)^{\mathrm{T}}]/\partial\delta^{\mathrm{T}}]\{\mathbf{Y}_n - \mathcal{B}(\delta)\lambda_n^0\}.
\end{aligned}$$

Let

$$\widehat{\Omega}_1(\delta) = -\mathbf{B}_1(\Phi_i^{\mathrm{T}}\delta)^{\mathrm{T}}\{\mathcal{B}(\delta)^{\mathrm{T}}\mathcal{B}(\delta)\}^{-1}\mathcal{B}(\delta)^{\mathrm{T}}\{\dot{g}(U_i)\Phi_i, 1 \leq i \leq n\}^{\mathrm{T}}.$$

Following similar reasoning as the proofs in Proposition 1, it can be shown that

$$\begin{aligned}
\|\Omega_1(\widetilde{\delta}_n^0)^{\mathrm{T}} - \widehat{\Omega}_1(\widetilde{\delta}_n^0)^{\mathrm{T}}\|_\infty &= O_p(J_{n,1}^{1-\alpha} + J_{n,2}^{-\alpha} + m_{\min}^{-1}), \\
\|\mathbf{Y}_n - \mathcal{B}(\delta)\lambda_n^0\|_\infty &= O_p(J_{n,1}^{-\alpha} + J_{n,2}^{-\alpha} + m_{\min}^{-1} + (\log n)^{1/2} J_{n,1}^{1/2} n^{-1/2}).
\end{aligned} \tag{22}$$

Denote

$$\mathbf{A}(\delta) = \{A_1(\delta), \ldots, A_{J_{n,1}}(\delta)\}^{\mathrm{T}} = \{\mathcal{B}(\delta)^{\mathrm{T}}\mathcal{B}(\delta)\}^{-1}\mathcal{B}(\delta)^{\mathrm{T}}\mathbf{1}_n.$$

By (19) and Berstein's inequality in Boor (2001), we have

$$\begin{aligned}
\sup_{1 \leq s \leq J_{n,1}} |A_s(\widetilde{\delta}_n^0)| &\leq \|\{n^{-1}\mathcal{B}(\widetilde{\delta}_n^0)^{\mathrm{T}}\mathcal{B}(\widetilde{\delta}_n^0)\}^{-1}\|_\infty \|n^{-1}\mathcal{B}(\widetilde{\delta}_n^0)^{\mathrm{T}}\mathbf{1}_n\|_\infty \\
&= O_p(J_{n,1})O_p(J_{n,1}^{-1}) = O_p(1),
\end{aligned}$$

and thus with probability approaching 1, $\sup_{1 \leq s \leq J_{n,1}} |\dot{A}_s(\widetilde{\delta}_n^0)| \leq C$ for some constant $0 < C < \infty$ by the fact that $B_{s,1}(u)$ and $\dot{B}_{s,1}(u)$ are functions with values bounded between 0 and 1. Hence, with probability approaching 1,

$$\sup_{1\le s\le J_{n,1}} \|\partial A_s(\widetilde{\delta}_n^0)/\partial\delta\|_\infty \le \sup_{1\le s\le J_{n,1}} \|\dot{A}_s(\widetilde{\delta}_n^0)\|_\infty \sup_{1\le i\le n} \|\Phi_i\|_\infty \le C'$$

for some constant $0 < C' < \infty$. By B-spline properties, we have $\sum_{1\le s\le J_{n,1}} |B_{s,1}(\Phi_i^{\mathrm{T}}\widetilde{\delta}_n^0)| = O(1)$. Therefore,

$$\begin{aligned}\|\Omega_2(\widetilde{\delta}_n^0)^{\mathrm{T}}\|_\infty &\le \sum_{1\le s\le J_{n,1}} |B_{s,1}(\Phi_i^{\mathrm{T}}\widetilde{\delta}_n^0)| \sup_{1\le s\le J_{n,1}} \|\partial A_s(\widetilde{\delta}_n^0)/\partial\delta\|_\infty \|\mathbf{Y}_n - \mathcal{B}(\delta)\lambda_n^0\|_\infty \\ &= O(1)O_p(1)O_p(J_{n,1}^{-\alpha} + J_{n,2}^{-\alpha} + m_{\min}^{-1} + (\log n)^{1/2} J_{n,1}^{1/2} n^{-1/2}) \\ &= O_p(J_{n,1}^{-\alpha} + J_{n,2}^{-\alpha} + m_{\min}^{-1} + (\log n)^{1/2} J_{n,1}^{1/2} n^{-1/2}). \end{aligned} \tag{23}$$

