ESTIMATION AND MODEL SELECTION PROCEDURES IN GENERALIZED FUNCTIONAL PARTIALLY ADDITIVE HYBRID MODEL WITH DIVERGING NUMBER OF COVARIATES

Supplementary Material

S1. The proof of theorems

This part contains the proofs of THEOREM 1 and THEOREM 2, which depend on two preliminary lemmas and first we will give the following Lemma 1, a direct result of de Boor (2001).

Lemma 1 There exists positive constants C, for any functions $g(z) \in \mathcal{H}(r)$ with $r < h-1$, there exists a function $g(z) \in \mathcal{S}_n^0$, such that

$$\sup_z |g(z) - g_0(z)| \leqslant CK_n^{-r}.$$

From Lemma 1, we can find $\boldsymbol{\zeta} = \{\zeta_{jk}, j = 1, \cdots, p_n, k = 1, \cdots, K_n\}^\top$, and an additive spline function $h(\boldsymbol{Z}) = \boldsymbol{B}^\top(\boldsymbol{Z})\boldsymbol{\zeta} \in \mathcal{G}_n$, such that

$$\sup_{\boldsymbol{Z}} |h(\boldsymbol{Z}) - h_0(\boldsymbol{Z})| \leqslant Cp_nK_n^{-r}. \tag{S1.1}$$

Denote $m_0(\boldsymbol{\mathcal{F}}) = \int_0^1 X(t)\alpha_0(t)\mathrm{d}t + h_0(\boldsymbol{Z})$, $\boldsymbol{\mathcal{T}} = (\hat{\boldsymbol{\xi}}, \boldsymbol{Z})$, $\tilde{\xi}_{ik} = \langle \hat{X}_i, v_k \rangle$ and $\tilde{m}_0(\boldsymbol{\mathcal{T}}) = \hat{\boldsymbol{\xi}}^\top \boldsymbol{\gamma}_0 + \boldsymbol{B}^\top(\boldsymbol{Z})\boldsymbol{\zeta}_0$, where $\boldsymbol{\gamma}_0$ and $\boldsymbol{\zeta}_0$ be the true parameter values of $\boldsymbol{\gamma}$ and $\boldsymbol{\zeta}$, and define

$$\boldsymbol{\Gamma}(\hat{\boldsymbol{\xi}}) = \frac{E[\boldsymbol{Z}\rho_2\{\tilde{m}_0(\boldsymbol{\mathcal{T}})\}|\hat{\boldsymbol{\xi}}]}{E[\rho_2\{\tilde{m}_0(\boldsymbol{\mathcal{T}})\}|\hat{\boldsymbol{\xi}}]}, \quad \tilde{\boldsymbol{B}}(\boldsymbol{Z}) = \boldsymbol{B}(\boldsymbol{Z}) - \boldsymbol{\Gamma}(\hat{\boldsymbol{\xi}}).$$

Lemma 2 Let $R_i = \tilde{m}_0(\boldsymbol{\mathcal{T}}_i) - m_0(\boldsymbol{\mathcal{F}}_i) = \hat{\boldsymbol{\xi}}_i^\top \boldsymbol{\gamma}_0 - \int_0^1 X_i(t)\alpha_0(t)\mathrm{d}t + \boldsymbol{B}^\top(\boldsymbol{Z}_i)\boldsymbol{\zeta}_0 - h_0(\boldsymbol{Z}_i), i = 1, 2, \ldots, n$. Suppose that conditions C1-C12 hold and the dimension of the spline space $\mathcal{S}_n^0$ fulfils $K_n = O_p(n^{1/(1+2r)})$, then we have

$$\|R_i\|^2 = O_p\left(n^{-\frac{a+2b-1}{a+2b}}\right), \quad i = 1, 2, \ldots, n.$$

Proof of Lemma 2.

$$\begin{aligned}
\|R_i\|^2 &= \|\hat{\boldsymbol{\xi}}_i^\top \boldsymbol{\gamma}_0 - \int_0^1 X_i(t)\alpha_0(t)\mathrm{d}t + \boldsymbol{B}^\top(\boldsymbol{Z}_i)\boldsymbol{\zeta}_0 - h_0(\boldsymbol{Z}_i)\|^2 \\
&= \|\sum_{k=1}^{m_n} \hat{\xi}_{ik}\gamma_{0k} - \sum_{k=1}^{\infty} \xi_{ik}\gamma_{0k} + \boldsymbol{B}^\top(\boldsymbol{Z}_i)\boldsymbol{\zeta}_0 - h_0(\boldsymbol{Z}_i)\|^2 \\
&= \|\sum_{k=1}^{m_n}(\hat{\xi}_{ik} - \xi_{ik})\gamma_{0k} - \sum_{k=m_n+1}^{\infty} \xi_{ik}\gamma_{0k} + \boldsymbol{B}^\top(\boldsymbol{Z}_i)\boldsymbol{\zeta}_0 - h_0(\boldsymbol{Z}_i)\|^2 \\
&= \|\sum_{k=1}^{m_n}(\hat{\xi}_{ik} - \tilde{\xi}_{ik})\gamma_{0k} + \sum_{k=1}^{m_n}(\tilde{\xi}_{ik} - \xi_{ik})\gamma_{0k} - \sum_{k=m_n+1}^{\infty} \xi_{ik}\gamma_{0k} + \boldsymbol{B}^\top(\boldsymbol{Z}_i)\boldsymbol{\zeta}_0 - h_0(\boldsymbol{Z}_i)\|^2 \\
&\leqslant 4\|\sum_{k=1}^{m_n}(\hat{\xi}_{ik} - \tilde{\xi}_{ik})\gamma_{0k}\|^2 + 4\|\sum_{k=1}^{m_n}(\tilde{\xi}_{ik} - \xi_{ik})\gamma_{0k}\|^2 \\
&\quad +4\|\sum_{k=m_n+1}^{\infty} \xi_{ik}\gamma_{0k}\|^2 + 4\|\boldsymbol{B}^\top(\boldsymbol{Z}_i)\boldsymbol{\zeta}_0 - h_0(\boldsymbol{Z}_i)\|^2 \\
&\triangleq 4A_1 + 4A_2 + 4A_3 + 4A_4,
\end{aligned}$$

