Received: 27 February 2022 | Revised: 27 October 2022 | Accepted: 5 December 2022

DOI: 10.1002/env.2784

RESEARCH ARTICLE

WILEY

Estimating functional single index models with compact support

Yunlong Nie | **Liangliang Wang** | **Jiguo Cao**

Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, British Columbia, Canada

Correspondence
Jiguo Cao, Department of Statistics and Actuarial Science Simon Fraser University Burnaby, BC, Canada.
Email: jiguo_cao@sfu.ca

Funding information
NSERC Discovery grants of Liangliang Wang and Jiguo Cao; Natural Sciences and Engineering Research Council of Canada

Abstract

The functional single index models are widely used to describe the nonlinear relationship between a scalar response and a functional predictor. The conventional functional single index model assumes that the coefficient function is nonzero in the entire time domain. In other words, the functional predictor always has a nonzero effect on the response all the time. We propose a new compact functional single index model, in which the coefficient function is only nonzero in a subregion. We also propose an efficient method that can simultaneously estimate the nonlinear link function, the coefficient function and also the nonzero region of the coefficient function. Hence, our method can identify the region in which the functional predictor is related to the response. Our method is illustrated by an application example in which the total number of daily bike rentals is predicted based on hourly temperature data. The finite sample performance of the proposed method is investigated by comparing it to the conventional functional single index model in a simulation study

KEYWORDS

bicycle rentals, functional data analysis, interpretability, weather conditions

1 | INTRODUCTION

Functional data analysis (FDA) is a growing statistical field for analyzing curves, surfaces, or any multidimensional functions, in which each random function is treated as a sample element Ferraty and Vieu (2006); Ramsay and Silverman (2005). In this article, we consider the problem of modeling the relationship between a functional predictor and a scalar response. Given a scalar response Y and a functional predictor $X(t)$ observed in $[0, T]$, the conventional functional linear regression model assumes that the scalar response Y is linked with a functional predictor in the following model:

$$E(Y|X) = \beta_0 + \int_0^T \beta(t)X(t)dt,$$

in which β_0 is the intercept and $\beta(t)$ is the coefficient function. The above functional linear regression model describes a linear relationship between the response and the functional predictor.

A natural way to extend the above functional linear regression model is to use a nonlinear functional single index model (FuSIM):

$$E(Y|X) = g\left(\int_0^T \beta(t)X(t)dt\right), \quad (1)$$

in which $\beta(t)$ is the index or weight function and $g(\cdot)$ is the link function describing the relationship between the scalar response Y and the integral of the functional predictor $X(t)$ and the index function $\beta(t)$. If $g(x)$ takes a known parametric form, such as the identity or logit function, then the functional single index model reduces to the conventional functional linear model Ramsay and Silverman (2005) or generalized functional linear model Müller and Stadtmüller (2005), respectively.

If we take a closer look at the functional single index model (1), the index function $\beta(t)$ determines weights of the contribution of the functional predictor $X(t)$ related to the scalar response through the function $g(\cdot)$ at each time point $t \in [0, T]$. In particular, if $\beta(t) \equiv 0$ on a subinterval $I \subset [0, T]$, then the functional predictor $X(t)$ is not related to the response Y on this subinterval. It is of great interest to estimate the support region of $\beta(t)$, denoted by $\mathcal{S} = \{t : \beta(t) \neq 0\}$, in which the functional predictor is related to the the response.

In this paper, we propose the following compact functional single index model (compact FuSIM):

$$E(Y|X) = g\left(\int_{\mathcal{S}} \beta(t)X(t)dt\right) \tag{2}$$

in which $\mathcal{S} \subset [0, T]$ is unknown. Our goal in this paper is to simultaneously estimate the index function $\beta(t)$, the nonlinear function $g(\cdot)$ without any parametric assumptions, and identify the support region $\mathcal{S}$ on which $X(t)$ is related to Y. Such a compact-supported estimate of $\beta(t)$ can not only increase the prediction accuracy of the response, but also lead to a better interpretation of the functional single index model.

Several methods have been proposed to estimate the functional single index model (1). For instance, Chen et al. (2011) estimated the unknown link function $g(\cdot)$ by nonparametric kernel smoothing with polynomial convergence rate Li (2018). Ma (2014) considered a functional single-index model with multiple functional predictors. They used B-spline basis functions to estimate both the link function $g(\cdot)$ and the index or slope function $\beta(t)$. They also showed uniform convergence rates of the proposed spline estimators. In addition, Ait-Saïdi et al. (2008) proposed an alternative cross-validation procedure Marron (1987) for model estimation. Jiang and Wang (2011) and Li et al. (2017) considered another type of index model in which the response is a function, and the covariates are scalar variables or functions. Other literature include Fan et al. (2015) and Goia and Vieu (2015). All the above methods assume that the support of the index function is the same as that of the functional predictors. That is, that $\mathcal{S} = [0, T]$.

In the conventional functional linear regression model, in which the link function $g(\cdot)$ takes a linear form, this compact support estimate has been studied by by James et al. (2009), Lin et al. (2017), and Zhou et al. (2013). The "FLiRTI" method proposed by James et al. (2009) approximates the coefficient function nonparametrically at some discrete grid points and penalizes the L_1 norm of $\beta(t)$ and its multiple derivatives such as $d\beta(t)/dt$ and $d^2\beta(t)/dt^2$ to determine whether the estimated function and its derivatives are zero at each point. As mentioned by Zhou et al. (2011), the coefficient function estimated by the "FLiRTI" method has a large variance when the number of points increases, and the model is often over-parameterized. To overcome this issue, a two-step procedure proposed by Zhou et al. (2011) is used to obtain an initial estimate for the support region of the coefficient function in the first step. The estimated support region is refined in the second step. However, the computation cost of this two-step procedure is quite high and the approach is hard to implement. A simple one-step procedure called "SloS" proposed by Lin et al. (2017) yields a smooth and locally sparse estimator of $\beta(t)$. The SloS method regularizes the sparseness and smoothness of the coefficient function using the functional SCAD penalty and the roughness penalty in a single optimization objective function. However, none of the above methods consider the scenario in which the functional predictor is related to the response by an unknown nonlinear link function $g(\cdot)$. To the best of our knowledge, this is the first study that considers the compact support problem within the functional single index model framework.

