

# ESTIMATION AND INFERENCE FOR DYNAMIC SINGLE-INDEX VARYING-COEFFICIENT MODELS

Xin Guan, Hua Liu, Jinhong You and Yong Zhou

*Zhongnan University of Economics and Law, Shanghai University of Finance and Economics, Shanghai Lixin University of Accounting and Finance and East China Normal University*

*Abstract:* Motivated by applications, we propose a class of dynamic single-index varying-coefficient models to explore the varying interaction effects on the response variable among a set of covariates. That is, the interaction effects are allowed to change with some factors of interest, such as time, spatial location, or other covariates. A spline-based approach is developed to estimate the index and varying-coefficient functions. The convergence rates and asymptotic normalities of the resulting estimators are established. It is also shown that the resulting estimators exhibit the oracle property. A penalized method is presented to select related covariates, and the consistency of the penalized estimator is proved. A test statistic is provided to check whether the interaction effect also varies with the factors of interest, and the asymptotic normality of the test statistic is established. Simulation studies and two real-data analyses illustrate the good performance of the proposed model and the corresponding statistical inference methods for finite samples.

*Key words and phrases:* Interaction effect, single-index varying-coefficient regression model, spline approximation, variable selection.

## 1. Introduction

Single-index varying-coefficient regression models (SICMs), proposed by Xia and Li (1999), have attracted much attention over the past few decades and have proven to be effective in practical applications; see Fan, Yao and Cai (2003), Ma and Song (2015), Liu, Cui and Li (2016), and the references therein. SICMs can capture the nonlinear relationship between a response and related covariates, and can achieve dimensional reduction by taking a linear combination of covariates as an index. A classical SICM is defined by

$$Y = \sum_{l=1}^d g_l(\mathbf{Z}^T \boldsymbol{\beta}) X_l + \varepsilon, \quad (1.1)$$

---

Corresponding author: Jinhong You, School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai, China. E-mail: [johnyou07@163.com](mailto:johnyou07@163.com).

where  $Y$  is the response variable,  $(\mathbf{Z}^T, \mathbf{X}^T)^T$  is a vector of covariates consisting of  $\mathbf{X} = (X_1, X_2, \dots, X_d)^T$ , with  $X_1 = 1$  and  $\mathbf{Z} \in \mathbb{R}^p$ ,  $\boldsymbol{\beta} \in \mathbb{R}^p$  is the coefficient parameter vector,  $g_l(\cdot)$ , for  $l = 1, \dots, d$ , are unknown smooth functions that describe the nonlinear relationship between  $Y$  and the covariates,  $\varepsilon$  is the error term with mean zero, and the superscript  $T$  denotes the transpose of a vector or matrix. From (1.1), it is easy to see that the effects of an interaction between  $\mathbf{X}$  and  $\mathbf{Z}$  in the SICM are static. However, this may be not true in many applications, where the interaction effects of some covariates may depend on factors of interest, such as time and spatial location, or on other covariates. For example, in an analysis of air pollution data in Guangzhou, China, we found that the interaction effects on the air quality index of the next day between four air pollutants (nitrogen dioxide, ozone, fine particles, and inhalable particles) and two environmental factors varied with wind direction; see Figures 3(a), 4(a), and 4(b). As another example, in the WeChat public accounts data, the interaction relationship between reading volume and the attributes of the posted articles varies with the proportion of articles posted during 18:00–02:00, as can be seen in Figures 3(b) and 5(a)–5(e). Thus, SICMs may be inappropriate in these situations with varying interaction effects. To explore the possible varying interaction effect patterns, we develop a class of dynamic single-index varying-coefficient models (DSICMs), defined by

$$Y = \sum_{l=1}^d g_l(\mathbf{Z}^T \boldsymbol{\beta}(U)) X_l + \varepsilon, \quad (1.2)$$

where  $U$  is a scalar, which may be one of the entries of  $\mathbf{Z}$  or  $\mathbf{X}$  or some other variable,  $\boldsymbol{\beta}(u) = (\beta_k(u), 1 \leq k \leq p)^T$  is a  $p$ -dimensional varying-coefficient function with nonparametric covariate  $U$ , and  $\varepsilon$  is an independent and identically distributed (i.i.d.) random error satisfying  $E(\varepsilon | \mathbf{Z}, \mathbf{X}, U) = 0$  and  $\text{Var}(\varepsilon | \mathbf{Z}, \mathbf{X}, U) = \sigma^2(\mathbf{Z}, \mathbf{X}, U)$ .

Obviously, the model (1.2) includes many important models as special cases. For example, when  $\boldsymbol{\beta}(U) \equiv \boldsymbol{\beta}$ , it reduces to the classical SICM (1.1), and when  $d = 1$  and  $X_l = 1$ , it reduces to the class of single-index varying-coefficient models (SIVCs) studied by Luo, Zhu and Zhu (2016) and Ma (2016) or the generalized varying-coefficient models with unspecified link functions studied by Chen, Hall and Müller (2011), Kuruwita, Kulasekera and Gallagher (2011), Zhang, Li and Xia (2015), and others. In fact, Chen, Hall and Müller (2011) were the first to study the generalized functional linear model with an unspecified link function. They approximated the link function by kernel smoothing, and established a polynomial convergence rate for it. Kuruwita, Kulasekera and Gallagher (2011)

developed a locally constant approximation to the varying-coefficient functions and the link function. They established pointwise-asymptotic properties of the estimated varying-coefficient functions. Zhang, Li and Xia (2015) provided an iterative procedure to approximate the varying-coefficient functions and the unspecified link function by using a two-dimensional kernel. Luo, Zhu and Zhu (2016) proposed an SIVC model and derived asymptotic properties of varying-coefficient functions and index functions by the optimal integration of information across all grid points. Ma (2016) considered a functional single-index model and established uniform convergence rates of coefficient functions and index functions, based on B-spline basis expansions. The convergence rates in their work were slower than the classical uniform nonparametric rates.

We propose a B-spline-based method to estimate the unknown functions in a DSICM with the model identification criterion from Ma (2016). The basic idea is to approximate the smooth index functions  $g_l(\cdot)$  using a linear assumption, and give initial estimates of the varying-coefficient functions  $\beta_k(\cdot)$ . We then fit two varying-coefficient models (VCMs) (Hastie and Tibshirani (1993)) to obtain the final spline estimators of the index functions and varying-coefficient functions, respectively. We show that the resulting estimators are consistent, and that each estimator attains the optimal convergence rate of a univariate nonparametric function. We derive the asymptotic normality of the estimators, and present asymptotic simultaneous confidence bands (SCBs) for each univariate function. Variable selection is an important aspect of nonparametric models. As in the case of the traditional single-index model (1.1), our primary concern is to select relevant predictors  $\mathbf{Z}$  that modify the relationship between  $Y$  and  $\mathbf{X}$ , that is, to pick out nonzero functions  $\beta_k(\cdot)$  in the model (1.2) and improve the estimation accuracy. Many methods and inferences for nonparametric model selection have been developed; see, for example, Wang, Li and Huang (2008) and Hu, Huang and You (2019). We propose an identification method based on a smoothly clipped absolute deviations (SCAD) penalty function for varying-coefficient functions, and estimate the varying-coefficient functions and the index functions simultaneously. We show that our estimation procedure is consistent; that is, the true model is correctly selected with probability tending to one. Another important contribution of this study is that we develop a test statistic to check whether the varying-coefficient function vector  $\beta(\cdot)$  varies with the factors of interest, and we derive the asymptotic distribution of the proposed test statistic.

The rest of this paper is organized as follows. The three-step estimation procedure is described in Section 2. The asymptotic theory of the proposed estimators is presented in Section 3. Section 4 describes the model identification

method and the hypothesis testing procedure for the varying-coefficient components. Sections 5 and 6 present simulation studies and two real-data analyses, respectively, to demonstrate the finite-sample performance of the DSICM. Some conclusions are discussed in Section 7. A description of the notation and all technical proofs are relegated to the Supplementary Material.

## 2. Spline Estimation Method

### 2.1. Model identifiability

A single-index model is not identifiable in the absence of constraints on its structure. For the semiparametric single-index model (1.1), Xia et al. (2002) and Fan, Yao and Cai (2003) developed a standard model identification method by assuming that the parameter vector has a positive first component and a norm equal to one. For the nonparametric single-index model (1.2), Kuruwita, Kulasekera and Gallagher (2011) and Luo, Zhu and Zhu (2016) employed a pointwise identification condition by simply borrowing the idea from model (1.1). Zhang, Li and Xia (2015) pointed out that the restriction to pointwise identifiability imposed some limitations. As such, they proposed a weaker model identification condition, based on a binary function that separated the direction and the norm of each varying-coefficient function. Here, similarly to Ma (2016), we adopt a model identification condition based on a spline approximation, as follows.

