
NONPARAMETRIC SMOOTHING FOR EXTREMAL QUANTILE REGRESSION WITH HEAVY TAILED DATA

Author: TAKUMA YOSHIDA 
– Graduate School of Science and Engineering, Kagoshima University,
Japan
yoshida@sci.kagoshima-u.ac.jp

Received: September 2017

Revised: April 2019

Accepted: April 2019

Abstract:

- In several different fields, it is interested in analyzing the upper or lower tail quantile of the underlying distribution rather than mean or center quantile. However, the investigation of the tail quantile is somewhat difficult because of data sparsity. This paper challenges to develop the nonparametric quantile regression for extremal quantile level. In the extremal quantile regression, there are two situation of technical conditions of order of convergence of the quantile level that intermediate order or extreme order. For the intermediate order quantile, the ordinary nonparametric estimator is used. On the other hand, for the extreme order quantile, we provide the new estimator by the extrapolating the intermediate order quantile estimator. The performance of the estimator is guaranteed by the asymptotic theory and the extreme value theory. As the result, we show the asymptotic normality and the rate of convergence of the nonparametric quantile regression estimator for both intermediate and extreme order quantile. Simulation is addressed to confirm the behavior of the proposed estimator. The data application is also assessed.

Keywords:

- *asymptotic normality; extrapolation; extremal quantile regression; extreme value theory; nonparametric estimator.*

AMS Subject Classification:

- 62G08, 62G20, 62G32.

1. INTRODUCTION

In a wide variety of areas, such as in the study of heavy rainfall, low birth weight, and high-risk finance, the tail behavior of the distribution of the target variable is of interest rather than the average or median. In these cases, we often investigate the upper or lower quantile of the data. However, the estimation of the tail quantile is difficult because of data sparsity. Therefore, the development of the mathematical properties of the tail quantile would be welcome. The theoretical performance of the tail behavior of the distribution function is provided by extreme value theory (EVT). The fundamental properties of EVT were surveyed by Coles (2001) [10], Beirlant *et al.* (2004a) [3], and de Haan and Ferreira (2006) [13]. On the other hand, the performance of the estimator is often guaranteed by a large sample or asymptotic theory in statistics. Thus, the mathematical properties of the estimator of the tail quantile are analyzed using a hybrid of EVT and asymptotic theory. In many cases, it is important to research the target variable along with the information of other variable. Then we should analyze the data in the literature of regression. In this paper, we consider the estimation of the extremal conditional quantile of the response Y given $X = x$.

Many researchers have developed the extremal conditional quantile estimation. Beirlant and Goegebeur (2004) [2] developed a Pareto distribution approach. Gardes *et al.* (2010) [20] and Gardes and Girard (2010) [19] studied the nearest-neighbor estimation. Daouia *et al.* (2011, 2013) [12, 11], El Methni *et al.* (2014) [18], and Girard and Louhichi (2015) [21] investigated the extremal quantile of the nonparametric estimator of the conditional distribution function of Y given $X = x$. The local-moment-type methods were studied by Goegebeur *et al.* (2017) [22]. Durrieu *et al.* (2015) [15] have developed the weighted quasi-log-likelihood method. On the other hand, quantile regression, which was pioneered by Koenker and Bassett (1978) [31], is a typical approach used to investigate the conditional quantile. For the center quantile, several authors have developed quantile regression methods. These fundamental developments have been summarized by Koenker (2005) [29]. However, much less work has been done on quantile regression for the extremal quantile. Chernozhukov (2005) [6], Chernozhukov and Fernández-val (2011) [7], Wang *et al.* (2012) [44], and He *et al.* (2016) [24] studied extremal quantile regression, but they focused only on linear models. For the tail quantile, the linear structure assumption is strong and its assumption is violated in several cases. Therefore, a nonparametric approach should be used in extremal quantile regression in such situations. Beirlant *et al.* (2004b) [4] studied extremal nonparametric quantile regression, but the theoretical property was not investigated. In this paper, we develop nonparametric quantile regression for the extremal quantile and mathematical properties.

Before we describe our study, we review extremal quantile regression with linear models in more detail. For extremal quantile regression, the quantile level τ approaches 0 or 1 as the sample size n increases. This paper treats only the upper quantile and, hence, $\tau \rightarrow 1$ as $n \rightarrow \infty$. Thus, there are two important types of the order of τ : the intermediate order quantile and the extreme order quantile. The former means that $\tau \rightarrow 1$ and $n(1 - \tau) \rightarrow \infty$ as $n \rightarrow \infty$, whereas in the latter $\tau \rightarrow 1$ and $n(1 - \tau) \rightarrow c \in [0, \infty)$ as $n \rightarrow \infty$. If τ is fixed, it is the so-called center quantile. According to Chernozhukov (2005) [6] and Chernozhukov and Fernández-val (2011) [7], the quantile regression estimator with linear models has asymptotic normality for the intermediate order quantile but it converges to a non-degenerated distribution (not normal) for the extreme order quantile. Thus, the extreme order quantile is difficult to handle.

Wang *et al.* (2012) [44] provided a nice approach to obtain the extreme order quantile estimator by extrapolation from the intermediate order quantile estimator. As a result, this extrapolated estimator has asymptotic normality. It seems that above results should be extended to the nonparametric quantile regression for many applications.

In this paper, we first construct the ordinary nonparametric estimator for the intermediate order quantile. We then use the B -spline method with ℓ_2 penalty. This approach was originally considered by O'Sullivan (1986) [37] and Eilers and Marx (1996) [16] in mean regression. Pratesi *et al.* (2009) [39], Reiss and Huang (2012) [40], and Yoshida (2013) [46] used the quantile regression for only the center quantile. We show the asymptotic bias and variance as well as the asymptotic normality of the penalized spline estimator. Next, the extrapolated estimator is obtained for the extreme order quantile. Similar to the approach of Wang *et al.* (2012) [44], we use the Weissman-type extrapolation method (see Weissman 1978 [45]). The asymptotic normality and the optimal rate of convergence of the extreme order quantile estimator are shown. To the best of our knowledge, this is the first time that the rate of convergence of the nonparametric estimator is revealed in the extremal quantile regression.

This paper is organized as follows. In Section 2, we coordinate the conditions of the true conditional quantile by EVT in nonparametric extremal quantile regression. Section 3 presents the nonparametric estimator and its asymptotic property for both intermediate and extreme order quantiles. In Section 4 the Monte Carlo simulation is conducted to confirm the performance of the proposed estimator. Section 5 addresses the application to Beijing's PM_{2.5} pollution data. The conclusions and future research are described in Section 6. In Appendices A and B, the computational aspects of the penalized spline estimator and the proofs of the mathematical results that appear in this paper are presented. Throughout the paper and without loss of generality, we focus on the conditional high quantile because a low quantile of the response can be viewed as a high quantile of the inverted sign of the response.

2. CONDITIONAL EXTREMAL QUANTILES

2.1. Extreme value theory

Let $\{(X_i, Y_i); i = 1, \dots, n\}$ be the independent copies of a random pair $(X, Y) \in \mathbb{R} \times \mathbb{R}$. We assume that the support of X is bounded on $[a, b]$, where $-\infty < a < b < \infty$. The conditional distribution function of Y given $X = x$ is denoted by $F_Y(y|x) = P(Y \leq y|X = x)$. Then the conditional 100 τ % quantile of Y given $X = x$ is

$$q_Y(\tau|x) \equiv F_Y^{-1}(\tau|x) = \inf\{t, F_Y(t|x) \geq \tau\}.$$

The main purpose of this study is to estimate $q_Y(\tau|x)$ for a high quantile level $\tau \simeq 1$. The tail behavior of the distribution or quantile function can be characterized by EVT. To analyze the conditional high quantile of Y given $X = x$, we introduce the EVT conditions of $F_Y(\cdot|x)$ and $q_Y(\cdot|x)$. Define $F(y)$ and $q(\tau)$ as the marginal distribution function and 100 τ % quantile of Y . Throughout the paper, we assume that F and $F(\cdot|x)$ belong to the maximum domain of attraction of an extreme value distribution G_γ , denoted by $F, F_Y(\cdot|x) \in D(G_\gamma)$.

The distribution function Q belongs to the maximum domain of attraction, which means that for the random sample Z_1, \dots, Z_n from Q , there exists a constant $\alpha_n > 0$ and $\beta_n \in \mathbb{R}$ such that for $1 + \gamma z \geq 0$,

$$P\left(\frac{\max_{1 \leq i \leq n} Z_i - \beta_n}{\alpha_n} \leq z\right) \rightarrow G_\gamma(z) = \exp[-(1 + \gamma z)^{-1/\gamma}]$$

as $n \rightarrow \infty$. Here, $\gamma \in \mathbb{R}$ is the extreme value index (EVI). The EVI is very important since this generally controls the tail behavior of the distribution function. For $Q \in D(G_\gamma)$, if $\gamma = 0$ or $\gamma < 0$, Q has a light tail or short tail. When $\gamma > 0$, Q has a heavy tail. This paper only discusses the heavy-tail case and, hence, we assume that $\gamma > 0$ from now on. The maximum domain of attraction is a very weak condition. For example, uniform, beta, Gaussian, t , Pareto, Cauchy, and many other distributions belong to the maximum domain of attraction with appropriately specified $\gamma \in \mathbb{R}$. The details of the maximum domain of attraction and EVI are given in fundamental books such as that by de Haan and Ferreira (2006) [13].

We now state the conditions to connect the tail behavior of F and $F(\cdot|x)$. For this, we need an additional definition. Let $RV(a) = \{A \in \mathbb{R}^+ \rightarrow \mathbb{R}^+ | A(mt)/A(t) \rightarrow m^a \text{ as } t \rightarrow \infty, m > 0\}$ be the set of regularly varying functions, where $\mathbb{R}^+ = (0, \infty)$. When $A \in RV(0)$, A is the so-called slowly varying function.

Conditions A

- A1.** There exists $L \in RV(0)$ such that the distribution function F satisfies $1 - F(y) = y^{-1/\gamma}L(y)\{1 + o(1)\}$ as $y \rightarrow \infty$.
- A2.** We have $q_0(\tau) = \partial q(t)/\partial t|_{t=1-\tau}$ regularly varying at 0 with index $-\gamma - 1$. That is, for $x > 0$,

$$\lim_{\tau \rightarrow 0} \frac{q_0(x\tau)}{q_0(\tau)} = x^{-\gamma-1}.$$

- A3.** There exists an auxiliary function $f(x)$ such that $V \equiv Y - f(x)$ has the distribution function $F_V(y|x)$ satisfying, as $y \rightarrow \infty$,

$$1 - F_V(y|x) = H(x)\{1 - F(y)\}(1 + o(1)),$$

where $H(x) > 0$ is a positive, continuous, and bounded function on $[a, b]$ and has $E[H(X)] = 1$.

- A4.** For $q_V(\cdot|x) = F_V^{-1}(\cdot|x)$, $\partial q_V(\tau|x)/\partial \tau = \partial q(1 - (1 - \tau)/H(x))/\partial \tau\{1 + o(1)\}$ uniformly in $x \in [a, b]$ as $\tau \rightarrow 1$.

Conditions A may not hold if either F or $F(\cdot|x)$ is not included in $D(G_\gamma)$. In other words, if $F, F(\cdot|x) \in D(G_\gamma)$, Conditions A are natural. Condition A1 is the formal notation of a Pareto-type tail (see Chernozhukov and Fernández-val 2011 [7]). The equivalent to condition A1 is

$$(2.1) \quad q(\tau) = (1 - \tau)^{-\gamma} \bar{L}(1/(1 - \tau))\{1 + o(1)\} \quad \text{as } \tau \rightarrow 1$$

with $\bar{L} \in RV(0)$. Therefore, if the distribution F is continuous and $\partial \bar{L}(1/(1 - \tau))/\partial \tau \rightarrow 0$ as $\tau \rightarrow 1$, condition A2 holds from condition A1. Actually, $q_0(\tau) = \partial q(t)/\partial t|_{t=1-\tau} = \gamma \tau^{-\gamma-1} \bar{L}(1/\tau) + \tau^{-\gamma} \partial \bar{L}(1/(1 - t))/\partial t|_{t=1-\tau} = \gamma \tau^{-\gamma-1} \bar{L}(1/\tau)\{1 + o(1)\}$ as $\tau \rightarrow 0$. Therefore,

we have $\partial q(t)/\partial t|_{t=1-\tau} = \gamma\tau^{-\gamma-1}\bar{L}(1/\tau)\{1 + o(1)\}$. Thus, condition A2 is weak. From A3, it is easy to show that $q_V(\tau|x) = H(x)^\gamma q(\tau)(1 + o(1))$ as $\tau \rightarrow 1$. Furthermore, since $\tau = F_Y(q_Y(\tau|x)|x) = P(Y < q_Y(\tau|x)|x) = P(Y - f(x) < q_Y(\tau|x) - f(x)|x) = F_V(q_Y(\tau|x) - f(x)|x)$, we obtain $q_V(\tau|x) = q_Y(\tau|x) - f(x)$. Consequently, we have

$$(2.2) \quad q_Y(\tau|x) = f(x) + h(x)q(\tau)\{1 + o(1)\} \quad \text{as } \tau \rightarrow 1,$$

where $h(x) = H(x)^\gamma$. Chernozhukov (2005) [6], Chernozhukov and Fernández-val (2011) [7] and Wang *et al.* (2012) [44] also provided this type of condition in multiple linear models. That is, they further assumed that $f(x) = x^T\beta$ and $h(x) = x^Tc$ for $x = (x_1, \dots, x_p)$, where β and c are unknown p -dimensional parameter vectors. Thus, A3 is the nonparametric model version of the above previous studies. Einmahl *et al.* (2016) [17] also considered the similar model to the survival function for more simple situation where time is the covariate. Condition A4 guarantees the existence of a conditional quantile density function (the derivative of the quantile function). Furthermore, the conditional quantile density function also behaves like a Pareto-type function by condition A4. Assumption A3 is strengthened by condition A4.

Remark 1. Let $U(t) = q(1 - 1/t) = \inf\{z|F(z) \geq 1 - 1/t\}$ and let $U_V(t|x) = q_V(1 - 1/t|x)$. In several articles (see, for example, de Haan and Ferreira 2006 [13]), the conditions of EVT are applied to $U(t)$ and $U_V(t|x)$ as $t \rightarrow \infty$. Since $q(\tau) = U(1/(1 - \tau))$, the condition (2.1) is similar to $U(t) = t^\gamma L(t)\{1 + o(1)\}$ with $t = 1/(1 - \tau)$. Condition A4 can also be expressed as $\partial U_V(t|x)/\partial t = \partial U(tH(x))/\partial t$ with $t = 1/(1 - \tau)$. Thus, we can reconsider the EVT conditions for quantiles by using U and U_V . In particular, the use of U is appropriate when using the second-order condition of EVT (see Section 3.2).