The main contribution of our paper is twofold. First, we propose a new compact functional single index model, in which the support region of the coefficient function $\beta(t)$ is a subregion of the entire domain. Second, we propose an efficient algorithm to estimate the coefficient function $\beta(t)$, its support region and the unknown nonlinear link function $g(\cdot)$ simultaneously. An R package `cFuSIM` has been developed to implement the proposed method. The R code for replicating the application results are available at https://github.com/caojiguo/cFuncSIM.

The remainder of this article is organized as follows. In Section 2, we present the details of compact FuSIM and an efficient algorithm for parameter estimation. The finite-sample performance of our method is evaluated in Section 3 using a carefully-designed simulation study. Section 4 applies the compact FuSIM method to analyze a bike rental dataset. Section 5 provides the concluding remarks.

2 | METHODOLOGY

2.1 | A compact functional single index model

Let $\{(x_i(t), y_i), i = 1, 2, \dots, n\}$ denote the i.i.d sample from model (2), where the noise of the model is assumed to follow the normal distribution $Normal(0, \sigma^2)$. The index function $\beta(t)$ is expressed as a linear combination of basis function: $\beta(t) = \mathbf{b}^\top \mathbf{B}(t)$, where $\mathbf{B}(t) = (B_1(t), \dots, B_L(t))^\top$ are basis functions and $\mathbf{b} = (b_1, \dots, b_L)^\top$ are the corresponding basis coefficients. We choose cubic B-splines as our basis functions, because they have the compact support property de Boor (2001), which is important for efficient computation and identification of the support region of $\beta(t)$. Suppose that the cubic B-splines are defined on $M + 1$ equally spaced knots $t_0, t_1, \dots, t_M$ in the domain $[0, T]$.

We propose to estimate the compact functional single index model (2) by minimizing the following criterion:

$$Q(\beta, g) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - g\left(\int_0^T \beta(t)x_i(t)dt\right)\right)^2 + \text{PEN}_\gamma(\beta) + \text{PEN}_\lambda(\beta), \tag{3}$$

where the L^2-norm of $\beta(t)$ is required to be 1 to ensure the identifiability of $g(\cdot)$ and $\beta(t)$, that is, $\int_0^T \beta^2(t)dt = 1$. Note that the integral inside the function $g(\cdot)$ is taken between $[0, T]$ instead of the true support region of $\beta(t)$, that is, S. The second term

$$\text{PEN}_\gamma(\beta) = \gamma \int_0^T \beta''(t)^2 dt,$$

which represents the roughness penalty on $\beta(\cdot)$, which is also introduced in Lin et al. (2017).

The third penalty term $\text{PEN}_\lambda(\beta)$ in (3) penalizes the L_1 norm of $\beta(t)$ to obtain a compact- supported estimate for $\beta(t)$. We employ the functional SCAD method proposed by Lin et al. (2017). which is a functional generalization of the SCAD method Fan and Li (2001). The functional SCAD method is proposed in Lin et al. (2017) to find a locally sparse estimator for the coefficient function in functional linear regression models. The nice shrinkage property of functional SCAD allows the proposed estimator to locate null subregions of the coefficient function without over shrinking nonzero values of the coefficient functions. More specifically, the functional SCAD penalty in the compact FuSIM method is defined as:

$$\text{PEN}_\lambda(\beta) = \frac{M}{T}\int_0^T p_\lambda(|\beta(t)|)dt, \tag{4}$$

where $p_\lambda(\cdot)$ is the SCAD function defined in Fan and Li (2001):

$$p_\lambda(u) = \begin{cases} \lambda u & \text{if } 0 \le u \le \lambda, \\ -\frac{u^2 - 2a\lambda u + \lambda^2}{2(a-1)} & \text{if } \lambda < u < a\lambda, \\ \frac{(a+1)\lambda^2}{2} & \text{if } u \ge a\lambda. \end{cases}$$

Here a is 3.7, as suggested by Fan and Li (2001), and λ is the tuning parameter. We refer to λ as the sparsity parameter for the rest of this paper. A large value of λ will penalize the nonzero region of the corresponding $\beta(t)$, hence leading to a compact estimation of $\beta(t)$. On the other hand, when the sparsity parameter, λ, is zero, the resulting $\beta(t)$ reduces to the conventional functional single index model.

2.2 | Summary of computing algorithm

Given a specific value of λ in the fSCAD penalty, we estimate $\beta(t)$ and the nonlinear function $g(\cdot)$ in an iterative fashion. The algorithm is summarized below:

Step I. Set the initial value for $\beta(t)$, denoted as $\beta^{(0)}(t)$. For example, the initial value for $\beta(t)$ can be a constant function.

Step II. Given the current value of $\beta^{(j)}(t)$, we estimate $g^{(j)}(\cdot)$ by minimizing

$$Q_1(g|\beta^{(j)}) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - g\left(\int_0^T \beta^{(j)}(t)x_i(t)dt\right)\right)^2;$$

Step III. Given the current value of $g^{(j)}(\cdot)$, we update the estimate of $\beta(t)$ to $\beta^{(j+1)}(t)$ by minimizing

$$Q_2(\beta|g^{(j)}) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - g^{(j)}\left(\int_0^T \beta(t)x_i(t)dt\right)\right)^2 + \text{PEN}_\lambda(\beta) + \gamma\int \beta''(t)^2dt;$$

Step IV. Repeat Step II and III until the algorithm coverages.