By lemma 1 of Kong et al. (2016), we can see that $\|\hat{X}_i\|^2 = O_p(1)$, $\|\hat{X}_i - X_i\|^2 =$

$o_p(n^{-1})$ and $\|\hat{v}_k - v_k\|^2 = O_{\mathrm{p}}(n^{-1}k^2)$, then we have

$$\begin{aligned}
A_1 &= \|\sum_{k=1}^{m_n}\langle \hat{X}_i, \hat{v}_k - v_k\rangle \gamma_{0k}\|^2 \\
&\leqslant m_n \sum_{k=1}^{m_n} |\gamma_{0k}|^2 \|\hat{X}_i\|^2 \, \|\hat{v}_k - v_k\|^2 \\
&\leqslant O_{\mathrm{p}}(n^{-1}m_n) \sum_{k=1}^{m_n} k^{2-2b} \\
&\leqslant O_{\mathrm{p}}\left(n^{-\frac{a+2b-1}{a+2b}}\right). \qquad \text{(S1.2)}
\end{aligned}$$

$$\begin{aligned}
A_2 &= \|\sum_{k=1}^{m_n}\langle \hat{X}_i - X_i, v_k\rangle \gamma_{0k}\|^2 \\
&\leqslant m_n \sum_{k=1}^{m_n} \|\hat{X}_i - X_i\|^2 \|\hat{v}_k\|^2 |\gamma_{0k}|^2 \\
&\leqslant m_n o_{\mathrm{p}}(n^{-1}) \sum_{k=1}^{m_n} k^{-2b} \\
&= o_{\mathrm{p}}(n^{-\frac{a+2b-1}{a+2b}}). \qquad \text{(S1.3)}
\end{aligned}$$

It follows that $\{\langle X_i, v_k\rangle\}_{k=m_n+1}^{\infty}$ are uncorrelated random variables and condition C2, we have

$$\begin{aligned}
\mathrm{E}\left(A_3\right) &= \mathrm{E}\left(\sum_{k=m_n+1}^{\infty} \langle X_i, v_k\rangle \gamma_{0k}\right)^2 = \sum_{k=m_n+1}^{\infty} \gamma_{0k}^2 \mathrm{E}\left(\langle X_i, v_k\rangle^2\right) = \sum_{k=m_n+1}^{\infty} \gamma_{0k}^2 \lambda_k \\
&\leqslant C \sum_{k=m_n+1}^{\infty} k^{-2b} k^{-a} = O\left(m_n^{-(2b+a-1)}\right) = O(n^{-\frac{a+2b-1}{a+2b}}).
\end{aligned}$$

So

$$A_3 = O_{\mathrm{p}}(n^{-\frac{a+2b-1}{a+2b}}). \qquad \text{(S1.4)}$$

By (S1.1), it is easy to see that

$$A_4 = O_{\mathrm{p}}(p_n^2 K_n^{-2r}) = o(1). \tag{S1.5}$$

Combining (S1.2)-(S1.5), we obtain that

$$\|R_i\|^2 = O_p\left(n^{-\frac{a+2b-1}{a+2b}}\right).$$

□

Proof of Theorem **??**. Let $\boldsymbol{\zeta} = \boldsymbol{\zeta}_0 + \delta_n \boldsymbol{T}_1, \boldsymbol{\gamma} = \boldsymbol{\gamma}_0 + \delta_n \boldsymbol{T}_2$ and $\boldsymbol{T} = (\boldsymbol{T}_1^\top, \boldsymbol{T}_2^\top)^\top$. It suffices to show that for any given $\varepsilon > 0$, there exists a large constant C such that

$$P\left\{\sup_{\|\boldsymbol{T}\|=C} PL(\boldsymbol{\zeta}_0 + \delta_n \boldsymbol{T}_1, \boldsymbol{\gamma}_0 + \delta_n \boldsymbol{T}_2) < PL\left(\boldsymbol{\zeta}_0, \boldsymbol{\gamma}_0\right)\right\} \geq 1 - \varepsilon. \tag{S1.6}$$

This implies that, with probability tending to one, there is local maximum $(\hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\gamma}})$ in the ball $\{(\boldsymbol{\zeta}_0 + \delta_n \boldsymbol{T}_1, \boldsymbol{\gamma}_0 + \delta_n \boldsymbol{T}_2) : \|(\boldsymbol{T}_1^\top, \boldsymbol{T}_2^\top)^\top\| \leq C\}$ such that $\|\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0\| = O_p(\delta_n)$, $\|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\| = O_p(\delta_n)$.

Denote $D_n(\boldsymbol{T}_1, \boldsymbol{T}_2) = PL(\boldsymbol{\zeta}_0 + \delta_n \boldsymbol{T}_1, \boldsymbol{\gamma}_0 + \delta_n \boldsymbol{T}_2) - PL\left(\boldsymbol{\zeta}_0, \boldsymbol{\gamma}_0\right)$, and $D_{n,1}(\boldsymbol{T}_1, \boldsymbol{T}_2) = L(\boldsymbol{\zeta}_0 + \delta_n \boldsymbol{T}_1, \boldsymbol{\gamma}_0 + \delta_n \boldsymbol{T}_2) - L\left(\boldsymbol{\zeta}_0, \boldsymbol{\gamma}_0\right), D_{n,2}(\boldsymbol{T}_1) = -n \sum_{j=1}^{s} [p_{\lambda_n}(\|\boldsymbol{\zeta}_{0j} + \delta_n \boldsymbol{T}_{1j}\|) - p_{\lambda_n}(\|\boldsymbol{\zeta}_{0j}\|)]$, where s is the number of nonzero additive components. Note that $p_{\lambda_n}(0) = 0$ and $p_{\lambda_n}(\|\boldsymbol{\zeta}\|) \geq 0$ for all $\boldsymbol{\zeta}$, thus $PL(\boldsymbol{\zeta}_0 + \delta_n \boldsymbol{T}_1, \boldsymbol{\gamma}_0 + \delta_n \boldsymbol{T}_2) - PL\left(\boldsymbol{\zeta}_0, \boldsymbol{\gamma}_0\right) \leq D_{n,1} + D_{n,2}$. We first deal with $D_{n,1}$. Let $\boldsymbol{U}_i = (\boldsymbol{B}^\top(\boldsymbol{Z}_i), \hat{\boldsymbol{\xi}}_i^\top)^\top$, then