2.2. B-spline model

The conditional quantile $q_Y(\tau|x)$ can be written as

$$(2.3) \quad q_Y(\tau|x) = \operatorname{argmin}_a E[\rho_\tau(Y - a)|X = x],$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$ is Koenker’s check function (Koenker 2005 [29]) and I is the indicator function. The estimator of q_Y is often obtained along with an empirical version of (2.3). To estimate $q_Y(\tau|x)$, we use the B -spline regression method as the nonparametric technique in this paper. Let $\{B_k(x) : k = 1, \dots, K + p\}$ be the p th degree of the B -spline basis with knots $a = \kappa_0 < \kappa_1 < \dots < \kappa_K = b$. In addition, other sets of $2p$ knots are defined as $\kappa_{-p} = \dots = \kappa_{-1} = a$ and $\kappa_{K+1} = \dots = \kappa_{K+p} = b$. We then define the B -spline model as

$$s(x) = \sum_{k=1}^{K+p} B_k(x)b_k = \mathbf{B}(x)^T \mathbf{b},$$

where $\mathbf{B}(x) = (B_1(x), \dots, B_{K+p}(x))^T$ and $\mathbf{b} = (b_1, \dots, b_{K+p})^T$ is an unknown parameter vector. We now describe the relationship between the B -spline model and EVT discussed in the previous section. Let $W^m[a, b] = \{g|g^{(k)}$ is continuous, $k = 0, 1, \dots, m - 1$ and $\int_a^b \{g^{(m)}(x)\}^2 dx < \infty\}$ be the m th-order Sobolev space. From Barrow and Smith (1978) [1], for any function $g \in W^m([a, b])$, there exists $\mathbf{b}_g \in \mathbb{R}^{K+p}$ such that $g(x) - \mathbf{B}(x)^T \mathbf{b}_g = K^{-d}g^{(d)}(x)O(1)$ as $K \rightarrow \infty$, where $d = \min\{m, p + 1\}$. For simplicity, we assume that $m \leq p + 1$, that is $d = m$. Actually, $(p, m) = (3, 2)$ is the standard condition of B -spline smoothing.

For $\tau \in (0, 1)$, let

$$\mathbf{b}_0(\tau) = \operatorname{argmin}_{\mathbf{b} \in \mathbb{R}^{K+p}} E[\rho_\tau(Y - s(x)) | X = x]$$

and let $s_0(\tau|x) = \mathbf{B}(x)^T \mathbf{b}_0(\tau)$. We then found that $q_Y(\tau|x) = s_0(\tau|x) + K^{-m} \{ \partial^m q_Y(\tau|x) / \partial x^m \} O(1)$ for $q_Y(\cdot|x) \in W^m[a, b]$. If $h \in W^m[a, b]$, Conditions A and (2.2) yield that $\{ \partial^m q_Y(\tau|x) / \partial x^m \} = h^{(m)}(x)q(\tau)(1 + o(1)) = O((1 - \tau)^{-\gamma})$ and, hence, $s_0(\tau|x) - q_Y(\tau|x) = O(K^{-m}(1 - \tau)^{-\gamma})$ as $\tau \rightarrow 1$ and $K \rightarrow \infty$, which indicates that the condition B4 below is required.

If f and h defined in (2.2) belong to $W^m[a, b]$, there exists $\mathbf{b}_f, \mathbf{b}_h \in \mathbb{R}^{K+p}$ such that $f(x) - \mathbf{B}(x)^T \mathbf{b}_f = O(K^{-m})$ and $h(x) - \mathbf{B}(x)^T \mathbf{b}_h = O(K^{-m})$. We then obtain $\mathbf{b}_0(\tau) = \mathbf{b}_f + q(\tau)\mathbf{b}_h + O(K^{-m}q(\tau))$ as $\tau \rightarrow 1$ and $K \rightarrow \infty$. Therefore, (2.1) and condition A4 indicate that $\partial s_0(\tau|x) / \partial \tau \sim \mathbf{B}(x)^T \mathbf{b}_h \partial q(\tau) / \partial \tau$ is satisfied since \mathbf{b}_f and \mathbf{b}_h are not dependent on τ . Thus, the B -spline model also holds (2.2) and condition A4 and, hence, the tail behavior of the B -spline model can be studied by using Conditions A. The following conditions are the fundamental assumptions for B -spline regression.

Conditions B

- B1.** For some constant $\nu > 0$, $E[|Y|^{2+\nu} | X = x] < \infty$.
- B2.** The functions f and h in (2.2) are included in $W^m[a, b]$.
- B3.** We have $\max_{1 \leq j \leq K} \{ \kappa_{j+1} - \kappa_j \} = O(K^{-1})$.
- B4.** For some $\alpha \in (0, 1)$, the number of knots $K = O(n^\alpha)$.
- B5.** As $\tau \rightarrow 1$ and $K \rightarrow \infty$, $K^m(1 - \tau)^\gamma \rightarrow \infty$.

Condition B1 is needed to that the estimator satisfies the Lyapunov condition of central limit theorem. When condition B2 holds, the B -spline model can approximate to $q_Y(\tau|x)$. Conditions B3 and B4 are standard conditions for B -spline models. Together with condition B2, the B -spline model and EVT are connected for high quantile level. Condition B5 guarantees that the model bias between the conditional quantile and B -spline model converges to 0.

3. PENALIZED B -SPLINE ESTIMATOR FOR EXTREMAL QUANTILES

In this section, we define the nonparametric B -spline estimator and develop the asymptotic result. Then, we consider two scenario of extremal quantile rate: (i) intermediate order quantiles that $\tau \rightarrow 1$ and $(1 - \tau)n \rightarrow \infty$ as $n \rightarrow \infty$ and (ii) extreme order quantiles that $\tau \rightarrow 1$ and $(1 - \tau)n \rightarrow c \in [0, \infty)$ as $n \rightarrow \infty$. We denote the intermediate order quantile level by τ_I and the extreme order quantile level by τ_E , respectively. That is, as $n \rightarrow \infty$, $\tau_I, \tau_E \rightarrow 1$, $n(1 - \tau_I) \rightarrow \infty$, $n(1 - \tau_E) \rightarrow c \in [0, \infty)$, and $(1 - \tau_I)/(1 - \tau_E) \rightarrow \infty$.

3.1. Estimation of intermediate order quantiles

The ordinary B -spline quantile estimator for $\tau \in (0, 1)$ is defined based on minimizing $\sum_{i=1}^n \rho_\tau(Y_i - s(x_i))$. However, it is known that the ordinary estimator tends to have a wiggly curve caused by data sparsity. To avoid this, we introduce the penalization method to control the behavior of the estimator. Although various types of penalties have been developed to prevent overfitting, we will use O’Sullivan’s (1986) [37] penalty. For $\tau \in (0, 1)$, the penalized spline estimator $\tilde{\mathbf{b}}(\tau) = (\tilde{b}_1(\tau), \dots, \tilde{b}_{K+p}(\tau))^T$ of vector $\mathbf{b}(\tau) = (b_1(\tau), \dots, b_{K+p}(\tau))^T$ is constructed by minimizing

$$(3.1) \quad \sum_{i=1}^n \rho_\tau(Y_i - s(x_i)) + \lambda \int_a^b \{s^{(m)}(x)\}^2 dx,$$

where $\lambda > 0$ is the smoothing parameter. Using $\tilde{\mathbf{b}}(\tau)$, for the intermediate order quantile level τ_I , we define

$$\tilde{q}_Y(\tau_I|x) = \mathbf{B}(x)^T \tilde{\mathbf{b}}(\tau_I).$$

We study the asymptotic theory for the penalized spline estimator $\tilde{q}_Y(\tau_I|x)$. Then, the conditions of the number of knots and the smoothing parameter included in $\tilde{q}_Y(\tau_I|x)$ are very important. The penalty $\int_a^b \{s^{(m)}(x)\}^2 dx$ can be written as $\mathbf{b}^T D_m^T R D_m \mathbf{b}$, where the $(K + p)$ -matrix R has elements $R_{ij} = \int_a^b B_i(x) B_j(x) dx$ and the $(K + p - m) \times (K + p)$ matrix D_m satisfies $\mathbf{b}^{(m)} = D_m \mathbf{b}$, where $\mathbf{b}^{(m)} = (b_1^{(m)}, \dots, b_{K+p-m}^{(m)})^T$, and for $m = 1, 2, \dots$,

$$b_j^{(1)} = p \frac{b_j - b_{j-1}}{\kappa_{j+p} - \kappa_j}, \quad b_j^{(m)} = (p + 1 - m) \frac{b_j^{(m-1)} - b_{j-1}^{(m-1)}}{\kappa_{j+p+1-m} - \kappa_j}.$$

From now on, we use the symbols D_m and R . Let $G(h)$ be the $(K + p)$ -matrix with elements $G_{ij} = \int_a^b h(x) B_i(x) B_j(x) dx$ and

$$\Lambda(h) = \Lambda(h, n, \tau_I) = \gamma^{-1} G(h^{-\gamma}) + \frac{\lambda q(\tau_I)}{(1 - \tau_I)n} D_m^T R D_m.$$

Let $G = G(1)$ be $G(h)$ with $h(x) \equiv 1$. Define

$$K(m, \tau_I) = K \left(\frac{\lambda q(\tau_I)}{n(1 - \tau_I)} \right)^{1/2m},$$

which controls the asymptotic scenario branch discussed in Remark 1 below.

Conditions C

- C1.** We have $K(m, \tau_I) \geq 1$.
- C2.** We have $K\{\lambda q(\tau_I)/n(1 - \tau_I)\}^{1/2m} \rightarrow \infty$ as $n \rightarrow \infty$.
- C3.** We have $\lambda = o(q(\tau_I)^{-1}\{n(1 - \tau_I)/q(\tau_I)\}^2)$ as $n \rightarrow \infty$.

Condition C concerns with the asymptotic property of the penalized spline estimator. C1 is detailed in Remark 2. C2 allows us to use the large K . If C3 fails, the asymptotic bias of the penalized spline estimator cannot be vanished. We now show the asymptotic distribution

of $\tilde{q}(\tau_I|x)$. First, we derive the two types of bias, model bias and shrinkage bias. Roughly speaking, the model bias is the bias between the B -spline model and the true function, and the shrinkage bias is the difference between the expectation of the penalized estimator and the unpenalized estimator. According to Section 2.2, the model bias is $b_a(\tau_I|x) = s_0(\tau_I|x) - q_Y(\tau_I|x) = O(K^{-m}q(\tau))$. This model bias becomes the negligible order from condition C2. That is, the bias is dominated by the shrinkage bias. Define

$$b_\lambda(\tau|x) = \frac{\lambda q(\tau)}{(1-\tau)n} \mathbf{B}(x)^T \Lambda(H^{-\gamma})^{-1} D_m^T R D_m \mathbf{b}_0(\tau),$$

$$v^2(\tau|x) = \frac{q(\tau)^2}{(1-\tau)n} \mathbf{B}(x)^T \Lambda(H^{-\gamma})^{-1} G \Lambda(H^{-\gamma})^{-1} \mathbf{B}(x).$$

As a result, $b_\lambda(\tau_I|x)$ is the asymptotic shrinkage bias and $v(\tau_I|x)$ is the asymptotic variance of $\tilde{q}_Y(\tau_I|x)$. The following theorem shows the asymptotic order of the asymptotic bias and variance of the intermediate order quantile estimator.

Theorem 3.1. *Under Conditions A–C, as $n \rightarrow \infty$,*

$$b_\lambda(\tau_I|x) = O\left(q(\tau) \left(\frac{\lambda q(\tau_I)}{(1-\tau_I)n}\right)^{1/2}\right), \quad v^2(\tau_I|x) = O\left(\frac{q(\tau_I)^2}{(1-\tau_I)n} \left(\frac{\lambda q(\tau_I)}{(1-\tau_I)n}\right)^{-1/2m}\right).$$

From condition C3 and Theorem 3.1, we see that the shrinkage bias and variance converge to 0 as $n \rightarrow \infty$. Using the central limit theorem, Lyapunov's condition, and a Cramér–Wold device, the asymptotic normality of the estimator $\tilde{q}_Y(\tau_I|x)$ can be shown.

Theorem 3.2. *Suppose that Conditions A–C hold. As $n \rightarrow \infty$, $b_\lambda(\tau|x)$ and $v^2(\tau|x)$ are the asymptotic bias and variance of $\tilde{q}_Y(\tau|x)$ and*

$$\left(\frac{v(\tau_I|x)}{q_Y(\tau_Y|x)}\right)^{-1} \left\{ \frac{\tilde{q}_Y(\tau_I|x)}{q_Y(\tau_Y|x)} - 1 - \frac{b_\lambda(\tau_I|x)}{q_Y(\tau_Y|x)} \right\} \xrightarrow{D} N(0, 1).$$

Furthermore, under $\lambda = O(q(\tau_I)^2\{(1-\tau_I)n\}^{1/(2m+1)})$, the optimal rate of convergence of the mean integrated squared error (MISE) of $\tilde{q}_Y(\tau_I|x)/q_Y(\tau_I|x)$ is

$$E \left[\left\{ \frac{\tilde{q}_Y(\tau_I|x)}{q_Y(\tau_I|x)} - 1 \right\}^2 \right] = O(\{(1-\tau_I)n\}^{-2m/(2m+1)}).$$

Theorems 3.1 and 3.2 yield that the trade-off between bias and variance is controlled by λ . Thus, this indicates that the careful choice of K is not important in the penalized spline methods. According to Yoshida (2013) [46], for the center quantile level τ , the MISE of the penalized spline quantile estimator has the order $O(n^{-2m/(2m+1)})$. Thus, the rate of convergence of the MISE of the penalized spline estimator for the intermediate quantile level is slower than that for the center quantile level. This result is not surprising in the context of the difficulties of the estimation for the tail quantile.

When $\mathbf{B}(x) = \mathbf{x}$ and $\lambda = 0$, the estimator is reduced to the ordinary quantile regression with the linear model. In the linear regression, the model bias is 0 and, hence, the bias term vanishes. On the other hand, since $G = E[\mathbf{X}\mathbf{X}^T]$ and $\Lambda(H^{-\gamma})^{-1} = \gamma^{-1}E[H(X)^{-\gamma}\mathbf{X}\mathbf{X}^T]^{-1}$, the asymptotic variance becomes

$$v^2(\tau_I|\mathbf{x}) = \frac{q(\tau_I)^2}{(1-\tau_I)n} \gamma^2 \mathbf{x}^T E[H(X)^{-\gamma}\mathbf{X}\mathbf{X}^T]^{-1} E[\mathbf{X}\mathbf{X}^T] E[H(X)^{-\gamma}\mathbf{X}\mathbf{X}^T]^{-1} \mathbf{x},$$

which is similar in form to the asymptotic variance of the linear estimator of Lemma 3 of Wang *et al.* (2012) [44]. Then, the rate of convergence of the MISE of the linear estimator is $E[\{\tilde{q}_Y(\tau_I|x)/q_Y(\tau|x) - 1\}^2] = O(\{(1 - \tau_I)n\}^{-1})$. Thus, it can be considered that Theorem 3.2 is the generalization of the asymptotic result of the linear-type parametric estimator.

Remark 2. Claeskens *et al.* (2009) [9] have studied the asymptotic properties of the penalized spline mean estimator in two scenarios: roughly speaking, case (a) small K scenario and case (b) large K scenario. In case (a), the asymptotic behavior of penalized splines is similar to that of regression splines, which have the unpenalized estimator ($\lambda \equiv 0$). Case (b) results in the penalized splines nearing the smoothing splines. We briefly describe the asymptotic scenario branch along with the result of Claeskens *et al.* (2009) [9]. The penalized spline mean estimator is obtained as $\hat{f}(x) = \mathbf{B}(x)^T (Z^T Z + (\lambda/n) D_m^T R D_m)^{-1} Z^T \mathbf{y}$, where $Z = (\mathbf{B}(x_1), \dots, \mathbf{B}(x_n))^T$ is the design matrix and $\mathbf{y} = (y_1, \dots, y_n)$. Then, the two asymptotic scenarios are divided by the asymptotic order of the maximum eigenvalue of $(Z^T Z + (\lambda/n) D_m^T R D_m)^{-1}$, which is obtained as $K(m)^{2m}$

$$K(m) = K \left(\frac{\lambda}{n} \right)^{1/2m} (1 + o(1)).$$

If $K(m) < 1$ for a sufficiently large n , K , and λ , we achieve case (a). When $K(m) > 1$ for a sufficiently large n , K , and λ , we achieve case (b). Although Claeskens *et al.* (2009) [9] focused only on mean regression, these two scenarios can also be discussed with respect to quantile regression. The asymptotic scenario branch discussed in this section is dependent on the asymptotic order of $\Lambda(h)^{-1}$. Similar to Claeskens *et al.* (2009) [9], the order of the maximum eigenvalue of $\Lambda(h)^{-1}$ can be obtained as $K(m, \tau_I)^{2m}$, which corresponds to $K(m)$ in mean regression. Consequently, condition C1 indicates that the large K scenario should be studied. We finally note why we focus on the large K scenario. Ruppert (2002) [41] recommended that one should first set the knots with a large K to obtain the overfitted estimator and control λ to achieve smoothness and fitness. Therefore, the large K scenario matches the concept of Ruppert (2002) [41] and this motivates us to consider the large K scenario.