Moreover, by Condition (C3), for every $t_{ij} \in \mathcal{T}$, there exists $\zeta_{n,k}(t_{ij}) \in R^{J_{n,1}}$ such that $|E\{X_{i,k}(t_{ij})|U_i\} - \mathbf{B}_1(U_i)^{\mathrm{T}}\zeta_{n,k}(t_{ij})| = O(J_{n,1}^{-1})$, and thus for every s and k,

$$\begin{aligned}&\left|E(\dot{g}(U_i)\Phi_{i,sk}|U_i) - \dot{g}(U_i)\sum_{j=1}^{m_i}(t_{i,j+1} - t_{ij})B_{s,2}(t_{ij})\mathbf{B}_1(U_i)^{\mathrm{T}}\zeta_{n,k}(t_{ij})\right| \\ &= \sum_{j=1}^{m_i}(t_{i,j+1} - t_{ij})B_{s,2}(t_{ij})\dot{g}(U_i)O(J_{n,1}^{-1}) \\ &= \left(\int B_{s,2}(t)\mathrm{d}t\right)\dot{g}(U_i)O(J_{n,1}^{-1} + m_{\min}^{-1}) \\ &= O\{J_{n,2}^{-1}(J_{n,1}^{-1} + m_{\min}^{-1})\}. \end{aligned} \tag{24}$$

Let

$$\begin{aligned}\widetilde{\Omega}_1 &= \{\widetilde{\Omega}_{1,sk}\} \\ &= -\mathbf{B}_1(U_i)^{\mathrm{T}}(\mathcal{B}^{\mathrm{T}}\mathcal{B})^{-1}\mathcal{B}^{\mathrm{T}}\{\dot{g}(U_i)\Phi_i, 1\le i\le n\}^{\mathrm{T}},\end{aligned}$$

where $\mathcal{B} = [\{\mathbf{B}_1(U_1), \ldots, \mathbf{B}_1(U_n)\}^{\mathrm{T}}]_{n\times J_{n,1}}$. By (24) and Berstein's inequality, we have

$$\begin{aligned}&\sup_{s,k}\left|-\widetilde{\Omega}_{1,sk} - \dot{g}(U_i)\sum_{j=1}^{m_i}(t_{i,j+1} - t_{ij})B_{s,2}(t_{ij})\mathbf{B}_1(U_i)^{\mathrm{T}}\zeta_{n,k}(t_{ij})\right| \\ &= \sup_{s,k}|\mathbf{B}_1(U_i)^{\mathrm{T}}(\mathcal{B}^{\mathrm{T}}\mathcal{B})^{-1}\mathcal{B}^{\mathrm{T}}[\dot{g}(U_i)\Phi_{i,sk} - E(\dot{g}(U_i)\Phi_{i,sk}|U_i) \\ &\quad + O(J_{n,2}^{-1}J_{n,1}^{-1} + J_{n,2}^{-1}m_{\min}^{-1}), 1\le i\le n]| \\ &= (J_{n,2}^{-1} + m_{\min}^{-1})O_p((\log n)^{1/2}J_{n,1}^{1/2}n^{-1/2}) + O(J_{n,2}^{-1}J_{n,1}^{-1} + J_{n,2}^{-1}m_{\min}^{-1}) \\ &= O_p((\log n)^{1/2}J_{n,2}^{-1}J_{n,1}^{1/2}n^{-1/2} + (\log n)^{1/2}m_{\min}^{-1}J_{n,1}^{1/2}n^{-1/2} + J_{n,2}^{-1}J_{n,1}^{-1} + J_{n,2}^{-1}m_{\min}^{-1}),\end{aligned}$$

and thus

$$\sup_{s,k} |-\widetilde{\Omega}_{1,sk} - E(\dot{g}(U_i)\Phi_{i,sk}|U_i)|$$
$$= O_p((\log n)^{1/2} J_{n,2}^{-1} J_{n,1}^{1/2} n^{-1/2} + (\log n)^{1/2} m_{\min}^{-1} J_{n,1}^{1/2} n^{-1/2} + J_{n,2}^{-1} J_{n,1}^{-1} + J_{n,2}^{-1} m_{\min}^{-1}).$$

