

SUPPLEMENT TO “ESTIMATION IN FUNCTIONAL LINEAR QUANTILE REGRESSION”*

BY KENGO KATO

Hiroshima University

This supplementary file contains the additional discussion on the connection to nonlinear ill-posed inverse problems, technical proofs omitted in the main body, some useful technical tools and additional simulation results. In this supplementary file, we follow the notation, the numbering and the convention used in the main body. In the technical proofs, we use $C > 0$ to indicate a generic constant of which the value may change from line to line.

APPENDIX A: CONNECTION TO NONLINEAR ILL-POSED INVERSE PROBLEMS

This section discusses the connection of our problem of estimating the slope function to nonlinear ill-posed inverse problems. For any fixed $u \in \mathcal{U}$, our estimator $\hat{b}(\cdot, u)$ can be understood as a regularized solution to an empirical version of a nonlinear inverse problem that corresponds to the “normal equation” :

$$(A.1) \quad \mathcal{A}(u, b(\cdot, u)) = 0,$$

where the map $\mathcal{A} : \mathcal{U} \times L_2[0, 1] \rightarrow L_2[0, 1]$ is defined by

$$\begin{aligned} \mathcal{A}(u, g)(\cdot) &= \mathbb{E}[\{u - 1(Y \leq \int_0^1 g(t)X^c(t)dt)\}X^c(\cdot)] \\ &= \mathbb{E}[\{u - F_{Y|X}(\int_0^1 g(t)X^c(t)dt \mid X)\}X^c(\cdot)], \quad u \in \mathcal{U}, g \in L_2[0, 1]. \end{aligned}$$

Here $F_{Y|X}(y \mid X)$ is the conditional distribution function of Y given X . For the sake of simplicity, we have ignored the constant term. Observe that for any fixed $u \in \mathcal{U}$, the map $\mathcal{A}(u, \cdot) : L_2[0, 1] \rightarrow L_2[0, 1]$ is a nonlinear operator. In fact, using an approximation $X_i^c \approx \sum_{j=1}^m \hat{\xi}_{ij} \hat{\phi}_j =: \hat{X}_i^c$, our estimator $\hat{b}(\cdot, u)$ is an approximate solution to an empirical version of (A.1) over the linear subspace spanned by $\{\hat{\phi}_1, \dots, \hat{\phi}_m\}$:

$$(A.2) \quad \hat{\mathcal{A}}(u, \hat{b}(\cdot, u)) \approx 0,$$

*Supported by the Grant-in-Aid for Young Scientists (B) (22730179) from the JSPS.

where the map $\hat{\mathcal{A}} : \mathcal{U} \times L_2[0, 1] \rightarrow L_2[0, 1]$ is defined by

$$\hat{\mathcal{A}}(u, g)(\cdot) = \mathbb{E}_n[\{u - 1(Y_i \leq \int_0^1 g(t) \hat{X}_i^c(t) dt)\} \hat{X}_i^c(\cdot)], \quad u \in \mathcal{U}, g \in L_2[0, 1].$$

To see (A.2), observe that

$$\hat{\mathcal{A}}(u, \hat{b}(\cdot, u))(\cdot) = \sum_{j=1}^m \mathbb{E}_n[\{u - 1(Y_i \leq \sum_{k=1}^m \hat{\xi}_{ik} \hat{b}_k(u))\} \hat{\xi}_{ij}] \hat{\phi}_j(\cdot).$$

The first order condition to (2.4) (in the main body) implies that

$$\mathbb{E}_n[\{u - 1(Y_i \leq \sum_{k=1}^m \hat{\xi}_{ik} \hat{b}_k(u))\} \hat{\xi}_{ij}] \approx 0, \quad 1 \leq j \leq m,$$

which leads to (A.2) [the discussion here is informal to give an intuition behind our estimator]. Note that solving (2.4) (in the main body) is computationally more appealing than directly searching a solution to (A.2) as the former problem is convex while the latter is not.

Meanwhile, as long as the map $y \mapsto F_{Y|X}(y | x)$ is continuous, for any fixed $u \in \mathcal{U}$, the nonlinear inverse problem (A.1) is locally ill-posed at $b(\cdot, u)$ in the sense of Hofmann and Scherzer (1998, Definition 1.1), i.e., there exists a sequence of functions $\{g_n\}$ in a neighborhood of $b(\cdot, u)$ (in $L_2[0, 1]$) such that $\mathcal{A}(u, g_n) \rightarrow \mathcal{A}(u, b(\cdot, u))$ but $g_n \not\rightarrow b(\cdot, u)$ in the L_2 -norm. To see this, take a sequence of functions $\{g_n\}$ in a neighborhood of $b(\cdot, u)$ such that $g_n \xrightarrow{w} b(\cdot, u)$ but $g_n \not\rightarrow b(\cdot, u)$ in the L_2 -norm, where $\xrightarrow{w}$ means the weak convergence in $L_2[0, 1]$. Then, by the weak convergence, we have

$$(A.3) \quad \int_0^1 g_n(t) X^c(t) dt \rightarrow \int_0^1 b(t, u) X^c(t) dt.$$

By the continuity of the map $y \mapsto F_{Y|X}(y | X)$, (A.3) implies that $\mathcal{A}(u, g_n) \rightarrow \mathcal{A}(u, b(\cdot, u))$ despite $g_n \not\rightarrow b(\cdot, u)$. This suggests that any sensible estimation procedure based on the normal equation (A.1) has to involve some regularizations. In our case, the regularization is done by restricting the parameter space for $b(\cdot, u)$ to a sequence of finite dimensional subspaces, where the cut-off level m plays a role of regularization parameter.

APPENDIX B: PROOF OF THEOREM 3.2

Let X_{n+1} be a copy of X independent of the data

$$\mathcal{D}_n := \{(Y_1, X_1), \dots, (Y_n, X_n)\}.$$

Then

$$\mathcal{E}(\hat{Q}_{Y|X}, u) = \mathbb{E}[\{\hat{Q}_{Y|X}(u | X_{n+1}) - Q_{Y|X}(u | X_{n+1})\}^2 | \mathcal{D}_n].$$

Let $X_{n+1}^c = X_{n+1} - \mathbb{E}[X(t)] = \sum_{j=1}^{\infty} \xi_{n+1,j} \phi_j$. Observe that

$$\begin{aligned} \hat{Q}_{Y|X}(u | X_{n+1}) &= \hat{a}(u) + \sum_{j=1}^m \hat{b}_j(u) \xi_{n+1,j} \\ &+ \int_0^1 X_{n+1}^c(t) \sum_{j=1}^m \hat{b}_j(u) (\hat{\phi}_j - \phi_j)(t) dt \\ &+ \int_0^1 (\mathbb{E}[X(t)] - \bar{X}(t)) \hat{b}(t, u) dt \end{aligned}$$

and

$$Q_{Y|X}(u | X_{n+1}) = a(u) + \sum_{j=1}^m b_j(u) \xi_{n+1,j} + \sum_{j=m+1}^{\infty} b_j(u) \xi_{n+1,j}.$$

Letting $\eta_{n+1,j} = \kappa_j^{-1/2} \xi_{n+1,j}$, we have

$$\begin{aligned} &\{\hat{Q}_{Y|X}(u | X_{n+1}) - Q_{Y|X}(u | X_{n+1})\}^2 \\ &\leq C \left[(\hat{a} - a)^2(u) + \left\{ \sum_{j=1}^m (\hat{d}_j - d_j)(u) \eta_{n+1,j} \right\}^2 + \left\{ \sum_{j=m+1}^{\infty} b_j(u) \xi_{n+1,j} \right\}^2 \right. \\ &\quad + \int_0^1 X_{n+1}^c(t)^2 dt \int_0^1 \left\{ \sum_{j=1}^m \hat{b}_j(u) (\hat{\phi}_j - \phi_j)(t) \right\}^2 dt \\ &\quad \left. + \left\{ \int_0^1 (\mathbb{E}[X(t)] - \bar{X}(t)) \hat{b}(t, u) dt \right\}^2 \right]. \end{aligned}$$

Taking expectation with respect to X_{n+1} , we have

$$\begin{aligned} &\mathbb{E}[\{\hat{Q}_{Y|X}(u | X_{n+1}) - Q_{Y|X}(u | X_{n+1})\}^2 | \mathcal{D}_n] \\ &\leq C \left[\|\hat{d}^m(u) - d^m(u)\|_{\ell^2}^2 + \sum_{j=m+1}^{\infty} \kappa_j b_j^2(u) + \int_0^1 \left\{ \sum_{j=1}^m \hat{b}_j(u) (\hat{\phi}_j - \phi_j)(t) \right\}^2 dt \right. \\ &\quad \left. + \left\{ \int_0^1 (\mathbb{E}[X(t)] - \bar{X}(t)) \hat{b}(t, u) dt \right\}^2 \right]. \end{aligned}$$

