Statistica Sinica Preprint No: SS-2018-0316	
Title	Partial Functional Partially Linear Single-Index Models
Manuscript ID	SS-2018-0316
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202018.0316
Complete List of Authors	Qingguo Tang Linglong Kong David Ruppert and Rohana J. Karunamuni
Corresponding Author	Rohana J. Karunamuni
E-mail	R.J.Karunamuni@ualberta.ca
Notice: Accepted version subject to English editing.	

PARTIAL FUNCTIONAL PARTIALLY LINEAR SINGLE-INDEX MODELS

Qingguo Tang[1], Linglong Kong[2], David Ruppert[3], and Rohana J. Karunamuni[2]

[1] *Nanjing University of Science and Technology*

[2] *University of Alberta and* [3] *Cornell University*

Abstract: This paper studies a *partial functional partially linear single-index model* that consists of a functional linear component as well as a linear single-index component. This model generalizes many well-known existing models and is suitable for more complicated data structures. We develop a new estimation procedure which combines functional principal component analysis of the functional predictors, a B-spline model for parameters, and profile estimation of the unknown parameters and functions in the model. We establish the consistency and asymptotic normality of the parametric estimators. Furthermore, the global convergence rate of the proposed estimator of the linear slope function is also obtained, and it is optimal in the minimax sense. We implement a two-stage procedure to estimate the nonparametric link function of the single-index component of the model, and the resulting estimator possesses the optimal global rate of convergence. Further, we also obtain the convergence rate of the mean squared prediction error for a predictor. We study empirical properties of the proposed procedures through Monte Carlo simulations. The proposed method is illustrated by analyzing a diffusion tensor imaging (DTI) data set from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database.

Keywords: Functional data analysis, Single-index model, Principal component analysis, Consistency, Asymptotic normality.

1. Introduction

Functional data analysis has generated increasing interest in recent years in many areas, including biology, chemometrics, econometrics, geophysics, medical sciences, meteorology, etc. In neurosciences, there is great need in the analysis of complex neuroimaging data obtained by collecting structural, neurochemical, and functional images over both time and space. Functional data are made up

of repeated measurements taken as curves, surfaces or other objects varying over a continuum, such as the time and space. In many experiments, such as clinical diagnosis of neurological diseases from the brain imaging data, functional data appear as the basic unit of observations. As a natural extension of the multivariate data analysis, functional data analysis provides valuable insights into these experiments, taking into account the underlying smoothness of high-dimensional covariates and provides new approaches for solving inference problems. One may refer to the monographs of Ramsay and Silverman (2002, 2005), Ferraty and Vieu (2006), Horváth and Kokoszka (2012) and Hsing and Eubank (2015) for a general overview on functional data analysis.

In the analysis of complex neuroimaging data, motivated by more complicated data structures, which appeal to more comprehensive, flexible and adaptable models, in this paper we investigate the following *partial functional partially linear single-index model*:

$$Y = \int_{\mathcal{T}} a(t)X(t)dt + W^T\boldsymbol{\alpha}_0 + g(Z^T\boldsymbol{\beta}_0) + \varepsilon, \tag{1.1}$$

where $X(t)$ is a random function defined on some bounded interval $\mathcal{T}$, $a(t)$ is an unknown square integrable slope function on $\mathcal{T}$, W is a $q \times 1$ vector of covariates, $\boldsymbol{\alpha}_0$ is a $q \times 1$ unknown coefficient vectors, $Z \in R^d$ is a $d \times 1$ vector of covariates, $\boldsymbol{\beta}_0$ is a $d \times 1$ coefficient vector to be estimated, g is an unknown link function, ε is a random error with mean zero and variance σ^2, and ε is independent of the covariates $(X(t), W, Z)$.

Model (1.1) is more flexible and can deal with more complicated data structures than currently available models. To the best of our knowledge, a model containing a functional linear component as well as a linear single-index com-

ponent has not been fully studied in the literature yet. This model generalizes many well-known existing models and is suitable for more complicated data structures. However, its estimation inherits some difficulties and complexities from both components that create a challenging problem calling for new methodology. We propose a new estimation procedure which combines functional principal component analysis, B-spline methods and a profile method to estimate unknown parameters and functions in model (1.1). Using functional principal component analysis, the unknown slope function is approximated by an average value which includes the unknown parameters and link function. The estimators of the unknown parameters and link function are obtained by solving a series of minimization problems. The advantages of our method include that it can avoid ill-posed inverse problems that arise in functional data analysis, and the unknown parameters and functions in the model can be estimated efficiently. Since a principal component basis is very efficient for modeling functional predictors and is widely used in practice, our method for using this basis should be of interest in other contexts. In particular, our method can be generalized to the models which are formed by a linear functional model plus a general model with a finite-dimensional predictor variable, such as partially linear models, varying coefficient models and additive models.

Model (1.1) can be interpreted from two perspectives. First, it generalizes the partial functional linear models

$$Y = \int_{\mathcal{T}} a(t)X(t)dt + W^T\boldsymbol{\alpha}_0 + \varepsilon, \tag{1.2}$$

by adding a nonparametric component, $g(Z^T\boldsymbol{\beta}_0)$, with an unknown univariate link function g. This single-index term reduces the dimensionality from the mul-

tivariate predictors to a univariate index $Z^T\boldsymbol{\beta}_0$ and avoids the curse of dimensionality, while still capturing important features in high-dimensional data. Furthermore, since a nonlinear link function g is applied to the index $Z^T\boldsymbol{\beta}_0$, interactions between the covariates Z can be modeled. The standard functional linear model (Li and Hsing 2007, Cardot et al. 2007, Cai and Hall 2006, and Hall and Horowitz 2007) with scalar response Y has the same form as model (1.2) without the linear part. In general, $X(t)$ can be a multivariate functional variable, but here we will only focus on the univariate case. Our main interest is estimation of functional coefficient $a(t)$ based on a sample $(X_1, Y_1), ..., (X_n, Y_n)$ generated from the standard functional linear model. There are a number of articles in the literature discussing the slope estimation in model (1.2) using methods such as the penalized spline method (Cardot et al. 2007), the functional principal component analysis (Cai and Hall 2006, Hall and Horowitz 2007, Yuan and Cai 2010) and the functional partial least squares method (Delaigle and Hall 2012), among others.

Second, model (1.1) can be considered as a generalization of the partially linear single-index model (Carroll et al. 1997, Yu and Ruppert 2002),

$$Y = g(Z^T\boldsymbol{\beta}_0) + W^T\boldsymbol{\alpha}_0 + \varepsilon, \tag{1.3}$$

with an addition of functional covariates $X(t)$. The partially linear single-index model (1.3) was first explored by Carroll et al. (1997). In fact, they have considered a more generalized version, where a known link function is employed in the regression function, while model (1.3) assumes an identity link function. Model (1.3) has also been studied by many other authors, including Xia and Härdle (2006), Liang et al. (2010) and Wang et al. (2010).

In order to estimate unknown quantities in model (1.1), we develop a new

method of estimation, which is a combination of functional principal component analysis, B-spline method and a profile method. We believe our technique is new and is first to explore a combination of the functional principal component analysis and a profile method in functional linear models. More specifically, we estimate the unknown parameters $(\boldsymbol{\alpha}_0^T, \boldsymbol{\beta}_0^T)^T$ by employing a B-spline function to approximate the unknown link function g combine with the functional principal component analysis (FPCA) to estimate the slope function $a(t)$. Under some regularity conditions, we prove the consistency and asymptotic normality of the proposed estimators. We also establish a global rate of convergence of the estimator of $a(t)$, and it is shown to be optimal in the minimax sense of Hall and Horowitz (2007). Based on the estimators of parameters, another B-spline function is employed to approximate the function g and then the optimal global convergence rate of the approximation is established. We also obtain convergence rates of the mean squared prediction error for a predictor. We further apply our model and estimation method to analyze a diffusion tensor imaging (DTI) data set from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. The analyzing results indicate that model (1.1) is more flexible and efficient than model (1.2).

To gain more flexibility and partly motivated by applications, a number of other models based on the standard functional linear model have been studied in the literature, including the partial functional linear regression model (1.2) (Shen 2009, Shen and Lee 2012, Tang and Cheng 2014, Kong et al. 2016, Yao et al. 2017), generalized functional linear models (Li et al. 2010, Chen and Müller 2012), single and multiple index functional regression models (Chen et al. 2011, Ma 2016) and a functional partial linear single-index model (Wang et al. 2016),

among others.

The paper is organized as follows. Section 2 describes the proposed estimation method. Section 3 presents asymptotic results of our estimator. In Section 4, we conduct simulation studies to examine the finite sample performance of the proposed procedures. In Section 5, the proposed method is illustrated by analyzing a diffusion tensor imaging data set from the Alzheimer's Disease Neuroimaging Initiative database (adni.loni.ucla.edu). The proofs of the main results are given in the *Supplementary Material*.

2. Proposed estimation method

In this section we develop a new estimation procedure which combines functional principal component analysis, B-spline methods and a profile method to estimate unknown parameters and functions in model (1.1).

Let Y be a real-valued response variable and $\{X(t) : t \in \mathcal{T}\}$ be a mean zero second-order (i.e., $EX(t)^2 < \infty$ for all $t \in \mathcal{T}$) stochastic process with sample paths in $L_2(\mathcal{T})$, where $\mathcal{T}$ is a bounded closed interval and $L_2(\mathcal{T})$ denotes the set of all square integrable functions on $\mathcal{T}$. Let $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ denote the $L_2(\mathcal{T})$ inner product and norm, respectively. Denote the covariance function of the process $X(t)$ by $K(s,t) = \mathrm{cov}(X(s), X(t))$. We suppose that $K(s,t)$ is positive definite. Then $K(s,t)$ admits a spectral decomposition in terms of strictly positive eigenvalues λ_j:

$$K(s,t) = \sum_{j=1}^{\infty} \lambda_j \phi_j(s) \phi_j(t), \quad s, t \in \mathcal{T}, \tag{2.1}$$

where λ_j and ϕ_j are eigenvalue and eigenfunction pairs of the linear operator with kernel K, the eigenvalues are ordered so that $\lambda_1 > \lambda_2 > \cdots > 0$ and eigenfunctions $\phi_1, \phi_2, \cdots$ form an orthonormal basis for $L_2(\mathcal{T})$. This leads to the

Karhunen-Loéve representation $X(t) = \sum_{j=1}^{\infty} \xi_j \phi_j(t)$, where $\xi_j = \int_{\mathcal{T}} X(t)\phi_j(t)dt$ are uncorrelated random variables with mean zero and variance $E\xi_j^2 = \lambda_j$. Let $a(t) = \sum_{j=1}^{\infty} a_j \phi_j(t)$. Then model (1.1) can be written as

$$Y = \sum_{j=1}^{\infty} a_j \xi_j + W^T \boldsymbol{\alpha}_0 + g(Z^T \boldsymbol{\beta}_0) + \varepsilon. \tag{2.2}$$

By (2.2), we have

$$a_j = E\{[Y - (W^T \boldsymbol{\alpha}_0 + g(Z^T \boldsymbol{\beta}_0))]\xi_j\}/\lambda_j. \tag{2.3}$$

Let $(X_i(t), W_i, Z_i, Y_i), i = 1, \cdots, n$, be independent realizations of $(X(t), W, Z, Y)$ generated from model (1.1). Then the empirical versions of K and of its spectral decomposition are

$$\hat{K}(s,t) = \frac{1}{n}\sum_{i=1}^{n} X_i(s)X_i(t) = \sum_{j=1}^{\infty} \hat{\lambda}_j \hat{\phi}_j(s)\hat{\phi}_j(t), \quad s,t \in \mathcal{T}. \tag{2.4}$$