Furthermore, it can be proved that $\|\widehat{\Omega}_1(\widetilde{\delta}_n^0)^{\mathrm{T}} - \widetilde{\Omega}_1^{\mathrm{T}}\|_\infty = O(J_{n,2}^{-\alpha} + m_{\min}^{-1})$. Therefore, we have

$$\begin{aligned}
&\|\widehat{\Omega}_1(\widetilde{\delta}_n^0)^{\mathrm{T}} + E(\dot{g}(U_i)\Phi_i|U_i)\|_\infty \\
&\quad = O_p((\log n)^{1/2} J_{n,2}^{-1} J_{n,1}^{1/2} n^{-1/2} + J_{n,2}^{-1} J_{n,1}^{-1} + m_{\min}^{-1} + J_{n,2}^{-\alpha}).
\end{aligned} \tag{25}$$

By (21), (22), (23) and (25), we have

$$\begin{aligned}
&\|\{\widehat{\lambda}(\widetilde{\delta}_n^0)^{\mathrm{T}}/\partial\delta\}\mathbf{B}_1(\Phi_i^{\mathrm{T}}\widetilde{\delta}_n^0) + E(\dot{g}(U_i)\Phi_i|U_i)\|_\infty \\
&\quad = O_p(J_{n,2}^{-1} J_{n,1}^{-1} + m_{\min}^{-1} + J_{n,2}^{-\alpha} + J_{n,1}^{1-\alpha} + (\log n)^{1/2} J_{n,1}^{1/2} n^{-1/2}).
\end{aligned}$$

Let

$$\Delta_i = [\widehat{\dot{g}}(\Phi_i^{\mathrm{T}}\widetilde{\delta}_n^0, \widetilde{\delta}_n^0)\Phi_i + \{\widehat{\lambda}(\widetilde{\delta}_n^0)^{\mathrm{T}}/\partial\delta\}\mathbf{B}_1(\Phi_i^{\mathrm{T}}\widetilde{\delta}_n^0)] - [\dot{g}(U_i)\Phi_i - E(\dot{g}(U_i)\Phi_i|U_i)].$$

By the above result and Proposition (1), we have

$$\begin{aligned}
&\|\Delta_i\|_\infty \\
&\quad = O_p\{(J_{n,1}^{1-\alpha} + J_{n,1} J_{n,2}^{-\alpha} + J_{n,1} m_{\min}^{-1} + (\log n)^{1/2} J_{n,1}^{3/2} n^{-1/2})(J_{n,2}^{-1} + m_{\min}^{-1})\} \\
&\qquad + O_p(J_{n,2}^{-1} J_{n,1}^{-1} + m_{\min}^{-1} + J_{n,2}^{-\alpha} + J_{n,1}^{1-\alpha} + (\log n)^{1/2} J_{n,1}^{1/2} n^{-1/2}) \\
&\quad = O_p(J_{n,2}^{-1} J_{n,1}^{-1} + m_{\min}^{-1} + J_{n,2}^{-\alpha} + J_{n,1}^{1-\alpha} + (\log n)^{1/2} J_{n,1}^{1/2} n^{-1/2} \\
&\qquad + J_{n,1} J_{n,2}^{-\alpha-1} + J_{n,1} J_{n,2}^{-1} m_{\min}^{-1} + (\log n)^{1/2} J_{n,1}^{3/2} J_{n,2}^{-1} n^{-1/2}).
\end{aligned} \tag{26}$$