**Proposition 1.** *Assume that  $\|\beta(u)\|_{L_2} = 1$  and that  $\beta_1(u)$  is nonconstant and monotone nondecreasing over  $u \in S_U$ . Then, model (1.2) is identifiable.*

### 2.2. Estimation

Suppose  $\{(Y_i, \mathbf{X}_i^T, \mathbf{Z}_i^T, U_i), 1 \leq i \leq n\}$  are i.i.d. realizations of  $(Y, \mathbf{X}^T, \mathbf{Z}^T, U)$ . Let  $\mathbf{B}_1(w) = (B_{m,1}(w), 1 \leq m \leq J_{n,1})^T$  and  $\mathbf{B}_2(u) = (B_{s,2}(u), 1 \leq s \leq J_{n,2})^T$  be B-spline bases of orders  $q_1$  and  $q_2$ , respectively. The number of basis functions is defined as  $J_{n,t} = K_t + q_t$ , where  $K_t$  is the number of interior knots for a knot sequence  $\zeta_1 = \cdots = 0 = \zeta_{q_t} < \zeta_{q_t+1} < \cdots < \zeta_{q_t+K_t} < 1 = \zeta_{q_t+K_t+1} = \cdots = \zeta_{2q_t+K_t}$ , and  $K_t$  increases with the sample size  $n$ , for  $t = 1, 2$ . Here, we adopt a spline-based method to estimate the varying-coefficient functions and index functions in three steps:

**Step 0.** Obtain an initial spline approximation  $\hat{\beta}_k^{(0)}(\cdot) \approx \mathbf{B}_2(u)^T \hat{\boldsymbol{\delta}}_k^{(0)}$  by fitting a linear structure to the index functions, where  $\hat{\boldsymbol{\delta}}_k^{(0)} = (\hat{\delta}_{sk}^{(0)}, 1 \leq s \leq J_{n,2})^T$  are the estimated spline coefficients.

**Step 1.** Denote  $\Phi_{i,sk} = B_{s,2}(U_i)Z_{ik}$  and  $\Phi_i = (\Phi_{i,1}^T, \dots, \Phi_{i,p}^T)^T$ , with  $\Phi_{i,k} = (\Phi_{i,1k}, \dots, \Phi_{i,J_{n,2k}})^T$ . At iteration  $s$ , given  $\hat{\boldsymbol{\delta}}^{(s-1)} = (\hat{\boldsymbol{\delta}}_1^{(s-1),T}, \dots, \hat{\boldsymbol{\delta}}_p^{(s-1),T})^T$ , we obtain for model (1.2) the following approximated VCM:  $Y_i \approx \sum_{l=1}^d g_l(\Phi_i^T \hat{\boldsymbol{\delta}}^{(s-1)})X_{il} + \varepsilon_i$ . Suppose that  $\hat{\boldsymbol{\lambda}}^{(s)}(\hat{\boldsymbol{\delta}}^{(s-1)}) = (\hat{\lambda}_{m,l}^{(s)}, 1 \leq m \leq J_{n,1}, 1 \leq l \leq d)^T$  is obtained by minimizing  $\sum_{i=1}^n [Y_i - \sum_{l=1}^d \sum_{m=1}^{J_{n,1}} B_{m,1}(\Phi_i^T \hat{\boldsymbol{\delta}}^{(s-1)})\lambda_{m,l} X_{il}]^2$ . Denote  $W_i = \mathbf{Z}_i^T \boldsymbol{\beta}(U_i)$  and  $\widehat{W}_i^{(s)} = \mathbf{Z}_i^T \widehat{\boldsymbol{\beta}}^{(s)}(U_i) = \Phi_i^T \hat{\boldsymbol{\delta}}^{(s)}$ . Let  $\mathbf{D}_i(W_i) = \{D_{i,ml}(W_i), 1 \leq l \leq d, 1 \leq m \leq J_{n,1}\}^T$ , with  $D_{i,ml}(W_i) = B_{m,1}(W_i)X_{i,l}$ . Similarly, we define  $\mathbf{D}_i(\widehat{W}_i^{(s)})$  by replacing  $W_i$  with  $\widehat{W}_i^{(s)}$ . Let  $\mathbf{W} = (W_1, \dots, W_n)^T$  and  $\mathbf{D}(\widehat{\mathbf{W}}^{(s)}) = [(\mathbf{D}_1(\widehat{W}_1^{(s)}), \dots, \mathbf{D}_n(\widehat{W}_n^{(s)}))^T]_{n \times J_{n,1}d}$ . Thus,

$$\hat{\boldsymbol{\lambda}}^{(s)}(\hat{\boldsymbol{\delta}}^{(s-1)}) = [\mathbf{D}(\widehat{\mathbf{W}}^{(s-1)})^T \mathbf{D}(\widehat{\mathbf{W}}^{(s-1)})]^{-1} \mathbf{D}(\widehat{\mathbf{W}}^{(s-1)})^T \mathbf{Y}. \quad (2.1)$$

The spline estimator of the index function is then given by  $\hat{g}_l^{(s)}(w, \hat{\boldsymbol{\delta}}^{(s-1)}) = \sum_{m=1}^{J_{n,1}} B_{m,1}(w) \hat{\lambda}_{m,l}^{(s)}(\hat{\boldsymbol{\delta}}^{(s-1)})$ ,  $l = 1, \dots, d$ .

**Step 2.** At iteration  $s$ , when the index components are known, the model (1.2) is specified as a VCM. Suppose that  $\hat{\boldsymbol{\delta}}^{(s)}$  minimizes

$$L(\boldsymbol{\delta}) = \frac{1}{2} \sum_{i=1}^n \left[ Y_i - \sum_{l=1}^d \mathbf{B}_1(\Phi_i^T \boldsymbol{\delta})^T \hat{\boldsymbol{\lambda}}_l^{(s)}(\hat{\boldsymbol{\delta}}^{(s-1)}) X_{il} \right]^2, \quad (2.2)$$

subject to  $\delta_{11} \leq \dots \leq \delta_{J_{n,2,1}}$ . Then,  $\hat{\beta}_k^{(s)}(u) = \sum_{s=1}^{J_{n,2}} B_{s,2}(u) \hat{\delta}_{sk}^{(s)}$  for  $k = 1, \dots, p$ , and we further identify it according to Proposition 1. The constraint condition  $\delta_{11} \leq \dots \leq \delta_{J_{n,2,1}}$  is only used to ensure that  $\beta_1(u)$  is nondecreasing. We use the “constrOptim” function in R to deal with the constrained optimization problem (2.2); the gradient function is given in the Supplementary Material.

**Step 3.** Repeat Steps 1 and 2 until convergence. We then obtain  $\hat{\beta}_k(\cdot)$  and  $\hat{g}_l(\cdot)$ .

### 3. Asymptotic Properties of the Estimation Method

In this section, we obtain asymptotic results for the three-step spline estimators of the varying-coefficient components and the index components. Like Li and Lv (2022) and Ma and Song (2015), we focus on the space  $\mathcal{M}$  as the following collection of dynamic single-index varying-coefficient functions with finite  $L_2$  norm:  $\mathcal{M} = \{f : f(\mathbf{x}, \mathbf{z}, u) = \sum_{l=1}^d f_l(\mathbf{z}^T \boldsymbol{\beta}(u)) x_l, E f^2(\mathbf{x}, \mathbf{z}, u) < \infty\}$ , with  $\mathbf{z} = (z_1, \dots, z_p)^T$ ,  $\mathbf{x} = (x_1, \dots, x_d)^T$ , and  $u \in [0, 1]$ . For  $1 \leq k \leq p$ ,  $1 \leq s \leq J_{n,2}$ , we assume that  $f_{sk}^0$  satisfies

$$\mathbb{P}(\Phi_{sk}) = \sum_{l=1}^d f_{l,sk}^0(\mathbf{z}^T \boldsymbol{\beta}^0(u)) x_l = \underset{f \in \mathcal{M}}{\operatorname{argmin}} E\{\Phi_{sk} - f(\mathbf{z}^T \boldsymbol{\beta}^0(u), \mathbf{x})\}^2, \quad (3.1)$$

which implies a projection of the variable  $\boldsymbol{\Phi}$  onto the space  $\mathcal{M}$ . Let  $\mathbb{P}(\boldsymbol{\Phi}) = \{\mathbb{P}(\boldsymbol{\Phi}_1)^T, \dots, \mathbb{P}(\boldsymbol{\Phi}_p)^T\}^T$ , with  $\mathbb{P}(\boldsymbol{\Phi}_k) = \{\mathbb{P}(\boldsymbol{\Phi}_{1k}), \dots, \mathbb{P}(\boldsymbol{\Phi}_{J_{n,2k}})\}^T$ , for  $k = 1, \dots, p$ . Denote  $\tilde{\boldsymbol{\Phi}} = \boldsymbol{\Phi} - \mathbb{P}(\boldsymbol{\Phi})$ .

### 3.1. Assumptions

We make the following assumptions:

- (C1) The density function  $f_W(\cdot)$  of the random variable  $W = \mathbf{Z}^T \boldsymbol{\beta}(U)$  is bounded away from zero on  $S_W$ , and  $f_W(\cdot) \in C^{0,1}(S_W, c)$  for  $\boldsymbol{\beta}(u)$  in the neighborhood of  $\boldsymbol{\beta}^0(u)$ , with  $S_W = \{\mathbf{Z}^T \boldsymbol{\beta}(U), \mathbf{Z} \in S_Z, U \in S_U\}$ , where  $S_Z$  and  $S_U$  are compact support sets of  $\mathbf{Z}$  and  $U$ , respectively. Without loss of generality, we assume that  $S_U = [0, 1]$  and  $S_W = [a, b]$ , and that  $\mathbf{X}$  is bounded.
- (C2)  $\beta_k(\cdot) \in C^{(r)}[0, 1]$ , for  $1 \leq k \leq p$ , and  $g_l(\cdot) \in C^{(r)}[a, b]$ , for  $1 \leq l \leq d$  and some integer  $r \geq 2$ , and the spline order satisfies  $q \geq r$ .
- (C3) The conditional variance function  $\sigma^2(\mathbf{x}, \mathbf{z}, u)$  is measurable and bounded away from some constant  $0 < C_\sigma < \infty$ .
- (C4) There exist constants  $0 < c_Q \leq C_Q < \infty$  such that  $c_Q \leq Q(\mathbf{z}, u) = E(\mathbf{X}\mathbf{X}^T | \mathbf{Z} = \mathbf{z}, U = u) \leq C_Q$ , for all  $\mathbf{z} \in S_Z$  and  $u \in [0, 1]$ .
- (C5) For  $1 \leq k \leq p$ ,  $1 \leq s \leq J_{n,2}$ ,  $f_{sk}^0(\cdot) \in C^{(2)}[a, b]$ .
- (C6)  $\Xi_1 = E\{\mathbf{D}_i(W_i)\mathbf{D}_i(W_i)^T\}$ , and the eigenvalues of  $\Xi_1$  are bounded away from zero and infinity.
- (C7)  $\Xi_2 = E\{[\sum_{l=1}^d \dot{g}_l(W_i) X_{il} \tilde{\boldsymbol{\Phi}}_i]^{\otimes 2}\}$ , and the eigenvalues of  $\Xi_2$  are bounded away from zero and infinity.