Analogously to the case of K, $(\hat{\lambda}_j, \hat{\phi}_j)$ are (eigenvalue, eigenfunction) pairs for the linear operator with kernel $\hat{K}$, ordered such that $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq 0$. We take $(\hat{\lambda}_j, \hat{\phi}_j)$ and $\hat{\xi}_{ij} = \langle X_i, \hat{\phi}_j \rangle$ to be the estimators of (λ_j, ϕ_j) and $\xi_{ij} = \langle X_i, \phi_j \rangle$, respectively, and set

$$\tilde{a}_j = \frac{1}{n\hat{\lambda}_j}\sum_{i=1}^{n} \left[Y_i - (W_i^T \boldsymbol{\alpha}_0 + g(Z_i^T \boldsymbol{\beta}_0))\right] \hat{\xi}_{ij}. \tag{2.5}$$

In order to estimate g, we adapt spline approximations. We assume that $\|\boldsymbol{\beta}_0\| = 1$ and that the last element β_{0d} of $\boldsymbol{\beta}_0$ is positive, to ensure identifiability. Let $\boldsymbol{\beta}_{-d} = (\beta_1, \ldots, \beta_{d-1})^T$ and $\boldsymbol{\beta}_{0,-d} = (\beta_{01}, \ldots, \beta_{0(d-1)})^T$. Since $\beta_{0d} = \sqrt{1 - (\beta_{01}^2 + \cdots + \beta_{0(d-1)}^2)} > 0$, there exists a small constant $\rho_0 \in (0,1)$ such that $\boldsymbol{\beta}_0 \in \Theta_{\rho_0} = \{\boldsymbol{\beta} = (\beta_1, \ldots, \beta_d)^T : \beta_d = \sqrt{1 - (\beta_1^2 + \cdots + \beta_{d-1}^2)} \geq \rho_0\}$. Let $\mathcal{D}$ denote the convex hull of the discrete set of the observed Z_i, $i = 1, \ldots, n$. Denote

$U_* = \inf_{z\in\mathcal{D},\boldsymbol{\beta}\in\Theta_{\rho_0}} z^T\boldsymbol{\beta}$ and $U^* = \sup_{z\in\mathcal{D},\boldsymbol{\beta}\in\Theta_{\rho_0}} z^T\boldsymbol{\beta}$. We first split the interval $[U_*, U^*]$ into k_n subintervals with knots $\{U_* = u_{n0} < u_{n1} < \cdots < u_{nk_n} = U^*\}$. For fixed $\boldsymbol{\beta}$, there exist positive integers l and $k_{\boldsymbol{\beta}}$ such that $u_{n(l-1)} < \inf_{z\in\mathcal{D}} z^T\boldsymbol{\beta} \le u_{nl} < u_{n(l+k_{\boldsymbol{\beta}})} \le \sup_{z\in\mathcal{D}} z^T\boldsymbol{\beta} < u_{n(l+k_{\boldsymbol{\beta}}+1)}$. Let $U_{\boldsymbol{\beta}} = u_{nl}$ and $U^{\boldsymbol{\beta}} = u_{n(l+k_{\boldsymbol{\beta}})}$. For any fixed integer $s \ge 1$, let $S^s_{k_{\boldsymbol{\beta}}}(u)$ be the set of spline functions of degree s with knots $\{U_{\boldsymbol{\beta}} = u_{nl} < u_{n(l+1)} < \cdots < u_{n(l+k_{\boldsymbol{\beta}})} = U^{\boldsymbol{\beta}}\}$; that is, a function $f(u)$ belongs to $S^s_{k_{\boldsymbol{\beta}}}(u)$ if and only if $f(u)$ belongs to $C^{s-1}[u_{nl}, u_{n(l+k_{\boldsymbol{\beta}})}]$ and its restriction to each $[u_{nk}, u_{n(k+1)})$ is a polynomial of degree at most s. Let $\{B_{k\boldsymbol{\beta}}(u)\}_{k=1}^{K_{\boldsymbol{\beta}}}$ be a basis for $S^s_{k_{\boldsymbol{\beta}}}(u)$, where $K_{\boldsymbol{\beta}} = k_{\boldsymbol{\beta}} + s$. See Schumaker (1981) for the construction of the spline basis.

For fixed $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, we use $\sum_{j=1}^m \tilde{a}_j\hat{\xi}_j$ to approximate $\sum_{j=1}^\infty a_j\xi_j$ in (2.2) and use $\sum_{k=1}^{K_{\boldsymbol{\beta}}} b_k B_{k\boldsymbol{\beta}}(u)$ to approximate $g(u)$ for $u \in [U_{\boldsymbol{\beta}}, U^{\boldsymbol{\beta}}]$. We then estimate $g(\cdot)$ by minimizing

$$\sum_{i=1}^n \left\{ Y_i - \sum_{j=1}^m \frac{\hat{\xi}_{ij}}{n\hat{\lambda}_j} \sum_{l=1}^n \left[Y_l - W_l^T\boldsymbol{\alpha} - \sum_{k=1}^{K_{\boldsymbol{\beta}}} b_k B_{k\boldsymbol{\beta}}(Z_l^T\boldsymbol{\beta}) \right] \hat{\xi}_{lj} - W_i^T\boldsymbol{\alpha} - \sum_{k=1}^{K_{\boldsymbol{\beta}}} b_k B_{k\boldsymbol{\beta}}(Z_i^T\boldsymbol{\beta}) \right\}^2 \tag{2.7}$$

with respect to $b_1, \ldots, b_{K_{\boldsymbol{\beta}}}$, where m is a smoothing parameter which denotes a frequency cut-off. Define $\tilde{\xi}_{il} = \sum_{j=1}^m \hat{\xi}_{ij}\hat{\xi}_{lj}/\hat{\lambda}_j$, $\tilde{Y}_i = Y_i - \frac{1}{n}\sum_{l=1}^n Y_l\tilde{\xi}_{il}$, $\tilde{W}_i = W_i - \frac{1}{n}\sum_{l=1}^n W_l\tilde{\xi}_{il}$ and $\tilde{B}_{k\boldsymbol{\beta}}(Z_i^T\boldsymbol{\beta}) = B_{k\boldsymbol{\beta}}(Z_i^T\boldsymbol{\beta}) - \frac{1}{n}\sum_{l=1}^n B_{k\boldsymbol{\beta}}(Z_l^T\boldsymbol{\beta})\tilde{\xi}_{il}$. Then (2.7) can be written as

$$\sum_{i=1}^n \left\{ \tilde{Y}_i - \tilde{W}_i^T\boldsymbol{\alpha} - \sum_{k=1}^{K_{\boldsymbol{\beta}}} b_k \tilde{B}_{k\boldsymbol{\beta}}(Z_i^T\boldsymbol{\beta}) \right\}^2. \tag{2.8}$$

Denote $\tilde{\boldsymbol{B}}_{\boldsymbol{\beta}}(Z_i^T\boldsymbol{\beta}) = (\tilde{B}_{1\boldsymbol{\beta}}(Z_i^T\boldsymbol{\beta}), \ldots, \tilde{B}_{K_{\boldsymbol{\beta}}\boldsymbol{\beta}}(Z_i^T\boldsymbol{\beta}))^T$, $\tilde{\boldsymbol{B}}(\boldsymbol{\beta}) = (\tilde{\boldsymbol{B}}_{\boldsymbol{\beta}}(Z_1^T\boldsymbol{\beta}), \ldots, \tilde{\boldsymbol{B}}_{\boldsymbol{\beta}}(Z_n^T\boldsymbol{\beta}))^T$, $\tilde{\boldsymbol{Y}} = (\tilde{Y}_1, \ldots, \tilde{Y}_n)^T$, $\tilde{\boldsymbol{W}} = (\tilde{W}_1, \ldots, \tilde{W}_n)^T$ and $\boldsymbol{b} = (b_1, \ldots, b_{K_{\boldsymbol{\beta}}})^T$. If $\tilde{\boldsymbol{B}}^T(\boldsymbol{\beta})\tilde{\boldsymbol{B}}(\boldsymbol{\beta})$ is invertible, then the estimator $\tilde{\boldsymbol{b}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = (\tilde{b}_1(\boldsymbol{\alpha}, \boldsymbol{\beta}), \ldots, \tilde{b}_{K_{\boldsymbol{\beta}}}(\boldsymbol{\alpha}, \boldsymbol{\beta}))^T$ of $\boldsymbol{b}$ is given

by

$$\tilde{\boldsymbol{b}}(\boldsymbol{\alpha},\boldsymbol{\beta}) = \left\{\tilde{\boldsymbol{B}}^T(\boldsymbol{\beta})\tilde{\boldsymbol{B}}(\boldsymbol{\beta})\right\}^{-1}\tilde{\boldsymbol{B}}^T(\boldsymbol{\beta})(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\boldsymbol{\alpha}). \tag{2.9}$$

We solve the following minimization problem

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \left\{\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\boldsymbol{\alpha} - \tilde{\boldsymbol{B}}(\boldsymbol{\beta})\tilde{\boldsymbol{b}}(\boldsymbol{\alpha},\boldsymbol{\beta})\right\}^T \left\{\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\boldsymbol{\alpha} - \tilde{\boldsymbol{B}}(\boldsymbol{\beta})\tilde{\boldsymbol{b}}(\boldsymbol{\alpha},\boldsymbol{\beta})\right\} \tag{2.10}$$

to obtain the estimators $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$. A Newton-Raphson algorithm can be applied for the minimization. An estimator of $\boldsymbol{b}$ is obtained by solving the following minimization problem

$$\hat{\boldsymbol{b}} = \min_{\boldsymbol{b}} \sum_{i=1}^{n} \left\{\tilde{Y}_i - \tilde{W}_i^T\hat{\boldsymbol{\alpha}} - \boldsymbol{b}^T\tilde{\boldsymbol{B}}_{\hat{\boldsymbol{\beta}}}(Z_i^T\hat{\boldsymbol{\beta}})\right\}^2, \tag{2.11}$$

and then $\hat{\boldsymbol{b}}$ is given by

$$\hat{\boldsymbol{b}} = \tilde{\boldsymbol{b}}(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}) = \left\{\tilde{\boldsymbol{B}}^T(\hat{\boldsymbol{\beta}})\tilde{\boldsymbol{B}}(\hat{\boldsymbol{\beta}})\right\}^{-1}\tilde{\boldsymbol{B}}^T(\hat{\boldsymbol{\beta}})(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}^T\hat{\boldsymbol{\alpha}}). \tag{2.12}$$

Let $\tilde{g}(u) = \sum_{k=1}^{K_{\hat{\boldsymbol{\beta}}}} \hat{b}_k B_{k\hat{\boldsymbol{\beta}}}(u)$ for $u \in [U_{\hat{\boldsymbol{\beta}}}, U^{\hat{\boldsymbol{\beta}}}]$. We then choose a new tuning parameter $\tilde{m}$ and the estimator of $a(t)$ is given by

$$\hat{a}(t) = \sum_{j=1}^{\tilde{m}} \hat{a}_j\hat{\phi}_j(t) \tag{2.13}$$

with

$$\hat{a}_j = \frac{1}{n\hat{\lambda}_j}\sum_{i=1}^{n}\left\{Y_i - W_i^T\hat{\boldsymbol{\alpha}} - \tilde{g}(Z_i^T\hat{\boldsymbol{\beta}})\right\}\hat{\xi}_{ij}. \tag{2.14}$$

In order to construct an estimator of g that achieves the optimal rate of convergence, we select new knots and new B-spline basis based on the estimators $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$. Let $\{U_{\hat{\boldsymbol{\beta}}} = \bar{u}_{n0} < \bar{u}_{n1} < \cdots < \bar{u}_{nk_n^*)} = U^{\hat{\boldsymbol{\beta}}}\}$ be new knots and $\{B_k^*(u)\}_{k=1}^{K_n^*}$ be a new basis, where $K_n^* = k_n^* + s$. Then $B_{k\boldsymbol{\beta}}^*(Z_i^T\boldsymbol{\beta})$, $\boldsymbol{B}_{\boldsymbol{\beta}}^*(Z_i^T\boldsymbol{\beta})$ and $\boldsymbol{B}^*(\boldsymbol{\beta})$ are defined similarly as $\tilde{B}_{k\boldsymbol{\beta}}(Z_i^T\boldsymbol{\beta})$, $\tilde{\boldsymbol{B}}_{\boldsymbol{\beta}}(Z_i^T\boldsymbol{\beta})$ and $\tilde{\boldsymbol{B}}(\boldsymbol{\beta})$, respectively. We then solve

the following minimization problem

$$\min_{\boldsymbol{b}^*} \sum_{i=1}^{n} \left\{ \tilde{Y}_i - \tilde{W}_i^T \hat{\boldsymbol{\alpha}} - \boldsymbol{b}^{*T} \boldsymbol{B}^*_{\hat{\boldsymbol{\beta}}}(Z_i^T \hat{\boldsymbol{\beta}}) \right\}^2 \tag{2.15}$$

to obtain an estimator of $\boldsymbol{b}^*$, where $\boldsymbol{b}^* = (b_1, \ldots, b_{K_n^*})^T$. If $\boldsymbol{B}^{*T}(\hat{\boldsymbol{\beta}})\boldsymbol{B}^*(\hat{\boldsymbol{\beta}})$ is invertible, then an estimator of $\boldsymbol{b}^*$ is given by

$$\hat{\boldsymbol{b}}^* = \boldsymbol{b}^*(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) = \left\{ \boldsymbol{B}^{*T}(\hat{\boldsymbol{\beta}})\boldsymbol{B}^*(\hat{\boldsymbol{\beta}}) \right\}^{-1} \boldsymbol{B}^{*T}(\hat{\boldsymbol{\beta}})(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}^T \hat{\boldsymbol{\alpha}}). \tag{2.16}$$

The second stage estimator of $g(u)$ is then equal to $\hat{g}(u) = \sum_{k=1}^{K_n^*} \hat{b}_k^* B_k^*(u)$ for $u \in [U_{\hat{\boldsymbol{\beta}}}, U^{\hat{\boldsymbol{\beta}}}]$.