$$\begin{aligned}
&\partial\widetilde{L}_n(\widetilde{\delta}_n^0)/\partial\delta \\
&\quad = -2\sum_{i=1}^n \{Y_i - \widehat{g}(\Phi_i^{\mathrm{T}}\delta_n^0, \delta_n^0)\} \\
&\qquad \times [\widehat{\dot{g}}(\Phi_i^{\mathrm{T}}\delta_n^0, \delta_n^0)\Phi_i + \{\widehat{\lambda}(\delta_n^0)^{\mathrm{T}}/\partial\delta_n\}\mathbf{B}_1(\Phi_i^{\mathrm{T}}\delta_n^0)](1 + o_p(1)) \\
&\quad = -2(\Theta_1 + \Theta_2 + \Theta_3 + \Theta_4 + \Theta_5)(1 + o_p(1))
\end{aligned}$$

where

$$\Theta_1 = \sum_{i=1}^n \{Y_i - g(U_i)\}\dot{g}(U_i)\{\Phi_i - E(\Phi_i|U_i)\},$$
$$\Theta_2 = \sum_{i=1}^n \{g(U_i) - \widehat{g}(\Phi_i^{\mathrm{T}}\delta_n^0, \delta_n^0)\}\dot{g}(U_i)\{\Phi_i - E(\Phi_i|U_i)\},$$

$$\Theta_3 = \sum_{i=1}^{n}\{Y_i - g(U_i)\}\Delta_i,$$
$$\Theta_4 = \sum_{i=1}^{n}\{g(U_i) - \widehat{g}(\Phi_i^{\mathrm{T}}\delta_n^0, \delta_n^0)\}\Delta_i. \tag{27}$$

In the following, we will prove that $\|\Theta_i\|_\infty = o_p(n^{1/2}J_{n,2}^{-1/2})$ for $i = 2, 3, 4$. By (8), we have
$|\widehat{g}(\Phi_i^{\mathrm{T}}\delta_n^0, \delta_n^0) - \widehat{g}(U_i)| = O(J_{n,2}^{-\alpha} + m_{\min}^{-1})$. Moreover, we have $\|\Phi_i\|_\infty = O(J_{n,2}^{-1} + m_{\min}^{-1})$. Thus,

$$\|\Theta_2 - \widetilde{\Theta}_2\|_\infty = O\{n(J_{n,2}^{-\alpha} + m_{\min}^{-1})(J_{n,2}^{-1} + m_{\min}^{-1})\},$$

where $\widetilde{\Theta}_2 = \widetilde{\Theta}_{12} + \widetilde{\Theta}_{22}$,

$$\widetilde{\Theta}_{12} = \sum_{i=1}^{n}\{g(U_i) - \widehat{g}_g(U_i)\}\dot{g}(U_i)\{\Phi_i - E(\Phi_i|U_i)\},$$
$$\widetilde{\Theta}_{22} = -\sum_{i=1}^{n}\widehat{g}_e(U_i)\dot{g}(U_i)\{\Phi_i - E(\Phi_i|U_i)\},$$

in which $\widehat{g}_g(U_i) = \mathbf{B}_1(U_i)^{\mathrm{T}}(\mathcal{B}^{\mathrm{T}}\mathcal{B})^{-1}\mathcal{B}^{\mathrm{T}}\mathbf{g}_n$ and $\widehat{g}_g(U_i) = \mathbf{B}_1(U_i)^{\mathrm{T}}(\mathcal{B}^{\mathrm{T}}\mathcal{B})^{-1}\mathcal{B}^{\mathrm{T}}\varepsilon_n$. By law of large numbers and $|g(U_i) - \widehat{g}_g(U_i)| = O_p(J_{n,1}^{-\alpha} + J_{n,1}^{1/2}/n^{1/2})$, we have $\|\widetilde{\Theta}_{12}\|_\infty = o_p(n^{1/2}J_{n,2}^{-1/2})$. Moreover,

$$\begin{aligned}\|\widetilde{\Theta}_{22}\|_\infty &\le \|\sum_{i=1}^{n}\mathbf{B}_1(U_i)\dot{g}(U_i)\{\Phi_i - E(\Phi_i|U_i)\}\|_\infty\|(\mathcal{B}^{\mathrm{T}}\mathcal{B})^{-1}\mathcal{B}^{\mathrm{T}}\varepsilon_n\|_\infty \\ &= O_p((\log n)^{1/2}J_{n,1}^{-1/2}n^{1/2})O_p((\log n)^{1/2}J_{n,1}^{1/2}n^{-1/2}) = O_p(\log n).\end{aligned}$$