By the proof of Theorem 3.1, we have

$$\sup_{u \in \mathcal{U}} \|\hat{d}^m(u) - d^m(u)\|_{\ell^2}^2 = O_P(m/n) = O_P(n^{-(\alpha+2\beta-1)/(\alpha+2\beta)}), \text{ and}$$

$$\sum_{j=m+1}^{\infty} \kappa_j b_j^2(u) \leq C \sum_{j=m+1}^{\infty} j^{-\alpha-2\beta} = O(m^{-\alpha-2\beta+1}) = O(n^{-(\alpha+2\beta-1)/(\alpha+2\beta)}).$$

Observe that

$$\begin{aligned} & \left\{ \sum_{j=1}^m \hat{b}_j(u)(\hat{\phi}_j - \phi_j) \right\}^2 \\ & \leq 2 \left\{ \sum_{j=1}^m b_j(u)(\hat{\phi}_j - \phi_j) \right\}^2 + 2 \left\{ \sum_{j=1}^m (\hat{b}_j - b_j)(u)(\hat{\phi}_j - \phi_j) \right\}^2 \\ & = 2 \left\{ \sum_{j=1}^m b_j(u)(\hat{\phi}_j - \phi_j) \right\}^2 + 2 \left\{ \sum_{j=1}^m \kappa_j^{-1/2}(\hat{d}_j - d_j)(u)(\hat{\phi}_j - \phi_j) \right\}^2 \\ & \leq 2m \sum_{j=1}^m b_j^2(u)(\hat{\phi}_j - \phi_j)^2 + 2\|\hat{d}^m(u) - d^m(u)\|_{\ell^2}^2 \sum_{j=1}^m \kappa_j^{-1}(\hat{\phi}_j - \phi_j)^2, \end{aligned}$$

by which we have

$$\begin{aligned} & \int_0^1 \left\{ \sum_{j=1}^m \hat{b}_j(u)(\hat{\phi}_j - \phi_j)(t) \right\}^2 dt \\ & \leq 2m \sum_{j=1}^m b_j^2(u) \|\hat{\phi}_j - \phi_j\|^2 + 2\|\hat{d}^m(u) - d^m(u)\|_{\ell^2}^2 \sum_{j=1}^m \kappa_j^{-1} \|\hat{\phi}_j - \phi_j\|^2. \end{aligned}$$

By the proof of Theorem 3.1, we see that

$$\begin{aligned} m \sum_{j=1}^m b_j^2(u) \|\hat{\phi}_j - \phi_j\|^2 & \leq Cm \sum_{j=1}^m j^{-2\beta} \|\hat{\phi}_j - \phi_j\|^2 = O_P(mn^{-1}) \\ & = O_P(n^{-(\alpha+2\beta-1)/(\alpha+2\beta)}), \end{aligned}$$

while by the proof of (5.4) in Appendix C, we have $\sum_{j=1}^m \kappa_j^{-1} \|\hat{\phi}_j - \phi_j\|^2 = o_P(1)$. Hence we conclude that

$$\sup_{u \in \mathcal{U}} \int_0^1 \left\{ \sum_{j=1}^m \hat{b}_j(u)(\hat{\phi}_j - \phi_j)(t) \right\}^2 dt = O_P(n^{-(\alpha+2\beta-1)/(\alpha+2\beta)}).$$

Finally, using assumption (A8), we have

$$\begin{aligned} \left\{ \int_0^1 (\mathbb{E}[X(t)] - \bar{X}(t)) \hat{b}(t, u) dt \right\}^2 &\leq \int_0^1 (\mathbb{E}[X(t)] - \bar{X}(t))^2 dt \int_0^1 \hat{b}^2(t, u) dt \\ &= O_P(n^{-1} + \Delta^\gamma) \times O_P(1) = O_P(n^{-1}), \end{aligned}$$

uniformly in $u \in \mathcal{U}$. Taking these together, we conclude that

$$\sup_{u \in \mathcal{U}} \mathbb{E}[\{\hat{Q}_{Y|X}(u | X_{n+1}) - Q_{Y|X}(u | X_{n+1})\}^2 | \mathcal{D}_n] = O_P(n^{-(\alpha+2\beta-1)/(\alpha+2\beta)}).$$

This completes the proof. $\square$

APPENDIX C: PROOF OF PROPOSITION 3.1

We here provide a proof for (3.4). Consider the same construction as in Hall and Horowitz (2007). Let $\phi_1(t) \equiv 1$ and $\phi_{j+1}(t) = 2^{1/2} \cos(j\pi t)$ for $j \geq 1$. Put $\varrho_j = \theta_j j^{-\beta}$ for $[n^{1/(\alpha+2\beta)}] + 1 \leq j \leq 2[n^{1/(\alpha+2\beta)}]$ and $\varrho_j = 0$ otherwise where $[y]$ denotes the integer part of $y \in \mathbb{R}$ and each θ_j is either 0 or 1. Let $Z_1, Z_2, \dots \sim U[-3^{1/2}, 3^{1/2}]$ i.i.d. Take $X(t) = \sum_{j=1}^{\infty} j^{-\alpha/2} Z_j \phi_j(t)$ and $\varrho(t) = \sum_{j=[n^{1/(\alpha+2\beta)}]+1}^{2[n^{1/(\alpha+2\beta)}]} \varrho_j \phi_j(t)$. Note that the former sum is almost surely uniformly convergent in $t \in [0, 1]$ and hence X has sample paths almost surely continuous (see for example Marcus, 1975, Theorem 1.1). Consider a sequence of data generating processes

$$Y = \int_0^1 \varrho(t) X(t) dt + \epsilon = \sum_{j=[n^{1/(\alpha+2\beta)}]+1}^{2[n^{1/(\alpha+2\beta)}]} \theta_j j^{-(\alpha+2\beta)/2} Z_j + \epsilon, \quad \epsilon \sim N(0, 1), \quad \epsilon \perp X.$$

Then we have

$$Q_{Y|X}(u | X) = a(u) + \int b(t, u) X(t) dt,$$

where

$$a(u) = \Phi^{-1}(u), \quad b(t, u) \equiv \varrho(t), \quad f_{Y|X}(y|X) = \phi(y - \int_0^1 \varrho(t) X(t) dt),$$

by which one sees that assumptions (A4)-(A6) are satisfied. Here $\phi(\cdot)$ and $\Phi(\cdot)$ are the density and the distribution function of the standard normal distribution, respectively. Suppose that $\alpha \leq 3$. Then, since for any $0 < \gamma < \alpha - 1$, the function $t \mapsto \cos(t)$ is $\gamma/2$ -Hölder continuous (by the periodicity of the cosine function), we have

$$\mathbb{E}[(X(s) - X(t))^2] \leq C |s - t|^\gamma \sum_{j=1}^{\infty} j^{-\alpha+\gamma} \leq C' |s - t|^\gamma, \quad \forall s, t \in [0, 1],$$

where C and C' are some constants. This shows that assumption (A8) is satisfied with $0 < \gamma < \alpha - 1$ when $\alpha \leq 3$. For $\alpha > 3$, $K(s, t)$ is twice continuously differentiable, so that assumption (A8) is satisfied with $0 < \gamma \leq 2$. Finally, since $\mathbb{E}[Y | X] = \int_0^1 \varrho(t)X(t)dt$, by Hall and Horowitz (2007), for any estimator $(t, u) \mapsto \bar{b}(t, u)$,

$$\begin{aligned} & \sup^* \sup_{u \in \mathcal{U}} \int_0^1 \mathbb{E}[(\bar{b}(t, u) - b(t, u))^2] dt \\ & \geq \sup^* \int_0^1 \mathbb{E}[(\bar{b}(t, u_0) - \varrho(t))^2] dt \quad (u_0 \text{ is any point in } \mathcal{U}) \\ & \geq Dn^{-(2\beta-1)/(\alpha+2\beta)}, \end{aligned}$$

where $\sup^*$ denotes the supremum over all $2^{\lfloor n^{1/(\alpha+2\beta)} \rfloor}$ different distributions of (Y, X) obtained by taking different choices of $\theta_{\lfloor n^{1/(\alpha+2\beta)} \rfloor + 1}, \dots, \theta_{2^{\lfloor n^{1/(\alpha+2\beta)} \rfloor}}$, and $D > 0$ is a constant. The “in-probability” version of the lower bound (3.4) follows from the same reasoning as in the proof of Yuan and Cai (2010, p. 3442) and the Paley-Zygmund inequality (which states that for any nonnegative random variable Z with $\mathbb{E}[Z^2] < \infty$, $\mathbb{P}(Z > \lambda \mathbb{E}[Z]) \geq (1 - \lambda)^2 (\mathbb{E}[Z])^2 / (\mathbb{E}[Z^2])$ for all $\lambda \in (0, 1)$). The other assertions follow similarly. This completes the proof. $\square$

APPENDIX D: PROOFS OF (5.2), (5.4), (5.6), (5.7) AND (5.9)

