Estimation and Inference for Dynamic Single-Index Varying-Coefficient Models

Xin Guan[1], Hua Liu[2], Jinhong You[2,3], and Yong Zhou[4]

[1]*Zhongnan University of Economics and Law*

[2] *Shanghai University of Finance and Economics*

[3]*Shanghai Lixin University of Accounting and Finance*

[4] *East China Normal University*

Supplementary Material

This part includes notations, the implementation of the gradient function, technical proofs in the manuscript, an additional numerical example, and two additional real data analyses.

S1 Notation

Throughout this article, $\otimes$ denotes the Kronecker product. For any matrix A, $A^{\otimes 2} = AA^T$. $\|f\|_{L_2} = [\int f^2(u)\,du]^{1/2}$ is the L_2 norm of any function such that $\int f^2(u)\,du < \infty$. For any vector $v = (v_1, \ldots, v_s)^T \in \mathbb{R}^s$, let $\|v\|_\infty = \max_{1\le l\le s} |v_l|$ and let $\|v\|_2$ be the Euclidean norm. For any symmetric matrix $A_{s\times s}$, denote its L_r norm by $\|A\|_r = \max_{v\in\mathbb{R}^s, v\neq 0} \|Av\|_r \|v\|_r^{-1}$.

For any matrix $A = (A_{ij})_{i=1,j=1}^{s,t}$, denote $\|A\|_\infty = \max_{1\le i\le s}\sum_{j=1}^t |A_{ij}|$. Given positive numbers a_n and b_n, $a_n \ll b_n$ means $\lim_{n\to\infty} a_n/b_n = 0$, $a_n \lesssim b_n$ means a_n/b_n is bounded, and $a_n \asymp b_n$ means $\lim_{n\to\infty} a_n/b_n = c$, where c is some nonzero constant. $\xrightarrow{d}$ indicates convergence in distribution. Denote by $C^{(r)}[0,1] = \{\phi : \phi^{(r)} \in C[0,1]\}$ the space of rth-order smooth functions. Let $C^{0,1}([a,b],c)$ be the space of Lipschitz-continuous functions for any fixed constant c; that is, $C^{0,1}([a,b],c) = \{\phi : |\phi(x_1)-\phi(x_2)| \le c|x_1-x_2|\ \forall x_1,x_2 \in [a,b]\}$.

S2 Algorithm Implementation

By de Boor (2001), the first-derivative function $\dot{g}_l(\cdot)$ can be approximated by the spline basis function one order lower than that of $g_l(\cdot)$. That is, $\hat{\dot{g}}_l(w,\hat{\boldsymbol{\delta}}) = \sum_{m=1}^{J_{n,1}} \dot{B}_{m,1}(w)\hat{\lambda}_{m,l}(\hat{\boldsymbol{\delta}}) = \mathbf{B}_1^{q_1-1}(w)^T\mathbb{E}\hat{\boldsymbol{\lambda}}(\hat{\boldsymbol{\delta}})$, where $\mathbf{B}_1^{q_1-1}(w) = \{B_{m,1}^{q_1-1}(w), 2\le m\le J_{n,1}\}^T$ is the (q_1-1)th-order B-spline basis function and

$$\mathbb{E} = (q_1-1)\begin{bmatrix} \frac{-1}{\zeta_{q_1+1}-\zeta_2} & \frac{1}{\zeta_{q_1+1}-\zeta_2} & 0 & \cdots & 0 \\ 0 & \frac{-1}{\zeta_{q_1+2}-\zeta_3} & \frac{1}{\zeta_{q_1+2}-\zeta_3} & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{-1}{\zeta_{K_1+2q_1-1}-\zeta_{K_1+q_1}} & \frac{1}{\zeta_{K_1+2q_1-1}-\zeta_{K_1+q_1}} \end{bmatrix}_{(J_{n,1}-1)\times J_{n,1}}.$$

Therefore, the gradient of the objective function (2.2) is written as

$$\frac{\partial L(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}} = -\sum_{i=1}^{n}\left[Y_i - \sum_{l=1}^{d}\hat{g}_l(\boldsymbol{\Phi}_i^T\boldsymbol{\delta})X_{il}\right]\left[\sum_{l=1}^{d}\hat{g}_l(\boldsymbol{\Phi}_i^T\boldsymbol{\delta})X_{il}\boldsymbol{\Phi}_i + \frac{\partial\hat{\boldsymbol{\lambda}}(\boldsymbol{\delta})^T}{\partial\boldsymbol{\delta}}\mathbf{D}_i(W_i^0)\right],$$

where $\mathbf{D}_i(W_i^0) = (D_{i,ml}(W_i^0), 1 \le m \le J_{n,1}, 1 \le l \le d)^T$ with $W_i^0 = \boldsymbol{\Phi}_i^T\boldsymbol{\delta}$.

In fact, the above gradient function can be approximated by

$$\frac{\partial L(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}} \approx -\sum_{i=1}^{n}\left[Y_i - \sum_{l=1}^{d}\sum_{m=1}^{J_{n,1}}B_{m,1}(\boldsymbol{\Phi}_i^T\boldsymbol{\delta})X_{il}\hat{\boldsymbol{\lambda}}_{m,l}\right]\left[\sum_{l=1}^{d}\hat{g}_l(\boldsymbol{\Phi}_i^T\boldsymbol{\delta})X_{il}\hat{\boldsymbol{\Phi}}_i\right].$$

This asymptotic equivalence is shown in lemma 4. $\hat{\boldsymbol{\Phi}}_i$ is defined as $\hat{\boldsymbol{\Phi}}_i = \boldsymbol{\Phi}_i - \mathbb{P}_n(\boldsymbol{\Phi}_i)$, with $\mathbb{P}_n(\boldsymbol{\Phi}_i) = \{\mathbb{P}_n(\Phi_{i,sk}), 1 \le k \le p, 1 \le s \le J_{n,2}\}^T$ and $\mathbb{P}_n(\Phi_{i,sk}) = \sum_{l=1}^{d}\hat{f}_{l,sk}^0(\boldsymbol{\Phi}_i^T\hat{\boldsymbol{\delta}})X_{il}$. $\hat{f}_{l,sk}^0(\cdot)$ can be obtained in the same way as $\hat{g}_l(\cdot)$, but with $\Phi_{i,sk}$ replacing Y_i.

S3 The Wild Bootstrap of the Hypothesis Testing

Step 1: Perform the proposed three-step approach to estimate $\boldsymbol{\beta}(u)$ under the model (1.2) and calculate the residuals $\hat{\varepsilon}_i = Y_i - \sum_{l=1}^{d}\hat{g}_l(\mathbf{Z}^T\hat{\boldsymbol{\beta}}(U_i)) X_{il}$. Denote the centered residuals by $\hat{\varepsilon}_i^{\text{cent}}$.

Step 2: Under H_0, obtain the estimator $\hat{\boldsymbol{\beta}}$ by applying the profile least squares method from Ma and Song (2015). Then compute the observed test statistic $\mathcal{T}$ by (4.2).

Step 3: Generate the wild bootstrap samples $\{Y_i^*, \mathbf{X}_i\}$, with $Y_i^* = \sum_{l=1}^{d} \hat{g}_l(\mathbf{Z}^T\hat{\boldsymbol{\beta}})X_{il} + \varepsilon_i^*$, where $\varepsilon_i^* = \hat{\varepsilon}_i^{\text{cent}}\varsigma_i$, and $\{\varsigma_i\}$ follows an i.i.d. standard normal distribution. Then compute the test statistic $\mathcal{T}^*$ based on the wild bootstrap samples $\{Y_i^*, \mathbf{X}_i\}$, that is, $\mathcal{T}^* = \frac{n}{J_{n,2}} \int_0^1 \|\hat{\boldsymbol{\beta}}^*(u) - \hat{\boldsymbol{\beta}}^*\|^2 \, du$, where $\hat{\boldsymbol{\beta}}^*(u)$ and $\hat{\boldsymbol{\beta}}^*$ are the estimates with the wild bootstrap samples.

Step 4: Repeat Step 3 B times to obtain $\{\mathcal{T}_b^*, 1 \le b \le B\}$. The p-value of the test is defined as $B^{-1} \sum_{b=1}^{B} I(\mathcal{T}_b^* > \mathcal{T})$. Reject the null hypothesis H_0 at level α if the p-value is smaller than a given significant level α.

S4 Proofs

S4.1 Proof of Proposition 1.

Proof. Without loss of generality, let $l = 1$. Assume the model is not identifiable. Then there exists $\{g(\cdot), \beta_k(\cdot)\} \neq \{h(\cdot), \alpha_k(\cdot)\}$ such that

$$g\left(\sum_{k=1}^{p} Z_{ik}\beta_k(u)\right) = h\left(\sum_{k=1}^{p} Z_{ik}\alpha_k(u)\right).$$

Under the condition (C2), we have $\beta_k(u) = \sum_{s=1}^{J_{n,2}} B_{s,2}(u)\delta_{sk}$ and $\alpha_k(u) = \sum_{s=1}^{J_{n,2}} B_{s,2}(u)\gamma_{sk}$, with spline coefficients $\boldsymbol{\delta}_k = \{\delta_{sk}\}$ and $\boldsymbol{\alpha}_k = \{\gamma_{sk}\}$. Thus, $g\big(\sum_{k=1}^{p} \boldsymbol{\Phi}_{ik}^T \boldsymbol{\delta}_k\big) = h\big(\sum_{k=1}^{p} \boldsymbol{\Phi}_{ik}^T \boldsymbol{\alpha}_k\big)$. Suppose that $\delta_{11} \neq 0$ and $\gamma_{11} \neq 0$. Then

$$\frac{\partial g/\partial \Phi_{i,sk}}{\partial g/\partial \Phi_{i,11}} = \frac{\delta_{sk}}{\delta_{11}} = \frac{\gamma_{sk}}{\gamma_{11}} = \frac{\partial h/\partial \Phi_{i,sk}}{\partial h/\partial \Phi_{i,11}}.$$

If $\|\boldsymbol{\beta}(u)\|_{L_2} = 1$, then it holds that $\sum_{k=1}^{p}\int_0^1 \boldsymbol{\delta}_k^T \mathbf{B}_2(u)\mathbf{B}_2(u)^T \boldsymbol{\delta}_k \, du = 1$. By the properties of B-spline basis functions, $\|\boldsymbol{\beta}(u)\|_{L2}^2 \asymp J_{n,2}^{-1}\|\boldsymbol{\delta}\|^2 = J_{n,2}^{-1}\sum_{k=1}^{p}\|\boldsymbol{\delta}_k\|^2 = J_{n,2}^{-1}\sum_{k=1}^{p}\sum_{s=1}^{J_{n,2}}\delta_{sk}^2$. Let $\delta'_{sk} = J_{n,2}^{1/2}\delta_{sk}$ and $\gamma'_{sk} = J_{n,2}^{1/2}\gamma_{sk}$. Then $\sum_{k=1}^{p}\sum_{s=1}^{J_{n,2}}(\delta'_{sk})^2 = O(1)$, and therefore $\delta'_{11} = \pm\gamma'_{11}$; furthermore, we have $\delta_{sk} = \pm\gamma_{sk}$ for any s, k. This yields $\beta_k(u) = \alpha_k(u)$ or $\beta_k(u) = -\alpha_k(u)$ for all $u \in [0,1]$ and each $k = 1, \ldots, p$. Assume that $\beta_1(u)$ is monotone nondecreasing. Then $\beta_k(u) = \alpha_k(u)$, and therefore $g(\cdot) = h(\cdot)$, which contradicts the assumption $\{g(\cdot), \beta_k(\cdot)\} \neq \{h(\cdot), \alpha_k(\cdot)\}$. Hence, the model (1.2) is indeed identifiable. $\square$

S4.2 Proof of Proposition 2.

Lemma 1 (Ma and Yang 2011). *Let $\xi_1^{(n)}, \ldots, \xi_n^{(n)}$ have a joint normal distribution with $E\xi_i^{(n)} \equiv 0$, $E(\xi_i^{(n)})^2 \equiv 1$, $1 \le i \le n$, and let there exist constants $C > 0$, $a > 1$, $r \in (0,1)$ such that the correlations $r_{ij} = r_{ij}^n = E\xi_i^{(n)}\xi_j^{(n)}$, $1 \le i \neq j \le n$ satisfy $|r_{ij}| \le \min(r, Ca^{-|i-j|})$ for $1 \le i \neq j \le n$. Then the absolute maximum $M_{n,\xi} = \max\{|\xi_1^{(n)}|, \ldots, |\xi_n^{(n)}|\}$ satisfies, for any $\tau \in \mathbb{R}$, $P(M_{n,\xi} \le \tau/a_n + b_n) \to \exp(-2e^{-\tau})$, as $n \to \infty$, where $a_n = (2\log n)^{1/2}$ and $b_n = a_n - (1/2)a_n^{-1}(\log\log n + \log 4\pi)$.*

Lemma 2. *Let $\mathbf{V}(\boldsymbol{\delta}) = E[\mathbf{D}_i(W_i^0)\mathbf{D}_i(W_i^0)^T]$ and $\hat{\mathbf{V}}(\boldsymbol{\delta}) = n^{-1}\mathbf{D}(\boldsymbol{W}^0)^T\mathbf{D}(\boldsymbol{W}^0)$. Then, under Conditions (C1) and (C4), for any vector $\boldsymbol{\alpha} = \{(\boldsymbol{\alpha}_1^T, \ldots, \boldsymbol{\alpha}_d^T)^T\}_{dJ_{n,1}\times 1}$*

with $\boldsymbol{\alpha}_l = (\alpha_{m,l} : 1 \leq m \leq J_{n,1})^T$ *and* $\|\boldsymbol{\alpha}_l\| = 1$, $1 \leq l \leq d$, *there are constants* $0 < c_V < C_V < \infty$ *such that, for large enough* n,

$$c_V J_{n,1}^{-1} \leq \boldsymbol{\alpha}^T \mathbf{V}(\boldsymbol{\delta})\boldsymbol{\alpha} \leq C_V J_{n,1}^{-1}, \qquad C_V^{-1} J_{n,1} \leq \boldsymbol{\alpha}^T \mathbf{V}(\boldsymbol{\delta})^{-1}\boldsymbol{\alpha} \leq c_V^{-1} J_{n,1}, \tag{S4.1}$$

and, with probability approaching one,

$$c_V J_{n,1}^{-1} \leq \boldsymbol{\alpha}^T \hat{\mathbf{V}}(\boldsymbol{\delta})\boldsymbol{\alpha} \leq C_V J_{n,1}^{-1}, \qquad C_V^{-1} J_{n,1} \leq \boldsymbol{\alpha}^T \hat{\mathbf{V}}(\boldsymbol{\delta})^{-1}\boldsymbol{\alpha} \leq c_V^{-1} J_{n,1}. \tag{S4.2}$$

