Statistica Sinica **32** (2022), 1789-1809
doi:https://doi.org/10.5705/ss.202021.0025

ESTIMATION FOR FUNCTIONAL SINGLE INDEX MODELS WITH UNKNOWN LINK FUNCTIONS

Yunxiang Huang and Qihua Wang

Chinese Academy of Sciences

Abstract: This study examines an estimating problem in single index models with functional predictors. An estimating approach is developed to estimate the slope function in the single-index and the nonparametric link function. Optimal convergence rates for the estimator of the slope function are established in a minimax sense, under mild conditions, using a functional principal component analysis and the estimating equation technique. For the estimator of the nonparametric link function, both the uniform and mean squared convergence rates are obtained. An error variance estimator is also defined and is proved to be asymptotically normal. The finite-sample performance of the proposed estimators is illustrated by simulations and a real-data application.

Key words and phrases: Functional data analysis, functional principal components analysis, kernel smoother, local linear smoothing, nonparametric models, semiparametric models.

1. Introduction

Functional data analysis has received substantial interest in recent decades. There have been extensive studies on functional linear models with a scalar response (Cai and Hall (2006); Hall and Horowitz (2007); Yuan and Cai (2010); Cai and Yuan (2012); Delaigle and Hall (2012); Hilgert, Mas and Verzelen (2013)) and nonlinear models with known link functions (James (2002); Dou, Pollard and Zhou (2012)). For nonparametric models with functional predictors, Ferraty and Vieu (2006) and Goia and Vieu (2016) provide comprehensive discussions on this topic. However, because functional data are inherently infinite dimensional, the statistical performance of full nonparametric methods is unfavorable owing to the so-called curse of dimensionality.

To avoid the curse, we model scalar responses with functional covariates as single-index models, which include a simple linear term and a flexible nonparametric link function. Early works assumed that either the link function is monotonic (Müller and Stadtmüller (2005)) or that the slope function lies in a

Corresponding author: Qihua Wang, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100080, People's Republic of China. E-mail: qhwang@amss.ac.cn.

predefined index set (Ait-Saïdi et al. (2008)), limiting the applicability of those methods. Chen, Hall and Müller (2011) considered single-index functional regression models with unknown slope functions and general nonparametric link functions. They proposed a direct kernel-based estimating method and established a polynomial rate of convergence of the estimator. However, as noted by the authors, the convergence rate for the slope function in Chen, Hall and Müller (2011) is not optimal. Therefore, it is natural to ask whether we can develop a new estimating approach such that the convergence rate of the slope function estimator in functional single-index models can achieve the rate of functional linear models. For an alternative approach based on sufficient dimension reduction with functional predictors, see Ferré and Yao (2003), Ferré and Yao (2005), Cook, Forzani and Yao (2010), Jiang, Yu and Wang (2014), Yao, Lei and Wu (2015), Yao, Wu and Zou (2016), and Zhang, Wang and Wu (2018). Although we can apply those approaches to functional single-index models, it may be difficult to obtain the optimal rate.

In this paper, we introduce a functional version of the kernel estimators considered in Wang et al. (2010) and Cui, Härdle and Zhu (2011). The most attractive feature of the proposed estimator is that the convergence rate of the slope function estimator and that of the mean squared prediction error of the scalar single index are optimal in a minimax sense. The estimating procedure includes two steps. We first use a functional principal component analysis (FPCA) to approximate the slope function using the leading empirical functional principal components (FPCs). For more details about the FPCA, see, for example, Ramsay and Silverman (2005), Horváth and Kokoszka (2012), Hsing and Eubank (2015). Next, we use an alternating estimation procedure to estimate the slope function and the link function. The proposed estimator is based on estimating equation methods that take advantage of the unitary constraint on the slope terms of the scalar single index and have a close relationship with the estimators in Wang et al. (2010) and Cui, Härdle and Zhu (2011) for single-index models. However, there are three main differences between the proposed method and the approaches in Wang et al. (2010); Cui, Härdle and Zhu (2011). First, the number of FPC scores in our estimation procedure diverges as the sample size increases. Second, we include the variances of the FPC scores in the estimating equations. Third, we truncate the denominators of the kernel estimators to avoid those closing to zero. These modifications help us to develop the optimal convergence rate of the slope function estimator under mild conditions. We obtain the uniform convergence rate and the mean squared convergence rate of the link function estimation, where the latter is faster than that in Chen, Hall and Müller (2011).

We also provide the asymptotic normality of the error variance estimator.

The proof is challenging from a technical point of view. The approximation step for the slope function leads to a cut-off bias. A plug-in error is also introduced, because the true FPC basis functions and scores are not known, and have to be replaced by their estimates. To overcome these difficulties, we introduce a truncated version of the slope function that satisfies the unitary constraint, which serves as a bridge between in the proof the true underlying slope function and its estimator. The proof of the asymptotic results is based on the theory of M-estimators with diverging numbers of covariates (e.g., see, Wang (2011)). We also carefully verify whether the existing results for kernel estimators and single-index models remain valid with a diverging number of covariates. Additionally, we do not require the compact support assumption that the scalar single index is defined on a compact support, which is widely used in the literature on kernel smoothing methods. In our approach, the scalar single index may have positive density on the whole real line. Removing the compact support assumption also brings technical complications.

The rest of the paper is organized as follows. We introduce our estimation methodology in Section 2, and then present the asymptotic results in Section 3. The results of the simulation studies and an application to spectral data are reported in Sections 4 and 5, respectively. Detailed proofs are deferred to the supplementary material.

2. Methodology

2.1. Model and FPCA approximation

Suppose that (X_i, Y_i), for $i = 1, \ldots, n$, are independent and identically distributed (i.i.d.) observations of (X, Y), where Y is a scalar, and X is a square integrable random function defined on a compact interval $\mathcal{I}$. We model the relationship between Y and X by imposing a single-index structure on the conditional mean. Specifically, the model takes the form

$$Y_i = \eta(U_i(\beta_0)) + e_i = \eta\left(\int_{\mathcal{I}}[X_i(t) - \mu(t)]\beta_0(t)\,dt\right) + e_i, \quad i = 1, \ldots, n, \tag{2.1}$$

where $\mu(\cdot) = \mathrm{E}\,X(\cdot)$ is the mean function, β_0 is an unknown square integrable function defined on $\mathcal{I}$, η is an unknown link function for the conditional mean, and e_i, for $i = 1, \ldots, n$, are i.i.d. error terms with mean zero and variance σ_e^2 that are independent of X_i.

Letting $\Sigma(t_1, t_2) = \mathrm{Cov}(X(t_1), X(t_2))$ be the covariance function of X, it admits the following spectral decomposition:

$$\Sigma(t_1, t_2) = \sum_{j=1}^{\infty} \lambda_j \phi_j(t_1)\phi_j(t_2),\ (t_1, t_2) \in \mathcal{I}^2,$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ are eigenvalues and (λ_j, ϕ_j), for $j = 1, 2, \ldots,$ are (eigenvalue, eigenfunction) pairs for Σ. Then, the Karhunen–Loève expansion of X_i is given by

$$X_i(t) = \mu(t) + \sum_{j=1}^{\infty} x_{ij}\phi_j(t),\ t \in \mathcal{I}, \tag{2.2}$$

where the FPC scores

$$x_{ij} = \int_{\mathcal{I}} [X_i(t) - \mu(t)]\phi_j(t)\, dt\,,\ j = 1, 2, \ldots, \tag{2.3}$$

have zero means and variances $\mathrm{E}\, x_{ij}^2 = \lambda_j$, and are uncorrelated. Because the sequences of eigenfunctions $\{\phi_j\}_{j=1}^{\infty}$ are complete in the class of square integrable functions on $\mathcal{I}$, the slope function β_0 can be represented by

$$\beta_0(t) = \sum_{j=1}^{\infty} b_{0j}\phi_j(t),\ t \in \mathcal{I}, \tag{2.4}$$

where

$$b_{0j} = \int_{\mathcal{I}} \beta_0(t)\phi_j(t)\, dt\,, \quad j = 1, 2, \ldots, \tag{2.5}$$

are expansion coefficients. Because the eigenfunctions are orthogonal, by substituting (2.2) and (2.4) into (2.1), we have

$$Y_i = \eta(U(\beta_0)) + e_i = \eta\left(\sum_{j=1}^{\infty} x_{ij} b_{0j}\right) + e_i. \tag{2.6}$$

Therefore, the scalar single index in parentheses depends only on the FPC scores x_{ij} and the expansion coefficients b_{0j}. Note that x_{ij} depends on the unknown eigenfunction ϕ_j, and hence needs to be estimated from samples. To ensure identifiability, we set an additional constraint on $\beta_0(t)$ that $\int_{\mathcal{I}} [\beta_0(t)]^2\, dt = \sum_{j=1}^{\infty} b_{0j}^2 = 1$. We also assume $b_{01} > 0$ (or $b_{0r} > 0$, for some $b_{0r} \neq 0$, if $b_{01} = 0$).