**Remark 1.** Condition (C1) is the same as in Tang and Cheng (2008), Ma and Song (2015), and Zhao, Lian and Liang (2017). The boundedness of  $X$  is assumed mainly for convenience in the proof, and can be replaced by a weaker assumption; see Zhao, Lian and Liang (2017). Conditions (C2)–(C4) are common conditions for nonparametric models; see Ma and Song (2015), Ma (2016), and Liu, Cui and Li (2016). Condition (C5) is a smoothness condition on the function  $f_{sk}^0$  defined in (3.1). Conditions (C6) and (C7) are necessary to derive the asymptotic theorem.

### 3.2. Asymptotic results

According to de Boor (2001), for  $\beta_k(\cdot)$  and  $g_l(\cdot)$  satisfying Condition (C2), there exist spline functions  $g_l^0(w) = \sum_{m=1}^{J_{n,1}} B_{m,1}(w)\lambda_{m,l}^0 = \mathbf{B}_1(w)^T \boldsymbol{\lambda}_l^0$  with  $\boldsymbol{\lambda}_l^0 \in \mathbb{R}^{J_{n,1}}$  and  $\beta_k^0(u) = \sum_{s=1}^{J_{n,2}} B_{s,2}(u)\delta_{s,k}^0 = \mathbf{B}_2(u)^T \boldsymbol{\delta}_k^0$  with  $\boldsymbol{\delta}_k^0 \in \mathbb{R}^{J_{n,2}}$ , such that

$$\sup_{w \in S_W} |g_l^0(w) - g_l(w)| = O(J_{n,1}^{-r}), \quad \sup_{u \in S_U} |\beta_k^0(u) - \beta_k(u)| = O(J_{n,2}^{-r}). \quad (3.2)$$

Denote  $\boldsymbol{\delta}^0 = (\boldsymbol{\delta}_1^{0T}, \dots, \boldsymbol{\delta}_p^{0T})^T$ . The following proposition gives the uniform convergence rates of the spline estimators  $\hat{g}_l(w; \boldsymbol{\delta})$  and  $\hat{\beta}_k(u; \boldsymbol{\delta})$  near  $\boldsymbol{\delta}^0$ .

**Proposition 2.** *Under Conditions (C1) and (C2), when  $J_{n,1}^3 n^{-1} \log n = o(1)$ ,  $J_{n,1} J_{n,2}^{-r} = o(1)$ , and  $a_n = o(J_{n,1}^{-1})$ , for  $l = 1, \dots, d$ , we have (i)  $\sup_{\|\boldsymbol{\delta} - \boldsymbol{\delta}^0\|_\infty \leq a_n} \sup_{w \in S_W} |\hat{g}_l(w, \boldsymbol{\delta}) - g_l(w)| = O_p(a_n + J_{n,1}^{-r} + J_{n,2}^{-r} + (\log n)^{1/2} J_{n,1}^{1/2} n^{-1/2})$ ; (ii)  $\sup_{\|\boldsymbol{\delta} - \boldsymbol{\delta}^0\|_\infty \leq a_n} \sup_{w \in S_W} |\hat{g}_l(w, \boldsymbol{\delta}) - \dot{g}_l(w)| = O_p(J_{n,1} a_n + J_{n,1} J_{n,2}^{-r} + J_{n,1}^{1-r} + (\log n)^{1/2} J_{n,1}^{3/2} n^{-1/2})$ .*

**Remark 2.** Let  $\hat{\boldsymbol{\delta}}$  be the minimizer of  $L(\boldsymbol{\delta})$  given in (2.2) satisfying  $\|\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}^0\| \leq a_n$  with probability tending to one. This implies that we are assuming that there exists a local minimizer in the neighborhood of  $\boldsymbol{\delta}^0$  in probability. Similar restrictions are made in Xia and Li (1999) and Ma (2016).

Let  $\mathbb{B}_1(w) = \mathbf{I}_d \otimes \mathbf{B}_1(w)^T$  and  $\mathbb{B}_2(u) = \mathbf{I}_p \otimes \mathbf{B}_2(u)^T$ , where  $\mathbf{I}_d$  and  $\mathbf{I}_p$  are the  $d \times d$  and  $p \times p$  identity matrices, respectively. Let  $\mathbf{e}_l$  be the  $d \times 1$  vector with the  $l$ th element being one and all other elements being zero and let  $\mathbf{b}_k$  be the  $p \times 1$  vector with the  $k$ th element being one and all other elements being zero. Define  $d_{1n} = 1 - [2 \log(K_1 + 1)]^{-1} \{\log[-0.5 \log(1 - \alpha)] + 0.5 \{\log[\log(K_1 + 1)] + \log(4\pi)\}\}$ , and let  $d_{2n}$  be the same as  $d_{1n}$ , except that  $K_1$  is replaced by  $K_2$ . Let  $Q_{n1}(\alpha) = [2 \log(K_1 + 1)]^{1/2} d_{1n}$  and  $Q_{n2}(\alpha) = [2 \log(K_2 + 1)]^{1/2} d_{2n}$ .

The following theorems give the asymptotic results and construct asymptotic SCBs for the varying-coefficient component  $\hat{\beta}_k(\cdot)$  and the index component  $\hat{g}_l(\cdot)$ .

**Theorem 1.** *Under Conditions (C1)–(C7), if  $n^{1/(2r+2)} \ll J_{n,1} \ll n^{1/4}$  and  $n^{1/(2r+2)} \ll J_{n,2} \ll n^{1/4}$ , then we have (i)  $\|\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}(u)\|_{L_2} = O_p(J_{n,2}^{-r} + J_{n,2}^{1/2} n^{-1/2})$ ; (ii)  $\sup_{u \in S_U} |\hat{\beta}_k(u) - \beta_k(u)| = O_p(J_{n,2}^{-r} + (\log n)^{1/2} J_{n,2}^{1/2} n^{-1/2})$ ; (iii) if  $n J_{n,2}^{-(2r+1)} = o(1)$ , then  $\Omega_2^{-1}(u) [\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}(u)] \xrightarrow{d} N(0, 1)$ , where  $\Omega_2(u) = n^{-1} \mathbb{B}_2(u) \Xi_2^{-1} E\{\sigma^2(U_i, \mathbf{Z}_i, \mathbf{X}_i) [\sum_{l=1}^d \dot{g}_l(W_i) X_{il} \tilde{\boldsymbol{\Phi}}_i]^{\otimes 2}\} \Xi_2^{-1} \mathbb{B}_2(u)^T$ . Furthermore, for  $k = 1, \dots, p$ , we have  $\sigma_{2,k}^{-1}(u) [\hat{\beta}_k(u) - \beta_k(u)] \xrightarrow{d} N(0, 1)$ , where  $\sigma_{2,k}^2(u) = \mathbf{b}_k^T \Omega_2(u) \mathbf{b}_k$ .*

**Remark 3.** Let  $J_{n,2} \asymp n^{1/(2r+1)}$ . Then the spline estimates of the varying-coefficient components attain the optimal  $L_2$ -norm nonparametric rate  $O_p(n^{-r/(2r+1)})$  and the optimal uniform nonparametric rate  $O_p((\log n)^{1/2}n^{-r/(2r+1)})$ .

**Theorem 2.** Under Conditions (C1)–(C7), if  $n^{1/(2r+2)} \ll J_{n,1} \ll n^{1/4}$  and  $n^{1/(2r+2)} \ll J_{n,2} \ll J_{n,1}$ , then, for  $l = 1, \dots, d$ , we have (i)  $\sup_{w \in S_W} |\hat{g}_l(w; \hat{\boldsymbol{\delta}}) - g_l(w)| = O_p(J_{n,1}^{-r} + J_{n,2}^{-r} + (\log n)^{1/2} J_{n,1}^{1/2} n^{-1/2} + (\log n)^{1/2} J_{n,2}^{1/2} n^{-1/2})$ ; (ii) if  $J_{n,2} \ll J_{n,1}$  and  $nJ_{n,2}^{-(2r+1)} = o(1)$ , then  $\sigma_{1,l}^{-1}(w)[\hat{g}_l(w; \hat{\boldsymbol{\delta}}) - g_l(w)] \xrightarrow{d} N(0, 1)$ , where  $\sigma_{1,l}^2(w) = n^{-1} \mathbf{e}_l^T \mathbb{B}_1(w) \Xi_1^{-1} [E(\sigma^2(U_i, \mathbf{Z}_i, \mathbf{X}_i) \mathbf{D}_i(W_i)^{\otimes 2})] \Xi_1^{-1} \mathbb{B}_1^T(w) \mathbf{e}_l$ .

