SUPPLEMENT FOR "ESTIMATION OF CONDITIONAL EXTREMILES IN REPRODUCING KERNEL HILBERT SPACES WITH APPLICATION TO LARGE COMMERCIAL BANKS DATA"

Fang Chen[1], Caixing Wang[2]

[1]*Nanjing Forestry University and* [2]*Southeast University*

In this Supplementary Material, Section S1 presents several useful lemmas, while Section S2 provides the detailed proofs for Proposition 1 as well as Theorems 1 and 2.

S1 Some Useful Lemmas

Proposition 2. *Suppose that $\mathbb{E}\left(y \mid \mathbf{x}\right)$ is uniformly bounded,*

(i) If $F \in \mathrm{D}\left(\Phi_{\gamma(\mathbf{x})}\right)$ with $\gamma(\mathbf{x})<1$, as $\tau \rightarrow 1$ then

$$\frac{\xi_\tau(\mathbf{x})}{q_\tau(\mathbf{x})} \sim \Gamma(1-\gamma(\mathbf{x}))\{\log 2\}^{\gamma(\mathbf{x})}.$$

(ii) If $F \in \mathrm{D}\left(\Psi_{\gamma(\mathbf{x})}\right)$ as $\tau \rightarrow 1$ then $y_u<\infty$

$$\frac{y_u-\xi_\tau(\mathbf{x})}{y_u-q_\tau(\mathbf{x})} \sim \Gamma(1-\gamma(\mathbf{x}))\{\log 2\}^{\gamma(\mathbf{x})}.$$

(iii) If $F \in \mathrm{DA}(\Lambda)$ and $y_u = \infty$ as $\tau \to 1$, when $y_u = \infty$, then $\xi_\tau(\mathbf{x}) \sim q_\tau(\mathbf{x})$. If $y_u < \infty$, we have

$$y_u - \xi_\tau(\mathbf{x}) \sim y_u - q_\tau(\mathbf{x}).$$

Proof of Proposition 2: The proof is entirely similar to that of Proposition 3 in Daouia et al. (2019). ■

We define $\widetilde{q}_{\tau_n} = \operatorname{argmin}_{q_{\tau_n} \in \mathcal{H}_K} \mathbb{E}[\rho_{\tau_n}(y - q_{\tau_n}(\mathbf{x}))] + \lambda_{2n} \|q_{\tau_n}\|_K$. The following lemma shows the convergence result of the difference between $\widehat{q}_{\tau_n}$ and $\widetilde{q}_{\tau_n}$.

Lemma 1. *Suppose that Assumption 1 is satisfied. For any $\delta_{1,n} > 4(\log n)^{-2}\mathbb{E}(y^2)$, with probability at least $1 - \delta_{1,n}/2$, there holds*

$$\|\widehat{q}_{\tau_n} - \widetilde{q}_{\tau_n}\|_K \leq C \left(\log \frac{8}{\delta_{1,n}}\right)^{1/4} \frac{(\log n)^{1/2} \left(1 + \kappa \lambda_{2n}^{-1/2}\right)^{1/2}}{\lambda_{2n}^{1/2} n^{1/4}},$$

where $C = 4 \max\{1, \kappa\}$.

Proof of Lemma 1: We first set two events that

$$\mathcal{C}_1 = \left\{ \mathcal{Z}^n : \|\widehat{q}_{\tau_n} - \widetilde{q}_{\tau_n}\|_K \geq c_2 \left(\log \frac{8}{\delta_{1,n}}\right)^{1/4} \frac{(\log n)^{1/2} \left(1 + \kappa \lambda_{2n}^{-1/2}\right)^{1/2}}{\lambda_{2n}^{1/2} n^{1/4}} \right\},$$

and

$$\mathcal{C}_2 = \{y : |y| > \log n\},$$

and denote $\mathcal{C}_2^c$ as the complement of $\mathcal{C}_2$. Clearly, $\mathbb{P}(\mathcal{C}_1)$ can be decomposed as

$$\mathbb{P}(\mathcal{C}_1) = \mathbb{P}(\mathcal{C}_1 \cap \mathcal{C}_2) + \mathbb{P}(\mathcal{C}_1 \cap \mathcal{C}_2^c) \leq \mathbb{P}(\mathcal{C}_2) + \mathbb{P}(\mathcal{C}_1 \mid \mathcal{C}_2^c) := P_1 + P_2.$$

To bound P_1, use the Markov's inequality, there holds

$$\mathbb{P}\left(|y| > \log n\right) \leq \frac{\mathbb{E}\left(y^2\right)}{\log^2 n}.$$

For simplicity, denote

$$\mathcal{E}\left(q_{\tau_n}\right) = \mathbb{E}_{\mathcal{Z}^n} \rho_{\tau_n}\left(y - q_{\tau_n}(\mathbf{x})\right),$$

$$\mathcal{E}^{\lambda_{2n}}\left(q_{\tau_k}\right) = \mathcal{E}\left(q_{\tau_n}\right) + \lambda_{2n} \left\|q_{\tau_n}\right\|_K^2,$$

$$\mathcal{E}_{\mathcal{Z}^n}^{\lambda_{2n}}\left(q_{\tau_k}\right) = \mathcal{E}_{\mathcal{Z}^n}\left(q_{\tau_n}\right) + \lambda_{2n} \left\|q_{\tau_n}\right\|_K^2.$$

Conditioning on the event $\{\mathcal{Z}^n : \max_{i=1,\ldots,n} |y_i| \leq M_n\}$, where $\mathcal{Z} := (\mathbb{R}, \mathcal{X})$, we consider the functional space

$$\mathcal{F}_{M_n} = \left\{f \in \mathcal{H}_K : \|f\|_K^2 \leq \lambda_{2n}^{-1} M_n\right\}.$$

Note that $\mathcal{F}_{M_n}$ is fairly large in the sense that the minimizer $\widehat{q}_{\tau_n}$ is contained in $\mathcal{F}_{M_n}$ by the fact that

$$\lambda_{2n} \left\|\widehat{q}_{\tau_n}\right\|_K^2 \leq \mathcal{E}_{\mathcal{Z}^n}\left(\widehat{q}_{\tau_n}\right) + \lambda_{2n} \left\|\widehat{q}_{\tau_n}\right\|_K^2 \leq \mathcal{E}_{\mathcal{Z}^n}(0) + \lambda_{2n} \|0\|_K^2 \leq \max_{i=1,\ldots,n} |y_i| \leq M_n.$$