Proof. According to Theorem 5.4.2 of DeVore and Lorentz (1993) and Condition (C1), there are constants $0 < c_l \leq C_l < \infty$ such that for large enough n,

$$c_l J_{n,1}^{-1} \leq \boldsymbol{\alpha}_l^T E[\mathbf{B}_1(\boldsymbol{\Phi}_i^T\boldsymbol{\delta})\mathbf{B}_1(\boldsymbol{\Phi}_i^T\boldsymbol{\delta})^T]\boldsymbol{\alpha}_l \leq C_l J_{n,1}^{-1}.$$

By Conditions (C1) and (C4), it follows that

$$\begin{aligned}
\boldsymbol{\alpha}^T E[\mathbf{D}_i(W_i^0)\mathbf{D}_i(W_i^0)^T]\boldsymbol{\alpha} &= \sum_{l,l'} \sum_{m,m'} E[\alpha_{m,l}\alpha_{m',l'} B_{m,1}(\boldsymbol{\Phi}_i^T\boldsymbol{\delta}) B_{m',1}(\boldsymbol{\Phi}_i^T\boldsymbol{\delta}) X_{il} X_{il'}] \\
&\geq c_Q \sum_l \boldsymbol{\alpha}_l^T E[\mathbf{B}_1(\boldsymbol{\Phi}_i^T\boldsymbol{\delta})\mathbf{B}_1(\boldsymbol{\Phi}_i^T\boldsymbol{\delta})^T]\boldsymbol{\alpha}_l \\
&\geq d c_Q \min(c_l) J_{n,1}^{-1}.
\end{aligned}$$

Similarly, we have $\boldsymbol{\alpha}^T E[\mathbf{D}_i(W_i^0)\mathbf{D}_i(W_i^0)^T]\boldsymbol{\alpha} \leq dC_Q \max(c_l) J_{n,1}^{-1}$. The second result in (S4.1) follows by replacing $\boldsymbol{\alpha}$ as $\mathbf{V}(\boldsymbol{\delta})^{-1/2}\boldsymbol{\alpha}$. Results in (S4.2) can be derived from (S4.1) and the Bernstein's inequality in Bosq

(1998). $\square$

Furthermore, by Lemma 2 and Demko (1986), it can be proved that $\|\hat{\mathbf{V}}(\boldsymbol{\delta})^{-1}\|_\infty = O_p(J_{n,1})$.

Proof of Proposition 2.

Proof. (i) Denote $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^T$, $\mathbf{g} = (g(W_1), \ldots, g(W_n))^T$. By (2.1), $\hat{\boldsymbol{\lambda}}(\boldsymbol{\delta})$ can be decomposed into $\hat{\boldsymbol{\lambda}}(\boldsymbol{\delta}) = \hat{\boldsymbol{\lambda}}_{\mathbf{g}}(\boldsymbol{\delta}) + \hat{\boldsymbol{\lambda}}_{\boldsymbol{\varepsilon}}(\boldsymbol{\delta})$, where $\hat{\boldsymbol{\lambda}}_{\mathbf{g}}(\boldsymbol{\delta}) = [\mathbf{D}(\boldsymbol{W}^0)^T\mathbf{D}(\boldsymbol{W}^0)]^{-1}\mathbf{D}(\boldsymbol{W}^0)^T\mathbf{g}$ and $\hat{\boldsymbol{\lambda}}_{\boldsymbol{\varepsilon}}(\boldsymbol{\delta}) = [\mathbf{D}(\boldsymbol{W}^0)^T\mathbf{D}(\boldsymbol{W}^0)]^{-1}\mathbf{D}(\boldsymbol{W}^0)^T\boldsymbol{\varepsilon}$. Let $\hat{\boldsymbol{\lambda}}_{\boldsymbol{\varepsilon}}(\boldsymbol{\delta}) = \{\hat{\boldsymbol{\lambda}}_{1,\boldsymbol{\varepsilon}}(\boldsymbol{\delta})^T, \ldots, \hat{\boldsymbol{\lambda}}_{d,\boldsymbol{\varepsilon}}(\boldsymbol{\delta})^T\}^T$, with $\hat{\boldsymbol{\lambda}}_{l,\boldsymbol{\varepsilon}}(\boldsymbol{\delta}) = \{\hat{\boldsymbol{\lambda}}_{m,l,\boldsymbol{\varepsilon}}(\boldsymbol{\delta}) : 1 \le m \le J_{n,1}\}$ and $\hat{\boldsymbol{\lambda}}_{\mathbf{g}}(\boldsymbol{\delta}) = \{\hat{\boldsymbol{\lambda}}_{1,\mathbf{g}}(\boldsymbol{\delta})^T, \ldots, \hat{\boldsymbol{\lambda}}_{d,\mathbf{g}}(\boldsymbol{\delta})^T\}^T$ with $\hat{\boldsymbol{\lambda}}_{l,\mathbf{g}}(\boldsymbol{\delta}) = \{\hat{\boldsymbol{\lambda}}_{m,l,\mathbf{g}}(\boldsymbol{\delta}) : 1 \le m \le J_{n,1}\}$ for $1 \le l \le d$. Thus, $\hat{g}_l(w;\boldsymbol{\delta}) = \hat{g}_{l,\mathbf{g}}(w;\boldsymbol{\delta}) + \hat{g}_{l,\boldsymbol{\varepsilon}}(w;\boldsymbol{\delta})$, where $\hat{g}_{l,\mathbf{g}}(w;\boldsymbol{\delta}) = \mathbf{B}_1(w)^T\hat{\boldsymbol{\lambda}}_{l,\mathbf{g}}(\boldsymbol{\delta})$ and $\hat{g}_{l,\boldsymbol{\varepsilon}}(w;\boldsymbol{\delta}) = \mathbf{B}_1(w)^T\hat{\boldsymbol{\lambda}}_{l,\boldsymbol{\varepsilon}}(\boldsymbol{\delta})$.

Rewrite $\hat{g}_{l,\mathbf{g}}(w;\boldsymbol{\delta}) = \mathbf{e}_l^T\mathbb{B}_1(w)\hat{\boldsymbol{\lambda}}_{\mathbf{g}}(\boldsymbol{\delta})$ and $\hat{g}_{l,\boldsymbol{\varepsilon}}(w;\boldsymbol{\delta}) = \mathbf{e}_l^T\mathbb{B}_1(w)\hat{\boldsymbol{\lambda}}_{\boldsymbol{\varepsilon}}(\boldsymbol{\delta})$. Notice that

$$\begin{aligned}
\hat{g}_{l,\mathbf{g}}(w;\boldsymbol{\delta}) - g_l^0(w) &= \mathbf{e}_l^T\mathbb{B}_1(w)[\hat{\boldsymbol{\lambda}}_{\mathbf{g}}(\boldsymbol{\delta}) - \boldsymbol{\lambda}^0] \\
&= \mathbf{e}_l^T\mathbb{B}_1(w)[\mathbf{D}(\boldsymbol{W}^0)^T\mathbf{D}(\boldsymbol{W}^0)]^{-1}\mathbf{D}(\boldsymbol{W}^0)^T[\mathbf{g} - \mathbf{D}(\boldsymbol{W}^0)\boldsymbol{\lambda}^0] \\
&= \psi_1(w) + \psi_2(w),
\end{aligned}$$

with

$$\psi_1(w) = \mathbf{e}_l^T \mathbb{B}_1(w)[\mathbf{D}(\boldsymbol{W}^0)^T \mathbf{D}(\boldsymbol{W}^0)]^{-1} \mathbf{D}(\boldsymbol{W}^0)^T \left[\sum_{l=1}^{d} [g_l(W_i) - g_l(\boldsymbol{\Phi}_i^T \boldsymbol{\delta})] X_{il} \right]_{i=1}^{n},$$

$$\psi_2(w) = \mathbf{e}_l^T \mathbb{B}_1(w)[\mathbf{D}(\boldsymbol{W}^0)^T \mathbf{D}(\boldsymbol{W}^0)]^{-1} \mathbf{D}(\boldsymbol{W}^0)^T \left\{ \left[\sum_{l=1}^{d} g_l(\boldsymbol{\Phi}_i^T \boldsymbol{\delta}) X_{il} \right]_{i=1}^{n} - \mathbf{D}(\boldsymbol{W}^0) \boldsymbol{\lambda}^0 \right\}.$$

Since $E[D_{i,ml}(\boldsymbol{\delta})] = O(J_{n,1}^{-1})$ and $\sup_{m,l} E[D_{i,ml}(\boldsymbol{\delta})]^2 = O(J_{n,1}^{-1})$, it is easy to show that $\|n^{-1}\mathbf{D}(\boldsymbol{W}^0)^T 1_n\|_\infty = O_p(J_{n,1}^{-1})$ by Bernstein's inequality. Therefore, we have

$$\begin{aligned}
\sup_{w \in S_W} |\psi_1(w)| &\le \left| \mathbf{e}_l^T \mathbb{B}_1(w)[n^{-1}\mathbf{D}(\boldsymbol{W}^0)^T \mathbf{D}(\boldsymbol{W}^0)]^{-1} n^{-1} \mathbf{D}(\boldsymbol{W}^0)^T \mathbf{1}_n \right| \left\| \left[\sum_{l=1}^{d} [g_l(W_i) - g_l(\boldsymbol{\Phi}_i^T \boldsymbol{\delta})] X_{il} \right]_{i=1}^{n} \right\|_\infty \\
&\le \sup_{w \in S_W} \left| \sum_m B_{m,1}(w) \right| \left\| \hat{\mathbf{V}}(\boldsymbol{\delta})^{-1} \right\|_\infty \left\| n^{-1} \mathbf{D}(\boldsymbol{W}^0)^T \mathbf{1}_n \right\|_\infty O(J_{n,2}^{-r} + a_n) \\
&= O_p(J_{n,1}) O_p(J_{n,1}^{-1}) O(J_{n,2}^{-r} + a_n) \\
&= O(J_{n,2}^{-r} + a_n)
\end{aligned}$$

and

$$\begin{aligned}
\sup_{w \in S_W} |\psi_2(w)| &\le \left| \mathbf{e}_l^T \mathbb{B}_1(w)[n^{-1}\mathbf{D}(\boldsymbol{W}^0)^T \mathbf{D}(\boldsymbol{W}^0)]^{-1} n^{-1} \mathbf{D}(\boldsymbol{W}^0)^T \mathbf{1}_n \right| \left\| \left[\sum_{l=1}^{d} g_l(\boldsymbol{\Phi}_i^T \boldsymbol{\delta}) X_{il} \right]_{i=1}^{n} - \mathbf{D}(\boldsymbol{W}^0) \boldsymbol{\lambda}^0 \right\|_\infty \\
&\le \sup_{w \in S_W} \left| \sum_m B_{m,1}(w) \right| \left\| \hat{\mathbf{V}}(\boldsymbol{\delta})^{-1} \right\|_\infty \left\| n^{-1} \mathbf{D}(\boldsymbol{W}^0)^T 1_n \right\|_\infty O(J_{n,1}^{-r}) \\
&= O_p(J_{n,1}) O_p(J_{n,1}^{-1}) O(J_{n,1}^{-r}) \\
&= O(J_{n,1}^{-r}).
\end{aligned}$$

We then obtain that $\sup_{w \in S_W} |\hat{g}_{l,\mathbf{g}}(w; \boldsymbol{\delta}) - g_l^0(w)| = O_p(a_n + J_{n,1}^{-r} + J_{n,2}^{-r})$. It is obvious that $E\{\hat{g}_{l,\boldsymbol{\varepsilon}}(w;\boldsymbol{\delta}) \mid \mathbf{X},\mathbf{Z}\} = 0$. Moreover, by Condition (C3) and (S4.1), it holds that

$$\begin{aligned}
E\{\hat{g}_{l,\boldsymbol{\varepsilon}}(w;\boldsymbol{\delta}) \mid \mathbf{X},\mathbf{Z}\}^2 &= n^{-2}\mathbf{e}_l^T\mathbb{B}_1(w)\hat{\mathbf{V}}(\boldsymbol{\delta})^{-1}\mathbf{D}(\boldsymbol{W}^0)^T E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T \mid \mathbf{X},\mathbf{Z})\mathbf{D}(\boldsymbol{W}^0)\hat{\mathbf{V}}(\boldsymbol{\delta})^{-1}\mathbb{B}_1(w)^T\mathbf{e}_l \\
&\le n^{-1}C_\sigma\mathbf{e}_l^T\mathbb{B}_1(w)\hat{\mathbf{V}}(\boldsymbol{\delta})^{-1}\mathbb{B}_1(w)^T\mathbf{e}_l \\
&\le n^{-1}C_\sigma\left\|\mathbb{B}_1(w)^T\mathbf{e}_l\right\|_2^2\left\|\hat{\mathbf{V}}(\boldsymbol{\delta})^{-1}\right\|_2 \\
&= O(J_{n,1}/n).
\end{aligned}$$

Again it can be proved that $\sup_{w\in S_W}|\hat{g}_{l,\boldsymbol{\varepsilon}}(w;\boldsymbol{\delta})| = O_p((\log n)^{1/2}J_{n,1}^{1/2}n^{-1/2})$ by Bernstein's inequality. Finally, we have

$$\sup_{w\in S_W}|\hat{g}_l(w;\boldsymbol{\delta}) - g_l(w)| = O_p(a_n + J_{n,1}^{-r} + J_{n,2}^{-r} + (\log n)^{1/2}J_{n,1}^{1/2}n^{-1/2}).$$

(ii) Rewrite $\hat{\dot{g}}_l(w,\boldsymbol{\delta}) = \hat{\dot{g}}_{l,\mathbf{g}}(w,\boldsymbol{\delta}) + \hat{\dot{g}}_{l,\boldsymbol{\varepsilon}}(w,\boldsymbol{\delta})$ with $\hat{\dot{g}}_{l,\mathbf{g}}(w,\boldsymbol{\delta}) = \mathbf{B}_1^{q_1-1}(w)^T\mathbb{E}\hat{\boldsymbol{\lambda}}_{l,\mathbf{g}}(\boldsymbol{\delta})$ and $\hat{\dot{g}}_{l,\boldsymbol{\varepsilon}}(w,\boldsymbol{\delta}) = \mathbf{B}_1^{q_1-1}(w)^T\mathbb{E}\hat{\boldsymbol{\lambda}}_{l,\boldsymbol{\varepsilon}}(\boldsymbol{\delta})$. It is easy to show that $\|\mathbb{E}\|_\infty = O(J_{n,1})$. Applying the same approach as in the proof of (i), we get

$$\sup_{w\in S_W}|\hat{\dot{g}}_l(w;\boldsymbol{\delta}) - \dot{g}_l(w)| = O_p(a_nJ_{n,1} + J_{n,1}^{1-r} + J_{n,1}J_{n,2}^{-r} + (\log n)^{1/2}J_{n,1}^{3/2}n^{-1/2}).$$

□

S4.3 Proof of Theorem 1.