2.3 | Estimating the link function $g(\cdot)$

We now give the details of estimating the link function $g(\cdot)$ in Step II of the computing algorithm. For simplicity, we omit the index for iteration in this subsection. Given the current estimate of $\hat{\beta}(t)$, the criterion given in Step II becomes

$$Q_1(g|\hat{\beta}(t)) = \frac{1}{n}\sum_{i=1}^{n}(y_i - g(\hat{z}_i))^2,$$

in which $\hat{z}_i = \int_0^T \hat{\beta}(t)x_i(t)dt$. We propose to estimate g by the local linear regression Wand and Jones (1994). At any target point z_0, the local linear regression estimator of g is given as

$$\hat{g}(z_0) = \hat{m}_0(z_0) + \hat{m}_1(z_0)z_0,$$

where $\hat{m}_0(z_0), \hat{m}_1(z_0)$ is obtained by minimizing

$$\frac{1}{n}\sum_{i=1}^{n}\{y_i - m_0(z_0) - m_1(z_0)\hat{z}_i\}^2 K\left(\frac{\hat{z}_i - z_0}{h}\right)$$

with respect to $m_0(z_0)$ and $m_1(z_0)$. Here $K(\cdot)$ is a unimodal nonnegative kernel function, and h is the bandwidth. The bandwidth is selected by cross-validation as suggested by Fan and Gijbels (1995). Because m_0 and m_1 are estimated at any target point z_0, we express m_0 and m_1 as functions of z_0.

2.4 | Estimating the index function $\beta(t)$

This subsection covers the details of estimating the index function $\beta(t)$ in Step III of the computing algorithm. We also omit the index for iteration in this subsection for simplicity. Given the current estimate of $\hat{g}(\cdot)$, the criterion given in Step III becomes

$$Q_2(\beta|\hat{g}) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{g}\left(\int_0^T \beta(t)x_i(t)dt\right)\right)^2 + \text{PEN}_\gamma(\beta) + \text{PEN}_\lambda(\beta). \quad (5)$$

The integral inside the least squares term in (5)is expressed as

$$\int_0^T \beta(t)x_i(t)dt = \mathbf{b}^\top \int_0^T \mathbf{B}(t)x_i(t)dt = \mathbf{Z}_i^\top \mathbf{b},$$

in which $\mathbf{Z}_i = (\int B_1(t)x_i(t)dt, \ldots, \int B_L(t)x_i(t)dt)^\top$. Because $\hat{g}(\cdot)$ is a nonlinear function, minimizing $Q_2(\beta)$ in (5) is a nonlinear optimization problem. To solve this nonlinear optimization problem, we propose to apply local approximation and obtain the minimizer in an iterative manner. Based on the current estimates $\beta^{(j)}(t) = \mathbf{b}^{(j)}\mathbf{B}(t)$ in the jth iteration step, $j = 1, 2, \ldots$, the least squares term in (5) can be approximated by

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{g}\left(\int_0^T \beta(t)x_i(t)dt\right)\right)^2 \\
&\approx \frac{1}{n}\sum_{i=1}^{n}\left[y_i - \hat{g}(\mathbf{Z}_i^\top\mathbf{b}^{(j)}) - \hat{g}'(\mathbf{Z}_i^\top\mathbf{b}^{(j)})(\mathbf{Z}_i^\top\mathbf{b} - \mathbf{Z}_i^\top\mathbf{b}^{(j)})\right]^2 \\
&= (\mathbf{P}^{(j)} - \mathbf{T}^{(j)}\mathbf{b})^\top(\mathbf{P}^{(j)} - \mathbf{T}^{(j)}\mathbf{b}),
\end{aligned} \tag{6}$$

in which $\mathbf{P}^{(j)}$ is a $n \times 1$ vector with the ith element being $\frac{1}{\sqrt{n}}(y_i - \hat{g}(\mathbf{Z}_i^\top\mathbf{b}^{(j)}) - \hat{g}'(\mathbf{Z}_i^\top\mathbf{b}^{(j)})\mathbf{Z}_i^\top\mathbf{b}^{(j)})$ and $\mathbf{T}^{(j)}$ is a $n \times L$ matrix with the ith row being $\frac{1}{\sqrt{n}}\hat{g}'(\mathbf{Z}_i^\top\mathbf{b}^{(j)})\mathbf{Z}_i^\top$.

The roughness penalty term $\mathrm{PEN}_\gamma(\beta)$ in the loss function can be expressed as

$$\mathrm{PEN}_\gamma(\beta) = \gamma\int_0^T \left[\beta''(t)\right]^2 dt = \gamma\mathbf{b}^\top\Gamma\mathbf{b} \tag{7}$$

in which Γ denotes a $L \times L$ matrix with the (i,j)th entry $\int b_i''(t)b_j''(t)dt$ for $1 \le i,j \le L$. We can also use the local quadratic approximation (LQA) proposed in Fan and Li (2001) to approximate the penalty term $\mathrm{PEN}_\lambda(\beta)$ in (5) given the current estimate $\beta^{(j)}(t)$:

$$\mathrm{PEN}_\lambda(\beta) = \frac{M}{T}\int_0^T p_\lambda(|\beta(t)|)dt \approx \sum_{\ell=1}^{M} p_\lambda\left(\sqrt{\int_{t_{\ell-1}}^{t_\ell}\beta^2(t)dt}\right), \tag{8}$$

in which $t_0, t_1, \ldots, t_M$ denote the sequence of the knots of the B-spline basis functions $\mathbf{B}(t)$. As mentioned previously, B-spline basis is selected because each B-spline basis function is only nonzero for no more than d consecutive subintervals, which is called the compact support property de Boor (2001). In other words, if consecutive coefficients of the B-spline basis functions are zero, the corresponding estimated $\beta(t)$ would become strictly zero in those subintervals. Therefore, our motivation to define the functional SCAD penalty $\mathrm{PEN}_\lambda(\beta)$, instead of $p_\lambda(|\beta(t)|)$, is that we would like to shrink the estimator of $\beta(t)$ to be completely zero at a whole subinterval $[t_{\ell-1}, t_\ell]$, not at any single time point.