Because the model (2.6) is infinite dimensional, we approximate β_0 by truncation

$$\beta_0(t) \approx \sum_{j=1}^{p} b_{0j}\phi_j(t),\ t \in \mathcal{I},$$

where p denotes the frequency cut-off. In practice, the mean function μ, the covariance function Σ, and its eigenfunctions $\{\phi_j\}$ are unknown, and thus need to be estimated from samples. The estimate of the mean function μ is given by

$$\hat{\mu}(t) = \frac{1}{n}\sum_{i=1}^{n} X_i(t),\ t \in \mathcal{I}.$$

The sample versions of Σ and its spectral decomposition are

$$\begin{aligned}\hat{\Sigma}(t_1, t_2) &= \frac{1}{n}\sum_{i=1}^{n}[X_i(t_1) - \hat{\mu}(t_1)][X_i(t_1) - \hat{\mu}(t_2)]\\ &= \sum_{j=1}^{\infty} \hat{\lambda}_j \hat{\phi}_j(t_1)\hat{\phi}_j(t_2),\ (t_1, t_2) \in \mathcal{I}^2,\end{aligned}$$

where $(\hat{\lambda}_j, \hat{\phi}_j)$ are (eigenvalue, eigenfunction) pairs for $\hat{\Sigma}$, ordered such that $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots$ and $\hat{\lambda}_j = 0$, for $j \geq n+1$. By replacing (μ, ϕ_j) in (2.3) with its estimate $(\hat{\mu}, \hat{\phi}_j)$, we obtain the estimates of the FPC scores as follows:

$$\hat{x}_{ij} = \int_{\mathcal{I}} [X_i(t) - \hat{\mu}(t)]\hat{\phi}_j(t)\, dt\,.$$

The nonlinear least squares estimator of $(\eta, \boldsymbol{b}_p)$ is defined by minimizing

$$S(\boldsymbol{b}_p, \eta) = \sum_{i=1}^{n} \left[Y_i - \eta\big(\hat{\boldsymbol{x}}_{ip}^T \boldsymbol{b}_p\big)\right]^2, \quad \text{subject to } \|\boldsymbol{b}_p\|^2 = 1 \text{ and } b_1 > 0, \tag{2.7}$$

with respect to $(\eta, \boldsymbol{b}_p)$, where $\hat{\boldsymbol{x}}_{ip} = (\hat{x}_{i1}, \ldots, \hat{x}_{ip})^T$. As long as we obtain $\hat{\boldsymbol{b}}_p$, the minimizer of (2.7), the estimator of β is given by

$$\hat{\beta}_p(t) = \sum_{j=1}^{p} \hat{b}_j \hat{\phi}_j(t),\, t \in \mathcal{I}. \tag{2.8}$$

In practice, it is difficult to solve the minimization problem (2.7) because η is an infinite-dimensional parameter. We hence develop the following two-step estimating approach. Because η is not known, we first use the local linear estimator to represent η as a function of $\boldsymbol{b}_p$ in Section 2.2. Then, we estimate $\boldsymbol{b}_p$ by solving the estimating equations to obtain the estimator of (η, β); see Section 2.3. The computation procedure is illustrated in Section 2.4.

2.2. The local linear estimator for the link function

For a given $\boldsymbol{b}_p$, we can estimate η and its first derivative η' at a point u by minimizing the weighted sum of squares

$$\sum_{i=1}^{n}\Big[Y_i-\zeta-\vartheta(\hat{U}_i(\boldsymbol{b}_p)-u)\Big]^2 K_h\Big[\hat{U}_i(\boldsymbol{b}_p)-u\Big], \tag{2.9}$$

with respect to the parameters $(\zeta,\,\vartheta)$, where $K_h(\cdot)=h^{-1}K(\cdot/h)$, $K(\cdot)$ is a kernel function, $h>0$ is a bandwidth, and

$$\hat{U}_i(\boldsymbol{b}_p)=\int_{\mathcal{I}}[X_i(t)-\hat{\mu}(t)][\boldsymbol{b}_p^T\hat{\boldsymbol{\phi}}_p(t)]\,dt=\hat{\boldsymbol{x}}_{ip}^T\boldsymbol{b}_p,$$

where $\hat{\boldsymbol{\phi}}_p=(\hat{\phi}_1,\ldots,\,\hat{\phi}_p)^T$. The minimizer $(\hat{\zeta},\,\hat{\vartheta})$ of (2.9) is defined as the estimators of $(\eta,\,\eta')$. For the estimator of $\hat{\eta'}$, we use a different bandwidth h^* in (2.9) instead of h. We have the following explicit expressions for $\hat{\eta}$ and $\hat{\eta'}$:

$$\hat{\eta}(u\mid\boldsymbol{b}_p)=\sum_{i=1}^{n}W_{ni}(u\mid\boldsymbol{b}_p)Y_i, \tag{2.10}$$

and

$$\hat{\eta'}(u\mid\boldsymbol{b}_p)=\sum_{i=1}^{n}\tilde{W}_{ni}(u\mid\boldsymbol{b}_p)Y_i, \tag{2.11}$$

where

$$\begin{aligned}
W_{ni}(u\mid\boldsymbol{b}_p)&=\frac{\xi_i(u\mid\boldsymbol{b}_p,\,h)}{\sum_{k=1}^{n}\xi_k(u\mid\boldsymbol{b}_p,\,h)},\\
\tilde{W}_{ni}(u\mid\boldsymbol{b}_p)&=\frac{\tilde{\xi}_i(u\mid\boldsymbol{b}_p,\,h^*)}{\sum_{k=1}^{n}\xi_k(u\mid\boldsymbol{b}_p,\,h^*)},\\
\xi_i(u\mid\boldsymbol{b}_p,\,h)&=[s_2(u\mid\boldsymbol{b}_p,\,h)-s_1(u\mid\boldsymbol{b}_p,\,h)(\hat{U}_i(\boldsymbol{b}_p)-u)]K_h\Big(\hat{U}_i(\boldsymbol{b}_p)-u\Big),\\
\tilde{\xi}_i(u\mid\boldsymbol{b}_p,\,h^*)&=[s_0(u\mid\boldsymbol{b}_p,\,h^*)(\hat{U}_i(\boldsymbol{b}_p)-u)-s_1(u\mid\boldsymbol{b}_p,\,h^*)]K_{h^*}\Big(\hat{U}_i(\boldsymbol{b}_p)-u\Big),
\end{aligned}$$

and

$$s_l(u\mid\boldsymbol{b}_p,\,h)=\frac{1}{n}\sum_{i=1}^{n}(\hat{U}_i(\boldsymbol{b}_p)-u)^l K_h\Big[\hat{U}_i(\boldsymbol{b}_p)-u\Big],\ l=0,\,1,\ldots.$$