3.2. Estimation of extreme order quantiles

For the extreme order quantile, the estimator $\tilde{q}_Y(\tau_E|x)$ discussed in the previous section would not have asymptotic normality (Chernozhukov and Fernández-val 2011 [7]). In this paper, we try to approximate the extreme conditional quantile from intermediate quantile. According to Weissman (1978) [45], the following holds:

$$q_Y(\tau_E|x) \approx \left(\frac{1 - \tau_I}{1 - \tau_E} \right)^\gamma q_Y(\tau_I|x).$$

From this, using the estimator of the intermediate order quantile $\tilde{q}_Y(\tau|x)$, we define the extrapolated estimator of the extreme order quantile. To achieve this, we need to estimate the EVI γ .

Let $\tau_1 > \dots > \tau_k$ be the sequence of quantile levels, where $\tau_j = 1 - ([n^\eta] + j)/(n + 1)$, $\eta \in (0, 1)$ and $[a]$ is the integer part of a . Then, since $(1 - \tau_j)n = n([n^\eta] + j)/(n + 1) \rightarrow \infty$ as

$n \rightarrow \infty$, all τ_j are intermediate order quantiles. Using this sequence, we define the Hill-types estimator of γ as

$$\hat{\gamma}(x) = \frac{1}{k-1} \sum_{j=1}^{k-1} \log \left(\frac{\tilde{q}_Y(\tau_j|x)}{\tilde{q}_Y(\tau_k|x)} \right).$$

In this paper, we assume that the tail behavior of $F_Y(\cdot|x)$ and $F(\cdot)$ are equivalent (see Condition A1). Therefore, it is somewhat unnatural that the estimator of γ varies with x . Nevertheless, we define the extrapolated estimator with $\hat{\gamma}(x)$ and investigate the mathematical property. For $x \in [a, b]$, using $\hat{\gamma}(x)$, we define the estimator of the extreme order quantile as

$$\hat{q}_Y(\tau_E|x) = \left(\frac{1 - \tau_I}{1 - \tau_E} \right)^{\hat{\gamma}(x)} \tilde{q}_Y(\tau_I|x).$$

We next consider the EVI estimator along with condition A1. Define the common index (pooled) estimator

$$\hat{\gamma}^C = \frac{1}{n} \sum_{i=1}^n \hat{\gamma}(x_i)$$

and the extrapolated estimator with common index estimator $\hat{\gamma}^C$ as

$$\hat{q}_Y^C(\tau_E|x) = \left(\frac{1 - \tau_I}{1 - \tau_E} \right)^{\hat{\gamma}^C} \tilde{q}_Y(\tau_I|x).$$

To investigate the asymptotic distribution of $\hat{q}_Y(\tau_E|x)$ and $\hat{q}_Y^C(\tau_E|x)$, we impose the second-order condition in Conditions A:

- A5.** The function $U(t) = F^{-1}(1 - 1/t)$ satisfies the second-order condition with (γ, ρ, A) . That is, there exist $\rho < 0$ and $A(t) \in RV(\rho)$ such that as $t \rightarrow \infty$,

$$A(t)^{-1} \left\{ \frac{U(tz)}{U(t)} - z^\gamma \right\} \rightarrow \frac{(z^\gamma)(z^\rho - 1)}{\rho}.$$

Furthermore, $A(t) = \gamma dt^\rho$ with $d \neq 0$.

- A3'.** There exist $\delta > 0$ and positive, continuous and bounded function $H_1(x)$ such that as $y \rightarrow \infty$,

$$1 - F_V(y|x) = H(x)\{1 - F(y)\} + H_1(x)(1 - F(y))^{1+\delta}(1 + o(1))$$

Condition A5 is the standard second-order condition of EVT and is detailed in de Haan and Ferreira (2006) [13]. Condition A3' provides the second order of tail behavior of $F(y)$. From conditions A5 and A3', we see that $U_Y(\cdot|x)$ also satisfy the second-order condition with $(\gamma, \rho^* = \min\{\rho, -\delta\}, A^*(\cdot|x))$ and $A^*(t|x) = \gamma d^*(x)t^{\rho^*}$ with $d^*(x) \neq 0$, which were proven in Lemma 2 of Wang *et al.* (2012) [44]. Using this, we show the asymptotic property of the Hill-type estimator of the EVI in the following.

Theorem 3.3. Suppose that the smoothing parameter included in $\tilde{q}(\tau_I|x)$ satisfies $\lambda = O(q(\tau_I)^2\{(1 - \tau_I)n\}^{1/(2m+1)})$. Furthermore, suppose that $k \rightarrow \infty$, $k/n \rightarrow 0$, $n^\eta \log(k)/k^{m/(2m+1)} \rightarrow 0$ and $k^{m/(2m+1)}(n/k)^{\max\{-\gamma, -\delta, \rho\}} \rightarrow 0$ as $n \rightarrow \infty$. Under Conditions A–C, as $n \rightarrow \infty$,

$$\frac{\hat{\gamma}(x) - \gamma - b(k|x)}{v(k|x)} \xrightarrow{D} N(0, 1),$$

and

$$\frac{\hat{\gamma}^C - \gamma - E[b(k|X)]}{E[v(k|X)]} \xrightarrow{D} N(0, 1),$$

where $b(k|x)$ and $v(k|x)$ are defined in (B.4) of Appendix B and have an asymptotic order $O(k^{-m/(2m+1)})$. Furthermore,

$$E[\{\hat{\gamma}(x) - \gamma\}^2] = O\left(k^{-2m/(2m+1)}\right)$$

and

$$E[\{\hat{\gamma}^C - \gamma\}^2] = O\left(k^{-2m/(2m+1)}\right).$$

Using Theorem 3.3, we obtain the asymptotic normality of the ratio of $\hat{q}_Y(\tau_E|x)$ and $\hat{q}_Y^C(\tau_E|x)$.

Theorem 3.4. Suppose that the same conditions as Theorem 3.3. Furthermore, assume that $k^{-m/(2m+1)} \log\{(1 - \tau_I)/(1 - \tau_E)\} \rightarrow 0$ as $k, n \rightarrow \infty$, $\tau_I, \tau_E \rightarrow 1$. As $n \rightarrow \infty$, $\tau_E \rightarrow 1$ and $n(1 - \tau_E) \rightarrow c \in [0, \infty)$,

$$\frac{\frac{\hat{q}_Y(\tau_E|x)}{q_Y(\tau_E|x)} - 1 - \text{bias}(\tau_E|x)}{s(\tau_E|x)} \xrightarrow{D} N(0, 1)$$

and

$$\frac{\frac{\hat{q}_Y^C(\tau_E|x)}{q_Y(\tau_E|x)} - 1 - \text{bias}^C(\tau_E|x)}{s^C(\tau_E|x)} \xrightarrow{D} N(0, 1)$$

where $\text{bias}(\tau_E|x)$, $s(\tau_E|x)$, $\text{bias}^C(\tau_E|x)$ and $s^C(\tau_E|x)$ are defined in (B.5), (B.6), (B.7) and (B.8) of Proof of Theorem 3.4 on Appendix B. Furthermore,

$$E\left[\left\{\frac{\hat{q}_Y(\tau_E|x)}{q_Y(\tau_E|x)} - 1\right\}^2\right] = O\left(\max\left\{k^{-\frac{2m}{2m+1}} \log^2\left(\frac{1 - \tau_I}{1 - \tau_E}\right), \{(1 - \tau_I)n\}^{-\frac{2m}{2m+1}}\right\}\right).$$

and

$$E\left[\left\{\frac{\hat{q}_Y^C(\tau_E|x)}{q_Y(\tau_E|x)} - 1\right\}^2\right] = O\left(\max\left\{k^{-\frac{2m}{2m+1}} \log^2\left(\frac{1 - \tau_I}{1 - \tau_E}\right), \{(1 - \tau_I)n\}^{-\frac{2m}{2m+1}}\right\}\right).$$

For the asymptotic order in Theorem 3.4, the term $O(k^{-2m/(2m+1)} \log^2\{(1 - \tau_I)/(1 - \tau_E)\})$ is derived from $\hat{\gamma}(x)$ and the another term is derived from $\tilde{q}_Y(\tau_I|x)$. If we use $\tau_I = \tau_1 = (n - [n^\eta])/(n + 1)$ or $\tau_I = O(\tau_1)$, we have $(1 - \tau_I)n = O([n^\eta]) = o(k)$ since $[n^\eta]/k \rightarrow 0$.

That is, the asymptotic inference of $\hat{q}_Y(\tau_E|x)$ is dominated by that of $\tilde{q}_Y(\tau_I|x)$ and hence, the rate of convergence of the estimator is

$$E \left[\left\{ \frac{\hat{q}_Y(\tau_E|x)}{q_Y(\tau_E|x)} - 1 \right\}^2 \right] = O \left(\{(1 - \tau_I)n\}^{-\frac{2m}{2m+1}} \right).$$

One may have sense of discomfort with this result since the extreme order quantile estimator and the intermediate order quantile estimator has same rate of convergence. Indeed, the leading terms of $\tilde{q}_Y(\tau_I|x)$ and $\hat{q}_Y(\tau_E|x)$ are similar. However, the convergence speed of the subsequent term of $\hat{q}_Y(\tau_E|x)$ is obviously slower than that of $\tilde{q}_Y(\tau_I|x)$ because of the influence of $\hat{\gamma}(x)$. Therefore, for the application with a finite sample, the behavior of $\tilde{q}_Y(\tau_I|x)$ would be more stable than $\hat{q}_Y(\tau_E|x)$. On the other hand, when $\tau_I = \tau_k$ or $\tau_I = O(\tau_k)$, which leads to $n(1 - \tau_I) = O(k)$, is adopted, $\hat{q}_Y(\tau_E|x)$ is heavily affected by $\hat{\gamma}(x)$ but not by $\tilde{q}_Y(\tau_I|x)$. Actually, since $n(1 - \tau_E) \rightarrow c \in [0, \infty)$ and $\log(\{1 - \tau_I\}/\{1 - \tau_E\}) = \log(\{n(1 - \tau_I)\}/\{n(1 - \tau_E)\}) = O(\log[k/\{n(1 - \tau_E)\}])$, we have

$$\begin{aligned} E \left[\left\{ \frac{\hat{q}_Y(\tau_E|x)}{q_Y(\tau_E|x)} - 1 \right\}^2 \right] &= O \left(\max \left\{ k^{-\frac{2m}{2m+1}} \log^2 \left(\frac{k}{n(1 - \tau_E)} \right), k^{-\frac{2m}{2m+1}} \right\} \right) \\ &= O \left(k^{-\frac{2m}{2m+1}} \log^2 \left(\frac{k}{n(1 - \tau_E)} \right) \right). \end{aligned}$$

For the common index quantile estimator $\hat{q}_Y^C(\tau_E|x)$ with $\tau_I = O(\tau_k)$, the asymptotic order of $\hat{q}_Y^C(\tau_E|x)/q_Y(\tau_E|x)$ are dominated by the term of $\hat{\gamma}^C$, and this do not vary with x . This result is quite unnatural in the quantile regression although $O(\log^2(k/\{n(1 - \tau_E)\}))$, which is the difference between the asymptotic inference of $\hat{\gamma}^C$ and $\tilde{q}_Y(\tau_I|x)$, is quite small. Therefore, if the common index quantile estimator is mainly used, we may have to choose the baseline quantile τ_I so that $\tau_I > \tau_k$. Thus, the balance of τ_I and k controls the asymptotic behavior of $\hat{q}_Y(\tau_E|x)$. The same is true of $\hat{q}_Y^C(\tau_E|x)$.

Wang *et al.* (2012) [44] obtained the extrapolated estimator in the linear model with $\tau_I = \tau_k$. From their result, the rate of convergence of the MISE of the linear estimator is $E[\{\hat{q}_Y(\tau_E|x)/q_Y(\tau_E|x) - 1\}^2] = O(k^{-1} \log^2(\{1 - \tau_I\}/\{1 - \tau_E\}))$. That is, the difference in the rate of convergence between the parametric estimator and the nonparametric estimator is k^{-1} and $k^{-2m/(2m+1)}$, which could be intuitively derived from the classical works on parametric and nonparametric regression.

Remark 3. The intermediate order quantile and the extreme order quantile are separated mathematically by the rate of the quantile level. However, in data analysis, the distinction between these two rates should be drawn for fixed n . Define $\xi = \xi(\tau, n) = (1 - \tau)n$. Using ξ , Chernozhukov and Fernández-val (2011) [7] suggested the following rule of thumb. For the quantile level τ , if $\xi < 30$, it is the extreme order inference, that is, $\tau = \tau_E$ and we should use $\hat{q}_Y(\tau|x)$. When $\xi \geq 30$, it is sufficient to use the intermediate order quantile estimator $\tilde{q}_Y(\tau|x)$. If the predictor is the continuous, this threshold is $\xi = 15-20$. However, they noted that the above rule is conservative. In this paper, we treat $\tau_1 = \frac{n - [n^\eta]}{n+1}$ as the intermediate order quantile. For example, when $n = 200$, $\tau = 0.925$ leads to $\xi = 15$. Then, we have $\eta \approx 0.5$. For $n = 1000$, $\tau = 0.985$ and $\eta = 0.4$ correspond to $(1 - \tau_1)n \approx (1 - \tau)n = 15$.

On the other hand, Wang *et al.* (2012) [44] suggested to use $\eta = 0.1$. In their rule, we have $\xi(\tau_1, n) = 3$ for $n < 1025$. Thus, it seems that the rule of Chernozhukov and Fernández-val (2011) [7] is more conservative rather than that of Wang *et al.* (2012) [44]. In our experience, the rule of Wang *et al.* (2012) [44] worked well for $n \leq 1000$. Therefore, in the simulation study of the next section, we also use $\eta = 0.1$. However, the determination of the split of the intermediate order and the extreme order is still a difficult problem and further study would be welcomed.