Lemma 3 (Ma 2016)**.** *Under Condition (C2), there exists* $\tilde{\boldsymbol{\delta}}_1^0 = (\tilde{\delta}_{s1}^0, 1 \le s \le J_{n,2})^T \in \mathbb{R}^{J_{n,2}}$ *with* $\tilde{\delta}_{11}^0 \le \cdots \le \tilde{\delta}_{J_{n,2}1}^0$ *such that* $\sup_{u\in S_U}|\beta_1(u) - \tilde{\beta}_1(u)| = O_p(J_{n,2}^{-r})$*, where* $\tilde{\beta}_1(u) = \mathbf{B}_2(u)^T\tilde{\boldsymbol{\delta}}_1^0$.

Lemma 4. *Let* $\tilde{\boldsymbol{\delta}}^0 = (\tilde{\boldsymbol{\delta}}_1^0, \boldsymbol{\delta}_2^0, \ldots, \boldsymbol{\delta}_p^0)^T$. *Under the assumptions of Theorem 1, we have*

$$\left\| \frac{\partial L(\tilde{\boldsymbol{\delta}}^0)}{\partial \boldsymbol{\delta}} - \sum_{i=1}^{n} \left[Y_i - \sum_{l=1}^{d} g_l(W_i) X_{il} \right] \sum_{l=1}^{d} \dot{g}_l(W_i) X_{il} \tilde{\boldsymbol{\Phi}}_i \right\|_\infty = o_p\left(\sqrt{n/J_{n,2}} \right).$$

Proof. For $\hat{\boldsymbol{\lambda}}(\boldsymbol{\delta})$ defined in (2.1), we have

$$\begin{aligned}
\mathbf{D}_i(W_i^0)^T \frac{\partial \hat{\boldsymbol{\lambda}}(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}^T} &= \mathbf{D}_i(W_i^0)^T \frac{\partial [\hat{\boldsymbol{\lambda}}(\boldsymbol{\delta}) - \boldsymbol{\lambda}^0]}{\partial \boldsymbol{\delta}^T} \\
&= \mathbf{D}_i(W_i^0)^T \frac{\partial \{ [\mathbf{D}(\boldsymbol{W}^0)^T \mathbf{D}(\boldsymbol{W}^0)]^{-1} \mathbf{D}(\boldsymbol{W}^0)^T [\mathbf{Y} - \mathbf{D}(\boldsymbol{W}^0)\boldsymbol{\lambda}^0] \}}{\partial \boldsymbol{\delta}^T} \\
&= \Theta_1(\boldsymbol{\delta}) + \Theta_2(\boldsymbol{\delta}), \qquad \text{(S4.3)}
\end{aligned}$$

where

$$\Theta_1(\boldsymbol{\delta}) = -\mathbf{D}_i(W_i^0)^T [\mathbf{D}(\boldsymbol{W}^0)^T \mathbf{D}(\boldsymbol{W}^0)]^{-1} \mathbf{D}(\boldsymbol{W}^0)^T \left[\sum_{l=1}^{d} \dot{g}_l^0 (\boldsymbol{\Phi}_i^T \boldsymbol{\delta}) X_{il} \boldsymbol{\Phi}_i^T \right]_{i=1}^{n},$$

$$\Theta_2(\boldsymbol{\delta}) = \mathbf{D}_i(W_i^0)^T \frac{\partial \{ [\mathbf{D}(\boldsymbol{W}^0)^T \mathbf{D}(\boldsymbol{W}^0)]^{-1} \mathbf{D}(\boldsymbol{W}^0)^T \}}{\partial \boldsymbol{\delta}^T} [\mathbf{Y} - \mathbf{D}(\boldsymbol{W}^0)\boldsymbol{\lambda}^0].$$

Define $\widetilde{W}_i = \boldsymbol{Z}_i^T \widetilde{\boldsymbol{\beta}}(U_i)$, $\mathbf{D}_i(\widetilde{W}_i) = \{ D_{i,ml}(\widetilde{W}_i), 1 \le l \le d, 1 \le m \le J_{n,1} \}^T$ with $D_{i,ml}(\widetilde{W}_i) = B_{m,1}(\widetilde{W}_i) X_{i,l}$,

$$\hat{\boldsymbol{\kappa}} = \underset{\boldsymbol{\kappa} \in \mathbb{R}^{dJ_{n,1} \times pJ_{n,2}}}{\arg\min} \sum_{i=1}^{n} [\boldsymbol{\Phi}_i - \mathbf{D}_i(\widetilde{W}_i)^T \boldsymbol{\kappa}]^T [\boldsymbol{\Phi}_i - \mathbf{D}_i(\widetilde{W}_i)^T \boldsymbol{\kappa}],$$

$$\begin{aligned}
\mathbb{P}_n\{\boldsymbol{\Phi}_i\} &= \mathbf{D}_i(\widetilde{W}_i)^T \hat{\boldsymbol{\kappa}} \\
&= \mathbf{D}_i(\widetilde{W}_i)^T [\mathbf{D}(\widetilde{\boldsymbol{W}})^T \mathbf{D}(\widetilde{\boldsymbol{W}})]^{-1} \mathbf{D}(\widetilde{\boldsymbol{W}})^T (\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_n)^T.
\end{aligned}$$

From the definition of $\mathbb{P}(\cdot)$ given in Section 2 and Condition (C5), it can be proved that $\|\mathbb{P}_n\{\boldsymbol{\Phi}_i\} - \mathbb{P}\{\boldsymbol{\Phi}_i\}\|_\infty = O_p(J_{n,1}^{-2} + J_{n,1}^{1/2} n^{-1/2})$. Thus, we can rewrite $\Theta_1(\tilde{\boldsymbol{\delta}}^0)$ as

$$\begin{aligned}\Theta_1(\tilde{\boldsymbol{\delta}}^0) &= -\mathbb{P}_n\left\{\sum_{l=1}^{d}\dot{g}_l^0(\boldsymbol{\Phi}_i^T\tilde{\boldsymbol{\delta}}^0)X_{il}\boldsymbol{\Phi}_i^T\right\}\\ &= -\mathbb{P}\left\{\sum_{l=1}^{d}\dot{g}_l^0(\boldsymbol{\Phi}_i^T\tilde{\boldsymbol{\delta}}^0)X_{il}\boldsymbol{\Phi}_i^T\right\} + O_p(J_{n,1}^{-2} + J_{n,1}^{1/2}n^{-1/2})\\ &= -\sum_{l=1}^{d}\dot{g}_l^0(\boldsymbol{\Phi}_i^T\tilde{\boldsymbol{\delta}}^0)X_{il}\mathbb{P}\{\boldsymbol{\Phi}_i^T\} + O_p(J_{n,1}^{-2} + J_{n,1}^{1/2}n^{-1/2}). \qquad \text{(S4.4)}\end{aligned}$$

Decompose $\Theta_2(\tilde{\boldsymbol{\delta}}^0)$ as

$$\begin{aligned}\Theta_2(\tilde{\boldsymbol{\delta}}^0) &= \mathbf{D}_i(\widetilde{W}_i)^T\left[\frac{\partial\{[\mathbf{D}(\widetilde{\boldsymbol{W}})^T\mathbf{D}(\widetilde{\boldsymbol{W}})]^{-1}\mathbf{D}(\widetilde{\boldsymbol{W}})^T\}}{\partial\boldsymbol{\delta}^T}\right]\left\{\mathbf{Y} - \left[\sum_{l=1}^{d}g_l(W_i)X_{il}\right]_{i=1}^{n}\right\}\\ &\quad + \mathbf{D}_i(\widetilde{W}_i)^T\left[\frac{\partial\{[\mathbf{D}(\widetilde{\boldsymbol{W}})^T\mathbf{D}(\widetilde{\boldsymbol{W}})]^{-1}\mathbf{D}(\widetilde{\boldsymbol{W}})^T\}}{\partial\boldsymbol{\delta}^T}\right]\left\{\left[\sum_{l=1}^{d}g_l(W_i)X_{il}\right]_{i=1}^{n} - \mathbf{D}(\widetilde{\boldsymbol{W}})\boldsymbol{\lambda}^0\right\}\\ &= \Theta_{21}(\tilde{\boldsymbol{\delta}}^0) + \Theta_{22}(\tilde{\boldsymbol{\delta}}^0). \qquad \text{(S4.5)}\end{aligned}$$

Let $\mathbf{A}(\tilde{\boldsymbol{\delta}}^0) = [\hat{\mathbf{V}}(\tilde{\boldsymbol{\delta}}^0)^{-1}\mathbf{D}(\widetilde{\boldsymbol{W}})^T]_{J_{n,1}d\times n}$, and let $\partial\mathbf{A}(\tilde{\boldsymbol{\delta}}^0)/\partial\boldsymbol{\delta}^T = (\dot{A}_{i,ml,sk})_{sk}$ be a $J_{n,1}d\times n\times J_{n,2}p$ array. Similar to Lemma A.3 of Ma and Song (2015), we have

$$\begin{aligned}\|\Theta_{21}(\tilde{\boldsymbol{\delta}}^0)\|_\infty &= \left\|n^{-1}\mathbf{D}_i(\widetilde{W}_i)^T\frac{\partial\mathbf{A}(\tilde{\boldsymbol{\delta}}^0)}{\partial\boldsymbol{\delta}^T}\left\{\mathbf{Y} - \left[\sum_{l=1}^{d}g_l(W_i)X_{il}\right]_{i=1}^{n}\right\}\right\|_\infty\\ &= \left\|n^{-1}\sum_{i'=1}^{n}\sum_{l=1}^{d}\sum_{m=1}^{J_{n,1}}D_{i,ml}(\widetilde{W}_i)(\dot{A}_{i',ml,sk})_{sk}\varepsilon_{i'}\right\|_\infty\\ &= O_p((\log n)^{1/2}n^{-1/2}).\end{aligned}$$

By B-spline properties, the basis functions $\mathbf{B}_1(\cdot)$, $\mathbf{B}_2(\cdot)$, and $\dot{\mathbf{B}}_1(\cdot)$ are bounded between zero and one. Notice that

$$
\begin{aligned}
&\big\|[\mathbf{D}(\widetilde{\boldsymbol{W}})^T\mathbf{D}(\widetilde{\boldsymbol{W}})]^{-1}\mathbf{D}(\widetilde{\boldsymbol{W}})^T\mathbf{1}_n\big\|_\infty \\
&\quad\leq \big\|[\mathbf{D}(\widetilde{\boldsymbol{W}})^T\mathbf{D}(\widetilde{\boldsymbol{W}})]^{-1}\big\|_\infty\big\|\mathbf{D}(\widetilde{\boldsymbol{W}})^T\mathbf{1}_n\big\|_\infty = O_P(J_{n,1})O_p(J_{n,1}^{-1}) = O_p(1),
\end{aligned}
$$

and so, together with Condition (C1), we can directly derive

$$
\begin{aligned}
\|\Theta_{22}(\tilde{\boldsymbol{\delta}}^0)\|_\infty &= \left\|\mathbf{D}_i(\widetilde{W}_i)^T\left[\frac{\partial\{[\mathbf{D}(\widetilde{\boldsymbol{W}})^T\mathbf{D}(\widetilde{\boldsymbol{W}})]^{-1}\mathbf{D}(\widetilde{\boldsymbol{W}})^T\}}{\partial\boldsymbol{\delta}^T}\right]\left\{\mathbf{D}(\widetilde{\boldsymbol{W}})\boldsymbol{\lambda}^0 - \left[\sum_{l=1}^d g_l(W_i)X_{il}\right]_{i=1}^n\right\}\right\|_\infty \\
&= \left|\sum_{m,l} D_{i,ml}(\widetilde{W}_i)\right|\left\|\frac{\partial\{[\mathbf{D}(\widetilde{\boldsymbol{W}})^T\boldsymbol{D}(\widetilde{\boldsymbol{W}})]^{-1}\mathbf{D}(\widetilde{\boldsymbol{W}})^T\mathbf{1}_n\}}{\partial\boldsymbol{\delta}}\right\|_\infty O_p(J_{n,1}^{-r}+J_{n,2}^{-r}) \\
&= O_p(J_{n,1}^{-r}+J_{n,2}^{-r}).
\end{aligned}
$$