**Remark 4.** Let  $J_{n,1} \asymp J_{n,2} \asymp n^{1/(2r+1)}$ . Then, the spline estimates of the index components attain the optimal uniform nonparametric rate  $O_p((\log n)^{1/2}n^{-r/(2r+1)})$ . The conditions  $J_{n,2} \ll J_{n,1}$  and  $nJ_{n,2}^{-(2r+1)} = o(1)$  are only imposed to make the bias asymptotically negligible. It can be seen that the asymptotic variance of  $\hat{\beta}_k(\cdot)$  obtained from Theorem 1 is the same as would be obtained if all the index functions  $g_l(\cdot)$ , for  $l = 1, \dots, d$ , were known. Analogously, the asymptotic variance of  $\hat{g}_l(\cdot)$  obtained from Theorem 2 is the same as would be obtained if all the  $\beta_k(\cdot)$ , for  $k = 1, \dots, p$ , were known. Hence, the proposed spline estimator has the oracle property. If  $\sigma^2(\mathbf{Z}, \mathbf{X}, U) = \sigma^2$ , then the asymptotic variance matrices reduce to  $\sigma_{1,l}^2(w) = n^{-1} \sigma^2 \mathbf{e}_l^T \mathbb{B}_1(w) \Xi_1^{-1} \mathbb{B}_1^T(w) \mathbf{e}_l$  and  $\sigma_{2,k}^2(u) = n^{-1} \sigma^2 \mathbf{b}_k^T \mathbb{B}_2(u) \Xi_2^{-1} \mathbb{B}_2^T(u) \mathbf{b}_k$ .

**Theorem 3.** Under Conditions (C1)–(C7), if  $n^{1/(2r+2)} \ll J_{n,1} \ll n^{1/4}$ ,  $n^{1/(2r+2)} \ll J_{n,2} \ll n^{1/4}$ , and  $nJ_{n,2}^{-(2r+1)} = o(1)$ , then, for any  $\alpha \in (0, 1)$ , we have  $\lim_{n \rightarrow \infty} P[\sup_{u \in S_{n,2}} |\sigma_{2,k}^{-1}(u) \{\hat{\beta}_k(u) - \beta_k(u)\}| \leq Q_{n2}(\alpha)] = 1 - \alpha$ , where  $S_{n,2}$  is a subset of  $S_U$  that becomes denser as  $n \rightarrow \infty$ . Therefore, an asymptotic  $100(1 - \alpha)\%$  SCB for  $\beta_k(u)$  over  $u \in S_{n,2}$  is  $\hat{\beta}_k(u) \pm \sigma_{2,k}(u) Q_{n2}(\alpha)$ .

**Theorem 4.** Under Conditions (C1)–(C7), if  $n^{1/(2r+2)} \ll J_{n,1} \ll n^{1/4}$ ,  $n^{1/(2r+2)} \ll J_{n,2} \ll n^{1/4}$ ,  $J_{n,2} \ll J_{n,1}$ , and  $nJ_{n,2}^{-(2r+1)} = o(1)$ , then, for any  $\alpha \in (0, 1)$ , we have  $\lim_{n \rightarrow \infty} P[\sup_{w \in S_{n,1}} |\sigma_{1,l}^{-1}(w) \{\hat{g}_l(w; \hat{\boldsymbol{\delta}}) - g_l(w)\}| \leq Q_{n1}(\alpha)] = 1 - \alpha$ , where  $S_{n,1}$  is a subset of  $S_W$  that becomes denser as  $n \rightarrow \infty$ . Therefore, an asymptotic  $100(1 - \alpha)\%$  SCB for  $\hat{g}_l(w; \hat{\boldsymbol{\delta}})$  over  $w \in S_{n,1}$  is  $\hat{g}_l(w; \hat{\boldsymbol{\delta}}) \pm \sigma_{1,l}(w) Q_{n1}(\alpha)$ .

**Remark 5.** Compared with the asymptotic  $100(1 - \alpha)\%$  pointwise confidence bands (PCBs),  $\hat{g}_l(u) \pm \sigma_{1,l}(u) z_\alpha$  and  $\hat{\beta}_k(u) \pm \sigma_{2,k}(u) z_\alpha$ , the widths of the SCBs are inflated by rates  $[2 \log(K_1 + 1)]^{1/2} d_{1n}/z_\alpha$  and  $[2 \log(K_2 + 1)]^{1/2} d_{2n}/z_\alpha$ , respectively, where  $z_\alpha$  is the  $\alpha$ -quantile of the standard normal distribution. To establish the asymptotic confidence bands of  $\hat{g}_l(\cdot)$  and  $\hat{\beta}_k(\cdot)$ , the key issue is to obtain accurate variance estimates  $\hat{\sigma}_{1,l}$  and  $\hat{\sigma}_{2,k}$ . On the suggestion of Yang (2008) and



Ma et al. (2015), we adopt a wild bootstrap procedure to approximate the asymptotic variances, because the plug-in method may lead to inaccurate results for finite samples.

## 4. Model Selection

### 4.1. Model identification

In this subsection, we present a model identification procedure, based on a penalized method, for selecting significant variables  $Z_k$  with nonzero varying-coefficient functions  $\beta_k(\cdot)$ . Without loss of generality, we assume that  $\beta_k(u)$ , for  $k = 1, \dots, p_1$ , are nonzero varying-coefficient functions, and  $\beta_k(u) = 0$  for  $k = p_1 + 1, \dots, p$  in the true model.

Denote  $\mathbf{H} = (h_{ij})_{J_{n,2} \times J_{n,2}}$  with entries  $h_{ij} = \int_U B_{i,2}(u)B_{j,2}(u) du$ . Let  $\boldsymbol{\pi} = (\boldsymbol{\pi}_1^T, \dots, \boldsymbol{\pi}_p^T)^T$ , with  $\boldsymbol{\pi}_k = (\pi_{1k}, \dots, \pi_{J_{n,2}k})^T$ , for  $1 \leq k \leq p$ . For  $\beta_k(u) = \mathbf{B}_2(u)^T \boldsymbol{\pi}_k$ , let  $\|\beta_k(u)\|_{L_2}^2 = \boldsymbol{\pi}_k^T \mathbf{H} \boldsymbol{\pi}_k = \|\boldsymbol{\pi}_k\|_{\mathbf{H}}^2$ . We adopt the penalized approach for varying-coefficient functions presented in Wang, Li and Huang (2008). Define the objective function with penalization as

$$\begin{aligned} Q(\boldsymbol{\pi}) &= \frac{1}{2} \sum_{i=1}^n \left[ Y_i - \sum_{l=1}^d \hat{g}_l(\boldsymbol{\Phi}_i^T \boldsymbol{\pi}) X_{il} \right]^2 + n \sum_{k=2}^p p_{\alpha_n}(\|\boldsymbol{\pi}_k\|_{\mathbf{H}}) \\ &= L(\boldsymbol{\pi}) + n \sum_{k=2}^p p_{\alpha_n}(\|\boldsymbol{\pi}_k\|_{\mathbf{H}}), \end{aligned} \quad (4.1)$$

where the functions  $\hat{g}_l(\cdot)$  are known,  $p_{\alpha_n}$  is the SCAD penalty function proposed by Fan and Li (2001), and  $\alpha_n$  is the tuning parameter. We find the minimizer of (4.1) as follows:

**Step 1:** Initialize  $\boldsymbol{\pi} = \boldsymbol{\pi}^{(1)}$ .

**Step 2:** Given  $\boldsymbol{\pi}^{(m)}$ , update  $\boldsymbol{\pi}^{(m+1)}$  by minimizing (4.1).

**Step 3:** Iterate Step 2 until  $\|\boldsymbol{\pi}_k^{(m+1)} - \boldsymbol{\pi}_k^{(m)}\|_{\mathbf{H}} < \epsilon_1$ .

In the implementation of the algorithm, the initial value  $\boldsymbol{\pi}^{(1)}$  is given by the unpenalized estimators. To update  $\boldsymbol{\pi}^{(m+1)}$  in Step 2, we apply an iterative process similar to Step 2 in Section 2, but with an additional penalty term. If the  $m$ th iterative estimator  $\boldsymbol{\pi}^{(m)}$  is close to zero, that is,  $\|\boldsymbol{\pi}^{(m)}\|_{\mathbf{H}}$  is smaller than a positive threshold value  $\epsilon_2$ , then we set  $\hat{\beta}_k(u) = 0$ . We set  $\epsilon_1 = \epsilon_2 = 0.01$  in the simulation.

The penalized estimator of  $\boldsymbol{\pi}$  is denoted by  $\hat{\boldsymbol{\pi}} = \{\hat{\pi}_{sk}, 1 \leq s \leq J_{n,2}, 1 \leq k \leq p\}^T = \operatorname{argmin}_{\boldsymbol{\pi}} Q(\boldsymbol{\pi})$ , and the penalized spline estimators are given by  $\hat{\beta}_k^P(u) = \mathbf{B}_2(u)^T \hat{\boldsymbol{\pi}}_k$ , for  $k = 1, \dots, p$ . We then have the following results.