4. SIMULATION

The practical performance of the proposed estimator is confirmed by Monte Carlo simulation. Define the true regression model as

$$(4.1) \quad Y_i = f(X_i) + \sigma(X_i)\varepsilon_i(X_i), \quad i = 1, \dots, n,$$

where

$$f(x) = \sqrt{x(1-x)} \sin\left(\frac{2\pi(1+2^{-7/5})}{x+2^{-7/5}}\right)$$

and $\sigma(x) = 10^{-1}(1+x)$. The predictor X_i is independently generated from the standard uniform distribution. This setting was introduced by Daouia *et al.* (2013) [11]. We consider two types of error distribution: (a) $\varepsilon_i(X_i) = \varepsilon_i \sim t_5$ and (b) $\varepsilon_i(X_i) \sim t_{s(x)}$, where

$$s(x) = [\nu(x)] + 1, \quad \nu(x) = [\{1.1 - 0.5 \exp[-64(x - 0.5)^2]\} \{0.1 + \sin(\pi x)\}]^{-1}.$$

The error type (b) is also used by Daouia *et al.* (2013) [11]. For the t_ν distribution, the EVI is $\gamma = 1/\nu$ and hence, $\gamma = 0.2$ and $\gamma(x) = 1/s(x)$ for (a) and (b), respectively. For both cases, the EVI is larger than 0, which indicates that the distribution of Y_i has a heavy tail. In (4.1), the conditional τ th quantile of Y given $X = x$ is $q_Y(\tau|x) = f(x) + \sigma(x)q_\varepsilon(\tau|x)$, where $q_\varepsilon(\tau|x)$ is the τ th quantile of $\varepsilon_i(x)$. For the case (a), $q_\varepsilon(\tau|x) = q(\tau)$ is the τ th quantile of t_5 and is not dependent on x . Thus, the model (4.1) with (a) is the location-scale shifted model and is of the form of (2.2). In the case of (b), $\gamma(x) = 1/2$ for $x \in [0.12, 0.88]$ and $\gamma(x) \in (0, 1/3)$ otherwise. That is, the model (4.1) with (b) has high EVI at the center and low (but larger than 0) EVI at the boundary. The conditional quantile with (b) fail due to Conditions A. However, it is important to confirm the performance of the estimator under (b).

We construct the intermediate order quantile estimator $\tilde{q}_Y(\tau|x)$, the Hill-type estimator $\hat{\gamma}(x)$, $\hat{\gamma}^C$, and the extreme order quantile estimator $\hat{q}_Y(\tau|x)$ and $\hat{q}_Y^C(\tau|x)$. For the intermediate order quantile estimator, we use the number of knots $K = 40$ and the smoothing parameter selected via generalized approximated cross-validation (Yuan 2006 [47]). To obtain $\hat{\gamma}$, $\hat{\gamma}^C$, \hat{q}_Y and \hat{q}_Y^C , we need to determine $\tau_j = 1 - \frac{[n^\eta]+j}{n+1}$ ($j = 1, \dots, k$). In this simulation, η is chosen so that $(1 - \tau_1)n = \xi = 3$ and $k = [7.5n^{1/3}]$. Such η and k are selected from a pilot study. Wang *et al.* (2012) [44] used $\eta = 0.1$ and $k = [4.5n^{1/3}]$ in the linear regression. Thus, our k is somewhat larger than that of Wang *et al.* (2012) [44].

For the estimator $\hat{f}(x)$ of the true function $f(x)$, the Mean Integrated Squared Error (MISE):

$$MISE(\hat{f}) = \int_0^1 E[\{\hat{f}(x) - f(x)\}^2] dx$$

is used as the accuracy measure of the estimator. We calculate the estimated MISE of $\tilde{q}_Y(\tau|x)$, $\hat{q}_Y(\tau|x)$ and $\hat{q}_Y^C(\tau|x)$ over 400 replications. The estimators $\tilde{q}_Y(\tau|x)$, $\hat{q}_Y(\tau|x)$ and $\hat{q}_Y^C(\tau|x)$ are denoted by PSE-I, PSE-E and PSE-Ep. As the competitors, we consider the functional nonparametric estimator (Gardes *et al.* 2010 [20]) and the kernel smoothing estimator (Daouia *et al.* 2013 [11]). The estimators $\hat{q}_1(\tau, x)$ and $\hat{q}_2(\tau, x)$ defined in Gardes *et al.* (2010) [20] are denoted by FNS-I and FNS-E, respectively. Furthermore, the estimators $\hat{q}_n(\tau|x)$ and $\tilde{q}_n^{RP}(\tau|x)$ defined by Daouia *et al.* (2013) [11] are labeled by KSE-I and KSE-E in this section. Thus, FNS-I, FNS-E, KSE-I and KSE-E are also demonstrated in simulation.

We report the simulation results for the case (a). Figure 1 shows the true conditional quantiles and the intermediate order quantile estimators for one dataset with $n = 200, 600$ and 1000 . For $\tau = 0.8, 0.85$ and 0.9 , the estimator behaved well, but for $\tau \geq 0.95$, there was a significant difference between the true function and the intermediate order quantile estimator.

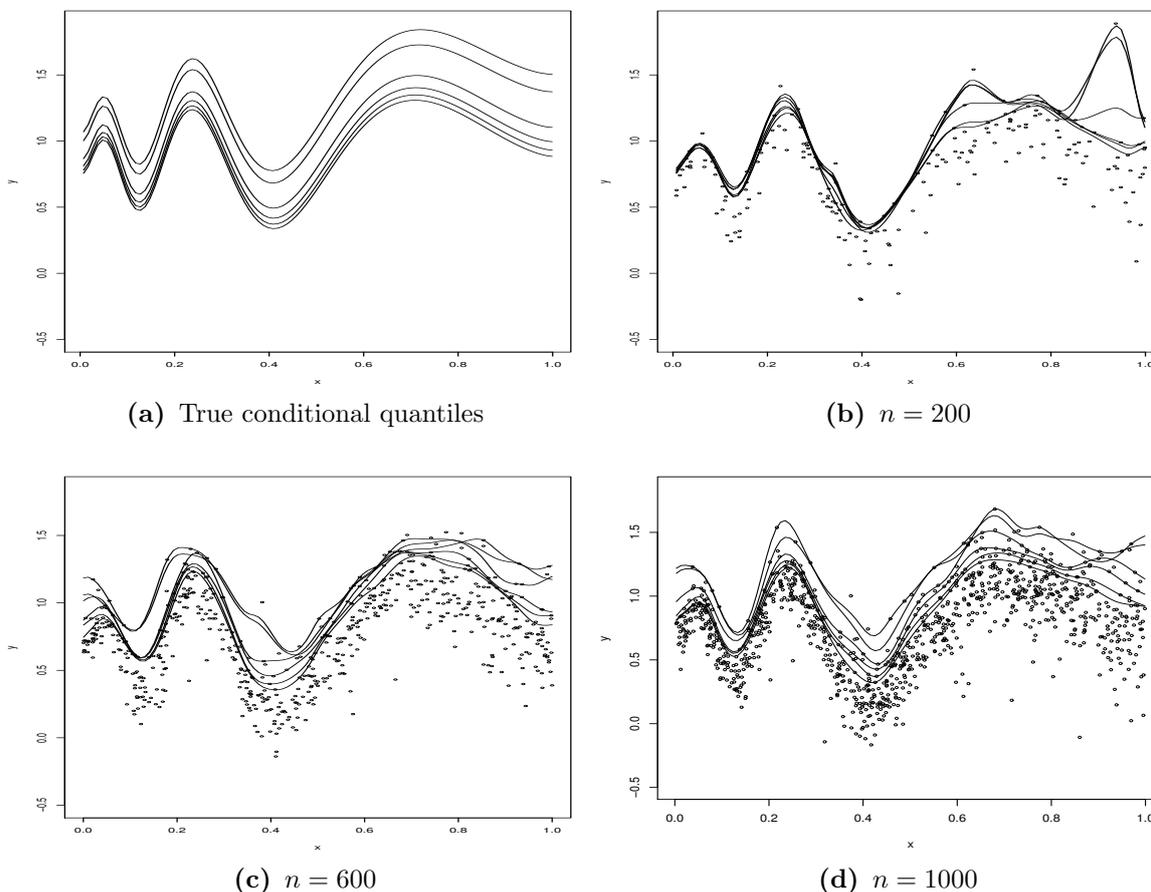


Figure 1: True conditional quantiles for $\tau = 0.8, 0.85, 0.9, 0.95, 0.99$ and 0.995 , and these intermediate order quantile estimators for one dataset.

In Figure 2, the MISEs of the estimators for $\tau \in [0.5, 0.995]$ are illustrated. We can observe that the proposed estimator behaves better than the competitors. From Figure 2 (d), we can find that the estimator behaves well as n increases. This indicates that the estimator has a consistency property. However, as τ increases, the performance of the estimator becomes drastically decreases. Therefore, for $\tau \approx 1$, it is difficult to predict the conditional quantile using the intermediate order quantile estimator.

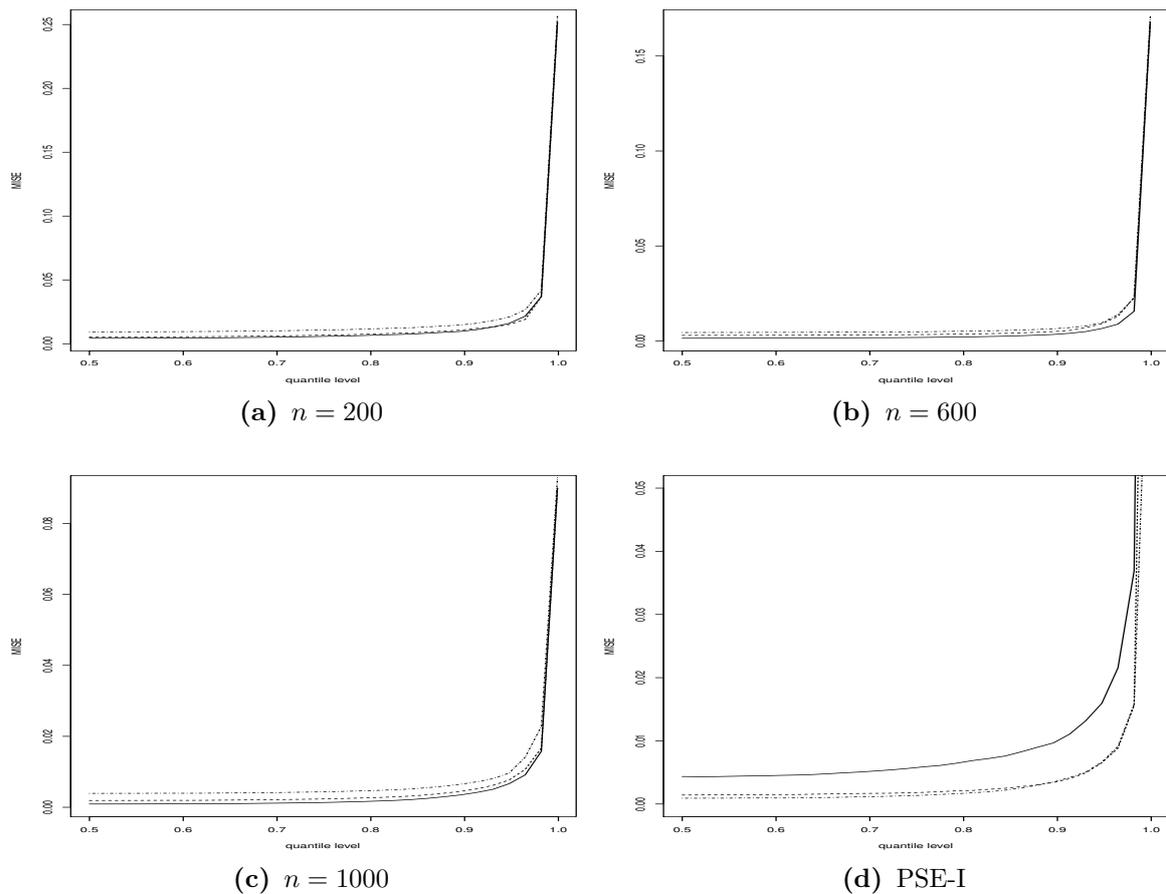


Figure 2: MISE of the intermediate order quantile estimators for $\tau \in [0.5, 0.995]$. In (a), (b) and (c), the solid line is PSE-I. The dashed line and dot-dashed lines are KSE-I and FNS-I, respectively. In (d), the solid, dashed and dot-dashed lines are PSE for $n = 200$, $n = 600$ and $n = 1000$.

We next show the performance of the EVI estimator. Figure 3 shows the behavior of $\hat{\gamma}^C$ over k for one dataset and the distribution of $\hat{\gamma}^C$ using $k = \lceil 7.5n^{1/3} \rceil$ by Monte Carlo simulation. From the results, we see that the suggested $k = \lceil 7.5n^{1/3} \rceil$ is good choice. When $n = 1000$, the behavior of $\hat{\gamma}^C$ is stable from (c) and (d).

Figure 4 shows the extreme order quantile estimators PSE-E and PSE-Ep for one dataset and the MISE of the extreme order quantile estimators for $\tau \in [0.95, 0.999]$. From (a–c), we can observe that the estimator behaves well. We can see that the behavior of the PSE-Ep is stable than the PSE-E. This is not a surprising result since the estimator of EVI included in PSE-Ep is not dependent on x unlike PSE-E. It can be recognized from Figure 4 (d–f) that the proposed estimator has better behavior than the competitors although the differences are not large. Furthermore, the performance of the PSE-Ep was superior to that of PSE-E. We think that this is a result of the stability of $\hat{\gamma}^C$. It can be recognized from Figure 5 that the extrapolated estimator has consistency.

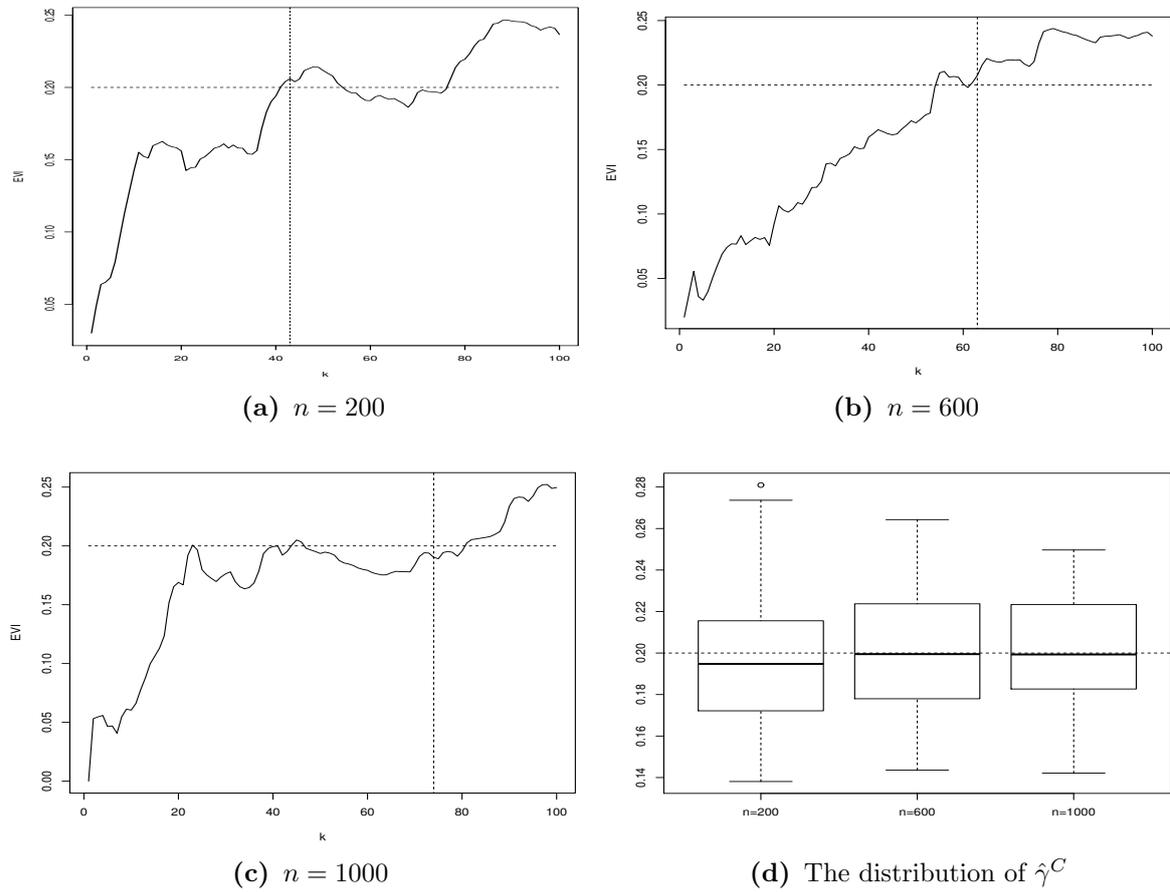


Figure 3: (a-c) The sample path of the common index estimator of EVI with k for one dataset. The dataset is similar to Figure 1 for each n . The dashed line is $k = \lceil 7.5n^{1/3} \rceil$ and $\gamma = 0.2$. (d) Box plot of $\hat{\gamma}^C$ with $k = \lceil 7.5n^{1/3} \rceil$ from 400 replications.