In this section, we provide proofs of (5.2), (5.4), (5.6), (5.7) and (5.9) omitted in Section 5. Throughout the section, we assume all the conditions of Theorem 3.1. Define the (infeasible) empirical covariance kernel

$$\hat{K}^*(s, t) = \mathbb{E}_n[(X_i(s) - \bar{X}(s))(X_i(t) - \bar{X}(t))],$$

where $\bar{X}(t) = n^{-1} \sum_{i=1}^n X_i(t)$. Let $\hat{K}^*(s, t) = \sum_{j=1}^\infty \hat{\kappa}_j^* \hat{\phi}_j^*(s) \hat{\phi}_j^*(t)$ be the spectral expansion of $\hat{K}^*(s, t)$ where $\hat{\kappa}_1^* \geq \hat{\kappa}_2^* \geq \dots \geq 0$ and $\{\hat{\phi}_j^*\}_{j=1}^\infty$ is an orthonormal basis of $L_2[0, 1]$. Without loss of generality, we may assume that

$$\int \hat{\phi}_j^* \phi_j \geq 0, \quad \int \hat{\phi}_j \hat{\phi}_j^* \geq 0, \quad \forall j \geq 1,$$

where, to ease the notation, $\int_0^1 f(t)dt$ is abbreviated as $\int f$ for any function $f : [0, 1] \rightarrow \mathbb{R}$. Define

$$\hat{\xi}_{ij}^* = \int (X_i - \bar{X}) \hat{\phi}_j^*, \quad \hat{\eta}_{ij}^* = \kappa_j^{-1/2} \hat{\xi}_{ij}^*, \quad i = 1, \dots, n; j \geq 1.$$

Recall that $\eta_{ij} = \kappa_j^{-1/2} \xi_{ij} = \kappa_j^{-1/2} \int X_i^c \phi_j$ (where $X_i^c(t) = X_i(t) - \mathbb{E}[X_i(t)]$) and $\hat{\eta}_{ij} = \kappa_j^{-1/2} \hat{\xi}_{ij} = \kappa_j^{-1/2} \int (\hat{X}_i - \bar{X}) \hat{\phi}_j$ for $i = 1, \dots, n$ and $j \geq 1$. We will frequently use the following decomposition: for $i = 1, \dots, n$ and $j \geq 1$,

$$\begin{aligned} \hat{\eta}_{ij} - \eta_{ij} &= \hat{\eta}_{ij} - \hat{\eta}_{ij}^* + \hat{\eta}_{ij}^* - \eta_{ij}, \\ \hat{\eta}_{ij} - \hat{\eta}_{ij}^* &= \kappa_j^{-1/2} \int (\hat{X}_i - X_i) \hat{\phi}_j + \kappa_j^{-1/2} \int X_i (\hat{\phi}_j - \hat{\phi}_j^*) \\ &\quad - \kappa_j^{-1/2} \int (\bar{X} - \bar{X}) \hat{\phi}_j - \kappa_j^{-1/2} \int \bar{X} (\hat{\phi}_j - \hat{\phi}_j^*) \\ &=: \hat{\Delta}_{ij1} + \hat{\Delta}_{ij2} - \hat{\Delta}_{j3} - \hat{\Delta}_{j4}, \\ \hat{\eta}_{ij}^* - \eta_{ij} &= \kappa_j^{-1/2} \int X_i^c (\hat{\phi}_j^* - \phi_j) - \kappa_j^{-1/2} \int \bar{X}^c \phi_j - \kappa_j^{-1/2} \int \bar{X}^c (\hat{\phi}_j^* - \phi_j) \\ &=: \hat{\Delta}_{ij5} - \hat{\Delta}_{j6} - \hat{\Delta}_{j7}, \end{aligned}$$

where $\bar{X}^c(t) = n^{-1} \sum_{i=1}^n X_i^c(t)$.

We prepare some lemmas. For any function $R : [0, 1]^2 \rightarrow \mathbb{R}$, define $|||R||| = (\int \int R^2(s, t) ds dt)^{1/2}$. Recall that $\|f\|^2 = \int_0^1 f^2(t) dt$ for any $f : [0, 1] \rightarrow \mathbb{R}$.

LEMMA D.1. *We have*

$$\mathbb{E}_n[\|\hat{X}_i - X_i\|^2] = O_P(\Delta^\gamma), \quad |||\hat{K} - \hat{K}^*|||^2 = O_P(\Delta^\gamma).$$

Furthermore, as $n \rightarrow \infty$, with probability approaching one,

$$\|\hat{\phi}_j - \hat{\phi}_j^*\| \leq C j^{\alpha+1} |||\hat{K} - \hat{K}^*|||, \quad 1 \leq j \leq m.$$

PROOF. Observe that

$$\hat{X}_i(t) - X_i(t) = \sum_{l=1}^{L_i} (X_i(t_{il}) - X_i(t)) 1(t \in [t_{il}, t_{i,l+1})), \quad t \in [0, 1),$$

by which we have

$$\|\hat{X}_i - X_i\|^2 = \sum_{l=1}^{L_i} \int_{t_{il}}^{t_{i,l+1}} (X_i(t_{il}) - X_i(t))^2 dt.$$

Taking expectation, we have

$$\begin{aligned} \mathbb{E}[\|\hat{X}_i - X_i\|^2] &= \sum_{l=1}^{L_i} \int_{t_{il}}^{t_{i,l+1}} \mathbb{E}[(X(t) - X(t_{il}))^2] dt \\ &\leq C \sum_{l=1}^{L_i} (t_{i,l+1} - t_{il})^{\gamma+1} \leq C \Delta^\gamma, \end{aligned}$$

where we have used assumption (A8). This leads to the first assertion. The second assertion follows from the Schwarz inequality and the first assertion. The third assertion needs some effort. By Bosq (2000, Lemmas 4.2 and 4.3; see also the remark below), we have

$$(D.1) \quad \sup_{j \geq 1} |\hat{\kappa}_j^* - \kappa_j| \leq |||\hat{K}^* - K|||, \quad \sup_{j \geq 1} \hat{\chi}_j \|\hat{\phi}_j - \hat{\phi}_j^*\| \leq 8^{1/2} |||\hat{K} - \hat{K}^*|||,$$

where $\hat{\chi}_j = \min\{\hat{\kappa}_{j-1}^* - \hat{\kappa}_j^*, \hat{\kappa}_j^* - \hat{\kappa}_{j+1}^*\}$ for $j \geq 2$ and $\hat{\chi}_1 = \hat{\kappa}_1^* - \hat{\kappa}_2^*$. For some small constant $c > 0$, define the event

$$\mathcal{E}_n = \{\hat{\chi}_j \geq cj^{-\alpha-1}, 1 \leq j \leq m\}.$$

It suffices to show that $\mathbb{P}(\mathcal{E}_n) \rightarrow 1$. By the first inequality in (D.1),

$$\hat{\kappa}_k^* - \hat{\kappa}_{k+1}^* \geq \kappa_k - \kappa_{k+1} - 2|||\hat{K}^* - K||| \geq C^{-1}k^{-\alpha-1} - 2|||\hat{K}^* - K|||.$$

Since $k^{-\alpha-1} \geq m^{-\alpha-1} \asymp n^{-(\alpha+1)/(\alpha+2\beta)}$, $|||\hat{K}^* - K||| = O_P(n^{-1/2})$ (which follows by a simple calculation), and $n^{-1/2} = o(n^{-(\alpha+1)/(\alpha+2\beta)})$ (which follows by $\beta > \alpha/2 + 1$), we have uniformly in $1 \leq k \leq m$,

$$\hat{\kappa}_k^* - \hat{\kappa}_{k+1}^* \geq C^{-1}(1 - o_P(1))k^{-\alpha-1},$$

which leads to that $\mathbb{P}(\mathcal{E}_n) \rightarrow 1$ by taking c sufficiently small. $\square$

REMARK D.1. Lemma 4.3 of Bosq (2000) reads as follows: for functions $Q, R : [0, 1]^2 \rightarrow \mathbb{R}$ having the spectral expansions in $L_2[0, 1]^2$ of the form $Q(s, t) = \sum_{j=1}^{\infty} \lambda_j \psi_j(s) \psi_j(t)$ and $R(s, t) = \sum_{j=1}^{\infty} \nu_j \varphi_j(s) \varphi_j(t)$, where $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$, $\nu_1 \geq \nu_2 \geq \dots \geq 0$, and $\{\psi_j\}_{j=1}^{\infty}$ and $\{\varphi_j\}_{j=1}^{\infty}$ are orthonormal bases for $L_2[0, 1]$, we have: $\chi_j \|\varphi_j - \psi_j\| \leq 8^{1/2} |||R - Q|||$ for all $j \geq 1$ such that $\chi_j > 0$, where $\chi_j = \min\{\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1}\}$ for $j \geq 2$ and $\chi_1 = \lambda_1 - \lambda_2$. Here, we have assumed that $\int \varphi_j \psi_j \geq 0$ for all $j \geq 1$. This lemma actually holds with $\sup_{j \geq 1} \chi_j \|\varphi_j - \psi_j\| \leq 8^{1/2} |||R - Q|||$ since the inequality trivially holds in case of $\chi_j = 0$.

