Generalized functional feature regression models

Qingzhi Zhong[1], Wei Liu[2], Li Liu[3], Hua Liang[4], Huazhen Lin[2,*]

[1] *Department of Statistics and Data Science, School of Economics,*

Jinan University, Guangzhou, China.

[2] *Center of Statistical Research and School of Statistics, New Cornerstone Science Laboratory,*

Southwestern University of Finance and Economics, Chengdu, Sichuan, China.

[3] *School of Mathematics and Statistics, Wuhan University, Wuhan, China.*

[4] *Department of Statistics, George Washington University, Washington, D.C., USA*

Supplementary Material

Before proving the main results, we first introduce some notations, define the covering number and present several preliminary lemmas, which will be frequently used for proving the main results.

*Corresponding author. Email: *linhz@swufe.edu.cn*. The research was supported by National Key R&D Program of China (No.2022YFA1003702), National Natural Science Foundation of China (Nos. 11931014 and 12171374), and New Cornerstone Science Foundation.

S1. Notations

Denote $\mathbf{B}_{np}(t)$ the $pk_n \times p$ block diagonal matrix with block elements $\mathbf{B}_n(t)$, and $\mathbf{B}_{ni}$ is $pk_n \times pn_i$ block diagonal matrix with block elements $\mathbf{B}_n(\mathbf{t}_i)'$, $(\breve{\boldsymbol{\alpha}}, \breve{\boldsymbol{\Gamma}}, \breve{\mathbf{U}}, \breve{\mathbf{H}}, \breve{\boldsymbol{\delta}}, \breve{\boldsymbol{\vartheta}})$ is the maximizer of (2.3) without imposing a sparsity penalty. Write

$$\mathcal{G}_n = \left\{ \boldsymbol{\delta}'\mathbf{S}_n(x) : \boldsymbol{\delta} = (\delta_1, \cdots, \delta_{\widetilde{k}_n})' \in R^{\widetilde{k}_n}, \max_{1 \le i \le \widetilde{k}_n} |\delta_i| \le M, x \in [0,1] \right\},$$

$$\boldsymbol{\Theta}_n^* = \{\boldsymbol{\Theta}_n = (\overrightarrow{\mathbf{H}}', g, \boldsymbol{\psi}')' \in R^{dK_n} \otimes \mathcal{G}_n \otimes \prod_{j=1}^{d} \mathcal{G}_n, \|\mathbf{H}\| = 1\},$$

$$\mathbf{V}_{i1} = \mathbf{B}_{ni}\mathbf{Z}_i,\ \mathbf{V}_{i2} = \mathbf{B}_{ni}\mathbf{B}_{ni}',\ \mathbf{X}_i = \frac{1}{n_i}(\mathbf{V}_{i1} - \mathbf{V}_{i2}\overrightarrow{\breve{\boldsymbol{\alpha}}}),$$

$$\mathbf{e}_i = (\mathbf{e}_{i1}', \cdots, \mathbf{e}_{ip}')',\ \mathbf{e}_{iq} = e_{iq}(\mathbf{t}_i),$$

$$P\ell(\boldsymbol{\Theta}_n; \mathbf{u}) = E\ell(\boldsymbol{\Theta}_n; \mathbf{u}_i),\ P_n\ell(\boldsymbol{\Theta}_n; \mathbf{u}) = \frac{1}{n}\sum_{i=1}^{n} \ell(\boldsymbol{\Theta}_n; \mathbf{u}_i).$$

Let $\sqrt{n}(P_n - P)\ell(\boldsymbol{\Theta}_n; \mathbf{u})$ be the empirical process indexed by $\ell(\boldsymbol{\Theta}_n; \mathbf{u}_i)$. $\widehat{\mathbf{u}}_i(\boldsymbol{\Theta}_n)$ is the estimator of $\mathbf{u}_i$ given $\boldsymbol{\Theta}_n$, and $\widehat{\mathbf{u}}_i \equiv \widehat{\mathbf{u}}_i(\widehat{\boldsymbol{\Theta}}_n)$, $\mathbf{u}_i \equiv \mathbf{u}_i(\boldsymbol{\Theta}_0)$. Let $N(\epsilon, \mathcal{L}_n, L_1(P_n))$ denote the covering number of the function class $\mathcal{L}_n = \{\ell(\boldsymbol{\Theta}_n; \widehat{\mathbf{u}}_i(\boldsymbol{\Theta}_n)) : \boldsymbol{\Theta}_n \in \boldsymbol{\Theta}_n^*\}$. For any $\epsilon > 0$, define the covering number $N(\epsilon, \mathcal{L}_n, L_1(P_n))$ as the smallest value of κ for which there exist $\{\boldsymbol{\Theta}_{n,j} \in \boldsymbol{\Theta}_n^*, k = 1, \cdots, \kappa\}$, such that for any $\boldsymbol{\Theta}_n \in \boldsymbol{\Theta}_n^*$,

$$\min_{k \in \{1, \cdots, \kappa\}} \frac{1}{n}\sum_{i=1}^{n} |\ell(\boldsymbol{\Theta}_n; \widehat{\mathbf{u}}_i(\boldsymbol{\Theta}_n)) - \ell(\boldsymbol{\Theta}_{n,k}; \widehat{\mathbf{u}}_i(\boldsymbol{\Theta}_{n,j}))| < \epsilon.$$

If no such κ exists, define $N(\epsilon, \mathcal{L}_n, L_1(P_n)) = \infty$.

S2. Proposition 1 and its proof

Let $\mathcal{O}$ be the parameter space of the parameter $\mathbf{\Omega}_n$ satisfying Condition (C1').

Proposition 1. *Suppose* $\mathcal{O}_{\mathbf{\Omega}_n} = \{\mathbf{\Omega}_n \in \mathcal{O} : L_p(\mathbf{\Omega}_n) \geq L_p(\mathbf{\Omega}_n^{(0)})\}$ *is compact for the initial value* $\mathbf{\Omega}_n^{(0)}$*. Given the proposed iterative algorithm based on (3.5)–(3.10), we have that all the limit points of* $\mathbf{\Omega}_n^{(o)}$ *are local maxima of* $L_p(\mathbf{\Omega}_n)$ *in the parameter space* $\mathcal{O}$*, and* $L_p(\mathbf{\Omega}_n^{(o)})$ *converges monotonically to* $L^* = L_p(\mathbf{\Omega}_n^*)$ *for some* $\mathbf{\Omega}_n^* \in \mathcal{Q}$*, where* $\mathcal{Q}$ *is the set of local maxima in the interior of* $\mathcal{O}$*.*

Before proving Proposition 1, we first give a definition and a lemma, which will be used in the proof of Proposition 1. Let $\mathcal{Q} = \{$ set of local maxima in the interior of the parameter space of the parameter $\mathbf{\Omega}_n\}$.

Definition 1. (David G. Luenberger (2016), page 199) A point-to-set mapping G from $\mathcal{X}$ to $\mathcal{Y}$ is said to be closed at $x \in \mathcal{X}$ if the assumptions 1) $\lim_{k\to\infty} x_k \to x$, 2) $\lim_{k\to\infty} y_k \to y$ and 3) $y_k \in G(x_k)$ imply $y \in G(x)$. Moreover, G is said to be closed over $\mathcal{X}$ if G is closed at every point of $\mathcal{X}$.

Lemma S.1. *(Wu, 1983) Let* $\{\mathbf{\Omega}_n^{(o)}, o = 1, 2, \cdots\}$ *be a sequence generated by* $\mathbf{\Omega}_n^{(o)} = F(\mathbf{\Omega}_n^{(o-1)})$*, and suppose that (i)* F *is a closed point-to-set map over the complement of* $\mathcal{Q}$*; (ii)* $L(\mathbf{\Omega}_n^{(o)}) > L(\mathbf{\Omega}_n^{(o-1)})$ *for all* $\mathbf{\Omega}_n^{(o)} \notin \mathcal{Q}$*. Then* *all the limit points of* $\mathbf{\Omega}_n^{(o)}$ *are local maxima of* L*, and* $L(\mathbf{\Omega}_n^{(o)})$ *converges monotonically to* $L^* = L(\mathbf{\Omega}_n^*)$ *for some* $\mathbf{\Omega}_n^* \in \mathcal{Q}$*.*

Recalling $\mathbf{\Omega}_n = (\boldsymbol{\alpha}, \mathbf{\Gamma}, \mathbf{U}, \mathbf{H}, \boldsymbol{\delta}, \boldsymbol{\vartheta})$ and $\mathbf{\Omega}_n^{(o-1)} = (\boldsymbol{\alpha}^{(o-1)}, \mathbf{\Gamma}^{(o-1)}, \mathbf{U}^{(o-1)}, \mathbf{H}^{(o-1)}, \boldsymbol{\delta}^{(o-1)}, \boldsymbol{\vartheta}^{(o-1)})$ being the $(o-1)$th iterative value of $\mathbf{\Omega}_n$. Let F be the map satisfying $\mathbf{\Omega}_n^{(o)} = F(\mathbf{\Omega}_n^{(o-1)}), o \geq 1$. In the following, we apply Definition 1 and Lemma S.1 to prove the desired results.

Proof of Proposition 1. The proof includes three steps as follows:

Step 1: Show that $F(\mathbf{\Omega}_n)$ is a point-to-point mapping function, a special case of the point-to-set mapping and is closed over the complement of $\mathcal{Q}$. By the explicit iterative equations (3.5)-(3.10) in the main text, we know that $F(\mathbf{\Omega}_n)$ consists of the deterministic combinations and compositions of a series of elementary functions of $\mathbf{\Omega}_n$. Thus, $F(\mathbf{\Omega}_n)$ is continuous and a unique iterative value of $\mathbf{\Omega}_n^{(o)}$ can be obtained. That is, there exists a unique $\mathbf{\Omega}_n^{(o)}$ such that $\mathbf{\Omega}_n^{(o)} = F(\mathbf{\Omega}_n^{(o-1)})$. Therefore, $F(\mathbf{\Omega}_n)$ is a point-to-point mapping function and closed over the complement of $\mathcal{Q}$ by Definition 1.

Step 2: Show that $L_p(\mathbf{\Omega}_n)$ is nondecreasing with respect to the sequence $\{\mathbf{\Omega}_n^{(o)}\}$, i.e., $L_p((\mathbf{\Omega}_n^{(o-1)}) \leq L_p(\mathbf{\Omega}_n^{(o)})$. By the estimation equations in Section

3, we have

$$\boldsymbol{\alpha}^{(o)} = \text{argmax}_{\boldsymbol{\alpha}} L_p(\boldsymbol{\alpha}, \boldsymbol{\Gamma}^{(o-1)}, \mathbf{U}^{(o-1)}, \mathbf{H}^{(o-1)}, \boldsymbol{\delta}^{(o-1)}, \boldsymbol{\vartheta}^{(o-1)}), \quad \text{(S2.1)}$$

$$\boldsymbol{\Gamma}^{(o)} = \text{argmax}_{\boldsymbol{\Gamma}} L_p(\boldsymbol{\alpha}^{(o)}, \boldsymbol{\Gamma}, \mathbf{U}^{(o-1)}, \mathbf{H}^{(o-1)}, \boldsymbol{\delta}^{(o-1)}, \boldsymbol{\vartheta}^{(o-1)}), \quad \text{(S2.2)}$$

$$\mathbf{U}^{(o)} = \text{argmax}_{\mathbf{U}} L_p(\boldsymbol{\alpha}^{(o)}, \boldsymbol{\Gamma}^{(o)}, \mathbf{U}, \mathbf{H}^{(o-1)}, \boldsymbol{\delta}^{(o-1)}, \boldsymbol{\vartheta}^{(o-1)}), \quad \text{(S2.3)}$$

$$\mathbf{H}^{(o)} = \text{argmax}_{\mathbf{H}} L_p(\boldsymbol{\alpha}^{(o)}, \boldsymbol{\Gamma}^{(o)}, \mathbf{U}^{(o)}, \mathbf{H}, \boldsymbol{\delta}^{(o-1)}, \boldsymbol{\vartheta}^{(o-1)}), \quad \text{(S2.4)}$$

$$\boldsymbol{\delta}^{(o)} = \text{argmax}_{\boldsymbol{\delta}} L_p(\boldsymbol{\alpha}^{(o)}, \boldsymbol{\Gamma}^{(o)}, \mathbf{U}^{(o)}, \mathbf{H}^{(o)}, \boldsymbol{\delta}, \boldsymbol{\vartheta}^{(o-1)}), \quad \text{(S2.5)}$$

$$\boldsymbol{\vartheta}^{(o)} = \text{argmax}_{\boldsymbol{\vartheta}} L_p(\boldsymbol{\alpha}^{(o)}, \boldsymbol{\Gamma}^{(o)}, \mathbf{U}^{(o)}, \mathbf{H}^{(o)}, \boldsymbol{\delta}^{(o)}, \boldsymbol{\vartheta}) \quad \text{(S2.6)}$$

Thus, by (S2.1)–(S2.6) we obtain following inequalities, respectively.

$$L_p(\boldsymbol{\Omega}^{(o-1)}) \leq L_p(\boldsymbol{\alpha}^{(o)}, \boldsymbol{\Gamma}^{(o-1)}, \mathbf{U}^{(o-1)}, \mathbf{H}^{(o-1)}, \boldsymbol{\delta}^{(o-1)}, \boldsymbol{\vartheta}^{(o-1)}), \quad \text{(S2.7)}$$

$$L_p(\boldsymbol{\alpha}^{(o)}, \boldsymbol{\Gamma}^{(o-1)}, \mathbf{U}^{(o-1)}, \mathbf{H}^{(o-1)}, \boldsymbol{\delta}^{(o-1)}, \boldsymbol{\vartheta}^{(o-1)})$$
$$\leq L_p(\boldsymbol{\alpha}^{(o)}, \boldsymbol{\Gamma}^{(o)}, \mathbf{U}^{(o-1)}, \mathbf{H}^{(o-1)}, \boldsymbol{\delta}^{(o-1)}, \boldsymbol{\vartheta}^{(o-1)}), \quad \text{(S2.8)}$$

$$L_p(\boldsymbol{\alpha}^{(o)}, \boldsymbol{\Gamma}^{(o)}, \mathbf{U}^{(o-1)}, \mathbf{H}^{(o-1)}, \boldsymbol{\delta}^{(o-1)}, \boldsymbol{\vartheta}^{(o-1)})$$
$$\leq L_p(\boldsymbol{\alpha}^{(o)}, \boldsymbol{\Gamma}^{(o)}, \mathbf{U}^{(o)}, \mathbf{H}^{(o-1)}, \boldsymbol{\delta}^{(o-1)}, \boldsymbol{\vartheta}^{(o-1)}), \quad \text{(S2.9)}$$

$$L_p(\boldsymbol{\alpha}^{(o)}, \boldsymbol{\Gamma}^{(o)}, \mathbf{U}^{(o)}, \mathbf{H}^{(o-1)}, \boldsymbol{\delta}^{(o-1)}, \boldsymbol{\vartheta}^{(o-1)})$$
$$\leq L_p(\boldsymbol{\alpha}^{(o)}, \boldsymbol{\Gamma}^{(o)}, \mathbf{U}^{(o)}, \mathbf{H}^{(o)}, \boldsymbol{\delta}^{(o-1)}, \boldsymbol{\vartheta}^{(o-1)}), \quad \text{(S2.10)}$$

$$L_p(\boldsymbol{\alpha}^{(o)}, \boldsymbol{\Gamma}^{(o)}, \mathbf{U}^{(o)}, \mathbf{H}^{(o)}, \boldsymbol{\delta}^{(o-1)}, \boldsymbol{\vartheta}^{(o-1)})$$
$$\leq L_p(\boldsymbol{\alpha}^{(o)}, \boldsymbol{\Gamma}^{(o)}, \mathbf{U}^{(o)}, \mathbf{H}^{(o)}, \boldsymbol{\delta}^{(o)}, \boldsymbol{\vartheta}^{(o-1)}), \quad \text{(S2.11)}$$

$$L_p(\boldsymbol{\alpha}^{(o)}, \boldsymbol{\Gamma}^{(o)}, \mathbf{U}^{(o)}, \mathbf{H}^{(o)}, \boldsymbol{\delta}^{(o)}, \boldsymbol{\vartheta}^{(o-1)}) \leq L_p(\boldsymbol{\Omega}^{(o)}). \quad \text{(S2.12)}$$

Combining (S2.7)–(S2.12), we conclude

$$L_p(\mathbf{\Omega}_n^{(o-1)}) \leq L_p(\mathbf{\Omega}_n^{(o)}),$$

which shows that $L_p(\mathbf{\Omega}_n)$ indeed does not decrease as the iteration step o increases.