2.3. The estimating equations for the expansion coefficients

In order to obtain $\hat{\boldsymbol{b}}_p$, we take advantage of the constraint $\|\boldsymbol{b}_p\|^2=1$ to transfer the restricted least squares (2.7) to an unrestricted least squares by reparametrization. Let

$$\check{\boldsymbol{b}}_p = (\check{b}_2, \ldots, \check{b}_p)^T = (\hat{\lambda}_2^{1/2} b_2, \ldots, \ldots, \hat{\lambda}_p^{1/2} b_p)^T.$$

We can write

$$\boldsymbol{b}_p = \boldsymbol{b}_p(\check{\boldsymbol{b}}_p) = \left(\left(1 - \sum_{j=2}^{p} \hat{\lambda}_j^{-1} \check{b}_j^2 \right)^{1/2}, \hat{\lambda}_2^{-1/2} \check{b}_2, \ldots, \hat{\lambda}_p^{-1/2} \check{b}_p \right)^T. \tag{2.12}$$

This "remove-one-component" method is also used in Yu and Ruppert (2002), Wang et al. (2010), and Cui, Härdle and Zhu (2011). However, the proposed estimator differs from the latter methods in that we include the variance estimates of the FPC scores, $\hat{\lambda}_j$, in the reparameterization (2.12). This modification is essential for functional single-index models in both computation and theory, because λ_j decreases to zero as j increases. We obtain $\hat{\boldsymbol{b}}_p$ by minimizing $S[\boldsymbol{b}_p(\check{\boldsymbol{b}}_p)]$ in (2.7) with respect to $\check{\boldsymbol{b}}_p$, which is an unrestricted least squares problem, and then computing $\hat{\boldsymbol{b}}_p$ using (2.12). Let $\hat{\boldsymbol{J}}_p(\check{\boldsymbol{b}}_p) = \partial \boldsymbol{b}_p / \partial \check{\boldsymbol{b}}_p$ be the Jacobian matrix of $\boldsymbol{b}_p$ with respect to $\check{\boldsymbol{b}}_p$. A simple calculation yields

$$\hat{\boldsymbol{J}}_p(\check{\boldsymbol{b}}_p) = \hat{\boldsymbol{\Lambda}}_p^{-1/2} \begin{pmatrix} -\hat{\lambda}_1^{1/2} b_1^{-1} (\hat{\lambda}_2^{-1} \check{b}_2, \ldots, \hat{\lambda}_p^{-1} \check{b}_p) \\ \boldsymbol{I}_{p-1} \end{pmatrix},$$

where $\hat{\boldsymbol{\Lambda}}_p = \mathrm{diag}\{\hat{\lambda}_1, \ldots, \hat{\lambda}_p\}$. We have the following estimation equations with respect to $\check{\boldsymbol{b}}_p$:

$$\sum_{i=1}^{n} \left[Y_i - \hat{\eta}\big((\hat{\boldsymbol{x}}_{ip})^T \boldsymbol{b}_p(\check{\boldsymbol{b}}_p) \mid \boldsymbol{b}_p(\check{\boldsymbol{b}}_p)\big) \right] \hat{\eta'}\big((\hat{\boldsymbol{x}}_{ip})^T \boldsymbol{b}_p(\check{\boldsymbol{b}}_p) \mid \boldsymbol{b}_p(\check{\boldsymbol{b}}_p)\big) \hat{\boldsymbol{J}}_p^T(\check{\boldsymbol{b}}_p) \hat{\boldsymbol{x}}_{ip} = 0, \tag{2.13}$$

where $\hat{\eta}(\cdot|\cdot)$ and $\hat{\eta'}(\cdot|\cdot)$ are defined in (2.10) and (2.11), respectively.

To avoid the boundary effect caused by the small values in the denominators of $\hat{\eta}$ and $\hat{\eta'}$, we introduce a truncated version of $(\hat{\eta}, \hat{\eta'})$. For some positive constant sequence d_n tending to zero, let

$$g_n(u \mid \boldsymbol{b}_p, h) = \frac{1}{nh^2} \sum_{i=1}^{n} \xi_i(u \mid \boldsymbol{b}_p, h),$$

and

$$g_{d_n}(u \mid \boldsymbol{b}_p, h) = \max\{g_n(u \mid \boldsymbol{b}_p, h), \nu_2 d_n^2\},$$

where $\nu_l = \int u^l K(u)\, du$, for integer $l \geq 0$. The truncated version of $\hat{\eta}(u)$ is defined by

$$\hat{\eta}_{d_n}(u \mid \boldsymbol{b}_p) = \hat{\eta}(u \mid \boldsymbol{b}_p)\frac{g_n(u \mid \boldsymbol{b}_p, h)}{g_{d_n}(u \mid \boldsymbol{b}_p, h)}. \tag{2.14}$$

Similarly, the truncated version of $\hat{\eta'}(u)$ is defined by

$$\hat{\eta'}_{d_n}(u \mid \boldsymbol{b}_p) = \hat{\eta'}(u \mid \boldsymbol{b}_p)\frac{g_n(u \mid \boldsymbol{b}_p, h^*)}{g_{d_n}(u \mid \boldsymbol{b}_p, h^*)}. \tag{2.15}$$

Replacing $(\hat{\eta}, \hat{\eta}')$ in (2.13) with $(\hat{\eta}_{d_n}, \hat{\eta'}_{d_n})$ yields

$$\sum_{i=1}^{n}\left[Y_i - \hat{\eta}_{d_n}\left((\hat{\boldsymbol{x}}_{ip})^T\boldsymbol{b}_p(\check{\boldsymbol{b}}_p) \mid \boldsymbol{b}_p(\check{\boldsymbol{b}}_p)\right)\right]\hat{\eta'}_{d_n}\left((\hat{\boldsymbol{x}}_{ip})^T\boldsymbol{b}_p(\check{\boldsymbol{b}}_p) \mid \boldsymbol{b}_p(\check{\boldsymbol{b}}_p)\right)\hat{\boldsymbol{J}}_p^T(\check{\boldsymbol{b}}_p)\hat{\boldsymbol{x}}_{ip} = 0. \tag{2.16}$$

Denote the solution of (2.16) by $\check{\boldsymbol{b}}_p$. The resulting estimator $\hat{\boldsymbol{b}}_p = \boldsymbol{b}_p(\check{\boldsymbol{b}}_p)$ is our target estimator of $\boldsymbol{b}_p$. By replacing $\boldsymbol{b}_p$ in (2.8) and (2.14) with the estimator $\hat{\boldsymbol{b}}_p$, we obtain the estimators of β and η.

The selection of the tuning parameters (p, h, h^*, d_n) is important for the estimators of η and $\boldsymbol{b}_p$. We discuss this issue in Section 2.4.

2.4. Algorithm

We compute the estimates of η and $\boldsymbol{b}_p$ following an alternating estimating strategy after initializing β. The algorithm is described as follows.

Algorithm 1

Step 1 Apply a functional dimension-reduction method for the regression of Y_i versus X_i to find an initial estimate $\hat{\beta}$.

Step 2 Set $h = n^{-1/4}$, $h^* = n^{-1/6}$, and $d_n = n^{-0.55}$. Smooth Y_i versus $\int_{\mathcal{I}} X_i(t)\hat{\beta}(t)\,dt$ using the truncated local linear estimator in (2.14) and (2.15) to obtain the initial estimates of η and its first derivative η'.

Step 3 Given $\hat{\eta}_{d_n}$ and $\hat{\eta'}_{d_n}$, compute $\check{\boldsymbol{b}}_p$ by solving the estimating equations (2.16), and then substitute $\check{\boldsymbol{b}}_p$ into (2.12) to obtain $\hat{\boldsymbol{b}}_p$.

Step 4 Given $\hat{\boldsymbol{b}}_p$, select (p, h, h^*, d_n) and smooth Y_i versus $\hat{U}_i(\boldsymbol{b}_p) = \hat{\boldsymbol{x}}_{ip}^T\hat{\boldsymbol{b}}_p$ using the truncated local linear estimator in (2.14) and (2.15) to update the estimates of η and η'.

Step 5 Iterate Steps 3 and 4 until $\hat{\boldsymbol{b}}_p$ fails to change or the number of interactions reaches a predefined limit.