The following useful result was established in Hall and Horowitz (2007).

LEMMA D.2. *As $n \rightarrow \infty$, with probability approaching one,*

$$\|\hat{\phi}_j^* - \phi_j\|^2 \leq 10 \sum_{k: k \neq j} (\kappa_j - \kappa_k)^{-2} \left\{ \iint (\hat{K}^* - K)(s, t) \phi_j(s) \phi_k(t) ds dt \right\}^2,$$

for all $1 \leq j \leq m$. Furthermore, we have

$$\sum_{k:k \neq j} (\kappa_j - \kappa_k)^{-2} \mathbb{E} \left[\left\{ \iint (\hat{K}^* - K)(s, t) \phi_j(s) \phi_k(t) ds dt \right\}^2 \right] = O(j^2 n^{-1}),$$

uniformly in $1 \leq j \leq m$.

PROOF. See Hall and Horowitz (2007, p.83-84). $\square$

D.1. Proofs of (5.2) and (5.4). We first prove (5.4). By Lemmas D.1 and D.2, we have

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m \hat{\Delta}_{ij1}^2 &\leq \mathbb{E}_n[\|\hat{X}_i - X_i\|^2] \left\{ \sum_{j=1}^m \kappa_j^{-1} \right\} = O_P(m^{\alpha+1} \Delta^\gamma) = o_P(1), \\ \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m \hat{\Delta}_{ij2}^2 &\leq \mathbb{E}_n[\|X_i\|^2] \left\{ \sum_{j=1}^m \kappa_j^{-1} \|\hat{\phi}_j - \hat{\phi}_j^*\|^2 \right\} \\ &= O_P(1) \times O_P(\Delta^\gamma \sum_{j=1}^m j^{3\alpha+2}) = O_P(m^{3\alpha+3} \Delta^\gamma) = o_P(1), \\ \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m \hat{\Delta}_{ij5}^2 &\leq \mathbb{E}_n[\|X_i^c\|^2] \left\{ \sum_{j=1}^m \kappa_j^{-1} \|\hat{\phi}_j^* - \phi_j\|^2 \right\} \\ &= O_P(1) \times O_P(n^{-1} \sum_{j=1}^m j^{\alpha+2}) = O_P(m^{\alpha+3} n^{-1}) = o_P(1). \end{aligned}$$

Similarly, we have

$$\begin{aligned} \sum_{j=1}^m \hat{\Delta}_{j3}^2 &= O_P(m^{\alpha+1} \Delta^\gamma) = o_P(1), \\ \sum_{j=1}^m \hat{\Delta}_{j4}^2 &= O_P(m^{3\alpha+3} \Delta^\gamma) = o_P(1), \\ \sum_{j=1}^m \hat{\Delta}_{j7}^2 &= O_P(m^{\alpha+3} n^{-2}) = o_P(1). \end{aligned}$$

Using the decomposition $X_i^c(t) = \sum_{k=1}^{\infty} \xi_{ik} \phi_k(t)$, we have

$$\hat{\Delta}_{j6} = \kappa_j^{-1/2} \int \bar{X}^c \phi_j = \kappa_j^{-1/2} \bar{\xi}_j = \bar{\eta}_j,$$

by which we have

$$\sum_{j=1}^m \hat{\Delta}_{j6}^2 = O_P(\mathbb{E}[\sum_{j=1}^m \hat{\Delta}_{j6}^2]) = O_P(\sum_{j=1}^m \mathbb{E}[\bar{\eta}_j^2]) = O_P(mn^{-1}) = o_P(1).$$

Finally, by a direct calculation, we have

$$\mathbb{E}_n[\|\eta_i^m\|^2] = O_P(m).$$

Taking these together, we obtain (5.4).

We now turn to prove (5.2). Observe that

$$\begin{aligned} \max_{1 \leq i \leq n} \sum_{j=1}^m \hat{\Delta}_{ij1}^2 &\leq \max_{1 \leq i \leq n} \|\hat{X}_i - X_i\|^2 \times O_P(m^{\alpha+1}), \\ \max_{1 \leq i \leq n} \sum_{j=1}^m \hat{\Delta}_{ij2}^2 &\leq \max_{1 \leq i \leq n} \|X_i\|^2 \times O_P(m^{3\alpha+3}\Delta^\gamma), \\ \max_{1 \leq i \leq n} \sum_{j=1}^m \hat{\Delta}_{ij5}^2 &\leq \max_{1 \leq i \leq n} \|X_i^c\|^2 \times O_P(m^{\alpha+3}n^{-1}). \end{aligned}$$

Since $\int \mathbb{E}[X^4] \leq C$, we have

$$\max_{1 \leq i \leq n} \|X_i^c\|^2 = O_P(n^{1/2}),$$

which leads to that

$$\max_{1 \leq i \leq n} \sum_{j=1}^m \hat{\Delta}_{ij2}^2 = O_P(n^{1/2}m^{3\alpha+3}\Delta^\gamma), \quad \max_{1 \leq i \leq n} \sum_{j=1}^m \hat{\Delta}_{ij5}^2 = O_P(m^{\alpha+3}n^{-1/2}).$$

Using the trivial bound $\max_{1 \leq i \leq n} \|\hat{X}_i - X_i\|^2 \leq \sum_{i=1}^n \|\hat{X}_i - X_i\|^2$, we also have

$$\max_{1 \leq i \leq n} \sum_{j=1}^m \hat{\Delta}_{ij1}^2 = O_P(nm^{\alpha+1}\Delta^\gamma).$$

Similarly, since $\mathbb{E}[\eta_{ij}^4] = \kappa_j^{-2}\mathbb{E}[\xi_{ij}^4] \leq C$ for all $j \geq 1$ by assumption (A2), we have

$$\max_{1 \leq i \leq n} \sum_{j=0}^m \eta_{ij}^2 = O_P(mn^{1/2}).$$

Taking these together, we have

$$\max_{1 \leq i \leq n} \|\hat{\eta}_i^m\|_{\ell^2}^2 = O_P(nm^{\alpha+1}\Delta^\gamma + n^{1/2}m^{3\alpha+3}\Delta^\gamma + m^{\alpha+3}n^{-1/2} + mn^{1/2}).$$

Since $\alpha > 1$, $\beta > \alpha/2 + 1$ and $m^{3\alpha+3}\Delta^\gamma \rightarrow 0$, there exists a small constant $c > 0$ (depending on α and β) such that the right side is $O_P(n^{-c}(n/m))$. This implies (5.2), completing the proof. $\square$

D.2. Proofs of (5.6) and (5.7). We first prove (5.6). Observe that

$$(h^m \cdot \hat{\eta}_i^m)^2 - (h^m \cdot \eta_i^m)^2 = \{h^m \cdot (\hat{\eta}_i^m - \eta_i^m)\}^2 + 2(h^m \cdot \eta_i^m)\{h^m \cdot (\hat{\eta}_i^m - \eta_i^m)\},$$

by which we have for all $h^m \in \mathbb{S}^m$,

$$\begin{aligned} |\mathbb{E}_n[(h^m \cdot \hat{\eta}_i^m)^2] - \mathbb{E}_n[(h^m \cdot \eta_i^m)^2]| &\leq \mathbb{E}_n[\|\hat{\eta}_i^m - \eta_i^m\|_{\ell^2}^2] \\ &\quad + 2(\mathbb{E}_n[(h^m \cdot \eta_i^m)^2])^{1/2}(\mathbb{E}_n[\|\hat{\eta}_i^m - \eta_i^m\|_{\ell^2}^2])^{1/2}. \end{aligned}$$

By the proof of (5.4), we have

$$\mathbb{E}_n[\|\hat{\eta}_i^m - \eta_i^m\|_{\ell^2}^2] = o_P(1).$$

While by Rudelson's inequality (Theorem E.1 in Appendix E), we have

$$\mathbb{E} \left[\sup_{h^m \in \mathbb{S}^m} |\mathbb{E}_n[(h^m \cdot \eta_i^m)^2] - 1| \right] \leq C \sqrt{\frac{\log n}{n}} \mathbb{E}[\max_{1 \leq i \leq n} \|\eta_i^m\|_{\ell^2}^2],$$

provided that the right side is smaller than 1. Since $\mathbb{E}[\eta_{ij}^4] = \kappa_j^{-2} \mathbb{E}[\xi_{ij}^4] \leq C$ for all $j \geq 1$ by assumption (A2), by Lemma E.1 in Appendix E, we have

$$\mathbb{E}[\max_{1 \leq i \leq n} \|\eta_i^m\|_{\ell^2}^2] = O(mn^{1/2}).$$

Therefore, we conclude that

$$\sup_{h^m \in \mathbb{S}^m} |\mathbb{E}_n[(h^m \cdot \eta_i^m)^2] - 1| = O_P(n^{-1/4} m^{1/2} (\log n)^{1/2}) = o_P(1),$$

so that uniformly in $h^m \in \mathbb{S}^m$,

$$\begin{aligned} \mathbb{E}_n[(h^m \cdot \hat{\eta}_i^m)^2] &= \mathbb{E}_n[(h^m \cdot \eta_i^m)^2] + o_P(1) + O_P(1) \times o_P(1) \\ &= 1 + o_P(1). \end{aligned}$$

This completes the proof of (5.6).