To implement our estimation method, some appropriate values for m, k_n, $\tilde{m}$ and K_n^* are necessary. The values for tuning parameter m and k_n can be selected by the Bayesian information criterion (BIC) given by

$$\begin{aligned} &BIC(m, k_n) \\ &= \log \left\{ \tfrac{1}{n} \textstyle\sum_{i=1}^{n} \left(Y_i - W_i \hat{\boldsymbol{\alpha}}_{m,k_n} - \sum_{j=1}^{m} \hat{a}_j \hat{\xi}_{ij} - \tilde{g}(Z_i^T \hat{\boldsymbol{\beta}}_{m,k_n}) \right)^2 \right\} + \tfrac{\log(n)(m+k_n+s)}{n}, \end{aligned}$$

where $\hat{\boldsymbol{\alpha}}_{m,k_n}$ and $\hat{\boldsymbol{\beta}}_{m,k_n}$ depend on m and k_n. Large values of BIC indicate poor fitting. m and k_n are used to estimate the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. From our simulation in Section 4 below, we observe that the parametric estimators $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ are not sensitive to the choices of m and k_n, for simplicity, we can choose $k_n = c_0 n^{1/(2s-1)}$ with some positive constant c_0.

After m and k_n are determined, the value for tuning parameter $\tilde{m}$ can be selected by the following BIC information criterion given by

$$BIC(\tilde{m}) = \log \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - W_i \hat{\boldsymbol{\alpha}} - \sum_{j=1}^{\tilde{m}} \hat{a}_j \hat{\xi}_{ij} - \tilde{g}(Z_i^T \hat{\boldsymbol{\beta}}) \right)^2 \right\} + \frac{\log(n)\tilde{m}}{n}.$$

A value for K_n^* can also be selected by the following BIC information criterion:

$$BIC(K_n^*) = \log \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(\tilde{Y}_i - \tilde{W}_i \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{b}}^{*T} \boldsymbol{B}^*_{\hat{\boldsymbol{\beta}}}(Z_i^T \hat{\boldsymbol{\beta}}) \right)^2 \right\} + \frac{\log(n) K_n^*}{n}.$$

In practice, the proposed estimation method is implemented using the following steps:

Step 1. Choose an m and fit a partial functional linear model; that is, solve the minimization problem (2.8) with the link function g replaced by a linear function to obtain initial values $\hat{\boldsymbol{\alpha}}^{(0)}$ and $\hat{\boldsymbol{\beta}}_1^{(0)}$. Then set $\hat{\boldsymbol{\beta}}^{(0)} = \hat{\boldsymbol{\beta}}_1^{(0)}/\|\hat{\boldsymbol{\beta}}_1^{(0)}\|$, and multiply it by -1 if necessary.

Step 2. Compute $U_{\hat{\boldsymbol{\beta}}^{(0)}}$ and $U^{\hat{\boldsymbol{\beta}}^{(0)}}$, and construct the B-spline basis $\{B_{k\hat{\boldsymbol{\beta}}^{(0)}}(u)\}_{k=1}^{K_{\hat{\boldsymbol{\beta}}^{(0)}}}$. Then obtain $\tilde{b}(\hat{\boldsymbol{\alpha}}^{(0)}, \hat{\boldsymbol{\beta}}^{(0)})$ from (2.9) and solve the minimizing problem (2.10) to obtain the estimators $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$.

Step 3. Compute $\hat{b}$ and $\hat{a}_j$ from (2.12) and (2.14), respectively, and obtain the estimator $\hat{a}(t)$.

Step 4. Compute $U_{\hat{\boldsymbol{\beta}}}$ and $U^{\hat{\boldsymbol{\beta}}}$, and construct the basis $\{B_k^*(u)\}_{k=1}^{K_n^*}$. Then obtain the estimator $\hat{\mathbf{b}}^*$ from (2.16) and obtain the estimator $\hat{g}(u)$.

Remark 2.1. In practical applications, $X(t)$ is only discretely observed. Without loss of generality, suppose $X_i(t)$ is observed at n_i discrete points $0 = t_{i1} < \ldots < t_{in_i} = 1$ for each $i = 1, \ldots, n$. Then linear interpolation functions or spline interpolation functions can be used for estimation of $X_i(t)$.

Remark 2.2 Though the basis function $B_{k\boldsymbol{\beta}}(u)$ depends on $\boldsymbol{\beta}$, we see from (2.6) that the total number of all different $B_{k\boldsymbol{\beta}}(u)$ is not more than $(s+1)k_n$. In certain practical applications where the sample size n is not large enough and k_n is not large, one can choose $U_{\boldsymbol{\beta}} = \inf_{z\in\mathcal{D}} z^T\boldsymbol{\beta}$ and $U^{\boldsymbol{\beta}} = \sup_{z\in\mathcal{D}} z^T\boldsymbol{\beta}$ and construct the basis $\{B_{k\boldsymbol{\beta}}(u)\}_{k=1}^{K_{\boldsymbol{\beta}}}$ with knots $\{U_{\boldsymbol{\beta}} < u_{n(l+1)} < \cdots < u_{n(l+k_{\boldsymbol{\beta}}-1)} < U^{\boldsymbol{\beta}}\}$ to make full use of the data. That is, the intervals $[u_{nl}, u_{n(l+1)}]$ and $[u_{n(l+k_{\boldsymbol{\beta}}-1)}, u_{n(l+k_{\boldsymbol{\beta}})}]$ are replaced by $[U_{\boldsymbol{\beta}}, u_{n(l+1)}]$ and $[u_{n(l+k_{\boldsymbol{\beta}}-1)}, U^{\boldsymbol{\beta}}]$, respectively.

3. Asymptotic properties

In this section we state the main results on the asymptotic normality and convergence rates of the estimators proposed in the previous section. Before stating main results, we first state a few assumptions that are necessary to prove the theoretical results.

Assumption 1. $E(Y^4) < +\infty$, $E(\|W\|^4) < +\infty$ and $\int_{\mathcal{T}} E(X^4(t))dt < \infty$. $E(\xi_j|Z^T\boldsymbol{\beta}) = 0$ and $E(\xi_i\xi_j|Z^T\boldsymbol{\beta}) = 0$ for $i \neq j,\ i,j = 1,2,\ldots$; and $\boldsymbol{\beta} \in \Theta_{\rho_0}$. For each $j \geq 1$, $E(\xi_j^{2r}|Z^T\boldsymbol{\beta}) \leq C_1\lambda_j^r$ for $r = 1,2$, where $C_1 > 0$ is a constant. For any sequence $j_1,\ldots,j_4$, $E(\xi_{j_1}\ldots\xi_{j_4}|Z^T\boldsymbol{\beta}) = 0$ unless each index j_k is repeated.

Assumption 2. There exists a convex function φ defined on the interval $[0,1]$ such that $\varphi(0) = 0$ and $\lambda_j = \varphi(1/j)$ for $j \geq 1$.

Assumption 3. For Fourier coefficients a_j, there exist constants $C_2 > 0$ and $\gamma > 3/2$ such that $|a_j| \leq C_2 j^{-\gamma}$ for all $j \geq 1$.

Assumption 4. The function $g(u)$ is a s-times continuously differentiable function such that $|g^{(s)}(u') - g^{(s)}(u)| \leq C_3|u' - u|^{\varsigma}$, for $U_* \leq u', u \leq U^*$ and $p = s + \varsigma > 3$, with constants $0 < \varsigma \leq 1$ and $C_3 > 0$. The knots $\{U_* = u_{n0} < u_{n1} < \cdots < u_{nk_n} = U^*\}$ satisfy that $h_0/\min_{1\leq k\leq k_n} h_{nk} \leq C_4$, where $h_{nk} = u_{nk} - u_{n(k-1)}, h_0 = \max_{1\leq k\leq k_n} h_{nk}$ and $C_4 > 0$ is a constant.

Assumption 5. $nh_0^{2p} \to 0$, $n^{-1/2}m\lambda_m^{-1} \to 0$, $n^{-1}m^4\lambda_m^{-1}h_0^{-6}\log m \to 0$ and $m^{-2\gamma}h_0^{-2} \to 0$.

Assumption 5'. $m \to \infty$, $h_0 \to 0$, $n^{-1/2}m\lambda_m^{-1} \to 0$, $n^{-1}m^4\lambda_m^{-1}h_0^{-2}\log m \to 0$ and $(nh_0^3)^{-1}(\log n)^2 \to 0$.

Assumption 6. The marginal density function $f_{\boldsymbol{\beta}}(u)$ of $Z^T\boldsymbol{\beta}$ is bounded away from zero and infinity for $u \in [U_{\boldsymbol{\beta}}, U^{\boldsymbol{\beta}}]$ and satisfies that $0 < c_1 \leq f_{\boldsymbol{\beta}}(u) \leq C_5 < +\infty$ for $\boldsymbol{\beta}$ in a small neighborhood of $\boldsymbol{\beta}_0$ and $u \in [U_{\boldsymbol{\beta}_0}, U^{\boldsymbol{\beta}_0}]$, where c_1 and

C_5 are two positive constants.

Let $\mathcal{A}$ denote the class of the random variables such that $V \in \mathcal{A}$ if $V = \sum_{j=1}^{\infty} v_j \xi_j$ and $|v_j| \leq C_0 j^{-\gamma}$ for all $j \geq 1$, where γ is defined in Assumption 3 and $C_0 > 0$ is a constant. To derive the asymptotic distribution of the parametric estimators, we first adjust for the dependence of $W = (W_1, \ldots, W_q)^T$ and $X(t)$, which is a common complication in semiparametric models. Denote $V_r = \sum_{j=1}^{\infty} v_{rj} \xi_j$. Let $V_r^* = \sum_{j=1}^{\infty} v_{rj}^* \xi_j$ such that

$$V_r^* = \text{arginf}_{V_r \in \mathcal{A}} E[(W_r - \sum_{j=1}^{\infty} v_{rj} \xi_j)^2].$$

Since

$$E[(W_r - \textstyle\sum_{j=1}^{\infty} v_{rj} \xi_j)^2] = E[(W_r - E(W_r|X))^2] + E[(E(W_r|X) - \textstyle\sum_{j=1}^{\infty} v_{rj} \xi_j)^2],$$

we have

$$V_r^* = \text{arginf}_{V_r \in \mathcal{A}} E[(E(W_r|X) - \sum_{j=1}^{\infty} v_{rj} \xi_j)^2].$$

Thus, V_r^* are the projections of $E(W_r|X)$ onto the space $\mathcal{A}$. In other words, V_r^* is an element that belongs to $\mathcal{A}$, and it is the closest to $E(W_r|X)$ among all the random variables in $\mathcal{A}$. Let $\breve{V}_r = W_r - V_r^*$ for $r = 1, \ldots, d$, and $\breve{V} = (\breve{V}_1, \ldots, \breve{V}_d)^T$.