A large number of knots would make the sparsity detection more granular at the cost of smoothness of the resulting $\beta(t)$. However, with a roughness penalty, we suggest users choose a large number of knots and use the roughly penalty to control the smoothness of the estimated $\beta(t)$. We further express

$$\int_{t_{\ell-1}}^{t_\ell}\beta^2(t)dt = \mathbf{b}^\top\mathbf{B}_\ell\mathbf{b},$$

in which $\mathbf{B}_\ell$ denotes a $L \times L$ matrix with the (p,q)-entry as $\int_{t_{j-1}}^{t_j} b_p(t)b_q(t)dt$ when $\ell \le p, q \le \ell + d$ and zero elsewhere. Then we can derive that

$$\mathrm{PEN}_\lambda(\beta) = \frac{M}{T}\int_0^T p_\lambda(|\beta(t)|)dt \approx \left[\mathbf{b}^\top\boldsymbol{W}^{(j)}\mathbf{b} + C(\mathbf{b}^{(j)})\right], \tag{9}$$

where

$$\boldsymbol{W}^{(j)} = \frac{1}{2}\sum_{\ell=1}^{M}\left(\frac{p_\lambda'(||\beta_{[\ell]}(t)||_2\sqrt{M/T})}{||\beta_{[\ell]}(t)||_2\sqrt{T/M}}\mathbf{B}_\ell\right),$$

and

$$C(\mathbf{b}^{(j)}) \equiv \sum_{\ell=1}^{L} p_\lambda \left(\frac{||\beta^{(j)}{}_{[\ell]}||_2}{\sqrt{T/M}} \right) - \frac{1}{2} \sum_{\ell=1}^{L} p'_\lambda \left(\frac{||\beta^{(j)}{}_{[\ell]}||_2}{\sqrt{T/M}} \right) \frac{||\beta^{(j)}{}_{[\ell]}||_2}{\sqrt{T/M}}.$$

Putting (6), (7), and (9) together, given the current estimate $\beta^{(j)}(t)$, the local quadratic approximation of Q_2 in (5) can be expressed as

$$Q_2(\mathbf{b}) \approx (\mathbf{P}^{(j)} - \mathbf{T}^{(j)}\mathbf{b})^\top (\mathbf{P}^{(j)} - \mathbf{T}^{(j)}\mathbf{b}) + \left[\mathbf{b}^\top \boldsymbol{W}^{(j)}\mathbf{b} + C(\mathbf{b}^{(j)})\right] + \gamma \mathbf{b}^\top \Gamma \mathbf{b}. \quad (10)$$

Taking the derivative of Q_2 with respect to $\mathbf{b}$, we have

$$\frac{\partial Q_2(\mathbf{b})}{\partial \mathbf{b}} = -2\mathbf{T}^{(j)\top}(\mathbf{P}^{(j)} - \mathbf{T}^{(j)}\mathbf{b}) + 2\boldsymbol{W}^{(j)}\mathbf{b} + 2\gamma\Gamma\mathbf{b}.$$

Therefore, the updated estimate is $\widehat{\mathbf{b}'} = \left(\mathbf{T}^{(j)\top}\mathbf{T}^{(j)} + \boldsymbol{W}^{(j)} + \gamma\Gamma\right)^{-1} \mathbf{T}^{(j)\top}\mathbf{P}^{(j)}$. To enforce the identifiability constraint, that is, $\int_0^T \beta^2(t)dt = 1$, we normalized the estimated coefficients as $\widehat{\mathbf{b}} = \widehat{\mathbf{b}'}/\sqrt{\widehat{\mathbf{b}'}^\top \widehat{\mathbf{b}'}}$. The estimate for the index function $\beta(t)$ is $\hat{\beta}(t) = \widehat{\mathbf{b}}^\top \mathbf{B}(t)$.

In total, the algorithm contains five tuning parameters: the sparsity parameter λ, the smoothing parameter γ, the number of knots in estimating $\beta(\cdot)$, the order of B-spline basis functions, and the kernel bandwidth h in estimating g$(\cdot)$. We recommend users choose a large number of knots and use the smoothing parameter γ to control the roughness of the resulting index function $\beta(\cdot)$. For the remaining four tuning parameters, we apply a standard 10-fold cross-validation procedure to select the optimal tuning parameters both in the real data application and simulation studies.

3 | SIMULATIONS

To assess the performance of our method, we conduct several simulation studies under various scenarios. The response $y_i, i = 1, \dots, n$, is generated with the following model

$$y_i = g\left(\int_0^1 \beta(t)X_i(t)dt\right) + \varepsilon_i,$$

where $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$. We consider three different types of index function $\beta(t)$:

Case I. There is a zero region in $\beta(t)$. That is, the true $\mathcal{S}$ is a subinterval of $[0, 1]$ The following function

$$\beta(t) = \begin{cases} 2.83\sin(4\pi(t-0.2)), & t \in [0.2, 0.45], \\ 0, & \text{elsewhere}. \end{cases} \quad (11)$$

Case II. There is no zero region in $\beta(t)$. The following function

$$\beta(t) = -\sqrt{2}\sin(4\pi(t + 0.2))), t \in [0, 1] \quad (12)$$

is considered.