We now turn to prove (5.7). Observe that

$$\hat{r}_i^2(u) \leq 2\{(\hat{\eta}_i^m - \eta_i^m) \cdot d^m(u)\}^2 + 2 \left\{ \sum_{j=m+1}^{\infty} d_j(u) \eta_{ij} \right\}^2.$$

Since $\mathbb{E}[\eta_{ij}] = 0$, $\mathbb{E}[\eta_{ij}^2] = 1$ and $\mathbb{E}[\eta_{ij}\eta_{ik}] = 0$ for all $j \neq k$, we have

$$\begin{aligned} \mathbb{E} \left[\left\{ \sum_{j=m+1}^{\infty} d_j(u) \eta_{ij} \right\}^2 \right] &= \sum_{j=m+1}^{\infty} d_j^2(u) = \sum_{j=m+1}^{\infty} \kappa_j b_j^2(u) \\ &\leq C \sum_{j=m+1}^{\infty} j^{-\alpha-2\beta} = O(mn^{-1}). \end{aligned}$$

Hence, to prove that

$$\sup_{u \in \mathcal{U}} \mathbb{E}_n \left[\left\{ \sum_{j=m+1}^n d_j(u) \eta_{ij} \right\}^2 \right] = O_P(m/n),$$

it suffices to prove that

$$\sup_{u \in \mathcal{U}} \left| \mathbb{E}_n \left[\left\{ \sum_{j=m+1}^n d_j(u) \eta_{ij} \right\}^2 \right] - \mathbb{E} \left[\left\{ \sum_{j=m+1}^n d_j(u) \eta_{ij} \right\}^2 \right] \right| = O_P(m/n).$$

Defining $f_u(X_i) = \sum_{j=m+1}^{\infty} d_j(u) \eta_{ij}$, we wish to show that

$$\mathbb{E} \left[\sup_{u \in \mathcal{U}} \left| \sum_{i=1}^n \{f_u(X_i)^2 - \mathbb{E}[f_u(X_i)^2]\} \right| \right] = O(m).$$

By the symmetrization inequality, the left side is bounded by

$$2\mathbb{E} \left[\sup_{u \in \mathcal{U}} \left| \sum_{i=1}^n \sigma_i f_u(X_i)^2 \right| \right],$$

where $\sigma_1, \dots, \sigma_n$ are i.i.d. Rademacher random variables independent of $X_1, \dots, X_n$. Observe that $|f_u(X_i)| \leq C \sum_{j=m+1}^{\infty} j^{-\beta-\alpha/2} |\eta_{ij}| = F(X_i)$; then by the contraction principle (see van der Vaart and Wellner, 1996, Proposition A.3.2), the above term is further bounded by

$$\begin{aligned} & 8\mathbb{E} \left[\max_{1 \leq i \leq n} F(X_i) \cdot \sup_{u \in \mathcal{U}} \left| \sum_{i=1}^n \sigma_i f_u(X_i) \right| \right] \\ & \leq 8 \sqrt{\mathbb{E}[\max_{1 \leq i \leq n} F(X_i)^2]} \sqrt{\mathbb{E} \left[\sup_{u \in \mathcal{U}} \left| \sum_{i=1}^n \sigma_i f_u(X_i) \right|^2 \right]} \\ & \quad (D.2) \\ & \leq O(1) \left\{ \mathbb{E}[\max_{1 \leq i \leq n} F(X_i)^2] + \sqrt{\mathbb{E}[\max_{1 \leq i \leq n} F(X_i)^2]} \cdot \mathbb{E} \left[\sup_{u \in \mathcal{U}} \left| \sum_{i=1}^n \sigma_i f_u(X_i) \right| \right] \right\}, \end{aligned}$$

where the first inequality follows from the Schwarz inequality and the second inequality follows from the Hoffmann-Jorgensen inequality (see van der Vaart and Wellner, 1996, Proposition A.1.6). Observe that

$$\max_{1 \leq i \leq n} F(X_i) \leq C \sum_{j=m+1}^{\infty} j^{-\beta-\alpha/2} \max_{1 \leq i \leq n} |\eta_{ij}|,$$

and $\mathbb{E}[\max_{1 \leq i \leq n} \eta_{ij}^2] \leq (\mathbb{E}[\max_{1 \leq i \leq n} \eta_{ij}^4])^{1/2} \leq (\sum_{i=1}^n \mathbb{E}[\eta_{ij}^4])^{1/2} \leq Cn^{1/2}$, so that

$$\begin{aligned} \mathbb{E}[\max_{1 \leq i \leq n} F(X_i)^2] &\leq O(1) \left\{ \sum_{j=m+1}^{\infty} j^{-\beta-\alpha/2} \right\} \left\{ \sum_{j=m+1}^{\infty} j^{-\beta-\alpha/2} \mathbb{E}[\max_{1 \leq i \leq n} \eta_{ij}^2] \right\} \\ &= O(n^{1/2} m^{-2\beta-\alpha+2}), \end{aligned}$$

which is $o(1)$ because $\beta + \alpha/2 > \alpha + 1 > 2$ and $m \asymp n^{1/(2\beta+\alpha)}$. Further, because

$$\sum_{i=1}^n \sigma_i f_u(X_i) = \sum_{j=m+1}^{\infty} d_j(u) \sum_{i=1}^n \sigma_i \eta_{ij},$$

we have

$$\begin{aligned} \mathbb{E} \left[\sup_{u \in \mathcal{U}} \left| \sum_{i=1}^n \sigma_i f_u(X_i) \right| \right] &\leq C \sum_{j=m+1}^{\infty} j^{-\beta-\alpha/2} \mathbb{E} \left[\left| \sum_{i=1}^n \sigma_i \eta_{ij} \right| \right] \\ &\leq Cn^{1/2} \sum_{j=m+1}^{\infty} j^{-\beta-\alpha/2} = O(n^{1/2} m^{-\beta-\alpha/2+1}). \end{aligned}$$

Hence the second term in (D.2) is

$$O(n^{3/4} m^{-2\beta-\alpha+2}) = O(m)$$

under our assumption.

On the other hand, by the proof of (5.4), we have

$$\begin{aligned} \sup_{u \in \mathcal{U}} \mathbb{E}_n[\{(\hat{\eta}_i^m - \eta_i^m) \cdot d^m(u)\}^2] &\leq C \sum_{j=1}^m j^{-\alpha-2\beta} \mathbb{E}_n[(\hat{\eta}_{ij} - \eta_{ij})^2] \\ &= O_P(\sum_{j=1}^m j^{-\alpha-2\beta} (n^{-1} j^{\alpha+2} + \Delta^\gamma j^{3\alpha+2})), \end{aligned}$$

where

$$\sum_{j=1}^m j^{-2\beta+2} = O(1)$$

and

$$\sum_{j=1}^m j^{-2\beta+2\alpha+2} = \begin{cases} O(1) & \text{if } -2\beta + 2\alpha + 2 < -1 \\ O(\log n) & \text{if } -2\beta + 2\alpha + 2 = -1 \\ O(m^{-2\beta+2\alpha+3}) & \text{if } -2\beta + 2\alpha + 2 > -1. \end{cases}$$

Since $m^{-2\beta+2\alpha+3} \asymp n^{-1}m^{3\alpha+3}$ and $m^{3\alpha+3} \asymp n$ when $-2\beta + 2\alpha + 2 = -1$, we have

$$\begin{aligned} \sum_{j=1}^m j^{-\alpha-2\beta} (n^{-1}j^{\alpha+2} + \Delta^\gamma j^{3\alpha+2}) &= O(n^{-1} + \Delta^\gamma + n^{-1}(\log n)m^{3\alpha+3}\Delta^\gamma) \\ &= O(n^{-1}). \end{aligned}$$

Taking these together, we obtain (5.7). This completes the proof. $\square$

D.3. Proof of (5.9). Consider the classes of functions

$$\begin{aligned} \mathcal{G}_1 = \Big\{ \mathbb{R} \times D[0, 1] \times \mathbb{R}^{m+1} \ni (y, x, \eta^m) \mapsto 1(y \leq \eta^m \cdot (d^m(u) + \delta^m))(h^m \cdot \eta^m) \\ : u \in \mathcal{U}, h^m \in \mathbb{S}^m, \delta^m = M\sqrt{m/n}h^m \Big\}, \end{aligned}$$

and

$$\begin{aligned} \mathcal{G}_2 = \Big\{ \mathbb{R} \times D[0, 1] \times \mathbb{R}^{m+1} \ni (y, x, \eta^m) \mapsto 1(y \leq Q_{Y|X}(u | x))(h^m \cdot \eta^m) \\ : u \in \mathcal{U}, h^m \in \mathbb{S}^m \Big\}. \end{aligned}$$

It is relatively standard to see that $\mathcal{G}_1$ is a VC subgraph class with VC index bounded by cm for some constant $c \geq 1$ (see Belloni et al., 2011, Lemma 18). For $\mathcal{G}_2$, observe first that $1(y \leq Q_{Y|X}(u | x)) = 1(F_{Y|X}(y|x) \leq u)$. Since $F_{Y|X}(y|x)$ is a fixed function, it is also shown that $\mathcal{G}_2$ is a VC subgraph class with VC index bounded by $c'm$ for some constant $c' \geq 1$. The conclusion now follows from an application of Theorem 2.6.7 of van der Vaart and Wellner (1996) and a simple covering number calculation. $\square$

APPENDIX E: USEFUL INEQUALITIES

We introduce some useful inequalities.