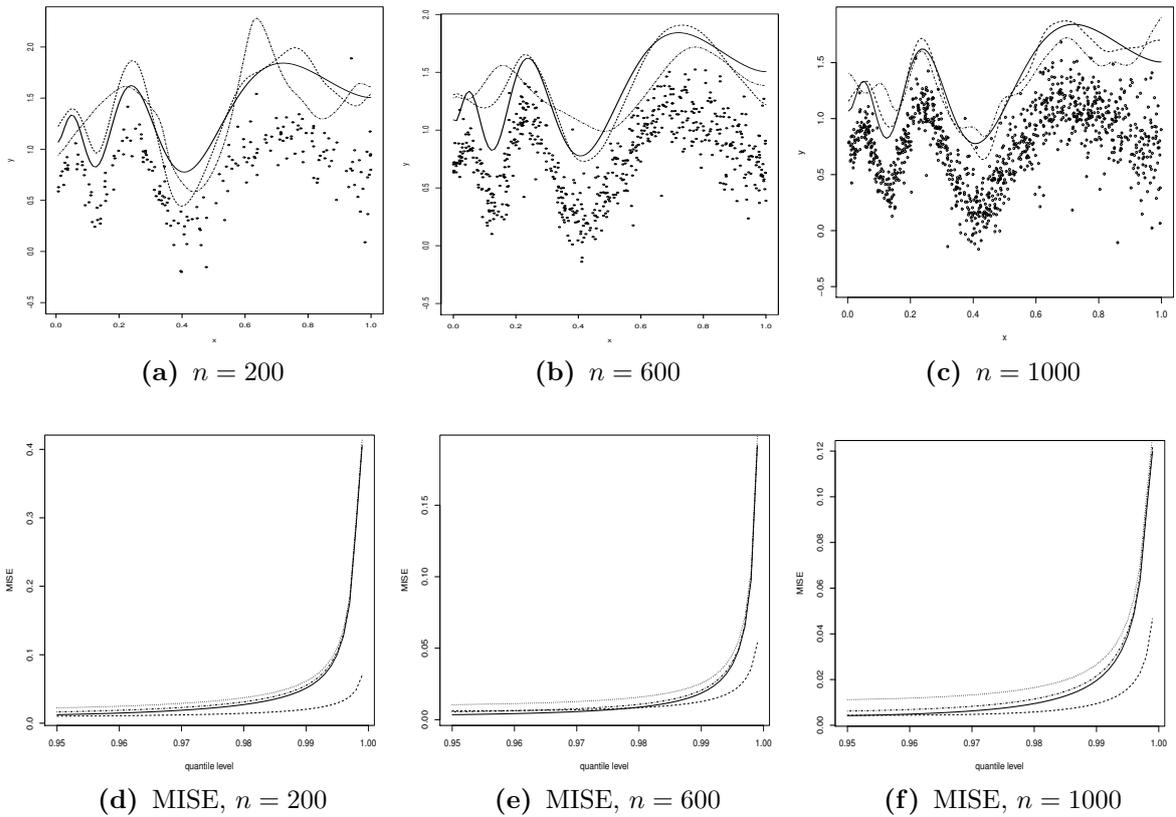


Figure 4: (a–c) The true conditional quantiles (solid) and the extreme order quantile estimator PSE-E (dot-dashed) and PSE-Ep (dashed) for $\tau = 0.995$ for one dataset. The dataset is similar to that given in Figure 1 for each n . (d–f) MISE of the estimators for $\tau \in [0.95, 0.999]$. The solid, dashed, dotted and dot-dashed lines are PSE-E, PSE-Ep, FNS-E and KSE-E, respectively.

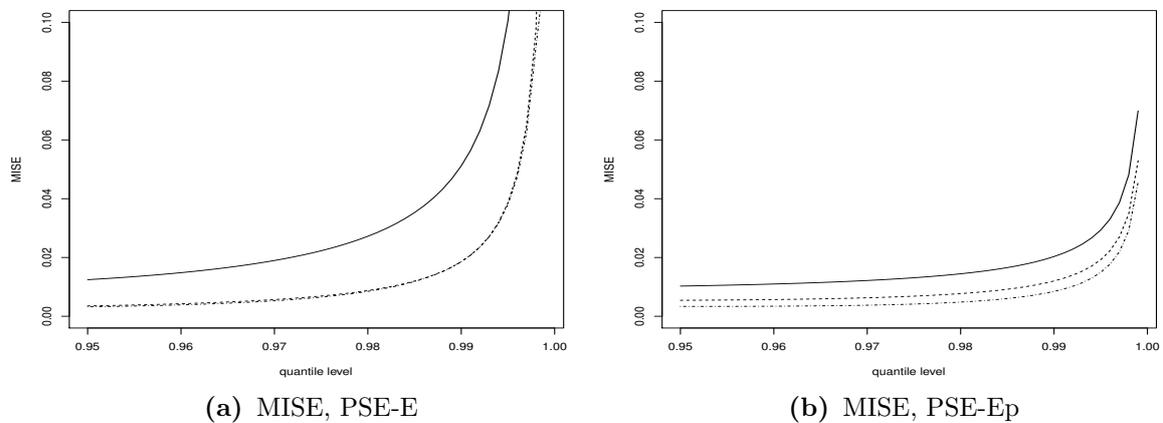


Figure 5: (a) and (b) are the MISE of the PSE-E and the PSE-Ep, respectively. The results are similar to Figure 4 (d–f). For both panels, solid, dashed and dot-dashed lines are for $n = 200, 600$ and 1000 , respectively.

From now on, we describe the simulation results for the model (b). Figure 6 shows the true conditional quantiles and the intermediate order quantile estimators for $\tau \in [0.8, 0.995]$ for one dataset. It appears that for $\tau \in [0.8, 0.9]$, the estimator can capture the true conditional quantile even for $n = 200$. However, for $\tau \geq 0.95$, the estimator has a wiggly curve. In Figure 7, the results of MISE of the intermediate order quantile estimators for each n are illustrated. We found that the proposed estimator performs well for $\tau \in [0.5, 0.95)$. However, the MISE drastically grows as τ increases. The behaviors of PSE-E and PSE-Ep for one dataset are described in Figure 8 (a–c). It can be seen from Figures 6 and 8 (a–c) that the PSE-E and PSE-Ep performed better than PSE-I. Figure 8 (d)–(f) shows the MISE of the extreme order quantile estimators. It can be confirmed that the performance of PSE-E is slightly better than that of PSE-Ep. We see that the proposed estimators have better behavior than the competitors. Figure 9, the consistency of the PSE-E and the PSE-Ep can be observed in numerically. Although the performance of the proposed estimator is drastically superior to that of Daouia *et al.* (2013) [11], this simulation result indicates that our method is one of useful tools to the problem of extremal quantile regression.

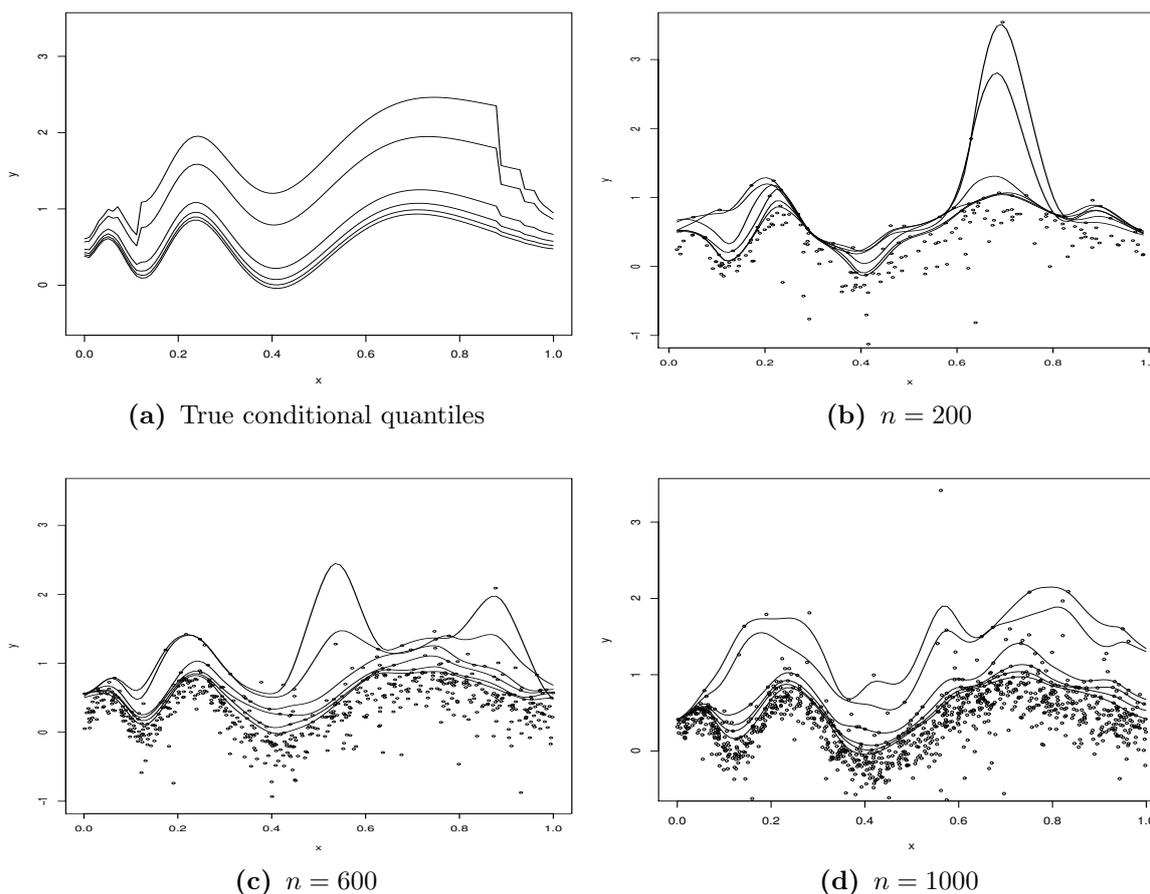


Figure 6: True conditional quantiles for $\tau = 0.8, 0.85, 0.9, 0.95, 0.99, 0.995$ and these intermediate order quantile estimators for one dataset with model (b).

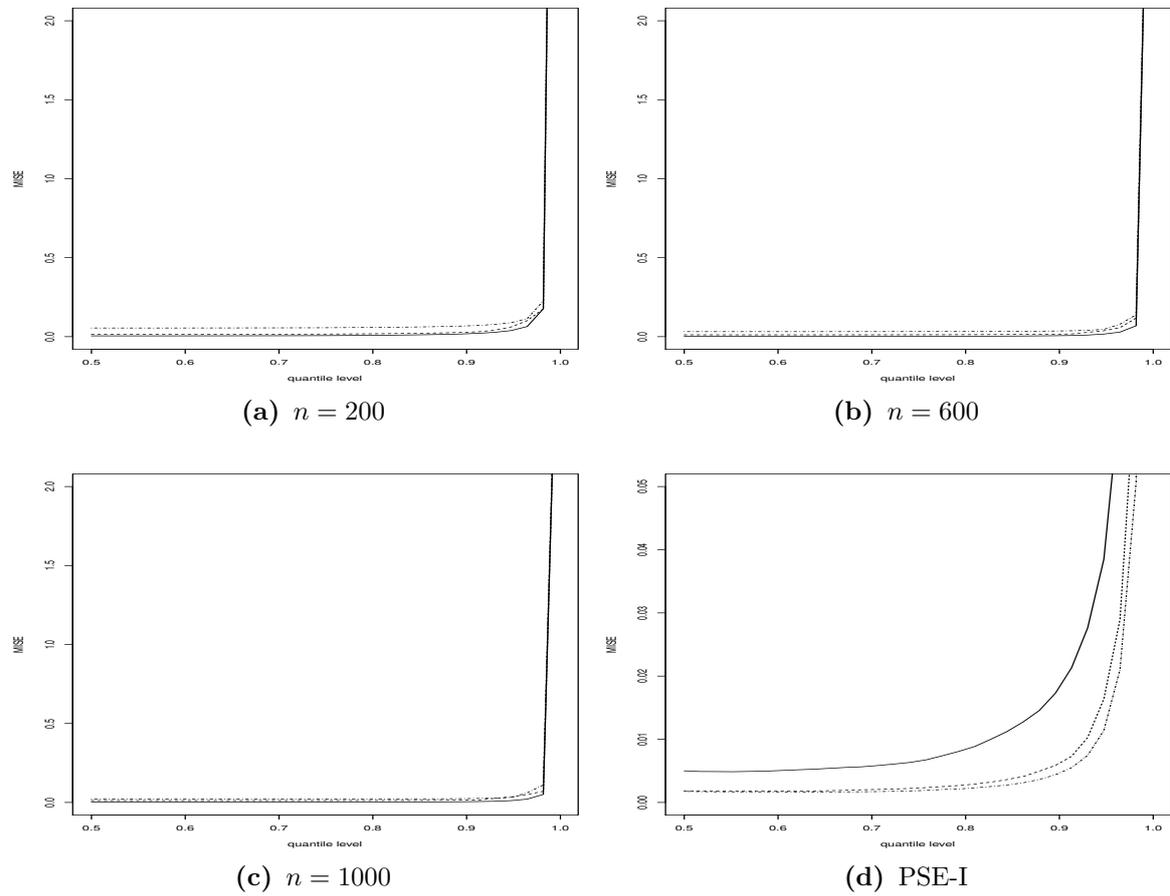


Figure 7: MISE of the intermediate order quantile estimators for $\tau \in [0.5, 0.995]$. The description is similar to Figure 2.