**Theorem 5.** *Under Conditions (C1)–(C7), if  $\alpha_n \rightarrow 0$ ,  $J_{n,2}^{-r}/\alpha_n \rightarrow 0$ , and  $J_{n,2}^{1/2}n^{-1/2}/\alpha_n \rightarrow 0$  as  $n \rightarrow \infty$ , then we have (i)  $\hat{\beta}_k^P(u) = 0$ , for  $k = p_1 + 1, \dots, p$ , with probability tending to one; (ii)  $\|\hat{\beta}_k^P(u) - \beta_k(u)\|_{L_2} = O_p(J_{n,2}^{1/2}n^{-1/2} + J_{n,2}^{-r})$ , for  $k = 1, \dots, p_1$ .*

## 4.2. Hypothesis testing

The model (1.2) is quite general, because it does not depend on any specific parametric structure. However, it is worth considering whether it can be reduced to a simpler form. In this subsection, we propose an  $L_2$ -distance statistic to complete the testing problem.

We are interested in testing whether the model (1.2) can degenerate to the single-index model given in (1.1), that is,

$$H_0 : \boldsymbol{\beta}(u) = \boldsymbol{\beta}, \quad \text{a.s. on } [0, 1].$$

The null hypothesis can be rewritten as  $H_0 : \int_0^1 \|\boldsymbol{\beta}(u) - \boldsymbol{\beta}\|^2 du = 0$ . Let  $\hat{\boldsymbol{\beta}}(u)$  be the spline estimators of model (1.2), and let  $\hat{\boldsymbol{\beta}}$  be the parametric estimator of model (1.1). We define the  $L_2$ -distance test statistic as

$$\mathcal{T} = \frac{n}{J_{n,2}} \int_0^1 \|\hat{\boldsymbol{\beta}}(u) - \hat{\boldsymbol{\beta}}\|^2 du. \quad (4.2)$$

The following theorem presents the asymptotic behavior of the test statistic under the null hypothesis.

**Theorem 6.** *Under Conditions (C1)–(C7), if  $n^{1/(2r+1)} \ll J_{n,1} \ll n^{1/4}$  and  $n^{1/(2r+1)} \ll J_{n,2} \ll n^{1/4}$ , then, under  $H_0$ ,  $\mathcal{T} - B \xrightarrow{d} N(0, V)$ , where  $B = (n/J_{n,2}) \int_0^1 \Omega_2(u) du$  and  $V = (8n^2/J_{n,2}^2) \int_0^1 \int_0^1 \Omega_2(u)\Omega_2(u') du du'$ .*

To demonstrate the power of the test, we define a sequence of local alternatives that converge to the null hypothesis with increasing sample size. We consider  $H_1 : \boldsymbol{\beta}_n(u) = \boldsymbol{\beta} + c_n \Delta(u)$ , where  $c_n \rightarrow 0$  as  $n \rightarrow \infty$ , and  $\Delta(u)$  is a continuous function. We show the asymptotic distribution of the test statistic under  $H_1$  in the following theorem.

**Theorem 7.** *Under Conditions (C1)–(C7), if  $n^{1/(2r+1)} \ll J_{n,1} \ll n^{1/4}$  and  $n^{1/(2r+1)} \ll J_{n,2} \ll n^{1/4}$ , and we let  $c_n^2 = J_{n,2}/n$ , then, under  $H_1$ ,  $\mathcal{T} - B \xrightarrow{d}$*

$N(\tilde{\Delta}, V)$ , where  $\tilde{\Delta} = \int_0^1 \Delta^2(u) du$ .

Although a plug-in procedure can be used to evaluate the asymptotic distribution of the test statistic, we adopt a wild bootstrap method to improve the performance of the test. The steps can be found in the Supplementary Material.

## 5. Numerical Studies

In this section, we conduct simulation studies to illustrate the good performance of the proposed model and the corresponding statistical inference methods for finite samples.

### 5.1. Selecting the smoothing parameter

We use spline order  $q_1 = q_2 = 3$  throughout the numerical simulation studies. For each univariate function, the positions of the interior knots are chosen as equally spaced sample quantiles. The smoothing parameters  $K_1$  and  $K_2$  are selected by minimizing the following Bayesian information criterion (BIC):  $\text{BIC}(K_1, K_2) = \log\{n^{-1} \sum_{i=1}^n [Y_i - \sum_{l=1}^d \hat{g}_l(\mathbf{Z}_i^T \hat{\boldsymbol{\beta}}(U_i)) X_{il}]^2\} + df \log(n)/n$ , where  $df = p(q + K_1) + d(q + K_2)$ . That is, the optimal number of interior knots is given by  $(\hat{K}_1, \hat{K}_2) = \text{argmin}_{K_1, K_2 \in K_T} \text{BIC}(K_1, K_2)$ , where  $K_T$  is the potential number of interior knots.

For the model identification procedure, we set  $a = 3.7$  in the SCAD function, as in Fan and Li (2001). The tuning parameter  $\alpha_n$  is chosen by minimizing  $\text{BIC}(\alpha_n) = \log\{n^{-1} \sum_{i=1}^n [Y_i - \sum_{l=1}^d \hat{g}_l(\mathbf{Z}_i^T \hat{\boldsymbol{\beta}}^P(U_i)) X_{il}]^2\} + (p - p_1)(q + K_1) \log(n)/n$ . The optimal choice is obtained as  $\hat{\alpha}_n = \text{argmin} \text{BIC}(\alpha_n)$ .

### 5.2. Numerical examples

**Example 1.** We consider the following model:

$$Y_i = g_1(\mathbf{Z}_i^T \boldsymbol{\beta}(U_i)) X_{i1} + g_2(\mathbf{Z}_i^T \boldsymbol{\beta}(U_i)) X_{i2} + \varepsilon_i, \quad i = 1, \dots, n. \quad (5.1)$$

We independently generate covariates  $X_1$  and  $X_2$  from a bivariate normal distribution with mean zero, variance one, and covariance 0.3. Let  $(Z_{i1}^*, Z_{i2}^*, Z_{i3}^*)^T$  follow a multivariate normal distribution with mean zero, variance one, and covariance 0.2, and then scale it as  $Z_{ik} = \Phi(Z_{ik}^*) - 0.5$ , where  $\Phi(\cdot)$  is the cumulative distribution function of the standard normal distribution. In addition,  $U_i$  is generated from the uniform distribution  $U(0, 1)$ . The varying coefficients  $\boldsymbol{\beta}(u) = (\beta_1(u), \beta_2(u), \beta_3(u))^T$  are given by  $\beta_1(u) = 1 + u^2$ ,  $\beta_2(u) = (1 - u)^2$ , and  $\beta_3(u) = -1 + 4(u - 0.5)^2$ , and they are then identified as  $\boldsymbol{\beta}(u)/\|\boldsymbol{\beta}(u)\|_{L_2}$ . Set the index functions as  $g_1(w) = 10 \exp(5w)/\{1 + \exp(5w)\} - E[10 \exp(5W)/\{1 + \exp(5W)\}]$ ,

Table 1. Means and standard deviations (in parentheses) of the RMSEs of the estimators for each component function.

$n$	$\beta_1$	$\beta_2$	$\beta_3$	$g_1$	$g_2$
300	0.0682(0.0313)	0.0711(0.0270)	0.0717(0.0258)	0.4450(0.1076)	0.4923(0.1132)
500	0.0523(0.0222)	0.0509(0.0186)	0.0508(0.0181)	0.3210(0.0698)	0.3546(0.0761)
700	0.0430(0.0167)	0.0397(0.0134)	0.0402(0.0143)	0.2619(0.0546)	0.2872(0.0557)

$g_2(w) = 5 \sin(\pi w) - E[5 \sin(\pi W)]$ . The error term  $\varepsilon_i$  follows an independent normal distribution  $N(0, \sigma^2(\mathbf{X}_i, \mathbf{Z}_i, U_i))$  with  $\sigma^2(\mathbf{X}_i, \mathbf{Z}_i, U_i) = 0.2 \text{Var}\{E(Y_i | \mathbf{X}_i, \mathbf{Z}_i, U_i)\}$ .

To evaluate the performance of the proposed spline estimators, we use the root mean squared error (RMSE) criterion. For any function  $f(u)$ , the RMSE is given as  $\text{RMSE}(f) = \{n^{-1} \sum_{i=1}^n [\hat{f}(u_i) - f(u_i)]^2\}^{1/2}$ . Table 1 lists the means and standard deviations (in parentheses) of the RMSEs of the estimators for each component function with sample size  $n = 300, 500, \text{ or } 700$ , based on 500 Monte Carlo replications. It is obvious that the mean and standard deviation of the RMSEs of all estimators decrease as the sample size increases.

Figures 1(a)–1(e) show the estimated functions (dashed), 95% SCBs (dash-dotted), and PCBs (dotted) for each estimator with sample size  $n = 300$ . The estimated functions are close to the true curves (solid), and they all fall into the PCBs and SCBs. We also compare the 95% PCBs based on the asymptotic results and the wild bootstrap method described in Section 4.2. It can be seen from Figure 1 in the Supplementary Material that for all component functions, these two types of confidence bands are close to each other, which means that the wild bootstrap method performs well.

**Example 2.** To evaluate the power of the test presented in Section 4.2, we test whether the varying-coefficient components  $\beta(u)$  in the model (5.1) are constant. A sequence of alternative models is specified as  $H_1 : \beta(u, \theta) = \mathbf{c} + \theta\beta(u)$ , where  $\mathbf{c} = (c_k, 1 \leq k \leq 3)^T$ , with  $c_k = \int_0^1 \beta_k(u) du$  and  $\theta = \{0, 0.1, 0.2, 0.3, 0.5, 0.8\}$ . Let the index components  $\mathbf{g}(\cdot)$  and covariates  $\mathbf{Z}$  and  $\mathbf{X}$  be the same as in Example 1. Let  $U$  be equidistant sequences from zero to one. The error term follows the same distribution as in Example 1. When  $\theta = 0$ , it reduces to the null-hypothesis model.