Step 6 Given the final estimate $\hat{\boldsymbol{b}}_p$, compute $\hat{\beta}_p$ in (2.8) as the final estimate of β. Select the optimal (h, d_n) and compute $\hat{\eta}_{d_n}$ in (2.14) as the final estimate of η.

In Step 1, one can use any existing method to initialize $\hat{\beta}$. We suggest using the functional cumulative slicing method of Yao, Lei and Wu (2015), which is fast and performed well in our simulations. In Step 2, we use larger tuning parameters to initialize η and η', which helps to reduce the variance of the initial estimators. For the final estimator $\hat{\eta}_{d_n}$ in Step 6, note that $\hat{\eta}_{d_n}(u_0) = 0$ for a given u_0 if $|U_i(\hat{\boldsymbol{b}}_p) - u_0| \geq h$, for all $i = 1, \ldots, n$. In this case, we may adopt a linear extrapolation, or other extrapolation method to extend $\hat{\eta}_{d_n}$ to u_0 if we are concerned about the value of $\eta(u_0)$, although there is no theoretical assurance of the extrapolation accuracy.

In Steps 4 and 6, we use K-fold cross-validation repeated over a few random partitions to determine the optimal (p, h) that minimizes the cross-validated mean squared prediction error. Because the proposed estimator is not sensitive to d_n, we simply set $d_n = p^{1/2}h^{-1/2}n^{-3/4}$ for each (p, h). We use the selected $(\hat{p}_{opt}, \hat{h}_{opt})$ and the corresponding d_n to compute η in Step 6. In Step 4, we set

$$\hat{h} = \hat{h}_{opt}n^{-0.05}, \quad \hat{h}^* = \hat{h}^{2/3}, \text{ and } \hat{d}_n = \hat{p}_{opt}^{1/2}\hat{h}^{-1/2}n^{-3/4},$$

because this ensures that the tuning parameters (h, h^*, d_n) are in the correct orders in Condition (A4) in Section 3 for optimal asymptotic performance. According to the discussion after Theorem 2, the optimal bandwidth $h_{opt} \sim (n^{-1}pu_n \log n)^{1/5}$, where $u_n = O(\log n)$. This approach can also be found in Carroll et al. (1997), Stute and Zhu (2005), and Wang et al. (2010).

3. Theoretical Properties

Our first aim is to show that the proposed estimator for the slope function β achieves the optimal rates of convergence for both estimation and prediction of the scalar single index under mild conditions. Suppose that i.i.d. data (X_i, Y_i), for $i = 1, \ldots, n$, are generated by (2.1). Let X be an independent copy of X_1. Denote the jth PCA score of X by $x_{\cdot j} = \int_{\mathcal{I}} X(t)\phi_j(t)\, dt$, and write $\boldsymbol{x}_{\cdot p} = (x_{\cdot 1}, \ldots, x_{\cdot p})^T$. Below, we use c to denote a positive constant that may change depending in the context. The following regularity conditions are required:

(A1) (i) For each $r \geq 2$ and $j \geq 1$, $\mathrm{E}\, x_{\cdot j}^{2r} \leq c(r)\lambda_j^r$, where $c(r)$ does not depend on j.

(ii) Of the eigenvalues λ_j, it is required that $\lambda_j - \lambda_{j+1} \geq cj^{-\alpha_0 - 1}$, for some $\alpha_0 > 1$ and all $j \geq 1$.

(iii) $\lambda_j \leq cj^{-\alpha_0}$, for all $j > 0$.

(iv) The expansion coefficients $|b_{0j}|$ satisfy $|b_{0j}| \leq cj^{-\alpha_1}$, for some $\alpha_1 > \alpha_0/2 + 1$ and all $j \geq 1$.

(A2) The cut-off frequency p satisfies $p \sim n^{1/(\alpha_0+2\alpha_1)}$.

Conditions (A1)(ii) and (iv) are fundamental in the literature on functional regression models based on the FPCA, and the rate of the cut-off frequency p in condition (A2) is known as the optimal rate; see Hall and Horowitz (2007) for details. Note that $\lambda_j - \lambda_{j+1} \geq cj^{-\alpha_0-1}$, for all $j \geq 1$, implies $\lambda_j \geq cj^{-\alpha_0}$, for some constant $c \geq 0$ not depending on j. Conditions (A1)(i) and (A1)(iii) are extracted from equations (4.3) and (4.2) in Cai and Hall (2006), respectively, and are required to control $\mathrm{E}[\hat{x}_{\cdot j} - x_{\cdot j}]^2$ and $\mathrm{E}[\int_{\mathcal{I}} X(t)[\beta_0(t) - \hat{\beta}(t)]\, dt]^2$.

Before proceeding further, we first introduce a truncated version of the true underlying slope function, which plays a central role in the derivation of the theoretical results. For a given p, let $\tilde{\boldsymbol{b}}_p = (\tilde{b}_1, \ldots, \tilde{b}_p)$, where

$$\tilde{b}_j = b_{0j},\ 1 \leq j \leq p-1, \text{ and } |\tilde{b}_p| = \left(\sum_{j=p}^{\infty} b_{0j}^2\right)^{1/2},$$

such that $\tilde{b}_p b_{0p} \geq 0$, and

$$\tilde{\beta}_p(t) = \sum_{i=1}^{p} \tilde{b}_j \hat{\phi}_j(t),\ t \in \mathcal{I}.$$

Clearly, $\|\tilde{\beta}_p\|_2 = \|\tilde{\boldsymbol{b}}_p\| = 1$. We have the following result for $\tilde{\beta}_p$.

Lemma 1. *Under conditions* (A1) *and* (A2)*, we have*

$$\mathrm{E}\,\|\tilde{\beta}_p - \beta_0\|_2^2 = O\Big(n^{-(2\alpha_1-1)/(\alpha_0+2\alpha_1)}\Big), \tag{3.1}$$

and

$$\mathrm{E}\left[\int_{\mathcal{I}}(X(t) - \hat{\mu}(t))\tilde{\beta}_p(t)\, dt - \int_{\mathcal{I}}(X(t) - \mu(t))\beta_0(t)\, dt\right]^2 = O\Big(\frac{p}{n}\Big). \tag{3.2}$$

It is known that the convergence rates on the right sides of (3.1) and (3.2) achieve the minimax rates for the estimation error and the prediction error, respectively, in functional linear models; see Cai and Hall (2006), Hall and Horowitz (2007), Yuan and Cai (2010), and Cai and Yuan (2012) for details. Additionally, define

$$\mathcal{B}_p = \mathcal{B}_p(n, C_1) = \left\{\boldsymbol{b}_p \in \mathbb{R}^p : \|\boldsymbol{b}_p\| = 1,\ \sum_{j=1}^{p} \lambda_j (b_j - \tilde{b}_j)^2 \leq \frac{C_1 p}{n}\right\}, \tag{3.3}$$

where $C_1 < \infty$ is a positive constant. With a slight abuse of notation, we also write $\beta_p \in \mathcal{B}_p$ if $\beta_p = \sum_{j=1}^{p} b_j \hat{\phi}_j$, for some $\boldsymbol{b}_p \in \mathcal{B}_p$. We have the following result.

Lemma 2. *Under conditions* (A1) *and* (A2)*, we have*

$$\sup_{\beta_p \in \mathcal{B}_p} \|\beta_p - \tilde{\beta}\|_2^2 = O_p\Big(n^{-(2\alpha_1 - 1)/(\alpha_0 + 2\alpha_1)}\Big), \text{ and} \tag{3.4}$$

$$\begin{aligned} &\mathrm{E}\left\{\sup_{\beta_p \in \mathcal{B}_p} \left[\int_{\mathcal{I}} (X(t) - \hat{\mu}(t))(\beta_p(t) - \tilde{\beta}(t))\, dt\right]^2\right\} \\ &= \mathrm{E} \sup_{\boldsymbol{b}_p \in \mathcal{B}_p} [\hat{\boldsymbol{x}}_{\cdot p}^T(\boldsymbol{b}_p - \tilde{\boldsymbol{b}}_p)]^2 = O\Big(n^{-(\alpha_0 + 2\alpha_1 - 1)/(\alpha_0 + 2\alpha_1)}\Big). \end{aligned} \tag{3.5}$$

By using Lemmas 1 and 2, we find that in order to establish the minimax rates for the proposed estimator, it suffices to show that there exists a $\boldsymbol{b}_p \in \mathcal{B}_p$ with $C_1 < \infty$ satisfying (2.16) in probability. To prove this, the following conditions are also required. We use $c_1, c_2, \ldots$ to denote positive constants not depending on n.