Under Assumption 4, according to Corollary 6.21 of Schumaker (1981, p.227), there exists a spline function $g_0(u) = \sum_{k=1}^{K_{\boldsymbol{\beta}_0}} b_{0k} B_{k\boldsymbol{\beta}_0}(u)$ and a constant $C_6 > 0$ such that

$$\sup_{u \in [U_{\boldsymbol{\beta}_0}, U^{\boldsymbol{\beta}_0}]} |R^{(k)}(u)| \leq C_6 h_0^{p-k} \tag{3.1}$$

for $k = 0, 1, \ldots, s$, where $R(u) = g(u) - g_0(u)$ and $R^{(k)}(u) = d^k R / du^k$. Let $\boldsymbol{B}_{\boldsymbol{\beta}}(u) = (B_{1\boldsymbol{\beta}}(u), \ldots, B_{K_{\boldsymbol{\beta}}\boldsymbol{\beta}}(u))^T$ and $\boldsymbol{b}_0 = (b_{01}, \ldots, b_{0K_{\boldsymbol{\beta}_0}})^T$. Define

$$\begin{aligned} G(\boldsymbol{\alpha}, \boldsymbol{\beta}) &= (\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)^T E(\breve{V}\breve{V}^T)(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) - 2\boldsymbol{b}_0^T E[\boldsymbol{B}_{\boldsymbol{\beta}_0}(Z^T \boldsymbol{\beta}_0)\breve{V}^T](\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) \\ &\quad + \boldsymbol{b}_0^T \Gamma(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0) \boldsymbol{b}_0 - \Pi^T(\boldsymbol{\alpha}, \boldsymbol{\beta}) \Gamma^{-1}(\boldsymbol{\beta}, \boldsymbol{\beta}) \Pi(\boldsymbol{\alpha}, \boldsymbol{\beta}) + \sigma^2, \end{aligned} \tag{3.2}$$

where $\Gamma(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = (\gamma_{kk'}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2))_{K_{\boldsymbol{\beta}_1} \times K_{\boldsymbol{\beta}_2}}$ with $\gamma_{kk'}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = E[B_{k\boldsymbol{\beta}_1}(Z^T\boldsymbol{\beta}_1)B_{k'\boldsymbol{\beta}_2}(Z^T\boldsymbol{\beta}_2)]$ and $\Pi(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \Gamma(\boldsymbol{\beta}, \boldsymbol{\beta}_0)\boldsymbol{b}_0 - E[\boldsymbol{B}_{\boldsymbol{\beta}}(Z^T\boldsymbol{\beta})\breve{V}^T](\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)$. Put $\boldsymbol{\theta} = (\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)^T$, $\boldsymbol{\theta}_{-d} = (\boldsymbol{\alpha}^T, \boldsymbol{\beta}_{-d}^T)^T$, $\hat{\boldsymbol{\theta}}_{-d} = (\hat{\boldsymbol{\alpha}}^T, \hat{\boldsymbol{\beta}}_{-d}^T)^T$ and $\boldsymbol{\theta}_{0,-d} = (\boldsymbol{\alpha}_0^T, \boldsymbol{\beta}_{0,-d}^T)^T$. Define

$$G^*(\boldsymbol{\theta}_{-d}) = G^*(\boldsymbol{\alpha}, \boldsymbol{\beta}_{-d}) = G(\boldsymbol{\alpha}, \beta_1, \ldots, \beta_{d-1}, \sqrt{1 - \|\boldsymbol{\beta}_{-d}\|^2})$$

and its Hessian matrix $H^*(\boldsymbol{\theta}_{-d}) = \frac{\partial^2}{\partial \boldsymbol{\theta}_{-d} \partial \boldsymbol{\theta}_{-d}^T} G^*(\boldsymbol{\theta}_{-d})$.

Assumption 7. $G^*(\boldsymbol{\theta}_{-d})$ is locally convex at $\boldsymbol{\theta}_{0,-d}$ such that for any $\varepsilon > 0$, there exists some $\epsilon > 0$ such that $\|\boldsymbol{\theta}_{-d} - \boldsymbol{\theta}_{0,-d}\| < \varepsilon$ holds whenever $|G^*(\boldsymbol{\theta}_{-d}) - G^*(\boldsymbol{\theta}_{0,-d})| < \epsilon$. Furthermore, the Hessian matrix $H^*(\boldsymbol{\theta}_{-d})$ is continuous in some neighborhood of $\boldsymbol{\theta}_{0,-d}$ and $H^*(\boldsymbol{\theta}_{0,-d}) > 0$.

Assumption 8. The knots $\{U_{\hat{\boldsymbol{\beta}}} = \bar{u}_{n0} < \bar{u}_{n1} < \cdots < \bar{u}_{n\vec{k}_n)} = U^{\hat{\boldsymbol{\beta}}}\}$ satisfy that $h/\min_{1 \le k \le \vec{k}_n} \bar{h}_{nk} \le C_7$, where $\bar{h}_{nk} = \bar{u}_{nk} - \bar{u}_{n(k-1)}, h = \max_{1 \le k \le \vec{k}_n} \bar{h}_{nk}$ and $C_7 > 0$ is a constant. Further, $h \to 0$ and $n^{-1}m^4\lambda_m^{-1}h^{-4}\log m \to 0$.

Assumptions 1 and 3 are standard conditions for functional linear models; see, e.g., Cai and Hall (2006) and Hall and Horowitz (2007). Assumption 2 is slightly less restrictive than (3.2) of Hall and Horowitz (2007). The quantity p in Assumption 4 is the order of smoothness of the function $g(u)$. Assumptions 5 can be easily verified and will be further discussed below. Assumption 6 ensures the existence and uniqueness of the spline estimator of the function $g(u)$. Assumption 7 ensures the existence and uniqueness of the estimator of $\boldsymbol{\theta}_{0,-d}$ in a neighborhood of $\boldsymbol{\theta}_{0,-d}$.

Remark 3.1. If $\lambda_j \sim j^{-\delta}$, $m \sim n^{\iota}$ and $h_0 \sim n^{-\tau}$, then Assumption 5 holds when $\iota < \min(1/(2(1+\delta)), 1/(\delta+4))$ and $1/(2p) < \tau < (1 - \iota(\delta+4))/6$, where $\delta > 1$, $\iota > 0$ and $\tau > 0$ are constants, and the notation $a_n \sim b_n$ means that the ratio a_n/b_n is bounded away from zero and infinity.

Theorem 3.1. (i) Suppose that Assumptions 1 to 4, 5', 6 and 7 hold. Then, as $n \to \infty$,

$$\hat{\boldsymbol{\alpha}} \xrightarrow{P} \boldsymbol{\alpha}_0, \quad \hat{\boldsymbol{\beta}}_{-d} \xrightarrow{P} \boldsymbol{\beta}_{0,-d}, \tag{3.3}$$

where $\xrightarrow{P}$ means convergence in probability.

(ii) Suppose that Assumptions 1 to 7 hold. Then

$$\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 = o_p(h_0), \quad \hat{\boldsymbol{\beta}}_{-d} - \boldsymbol{\beta}_{0,-d} = o_p(h_0). \tag{3.4}$$

In order to establish the asymptotic distributions of the estimators $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}_{-d}$, we first introduce some notation. Define

$$\tilde{G}_n(\boldsymbol{\theta}) = \tilde{G}_n(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \tilde{Y}_i - \tilde{W}_i^T \boldsymbol{\alpha} - \sum_{k=1}^{K_{\boldsymbol{\beta}}} \tilde{b}_k(\boldsymbol{\alpha}, \boldsymbol{\beta}) \tilde{B}_{k\boldsymbol{\beta}}(Z_i^T \boldsymbol{\beta}) \right\}^2. \tag{3.5}$$

Note that (3.5) is related to (2.10). By (3.4), if $u_{n(l-1)} < \inf_{z \in \mathcal{D}} z^T \boldsymbol{\beta}_0 < u_{nl}$, we have $U_{\hat{\boldsymbol{\beta}}} = U_{\boldsymbol{\beta}_0} = u_{nl}$ for sufficiently large n. If $\inf_{z \in \mathcal{D}} z^T \boldsymbol{\beta}_0 = u_{nl}$, then we modify u_{nl} such that $\inf_{z \in \mathcal{D}} z^T \boldsymbol{\beta}_0 < u_{nl}$, and also we then have $U_{\hat{\boldsymbol{\beta}}} = U_{\boldsymbol{\beta}_0} = u_{nl}$. Similarly, if $\sup_{z \in \mathcal{D}} z^T \boldsymbol{\beta}_0 = u_{n(l+k_{\boldsymbol{\beta}})}$, then we modify $u_{n(l+k_{\boldsymbol{\beta}})}$ such that $u_{n(l+k_{\boldsymbol{\beta}})} < \sup_{z \in \mathcal{D}} z^T \boldsymbol{\beta}_0$, and then we have $U^{\hat{\boldsymbol{\beta}}} = U^{\boldsymbol{\beta}_0} = u_{n(l+k_{\boldsymbol{\beta}})}$. Therefore, if necessary, we first modify the knots $\{u_{nk}\}_{k=0}^{k_n}$, so that there exists a neighborhood $\delta^*(\boldsymbol{\beta}_{0,-d}; r^*)$ of $\boldsymbol{\beta}_{0,-d}$ such that $U_{\boldsymbol{\beta}} = U_{\boldsymbol{\beta}_0}$, $U^{\boldsymbol{\beta}} = U^{\boldsymbol{\beta}_0}$ for $\boldsymbol{\beta} \in \delta^*(\boldsymbol{\beta}_{0,-d}; r^*)$ and $\hat{\boldsymbol{\beta}} \in \delta^*(\boldsymbol{\beta}_{0,-d}; r^*)$ for sufficiently large n. Let $K_n = K_{\boldsymbol{\beta}_0}$, $B_k(u) = B_{k\boldsymbol{\beta}_0}(u)$ and $\tilde{B}_k(u) = \tilde{B}_{k\boldsymbol{\beta}_0}(u)$. For $\boldsymbol{\beta} \in \delta^*(\boldsymbol{\beta}_{0,-d}; r^*)$, we have $K_{\boldsymbol{\beta}} = K_n$, $B_k(u) = B_{k\boldsymbol{\beta}}(u)$ and $\tilde{B}_k(u) = \tilde{B}_{k\boldsymbol{\beta}}(u)$. Further, we have

$$\begin{aligned} \tilde{G}_n(\boldsymbol{\alpha}, \boldsymbol{\beta}) &= \tfrac{1}{n} \textstyle\sum_{i=1}^{n} \left\{ \tilde{Y}_i - \tilde{W}_i^T \boldsymbol{\alpha} - \sum_{k=1}^{K_n} \tilde{b}_k(\boldsymbol{\alpha}, \boldsymbol{\beta}) \tilde{B}_k(Z_i^T \boldsymbol{\beta}) \right\}^2 \\ &= \tfrac{1}{n} \left\{ \tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}} \boldsymbol{\alpha} - \tilde{\boldsymbol{B}}(\boldsymbol{\beta}) \tilde{\boldsymbol{b}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \right\}^T \left\{ \tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}} \boldsymbol{\alpha} - \tilde{\boldsymbol{B}}(\boldsymbol{\beta}) \tilde{\boldsymbol{b}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \right\}. \end{aligned}$$