$$\begin{aligned}
& D_{n,1}(\boldsymbol{T}_1, \boldsymbol{T}_2) \\
= & \sum_{i=1}^{n} [Q\{g^{-1}(\boldsymbol{B}^\top(\boldsymbol{Z}_i)(\boldsymbol{\zeta}_0 + \delta_n \boldsymbol{T}_1) + \hat{\boldsymbol{\xi}}_i^\top(\gamma_0 + \delta_n \boldsymbol{T}_2)), Y_i\} \\
& - Q\{g^{-1}(\boldsymbol{B}^\top(\boldsymbol{Z}_i)\boldsymbol{\zeta}_0 + \hat{\boldsymbol{\xi}}_i^\top \boldsymbol{\gamma}_0), Y_i\}] \\
= & \sum_{i=1}^{n} [Q\{g^{-1}(\tilde{m}_0(\boldsymbol{\mathcal{T}}_i) + \delta_n(\boldsymbol{B}^\top(\boldsymbol{Z}_i)\boldsymbol{T}_1 + \hat{\boldsymbol{\xi}}_i^\top \boldsymbol{T}_2)), Y_i\} - Q\{g^{-1}(\tilde{m}_0(\boldsymbol{\mathcal{T}}_i), Y_i\}] \\
= & \sum_{i=1}^{n} [Q\{g^{-1}(\tilde{m}_0(\boldsymbol{\mathcal{T}}_i) + \delta_n \boldsymbol{T}^\top \boldsymbol{U}_i), Y_i\} - Q\{g^{-1}(\tilde{m}_0(\boldsymbol{\mathcal{T}}_i), Y_i\}].
\end{aligned}$$

By means of a Taylor expansion, we obtain

$$D_{n,1} = \sum_{i=1}^{n} q_1\left(\tilde{m}_0(\boldsymbol{\mathcal{T}}_i), Y_i\right) \delta_n \boldsymbol{T}^\top \boldsymbol{U}_i + \frac{n}{2}\delta_n^2 \boldsymbol{T}^\top \mathbf{B}_n \boldsymbol{T},$$

where the random matrix

$$\begin{aligned}
\mathbf{B}_n &= \frac{1}{n}\sum_{i=1}^{n} q_2(\tilde{m}_0(\boldsymbol{\mathcal{T}}_i) + \varsigma_i, Y_i)\boldsymbol{U}_i \boldsymbol{U}_i^\top \\
&= \frac{1}{n}\sum_{i=1}^{n} [(Y_i - g^{-1}(\tilde{m}_0(\boldsymbol{\mathcal{T}}_i) + \varsigma_i))\rho_1'(\tilde{m}_0(\boldsymbol{\mathcal{T}}_i) + \varsigma_i) - \rho_2(\tilde{m}_0(\boldsymbol{\mathcal{T}}_i) + \varsigma_i)]\boldsymbol{U}_i \boldsymbol{U}_i^\top,
\end{aligned}$$

with ς_i between 0 and $\delta_n \boldsymbol{T}^\top \boldsymbol{U}_i$, for $i = 1, \cdots, n$, independent of Y_i. By Theorem 19.24 of van Der Vaart (1998), it can be shown that

$$\begin{aligned}
\mathbf{B}_n &= -\frac{1}{n}\sum_{i=1}^{n} \rho_2(m_0(\boldsymbol{\mathcal{F}}_i))\boldsymbol{U}_i \boldsymbol{U}_i^\top + o_p(1) \\
&= -E[\rho_2(m_0(\boldsymbol{\mathcal{F}}))\boldsymbol{U}\boldsymbol{U}^\top] + o_p(1) \equiv -\mathbf{B} + o_p(1). \qquad \text{(S1.7)}
\end{aligned}$$

So

$$D_{n,1} = \delta_n \boldsymbol{T}^\top \textstyle\sum_{i=1}^{n} q_1\left(\tilde{m}_0(\boldsymbol{\mathcal{T}}_i), Y_i\right) \boldsymbol{U}_i - \frac{n}{2}\delta_n^2 \boldsymbol{T}^\top \mathbf{B} \boldsymbol{T} + o_p(1).$$

By means of a Taylor expansion, we have

$$
\begin{aligned}
& \delta_n \sum_{i=1}^{n} q_1\left(\tilde{m}_0(\boldsymbol{\mathcal{T}}_i), Y_i\right) \boldsymbol{U}_i \\
= \;& \delta_n \sum_{i=1}^{n} q_1\left(m_0(\boldsymbol{\mathcal{F}}_i), Y_i\right) \boldsymbol{U}_i + \delta_n \sum_{i=1}^{n} q_2\left(m_0(\boldsymbol{\mathcal{F}}_i), Y_i\right)\left(\tilde{m}_0(\boldsymbol{\mathcal{T}}_i) - m_0(\boldsymbol{\mathcal{F}}_i)\right) \boldsymbol{U}_i + O_p(n\delta_n \|\tilde{m}_0 - m_0\|^2) \\
= \;& \delta_n \sum_{i=1}^{n} q_1\left(m_0(\boldsymbol{\mathcal{F}}_i), Y_i\right) \boldsymbol{U}_i + \delta_n \sum_{i=1}^{n} q_2\left(m_0(\boldsymbol{\mathcal{F}}_i), Y_i\right)\left(\tilde{m}_0(\boldsymbol{\mathcal{T}}_i) - m_0(\boldsymbol{\mathcal{F}}_i)\right) \boldsymbol{U}_i + o_p(1).
\end{aligned}
$$