(A3) The kernel function K is a symmetric probability density function satisfying the Lipschitz condition of order one with support $[-1, 1]$.

(A4) (i) $nh^3/p \to \infty$, $n^{(\alpha_1 - 1/2)/(\alpha_0 + 2\alpha_1) - c_1} h \to \infty$, and $nh^8 \to 0$.

(ii) There is a $c_2 > 0$ such that $d_n n^{1/2 + c_2} \to 0$, $n^{1 - c_2} h d_n \to \infty$, $h/h^* = o(n^{-c_2})$, and $(h^*)^2/h = o(n^{-c_2})$.

(iii) $nh^4/p = o(1)$. There is a $c_3 > 0$ such that $[nh(h^*)^3]^{-1} = o(n^{-c_3})$, $p^2[n(h^*)^3]^{-1} = o(n^{-c_3})$, and $p^3[n^2 h(h^*)^3 d_n]^{-1} = o(n^{-c_3})$.

(A5) (i) The density function $f_{\beta_0}(\cdot)$ of $U(\beta_0) = \int_{\mathcal{I}} [X(t) - \mu(t)]\beta_0(t)\, dt$ has a connected support and satisfies the Lipschitz condition of order one on its support.

(ii) There exist c_4, $c_5 > 0$ and a constant u_0 such that $f_{\beta_0}(u) \le c_4 \exp\{-c_5|u|\}$, for $|u| \ge u_0$.

(iii) There exists a $c_6 > 0$ such that $(h^*|f'_{\beta_0}(u + h^* v)|)/(\max\{f_{\beta_0}(u), d_n\}) = o(n^{-c_6})$ holds uniformly over $u \in \mathbb{R}$ and $v \in [-1, 1]$, where d_n and h^* satisfy conditions (A4)(ii) and (iii).

(A6) (i) Let $\mathcal{A}_x$ be the functional space such that $X \in \mathcal{A}_x$ almost surely. Then, $\eta(U(\beta_0))$ has bounded continuous derivatives up to the second order, and $\eta'(U(\beta_0))$ is not identical to zero for $X \in \mathcal{A}_x$.

(ii) Define $\eta_{2j}(u) = \lambda_j^{-1/2}\,\mathrm{E}(x_{\cdot j} \mid U(\beta_0) = u)$. For any $u_1, u_2 \in \mathbb{R}$ such that $f_{\beta_0}(u_1) > 0$ and $f_{\beta_0}(u_2) > 0$, it follows that $|\eta_{2j}(u_2) - \eta_{2j}(u_1)| \le c|u_2 - u_1|$ holds for all $j \ge 1$, where c is a constant not depending on j. In other words, for all $j \ge 1$, $\eta_{2j}(\cdot)$ satisfies the Lipschitz condition of order one on $\{u : f_{\beta_0}(u) > 0\}$ with the Lipschitz constant not larger than c.

(A7) $\mathrm{E}\, e_i = 0$, $\mathrm{E}\, e_i^4 < \infty$.

(A8) Let

$$\boldsymbol{V}_p = \mathrm{E}\left\{[\eta'(U(\beta_0))]^2 \tilde{\boldsymbol{J}}_p^T(\boldsymbol{x}_{\cdot p} - \mathrm{E}[\boldsymbol{x}_{\cdot p} \mid U(\beta_0)])(\boldsymbol{x}_{\cdot p} - \mathrm{E}[\boldsymbol{x}_{\cdot p} \mid U(\beta_0)])^T \tilde{\boldsymbol{J}}_p\right\},$$

where

$$\tilde{\boldsymbol{J}}_p = \boldsymbol{\Lambda}_p^{-1/2}\begin{pmatrix} \lambda_1^{1/2}\tilde{b}_1^{-1}(\lambda_2^{-1/2}\tilde{b}_2, \ldots, \lambda_p^{-1/2}\tilde{b}_p) \\ \boldsymbol{I}_{p-1} \end{pmatrix}.$$

Denoting the minimum eigenvalue of $\boldsymbol{V}_p$ by $\lambda_{min}(\boldsymbol{V}_p)$, there exists a constant c not depending on p such that $\lambda_{min}(\boldsymbol{V}_p) \ge c$, for all $p \ge 2$.

Condition (A3) is a standard second-order kernel condition. The bandwidth in condition (A4)(i) is used in the final step of Algorithm 1 to calculate $\hat{\eta}_{d_n}$. Conditions (A4)(ii) and (iii) put some restrictions on d_n and h^*. Here, we use two bandwidths h and h^* to estimate η and its derivative η', because the convergence rate of the estimator of η' is slower than that of the estimator of η. See Chiou and Müller (1998) and Wang et al. (2010) for more discussion on using different bandwidths in relevant models. Given conditions (A1) and (A2), we may set $h \sim n^{-1/4}$, $h^* \sim n^{-1/8}\sqrt{ph}$ and $d_n \sim p/(nh)$ to ensure that condition (A4) is satisfied. Condition (A5)(i) is a standard smoothness condition for the density function. Condition (A5)(ii) implies that the density function $f_{\beta_0}(u)$ is sub-exponential. Given condition (A5)(ii), it is not difficult to show

$$\mathrm{P}[f_{\beta_0}(u) \le n^{-c}] = O(n^{-c}\log n), \tag{3.6}$$

for any $c > 0$, which has a close relationship to (3.6) of Chen, Hall and Müller (2011). Condition (A5)(iii) focuses on the infinitesimal relative change in the density function f_{β_0}. It is easy to verify that both conditions (A5)(ii) and (iii) are obtained if X is a Gaussian process. In the case that $\|X\|_2$ is bounded almost surely, which ensures condition (A5)(ii), if the FPC scores $x_{\cdot j}$ are all independent and have continuous probability densities and β_0 has at least $(\log d_n)/(\log h^*)$ nonzero expansion coefficients b_{0j}, condition (A5)(iii) also holds. This can be verified simply by using the formulae for the convolution of probability distributions. Condition (A6) is a standard smoothness condition for η and η_{2j}. Condition (A7) is a sufficient condition to derive the asymptotic results. Condition (A8) parallels the condition in Lemma A.7 of Li et al. (2011). It is not difficult to show that, given that $\eta'(U(\beta_0))$ is not identical to zero for $X \in \mathcal{A}_x$ in condition (A6), condition (A8) holds if X has an elliptically contoured distribution, which is satisfied if X is a Gaussian process, although the latter is not necessary. See Li and Hsing (2010) for more details. Therefore, condition (A8) is not very restrictive.

Theorem 1 provides the convergence rate of the estimator $\hat{\beta}$ and that of the mean squared prediction error of the scalar single index, which achieve the minimax lower bounds for the estimation of the slope function and the prediction of the response, respectively, in functional linear regression models.