Thus,

$$
\|\Theta_2(\tilde{\boldsymbol{\delta}}^0)\|_\infty = O_p((\log n)^{1/2}n^{-1/2} + J_{n,1}^{-r} + J_{n,2}^{-r}). \qquad \text{(S4.6)}
$$

Combining (S4.3)–(S4.6), we have

$$
\begin{aligned}
&\mathbf{D}_i(\widetilde{W}_i)^T\,\frac{\partial\hat{\boldsymbol{\lambda}}(\tilde{\boldsymbol{\delta}}^0)}{\partial\boldsymbol{\delta}^T} \\
&= -\sum_{l=1}^d \dot{g}_l^0(\boldsymbol{\Phi}_i^T\tilde{\boldsymbol{\delta}}^0)X_{il}\mathbb{P}\{\boldsymbol{\Phi}_i^T\} + O_p(J_{n,1}^{1/2}n^{-1/2} + J_{n,1}^{-2} + J_{n,2}^{-r} + (\log n)^{1/2}n^{-1/2}).
\end{aligned}
$$

Based on the results in Proposition 2, it is easy to see that

$$
\hat{\dot{g}}_l(\boldsymbol{\Phi}_i^T\tilde{\boldsymbol{\delta}}^0)X_{il} = \dot{g}_l(W_i)X_{il} + O_p(J_{n,1}^{3/2}n^{-1/2} + J_{n,1}^{1-r} + J_{n,1}J_{n,2}^{-r}).
$$

Therefore,

$$
\begin{aligned}
&\sum_{l=1}^{d} \hat{\dot{g}}_l(\mathbf{\Phi}^T\tilde{\boldsymbol{\delta}}^0)X_{il}\mathbf{\Phi}_i + \frac{\partial\hat{\boldsymbol{\lambda}}(\tilde{\boldsymbol{\delta}}^0)^T}{\partial\boldsymbol{\delta}}\mathbf{D}_i(\widetilde{W}_i)\\
&= \left[\sum_{l=1}^{d} \dot{g}_l(W_i)X_{il} + O_p(J_{n,1}^{3/2}n^{-1/2} + J_{n,1}^{1-r} + J_{n,1}J_{n,2}^{-r})\right]\mathbf{\Phi}_i\\
&\quad - \left[\sum_{l=1}^{d} \dot{g}_l(W_i)X_{il} + O_p(J_{n,1}^{1-r} + J_{n,1}J_{n,2}^{-r})\right]\mathbb{P}\{\mathbf{\Phi}_i\} + O_p(J_{n,1}^{1/2}n^{-1/2} + J_{n,1}^{-2} + J_{n,2}^{-r} + (\log n)^{1/2}n^{-1/2})\\
&= \sum_{l=1}^{d} \dot{g}_l(W_i)X_{il}\tilde{\mathbf{\Phi}}_i + \mathbf{\Phi}_i O_p(J_{n,1}^{3/2}n^{-1/2} + J_{n,1}^{1-r} + J_{n,1}J_{n,2}^{-r}) - \mathbb{P}\{\mathbf{\Phi}_i\}O_p(J_{n,1}^{1-r} + J_{n,1}J_{n,2}^{-r})\\
&\quad + O_p(J_{n,1}^{1/2}n^{-1/2} + J_{n,1}^{-2} + J_{n,2}^{-r} + (\log n)^{1/2}n^{-1/2}). \qquad \text{(S4.7)}
\end{aligned}
$$

By (2.2) and the above results, we obtain

$$
\begin{aligned}
\frac{\partial L(\tilde{\boldsymbol{\delta}}^0)}{\partial\boldsymbol{\delta}} &= -\sum_{i=1}^{n}\left[Y_i - \sum_{l=1}^{d}\sum_{m=1}^{J_{n,1}} B_{m,1}(\mathbf{\Phi}_i^T\tilde{\boldsymbol{\delta}}^0)X_{il}\hat{\boldsymbol{\lambda}}_{m,l}(\tilde{\boldsymbol{\delta}}^0)\right]\left[\sum_{l=1}^{d}\hat{\dot{g}}_l(\Phi_i^T\tilde{\boldsymbol{\delta}}^0)X_{il}\mathbf{\Phi}_i + \frac{\partial\hat{\boldsymbol{\lambda}}(\tilde{\boldsymbol{\delta}}^0)^T}{\partial\boldsymbol{\delta}}\mathbf{D}_i(\widetilde{W}_i)\right]\\
&= -\sum_{i=1}^{n}\left[Y_i - \sum_{l=1}^{d} g_l(W_i)X_{il}\right]\sum_{l=1}^{d}\dot{g}_l(W_i)X_{il}\tilde{\mathbf{\Phi}}_i - (\Lambda_1+\Lambda_2+\Lambda_3+\Lambda_4),
\end{aligned}
$$

where

$$
\begin{aligned}
\Lambda_1 &= \sum_{i=1}^{n}\left[\sum_{l=1}^{d} g_l(W_i)X_{il} - \mathbf{D}_i(\widetilde{W}_i)^T\hat{\boldsymbol{\lambda}}(\tilde{\boldsymbol{\delta}}^0)\right]\sum_{l=1}^{d}\dot{g}_l(W_i)X_{il}\tilde{\mathbf{\Phi}}_i,\\
\Lambda_2 &= \sum_{i=1}^{n}[Y_i - \mathbf{D}_i(\widetilde{W}_i)^T\hat{\boldsymbol{\lambda}}(\tilde{\boldsymbol{\delta}}^0)]\mathbf{\Phi}_i O_p(J_{n,1}^{3/2}n^{-1/2} + J_{n,1}^{1-r} + J_{n,1}J_{n,2}^{-r}),\\
\Lambda_3 &= -\sum_{i=1}^{n}[Y_i - \mathbf{D}_i(\widetilde{W}_i)^T\hat{\boldsymbol{\lambda}}(\tilde{\boldsymbol{\delta}}^0)]\mathbb{P}\{\mathbf{\Phi}_i\}O_p(J_{n,1}^{1-r} + J_{n,1}J_{n,2}^{-r}),\\
\Lambda_4 &= \sum_{i=1}^{n}[Y_i - \mathbf{D}_i(\widetilde{W}_i)^T\hat{\boldsymbol{\lambda}}(\tilde{\boldsymbol{\delta}}^0)]O_p(J_{n,1}^{1/2}n^{-1/2} + J_{n,1}^{-2} + J_{n,2}^{-r} + (\log n)^{1/2}n^{-1/2}).
\end{aligned}
$$

We will show that $\|\Lambda_i\|_\infty = o_p(n^{1/2}J_{n,2}^{-1/2})$ for each $i = 1,2,3,4$. Based on Bernstein's inequality and the results in Proposition 2, we can directly derive that

$$
\begin{aligned}
\|\Lambda_1\|_\infty &= \left\|\sum_{i=1}^{n}\left[\sum_{l=1}^{d}\dot{g}_l(W_i)X_{il}\tilde{\mathbf{\Phi}}_i\right]\right\|_\infty O_p(J_{n,1}^{-r}+J_{n,2}^{-r}+n^{-1/2}J_{n,1}^{1/2}) \\
&= O_p(n^{1/2}J_{n,2}^{-1/2}(\log n)^{1/2})O_p(J_{n,1}^{-r}+J_{n,2}^{-r}+n^{-1/2}J_{n,1}^{1/2}) \\
&= o_p(n^{1/2}J_{n,2}^{-1/2}).
\end{aligned}
$$

Note that $\sup_{s,k} E(\mathbf{\Phi}_{i,sk}\varepsilon_i) = 0$ and $\sup_{s,k} E(\mathbf{\Phi}_{i,sk}\varepsilon_i)^2 = O(J_{n,2}^{-1})$, and, by Bernstein's inequality, it holds that $\|\sum_{i=1}^{n}\mathbf{\Phi}_i\varepsilon_i\|_\infty = O_p((\log n)^{1/2}J_{n,2}^{-1/2}n^{1/2})$. Similarly, we have $\|\sum_{i=1}^{n}\mathbf{\Phi}_i\|_\infty = O_p(nJ_{n,2}^{-1})$. By the assumptions $n^{1/(2r+2)} \ll J_{n,1} \ll n^{1/4}$ and $n^{1/(2r+2)} \ll J_{n,2} \ll n^{1/4}$, it is obvious that

$$
\begin{aligned}
\|\Lambda_2\|_\infty = &\left\|\sum_{i=1}^{n}\mathbf{\Phi}_i\varepsilon_i\right\|_\infty O_p(J_{n,1}^{3/2}n^{-1/2}+J_{n,1}^{1-r}+J_{n,1}J_{n,2}^{-r}) \\
&+\left\|\sum_{i=1}^{n}\left[\sum_{l=1}^{d}g_l(W_i)X_{il}-\mathbf{D}_i(\widetilde{W}_i)^T\hat{\boldsymbol{\lambda}}(\tilde{\boldsymbol{\delta}}^0)\right]\mathbf{\Phi}_i\right\|_\infty O_p(J_{n,1}^{3/2}n^{-1/2}+J_{n,1}^{1-r}+J_{n,1}J_{n,2}^{-r}) \\
= &O_p((\log n)^{1/2}J_{n,2}^{-1/2}n^{1/2})O_p(J_{n,1}^{3/2}n^{-1/2}+J_{n,1}^{1-r}+J_{n,1}J_{n,2}^{-r}) \\
&+O_p(nJ_{n,2}^{-1})O_p(J_{n,1}^{-r}+J_{n,2}^{-r}+n^{-1/2}J_{n,1}^{1/2})O_p(J_{n,1}^{3/2}n^{-1/2}+J_{n,1}^{1-r}+J_{n,1}J_{n,2}^{-r}) \\
= &o_p(n^{1/2}J_{n,2}^{-1/2}),
\end{aligned}
$$

Similarly, we have $\|\Lambda_3\|_\infty = o_p(n^{1/2}J_{n,2}^{-1/2})$. Again according to the assumptions $n^{1/(2r+2)} \ll J_{n,1} \ll n^{1/4}$ and $n^{1/(2r+2)} \ll J_{n,2} \ll n^{1/4}$, we obtain

$$\begin{aligned}
\|\Lambda_4\|_\infty &= O_p(n^{1/2})O_p(J_{n,1}^{1/2}n^{-1/2} + J_{n,1}^{-2} + J_{n,2}^{-r} + (\log n)^{1/2}n^{-1/2}) \\
&\quad + nO_p(J_{n,1}^{-r} + J_{n,2}^{-r} + n^{-1/2}J_{n,1}^{1/2})O_p(J_{n,1}^{1/2}n^{-1/2} + J_{n,1}^{-2} + J_{n,2}^{-r} + (\log n)^{1/2}n^{-1/2}) \\
&= o_p(n^{1/2}J_{n,2}^{-1/2}).
\end{aligned}$$

□

Lemma 5. *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. Then, under the assumptions of Theorem 1, we have*

$$\|\hat{\boldsymbol{\delta}} - \tilde{\boldsymbol{\delta}}^0\|_2 = O_p(n^{-1/2}J_{n,2}). \tag{S4.8}$$

Proof. By Lemma 4, it is straightforward to prove that

$$\frac{\partial^2 L(\tilde{\boldsymbol{\delta}}^0)}{\partial \boldsymbol{\delta}\, \partial \boldsymbol{\delta}^T} = \sum_{i=1}^{n} \left[\sum_{l=1}^{d} \dot{g}_l(W_i)X_{il}\tilde{\boldsymbol{\Phi}}_i\right]^{\otimes 2} + o_p(nJ_{n,2}^{-1}).$$

Suppose that $\hat{\boldsymbol{\delta}}$ minimizes $L(\boldsymbol{\delta})$ given in (2.2) satisfying $\|\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}^0\| \leq a_n$ with probability tending to one. Then we assume that the minimizer is obtained in a neighborhood of $\boldsymbol{\delta}^0$ in probability. To prove (S4.8), it suffices to show that for any $\epsilon > 0$, there exists a constant $C > 0$ such that for large enough n,

$$P\left\{\sup_{\|\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}}^0\|_2 = Cn^{-1/2}J_{n,2}} (\boldsymbol{\delta} - \tilde{\boldsymbol{\delta}}^0)^T \frac{\partial L(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}} > 0\right\} \geq 1 - \epsilon. \tag{S4.9}$$

By Taylor expansion, there exists $\boldsymbol{\delta}^*$ lying between $\tilde{\boldsymbol{\delta}}^0$ and $\boldsymbol{\delta}$ such that

$$\begin{aligned}(\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}}^0)^T\,\frac{\partial L(\boldsymbol{\delta})}{\partial\boldsymbol{\delta}} &= (\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}}^0)^T\,\frac{\partial L(\tilde{\boldsymbol{\delta}}^0)}{\partial\boldsymbol{\delta}} + (\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}}^0)^T\,\frac{\partial^2 L(\boldsymbol{\delta}^*)}{\partial\boldsymbol{\delta}\,\partial\boldsymbol{\delta}^T}(\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}}^0)\\ &= (\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}}^0)^T\,\frac{\partial L(\tilde{\boldsymbol{\delta}}^0)}{\partial\boldsymbol{\delta}} + (\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}}^0)^T\,\frac{\partial^2 L(\tilde{\boldsymbol{\delta}}^0)}{\partial\boldsymbol{\delta}\,\partial\boldsymbol{\delta}^T}(\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}}^0)[1+o_p(1)].\end{aligned}$$

According to Lemma 4, we have $\partial L(\tilde{\boldsymbol{\delta}}^0)/\partial\boldsymbol{\delta} = -\sum_{i=1}^n\sum_{l=1}^d \dot{g}_l(W_i)X_{il}\tilde{\boldsymbol{\Phi}}_i\varepsilon_i[1+o_p(1)]$. Thus, for any $\boldsymbol{\delta}$ satisfying $\|\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}}^0\|_2 = Cn^{-1/2}J_{n,2}$, we employ the Cauchy–Schwartz inequality to derive

$$\left|(\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}}^0)^T\,\frac{\partial L(\tilde{\boldsymbol{\delta}}^0)}{\partial\boldsymbol{\delta}}\right| \le Cn^{-1/2}J_{n,2}\left\|\sum_{i=1}^n\sum_{l=1}^d \dot{g}_l(W_i)X_{il}\tilde{\boldsymbol{\Phi}}_i\varepsilon_i\right\|_2. \tag{S4.10}$$