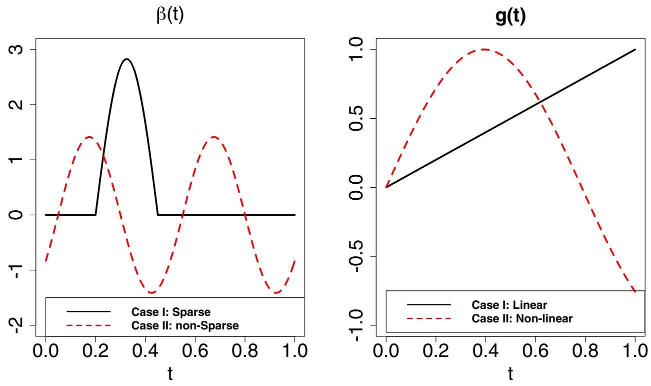

FIGURE 1 The form of $\beta(\cdot)$ on the left panel and $g(\cdot)$ in the right panel used in the simulation study. More specifically, the sparse case of index function $\beta(\cdot)$ on the left panel is defined in (11) and the nonsparse case of $\beta(\cdot)$ is defined in (12); the linear and nonlinear forms of link function are $g(x) = x$ and $g(x) = \sin(4x)$, respectively

Figure 1 shows the form of $\beta(\cdot)$ in both case I and case II on the left panel. In both cases, $\beta(t)$ is normalized to such that $||\beta||^2 = 1$. In addition, we also consider two different types of link function $g(t)$:

Case I. $g(\cdot)$ is a linear function: $g(x) = x$.
Case II. $g(\cdot)$ is a nonlinear function: $g(x) = \sin(4x)$.

Figure 1 shows the actual form of $g(\cdot)$ in both linear and nonlinear cases on the right panel. The variance of random noise, σ_ε, is taken such that the noise to signal ratio is either 10% or 20%. In each simulation setting, we generate 500 functional predictors with the corresponding responses.

We estimate both the link function $g(\cdot)$ and the index function $\beta(\cdot)$ using the proposed compact functional single index model (cFuSIM). We also apply three other existing methods, including the conventional functional single index model (FSIM), the functional nonparametric model (NPFM) proposed in Ferraty and Vieu (2006), functional linear regression model (FLR) proposed in Ramsay and Silverman (2005), locally sparse functional linear model (Slos) proposed in Lin et al. (2017). The number of B-spline basis functions are taken to be 50 and the knots are evenly distributed between 0 and 1. We compared the estimated link function $\beta(t)$ with its true value using integrated squared error (ISE) on both the zero subintervals and nonzero subintervals, respectively, by

$$\text{ISE}_1 = \frac{1}{|S|}\int_S (\hat{\beta}(t) - \beta_0(t))^2 dt,$$

and

$$\text{ISE}_0 = \frac{1}{1-|S|}\int_{t\notin S} (\hat{\beta}(t) - \beta_0(t))^2 dt,$$

where $|S|$ is the length of the support of $\beta(\cdot)$ within $[0, 1]$. In addition, prediction performance is accessed by prediction mean square errors (PMSE) on an independent test set. More specifically, the PMSE is computed as follows:

$$\text{PMSE} = \frac{1}{N}\sum_{y\in\text{test}} (y - \hat{y})^2,$$

where "test" represents the test set of size N, and $\hat{y}$ is the predicted value given the corresponding function predictor X in the test set using different methods. Due to the space limit, we only present the results with 10% noise-to-signal ratio in the manuscript. The results with 20% noise-to-signal ratio are similar, and are included in the Supplementary Material.

Table 1 displays the PMSEs, ISE_1 and ISE_0 when the true index $\beta(\cdot)$, defined in (11), is only nonzero between $[0.2, 0.45]$ and the true link function $g(\cdot)$ takes a nonlinear form: $g(x) = \sin(4x)$ with 100 repetitions. As expected, the proposed compact functional single index model (cFuSIM) is favorable under this setting in terms of both predicting the responses

TABLE 1 The prediction mean square error (PMSE), the integrated error, ISE_1 and ISE_0, defined on the nonzero subinterval and zero subintervals of the true index function $\beta(t)$ respectively on the test dataset for 100 simulation repetitions with 10% signal-to-noise ratio

Methods	PMSE (10^{-4})	ISE_1 (10^{-4})	ISE_0 (10^{-4})
cFuSIM	12.6 (3.7)	4.1 (10.6)	2.2 (2.4)
FSIM	24.6 (5.3)	9.0 (2.0)	13.9 (4.4)
Slos	3985.0 (695.3)	3998.2 (4079.8)	974.4 (1913.4)
FLR	4014.6 (501.5)	1819.3 (1794.2)	4046.9 (2881.7)
NPFM	3454.8 (274.7)	NA	NA

Note: Here, the true $\beta(\cdot)$ is defined in (11) and the true $g(x) = \sin(4x)$. Five different methods are used: the proposed compact functional single index model (cFuSIM), the traditional functional single index model (FSIM), the locally sparse functional linear model (Slos), the functional linear regression model (FLR) and the functional nonparametric method (NPFM).

TABLE 2 The prediction mean square error (PMSE), the integrated error, ISE_1 and ISE_0, defined on the nonzero subinterval and zero subintervals of the true index function $\beta(t)$ respectively on the test dataset for 100 simulation repetitions with 10% signal-to-noise ratio

Methods	PMSE (10^{-4})	ISE_1 (10^{-4})	ISE_0 (10^{-4})
cFuSIM	2.9 (0.4)	9.0 (5.1)	1.3 (1.8)
FSIM	3.0 (0.4)	9.2 (2.5)	12.8 (3.7)
Slos	2.9 (0.4)	5.3 (3.3)	0.2 (0.3)
FLR	11.4 (1.7)	180.2 (8.8)	287.2 (12.3)
NPFM	218.0 (30.4)	NA	NA

Note: Here, the true $\beta(\cdot)$ is defined in (11) and the true $g(x) = x$. Five different methods are used: the proposed compact functional single index model (cFuSIM), the traditional functional single index model (FSIM), the locally sparsed functional linear model (Slos), the functional linear regression model (FLR) and the functional nonparametric method (NPFM).

and estimating the index function. Compared to the traditional functional single index model (FSIM), cFuSIM reduces the PMSE by around 50%, ISE_1 by 55% and ISE_0 by 84%. In addition, the performance of the functional linear models such as the locally sparsed functional linear model (Slos) and the functional linear regression (FLR) is very poor compared to both cFuSIM and FSIM due to the nonlinearity of the link function $g(\cdot)$.