Set $G_n(\boldsymbol{\theta}_{-d}, \boldsymbol{b}) = G_n(\boldsymbol{\alpha}, \boldsymbol{\beta}_{-d}, \boldsymbol{b}) = \frac{1}{n} \left\{ \tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}} \boldsymbol{\alpha} - \tilde{\boldsymbol{B}}(\boldsymbol{\beta}_{-d}) \boldsymbol{b} \right\}^T \left\{ \tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}} \boldsymbol{\alpha} - \tilde{\boldsymbol{B}}(\boldsymbol{\beta}_{-d}) \boldsymbol{b} \right\}$,

where

$$\tilde{\boldsymbol{B}}(\boldsymbol{\beta}_{-d}) = \tilde{\boldsymbol{B}}(\beta_1, \ldots, \beta_{d-1}, \sqrt{1-(\beta_1^2+\ldots+\beta_{d-1}^2)}).$$

Since $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ is the minimizer of $\tilde{G}_n(\boldsymbol{\alpha}, \boldsymbol{\beta})$, then $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_{-d}, \hat{\boldsymbol{b}})$ is the minimizer of $G_n(\boldsymbol{\alpha}, \boldsymbol{\beta}_{-d}, \boldsymbol{b})$, where $\hat{\boldsymbol{b}} = \tilde{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}_{-d}) = \tilde{\boldsymbol{b}}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_{-d}) = \left\{\tilde{\boldsymbol{B}}^T(\hat{\boldsymbol{\beta}}_{-d})\tilde{\boldsymbol{B}}(\hat{\boldsymbol{\beta}}_{-d})\right\}^{-1}(\tilde{\boldsymbol{B}}^T\hat{\boldsymbol{\beta}}_{-d})(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\hat{\boldsymbol{\alpha}})$. Hence,

$$\left.\frac{\partial G_n(\boldsymbol{\alpha}, \boldsymbol{\beta}_{-d}, \boldsymbol{b})}{\partial \boldsymbol{\alpha}}\right|_{(\boldsymbol{\alpha}, \boldsymbol{\beta}_{-d}, \boldsymbol{b}) = (\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_{-d}, \hat{\boldsymbol{b}})} = -\frac{2}{n}\tilde{\boldsymbol{W}}^T\left\{\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{B}}(\hat{\boldsymbol{\beta}}_{-d})\hat{\boldsymbol{b}}\right\} = 0 \quad (3.6)$$

$$\left.\frac{\partial G_n(\boldsymbol{\alpha}, \boldsymbol{\beta}_{-d}, \boldsymbol{b})}{\partial \beta_r}\right|_{(\boldsymbol{\alpha}, \boldsymbol{\beta}_{-d}, \boldsymbol{b}) = (\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_{-d}, \hat{\boldsymbol{b}})} = -\frac{2}{n}\left\{\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{W}}\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{B}}(\hat{\boldsymbol{\beta}}_{-d})\hat{\boldsymbol{b}}\right\}^T \dot{\tilde{B}}_r(\hat{\boldsymbol{\beta}}_{-d})\hat{\boldsymbol{b}} = 0 \quad (3.7)$$

for $r = 1, \ldots, d-1$, where $\dot{\tilde{B}}_r(\boldsymbol{\beta}_{-d}) = \frac{\partial \tilde{\boldsymbol{B}}(\boldsymbol{\beta}_{-d})}{\partial \boldsymbol{\beta}_r}$. Set $\dot{G}_n(\boldsymbol{\theta}_{-d}, \boldsymbol{b}) = \frac{\partial G_n(\boldsymbol{\theta}_{-d}, \boldsymbol{b})}{\partial \boldsymbol{\theta}_{-d}} = (\frac{\partial}{\partial \boldsymbol{\alpha}} G_n(\boldsymbol{\alpha}, \boldsymbol{\beta}_{-d}, \boldsymbol{b})^T, \frac{\partial}{\partial \boldsymbol{\beta}_{-d}} G_n(\boldsymbol{\alpha}, \boldsymbol{\beta}_{-d}, \boldsymbol{b})^T)^T$. Then from (3.6) and (3.7) and using a Taylor expansion, we obtain

$$\dot{G}_n(\boldsymbol{\theta}_{0,-d}, \tilde{\boldsymbol{b}}(\boldsymbol{\theta}_{0,-d})) + \ddot{G}_n(\boldsymbol{\theta}^*_{-d}, \tilde{\boldsymbol{b}}(\boldsymbol{\theta}^*_{-d}))(\hat{\boldsymbol{\theta}}_{-d} - \boldsymbol{\theta}_{0,-d}) = 0, \quad (3.8)$$

where $\ddot{G}_n(\boldsymbol{\theta}_{-d}, \tilde{\boldsymbol{b}}(\boldsymbol{\theta}_{-d})) = \frac{\partial}{\partial \boldsymbol{\theta}_{-d}} \dot{G}_n(\boldsymbol{\theta}_{-d}, \tilde{\boldsymbol{b}}(\boldsymbol{\theta}_{-d}))$ is a $(q+d-1) \times (q+d-1)$ matrix and $\boldsymbol{\theta}^*_{-d}$ is between $\hat{\boldsymbol{\theta}}_{-d}$ and $\boldsymbol{\theta}_{0,-d}$. Let

$$\Omega_0 = (\varpi_{kr})_{(q+d-1)\times(q+d-1)}, \quad (3.9)$$

$$\varpi_{kr} = E(\breve{V}_k \breve{V}_r) - E[\boldsymbol{B}(Z^T\boldsymbol{\beta}_0)\breve{V}_k]^T \Gamma^{-1}(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0) E[\boldsymbol{B}(Z^T\boldsymbol{\beta}_0)\breve{V}_r], \quad k, r = 1, \ldots, q,$$

$$\varpi_{k(q+r)} = E[\dot{\boldsymbol{B}}_r(Z^T\boldsymbol{\beta}_0)\breve{V}_k]^T \boldsymbol{b}_0 - E[\boldsymbol{B}(Z^T\boldsymbol{\beta}_0)\breve{V}_k]^T \Gamma^{-1}(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0) H_r(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0)\boldsymbol{b}_0,$$

$\varpi_{(q+r)k} = \varpi_{k(q+r)}$ for $k, = 1, \ldots, q; r = 1, \ldots, d-1$, and

$$\varpi_{(q+k)(q+r)} = \boldsymbol{b}_0^T \left\{R_{rk}(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0) - H_r^T(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0)\Gamma^{-1}(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0) H_k(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0)\right\}\boldsymbol{b}_0$$

for $k, r = 1, \ldots, d-1$, where $\boldsymbol{B}(Z^T\boldsymbol{\beta}) = (B_1(Z^T\boldsymbol{\beta}), \ldots, B_{K_n}(Z^T\boldsymbol{\beta}))^T$, $\dot{\boldsymbol{B}}_r(Z^T\boldsymbol{\beta}) = \frac{\partial \boldsymbol{B}(Z^T\boldsymbol{\beta})}{\partial \beta_r}$, $\Gamma(\boldsymbol{\beta}, \boldsymbol{\beta}')$, $H_r(\boldsymbol{\beta}, \boldsymbol{\beta}')$ and $R_{rk}(\boldsymbol{\beta}, \boldsymbol{\beta}')$ are $K_n \times K_n$ matrices whose (l, l')th el-

ements are $E[B_l(Z^T\boldsymbol{\beta})B_{l'}(Z^T\boldsymbol{\beta}')]$, $E[B_l(Z^T\boldsymbol{\beta})\dot{B}_{l'r}(Z^T\boldsymbol{\beta}')]$ and $E[\dot{B}_{lr}(Z^T\boldsymbol{\beta})\dot{B}_{l'k}(Z^T\boldsymbol{\beta}')]$, respectively, and $\dot{B}_{lr}(Z^T\boldsymbol{\beta}) = \frac{\partial B_l(Z^T\boldsymbol{\beta})}{\partial \beta_r}$.

Theorem 3.2. Suppose that Assumptions 1 to 7 hold and that Ω_0 is invertible. Then we have

$$\sqrt{n}\Omega_0^{1/2}(\hat{\boldsymbol{\theta}}_{-d} - \boldsymbol{\theta}_{0,-d}) \to^{\mathcal{D}} N(0, \sigma^2 I_{q+d-1}), \tag{3.10}$$

where $\to^{\mathcal{D}}$ means convergence in distribution and I_{q+d-1} is the $(q+d-1)\times(q+d-1)$ identity matrix.

Next we establish the convergence rates of the estimators $\hat{a}(t)$ and $\hat{g}(u)$.

Theorem 3.3. Assume that Assumptions 1 to 7 hold. Further assume that the tuning parameter $\tilde{m}$ used in (2.13) satisfies $\tilde{m} \to \infty$ and $n^{-1}\tilde{m}^2\lambda_{\tilde{m}}^{-1}\log\tilde{m} \to 0$. Then

$$\int_{\mathcal{T}} \{\hat{a}(t) - a(t)\}^2\, dt = O_p\Big(\frac{\tilde{m}}{n\lambda_{\tilde{m}}} + \frac{\tilde{m}}{n^2\lambda_{\tilde{m}}^2}\sum_{j=1}^{\tilde{m}}\frac{j^3a_j^2}{\lambda_j^2} + \frac{1}{n\lambda_{\tilde{m}}}\sum_{j=1}^{\tilde{m}}\frac{a_j^2}{\lambda_j} + \tilde{m}^{-2\gamma+1}\Big). \tag{3.11}$$

If $\lambda_j \sim j^{-\delta}$, $\delta > 1$, $\tilde{m} \sim n^{1/(\delta+2\gamma)}$, $\gamma > 2$ and $\gamma > 1+\delta/2$, then $\sum_{j=1}^{\tilde{m}} j^3a_j^2\lambda_j^{-2} \le \bar{C}(\log\tilde{m} + \tilde{m}^{2\delta+4-2\gamma})$ and $\sum_{j=1}^{\tilde{m}} a_j^2\lambda_j^{-1} < +\infty$, where $\bar{C}$ is a positive constant. Then we have the following corollary.

Corollary 3.1. Under Assumptions 1 to 7, if $\lambda_j \sim j^{-\delta}$, $\delta > 1$, $\tilde{m} \sim n^{1/(\delta+2\gamma)}$ and $\gamma > \min(2, 1+\delta/2)$, then it follows that

$$\int_{\mathcal{T}} \{\hat{a}(t) - a(t)\}^2\, dt = O_p\left(n^{-(2\gamma-1)/(\delta+2\gamma)}\right). \tag{3.12}$$

The global convergence result (3.12) indicates that the estimator $\hat{a}(t)$ attains the same convergence rate as those of the estimators of Hall and Horowitz (2007), which are optimal in the minimax sense.

Remark 3.2. Note that the tuning parameter $\tilde{m}$ is used only to obtain the

estimator $\hat{a}(t)$ defined by (2.13). Whereas, the tuning parameter m is used in estimation of the unknown coefficient vectors $\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}_0$. Corollary 3.1 shows that the estimator $\hat{a}(t)$ attains the optimal convergence rate whenever $\tilde{m} \sim n^{1/(\delta+2\gamma)}$. From Remark 3.1 we note that the asymptotic normality of the estimator $\hat{\boldsymbol{\theta}}_{-d}$ can be derived whenever $m \sim n^{\iota}$ with $0 < \iota \leq n^{1/(\delta+2\gamma)}$. If $\tilde{m} = m$ then (3.11) still holds with $\tilde{m}$ replaced by m, provided Assumptions 1-7 hold.

Theorem 3.4. Suppose that Assumptions 1 to 8 hold. Then,

$$\int_{U_{\boldsymbol{\beta}_0}}^{U^{\boldsymbol{\beta}_0}} \{\hat{g}(u) - g(u)\}^2 \, du = O_p\left((nh)^{-1} + h^{2p}\right). \tag{3.13}$$

Further, if $h = O(n^{-1/(2p+1)})$ in Assumption 8, then

$$\int_{U_{\boldsymbol{\beta}_0}}^{U^{\boldsymbol{\beta}_0}} \{\hat{g}(u) - g(u)\}^2 \, du = O_p\left(n^{-2p/(2p+1)}\right). \tag{3.14}$$

The global convergence result (3.14) indicates that the estimator $\hat{g}(u)$ attains the optimal convergence rate.

