AN RKHS APPROACH FOR PIVOTAL INference
IN FUNCTIONAL LINEAR REGRESSION

Holger Dette and Jiajun Tang

Supplementary Material

Section S1 contains the proofs of the theorems in our main article, and Section S2 contains the supporting lemmas.

S1. Theoretical details of main results

S1.1 Proof of Theorem 1

In the sequel, we use $c$ to denote a generic positive constant that might differ from line to line. We first prove in Lemma 1 below the uniform convergence rate of the sequential RKHS estimator $\hat{\beta}_{n,\lambda}(\cdot, \nu)$ for the slope function $\beta_0$ defined in (3.2) w.r.t. the $\| \cdot \|_K$-norm.

Lemma 1. Under Assumptions 1–5, we have, for any fixed (but arbitrary) $\nu_0 \in (0, 1]$,

$$\sup_{\nu \in [\nu_0, 1]} \| \hat{\beta}_{n,\lambda}(\cdot, \nu) - \beta_0 \|_K = O_p(\lambda^{1/2} + n^{-1/2} \lambda^{-(2a+1)/(4D)}) .$$

Proof. Define $S_{\lambda,\nu}(\beta) = \text{E}\{S_{n,\lambda,\nu}(\beta)\}$ and $\mathcal{D}S_{\lambda,\nu}(\beta) = \text{E}\{\mathcal{D}S_{n,\lambda,\nu}(\beta)\}$. In view of (4.7),

$$S_{\lambda,\nu}(\beta) = \text{E}\{S_{n,\lambda,\nu}(\beta)\} = -\text{E}\left\{ Y_0 - \int_0^1 X_0(s) \beta(s) \, ds \right\} \tau_\lambda(X_0) + W_\lambda(\beta),$$

$$\mathcal{D}S_{\lambda,\nu}(\beta)\beta_1 = \text{E}\{\mathcal{D}S_{n,\lambda,\nu}(\beta)\beta_1\} = \text{E}\left\{ \int_0^1 X_0(s) \beta(s) \, ds \right\} \tau_\lambda(X_0) + W_\lambda(\beta).$$

Recall from (4.7) that, for $\nu \in [0, 1]$ and for any $\beta_1, \beta_2 \in \mathcal{H},$

$$\langle \mathcal{D}S_{\lambda,\nu}(\beta)\beta_1, \beta_2 \rangle_K = \text{E}\{ (\tau_\lambda(X_1), \beta_1)_K (\tau_\lambda(X_1), \beta_2)_K \} + \langle W_\lambda(\beta_1), \beta_2 \rangle_K$$

$$= \text{E}\{ (X_i, \beta_1)_{L^2} (X_i, \beta_2)_{L^2} \} + \langle W_\lambda(\beta_1), \beta_2 \rangle_K = V(\beta_1, \beta_2) + \lambda J(\beta_1, \beta_2) = \langle \beta_1, \beta_2 \rangle_K = \langle \text{id}(\beta_1), \beta_2 \rangle_K ,$$
which implies that $DS_{\lambda,\nu}(\beta) = id$ where $id$ denotes the identity operator on $\mathcal{H}$. Since $D^2 S_{\lambda,\nu}$ vanishes, there exists a unique solution to the estimating equation $S_{\lambda,\nu}(\beta) = 0$. In addition, by the mean value theorem, for any $\beta \in \mathcal{H}$, $S_{\lambda,\nu}(\beta) = S_{\lambda,\nu}(\beta_0) + D S_{\lambda,\nu}(\beta - \beta_0) = S_{\lambda,\nu}(\beta_0) + (\beta - \beta_0)$. Let $\beta_{\lambda,\nu} = \beta_0 - S_{\lambda,\nu}(\beta_0)$. We deduce that $S_{\lambda,\nu}(\beta_{\lambda,\nu}) = S_{\lambda,\nu}(\beta_0) + (\beta_{\lambda,\nu} - \beta_0) = 0$, so that $\beta_{\lambda,\nu}$ is the unique solution to the estimating equation $S_{\lambda,\nu}(\beta) = 0$. Moreover, since $E[\{Y_0 - \int_0^1 X_0(s)\beta_0(s)ds\} \tau_\lambda(X_0)] = 0$, in view of (S1.1), for any $\nu \in [\nu_0, 1]$,

$$\|\beta_{\lambda,\nu} - \beta_0\|_K = \|S_{\lambda,\nu}(\beta_0)\|_K = \|W_\lambda(\beta_0)\|_K. \tag{S1.2}$$

Therefore, by the Cauchy-Schwarz inequality, we deduce that

$$\sup_{\nu \in [\nu_0, 1]} \|\beta_{\lambda,\nu} - \beta_0\|_K = \|W_\lambda(\beta_0)\|_K = \sup_{\|\gamma\|_K = 1} |\langle W_\lambda(\beta_0), \gamma \rangle_K| = \sup_{\|\gamma\|_K = 1} \lambda |J(\beta_0, \gamma)| \leq \sup_{\|\gamma\|_K = 1} \left\{ \sqrt{\lambda J(\beta_0, \beta_0)} \sqrt{\lambda J(\gamma, \gamma)} \right\} \leq \sup_{\|\gamma\|_K = 1} \left\{ \sqrt{\lambda J(\beta_0, \beta_0)} \|\gamma\|_K \right\} = O(\lambda^{1/2}). \tag{S1.3}$$

Since $\|\hat{\beta}_{n,\lambda}(\cdot, \nu) - \beta_0\|_K \leq \|\beta_{\lambda,\nu} - \beta_0\|_K + \|\hat{\beta}_{n,\lambda}(\cdot, \nu) - \beta_{\lambda,\nu}\|_K$, we then proceed to show the rate of $\|\hat{\beta}_{n,\lambda}(\cdot, \nu) - \beta_{\lambda,\nu}\|_K$. For $\nu \in [\nu_0, 1]$, let $F_{n,\nu}(\beta) = \beta - S_{n,\lambda,\nu}(\beta_{\lambda,\nu} + \beta)$. Observing that $DS_{\lambda,\nu}(\beta) = id$, since $D^2 S_{\lambda,\nu}$ vanishes, we obtain

$$F_{n,\nu}(\beta) = DS_{\lambda,\nu}(\beta_{\lambda,\nu})\beta - S_{n,\lambda,\nu}(\beta_{\lambda,\nu} + \beta) = I_{1,n,\nu}(\beta) + I_{2,n,\nu}(\beta) - S_{n,\lambda,\nu}(\beta_{\lambda,\nu}), \tag{S1.4}$$

where $DS_{n,\lambda,\nu}$ is defined in (4.7) and

$$I_{1,n,\nu}(\beta) = -\{S_{n,\lambda,\nu}(\beta_{\lambda,\nu} + \beta) - S_{n,\lambda,\nu}(\beta_{\lambda,\nu}) - DS_{n,\lambda,\nu}(\beta_{\lambda,\nu})\},$$

$$I_{2,n,\nu}(\beta) = -\{DS_{n,\lambda,\nu}(\beta_{\lambda,\nu})\beta - DS_{\lambda,\nu}(\beta_{\lambda,\nu})\}. \tag{S1.5}$$

First, for $I_{1,n,\nu}(\beta)$ in (S1.5), in view of $S_{n,\lambda,\nu}$ and $DS_{n,\lambda,\nu}$ defined in (4.7), we find

$$I_{1,n,\nu}(\beta) = \frac{1}{n\nu} \sum_{i=1}^{n\nu} \left[ Y_i - \int_0^1 \{\beta_{\lambda,\nu}(s) + \beta(s)\} X_i(s) ds \right] \tau_\lambda(X_i)$$
For the second term $I_{2,n,\nu}(\beta)$ in (S1.5), define the event $E_n(c) = \{ \max_{1 \leq i \leq n} \|X_i\|_{L^2} \leq c \log n \}$. By Assumption 4 and Markov’s inequality, if we take $c > 3/\varpi > 0$, we have $P\{E_n(c)\} \leq n P(\|X_0\|_{L^2} \leq c \log n) \leq n^{1-c\varpi} E\{\exp(\varpi \|X_0\|_{L^2})\} = o(n^{-2})$. Then, it suffices to confine the proof on the event $E_n(c)$. In view of (S1.7) and (S1.1),

$$I_{2,n,\nu}(\beta) = DS_{n,\nu}(\beta_{\lambda,\nu}) - DS_{\nu}(\beta_{\lambda,\nu}) = I_{2,1,n,\nu}(\beta) + I_{2,2,n,\nu}(\beta), \quad \text{(S1.7)}$$

where

$$I_{2,1,n,\nu}(\beta) = E \left[ \tau_\lambda(X_i) \int_0^1 \beta(s)X_i(s)ds \times 1\{E^n(c)\} \right], \quad \text{(S1.8)}$$

$$I_{2,2,n,\nu}(\beta) = -\frac{1}{[n\nu]} \sum_{i=1}^{[n\nu]} \left( \tau_\lambda(X_i) \int_0^1 \beta(s)X_i(s)ds \times 1\{E^n(c)\} - E \left[ \tau_\lambda(X_i) \int_0^1 \beta(s)X_i(s)ds \times 1\{E^n(c)\} \right] \right).$$

For the first term $I_{2,1,n,\nu}$ in (S1.7), by the Cauchy-Schwarz inequality and Lemma 4,

$$\|I_{2,1,n,\nu}(\beta)\|_K = \left\| E \left[ \tau_\lambda(X_0) \int_0^1 \beta(s)X_0(s)ds \times 1\{E^n(c)\} \right] \right\|_K \leq \sup_{\|\gamma\|_K=1} \left\langle \gamma, E \left[ \tau_\lambda(X_0) \int_0^1 \beta(s)X_0(s)ds \times 1\{E^n(c)\} \right] \right\rangle_K \leq \left[ P\{E^n(c)\} \right]^{1/2} E\left( \langle X_0, \beta \rangle_{L^2} \right)^{1/4} \sup_{\|\gamma\|_K=1} E\left( \langle X_0, \gamma \rangle_{L^2} \right)^{1/4} \leq c \left[ P\{E^n(c)\} \right]^{1/2} E\left( \langle X_0, \beta \rangle_{L^2} \right)^{1/2} \sup_{\|\gamma\|_K=1} E\left( \langle X_0, \gamma \rangle_{L^2} \right)^{1/2} = c \left[ P\{E^n(c)\} \right]^{1/2} E\left( \langle \tau_\lambda(X_0), \beta \rangle_K \right)^{1/2} \sup_{\|\gamma\|_K=1} E\left( \langle \tau_\lambda(X_0), \gamma \rangle_K \right)^{1/2} \leq c \left[ P\{E^n(c)\} \right]^{1/2} \|\tau_\lambda(X_0)\|_K^2 \|\beta\|_K \leq o(n^{-1}\lambda^{-1/(2D)}) \|\beta\|_K = o(1) \|\beta\|_K. \quad \text{(S1.9)}$$