Directly by Proposition 2 and Theorems 2.6 and 2.7 in Villa et al. (2012), there holds

$$\begin{aligned}
\Psi_{\lambda_{2n}}^{\circ}\left(\|\widehat{q}_{\tau_n} - \widetilde{q}_{\tau_n}\|_K\right) &\leq 4 \sup_{q \in \mathcal{F}_{M_n}} \left| t_{\mathcal{E}^{\lambda_{2n}}} \mathcal{E}^{\lambda_{2n}}\left(q_{\tau_n}\right) - t_{\mathcal{E}^{\lambda_{2n}}} \mathcal{E}_{\mathcal{Z}^n}^{\lambda_{2n}}\left(q_{\tau_n}\right) \right| \\
&\leq 4 \sup_{q \in \mathcal{F}_{M_n}} \left| \mathcal{E}\left(q_{\tau_n}\right) - \mathcal{E}_{\mathcal{Z}^n}\left(q_{\tau_n}\right) \right|,
\end{aligned}$$

where $\Psi^{\circ}_{\lambda_{2n}}(t) = \inf\left\{\frac{\lambda_{2n}s^2}{2} + |t-s| : s \in [0,\infty)\right\}$ and $t_{\mathcal{E}^{\lambda_{2n}}}$ is the translation map defined as $t_{\mathcal{E}^{\lambda_{2n}}} G\left(q_{\tau_n}\right) = G\left(q_{\tau_n} + \widetilde{q}_{\tau_n}\right) - \mathcal{E}^{\lambda_{2n}}\left(\widetilde{q}_{\tau_n}\right)$ for all $G : \mathcal{H}_K \to \mathbb{R}$. Since is invertible and increasing, we can write its inverse $\left(\Psi^{\diamond}_{\lambda_{2n}}\right)^{-1}$ explicitly as

$$\left(\Psi^{\diamond}_{\lambda_{2n}}\right)^{-1}(t) = \begin{cases} \sqrt{2t/\lambda_{2n}}, & \text{if } t < 1/(2\lambda_{2n}); \\ t + 1/(2\lambda_{2n}), & \text{otherwise }. \end{cases}$$

When the upper bound of $\Psi^{\circ}_{\lambda_{2n}}\left(\|\widehat{q}_{\tau_n} - \tilde{q}_{\tau_n}\|_K\right)$ is sufficiently small, we have

$$\|\widehat{q}_{\tau_n} - \widetilde{q}_{\tau_n}\|_{\mathcal{H}_K} \le \frac{2\sqrt{2}}{\lambda_{2n}^{1/2}} \left(\sup_{q \in \mathcal{F}_{M_n}} \left|\mathcal{E}\left(q_{\tau_n}\right) - \mathcal{E}_{\mathcal{Z}^n}\left(q_{\tau_n}\right)\right|\right)^{1/2}.$$

Followed by the Lemma 2 and 3 of Chen et al. (2021), with probability at least $1 - \delta_{1,n}/2$, for some constant, it holds

$$\|\widehat{q}_{\tau_n} - \widetilde{q}_{\tau_n}\|_K \le C\left(\log\frac{8}{\delta_{1,n}}\right)^{1/4} \frac{(\log n)^{1/2}\left(1 + \kappa\lambda_{2n}^{-1/2}\right)^{1/2}}{\lambda_{2n}^{1/2} n^{1/4}},$$

where $C = 4\max\{1,\kappa\}$. Therefore, we have $P_2 \le \delta_{1,n}/4$, and thus for any $\delta_{1,n} > 4(\log n)^{-2}\mathbb{E}\left(y^2\right)$, $\mathbb{P}\left(\mathcal{C}_1\right) \le P_1 + P_2 \le \delta/2$. The desired results follow immediately. ■

S2 Proofs of Main Results

S2.1 Proof of Proposition 1

Define a piece-wise linear function $\widehat{q}(\cdot)$, satisfying $\widehat{q}\,(\tau_k) = \hat{q}_{\tau_k}, \quad \forall \tau_k \in \Omega$. For brevity, we represent $\widehat{\boldsymbol{\theta}} = \left(\widehat{q}_{\tau_1}, \widehat{q}_{\tau_2}, \ldots, \widehat{q}_{\tau_{s_n}}\right)^\top$ as a s_n dimension estimated quantile vector. $\widetilde{\boldsymbol{\theta}} = \left(\tilde{q}_{\tau_1}, \widetilde{q}_{\tau_2}, \ldots, \widetilde{q}_{\tau_{s_n}}\right)^\top$ is the quantile vector obtained by minimizing the expected loss function. Denote $\boldsymbol{\theta}^* = \left(q^*_{\tau_1}, q^*_{\tau_2}, \ldots, q^*_{\tau_{s_n}}\right)^\top$ as the ture quantile vector. Define a tight set $B_\tau = \{q \in \mathcal{H}_K : \|q - q^*_\tau\|_K < \eta\}$ for arbitrary $\eta > 0$. In terms of that the true quantile function belongs to the RKHS, so we have $\lim_{\lambda_{2n} \to 0} \|\tilde{q}_\tau - q^*_\tau\| = 0$. Then as $s_n \to \infty$, it holds that $s_n^{-1}\|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\| \to 0$, which means there exists S_η such that for any $s_n > S_\eta$, $s_n^{-1}\|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\| < \eta/2$. Then it can be inferred that $\widetilde{\boldsymbol{\theta}} \in B_\tau \times \Omega$. As $\|\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}}\| < s_n^{1/2}\kappa \max_{\tau \in \Omega} \|\widehat{q}_\tau - \widetilde{q}_\tau\|_K$, based on lemma 1, when $\frac{s_n^{1/2}(\log\log n)^{1/4}(1+\kappa\lambda_{2n}^{-\frac{1}{2}})^{1/2}}{n^{1/4}} \to 0$, it holds that $\lim_{n\to\infty} \|\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}}\| = 0$, $\lim_{n\to\infty} \mathbb{P}\left(\widehat{\boldsymbol{\theta}} \in B_\tau \times \Omega\right) = 1$. Then we have

$$\sup_{\tau \in \left[\frac{1}{s_n+1}, \frac{s_n}{s_n+1}\right]} \|\widehat{q}_\tau - q^*_\tau\|_K = o_p(1).$$