For different sample sizes and  $\theta$ , we carry out the test procedure 300 times with sample sizes  $n = 300, 500, \text{ and } 700$ . For each replication, we draw  $B = 500$  bootstrap samples. The significance level  $\alpha$  is 0.05. We give the simulated powers in Figure 1(f). We can conclude that under the null hypothesis, the

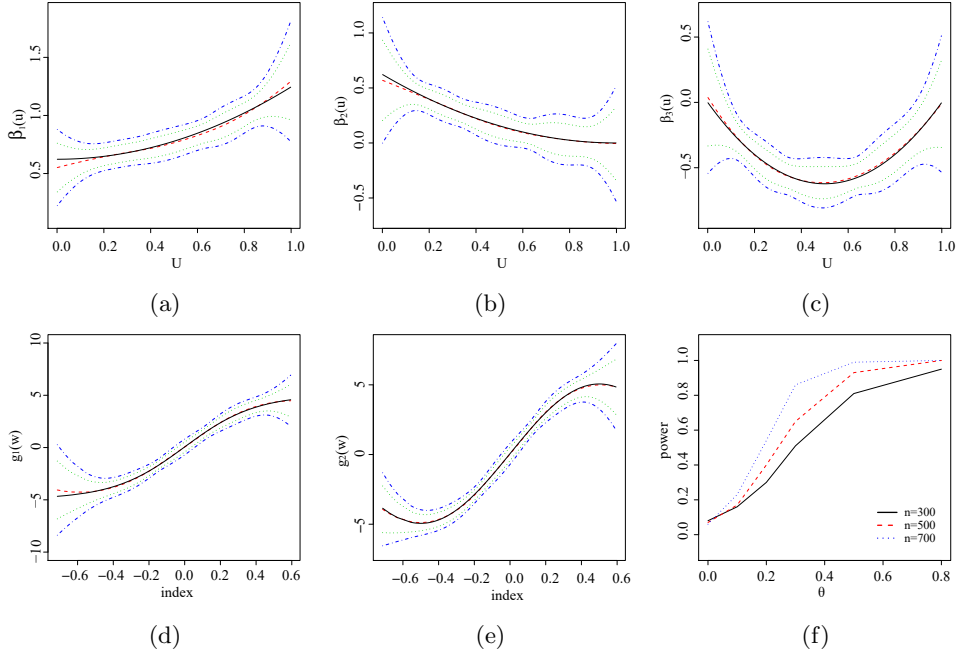


Figure 1. (a)–(c) Curves of varying-coefficient functions  $\beta_k(\cdot)$ , for  $k = 1, 2, 3$ . (d) and (e) Curves of index functions  $g_l(\cdot)$ , for  $l = 1, 2$ . The true functions are shown by solid curves, the spline estimators by dashed curves, the 95% PCBs by dotted curves, and the 95% SCBs by dash-dotted curves. (f) Simulated powers for different sample sizes.

estimated powers are close to 0.05, which is consistent with the theoretical result. Furthermore, the estimated powers increase to one as  $\theta$  increases, indicating that the proposed test statistic works well.

### 5.3. Comparison with the random forest approach

Prediction is an important concern in statistical modeling. To investigate the prediction performance of the proposed model and estimation, we compare the coverage probability and width of the prediction confidence intervals with three random forest approaches discussed in Zhang et al. (2020), which are out-of-the-bag prediction intervals (OOB), quantile regression forests (QRF), and split conformal prediction intervals (SC). We generate the data according to the settings in Example 1, and rewrite model (5.1) as  $Y = \mathbf{g}(\tilde{\mathbf{X}}) + \epsilon$ , where the predictor  $\tilde{\mathbf{X}} = (\tilde{X}_1, \dots, \tilde{X}_{\tilde{p}})^T = (\mathbf{X}^T, \mathbf{Z}^T, U)^T$  with  $\tilde{p} = 6$ .

We set the nominal level at  $(1 - \tilde{\alpha})$ , with  $\tilde{\alpha} = 0.1$ . Let the training sample sizes  $n = 200, 300, 500$ , and  $700$  and the test sample sizes  $n_0 = 20$ . Based on 200 replications, Figures 2(a) and 2(b) present the estimated coverage probability

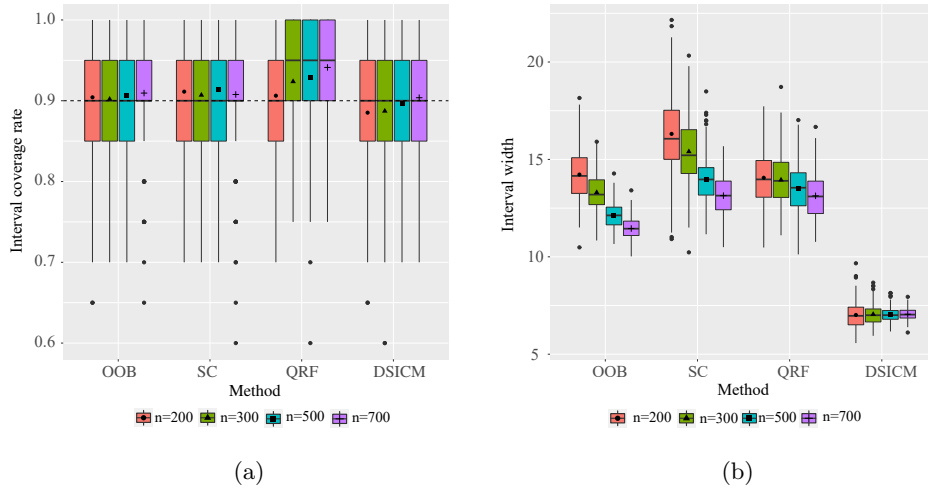


Figure 2. (a) and (b) show the estimated coverage probability and width, respectively, of the prediction confidence interval for OOB, QRF, SC, and DSICM in box plots. The points of different shapes in box plots represent the mean value of the coverage probability or width of the prediction confidence interval under different sample sizes.

and width, respectively, of the prediction confidence interval for OOB, QRF, SC, and the proposed model (DSICM). The points of different shapes in box plots represent the mean value of the coverage probability or width of the prediction confidence interval under different sample sizes. As shown in Figure 2(a), the coverage probabilities of OOB, SC, and DSICM are close to the nominal level of 0.9, while the QRF estimators are over covered. Figure 2(b) shows that the width of the prediction confidence interval of DSICM is shorter than those based on the other methods.

## 6. Real-Data Analysis

In this section, we illustrate the proposed model and statistical inference methods by analyzing two real-data applications.

### 6.1. Air pollution data for Guangzhou, China

This data set consists of 2,174 observations between January 1, 2014, and December 31, 2019, after the exclusion of several missing samples. The data are available on the website: <https://www.aqistudy.cn>. The aim of this example is to study the potential relationship between the air quality index (AQI) and predictors of the previous day.

Let  $Y$  be the logarithm of the AQI for the next day. The predictors include

the daily average levels of pollutants commonly used in the literature, namely, ozone  $Z_1$  ( $\mu\text{g}/\text{m}^3$ ), nitrogen dioxide  $Z_2$  ( $\mu\text{g}/\text{m}^3$ ), fine particles  $Z_3$  ( $\mu\text{g}/\text{m}^3$ ), and inhalable particles  $Z_4$  ( $\mu\text{g}/\text{m}^3$ ), and two environmental factors, namely, temperature  $X_2$  and humidity  $X_3$ . Owing to the particular location and landform of Guangzhou, its climate is mainly influenced by the East Asian monsoon, and so the interaction effect of air pollutants and environmental factors may vary with wind direction.

To begin, we split the data into three groups according to the angles of wind direction. The angles of the first group are  $1^\circ$ – $120^\circ$ , those of the second group are  $120^\circ$ – $240^\circ$ , and those of the third group are  $240^\circ$ – $360^\circ$ . We then fit each group using the SICM

$$Y = g_1(\mathbf{Z}^T \boldsymbol{\beta}) + g_2(\mathbf{Z}^T \boldsymbol{\beta})X_2 + g_3(\mathbf{Z}^T \boldsymbol{\beta})X_3 + \varepsilon, \quad (6.1)$$

where  $g_l(\cdot)$ ,  $l = 1, 2, 3$ , are unknown smooth functions,  $\mathbf{Z} = (Z_1, Z_2, Z_3, Z_4)^T$ , and  $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)^T$  is the coefficient vector. The covariates  $\mathbf{Z}$  are normalized. Figure 3(a) shows the estimated coefficient vectors  $(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4)^T$  for three groups: Group 1 (solid lines), Group 2 (long dashed lines), and Group 3 (dashed lines). Note that the parameters  $\boldsymbol{\beta}$  are different in the three groups, and so the interaction pattern may be affected by the wind direction. To further explore the dynamic interaction between the response and the covariates, we let  $U = \text{Direction}/360$ , and consider the following model:

$$Y = g_1(\mathbf{Z}^T \boldsymbol{\beta}(U)) + g_2(\mathbf{Z}^T \boldsymbol{\beta}(U))X_2 + g_3(\mathbf{Z}^T \boldsymbol{\beta}(U))X_3 + \varepsilon, \quad (6.2)$$

where  $\boldsymbol{\beta}(\cdot) = (\beta_k(\cdot), 1 \leq k \leq 4)^T$ .