Therefore, we deduce that

$$\sup_{\nu \in [\nu_0, 1]} \|I_{2,1,n,\nu}(\beta)\|_K = o(1) \|\beta\|_K. \quad \text{(S1.10)}$$
For the second term $I_{2,2,n,\nu}(\beta)$ in (S1.7), for $a, D$ in Assumption 2 and $c_K$ in Lemma 5 in Section S2 let $p_n = c_K^{-2} \lambda^{(2a+1)/(2D) - 1}$. In order to apply Lemma 5 in Section S2, we shall rescale $\beta$ such that the $L^2$-norm of its rescaled version is bounded by 1, that is
\[
\tilde{\beta} = \begin{cases} 
(c_K \lambda^{-(2a+1)/(4D)} \|\beta\|_K)^{-1} \beta & \text{if } \beta \neq 0, \\
0 & \text{if } \beta = 0
\end{cases} \tag{S1.11}
\]
where $c_K$ is the constant in Lemma 5. We have $\|\tilde{\beta}\|_{L^2} \leq c_K \lambda^{-(2a+1)/(4D)} \|\beta\|_K \leq 1$, since $\|\tilde{\beta}\|_K \leq (c_K \lambda^{-(2a+1)/(4D)})^{-1}$ in view of Lemma 5. In addition, observing (4.2), it follows that
\[
J(\tilde{\beta}, \beta) \leq \lambda^{-1} \|\tilde{\beta}\|^2_K \leq c_K^{-2} \lambda^{(2a+1)/(2D) - 1} = p_n.
\]
Therefore,
\[
\tilde{\beta} \in \mathcal{F}_{p_n} := \{\beta \in \mathcal{H} : \|\beta\|_{L^2} \leq 1, J(\beta, \beta) \leq p_n\}. \tag{S1.12}
\]
For the event $\mathcal{E}_n(\nu)$ defined below equation (S1.6) and for any $\beta \in \mathcal{H}$, let
\[
\tilde{H}_{n,\nu}(\beta) = \frac{1}{\sqrt{|\nu| n}} \sum_{i=1}^{|\nu|} \left( \tau_{\lambda}(X_i) \langle \beta, X_i \rangle_{L^2} 1\{\mathcal{E}_n(\nu)\} - \mathbb{E} \left[ \tau_{\lambda}(X_i) \langle \beta, X_i \rangle_{L^2} 1\{\mathcal{E}_n(\nu)\} \right] \right). \tag{S1.13}
\]
Note that, for $\nu \in [\nu_0, 1]$,
\[
\sup_{\beta \in \mathcal{F}_{p_n}} \|\tilde{H}_{n,\nu}(\beta)\|_K = \frac{1}{\sqrt{|\nu| n}} \sup_{\beta \in \mathcal{F}_{p_n}} \|H_{n,|\nu|}(\beta)\|_K \leq c\nu_0^{-1/2} \max_{1 \leq k \leq n} \sup_{\beta \in \mathcal{F}_{p_n}} \|H_{n,k}(\beta)\|_K, \tag{S1.14}
\]
where, for the $\mathcal{E}_n(\nu)$ defined below equation (S1.6), $H_{n,k}$ is defined by
\[
H_{n,k}(\beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^k \left( \tau_{\lambda}(X_i) \int_0^1 \beta(s) X_i(s) ds 1\{\mathcal{E}_n(\nu)\} - \mathbb{E} \left[ \tau_{\lambda}(X_i) \int_0^1 \beta(s) X_i(s) ds 1\{\mathcal{E}_n(\nu)\} \right] \right). \tag{S1.15}
\]
Therefore, observing that $\nu n^{-1/2} = o(p_n^{1/(2m)})$ by Assumption 5 combining (S1.14) and Lemma 8 yields with probability tending to one,
\[
\sup_{\nu \in [\nu_0, 1]} \sup_{\beta \in \mathcal{F}_{p_n}} \|\tilde{H}_{n,\nu}(\beta)\|_K \leq c (p_n^{1/(2m) + n^{-1/2}} \lambda^{-1/(2D)} \log n)^{1/2} \leq c p_n^{1/(2m)} \lambda^{-1/(4D)} (\log n)^{1/2},
\]
where $c > 0$ depends on $\nu_0$. In view of (S1.11), we deduce from the above equation that, for the $\beta$ in (S1.11), with probability tending to one,
\[
\sup_{\nu \in [\nu_0, 1]} \|\tilde{H}_{n,\nu}(\beta)\|_K \leq (c_K \lambda^{-(2a+1)/(4D)} \|\beta\|_K) \sup_{\nu \in [\nu_0, 1]} \sup_{\beta \in \mathcal{F}_{p_n}} \|\tilde{H}_{n,\nu}(\beta)\|_K.
\]
\[ \leq c \, n^{1/2} \lambda^{-(a+1)/(2D)} (\log n)^{1/2} \| \beta \|_K. \]

Observing that \( p_n = O(\lambda^{(2a+1)/(2D)-1}) \) and (S1.8), we have with probability tending to one,

\[ \sup_{\nu \in [\nu_0, 1]} \| I_{2, n, \nu}(\beta) \|_K \leq n^{-1/2} \sup_{\nu \in [\nu_0, 1]} \| \widetilde{H}_{n, \nu}(\beta) \|_K \leq c \, n^{-1/2} p_n^{1/(2m)} \lambda^{-(a+1)/(2D)} (\log n)^{1/2} \| \beta \|_K \]

\[ \leq c \, n^{-1/2} \lambda^{-\epsilon} (\log n)^{1/2} \| \beta \|_K = o(1) \| \beta \|_K, \tag{S1.16} \]

where we used Assumption 5 in the last step. Therefore, combining (S1.7), (S1.10) and (S1.16) yields that, as \( n \to \infty \),

\[ \sup_{\nu \in [\nu_0, 1]} \| I_{2, n, \nu}(\beta) \|_K = o(1) \| \beta \|_K. \tag{S1.17} \]

We now consider the term \( -S_{n, \lambda, \nu}(\beta_{\lambda, \nu}) \) in (S1.4). Recalling the definition of \( \tau_{\lambda} \) in (3.8) and observing that \( S_{\lambda, \nu}(\beta_{\lambda, \nu}) = 0 \) and \( E\{\varepsilon_0 \tau_{\lambda}(X_0)\} = 0 \), in view of (4.7), we find

\[ -S_{n, \lambda, \nu}(\beta_{\lambda, \nu}) = -\{S_{n, \lambda, \nu}(\beta_{\lambda, \nu}) - S_{\lambda, \nu}(\beta_{\lambda, \nu})\} = \frac{1}{[n\nu]} \sum_{i=1}^{[n\nu]} \varepsilon_i \tau_{\lambda}(X_i) - I_{2, n, \nu}(\beta_0 - \beta_{\lambda, \nu}), \]

where \( I_{2, n, \nu} \) is defined in (S1.7). We deduce from the above equation and (S1.17) that

\[ \sup_{\nu \in [\nu_0, 1]} \| S_{n, \lambda, \nu}(\beta_{\lambda, \nu}) \|_K^2 \leq 2 \sup_{\nu \in [\nu_0, 1]} \left\| \frac{1}{[n\nu]} \sum_{i=1}^{[n\nu]} \varepsilon_i \tau_{\lambda}(X_i) \right\|_K^2 + o(1) \| \beta_0 - \beta_{\lambda, \nu} \|_K^2. \tag{S1.18} \]

For the first term in (S1.18), by direct calculations, we find

\[ \left\| \frac{1}{[n\nu]} \sum_{i=1}^{[n\nu]} \tau_{\lambda}(X_i) \varepsilon_i \right\|_K^2 = \frac{1}{[n\nu]^2} \sum_{i=1}^{[n\nu]} \sum_{i=1}^{[n\nu]} \sum_{k=1}^{\infty} \sum_{\ell=1}^{\infty} \left\langle \frac{\varepsilon_i \varepsilon_{i_1} X_{i_1, \varphi_{k}} L^2}{1 + \lambda \rho_k} \varphi_{k, \lambda} \varphi_{k}, \frac{\varepsilon_{i_2} X_{i_2, \varphi_{k}} L^2}{1 + \lambda \rho_{\ell}} \varphi_{\ell} \right\rangle_K \]

\[ = \frac{1}{[n\nu]^2} \sum_{i=1}^{[n\nu]} \sum_{i_1=1}^{[n\nu]} \sum_{k=1}^{\infty} \sum_{\ell=1}^{\infty} \frac{1}{1 + \lambda \rho_k} \left\langle \varepsilon_i X_{i_1, \varphi_{k}} L^2 \varepsilon_{i_2} X_{i_2, \varphi_{k}} L^2 \right\rangle \sum_{k=1}^{\infty} \frac{1}{1 + \lambda \rho_k} \frac{1}{[n\nu]} \sum_{i=1}^{[n\nu]} \left\langle \varepsilon_i X_i, \varphi_k \right\rangle L^2 \]

\[ = \frac{1}{[n\nu]} \sum_{k=1}^{\infty} \sum_{i=1}^{[n\nu]} \frac{1}{1 + \lambda \rho_k} \left\langle \frac{1}{[n\nu]} \sum_{i=1}^{[n\nu]} \varepsilon_i X_i \varphi_k \right\| L^2 \right\|^2 \leq \frac{1}{[n\nu]} \left\| \frac{1}{[n\nu]} \sum_{i=1}^{[n\nu]} \varepsilon_i X_i \right\| L^2 \sum_{k=1}^{\infty} \frac{1}{1 + \lambda \rho_k} \frac{1}{[n\nu]} \sum_{i=1}^{[n\nu]} \left\| \varphi_k \right\| L^2. \tag{S1.19} \]

Denote the long-run covariance function \( C_{X_\varepsilon}(s, t) = \sum_{\ell=\infty}^{\infty} \text{cov}\{\varepsilon_0 X_0(s), \varepsilon_\ell X_\ell(t)\} \). Observing Lemmas 6 and 7, we have that \( C_{X_\varepsilon} \in L^2([0, 1]^2) \) and \( \int_0^1 C_{X_\varepsilon}(s, s) ds < \infty \). By Assumption 1, we have that \( C_{X_\varepsilon} \) is positive definite. Let \( \{\xi_j\}_{j=1}^{\infty} \) and \( \{\psi_j\}_{j=1}^{\infty} \) denote the eigenvalues
and the corresponding eigenfunctions of the covariance kernel $C_{X\varepsilon}$, such that $\sum_{j=1}^{\infty} \tilde{\zeta}_j < \infty$.

In addition, since the $X_i$’s and the $\varepsilon_i$’s are independent, it is easy to see that the series $\{\varepsilon_i X_i\}_{i \in \mathbb{Z}}$ is $m$-approximable by $\{\varepsilon_{i,\ell} X_i\}_{i,\ell \in \mathbb{Z}}$. By Theorem 1.1 in [Berkes, Horváth, and Rice (2013)], there exists a Gaussian process $\{\Gamma_{X\varepsilon}(s, \nu)\}_{s \in [0,1], \nu \in [0,1]}$ in $\mathcal{F}$ defined in (4.11), given by $\Gamma_{X\varepsilon}(s, \nu) = \sum_{j=1}^{\infty} \tilde{\zeta}_j^{1/2} W_j(\nu) \tilde{\psi}_j(s)$, such that $\sup_{\nu \in [0,1]} \|\Gamma_{X\varepsilon}(\cdot, \nu) - n^{-1/2} \sum_{i=1}^{[n\nu]} \varepsilon_i X_i\|_{L^2} = o_p(1)$. Here, $\{W_j\}_{j=1}^{\infty}$ is a series of i.i.d. Wiener processes. Note that $E\{\sup_{\nu \in [0,1]} W_j^2(\nu)\} < \infty$, so that $E\{\sup_{\nu \in [0,1]} \|\Gamma_{X\varepsilon}(\cdot, \nu)\|_{L^2}^2\} \leq \sum_{j=1}^{\infty} \tilde{\zeta}_j E\{\sup_{\nu \in [0,1]} W_j^2(\nu)\} < \infty$. Therefore, in view of (S1.19), we deduce from the above finding that

\[
\sup_{\nu \in [0,1]} \left\| \frac{1}{[n\nu]} \sum_{i=1}^{[n\nu]} \tau_{\lambda}(X_i) \varepsilon_i \right\|_K^2 \leq \sup_{\nu \in [0,1]} \left\{ \frac{1}{[n\nu]} \sqrt{\frac{1}{[n\nu]} \sum_{i=1}^{[n\nu]} \varepsilon_i X_i} \right\}^2 \times \sum_{k=1}^{\infty} \frac{\|\varphi_k\|_{L^2}^2}{1 + \lambda \rho_k} 
\]

\[
\leq n^{-1} \nu_0^{-2} \sup_{\nu \in [0,1]} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{[n\nu]} \varepsilon_i X_i \right\|_{L^2}^2 \times \sum_{k=1}^{\infty} \frac{\|\varphi_k\|_{L^2}^2}{1 + \lambda \rho_k} \times \{1 + o_p(1)\} 
\]

\[
\leq n^{-1} \nu_0^{-2} \left\{ \sup_{\nu \in [0,1]} \left\| \Gamma_{X\varepsilon}(\cdot, \nu) - \frac{1}{\sqrt{n}} \sum_{i=1}^{[n\nu]} \varepsilon_i X_i \right\|_{L^2}^2 + \sup_{\nu \in [0,1]} \|\Gamma_{X\varepsilon}(\cdot, \nu)\|_{L^2}^2 \right\} \times \sum_{k=1}^{\infty} \frac{\|\varphi_k\|_{L^2}^2}{1 + \lambda \rho_k} \times \{1 + o_p(1)\} = O_p(n^{-1}\lambda^{-(2a+1)/(2D)}) .
\]