Note that $E\{\sum_{i=1}^n\sum_{l=1}^d \dot{g}_l(W_i)X_{il}\tilde{\boldsymbol{\Phi}}_i\varepsilon_i\} = 0$. Similar to the proof of Lemma 2, together with Conditions (C1), (C2), and (C4), we have

$$\begin{aligned}E\left\{\left\|\sum_{i=1}^n\sum_{l=1}^d \dot{g}_l(W_i)X_{il}\tilde{\boldsymbol{\Phi}}_i\varepsilon_i\right\|_2^2\right\} &= E\left\{\sum_{i=1}^n\varepsilon_i^2\,\mathrm{tr}\left(\left[\sum_{l=1}^d \dot{g}_l(W_i)X_{il}\tilde{\boldsymbol{\Phi}}_i\right]^{\otimes 2}\right)\right\}\\ &\le C_\sigma\,\mathrm{tr}\left(\sum_{i=1}^n\left[\sum_{l=1}^d \dot{g}_l(W_i)X_{il}\tilde{\boldsymbol{\Phi}}_i\right]^{\otimes 2}\right) = O(n),\end{aligned}$$

where $\mathrm{tr}(\cdot)$ is the matrix trace. Then $\|\sum_{i=1}^n\sum_{l=1}^d \dot{g}_l(W_i)X_{il}\tilde{\boldsymbol{\Phi}}_i\varepsilon_i\|_2 = O_p(n^{1/2})$ by the weak law of large numbers. Applying this result to (S4.10), we obtain

$$\left|(\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}}^0)^T\frac{\partial L(\tilde{\boldsymbol{\delta}}^0)}{\partial\boldsymbol{\delta}}\right| \lesssim CJ_{n,2}. \tag{S4.11}$$

On the other hand,

$$\begin{aligned}\left|(\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}}^0)^T\,\frac{\partial^2 L(\tilde{\boldsymbol{\delta}}^0)}{\partial\boldsymbol{\delta}\,\partial\boldsymbol{\delta}^T}(\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}}^0)\right| &\lesssim \|\boldsymbol{\delta}-\tilde{\boldsymbol{\delta}}^0\|_2^2\rho_{\max}\left(\sum_{i=1}^n\left[\sum_{l=1}^d \dot{g}_l(W_i)X_{il}\tilde{\boldsymbol{\Phi}}_i\right]^{\otimes 2}\right)\\ &\lesssim C^2J_{n,2}, \end{aligned}\tag{S4.12}$$

where $\rho_{\max}(\cdot)$ represents the largest eigenvalue of a matrix. The relationships (S4.10)–(S4.12) imply that $(\boldsymbol{\delta} - \tilde{\boldsymbol{\delta}}^0)^T \partial L(\tilde{\boldsymbol{\delta}}^0)/\partial \boldsymbol{\delta}$ is dominated by $(\boldsymbol{\delta} - \tilde{\boldsymbol{\delta}}^0)^T [\partial^2 L(\tilde{\boldsymbol{\delta}}^0)/\partial \boldsymbol{\delta}\, \partial \boldsymbol{\delta}^T](\boldsymbol{\delta} - \tilde{\boldsymbol{\delta}}^0)$, which is always positive for large n. Thus, (S4.9) holds. □

Proof of Theorem 1.

Proof. (i) According to the results of Lemma 5 and the properties of B-spline basis functions, we have

$$\|\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}^0(u)\|_{L_2}^2 \asymp J_{n,2}^{-1} \|\hat{\boldsymbol{\delta}} - \tilde{\boldsymbol{\delta}}^0\|_2^2 = O_p(J_{n,2} n^{-1}).$$

Combined with (3.3), we get

$$\begin{aligned} \|\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}(u)\|_{L_2} &\leq \|\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}^0(u)\|_{L_2} + \|\boldsymbol{\beta}^0(u) - \boldsymbol{\beta}(u)\|_{L_2} \\ &= O_p(J_{n,2}^{1/2} n^{-1/2} + J_{n,2}^{-r}). \end{aligned}$$

(ii) By Taylor expansion, we can directly derive

$$\hat{\boldsymbol{\delta}} - \tilde{\boldsymbol{\delta}}^0 = -\left[\frac{\partial^2 L(\tilde{\boldsymbol{\delta}}^0)}{\partial \boldsymbol{\delta}\, \partial \boldsymbol{\delta}^T}\right]^{-1} \frac{\partial L(\tilde{\boldsymbol{\delta}}^0)}{\partial \boldsymbol{\delta}} [1 + o_p(1)].$$

Similar to Lemma 2 of Ma (2016), it can be shown that $\partial^2 L(\tilde{\boldsymbol{\delta}}^0)/\partial \boldsymbol{\delta}\, \partial \boldsymbol{\delta}^T = O_p(n J_{n,2}^{-1})$ and $\|\partial L(\tilde{\boldsymbol{\delta}}^0)/\partial \boldsymbol{\delta}\|_\infty = O_p((\log n)^{1/2} n^{1/2} J_{n,2}^{-1/2})$. Then, applying the results of Demko (1986), we have

$$\|\hat{\boldsymbol{\delta}} - \tilde{\boldsymbol{\delta}}^0\|_\infty \leq \left\| \left[\frac{\partial^2 L(\tilde{\boldsymbol{\delta}}^0)}{\partial \boldsymbol{\delta}\, \partial \boldsymbol{\delta}^T}\right]^{-1} \right\|_\infty \left\| \frac{\partial L(\tilde{\boldsymbol{\delta}}^0)}{\partial \boldsymbol{\delta}} \right\|_\infty = O_p((\log n)^{1/2} J_{n,2}^{1/2} n^{-1/2}).$$

Therefore,

$$\sup_{u\in S_U}|\hat{\boldsymbol{\beta}}(u)-\boldsymbol{\beta}^0(u)|\lesssim\|\hat{\boldsymbol{\delta}}-\tilde{\boldsymbol{\delta}}^0\|_\infty=O_p((\log n)^{1/2}J_{n,2}^{1/2}n^{-1/2}),$$

and hence

$$\begin{aligned}\sup_{u\in S_U}|\hat{\boldsymbol{\beta}}(u)-\boldsymbol{\beta}(u)|&\le\sup_{u\in S_U}|\hat{\boldsymbol{\beta}}(u)-\boldsymbol{\beta}^0(u)|+\sup_{u\in S_U}|\boldsymbol{\beta}^0(u)-\boldsymbol{\beta}(u)|\\&=O_p((\log n)^{1/2}J_{n,2}^{1/2}n^{-1/2}+J_{n,2}^{-r}).\end{aligned}$$

(iii) By Lemma 4, it is easy to see

$$\begin{aligned}&\hat{\boldsymbol{\delta}}-\tilde{\boldsymbol{\delta}}^0\\&=\left\{n^{-1}\sum_{i=1}^n\left[\sum_{l=1}^d\dot{g}_l(W_i)X_{il}\tilde{\boldsymbol{\Phi}}_i\right]^{\otimes2}\right\}^{-1}\left[n^{-1}\sum_{i=1}^n\sum_{l=1}^d\dot{g}_l(W_i)X_{il}\tilde{\boldsymbol{\Phi}}_i\varepsilon_i\right]+o_p(n^{-1/2}J_{n,2}^{1/2}).\end{aligned}$$

According to the Lindeberg–Feller central limit theorem, it is straightforward to derive that

$$\Omega_2^{-1}(u)[\hat{\boldsymbol{\beta}}(u)-\boldsymbol{\beta}^0(u)]\xrightarrow{d}N(0,1).$$

By Slutsky's lemma and the condition $nJ_{n,2}^{-(2r+1)}=o(1)$, the result (iii) holds. Further, for $k=1,\dots,p$, we have

$$\sigma_{2,k}^{-1}(u)[\hat{\beta}_k(u)-\beta_k(u)]\xrightarrow{d}N(0,1).$$

□

Proof of Theorem 2.

Proof. (i) According to the results in Theorem 1(ii) and Proposition 2, it is easy to see that

$$\sup_{w\in S_W} |\hat{g}_l(w,\hat{\boldsymbol{\delta}})-g_l(w)| = O_p(J_{n,1}^{-r}+J_{n,2}^{-r}+(\log n)^{1/2}J_{n,1}^{1/2}n^{-1/2}+(\log n)^{1/2}J_{n,2}^{1/2}n^{-1/2}).$$

(ii) Under the assumptions of Theorem 2, and the conditions $J_{n,2}J_{n,1}^{-1} = o(1)$, $nJ_{n,2}^{-(2r+1)} = o(1)$ and (3.2), we obtain

$$\begin{aligned}
\hat{g}_l(w,\hat{\boldsymbol{\delta}}) - g_l(w) &= \hat{g}_l(w,\hat{\boldsymbol{\delta}}) - g_l^0(w) + g_l^0(w) - g_l(w) \\
&= \hat{g}_{l,\boldsymbol{\varepsilon}}(w,\hat{\boldsymbol{\delta}}) + \hat{g}_{l,\mathbf{g}}(w,\hat{\boldsymbol{\delta}}) - g_l^0(w) + O_p(J_{n,2}^{-r} + J_{n,2}^{1/2}n^{-1/2}) \\
&= \hat{g}_{l,\boldsymbol{\varepsilon}}(w,\hat{\boldsymbol{\delta}}) + o_p(J_{n,1}^{1/2}n^{-1/2}),
\end{aligned}$$

with

$$\hat{g}_{l,\boldsymbol{\varepsilon}}(w,\hat{\boldsymbol{\delta}}) = \mathbf{e}_l^T\mathbb{B}_1(w)\hat{\mathbf{V}}(\hat{\boldsymbol{\delta}})^{-1}n^{-1}\sum_{i=1}^n \mathbf{D}_i(\widehat{W}_i)\varepsilon_i.$$

Furthermore, we decompose $\hat{g}_{l,\boldsymbol{\varepsilon}}(w,\hat{\boldsymbol{\delta}}) = I_1 + I_2$, with

$$\begin{aligned}
I_1 &= \mathbf{e}_l^T\mathbb{B}_1(w)\hat{\mathbf{V}}(\hat{\boldsymbol{\delta}})^{-1}n^{-1}\sum_{i=1}^n \mathbf{D}_i(W_i)\varepsilon_i, \\
I_2 &= \mathbf{e}_l^T\mathbb{B}_1(w)\hat{\mathbf{V}}(\hat{\boldsymbol{\delta}})^{-1}n^{-1}\sum_{i=1}^n [\mathbf{D}_i(\widehat{W}_i) - \mathbf{D}_i(W_i)]\varepsilon_i.
\end{aligned}$$

Since $E\{[D_{i,ml}(\widehat{W}_i) - D_{i,ml}(W_i)]\varepsilon_i\}^2 \lesssim \|\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}(u)\|_{L_2}^2 = O_p(J_{n,2}n^{-1})$, it can be proved using Bernstein's inequality that

$$\sup_{1\le l\le d, 1\le m\le J_{n,1}} \left| n^{-1}\sum_{i=1}^n [D_{i,ml}(\widehat{W}_i) - D_{i,ml}(W_i)]\varepsilon_i \right| = O_p((\log n)^{1/2}J_{n,2}^{1/2}n^{-1}).$$

Thus, $I_2 = o_p(n^{-1/2}J_{n,1}^{1/2})$. Finally, we have

$$\begin{aligned}
&\hat{g}_l(w, \hat{\boldsymbol{\delta}}) - g_l(w) \\
&= \mathbf{e}_l^T \mathbb{B}_1(w) \left[n^{-1} \sum_{i=1}^{n} \mathbf{D}_i(W_i)\mathbf{D}_i(W_i)^T \right]^{-1} n^{-1} \sum_{i=1}^{n} \mathbf{D}_i(W_i)\varepsilon_i + o_p(n^{-1/2}J_{n,1}^{1/2}).
\end{aligned}$$

By checking the Lindeberg condition, it is straightforward to show that

$$\sigma_{1,l}^{-1}(w)[\hat{g}_l(w; \hat{\boldsymbol{\delta}}) - g_l(w)] \xrightarrow{d} N(0, 1).$$

□

S4.4 Proofs of Theorems 3 and 4.

Proof. We show only the proof of Theorem 3, since the proof of Theorem 4 is similar.

Let $\mathit{\Pi} = E\{\sigma^2(U_i, \mathbf{Z}_i, \mathbf{X}_i)[\sum_{l=1}^{d} \dot{g}_l(W_i) X_{il} \tilde{\boldsymbol{\Phi}}_i]^{\otimes 2}\}$ and let $\mathbf{M}_1, \ldots, \mathbf{M}_n$ be independent random variables from the multivariate normal distribution $N(\mathbf{0}, \mathbf{I}_{J_{n,2}p \times J_{n,2}p})$, where $\mathbf{M}_i = \{M_{i,sk}, 1 \le s \le J_{n,2}, 1 \le k \le p\}$ and $\mathbf{I}_{J_{n,2}p \times J_{n,2}p}$ is the $J_{n,2}p \times J_{n,2}p$ identity matrix. Define

$$\eta_k^0(u) = \sigma_{2,k}^{-1}(u)\mathbf{b}_k^T \mathbb{B}_2(u) \Xi_2^{-1} n^{-1} \sum_{i=1}^{n} \mathit{\Pi}^{1/2}\mathbf{M}_i.$$

By the strong approximation theorem (Csörgő and Révész 1981, Theorem 2.6.2), together with the same arguments as were used for Theorem 3 in Ma (2016), it can be proved that

$$\sup_{u\in S_U} |\sigma_{2,k}^{-1}(u)[\hat{\beta}_k(u) - \beta_k(u)] - \eta_k^0(u)| = o_p(J_{n,2}^{1/2} n^{-1/2}). \tag{S4.13}$$

Applying similar techniques as for Theorem 4 in Ma et al. (2015), we partition S_U into K_2 equally spaced intervals with $0 < \nu_0 < \nu_1 < \cdots < \nu_{K_2} < \nu_{K_2+1} = 1$, where $K_2 \to \infty$, and we construct SCBs over a subset of S_U, which is specified as $S_{n,2} = (\nu_0, \ldots, \nu_{K_2})$.