Theorem 1. *Suppose that conditions* (A1)–(A8) *are satisfied. Then, there exists a root* $\check{\boldsymbol{b}}_p$ *for the estimating equations* (2.16) *such that* $\mathrm{P}[\hat{\boldsymbol{b}}_p \in \mathcal{B}_p(n, C_1)] \to 1$ *with* $C_1 < \infty$, *where* $\hat{\boldsymbol{b}}_p = \boldsymbol{b}_p(\check{\boldsymbol{b}}_p)$ *is defined in* (2.12) *and* $\mathcal{B}_p$ *is defined in* (3.3). *Therefore, it follows that*

$$\|\hat{\beta} - \beta_0\|_2^2 = O_p\Big(n^{(-2\alpha_1+1)/(\alpha_0+2\alpha_1)}\Big) \tag{3.7}$$

and

$$\begin{aligned} &\mathrm{E}_X\left[\int_{\mathcal{I}}(X(t) - \hat{\mu}(t))\hat{\beta}(t)\,dt - \int_{\mathcal{I}}(X(t) - \mu(t))\beta_0(t)\,dt\right]^2 \\ &= O_p\Big(n^{(-\alpha_0-2\alpha_1+1)/(\alpha_0+2\alpha_1)}\Big), \end{aligned} \tag{3.8}$$

where the expectation E_X *is taken only over the new observation* X.

By an argument similar to the proof of Theorem 1 in Cai and Yuan (2012), the minimax lower bounds of the rates in Theorem 1 are not less than the corresponding rates in functional linear regression models. Therefore, the rates given in Theorem 1 are optimal in the minimax sense.

We now turn to the estimator for the link function $\hat{\eta}_{d_n}$ given in the final step of Algorithm 1. In Theorem 1, we have shown $\mathrm{P}(\hat{\boldsymbol{b}}_p \in \mathcal{B}_p) \to 1$. Therefore, we need only consider the asymptotic property of $\hat{\eta}_{d_n}$ with $\beta \in \mathcal{B}_p$. Recall that $U(\beta_0) = \int_{\mathcal{I}}[X(t) - \mu(t)]\beta_0(t)\,dt$. Write $\hat{U}(\beta) = \int_{\mathcal{I}}[X(t) - \hat{\mu}(t)]\beta(t)\,dt$. Given a positive sequence $\{u_n\}_{n=1}^{\infty}$, let $\mathcal{A}_x(u_n) = \{X \in \mathcal{A}_x : \|X - \mu\|_2 \leq u_n\}$ and $d'_n = \inf_{X\in\mathcal{A}_x(u_n)}[f_{\beta_0}(U(\beta_0))]$.

Theorem 2. *Suppose that conditions* (A1)–(A8) *except for* (A4)(iii) *are satisfied.*

If $u_n = O(\log n)$ *and* $\liminf_{n\to\infty} d_n'/d_n > 1$*, then we have*

$$\begin{aligned}&\sup_{(\beta,\,X)\in\mathcal{B}_p\times\mathcal{A}_x(u_n)}[\hat{\eta}_{d_n}(\hat{U}(\beta))-\eta(U(\beta_0))]^2\\&=O_p(h^4)+O_p\left(\frac{p\log n}{nhd_n'}\right)+O_p\left(n^{(-2\alpha_1+1)/(\alpha_0+2\alpha_1)}u_n^2\right),\end{aligned}\tag{3.9}$$

and

$$\begin{aligned}&\mathrm{E}_X\sup_{\beta\in\mathcal{B}_p}\left\{[\hat{\eta}_{d_n}(\hat{U}(\beta))-\eta(U(\beta_0))]^2\,\mathrm{I}[X\in\mathcal{A}_x(u_n)]\right\}\\&=O_p\big(h^4\big)+O_p\left(\frac{pu_n\log n}{nh}\right),\end{aligned}\tag{3.10}$$

where the expectation E_X *is taken only over the new observation* X*.*

In Theorem 2, we restrict $X \in \mathcal{A}_x(u_n)$ to the bound $|U(\beta_0)-\hat{U}(\hat{\beta})|$, which is required in the proof. In addition, we suppose $\liminf_{n\to\infty} d_n'/d_n > 1$. A simple explanation is that there are no sufficient observations in the region where the probability density f_{β_0} is near zero. However, we still allow $u_n \to \infty$ so that η can be consistently estimated on the whole real line. Equation (3.9) provides a uniform convergence rate for the squared perdition error of $\hat{\eta}_{d_n}$. Specifically, the first term on the right side of (3.9) is attributed to the bias of the local linear estimator. The second term is due to the variance of the local linear estimator. Here, we allow $p \to \infty$ and $d_n' \to 0$, which induce an additional penalty p/d_n'. Similar results can also be found in Hansen (2008) for the Nadaraya–Watson regression estimators and the local linear regression estimators. The last term on the right side of (3.9) results from the truncated error of the functional predictors, which can be improved if additional assumptions on the FPC scores of the new observation X are available. In practice, one may only be concerned with the estimate of η on a compact interval. In this case, both u_n and d_n' are constants, and the last two terms on the right side of (3.9) become $O_p[(p\log n)/nh]$ and $O_p[n^{(-2\alpha_1+1)/(\alpha_0+2\alpha_1)}]$, respectively.

For the mean squared prediction error in (3.10), the bias caused by the truncation error of the slope function is dominated by the second term on the right side of (3.10). Note that the right side of (3.10) does not depend on d_n'. Intuitively, although the variance of $\hat{\eta}_{d_n}(u)$ for a given u is large if the corresponding density $f_{\beta_0}(u)$ is low, the effect on the mean squared prediction error can be offset by its density when taking the expectation. When $u_n \sim \log n$, the optimal bandwidth that minimizes the right side of (3.10) is in the order of $(n^{-1}p\log^2 n)^{1/5}$. Recall that $p = o(n^{1/4})$ by conditions (A1) and (A2). By taking $h \sim (n^{-1}p\log^2 n)^{1/5}$, the mean squared convergence rate in (3.10) is faster than that in Theorem 1 of Chen, Hall and Müller (2011).

Let $\hat{e}_i = Y_i - \hat{\eta}_{d_n}(\hat{U}_i(\hat{\beta}))$ be the estimate of the regression error. Define the estimator of σ_e^2 by $\hat{\sigma}_e^2 = n^{-1}\sum_{i=1}^n \hat{e}_i^2$. The next theorem provides the asymptotic normality of $\hat{\sigma}_e^2$.

Theorem 3. *Under the conditions of Theorem* 2*, we have*

$$\sqrt{n}(\hat{\sigma}_e^2 - \sigma_e^2)(\widehat{\mathrm{Var}}(e_1^2))^{-1/2} \xrightarrow{d} N(0,\, 1), \tag{3.11}$$

where

$$\widehat{\mathrm{Var}}(e_1^2) = \frac{1}{n}\sum_{i=1}^n \hat{e}_i^4 - \left[\frac{1}{n}\sum_{i=1}^n \hat{e}_i^2\right]^2.$$

4. Simulation Studies

In this section, we illustrate the finite-sample performance of the proposed estimating procedure for functional single-index models using simulations. We use three different sample sizes $n = 100$, 200, and 400 in the simulations. The functional predictors are generated from the process $X_i(t) = t + \sum_{j=1}^{50} x_{ij}\phi_j(t)$ with $t \in [0,\, 1]$, where $\phi_j(t) = 2^{1/2}\sin(2j\pi t)$ for odd j and $\phi_j(t) = 2^{1/2}\cos(2j\pi t)$ for even j, and the FPC scores are i.i.d. as $N(0,\, j^{-1.5})$. The slope function is generated by $\beta(t) = \sum_{i=1}^{50} b_j\phi_j(t)$, with $b_j = 1$ for $j = 1,\, 2$, and, 3 and $b_j = j^{-2}$ for $j = 4, \ldots, 50$. The following link functions are considered:

Model I $\eta(U_i) = \cos(\pi U_i) + U_i$, non-monotone link with a linear trend;

Model II $\eta(U_i) = (U_i + 1/2)\sin[2^{-1}\pi U_i(1/2 + \exp(U_i))/(1 + \exp(U_i))]$, period link with varying amplitude and frequency; and

Model III $\eta(U_i) = \arctan(2U_i)$, monotone link.

Here, $U_i = \int_0^1 X_i(t)\beta(t)\,dt$ in all models. The Response Y_i are obtained as $Y_i = \eta(U_i) + e_i$, where the errors e_i, for $i = 1, \ldots, n$, are i.i.d. as $N(0,\, \sigma_e^2)$ with $\sigma_e^2 = R\,\mathrm{Var}(g(U_i))$. Here, R is a measure of the noise-to-signal ratio, which is set to 0.1 or 1 in our simulations.