$\blacksquare$

S2.2 Proof of Theorem 1

Before giving the proof of Theorem 1, we first define the data-free and noise-free version of our estimator $\widehat{\xi}_{\tau_n}$,

$$\widetilde{\xi}_{\tau_n} = \underset{f \in \mathcal{H}_K}{\operatorname{argmin}} \left\{ \int_{\mathcal{X}} (f(\mathbf{x})) - \xi_{\tau_n}(\mathbf{x}))^2 d\rho_{\mathbf{x}} + \lambda_{1n} \|f\|_K^2 \right\}.$$

The following lemma shows the empirical error between $\widehat{\xi}_{\tau_n}$ and $\widetilde{\xi}_{\tau_n}$.

Lemma 2. *Suppose Assumption 1 is satisfied, then conditioning on the event* $\{\mathcal{Z}^n : \max_{i=1,\ldots,n} |y_i| \leq M_n\}$ *with* $M_n \geq (\kappa^2 |f_0|_K^2 + \sigma^2)^{1/2}$*, for any* $\delta_n \in (0,1)$*, with probability at least* $1 - \delta_n$*, there holds*

$$\|\widehat{\xi}_{\tau_n} - \widetilde{\xi}_{\tau_n}\|_K \leq \log \frac{2}{\delta_n} \left(\frac{6\kappa r_{\tau_n}}{\lambda_{1n} n^{1/2}} + \frac{C\kappa r_{\tau_n}^2}{\lambda_{1n}\alpha_n}(1 + \kappa\lambda_{1n}^{-1/2}) \right) \max\{M_n, \|\xi_{\tau_n}\|_2\},$$

where $\alpha_n > 0$.

Proof of Lemma 2: Define the sample operators $S_{\mathbf{x}} : \mathcal{H}_K \to \mathbb{R}^n$ and $S_{\mathbf{x}}^T : \mathbb{R}^n \to \mathbb{R}$ as $S_{\mathbf{x}}(f) = (f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n))^T$ and $S_{\mathbf{x}}^T \mathbf{c} = \sum_{i=1}^n c_i K_{\mathbf{x}_i}$, we denote $n \times n$ matrix

$$\mathbf{W}_{\tau_n} = \begin{pmatrix} J_{\tau_n}\left(\widehat{F}_{\mathbf{x}_1}(y_1)\right) & & & \\ & J_{\tau_n}\left(\widehat{F}_{\mathbf{x}_2}(y_2)\right) & & \\ & & \ddots & \\ & & & J_{\tau_n}\left(\widehat{F}_{\mathbf{x}_n}(y_n)\right) \end{pmatrix}.$$

Then

$$\widehat{\xi}_{\tau_n} := \arg\min_{f \in \mathcal{H}_k} \frac{1}{n} \sum_{i=1}^n J_{\tau_n}\left(\widehat{F}_{\mathbf{x}_i}(y_i)\right)(y_i - f(\mathbf{x}_i))^2 + \lambda_{1n}\|f\|_K^2$$

is equal to

$$\widehat{\xi}_{\tau_n} = \arg\min_{f\in\mathcal{H}_k} \frac{1}{n}\mathbf{y}^T\mathbf{W}_{\tau_n}\mathbf{y} - \frac{2}{n} < f, S_{\mathbf{x}}^T\mathbf{W}_{\tau_n}\mathbf{y} >_{\mathcal{H}_K} + \frac{1}{n} < f, S_{\mathbf{x}}^T\mathbf{W}_{\tau_n}S_{\mathbf{x}}f >_K$$
$$+ \lambda_{1n} < f, f >_K,$$

where $\mathbf{y} = (y_1, y_2, \ldots, y_n)^T$. We devirate from above to get the concert expression of $\widehat{\xi}_{\tau_n}$

$$\widehat{\xi}_{\tau_n} = \left(\frac{1}{n}S_{\mathbf{x}}^T\mathbf{W}_{\tau_n}S_{\mathbf{x}} + \lambda_{1n}I\right)^{-1}\frac{1}{n}S_{\mathbf{x}}^T\mathbf{W}_{\tau_n}\mathbf{y}\,.$$

Define $L_K f = S_{\mathbf{x}}^T S_{\mathbf{x}} f$, then similarly the data-free and noise-free version $\widetilde{\xi}_{\tau_n}$ is

$$\widetilde{\xi}_{\tau_n} = (L_K + \lambda_{1n}I)^{-1}L_K\xi_{\tau_n}.$$

Then we have

$$\begin{aligned}\widehat{\xi}_{\tau_n} - \widetilde{\xi}_{\tau_n} &= \left(\frac{1}{n}S_{\mathbf{x}}^T\mathbf{W}_{\tau_n}S_{\mathbf{x}} + \lambda_{1n}I\right)^{-1}\left(\frac{1}{n}S_{\mathbf{x}}^T\mathbf{W}_{\tau_n}Y - \frac{1}{n}S_{\mathbf{x}}^T\mathbf{W}_{\tau_n}S_{\mathbf{x}}\widetilde{\xi}_{\tau_n} - \lambda_{1n}\widetilde{\xi}_{\tau_n}\right)\\ &= \left(\frac{1}{n}S_{\mathbf{x}}^T\mathbf{W}_{\tau_n}S_{\mathbf{x}} + \lambda_{1n}I\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n J_{\tau_n}(\widehat{F}_{\mathbf{x}_i}(y_i))(y_i - \widetilde{\xi}_{\tau_n}(\mathbf{x}_i))K_{\mathbf{x}_i} - L_K(\xi_{\tau_n} - \widetilde{\xi}_{\tau_n})\right).\end{aligned}$$