Consequently, combining the above finding and (S1.3) and (S1.18), we obtain that

\[
\sup_{\nu \in [0,1]} \|S_{n,\lambda,\nu}(\beta_{\lambda,\nu})\|_K = O_p(n^{-1/2}\lambda^{-(2a+1)/(4D)}) + o(\lambda^{1/2}) .
\]

Next, let $q_n = c(n^{-1/2}\lambda^{-(2a+1)/(4D)} + \lambda^{1/2})$ and denote by $\mathcal{B}(r) = \{\gamma \in \mathcal{H}, \|\gamma\|_K \leq r\}$ denote the $\|\cdot\|_K$-ball with radius $r > 0$ in $\mathcal{H}$. In view of (S1.16), for any $\beta \in \mathcal{B}(q_n)$, with probability tending to one, $\|I_{2,\nu,\nu}(\beta)\|_K \leq \|\beta\|_K/2 \leq q_n/2$. Therefore, in view of (S1.6), (S1.17) and (S1.21), for $F_n(\beta)$ defined in (S1.4), with probability tending to one, for any $\beta \in \mathcal{B}(q_n)$, $\sup_{\nu \in [0,1]} \|F_{n,\nu}(\beta)\|_K \leq \sup_{\nu \in [0,1]} \|I_{2,\nu,\nu}(\beta)\|_K + \sup_{\nu \in [0,1]} \|S_{n,\lambda,\nu}(\beta_{\lambda,\nu})\|_K \leq c n^{-1/2}\lambda^{-(2a+1)/(4D)} + q_n/2 \leq q_n$, which implies that $F_n,\nu(\mathcal{B}(q_n)) \subset \mathcal{B}(q_n)$ uniformly in $\nu \in [\nu_0,1]$. Observing (S1.4)–(S1.6), we have, for any $\beta_1, \beta_2 \in \mathcal{B}(q_n)$, $F_{n,\nu}(\beta_1) - F_{n,\nu}(\beta_2) =$
\[ I_{2,n,\nu}(\beta_1) - I_{2,n,\nu}(\beta_2). \] Due to (S1.17), with probability tending to one,

\[ \sup_{\nu\in[\nu_0,1]} \|F_{n,\nu}(\beta_1) - F_{n,\nu}(\beta_2)\|_K = \sup_{\nu\in[\nu_0,1]} \|I_{2,n,\nu}(\beta_1) - I_{2,n,\nu}(\beta_2)\|_K \leq \|\beta_1 - \beta_2\|_K/2, \]

which implies that \( F_{n,\nu} \) is a contraction mapping on \( B(q_n) \) uniformly in \( \nu \in [\nu_0,1] \). By the Banach contraction mapping theorem, there exists a unique element \( \beta^*_\nu \in B_n \) such that

\[ \beta^*_\nu = F_{n,\nu}(\beta^*_\nu) = \beta^*_\nu - S_{n,\lambda,\nu}(\beta_{\lambda,\nu} + \beta^*_\nu). \]

Letting \( \hat{\beta}_{n,\lambda}(\cdot, \nu) = \beta_{\lambda,\nu} + \beta^*_\nu \), we have \( S_{n,\lambda,\nu}\{\hat{\beta}_{n,\lambda}(\cdot, \nu)\} = 0 \), which implies that \( \hat{\beta}_{n,\lambda}(\cdot, \nu) \) is the estimator defined by (3.2). Moreover, we have, with probability tending to one, \( \sup_{\nu\in[\nu_0,1]} \|\hat{\beta}_{n,\lambda}(\cdot, \nu) - \beta_{\lambda,\nu}\|_K = \sup_{\nu\in[\nu_0,1]} \|\beta^*_\nu\|_K \leq q_n \). In view of (S1.3), we conclude

\[ \sup_{\nu\in[\nu_0,1]} \|\hat{\beta}_{n,\lambda}(\cdot, \nu) - \beta_0\|_K \leq \sup_{\nu\in[\nu_0,1]} \|\beta_{\lambda,\nu} - \beta_0\|_K + \sup_{\nu\in[\nu_0,1]} \|\hat{\beta}_{n,\lambda}(\cdot, \nu) - \beta_{\lambda,\nu}\|_K \]

\[ = O_p(\lambda^{1/2} + q_n) = O_p(\lambda^{1/2} + n^{-1/2}\lambda^{-(2a+1)/(4D)}). \]

**Proof of Theorem 1.** Let

\[ S_{n,\nu}(\beta) = -\frac{1}{[n\nu]} \sum_{i=1}^{[n\nu]} \left\{ Y_i - \int_0^1 X_i(s) \beta(s) \, ds \right\} \tau_\lambda(X_i), \]

(S1.22)

\[ S_{\nu}(\beta) = -E \left[ \left\{ Y_0 - \int_0^1 X_0(s) \beta(s) \, ds \right\} \tau_\lambda(X_0) \right], \]

and \( \Delta_{\nu}\beta = \hat{\beta}_{n,\lambda}(\cdot, \nu) - \beta_0 + W_{\lambda}(\beta_0) \) for the sake of notational convenience. Since \( D^2 S_{\lambda,\nu} \) vanishes and \( DS_{\lambda,\nu}(\beta_0) = id \), we have \( S_{\lambda,\nu}\{\hat{\beta}_{n,\lambda}(\cdot, \nu)\} - S_{\lambda,\nu}(\beta_0) = DS_{\lambda,\nu}(\beta_0)\Delta_{\nu}\beta = \Delta_{\nu}\beta \).

Since \( S_{n,\lambda,\nu}\{\hat{\beta}_{n,\lambda}(\cdot, \nu)\} = 0 \), we deduce from this equation that

\[ \hat{\beta}_{n,\lambda}(\cdot, \nu) - \beta_0 + S_{n,\lambda,\nu}(\beta_0) = \Delta_{\nu}\beta + S_{n,\lambda,\nu}(\beta_0) \]

\[ = -S_{n,\nu}\{\hat{\beta}_{n,\lambda}(\cdot, \nu)\} + S_{n,\nu}(\beta_0) + S_{\nu}\{\hat{\beta}_{n,\lambda}(\cdot, \nu)\} - S_{\nu}(\beta_0), \]

(S1.23)

where \( S_{n,\nu} \) and \( S_{\nu} \) are defined in (S1.22). Let \( r_n = \lambda^{1/2} + n^{-1/2}\lambda^{-(2a+1)/(4D)} \). For \( c_1 > 0 \), consider the event \( \mathcal{M}_n = \{ \sup_{\nu\in[\nu_0,1]} \|\Delta_{\nu}\beta\|_K \leq c_1 r_n \} \). By Lemma 1 we obtain that \( P(\mathcal{M}_n) \)
tends to one if the constant $c_1 > 0$ is chosen sufficiently large. For $c_K > 0$ in Lemma 5, let 
$q_n = c_1 c_K \lambda^{-(2a+1)/(4D)} r_n$ and let $p_n = c_1^2 q_n^{-2} \lambda^{-1} r_n = c_K^{-2} \lambda^{-(2D+2a+1)/(2D)}$. Note that $p_n \geq 1$ for $n$ large enough. In order to apply Lemma 8, we shall rescale $\Delta_\nu \beta$ such that the $L^2$-norm of its rescaled version is bounded by 1. Let $\tilde{\Delta}_\nu \beta = q_n^{-1} \Delta_\nu \beta$. By Lemma 5, we have that, on the event $\mathcal{M}_n$, 
\[ \|\tilde{\Delta}_\nu \beta\|_{L^2} \leq c_K \lambda^{-(2a+1)/(4D)} \|\Delta_\nu \beta\|_{K} \leq c_K q_n^{-1} \lambda^{-(2a+1)/(4D)} \|\Delta_\nu \beta\|_{K} \leq c_1 c_K q_n^{-1} \lambda^{-(2a+1)/(4D)} r_n \leq 1. \]

In addition, since $J(\Delta_\nu \beta, \Delta_\nu \beta) \leq \lambda^{-1} \|\Delta_\nu \beta\|_{K}^2$, we have 
\[ J(\tilde{\Delta}_\nu \beta, \tilde{\Delta}_\nu \beta) \leq q_n^{-2} J(\Delta_\nu \beta, \Delta_\nu \beta) \leq q_n^{-2} \lambda^{-1} \|\Delta_\nu \beta\|_{K}^2 \leq c_1^2 q_n^{-2} \lambda^{-1} r_n = p_n. \]

Hence, we have shown that $\tilde{\Delta}_\nu \beta \in \mathcal{F}_{p_n}$, where $\mathcal{F}_{p_n}$ is defined in (S1.12).