We compare our method with the functional cumulative slicing (FCS) method in Yao, Lei and Wu (2015). We use the truncated local linear estimator described in (2.14) to estimate the link function η for the FCS method. The quartic kernel $K(u) = (15/16)(1 - u^2)^2$, $|u| \leq 1$ is used for all the smoothing steps. To select the optimal number of FPC scores p and the bandwidth h, we employ five-fold

Table 1. Simulation results for the angles between $\hat{\beta}$ and β and the mean squared prediction errors of the scalar single index $\mathrm{MSPE_I}$, obtained from 100 Monte Carlo repetitions. The standard errors are provided in parentheses. All entries are multiplied by 100.

			Proposed		FCS	
Model	n	R	Angle	$\mathrm{MSPE_I}$	Angle	$\mathrm{MSPE_I}$
I	100	0.1	18.4 (12.5)	0.54 (1.31)	27.5 (14.0)	1.49 (1.18)
		1	45.6 (22.7)	3.81 (4.30)	54.4 (26.4)	5.82 (5.25)
	200	0.1	9.47 (3.99)	0.09 (0.21)	16.7 (5.68)	0.58 (0.41)
		1	31.6 (17.6)	1.52 (2.07)	34.4 (19.6)	2.24 (2.11)
	400	0.1	6.56 (1.82)	0.03 (0.02)	11.1 (3.46)	0.27 (0.21)
		1	20.4 (10.9)	0.49 (0.94)	20.4 (12.2)	1.04 (0.99)
II	100	0.1	20.6 (12.4)	0.61 (1.30)	26.7 (12.1)	1.75 (1.97)
		1	50.1 (25.1)	4.46 (4.38)	58.0 (29.1)	7.08 (5.70)
	200	0.1	12.0 (4.72)	0.12 (0.11)	19.5 (8.30)	0.97 (1.12)
		1	33.7 (15.9)	1.96 (2.70)	34.2 (20.3)	2.69 (2.83)
	400	0.1	7.49 (2.04)	0.04 (0.03)	12.5 (4.11)	0.39 (0.32)
		1	23.3 (7.45)	0.45 (0.24)	20.2 (11.6)	1.06 (1.03)
III	100	0.1	21.9 (7.76)	0.44 (0.28)	25.9 (11.4)	0.93 (0.71)
		1	57.8 (18.8)	4.80 (3.03)	58.9 (21.3)	5.34 (3.43)
	200	0.1	14.7 (5.24)	0.18 (0.11)	17.5 (7.18)	0.38 (0.24)
		1	43.6 (13.2)	2.24 (1.30)	44.1 (16.0)	2.50 (1.67)
	400	0.1	9.99 (3.29)	0.07 (0.04)	11.3 (4.54)	0.17 (0.11)
		1	31.8 (10.7)	0.98 (0.71)	30.7 (14.4)	1.20 (0.90)

cross-validation, repeated over 10 random partitions. The accuracy of the slope function is quantified by the angle in radians between $\hat{\beta}$ and β. To assess the model estimation, we generate a validation sample of size $N = 1{,}000$ in each Monte Carlo run, and quantify the prediction error of the scalar single index by its mean squared prediction error $\mathrm{MSPE_I} = N^{-1}\sum_{i=1}^{N}\int_0^1[(X_i(t)-\hat{\mu}(t))(\beta_0(t)-\hat{\beta}(t))]^2\,dt$, where $\beta_0 = \beta/\|\beta\|_2$, and quantify the prediction error of the scalar response by $\mathrm{MSPE}_Y = N^{-1}\sum_{i=1}^{N}[Y_i^* - \hat{\eta}(\hat{U}_i)]^2$ where $Y_i^* = \eta(U_i)$.

We report the average values of the angles, $\mathrm{MSPE_I}$, and MSPE_Y over 100 Monte Carlo repetitions in Tables 1 and 2, along with their standard errors. The prediction errors of the scalar responses with known β are also reported, which serves as a benchmark. Tables 1 and 2 indicate that our method outperforms the FCS method by Yao, Lei and Wu (2015) when the noise-to-signal ratio $R = 0.1$. For the case $R = 1$, the two approaches perform comparably, which could be because the proposed method is more sensitive to larger regression errors. Our method has lower prediction errors of the scalar single indices.

Table 2. Simulation results for the mean squared prediction errors of the scalar responses MSPE_Y obtained from 100 Monte Carlo repetitions. The standard errors are provided in parentheses. All entries are multiplied by 100.

			Proposed	FCS	β is known
Model	n	R		MSPE_Y	
I	100	0.1	35.6 (15.4)	46.1 (14.4)	26.3 (7.48)
		1	75.8 (29.2)	79.1 (29.9)	34.3 (12.7)
	200	0.1	25.8 (8.21)	31.9 (7.63)	23.8 (6.19)
		1	48.1 (22.0)	52.5 (20.5)	27.7 (8.34)
	400	0.1	7.67 (3.70)	10.8 (4.47)	5.50 (2.83)
		1	26.9 (11.7)	31.8 (11.8)	18.2 (4.02)
II	100	0.1	13.8 (7.56)	19.8 (9.15)	9.29 (2.35)
		1	31.9 (14.8)	35.5 (15.9)	11.5 (4.38)
	200	0.1	5.10 (2.54)	9.17 (4.38)	3.56 (1.40)
		1	19.8 (9.95)	21.2 (9.22)	9.14 (2.88)
	400	0.1	1.95 (0.93)	3.78 (1.94)	1.51 (0.85)
		1	8.12 (3.34)	11.1 (5.78)	4.43 (1.87)
III	100	0.1	5.29 (2.28)	6.45 (2.32)	3.60 (1.34)
		1	24.5 (9.64)	22.1 (10.1)	9.72 (5.49)
	200	0.1	3.04 (1.16)	3.63 (1.25)	2.15 (0.73)
		1	13.6 (5.99)	12.1 (5.03)	5.72 (2.84)
	400	0.1	1.53 (0.70)	1.84 (0.76)	1.16 (0.55)
		1	6.50 (2.76)	6.19 (2.36)	3.34 (1.49)

5. Real-Data Application

In this section, we apply the functional single-index model to the Tecator spectrometric data set available on `http://lib.stat.cmu.edu/datasets/tecator`. The data set consists of 215 absorbance spectrum curves of finely chopped pure meat samples, measured at 100 wavelengths from 850 nm to 1,050 nm. The absorbance is $-\log_{10}$ of the transmittance. The spectrum curves are displayed in Figure 1(a). The fat content of each meat sample is determined by analytical chemistry. Our aim is to predict the percentage of fat content Y_i from the spectrometric curve $X_i(\cdot)$. According to the Beer–Lambert Law in analytical chemistry (Skoog et al. (2013)), there is a near linear relationship between the absorbance and the concentration of the absorbing species.