Step 3: Show that if $\mathbf{\Omega}_n^{(o-1)} \notin \mathcal{Q}$, then $L_p(\mathbf{\Omega}_n^{(o-1)}) < L_p(\mathbf{\Omega}_n^{(o)})$ for $\mathbf{\Omega}_n^{(o)} = F(\mathbf{\Omega}_n^{(o-1)})$; and if $\mathbf{\Omega}_n^{(o-1)} \in \mathcal{Q}$, then $L_p(\mathbf{\Omega}_n^{(o-1)}) = L_p(\mathbf{\Omega}_n^{(o)})$. Show the desired results. By the results in Step 2, it can be seen that the map F satisfies (i) if $\mathbf{\Omega}_n^{(o-1)}$ is the local maximum point of $L_p(\mathbf{\Omega}_n)$ in the parameter space $\mathcal{O}_{\mathbf{\Omega}_n}$, then $\mathbf{\Omega}_n^{(o)} = F(\mathbf{\Omega}_n^{(o-1)}) = \mathbf{\Omega}_n^{(o-1)}$, and (ii) if $\mathbf{\Omega}_n^{(o-1)}$ is not the local maximum point of $L_p(\mathbf{\Omega}_n)$ in $\mathcal{O}_{\mathbf{\Omega}_n}$, then $L_p(\mathbf{\Omega}_n^{(o-1)}) < L_p(\mathbf{\Omega}_n^{(o)})$.

Step 4: Show the desired results. By the results obtained in Steps 1–3, and Lemma S.1, we conclude that all the limit points of $\mathbf{\Omega}_n^{(o)}$ are local maxima of $L_p(\mathbf{\Omega}_n)$. Since $L_p(\mathbf{\Omega}_n)$ is a continuous function on the compact space $\mathcal{O}_{\mathbf{\Omega}_n}$, we obtain that $L_p(\mathbf{\Omega}_n^{(o)})$ converges monotonically to $L^* = L_p(\mathbf{\Omega}_n^*)$ for some $\mathbf{\Omega}_n^* \in \mathcal{O}$. Thus, we complete the proofs.

S3. Lemmas

Lemma S.2. *The covering number of the class* $\mathbf{\Theta}_n^*$ *satisfies*

$$N(\epsilon, \mathbf{\Theta}_n^*, L_2) \leq (Md + M)^{(d+1)\widetilde{k}_n} (\epsilon/10)^{-a_1}$$

where M is a finite positive constant and $a_1 = dK_n + (d+1)\widetilde{k}_n$.

Proof: The results follows by applying Lemma 2.5 and Corollary 2.6 in Van De Geer (2000). Concretely, for any $\boldsymbol{\Theta}_1 = (\mathbf{H}_1', g_1, \boldsymbol{\psi}_1')' \in \boldsymbol{\Theta}_n^*$ and $\boldsymbol{\Theta}_2 = (\mathbf{H}_2', g_2, \boldsymbol{\psi}_2')' \in \boldsymbol{\Theta}_n^*$,

$$\begin{aligned}\|\boldsymbol{\Theta}_1 - \boldsymbol{\Theta}_2\| &\leq \|\mathbf{H}_1 - \mathbf{H}_2\| + \|g_1 - g_2\| + \|\boldsymbol{\psi}_1 - \boldsymbol{\psi}_2\| \\ &\leq \|\mathbf{H}_1 - \mathbf{H}_2\| + \sum_{l=1}^{d+1} \|(\boldsymbol{\delta}_{1l} - \boldsymbol{\delta}_{2l})'\mathbf{S}_n(x)\|,\end{aligned}$$

where $\boldsymbol{\delta}_{jl}, j = 1, 2 \in R^{\widetilde{k}_n}$. Lemma 2.5 in Van De Geer (2000) shows that $\{\mathbf{H} \in R^{dK_n}, \|\mathbf{H}\| = 1\}$ is covered by $(5/\{\epsilon/2\})^{dK_n}$ balls with radius $\epsilon/2$. Corollary 2.6 in Van De Geer (2000) shows $\mathcal{G}_n$ is covered by $(5M/\{\epsilon/(2d+2)\})^{\widetilde{k}_n}$ balls with radius $\epsilon/(2d+2)$. Therefore,

$$N(\epsilon, \boldsymbol{\Theta}_n^*, L_2) \leq (Md + M)^{(d+1)\widetilde{k}_n}(\epsilon/10)^{-a_1}.$$

Lemma S.3. *Under Conditions (A1) and (A3), we have for some $0 < M < \infty$,*

(i) $E\|\frac{1}{\sqrt{Nk_n}}\sum_{i=1}^n \mathbf{B}_{ni}\mathbf{e}_i\|^2 \leq M$,

(ii) $E(\frac{1}{n}\sum_{i=1}^n \|\frac{1}{\sqrt{n_i k_n}}\mathbf{B}_{ni}\mathbf{e}_i\|^2) \leq M$,

(iii) $\left\{\int_0^1 \mathbf{B}_{np}(t)\boldsymbol{\phi}(t)'dt\right\}'\left\{\int_0^1 \mathbf{B}_{np}(t)\boldsymbol{\phi}(t)'dt\right\} = \boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0' + O_p(K_n k_n^{-r})$,

recall that $N = \sum_{i=1}^n n_i$.

Proof: (i). Recall $e_{iq,j} = e_{iq}(t_{ij})$ and it is the element of the vector $\mathbf{e}_i$. By Conditions (A3), we have, for any $i \leq n, q \leq p, j \leq n_i$, $E(e_{iq,j}^2) \leq M$.

Since the basis functions $\mathbf{B}_n(\cdot)$ are continuous on the bounded support, it holds that $k_n^{-1}\mathbf{B}_n(t_{ij})'\mathbf{B}_n(t_{ij}) \leq M$ for any t_{ij}. Furthermore, using the independence of the elements in $\mathbf{e}_i$, we obtain

$$\begin{aligned} E\|\frac{1}{\sqrt{Nk_n}}\sum_{i=1}^{n}\mathbf{B}_{ni}\mathbf{e}_i\|^2 &= \frac{1}{Nk_n}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\sum_{q=1}^{p}E(e_{iq,j}^2)\mathbf{B}_n(t_{ij})'\mathbf{B}_n(t_{ij}) \\ &\leq M\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\sum_{q=1}^{p}E(e_{iq,j}^2) \leq M^2. \end{aligned}$$

(ii). Using the same arguments as (i), we have

$$\begin{aligned} E(\|\frac{1}{\sqrt{n_ik_n}}\mathbf{B}_{ni}\mathbf{e}_i\|^2) &= \frac{1}{n_ik_n}\sum_{j=1}^{n_i}\sum_{q=1}^{p}E(e_{iq,j}^2)\mathbf{B}_n(t_{ij})'\mathbf{B}_n(t_{ij}) \\ &\leq M\frac{1}{n_i}\sum_{j=1}^{n_i}\sum_{q=1}^{p}E(e_{iq,j}^2) \leq M^2. \end{aligned}$$

Thus, the result is followed.

(iii). By Corollary 6.21 of Schumacker (1981), there exist $\phi_{nkq}(t) = \boldsymbol{\gamma}_{kq0}'\mathbf{B}_n(t)$, such that $\sup_{t\in[0,1]}|\phi_{nkq}(t) - \phi_{kq0}(t)| = O_p(k_n^{-r}), q = 1, \cdots, p$. Then by (5), we have

$$\begin{aligned} &\left\{\int_0^1 \mathbf{B}_{np}(t)\boldsymbol{\phi}(t)'dt\right\}'\left\{\int_0^1 \mathbf{B}_{np}(t)\boldsymbol{\phi}(t)'dt\right\} \\ &= \left[\int_0^1 \mathbf{B}_{np}(t)\{\mathbf{B}_{np}(t)'\boldsymbol{\Gamma}_0' + \mathbf{1}_{K_n\times p}'O_p(k_n^{-r})\}dt\right]'\left[\int_0^1 \mathbf{B}_{np}(t)\{\mathbf{B}_{np}(t)'\boldsymbol{\Gamma}_0' + \mathbf{1}_{K_n\times p}'O_p(k_n^{-r})\}dt\right] \\ &= \boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0' + O_p(k_n^{-r})\boldsymbol{\Gamma}_0\int_0^1 \mathbf{B}_{np}(t)\mathbf{1}_{K_n\times p}'dt + O_p(k_n^{-r})\int_0^1 \mathbf{1}_{K_n\times p}\mathbf{B}_{np}(t)'dt\boldsymbol{\Gamma}_0' \\ &\quad + O_p(k_n^{-2r})\int_0^1 \mathbf{1}_{K_n\times p}\mathbf{B}_{np}(t)'dt\int_0^1 \mathbf{B}_{np}(t)\mathbf{1}_{K_n\times p}'dt \\ &= \boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0' + O_p(K_nk_n^{-r}), \end{aligned}$$

where $\mathbf{1}_{K_n\times p}$ is the $K_n \times p$ matrix of ones.

Lemma S.4. *Under Conditions (C1') and (A1), we have*

$$\mathbf{X}_i = \left\{\int_0^1 \mathbf{B}_{np}(t)\boldsymbol{\phi}(t)'dt\right\}\mathbf{u}_i + \frac{1}{n_i}\mathbf{B}_{ni}\mathbf{e}_i - \frac{1}{N}\sum_{i=1}^{n}\mathbf{B}_{ni}\mathbf{e}_i + O_p(\frac{k_n^{3/2}}{n_i} + \frac{nk_n^{3/2}}{N} + k_n^{-r}).$$

Proof: Recall

$$\mathbf{X}_i = \frac{1}{n_i}(\mathbf{V}_{i1} - \mathbf{V}_{i2}\overrightarrow{\breve{\boldsymbol{\alpha}}}), \tag{S3.13}$$

where $\mathbf{V}_{i1} = \mathbf{B}_{ni}\mathbf{Z}_i$ and $\mathbf{V}_{i2} = \mathbf{B}_{ni}\mathbf{B}_{ni}'$. To prove the result, we first consider the asymptotical expression of $\overrightarrow{\breve{\boldsymbol{\alpha}}}$. Substituting $\mathbf{Z}_i(t) = \boldsymbol{\mu}(t) + \sum_{k=1}^{K_n} u_{ik}\boldsymbol{\phi}_k(t) + \mathbf{e}_i(t)$ into the estimation equation of $\breve{\boldsymbol{\alpha}}$, we have

$$\begin{aligned}
\overrightarrow{\breve{\boldsymbol{\alpha}}} &= \left(\sum_{i=1}^{n}\mathbf{V}_{i2}\right)^{-1}\sum_{i=1}^{n}\left(\mathbf{V}_{i1} - \mathbf{V}_{i2}\breve{\boldsymbol{\Gamma}}'\breve{\mathbf{u}}_i\right)\\
&= \left\{(\sum_{i=1}^{n}\mathbf{V}_{i2})^{-1} - N^{-1}\mathbf{I}_{k_n} + N^{-1}\mathbf{I}_{k_n}\right\}\cdot\sum_{i=1}^{n}\left\{\mathbf{B}_{ni}\boldsymbol{\mu}(\mathbf{t}_i) - n_i\int_0^1\mathbf{B}_{np}(t)\boldsymbol{\mu}(t)dt + n_i\int_0^1\mathbf{B}_{np}(t)\boldsymbol{\mu}(t)dt\right.\\
&\quad\left. + \sum_{k=1}^{K_n}u_{ik}\mathbf{B}_{ni}\boldsymbol{\phi}_k(\mathbf{t}_i) - n_iu_{ik}\int_0^1\mathbf{B}_{np}(t)\boldsymbol{\phi}_k(t)dt + n_iu_{ik}\int_0^1\mathbf{B}_{np}(t)\boldsymbol{\phi}_k(t)dt) + \mathbf{B}_{ni}\mathbf{e}_i - \mathbf{V}_{i2}\breve{\boldsymbol{\Gamma}}'\breve{\mathbf{u}}_i\right\}.
\end{aligned} \tag{S3.14}$$

By Lemma 6.2 of Cardot (2000), for any $f(t) \in \mathcal{H}_r$ we have $\|\frac{1}{n_i}\sum_{j=1}^{n_i} f(t_{ij})\mathbf{B}_n(t_{ij}) - \int_0^1 f(t)\mathbf{B}_n(t)dt\| = O_p(\frac{\sqrt{k_n}}{n_i})$. Then $(\sum_{i=1}^{n}\mathbf{V}_{i2})^{-1} = N^{-1}\mathbf{I}_{k_n} + O_p(\frac{k_n}{N^2})$ by the orthogonality of spline basis functions. Thus, by Condition (A1) and (S3.14) we have

$$\begin{aligned}
\overrightarrow{\breve{\boldsymbol{\alpha}}} &= \frac{1}{N}\sum_{i=1}^{n}\left(n_i\int_0^1 \mathbf{B}_{np}(t)\boldsymbol{\mu}(t)dt + \sum_{k=1}^{K_n} n_i u_{ik}\int_0^1 \mathbf{B}_{np}(t)\boldsymbol{\phi}_k(t)dt + \mathbf{B}_{ni}\mathbf{e}_i - \mathbf{V}_{i2}\breve{\boldsymbol{\Gamma}}'\breve{\mathbf{u}}_i\right) + O_p(\frac{nk_n^{3/2}}{N}) \\
&= \int_0^1 \mathbf{B}_{np}(t)\boldsymbol{\mu}(t)dt + \frac{1}{N}\sum_{i=1}^{n}\mathbf{B}_{ni}\mathbf{e}_i + N^{-1}\sum_{i=1}^{n}\mathbf{V}_{i2}\breve{\boldsymbol{\Gamma}}'\breve{\mathbf{u}}_i + O_p(\frac{nk_n^{3/2}}{N}). \qquad \text{(S3.15)}
\end{aligned}$$

Further by Condition (C1') that $N^{-1}\sum_{i=1}^{n} n_i\breve{\mathbf{u}}_i = \mathbf{0}$, we obtain $\|N^{-1}\sum_{i=1}^{n}\mathbf{V}_{i2}\breve{\boldsymbol{\Gamma}}'\breve{\mathbf{u}}_i\| \leq \|\breve{\boldsymbol{\Gamma}}\|\times\, \|N^{-1}\sum_{i=1}^{n} n_i\breve{\mathbf{u}}_i\| = 0$. Therefore, (S3.15) simplifies as

$$\overrightarrow{\breve{\boldsymbol{\alpha}}} = \int_0^1 \mathbf{B}_{np}(t)\boldsymbol{\mu}(t)dt + \frac{1}{N}\sum_{i=1}^{n}\mathbf{B}_{ni}\mathbf{e}_i + O_p(\frac{nk_n^{3/2}}{N}). \qquad \text{(S3.16)}$$

Plugging (S3.16) into (S3.13), we have

$$\begin{aligned}
\mathbf{X}_i &= \int_0^1 \mathbf{B}_{np}(t)\boldsymbol{\mu}(t)dt + \sum_{k=1}^{K_n} u_{ik}\int_0^1 \mathbf{B}_{np}(t)\boldsymbol{\phi}_k(t)dt + \frac{1}{n_i}\mathbf{B}_{ni}\mathbf{e}_i - \overrightarrow{\breve{\boldsymbol{\alpha}}} + O_p(\frac{k_n^{3/2}}{n_i} + k_n^{-r}) \\
&= \sum_{k=1}^{K_n} u_{ik}\int_0^1 \mathbf{B}_{np}(t)\boldsymbol{\phi}_k(t)dt + \frac{1}{n_i}\mathbf{B}_{ni}\mathbf{e}_i - \frac{1}{N}\sum_{i=1}^{n}\mathbf{B}_{ni}\mathbf{e}_i + O_p(\frac{k_n^{3/2}}{n_i} + \frac{nk_n^{3/2}}{N} + k_n^{-r}).
\end{aligned}$$

The result follows.

Lemma S.5. *Under Conditions (C1') and (A1)–(A6), we have*

$$\|\breve{\mathbf{u}}_i(\boldsymbol{\Theta}_n) - \mathbf{W}'\mathbf{u}_i\|^2 = O_p\Big\{K_n^{2c_0+5}n/N + K_n^{2c_0+3}k_n^2 n/N + K_n^{2c_0+4}/n + K_n^{2c_0+4}k_n^{-2r}\Big\}$$

uniformly for $\boldsymbol{\Theta}_n \in \boldsymbol{\Theta}_n^*$.

Proof: Recall that $\mathbf{P} = \mathrm{diag}\{\sqrt{n_1},\cdots,\sqrt{n_n}\}$ defined in the iterative algorithm of the main text. Let $\mathbf{X} = (\mathbf{X}_1,\cdots,\mathbf{X}_n)'$ be the $n\times pk_n$ matrix, $\breve{\mathbf{U}}(\boldsymbol{\Theta}_n) = (\breve{\mathbf{u}}_1(\boldsymbol{\Theta}_n),\cdots,\breve{\mathbf{u}}_n(\boldsymbol{\Theta}_n))'$, denote $\mathbf{V}_n$ the $K_n\times K_n$ diagonal matrix consisting of eigenvalues of the matrix $\mathbf{PXX'P}/N$ in decreasing order, $\mathbf{W} = \left\{\int_0^1 \mathbf{B}_{np}(t)\boldsymbol{\phi}(t)'dt\right\}'\left\{\int_0^1 \mathbf{B}_{np}(t)\boldsymbol{\phi}(t)'dt\right\} \times \left(\mathbf{U}'\mathbf{P}^2\breve{\mathbf{U}}(\boldsymbol{\Theta}_n)/N\right)\mathbf{V}_n^{-1}$.