THEOREM E.1 (Rudelson's (1999) inequality). *Let $Z_1, \dots, Z_n$ be i.i.d. random vectors in $\mathbb{R}^k$ with $\Sigma := E[Z_1 Z_1']$. Then, for all $k \geq 2$,*

$$\mathbb{E} \left[\left\| \frac{1}{n} \sum_{i=1}^n Z_i Z_i' - \Sigma \right\|_{\text{op}} \right] \leq \max\{\|\Sigma\|_{\text{op}}^{1/2} \delta, \delta^2\}, \quad \delta = C \sqrt{\frac{\log k}{n} \mathbb{E}[\max_{1 \leq i \leq n} \|Z_i\|_{\ell^2}^2]},$$

where $\|\cdot\|_{\text{op}}$ is the operator norm and C is a universal constant.

The expression of Theorem E.1 is slightly different from Rudelson's original form, but is directly deduced from his proof. Theorem E.1 gives moment bounds on the difference between empirical and population Gram matrices in the operator norm. Recall that for any $k \times k$ symmetric matrix A , $\|A\|_{\text{op}} = \max_{v \in \mathbb{S}^{k-1}} |v'Av|$. To apply Rudelson's inequality, we have to bound $\mathbb{E}[\max_{1 \leq i \leq n} \|Z_i\|_{\ell_2}^2]$, which is typically implemented by using the following lemma.

LEMMA E.1. *Let $X_1, \dots, X_n$ be arbitrary scalar random variables such that $\max_{1 \leq i \leq n} \mathbb{E}[|X_i|^r] < \infty$ for some $r \geq 1$. Then, we have*

$$\mathbb{E}[\max_{1 \leq i \leq n} |X_i|] \leq C_r n^{1/r},$$

where C_r is a constant depending only on r and $\max_{1 \leq i \leq n} \mathbb{E}[|X_i|^r]$.

For the proof, see van der Vaart and Wellner (1996, Lemma 2.2.2).

In what follows, we introduce “conditional” maximal inequalities. Below we assume the class of functions to be a “pointwise measurable class” to avoid a measurability complication. A class of measurable functions $\mathcal{G}$ on a measurable space S is said to be pointwise measurable if there exists a countable class of measurable functions $\mathcal{H}$ on S such that for any $g \in \mathcal{G}$, there exists a sequence $\{h_m\} \subset \mathcal{H}$ with $h_m(x) \rightarrow g(x)$ for all $x \in S$. See Chapter 2.3 of van der Vaart and Wellner (1996). This condition is satisfied in our application.

PROPOSITION E.1. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ denote the underlying probability space. Let $\mathcal{D}$ be a sub σ -field of $\mathcal{A}$. Let $\{(u_i, v_i)\}_{i=1}^n$ be a sequence of random variables taking values in some measurable space S such that $v_1, \dots, v_n$ are $\mathcal{D}$ -measurable, the regular conditional distribution of $(u_1, \dots, u_n)$ given $\mathcal{D}$ exists, and conditional on $\mathcal{D}$, $u_1, \dots, u_n$ are independent. Let $\mathcal{G}$ be a pointwise measurable class of functions on S such that for some $\mathcal{D}$ -measurable random variables $\hat{B}$ and $\hat{\tau}$,*

- (i) $\sup_{g \in \mathcal{G}} |g(u_i, v_i)| \leq \hat{B}, \quad 1 \leq i \leq n,$
- (ii) $\sup_{g \in \mathcal{G}} \mathbb{E}_n[\mathbb{E}[g^2(u_i, v_i) \mid \mathcal{D}]] \leq \hat{\tau}^2,$
- (iii) $\hat{\tau} \leq \hat{B},$

almost surely. Suppose that there exist constants $A \geq 3\sqrt{e}$ and $W \geq 1$ such that

$$(E.1) \quad N(\hat{B}\epsilon, \mathcal{G}, L_2(P_n)) \leq (A/\epsilon)^W, \quad 0 < \forall \epsilon \leq 1,$$

where P_n denotes the empirical distribution on S that assigns probability n^{-1} to each (u_i, v_i) , $i = 1, \dots, n$. Let $\sigma_1, \dots, \sigma_n$ be independent Rademacher random variables defined on another probability space. Extend the underlying probability space by the product probability space. Then, we have

$$\begin{aligned} \mathbb{E} \left[\sup_{g \in \mathcal{G}} |\mathbb{E}_n[\sigma_i g(u_i, v_i)]| \mid \mathcal{D} \right] \\ \leq 1(\hat{\tau} > 0)D \left[\sqrt{\frac{\hat{\tau}^2 W}{n} \log \frac{A\hat{B}}{\hat{\tau}}} + \frac{W\hat{B}}{n} \log \frac{A\hat{B}}{\hat{\tau}} \right], \text{ a.s.,} \end{aligned}$$

where D is a universal constant.

Proposition E.1 is a conditional version of Proposition 2.1 of Giné and Guillou (2001). Here, $\{(u_i, v_i)\}_{i=1}^n$ are not necessarily independent. However, conditional on $\mathcal{D}$, $\{u_i\}_{i=1}^n$ are independent.

PROOF OF PROPOSITION E.1. The proof is a modification of that of Giné and Guillou (2001, Proposition 2.1). For the sake of completeness, we provide the full proof. Without loss of generality we may assume that $\mathcal{G}$ contains the 0-function. Suppose that $\mathbb{E}[\sup_{g \in \mathcal{G}} \mathbb{E}_n[g^2(u_i, v_i)] \mid \mathcal{D}] \wedge \hat{\tau} > 0$. Otherwise the conclusion follows trivially. By Dudley's inequality (see van der Vaart and Wellner, 1996, Corollary 2.2.8), we have

$$\mathbb{E} \left[\sup_{g \in \mathcal{G}} |\sqrt{n} \mathbb{E}_n[\sigma_i g(u_i, v_i)]| \mid \{(u_i, v_i)\}_{i=1}^n \right] \leq D \int_0^\theta \sqrt{\log N(\epsilon, \mathcal{G}, L_2(P_n))} d\epsilon,$$

where $\theta := (\sup_{g \in \mathcal{G}} \mathbb{E}_n[g^2(u_i, v_i)])^{1/2}$ and D is a universal constant. Suppose that $\theta > 0$. Using changes of variables, we have

$$\begin{aligned} \int_0^\theta \sqrt{\log N(\epsilon, \mathcal{G}, L_2(P_n))} d\epsilon &= \hat{B} \int_0^{\theta/\hat{B}} \sqrt{\log N(\hat{B}\epsilon, \mathcal{G}, L_2(P_n))} d\epsilon \\ &\leq \hat{B} \sqrt{W} \int_0^{\theta/\hat{B}} \sqrt{\log(A/\epsilon)} d\epsilon \\ &\leq \hat{B} A \sqrt{W} \int_{A\hat{B}/\theta}^\infty \frac{\sqrt{\log \epsilon}}{\epsilon^2} d\epsilon. \end{aligned} \tag{E.2}$$

Integration by parts gives

$$\begin{aligned} \int_c^\infty \frac{\sqrt{\log \epsilon}}{\epsilon^2} d\epsilon &= \left[-\frac{\sqrt{\log \epsilon}}{\epsilon} \right]_c^\infty + \frac{1}{2} \int_c^\infty \frac{1}{\epsilon^2 \sqrt{\log \epsilon}} d\epsilon \\ &\leq \frac{\sqrt{\log c}}{c} + \frac{1}{2} \int_c^\infty \frac{\sqrt{\log \epsilon}}{\epsilon^2} d\epsilon, \end{aligned}$$

provided that $c \geq e$, by which we have

$$\int_c^\infty \frac{\sqrt{\log \epsilon}}{\epsilon^2} d\epsilon \leq \frac{2\sqrt{\log c}}{c}, \text{ if } c \geq e.$$