Note that for $0 \le j \le K_2$, $\eta_k^0(u)$ is a Gaussian process with $E[\eta_k^0(u)] = 0$ and $\mathrm{Var}[\eta_k^0(u)] = 1$. By the fact that $|\sigma_{2,k}(u)| = O_p(J_{n,2}^{1/2} n^{-1/2})$ and using Conditions (C3) and (C7), we find that for any $j < j'$, the covariance matrix of the Gaussian process is

$$\begin{aligned}
|\mathrm{Cov}\{\eta_k^0(\nu_j), \eta_k^0(\nu_{j'})\}| &= |n^{-1}\sigma_{2,k}^{-1}(\nu_j)\sigma_{2,k}^{-1}(\nu_{j'})\mathbf{b}_k^T \mathbb{B}_2(\nu_j)\Xi_2^{-1} \mathit{\Pi} \Xi_2^{-1}\mathbb{B}_2(\nu_{j'})^T \mathbf{b}_k| \\
&\asymp J_{n,2}^{-1}|\mathbf{b}_k^T \mathbb{B}_2(\nu_j)\Xi_2^{-1}\mathbb{B}_2(\nu_{j'})^T \mathbf{b}_k| \\
&\asymp |\mathbf{b}_k^T \mathbb{B}_2(\nu_j)\mathbb{B}_2(\nu_{j'})^T \mathbf{b}_k| \asymp \left|\sum_{s=1}^{J_{n,2}} B_{s,2}(\nu_j) B_{s,2}(\nu_{j'})\right|.
\end{aligned}$$

According to the properties of B-spline functions, we have $|\mathrm{Cov}\{\eta_k^0(\nu_j), \eta_k^0(\nu_{j'})\}| \le C^{-|j-j'|}$. Based on the above results and Lemma 1, it is easy to see that

$$\lim_{n\to\infty} P\left\{\sup_{0\le j\le K_2} \left|[2\log(K_2+1)]^{-1/2}\eta_k^0(\nu_j)\right| \le d_{2n}(\alpha)\right\} = 1-\alpha. \tag{S4.14}$$

Combining (S4.13) and (S4.14), we have

$$\lim_{n\to\infty} P\left\{\sup_{u\in S_{n,2}} \left|[2\log(K_2+1)]^{-1/2}\sigma_{2,k}^{-1}(u)[\hat{\beta}_k(u) - \beta_k(u)]\right| \le d_{2n}(\alpha)\right\} = 1-\alpha.$$

□

S4.5 Proof of Theorem 5.

Proof. (i) By (3.2) and Condition (C2), there exists $\beta_k^0(u) = \mathbf{B}_2(u)^T \boldsymbol{\pi}_k^0$ such that $\|\beta_k^P(u) - \beta_k^0(u)\|_{L_2} = O(J_{n,2}^{-r})$ with $\boldsymbol{\pi}_k^0 = (\pi_{1k}^0, \ldots, \pi_{J_{n,2}k}^0)^T$. Let $\hat{\boldsymbol{\pi}}$ be the minimizer of $Q(\boldsymbol{\pi})$ given in (4.1) and let this minimizer be obtained in a neighborhood of $\boldsymbol{\pi}^0$ with probability tending to one. We first show that $\|\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}^0\|_2 = r_n$, with $r_n = O_p(J_{n,2}n^{-1/2} + J_{n,2}^{(1-r)/2}\alpha_n^{1/2})$.

By Taylor expansion, there exists $\boldsymbol{\pi}^*$ that lies between $\hat{\boldsymbol{\pi}}$ and $\boldsymbol{\pi}^0$ such that

$$Q(\hat{\boldsymbol{\pi}}) - Q(\boldsymbol{\pi}^0) = (\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}^0)^T \frac{\partial L(\boldsymbol{\pi}^0)}{\partial \boldsymbol{\pi}} + \frac{1}{2}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}^0)^T \frac{\partial^2 L(\boldsymbol{\pi}^*)}{\partial \boldsymbol{\pi}\, \partial \boldsymbol{\pi}^T}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}^0) + n\sum_{k=2}^{p}[p_{\alpha_n}(\|\hat{\boldsymbol{\pi}}_k\|_{\mathrm{H}}) - p_{\alpha_n}(\|\boldsymbol{\pi}_k^0\|_{\mathrm{H}})]. \tag{S4.15}$$

According to the results of Lemma 5,

$$\begin{aligned}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}^0)^T \frac{\partial^2 L(\boldsymbol{\pi}^*)}{\partial \boldsymbol{\pi}\, \partial \boldsymbol{\pi}^T}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}^0) &= (\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}^0)^T \frac{\partial^2 L(\boldsymbol{\pi}^0)}{\partial \boldsymbol{\pi}\, \partial \boldsymbol{\pi}^T}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}^0)[1 + o_p(1)] \\ &= \|\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}^0\|_2^2 O_p(nJ_{n,2}^{-1}) = O_p(r_n^2 J_{n,2}^{-1} n)\end{aligned} \tag{S4.16}$$

and

$$\left|(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}^0)^T \frac{\partial L(\boldsymbol{\pi}^0)}{\partial \boldsymbol{\pi}}\right| \le \left\|\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}^0\right\|_2 \left\|\frac{\partial L(\boldsymbol{\pi}^0)}{\partial \boldsymbol{\pi}}\right\|_2 = O_p(r_n n^{1/2}). \tag{S4.17}$$

For $k = 2, \ldots, p$, by Minkowski's inequality, it follows that $|\|\hat{\boldsymbol{\pi}}_k\|_{\mathrm{H}} - \|\boldsymbol{\pi}_k^0\|_{\mathrm{H}}| \le \|\hat{\boldsymbol{\pi}}_k - \boldsymbol{\pi}_k^0\|_{\mathrm{H}} = o_p(1)$. Thus, it holds that $\|\hat{\boldsymbol{\pi}}_k\|_{\mathrm{H}} \to \|\beta_k^0\|_{L_2}$ with probability. Note that $\|\beta_k^0\|_{L_2} > 0$ for $k = 2, \ldots, p_1$ and $\alpha_n \to 0$, then, with probability approaching one, we have $\|\hat{\boldsymbol{\pi}}_k\|_{\mathrm{H}} > a\alpha_n$ and $\|\boldsymbol{\pi}_k^0\|_{\mathrm{H}} > a\alpha_n$, which implies $P\{p_{\alpha_n}(\|\hat{\boldsymbol{\pi}}_k\|_{\mathrm{H}}) = p_{\alpha_n}(\|\boldsymbol{\pi}_k^0\|_{\mathrm{H}})\} \to 1, k = 2, \ldots, p_1$. Since $\|\beta_k^0\|_{L_2} = 0$ for $k = p_1 + 1, \ldots, p$, we have $\|\boldsymbol{\pi}_k^0\|_{\mathrm{H}} = O_p(J_{n,2}^{-r})$. Under the condition $J_{n,2}^{-r}/\alpha_n \to 0$, it follows directly that $P\{p_{\alpha_n}(\|\boldsymbol{\pi}_k^0\|_{\mathrm{H}}) = \alpha_n\|\boldsymbol{\pi}_k^0\|_{\mathrm{H}}\} \to 1, k = p_1 + 1, \ldots, p$. Then

$$n\sum_{k=2}^{p}[p_{\alpha_n}(\|\hat{\boldsymbol{\pi}}_k\|_{\mathrm{H}}) - p_{\alpha_n}(\|\boldsymbol{\pi}_k^0\|_{\mathrm{H}})] = -n\alpha_n \sum_{k=p_1+1}^{p} \|\boldsymbol{\pi}_k^0\|_{\mathrm{H}} = -O_p(n\alpha_n J_{n,2}^{-r}). \tag{S4.18}$$

Combining (S4.15)–(S4.18), we have $Q(\hat{\boldsymbol{\pi}}) - Q(\boldsymbol{\pi}^0) = O_p(r_n n^{1/2} + r_n^2 J_{n,2}^{-1} n) - O_p(n\alpha_n J_{n,2}^{-r})$, which implies that $r_n = O_p(J_{n,2} n^{-1/2} + J_{n,2}^{(1-r)/2}\alpha_n^{1/2})$. Furthermore, $\|\hat{\boldsymbol{\beta}}^P(u) - \boldsymbol{\beta}^0(u)\|_{L_2} = O_p(J_{n,2}^{1/2} n^{-1/2} + J_{n,2}^{-r/2}\alpha_n^{1/2})$ holds.

Next we will show that, with probability tending to one, $\hat{\boldsymbol{\pi}}_k = 0$ for $k = p_1 + 1, \ldots, p$. Suppose there exists $k_0 > p_1$ such that $\hat{\boldsymbol{\pi}}_{k_0} \neq 0$. Let $\bar{\boldsymbol{\pi}}$ be a vector constructed by replacing $\hat{\boldsymbol{\pi}}_{k_0}$ with 0 in $\hat{\boldsymbol{\pi}}$. Similar to (S4.15), we have

$$\begin{aligned} Q(\hat{\boldsymbol{\pi}}) - Q(\bar{\boldsymbol{\pi}}) = (\hat{\boldsymbol{\pi}} - \bar{\boldsymbol{\pi}})^T \frac{\partial L(\bar{\boldsymbol{\pi}})}{\partial \boldsymbol{\pi}} + \frac{1}{2}(\hat{\boldsymbol{\pi}} - \bar{\boldsymbol{\pi}})^T \frac{\partial^2 L(\boldsymbol{\pi}^0)}{\partial \boldsymbol{\pi}\, \partial \boldsymbol{\pi}^T}(\hat{\boldsymbol{\pi}} - \bar{\boldsymbol{\pi}}) \\ + n p_{\alpha_n}(\|\hat{\boldsymbol{\pi}}_{k_0}\|_{\mathrm{H}}) + o_p(1). \end{aligned}$$

By definition, $\|\hat{\boldsymbol{\pi}} - \bar{\boldsymbol{\pi}}\|_{\mathrm{H}} = \|\hat{\boldsymbol{\pi}}_{k_0}\|_{\mathrm{H}} > 0$. Noting that $\beta_{k_0}(u) = 0$, we have $\|\hat{\boldsymbol{\pi}}_{k_0}\|_{\mathrm{H}} = \|\hat{\beta}^P_{k_0}(u)\|_{L_2} = O_p(J_{n,2}^{1/2} n^{-1/2} + J_{n,2}^{-r/2}\alpha_n^{1/2})$. Then, under the condition $J_{n,2}^{1/2} n^{-1/2}/\alpha_n \to 0$, $J_{n,2}^{-r}/\alpha_n \to 0$, $p_{\alpha_n}(\|\hat{\boldsymbol{\pi}}_{k_0}\|_{\mathrm{H}}) = \alpha_n \|\hat{\boldsymbol{\pi}}_{k_0}\|_{\mathrm{H}}$ holds. Along the same line of argument as above, we have

$$Q(\hat{\boldsymbol{\pi}}) - Q(\bar{\boldsymbol{\pi}}) = O_p(\|\hat{\boldsymbol{\pi}}_{k_0}\|_{\mathrm{H}} J_{n,2}^{1/2} n^{1/2} + \|\hat{\boldsymbol{\pi}}_{k_0}\|^2_{\mathrm{H}} n) + O_p(n\alpha_n \|\hat{\boldsymbol{\pi}}_{k_0}\|_{\mathrm{H}}). \tag{S4.19}$$

It is easy to see that the first two terms of (S4.19) are dominated by the third term, which is always positive. This contradicts the fact that $Q(\hat{\boldsymbol{\pi}}) - Q(\bar{\boldsymbol{\pi}}) \leq 0$. Thus, (i) holds.

(ii) Let $\boldsymbol{\beta}(u) = (\boldsymbol{\beta}^{(1)}(u)^T, \boldsymbol{\beta}^{(2)}(u)^T)^T$, with $\boldsymbol{\beta}^{(1)}(u) = (\beta_1(u), \ldots, \beta_{p_1}(u))^T$ and $\boldsymbol{\beta}^{(2)}(u) = (\beta_{p_1+1}(u), \ldots, \beta_p(u))^T$. Denote $\boldsymbol{\pi} = ((\boldsymbol{\pi}^{(1)})^T, (\boldsymbol{\pi}^{(2)})^T)^T$, with $\boldsymbol{\pi}^{(1)} = (\boldsymbol{\pi}_1^T, \ldots, \boldsymbol{\pi}_{p_1}^T)^T$ and $\boldsymbol{\pi}^{(2)} = (\boldsymbol{\pi}_{p_1+1}^T, \ldots, \boldsymbol{\pi}_p^T)^T$. By (3.3) and Condition (C2), there exists a B-spline function $\boldsymbol{\beta}_{\text{oracle}}(u) = (\boldsymbol{\beta}^{(1)}_{\text{oracle}}(u)^T, \boldsymbol{\beta}^{(2)}_{\text{oracle}}(u)^T)^T$ satisfying $\|\boldsymbol{\beta}_{\text{oracle}}(u) - \boldsymbol{\beta}(u)\|_{L_2} = O_p(J_{n,2}^{-r})$. Denote the corresponding B-spline coefficients by $\boldsymbol{\pi}_{\text{oracle}} = ((\boldsymbol{\pi}^{(1)}_{\text{oracle}})^T, (\boldsymbol{\pi}^{(2)}_{\text{oracle}})^T)^T$. Since the true function $\boldsymbol{\beta}^{(2)}(u) = 0$, we define $\boldsymbol{\beta}^{(2)}_{\text{oracle}}(u) = \mathbf{0}$ and $\boldsymbol{\pi}^{(2)}_{\text{oracle}} = \mathbf{0}$.