This gives a bound of its $\mathcal{H}_K$-norm as

$$\begin{aligned}\|\widehat{\xi}_{\tau_n} - \widetilde{\xi}_{\tau_n}\|_K &\le \lambda_{1n}^{-1}\left\|\frac{1}{n}\sum_{i=1}^n J_{\tau_n}(\widehat{F}_{\mathbf{x}_i}(y_i))(y_i - \widetilde{\xi}_{\tau_n}(\mathbf{x}_i))K_{\mathbf{x}_i} - L_K(\xi_{\tau_n} - \widetilde{\xi}_{\tau_n})\right\|_K\\ &\le \lambda_{1n}^{-1}(\Delta_1 + \Delta_2),\end{aligned} \tag{S1.1}$$

where

$$\Delta_1 = \left\|\frac{1}{n}\sum_{i=1}^n J_{\tau_n}(F_{\mathbf{x}_i}(y_i))(y_i - \widetilde{\xi}_{\tau_n}(\mathbf{x}_i))K_{\mathbf{x}_i} - L_K(\xi_{\tau_n} - \widetilde{\xi}_{\tau_n})\right\|_K,$$

$$\Delta_2 = \left\| \frac{1}{n} \sum_{i=1}^{n} \left(J_{\tau_n}(\widehat{F}_{\mathbf{x}_i}(y_i)) - J_{\tau_n}(F_{\mathbf{x}_i}(y_i)) \right) (y_i - \widetilde{\xi}_{\tau_n}(\mathbf{x}_i)) K_{\mathbf{x}_i} \right\|_K .$$

To bound Δ_1, we define random variable $\eta_i = J_{\tau_n}(F_{\mathbf{x}_i}(y_i))(y_i - \widetilde{\xi}_{\tau_n}(\mathbf{x}_i))K_{\mathbf{x}_i}$, then it's easy to verify that

$$\begin{aligned}
\mathbb{E}\eta &= \int_{\mathbb{R}^p} K_{\mathbf{x}} \int_{\mathbb{R}} J_{\tau_n}(F_{\mathbf{x}}(y))(y - \widetilde{\xi}_{\tau_n}(\mathbf{x})) d\rho(y|\,\mathbf{x}) d\rho(\mathbf{x}) \\
&= \int_{\mathbb{R}^p} K_{\mathbf{x}} \int_{\mathbb{R}} J_{\tau_n}(F_{\mathbf{x}}(y)) y d\rho(y|\,\mathbf{x}) d\rho(\mathbf{x}) \\
&- \int_{\mathbb{R}^p} K_{\mathbf{x}} \widetilde{\xi}_{\tau_n}(\mathbf{x}) \int_{\mathbb{R}} J_{\tau_n}(F_{\mathbf{x}}(y)) d\rho(y|\,\mathbf{x}) d\rho(\mathbf{x}) \\
&= L_K(\xi_{\tau_n} - \widetilde{\xi}_{\tau_n}),
\end{aligned}$$

where the last equality is from the definition that $\xi_\tau(\mathbf{x}) = \mathbb{E}\left[y J_\tau \left(F_{\mathbf{x}}(y) \right) | \, \mathbf{x} \right]$ and the fact that $\mathbb{E}\left[J_\tau \left(F_{\mathbf{x}}(y) \right) | \, \mathbf{x} \right] = 1$. And $\|\eta\|_K = |J_{\tau_n}(F_{\mathbf{x}}(y))(y - \widetilde{\xi}_{\tau_n}(\mathbf{x}))|\sqrt{K_{\mathbf{x}}}$. Recall that the definition of $J_{\tau_n}(\cdot)$ is

$$J_\tau(t) = \begin{cases} s_{\tau_n}(1-t)^{s_{\tau_n}-1}; & 0 < t < \frac{1}{2}; \\ r_{\tau_n} t^{r_{\tau_n}-1} & \frac{1}{2} < t < 1, \end{cases}$$

where $r_{\tau_n} = \log(1/2)/\log(\tau_n)$ and $s_{\tau_n} = \log(1/2)/\log(1-\tau_n)$. Thus when $\tau_n \to 1$, by Assumption 1 we have

$$\|\eta\|_K \leq \kappa r_{\tau_n}(M_n + \|\widetilde{\xi}_{\tau_n}\|_\infty) \quad \text{and} \quad \mathbb{E}\|\eta\|_K^2 \leq \kappa^2 r_{\tau_n}^2 \int (y - \widetilde{\xi}_{\tau_n}(\mathbf{x}))^2 d\rho_{\mathbf{x},y}.$$

It follows from Lemma 2 of Smale and Zhou (2007) and Assumption 1 in the main text, with probability at least $1-\delta_n$, there holds

$$\begin{aligned}\Delta_1 \leq & 2n^{-1}\kappa r_{\tau_n}\log\frac{2}{\delta_n}(M_n+\|\widetilde{\xi}_{\tau_n}\|_\infty)+n^{-1/2}\kappa r_{\tau_n}(2\log\frac{2}{\delta_n})^{1/2} \\ & \cdot\left(\int (y-\widetilde{\xi}_{\tau_n}(\mathbf{x}))^2 d\rho_{\mathbf{x},y}\right)^{1/2}. \end{aligned} \tag{S1.2}$$

For $\|\widetilde{\xi}_{\tau_n}\|_\infty$, by the definition of $\widetilde{\xi}_{\tau_n}$, we have

$$\left\|\widetilde{\xi}_{\tau_n}-\xi_{\tau_n}\right\|_2^2+\lambda_{1n}\|\widetilde{\xi}_{\tau_n}\|_K^2 \leq \|0-\xi_{\tau_n}\|_2^2+\lambda_{1n}\|0\|_K^2 \leq \|\xi_{\tau_n}\|_2^2, \tag{S1.3}$$

where $\|\xi_{\tau_n}\|_2^2$ is bounded. Hence, we have

$$\|\widetilde{\xi}_{\tau_n}\|_\infty \leq \kappa\|\widetilde{\xi}_{\tau_n}\|_K \leq \kappa\lambda_{1n}^{-1/2}\|\xi_{\tau_n}\|_2. \tag{S1.4}$$

For $\int (y-\widetilde{\xi}_{\tau_n}(\mathbf{x}))^2 d\rho_{\mathbf{x},y}$, by the definition of regression function, we have

$$\int (y-f(\mathbf{x}))^2 d\rho_{\mathbf{x},y}-\int\left(y-\xi_{\tau_n}(\mathbf{x})\right)^2 d\rho_{\mathbf{x},y}=\|f-\xi_{\tau_n}\|_2^2, \quad \text{for any } f.$$