By Lemma 2, the second term on the right hand side of the above is

$$
\begin{aligned}
& \delta_n \sum_{i=1}^{n} q_2\left(m_0(\boldsymbol{\mathcal{F}}_i), Y_i\right)\left(\tilde{m}_0(\boldsymbol{\mathcal{T}}_i) - m_0(\boldsymbol{\mathcal{F}}_i)\right) \boldsymbol{U}_i \\
& = \delta_n \sum_{i=1}^{n} \rho_1'(m_0(\boldsymbol{\mathcal{F}}_i))\boldsymbol{U}_i\left(\tilde{m}_0(\boldsymbol{\mathcal{T}}_i) - m_0(\boldsymbol{\mathcal{F}}_i)\right)\varepsilon_i - \delta_n \sum_{i=1}^{n} \rho_2(m_0(\boldsymbol{\mathcal{F}}_i))\boldsymbol{U}_i\left(\tilde{m}_0(\boldsymbol{\mathcal{T}}_i) - m_0(\boldsymbol{\mathcal{F}}_i)\right) \\
& \leq \delta_n \sum_{i=1}^{n} \mid \rho_1'(m_0(\boldsymbol{\mathcal{F}}_i))\boldsymbol{U}_i\varepsilon_i \mid \cdot \parallel \tilde{m}_0 - m_0 \parallel + O_p(n\delta_n \|\tilde{m}_0 - m_0\|) = o_p(1),
\end{aligned}
$$

where $\varepsilon_i = Y_i - g^{-1}(m_0(\boldsymbol{\mathcal{F}}_i))$. Therefore, we have

$$
D_{n,1} \quad = \delta_n \boldsymbol{T}^\top \textstyle\sum_{i=1}^{n} q_1\left(m_0(\boldsymbol{\mathcal{F}}_i), Y_i\right) \boldsymbol{U}_i - \frac{n}{2}\delta_n^2 \boldsymbol{T}^\top \mathbf{B} \boldsymbol{T} + o_p(1), \tag{S1.8}
$$

where the orders of the first term and the second term on the right-hand side of (S1.8) are $O_p(n^{1/2}\delta_n(m_n+K_n))$ and $O_p(n\delta_n^2(m_n+K_n)^2)$, respectively. By choosing a sufficiently large C, the second term dominates the first term uniformly in $\|\boldsymbol{T}\| = C$. Furthermore, according to known conditions, we have

$$
D_{n,2} \leq -n \sum_{j=1}^{s} [\delta_n p'_{\lambda_n}(\|\boldsymbol{\zeta}_{0j}\|) \frac{\boldsymbol{\zeta}_{0j}^\top}{\|\boldsymbol{\zeta}_{0j}\|} \boldsymbol{T}_{1j} + \frac{1}{2}\delta_n^2 p''_{\lambda_n}(\|\boldsymbol{\zeta}_{0j}\|) \boldsymbol{T}_{1j}^\top \boldsymbol{T}_{1j}\{1 + o(1)\}].
$$

From known conditions, we get that $D_{n,2}$ is bounded by

$$
\sqrt{s} n \delta_n a_n \|\boldsymbol{T}_1\| + n\delta_n^2 b_n \|\boldsymbol{T}_1\|^2 = Cn\delta_n^2(\sqrt{s} + b_n C).
$$

As $b_n \to 0$, this is also dominated by the second term of (S1.8) in $\|\boldsymbol{T}_1\| = C$. Hence, (S1.6) holds for sufficiently large C.

Next, we show that

$$
\begin{aligned}
\| \hat{h}(\boldsymbol{Z}) - h_0(\boldsymbol{Z}) \|^2 &= \left\| (\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0)^\top \boldsymbol{B} \right\|^2 \\
&= (\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0)^\top E \left[n^{-1} \sum_{i=1}^{n} \mathbf{B}(\boldsymbol{Z}_i)\mathbf{B}^\top(\boldsymbol{Z}_i) \right] (\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0) \\
&\leq C \|\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0\|^2,
\end{aligned}
$$

thus $\| \hat{h}(\boldsymbol{Z}) - h_0(\boldsymbol{Z}) \|^2 = O_p(\delta_n^2)$.

$$
\begin{aligned}
\|\hat{\alpha}(t) - \alpha_0(t)\|^2 &= \| \sum_{j=1}^{m_n} \hat{\gamma}_j \hat{v}_j(t) - \sum_{j=1}^{\infty} \gamma_{0j} v_j(t) \|^2 \\
&\leq 2\| \sum_{j=1}^{m_n} \hat{\gamma}_j \hat{v}_j(t) - \sum_{j=1}^{m_n} \gamma_{0j} v_j(t) \|^2 + 2\| \sum_{j=m_n+1}^{\infty} \gamma_{0j} v_j(t) \|^2 \\
&\leq 4\| \sum_{j=1}^{m_n} (\hat{\gamma}_j - \gamma_{0j})\, \hat{v}_j(t) \|^2 + 4\| \sum_{j=1}^{m_n} \gamma_{0j}\, (\hat{v}_j(t) - v_j(t))\, \|^2 + 2 \sum_{j=m_n+1}^{\infty} {\gamma_{0j}}^2 \\
&\leq 4\|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\|^2 + 4\| \sum_{j=1}^{m_n} \gamma_{0j}\, (\hat{v}_j(t) - v_j(t))\, \|^2 + 2 \sum_{j=m_n+1}^{\infty} {\gamma_{0j}}^2 \\
&= 4\|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\|^2 + 4I_1 + 2I_2. \qquad \text{(S1.9)}
\end{aligned}
$$