Since $A\hat{B}/\theta \geq A \geq 3\sqrt{e} > e$, we have

$$(E.2) \leq 2\sqrt{W}\theta\sqrt{\log(A\hat{B}/\theta)},$$

by which we have, using Hölder's inequality,

$$\mathbb{E}[(E.2) \mid \mathcal{D}] \leq \sqrt{2W} \sqrt{\mathbb{E} \left[1(\theta > 0) \theta^2 \log \frac{A^2 \hat{B}^2}{\theta^2} \mid \mathcal{D} \right]}.$$

For any fixed $c > 0$, define $f(u) = u \log(c/u)$ if $u > 0$ and $f(0) = 0$. Then, $f(u)$ is concave on $[0, \infty)$. Thus, by Jensen's inequality, the last expression is bounded by

$$\sqrt{2W} \sqrt{\mathbb{E}[\sup_{g \in \mathcal{G}} \mathbb{E}_n[g^2(u_i, v_i)] \mid \mathcal{D}] \times \log \frac{A^2 \hat{B}^2}{\mathbb{E}[\sup_{g \in \mathcal{G}} \mathbb{E}_n[g^2(u_i, v_i)] \mid \mathcal{D}]}}.$$

Using the decomposition

$$g^2(u_i, v_i) = \mathbb{E}[g^2(u_i, v_i) \mid \mathcal{D}] + \{g^2(u_i, v_i) - \mathbb{E}[g^2(u_i, v_i) \mid \mathcal{D}]\},$$

and the symmetrization inequality conditional on $\mathcal{D}$, we have

$$\begin{aligned} & \mathbb{E} \left[\sup_{g \in \mathcal{G}} \mathbb{E}_n[g^2(u_i, v_i)] \mid \mathcal{D} \right] \\ & \leq \sup_{g \in \mathcal{G}} \mathbb{E}_n[\mathbb{E}[g^2(u_i, v_i) \mid \mathcal{D}]] + 2\mathbb{E} \left[\sup_{g \in \mathcal{G}} |\mathbb{E}_n[\sigma_i g^2(u_i, v_i)]| \mid \mathcal{D} \right] \\ & \leq \hat{\tau}^2 + 2\mathbb{E} \left[\sup_{g \in \mathcal{G}} |\mathbb{E}_n[\sigma_i g^2(u_i, v_i)]| \mid \mathcal{D} \right]. \end{aligned}$$

Using now the contraction principle (see van der Vaart and Wellner, 1996, Proposition A.3.2), we have

$$\begin{aligned} & \mathbb{E} \left[\sup_{g \in \mathcal{G}} |\mathbb{E}_n[\sigma_i g^2(u_i, v_i)]| \mid \{(u_i, v_i)\}_{i=1}^n \right] \\ & \leq 4\hat{B} \mathbb{E} \left[\sup_{g \in \mathcal{G}} |\mathbb{E}_n[\sigma_i g(u_i, v_i)]| \mid \{(u_i, v_i)\}_{i=1}^n \right], \end{aligned}$$

so that

$$\mathbb{E} \left[\sup_{g \in \mathcal{G}} \mathbb{E}_n[g^2(u_i, v_i)] \mid \mathcal{D} \right] \leq \hat{\tau}^2 + 8\hat{B} \mathbb{E} \left[\sup_{g \in \mathcal{G}} |\mathbb{E}_n[\sigma_i g(u_i, v_i)]| \mid \mathcal{D} \right].$$

Note that the right side is at most $9\hat{B}^2$. Since for any given $c > 0$, the map $u \mapsto u \log(c/u)$ is non-decreasing for $0 < u \leq c/e$, and $A \geq 3\sqrt{e}$, we have

$$\begin{aligned} & \mathbb{E}[\sup_{g \in \mathcal{G}} \mathbb{E}_n[g^2(u_i, v_i)] \mid \mathcal{D}] \times \log \frac{A^2 \hat{B}^2}{\mathbb{E}[\sup_{g \in \mathcal{G}} \mathbb{E}_n[g^2(u_i, v_i)] \mid \mathcal{D}]} \\ & \leq (\hat{\tau}^2 + 8\hat{B}Z) \log \frac{A^2 \hat{B}^2}{\hat{\tau}^2 + 8\hat{B}Z} \\ & \leq (\hat{\tau}^2 + 8\hat{B}Z) \log \frac{A^2 \hat{B}^2}{\hat{\tau}^2} \\ & = 2(\hat{\tau}^2 + 8\hat{B}Z) \log \frac{A\hat{B}}{\hat{\tau}}, \end{aligned}$$

where

$$Z := \mathbb{E} \left[\sup_{g \in \mathcal{G}} |\mathbb{E}_n[\sigma_i g(u_i, v_i)]| \mid \mathcal{D} \right].$$

Taking these together, we have

$$\sqrt{n}Z \leq 2D\sqrt{W} \sqrt{(\hat{\tau}^2 + 8\hat{B}Z) \log \frac{A\hat{B}}{\hat{\tau}}}.$$

Solving this inequality with respect to Z gives the desired bound. $\square$

PROPOSITION E.2. *Consider the same setting as in Proposition E.1. Instead of (i)-(iii) and (E.1), suppose that there is an envelope function G for $\mathcal{G}$ such that for some constants $A \geq e$ and $W \geq 1$,*

$$N(\epsilon \|G\|_{L_2(P_n)}, \mathcal{G}, L_2(P_n)) \leq (A/\epsilon)^W, \quad 0 < \forall \epsilon \leq 1.$$

Then, we have for all $q \in [1, \infty)$,

$$\begin{aligned} & \left(\mathbb{E} \left[\sup_{g \in \mathcal{G}} |\mathbb{E}_n[\sigma_i g(u_i, v_i)]|^q \mid \mathcal{D} \right] \right)^{1/q} \\ & \leq Dn^{-1/2} (\mathbb{E}_n[\mathbb{E}[G^{\bar{q}}(u_i, v_i) \mid \mathcal{D}]]^{1/\bar{q}} \sqrt{W \log A} \text{ a.s.}, \end{aligned}$$

where $\bar{q} = q \vee 2$ and D is a constant depending only on q .

PROOF. Without loss of generality we may assume that $\mathcal{G}$ contains the 0-function. By Dudley's inequality,

$$\begin{aligned} & \left(\mathbb{E} \left[\sup_{g \in \mathcal{G}} |\sqrt{n} \mathbb{E}_n[\sigma_i g(u_i, v_i)]|^q \mid \{(u_i, v_i)\}_{i=1}^n \right] \right)^{1/q} \\ & \leq D \int_0^\theta \sqrt{\log N(\epsilon, \mathcal{G}, L_2(P_n))} d\epsilon, \end{aligned}$$

where $\theta := (\sup_{g \in \mathcal{G}} \mathbb{E}_n[g^2(u_i, v_i)])^{1/2} \leq (\mathbb{E}_n[G^2(u_i, v_i)])^{1/2}$ and D is a constant depending only on q . Using changes of variables, we have that the right side is bounded by

$$D(\mathbb{E}_n[G^2(u_i, v_i)])^{1/2} \sqrt{W} \int_0^1 \sqrt{\log(A/\epsilon)} d\epsilon.$$

If $q \geq 2$, then by Hölder's inequality,

$$\mathbb{E}[(\mathbb{E}_n[G^2(u_i, v_i)])^{q/2} \mid \mathcal{D}] \leq \mathbb{E}[\mathbb{E}_n[G^q(u_i, v_i)] \mid \mathcal{D}] = \mathbb{E}_n[\mathbb{E}[G^q(u_i, v_i) \mid \mathcal{D}]].$$

On the other hand, if $q \in [1, 2)$,

$$\mathbb{E}[(\mathbb{E}_n[G^2(u_i, v_i)])^{q/2} \mid \mathcal{D}] \leq (\mathbb{E}_n[\mathbb{E}[G^2(u_i, v_i) \mid \mathcal{D}]])^{q/2}.$$

This leads to the desired inequality. $\square$

APPENDIX F: TABLE AND FIGURES FOR CASE (B) IN SECTION 4

REFERENCES

- Belloni, A., Chernozhukov, V. and Fernández-Val, I. (2011). Conditional quantile processes based on series or many regressors. arXiv:1105.6154.
- Bosq, D. (2000). *Linear Processes in Function Spaces*. Lecture Notes in Statistics. Springer.
- Giné, E. and Guillou, A. (2001). On consistency of kernel density estimators for randomly censored data: rates holding uniformly over adaptive intervals. *Ann. Inst. H. Poincaré Probab. Statist.* **37** 503-522.
- Hall, P. and Horowitz, J.L. (2007). Methodology and convergence rates for functional linear regression. *Ann. Statist.* **35** 70-91.
- Hofmann, B. and Scherzer, O. (1998). Local ill-posedness and source conditions of operator equations in Hilbert space. *Inverse Problems* **14** 1189-1206.
- Marcus, M.B. (1975). Uniform convergence of random Fourier series. *Ark. Mat.* **13** 107-122.
- Rudelson, M. (1999). Random vectors in the isotropic position. *J. Functional Anal.* **164** 60-72.