Remark 3.3. Under Assumptions 1-7 and from a proof of similar to that of Theorem 3.4, one can obtain

$$\int_{U_{\boldsymbol{\beta}_0}}^{U^{\boldsymbol{\beta}_0}} \{\tilde{g}(u) - g(u)\}^2 \, du = O_p\left((nh_0)^{-1} + h_0^{2p}\right) = O_p\left((nh_0)^{-1}\right).$$

Due to the fact that $nh_0^{2p} \to 0$, $\tilde{g}(u)$ does not attain the global convergence rate of $O_p(n^{-2p/(2p+1)})$, which is the optimal rate for nonparametric models. In fact, the assumption that $nh_0^{2p} \to 0$ is made in order to make the bias of the estimator $\hat{\boldsymbol{\beta}}_{-d}$ in Theorem 3.2 negligible. This results in slower global convergence rate for the estimator $\tilde{g}(u)$.

Let $\mathcal{S} = \{(Y_i, X_i, W_i, Z_i) : i = 1, \ldots, n\}$. Suppose that $(Y_{n+1}, X_{n+1}, W_{n+1}, Z_{n+1})$ is a new vector of outcome and predictor variables taken from the same population as that of the data $\mathcal{S}$ and is independent of $\mathcal{S}$. Then the *mean squared*

prediction error (MSPE) of $\hat{Y}_{n+1}$ is given by

$$\begin{aligned}\text{MSPE} &= E\Big[\Big\{ \textstyle\int_{\mathcal{T}} \hat{a}(t)X_{n+1}(t)dt + W_{n+1}^T\hat{\boldsymbol{\alpha}} + \hat{g}(Z_{n+1}\hat{\boldsymbol{\beta}}) \\ &\quad - \big(\textstyle\int_{\mathcal{T}} a(t)X_{n+1}(t)dt + W_{n+1}^T\boldsymbol{\alpha}_0 + g(Z_{n+1}\boldsymbol{\beta}_0)\big) \Big\}^2 \Big| \mathcal{S}\Big].\end{aligned}$$

Theorem 3.5. Under Assumptions 1 to 4 and 6 to 8, if $\lambda_j \sim j^{-\delta}$, $\tilde{m} \sim n^{1/(\delta+2\gamma)}$, where $\gamma > \min(2, 1+\delta/2)$, $h_0 \sim n^{-\tau}$ with $1/(2p) < \tau < (\gamma-2)/(3(\delta+2\gamma))$ and $h = O(n^{-1/(2p+1)})$, then it follows that

$$\text{MSPE} = O_p\left(n^{-(\delta+2\gamma-1)/(\delta+2\gamma)}\right) + O_p(n^{-2p/(2p+1)}\Big). \tag{3.15}$$

Furthermore, if $\delta + 2\gamma = 2p+1$ then

$$\text{MSPE} = O_p\left(n^{-(\delta+2\gamma-1)/(\delta+2\gamma)}\right). \tag{3.16}$$

Remark 3.4. In Theorem 3.5, it is assumed that $h_0 \sim n^{-\tau}$ and $1/(2p) < \tau < (\gamma-2)/(3(\delta+2\gamma))$. If $\delta+2\gamma = 2p+1$, then the conditions that $p > \gamma$ and $\gamma > 5 + 3/(2p)$ are required. The preceding conditions hold when $p > \gamma \geq 5.3$.

4. Simulation results

In this section we present two Monte Carlo simulation studies to evaluate the finite-sample performance of the proposed estimator. The data are generated from the following models

$$Y_i = \int_{\mathcal{T}} a(t)X_i(t)dt + \alpha_0 W_i + \sin\left(\pi(Z_i^T\boldsymbol{\beta}_0 - E)/(F-E)\right) + \varepsilon_i, \tag{4.1}$$

$$Y_i = \int_{\mathcal{T}} a(t)X_i(t)dt + \alpha_1 W_{i1} + \alpha_2 W_{i2} - 2Z_i^T\boldsymbol{\beta}_0 + 5 + \varepsilon_i, \tag{4.2}$$

with $\mathcal{T} = [0,1]$ and the trivariate random vectors Z_i's have independent components following the uniform distribution on $[0,1]$. In model (4.1), $\alpha_0 = 0.3$, $\boldsymbol{\beta}_0 = (1,1,1)^T/\sqrt{3}$, $E = \sqrt{3}/2 - 1.645/\sqrt{12}$ and $F = \sqrt{3}/2 + 1.645/\sqrt{12}$. We let $W_i = 0$ for odd i and $W_i = 1$ for even i, and the ε_i's are independent errors

following $N(0, 0.5^2)$. We take $a(t) = \sum_{j=1}^{50} a_j\phi_j(t)$ and $X_i(t) = \sum_{j=1}^{50} \xi_{ij}\phi_j(t)$, where $a_1 = 0.3$ and $a_j = 4(-1)^{j+1}j^{-2}$, $j \geq 2$; $\phi_1(t) \equiv 1$ and $\phi_j(t) = 2^{1/2}\cos((j-1)\pi t)$, $j \geq 2$; the ξ_{ij}'s are independently and normally distributed with $N(0, j^{-\delta})$. In model (4.2), $\alpha_1 = -2$, $\alpha_2 = 1.5$, $\boldsymbol{\beta}_0 = (1, 2, 2)^T/3$ and $X_i(t) = \sum_{j=1}^{50} \xi_{ij}\phi_j(t)$, the ξ_{ij}'s are independently and normally distributed with $N(0, \lambda_j)$, where $\lambda_1 = 1$, $\lambda_j = 0.22^2(1 - 0.0001j)^2$ if $2 \leq j \leq 4$, $\lambda_{5j+k} = 0.22^2((5j)^{-\delta/2} - 0.0001k)^2$ for $j \geq 1$ and $0 \leq k \leq 4$. Further, $W_{ik} = \check{W}_{ik} + V_{ik}$ and $\check{W}_{ik} = \sum_{j=1}^{50} kj^{-2}\xi_{ij}$ for $k = 1, 2$. The V_{ik}'s are independently and normally distributed with $N(-1, 2^2)$ and $N(2, 3^2)$, respectively, and independent of ξ_{ij}. Finally, the error ε_i's in (4.2) are independent $N(0, 1)$ random variables.

For the functional linear part of model (4.1), the eigenvalues of the operator K are well-spaced, while the latter part of model (4.1) was investigated by Carroll et al. (1997) and Yu and Ruppert (2002). In model (4.2), the eigenvalues of the operator K are closely spaced, while the link function $g(u) = -2u + 5$ is a linear function. All our results are reported based on the average over 500 replications for each setting. In each sample, we first use a linear function to replace $g(u)$ and use the least squares estimates for the partial functional linear model as an initial estimator. The function $g(u)$ is approximated using a cubic spline with equally spaced knots. We note from our simulation results (see Table 3) that parametric estimators are not sensitive to the choices of parameters m and h_0 which is $O(k_n^{-1})$. Here we take $m = 5$ and $h_0 = cn^{-1/5}$ with $c = 1$. When we compute the estimators of $g(u)$ and $a(t)$, the parameter K_n and the tuning parameter m are selected respectively by the BIC given in Section 2.

Table 1 reports the biases and standard deviations (sd) of the estimators $\hat{\alpha}_0$, $\hat{\boldsymbol{\beta}}_0 = (\hat{\beta}_{01}, \hat{\beta}_{02}, \hat{\beta}_{03})^T$ obtained by the proposed method in Section 2 and the

mean integrated squared errors (MISE) of the estimators $\hat{g}(u)$ and $\hat{a}(t)$ for model (4.1) based on $\delta = 1.5$ and sample sizes $n = 100$, 200. Figure 1 displays the true curves and the mean estimated curves over 500 simulations with sample size $n = 100$ of $g(u)$, $a(t)$ and their 95% pointwise confidence bands. Table 2 reports the biases and standard deviations (sd) of the estimators $\hat{\alpha}_k$ for $k = 1, 2$ and $\hat{\boldsymbol{\beta}}_1 = (\hat{\beta}_{11}, \hat{\beta}_{12}, \hat{\beta}_{13})^T$, and the mean integrated squared errors (MISE) of the estimators $\hat{g}(u)$ and $\hat{a}(t)$ for model (4.2) with $\delta = 1.5$ and $n = 100, 200$. For comparison purposes, Tables 1 and 2 also list the simulation results based on the least squares partial functional linear (LSPFL) estimators, which are obtained by using a linear function to approximate the link function g. Further, Table 1 also lists the simulation results based on the nonlinear least squares (ORACLE) estimation method when the exact form of sinusoidal model is known.

Table 1. Results of Monte Carlo experiments for model (4.1).

		n=100			n=200		
		LSPFL	ORACLE	Proposed method	LSPFL	ORACLE	Proposed method
$\hat{\alpha}_0$	bias	−0.0019	0.0034	−0.0008	−0.0025	0.0002	0.0002
	sd	0.0836	0.0330	0.0307	0.0565	0.0159	0.0122
$\hat{\beta}_{01}$	bias	−0.3678	−0.0066	−0.0056	−0.3365	−0.0037	0.0006
	sd	0.5445	0.0441	0.0464	0.5141	0.0202	0.0206
$\hat{\beta}_{02}$	bias	−0.3780	−0.0075	−0.0031	−0.3283	−0.0041	−0.0018
	sd	0.5449	0.0457	0.0479	0.5201	0.0263	0.0178
$\hat{\beta}_{03}$	bias	−0.0771	0.0082	0.0016	−0.0553	0.0058	-0.0001
	sd	0.2695	0.0506	0.0599	0.2694	0.0307	0.0239
$\hat{g}(u)$	MISE			0.0090			0.0007
$\hat{a}(t)$	MISE	0.1205	0.0189	0.0218	0.0756	0.0082	0.0084

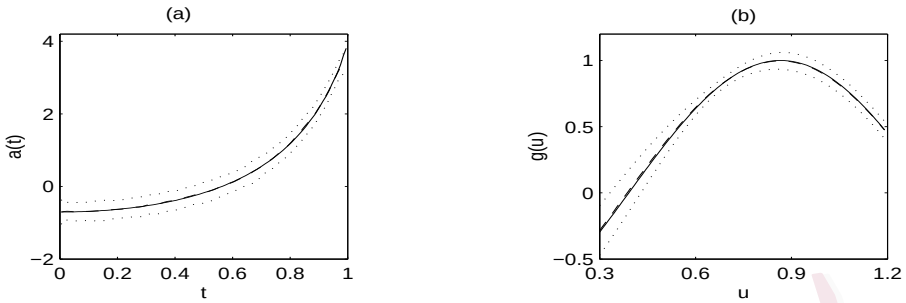

Figure 1. The actual and the mean estimated curves for $g(u)$ and $a(t)$ in model (4.1) with $n = 100$ and the 95% pointwise confidence bands. (a) is the figure for $a(t)$ and (b) is the figure for $g(u)$. —, true curves; - - -, mean estimated curves; ..., 95% pointwise confidence bands.

Table 2. Results of Monte Carlo experiments for model (4.2). The biases ($\times 10^{-4}$) and sds ($\times 10^{-4}$) of parametric estimators and MISE ($\times 10^{-4}$) of $\hat{g}(u)$ and MISE of $\hat{a}(t)$.

		n=100		n=200	
		LSPFL	Proposed method	LSPFL	Proposed method
$\hat{\alpha}_1$	bias (sd)	0.078(6.815)	0.100(6.870)	0.186(4.415)	0.173(4.435)
$\hat{\alpha}_2$	bias (sd)	−0.071(4.612)	−0.085(4.666)	0.359(3.038)	0.373(3.056)
$\hat{\beta}_{11}$	bias (sd)	−0.707(22.725)	−0.753 (23.162)	0.942(14.762)	0.816(14.896)
$\hat{\beta}_{12}$	bias (sd)	−1.670(18.370)	−1.720(18.347)	0.711(11.939)	0.735(11.906)
$\hat{\beta}_{13}$	bias (sd)	1.936(17.630)	2.007(17.655)	−1.220(11.944)	−1.181(11.919)
$\hat{g}(u)$	MISE		3.852		2.503
$\hat{a}(t)$	MISE	0.0087	0.0096	0.0047	0.0044

We observe from Table 1 that the least squares partial functional linear (LSPFL) method gives poor estimates, while our proposed estimates are far more accurate than the LSPFL estimates, and they can be as accurate as those obtained from the ORACLE when the exact form of sinusoidal model is known. Figure 1 shows that the difference between the true curves and the mean estimated curves are barely visible, and it shows that the bias is very small in the estimates. Furthermore, the 95% pointwise confidence bands are reasonably close to the true curve, showing a very little variation in the estimates. Table 2 shows that, even if the unknown link function $g(u)$ is a linear function, our proposed estimates behave as good as the least squares partial functional linear estimates.