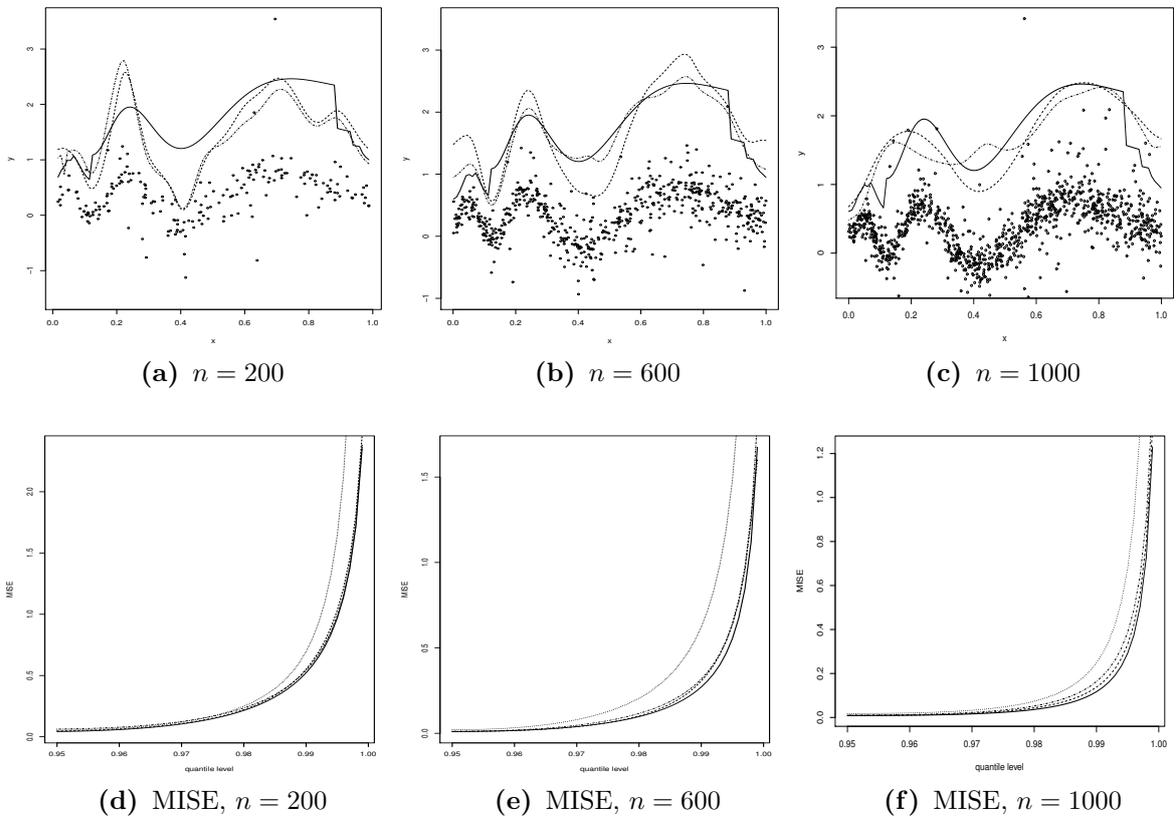


Figure 8: (a–c) The true conditional quantiles the extreme order quantile estimators for $\tau = 0.995$ for one dataset with model (b). The dataset is similar to that given in Figure 6 for each n . (d–f) MSE of the estimators for $\tau \in [0.95, 0.999]$. The description is similar to Figure 4.

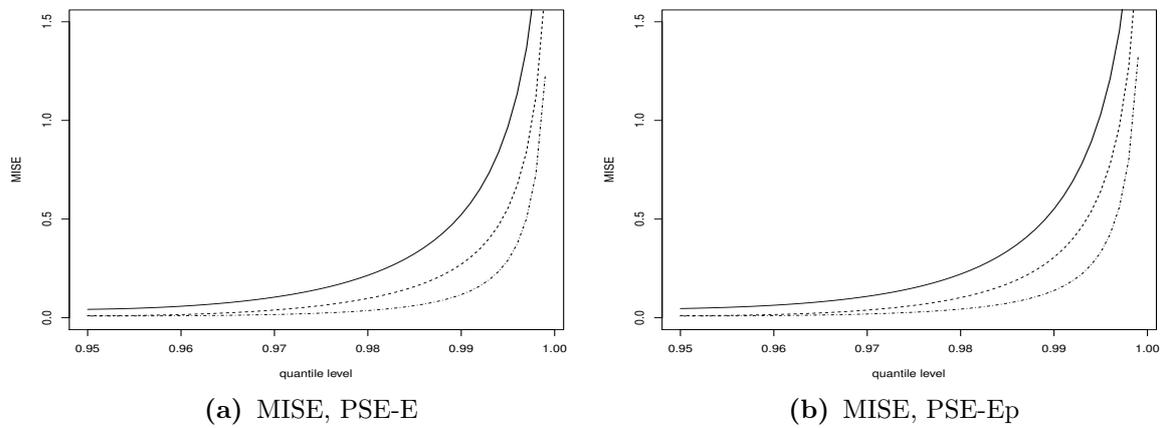


Figure 9: (a) and (b) are the MISE of the PSE-E and the PSE-Ep, respectively. The description is similar to Figure 5.

5. DATA EXAMPLE

In this section, we apply the proposed methods to Beijing’s PM_{2.5} Pollution data. The data is available from the website of the UCI Machine Learning Repository and Liang *et al.* (2015) [34] provided several analyses for this data. One of fundamental purposes of this data is to analyze the relationship between PM_{2.5} concentration and other meteorological variables. Our particular interest here is the prediction of high conditional quantiles of Y , PM_{2.5} concentration ($\mu\text{g}/\text{m}^3$), with the predictor x , temperature (degrees Celsius). We can observe from the scatter plot of y and x (see Figure 10) that this relationship is not linear for the upper quantile. Therefore, the nonparametric approach is suitable for this data. We demonstrate the analysis for each year from 2011 to 2014. We then omit the missing data and hence the sample size is $n = 8032, 8295, 8678,$ and 8661 in 2011, 2012, 2013, and 2014. We construct the extreme order quantile estimator for $\tau = 0.999$. The quantile level $\tau = 0.999$ indicates that about only eight events occur each year.

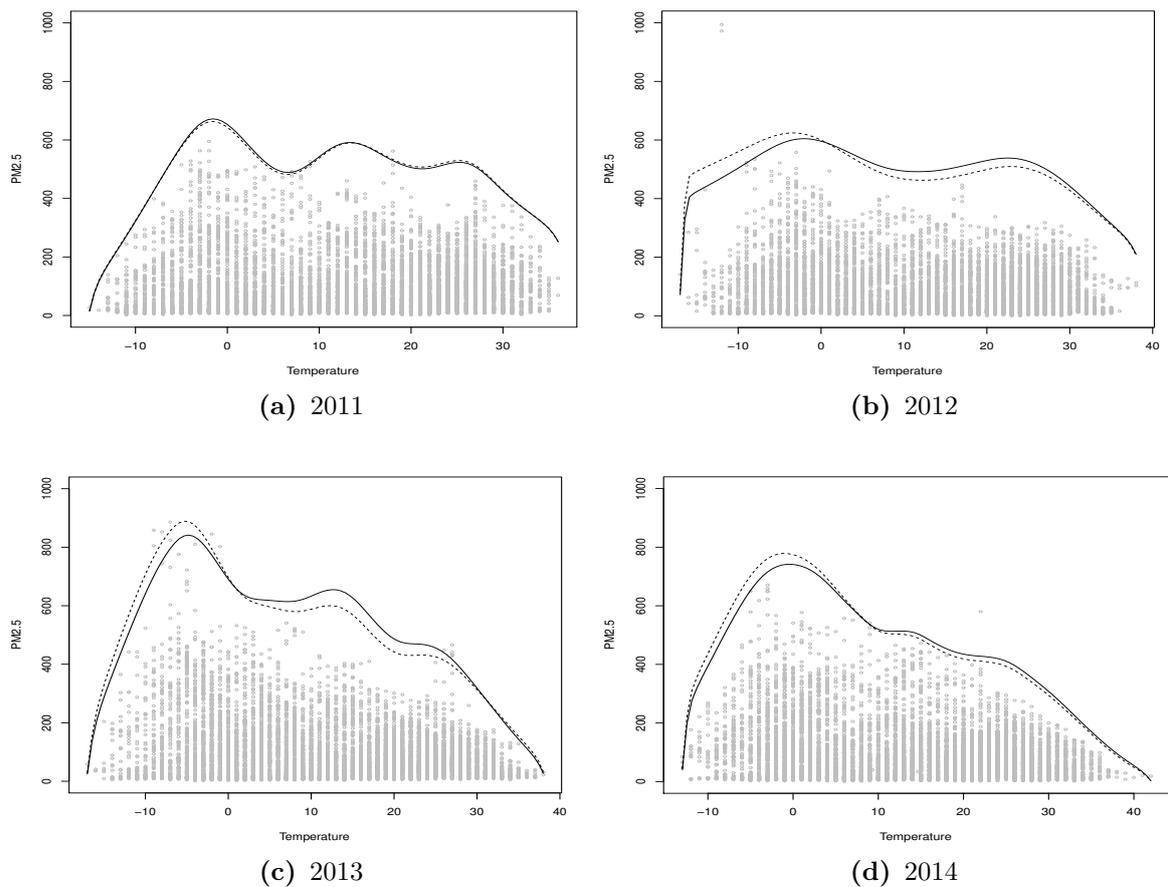


Figure 10: The proposed extreme order quantile estimator for $\tau = 0.999$ for years from 2011 to 2014. The solid line is $\hat{q}_Y^C(0.999|x)$ and the dashed line is $\hat{q}_Y(0.999|x)$ for each panel.

Figure 10 shows the extreme order quantile estimators for $\tau = 0.999$. It seems that the tail behavior is stable in 2011 and 2012. In 2013 and 2014, the estimator of the conditional quantile at $x \in [0, 10]$ has a large value compared with those in 2011 and 2012. Thus, in the cold season of 2013 and 2014, the risk of PM_{2.5} of pollution was increased. We can observe that $\hat{q}_Y(0.999|x)$ and $\hat{q}_Y^C(0.999|x)$ are quite similar. This indicates that the estimator of EVI $\hat{\gamma}(x)$ hardly changes with x . We report the EVI estimator used for constructing the extrapolated estimator \hat{q}_Y and \hat{q}_Y^{pool} . To obtain the estimator of EVI, we utilized $\eta = 0.4$ as $\tau_1 = n - [n^\eta]/(n + 1)$ so that about $\tau_1 = 0.995$. Then, EVI is estimated by using the intermediate order quantiles estimators for $\tau \in (0.978, 0.995)$. This choice leads to $\xi = n(1 - \tau_1) \approx 37$ and, hence, this is a very conservative situation in the study of Chernozhukov and Fernández-val (2011) [7]. Furthermore, we then adopted to use k so that the sample path of the pooled EVI estimator is stable in each year. Figure 11 shows the sample path of $\hat{\gamma}^C$ and selected k . As a result, $\hat{\gamma} = 0.220, 0.226, 0.260,$ and 0.231 in 2011, 2012, 2013, and 2014. Thus, the pooled EVI estimators are the same in 2011, 2012, and 2014, and it is only slightly larger in 2013.

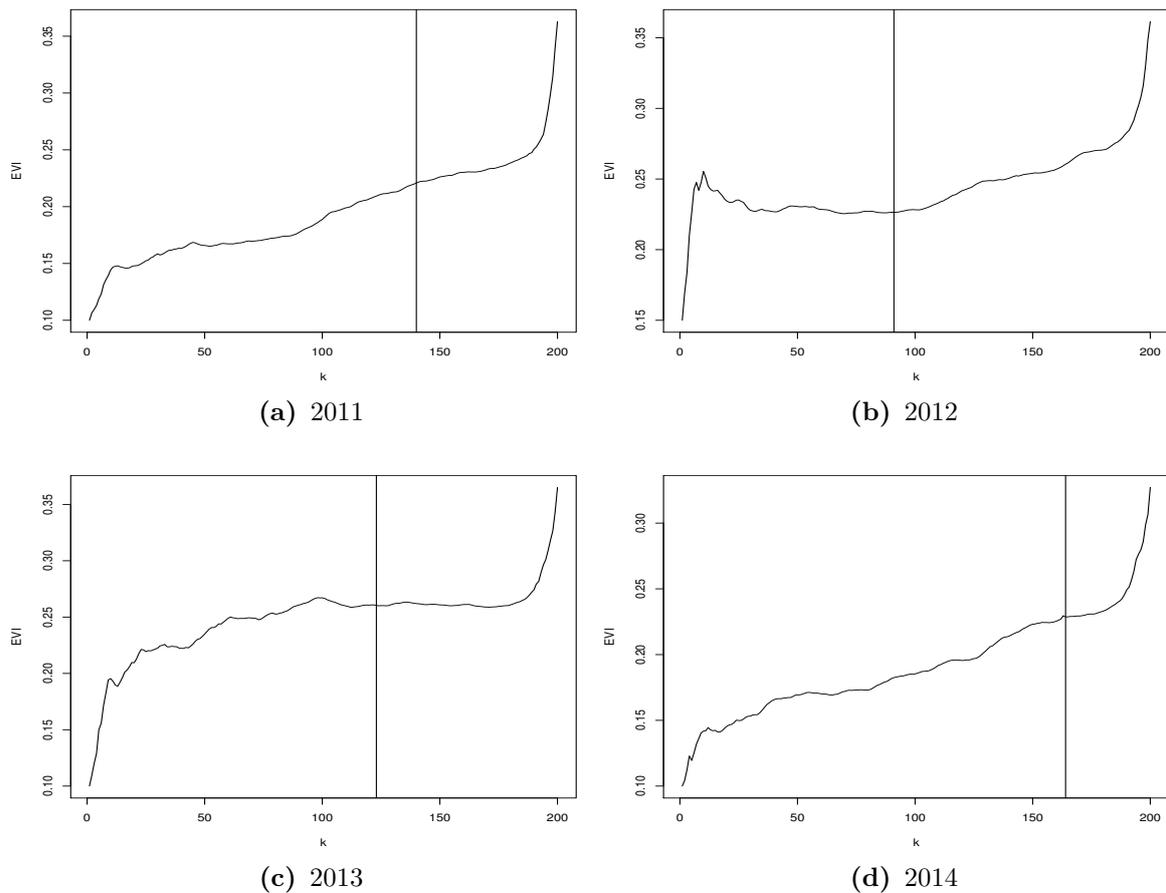


Figure 11: The sample path of EVI estimates $\hat{\gamma}^C$ with k for each year.

Figure 12 illustrates $\hat{\gamma}(x)$ with selected k for each year. We can observe that $\hat{\gamma}(x)$ has a narrow curve with x in 2011, 2012, and 2014. On the other hand, in 2013, $\hat{\gamma}(x)$ at the boundary is rather larger than at the center. Indeed, it can be seen from Figure 10 that the extreme point can be observed at $x < 0$ in 2013.

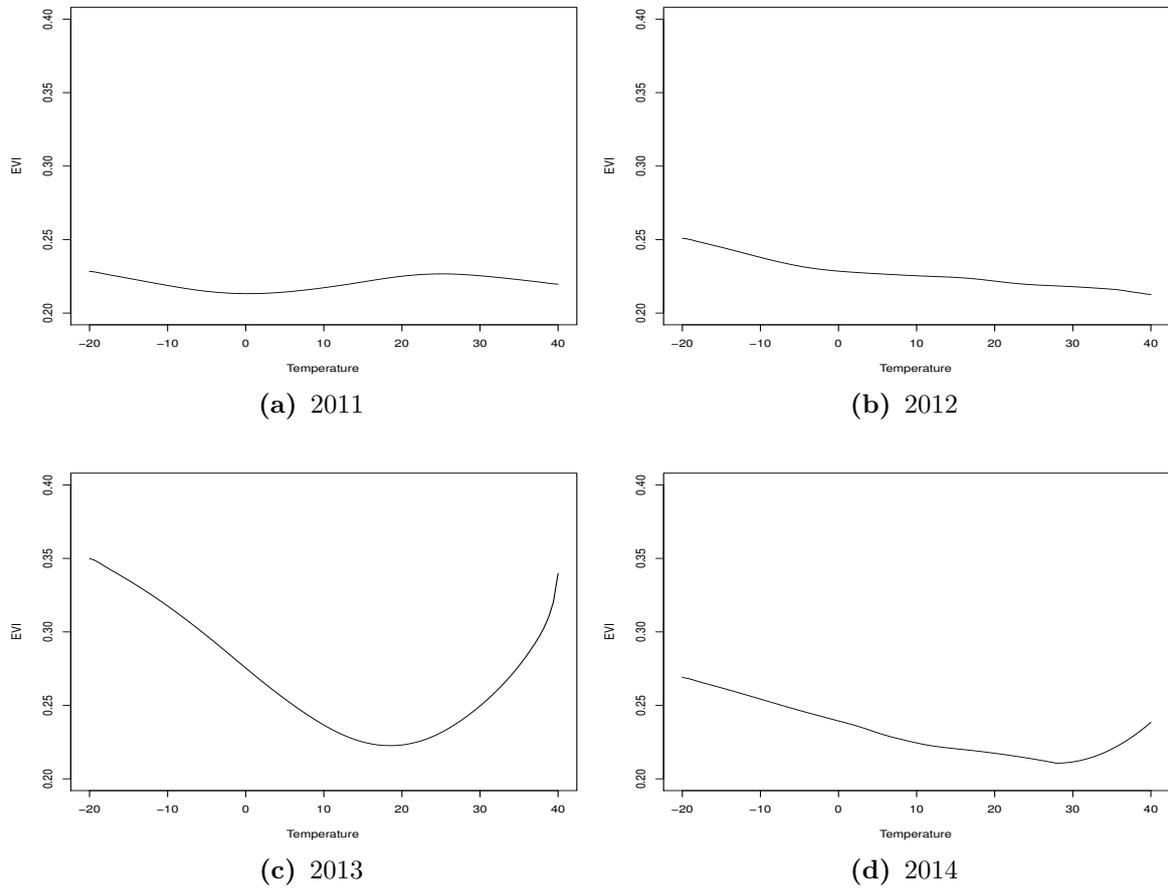


Figure 12: The EVI estimates $\hat{\gamma}(x)$ with selected k versus x for each year.