By means of $\|\hat{v}_j - v_j\|^2 = O_p(n^{-1}j^2)$ and Cauchy-Shwartz inequality and condition C2, we obtain that

$$
\begin{aligned}
I_1 &\leq m_n \sum_{j=1}^{m_n} \|\hat{v}_j - v_j\|^2 {\gamma_{0j}}^2 \leq \frac{m_n}{n} \sum_{j=1}^{m_n} j^2 {\gamma_{0j}}^2 \\
&= O_p(n^{-1} m_n \sum_{j=1}^{m_n} j^{2-2b}) = O_p(n^{-1} m_n) \\
&= O_p(n^{-\frac{a+2b-1}{a+2b}}), \qquad \text{(S1.10)}
\end{aligned}
$$

$$I_2 \;\leq\; C \sum_{j=m_n+1}^{\infty} j^{-2b} = O(m_n^{-(2b-1)}) = O(n^{-\frac{2b-1}{a+2b}}). \tag{S1.11}$$

From (S1.9)-(S1.11), we can get that

$$\|\hat{\alpha}(t) - \alpha_0(t)\|^2 = O_p\left(\delta_n^2\right).$$

This completes the proof of the Theorem **??**. □

Proof of Theorem **??***(a).* We use proof by contradiction. Suppose that there exists a $s+1 \leq k_0 \leq p_n$ such that the probability of $h_{k_0}(z)$ being a zero function does not converge to one. Then, there exists $\eta > 0$ such that, for infinitely many n, $P(\hat{\boldsymbol{\zeta}}_{k_0} \neq \mathbf{0}) = P(\hat{h}_{k_0}(z) \neq 0) \geq \eta$. Let $\hat{\boldsymbol{\zeta}}^*$ be the vector obtained from $\hat{\boldsymbol{\zeta}}$ with $\hat{\boldsymbol{\zeta}}_{k_0}$ being replaced by $\mathbf{0}$. It will be shown that there exists a $\delta > 0$ such that $PL(\hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\gamma}}) - PL(\hat{\boldsymbol{\zeta}}^*, \hat{\boldsymbol{\gamma}}) < 0$ with probability at least δ for infinitely many n, which contradicts with the fact that $PL(\hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\gamma}}) - PL(\hat{\boldsymbol{\zeta}}^*, \hat{\boldsymbol{\gamma}}) \geq 0$.

By Theorem **??**, we have $\|\hat{\boldsymbol{\zeta}}_k - \boldsymbol{\zeta}_k\| = O_p(\sqrt{(m_n + K_n) n^{-(\frac{2r}{1+2r} - \frac{2b-1}{a+2b})}})$. Since $\boldsymbol{\zeta}_k = \mathbf{0}$ for $k = s+1, \cdots, p_n$, we have $\|\hat{\boldsymbol{\zeta}}_k\| = O_p(\sqrt{(m_n + K_n) n^{-(\frac{2r}{1+2r} - \frac{2b-1}{a+2b})}})$ for $k = s+1, \cdots, p_n$. So $\|\hat{\boldsymbol{\zeta}}_{k_0}\| = O_p(\sqrt{(m_n + K_n) n^{-(\frac{2r}{1+2r} - \frac{2b-1}{a+2b})}})$. With probability tending to one, $\|\hat{\boldsymbol{\zeta}}_{k_0}\| \leq \lambda_n$, since

$\lambda_n (n^{-(\frac{2r}{1+2r} - \frac{2b-1}{a+2b})})^{-\frac{1}{2}} / \sqrt{m_n + J_n} \to \infty$. By the definition of $p_{\lambda_n}(\cdot)$, we have $P\{p_{\lambda_n}(\|\hat{\boldsymbol{\zeta}}_{k_0}\|) =$

$\lambda_n\|\hat{\boldsymbol{\zeta}}_{k_0}\|\} \to 1$. Mimicking the proof for Theorem **??** indicates that

$$\begin{aligned}
& PL(\hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\gamma}}) - PL(\hat{\boldsymbol{\zeta}}^*, \hat{\boldsymbol{\gamma}}) \\
= & \sum_{i=1}^{n}[Q\{g^{-1}(\boldsymbol{B}^\top(\boldsymbol{Z}_i)\hat{\boldsymbol{\zeta}} + \hat{\boldsymbol{\xi}}_i^\top\hat{\boldsymbol{\gamma}}), Y_i\} - Q\{g^{-1}(\boldsymbol{B}^\top(\boldsymbol{Z}_i)\hat{\boldsymbol{\zeta}}^* + \hat{\boldsymbol{\xi}}_i^\top\hat{\boldsymbol{\gamma}}), Y_i\}] \\
& -n\sum_{k=1}^{s}\{p_{\lambda_n}(\|\hat{\boldsymbol{\zeta}}_k\|) - p_{\lambda_n}(\|\hat{\boldsymbol{\zeta}}_k^*\|)\} \\
= & \sum_{i=1}^{n} q_1(m_0(\boldsymbol{\mathcal{F}}_i), Y_i)(\hat{\boldsymbol{\zeta}} - \hat{\boldsymbol{\zeta}}^*) \\
& -\frac{n}{2}(\hat{\boldsymbol{\zeta}} - \hat{\boldsymbol{\zeta}}^*)^\top E(\rho_2(m_0(\boldsymbol{\mathcal{F}})))(\hat{\boldsymbol{\zeta}} - \hat{\boldsymbol{\zeta}}^*) - n\lambda_n\|\hat{\boldsymbol{\zeta}}_{k_0}\| + o_p(1) \\
= & O_p(n^{1/2}\|\hat{\boldsymbol{\zeta}}_{k_0}\|) + O_p(n\|\hat{\boldsymbol{\zeta}}_{k_0}\|^2) - n\lambda_n\|\hat{\boldsymbol{\zeta}}_{k_0}\| \\
= & n\lambda_n^2(-{\lambda_n}^{-1}\|\hat{\boldsymbol{\zeta}}_{k_0}\| + O_p(\frac{\|\hat{\boldsymbol{\zeta}}_{k_0}\|^2}{\lambda_n^2})) \\
= & n\lambda_n^2(-{\lambda_n}^{-1}\|\hat{\boldsymbol{\zeta}}_{k_0}\| + O_p(\frac{(m_n + K_n)(n^{-(\frac{2r}{1+2r} - \frac{2b-1}{a+2b})})}{\lambda_n^2})). \qquad \text{(S1.12)}
\end{aligned}$$