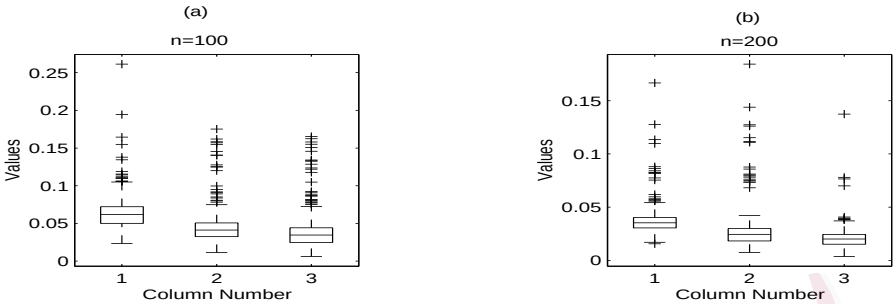

Figure 2. Boxplots of MAE for model (4.1). Label 1 is boxplot for $\delta = 1.1$, 2 is boxplot for $\delta = 1.5$ and 3 is boxplot for $\delta = 2$.

Both tables indicate that the proposed method yields accurate estimates and outperforms the least squares partial functional linear estimates when the link function is nonlinear, and it is comparable to the least squares partial functional linear estimates when the link function is a linear function.

To study the prediction performance of the proposed method, we generated samples of $n = 100, 200$ from models (4.1) and (4.2) with $\delta \in \{1.1, 1.5, 2\}$ for estimation, where δ is related to the eigenvalue of the operator with kernel K. We also generated test samples of size 300 to compute the prediction mean absolute error (MAE) defined by $MAE = \frac{1}{N}\sum_{i=1}^{N} |\tilde{Y}_{n+i} - \hat{Y}_{n+i}|$, where $\tilde{Y}_{n+i} = \int_{\mathcal{T}} a(t)X_{n+i}(t)dt + W_{n+i}^T\boldsymbol{\alpha}_0 + g_0(Z_{n+i}^T\boldsymbol{\beta}_0)$ and $\hat{Y}_{n+i} = \int_{\mathcal{T}} \hat{a}(t)X_{n+i}(t)dt + W_{n+i}^T\hat{\boldsymbol{\alpha}} + \hat{g}(Z_{n+i}^T\hat{\boldsymbol{\beta}})$. Figures 2 and 3 display the boxplots of MAE based on 500 replications and $N = 300$. We observe that the proposed method shows good prediction performances for both models and the MAEs are quite small even if $n = 100$. Figures 2 and 3 also show that the MAE decreases as n increases or as δ increases.

For different m and h_0, Table 3 exhibits the MSEs of the estimators $\hat{\alpha}_0$ and $\hat{\beta}_{01}$ for model (4.1) with $\delta = 1.5$ and sample size $n = 200$. We observe from Table

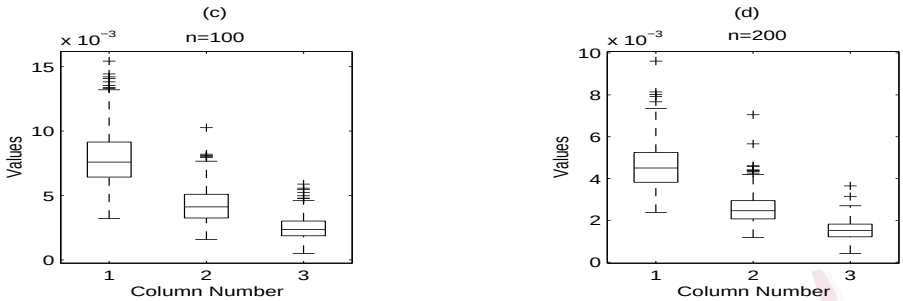

Figure 3. Boxplots of MAE for model (4.2). 1 is boxplot for $\delta = 1.1$, 2 is boxplot for $\delta = 1.5$ and 3 is boxplot for $\delta = 2$.

3 that MSEs of $\hat{\alpha}_0$ and $\hat{\beta}_{01}$ are not very sensitive to the change of m and h_0, and the estimators of α_0 and β_{01} are efficient under a broad range of values for m and h_0. The MSEs of $\hat{\beta}_{02}$ and $\hat{\beta}_{03}$ also show similar behaviors and are omitted here.

Table 3. MSE ($\times 10^{-3}$) of $\hat{\alpha}_0$ and $\hat{\beta}_{01}$ in model (4.1). The sample size is $n = 200$.

	h_0	$m=2$	$m=3$	$m=4$	$m=5$	$m=6$	$m=7$	$m=8$	$m=9$	$m=10$
$\hat{\alpha}_0$	0.2	1.4	0.9	0.5	0.3	0.4	0.5	0.6	0.4	0.6
	0.3	1.4	0.8	0.3	0.3	0.4	0.4	0.6	0.2	0.3
	0.4	1.4	0.8	0.3	0.2	0.4	0.3	0.6	0.4	0.6
	0.5	1.4	0.6	0.3	0.3	0.3	0.4	0.6	0.6	0.4
$\hat{\beta}_{01}$	0.2	1.8	2.0	1.0	0.5	0.9	1.1	0.8	0.9	1.6
	0.3	1.5	1.5	0.7	0.6	1.4	0.3	1.1	0.7	1.4
	0.4	1.5	1.5	0.7	0.6	1.1	0.3	1.1	1.1	1.4
	0.5	2.5	1.6	1.5	0.8	0.6	1.0	1.6	1.2	1.0

5. Real data example

In this section we analyze a real data set using the proposed method. For this purpose we use the diffusion tensor imaging (DTI) data with 217 subjects from the NIH Alzheimer's Disease Neuroimaging Initiative (ADNI) study. For more information on how this data were collected etc., see http://www.adni-info.org. The DTI data were processed by two key steps including a weighted least squares estimation method (Basser et al. 1994, Zhu et al. 2007) to construct the diffusion

tensors and a TBSS pipeline in FSL (Smith et al. 2006) to register DTIs from multiple subjects to create a mean image and a mean skeleton. This data have been recently analyzed by many authors using different models; see, e.g., Yu et al. (2016), Li et al. (2016) and the references therein.

Our interest is to predict mini-mental state examination (MMSE) scores, one of the most widely used screening tests to provide brief and objective measures of cognitive functioning for a long time. The MMSE scores have been seen as a reliable and valid clinical measure quantitatively assessing the severity of cognitive impairment. It was believed that the MMSE scores to be affected by demographic features such as age, education and cultural background (Tombaugh and McIntyre 1992), gender (Pöysti et al. 2012, O'Bryant et al. 2008), and possibly some genetic factors, for example, AOPE polymorphic alleles (Liu et al. 2013).

After cleaning the raw data that failed in quality control or had missing data, we include totally 196 individuals in our analysis. The response of interest Y is the MMSE scores. The functional covariate is fractional anisotropy (FA) values along the corpus callosum (CC) fiber tract with 83 equally spaced grid points, which can be treated as a function of arc-length. FA measures the inhomogeneous extent of local barriers to water diffusion and the averaged magnitude of local water diffusion (Basser et al. 1996). The scalar covariates of primary interests include gender (W_1), handedness (W_2), education level (W_3), genotypes for apoe4 (W_4, W_5, categorical data with 3 levels), age (W_6), ADAS13 (Z_1) and ADAS11 (Z_2). The genotypes apoe4 is one of three major alleles of apolipoprotein E (ApoE), a major cholesterol carrier that supports lipid transport and injury repair in the brain. ApoE polymorphic alleles are the main genetic determinants of Alzheimer disease risk (Liu et al. 2013). ADAS11 and ADAS13 are respectively

the 11-item and 13-item versions of the Alzheimer's Disease Assessment Scale-Cognitive subscale (ADAS-Cog), which were originally developed to measure cognition in patients within various stages of Alzheimer's Disease (Llano et al. 2011, Zhou et al. 2012, Podhorna et al. 2016).

We study the following two models

$$\begin{aligned} Y &= \int_0^1 a(t)X(t)dt + \alpha_0 + \alpha_1 W_1 + \alpha_2 W_2 + \alpha_3 W_3 + \alpha_4 W_4 + \alpha_5 W_5 \\ &\quad + \alpha_6 W_6 + \beta_1 Z_1 + \beta_2 Z_2 + \varepsilon, \end{aligned} \tag{5.1}$$

$$\begin{aligned} Y &= \int_0^1 a(t)X(t)dt + \alpha_1 W_1 + \alpha_2 W_2 + \alpha_3 W_3 + \alpha_4 W_4 + \alpha_5 W_5 \\ &\quad + \alpha_6 W_6 + g(\beta_1 Z_1 + \beta_2 Z_2) + \varepsilon, \end{aligned} \tag{5.2}$$

where $W_1 = 1$ stands for male and $W_1 = 0$ stands for female, $W_2 = 1$ denotes right-handed and $W_2 = 0$ denotes left-handed, $W_4 = 1$ and $W_5 = 0$ indicates type 0 for apoe4, $W_4 = 0$ and $W_5 = 1$ indicates type 1 for apoe4 and both $W_4 = 0$ and $W_5 = 0$ indicates type 2 for apoe4. The functional component $X(t)$ is chosen as the centered fractional anisotropy (FA) values so that $E[X(t)] = 0$. Model (5.1) is a partial functional linear model, while model (5.2) is partial functional linear single index model in which ADAS13 (Z_1) and ADAS11 (Z_2) are index variables.

Table 4. The parametric estimators for models (5.1) and (5.2).

model	α_1	α_2	α_3	α_4	α_5	α_6	β_1	β_2
(5.1)	0.0758	0.4317	0.1105	0.6875	0.5581	-0.0239	-0.0429	-0.1865
(5.2)	-0.0754	0.1814	0.1138	0.5961	0.5245	-0.0305	0.1957	0.9807

The parametric and nonparametric components in the models are computed by the procedure given in Section 2, with the nonparametric function $g(u)$ being approximated by a cubic spline with equally spaced knots. Since the values of Z_1 and Z_2 are large, we choose $h_0 = 5.0$ for model (5.2) and $m = 3$ for parametric estimation. Table 4 exhibits the parametric estimators, and Figure 4 shows the estimated curves of $a(t)$ and $g(u)$. For model (5.1), $\hat{\alpha}_0 = 28.9388$. The MSE

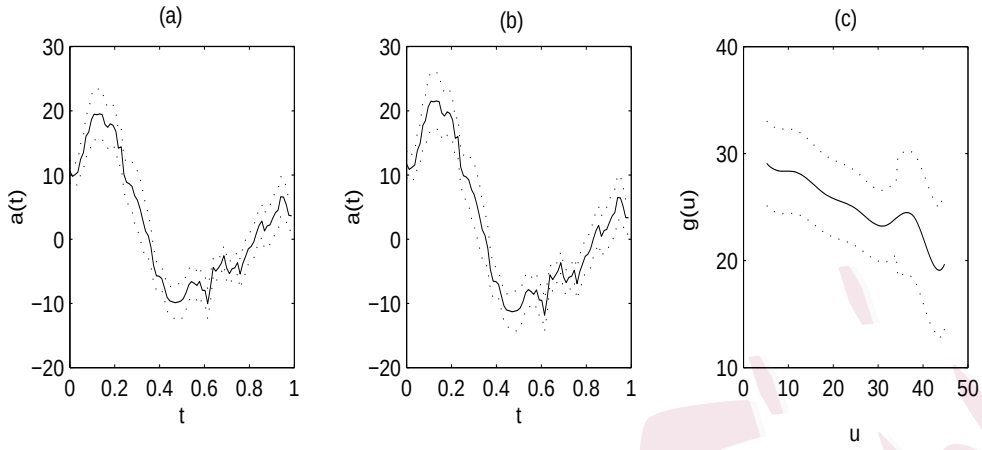

Figure 4. The solid lines are the estimated curves of $a(t)$ in model (5.1) (a), $a(t)$ in model (5.2) (b), and $g(u)$ in model (5.2) (c); The doted lines are their corresponding 95% point-wise confidence intervals.

of Y for models (5.1) and (5.2) are 2.8684 and 2.7782, respectively, and can be further reduced for model (5.2) as the number of knots increases.