Taking $f=0$ and $f=\widetilde{\xi}_{\tau_n}$ yield that

$$\int\left(y-\xi_{\tau_n}(\mathbf{x})\right)^2 d\rho_{\mathbf{x},y}+\|\xi_{\tau_n}\|_2^2=\int y^2 d\rho_{\mathbf{x},y} \leq \kappa^2\|f_0\|_K^2+\sigma^2 \leq M_n^2, \tag{S1.5}$$

$$\int\left(y-\widetilde{\xi}_{\tau_n}(\mathbf{x})\right)^2 d\rho_{\mathbf{x},y}=\|\widetilde{\xi}_{\tau_n}-\xi_{\tau_n}\|_2^2+\int\left(y-\xi_{\tau_n}(\mathbf{x})\right)^2 d\rho_{\mathbf{x},y} \leq 2M_n^2, \tag{S1.6}$$

where the last inequality is from (S1.3) and (S1.5).

Combine (S1.2)-(S1.6), we can conclude that with probability at least

$1-\delta_n$, there holds

$$\begin{aligned}\Delta_1 &\le 2n^{-1}\kappa r_{\tau_n}\log\frac{2}{\delta_n}(M_n+\kappa\lambda_{1n}^{-1/2}\left\|\xi_{\tau_n}\right\|_2)+2n^{-1/2}\kappa r_{\tau_n}(\log\frac{2}{\delta_n})^{1/2}M_n\\ &\le \frac{2\kappa r_{\tau_n}M_n}{n}\log\frac{2}{\delta_n}+\frac{2\kappa\left\|\xi_{\tau_n}\right\|_2}{n^{1/2}}\log\frac{2}{\delta_n}\frac{\kappa}{\lambda_{1n}^{1/2}n^{1/2}}+\frac{2\kappa r_{\tau_n}M_n}{n^{1/2}}\left(\log\frac{2}{\delta_n}\right)^{1/2}.\end{aligned} \tag{S1.7}$$

Note that when $\frac{\kappa}{\lambda_{1n}^{1/2}n^{1/2}}\le\left(3\log\frac{2}{\delta_n}\right)^{-1}$, the above bound can be simplified to

$$\lambda_{1n}^{-1}\Delta_1\le\frac{6\kappa r_{\tau_n}M_n^{\xi}}{\lambda_{1n}n^{1/2}}\log\frac{2}{\delta_n},$$

where $M_n^{\xi}=\max\{M_n,\|\xi_{\tau_n}\|_2\}$. When $\frac{\kappa}{\lambda_{1n}^{1/2}n^{1/2}}>\left(3\log\frac{2}{\delta_n}\right)^{-1}$, we have

$$\|\widehat{\xi}_{\tau_n}-\widetilde{\xi}_{\tau_n}\|_K\le\|\widehat{\xi}_{\tau_n}\|_K+\|\widetilde{\xi}_{\tau_n}\|_K\le\frac{2r_{\tau_n}M_n}{\lambda_{1n}^{1/2}}\le\frac{6\kappa r_{\tau_n}M_n}{\lambda_{1n}n^{1/2}}\log\frac{2}{\delta_n}, \tag{S1.8}$$

where the second inequality is from (S1.4) and (S1.5) and the definition of $\widehat{\xi}_{\tau_n}$ that

$$\frac{1}{n}\sum_{i=1}^{n}J_{\tau_n}(\widehat{F}_{\mathbf{x}_i}(y_i))\left(y_i-\widehat{\xi}_{\tau_n}(\mathbf{x}_i)\right)^2+\lambda_{1n}\|\widehat{\xi}_{\tau_n}\|_{\mathcal{H}_K}^2<\frac{1}{n}\sum_{i=1}^{n}r_{\tau_n}^2y_i^2\le r_{\tau_n}^2M_n^2.$$

To bound Δ_2, for any $y_i>0$, there exists a s_i such $q_{\tau_{s_i}}\le y_i\le q_{\tau_{s_{i+1}}}$, then when $\tau_n\to1$, we have

$$\begin{aligned}\max_i|J_{\tau_n}(\widehat{F}_{\mathbf{x}_i}(y_i))-J_{\tau_n}(F_{\mathbf{x}_i}(y_i))| &\le Cr_{\tau_n}^2\max_i|\widehat{F}_{\mathbf{x}_i}(y_i)-F_{\mathbf{x}_i}(y_i)|\\ &\le Cr_{\tau_n}^2\max_i|\tau_i^*-\inf\{\tau:q_\tau\ge y_i\}|\\ &\le Cr_{\tau_n}^2\alpha_n^{-1},\end{aligned}$$

with $\tau_i^* \in (\frac{s_i}{s_n+1}, \frac{s_{i+1}}{s_n+1})$ and $\alpha_n = O\left(\frac{s_n \lambda_{2n}^{1/2} n^{1/4}}{(\log n)^{1/2}\left(1+\kappa\lambda_{2n}^{-1/2}\right)^{1/2}}\right)$. The first inequality is from the Taylor expansion of $J_{\tau_n}(\cdot)$, the second inequality is from the definition of $\widehat{F}_{\mathbf{x}_i}(y_i)$ and $F_{\mathbf{x}_i}(y_i)$, and the last inequality is from Lemma 1 and Proposition 1. So we have

$$\Delta_2 \leq C r_{\tau_n}^2 \alpha_n (M_n + \|\widetilde{\xi}_{\tau_n}\|_\infty) \leq C r_{\tau_n}^2 \alpha_n^{-1} M_n (1 + \kappa \lambda_{1n}^{-1/2}), \tag{S1.9}$$

Combine (S1.1)-(S1.9), we get

$$\|\widehat{\xi}_{\tau_n} - \widetilde{\xi}_{\tau_n}\|_K \leq \log\frac{2}{\delta_n}\left(\frac{6\kappa r_{\tau_n} M_n^\xi}{\lambda_{1n} n^{1/2}} + \frac{C\kappa r_{\tau_n}^2 M_n^\xi}{\lambda_{1n}\alpha_n}(1+\kappa\lambda_{1n}^{-1/2})\right).$$

Thus we complete the proof. ■

Now we are ready to prove Theorem 1.