Let $\hat{\boldsymbol{\pi}}$ be the minimizer of $Q(\boldsymbol{\pi})$ given in (4.1). We assume that the minimizer can be obtained in a neighborhood of $\boldsymbol{\pi}_{\text{oracle}}$ with probability tending to one. We will show that $\|\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}_{\text{oracle}}\|_2 = \tilde{r}_n$, with $\tilde{r}_n = O_p(J_{n,2} n^{-1/2})$.

According to the proof of (i), $\|\hat{\boldsymbol{\beta}}^P(u) - \boldsymbol{\beta}(u)\|_{L_2} = o_p(1)$ holds, where

$\hat{\boldsymbol{\beta}}^P(u) = \mathbb{B}_2(u)^T\hat{\boldsymbol{\pi}} = (\hat{\boldsymbol{\beta}}^{(1)}(u)^T, \hat{\boldsymbol{\beta}}^{(2)}(u)^T)^T$. Then $\|\hat{\boldsymbol{\beta}}^{(1)}(u)\|_{L_2} \to \|\boldsymbol{\beta}^{(1)}(u)\|_{L_2} > 0$ with probability. By definition, we have $\|\boldsymbol{\beta}^{(1)}_{\text{oracle}}(u)\|_{L_2} \to \|\boldsymbol{\beta}^{(1)}(u)\|_{L_2} > 0$ with probability. On the other hand, by the results of (i), we have $\|\hat{\boldsymbol{\beta}}^{(2)}(u)\|_{L_2} = 0$ with probability. Thus,

$$\sum_{k=2}^{p} p_{\alpha_n}(\|\boldsymbol{\beta}_{k,\text{oracle}}(u)\|_{L_2}) = \sum_{k=2}^{p} p_{\alpha_n}(\|\hat{\boldsymbol{\beta}}^P_k(u)\|_{L_2})$$

holds with probability tending to one. Similar to (S4.15), we have

$$\begin{aligned} & Q(\hat{\boldsymbol{\pi}}) - Q(\boldsymbol{\pi}_{\text{oracle}}) \\ &= (\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}_{\text{oracle}})^T \frac{\partial L(\boldsymbol{\pi}_{\text{oracle}})}{\partial \boldsymbol{\pi}} + \frac{1}{2}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}_{\text{oracle}})^T \frac{\partial^2 L(\boldsymbol{\pi}_{\text{oracle}})}{\partial \boldsymbol{\pi}\partial \boldsymbol{\pi}^T}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}_{\text{oracle}}) + o_p(1) \\ &= O_p(\tilde{r}_n n^{1/2} + \tilde{r}_n^2 J_{n,2}^{-1} n), \end{aligned} \tag{S4.20}$$

which implies that $\tilde{r}_n = O_p(J_{n,2}n^{-1/2})$ by the fact that the second term of (S4.20) is always positive. Further, we have $\|\hat{\boldsymbol{\beta}}^P(u) - \boldsymbol{\beta}_{\text{oracle}}(u)\|_{L_2} = J_{n,2}^{1/2}n^{-1/2}$. Finally, $\|\hat{\boldsymbol{\beta}}^P(u) - \boldsymbol{\beta}(u)\|_{L_2} = O_p(J_{n,2}^{1/2}n^{-1/2} + J_{n,2}^{-r})$ holds. □

S4.6 Proofs of Theorems 6 and 7.

Proof. We only give the proof of Theorem 7, since Theorem 6 can be obtained by setting $\Delta(u) = 0$ in this proof.

The test statistic can be decomposed into $\mathcal{T} = \mathcal{T}_1 + \mathcal{T}_2 + \mathcal{T}_3$, where

$$\mathcal{T}_1 = \frac{n}{J_{n,2}} \int_0^1 \|\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}_n(u)\|^2 \, du,$$
$$\mathcal{T}_2 = \frac{n}{J_{n,2}} \int_0^1 \|\boldsymbol{\beta}_n(u) - \hat{\boldsymbol{\beta}}\|^2 \, du,$$
$$\mathcal{T}_3 = \frac{2n}{J_{n,2}} \int_0^1 [\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}_n(u)]^T [\boldsymbol{\beta}_n(u) - \hat{\boldsymbol{\beta}}] \, du.$$

By Condition (C2), there exists a B-spline function $\boldsymbol{\beta}_n^0(u)$ satisfying $\|\boldsymbol{\beta}_n^0(u) - \boldsymbol{\beta}_n(u)\| = O_p(J_{n,2}^{-r})$. Rewrite $\mathcal{T}_1 = \mathcal{T}_{11} + \mathcal{T}_{12} + \mathcal{T}_{13}$, with

$$\mathcal{T}_{11} = \frac{n}{J_{n,2}} \int_0^1 \|\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}_n^0(u)\|^2 \, du,$$
$$\mathcal{T}_{12} = \frac{n}{J_{n,2}} \int_0^1 \|\boldsymbol{\beta}_n^0(u) - \boldsymbol{\beta}_n(u)\|^2 \, du,$$
$$\mathcal{T}_{13} = \frac{2n}{J_{n,2}} \int_0^1 [\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}_n^0(u)]^T [\boldsymbol{\beta}_n^0(u) - \boldsymbol{\beta}_n(u)] \, du.$$

Denote $\boldsymbol{\Psi}_i = \sum_{l=1}^d \dot{g}_l(W_i) X_{il} \tilde{\boldsymbol{\Phi}}_i$. Following the proof of Theorem 1, we have

$$\begin{aligned}
\mathcal{T}_{11} &= \frac{n}{J_{n,2}} \int_0^1 [\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}_n^0(u)]^T [\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}_n^0(u)] \, du \\
&= \frac{n}{J_{n,2}} \int_0^1 \mathbb{B}_2(u) \Xi_2^{-1} \left(\frac{1}{n} \sum_{i=1}^n \boldsymbol{\Psi}_i \varepsilon_i \right) \left(\frac{1}{n} \sum_{i=1}^n \boldsymbol{\Psi}_i^T \varepsilon_i \right) \Xi_2^{-1} \mathbb{B}_2^T(u) \, du + o_p(1) \\
&= \frac{1}{J_{n,2}} \int_0^1 \mathbb{B}_2(u) \Xi_2^{-1} \left(\frac{1}{n} \sum_{i=1}^n \boldsymbol{\Psi}_i \boldsymbol{\Psi}_i^T \varepsilon_i^2 \right) \Xi_2^{-1} \mathbb{B}_2^T(u) \, du \\
&\quad + \frac{2}{J_{n,2}} \int_0^1 \mathbb{B}_2(u) \Xi_2^{-1} \left(\frac{1}{n} \sum_{i=1}^n \sum_{i<j} \boldsymbol{\Psi}_i \boldsymbol{\Psi}_j^T \varepsilon_i \varepsilon_j \right) \Xi_2^{-1} \mathbb{B}_2^T(u) \, du + o_p(1) \\
&= \mathcal{T}_{11}^{(1)} + \mathcal{T}_{11}^{(2)} + o_p(1).
\end{aligned}$$

Similar to the proof of Theorem 4.3 in Vogt (2015) and Hu, Huang, and You (2019), it can be proved that $\mathcal{T}_{11}^{(1)} = B + o_p(1)$ and $\mathcal{T}_{11}^{(2)} \xrightarrow{d} N(0, V)$. By definition, we have $\mathcal{T}_{12} = o_p(1)$. Using the Cauchy–Schwarz inequality, we have $\mathcal{T}_{13} = o_p(1)$. Note that $\mathcal{T}_2 = \mathcal{T}_{21} + \mathcal{T}_{22} + \mathcal{T}_{23}$ with

$$\begin{aligned}
\mathcal{T}_{21} &= \frac{n}{J_{n,2}} \int_0^1 \|\boldsymbol{\beta}_n(u) - \boldsymbol{\beta}\|^2 \, du, \\
\mathcal{T}_{22} &= \frac{n}{J_{n,2}} \int_0^1 \|\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\|^2 \, du, \\
\mathcal{T}_{23} &= \frac{2n}{J_{n,2}} \int_0^1 [\boldsymbol{\beta}_n(u) - \boldsymbol{\beta}]^T [\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}] \, du.
\end{aligned}$$

Then $\mathcal{T}_{21} = \tilde{\Delta}$. Since $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_2 = o_p(n^{-1/2})$, we have $\mathcal{T}_{22} = o_p(1)$. Furthermore, it is easy to see $\mathcal{T}_{23} = o_p(1)$ and $\mathcal{T}_3 = o_p(1)$ by the Cauchy–Schwarz inequality. Thus, we finally complete the proof. $\square$

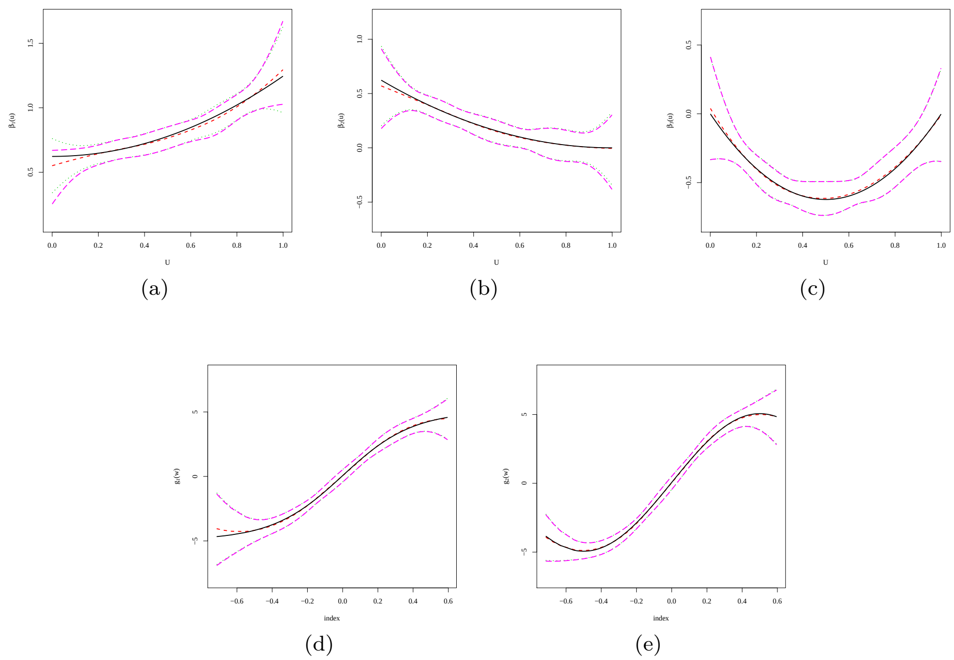


Figure 1: (a)-(c) give the curves of varying coefficient functions $\beta_k(\cdot)$, for $k = 1, \ldots, 3$. Curves are the true functions (solid), the three step spline estimators (dashed), the 95% PCBs (dotted) and the 95% wild bootstrap intervals (long dashed). (d)-(e) give the curves of index functions $g_l(\cdot)$, for $l = 1, 2$. Curves are the true function (solid), the three step spline estimators (dashed), the 95% PCBs (dotted) and the 95% wild bootstrap intervals (long dashed).

S5 Numerical Example

Example 3.(Continued) In this example, we investigate the finite-sample performance of the proposed model identification strategy presented in

Section 4.1. The data are generated from the model (1.2) with varying-coefficient components $\beta_1(u) = 1 + u^2$, $\beta_3(u) = (1-u)^2$, $\beta_5(u) = -1 + 4(u-0.5)^2$, and $\beta_2(u) = \beta_4(u) = \beta_6(u) = 0$. The index components are the same as in Example 1. Covariates X_l and Z_k with $l = 1, 2$ and $k = 1, 2, 3$ are generated by the same way as in Example 1. Covariates Z_4, Z_5, and Z_6 are simulated from an independent $U(0, 1)$. The interior variable U is generated from $U(0, 1)$, and the error term is given as in Example 1.

As in Wang, Li, and Huang (2008) and Hu, Huang, and You (2019), we represent the percentage of selections of the correct model (CF), the percentage of under-fitting (UF), the percentage of over-fitting (OF), the average number of selections of relevant variables in the model (AR), and the average number of selections of irrelevant variables in the model (AI) based on 500 replications, with sample size n = 500, 700, 900. From Table 1, it can be seen that the percentage of selections of the correct model (CF) increases with sample size. Meanwhile, the average number of selections of the relevant variables is approximately equal to the number of nonzero varying-coefficient components, and the average number of selections of irrelevant variables (AI) decreases with increasing sample size.

Table 1: Simualtion results of variable selection in example 2.

n	UF	CF	OF	AR	AI
500	0.000	0.506	0.494	3.000	0.628
700	0.000	0.660	0.340	3.000	0.418
900	0.000	0.816	0.184	3.000	0.210

S6 Real Data Analysis(Continued)

In this section, we illustrate the proposed model via the analysis of two real data applications.

S6.1 Body Fat Dataset

The body fat dataset contains 252 observations on subjects from 22 to 81 years old. These data are available at SatLib. The purpose of this study is to explore the relationship between body fat percentage and the predictors, including age, weight, height and ten body circumference measurements. We exclude six outliers as suggested by Peng and Huang (2011).