By the fact that $\lambda_n(n^{-(\frac{2r}{1+2r}-\frac{2b-1}{a+2b})})^{-\frac{1}{2}}/\sqrt{m_n + K_n} \to \infty$, we can conclude that (S1.12) is dominated by $\lambda_n^{-1}\|\hat{\boldsymbol{\zeta}}_{k_0}\|$, so $PL(\hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\gamma}}) - PL(\hat{\boldsymbol{\zeta}}^*, \hat{\boldsymbol{\gamma}}) < 0$, which contradicts to $PL(\hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\gamma}}) - PL(\hat{\boldsymbol{\zeta}}^*, \hat{\boldsymbol{\gamma}}) \geq 0$. The proof of the theorem **??** (a) is complete.

Now we prove the Theorem **??**(b). By Theorem **??**(a), we know that, with probability tending to one, it follows that $\hat{\boldsymbol{\zeta}}_b = \mathbf{0}$. We next establish the asymptotic normality of $\hat{\boldsymbol{\zeta}}_{\mathcal{S}}$. Let $\hat{\boldsymbol{\theta}} = \delta_n^{-1}(\hat{\boldsymbol{\zeta}}_{\mathcal{S}} - \boldsymbol{\zeta}_{0\mathcal{S}})$ and $\tilde{m}_0(\boldsymbol{\mathcal{T}}_{i\mathcal{S}}) = \boldsymbol{B}^\top(\boldsymbol{Z}_{i\mathcal{S}})\boldsymbol{\zeta}_{0\mathcal{S}} + \hat{\boldsymbol{\xi}}_i^\top\boldsymbol{\gamma}_0$. For any $\boldsymbol{v} \in R^{s_n}$, define $\tilde{m}_{\boldsymbol{v}} = \tilde{m}(\boldsymbol{\mathcal{T}}_{\mathcal{S}}) + \boldsymbol{v}^\top\tilde{\boldsymbol{B}}(\boldsymbol{Z}_{\mathcal{S}}) = \hat{\boldsymbol{\xi}}^\top\boldsymbol{\gamma} - \boldsymbol{v}^\top\boldsymbol{\Gamma}_1(\hat{\boldsymbol{\xi}}) + (\hat{\boldsymbol{\zeta}}_{\mathcal{S}} + \boldsymbol{v})^\top\boldsymbol{B}(\boldsymbol{Z}_{\mathcal{S}})$. Note that $\tilde{m}_{\boldsymbol{v}}$ maximizes the function $l_n(\tilde{m}) = \frac{1}{n}\sum_{i=1}^{n} Q[g^{-1}(\tilde{m}(\boldsymbol{\mathcal{T}}_{i\mathcal{S}})), Y_i] - \sum_{j=1}^{s} p_{\lambda_n}(\|\hat{\boldsymbol{\zeta}}_j + v_j\|)$

when $\boldsymbol{v} = \mathbf{0}$, thus we have

$$
\begin{aligned}
\mathbf{0} &= \frac{\partial}{\partial \boldsymbol{v}} l_n(\tilde{m}_{\boldsymbol{v}})|_{\boldsymbol{v}=\mathbf{0}} \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[q_1\left\{\tilde{m}(\boldsymbol{\mathcal{T}}_{i\mathcal{S}}), Y_i\right\}\tilde{\boldsymbol{B}}(\boldsymbol{Z}_{i\mathcal{S}})\right] \\
&\quad -\left\{p'_{\lambda_n}\left(\left\|\hat{\boldsymbol{\zeta}}_j\right\|\right)\frac{\hat{\boldsymbol{\zeta}}_j^{\top}}{\left\|\hat{\boldsymbol{\zeta}}_j\right\|}\right\}_{j=1}^{s} + o_p(1).
\end{aligned} \tag{S1.13}
$$

For the first term in (S1.13), we get

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left[q_1\left\{\tilde{m}(\boldsymbol{\mathcal{T}}_{i\mathcal{S}}), Y_i\right\}\tilde{\boldsymbol{B}}(\boldsymbol{Z}_{i\mathcal{S}})\right] \\
=\ &\frac{1}{n}\sum_{i=1}^{n}\left[q_1\left\{\tilde{m}_0(\boldsymbol{\mathcal{T}}_{i\mathcal{S}}), Y_i\right\}\tilde{\boldsymbol{B}}(\boldsymbol{Z}_{i\mathcal{S}})\right] \\
&+\frac{1}{n}\sum_{i=1}^{n}\left[q_2\left\{\tilde{m}_0(\boldsymbol{\mathcal{T}}_{i\mathcal{S}}), Y_i\right\}(\tilde{m}(\boldsymbol{\mathcal{T}}_{i\mathcal{S}}) - \tilde{m}_0(\boldsymbol{\mathcal{T}}_{i\mathcal{S}}))\tilde{\boldsymbol{B}}(\boldsymbol{Z}_{i\mathcal{S}})\right] \\
&+\frac{1}{n}\sum_{i=1}^{n}\left[q_2'(\bar{m}(\boldsymbol{\mathcal{T}}_{i\mathcal{S}}), Y_i)(\tilde{m}(\boldsymbol{\mathcal{T}}_{i\mathcal{S}}) - \tilde{m}_0(\boldsymbol{\mathcal{T}}_{i\mathcal{S}}))^2\tilde{\boldsymbol{B}}(\boldsymbol{Z}_{i\mathcal{S}})\right] \\
=\ & I_1 + I_2 + I_3.
\end{aligned} \tag{S1.14}
$$