Proof of Theorem 1: We denote two events

$$\mathcal{C}_1 = \left\{\left\|\widehat{\xi}_{\tau_n} - \xi_{\tau_n}\right\|_K \leq \log\frac{4}{\delta_n}\left(\frac{6\kappa r_{\tau_n} M_n^\xi}{\lambda_{1n} n^{1/2}} + \frac{C\kappa r_{\tau_n}^2 M_n^\xi}{\lambda_{1n}\alpha_n}(1+\kappa\lambda_{1n}^{-1/2}) \right.\right.$$
$$\left.\left. + \lambda_{1n}^{2r-1}\left\|L_K^{-r}\xi_{\tau_n}\right\|_2^2\right)\right\},$$

$$\mathcal{C}_2 = \left\{\max_{i=1,\ldots,n} |y_i| > \kappa \|f_0\|_K + \sqrt{\frac{2\sum_{i=1}^n \sigma_i^2}{\delta_n}}\right\},$$

where $M_n^\xi = \max\left\{\kappa\|f_0\|_K + \sqrt{\frac{2\sum_{i=1}^n \sigma_i^2}{\delta_n}}, \|\xi_{\tau_n}\|_2\right\}$. Then we have

$$\mathbb{P}(\mathcal{C}_1) = \mathbb{P}(\mathcal{C}_1 \cap \mathcal{C}_2) + \mathbb{P}(\mathcal{C}_1 \cap \mathcal{C}_2^c) \leq \mathbb{P}(\mathcal{C}_2) + \mathbb{P}(\mathcal{C}_1 | \mathcal{C}_2^c) = P_1 + P_2, \tag{T1.1}$$

where $\mathcal{C}_2^c$ denotes the complement of $\mathcal{C}_2$.

For P_1, by Chebyshev's inequality, we have

$$\mathbb{P}\left(\max_{i=1,\ldots,n}|\varepsilon_i|\geq t\right)=\mathbb{P}\left(\cup_{i=1}^n|\varepsilon_i|\geq t\right)\leq\sum_{i=1}^n\mathbb{P}\left(|\varepsilon_i|\geq t\right)\leq\sum_{i=1}^n\frac{\mathbb{E}(\varepsilon_i^2)}{t^2}=\frac{\sum_{i=1}^n\sigma_i^2}{t^2}.$$

Then by Assumption 1, for any $\delta_n\in(0,1)$, with probability at least $1-\frac{\delta_n}{2}$, there holds

$$\max_{i=1,\ldots,n}|y_i|\leq\kappa\left\|f_0\right\|_K+\max_{i=1,\ldots,n}|\varepsilon_i|\leq\kappa\left\|f_0\right\|_K+\sqrt{\frac{2\sum_{i=1}^n\sigma_i^2}{\delta_n}},$$

implying that $\mathbb{P}\left(\mathcal{C}_2\right)\leq\frac{\delta_n}{2}$.

For P_2, we have

$$\|\widehat{\xi}_{\tau_n}-\xi_{\tau_n}\|_K\leq\|\widehat{\xi}_{\tau_n}-\widetilde{\xi}_{\tau_n}\|_K+\|\widetilde{\xi}_{\tau_n}-\xi_{\tau_n}\|_K.$$

To bound the second term, suppose that $\{\mu_i,\varphi_i\}_{i\geq1}$ are the normalized eigenpairs of the integral operator $L_K:L^2_{\rho_{\mathbf{x}}}\to L^2_{\rho_{\mathbf{x}}}$. Thus by Assumption 2, there exists some function $h_{\tau_n}\in L^2_{\rho_{\mathbf{x}}}$ such that $\xi_{\tau_n}=L_K^r h_{\tau_n}=\sum_{i\geq1}\mu_i^r\left\langle h_{\tau_n},\varphi_i\right\rangle_2\varphi_i\in\mathcal{H}_K$. Then we have

$$\begin{aligned}\widetilde{\xi}_{\tau_n}-\xi_{\tau_n}&=\left(L_K+\lambda_{1n}I\right)^{-1}L_K\xi_{\tau_n}-\xi_{\tau_n}=\left(L_K+\lambda_{1n}I\right)^{-1}\left(-\lambda_{1n}\xi_{\tau_n}\right)\\&=-\sum_{i\geq1}\frac{\lambda_{1n}}{\lambda_{1n}+\mu_i}\mu_i^r\left\langle h_{\tau_n},\varphi_i\right\rangle_2\varphi_i,\end{aligned}$$

which implies that

$$
\begin{aligned}
\|\widetilde{\xi}_{\tau_n} - \xi_{\tau_n}\|_K^2 &= \sum_{i\geq 1} \left(\frac{\lambda_{1n}}{\lambda_{1n}+\mu_i} \mu_i^{r-1/2} \langle h_{\tau_n}, \varphi_i \rangle_2 \right)^2 \left\| \mu_i^{1/2} \varphi_i \right\|_K^2 \\
&= \sum_{i\geq 1} \left(\frac{\lambda_{1n}}{\lambda_{1n}+\mu_i} \mu_i^{r-1/2} \langle h_{\tau_n}, \varphi_i \rangle_2 \right)^2 \\
&= \lambda_{1n}^{2r-1} \sum_{i\geq 1} \left(\frac{\lambda_{1n}}{\lambda_{1n}+\mu_i} \right)^{3-2r} \left(\frac{\mu_i}{\lambda_{1n}+\mu_i} \right)^{2r-1} \langle h_{\tau_n}, \varphi_i \rangle_2^2 \\
&\leq \lambda_{1n}^{2r-1} \sum_{i\geq 1} \langle h_{\tau_n}, \varphi_i \rangle_2^2 = \lambda_{1n}^{2r-1} \|h_{\tau_n}\|_2^2 = \lambda_{1n}^{2r-1} \left\| L_K^{-r} \xi_{\tau_n} \right\|_2^2 .
\end{aligned}
\tag{T1.2}
$$