From Table 4 and Figure 4, we observe that in both models MMSE is decreasing in terms of ADAS13 and ADAS11. However, in Figure 4 this decline is found to be nonlinear evidenced by the nonlinear trends of $g(u)$ in model (5.2). In single index models (5.2), we found that MMSE is higher for female than male, which is consistent with the results in the literature (Pöysti et al. 2012, O'Bryant et al. 2008), while model (5.1) incorrectly finds the opposite. Although we may not able to perform a formal test on model fitting, these observations show the superiority of the single index model (5.2).

To evaluate the prediction performance of the two models, we applied a combination of the bootstrap and the cross-validation method to the data set. For each bootstrap sample, we randomly divided the data into ten partitions. Since the number of individuals is not large, we used nine folds of the data to

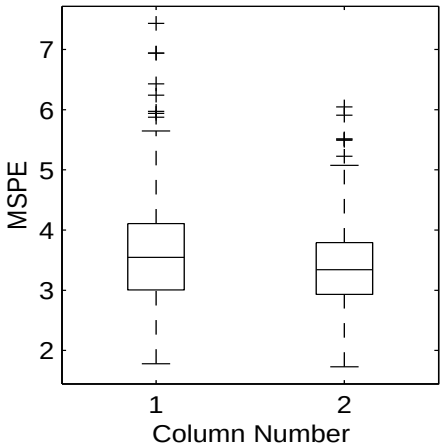

Figure 5. Boxplots for the mean squared prediction error (MSPE) for three models. Label 1 is boxplot for model (5.1) and 2 is boxplot for model (5.2).

estimate the model and the remaining fold for the testing data set. We calculated the mean squared prediction error (MSPE) for the testing data set. The MSPEs for the two models over the 200 replications are reported as boxplots in Figure 5. The means for MSPEs of the 200 replications for models (5.1) and (5.2) are 3.6996 and 3.4249, respectively. The medians for MSPEs of the 200 replications for models (5.1) and (5.2) are 3.5464 and 3.3421, respectively. This figure shows that model (5.2) fits the data better than model (5.1). We also calculated 95% point-wise confidence intervals of the estimated curves of $a(t)$ in model (5.1), $a(t)$ in model (5.2), and $g(u)$ in model (5.2), which are shown as (a), (b) and (c), respectively, in Figure 4. From Figure 4, it is evidenced that the functional slope for both models are very similar in shape, while $g(u)$ has a clear nonlinear feature. This also confirms that model (5.2) is more flexible than model (5.1).

Supplementary Material

Title: Supplementary Material for *Partial Functional Partially Linear Single-*

Index Models (a PDF file). There we give the proofs of Theorems 3.1 to 3.5, which depend on a number of preliminary lemmas.

Acknowledgments. The authors sincerely thank the Co-Editor Dr. Yazhen Wang, Associate Editor, and two referees for their helpful comments given toward improvement of the paper.

Q. Tang's research was supported by the National Social Science Foundation of China (16BTJ019) and Natural Science Foundation of Jiangsu Province of China (Grant No. BK20151481). Research of L. Kong was supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canadian Statistical Sciences Institute Collaborative Research Team (CANSSI-CRT). He also thanks the support of the Program on Challenges in Computational Neuroscience (CCNS) at the Statistical and Applied Mathematical Sciences Institute (SAMSI) during his visit in 2016. D. Ruppert's research was supported by NSF grant AST-1312903 and NIH grants P30 AG010129, K01 AG030514. R.J. Karunamuni's research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

Part of data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affilated company Genentech, Inc.; GE Health-

care; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles.

References

Basser, P. J., Mattiello, J. and LeBihan, D. (1994). MR diffusion tensor spectroscopy and imaging. *Biophysical journal*, **66** (1), 259–267.

Basser, P.J., and Pierpaoli, C. (1996). Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *Journal of Magnetic Resonance, Series B,* **111**, 209–219.

de Boor, C. (1978). *A Practical Guide to Splines.* Springer, New York.

Cai, T. T. and Hall, P. (2006). Prediction in functional linear regression. *Ann. Statist.,* **34**, 2159–2179.

Carroll, R. J., Fan, J., Gijbels, I. and Wand, M. P. (1997). Generalized partially linear single-index models. *J. Amer. Statist. Assoc.,* **92,** 477-489.

Cardot, H., Mas, A. and Sarda, P. (2007). CLT in functional linear models. *Probab. Theory Relat. Fields,* **138**, 325–361.

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

Chen, K. and Müller, H.-G. (2012). Conditional quantile analysis when covariates are functions, with application to growth data. *J. R. Statist. Soc. B,* **74**, 67–89.

Delaigle, A. and Hall, P. (2012). Methodology and theory for partial least squares applied to functional data. *Ann. Statist.,* **40,** 322–352.

Ferraty, F. and Vieu, P. (2006), *Nonparametric Functional Data Analysis.* Springer, New York.

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

Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications.* Springer, New York.

Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators.* Wiley, New York.

Kengo Kato (2012). Estimation in functional linear quantile regression. *Ann. Statist.* **40**, 3108–3136.

Kong, D., Xue, K., Yao, F. and Zhang, H. H. (2016). Partially functional linear regression in high dimensions. *Biometrika*, **103**, 147–159.

Li, J., Huang, C. and Zhu, H. (2016). A functional varying-coefficient single index model for functional response data. *J. Amer. Statist. Assoc.,* **112,** 1169–1181.

Li, Y. and Hsing, T. (2007). On the rates of convergence in functional linear regression. *J. Multivariate Anal.,* **98,** 1782–1804.

Li, Y., Wang, N. and Carroll, R. J. (2010). Generalized functional linear models with semiparametric single-index interactions. *J. Amer. Statist. Assoc.,* **105,** 621–633.

Liang, H., Liu, X., Li, R. and Tsai, C. L. (2010). Estimation and testing for partially linear single-index models. *Ann. Statist.,* **38,** 3811–3836.

Llano, D. A., Laforet, G. and Devanarayan, V. (2011). Derivation of a new ADAS-cog composite using tree-based multivariate analysis: prediction of conversion from mild cognitive impairment to Alzheimer disease. *Alzheimer Disease & Associated Disorders*, **25**(1), 73–84.

Liu, C. C., Kanekiyo, T., Xu, H. and Bu, G. (2013). Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nature Reviews Neurology,* **9**(2), 106–118.

Ma, S. (2016). Estimation and inference in functional single-index models. *Ann. Inst. Stat. Math.,* **68**, 181–208.

O'Bryant, S. E., Humphreys, J. D., Smith, G. E., Ivnik, R. J., Graff-Radford, N. R., Petersen, R. C. and Lucas, J. A. (2008). Detecting dementia with the mini-mental state examination (MMSE) in Highly Educated Individuals. *Archives of Neurology*, **65**(7), 963–967.

Podhorna, J., Krahnke, T., Shear, M. and Harrison, J. (2016). AlzheimerÕs disease assessment scale cognitive subscale variants in mild cognitive impairment and mild AlzheimerÕs disease: change over time and the effect of enrichment strategies. *Alzheimer's research & therapy*, **8**(1), 8.

Pöysti, M. M., Laakkonen, M. L., Strandberg, T., Savikko, N., Tilvis, R. S., Eloniemi-Sulkava, U. and Pitkälä, K. H. (2012). Gender differences in dementia spousal caregiving. *Inter. J. Alzheimer's disease*, http://dx.doi.org/10.1155/2012/162960.

Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. Springer, New York.

Ramsay, J. O. and Silverman, B. W. (2005) *Functional Data Analysis*. Springer, New York.

Schumaker, L. L. (1981). *Spline Functions: Basic Theory*. Wiley, New York.

Shin, H. (2009). Partial functional linear regression. *J. Statist. Plan. Infer.*, **139**, 3405–3418.

Shin, H. and Lee, M. H. (2012). On prediction rate in partial functional linear regression. *J. Multivariate Anal.*, **103**, 93–106.

Smith, S. M., Jenkinson, Johansen-Berg, Rueckert, D., Nichols, T. E., Mackay, C. E., Watkins, K. E., Ciccarelli, O., Cader, M. Z., Matthews, P. M., et al. (2006). Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroimage*, **31**(4), 1487–1505.

Tang, Q. (2013). B-spline estimation for semiparametric varying-coefficient partially linear regression with spatial data. *J. Nonparam. Stat.*, **25**, 361–378.

Tang, Q. (2015). Estimation for semi-functional linear regression. *Statistics*, **49,** 1262–1278.

Tang, Q. and Cheng, L. (2014). Partial functional linear quantile regression. *Science China Mathematics,* **57**(12), 2589–2608.

Tombaugh, T. N. and McIntyre, N.J. (1992). The mini-mental state examination:a comprehensive review. *J. Amer. Geriatr. Soc.,* **40**, 922–935.

Wang, Guochang, Fang, Xiang-Nan, and Chen, Min (2016). Functional partial linear single index model. *Scand. J. Statist.,* **43**, 261–274.

Wang, J. L., Xue, L. G., Zhu, L. X. and Chong, Y. S. (2010). Partial-linear single-index model with noised variable. *Ann. Statist.,* **38,** 246-272.

Xia, Y. C. and Härdle, W. (2006). Semi-parametric estimation of partially linear single-index models. *J. Multivariate Anal.,* **97,** 1162–1184.

Yao, F., Sue-Chee, S. and Wang, F. (2017). Regularized partially functional quantile regression. *J. Multivariate Anal.,* **156,** 39–56.

Yu, D. Kong, L. and Mizera, I. (2016). Partial functional linear quantile regression for neuroimaging data analysis. *Neurocomputing*, **195**, 74–87.

Yu, Y. and Ruppert, D. (2002). Penalized spline estimation for partially linear single-index models. *J. Amer. Statist. Assoc.,* **97,** 1042–1054.

Yuan, M. and Cai, T. T. (2010). A reproducing kernel hilbert space approach to functional linear regression. *Ann. Statist.,* **38,** 3412–3444.

Zhang, D., Lin, X. and Sowers, M. F. (2007). Two-stage functional mixed models for evaluating the effect of longitudinal covariate profiles on a scalar outcome. *Biometrics*, **63**, 351–362.

Zhu, H., Zhang, H., Ibrahim, J. G. and Peterson, B. S. (2006). Statistical analysis of diffusion tensors in diffusion-weighted magnetic resonance imaging data. *J. Amer. Statist. Assoc.,* **102**, 1085–1102.

Zhu, H., Li, R. and Kong, L. (2012). Multivariate varying coefficient model for functional responses. *Ann. Statist.,* **40,** 2634–2666.

Zhou, B., Nakatani, E., Teramukai, S., Nagai, Y. and Fukushima, M. (2012). Risk classification in mild cognitive impairment patients for developing Alzheimer's disease. *J. Alzheimer's Disease*, **30**(2), 367–375.

Zhou, H., Li, L. and Zhu, H. (2013). Tensor regression with applications in neuroimaging data analysis. *J. Amer. Statist. Assoc.,* **108,** 540–552.

School of Economics and Management, Nanjing University of Science and Technology, Nanjing, China

E-mail: (tangqguo@163.com)

Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta, Canada

E-mail: (lkong@ualberta.ca)

School of Operations Research and Information Engineering, Cornell University, Ithaca, NY 14853 USA

E-mail: (dr24@cornell.edu)

Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta, Canada

E-mail: (R.J.Karunamuni@ualberta.ca)