We decompose I_2 into two terms I_{21} and I_{22} as follows:

$$
\begin{aligned}
I_2 &= \frac{1}{n}\sum_{i=1}^{n} q_2\left\{\tilde{m}_0(\boldsymbol{\mathcal{T}}_{i\mathcal{S}}), Y_i\right\}\tilde{\boldsymbol{B}}(\boldsymbol{Z}_{i\mathcal{S}})\hat{\boldsymbol{\xi}}_i^{\top}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) \\
&\quad +\frac{1}{n}\sum_{i=1}^{n} q_2\left\{\tilde{m}_0(\boldsymbol{\mathcal{T}}_{i\mathcal{S}}), Y_i\right\}\tilde{\boldsymbol{B}}(\boldsymbol{Z}_{i\mathcal{S}})\boldsymbol{B}^{\top}(\boldsymbol{Z}_{i\mathcal{S}})(\hat{\boldsymbol{\zeta}}_{\mathcal{S}} - \boldsymbol{\zeta}_{0\mathcal{S}}) \\
&= I_{21} + I_{22}.
\end{aligned}
$$

Mimicking the proof for (S1.7) indicates that

$$\begin{aligned} I_{21} &= -\frac{1}{n}\sum_{i=1}^{n}\rho_2\left\{\tilde{m}_0(\boldsymbol{\mathcal{T}}_{i\mathcal{S}})\right\}\tilde{\boldsymbol{B}}(\boldsymbol{Z}_{i\mathcal{S}})\hat{\boldsymbol{\xi}}_i^{\top}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0)+o_p(1) \\ &= -E[\rho_2\left\{\tilde{m}_0(\boldsymbol{\mathcal{T}}_{\mathcal{S}})\right\}\tilde{\boldsymbol{B}}(\boldsymbol{Z}_{\mathcal{S}})\hat{\boldsymbol{\xi}}^{\top}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0)]+o_p(1). \end{aligned}$$

By the definition of $\tilde{\boldsymbol{B}}(\boldsymbol{Z}_{\mathcal{S}})$, for any measurable function ϕ, $E[\phi(\hat{\boldsymbol{\xi}})\rho_2\left\{\tilde{m}_0(\boldsymbol{\mathcal{T}}_{\mathcal{S}})\right\}\tilde{\boldsymbol{B}}(\boldsymbol{Z}_{\mathcal{S}})]=\mathbf{0}$. Hence $I_{21}=o_p(1)$. Using similar arguments, we can show that

$$\begin{aligned} I_{22} &= -\frac{1}{n}\sum_{i=1}^{n}\rho_2\left\{\tilde{m}_0(\boldsymbol{\mathcal{T}}_{i\mathcal{S}}),Y_i\right\}\tilde{\boldsymbol{B}}(\boldsymbol{Z}_{i\mathcal{S}})\boldsymbol{B}^{\top}(\boldsymbol{Z}_{i\mathcal{S}})(\hat{\boldsymbol{\zeta}}_{\mathcal{S}}-\boldsymbol{\zeta}_{0\mathcal{S}})+o_p(1) \\ &= -\frac{1}{n}\sum_{i=1}^{n}\rho_2\left\{\tilde{m}_0(\boldsymbol{\mathcal{T}}_{i\mathcal{S}}),Y_i\right\}\tilde{\boldsymbol{B}}(\boldsymbol{Z}_{i\mathcal{S}})\tilde{\boldsymbol{B}}^{\top}(\boldsymbol{Z}_{i\mathcal{S}})(\hat{\boldsymbol{\zeta}}_{\mathcal{S}}-\boldsymbol{\zeta}_{0\mathcal{S}})+o_p(1). \end{aligned}$$

According to (S1.6) and condition (C6), we have

$$\begin{aligned} I_3 &= \frac{1}{n}\sum_{i=1}^{n}\left[q_2'(\bar{m}(\boldsymbol{\mathcal{T}}_{i\mathcal{S}}),Y_i)(\tilde{m}(\boldsymbol{\mathcal{T}}_{i\mathcal{S}})-\tilde{m}_0(\boldsymbol{\mathcal{T}}_{i\mathcal{S}}))^2\tilde{\boldsymbol{B}}(\boldsymbol{Z}_{i\mathcal{S}})\right] \\ &\leq C\parallel\tilde{m}(\boldsymbol{\mathcal{T}}_{\mathcal{S}}))-\tilde{m}_0(\boldsymbol{\mathcal{T}}_{\mathcal{S}}))\parallel^2=O_p(\delta_n^2). \end{aligned}$$

For the second term in (S1.13), we get

$$\begin{aligned} &\left\{p'_{\lambda_n}\left(\left\|\hat{\boldsymbol{\zeta}}_j\right\|\right)\frac{\hat{\boldsymbol{\zeta}}_j^{\top}}{\left\|\hat{\boldsymbol{\zeta}}_j\right\|}\right\}_{j=1}^{s} \\ &= \left\{p'_{\lambda_n}\left(\|\boldsymbol{\zeta}_{0j}\|\right)\frac{\boldsymbol{\zeta}_{0j}^{\top}}{\|\boldsymbol{\zeta}_{0j}\|}\right\}_{j=1}^{s}+\left(\sum_{j=1}^{s}p''_{\lambda_n}(\|\boldsymbol{\zeta}_{0j}\|)+o_p(1)\right)(\hat{\boldsymbol{\zeta}}_j-\boldsymbol{\zeta}_{0j}). \end{aligned} \quad \text{(S1.15)}$$