where the second equality is from $\|\mu_i^{1/2}\varphi_i\|_K = \left(\sum_{j\geq 1} \frac{\langle \mu_i^{1/2}\varphi_i, \varphi_j \rangle_2^2}{\mu_j} \right)^{1/2} = \langle \varphi_i, \varphi_i \rangle_2 = 1$. From Proposition 1, we have

$$
\begin{aligned}
\mathbb{P}(\mathcal{C}_1 | \mathcal{C}_2^c) \leq & \mathbb{P} \left(\left\| \widehat{\xi}_{\tau_n} - \widetilde{\xi}_{\tau_n} \right\|_K \leq \log \frac{4}{\delta_n} \left(\frac{6\kappa r_{\tau_n} M_n^{\xi}}{\lambda_{1n} n^{1/2}} \right.\right. \\
& \left.\left. + \frac{C\kappa r_{\tau_n}^2 M_n^{\xi}}{\lambda_{1n}\alpha_n} (1+\kappa \lambda_{1n}^{-1/2}) \right) \middle| \mathcal{C}_2^c \right) \leq \delta_n/2.
\end{aligned}
$$

Thus we have $P_1 + P_2 \leq \delta_n/2 + \delta_n/2 = \delta_n$, then from (T1.1), we conclude that with probability at least $1-\delta_n$, there holds

$$
\|\widehat{\xi}_{\tau_n} - \xi_{\tau_n}\|_K \leq \log \frac{4}{\delta_n} \left(\frac{6\kappa r_{\tau_n} M_n^{\xi}}{\lambda_{1n} n^{1/2}} + \frac{C\kappa r_{\tau_n}^2 M_n^{\xi}}{\lambda_{1n}\alpha_n} (1+\kappa\lambda_{1n}^{-1/2}) + \lambda_{1n}^{r-1/2} \left\| L_K^{-r} \xi_{\tau_n} \right\|_2 \right).
$$

Thus we complete the proof. ■

S2.3 Proof of Theorem 2

Proof of Theorem 2: As we know,

$$
\begin{aligned}
\widehat{\gamma}(x) &= \frac{1}{k}\sum_{j=1}^{k}\log\frac{\widehat{\xi}_{\tau_{n-j}}(x)}{\widehat{\xi}_{\tau_{n-k}}(x)} \\
&= \frac{1}{k}\sum_{j=1}^{k}\log\frac{\xi_{\tau_{n-j}}\left(1+\frac{\widehat{\xi}_{\tau_{n-j}}-\xi_{\tau_{n-j}}}{\xi_{\tau_{n-j}}}\right)}{\xi_{\tau_{n-k}}\left(1+\frac{\widehat{\xi}_{\tau_{n-k}}-\xi_{\tau_{n-k}}}{\xi_{\tau_{n-k}}}\right)} \\
&= \frac{1}{k}\sum_{j=1}^{k}\log\frac{\xi_{\tau_{n-j}}}{\xi_{\tau_{n-k}}}+\frac{1}{k}\sum_{j=1}^{k}\frac{\widehat{\xi}_{\tau_{n-j}}-\xi_{\tau_{n-j}}}{\xi_{\tau_{n-j}}}\left\{1+o_p(1)\right\}. \\
&- \frac{1}{k}\sum_{j=1}^{k}\frac{\widehat{\xi}_{\tau_{n-k}}-\xi_{\tau_{n-k}}}{\xi_{\tau_{n-k}}}\left\{1+o_p(1)\right\} \\
&= \frac{1}{k}\sum_{j=1}^{k}\log\frac{q_{\tau_{n-j}}}{q_{\tau_{n-k}}}+\frac{1}{k}\sum_{j=1}^{k}\frac{\widehat{\xi}_{\tau_{n-j}}-\xi_{\tau_{n-j}}}{\xi_{\tau_{n-j}}}\left\{1+o_p(1)\right\} \\
&- \frac{1}{k}\sum_{j=1}^{k}\frac{\widehat{\xi}_{\tau_{n-k}}-\xi_{\tau_{n-k}}}{\xi_{\tau_{n-k}}}\left\{1+o_p(1)\right\} \\
&:= E_{1n}+E_{2n}-E_{3n},
\end{aligned}
$$

where the first term in the fourth equality is from Proposition 2.

Similar with the proof of Theorem 2.3 in Wang et al. (2012), we have

$$
E_{1n}=\gamma+\frac{A(n/k)}{1-\varrho}+o\left(A(n/k)\right),
$$

where ϱ and $A(\cdot)$ are defined in Section 4.1.

According to Proposition 1 in Daouia et al. (2022) and the assumption that $\mathbb{E}(Y|X)<\infty$, we have that $\xi_{\tau_{n-j}}$ is lower bounded by some constant

C, for $j = 1, \dots, k$. Therefore, we have

$$\|E_{2n}\|_K \leq \max_{j=1,\dots,k} \left\| \widehat{\xi}_{\tau_{n-j}} - \xi_{\tau_{n-j}} \right\|_K /C,$$

Denote the upper bound of $\|\widehat{\xi}_{\tau_{n-j}} - \xi_{\tau_{n-j}}\|_K$ proved in the last theorem as $B_{n,j}$.Thus we have $\|E_{2n}\|_K \leq \frac{\max_{j=1,\dots,k} B_{n,j}}{C}$. Similarly, there holds $\|E_{3n}\|_K \leq \frac{B_{n,k}}{C}$. Combine the results of E_{1n},E_{2n},E_{3n} together to get the upper bound of $\hat{\gamma}$

$$\|\hat{\gamma} - \gamma\|_K \leq \frac{A(n/k)}{1-\varrho} + 2\frac{B_{n,k}}{C} + o_p(A(n/k)). \tag{T2.1}$$