Therefore, for $n^{1/2}J_{n,2}^{-\alpha-1/2} = o(1)$, $n^{1/2}m_{\min}^{-1}J_{n,2}^{-1/2} = o(1)$ and $n^{1/2}m_{\min}^{-2} = o(1)$, we have $\|\Theta_2\|_\infty = o_p(n^{1/2}J_{n,2}^{-1/2})$. Similarly, it can be proved that $\|\Theta_3\|_\infty = o_p(n^{1/2}J_{n,2}^{-1/2})$. By Proposition 1 and (26), for $n^{1/(2\alpha+1)} \ll J_{n,2} \ll n^{1/3}(\log n)^{-1}$, $n^{1/(2\alpha+3)} \ll J_{n,1} \ll J_{n,2} \ll J_{n,1}^2$, and $n^{1/2}m_{\min}^{-1}J_{n,2}^{-1/2} = o(1)$, we have

$$\begin{aligned}\|\Theta_4\|_\infty &= n \times O_p(J_{n,1}^{-\alpha} + J_{n,2}^{-\alpha} + m_{\min}^{-1} + (\log n)^{1/2}J_{n,1}^{1/2}n^{-1/2}) \\ &\quad\times O_p(J_{n,2}^{-1}J_{n,1}^{-1} + m_{\min}^{-1} + J_{n,2}^{-\alpha} + J_{n,1}^{1-\alpha} + (\log n)^{1/2}J_{n,1}^{1/2}n^{-1/2} \\ &\quad+ J_{n,1}J_{n,2}^{-\alpha-1} + J_{n,1}J_{n,2}^{-1}m_{\min}^{-1} + (\log n)^{1/2}J_{n,1}^{3/2}J_{n,2}^{-1}n^{-1/2}) \\ &= o_p(n^{1/2}J_{n,2}^{-1/2}).\end{aligned}$$

Proof of Theorem 1 By (20), we have

$$\sup_{t\in\mathcal{T}} |\widehat{\beta}_k(t) - \beta_{k,n}(t)| \asymp \|\widehat{\delta} - \widetilde{\delta}_n^0\|_\infty = O_p\{(\log n)^{1/2} J_{n,2}^{1/2} n^{-1/2}\},$$

and by (8) and $n^{1/2} J_{n,2}^{-\alpha-1/2} = o(1)$,

$$\begin{aligned}\sup_{t\in\mathcal{T}} |\widehat{\beta}_k(t) - \beta_k(t)| &\leq \sup_{t\in\mathcal{T}} |\widehat{\beta}_k(t) - \beta_{k,n}(t)| + \sup_{t\in\mathcal{T}} |\beta_{k,n}(t) - \beta_k(t)| \\ &= O_p\{(\log n)^{1/2} J_{n,2}^{1/2} n^{-1/2} + J_{n,2}^{-\alpha}\} = O_p\{(\log n)^{1/2} J_{n,2}^{1/2} n^{-1/2}\}.\end{aligned}$$

Therefore, result (i) in Theorem 1 is proved. By (27), $\partial \widetilde{L}_n(\widetilde{\delta}_n^0)/\partial\delta_n = -2(\Pi_1 + \Theta_3 + \Theta_4)(1 + o_p(1))$, where

$$\Pi_1 = \sum_{i=1}^n \{Y_i - \widehat{g}(\Phi_i^{\mathrm{T}}\delta_n^0, \delta_n^0)\}\dot{g}(U_i)\{\Phi_i - E(\Phi_i|U_i)\}.$$

By (26),

$$\partial\Pi_1/\partial\delta^{\mathrm{T}} = -\sum\nolimits_{i=1}^n \dot{g}(U_i)^2\{\Phi_i - E(\Phi_i|U_i)\}^{\otimes 2} + o_p(nJ_{n,2}^{-1}).$$

Therefore,

$$\partial\widetilde{L}_n(\widetilde{\delta}_n^0)/\partial\delta\partial\delta^{\mathrm{T}} = 2\sum_{i=1}^n \dot{g}(U_i)^2\{\Phi_i - E(\Phi_i|U_i)\}^{\otimes 2} + o_p(nJ_{n,2}^{-1}).$$