We obtain the body fat percentage by Brozek's equation ($457/\text{density} - 414.2$). Let Y be the logarithm of the body fat percentage. According to the Pearson correlation analysis, we select three body circumference measurements with the strongest Pearson correlation coefficients with the response. These are abdomen (Z_1), chest (Z_2), and hip (Z_3). All predictors are centered and standardized. Note that the body fat percentage may depend on weight, and so we divide the data into three groups according to weight: light-weight (lighter than the first quantile of weight), medium-weight (between the first and second quantiles of weight), and high-weight (heavier than the second quantile of weight). For each group, we run a regression analysis using the single-index model

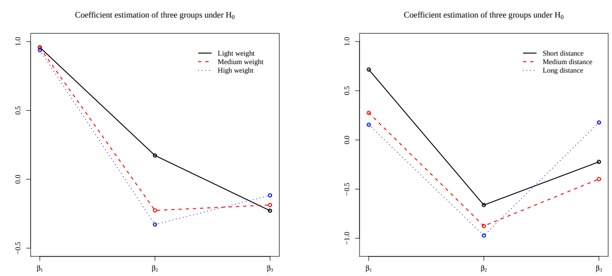
$$Y = g(\mathbf{Z}^T \boldsymbol{\beta}) + \varepsilon, \tag{S6.21}$$

where $g(\cdot)$ is an unknown smooth function, $\mathbf{Z} = (Z_1, Z_3, Z_3)^T$, and $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^T$ is a loading parameter vector. Figure 2(a) depicts three broken lines connected by variant estimated coefficient vectors $(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)$, which represent light-weight (solid), medium-weight (long dashed), and high-weight (dashed), respectively. It is clear that abdomen circumference is the most important factor in each group, while the effect of chest circumference varied quite a lot. The parameters β_1, β_2 and β_3 change with weight, which inspires us to further investigate the possible dynamic effect of body circumference measurements.

Let $U = (1 -$ fraction of body fat$) \times$ weight be the fat-free weight. We focus on the model

$$Y = g(\mathbf{Z}^T \boldsymbol{\beta}(U)) + \varepsilon, \tag{S6.22}$$

where $\mathbf{Z}^T \boldsymbol{\beta}(U) = Z_1\beta_1(U) + Z_2\beta_2(U) + Z_3\beta_3(U)$. The interior index U is



(a) (b)

Figure 2: Broken lines connected by the estimated coefficient vectors of different groups for two real data. (a) Light weight (solid), Medium weight (long dashed) and High weight (dashed); (b)Short distance (solid), Medium distance (long dashed) and Long distance (dashed).

normalized. To check the applicability of the model (S6.22), we conduct the hypothesis procedure described in Section 4.2 with 500 wild bootstrap resamples, and the resulting p-value is less than 0.01. Thus, there is sufficient evidence to reject H_0 (the model (S6.21)) at a 0.05 significance level. Based on the BIC in the simulation study, we choose as the optimal parameters $(q_1, q_2, K_1, K_2) = (3, 3, 2, 2)$. The residual standard deviation is 0.2141. In addition, the R^2 of the model (S6.22) is 0.7777, and thus this model outperforms the model (S6.21), whose R^2 is 0.6569.

Figures 3(a)–(c) give the spline estimates (solid) and the corresponding 95% SCBs (dashed) for the loading functions $\beta_k(\cdot)$, $k = 1, 2, 3$, based on 500 wild bootstrap samples. Figure 3(a) shows that the interaction effect between abdomen circumference and fat-free weight increases as the fat-free weight grows, and that their interaction has the most important positive effect on body fat percentage among the selected circumferences. Figure 3(b) represents that the interaction of chest circumference changes slightly when the fat-free weight is less than its mean value, and decreases monotonically when the fat-free weight exceeds its mean. In addition, their interaction results in a negative trend with increasing body fat percentage. Figure 3(c) shows that the interaction effect between hip circumference and fat-free weight is nearly linear when the fat-free weight is less than zero, and shows a decreasing trend as the fat-free weight increases. Moreover, their interaction has a negative influence on body fat percentage. Figure 3(d) shows the estimated g function (solid) versus the estimated index, together with the corresponding 95% SCBs (long dashed), as well as the response variables (scatter points). It can be seen that the estimated g function shows an increasing trend as the estimated index $\mathbf{Z}^T\hat{\boldsymbol{\beta}}(U)$ increases and that the combination of the three selected circumferences and fat-free weight has a positive impact on body fat percentage. The QQ plot of the residuals is shown in Figure 3(e).

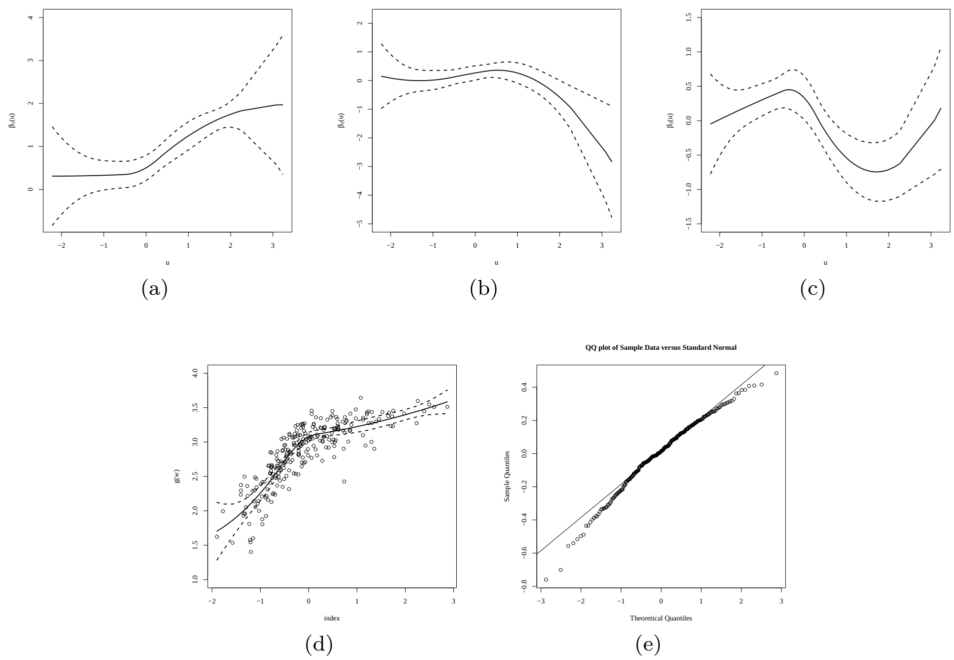


Figure 3: (a)-(c) The estimators (solid) of varying coefficient functions and the 95% SCB (dashed) for β_k, $k = 1, 2, 3$, respectively; (d)The spline estimators (solid), the corresponding 95% SCBs (dashed) of index function $g(\cdot)$ and the response data Y (scatter points); (e) The QQ plot of the residuals.

S6.2 Boston Housing Data

The Boston house-price data are for 506 different houses in the Boston Standard Metropolitan Statistical Area in 1970 and include the median value of owner-occupied homes and 13 sociodemographic variables. The aim of using these data is to study the potential relationship between response and predictors.

We consider several predictors commonly used in the literature, namely, the average number of rooms per dwelling (RM, Z_1), the percentage of the population who are of lower status (LSTA, Z_2), and the full-value property tax per \$10,000 (TAX, Z_3). On the other hand, we are interested in the impact of the per capita crime rate by town (CRIM), so we take $X_1 = 1$ as the intercept term and X_2 as the logarithmic transformation of CRIM. Chaudhuri, Doksum and Samarov (1997) and Wu, Yu and Yu (2010) found that the effect of weighted distances to five Boston employment centers (DIS) varies wildly at different quantiles, which motivates us to study the dynamic interaction between DIS and other predictors.

To begin, we similarly split the data into three groups according to the quantiles of distance, which are short-distance, medium-distance, and long-distance groups. Then we fit each group by the single-index model

$$Y = g_1(\mathbf{Z}^T\boldsymbol{\beta}) + g_2(\mathbf{Z}^T\boldsymbol{\beta})X_2 + \varepsilon, \tag{S6.23}$$

where $g_l(\cdot)$, $l = 1, 2$, are unknown smooth functions, $\mathbf{Z} = (Z_1, Z_2, Z_3)^T$, and $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^T$ is the coefficient vector. Covariates $\mathbf{Z}$ are standardized, and the response Y is centered around zero. Figure 2(b) shows the estimated coefficient vectors of the three groups: short-distance (solid), medium-distance (long dashed), and long-distance (dashed). It is worth noting that DIS does indeed modify the interaction pattens. Therefore, we select $U=$ DIS and consider the model

$$Y = g_1(\mathbf{Z}^T\boldsymbol{\beta}(U)) + g_2(\mathbf{Z}^T\boldsymbol{\beta}(U))X_2 + \varepsilon, \tag{S6.24}$$

where $\boldsymbol{\beta}(\cdot) = (\beta_k(\cdot), 1 \leq k \leq 3)^T$. We perform the hypothesis test from Section 4.2 with 500 wild bootstrap resamples, and we find that the test p-value is 0.058. Thus, we reject H_0 (the model (S6.23)) at a 0.1 significance level. We choose the optimal parameters $(q_1, q_2, K_1, K_2) = (3, 3, 4, 2)$ according to the BIC. Figure 4 shows the resulting spline estimators (solid) and the corresponding 95% SCBs (dashed) based on 500 wild bootstrap samples for $\beta_k(u)$, $k = 1, 2, 3$. The R^2 of the model (S6.24) is 0.8857, while the R^2 under the null hypothesis is 0.8146. The residual standard deviation of the model (S6.24) is 3.1096.

Figure 4(a) shows that the interaction relationship between RM and DIS becomes stronger with increasing DIS and that their combination has a significant positive nonlinear impact on the median value of owner-occupied homes since the 95% SCBs are above zero. Figure 4(b) shows that there is an increasing nonlinear interaction relationship between LSTA and DIS and that their interactions have a negative impact on the response. Figure 4(c) shows that the interaction effect of TAX decreases with increasing DIS, and that their interaction has a significant negative effect on the response. The spline estimators of the index functions $g_l(\cdot)$, $l = 1, 2$, and the corresponding

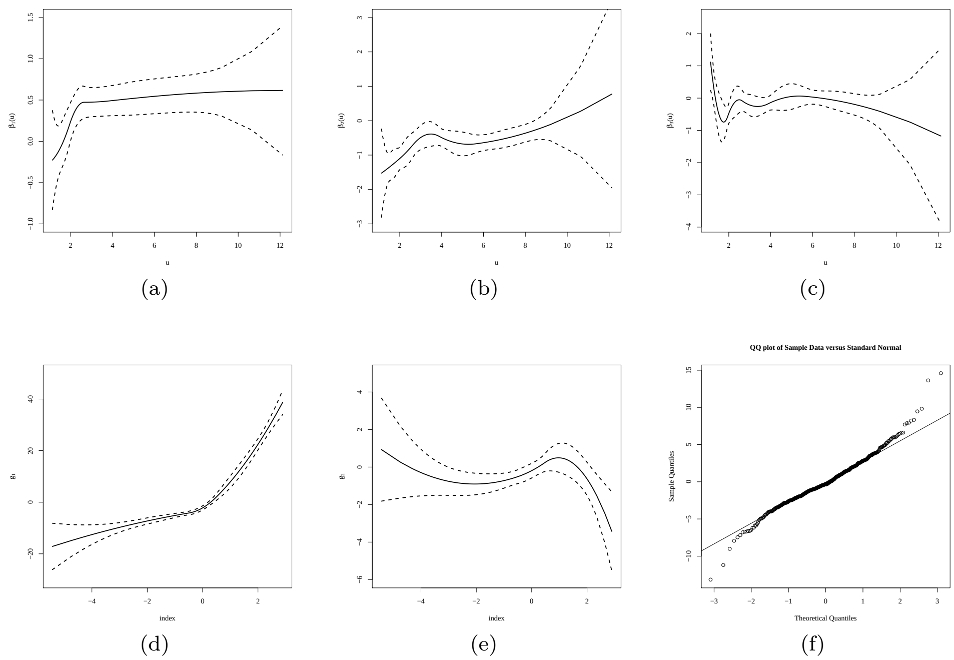


Figure 4: (a)-(c) The estimators (solid) of varying coeffcient functions and the 95% SCB (dashed) for $\beta_k(\cdot)$, $k = 1, \cdots, 3$, respectively; (d-e)The spline estimates (solid) and the 95% SCBs (dashed) for index functions $g_l(\cdot)$, $l = 1, 2$; (f) The QQ plot of the residuals.

95% SCBs are depicted in Figures 4(d) and (e), respectively. Figure 4(d) shows that there is a truly nonlinear relationship between the median value of owner-occupied homes and the predictors, and that the combined effect of DIS and other sociodemographic variables is monotonically increasing and changes from negative to positive. Figure 4(e) shows that the modification due to CRIM is altered by the admixture of other variables, and that this influence is negative with a decreasing trend. In addition, the QQ plot of the residuals in Figure 4(f) shows that the model (S6.24) is a a reasonable option for this dataset.

Bibliography

Chaudhuri, P., Doksum, K. and Samarov, A. (1997). On average derivative quantile regression. *Ann. Stat.* **25**, 715-744.

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes.* Springer-Verlag, New York.

Csörgő, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics.* Academic Press, New York.

de Boor, C. (2001). *A Practical Guide to Splines.* Springer, New York.

Demko, S. (1986). Spectral bounds for $\|a^{-1}\|_\infty$. *J. Approx. theory* **48**, 207-212.

DeVore, R. and Lorentz, G. (1993). *Constructive Approximation.* Springer-Verlag, Berlin.

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.

Ma, S., Carroll, R. J., Liang, H. and Xu, S. (2016). Estimation and inference in generalized additive coefficient models for nonlinear interactions with high-dimensional covariates. *Ann. Stat.* **43**, 2102-2131.

Ma, S. and Song, P. (2015). Varying index coefficient models. *J. Amer. Statist. Assoc.* **110**, 341-356.

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

Ma, S. and Yang, L. (2011). A jump-detecting procedure based on spline estimation. *J. Nonparametr. Stat.* **23**, 67-81.

Peng, H. and Huang, T. (2011). Penalized least squares for single index models. *J. Stat. Plan. Infer.* **141**, 1362-1379.

Vogt, M. (2015). Testing for structural change in time-varying nonparametric regression models. *Economet. Theor.* **31**, 811-859.

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.

Wu, T., Yu, K., and Yu, Y. (2010). Single-index quantile regression. *J. Multivariate Anal.* **101**, 1607-1621.