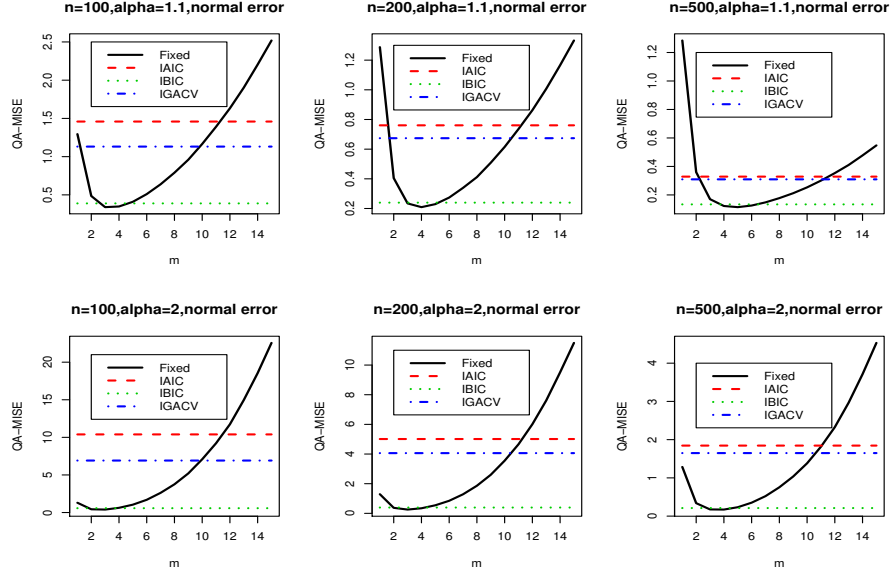
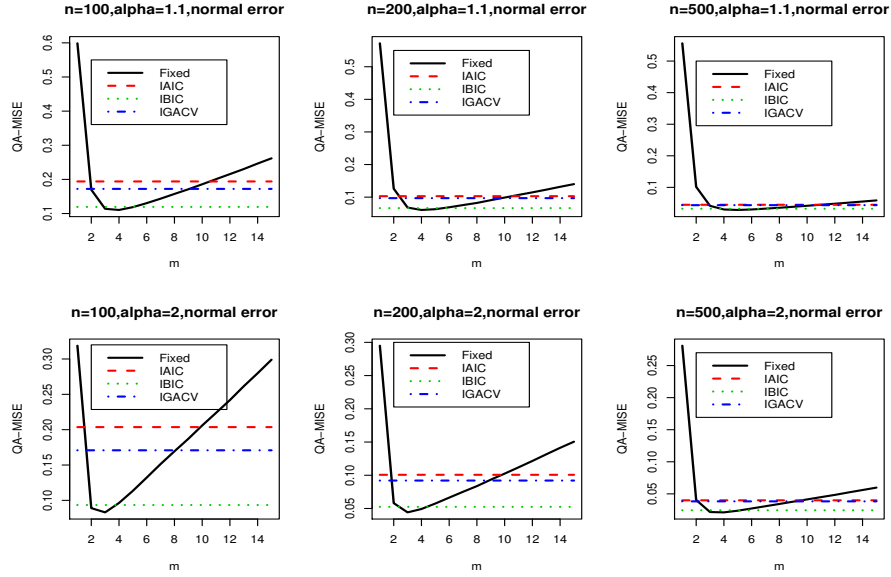
FIG 1. *Performance of selection criteria. Case (b). Estimation of $(t, u) \mapsto b(t, u)$.*FIG 2. *Performance of selection criteria. Case (b). Estimation of $Q_{Y|X}(u | x)$.*

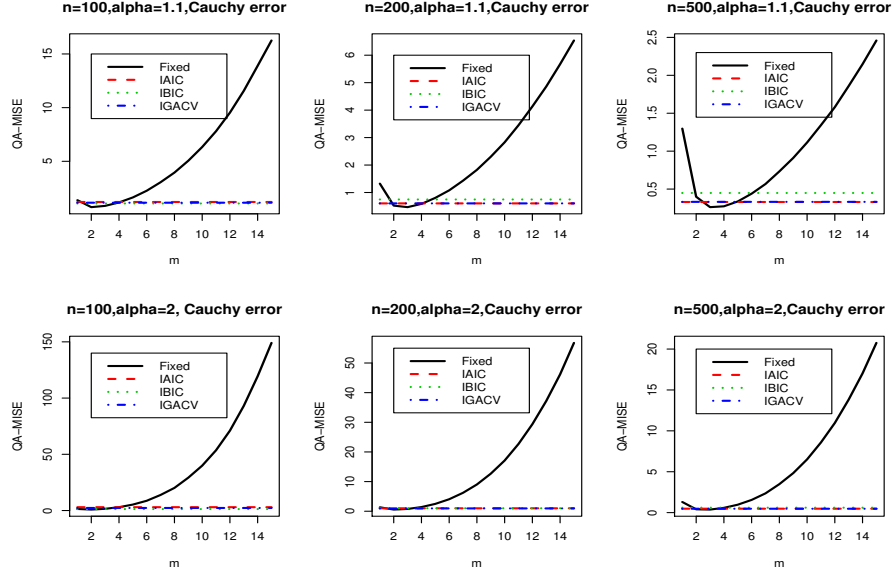
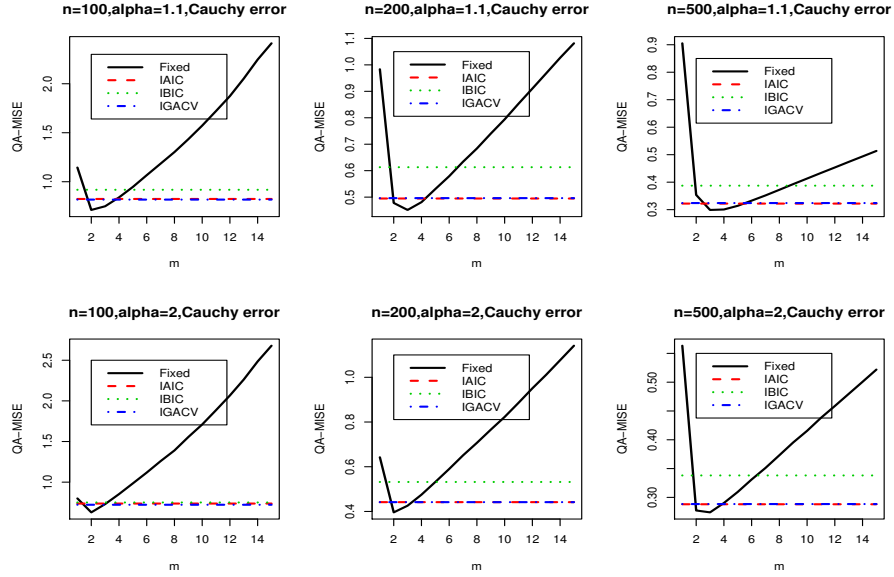
FIG 3. *Performance of selection criteria. Case (b). Estimation of $(t, u) \mapsto b(t, u)$.*FIG 4. *Performance of selection criteria. Case (b). Estimation of $Q_{Y|X}(u | x)$.*

TABLE 1
*Average numbers of m selected by three criteria in case (b).
Standard deviations are given in parentheses.*

n	α	error dist.	AIC	BIC	GACV
100	1.1	normal	8.76 (4.33)	3.12 (1.17)	7.25 (3.95)
200	1.1	normal	8.54 (4.07)	3.47 (1.11)	7.78 (3.86)
500	1.1	normal	9.09 (3.69)	4.01 (0.99)	8.66 (3.60)
100	2	normal	8.06 (4.59)	2.52 (0.84)	6.37 (4.12)
200	2	normal	7.75 (4.30)	2.72 (0.87)	6.82 (4.03)
500	2	normal	7.69 (3.96)	3.09 (0.67)	7.23 (3.82)
100	1.1	Cauchy	2.36 (1.53)	1.50 (0.59)	2.25 (1.34)
200	1.1	Cauchy	2.43 (1.03)	1.76 (0.56)	2.42 (1.01)
500	1.1	Cauchy	2.86 (0.92)	2.06 (0.48)	2.83 (0.90)
100	2	Cauchy	1.99 (1.42)	1.27 (0.47)	1.86 (1.14)
200	2	Cauchy	2.06 (0.88)	1.44 (0.53)	2.04 (0.83)
500	2	Cauchy	2.30 (0.66)	1.80 (0.45)	2.28 (0.63)

van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer.

Yuan, M. and Cai, T. (2010). A reproducing kernel Hilbert space approach to functional linear regression. *Ann. Statist.* **38** 3412-3444.

DEPARTMENT OF MATHEMATICS
GRADUATE SCHOOL OF SCIENCE
HIROSHIMA UNIVERSITY
1-3-1 KAGAMIYAMA, HIGASHI-HIROSHIMA
HIROSHIMA 739-8526, JAPAN.
E-MAIL: kkato@hiroshima-u.ac.jp