By Taylor expansion, Lemma 1, Berstein's inequality in Boor (2001) and the above result, we have

$$\begin{aligned}\widehat{\delta} - \widetilde{\delta}_n^0 &= -\{\partial L_n(\widetilde{\delta}_n^0)/\partial\delta\partial\delta^{\mathrm{T}}\}^{-1}\{\partial\widetilde{L}_n(\widetilde{\delta}_n^0)/\partial\delta\}\{1 + o_p(1)\} \\ &= \left\{\sum_{i=1}^n E(\Psi_i^{\otimes 2})\right\}^{-1}\sum_{i=1}^n \varepsilon_i\Psi_i + o_p(J_{n,2}^{1/2}n^{-1/2}).\end{aligned} \tag{28}$$

Result (ii) follows from Lindeberg–Feller Central Limit Theorem and Slutsky's Theorem. □

Proof of Theorem 2 By (20), Proposition 1 and the conditions in Theorem 2, we have $\sup_{u\in\mathcal{I}} |\widehat{g}(u;\widehat{\delta}) - g(u)| = O_p\{(\log n)^{1/2} J_{n,2}^{1/2} n^{-1/2} + J_{n,1}^{-\alpha}\}$. □

Proof of Theorem 3 Let $\Xi_i = E(\Psi_i^{\otimes 2})$ and $\Pi_i = E(\sigma^2(U_i)\Psi_i^{\otimes 2})$. Let $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ be independent random variables from $\mathrm{MVN}(\mathbf{0}, \mathbf{I}_{pJ_{n,2}\times pJ_{n,2}})$, where $\mathbf{Z}_i = \{Z_{i,sk}\}$.

Define

$$\eta_k(t) = \sigma_{n,k}^{-1}(t)\mathbf{B}_2(t)^{\mathrm{T}}\mathbf{\Lambda}_k \left\{ n^{-1}\sum_{i=1}^{n} \Xi_i \right\}^{-1} n^{-1}\sum_{i=1}^{n} \varepsilon_i \Psi_i,$$

$$\eta_k^0(t) = \sigma_{n,k}^{-1}(t)\mathbf{B}_2(t)^{\mathrm{T}}\mathbf{\Lambda}_k \left\{ n^{-1}\sum_{i=1}^{n} \Xi_i \right\}^{-1} n^{-1}\sum_{i=1}^{n} \Pi_i^{1/2}\mathbf{Z}_i.$$

By the fact that $\widehat{\beta}_k - \beta_{k,n} = \sigma_{n,k}^{-1}(t)\mathbf{B}_2(t)^{\mathrm{T}}\mathbf{\Lambda}_k(\widehat{\delta} - \widetilde{\delta}_n^0)$, (8) and (28), we have

$$\sup_{t\in\mathcal{T}} |\widehat{\beta}_k(t) - \beta_k(t) - \eta_k(t)| = o_p(J_{n,2}^{1/2} n^{-1/2}). \tag{29}$$

It is apparent that $\eta_k^0(t)$ is a Gaussian process with $E\{\eta_k^0(t)\} \equiv 0$, $\mathrm{Var}\{\eta_k^0(t)\} \equiv 1$, and covariance matrix given in (9). Therefore, we have

$$P\{\sup_{t\in T} |\eta_k^0(t)| \le Q_k(\alpha)\} = 1 - \alpha. \tag{30}$$

Next, we will prove that $\sup_{t\in\mathcal{T}} |\eta_k(t) - \eta_k^0(t)| = o_p(1)$. Let $\mathbf{e}_i = \{e_{i,sk}\} = \Pi_i^{-1/2}\varepsilon_i\Psi_i$. Denote $\Pi_i^{1/2} = \{\xi_{i,s'k',sk}\}$. There exists a constant $0 < C < \infty$, such that $\sup |\xi_{i,s'k',sk}| \le C J_{n,2}^{-1/2}$. Then, $E(\mathbf{e}_i) = \mathbf{0}$ and $\mathrm{Var}(\mathbf{e}_i) = \mathbf{I}_{pJ_{n,2}\times pJ_{n,2}}$.There exist s, s', k, k' such that

$$\left\| n^{-1}\sum_{i=1}^{n} \Pi_i^{1/2}(\mathbf{e}_i - \mathbf{Z}_i) \right\|_\infty \le n^{-1} p J_{n,2} \left| \sum_{i=1}^{n} \xi_{i,s'k',s,k}(e_{i,sk} - Z_{i,sk}) \right|.$$