The main proof is in three steps, we first prove

$$\begin{aligned}
\breve{\mathbf{u}}_i(\boldsymbol{\Theta}_n) - \mathbf{W}'\mathbf{u}_i &= \mathbf{V}_n^{-1}N^{-1}\sum_{j=1}^n n_j\breve{\mathbf{u}}_j(\boldsymbol{\Theta}_n)\mathbf{u}_j'\left\{\int_0^1 \mathbf{B}_{np}(t)\boldsymbol{\phi}(t)'dt\right\}'\zeta_i \\
&\quad + \mathbf{V}_n^{-1}N^{-1}\sum_{j=1}^n n_j\breve{\mathbf{u}}_j(\boldsymbol{\Theta}_n)\mathbf{u}_i'\left\{\int_0^1 \mathbf{B}_{np}(t)\boldsymbol{\phi}(t)'dt\right\}'\zeta_j \\
&\quad + \mathbf{V}_n^{-1}N^{-1}\sum_{j=1}^n n_j\breve{\mathbf{u}}_j(\boldsymbol{\Theta}_n)\zeta_j'\zeta_i \\
&= \mathbf{V}_n^{-1}(\mathbf{I}_{i1} + \mathbf{I}_{i2} + \mathbf{I}_{i3}), \qquad \text{(S3.17)}
\end{aligned}$$

uniformly in $\boldsymbol{\Theta}_n \in \boldsymbol{\Theta}_n^*$, where $\zeta_i = n_i^{-1}\mathbf{B}_{ni}\mathbf{e}_i - N^{-1}\sum_{i=1}^n \mathbf{B}_{ni}\mathbf{e}_i + O_p(\frac{k_n^{3/2}}{n_i} + \frac{nk_n^{3/2}}{N} + k_n^{-r})$.

By the estimation equations of $\breve{\boldsymbol{\alpha}}, \breve{\boldsymbol{\Gamma}} = (\overrightarrow{\breve{\boldsymbol{\gamma}}_1}, \cdots, \overrightarrow{\breve{\boldsymbol{\gamma}}_{K_n}})'$ and $\breve{\mathbf{u}}_i(\boldsymbol{\Theta}_n)$, we have

$$\begin{aligned}
&\sum_{i=1}^n \left\{\mathbf{V}_{i1} - \mathbf{V}_{i2}\breve{\boldsymbol{\Gamma}}'\breve{\mathbf{u}}_i(\boldsymbol{\Theta}_n) - \mathbf{V}_{i2}\overrightarrow{\breve{\boldsymbol{\alpha}}}\right\} = \mathbf{0}, \\
&\sum_{i=1}^n \left\{\mathbf{V}_{i1} - \mathbf{V}_{i2}\overrightarrow{\breve{\boldsymbol{\alpha}}} - \mathbf{V}_{i2}\sum_{r=1}^{K_n}\breve{u}_{ir}(\boldsymbol{\Theta}_n)\overrightarrow{\breve{\boldsymbol{\gamma}}_r}\right\}\breve{u}_{ik}(\boldsymbol{\Theta}_n) = \mathbf{0}, \quad k = 1, \cdots, K_n, \\
&\frac{n^{-1}\sum_{i=1}^n \partial\ell(\boldsymbol{\Theta}_n; \breve{\mathbf{u}}_i(\boldsymbol{\Theta}_n))}{2\omega n_i\partial\mathbf{u}_i(\boldsymbol{\Theta}_n)} + \frac{\breve{\boldsymbol{\Gamma}}(\mathbf{V}_{i1} - \mathbf{V}_{i2}\overrightarrow{\breve{\boldsymbol{\alpha}}})}{n_i} - \frac{\breve{\boldsymbol{\Gamma}}\mathbf{V}_{i2}\breve{\boldsymbol{\Gamma}}'}{n_i}\breve{\mathbf{u}}_i(\boldsymbol{\Theta}_n) = \mathbf{0}, i = 1, \cdots, n.
\end{aligned}$$

By Lemma S.2, conditions (A2) and $0 < v < 1 - 1/(2\varepsilon)$, we have $\frac{n^{-1}\sum_{i=1}^n \partial\ell(\boldsymbol{\Theta}_n; \breve{\mathbf{u}}_i(\boldsymbol{\Theta}_n))}{2\omega n_i\partial\mathbf{u}_i(\boldsymbol{\Theta}_n)} = o_p(1)$ as $n_i \to \infty$, uniformly for $\boldsymbol{\Theta}_n \in \boldsymbol{\Theta}_n^*$. Then the last n equations are equivalent to

$$\frac{\breve{\mathbf{\Gamma}}(\mathbf{V}_{i1} - \mathbf{V}_{i2}\overrightarrow{\breve{\boldsymbol{\alpha}}})}{n_i} - \frac{\breve{\mathbf{\Gamma}}\mathbf{V}_{i2}\breve{\mathbf{\Gamma}}'}{n_i}\breve{\mathbf{u}}_i(\mathbf{\Theta}_n) = \mathbf{0}, i = 1, \cdots, n.$$

Since $\mathbf{B}_n(t)$ is orthogonal and $\breve{\mathbf{\Gamma}}\breve{\mathbf{\Gamma}}'$ is diagonal, then $\breve{\mathbf{u}}_i(\mathbf{\Theta}_n)$ can be estimated by applying the principal components method to the covariance matrix $\mathbf{X}'\mathbf{P}^2\mathbf{X}/N$. Further the factors of the matrix $\mathbf{PX}$ also are the eigenvectors of $\mathbf{PXX}'\mathbf{P}/N$, then $\breve{\mathbf{U}}(\mathbf{\Theta}_n)$ is the $n \times K_n$ matrix consisting of K_n unitary eigenvectors (multiplied by $\sqrt{N}\mathbf{P}^{-1}$), associated with the K_n largest eigenvalues of the matrix $\mathbf{PXX}'\mathbf{P}/N$.

By the eigenvalue equation, we know $\mathbf{PXX}'\mathbf{P}^2\breve{\mathbf{U}}(\mathbf{\Theta}_n)/N = \mathbf{P}\breve{\mathbf{U}}(\mathbf{\Theta}_n)\mathbf{V}_n$, which implies $\breve{\mathbf{U}}(\mathbf{\Theta}_n) = \mathbf{XX}'\mathbf{P}^2\breve{\mathbf{U}}(\mathbf{\Theta}_n)/N\mathbf{V}_n^{-1}$. Then

$$\breve{\mathbf{u}}_i(\mathbf{\Theta}_n) = \mathbf{V}_n^{-1}\breve{\mathbf{U}}(\mathbf{\Theta}_n)'\mathbf{P}^2\mathbf{XX}_i/N. \tag{S3.18}$$

Let $\boldsymbol{\zeta} = (\zeta_1, \cdots, \zeta_n)'$. By Lemma S.4, we have $\mathbf{X}_i = \left\{\int_0^1 \mathbf{B}_{np}(t)\boldsymbol{\phi}(t)'dt\right\}\mathbf{u}_i +$ ζ_i. Then

$$\begin{aligned}
\mathbf{XX}_i &= \mathbf{U}\left\{\int_0^1 \mathbf{B}_{np}(t)\boldsymbol{\phi}(t)'dt\right\}'\left\{\int_0^1 \mathbf{B}_{np}(t)\boldsymbol{\phi}(t)'dt\right\}\mathbf{u}_i \qquad \text{(S3.19)}\\
&+ \mathbf{U}\left\{\int_0^1 \mathbf{B}_{np}(t)\boldsymbol{\phi}(t)'dt\right\}'\zeta_i + \mathbf{U}\left\{\int_0^1 \mathbf{B}_{np}(t)\boldsymbol{\phi}(t)'dt\right\}'\zeta_j + \boldsymbol{\zeta}\zeta_i.
\end{aligned}$$

Plugging (S3.19) into (S3.18) yields

$$
\begin{aligned}
\breve{\mathbf{u}}_i(\boldsymbol{\Theta}_n) = \mathbf{W}'\mathbf{u}_i + \mathbf{V}_n^{-1}N^{-1}\sum_{j=1}^{n} n_j \breve{\mathbf{u}}_j(\boldsymbol{\Theta}_n)\mathbf{u}_j' \left\{\int_0^1 \mathbf{B}_{np}(t)\boldsymbol{\phi}(t)'dt\right\}' \zeta_i \\
+\mathbf{V}_n^{-1}N^{-1}\sum_{j=1}^{n} n_j \breve{\mathbf{u}}_j(\boldsymbol{\Theta}_n)\mathbf{u}_i' \left\{\int_0^1 \mathbf{B}_{np}(t)\boldsymbol{\phi}(t)'dt\right\}' \zeta_j \\
+\mathbf{V}_n^{-1}N^{-1}\sum_{j=1}^{n} n_j \breve{\mathbf{u}}_j(\boldsymbol{\Theta}_n)\zeta_j'\zeta_i \\
= \mathbf{W}'\mathbf{u}_i + \mathbf{V}_n^{-1}(\mathbf{I}_{i1} + \mathbf{I}_{i2} + \mathbf{I}_{i3}).
\end{aligned}
\tag{S3.20}
$$

Thus, we obtain the desired result in (S3.17).

Next, we show that

$$
\|\mathbf{V}_n^{-1}\| = O_p(K_n^{c_0+1/2}). \tag{S3.21}
$$

By Lemma S.4 and $\frac{1}{N}\sum_{i=1}^{n} n_i \mathbf{u}_i\mathbf{u}_i' = \mathbf{I}_{K_n}$, we have

$$
\begin{aligned}
\mathbf{X}'\mathbf{P}^2\mathbf{X}/N &= \sum_{i=1}^{n}\frac{n_i}{N}\left\{\int_0^1 \mathbf{B}_{np}(t)\boldsymbol{\phi}(t)'dt\cdot\mathbf{u}_i + \zeta_i\right\}\left\{\int_0^1 \mathbf{B}_{np}(t)\boldsymbol{\phi}(t)'dt\cdot\mathbf{u}_i + \zeta_i\right\}' \\
&= \int_0^1 \mathbf{B}_{np}(t)\boldsymbol{\phi}(t)'dt\int_0^1 \boldsymbol{\phi}(t)\mathbf{B}_{np}(t)'dt + \sum_{i=1}^{n}\frac{n_i}{N}\int_0^1 \mathbf{B}_{np}(t)\boldsymbol{\phi}(t)'dt\cdot\mathbf{u}_i\zeta_i' \\
&\quad + \sum_{i=1}^{n}\frac{n_i}{N}\zeta_i\mathbf{u}_i'\int_0^1 \mathbf{B}_{np}(t)'\boldsymbol{\phi}(t)dt + \sum_{i=1}^{n}\frac{n_i}{N}\zeta_i\zeta_i'.
\end{aligned}
\tag{S3.22}
$$

Moreover, by Lemma S.3 and Condition (C1'),

$$
\begin{aligned}
\frac{1}{N}\sum_{j=1}^{n} n_j \|\int_0^1 \mathbf{B}_{np}(t)\boldsymbol{\phi}(t)'dt\cdot\mathbf{u}_j\|^2 &\le \|\int_0^1 \mathbf{B}_{np}(t)\boldsymbol{\phi}(t)'dt\|^2\cdot\frac{1}{N}\sum_{j=1}^{n} n_j\|\mathbf{u}_j\|^2 \\
&= O_p(K_n^2).
\end{aligned}
\tag{S3.23}
$$

Using Lemma S.3 again yields

$$\begin{aligned}\|\zeta_i\|^2 &= \|\frac{1}{n_i}\mathbf{B}_{ni}\mathbf{e}_i - \frac{1}{N}\sum_{i=1}^{n}\mathbf{B}_{ni}\mathbf{e}_i + O_p(\frac{k_n^{3/2}}{n_i} + \frac{nk_n^{3/2}}{N} + k_n^{-r})\|^2 \\ &\leq M\|\frac{1}{n_i}\mathbf{B}_{ni}\mathbf{e}_i\|^2 + M\|\frac{1}{N}\sum_{i=1}^{n}\mathbf{B}_{ni}\mathbf{e}_i\|^2 + O_p\{(\frac{k_n^{3/2}}{n_i} + \frac{nk_n^{3/2}}{N} + k_n^{-r})^2\} \\ &= O_p\left(k_n/n_i + 1/n + k_n^{-2r}\right). \qquad \text{(S3.24)}\end{aligned}$$

Thus, by (S3.22)-(S3.24) and the Cauchy-Schwarz inequality, we have

$$\|\mathbf{X}'\mathbf{P}^2\mathbf{X}/N - \int_0^1 \mathbf{B}_{np}(t)\boldsymbol{\phi}(t)'dt\int_0^1 \boldsymbol{\phi}(t)\mathbf{B}_{np}(t)'dt\| = O_p(K_n\sqrt{k_nn/N} + K_n/\sqrt{n} + K_nk_n^{-r}). \qquad \text{(S3.25)}$$

Since the eigenvalues of matrices $\mathbf{PXX'P}/N$ and $\mathbf{X'P}^2\mathbf{X}/N$ are the same, and $\mathbf{V}_n$ is a $K_n \times K_n$ diagonal matrix consisting of eigenvalues of the matrix $\mathbf{PXX'P}/N$ in decreasing order. Then $\mathbf{V}_n$ is a diagonal matrix of diagonal elements being eigenvalues of $\mathbf{X'P}^2\mathbf{X}/N$. Similarly, by Lemma S.3 and the definition of $\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0' = diag\{\lambda_1, \ldots, \lambda_{K_n}\}$, we know that $\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0'$ is a diagonal matrix of diagonal elements being eigenvalues of $\int_0^1 \mathbf{B}_{np}(t)\boldsymbol{\phi}(t)'dt\int_0^1 \boldsymbol{\phi}(t)\mathbf{B}_{np}(t)'dt$ up to an $O_p(K_nk_n^{-r})$ term. Therefore, by Weyl's Theorem, we have

$$\begin{aligned}|v_{nk} - \lambda_k| &\leq \|\mathbf{X}'\mathbf{P}^2\mathbf{X}/N - \int_0^1 \mathbf{B}_{np}(t)\boldsymbol{\phi}(t)'dt\int_0^1 \boldsymbol{\phi}(t)\mathbf{B}_{np}(t)'dt\| \\ &= O_p(K_n\sqrt{k_nn/N} + K_n/\sqrt{n} + K_nk_n^{-r}), \qquad \text{(S3.26)}\end{aligned}$$

where $\mathbf{V}_n = diag\{v_{n1}, \ldots, v_{nK_n}\}$. Hence,

$$\|\mathbf{V}_n - \boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0'\| = O_p(a_n),$$

where $a_n = K_n\sqrt{k_n n/N} + K_n/\sqrt{n} + K_n k_n^{-r}$. Denote $b_n = K_n^{c_0+1/2}$, then $a_n b_n \to 0$ by Conditions (A5) and (A6). This implies that $\|\mathbf{V}_n - \boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0'\| \leq 1/(2b_n)$ for large n with large probability. Applying an error bound for matrix inversion (Horn and Johnson, 2012) (page 381), it follows that

$$\frac{\|\mathbf{V}_n^{-1} - (\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0')^{-1}\|}{b_n} \leq \frac{b_n\|\mathbf{V}_n - (\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0')\|}{1 - b_n\|\mathbf{V}_n - (\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0')\|} \leq 1$$

in probability. In addition, (A4) implies that $\lambda_k \geq Mk^{-c_0}$, and then $\lambda_k^{-1} \leq Mk^{c_0}$ for $k = 1, \ldots, K_n$. This gives $\|(\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0')^{-1}\| \leq CK_n^{c_0+1/2} = O_p(b_n)$. Therefore,

$$\|\mathbf{V}_n^{-1}\| \leq \|(\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0')^{-1}\| + \|\mathbf{V}_n^{-1} - (\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0')^{-1}\| = O_p(K_n^{c_0+1/2}).$$

Last, we derive the bounds for $\mathbf{I}_{i1}, \mathbf{I}_{i2}$ and $\mathbf{I}_{i3}$ separately. By the Cauchy-Schwarz inequality, (S3.23), (S3.24) and $\frac{1}{N}\sum_{i=1}^{n} n_i \breve{\mathbf{u}}_i(\boldsymbol{\Theta}_n)\breve{\mathbf{u}}_i(\boldsymbol{\Theta}_n)' = \mathbf{I}_{K_n}$, we have

$$\begin{aligned}\|\mathbf{I}_{i1}\|^2 &\leq \left(\frac{1}{N}\sum_{j=1}^{n} n_j\|\breve{\mathbf{u}}_j(\boldsymbol{\Theta}_n)\|^2\right)\left(\frac{1}{N}\sum_{j=1}^{n} n_j\|\int_0^1 \mathbf{B}_{np}(t)\boldsymbol{\phi}(t)'dt \cdot \mathbf{u}_j\|^2\right)\|\zeta_i\|^2 \\ &= O_p(K_n)O_p(K_n^2)O_p(k_n/n_i + 1/n + k_n^{-2r}). \qquad \text{(S3.27)}\end{aligned}$$

Following the arguments analogously to those of $\mathbf{I}_{i1}$,

$$\|\mathbf{I}_{i2}\|^2 \leq \left(\frac{1}{N}\sum_{j=1}^{n} n_j \|\breve{\mathbf{u}}_j(\mathbf{\Theta}_n)\|^2\right) \| \int_0^1 \mathbf{B}_{np}(t)\boldsymbol{\phi}(t)' dt \cdot \mathbf{u}_i\|^2 \frac{1}{N}\sum_{j=1}^{n} n_j \|\zeta_j\|^2$$
$$\leq O_p(K_n^4 n/N + K_n^2 k_n^2 n/N + K_n^3/n + K_n^3 k_n^{-2r}). \quad \text{(S3.28)}$$

For $\mathbf{I}_{i3}$, by Condition (A3) we have

$$\|\mathbf{I}_{i3}\|^2 = \left(\frac{1}{N}\sum_{j=1}^{n} n_j \|\breve{\mathbf{u}}_j(\mathbf{\Theta}_n)\|^2\right)\left(\frac{1}{N}\sum_{j=1}^{n} n_j \|\zeta_j\|^2\right)\|\zeta_i\|^2$$
$$= O_p(K_n) O_p(k_n n/N + 1/n + k_n^{-2r}) O_p(k_n/n_i + 1/n + k_n^{-2r}). \quad \text{(S3.29)}$$

Combine (S3.27)-(S3.29), we have

$$\|\mathbf{I}_{i1}\|^2 + \|\mathbf{I}_{i2}\|^2 + \|\mathbf{I}_{i3}\|^2 = O_p(K_n^4 n/N + K_n^2 k_n^2 n/N + K_n^3/n + K_n^3 k_n^{-2r}). \quad \text{(S3.30)}$$

Note that $\|\breve{\mathbf{u}}_i(\mathbf{\Theta}_n) - \mathbf{W}'\mathbf{u}_i\|^2 \leq 3\|\mathbf{V}_n^{-1}\|^2(\|\mathbf{I}_{i1}\|^2 + \|\mathbf{I}_{i2}\|^2 + \|\mathbf{I}_{i3}\|^2)$. Hence, by (S3.17), (S3.21) and (S3.30), we have $\|\breve{\mathbf{u}}_i(\mathbf{\Theta}_n) - \mathbf{W}'\mathbf{u}_i\|^2 = O_p\Big\{K_n^{2c_0+5} n/N + K_n^{2c_0+3} k_n^2 n/N + K_n^{2c_0+4}/n + K_n^{2c_0+4} k_n^{-2r}\Big\}$ uniformly in $\mathbf{\Theta}_n \in \mathbf{\Theta}_n^*$.