Recall from (S1.13) that, for the event $\mathcal{E}_n(c)$ defined below equation (S1.6), for any $\beta \in \mathcal{H}$,
\[ \tilde{H}_{n,\nu}(\beta) = \frac{1}{\sqrt{[n^D]}} \sum_{i=1}^{[n^D]} \left( \tau_\lambda(X_i) \langle \beta, X_i \rangle_{L^2} 1\{\mathcal{E}_n(c)\} - E\left[ \tau_\lambda(X_i) \langle \beta, X_i \rangle_{L^2} 1\{\mathcal{E}_n(c)\} \right] \right). \]

In view of (S1.22) and (S1.23) we thus obtain on the event $\mathcal{E}_n(c)$ that 
\[ \hat{\beta}_{n,\lambda}(\cdot, \nu) - \beta_0 + S_{n,\lambda,\nu}(\beta_0) = -S_{n,\nu}(\hat{\beta}_{n,\lambda}(\cdot, \nu)) + S_{n,\nu}(\beta_0) + S(\hat{\beta}_{n,\lambda}(\cdot, \nu)) - S_{\nu}(\beta_0) \]
\[ = \frac{1}{[n^D]} \sum_{i=1}^{[n^D]} \left[ \tau_\lambda(X_i) \int_0^1 X_i(s) \Delta_\nu \beta(s) ds - E\left\{ \tau_\lambda(X_i) \int_0^1 X_i(s) \Delta_\nu \beta(s) ds \right\} \right] \]
\[ \leq \frac{1}{\sqrt{[n^D]}} \tilde{H}_{n,\nu}(\Delta_\nu \beta) - E\left\{ \tau_\lambda(X_i) \int_0^1 X_i(s) \Delta_\nu \beta(s) ds 1\{\mathcal{E}(c)^c\} \right\}. \quad (S1.24) \]

Note that following arguments similar to the ones used in (S1.9), we deduce that 
\[ \left\| E\left[ 1\{\mathcal{E}(c)^c\} \tau_\lambda(X_i) \int_0^1 X_i(s) \Delta_\nu \beta(s) ds \right] \right\|_{K} \leq o(n^{-1} \lambda^{-1/(2D)}) \|\Delta_\nu \beta\|_{K} = o(\nu_n). \quad (S1.25) \]

Since $\tilde{\Delta}_\nu \beta \in \mathcal{F}_{p_n}$, by applying Lemma 8, observing (S1.14), we deduce that 
\[ \sup_{\nu \in [\nu_0, 1]} \sup_{\Delta_\nu \beta \in \mathcal{F}_{p_n}} \|\tilde{H}_{n,\nu}(\Delta_\nu \beta)\|_{K} = O_p\left\{ \left( p_n^{1/(2m)} + n^{-1/2} \left( \lambda^{-1/(2D)} \log n \right)^{1/2} \right) \right\} = O_p\left\{ \left( p_n^{1/(2m)} \lambda^{-1/(4D)} (\log n)^{1/2} \right) \right\}. \]
Consequently, for the $\Delta_{\nu,\beta}$ in \[S1.24\], it follows with probability tending to one,

$$n^{-1/2} \sup_{\nu \in [0,1]} \| \tilde{H}_{n,\nu}(\Delta_{\nu,\beta}) \|_K \leq n^{-1/2} q_n \sup_{\nu \in [0,1]} \sup_{\Delta_{\nu,\beta} \in \mathcal{F}_{\rho_n}} \| \tilde{H}_{n,\nu}(\Delta_{\nu,\beta}) \|_K$$

\[\leq c n^{-1/2} q_n p_n^{1/(2m)} \lambda^{-1/(4D)} (\log n)^{1/2} \leq c n^{-1/2} \lambda^{-2(2a+1)/(4D)} r_n \lambda^{-2(2D+2a+1)/(4Dm)} \lambda^{-1/(4D)} (\log n)^{1/2} \]

\[= c n^{-1/2} \lambda^{-\varsigma} \left( \lambda^{1/2} + n^{-1/2} \lambda^{-2(2a+1)/(4D)} \right) (\log n)^{1/2},\]

for the constant $\varsigma > 0$ in Assumption 5. Combining this with \[S1.24\] and \[S1.25\] yields

$$\sup_{\nu \in [0,1]} \left\| \tilde{\beta}_{n,\lambda}(\cdot, \nu) - \beta_0 + W_\lambda(\beta_0) + S_{n,\lambda,\nu}(\beta_0) \right\|_K = O_P \left\{ n^{-1/2} \lambda^{-\varsigma} \left( \lambda^{1/2} + n^{-1/2} \lambda^{-2(2a+1)/(4D)} \right) (\log n)^{1/2} \right\} = O_P(v_n). \tag{S1.26}$$

Observing $S_{n,\lambda,\nu}(\beta_0)$ defined in \[4.7\], we therefore deduce from the above equation that

$$n^{-1/2} \sup_{\nu \in [0,1]} \left\| \frac{\nu}{n} \tilde{\beta}_{n,\lambda}(\cdot, \nu) - \beta_0 + W_\lambda(\beta_0) \right\|_K = O_P \left\{ n^{-1/2} \lambda^{-\varsigma} \left( \lambda^{1/2} + n^{-1/2} \lambda^{-2(2a+1)/(4D)} \right) (\log n)^{1/2} \right\} = O_P(v_n). \tag{S1.27}$$

Observing \[S1.20\] and Assumption 5 we find

$$n^{-1} \sup_{\nu \in [0,1]} \left\| \frac{1}{n^\nu} \sum_{i=1}^{[n^\nu]} \varepsilon_i \tau_\lambda(X_i) \right\|_K = O_P \left( n^{-3/2} \lambda^{-2(2a+1)/(4D)} \right) = o_P(v_n).$$

The proof is therefore complete by combining the above equation with \[S1.26\] and \[S1.27\].

### S1.2 Proof of Theorem 2

The proof is now performed in two steps. First, in Lemma 2, we will show that the $U_i$’s are $L^2$-$m$-approximable (see Assumptions (1.1)–(1.4) in Berkes, Horváth, and Rice, 2013). Second,
in Lemma 3 we will show that $C_{U,\lambda}$ defined in (3.7) satisfies $\sup_{\lambda>0} \int_0^1 \int_0^1 |C_{U,\lambda}(s, t)| ds dt < \infty$. Then, Theorem 2 is proved by the arguments as given in the proof of Theorem 1.1 in Berkes, Horváth, and Rice (2013), which are omitted for the sake of brevity.

**Lemma 2.** Under the assumptions of Theorem 2, the series $\{U_i\}_{i \in \mathbb{Z}}$ defined in (4.10) is $L^2$-m-approximable w.r.t. the series $\{U_{i,\ell}\}_{i,\ell \in \mathbb{Z}}$ uniformly in $\lambda > 0$, where

$$U_{i,\ell} = \lambda^{2a+1}/(2D) \varepsilon_{i,\ell} \tau_\lambda(X_{i,\ell}) = \lambda^{2a+1}/(2D) \varepsilon_{i,\ell} \sum_{k=1}^{\infty} \frac{\langle X_{i,\ell}, \varphi_k \rangle_{L^2}}{1 + \lambda \rho_k} \varphi_k, \quad (i, \ell \in \mathbb{Z}).$$

**Proof.** By Lemmas 4 and 5 in Section S2, there exists a constant $L$ such that

$$E\|U_i\|_L^2 \leq 2 \varepsilon_i \cdot \|\tau_\lambda(X_i)\|_L \leq c \lambda^{-(2a+1)/(4D)} \|\tau_\lambda(X_i)\|_K \leq c \lambda^{-(2a+1)/(2D)} \|X_i\|_L^2.$$  

(S1.28)

This together with the fact that $\|U_i\|_L^2 \leq \lambda^{2a+1}/(2D) \varepsilon_i \cdot \|\tau_\lambda(X_i)\|_L$ implies that $U_i \in L^2([0, 1])$ uniformly in $\lambda > 0$. In addition, $E(U_i) \equiv 0$, and, by (S1.28), for any $\delta \in (0, 1)$,

$$E\|U_i\|_{L^2}^{2+\delta} \leq \lambda^{2a+1}/(2D) E|\varepsilon_i|^{2+\delta} E\|\tau_\lambda(X_i)\|_{L^2}^{2+\delta} \leq c \lambda^{2a+1}/(1+\delta) \lambda^{2+\delta} \|X_i\|_{L^2}^{2+\delta} < \infty,$$

where in the last step we have used Assumptions 4.1 and 3.2. Moreover, note that $m$-approximable series are strictly stationary (see, for example, Hörmann and Kokoszka 2010).

Hence, by applying Assumption 3 and (S1.28), we find that, uniformly in $\lambda > 0$,

$$\sum_{\ell=1}^{\infty} \left( E\|U_i - U_{i,\ell}\|_{L^2}^{2+\delta} \right)^{1/\kappa} \leq \sum_{\ell=1}^{\infty} \left\{ \lambda^{2a+1}/(2+\delta) \lambda^{2a+1}/(2D) \lambda^{2+\delta} \|\tau_\lambda(X_i) - \varepsilon_{i,\ell} \tau_\lambda(X_{i,\ell})\|_{L^2}^{2+\delta} \right\}^{1/\kappa}$$

$$\leq 2^{(1+\delta)/\kappa} \sum_{\ell=1}^{\infty} \left\{ \lambda^{2a+1}/(2D) \lambda^{2a+1}/(2+\delta) \lambda^{2+\delta} \|\tau_\lambda(X_i)\|_{L^2}^{2+\delta} \right\}^{1/\kappa}$$

$$+ 2^{(1+\delta)/\kappa} \sum_{\ell=1}^{\infty} \left\{ \lambda^{2a+1}/(2D) \lambda^{2a+1}/(2+\delta) \lambda^{2+\delta} \|\tau_\lambda(X_i - \varepsilon_{i,\ell}\tau_\lambda(X_{i,\ell})\|_{L^2}^{2+\delta} \right\}^{1/\kappa}$$

$$= 2^{(1+\delta)/\kappa} \left\{ \lambda^{2a+1}/(2+\delta) \lambda^{2+\delta} \|\tau_\lambda(X_i)\|_{L^2}^{2+\delta} \right\}^{1/\kappa} \sum_{\ell=1}^{\infty} \left( E|\varepsilon_i - \varepsilon_{i,\ell}|^{2+\delta} \right)^{1/\kappa}$$

$$+ 2^{(1+\delta)/\kappa} \left( E|\varepsilon_0|^{2+\delta} \right)^{1/\kappa} \sum_{\ell=1}^{\infty} \left\{ \lambda^{2a+1}/(2+\delta) \lambda^{2+\delta} \|\tau_\lambda(X_i - \varepsilon_{i,\ell}\tau_\lambda(X_{i,\ell})\|_{L^2}^{2+\delta} \right\}^{1/\kappa}$$
Now, we have shown that the series \( \{U_i\}_{i \in \mathbb{Z}} \) is \( L^2 \)-approximable uniformly in \( \lambda > 0 \). \( \square \)

**Lemma 3.** Under the assumptions of Theorem 2, we have \( \sup_{\lambda > 0} \int_0^1 \int_0^1 \{C_{U,\lambda}(s,t)\}^2 \, ds \, dt < \infty \), where \( C_{U,\lambda} \) is defined in (3.7).

**Proof.** Note that by Assumption 3, for \( \ell \geq 1 \), \( \varepsilon_{0,\ell} \) and \( \varepsilon_{-\ell} \) are independent; \( \tau_\lambda(X_{0,\ell}) \) and \( \tau_\lambda(X_{-\ell}) \) are independent. Note that \( E(\varepsilon_i) = 0 \) for any \( \ell \in \mathbb{Z} \). Hence we deduce that, for \( \ell \geq 1 \), \( E(\varepsilon_{0,\ell}\varepsilon_{-\ell}) = 0 \), so that \( E(\varepsilon_{0}\varepsilon_{-\ell}) = E\{ (\varepsilon_0 - \varepsilon_{0,\ell})\varepsilon_{-\ell} \} + E(\varepsilon_{0,\ell}\varepsilon_{-\ell}) = E\{ (\varepsilon_0 - \varepsilon_{0,\ell})\varepsilon_{-\ell} \} \).