By conducting the variable selection described in Section 4.1, we find that the covariates  $Z_1$  and  $Z_2$  are significant. The optimal parameters are chosen as  $(q_1, q_2, K_1, K_2) = (3, 3, 2, 2)$  using the BIC. We perform the hypothesis test developed in Section 4.2 with 500 wild bootstrap resamples, and find that the test  $p$ -value is 0.036. Thus, we reject  $H_0$  [the model (6.1)] at the 0.05 significance level.

Figures 4(a) and 4(b) show the resulting spline estimators (solid curves) and the corresponding 95% SCBs (dashed curves) based on 500 wild bootstrap samples for  $\beta_k(u)$ ,  $k = 1, 2$ . From Figure 4(a), it can be seen that the interaction effect of ozone increases with increasing wind direction, and that their combination has a positive effect on the response. Figure 4(b) reveals that the interaction relationship between nitrogen dioxide and wind direction has a slowly decreasing trend as the wind direction increases, and that their interaction has

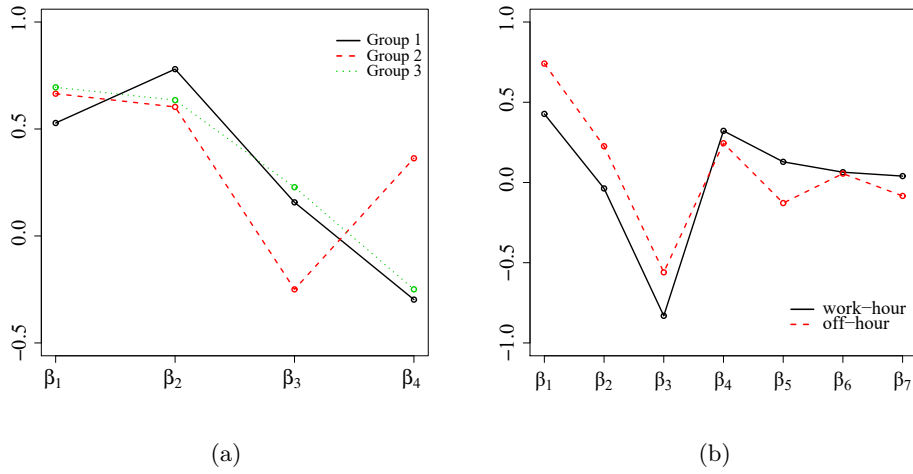


Figure 3. The lines connect the estimated coefficient vectors of different groups for two sets of real data: (a) Group 1 (solid lines), Group 2 (long dashed lines), and Group 3 (dashed lines); (b) work-hour (solid lines) and off-hour (dashed lines).

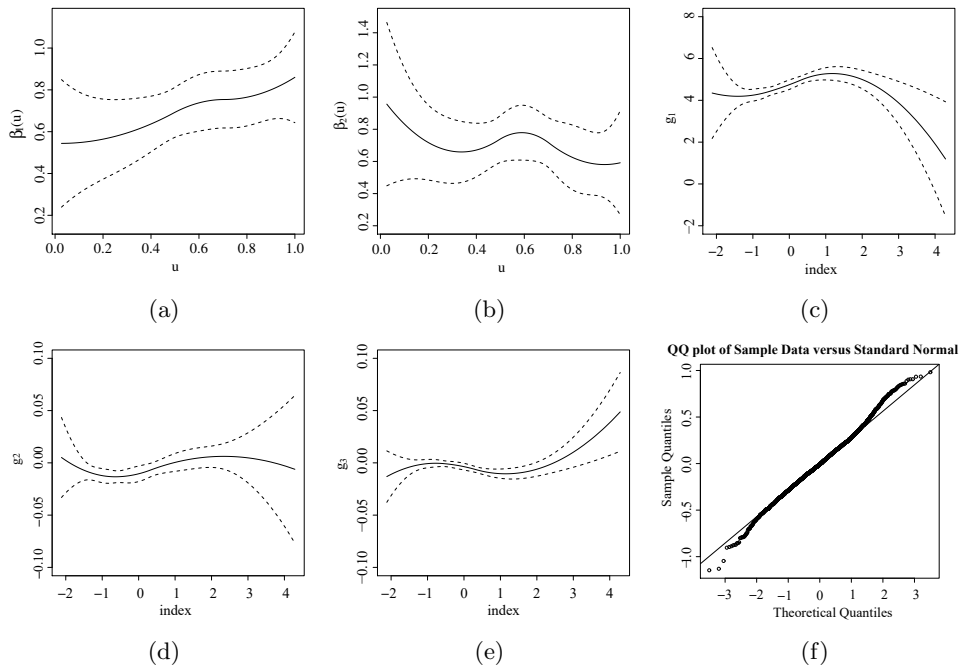


Figure 4. (a) and (b) Estimators (solid curves) of varying-coefficient functions and 95% SCBs (dashed curves) for  $\beta_k(\cdot)$ ,  $k = 1, 2$ , respectively. (c)–(e) Spline estimates (solid curves) and 95% SCBs (dashed curves) for index functions  $g_l(\cdot)$ ,  $l = 1, 2, 3$ . (f) *QQ* plot of residuals.



a positive nonlinear impact on the response. The spline estimators of the index functions  $g_l(\cdot)$ , for  $l = 1, 2, 3$ , and the corresponding 95% SCBs are depicted in Figures 4(c)–4(e). Figure 4(c) reveals that there is a truly nonlinear relationship between the median value  $\log(\text{AQI})$  and the predictors. Figures 4(d) and 4(e) show that the modifications due to the temperature and humidity of the previous day are altered by the admixture of other variables. In addition, the  $QQ$  plot of the residuals in Figure 4(f) shows that the model (6.2) is a reasonable option for this data set.

## 6.2. WeChat public accounts data

WeChat has become an extremely popular mobile application in recent years. The public account is one of the most important and most popular functions of WeChat, providing users with information and services. On the other hand, WeChat public accounts have a mature profit model, which is related to the amount of reading. The reading volume of a public account reflects its number of active fans and directly affects the income of the official accounts. The purpose of this study is to analyze the association between reading volume and related covariates.

The data set is obtained from a public account analysis website (<http://top.askci.com>), and contains information of 624 public accounts in March 2017. The response  $Y$  is the logarithm of reading volume. The covariates in the index  $\mathbf{Z} = (Z_1, Z_2, Z_3, Z_4, Z_5, Z_6, Z_7)^T$  include the type of public account ( $Z_1 = 1$  for enterprises and  $Z_1 = 0$  for individuals), the average number of daily posts ( $Z_2$ ), the ratio of original posts ( $Z_3$ ), the proportion of posts with video ( $Z_4$ ), the average title length of the last 10 articles ( $Z_5$ ), the title punctuation index of the last 10 articles ( $Z_6$ ), and the length of the account name ( $Z_7$ ). Covariates of interest also include the positive emotion score of titles in the last 10 articles ( $X_2$ ). Here, we take  $X_1 = 1$  as the intercept term and let  $\mathbf{X} = (X_1, X_2)^T$ .

This data set records the proportion of articles posted during six periods (from 18:00, every four hours is a period). We found that the time of posting articles is concentrated in two periods, 18:00–22:00 and 22:00–02:00. Note that the interaction effects between  $\mathbf{X}$  and the index  $\mathbf{Z}$  may vary with the time of posting articles. Let  $U$  be the proportion of articles posted during 18:00–02:00. We first divide the data into two groups according to  $U$ , namely work-hour ( $U < 0.5$ ) and off-hour ( $U \geq 0.5$ ). For each group, we run a regression analysis using the SICM

$$Y = g_1(\mathbf{Z}^T \boldsymbol{\beta}) + g_2(\mathbf{Z}^T \boldsymbol{\beta})X_2 + \varepsilon, \quad (6.3)$$

where  $g(\cdot)$  is an unknown smooth function and  $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7)^T$  is a loading parameter vector. All predictors are centered and standardized. The lines in Figure 3(b) connect the variant estimated coefficient vectors  $(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5, \hat{\beta}_6, \hat{\beta}_7)$ , with solid lines corresponding to the work-hour group and dashed lines to the off-hour group. It is clear that the proportion of articles posted during 18:00–02:00 does indeed change the interaction patterns, which inspires us to further consider the model  $Y = g_1(\mathbf{Z}^T \boldsymbol{\beta}(U)) + g_2(\mathbf{Z}^T \boldsymbol{\beta}(U))X_2 + \varepsilon$ .

First, we select significant variables  $Z_1, Z_2, Z_3, Z_4$ , and  $Z_5$  in the index using a variable-selection procedure, and choose the optimal parameters  $(q_1, q_2, K_1, K_2) = (3, 3, 2, 2)$  according to the BIC. Next, we conduct the hypothesis test from Section 4.2 with 500 wild bootstrap resamples, and the test  $p$ -value is 0.008. We reject  $H_0$  [the model (6.3)] at the 0.05 significance level.