Lemma S.6. *Under Conditions in Lemma S.5, we have*

$$\mathbf{W} = \mathbf{I}_{K_n} + O_p\left\{K_n^{c_0+1}\left(K_n^2\sqrt{\frac{n}{N}} + K_n k_n \sqrt{\frac{n}{N}} + \frac{K_n^{3/2}}{\sqrt{n}} + K_n^{3/2} k_n^{-r}\right)\right\},$$

uniformly for $\mathbf{\Theta}_n \in \mathbf{\Theta}_n^*$.

Proof: We start with an identity on $\mathbf{W}$,

$$(\mathbf{P}\breve{\mathbf{U}}(\mathbf{\Theta}_n))'\mathbf{P}\mathbf{U}/N = (\mathbf{P}\mathbf{U}\mathbf{W})'\mathbf{P}\mathbf{U}/N + (\mathbf{P}\breve{\mathbf{U}}(\mathbf{\Theta}_n) - \mathbf{P}\mathbf{U}\mathbf{W})'\mathbf{P}\mathbf{U}/N, \quad \text{(S3.31)}$$

where the definition of $\mathbf{W}$ see Lemma S.5. Then we take right multiplication of $\mathbf{W}$ on both sides of (S3.31), which yields

$$\begin{aligned}(\mathbf{P}\breve{\mathbf{U}}(\boldsymbol{\Theta}_n))'\mathbf{PUW}/N &= (\mathbf{PUW})'\mathbf{PUW}/N + (\mathbf{P}\breve{\mathbf{U}}(\boldsymbol{\Theta}_n) - \mathbf{PUW})'\mathbf{PUW}/N \\ &= \mathbf{W}'\mathbf{W} + (\mathbf{P}\breve{\mathbf{U}}(\boldsymbol{\Theta}_n) - \mathbf{PUW})'(\mathbf{PUW} - \mathbf{P}\breve{\mathbf{U}}(\boldsymbol{\Theta}_n))/N \\ &\quad + (\mathbf{P}\breve{\mathbf{U}}(\boldsymbol{\Theta}_n) - \mathbf{PUW})'\mathbf{P}\breve{\mathbf{U}}(\boldsymbol{\Theta}_n)/N, \qquad \text{(S3.32)}\end{aligned}$$

where the second equality follows from the fact that $\mathbf{U}'\mathbf{P}^2\mathbf{U}/N = \mathbf{I}_{K_n}$. By Lemma S.5, we have the second term is dominated by the third term in (S3.32) . Moreover, by Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}&(\mathbf{P}\breve{\mathbf{U}}(\boldsymbol{\Theta}_n) - \mathbf{PUW})'\mathbf{P}\breve{\mathbf{U}}(\boldsymbol{\Theta}_n)/N \qquad \text{(S3.33)}\\ &= O_p\left\{\sqrt{K_n^{2c_0+6}n/N + K_n^{2c_0+4}k_n^2 n/N + K_n^{2c_0+5}/n + K_n^{2c_0+5}k_n^{-2r}}\right\}.\end{aligned}$$

Combing (S3.32) and (S3.33) we get

$$\begin{aligned}&(\mathbf{P}\breve{\mathbf{U}}(\boldsymbol{\Theta}_n))'\mathbf{PUW}/N \qquad \text{(S3.34)}\\ &= \mathbf{W}'\mathbf{W} + O_p\left\{\sqrt{K_n^{2c_0+6}n/N + K_n^{2c_0+4}k_n^2 n/N + K_n^{2c_0+5}/n + K_n^{2c_0+5}k_n^{-2r}}\right\}.\end{aligned}$$

Note that the left hand side of (S3.34) can be expressed as

$$\begin{aligned}&(\mathbf{P}\breve{\mathbf{U}}(\boldsymbol{\Theta}_n))'\mathbf{PUW}/N \qquad \text{(S3.35)}\\ &= (\mathbf{P}\breve{\mathbf{U}}(\boldsymbol{\Theta}_n))'(\mathbf{PUW} - \mathbf{P}\breve{\mathbf{U}}(\boldsymbol{\Theta}_n))/N + (\mathbf{P}\breve{\mathbf{U}}(\boldsymbol{\Theta}_n))'\mathbf{P}\breve{\mathbf{U}}(\boldsymbol{\Theta}_n))/N \\ &= \mathbf{I}_{K_n} + O_p\left\{\sqrt{K_n^{2c_0+6}n/N + K_n^{2c_0+4}k_n^2 n/N + K_n^{2c_0+5}/n + K_n^{2c_0+5}k_n^{-2r}}\right\},\end{aligned}$$

where the second equality follows from (S3.33) and the fact that $(\mathbf{P}\breve{\mathbf{U}}(\mathbf{\Theta}_n))'\mathbf{P}\breve{\mathbf{U}}(\mathbf{\Theta}_n))/N = \mathbf{I}_{K_n}$. Then (S3.34) and (S3.35) imply that

$$\mathbf{I}_{K_n} = \mathbf{W}'\mathbf{W} + O_p\left\{\sqrt{K_n^{2c_0+6}n/N + K_n^{2c_0+4}k_n^2 n/N + K_n^{2c_0+5}/n + K_n^{2c_0+5}k_n^{-2r}}\right\}. \tag{S3.36}$$

So far, we proved that $\mathbf{W}$ is an orthogonal matrix in probability uniformly for $\mathbf{\Theta}_n \in \mathbf{\Theta}_n^*$, and its eigenvalues are either 1 or -1. We further show that $\mathbf{W}$ is a diagonal matrix.

By the definition of $\mathbf{W}$ and Lemma S.3, we have

$$\mathbf{\Gamma}_0\mathbf{\Gamma}_0'\mathbf{U}'\mathbf{P}^2\breve{\mathbf{U}}(\mathbf{\Theta}_n)/N = \mathbf{W}\mathbf{V}_n + O_p(K_n^{3/2}k_n^{-r}). \tag{S3.37}$$

At the same time, by (S3.31), we obtain

$$\begin{aligned} &\mathbf{U}'\mathbf{P}^2\breve{\mathbf{U}}(\mathbf{\Theta}_n)/N \qquad \text{(S3.38)}\\ &= \mathbf{W} + O_p\left\{\sqrt{K_n^{2c_0+6}n/N + K_n^{2c_0+4}k_n^2 n/N + K_n^{2c_0+5}/n + K_n^{2c_0+5}k_n^{-2r}}\right\}. \end{aligned}$$

Substituting (S3.38) into (S3.37), we get

$$(\mathbf{\Gamma}_0\mathbf{\Gamma}_0')\mathbf{W} = \mathbf{W}\mathbf{V}_n + O_p\left\{\sqrt{K_n^{2c_0+6}n/N + K_n^{2c_0+4}k_n^2 n/N + K_n^{2c_0+5}/n + K_n^{2c_0+5}k_n^{-2r}}\right\},$$

which implies that $\mathbf{W}$ is the eigenvector matrix of $\mathbf{\Gamma}_0\mathbf{\Gamma}_0'$ in probability, uniformly for $\mathbf{\Theta}_n \in \mathbf{\Theta}_n^*$. By Condition (C1'), $\mathbf{\Gamma}_0\mathbf{\Gamma}_0'$ is a diagonal matrix with distinct eigenvalues, then each eigenvalue is associated with a unique eigenvector and each eigenvector has a single nonzero eigenvalue. This implies that $\mathbf{W}$ is a diagonal matrix up to an order $O_p\left\{\sqrt{K_n^{2c_0+6}n/N + K_n^{2c_0+4}k_n^2 n/N + K_n^{2c_0+5}/n + K_n^{2c_0+5}k_n^{-2r}}\right\}$, and we knew that the eigenvalues of $\mathbf{W}$ are either 1 or -1 by (S3.36). With loss of generality, assume all eigenvalues of $\mathbf{W}$ are 1. Thus,

$$\mathbf{W} = \mathbf{I}_{K_n} + O_p\left\{K_n^{c_0+1}(K_n^2\sqrt{\frac{n}{N}} + K_n k_n\sqrt{\frac{n}{N}} + \frac{K_n^{3/2}}{\sqrt{n}} + K_n^{3/2}k_n^{-r})\right\},$$

uniformly in $\mathbf{\Theta}_n \in \mathbf{\Theta}_n^*$, which completes the proof.

Lemma S.7. *Under Conditions in Lemma S.5, we have that*

$$\|\breve{\mathbf{u}}_i(\mathbf{\Theta}_n) - \mathbf{u}_i\| = O_p\left\{K_n^{c_0+3/2}(K_n^2\sqrt{\frac{n}{N}} + K_n k_n\sqrt{\frac{n}{N}} + \frac{K_n^{3/2}}{\sqrt{n}} + K_n^{3/2}k_n^{-r})\right\},$$

uniformly for $\mathbf{\Theta}_n \in \mathbf{\Theta}_n^*$.

Proof: Using the triangle inequality, Lemma S.5 and Lemma S.6, we have

$$\begin{aligned}\|\breve{\mathbf{u}}_i(\mathbf{\Theta}_n) - \mathbf{u}_i\| &\leq \|\breve{\mathbf{u}}_i(\mathbf{\Theta}_n) - \mathbf{W}'\mathbf{u}_i\| + \|\mathbf{u}_i - \mathbf{W}'\mathbf{u}_i\| \\ &= O_p\left\{K_n^{c_0+3/2}(K_n^2\sqrt{\frac{n}{N}} + K_n k_n\sqrt{\frac{n}{N}} + \frac{K_n^{3/2}}{\sqrt{n}} + K_n^{3/2}k_n^{-r})\right\},\end{aligned}$$

uniformly for $\mathbf{\Theta}_n \in \mathbf{\Theta}_n^*$.

Lemma S.8. *Under Conditions in Lemma S.5, the covering number of* $\breve{\mathcal{L}}_n \hat{=} \{\ell(\mathbf{\Theta}_n; \breve{\mathbf{u}}_i(\mathbf{\Theta}_n)) : \mathbf{\Theta}_n \in \mathbf{\Theta}_n^*\}$ *satisfies*

$$N(\epsilon, \breve{\mathcal{L}}_n, L_1(P_n)) \leq (Md + M)^{(d+1)\widetilde{k}_n}(\epsilon/10)^{-a_1}.$$

Proof: By Lemma S.7 and Condition (A2), we have

$$\begin{aligned}P_n\ell(\mathbf{\Theta}; \breve{\mathbf{u}}(\mathbf{\Theta})) &= P_n\ell(\mathbf{\Theta}; \breve{\mathbf{u}}(\mathbf{\Theta})) - P_n\ell(\mathbf{\Theta}; \mathbf{u}(\mathbf{\Theta})) + P_n\ell(\mathbf{\Theta}; \mathbf{u}(\mathbf{\Theta})) \\ &= P_n\ell(\mathbf{\Theta}; \mathbf{u}(\mathbf{\Theta})) + o_p(1). \qquad (S3.39)\end{aligned}$$

Then by (S3.39) and the definition of covering number,

$$N(\epsilon, \breve{\mathcal{L}}_n, L_1(P_n)) = N(\epsilon, \mathcal{L}_{0n}, L_1(P_n)) + o_p(1), \tag{S3.40}$$

where $\mathcal{L}_{0n} = \{\ell(\mathbf{\Theta}_n; \mathbf{u}_i(\mathbf{\Theta}_n)) : \mathbf{\Theta}_n \in \mathbf{\Theta}_n^*\}$. Thus, in the following, we only need to prove

$$N(\epsilon, \mathcal{L}_{0n}, L_1(P_n)) \leq (Md + M)^{(d+1)\widetilde{k}_n} (\epsilon/10)^{-a_1}.$$

Note that for any $\mathbf{\Theta}^{(1)} = (\overrightarrow{\mathbf{H}}^{(1)}, g^{(1)}, \boldsymbol{\psi}^{(1)}), \mathbf{\Theta}^{(2)} = (\overrightarrow{\mathbf{H}}^{(2)}, g^{(2)}, \boldsymbol{\psi}^{(2)}) \in \mathbf{\Theta}_n^*$, using the Taylors series expansion and Condition (A2), we have

$$\begin{aligned} |\ell(\mathbf{\Theta}^{(1)}; \mathbf{u}_i(\mathbf{\Theta}^{(1)})) - \ell(\mathbf{\Theta}^{(2)}; \mathbf{u}_i(\mathbf{\Theta}^{(2)}))| &\leq M(\|\overrightarrow{\mathbf{H}}^{(1)} - \overrightarrow{\mathbf{H}}^{(2)}\| \\ &+ \|g^{(1)} - g^{(2)}\|_\infty + \sum_{j=1}^d \|\psi_j^{(1)} - \psi_j^{(2)}\|_\infty), \end{aligned} \tag{S3.41}$$

where M is a positive constant. Let $\boldsymbol{\delta}^{(k)} = (\delta_1^{(k)}, \cdots, \delta_{\widetilde{k}_n}^{(k)})$ and $\boldsymbol{\vartheta}_j^{(k)} = (\vartheta_{j1}^{(k)}, \cdots, \vartheta_{j\widetilde{k}_n}^{(k)})$ be the spline coefficients of $g^{(k)}$ and $\psi_j^{(k)}, k = 1, 2, j = 1, \cdots, d$, respectively. Then we can show that

$$\|g^{(1)} - g^{(2)}\|_\infty \leq M \max_{1 \leq i \leq \widetilde{k}_n} |\delta_i^{(1)} - \delta_i^{(2)}| := M\|\boldsymbol{\delta}^{(1)} - \boldsymbol{\delta}^{(2)}\|_\infty.$$

Similarly, we have $\|\psi_j^{(1)} - \psi_j^{(2)}\|_\infty \leq M\|\boldsymbol{\vartheta}_j^{(1)} - \boldsymbol{\vartheta}_j^{(2)}\|_\infty$. Therefore, (S3.41) reduces to

$$\begin{aligned} |\ell(\mathbf{\Theta}^{(1)}; \mathbf{u}_i(\mathbf{\Theta}^{(1)})) - \ell(\mathbf{\Theta}^{(2)}; \mathbf{u}_i(\mathbf{\Theta}^{(2)}))| &\leq M(\|\overrightarrow{\mathbf{H}}^{(1)} - \overrightarrow{\mathbf{H}}^{(2)}\| \\ &+ \|\boldsymbol{\delta}^{(1)} - \boldsymbol{\delta}^{(2)}\|_\infty + \sum_{j=1}^d \|\boldsymbol{\vartheta}_j^{(1)} - \boldsymbol{\vartheta}_j^{(2)}\|_\infty). \end{aligned} \tag{S3.42}$$

It follows that for any $\mathbf{\Theta}_n \in \mathbf{\Theta}_n^*$, there exist $\mathbf{\Theta}_{n,k}$ with $\mathbf{\Theta}_{n,k} = (\overrightarrow{\mathbf{H}}^{(k)}, g^{(k)}, \boldsymbol{\psi}^{(k)})$, such that

$$\frac{1}{n}\sum_{i=1}^{n} |\ell(\mathbf{\Theta}_n; \mathbf{u}_i(\mathbf{\Theta}_n)) - \ell(\mathbf{\Theta}_{n,k}; \mathbf{u}_i(\mathbf{\Theta}_{n,k}))| \leq M(\|\overrightarrow{\mathbf{H}} - \overrightarrow{\mathbf{H}}^{(k)}\|$$
$$+\|\boldsymbol{\delta} - \boldsymbol{\delta}^{(k)}\|_\infty + \sum_{j=1}^{d} \|\boldsymbol{\vartheta}_j - \boldsymbol{\vartheta}_j^{(k)}\|_\infty).$$

By Lemma S.2 and following the calculation in Lemma 1 of Ma et al. (2015), we have

$$N(\epsilon, \mathcal{L}_{0n}, L_1(P_n)) \leq (Md + M)^{(d+1)\widetilde{k}_n} (\epsilon/10)^{-a_1}. \tag{S3.43}$$

Therefore, combine (S3.40) and (S3.43) completes the proof.