In addition, for \( \ell \geq 1 \), we have

\[
E\{ \tau_\lambda(X_0)(s)\tau_\lambda(X_{-\ell})(t) \} = E\{ [\tau_\lambda(X_0)(s) - \tau_\lambda(X_{0,\ell})(s)]\tau_\lambda(X_{-\ell})(t) \} + E\{ \tau_\lambda(X_{0,\ell})(s)\tau_\lambda(X_{-\ell})(t) \}
= E\{ [\tau_\lambda(X_0)(s) - \tau_\lambda(X_{0,\ell})(s)]\tau_\lambda(X_{-\ell})(t) \} + E\{ \tau_\lambda(X_0)(s) \} E\{ \tau_\lambda(X_{0,\ell})(t) \}.
\]

Since \( E\{ \varepsilon_\ell \tau_\lambda(X_{0,\ell}) \} \equiv 0 \), this equation implies that, for \( \ell \geq 1 \),

\[
\text{cov}\{ \varepsilon_{0,\ell}\tau_\lambda(X_0)(s), \varepsilon_{-\ell}\tau_\lambda(X_{-\ell})(t) \} = E\{ (\varepsilon_0 - \varepsilon_{0,\ell})\varepsilon_{-\ell} \} E\{ [\tau_\lambda(X_0)(s) - \tau_\lambda(X_{0,\ell})(s)]\tau_\lambda(X_{-\ell})(t) \}
+ E\{ (\varepsilon_0 - \varepsilon_{0,\ell})\varepsilon_{-\ell} \} E\{ \tau_\lambda(X_0)(s) \} \times E\{ \tau_\lambda(X_0)(t) \}.
\]

Therefore, we deduce from the above equation that

\[
\int_0^1 \int_0^1 \{C_{U,\lambda}(s,t)\}^2 \, ds \, dt = \lambda^{2(2a+1)/D} \int_0^1 \int_0^1 \left[ \text{cov}\{ \varepsilon_{0,\ell}\tau_\lambda(X_0)(s), \varepsilon_{0,\ell}\tau_\lambda(X_0)(t) \} \right. \\
+ 2 \sum_{\ell=1}^{+\infty} \text{cov}\{ \varepsilon_{0,\ell}\tau_\lambda(X_0)(s), \varepsilon_{-\ell}\tau_\lambda(X_{-\ell})(t) \} \\
\left. + 2 \sum_{\ell=1}^{+\infty} \text{cov}\{ \varepsilon_{0,\ell}\tau_\lambda(X_0)(s), \varepsilon_{-\ell}\tau_\lambda(X_{-\ell})(t) \} \right]^2 \, ds \, dt \leq 3I_1 + 12I_2 + 12I_3, \tag{S1.29}
\]

where

\[
I_1 = \lambda^{2(2a+1)/D} \left\{ E(\varepsilon_0^2) \right\}^2 \int_{[0,1]^2} \left[ E\{ \tau_\lambda(X_0)(s) \times \tau_\lambda(X_0)(t) \} \right]^2 \, ds \, dt,
\]

\[
I_2 = \lambda^{2(2a+1)/D} \int_{[0,1]^2} \left( \sum_{\ell=1}^{+\infty} E\{ (\varepsilon_0 - \varepsilon_{0,\ell})\varepsilon_{-\ell} \} E\{ [\tau_\lambda(X_0)(s) - \tau_\lambda(X_{0,\ell})(s)]\tau_\lambda(X_{-\ell})(t) \} \right)^2 \, ds \, dt,
\]

and

\[
I_3 = \lambda^{2(2a+1)/D} \int_{[0,1]^2} \left( \sum_{\ell=1}^{+\infty} E\{ (\varepsilon_0 - \varepsilon_{0,\ell})\varepsilon_{-\ell} \} E\{ \tau_\lambda(X_0)(s) \} \times E\{ \tau_\lambda(X_{0,\ell})(t) \} \right)^2 \, ds \, dt.
\]
For the first term $I_1$, note that
\[
E\{\tau_\lambda(X_0)(s) \times \tau_\lambda(X_0)(t)\} = E\left[ \sum_{k_1=1}^{\infty} \frac{\langle X_0, \varphi_{k_1} \rangle_{L^2}}{1 + \lambda \rho_{k_1}} \varphi_{k_1}(s) \right] \cdot \sum_{k_2=1}^{\infty} \frac{\langle X_0, \varphi_{k_2} \rangle_{L^2}}{1 + \lambda \rho_{k_2}} \varphi_{k_2}(t)
\]
\[
= \sum_{k_1=1}^{\infty} \sum_{k_2=1}^{\infty} \frac{\varphi_{k_1}(s) \varphi_{k_2}(t)}{(1 + \lambda \rho_{k_1})(1 + \lambda \rho_{k_2})} E\left( \langle X_0, \varphi_{k_1} \rangle_{L^2} \langle X_0, \varphi_{k_2} \rangle_{L^2} \right)
\]
\[
= \sum_{k_1=1}^{\infty} \sum_{k_2=1}^{\infty} \frac{\varphi_{k_1}(s) \varphi_{k_2}(t)}{(1 + \lambda \rho_{k_1})(1 + \lambda \rho_{k_2})} \int_0^1 \int_0^1 C_X(t_1, t_2) \varphi_{k_1}(t_1) \varphi_{k_2}(t_2) dt_1 dt_2 = \sum_{k=1}^{\infty} \frac{\varphi_k(s) \varphi_k(t)}{(1 + \lambda \rho_k)^2}.
\]
Hence, by the Cauchy-Schwarz inequality, we find
\[
I_1 = \lambda^{2(2a+1)/D} \left( E(\varepsilon_0^2) \right)^2 \int_0^1 \int_0^1 \left\{ \sum_{k=1}^{\infty} \frac{\varphi_k(s) \varphi_k(t)}{(1 + \lambda \rho_k)^2} \right\}^2 dsdt
\]
\[
\leq \lambda^{2(2a+1)/D} \left( E(\varepsilon_0^2) \right)^2 \sum_{k=1}^{\infty} \frac{\|\varphi_k\|_{L^2}^2}{(1 + \lambda \rho_k)^2} = O(\lambda^{2(2a+1)/D}).
\]
For the second term $I_2$, by a direct application of the Cauchy-Schwarz inequality, we have
\[
\left( \sum_{\ell=1}^{+\infty} E(\varepsilon_0 - \varepsilon_{0,\ell}) E(\tau_\lambda(X_0)(s) - \tau_\lambda(X_0,\ell)(s)) \right)^2
\]
\[
\leq E(\varepsilon_0^2) \left( \sum_{\ell=1}^{+\infty} E(\varepsilon_0 - \varepsilon_{0,\ell})^2 \right) \left( \sum_{\ell=1}^{+\infty} E(\tau_\lambda(X_0)(s) - \tau_\lambda(X_0,\ell)(s))^2 \right) E(\tau_\lambda(X_0)(t))^2.
\]
Therefore, observing (S1.28) and Assumption 3, we deduce from the above equation that
\[
I_2 \leq \lambda^{2(2a+1)/D} E(\varepsilon_0^2) \left( \sum_{\ell=1}^{+\infty} E(\varepsilon_0 - \varepsilon_{0,\ell})^2 \right) \int_0^1 \int_0^1 \left( \sum_{\ell=1}^{+\infty} E(\tau_\lambda(X_0)(s) - \tau_\lambda(X_0,\ell)(s))^2 \right) ds \int_0^1 E(\tau_\lambda(X_0)(t))^2 dt
\]
\[
= \lambda^{2(2a+1)/D} E(\varepsilon_0^2) \left( \sum_{\ell=1}^{+\infty} E(\varepsilon_0 - \varepsilon_{0,\ell})^2 \right) \left[ \sum_{\ell=1}^{+\infty} E(\tau_\lambda(X_0) - \tau_\lambda(X_0,\ell))^2 \right] E(\tau_\lambda(X_0))^2_{L^2}
\]
\[
\leq c E(\varepsilon_0^2) \times \left( \sum_{\ell=1}^{+\infty} E(\varepsilon_0 - \varepsilon_{0,\ell})^2 \right) \times E(\|X_0\|_{L^2})^2 \times \left( \sum_{\ell=1}^{+\infty} E(\|X_0 - X_0,\ell\|_{L^2})^2 \right) < \infty.
\]
For the third term $I_3$, by the Cauchy-Schwarz inequality, (S1.28) and Assumption 3
\[
I_3 \leq \lambda^{2(2a+1)/D} \left( E(\tau_\lambda(X_0))^2 \right)^2 E(\varepsilon_0^2) \sum_{\ell=1}^{+\infty} E(\varepsilon_0 - \varepsilon_{0,\ell})^2 \leq c( E(\|X_0\|_{L^2})^2 ) E(\varepsilon_0^2) \sum_{\ell=1}^{+\infty} E(\varepsilon_0 - \varepsilon_{0,\ell})^2 < \infty.
\]
In conclusion, we deduce from (S1.29) that $\sup_{\lambda > 0} \int_0^1 \int_0^1 \{ C_{U,\lambda}(s, t) \}^2 ds dt < \infty$. \qed
S1.3 Proof of Theorem 3

We first deal with the bias term $W_\lambda(\beta_0)$ in (4.9). Observing (4.6), we deduce that

$$W_\lambda(\beta_0) = \sum_{k=1}^{\infty} V(\beta_0, \varphi_k) W_\lambda(\varphi_k) = \lambda \sum_{k=1}^{\infty} \frac{\rho_k \varphi_k V(\beta_0, \varphi_k)}{1 + \lambda \rho_k},$$

and, using Assumption 4.3 we conclude

$$\|W_\lambda(\beta_0)\|_K = \lambda \left\{ \sum_{k=1}^{\infty} \frac{\rho_k^2 V^2(\beta_0, \varphi_k)}{1 + \lambda \rho_k} \right\}^{1/2} \leq \lambda \left\{ \sum_{k=1}^{\infty} \rho_k^2 V^2(\beta_0, \varphi_k) \right\}^{1/2} = O(\lambda).$$

Note that by Lemma 5 in Section S2 and Assumption 5

$$\sqrt{n} \lambda^{(2a+1)/(2D)} \|W_\lambda(\beta_0)\|_2 \leq c \sqrt{n} \lambda^{(2a+1)/(4D)} \|W_\lambda(\beta_0)\|_K = o(1). \quad (S1.30)$$

Next, applying Theorem 1 and Lemma 5 in Section S2, we find

$$\sup_{\nu \in [0,1]} \left\| \sqrt{n} \lambda^{(2a+1)/(2D)} \nu \left\{ \hat{\beta}_{n,\lambda}(\cdot, \nu) - \beta_0 + W_\lambda(\beta_0) \right\} - \frac{1}{n} \sum_{i=1}^{[n\nu]} \varepsilon_i \tau_\lambda(X_i) \right\|_2^2 \leq c K n \lambda^{(2a+1)/(2D)} \sup_{\nu \in [0,1]} \left\| \nu \left\{ \hat{\beta}_{n,\lambda}(\cdot, \nu) - \beta_0 + W_\lambda(\beta_0) \right\} - \frac{1}{n} \sum_{i=1}^{[n\nu]} \varepsilon_i \tau_\lambda(X_i) \right\|_K^2 = O_p \left( n \lambda^{(2a+1)/(2D)} \nu^2 \right) = O_p \left( n \lambda^{(2a+1)/(2D)} \times n^{-1} \lambda^{-2c} \left( \lambda + n^{-1} \lambda^{-(2a+1)/(2D)} \right) (\log n) \right) = O_p \left( \lambda^{-2c+2(2D+2a+1)/(2D)} + n^{-1} \lambda^{-2c} (\log n) \right) = o_p(1), \quad (S1.31)$$

where we used Assumption 5 in the last step. By Theorem 2, there exists a Gaussian process

$$\{\Gamma(s, \nu)\}_{s \in [0,1], \nu \in [0,1]}$$

in $\mathcal{F}$ defined in the set (4.11) such that

$$\sup_{\nu \in [0,1]} \left\| n^{-1/2} \lambda^{(2a+1)/(2D)} \sum_{i=1}^{[n\nu]} \varepsilon_i \tau_\lambda(X_i) - \Gamma(\cdot, \nu) \right\|_2^2 = \sup_{\nu \in [0,1]} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{[n\nu]} U_i - \Gamma(\cdot, \nu) \right\|_2^2 = o_p(1).$$