6. CONCLUSION

We have developed the nonparametric extremal quantile regression methods for heavy-tailed data. To show the mathematical property of the proposed estimator, we have used the hybrid techniques of asymptotic theory for the nonparametric regression and EVT for the tail behavior of the conditional distribution. We then considered two quantile rates: (i) the intermediate order quantile that $(1 - \tau)n \rightarrow \infty$ as $n \rightarrow \infty$; and (ii) the extreme order quantile that $(1 - \tau)n \rightarrow \xi < \infty$. For the intermediate order quantile, the penalized spline estimator and its asymptotic normality have been developed. On the other hand, for the extremal order quantile, we have studied the Weissman-type extrapolated estimator using the intermediate order quantile estimator and its asymptotic normality. For the both intermediate and extreme order quantile, we show the asymptotic normality and the optimal rate of convergence of the proposed estimator. In particular, we found that the convergence speed of the estimator for extremal quantile is slower than that for center quantile. This result would be intuitively correct.

We now discuss some future directions of study. First, for technical reasons, we assumed that the tail behavior of the conditional distribution of Y given $X = x$ is equivalent across the predictor x (see Conditions A1). Since the estimation of the tail behavior is difficult due to data sparsity, this assumption is helpful in data analysis. However, if this assumption is violated, additional research is needed to explicate the performance of the estimator.

Second, in this paper, we focused on the spline smoothing with ℓ_2 penalty. On the other hand, Koenker *et al.* (1994) [32] and Koenker (2011) [30] studied the smoothing spline with the ℓ_1 -type penalty. That is, the penalty is defined as $\int_a^b |s^{(m)}(x)| dx$ instead of $\int_a^b \{s^{(m)}(x)\}^2 dx$. It is known that the estimator with ℓ_1 penalty has local adaptiveness. Therefore, for some cases, the performance of the estimator with ℓ_1 penalty would be better than that with ℓ_2 penalty. Recently, the ℓ_1 -type penalty has been rapidly developed in mean regression (Kim *et al.* 2009 [27]; Tibshirani 2014 [43]; Sadhanala and Tibshirani 2017 [42]). Although it is difficult to show the asymptotic distribution of an ℓ_1 penalized estimator, the developments of the ℓ_1 penalized smoothing to the extremal quantile regression is an interesting problem.

Finally, we can consider extending the proposed method to the multidimensional case. In particular, it is important to use the additive models (Hastie and Tibshirani 1990 [23]) that for $x = (x_1, \dots, x_d) \in \mathbb{R}^d$, the true function is can be decomposed as

$$f(x) = f_1(x_1) + \dots + f_d(x_d),$$

where each f_j is the univariate function. The additive model is known to enables us to avoid the problem of dimensionality. The nonparametric additive quantile regression (for center quantile) was studied by Lu and Yu (2004) [35], Horowitz and Lee (2005) [26], Cheng *et al.* (2011) [8], Koenker (2011) [30], Lee *et al.* (2010) [33] and references therein. However, the extremal inference of the additive quantile regression has not yet been studied until now. It seems that the developments of the extremal quantile regression with the additive model is an important issue.

A. APPENDIX – Computation of the intermediate order quantile estimator

We describe here the approximation algorithm to solve (3.1). Nychka *et al.* (1995) [36] and Reiss and Huang (2012) [40] proposed the penalized iteratively reweighted least squares algorithm. We use here the modified version of Nychka *et al.* (1995) [36]. Nychka *et al.* (1995) [36] proposed the following optimization:

$$(A.1) \quad \tilde{\mathbf{b}} = \underset{\mathbf{b}}{\operatorname{argmin}} \left\{ \sum_{i=1}^n \rho_{\tau,\alpha}(y_i - \mathbf{B}(x_i)^T \mathbf{b}) + \lambda \mathbf{b}^T D_m^T R D_m \mathbf{b} \right\},$$

where

$$\rho_{\tau,\alpha}(u) = \begin{cases} \rho_{\tau}(u), & |u| > \alpha, \\ \tau u^2 / \alpha, & 0 \leq u \leq \alpha, \\ (1 - \tau) u^2 / \alpha, & -\alpha \leq u \leq 0 \end{cases}$$

Obviously, the loss function $\rho_{\tau,\alpha}(u)$ tends to ρ_{τ} as $\alpha \rightarrow 0$. However, for the tail quantile ($\tau \approx 0, 1$), the above algorithm will not converge in our implementation. Therefore, we suggest using the slightly modified version of (A.1). The idea is similar to the proximal gradient method and it is very simple. The modified algorithm is defined as given the t th iteration estimate $\tilde{\mathbf{b}}^{(t)}$,

$$(A.2) \quad \tilde{\mathbf{b}}^{(t+1)} = \underset{\mathbf{b}}{\operatorname{argmin}} \left\{ \sum_{i=1}^n \rho_{\tau,\alpha_t}(y_i - \mathbf{B}(x_i)^T \mathbf{b}) + \lambda \mathbf{b}^T D_m^T R D_m \mathbf{b} + \eta_t \|\mathbf{b} - \tilde{\mathbf{b}}^{(t)}\|^2 \right\},$$

where a_t and η_t are the step sizes. For a_t , it is sufficient to use some sequence $\{\alpha_t\}$ satisfying $\alpha_t \rightarrow 0$ as $t \rightarrow \infty$. On the other hand, the choice of η_t is more important than a_t . If η_t is large, $\mathbf{b}^{(t+1)} \approx \mathbf{b}^{(t)}$ and hence the speed of convergence is very fast. When η_t is small, on the other hand, there is almost no difference between (A.2) and (A.1), that is, the algorithm does not converge in many cases. In Sections 4 and 5, we used $\alpha_t = 0.1 \times 2^{-t}$ and $\eta_{t+1} = 1.2\eta_t (\eta_0 = 1)$.

B. APPENDIX – Proof of theorems

Let Z be the $n \times (K+p)$ design matrix having elements $B_j(x_i)$, $G_n = n^{-1}Z^T \text{diag}[H(x_i)^{-\gamma}]Z$ and let $\Lambda_n = G_n + (\lambda q(\tau)/((1-\tau)n))D_m^T R D_m$. For any matrix $A = (a_{ij})_{ij}$, we denote $\|A\|_\infty = \max_{i,j} |a_{ij}|$. We first state the technical lemmas to prove theorems in this paper.

Lemma B.1. *Suppose that $K(m, \tau) \geq 1$. Under Conditions A–C, the following statements holds: $\|G(1)\|_\infty = O(K^{-1})$, $\|G(H^{-\gamma})\|_\infty = O(K^{-1})$, $\|\Lambda_n^{-1}\|_\infty = O(K(1 + K(m, \tau)^{2m})^{-1})$, $\|\Lambda(H^{-\gamma})^{-1}\|_\infty = O(K(1 + K(m, \tau)^{2m})^{-1})$, $\max_{i,j} \{ |G_n^{-1} - G(H^{-\gamma})^{-1} |_{ij} \} = o(K)$ and $\max_{i,j} \{ | \Lambda_n^{-1} - \Lambda(H^{-\gamma})^{-1} |_{ij} \} = o(K(1 + K(m, \tau)^{2m})^{-1})$.*

Lemma B.2. *Suppose that $K(m, \tau) \geq 1$. Under Assumptions 2–3, for any $K + p$ square matrix A satisfying $\|A\|_\infty = O(n^a)$ for $a \in \mathbb{R}$, $\|G(H^{-\gamma})A\|_\infty = O(n^a K^{-1})$ and $\|\Lambda(H^{-\gamma})^{-1}A\|_\infty = O(n^a K(1 + K(m, \tau)^{2m})^{-1})$.*

Lemma B.1 is proved by Lemma 6.3 and 6.4 of Zhou *et al.* (1998) [48] and Lemma A1 and A2 of Claeskens *et al.* (2009) [9]. Lemma B.2 says that the order of product of matrices is dependent only on the order of element of these matrices although the each element of GA and $\Lambda(h)^{-1}A$ is infinite sum as $n \rightarrow \infty$. The proof of Lemma B.2 can also be shown by Lemma A1 of Claeskens *et al.* (2009) [9]. Therefore, we only describe the outline here.

Proof of Lemma B.2: For any continuous and bounded function h , the matrix $G(h)$ is the band matrix from property of B -splines and hence $\|GA\|_\infty = O(n^a K^{-1})$ is obvious. Next, $\Lambda(h)^{-1}$ is the inverse of the band matrix. From Demko (1977) [14], there exists $c > 0$ and $r \in (0, 1)$ such that $| \{ \Lambda(h)^{-1} \}_{ij} | < cr^{|i-j|} K(1 + K(m, \tau)^{2m})^{-1}$. Thus, straightforward calculation yields that the infinite sum of each element of $\Lambda(h)^{-1}A$ is bounded by order of $\Lambda(h)^{-1}$ and the absolute of the maximum of element of A . \square

Proof of Theorem 3.1: Define $h(x) = H(x)^{-\gamma}$. We write $\tau \equiv \tau_I$ and hence $\tau \rightarrow 1$ and $(1 - \tau)n \rightarrow \infty$ as $n \rightarrow \infty$. By the fundamental property of B -spline basis, $s_0^{(m)}(\tau|x) = d^m s_0(\tau|x)/dx^m$ can be written as $s_0^{(m)}(\tau|x) = \mathbf{B}^{[p-m]}(x)^T D_m \mathbf{b}_0(\tau)$, where $\mathbf{B}^{[p-m]}(x)$ is the vector having element $\{ B_1^{[p-m]}(x), \dots, B_{K+p-m}^{[p-m]}(x) \}$ and $B_k^{[p-m]}(x)$'s are $(p - m)$ th degree B -spline bases. Therefore, the shrinkage bias can be expressed as

$$b_\lambda(\tau|x) = \frac{\lambda q(\tau)}{(1 - \tau)n} \mathbf{B}(x)^T \Lambda(H^{-\gamma})^{-1} D_m^T \int_a^b \mathbf{B}^{[p-m]}(x) s_0^{(m)}(\tau|x) dx.$$

From Conditions A–B, we have $s_0(\tau|x) = q_Y(\tau|x)(1 + o(1)) = h(x)q(\tau)(1 + o(1))$ and $s_0^{(m)}(\tau|x) = h^{(m)}(x)q(\tau)(1 + o(1))$ as $\tau \rightarrow 1$. Since h is bounded function, we get $\sup_{x \in [a,b]} |s_0^{(m)}(\tau|x)| = O(q(\tau))$. From the property of B -spline basis, we also obtain $\int_a^b B_k(x) ds = O(K^{-1})$. Thus, each element of $\int_a^b \mathbf{B}^{[p-m]}(x) s_0^{(m)}(\tau|x) dx$ has the order $O(K^{-1}q(\tau))$. Furthermore, the result of Cardot (2000) [5] provides $\|D_m\|_\infty = O(K^m)$. Therefore, Lemmas B.1, B.2 and the fact

that $K(m, \tau)^m / (1 + K(m, \tau)^{2m}) < 1/2$ yield that

$$\begin{aligned} b_\lambda(\tau|x) &= O\left(q(\tau) \frac{\lambda q(\tau) K^m}{(1-\tau)n} (1 + K(m, \tau)^{2m})^{-1}\right) \\ &= O\left(q(\tau) \left(\frac{\lambda q(\tau)}{(1-\tau)n}\right)^{1/2} \frac{K(m, \tau)^m}{1 + K(m, \tau)^{2m}}\right) \\ &\leq O\left(q(\tau) \left(\frac{\lambda q(\tau)}{(1-\tau)n}\right)^{1/2}\right). \end{aligned}$$

Next, we show the asymptotic order of $v(x|\tau)$. Since $\|G\|_\infty = O(K^{-1})$, from Lemmas B.1–B.2, we have

$$\begin{aligned} v(\tau|x) &= O\left(\frac{Kq(\tau)^2}{(1-\tau)n} \{1 + K(m, \tau)^{2m}\}^{-2}\right) \\ &= O\left(\frac{q(\tau)}{(1-\tau)n} \left(\frac{\lambda q(\tau)}{(1-\tau)n}\right)^{-1/2m} K(m, \tau) \{1 + K(m, \tau)^{2m}\}\right) \\ &\leq O\left(\frac{q(\tau)}{(1-\tau)n} \left(\frac{\lambda q(\tau)}{(1-\tau)n}\right)^{-1/2m}\right), \end{aligned}$$

which completes the proof. □

Proof of Theorem 3.2: We write $\tau \equiv \tau_I$ and hence $\tau \rightarrow 1$ and $(1 - \tau)n \rightarrow \infty$ as $n \rightarrow \infty$. Let $U_i = Y_i - \mathbf{B}(x)^T \mathbf{b}_0(\tau)$, $a_n = \sqrt{(1 - \tau)n}/q(\tau)$ and

$$\begin{aligned} Q_n(\boldsymbol{\delta}|\tau) &= \frac{a_n}{\sqrt{n(1-\tau)}} \sum_{i=1}^n \{\rho_\tau(U_i - \mathbf{B}(x_i)^T \boldsymbol{\delta}/a_n) - \rho_\tau(U_i)\} \\ &\quad + \frac{a_n \lambda}{2\sqrt{n(1-\tau)}} (\mathbf{b}_0(\tau) + \boldsymbol{\delta}/a_n)^T D_m^T R D_m (\mathbf{b}_0(\tau) + \boldsymbol{\delta}/a_n). \end{aligned}$$

Then the minimizer of Q_n is obtained as

$$\tilde{\boldsymbol{\delta}} = a_n(\tilde{\mathbf{b}}(\tau) - \mathbf{b}_0(\tau)).$$

Using Knight’s identity (Knight, 1998 [28]), we have

$$\rho_\tau(u - v) - \rho_\tau(u) = -v(\tau - I(u < 0)) + \int_0^v \{I(u \leq s) - I(u \leq 0)\} ds$$

and writing

$$Q_n(\boldsymbol{\delta}|\tau) = W_n(\tau)^T \boldsymbol{\delta} + G_n(\boldsymbol{\delta}|\tau) + \frac{a_n \lambda}{2\sqrt{(1-\tau)n}} (\mathbf{b}_0(\tau) + \boldsymbol{\delta}/a_n)^T D_m^T R D_m (\mathbf{b}_0(\tau) + \boldsymbol{\delta}/a_n),$$

where

$$\begin{aligned} W_n(\tau) &\equiv \frac{-1}{\sqrt{(1-\tau)n}} \sum_{i=1}^n (\tau - I(Y_i < \mathbf{B}(x_i)^T \mathbf{b}_0)) \mathbf{B}(x_i), \\ G_n(\boldsymbol{\delta}|\tau) &\equiv \sum_{i=1}^n \int_0^{\mathbf{B}(x_i)^T \boldsymbol{\delta}/a_n} I(U_i \leq s) - I(U_i \leq 0) ds. \end{aligned}$$

Since $\tau = P(Y < q_Y(\tau|x)|X = x) = P(Y < \mathbf{B}(x)^T \mathbf{b}_0(\tau)|X = x) + o(K^{-m})$ and $E[I(Y < \mathbf{B}(x)^T \mathbf{b}_0(\tau))] = P(Y < q_Y(\tau|x)|X = x) + o(K^{-m}(1 - \tau)^{-\gamma})$, we obtain $E[W_n(\tau)] = o(1)$. The variance of $W_n(\tau)$ can be evaluated as

$$V[W_n(\tau)^T \boldsymbol{\delta}] = \frac{\tau(1 - \tau)}{1 - \tau} \boldsymbol{\delta}^T \left(\frac{1}{n} Z^T Z \right) \boldsymbol{\delta} \xrightarrow{P} \boldsymbol{\delta}^T G \boldsymbol{\delta}$$

as $n \rightarrow \infty$ and $\tau \rightarrow 1$. Lyapunov’s condition for the central limit theorem and Cramèr-Wold device yield that $W_n(\tau)$ is asymptotically distributed as \mathbf{W} , which is the normal with mean $\mathbf{0}$ and variance G .