Lemma S.9. *Under Conditions in Lemma S.5, we have*

$$\sup_{\mathbf{\Theta}_n \in \mathbf{\Theta}_n^*} |P_n \ell(\mathbf{\Theta}_n; \check{\mathbf{u}}(\mathbf{\Theta}_n)) - P\ell(\mathbf{\Theta}_n; \mathbf{u}(\mathbf{\Theta}_n))| \to 0 \textit{ almost surely,}$$

Proof: By the triangular inequality, it holds that

$$\begin{aligned}\sup_{\mathbf{\Theta}_n \in \mathbf{\Theta}_n^*} |P_n \ell(\mathbf{\Theta}_n; \check{\mathbf{u}}(\mathbf{\Theta}_n)) - P\ell(\mathbf{\Theta}_n; \mathbf{u}(\mathbf{\Theta}_n))| &\leq \sup_{\mathbf{\Theta}_n \in \mathbf{\Theta}_n^*} |P_n \{\ell(\mathbf{\Theta}_n; \check{\mathbf{u}}(\mathbf{\Theta}_n)) - \ell(\mathbf{\Theta}_n; \mathbf{u}(\mathbf{\Theta}_n))\}| \\ &+ \sup_{\mathbf{\Theta}_n \in \mathbf{\Theta}_n^*} |P_n \ell(\mathbf{\Theta}_n; \mathbf{u}(\mathbf{\Theta}_n)) - P\ell(\mathbf{\Theta}_n; \mathbf{u}(\mathbf{\Theta}_n))|. \end{aligned} \tag{S3.44}$$

By Lemma S.7, we have

$$\sup_{\mathbf{\Theta}_n \in \mathbf{\Theta}_n^*} |P_n \{\ell(\mathbf{\Theta}_n; \check{\mathbf{u}}(\mathbf{\Theta}_n)) - \ell(\mathbf{\Theta}_n; \mathbf{u}(\mathbf{\Theta}_n))\}| \to 0 \text{ almost surely.} \tag{S3.45}$$

Thus, in the following, we only need to prove

$$\sup_{\mathbf{\Theta}_n \in \mathbf{\Theta}_n^*} |P_n \ell(\mathbf{\Theta}_n; \mathbf{u}(\mathbf{\Theta}_n)) - P\ell(\mathbf{\Theta}_n; \mathbf{u}(\mathbf{\Theta}_n))| \to 0 \text{ almost surely.} \tag{S3.46}$$

By Condition (A5), let $\nu/2 < \phi < 1/2$ and $\alpha_n = n^{-1/2+\phi}(\log n)^{1/2}$. Here, $\{\alpha_n\}$ is a non-increasing positive sequence. And for the given $\epsilon > 0$ in Lemma S.2, let $\epsilon_n = \epsilon\alpha_n$. Note that $P\ell^2(\mathbf{\Theta}_n; \mathbf{u}(\mathbf{\Theta}_n))$ is bounded under conditions (A1)-(A3) and $\mathrm{var}(Y_i|\mathbf{Z}_i) < \infty$. Then for any $\mathbf{\Theta}_n \in \mathbf{\Theta}_n^*$ and sufficiently large n, we have

$$\frac{\mathrm{var}(P_n\ell(\mathbf{\Theta}_n; \mathbf{u}(\mathbf{\Theta}_n)))}{(4\epsilon_n)^2} \le \frac{(1/n)P\ell^2(\mathbf{\Theta}_n; \mathbf{u}(\mathbf{\Theta}_n))}{16\epsilon^2\alpha_n^2} \le \frac{M}{16n\epsilon^2\alpha_n^2} \le \frac{1}{16\epsilon^2 n^{2\phi}\log n} \le \frac{1}{2},$$

where M is a finite constant independent of n. Furthermore, by the inequality (31) and Lemma 33 of Pollard (1984) and Lemma S.8 above, we have

$$\begin{aligned}
&P\{\sup_{\breve{\mathcal{L}}_n} |P_n\ell(\mathbf{\Theta}_n; \mathbf{u}(\mathbf{\Theta}_n)) - P\ell(\mathbf{\Theta}_n; \mathbf{u}(\mathbf{\Theta}_n))| > 8\epsilon_n\} \\
&\le 8N(\epsilon_n, \breve{\mathcal{L}}_n, L_1(P_n)) \exp(-n\epsilon_n^2/128) P\{\sup_{\breve{\mathcal{L}}_n} |P_n\ell^2(\mathbf{\Theta}_n; \mathbf{u}(\mathbf{\Theta}_n))| \le 64\} \\
&+P\{\sup_{\breve{\mathcal{L}}_n} |P_n\ell^2(\mathbf{\Theta}_n; \mathbf{u}(\mathbf{\Theta}_n))| > 64\} \\
&\le (Md + M)^{(d+1)\widetilde{k}_n} (\epsilon_n/10)^{-a_1} \exp(-n\epsilon_n^2/128) \\
&= \exp[(a_1 - dK_n)\log(Md + M) - a_1\log\{\epsilon n^{-1/2+\phi}(\log n)^{1/2}/10\} - n\epsilon^2 n^{-1+2\phi}\log n/128] \\
&\le \exp[a_1\{\log(Md + M) + (1/2 - \phi)\log n - \log\log n/2 - \log(\epsilon/10)\} - \epsilon^2 n^{2\phi}\log n/128] \\
&\le \exp\left(-c_1 n^{2\phi}\log n\right),
\end{aligned}$$

where c_1 is a finite positive constant, and the last inequality comes from $\nu/2 < \phi < 1/2$. Hence $\sum_{n=1}^{\infty} P\{\sup_{\breve{\mathcal{L}}_n} |P_n\ell(\mathbf{\Theta}_n; \mathbf{u}(\mathbf{\Theta}_n)) - P\ell(\mathbf{\Theta}_n; \mathbf{u}(\mathbf{\Theta}_n))| > 8\epsilon_n\} < \infty$. By the Borel-Cantelli Lemma, (S3.46) holds. Combing (S3.44), (S3.45) and (S3.46), we finish the proof.

S4. Proof of Lemma 1

The proof of Lemma 1 is divided into two steps. The first step proves the consistency of $\breve{\mathbf{\Theta}}_n$ and $\breve{\mathbf{u}}$. The second step establishes the convergence rates.

Step 1: consistency

Under Condition (A1) and by the Corollary 6.21 of Schumacker (1981), there exist $g_{n0}(x) = \boldsymbol{\delta}_0' \mathbf{B}_n(x)$ and $\psi_{nj0}(x) = \boldsymbol{\vartheta}_{j0}' \mathbf{B}_n(x)$ for $j = 1, \cdots, d$, such that

$$\sup_{x\in[0,1]} |g_{n0}(x) - g_0(x)| = O_p(\widetilde{k}_n^{-r}), \sup_{x\in[0,1]} |\psi_{nj0}(x) - \psi_{j0}(x)| = O_p(\widetilde{k}_n^{-r}),$$

where $g_0(x)$ and $\psi_{j0}(x)$ denote the true functions of $g(x)$ and $\psi_j(x)$, respectively, $j = 1, \cdots, d$. Let $\mathbf{\Theta}_{n0} = (\overrightarrow{\mathbf{H}}_0', g_{n0}, \boldsymbol{\psi}_{n0}')'$, $\mathbf{\Theta}_0 = (\overrightarrow{\mathbf{H}}_0', g_0, \boldsymbol{\psi}_0')'$. Then by Condition (A5),

$$d(\mathbf{\Theta}_{n0}, \mathbf{\Theta}_0) = O_p(n^{-r\widetilde{\nu}}). \tag{S4.47}$$

Let $M(\mathbf{\Theta}_n; \breve{\mathbf{u}}_i(\mathbf{\Theta}_n)) = -\ell(\mathbf{\Theta}_n; \breve{\mathbf{u}}_i(\mathbf{\Theta}_n))$, $K_\varsigma = \{\mathbf{\Theta}_n : d(\mathbf{\Theta}_n, \mathbf{\Theta}_{n0}) \geq \varsigma, \mathbf{\Theta}_n \in \mathbf{\Theta}_n^*\}$ for $\varsigma > 0$, $\zeta_{1n} = \sup_{\mathbf{\Theta}_n \in \mathbf{\Theta}_n^*} |P_n M(\mathbf{\Theta}_n; \breve{\mathbf{u}}(\mathbf{\Theta}_n)) - PM(\mathbf{\Theta}_n; \breve{\mathbf{u}}(\mathbf{\Theta}_n))|$ and $\zeta_{2n} = P_n M(\mathbf{\Theta}_{n0}; \breve{\mathbf{u}}(\mathbf{\Theta}_{n0})) - PM(\mathbf{\Theta}_{n0}; \breve{\mathbf{u}}(\mathbf{\Theta}_{n0}))$. Then, we have

$$\begin{aligned}\inf_{K_\varsigma} PM(\mathbf{\Theta}_n; \breve{\mathbf{u}}(\mathbf{\Theta}_n)) &= \inf_{K_\varsigma} \{PM(\mathbf{\Theta}_n; \breve{\mathbf{u}}(\mathbf{\Theta}_n)) - P_n M(\mathbf{\Theta}_n; \breve{\mathbf{u}}(\mathbf{\Theta}_n)) + P_n M(\mathbf{\Theta}_n; \breve{\mathbf{u}}(\mathbf{\Theta}_n))\} \\ &\leq \zeta_{1n} + \inf_{K_\varsigma} P_n M(\mathbf{\Theta}_n; \breve{\mathbf{u}}(\mathbf{\Theta}_n)). \end{aligned} \tag{S4.48}$$

If $\breve{\mathbf{\Theta}}_n \in K_\varsigma$, one can show that

$$\begin{aligned}\inf_{K_\varsigma} P_n M(\mathbf{\Theta}_n; \breve{\mathbf{u}}(\mathbf{\Theta}_n)) &= P_n M(\breve{\mathbf{\Theta}}_n; \breve{\mathbf{u}}(\breve{\mathbf{\Theta}}_n)) \\ &\leq P_n M(\mathbf{\Theta}_{n0}; \breve{\mathbf{u}}(\mathbf{\Theta}_{n0})) \\ &= \zeta_{2n} + PM(\mathbf{\Theta}_{n0}; \breve{\mathbf{u}}(\mathbf{\Theta}_{n0})). \qquad (S4.49)\end{aligned}$$

By (S4.48) and (S4.49), we have

$$\inf_{K_\varsigma} PM(\mathbf{\Theta}_n; \breve{\mathbf{u}}(\mathbf{\Theta}_n)) \leq \zeta_{1n} + \zeta_{2n} + PM(\mathbf{\Theta}_{n0}; \breve{\mathbf{u}}(\mathbf{\Theta}_{n0})) = \zeta_n + PM(\mathbf{\Theta}_{n0}; \breve{\mathbf{u}}(\mathbf{\Theta}_{n0}))$$

with $\zeta_n = \zeta_{1n} + \zeta_{2n}$. It is clear that $\zeta_n \geq \delta_\varsigma \widehat{=} \inf_{K_\varsigma} PM(\mathbf{\Theta}_n; \breve{\mathbf{u}}(\mathbf{\Theta}_n)) - PM(\mathbf{\Theta}_{n0}; \breve{\mathbf{u}}(\mathbf{\Theta}_{n0}))$, which is larger than zero when n is large enough. Hence,

$$\{\breve{\mathbf{\Theta}}_n \in K_\varsigma\} \subseteq \{\zeta_n \geq \delta_\varsigma\}. \qquad (S4.50)$$

By Lemma S.9 and the strong law of large numbers, we know $\zeta_{1n} \to 0$ and $\zeta_{2n} \to 0$ then $\zeta_n \to 0$ almost surely. Therefore, when n is large enough $\bigcup_{i=1}^{\infty} \bigcap_{n=i}^{\infty} \{\zeta_n \geq \delta_\varsigma\}$ is a null set. By (S4.50), we have $\bigcup_{i=1}^{\infty} \bigcap_{n=i}^{\infty} \{\breve{\mathbf{\Theta}}_n \in K_\varsigma\}$ is null set. Coupling with the definition of K_ς, we have

$$d(\breve{\mathbf{\Theta}}_n, \mathbf{\Theta}_{n0}) \to 0, \qquad (S4.51)$$

almost surely as $n \to \infty$. This together with (S4.47) yields $d(\breve{\mathbf{\Theta}}_n, \mathbf{\Theta}_0) \to 0$ in probability. Recall that $\breve{\mathbf{u}}_i \equiv \breve{\mathbf{u}}_i(\breve{\mathbf{\Theta}}_n)$, which with Lemma S.7, implies $\breve{\mathbf{u}}_i \to \mathbf{u}_i$ in probability.

Step 2: Convergence rate

To establish the convergence rate, for any $\eta > 0$, we define the class of functions $\mathcal{F}_\eta = \{\ell(\mathbf{\Theta}_n; \breve{\mathbf{u}}_i(\mathbf{\Theta}_n)) - \ell(\mathbf{\Theta}_{n0}; \breve{\mathbf{u}}_i(\mathbf{\Theta}_{n0})) : \mathbf{\Theta}_n \in \mathbf{\Theta}_n^*, \eta/2 \leq d(\mathbf{\Theta}_n, \mathbf{\Theta}_{n0}) \leq \eta\}$. By Lemma S.2, we know the complexity of space $\mathbf{\Theta}_n^*$ can be well controlled, which induces the control on the complexity of s-pace $\mathcal{F}_\eta$ by following the idea of the calculations of Shen and Wong (pp. 597; 1994). For $0 < \epsilon < \eta$, we can establish that $\log N_{[\,]}(\epsilon, \mathcal{F}_\eta, L_2(P)) \leq M(\widetilde{k}_n + K_n)\log(\eta/\epsilon)$. Moreover, for large n, we have $P(\ell(\mathbf{\Theta}_n; \breve{\mathbf{u}}(\mathbf{\Theta}_n)) - \ell(\mathbf{\Theta}_{n0}; \breve{\mathbf{u}}(\mathbf{\Theta}_{n0}))) \leq M\eta^2$, for any $\ell(\mathbf{\Theta}_n; \breve{\mathbf{u}}_i(\mathbf{\Theta}_n)) - \ell(\mathbf{\Theta}_{n0}; \breve{\mathbf{u}}_i(\mathbf{\Theta}_{n0})) \in \mathcal{F}_\eta$. Therefore, by Lemma 3.4.2 of Van der Vaart and Wellner (1996), we have

$$E_P\|n^{1/2}(P_n - P)\|_{\mathcal{F}_\eta} \leq MJ_{[\,]}(\eta, \mathcal{F}_\eta, L_2(P))\left\{1 + \frac{J_{[\,]}(\eta, \mathcal{F}_\eta, L_2(P))}{\eta^2\sqrt{n}}\right\}, \tag{S4.52}$$

where $J_{[\,]}(\eta, \mathcal{F}_\eta, L_2(P)) = \int_0^\eta \{1 + \log N_{[\,]}(\epsilon, \mathcal{F}_\eta, L_2(P))\}^{1/2} d\epsilon \leq M\sqrt{\widetilde{k}_n + K_n}\eta$. The right hand of (S4.52) yields that the key function in Theorem 3.2.5 of Van der Vaart and Wellner (1996) is given by $\phi_n(\eta) = \sqrt{\widetilde{k}_n + K_n}\eta + (\widetilde{k}_n + K_n)/\sqrt{n}$. Note that $\phi_n(\eta)/\eta$ is decreasing in η, and $r_n^2\phi_n(1/r_n) = r_n\sqrt{\widetilde{k}_n + K_n} + r_n^2(\widetilde{k}_n + K_n)/n^{1/2} \leq Mn^{1/2}$,where $r_n = (\widetilde{k}_n + K_n)^{-1/2}n^{1/2}$. Hence, by applying Theorem 3.4.1 of Van der Vaart and Wellner (1996), we have $d(\breve{\mathbf{\Theta}}_n, \mathbf{\Theta}_{n0}) = O_p\{\sqrt{\frac{\widetilde{k}_n + K_n}{n}} + K_n^{c_0+3/2}(K_n^2\sqrt{\frac{n}{N}} + K_nk_n\sqrt{\frac{n}{N}} + \frac{K_n^{3/2}}{\sqrt{n}} + K_n^{3/2}k_n^{-r})\}$. This with (S4.47) yields $d(\breve{\mathbf{\Theta}}_n, \mathbf{\Theta}_0) = O_p\{\sqrt{\frac{\widetilde{k}_n + K_n}{n}} + \widetilde{k}_n^{-r} + K_n^{c_0+3/2}(K_n^2\sqrt{\frac{n}{N}} + K_nk_n\sqrt{\frac{n}{N}} + \frac{K_n^{3/2}}{\sqrt{n}} + K_n^{3/2}k_n^{-r})\}$, together with Lemma S.7

completes the proof.