When $n/k \to \infty$,

$$\lim_{n/k\to\infty} \|\widehat{\gamma} - \gamma\|_K = 0.$$

Next we consider the upper bound of estimated error at quantile level of τ_n, note that

$$\begin{aligned}\frac{\widehat{\xi}_{\tau_n'}}{\xi_{\tau_n'}} &= \left(\frac{1-\tau_{n-k}}{1-\tau_n'}\right)^{\widehat{\gamma}} \frac{\widehat{\xi}_{\tau_{n-k}}}{\xi_{\tau_n'}} = \left(\frac{1-\tau_{n-k}}{1-\tau_n'}\right)^{\widehat{\gamma}} \frac{\widehat{\xi}_{\tau_{n-k}}}{\xi_{\tau_{n-k}}}\frac{\xi_{\tau_{n-k}}}{\xi_{\tau_n'}} = \left(\frac{1-\tau_{n-k}}{1-\tau_n'}\right)^{\widehat{\gamma}} \frac{\widehat{\xi}_{\tau_{n-k}}}{\xi_{\tau_{n-k}}}\frac{q_{\tau_{n-k}}}{q_{\tau_n'}} \\ &= \left(\frac{1-\tau_{n-k}}{1-\tau_n'}\right)^{\widehat{\gamma}} \frac{\widehat{\xi}_{\tau_{n-k}}}{\xi_{\tau_{n-k}}} \left\{ \frac{U\left(\frac{1}{1-\tau_n'}\right)}{U\left(\frac{1}{1-\tau_{n-k}}\right)} \right\}^{-1}, \end{aligned} \tag{T2.2}$$

by the second order condition of $U(\cdot)$, we have

$$\lim_{t\to\infty} \frac{1}{A(t)} \left\{ \frac{U(tz)}{U(t)} - z^\gamma \right\} = z^\gamma \frac{z^\varrho - 1}{\varrho}.$$

Let $z = \frac{1-\tau_{n-k}}{1-\tau_n'}$ and $t = \frac{1}{1-\tau_{n-k}}$, we get

$$\frac{U\left(\frac{1}{1-\tau_n'}\right)}{U\left(\frac{1}{1-\tau_{n-k}}\right)} = \left(1 + A\left(\frac{1}{1-\tau_{n-k}}\right) \frac{\left(\frac{1-\tau_{n-k}}{1-\tau_n'}\right)^\varrho - 1}{\varrho} + o_p\left(A(\frac{1}{1-\tau_{n-k}})\right)\right) \left(\frac{1-\tau_{n-k}}{1-\tau_n'}\right)^\gamma. \tag{T2.3}$$

Plug (T2.3) into (T2.2), it follows that

$$\frac{\widehat{\xi}'_{\tau_n}}{\xi'_{\tau_n}} = \left(\frac{1-\tau_{n-k}}{1-\tau'_n}\right)^{\widehat{\gamma}-\gamma} \frac{\widehat{\xi}_{\tau_{n-k}}}{\xi_{\tau_{n-k}}} \left(1 + A\left(\frac{1}{1-\tau_{n-k}}\right)\frac{\left(\frac{1-\tau_{n-k}}{1-\tau'_n}\right)^{\varrho}-1}{\varrho} + o_p\left(A(\frac{1}{1-\tau_{n-k}})\right)\right)^{-1}.$$

Recall that $p_n = 1-\tau'_n$ and $1-\tau_{n-k} = (k+1)/(n+1)$. Therefore,

$$\frac{\widehat{\xi}_{\tau'_n}}{\xi_{\tau'_n}} = \left\{\frac{k+1}{(n+1)p_n}\right\}^{\widehat{\gamma}-\gamma} \frac{\widehat{\xi}_{\tau_{n-k}}}{\xi_{\tau_{n-k}}} \left(1 - A\left(\frac{n}{k}\right)\frac{\left(\frac{k}{np_n}\right)^{\varrho}-1}{\varrho} + o_p\left(A(\frac{n}{k})\right)\right). \tag{T2.4}$$

For the first term in (T2.4), from (T2.1) we have

$$\begin{aligned}\left\{\frac{k+1}{(n+1)p_n}\right\}^{\widehat{\gamma}-\gamma} &= \exp\left[(\widehat{\gamma}-\gamma)\log\{(k+1)/((n+1)p_n)\}\right] \\ &= 1 + (\widehat{\gamma}-\gamma)\log\{(k+1)/((n+1)p_n)\}\{1+o_p(1)\} \\ &= 1 + \frac{\log(k/np_n)}{\sqrt{k}}\left\{\frac{\sqrt{k}A(n/k)}{1-\varrho} + \frac{2\sqrt{k}B_{n,k}}{C} + o_p(\sqrt{k}A(n/k))\right\} \\ &= 1 + \frac{\log(k/np_n)}{\sqrt{k}}\left\{\frac{\phi}{1-\varrho} + \frac{2\sqrt{k}B_{n,k}}{C} + o_p(1)\right\}.\end{aligned}$$

For the second term in (T2.4), from Theorem 1, we have

$$\left\|\frac{\widehat{\xi}_{\tau'_n}}{\xi_{\tau'_n}}\right\|_K \leq 1 + \frac{B_{n,k}}{\|\xi_{\tau_{n-k}}\|_K}.$$

Comparing the orders of the three terms in (T2.4) and using Taylor expansion, we get

$$\frac{\sqrt{k}}{\log(k/np_n)}\left\{\left\|\frac{\widehat{\xi}_{\tau'_n}}{\xi_{\tau'_n}}\right\|_K - 1\right\} = \frac{\phi}{1-\varrho} + \frac{\sqrt{k}B_{n,k}}{C}.$$

Thus we complete the proof. ■

Bibliography

Chen, F., He, X., and Wang, J. (2021). Learning sparse conditional distribution: An efficient kernel-based approach. *Electronic Journal of Statistics*, 15:1610–1635.

Daouia, A., Gijbels, I., and Stupfler, G. (2019). Extremiles: A new perspective on asymmetric least squares. *Journal of the American Statistical Association*, 114(527):1366–1381.

Daouia, A., Gijbels, I., and Stupfler, G. (2022). Extremile regression. *Journal of the American Statistical Association*, 117(539):1579–1586.

Smale, S. and Zhou, D. (2007). Learning theory estimates via integral operators and their approximations. *Constructive Approximation*, 26(2):153–172.

Villa, S., Rosasco, L., Mosci, S., and Verri, A. (2012). Consistency of learning algorithms using attouch-wets convergence. *Optimization*, 61:287–305.

Wang, H. J., Li, D., and He, X. (2012). Estimation of high conditional quantiles for heavy-tailed distributions. *Journal of the American Statistical Association*, 107(498):1453–1464.