Combining the above finding with (S1.30) and (S1.31) yields

$$\sup_{\nu \in [0,1]} \left\| \sqrt{n} \lambda^{(2a+1)/(2D)} \nu \left\{ \hat{\beta}_{n,\lambda}(\cdot, \nu) - \beta_0 \right\} - \Gamma(\cdot, \nu) \right\|_2^2 \leq 3 \sup_{\nu \in [0,1]} \left\| \sqrt{n} \lambda^{(2a+1)/(2D)} \nu \left\{ \hat{\beta}_{n,\lambda}(\cdot, \nu) - \beta_0 + W_\lambda(\beta_0) \right\} - \frac{1}{n} \sum_{i=1}^{[n\nu]} \varepsilon_i \tau_\lambda(X_i) \right\|_2^2 \quad (S1.32)$$
Recall the definition of $C_U$ in Assumption 

4.4 and let $\{\kappa_j\}_{j=1}^\infty$ and $\{\psi_j\}_{j=1}^\infty$ denote the eigenvalues and eigenfunctions of $C_U$, respectively, that is,

$$C_U(s, t) = \sum_{j=1}^\infty \kappa_j \psi_j(s) \psi_j(t), \quad \int_0^1 C_U(s, t) \psi_j(s) \, ds = \kappa_j \psi_j(t). \quad \text{(S1.33)}$$

By Theorem 1.1 in Berkes, Horváth, and Rice (2013), we have $\Gamma(\cdot, \nu) = \sum_{j=1}^\infty \sqrt{\kappa_j} \psi_j(s) W_j(\nu)$, for $s, \nu \in [0, 1]$, where the $W_j$'s are i.i.d. standard Brownian motions on $[0, 1]$. Since $E\left\{ \sup_{\nu \in [0, 1]} W_1^2(\nu) \right\}$ is finite, we obtain

$$E\left\{ \sup_{\nu \in [0, 1]} \|\Gamma(\cdot, \nu)\|_{L^2}^2 \right\} \leq \sum_{j=1}^\infty \kappa_j E\left\{ \sup_{\nu \in [0, 1]} W_j^2(\nu) \right\} < \infty. \quad \text{(S1.34)}$$

Furthermore, observing (S1.32), we deduce from direct calculations that

$$\hat{G}_n(\nu) = \sqrt{n \lambda^{(2a+1)/(2D)}} \nu^2 \int_0^1 \left\{ \hat{\beta}_{n, \lambda}^2(s, \nu) - \beta_0^2(s) \right\} \, ds$$

$$= 2\nu \int_0^1 \beta_0(s) \Gamma(s, \nu) \, ds + I_{1,n}(\nu) + I_{2,n}(\nu) + I_{3,n}(\nu) + I_{4,n}(\nu), \quad \text{(S1.35)}$$

where

$$I_{1,n}(\nu) = \frac{1}{\sqrt{n \lambda^{(2a+1)/(2D)}}} \int_0^1 \left[ \sqrt{n \lambda^{(2a+1)/(2D)}} \nu \left\{ \hat{\beta}_{n, \lambda}(s, \nu) - \beta_0(s) \right\} \right]^2 \, ds,$$

$$I_{2,n}(\nu) = \frac{1}{\sqrt{n \lambda^{(2a+1)/(2D)}}} \int_0^1 \Gamma^2(s, \nu) \, ds,$$

$$I_{3,n}(\nu) = \frac{2}{\sqrt{n \lambda^{(2a+1)/(2D)}}} \int_0^1 \left[ \sqrt{n \lambda^{(2a+1)/(2D)}} \nu \left\{ \hat{\beta}_{n, \lambda}(s, \nu) - \beta_0(s) \right\} - \Gamma(s, \nu) \right] \Gamma(s, \nu) \, ds,$$

$$I_{4,n}(\nu) = 2\nu \int_0^1 \beta_0(s) \left[ \sqrt{n \lambda^{(2a+1)/(2D)}} \nu \left\{ \hat{\beta}_{n, \lambda}(s, \nu) - \beta_0(s) \right\} - \Gamma(s, \nu) \right] \, ds.$$

Note that (S1.34) implies that $\sup_{\nu \in [0, 1]} \|\Gamma(\cdot, \nu)\|_{L^2}^2 = O_p(1)$. Therefore, observing that $n \lambda^{(2a+1)/D} \to \infty$ as $n \to \infty$ (see Assumption 5), (S1.32) and the Cauchy-Schwarz inequality,
it follows that \( \sup_{\nu \in [\nu_0, 1]} \{ |I_{1,n}(\nu)| + |I_{2,n}(\nu)| + |I_{3,n}(\nu)| + |I_{4,n}(\nu)| \} = o_p(1) \) as \( n \to \infty \).

Consequently, in view of (S1.35), we have that,

\[
\sup_{\nu \in [\nu_0, 1]} \left| \hat{G}_n(\nu) - 2\nu \int_0^1 \beta_0(s) \Gamma(s, \nu) \, ds \right| = o_p(1). \tag{S1.36}
\]

This proves the finite-dimensional convergence, i.e., for any \( k \in \mathbb{N}_+ \) and \( \nu_1, \ldots, \nu_k \in [\nu_0, 1] \),

\[
(\hat{G}_n(\nu_1), \ldots, \hat{G}_n(\nu_k)) \overset{d}{\to} \left( 2\nu_1 \int_0^1 \beta_0(s) \Gamma(s, \nu_1) \, ds, \ldots, 2\nu_k \int_0^1 \beta_0(s) \Gamma(s, \nu_k) \, ds \right). \tag{S1.37}
\]

Next, we shall show the tightness of the process \( \{\hat{G}_n(\nu)\}_{\nu \in [\nu_0, 1]} \). To achieve this, we shall show that the process \( \{\hat{G}_n(\nu)\}_{\nu \in [\nu_0, 1]} \) is asymptotically uniformly equicontinuous in probability (see Lemma 1.5.7 in [van der Vaart and Wellner, 1996]). By the Cauchy-Schwarz inequality and (S1.36), we deduce that

\[
\left| \hat{G}_n(\nu_1) - \hat{G}_n(\nu_2) \right| \leq \sup_{|\nu_1 - \nu_2| < \delta, \nu_1, \nu_2 \in [\nu_0, 1]} \left| 2\nu_1 \int_0^1 \beta_0(s) \Gamma(s, \nu_1) \, ds - 2\nu_2 \int_0^1 \beta_0(s) \Gamma(s, \nu_2) \, ds \right| + 2 \sup_{\nu \in [\nu_0, 1]} \left| \hat{G}_n(\nu) - 2\nu \int_0^1 \beta_0(s) \Gamma(s, \nu) \, ds \right| 
\]

\[
\leq \sup_{|\nu_1 - \nu_2| < \delta, \nu_1, \nu_2 \in [\nu_0, 1]} \left\{ 2|\nu_1 - \nu_2| \times \left| \int_0^1 \beta_0(s) \Gamma(s, \nu_1) \, ds \right| \right\} + 2 \sup_{\nu_1, \nu_2 \in [\nu_0, 1]} \left| \int_0^1 \beta_0(s) \{ \Gamma(s, \nu_1) - \Gamma(s, \nu_2) \} \, ds \right| + o_p(1)
\]

\[
\leq 2\delta \|\beta_0\|_{L^2} \sup_{\nu \in [\nu_0, 1]} \left| \int_0^1 \Gamma^2(s, \nu) \, ds \right|^{1/2} + 2 \|\beta_0\|_{L^2} \sup_{\nu_1, \nu_2 \in [\nu_0, 1]} \left| \int_0^1 \{ \Gamma(s, \nu_1) - \Gamma(s, \nu_2) \}^2 \, ds \right|^{1/2} + o_p(1).
\]

By Lemma 2.1 in [Berkes, Horváth, and Rice, 2013], we have \( \sup_{\nu \in [0, 1]} \int_0^1 \Gamma^2(s, \nu) \, ds < \infty \) a.s.

In addition, in view of (S1.33), note that the \( \psi_j \)'s are orthogonal in \( L^2([0, 1]) \), so that

\[
\sup_{|\nu_1 - \nu_2| < \delta, \nu_1, \nu_2 \in [0, 1]} \int_0^1 \{ \Gamma(s, \nu_1) - \Gamma(s, \nu_2) \}^2 \, ds = \sup_{|\nu_1 - \nu_2| < \delta, \nu_1, \nu_2 \in [0, 1]} \int_0^1 \left[ \sum_{j=1}^{\infty} \sqrt{\kappa_j} \psi_j(s) \{ W_j(\nu_1) - W_j(\nu_2) \} \right]^2 \, ds
\]

\[
= \sup_{|\nu_1 - \nu_2| < \delta, \nu_1, \nu_2 \in [0, 1]} \sum_{j=1}^{\infty} \kappa_j \{ W_j(\nu_1) - W_j(\nu_2) \}^2 \leq \sup_{|\nu_1 - \nu_2| < \delta, \nu_1, \nu_2 \in [0, 1]} \{ W_1(\nu_1) - W_1(\nu_2) \}^2 \sum_{j=1}^{\infty} \kappa_j = o_p(1), \tag{S1.39}
\]
where the last step is due to the modulus of continuity of Brownian motions and the fact that \( \sum_{j=1}^{\infty} \kappa_j < \infty \). Therefore, combining (S1.38)–(S1.39), we deduce that, for any \( \epsilon > 0 \),

\[
\lim_{\delta \downarrow 0} \limsup_{n \to \infty} P \left\{ \sup_{|\nu_1 - \nu_2| < \delta} \sup_{\nu_1, \nu_2 \in [0, 1]} |\widehat{G}_n(\nu_1) - \widehat{G}_n(\nu_2)| > \epsilon \right\} = 0.
\]

This proves the tightness of the process \( \{\widehat{G}_n(\nu)\}_{\nu \in [0, 1]} \). Together with (S1.37), by Lemma 1.5.4 in van der Vaart and Wellner (1996), this implies that \( \{\widehat{G}_n(\nu)\}_{\nu \in [0, 1]} \sim \{2\nu \int_0^1 \beta_0(s) \Gamma(s, \nu) ds\}_{\nu \in [0, 1]} \) in \( \ell^\infty([\nu_0, 1]) \). In addition, observing (S1.33), we have for the Gaussian process \( \{\Gamma(s, \nu)\}_{s \in [0, 1], \nu \in [0, 1]} \),

\[
\text{cov}\{\Gamma(s_1, \nu_1), \Gamma(s_2, \nu_2)\} = \text{cov}\left\{\sum_{j=1}^{\infty} \sqrt{\kappa_j} \psi_j(s_1) W_{j_1}(\nu_1), \sum_{j=1}^{\infty} \sqrt{\kappa_j} \psi_j(s_2) W_{j_2}(\nu_2)\right\} = \text{cov}\{W_{j_1}(\nu_1), W_{j_2}(\nu_2)\} \sum_{j=1}^{\infty} \kappa_j \psi_j(s_1) \psi_j(s_2) = (\nu_1 \wedge \nu_2) C_U(s_1, s_2),
\]

where \( C_U \) is defined in Assumption 4.4 and \( s_1, s_2 \in [0, 1] \) and \( \nu_1, \nu_2 \in [\nu_0, 1] \). Therefore,

\[
cov\left\{\int_0^1 \beta_0(s_1) \Gamma(s_1, \nu_1) ds, \int_0^1 \beta_0(s_2) \Gamma(s_2, \nu_2) ds \right\} = (\nu_1 \wedge \nu_2) \int_0^1 \int_0^1 C_U(s_1, s_2) \beta_0(s_1) \beta_0(s_2) ds_1 ds_2 = (\nu_1 \wedge \nu_2) \sigma_d^2,
\]

where \( \sigma_d^2 \) is defined in (3.6). This implies that \( \int_0^1 \beta_0(s) \Gamma(s, \nu) ds \overset{d}{=} 2\sigma_d \mathbb{B}(\nu) \),

where \( \mathbb{B} \) denotes a standard Brownian motion. Hence, combining the above finding with (S1.36) yields \( \{\widehat{G}_n(\nu)\}_{\nu \in [0, 1]} \sim \{2\sigma_d \mathbb{B}(\nu)\}_{\nu \in [0, 1]} \) in \( \ell^\infty([\nu_0, 1]) \), which completes the proof.