S5. Proof of Theorem 1

Proof of (i) in Theorem 1.

We prove (i) in two steps. First, we derive the convergence rates as follows.

$$\|\widehat{\mathbf{u}}_i - \mathbf{u}_i\| = O_p\{K_n^{c_0+3/2}(K_n^2\sqrt{\frac{n}{N}} + K_n k_n\sqrt{\frac{n}{N}} + \frac{K_n^{3/2}}{\sqrt{n}} + K_n^{3/2}k_n^{-r})\}, i = 1, \ldots, n, \tag{S5.53}$$

$$d(\widehat{\mathbf{\Theta}}_n, \mathbf{\Theta}_0) = O_p\{\sqrt{\frac{\widetilde{k}_n + K_n}{n}} + \widetilde{k}_n^{-r} + K_n^{c_0+3/2}(K_n^2\sqrt{\frac{n}{N}} + K_n k_n\sqrt{\frac{n}{N}} + \frac{K_n^{3/2}}{\sqrt{n}} + K_n^{3/2}k_n^{-r})\} \,. \tag{S5.54}$$

Under the conditions of Theorem 1, the proof of (S5.53) is the same as that of Lemma 1.

The proof of (S5.54) is similar to that of Lemma 1 with slightly different manipulations as follows. Let $M(\mathbf{\Theta}_n; \widehat{\mathbf{u}}_i(\mathbf{\Theta}_n)) = -\ell(\mathbf{\Theta}_n; \widehat{\mathbf{u}}_i(\mathbf{\Theta}_n)) + \sum_{k=1}^{K_n} p_\lambda(\|\mathbf{H}_{\cdot k}\|)$ and $\delta_\varsigma \widehat{=} \inf_{K_\varsigma} PM(\mathbf{\Theta}_n; \widehat{\mathbf{u}}(\mathbf{\Theta}_n)) - PM(\mathbf{\Theta}_{n0}; \widehat{\mathbf{u}}(\mathbf{\Theta}_{n0}))$. δ_ς is larger than

$$\begin{aligned} &\inf_{K_\varsigma} -P\ell(\mathbf{\Theta}_n; \widehat{\mathbf{u}}(\mathbf{\Theta}_n)) + \sum_{k=1}^{K_n} p_\lambda(\|\mathbf{H}_{\cdot k}\|) + P\ell(\mathbf{\Theta}_{n0}; \widehat{\mathbf{u}}(\mathbf{\Theta}_{n0})) - \sum_{k=1}^{K_n} p_\lambda(\|\mathbf{H}_{0,\cdot k}\|) \\ \geq\ &\inf_{K_\varsigma} -P\ell(\mathbf{\Theta}_n; \widehat{\mathbf{u}}(\mathbf{\Theta}_n)) + P\ell(\mathbf{\Theta}_{n0}; \widehat{\mathbf{u}}(\mathbf{\Theta}_{n0})) + \inf_{K_\varsigma} \sum_{k=1}^{K_n} p_\lambda(\|\mathbf{H}_{\cdot k}\|) - \sum_{k=1}^{K_n} p_\lambda(\|\mathbf{H}_{0,\cdot k}\|). \end{aligned}$$

Analogous to the proof of Lemma 1, we know that

$$\inf_{K_\varsigma} -P\ell(\mathbf{\Theta}_n; \widehat{\mathbf{u}}(\mathbf{\Theta}_n)) + P\ell(\mathbf{\Theta}_{n0}; \widehat{\mathbf{u}}(\mathbf{\Theta}_{n0})) > 0$$

$$\text{and} \qquad \inf_{K_\varsigma} \sum_{k=1}^{K_n} p_\lambda(\|\mathbf{H}_{\cdot k}\|) \geq 0.$$

In addition, under Condition (A7) that $\lambda = o(\inf_{1\leq k\leq s_0} \|\mathbf{H}_{0,\cdot k}\|)$ and using the explicit expression of the SCAD penalty, we can show $\sum_{k=1}^{K_n} p_\lambda(\|\mathbf{H}_{0,\cdot k}\|) = (a+1)\lambda^2 s_0/2 = o_p(1)$. Thus $\delta_\varsigma > 0$ when n is large enough. Following the lines of the proof of Lemma 1, we get the desired convergence rates.

Next, we demonstrate selection consistency. Suppose $\widehat{\mathbf{H}}_{\cdot k} \neq \mathbf{0}$ for some $k \geq s_0$. Let $\widehat{\mathbf{H}}^*$ be the same as $\widehat{\mathbf{H}}$ except that its kth column is replaced by the true parameter value 0 and $\mathbf{\Theta}_n^* = (\overrightarrow{\widehat{\mathbf{H}}}^{*\prime}, \widehat{g}, \widehat{\boldsymbol{\psi}}')'$. Then by Lemma S.7 and a Taylor expansion, we have

$$\begin{aligned} P_n M(\widehat{\mathbf{\Theta}}_n; \widehat{\mathbf{u}}(\widehat{\mathbf{\Theta}}_n)) \;-\; & P_n M(\mathbf{\Theta}_n^*; \widehat{\mathbf{u}}(\mathbf{\Theta}_n^*)) = P_n M(\widehat{\mathbf{\Theta}}_n; \mathbf{u}(\widehat{\mathbf{\Theta}}_n)) - P_n M(\mathbf{\Theta}_n^*; \mathbf{u}(\mathbf{\Theta}_n^*)) + o_p(1) \\ =\; & -(P_n \partial\ell(\widetilde{\mathbf{\Theta}}_n^*; \mathbf{u}(\widetilde{\mathbf{\Theta}}_n^*))/\partial \mathbf{H}_{\cdot k})'\widehat{\mathbf{H}}_{\cdot k} + p_\lambda(\|\widehat{\mathbf{H}}_{\cdot k}\|) + o_p(1) \\ \geq\; & -\|P_n \partial\ell(\widetilde{\mathbf{\Theta}}_n^*; \mathbf{u}(\widetilde{\mathbf{\Theta}}_n^*))/\partial \mathbf{H}_{\cdot k}\|\|\widehat{\mathbf{H}}_{\cdot k}\| + p_\lambda(\|\widehat{\mathbf{H}}_{\cdot k}\|) + o_p(1), \end{aligned}$$

where $\widetilde{\mathbf{\Theta}}_n^*$ lies between $\widehat{\mathbf{\Theta}}_n$ and $\mathbf{\Theta}_n^*$. Given the convergence rates in (S5.54), we know $\|P_n \partial\ell(\widetilde{\mathbf{\Theta}}_n^*; \mathbf{u}(\widetilde{\mathbf{\Theta}}_n^*))/\partial \mathbf{H}_{\cdot k}\| = O_p\{\sqrt{\frac{\widetilde{k}_n + K_n}{n}} + \widetilde{k}_n^{-r} + K_n^{c_0+3/2}(K_n^2\sqrt{\frac{n}{N}} + K_n k_n \sqrt{\frac{n}{N}} + \frac{K_n^{3/2}}{\sqrt{n}} + K_n^{3/2} k_n^{-r})\}$, $\|\widehat{\mathbf{H}}_{\cdot k}\| = O_p\{\sqrt{\frac{\widetilde{k}_n + K_n}{n}} + \widetilde{k}_n^{-r} + K_n^{c_0+3/2}(K_n^2\sqrt{\frac{n}{N}} + K_n k_n \sqrt{\frac{n}{N}} + \frac{K_n^{3/2}}{\sqrt{n}} + K_n^{3/2} k_n^{-r})\}$ and $p_\lambda(\|\widehat{\mathbf{H}}_{\cdot k}\|) = \lambda\|\widehat{\mathbf{H}}_{\cdot k}\|$ by the definition of the SCAD penalty. Thus, by Condition (A7), $P_n M(\widehat{\mathbf{\Theta}}_n; \widehat{\mathbf{u}}(\widehat{\mathbf{\Theta}}_n)) - P_n M(\mathbf{\Theta}_n^*; \widehat{\mathbf{u}}(\mathbf{\Theta}_n^*))$

is positive, which contradict against $\inf_{\mathbf{\Theta}_n\in\mathbf{\Theta}_n^*} P_nM(\mathbf{\Theta}_n;\widehat{\mathbf{u}}(\mathbf{\Theta}_n)) = P_nM(\widehat{\mathbf{\Theta}}_n;\widehat{\mathbf{u}}(\widehat{\mathbf{\Theta}}_n))$. This completes the proof.

Proof of (ii) in Theorem 1.

Given $\widehat{\mathbf{H}}_2 = \mathbf{0}$ with probability tending to 1, we only need to consider a correctly specified model without regularization. The establishment of convergence rates for $\widehat{\mathbf{u}}_i$ and $(\widehat{\mathbf{H}}_1,\widehat{g},\widehat{\boldsymbol{\psi}})$ follows by (S5.53), (S5.54) and (i) in Theorem 1.

Proof of (iii) in Theorem 1.

Let $\mathbf{\Theta}_1^*$ be the space of $(\mathbf{H}_1, g, \boldsymbol{\psi})$, and $\mathbf{\Theta}_{\backslash 0}^*$ denote $\mathbf{\Theta}_1^*$ excluding $\mathbf{\Theta}_{1,0}$, where $\mathbf{\Theta}_{1,0}$ is the true value of $(\mathbf{H}_1, g, \boldsymbol{\psi})$. By Condition (A8) and the results of (ii), we have $\|\widehat{\mathbf{H}}_1 - \mathbf{H}_{1,0}\| + \|\widehat{g} - g_0\|_2 + \|\widehat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0\|_2 = O_p\{\sqrt{\frac{\widetilde{k}_n+s_0}{n}} + \widetilde{k}_n^{-r}\}$. Write $\delta_n = \sqrt{\frac{\widetilde{k}_n+s_0}{n}} + \widetilde{k}_n^{-r}$. Let $\widetilde{V}$ denote the linear span of $\mathbf{\Theta}_{\backslash 0}^*$ and define the Fisher inner product on the space $\widetilde{V}$ as $<\mathbf{v},\check{\mathbf{v}}>= P\{\dot{\ell}(\mathbf{\Theta}_{1,0};\mathbf{u}_i(\mathbf{\Theta}_{1,0}))[\mathbf{v}]\dot{\ell}(\mathbf{\Theta}_{1,0};\mathbf{u}_i(\mathbf{\Theta}_{1,0}))[\check{\mathbf{v}}]\}$ for $\mathbf{v},\check{\mathbf{v}}\in\widetilde{V}$, the Fisher norm as $\|\mathbf{v}\|^2 =<\mathbf{v},\mathbf{v}>$, where $\dot{\ell}(\mathbf{\Theta}_1;\mathbf{u}_i(\mathbf{\Theta}_1))[\mathbf{v}] = \left.\frac{d\ell(\mathbf{\Theta}_1+s\mathbf{v};\mathbf{u}_i(\mathbf{\Theta}_1+s\mathbf{v}))}{ds}\right|_{s=0}$ is the first order directional derivative of $\ell(\mathbf{\Theta}_1;\mathbf{u}_i(\mathbf{\Theta}_1))$ at the direction $\mathbf{v}\in\widetilde{V}$, and $\mathbf{\Theta}_1\in\{\mathbf{\Theta}_1\in\mathbf{\Theta}_1^* : d(\mathbf{\Theta}_1,\mathbf{\Theta}_{1,0}) = O(\delta_n)\}$.

For a vector of ds_0-dimension $\mathbf{a}_n$ with $\|\mathbf{a}_n\|\leq 1$, define a smooth functional of $\mathbf{\Theta}_1$ and $\mathbf{U}(\mathbf{\Theta}_1)$ as $h(\mathbf{\Theta}_1;\mathbf{U}(\mathbf{\Theta}_1)) = \mathbf{a}_n'\overrightarrow{\mathbf{H}_1}$ and $\dot{h}(\mathbf{\Theta}_{1,0};\mathbf{U}(\mathbf{\Theta}_{1,0}))[\mathbf{v}] = \left.\frac{dh(\mathbf{\Theta}_{1,0}+s\mathbf{v};\mathbf{U}(\mathbf{\Theta}_{1,0}+s\mathbf{v}))}{ds}\right|_{s=0}$, where $\mathbf{U}(\mathbf{\Theta}_1) = (\mathbf{u}_1(\mathbf{\Theta}_1),\cdots,\mathbf{u}_n(\mathbf{\Theta}_1))'$. Note that $h(\mathbf{\Theta}_1;\mathbf{U}(\mathbf{\Theta}_1)) - h(\mathbf{\Theta}_{1,0};\mathbf{U}(\mathbf{\Theta}_{1,0})) = \dot{h}(\mathbf{\Theta}_{1,0};\mathbf{U}(\mathbf{\Theta}_{1,0}))[\mathbf{\Theta}_1 - \mathbf{\Theta}_{1,0}]$. Based on the Riesz representation theorem, for any given $\mathbf{v} \in \widetilde{V}$, there exists $\mathbf{v}^* \in \overline{V}$ such that $\dot{h}(\boldsymbol{\Theta}_{1,0}; \mathbf{U}(\boldsymbol{\Theta}_{1,0}))[\mathbf{v}] =< \mathbf{v}^*, \mathbf{v} >$. Thus according to Cramér-Wold device, in order to prove Theorem 1, it suffices to show that

$$\sqrt{n} < \mathbf{v}^*, \widehat{\boldsymbol{\Theta}}_{1,n} - \boldsymbol{\Theta}_{1,0} > \xrightarrow{d} N(0, \mathbf{a}_n' \boldsymbol{\Sigma}^{-1} \mathbf{a}_n), \tag{S5.55}$$

due to $\mathbf{a}_n'(\overrightarrow{\widehat{\mathbf{H}}_1} - \overrightarrow{\mathbf{H}}_{1,0}) = h(\widehat{\boldsymbol{\Theta}}_{1,n}; \mathbf{U}(\widehat{\boldsymbol{\Theta}}_{1,n})) - h(\boldsymbol{\Theta}_{1,0}; \mathbf{U}(\boldsymbol{\Theta}_{1,0})) = \dot{h}(\boldsymbol{\Theta}_{1,0}; \mathbf{U}(\boldsymbol{\Theta}_{1,0}))[\widehat{\boldsymbol{\Theta}}_{1,n} - \boldsymbol{\Theta}_{1,0}] =< \mathbf{v}^*, \widehat{\boldsymbol{\Theta}}_{1,n} - \boldsymbol{\Theta}_{1,0} >$. In fact, (S5.55) holds when (a) $\sqrt{n} < \mathbf{v}^*, \widehat{\boldsymbol{\Theta}}_{1,n} - \boldsymbol{\Theta}_{1,0} > \xrightarrow{d} N(0, \|\mathbf{v}^*\|^2)$ and (b) $\|\mathbf{v}^*\|^2 = \mathbf{a}_n' \boldsymbol{\Sigma}^{-1} \mathbf{a}_n$.