For notation simplicity, let $\xi_i = \xi_{i,s'k',s,k}$. Order all ξ_i, $1 \le i \le n$, from the largest to the smallest such that $\xi_{(1)} \ge \xi_{(2)} \ge \cdots \ge \xi_{(n)}$. Moreover, $Z_{i,sk}$ can be written as $Z_{i,sk} = W(i) - W(i-1)$, where $\{W(s), 0 \le s < \infty\}$ is a Wiener process that is a Borel function of $Z_{i,sk}$. Let $S_i = \sum_{i'=1}^{i} e_{i',sk}$ and $S_0 = 0$. Define $M_n = \max_{1\le s\le n} |S_s - W(s)|$. By Theorem 2.6.2 in Csőrgő and Révész (1981), we have $M_n = O_p(\log n)$. Then, allowdisplaybreaks

$$\begin{aligned}
&\left\| n^{-1}\sum_{i=1}^{n} \Pi_i^{1/2}(\mathbf{e}_i - \mathbf{Z}_i) \right\|_\infty \\
&\le n^{-1} p J_{n,2} \left\{ |\xi_n(S_n - W(n))| + \left| \sum_{i=1}^{n-1} (\xi_i - \xi_{i+1})(S_i - W(i)) \right| \right\} \\
&\le n^{-1} p J_{n,2} M_n \left(C J_{n,2}^{-1/2} + \sum_{i=1}^{n-1} |\xi_i - \xi_{i+1}| \right) \\
&= n^{-1} p J_{n,2} M_n (C J_{n,2}^{-1/2} + |\xi_1 - \xi_n|) \\
&\le 3C n^{-1} p J_{n,2}^{1/2} M_n = O_p(J_{n,2}^{1/2} n^{-1} \log n).
\end{aligned}$$

Therefore,

$$
\begin{aligned}
&\sup_{t\in\mathcal{T}} |\eta_k(t) - \eta_k^0(t)| \\
&\leq \sup_{t\in\mathcal{T}} \left\{ |\sigma_{n,k}^{-1}(t)| \sum_{r=1}^{J_{n,2}} |B_{r,2}(t)| \right\} \|\boldsymbol{\Lambda}_k\|_\infty \\
&\quad \times \left\| \left\{ n^{-1} \sum_{i=1}^{n} \Xi_i \right\}^{-1} \right\|_\infty \left\| n^{-1} \sum_{i=1}^{n} \Pi_i^{1/2} (\mathbf{e}_i - \mathbf{Z}_i) \right\|_\infty \\
&= O_p(n^{1/2} J_{n,2}^{-1/2}) O_p(J_{n,2}) O_p(J_{n,2}^{1/2} n^{-1} \log n) \\
&= O_p(J_{n,2} n^{-1/2} \log n) = o_p(1). \qquad (31)
\end{aligned}
$$

Thus, Theorem follows from (29), (30) and (31). □

Acknowledgments The author thanks the Editor, the Associate Editor and the two referees for their insightful comments and suggestions that lead to substantially improve the paper.

References

Bunea, F., Ivanescu, A.E., Wegkamp, M.H. (2011). Adaptive inference for the mean of a Gaussian process in functional data. *Journal of the Royal Statistical Society Series B*, 531–558.

Cai, T., Hall, P. (2006). Prediction in functional linear regression. *Annals of Statistics*, *34*, 2159–2179.

Cardot, H., Ferraty, F., Sarda, P. (2003). Spline estimators for the functional linear model. *Statistica Sinica*, *13*, 571–591.

Carroll, R. J., Fan, J., Gijbels, I., Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association*, *92*, 477–489.