Table 2 displays the PMSEs, ISE_1 and ISE_0 when the true index $\beta(\cdot)$ remains sparse but the true link function $g(\cdot)$ takes a linear form: $g(x) = x$ with 100 repetitions. Both Slos and the proposed cFuSIM are favorable under this setting by allowing $\beta(\cdot)$ to be compact. Both of these two methods yields the lowest PMSEs. Among these two methods, Slos estimates $\beta(t)$ better than the proposed cFuSIM method. Specifically, Slos reduces the ISE_1 and ISE_0 by 40% and 82% respectively compared to cFuSIM, which is expected as Slos does not need to estimate the link function $g(\cdot)$. In addition, we notice that the traditional functional single index model (FSIM) predicts the response variable well. However, the estimated $\hat{\beta}(\cdot)$ is worse than both Slos and cFuSIM, as it does not consider the possibility of $\beta(\cdot)$ only being nonzero in a subinterval.

Table 3 displays the PMSEs and ISE_1 when the true index $\beta(\cdot)$ remains nonsparse as defined in (12) and the true link function $g(\cdot)$ takes the nonlinear form: $g(x) = \sin(4x)$ with 100 repetitions. The traditional FSIM model is favorable under this setting and produces the lowest PMSEs and ISE_1. The performance of the proposed cFuSIM method is very similar to the traditional FSIM model, slightly worse in terms of prediction and comparable in terms of estimating index function $\beta(\cdot)$. Furthermore, we noticed that Slos and FLR perform poorly in predicting the response value and in estimating the index functions, since they assume a linear relationship between the functional predictor and the response, which is incorrect under this setting.

Table 4 displays the PMSEs and ISE_1 when the true index $\beta(\cdot)$ takes the nonsparse form defined in (12) and the true link function $g(\cdot)$ takes the linear form: $g(x) = x$ with 100 repetitions. The traditional functional linear model (FLR) is favorable under this setting and produces the lowest PMSEs and ISE_1. Both Slos and FSIM are similar to FLR in terms of predicting the response variable. However, these two methods estimate the index function $\beta(\cdot)$ poorly. The PMSEs and ISE1 using the proposed cFuSIM method are about 1.1 times and 3.6 times compared to the FLR method.

TABLE 3 The prediction mean square error (PMSE), the integrated error and ISE_1 defined on the nonzero subinterval [0, 1] the true index function $\beta(t)$ respectively on the test dataset for 100 simulation repetitions with 10% signal-to-noise ratio

Methods	PMSE (10^{-4})	ISE_1 (10^{-4})
cFuSIM	5.6 (1.2)	0.4 (0.2)
FSIM	5.3 (0.6)	0.9 (0.8)
Slos	426.3 (52.9)	2231.9 (881.1)
FLR	412.3 (52.9)	1958.0 (1116.9)
NPFM	352.1 (25.3)	NA

Note: Here, the true $\beta(\cdot)$ is defined in (12) and the true $g(x) = \sin(4x)$. Five different methods are used: the proposed compact functional single index model (cFuSIM), the traditional functional single index model (FSIM), the locally sparsed functional linear model (Slos), the functional linear regression model (FLR) and the functional nonparametric method (NPFM).

TABLE 4 The prediction mean square error (PMSE), the integrated error and ISE_1 defined on the nonzero subinterval [0, 1] the true index function $\beta(t)$ respectively on the test dataset for 100 simulation repetitions with 10% signal-to-noise ratio

Methods	PMSE (10^{-4})	ISE_1 (10^{-4})
cFuSIM	4.5 (0.5)	136.0 (35.5)
FSIM	4.6 (0.5)	139.1 (38.6)
Slos	3.0 (0.5)	795.9 (132.3)
FLR	4.0 (0.7)	37.7 (2.5)
NPFM	222.0 (37.2)	NA

Note: Here, the true $\beta(\cdot)$ is defined in (12) and the true $g(x) = x$. Five different methods are used: the proposed compact functional single index model (cFuSIM), the traditional functional single index model (FSIM), the locally sparsed functional linear model (Slos), the functional linear regression model (FLR) and the functional nonparametric method (NPFM).

4 | REAL DATA APPLICATION

There has been an increased demand for bicycle rentals in recent years, because renting is considered as a more economical and environmentally-friendly alternative to owning bicycles. Thus, it is of great interest to ensure a sufficient bike supply, which is critical for a successful business in this area. In this paper, we try to gain a better understanding of the relationship between weather conditions and the customers' rental behavior during the weekend days. More specifically, we focus on the capital bike share study Fanaee-T and Gama (2013), which are rentals to cyclists without membership in the Capital Bike Share program in Washington, D.C. The total counts of casual bike rentals are recorded from January 1, 2011 to December 31, 2012, for a total of 105 weeks. Weather information, such as temperature, is also collected on an hourly basis.

We restrict our analysis to rentals on sunny Saturdays, in which there is a high demand for casual bike rentals compared to weekdays' bike rentals. Our goal is to study how Saturday rentals relate to the hourly temperature. Understanding the nature of this association can help predict the casual rental demand based on the weather forecast. We apply the compact functional single index model (2) to address this problem, where the scalar response Y is the total number of Saturday rentals, and the functional covariate $X(t)$ is the hourly temperature.

The compact functional single index model (2) is estimated with our proposed method. The optimal values of the sparse parameter λ, smoothing parameter γ, the bandwidth h, and the order of B-spline basis functions are chosen from a four-dimensional grid: $\{10^2, 10^4, 10^6, 10^8\} \times \{10^1, 10^3, 10^5, 10^7\} \times \{0.3, 0.5, 0.8\} \times \{4, 8, 12\}$ by 10-fold cross-validation. The internal knots of the B-spline functions are $\{1, 2, \ldots, 23\}$, as it keeps the same resolution with the functional predictor. We use the Gaussian kernel function and find that the performance of our method is quite similar among different choices of the kernel function. The optimal value of the sparsity parameter is selected to be 10^6, the smoothing parameter is 10^3, the optimal bandwidth is selected to be 0.5, and the optimal order of the B-spline basis is 4.