We prove (S5.55) in two steps. First, we prove (a) $\sqrt{n} < \mathbf{v}^*, \widehat{\boldsymbol{\Theta}}_{1,n} - \boldsymbol{\Theta}_{1,0} > \xrightarrow{d} N(0, \|\mathbf{v}^*\|^2)$. Under the conditions $s_0 = o(n^{1/8})$, $r \geq 2$ and based on Corollary 6.21 of Schumacker (pp 227; 1981), there exists $\Pi_n \mathbf{v}^* \in \boldsymbol{\Theta}_{1,n}^*$ such that $\|\Pi_n \mathbf{v}^* - \mathbf{v}^*\| = o(1)$ and $\delta_n \|\Pi_n \mathbf{v}^* - \mathbf{v}^*\| = o(n^{-1/2})$. For any $\boldsymbol{\Theta}_1 \in \{\boldsymbol{\Theta}_1 \in \boldsymbol{\Theta}_1^* : d(\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_{1,0}) = O(\delta_n)\}$, define

$$\ddot{\ell}(\boldsymbol{\Theta}_1; \mathbf{u}_i(\boldsymbol{\Theta}_1))[\mathbf{v}, \check{\mathbf{v}}] = \left. \frac{d\dot{\ell}(\boldsymbol{\Theta}_1 + \check{s}\check{\mathbf{v}}; \mathbf{u}_i(\boldsymbol{\Theta}_1 + \check{s}\check{\mathbf{v}}))[\mathbf{v}]}{d\check{s}} \right|_{\check{s}=0},$$
$$r_i[\Pi_n \mathbf{v}^*, \boldsymbol{\Theta}_1 - \boldsymbol{\Theta}_{1,0}] = \dot{\ell}(\boldsymbol{\Theta}_1; \mathbf{u}_i(\boldsymbol{\Theta}_1))[\Pi_n \mathbf{v}^*] - \dot{\ell}(\boldsymbol{\Theta}_{1,0}; \mathbf{u}_i(\boldsymbol{\Theta}_{1,0}))[\Pi_n \mathbf{v}^*].$$

Since $P_n \dot{\ell}(\widehat{\boldsymbol{\Theta}}_{1,n}; \widehat{\mathbf{u}}(\widehat{\boldsymbol{\Theta}}_{1,n}))[\Pi_n \mathbf{v}^*] = 0$ by the definition of $\widehat{\boldsymbol{\Theta}}_{1,n}$, by Lemma

S.7, we have

$$\begin{aligned}
o_p(n^{-1/2}) &= P_n\dot{\ell}(\widehat{\mathbf{\Theta}}_{1,n};\mathbf{u}(\widehat{\mathbf{\Theta}}_{1,n}))[\Pi_n\mathbf{v}^*] \qquad \text{(S5.56)}\\
&= P_n\dot{\ell}(\mathbf{\Theta}_{1,0};\mathbf{u}(\mathbf{\Theta}_{1,0}))[\Pi_n\mathbf{v}^*] + P_n r[\Pi_n\mathbf{v}^*, \widehat{\mathbf{\Theta}}_{1,n} - \mathbf{\Theta}_{1,0}]\\
&= P_n\dot{\ell}(\mathbf{\Theta}_{1,0};\mathbf{u}(\mathbf{\Theta}_{1,0}))[\mathbf{v}^*]\\
&\quad + P_n\dot{\ell}(\mathbf{\Theta}_{1,0};\mathbf{u}(\mathbf{\Theta}_{1,0}))[\Pi_n\mathbf{v}^* - \mathbf{v}^*] + (P_n - P)r[\Pi_n\mathbf{v}^*, \widehat{\mathbf{\Theta}}_{1,n} - \mathbf{\Theta}_{1,0}]\\
&\quad + P(r[\Pi_n\mathbf{v}^*, \widehat{\mathbf{\Theta}}_{1,n} - \mathbf{\Theta}_{1,0}])\\
&= P_n\dot{\ell}(\mathbf{\Theta}_{1,0};\mathbf{u}(\mathbf{\Theta}_{1,0}))[\mathbf{v}^*] + I_1 + I_2 + I_3. \qquad \text{(S5.57)}
\end{aligned}$$

We now investigate the asymptotic properties of I_1, I_2, I_3. For I_1, by the Chebyshev inequality, $P\dot{\ell}(\mathbf{\Theta}_{1,0};\mathbf{u}(\mathbf{\Theta}_{1,0}))[\Pi_n\mathbf{v}^* - \mathbf{v}^*] = 0$ and $\|\Pi_n\mathbf{v}^* - \mathbf{v}^*\| = o(1)$, we have that

$$I_1 = o_p(n^{-1/2}). \qquad \text{(S5.58)}$$

For I_2, by the definition of $r[\Pi_n\mathbf{v}^*, \widehat{\mathbf{\Theta}}_{1,n} - \mathbf{\Theta}_{1,0}]$, we can get

$$I_2 = (P_n - P)\left\{\dot{\ell}(\widehat{\mathbf{\Theta}}_{1,n};\mathbf{u}(\widehat{\mathbf{\Theta}}_{1,n})) - \dot{\ell}(\mathbf{\Theta}_{1,0};\mathbf{u}(\mathbf{\Theta}_{1,0}))\right\}[\Pi_n\mathbf{v}^*].$$

By Theorem 2.8.3 of Van der Vaart and Wellner (1996), we know that $\{\dot{\ell}(\mathbf{\Theta}_1;\mathbf{u}(\mathbf{\Theta}_1))[\Pi_n\mathbf{v}^*] : \|\mathbf{\Theta}_1 - \mathbf{\Theta}_{1,0}\| = O_p(\delta_n)\}$ is a Donsker class. Hence by Theorem 2.11.23 of Van der Vaart and Wellner (1996), we have

$$I_2 = o_p(n^{-1/2}). \qquad \text{(S5.59)}$$

For I_3, by the definition of $r[\Pi_n \mathbf{v}^*, \widehat{\mathbf{\Theta}}_{1,n} - \mathbf{\Theta}_{1,0}]$, we have

$$
\begin{aligned}
I_3 &= P\big(r[\Pi_n \mathbf{v}^*, \widehat{\mathbf{\Theta}}_{1,n} - \mathbf{\Theta}_{1,0}]\big) \\
&= P\big(\dot{\ell}(\widehat{\mathbf{\Theta}}_{1,n}; \mathbf{u}(\widehat{\mathbf{\Theta}}_{1,n}))[\Pi_n \mathbf{v}^*] - \dot{\ell}(\mathbf{\Theta}_{1,0}; \mathbf{u}(\mathbf{\Theta}_{1,0}))[\Pi_n \mathbf{v}^*]\big).
\end{aligned}
$$

Let $\check{\mathbf{v}} = \widehat{\mathbf{\Theta}}_{1,n} - \mathbf{\Theta}_{1,0}$ and $f(s) = \dot{\ell}(\mathbf{\Theta}_{1,0} + s\check{\mathbf{v}}; \mathbf{u}(\mathbf{\Theta}_{1,0} + s\check{\mathbf{v}}))[\Pi_n \mathbf{v}^*]\big)$. Then $I_3 = P(f(1) - f(0)) = P\big(\dot{f}(\xi)\big)$ for some $\xi \in (0, 1)$. By the definition of $\ddot{\ell}$, we have $I_3 = P\ddot{\ell}(\widetilde{\mathbf{\Theta}}_1, \mathbf{u}(\widetilde{\mathbf{\Theta}}_1))[\Pi_n \mathbf{v}^*, \widehat{\mathbf{\Theta}}_{1,n} - \mathbf{\Theta}_{1,0}]$, where $\widetilde{\mathbf{\Theta}}_1 = \mathbf{\Theta}_{1,0} + \xi\check{\mathbf{v}}$ lies between $\widehat{\mathbf{\Theta}}_{1,n}$ and $\mathbf{\Theta}_{1,0}$. Thus,

$$
\begin{aligned}
I_3 &= P\left(\ddot{\ell}(\widetilde{\mathbf{\Theta}}_1; \mathbf{u}(\widetilde{\mathbf{\Theta}}_1)) - \ddot{\ell}(\mathbf{\Theta}_{1,0}; \mathbf{u}(\mathbf{\Theta}_{1,0}))\right)[\Pi_n \mathbf{v}^*, \widehat{\mathbf{\Theta}}_{1,n} - \mathbf{\Theta}_{1,0}] \\
&\quad + P\ddot{\ell}(\mathbf{\Theta}_{1,0}; \mathbf{u}(\mathbf{\Theta}_{1,0}))[\Pi_n \mathbf{v}^*, \widehat{\mathbf{\Theta}}_{1,n} - \mathbf{\Theta}_{1,0}] \\
&= o_p(n^{-1/2}) + P\ddot{\ell}(\mathbf{\Theta}_{1,0}; \mathbf{u}(\mathbf{\Theta}_{1,0}))[\mathbf{v}^*, \widehat{\mathbf{\Theta}}_{1,n} - \mathbf{\Theta}_{1,0}] \\
&\quad + P\ddot{\ell}(\mathbf{\Theta}_{1,0}; \mathbf{u}(\mathbf{\Theta}_{1,0}))[\Pi_n \mathbf{v}^* - \mathbf{v}^*, \widehat{\mathbf{\Theta}}_{1,n} - \mathbf{\Theta}_{1,0}] \\
&= o_p(n^{-1/2}) + P\ddot{\ell}(\mathbf{\Theta}_{1,0}; \mathbf{u}(\mathbf{\Theta}_{1,0}))[\mathbf{v}^*, \widehat{\mathbf{\Theta}}_{1,n} - \mathbf{\Theta}_{1,0}], \qquad \text{(S5.60)}
\end{aligned}
$$

where the second equality follows from (A1), (A2), $\|\Pi_n \mathbf{v}^*\|^2 \to \|\mathbf{v}^*\|^2$ and $\delta_n^2 = o(n^{-1/2})$, and the last equality holds since $\delta_n \|\Pi_n \mathbf{v}^* - \mathbf{v}^*\| = o(n^{-1/2})$.

Hence, by (S5.56), (S5.58), (S5.59), (S5.60), together with $P\dot{\ell}(\mathbf{\Theta}_{1,0}; \mathbf{u})[\mathbf{v}^*] = 0$ and the definition of $< \cdot, \cdot >$, we can establish

$$
\begin{aligned}
0 &= P_n \dot{\ell}(\mathbf{\Theta}_{1,0}; \mathbf{u})[\mathbf{v}^*] - < \mathbf{v}^*, \widehat{\mathbf{\Theta}}_{1,n} - \mathbf{\Theta}_{1,0} > + o_p(n^{-1/2}) \\
&= (P_n - P)\dot{\ell}(\mathbf{\Theta}_{1,0}; \mathbf{u})[\mathbf{v}^*] - < \mathbf{v}^*, \widehat{\mathbf{\Theta}}_{1,n} - \mathbf{\Theta}_{1,0} > + o_p(n^{-1/2}).
\end{aligned}
$$

By the central limit theorem, we have

$$\sqrt{n} < \mathbf{v}^*, \widehat{\mathbf{\Theta}}_{1,n} - \mathbf{\Theta}_{1,0} >= \sqrt{n}(P_n - P)\dot{\ell}(\mathbf{\Theta}_{1,0}; \mathbf{u})[\mathbf{v}^*] + o_p(1) \to N(0, \|\mathbf{v}^*\|^2),$$

in distribution, with $\|\mathbf{v}^*\|^2 = \|\dot{\ell}(\mathbf{\Theta}_{1,0}; \mathbf{u})[\mathbf{v}^*]\|^2$.

Next we calculate $\|\mathbf{v}^*\|^2$ to prove (b) $\|\mathbf{v}^*\|^2 = \mathbf{a}_n' \mathbf{\Sigma}^{-1} \mathbf{a}_n$. Rewrite $\overrightarrow{\mathbf{H}}_1 = (\beta_1, \cdots, \beta_{ds_0})$. For each component $\beta_q, q = 1, 2, \cdots, ds_0$, let $\psi_q^* = (b_{1,q}^*, b_{2,q}^*, \cdots, b_{d+1,q}^*)$ be the minimizer

$$E\{\ell_{\overrightarrow{\mathbf{H}}_1} \cdot e_q - \ell_1[b_{1,q}] - \ell_2[b_{2,q}] - \cdots - \ell_{d+1}[b_{d+1,q}]\}^2,$$

with respect to $\psi_q = (b_{1,q}, b_{2,q}, \cdots, b_{d+1,q})$, where e_q is a ds_0 dimensional vector of zeros except the $q-$th element being equal to 1, $\ell_{i,\overrightarrow{\mathbf{H}}_1} = (\ell_{\beta_1}, \cdots, \ell_{\beta_{ds_0}})'$, $\ell_{\beta_q} = \frac{\partial \ell(\mathbf{\Theta}_{1,0}; \mathbf{u}_i)}{\partial \beta_q}$, and $\ell_1[b_{1,q}], \ell_2[b_{2,q}], \cdots, \ell_{d+1}[b_{d+1,q}]$ are the directional derivatives with respect to $g, \psi_j, j = 1, \cdots, d$, respectively.

From the Riesz representation theorem, it follows

$$\|\mathbf{v}^*\|^2 = \sup_{\mathbf{v} \neq 0, \mathbf{v} \in \overline{V}} \frac{|\mathbf{a}_n' \mathbf{v}_{\overrightarrow{\mathbf{H}}_1}|^2}{\|\mathbf{v}\|^2},$$

where $\mathbf{v} = (\mathbf{v}_{\overrightarrow{\mathbf{H}}_1}', v_1, \ldots, v_{d+1})' := (\mathbf{v}_{\overrightarrow{\mathbf{H}}_1}', \mathbf{b}')' \in R^{ds_0} \times \overline{\mathbf{B}}$. Recalling we denote

$$\frac{\partial \ell(\mathbf{\Theta}_0, \mathbf{u})}{\partial \mathbf{\Theta}}[\mathbf{v}] = \ell_{\overrightarrow{\mathbf{H}}_1}[\mathbf{v}_{\overrightarrow{\mathbf{H}}_1}] + \ell_1[v_1] + \ldots + \ell_{d+1}[v_{d+1}],$$

it yields

$$\sup_{\mathbf{v} \neq 0, \mathbf{v} \in \overline{V}} \frac{|\mathbf{a}_n' \mathbf{v}_{\overrightarrow{\mathbf{H}}_1}|^2}{\|\mathbf{v}\|^2} = \sup_{\mathbf{b} \neq 0, \mathbf{b} \in \overline{\mathbf{B}}} \frac{|\mathbf{a}_n' \mathbf{v}_{\overrightarrow{\mathbf{H}}_1}|^2}{E\big(\ell_{\overrightarrow{\mathbf{H}}_1}[\mathbf{v}_{\overrightarrow{\mathbf{H}}_1}] + \ell_1[v_1] + \ldots + \ell_{d+1}[v_{d+1}]\big)^2}. \tag{S5.61}$$

The supremum reaches at $(v_1, \ldots, v_{d+1}) = (\mathbf{b}_1^{*\prime}\mathbf{c}_1^*, \ldots, \mathbf{b}_{d+1}^{*\prime}\mathbf{c}_{d+1}^*)$ for some $\mathbf{c}_k^* = (c_{k,1}^*, \ldots, c_{k,ds_0}^*)'$ depending on $\mathbf{v}_{\vec{\mathbf{H}}_1}$ and $\mathbf{b}_k^*$, where $\mathbf{b}_k^* = (b_{k,1}^*, \ldots, b_{k,ds_0}^*)'$ solves the following infinite-dimensional optimization problems for $k = 1, \ldots, d+1$ and $q = 1, \ldots, ds_0$,

$$\inf_{b_{k,q \in L_2([0,1])}} E\big(\ell_{\vec{\mathbf{H}}_1} \cdot e_q - \ell_1[b_{1,q}] - \ldots - \ell_{d+1}[b_{d+1,q}]\big)^2.$$

Replacing the optimal $\mathbf{b} = (v_1, \ldots, v_{d+1})'$ with the solution to (S5.61), we obtain $\mathbf{c}_l^* = -\mathbf{v}_{\vec{\mathbf{H}}_1}$, $l = 1, \ldots, d+1$ and the optimal value of $\mathbf{v}_{\vec{\mathbf{H}}_1}$, i.e., $\mathbf{v}_{\vec{\mathbf{H}}_1}^* = [E(S_{\vec{\mathbf{H}}_{1,0}} S'_{\vec{\mathbf{H}}_{1,0}})]^{-1}\mathbf{a}_n$, where $S_{\vec{\mathbf{H}}_{1,0}}$ is an ds_0-dimensional vector, with the $q-$th element being $\ell_{\vec{\mathbf{H}}_1} \cdot e_q - \ell_1[b_{1,q}^*] - \ldots - \ell_{d+1}[b_{d+1,q}^*]$. Therefore, $\mathbf{v}^* = (\mathbf{v}_{\vec{\mathbf{H}}_1}^*, -\mathbf{b}_1^{*\prime}\mathbf{v}_{\vec{\mathbf{H}}_1}^*, \ldots, -\mathbf{b}_{d+1}^{*\prime}\mathbf{v}_{\vec{\mathbf{H}}_1}^*)'$ and then $\|\mathbf{v}^*\|^2 = \mathbf{a}_n'\boldsymbol{\Sigma}^{-1}\mathbf{a}_n$. This completes the proof.

S6. Selection of tuning parameters and its simulation

The proposed estimation procedure involves the selection of several tuning parameters: the numbers of FPC K_n and splines k_n, the dimension of index d, the tuning parameters w and λ.

The selection of K_n has been extensively discussed in the literature. Compared to the traditional FPCA, the proposed estimation is less sensitive to the choice of K_n since we will further choose the scores by the SCAD-group penalty. Following the literature, we choose K_n by calculating the proportion of variability explained by each principal component (Happ and Greven, 2018). Since the directions that contain the important information on the relationship between $\mathbf{Z}_i(\cdot)$ and Y_i may be different from those for $\mathbf{Z}_i(\cdot)$, we take K_n to be large so that we can keep as much information of $\mathbf{Z}_i(\cdot)$ as possible. Recall that $\mathbf{\Gamma}\mathbf{\Gamma}'$ is the diagonal matrix of diagonal element being eigenvalues and decreasing, so the desired proportion is simply $(\mathbf{\Gamma}\mathbf{\Gamma}')_{kk}/\mathrm{tr}(\mathbf{\Gamma}\mathbf{\Gamma}')$. We choose K_n so that $\sum_{k=1}^{K_n}(\mathbf{\Gamma}^{(0)}\mathbf{\Gamma}^{(0)\prime})_{kk}/\mathrm{tr}(\mathbf{\Gamma}^{(0)}\mathbf{\Gamma}^{(0)\prime}) \geq 96\%$, where $\mathbf{\Gamma}^{(0)}$ is the initial values for $\mathbf{\Gamma}$ described in Section 3.