Combining (S1.13), (S1.14) and (S1.15), we have

$$\begin{aligned} \mathbf{0} &= \frac{1}{n}\sum_{i=1}^{n}\left[q_1\left\{\tilde{m}_0(\boldsymbol{\mathcal{T}}_{i\mathcal{S}}),Y_i\right\}\tilde{\boldsymbol{B}}(\boldsymbol{Z}_{i\mathcal{S}})\right]-\left\{p'_{\lambda_n}\left(\|\boldsymbol{\zeta}_{0j}\|\right)\frac{\boldsymbol{\zeta}_{0j}^{\top}}{\|\boldsymbol{\zeta}_{0j}\|}\right\}_{j=1}^{s} \\ &\quad -\{E[\rho_2\left(\tilde{m}_0(\boldsymbol{\mathcal{T}}_{\mathcal{S}})\right)\tilde{\boldsymbol{B}}(\boldsymbol{Z}_{\mathcal{S}})\tilde{\boldsymbol{B}}^{\top}(\boldsymbol{Z}_{\mathcal{S}})]+o_p(1)\}(\hat{\boldsymbol{\zeta}}_{\mathcal{S}}-\boldsymbol{\zeta}_{0\mathcal{S}}) \\ &\quad -\left(\sum_{j=1}^{s}p''_{\lambda_n}(\|\boldsymbol{\zeta}_{0j}\|)+o_p(1)\right)(\hat{\boldsymbol{\zeta}}_j-\boldsymbol{\zeta}_{0j})+O_p(\delta_n^2). \end{aligned}$$

Therefore,

$$\begin{aligned}\mathbf{0} &= \frac{1}{n}\sum_{i=1}^{n}\left[q_1\left\{\tilde{m}_0(\boldsymbol{\mathcal{T}}_{i\mathcal{S}}), Y_i\right\}\tilde{\boldsymbol{B}}(\boldsymbol{Z}_{i\mathcal{S}})\right] - \boldsymbol{c}_n \\ &\quad -(\boldsymbol{\Sigma}_{\lambda_n} + \Omega_{\mathcal{S}} + o_p(1))(\hat{\boldsymbol{\zeta}}_{\mathcal{S}} - \boldsymbol{\zeta}_{0\mathcal{S}}) + O_p(\delta_n^2).\end{aligned}$$

Note that $q_1\left\{\tilde{m}_0(\boldsymbol{\mathcal{T}}_{i\mathcal{S}}), Y_i\right\} = [Y - g^{-1}(\tilde{m}_0(\boldsymbol{\mathcal{T}}_{i\mathcal{S}}))]\rho_1(\tilde{m}_0(\boldsymbol{\mathcal{T}}_{i\mathcal{S}}))$ and

$$\begin{aligned}&E[\rho_1^2(\tilde{m}_0(\boldsymbol{\mathcal{T}}_{\mathcal{S}}))[Y - g^{-1}(\tilde{m}_0(\boldsymbol{\mathcal{T}}_{\mathcal{S}}))]^2\tilde{\boldsymbol{B}}(\boldsymbol{Z}_{\mathcal{S}})\tilde{\boldsymbol{B}}^{\top}(\boldsymbol{Z}_{\mathcal{S}})] \\ &= E[E([Y - g^{-1}(\tilde{m}_0(\boldsymbol{\mathcal{T}}_{\mathcal{S}}))]^2|\boldsymbol{\mathcal{T}}_{\mathcal{S}})\rho_1^2(\tilde{m}_0(\boldsymbol{\mathcal{T}}_{\mathcal{S}}))\tilde{\boldsymbol{B}}(\boldsymbol{Z}_{\mathcal{S}})\tilde{\boldsymbol{B}}^{\top}(\boldsymbol{Z}_{\mathcal{S}})] \\ &= E[\rho_2(\tilde{m}_0(\boldsymbol{\mathcal{T}}_{\mathcal{S}}))\tilde{\boldsymbol{B}}(\boldsymbol{Z}_{\mathcal{S}})\tilde{\boldsymbol{B}}^{\top}(\boldsymbol{Z}_{\mathcal{S}})] = E(\Omega_{\mathcal{S}}). \qquad \text{(S1.16)}\end{aligned}$$

Thus

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[q_1\left\{\tilde{m}_0(\boldsymbol{\mathcal{T}}_{i\mathcal{S}}), Y_i\right\}\tilde{\boldsymbol{B}}(\boldsymbol{Z}_{i\mathcal{S}})\right] \\ &= \sqrt{n}(\boldsymbol{\Sigma}_{\lambda_n} + \Omega_{\mathcal{S}})(\hat{\boldsymbol{\zeta}}_{\mathcal{S}} - \boldsymbol{\zeta}_{0\mathcal{S}}) + \sqrt{n}\boldsymbol{c}_n + o_p(1).\end{aligned}$$

By the Lindeberg-Feller central limit theorem, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[q_1\left\{\tilde{m}_0(\boldsymbol{\mathcal{T}}_{i\mathcal{S}}), Y_i\right\}\tilde{\boldsymbol{B}}(\boldsymbol{Z}_{i\mathcal{S}})\right] \xrightarrow{d} N(0, E[q_1^2\left\{\tilde{m}_0(\boldsymbol{\mathcal{T}}_{\mathcal{S}}), Y\right\}\tilde{\boldsymbol{B}}(\boldsymbol{Z}_{\mathcal{S}})\tilde{\boldsymbol{B}}^{\top}(\boldsymbol{Z}_{\mathcal{S}})]).$$

According to (S1.16), we have

$$\sqrt{n}(\boldsymbol{\Sigma}_{\lambda_n} + \Omega_{\mathcal{S}})\{\hat{\boldsymbol{\zeta}}_{\mathcal{S}} - \boldsymbol{\zeta}_{0\mathcal{S}} + (\boldsymbol{\Sigma}_{\lambda_n} + \Omega_{\mathcal{S}})^{-1}\boldsymbol{c}_n\} \xrightarrow{d} N(0, E(\Omega_{\mathcal{S}})).$$

The proof of the Theorem **??**(b) is complete. □

References

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

Kong, D. H., K. J. Xue, F. Yao, and H. H. Zhang (2016). Partially functional linear regression in high dimensions. *Biometrika 103*, 147–159.

van Der Vaart, A. W. (1998). Asymptotic statistics.