Figure 2 displays the estimated coefficient function $\hat{\beta}(t)$. It shows that the support of the estimated coefficient function $\beta(t)$ is between 1 to 7 pm, which indicates that the temperature mainly impacts the bike rental at this time. Besides, the

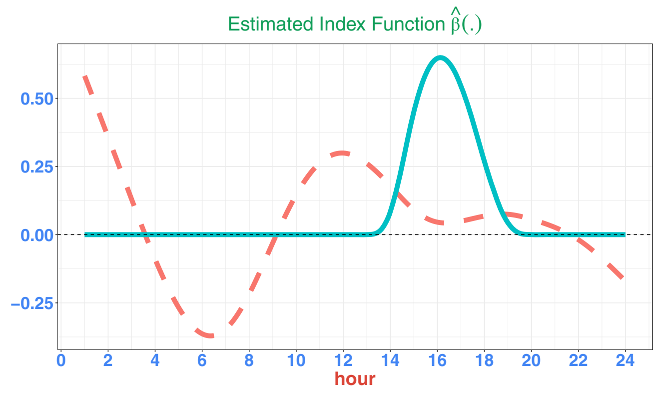

FIGURE 2 The estimated coefficient function $\hat{\beta}(t)$ using the compact FuSIM (solid line) and the conventional FuSIM (dashed line) respectively

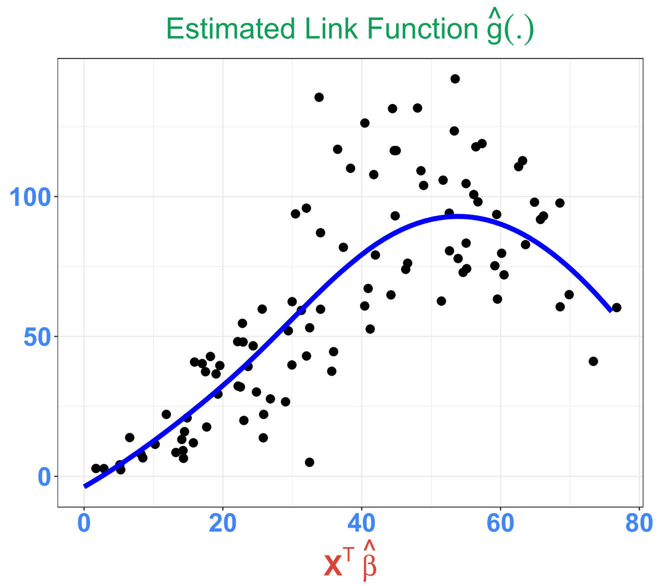

FIGURE 3 The estimated link function $\hat{g}(\cdot)$ for the compact FuSIM method (solid line). The dots represent the inner product between the daily temperature and the estimated index function $\hat{\beta}(\cdot)$. The label of the x-axis is defined as $\mathbf{X}^{\top}\hat{\beta} = \int_S X(t)\hat{\beta}(t)dt$.

estimated coefficient function starts increasing from 1 pm, reaches its peak at around 4 pm and decreases afterward until 7 pm. This observation suggests that the temperature around 4 pm is more influential to the bike rentals compared to other times throughout the day. In comparison, the estimated $\hat{\beta}(t)$ using the conventional FuSIM has no zero regions as expected. Instead, it is fluctuating around zero due to the lack of L_1 regulation.

Figure 3 shows the estimated link function $\hat{g}(t)$. The estimated link functions $\hat{g}(\cdot)$ has a nonlinear pattern. For interpretation, we plot the hourly temperature of three days, which correspond to three different levels of the integral value $\int X(t)\hat{\beta}(t)dt$ in Figure 4. It shows that when the temperature is around zero, the number of bike rentals are quite low, as shown in Day 53; when the temperature around 4 pm increases to around 20 degrees, the bike rental is at a high level as shown in Day 17; when the temperature around 4 pm is too high to around 38 degrees, the bike rental drops as shown in Day 77. Generally, the hourly temperature between 1 to 7 pm greatly affects the number of bike rentals.

Furthermore, we conduct a 10-fold cross-validation to access the prediction accuracy using the proposed compact FuSIM method. In each repetition, we select one fold of observations as the test set and leave the rest as the training set, then we estimate the compact FuSIM model using the training set only. After that, we obtain the predicted bike rentals in the test set and summarize the difference between the predicted values and the true values using mean squared prediction error (MSPE). Table 5 summarizes the MSPEs in the 10-fold cross-validation process. We also apply the same procedure on

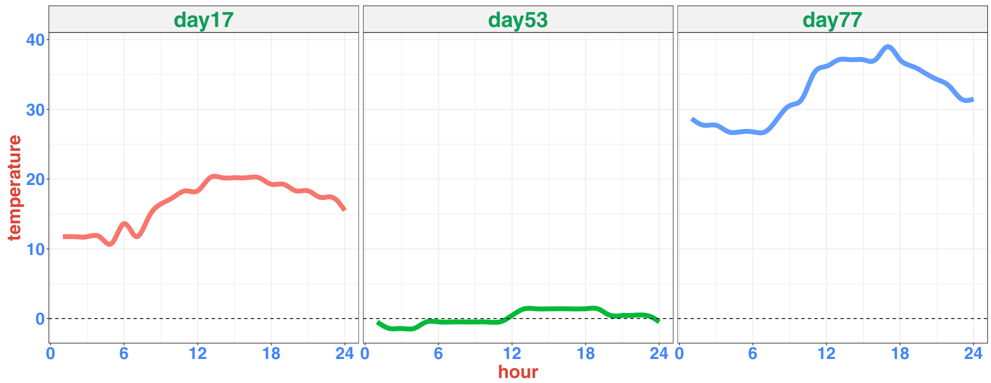

FIGURE 4 The hourly temperature for Day 17, 53, and 77. The estimated integral values, $\int_S X_i(t)\hat{\beta}(t)dt$, are 19.0, 0.8, and 35.9 and the corresponding predicted rentals are 80.8, 0.5, 56.7.