Figure 5 shows the resulting spline estimators (solid curves) and the corresponding 95% SCBs (dashed curves) based on 500 wild bootstrap samples for  $\beta_k(u)$ , for  $k = 1, \dots, 5$ . Figures 5(a)–5(e) show that the indices of  $Z_1, Z_2, Z_3, Z_4$ , and  $Z_5$  change with  $U$ , especially the index of  $Z_2$ . The spline estimators of the index functions  $g_l(\cdot)$ , for  $l = 1, 2$ , and the corresponding 95% SCBs are depicted in Figures 5(f) and 5(g), respectively. Figure 5(f) shows that there is a truly nonlinear relationship between the median value of the reading volume and the predictors. Figure 5(g) shows that the modification due to the positive emotion score of titles in the last 10 articles is altered by the admixture of other variables, and that this influence changes from positive to negative with a decreasing trend. In addition, the  $QQ$  plot of the residuals in Figure 5(h) shows that the model (6.2) is a reasonable option for this data set.

## 7. Conclusion

The dynamic single-index varying coefficient model proposed in this paper can be regarded as an extension of many existing nonparametric and semiparametric models. We have developed a spline-based method to approximate the unknown functions and establish the asymptotic properties of the resulting estimators. We have also proposed a model identification procedure to select significant covariates, and have developed a hypothesis test to check whether the proposed model can be reduced to a simpler form.

There are at least two directions in which this work could be extended. First, motivated by Ma and Song (2015), we could extend the proposed model to the form  $Y = \sum_{l=1}^d g_l(\mathbf{Z}^T \boldsymbol{\beta}_l(U))X_l + \varepsilon$ , where the coefficient functions  $\boldsymbol{\beta}_l(\cdot)$  vary across  $X_l$ . Second, we could consider a semiparametric model  $Y = \sum_{l=1}^d g_l(\mathbf{Z}^T \boldsymbol{\beta}(\cdot))X_l + \varepsilon$ .

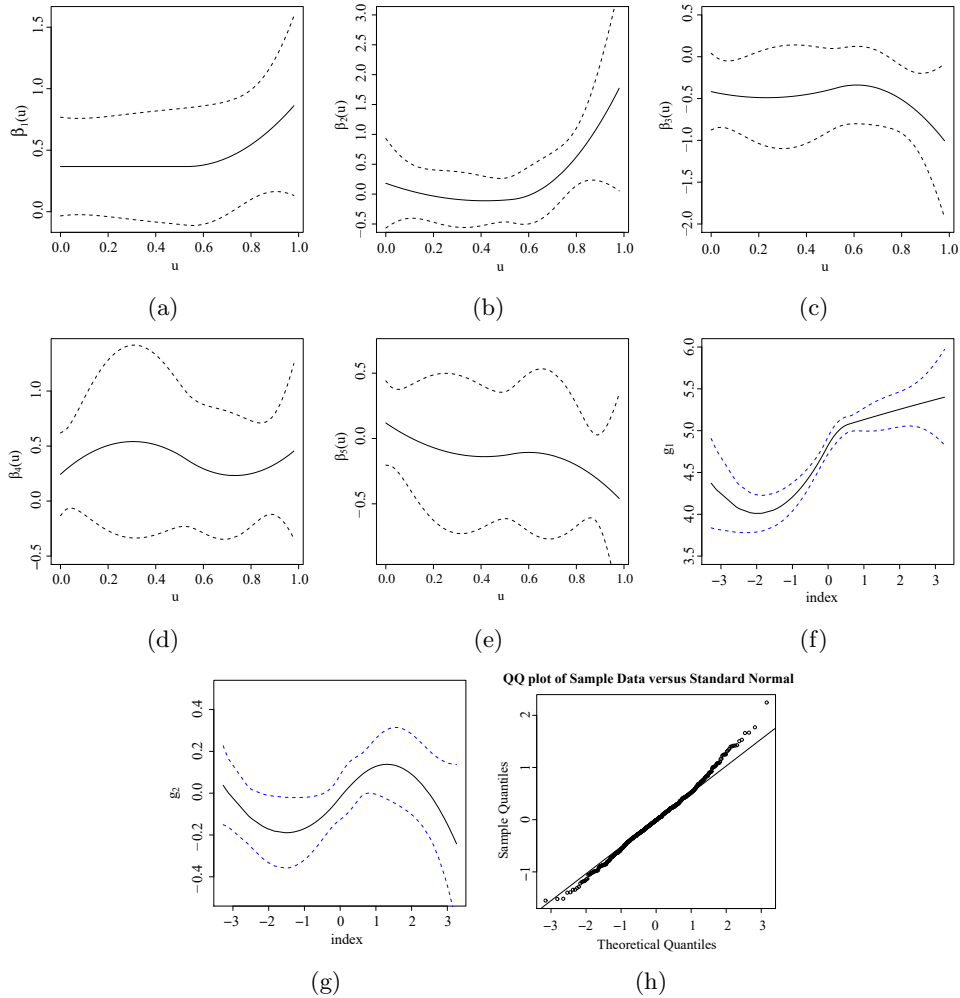


Figure 5. (a)–(e) Estimators (solid curves) of varying-coefficient functions and 95% SCBs (dashed curves) for  $\beta_k(\cdot)$ ,  $k = 1, \dots, 5$ . (f) and (g) Spline estimates (solid curves) and 95% SCBs (dashed curves) for index functions  $g_l(\cdot)$ ,  $l = 1, 2$ . (h) QQ plot of residuals.

$U) + \mathbf{G}^T \boldsymbol{\gamma}) X_l + \varepsilon$ , where the covariate  $\mathbf{G} \in \mathbb{R}^q$  and  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_q)^T$  is unknown.

## Supplementary Material

The online Supplementary Material includes the notation used in this paper, the implementation of the gradient function, the wild bootstrap procedure for hypothesis testing, technical proofs of the theorems, an additional numerical example, and two additional analyses of real data.

## Acknowledgments

The authors would like to thank the editor, associate editor, and two referees for their helpful comments and suggestions. Your research was supported in part by the National Natural Science Foundation of China (11971291). Zhou's work was supported by the State Key Program of the National Natural Science Foundation of China (71931004 and 92046005).

## References

- Chen, D., Hall, P. and Müller, H.-G. (2011). Single and multiple index functional regression models with nonparametric link. *Ann. Statist.* **39**, 1720–1747.
- de Boor, C. (2001). *A Practical Guide to Splines*. Springer, New York.
- Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96**, 1348–1360.
- Fan, J., Yao, Q. and Cai, Z. (2003). Adaptive varying-coefficient linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65**, 57–80.
- Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **55**, 757–796.
- Hu, L., Huang, T. and You, J. (2019). Estimation and identification of a varying-coefficient additive model for locally stationary processes. *J. Amer. Statist. Assoc.* **114**, 1191–1204.
- Kuruwita, C. N., Kulasekera, K. B. and Gallagher, C. M. (2011). Generalized varying coefficient models with unknown link function. *Biometrika* **98**, 701–710.
- Li, J. and Lv, J. (2022). High-dimensional varying index coefficient quantile regression model. *Statist. Sinica* **32**, 673–694.
- Liu, X., Cui, Y. and Li, R. (2016). Partial linear varying multi-index coefficient model for integrative gene-environment interactions. *Statist. Sinica* **26**, 1037–1060.
- Luo, X., Zhu, L. and Zhu, H. (2016). Single-index varying coefficient model for functional responses. *Biometrics* **72**, 1275–1284.
- Ma, S. (2016). Estimation and inference in functional single-index models. *Ann. Inst. Statist. Math.* **68**, 181–208.
- Ma, S., Carroll, R. J., Liang, H. and Xu, S. (2015). Estimation and inference in generalized additive coefficient models for nonlinear interactions with high-dimensional covariates. *Ann. Statist.* **43**, 2102–2131.
- Ma, S. and Song, P. X.-K. (2015). Varying index coefficient models. *J. Amer. Statist. Assoc.* **110**, 341–356.
- Tang, Q. and Cheng, L. (2008). M-estimation and B-spline approximation for varying coefficient models with longitudinal data. *J. Nonparam. Statist.* **20**, 611–625.
- Wang, L., Li, H. and Huang, J. (2008). Variable selection in nonparametric varying coefficient models for analysis of repeated measurements. *J. Amer. Statist. Assoc.* **103**, 1556–1569.
- Xia, Y. and Li, W. K. (1999). On single-index coefficient regression models. *J. Amer. Statist. Assoc.* **94**, 1275–1285.
- Xia, Y., Tong, H., Li, W. K. and Zhu, L. (2002). An adaptive estimation of dimension reduction space. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64**, 363–410.
- Yang, L. (2008). Confidence band for additive regression model. *J. Data Sci.* **6**, 207–217.

- Zhang, W., Li, D. and Xia, Y. (2015). Estimation in generalised varying-coefficient models with unspecified link functions. *J. Econometrics* **187**, 238–255.
- Zhang, H., Zimmerman, J., Nettleton, D. and Nordman, D. J. (2020). Random forest prediction intervals. *Amer. Statist.* **74**, 392–406.
- Zhao, W., Lian, H. and Liang, H. (2017). Quantile regression for the single-index coefficient model. *Bernoulli* **23**, 1997–2027.

Xin Guan

School of Statistics and Mathematics, Zhongnan University of Economics and Law, Wuhan, China.

E-mail: guanxin1112@163.com

Hua Liu

School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai, China.

E-mail: liuhua\_sufe@163.com

Jinhong You

School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai, China.

Shanghai Lixin University of Accounting and Finance, Shanghai, China.

E-mail: johnyou07@163.com

Yong Zhou

Key Laboratory of Advanced Theory and Application in Statistics and Data Science, MOE, and Academy of Statistics and Interdisciplinary Sciences, East China Normal University, Shanghai, China

E-mail: yzhou@amss.ac.cn

(Received December 2019; accepted April 2021)