### S1.4 Proof of Theorem 4

Observing \( \widehat{G}_n(\nu) \) defined in (4.12), we have

\[
\sqrt{n} \lambda^{(2a+1)/(2D)} \hat{\nu}_n = \sqrt{n} \lambda^{(2a+1)/(2D)} \left[ \int_{\nu_0}^{1} \left. \nu \right| \int_{\nu_0}^{1} \{\widehat{\beta}_{n,\lambda}^2(s, \nu) - \beta_{n,\lambda}^2(s, 1)\} ds \right|^2 \omega(d\nu) \right]^{1/2}
\]

\[
= \sqrt{n} \lambda^{(2a+1)/(2D)} \left[ \int_{\nu_0}^{1} \left. \nu \right| \int_{\nu_0}^{1} \{\widehat{\beta}_{n,\lambda}^2(s, \nu) - \beta_0^2(s)\} ds - \nu \int_{\nu_0}^{1} \{\widehat{\beta}_{n,\lambda}^2(s, 1) - \beta_0^2(s)\} ds \right|^2 \omega(d\nu) \right]^{1/2}
\]

\[
= \left\{ \int_{\nu_0}^{1} \left| \widehat{\nu}_n(\nu) - \nu \right|^2 \widehat{\nu}_n(1) \omega(d\nu) \right\}^{1/2}.
\]

In addition, for \( \widehat{T}_n \) defined in (3.4), \( \sqrt{n} \lambda^{(2a+1)/(2D)} (\widehat{T}_n - d_0) = \sqrt{n} \lambda^{(2a+1)/(2D)} \int_{\nu_0}^{1} \{\widehat{\beta}_{n,\lambda}^2(s, 1) - \beta_0^2(s)\} ds = \widehat{\beta}_n(1) \). Therefore, we obtain \( \sqrt{n} \lambda^{(2a+1)/(2D)} ((\widehat{T}_n - d_0), \nu_n) = (\widehat{\nu}_n(1), \int_{\nu_0}^{1} \left| \widehat{\nu}_n(\nu) - \right|\)
\[ \nu^2 \widehat{\mathcal{C}}_n(1)^2 \omega(d\nu)^{1/2} \]. Since \( \sigma_d^2 \neq 0 \), by Theorem 3 and the continuous mapping theorem,

\[ \frac{\widehat{T}_n - d_0}{\overline{V}_n} \xrightarrow{d} \frac{2\sigma_d \mathbb{B}(1)}{2\sigma_d \left\{ \int_{R} | \nu \mathbb{B}(\nu) - \nu^2 \mathbb{B}(1)^2 \omega(d\nu) \right\}^{1/2}} = \mathbb{W}, \]

where \( \mathbb{W} \) is defined in (4.14). This completes the proof.

### S1.5 Proof of Theorem 5

When \( d_0 = \int_{0}^{1} | \beta_0(t) |^2 dt = 0 \), we have \( \widehat{T}_n = \widehat{T}_n - d_0 = o_p(1) \) and \( \overline{V}_n = o_p(1) \), which implies that \( \widehat{T}_n - \mathcal{Q}_{1-\alpha}(\mathbb{W})\overline{V}_n = o_p(1) \), so that \( \lim_{n \to \infty} P\{ \widehat{T}_n > \mathcal{Q}_{1-\alpha}(\mathbb{W})\overline{V}_n + \Delta \} = \lim_{n \to \infty} P\{ \widehat{T}_n - \mathcal{Q}_{1-\alpha}(\mathbb{W})\overline{V}_n > \Delta \} = 0 \). When \( 0 < \int_{0}^{1} | \beta_0(t) |^2 dt < \Delta \), we use

\[ P\{ \widehat{T}_n > \mathcal{Q}_{1-\alpha}(\mathbb{W})\overline{V}_n + \Delta \} = P\left\{ \frac{\widehat{T}_n - d_0}{\overline{V}_n} > \mathcal{Q}_{1-\alpha}(\mathbb{W}) + \frac{\sqrt{n} \lambda^{(2a+1)/(2D)}}{\sqrt{n} \lambda^{(2a+1)/(2D)}} \overline{V}_n \right\} . \quad (S1.40) \]

Note that \( \sqrt{n} \lambda^{(2a+1)/(2D)}\overline{V}_n = o_p(1) \) and \( \sqrt{n} \lambda^{(2a+1)/(2D)} \to +\infty \) as \( n \to \infty \) according to Assumption 5, so that \( \sqrt{n} \lambda^{(2a+1)/(2D)}(\Delta - d_0) \to +\infty \). Hence, when \( d_0 < \Delta \), it follows that

\[ P\{ \widehat{T}_n > \mathcal{Q}_{1-\alpha}(\mathbb{W})\overline{V}_n + \Delta \} \to 0. \]

When \( d_0 = \int_{0}^{1} | \beta_0(t) |^2 dt = \Delta \), we have

\[ \lim_{n \to \infty} P\{ \widehat{T}_n > \mathcal{Q}_{1-\alpha}(\mathbb{W})\overline{V}_n + \Delta \} = \lim_{n \to \infty} P\left\{ \frac{\widehat{T}_n - d_0}{\overline{V}_n} > \mathcal{Q}_{1-\alpha}(\mathbb{W}) \right\} = \alpha. \]

When \( d_0 = \int_{0}^{1} | \beta_0(t) |^2 dt > \Delta \), we have \( \sqrt{n} \lambda^{(2a+1)/(2D)}(\Delta - d_0) \to -\infty \) as \( n \to \infty \), so that we obtain from (S1.40) that \( P\{ \widehat{T}_n > \mathcal{Q}_{1-\alpha}(\mathbb{W})\overline{V}_n + \Delta \} \to 0 \), which completes the proof.

### S2. Auxiliary lemmas

**Lemma 4.** Under Assumptions 4 and 2, there exists a constant \( c > 0 \) such that, for any \( x \in L^2([0, 1]) \), \( \| \tau_\lambda(x) \|^2_K \leq c \lambda^{-(2a+1)/(2D)} \| x \|^2_{L^2} \) and \( \mathbb{E} \| \tau_\lambda(X) \|^2_K \leq c \lambda^{-1/(2D)} \).

**Proof.** Recalling (3.8) and using the orthogonality of the functions \( \varphi_k \), we have

\[ \| \tau_\lambda(x) \|^2_K = \sum_{k=1}^{\infty} \frac{\langle x, \varphi_k \rangle_{L^2}^2}{1 + \lambda \rho_k} \leq \| x \|^2_{L^2} \sum_{k=1}^{\infty} \frac{\| \varphi_k \|^2_{L^2}}{1 + \lambda \rho_k} \leq c \lambda^{-(2a+1)/(2D)} \| x \|^2_{L^2}. \]
By Assumption 2, $\mathbb{E}((X, \varphi_k)^2) = 1$, so that $\mathbb{E}\|\tau_\lambda(X)\|^2_K = \sum_{k=1}^{\infty}(1 + \lambda \rho_k)^{-1} \leq c\lambda^{-1/(2D)}$.

**Lemma 5** (Lemma 3.1 in [Shang and Cheng, 2015]). Under Assumptions 1 and 2, there exists a constant $c_K > 0$ such that for any $\beta \in \mathcal{H}$, $\|\beta\|^2_{L^2} \leq c_K\lambda^{-(2a+1)/(2D)}\|\beta\|^2_K$.

Lemmas 6 and 7 below shows that the long-run covariance $C_{X_s}(s, t) = \sum_{i=-\infty}^{+\infty} \text{cov}\{\varepsilon_0X_0(s), \varepsilon_\ell X_\ell(t)\}$ of $\{X_\ell, \varepsilon_i\} \in \mathbb{Z}$ satisfies $C_{X_s} \in L^2([0, 1]^2)$ and $\int_0^1 C_{X_s}(t, t)dt < \infty$.

**Lemma 6.** Under Assumption 3, we have $C_{X_s} \in L^2([0, 1]^2)$.

*Proof.* Note that by Assumption 3, for $\ell \geq 1$, $\varepsilon_0, \varepsilon_\ell$ and $\varepsilon_{-\ell}$ are independent; $X_{0, \ell}$ and $X_{-\ell}$ are independent. Since $\mathbb{E}(X) \equiv 0$, we find that, for $\ell \geq 1$, $\mathbb{E}\{X_0(s)X_\ell(t)\} = \mathbb{E}\{X_0(s) - X_{0, \ell}(s)\}X_{-\ell}(t)\}$. Observing that $\mathbb{E}(\varepsilon_0\varepsilon_{-\ell}) = \mathbb{E}\{(\varepsilon_0 - \varepsilon_{0, \ell})\varepsilon_{-\ell}\} + \mathbb{E}(\varepsilon_{0, \ell}\varepsilon_{-\ell}) = \mathbb{E}\{(\varepsilon_0 - \varepsilon_{0, \ell})\varepsilon_{-\ell}\}$, for $\ell \geq 1$, we deduce from the above equation that, for $\ell \geq 1$,

$$\text{cov}\{\varepsilon_0X_0(s), \varepsilon_{-\ell}X_{-\ell}(t)\} = \mathbb{E}\{\varepsilon_0 - \varepsilon_{0, \ell}\}\mathbb{E}\{X_0(s) - X_{0, \ell}(s)\} \times X_{-\ell}(t)\}.$$  \hspace{1cm} (S2.1)

Therefore, we find from the above equation that $\int_0^1 \int_0^1 \{C_{X_s}(s, t)\}^2 dsdt \leq 2I_1 + 8I_2$, where

$$I_1 = \{\mathbb{E}(\varepsilon_0^2)\}^2 \int_0^1 \int_0^1 \left[\mathbb{E}\{X_0(s) \times X_0(t)\}\right]^2 dsdt,$$

$$I_2 = \int_0^1 \int_0^1 \left(\sum_{\ell=1}^{+\infty} \mathbb{E}\{(\varepsilon_0 - \varepsilon_{0, \ell})\varepsilon_{-\ell}\}\mathbb{E}\{X_0(s) - X_{0, \ell}(s)\} \times X_{-\ell}(t)\}\right)^2 dsdt.$$  \hspace{1cm} (S2.2)

For the first term $I_1$, we have $I_1 = \{\mathbb{E}(\varepsilon_0^2)\}^2 \times \|C_X\|^2_{L^2} < \infty$. For the second term $I_2$, by the Cauchy-Schwarz inequality, we find

$$\left(\sum_{\ell=1}^{+\infty} \mathbb{E}\{(\varepsilon_0 - \varepsilon_{0, \ell})\varepsilon_{-\ell}\}\mathbb{E}\{X_0(s) - X_{0, \ell}(s)\} \times X_{-\ell}(t)\}\right)^2 \leq \left[\sum_{\ell=1}^{+\infty} \mathbb{E}\{(\varepsilon_0 - \varepsilon_{0, \ell})\varepsilon_{-\ell}\}\right]^2 \times \left[\sum_{\ell=1}^{+\infty} \mathbb{E}\{X_0(s) - X_{0, \ell}(s)\}\right]^2 \times \mathbb{E}\{X_0(t)\}^2$$

$$= \mathbb{E}(\varepsilon_0^2) \times \left(\sum_{\ell=1}^{+\infty} \mathbb{E}|\varepsilon_0 - \varepsilon_{0, \ell}|^2\right) \times \left[\sum_{\ell=1}^{+\infty} \mathbb{E}\{X_0(s) - X_{0, \ell}(s)\}\right]^2 \times \mathbb{E}\{X_0(t)\}^2.$$  \hspace{1cm} (S2.3)
Therefore, by Assumption 3, we deduce from the above equation that $I_2 \leq E(\varepsilon_0)^2 \times (\sum_{\ell=1}^{+\infty} E|\varepsilon_0 - \varepsilon_{0,\ell}|^2) \times (\sum_{\ell=1}^{+\infty} E\|X_0 - X_{0,\ell}\|_{L^2}^2) \times E\|X_0\|_{L^2}^2 < \infty$. Finally, we deduce that $\int_0^1 \int_0^1 \{C_{X,\varepsilon}(s,t)\}^2 dsdt < \infty$ and complete the proof. \hfill \Box