We use the cubic B-spline basis approximation with $k_n = q_n + 4$, where q_n is the number of interior knots. 2-5 interior knots seem quite adequate and enough, which is what we recommend. We took $w = (\min\{n_i, i = 1, \cdots, n\})^{-v}, 0 < v < 1$ so that the assumption on w can be satisfied, and we tune v instead of w. The proposed algorithm is not sensitive to v and we take $v = 0.7$ in all of our numerical experiments and real data analysis. We use the Bayesian information criterion to select d and λ:

$$BIC(d,\lambda) = \sum_{i=1}^{n} \ell(\mathbf{\Theta}_n; \mathbf{u}_i) - w\sum_{i=1}^{n}\sum_{q=1}^{p} \|\mathbf{Z}_{iq} - \mathbf{B}_n(\mathbf{t}_i)\boldsymbol{\alpha}_q - \sum_{k=1}^{K_n} u_{ik}\mathbf{B}_n(\mathbf{t}_i)\boldsymbol{\gamma}_{kq}\|^2 - \frac{1}{2}df(\mathbf{\Omega}_n)\log(\sum_{i=1}^{n} n_i), \tag{S6.62}$$

where $df(\mathbf{\Omega}_n)$ is the number of nonzero parameters in $\mathbf{\Omega}_n$. In practice, we first select d without regularizing $\mathbf{H}$, that is, with $\lambda = 0$. Then, given d, we select λ.

We then use the setting of Example 1 to examine the performance of (S6.62) for selecting tuning parameters. Web Figure 4(a) is the average percentage of variance explained by each principal component and accumulated variances of the first thirteen principal components across 200 simulations. From Web Figure 4(a), we can see that the first four components account for the most important information of covariate curves, the variances of 5th and 6th components are small, but only the two components are associated with the response based on the setting of Example 1. We first take $K_n = 6$ so that the cumulative proportion of the first K_n is larger than 96%. We calculate $BIC(d, \lambda)$ when $d = 1, 2, 3, 4, 5$ and λ takes values in $\{0.1, 0.2, \cdots, 0.9\}$. Web Figure 4(b) displays the resulting BIC over various combinations of (d, λ). The results show that the optimal values of d, λ based on the BIC criterion are roughly independent, suggesting that we can separately select them. In addition, the largest BIC is achieved with $d = 2$, which is the true value. Web Figure 4(c), the barplot of the selected d based on the strategy and the BIC criterion 3.11, shows that the proposed method can correctly identify the dimension of the index. We also investigate the sensitivity of our proposed method to the number of interior knots q_n. Web Table 2 represents the RMSE of the proposed estimators with $q_n = 2, 3, 4, 5$. The results show that our method is insensitive to the number of interior knots.

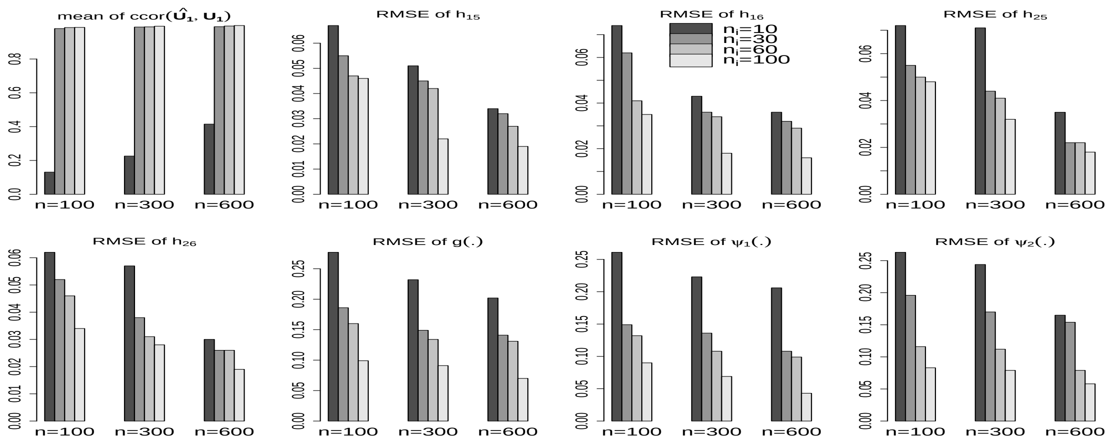
S7. Web Tables 1-2 and Figures 1-5

Web Table 1: The estimation results based on the proposed (Prop) method and the FR-FPCA model for Example 4.

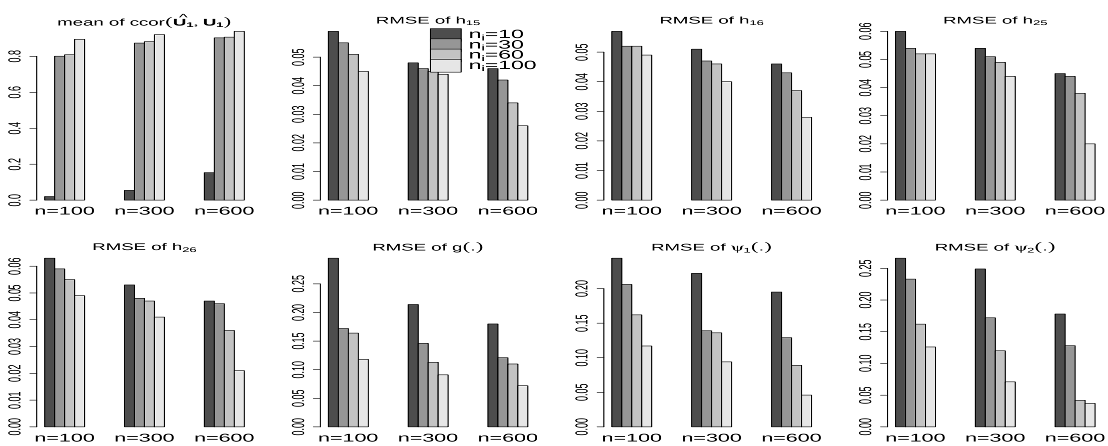
	Prop												FR-FPCA			
	bias	SD	RMSE		bias	SD	RMSE		bias	SD	RMSE		bias	SD	RMSE
$\mu_1(\cdot)$	0.005	0.071	0.071	$\phi_{62}(\cdot)$	0.009	0.058	0.059	$g(\cdot)$	0.087	0.154	0.177	$\mu_1(\cdot)$	0.008	0.117	0.117
$\mu_2(\cdot)$	0.005	0.078	0.078	h_{15}	0.007	0.030	0.031	$\psi_1(\cdot)$	0.032	0.063	0.071	$\mu_2(\cdot)$	0.014	0.124	0.125
$\phi_{51}(\cdot)$	0.009	0.084	0.084	h_{16}	0.006	0.017	0.018	$\psi_2(\cdot)$	0.018	0.094	0.096	$g(\cdot)$	0.213	0.468	0.514
$\phi_{52}(\cdot)$	0.011	0.062	0.063	h_{25}	0.010	0.034	0.035					$\psi_1(\cdot)$	0.113	0.697	0.706
$\phi_{61}(\cdot)$	0.010	0.071	0.072	h_{26}	0.007	0.019	0.020					$\psi_2(\cdot)$	0.134	0.474	0.492

Web Table 2: RMSE of the proposed estimators with $q_n = 2, 3, 4, 5$ for Example 1.

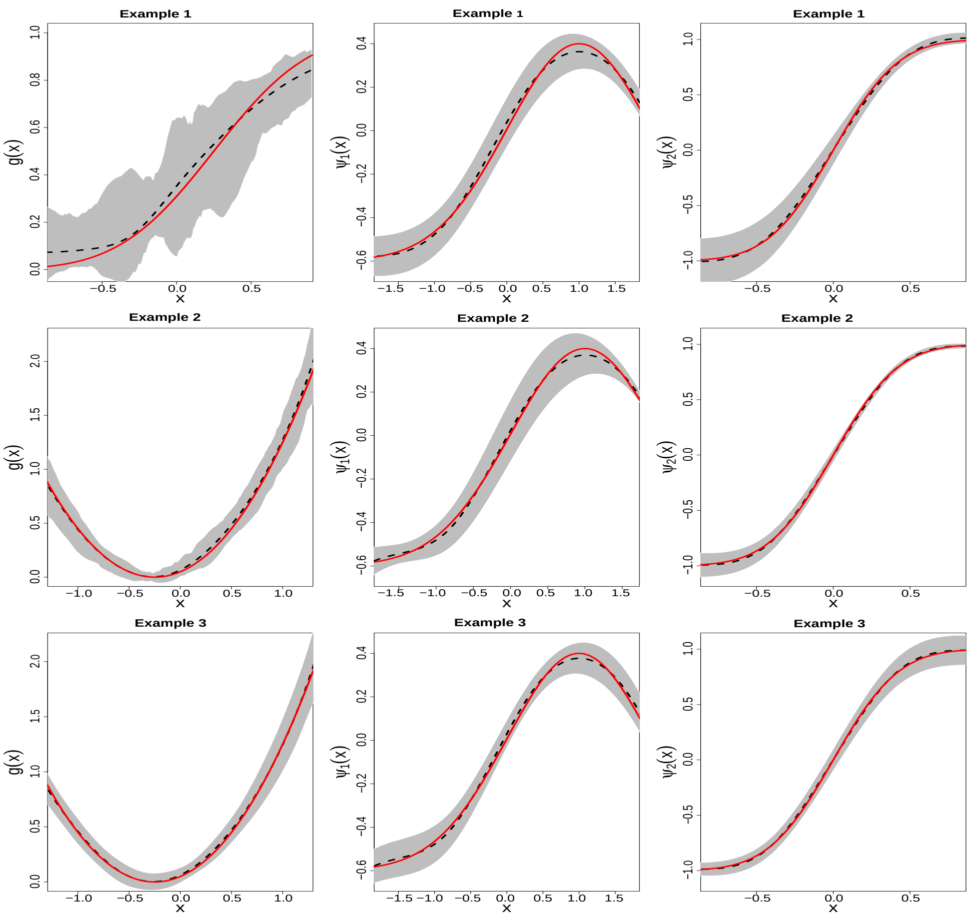
q_n	$\mu(\cdot)$	$\phi_5(\cdot)$	$\phi_6(\cdot)$	h_{15}	h_{16}	h_{25}	h_{26}	$g(\cdot)$	$\psi_1(\cdot)$	$\psi_2(\cdot)$
2	0.031	0.057	0.043	0.023	0.027	0.020	0.023	0.129	0.085	0.072
3	0.030	0.055	0.043	0.025	0.028	0.021	0.023	0.129	0.084	0.072
4	0.029	0.055	0.043	0.024	0.028	0.020	0.025	0.129	0.086	0.071
5	0.029	0.054	0.043	0.023	0.028	0.020	0.024	0.130	0.086	0.072



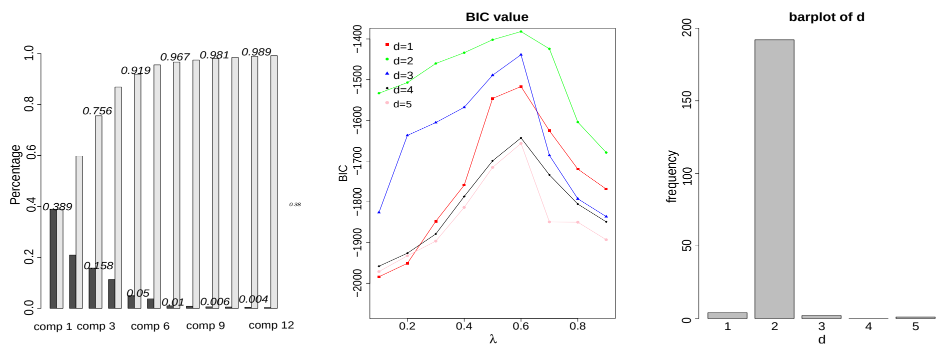
Web Figure 1: the results over the combinations of n and n_i for Example 1.



Web Figure 2: the results over the combinations of n and n_i for Example 2.

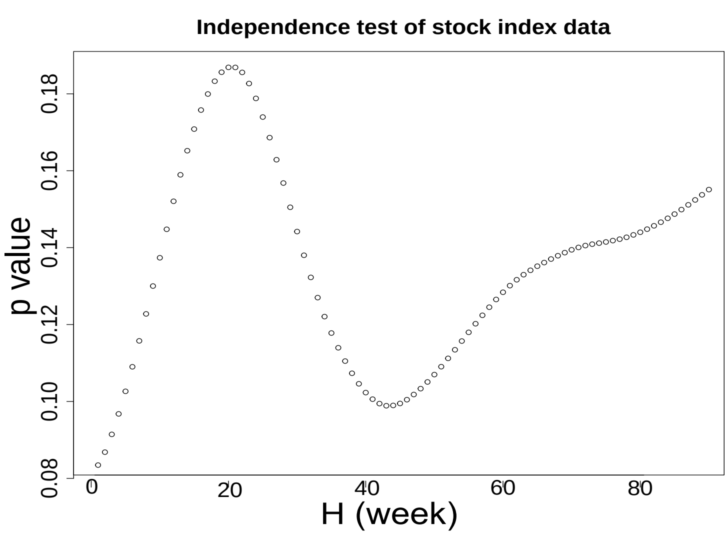


Web Figure 3: The average estimates of link and component functions using the Prop (red solid:true curves; black dotted:proposed estimator; gray shadow: 95% confidence band) for Examples 1–3.



(a) (b) (c)

Web Figure 4: (a) The barplot of principal components (dark grey: single percentage; light grey: accumulated percentage), (b) The BIC values under various d and λ, (c) The barplot of selected d based on the BIC criterion for Example 1.



Web Figure 5: The p-value for the independence test of stock index data.

Abbreviation and Notation

	meaning	first appearance
GLFR	generalized linear functional regression	page 2
GAFRM	generalized additive functional regression models	page 2
FR-FPCA	functional regression based on functional principle component analysis	page 3
DFR	direct functional regression	page 2
UGLFR	generalized linear functional regression with unknown link function	page 19
GFFR	generalized functional feature regression model	page 7
Prop	the proposed method	page 19
KGLFR	the GLFR with the logit link function	page 28
MIFR	the multiple index functional regression models	page 28
$\mathbf{u}_i$	$(u_{i1}, \cdots, u_{i,K_n})'$	page 4
$\mathbf{U}$	$(\mathbf{u}_1, \cdots, \mathbf{u}_n)'$	page 6
$V(m_i)$	$\text{var}(Y_i \mid \mathbf{Z}_i)$	page 7
$\Vert \cdot \Vert$	Euclidean norm for matrix	page 7
$\Vert \cdot \Vert_2$	L^2 norm for function	page 14
$\mathbf{H}$	$(\mathbf{h}_1, \cdots, \mathbf{h}_d)' = (\mathbf{H}_{\cdot 1}, \cdots, \mathbf{H}_{\cdot, K_n})$	page 4
$\mathbf{h}_j$	$(h_{j1}, \cdots, h_{jK_n})'$	page 7
$\boldsymbol{\alpha}$	the matrix of spline coefficient of mean functions	page 8
$\boldsymbol{\Gamma}$	the matrix of spline coefficient of eigenfunctions	page 9
$\boldsymbol{\delta}$	the vector of spline coefficient of link function	page 8
$\boldsymbol{\vartheta}$	the matrix of spline coefficient of additive component functions	page 9
$\boldsymbol{\Theta}_n$	$(\mathbf{H}, \boldsymbol{\delta}, \boldsymbol{\vartheta})$	page 9
$\boldsymbol{\Omega}_n$	$(\boldsymbol{\alpha}, \boldsymbol{\Gamma}, \mathbf{U}, \mathbf{H}, \boldsymbol{\delta}, \boldsymbol{\vartheta})$	page 9
$\mathbf{B}_{ni}$	$pk_n \times pn_i$ block diagonal matrix with block elements $\mathbf{B}_n(\mathbf{t}_i)'$	page 11
$\mathbf{V}_{i1}$	$\mathbf{B}_{ni}\mathbf{Z}_i$	page 11
$\mathbf{V}_{i2}$	$\mathbf{B}_{ni}\mathbf{B}'_{ni}$	page 11

Bibliography

Cardot, H. (2000). Nonparametric estimation of smoothed principal components analysis of sampled noisy functions. *J NONPARAMETR STAT 12*(4), 503–538.

David G. Luenberger, Y. Y. (2016). *Linear and Nonlinear Programming.* Springer International Publishing.

Happ, C. and S. Greven (2018). Multivariate functional principal component analysis for data observed on different (dimensional) domains. *J Am Stat Assoc 113*(522), 649–659.

Horn, R. A. and C. R. Johnson (2012). *Matrix Analysis*. Cambridge University Press, Second Edition.

Ma, L., T. Hu, and J. Sun (2015). Sieve maximum likelihood regression analysis of dependent current status data. *Biometrika 102*(3), 731–738.

Pollard, D. (1984). *Convergence of Stochastic Processes*. David Pollard.

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

Shen, X. and W. H. Wong (1994). Convergence rate of sieve estimates. *Ann Stat 22*(2), 580–615.

Van De Geer, S. (2000). *Empirical Processes in M-estimation*. Cambridge UP.

Van der Vaart, A. W. and J. A. Wellner (1996). *Weak convergence and empirical processes*. Springer.

Wu, C. J. (1983). On the convergence properties of the em algorithm. *Ann Stat 11*(1), 95–103.