Chen, D., Hall, P., Müller, H. G. (2011). Single and multiple index functional regression models with nonparametric link. *Annals of Statistics*, *39*, 1720–1747.

Claeskens, G., Krivobokova, T., Opsomer, J. D. (2009). Asymptotic properties of penalized spline estimators. *Biometrika*, *96*, 529–544.

Crainiceanu, C. M., Staicu, A.-M., Ray, S., Punjabi, N. (2012). Bootstrap-based inference on the difference in the means of two correlated functional processes. *Statistics in Medicine*, *31*, 3223–3240.

Csőrgő, M., Révész, P. (1981). *Strong approximations in probability and statistics*. New York: Academic Press.

de Boor, C. (2001). *A practical guide to splines*. New York: Springer.

Demko, S. (1986). Spectral bounds for $\left|a^{-1}\right|_\infty$. *Journal of Approximation Theory*, *48*, 207–212.

DeVore, R. A., Lorentz, G. G. (1993). *Constructive approximation*. Berlin: Springer-Verlag.

Gertheiss, J., Maity, A., Staicu, A.-M. (2013). Variable selection in generalized functional linear models. *Stat*, *2*, 86–101.

Goldsmith, J., Bobb, J., Crainiceanu, C. M., Caffo, B., Reich, D. (2011a). Penalized functional regression. *Journal of Computational and Graphical Statistics*, *20*, 830–851.

Goldsmith, J., Crainiceanu, C. M., Caffo, B., Reich, D. (2011b). Penalized functional regression analysis of white-matter tract profiles in multiple sclerosis. *NeuroImage*, *57*, 431–439.

Goldsmith, J., Crainiceanu, C., Caffo, B., Reich, D. (2012). Longitudinal penalized functional regression for cognitive outcomes on neuronal tract measurements. *Journal of the Royal Statistical Society Series C*, *61*, 453–469.

Hall, P., Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *Annals of Statistics*, *35*, 70–91.

Huang, J. (2003). Local asymptotics for polynomial spline regression. *Annals of Statistics*, *31*, 1600–1635.

Huang, J. Z., Wu, C. O., Zhou, L. (2004). Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statistica Sinica*, *14*, 763–788.

Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single index models. *Journal of Econometrics*, *58*, 71–120.

Jiang, C. R., Wang, J. L. (2011). Functional single index models for longitudinal data. *Annals of Statistics*, *39*, 362–388.

Li, Y., Hsing, T. (2007). On rates of convergence in functional linear regression. *Journal of Multivariate Analysis*, *98*, 1782–1804.

Ma, S., Yang, L., Carroll, R. (2012). Simultaneous confidence band for sparse longitudinal regression. *Statistica Sinica*, *22*, 95–122.

McLean, M. W., Hooker, G., Staicu, A.-M., Scheipl, F., Ruppert, D. (2014). Functional generalized additive models. *Journal of Computational and Graphical Statistics*, *23*, 249–269.

Müller, H. G., Stadtmüller, U. (2005). Generalized functional linear models. *Annals of Statistics*, *33*, 774–805.

Ramsay, J., Silverman, B. W. (2005). *Functional data analysis. Springer series in statistics* (2nd ed.). New York: Springer.

Wu, Y. (2010). *The partially monotone tensor spline estimation of joint distribution function with bivariate current status data*. In Technical report, University of Iowa.

Xia, Y., Tong, H., Li, W. K., Zhu, L. (2002). An adaptive estimation of dimension reduction space (with discussion). *Journal of the Royal Statistical Society Series B*, *64*, 363–410.

Xue, L., Qu, A., Zhou, J. (2010). Consistent model selection for marginal generalized additive model for correlated data. *Journal of American Statistical Association*, *105*, 1518–1530.

Xue, L., Yang, L. (2006). Additive coefficient modeling via polynomial spline. *Statistica Sinica*, *16*, 1423–1446.

Yao, F., Müller, H. G., Wang, J. L. (2005). Functional linear regression analysis for longitudinal data. *The Annals of Statistics*, *33*, 2873–2903.

Zhou, S., Shen, X., Wolfe, D. A. (1998). Local asymptotics for regression splines and confidence regions. *The Annals of Statistics*, *26*, 1760–1782.