**Lemma 7.** Under Assumption 3, we have $\int_0^1 C_{X,\varepsilon}(s,s)ds < \infty$.

**Proof.** By the arguments similar to the ones used to obtain (S2.1), it follows that, for $\ell \geq 1$, $\text{cov}\{\varepsilon_0 X_0(s), \varepsilon_{-\ell} X_{-\ell}(s)\} = E(\varepsilon_0 - \varepsilon_{0,\ell}) E\{X_0(s) - X_{0,\ell}(s)\} X_{-\ell}(s)$. Therefore, following the proof of Lemma 6, we deduce that

$$
\int_0^1 C_{X,\varepsilon}(s,s)ds = \int_0^1 \left[ \text{cov}\{\varepsilon_0 X_0(s), \varepsilon_0 X_0(s)\} + 2 \sum_{\ell=1}^{+\infty} \text{cov}\{\varepsilon_0 X_0(s), \varepsilon_{-\ell} X_{-\ell}(s)\} \right] ds
$$

$$
= E(\varepsilon_0^2) E\|X_0\|_{L^2}^2 + 2 \int_0^1 \left( \sum_{\ell=1}^{+\infty} E(\varepsilon_0 - \varepsilon_{0,\ell}) E\{X_0(s) - X_{0,\ell}(s)\} X_{-\ell}(s) \right) ds
$$

$$
\leq E(\varepsilon_0^2) E\|X_0\|_{L^2}^2 + 2 \int_0^1 \left\{ E(\varepsilon_0^2) \left( \sum_{\ell=1}^{+\infty} E|\varepsilon_0 - \varepsilon_{0,\ell}|^2 \right) \left( \sum_{\ell=1}^{+\infty} E\|X_0(s) - X_{0,\ell}\|_{L^2}^2 \right)^2 \right\}^{1/2} E\|X_0\|_{L^2}^2 ds,
$$

where in the last step we applied (S2.3) by taking $t = s$. Therefore, by the Cauchy-Schwarz inequality and Assumption 3, we conclude that $\int_0^1 C_{X,\varepsilon}(s,s)ds \leq E(\varepsilon_0^2) E\|X_0\|_{L^2}^2 + 2\{E(\varepsilon_0^2)\}^{1/2} \left( \sum_{\ell=1}^{+\infty} E|\varepsilon_0 - \varepsilon_{0,\ell}|^2 \right)^{1/2} \left( \sum_{\ell=1}^{+\infty} E\|X_0 - X_{0,\ell}\|_{L^2}^2 \right)^{1/2} < \infty$. \hfill \Box

Lemma 8 below is used to prove Theorem 1. Recall the definition of $H_{n,k}$ in (S1.15).

**Lemma 8.** For $p_n \geq 1$, let $\mathcal{F}_{p_n} = \{\beta \in \mathcal{H} : \|\beta\|_{L^2} \leq 1, J(\beta, \beta) \leq p_n\}$. Then, under Assumptions 7, 8 as $n \to \infty$,

$$
\max_{1 \leq k \leq n} \sup_{\beta \in \mathcal{F}_{p_n}} \frac{\|H_{n,k}(\beta)\|_K}{P_{n/2}^{1/2m} \|\beta\|_{L^2}^{(m-1)/m} + n^{-1/2}} = O_p(\lambda^{-1/(2D)} \log n)^{1/2}.
$$

**Proof.** The proof of Lemma 8 follows a modified argument of the proof of Lemma 3.4 in Shang and Cheng (2015). For any $x \in L^2([0,1])$, let $g(x, \beta) = \tau_\lambda(x) \int_0^1 \beta(s)x(s)ds$. We have

$$
H_{n,k}(\beta_1) - H_{n,k}(\beta_2) = \frac{1}{\sqrt{n}} \sum_{i=1}^k \left( g(X_i, \beta_1 - \beta_2) \times 1\{\mathcal{E}_n(c)\} - E[g(X_i, \beta_1 - \beta_2) \times 1\{\mathcal{E}_n(c)\}] \right).
$$
Note that, on the event $\mathcal{E}_n(c)$ defined below equation (S1.6), $\|\langle \beta_1 - \beta_2, X_i \rangle_{L^2} \mathbf{1}\{\mathcal{E}_n(c)\} - \mathbb{E}[\langle \beta_1 - \beta_2, X_i \rangle_{L^2} \mathbf{1}\{\mathcal{E}_n(c)\}]\| \leq c \|eta_1 - \beta_2\|_{L^2}^2$. Hence, we deduce that

$$\left\| g(X_i, \beta_1 - \beta_2) \mathbf{1}\{\mathcal{E}_n(c)\} - \mathbb{E}[g(X_i, \beta_1 - \beta_2) \mathbf{1}\{\mathcal{E}_n(c)\}] \right\|^2_K \leq 2(\log n)^2 \|eta_1 - \beta_2\|^2_{L^2} \lambda(X_i)^2.$$

For $1 \leq k \leq n$, let $W_{n,k}^2 = \frac{1}{n} \sum_{i=1}^k \|\tau_l(X_i)\|^2_K$ and $\mathcal{X}_n = \{\|\tau_l(X_i)\|^2_K\}_{i=1}^n$. By Lemma 4 of [Pinelis (1994)], for any $1 \leq j \leq n$, for any $\beta_1, \beta_2 \in \mathcal{H}$ and for $1 \leq j \leq n$,

$$\begin{align*}
P\left\{ \|H_{n,k}(\beta_1) - H_{n,k}(\beta_2)\|_{\mathcal{H}} \geq x \mid \mathcal{X}_n \right\} &= P\left\{ \left\| \frac{1}{\sqrt{k}} \sum_{i=1}^k \left[ g(X_i, \beta_1 - \beta_2) - \mathbb{E}[g(X_i, \beta_1 - \beta_2)] \right] \right\|_K \geq \sqrt{n/k} x \mid \mathcal{X}_n \right\} \\
&\leq 2 \exp \left( -\frac{nk^{-1}x^2}{2k^{-1} \sum_{i=1}^k \|\tau_l(X_i)\|^2_K \|eta_1 - \beta_2\|^2_{L^2}} \right) \leq 2 \exp \left( -\frac{x^2}{2 W_{n,k}^2 \|eta_1 - \beta_2\|^2_{L^2}} \right). \quad (S2.4)
\end{align*}$$

Following the proof of Lemma 3.4 in [Shang and Cheng (2015)] (see p. 13 of [Shang and Cheng (2015)]), we deduce that, for any $1 \leq k \leq n$,

$$P \left\{ \sup_{\beta \in \mathcal{F}_{n_0}, \|eta\|_{L^2} \leq \delta} \|H_{n,k}(\beta)\|^2_K \geq x \mid \mathcal{X}_n \right\} \leq 2 \exp \left( -c_1^{-2} W_{n,k}^{-2} p_n^{-1/(2m)} \delta^{-2 + 1/m} x^2 \right).$$

Taking $\gamma = 1 - 1/(2m)$, $b_n = \sqrt{n} p_n^{1/(4m)}$, $\theta_n = b_n^{-1}$, $Q_n = [-\log_2 \theta_n + \gamma - 1]$ and $T_n = c_2 (\lambda^{-1/(2D)} \log n)^{1/2}$, for some constant $c_2 > 0$ to be specified below, yields that

$$P \left\{ \max_{1 \leq k \leq n} \sup_{\beta \in \mathcal{F}_{n_0}, \|eta\|_{L^2} \leq 2} \frac{\sqrt{n} \|H_{n,k}(\beta)\|^2_K}{b_n \|eta\|^2_{L^2} + 1} \geq T_n \mid \mathcal{X}_n \right\} \leq \sum_{k=1}^n P \left\{ \sup_{\beta \in \mathcal{F}_{n_0}, \|eta\|_{L^2} \leq \theta_n^{1/\gamma}} \frac{\sqrt{n} \|H_{n,k}(\beta)\|^2_K}{b_n \|eta\|^2_{L^2} + 1} \geq T_n \mid \mathcal{X}_n \right\}
+ \sum_{k=1}^n \sum_{j=0}^{Q_n} P \left\{ \sup_{\beta \in \mathcal{F}_{n_0}, (\theta_n 2^j)^{1/\gamma} \leq \|eta\|_{L^2} \leq (\theta_n 2^{j+1})^{1/\gamma}} \frac{\sqrt{n} \|H_{n,k}(\beta)\|^2_K}{b_n \|eta\|^2_{L^2} + 1} \geq T_n \mid \mathcal{X}_n \right\}$$

$$\leq \sum_{k=1}^n P \left\{ \sup_{\beta \in \mathcal{F}_{n_0}, \|eta\|_{L^2} \leq \theta_n^{1/\gamma}} \sqrt{n} \|H_{n,k}(\beta)\|^2_K \geq T_n \mid \mathcal{X}_n \right\}
+ \sum_{k=1}^n \sum_{j=0}^{Q_n} P \left\{ \sup_{\beta \in \mathcal{F}_{n_0}, (\theta_n 2^j)^{1/\gamma} \leq \|eta\|_{L^2} \leq (\theta_n 2^{j+1})^{1/\gamma}} \sqrt{n} \|H_{n,k}(\beta)\|^2_K \geq (b_n \theta_n 2^j + 1) T_n \mid \mathcal{X}_n \right\}
\leq 2 \sum_{k=1}^n \exp \left( -c_1^{-2} W_{n,k}^{-2} p_n^{-1/(2m)} \theta_n^{-1}(1/m) \gamma^{-1} n^2 T_n^2 \right).$$
\[ + 2 \sum_{k=1}^{n} \sum_{j=0}^{Q_n} \exp \left\{ -c_1^{-2} W_{n,k}^{-2} T_{n}^2 \right\} \]

\[ \leq 2 \sum_{k=1}^{n} \exp \left( -c_1^{-2} W_{n,k}^{-2} T_{n}^2 \right) + 2(Q_n + 1) \sum_{k=1}^{n} \exp \left( -c_1^{-2} W_{n,k}^{-2} T_{n}^2 / 4 \right) \]

\[ \leq 2(Q_n + 2) \sum_{k=1}^{n} \exp \left( -c_1^{-2} W_{n,k}^{-2} T_{n}^2 / 4 \right) \leq 2(Q_n + 2) \exp \left( \log n - c_1^{-2} W_{n,n}^{-2} T_{n}^2 / 4 \right) = o(1), \]

which together with (S2.4) completes the proof.

References

Pinelis, I. (1994). Optimum bounds for the distributions of martingales in Banach spaces. Ann. Probab., 22, 1679–1706.

Shang, Z. and Cheng, G. (2015b) Supplement to “Nonparametric inference in generalized functional linear models”. DOI:10.1214/15-AOS1322SUPP.

Van der Vaart, A. W. and Wellner, J. A. (1996). Weak Convergence and Empirical Processes. Springer, New York.