TABLE 5 The means, standard deviations (SDs), medians, minimums and maximums of MSPEs for 10-fold cross-validation repetitions using the compact FuSIM (cFuSIM), the conventional FuSIM methods (FSIM), the functional nonparametric model (NPFM), functional linear regression model (FLR), locally sparse functional linear model (Slos)

Method	Mean	SD	Median	Minimum	Maximum
cFSIM	496.75	175.54	460.09	168.36	770.84
FSIM	548.80	233.42	493.28	168.36	1017.27
FLR	724.07	444.80	616.53	249.63	1542.59
Slos	739.85	453.64	623.37	287.76	1766.16
NPFM	1663.30	466.80	1717.90	916.38	2336.64

three other existing methods, including the conventional functional single index model (FSIM), the functional nonparametric model (NPFM) proposed in Ferraty and Vieu (2006), functional linear regression model (FLR) proposed in Ramsay and Silverman (2005), locally sparse functional linear model (Slos) proposed in Lin et al. (2017). The mean squared prediction errors using our proposed method is 8.7% lower than that using the conventional functional single index model (FSIM) and more than 30% lower than those using the functional regression model and locally sparsed estimator.

5 | CONCLUDING REMARKS

In this article, we propose a novel compact functional single index model, which studies the relationship between a scalar response and a functional predictor. Our method is not only able to estimate the nonlinear relationship between the functional predictor and the scalar response without any parametric assumption, but also able to identify the region in which the functional predictor is related to the response.

We apply the compact functional single index model to predict the daily bike rental counts using the temperature data. We find that bike rental is mainly affected by the temperature between 1 and 7 pm. Besides, the temperature has a nonlinear impact on bike rental. To be more specific, the bike rentals increase with the temperature from around 0 degrees to around 20 degrees and decreases with temperature from 20 degrees until 38 degrees. We also compare our compact functional single index model with the conventional functional single index model using a simulation study. The simulation study shows that our model yields better predictions to the response and also provides a satisfactory estimation of the active region of the functional predictor.

ACKNOWLEDGMENTS

The authors would like to thank the Editor, the Associate Editor and two reviewers for their constructive comments. Their comments are very helpful to improve this work. This research was supported by Yunlong Nie's Postgraduate Scholarship-Doctorial (PGS-D) from the Natural Sciences and Engineering Research Council of Canada (NSERC), and the NSERC Discovery grants of Liangliang Wang and Jiguo Cao.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available at https://github.com/caojiguo/cFuncSIM.

ORCID

Jiguo Cao https://orcid.org/0000-0001-7417-6330

REFERENCES

Ait-Saïdi, A., Ferraty, F., Kassa, R., & Vieu, P. (2008). Cross-validated estimations in the single-functional index model. *Statistics*, *42*(6), 475–494.

Chen, D., Hall, P., & Müller, H.-G. (2011). Single and multiple index functional regression models with nonparametric link. *The Annals of Statistics*, *39*(3), 1720–1747.

de Boor, C. (2001). *A practical guide to splines. Applied mathematical sciences*. Springer.

Fan, J., & Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting: Variable bandwidth and spatial adaptation. *Journal of the Royal Statistical Society Series B (Methodological)*, *57*, 371–394.

Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, *96*(456), 1348–1360.

Fan, Y., James, G. M., & Radchenko, P. (2015). Functional additive regression. *The Annals of Statistics*, *43*(5), 2296–2325.

Fanaee-T, H., & Gama, J. (2013). Event labeling combining ensemble detectors and background knowledge. *Progress in Artificial Intelligence*, *2*, 113–127.

Ferraty, F., & Vieu, P. (2006). *Nonparametric functional data analysis*. Springer.

Goia, A., & Vieu, P. (2015). A partitioned single functional index model. *Computational Statistics*, *30*(3), 673–692.

James, G. M., Wang, J., & Zhu, J. (2009). Functional linear regression that's interpretable. *The Annals of Statistics*, *37*(5A), 2083–2108.

Jiang, C.-R., & Wang, J.-L. (2011). Functional single index models for longitudinal data. *The Annals of Statistics*, *39*(1), 362–388.

Li, J., Huang, C., Hongtu, Z., & Initiative, A. D. N. (2017). A functional varying-coefficient single-index model for functional response data. *Journal of the American Statistical Association*, *112*(519), 1169–1181.

Li, Y. (2018). On the polynomial convergence rate to nonequilibrium steady states. *Annals of Applied Probability*, *28*(6), 3765–3812.

Lin, Z., Cao, J., Wang, L., & Wang, H. (2017). Locally sparse estimator for functional linear regression models. *Journal of Computational and Graphical Statistics*, *26*(2), 306–318.

Ma, S. (2014). Estimation and inference in functional single-index models. *Annals of the Institute of Statistical Mathematics*, *68*(1), 181–208.

Marron, J. S. (1987). A comparison of cross-validation techniques in density estimation. *The Annals of Statistics*, *15*(1), 152–162.

Müller, H.-G., & Stadtmüller, U. (2005). Generalized functional linear models. *The Annals of Statistics*, *33*(2), 774–805.

Ramsay, J. O., & Silverman, B. W. (2005). *Functional data analysis* (2nd ed.). Springer.

Wand, M. P., & Jones, M. C. (1994). *Kernel smoothing*. CRC Press.

Zhou, J., Wang, N.-Y., & Wang, N. (2013). Functional linear model with zero-value coefficient function at sub-regions. *Statistica Sinica*, *23*(1), 25–50.

Zhou, X.-H., Obuchowski, N. A., & McClish, D. K. (2011). *Statistical methods in diagnostic medicine*. John Wiley & Sons, Inc.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Nie, Y., Wang, L., & Cao, J. (2022). Estimating functional single index models with compact support. *Environmetrics*, e2784. https://doi.org/10.1002/env.2784