We first rescale the functional predictors on the interval [0, 1]. As suggested by Chen, Hall and Müller (2011), we normalize each spectrometric curve X_i by subtracting its area under the curve $\int_0^1 X_i(t)\,dt$. The normalized curves are displayed in Figure 1(b). In a preprocessing step, we remove three outliers by a

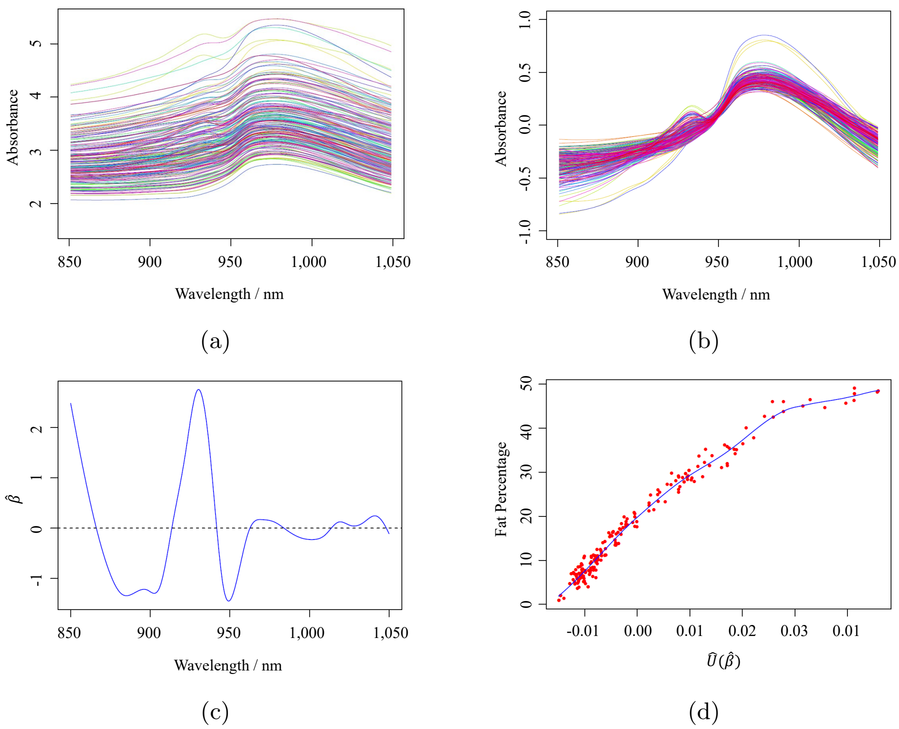

Figure 1. (a) (b) Spectra curves and normalized spectra curves of 215 meat samples. (c) The estimated slope function $\hat{\beta}$. (d) The estimated link function $\hat{\eta}_{d_n}$ (blue solid line) and the percentages of fat content of the samples Y_i (red dots).

simple visual inspection of the normalized curves. We then remove seven other curves according to the box plot of the residuals of the functional single-index model computed using the proposed method. In Chen, Hall and Müller (2011), the authors suggested an additive multiple-index model with two indices. Because the shapes of the slope functions of those indices were similar, it is natural to ask whether a single-index model is sufficient to capture the main relationship between the response and the functional predictor.

We use the proposed estimating procedure with the quartic kernel to fit the data and employ five-fold cross-validation, repeated over 10 random partitions, to select the number of FPC scores p and the bandwidth h. Because there is a significant difference between the cross-validated errors with $p = 3$ and 4, while including more FPCs in the model has little effect on the cross-validated error, we select $p = 4$ in this real-data application. The estimated slope functions $\hat{\beta}$ and the link function $\hat{\eta}_{d_n}$ are displayed in Figures 1(c) and 1(d), respectively. The shape of the slope function estimate $\hat{\beta}$ is similar to the shapes of the slope function estimates in Figure 3 of Chen, Hall and Müller (2011). The estimated link function $\hat{\eta}_{d_n}$ is near-linear, and bends down when the scalar index is large. The leave-one-out cross-validated mean squared prediction error of fat content as a percentage for the single-index model estimated by the proposed method is 2.27, while those for the additive multiple-index model with two indices estimated by the FCS method in Yao, Lei and Wu (2015) and reported in Chen, Hall and Müller (2011) are 3.20 and 2.39, respectively. Therefore, a single-index model is reasonable in this application.

We also apply the backfitting procedure described in Chen, Hall and Müller (2011) to check whether an additive multiple-index model with two indices is more appropriate. The leave-one-out cross-validated mean squared prediction error for the multiple-index model is 2.02, about 88% of that for the single-index model, showing a limited improvement. Although the additional index may provide some useful information for Y_i, we do not explore that here.

Supplementary Material

The online Supplementary Material contains proofs of Lemmas 1 and 2 and Theorems 1–3.

Acknowledgments

Wang's research was supported by the National Natural Science Foundation of China (General program 11871460, Key program 11331011, and program for Innovative Research Group in China 61621003), and a grant from the Key Lab of Random Complex Structure and Data Science, CAS, China.

References

Ait-Saïdi, A., Ferraty, F., Kassa, R. and Vieu, P. (2008). Cross-validated estimations in the single-functional index model. *Statistics* **42**, 475–494.

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

Cai, T. T. and Yuan, M. (2012). Minimax and adaptive prediction for functional linear regression. *J. Amer. Statist. Assoc.* **107**, 1201–1216.

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.

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

Chiou, J.-M. and Müller, H.-G. (1998). Quasi-likelihood regression with unknown link and variance functions. *J. Amer. Statist. Assoc.* **93**, 1376–1387.

Cook, R. D., Forzani, L. and Yao, A. F. (2010). Necessary and sufficient conditions for consistency of a method for smoothed functional inverse regression. *Statist. Sinica* **20**, 235–238.

Cui, X., Härdle, W. K. and Zhu, L. (2011). The EFM approach for single-index models. *Ann. Statist.* **39**, 1658–1688.

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

Dou, W. W., Pollard, D. and Zhou, H. H. (2012). Estimation in functional regression for general exponential families. *Ann. Statist.* **40**, 2421–2451.

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

Ferré, L. and Yao, A. F. (2003). Functional sliced inverse regression analysis. *Statistics* **37**, 475–488.

Ferré, L. and Yao, A.-F. (2005). Smoothed functional inverse regression. *Statist. Sinica* **15**, 665–683.

Goia, A. and Vieu, P. (2016). An introduction to recent advances in high/infinite dimensional statistics. *J. Multivariate Anal.* **146**, 1–6.

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

Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* **24**, 726–748.

Hilgert, N., Mas, A. and Verzelen, N. (2013). Minimax adaptive tests for the functional linear model. *Ann. Statist.* **41**, 838–869.

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

Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators.* Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester.

James, G. M. (2002). Generalized linear models with functional predictors. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64**, 411–432.

Jiang, C.-R., Yu, W. and Wang, J.-L. (2014). Inverse regression for longitudinal data. *Ann. Statist.* **42**, 563–591.

Li, T.-T., Yang, H., Wang, J.-L., Xue, L.-G. and Zhu, L.-X. (2011). Correction on estimation for a partial linear single-index model. *Ann. Statist.* **39**, 3441–3443.

Li, Y. and Hsing, T. (2010). Deciding the dimension of effective dimension reduction space for functional and high-dimensional data. *Ann. Statist.* **38**, 3028–3062.

Müller, H.-G. and Stadtmüller, U. (2005). Generalized functional linear models. *Ann. Statist.* **33**, 774–805.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis.* 2nd Edition. Springer Series in Statistics. Springer, New York.

Skoog, D. A., West, D. M., Holler, F. J. and Crouch, S. R. (2013). *Fundamentals of Analytical Chemistry.* 9th Edition. Nelson Education. Cengage Learning, Belmont.

Stute, W. and Zhu, L.-X. (2005). Nonparametric checks for single-index models. *Ann. Statist.* **33**, 1048–1083.

Wang, J.-L., Xue, L., Zhu, L. and Chong, Y. S. (2010). Estimation for a partial-linear single-index model. *Ann. Statist.* **38**, 246–274.

Wang, L. (2011). Gee analysis of clustered binary data with diverging number of covariates. *Ann. Statist.* **39**, 389–417.

Yao, F., Lei, E. and Wu, Y. (2015). Effective dimension reduction for sparse functional data. *Biometrika* **102**, 421–437.

Yao, F., Wu, Y. and Zou, J. (2016). Probability-enhanced effective dimension reduction for classifying sparse functional data. *TEST* **25**, 1–22.

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, X., Wang, C. and Wu, Y. (2018). Functional envelope for model-free sufficient dimension reduction. *J. Multivariate Anal.* **163**, 37–50.

Yunxiang Huang

Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100080, People's Republic of China.

University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China.

E-mail: yxhuang@amss.ac.cn

Qihua Wang

Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100080, People's Republic of China.

University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China.

Zhejiang Gongshang University, Hangzhou, Zhejiang 310018, People's Republic of China.

E-mail: qhwang@amss.ac.cn

(Received January 2021; accepted February 2021)