Next, we show that as $n \rightarrow \infty$ and $\tau \rightarrow 1$,

$$(B.1) \quad G_n(\boldsymbol{\delta}|\tau) \xrightarrow{P} \frac{1}{2} \gamma^{-1} \boldsymbol{\delta}^T G(H^{-\gamma}) \boldsymbol{\delta}.$$

Before that, we provide some differential results. Let $f_Y(y)$ and $f_Y(y|x)$ be the marginal density of Y and conditional density of Y given $X = x$, respectively. From A3, $f_Y(y|x) = f_Y(y|x)(1 + o(1)) = H(x)f_Y(y)(1 + o(1))$ as $y \rightarrow \infty$. In addition, since $1 = \partial F_V(q_V(\tau|x)|x)/\partial \tau = f_V(q_V(\tau|x)|x)\partial q_V(\tau|x)/\partial \tau$, we have $f_V(q_V(\tau|x)|x) = \{\partial q_V(\tau|x)/\partial \tau\}^{-1}$. Meanwhile, A4 and $q((1 - \tau)/H(x)) = \{(1 - \tau)/H(x)\}^{-\gamma} L(H(x)/(1 - \tau))(1 + o(1))$ yield that $\partial q((1 - \tau)/H(x))/\partial \tau = \gamma(1 - \tau)^{-\gamma-1} H(x)^\gamma L(H(x)/(1 - \tau))(1 + o(1))$. Consequently, as $\tau \rightarrow 1$,

$$f_Y(q_Y(\tau|x)|x) = \gamma^{-1} H(x)^{-\gamma} (1 - \tau)^{\gamma+1} L(H(x)/(1 - \tau))^{-1} (1 + o(1)).$$

Furthermore, by $L \in RV(0)$, $q(\tau)f_Y(q_Y(\tau|x)|x)/\{1 - \tau\} = \gamma^{-1} H(x)^{-\gamma} (1 + o(1))$.

We return to show (B.1). Since

$$G_n(\boldsymbol{\delta}|\tau) = \frac{1}{\sqrt{(1 - \tau)n}} \sum_{i=1}^n \left(\int_0^{\boldsymbol{\delta}^T \mathbf{B}(x_i)} I(U_i \leq s/a_n) - I(U_i \leq 0) ds \right),$$

we obtain

$$\begin{aligned} E[G_n(\boldsymbol{\delta}|\tau)] &= \frac{1}{\sqrt{(1 - \tau)n}} \sum_{i=1}^n E \left[\int_0^{\boldsymbol{\delta}^T \mathbf{B}(x_i)} I(U_i \leq s/a_n) - I(U_i \leq 0) ds \right] \\ &= \frac{1}{\sqrt{(1 - \tau)n}} \sum_{i=1}^n \left[\int_0^{\boldsymbol{\delta}^T \mathbf{B}(x_i)} F_Y(q_Y(\tau|x_i) + s/a_n|x_i) - F_Y(q_Y(\tau|x_i)|x_i) ds \right] \\ &= \sum_{i=1}^n \int_0^{\boldsymbol{\delta}^T \mathbf{B}(x_i)} \frac{f_Y(q_Y(\tau|x_i))}{a_n \sqrt{(1 - \tau)n}} s ds (1 + o(1)) \\ &= \frac{1}{n} \sum_{i=1}^n \int_0^{\boldsymbol{\delta}^T \mathbf{B}(x_i)} \frac{q(\tau) f_Y(q_Y(\tau|x_i))}{(1 - \tau)} s ds (1 + o(1)) \\ &= 2^{-1} \gamma^{-1} \boldsymbol{\delta}^T \left(\frac{1}{n} \sum_{i=1}^n H(x_i)^{-\gamma} \mathbf{B}(x_i) \mathbf{B}(x_i)^T \right) \boldsymbol{\delta}. \end{aligned}$$

From the simple but tedious calculation, $V[G_n(\boldsymbol{\delta}|\tau)] = o(1)$ can be evaluated. These results yield that

$$E[G_n(\boldsymbol{\delta}|\tau)] \xrightarrow{P} \frac{1}{2} \gamma^{-1} \boldsymbol{\delta}^T G(H^{-\gamma}) \boldsymbol{\delta}.$$

Thus, $Q_n(\boldsymbol{\delta}|\tau)$ is asymptotically equivalent to

$$Q_0(\boldsymbol{\delta}|\tau) = \left(W^T + \frac{\lambda}{a_n \sqrt{(1-\tau)n}} \mathbf{b}_0(\tau)^T D_m^T R D_m \right) \boldsymbol{\delta} + \frac{1}{2} \boldsymbol{\delta}^T \left(\gamma^{-1} G(H^{-\gamma}) + \frac{\lambda}{2a_n \sqrt{(1-\tau)n}} D_m^T R D_m \right) \boldsymbol{\delta}.$$

By the convexity lemma (see Pollard 1991 [38] and Knight 1998 [28]), the minimizer of Q_n and Q_0 are asymptotically equivalent and hence we have

$$\tilde{\boldsymbol{\delta}} = \left(\gamma^{-1} G(H^{-\gamma}) + \frac{\lambda}{2a_n \sqrt{(1-\tau)n}} D_m^T R D_m \right)^{-1} \left(W - \frac{\lambda}{\sqrt{a_n(1-\tau)n}} D_m^T R D_m \mathbf{b}_0(\tau) \right) + o_P(1).$$

Since $\tilde{q}_Y(\tau|x) - s_0(\tau|x) = \mathbf{B}(x)^T (\tilde{\mathbf{b}}(\tau) - \mathbf{b}_0(\tau))$, we obtain from $a_n = \sqrt{(1-\tau)n}/q(\tau)$ that

$$\begin{aligned} \frac{\sqrt{(1-\tau)n}}{q(\tau)} (\tilde{q}_Y(\tau|x) - s_0(\tau|x)) &= \\ &= \mathbf{B}(x)^T \left(\gamma^{-1} G(H^{-\gamma}) + \frac{\lambda q(\tau)}{2(1-\tau)n} D_m^T R D_m \right)^{-1} W \\ &\quad - \frac{\lambda}{\sqrt{(1-\tau)n}} \mathbf{B}(x)^T \left(\gamma^{-1} G(H^{-\gamma}) + \frac{\lambda q(\tau)}{2(1-\tau)n} D_m^T R D_m \right)^{-1} D_m^T R D_m \mathbf{b}_0(\tau) \\ \text{(B.2)} \quad &+ o_P(1). \end{aligned}$$

The second term of right hand side of (B.2) is the shrinkage bias. Consequently, as $n \rightarrow \infty$,

$$\frac{\sqrt{(1-\tau)n}}{q(\tau)} \frac{\tilde{q}_Y(\tau|x) - q_Y(\tau|x) - b_\lambda(\tau|x)}{\sqrt{\mathbf{B}(x)^T \Lambda(H^{-\gamma})^{-1} G \Lambda(H^{-\gamma})^{-1} \mathbf{B}(x)}} \xrightarrow{D} N(0, 1).$$

Finally, we obtain

$$\begin{aligned} E[\{\hat{q}_Y(\tau|x) - q_Y(\tau|x)\}^2] &= b_\lambda(\tau|x)^2 + v(\tau|x) \\ \text{(B.3)} \quad &= O\left(q(\tau)^2 \frac{\lambda q(\tau)}{(1-\tau)n} \right) + O\left(\frac{q(\tau)^2}{(1-\tau)n} \left(\frac{\lambda q(\tau)}{(1-\tau)n} \right)^{-1/2m} \right). \end{aligned}$$

We now derive the optimal rate of convergence of MISE of $\tilde{q}_Y(\tau|x)$. For the constant $C_1 > 0$, $C_2 > 0$, the solution of

$$C_1 q(\tau) \lambda - C_2 \left(\frac{q(\tau)}{(1-\tau)n} \right)^{-1/2m} \lambda^{-1/2m} = 0$$

is $\lambda = C q(\tau)^{-1} \{n(1-\tau)\}^{1/(2m+1)}$ for $C > 0$. By applying this λ in (B.3), we obtain

$$E \left[\left\{ \frac{\hat{q}_Y(\tau|x)}{q_Y(\tau|x)} - 1 \right\}^2 \right] = O(\{(1-\tau)n\}^{-2m/(2m+1)}),$$

which completes the proof. □

To improve the outlook, we now describe about the asymptotic bias and variance of $\hat{\gamma}(x)$ before prove Theorem 3.3. Define

$$(B.4) \quad b(k|x) = \frac{1}{k-1} \sum_{j=1}^{k-1} \frac{b_\lambda(\tau_j|x)}{q_Y(\tau_j|x)} - \frac{b_\lambda(\tau_k|x)}{q_Y(\tau_k|x)}, \quad \mathbf{v}(k|x) = \frac{1}{k-1} \sum_{j=1}^{k-1} \frac{\boldsymbol{\nu}(\tau_j|x)}{q_Y(\tau_j|x)} - \frac{\boldsymbol{\nu}(\tau_k|x)}{q_Y(\tau_k|x)},$$

where

$$\boldsymbol{\nu}(\tau|x) = \frac{q(\tau)}{\sqrt{(1-\tau)n}} G^{1/2} \Lambda(H^{-\gamma})^{-1} \mathbf{B}(x).$$

As the result, $b(k|x)$ and $\sqrt{\mathbf{v}(k|x)^T \mathbf{v}(k|x)}$ are the asymptotic bias and standard deviation of $\hat{\gamma}(x)$. We here get the asymptotic order of $b(k|x)$ and $v(k|x)$ from easy calculation.

Since $b_\lambda(\tau_j|x)/q_Y(\tau_j|x) = O(\{(1-\tau_j)n\}^{-m/(2m+1)})$ and each element of $\boldsymbol{\nu}(\tau_j|x)$ has $O(\{(1-\tau_j)n\}^{-m/(2m+1)})$, we have

$$b(k|x) = O\left(\frac{1}{k-1} \sum_{j=1}^{k-1} \{(1-\tau_j)n\}^{-\frac{m}{2m+1}} - \{(1-\tau_k)n\}^{-\frac{m}{2m+1}}\right)$$

and

$$\mathbf{v}(k|x) = O\left(\frac{1}{k-1} \sum_{j=1}^{k-1} \{(1-\tau_j)n\}^{-\frac{m}{2m+1}} - \{(1-\tau_k)n\}^{-\frac{m}{2m+1}}\right).$$

We then have from $[n^\eta]/k \rightarrow 0 (n, k \rightarrow \infty)$ that

$$\begin{aligned} \frac{1}{k} \sum_{j=1}^k \{(1-\tau_j)n\}^{-\frac{m}{2m+1}} &= \frac{1}{k} \sum_{j=1}^k \left\{ \frac{[n^\eta] + j}{n+1} n \right\}^{-\frac{m}{2m+1}} \\ &= k^{-\frac{m}{2m+1}} \frac{1}{k} \sum_{j=1}^k \left\{ \frac{[n^\eta] + j}{k+1} \right\}^{-\frac{m}{2m+1}} \\ &= k^{-\frac{m}{2m+1}} \int_0^1 u^{-\frac{m}{2m+1}} du (1 + o(1)) \\ &= \frac{2m+1}{m} k^{-\frac{m}{2m+1}} (1 + o(1)) \end{aligned}$$

and

$$\{(1-\tau_k)n\}^{-\frac{m}{2m+1}} = \left\{ \frac{[n^\eta] + k}{n+1} n \right\}^{-\frac{m}{2m+1}} = k^{-\frac{m}{2m+1}} (1 + o(1)).$$

This indicates that $b(k|x) = O(k^{-\frac{m}{2m+1}})$ and $\mathbf{v}(k|x) = O(k^{-\frac{m}{2m+1}})$.

Proof of Theorem 3.3: Theorem 3.1 indicates that $\tilde{q}_Y(\tau_k|x) - q_Y(\tau_k|x) = o_P(1)$ as $n \rightarrow \infty$. Therefore, the proposed estimator can be calculated as for $x \in \mathbb{R}$,

$$\begin{aligned} \hat{\gamma}(x) &= \frac{1}{k-1} \sum_{j=1}^{k-1} \log \frac{\hat{q}_Y(\tau_j|x)}{\hat{q}_Y(\tau_k|x)} \\ &= \frac{1}{k-1} \sum_{j=1}^{k-1} \log \frac{q_Y(\tau_j|x) \left\{ 1 + \frac{\hat{q}_Y(\tau_j|x) - q_Y(\tau_j|x)}{q_Y(\tau_j|x)} \right\}}{q_Y(\tau_k|x) \left\{ 1 + \frac{\hat{q}_Y(\tau_k|x) - q_Y(\tau_k|x)}{q_Y(\tau_k|x)} \right\}} \\ &= \frac{1}{k-1} \sum_{j=1}^{k-1} \log \frac{q_Y(\tau_j|x)}{q_Y(\tau_k|x)} + \frac{1}{k-1} \sum_{j=1}^{k-1} \log \frac{\left\{ 1 + \frac{\hat{q}_Y(\tau_j|x) - q_Y(\tau_j|x)}{q_Y(\tau_j|x)} \right\}}{\left\{ 1 + \frac{\hat{q}_Y(\tau_k|x) - q_Y(\tau_k|x)}{q_Y(\tau_k|x)} \right\}} \\ &= \frac{1}{k-1} \sum_{j=1}^{k-1} \log \frac{q_Y(\tau_j|x)}{q_Y(\tau_k|x)} + \frac{1}{k-1} \sum_{j=1}^{k-1} \log \left\{ 1 + \frac{\hat{q}_Y(\tau_j|x) - q_Y(\tau_j|x)}{q_Y(\tau_j|x)} \right\} \\ &\quad - \frac{1}{k-1} \sum_{j=1}^{k-1} \log \left\{ 1 + \frac{\hat{q}_Y(\tau_k|x) - q_Y(\tau_k|x)}{q_Y(\tau_k|x)} \right\} \\ &= \frac{1}{k-1} \sum_{j=1}^{k-1} \log \frac{q_Y(\tau_j|x)}{q_Y(\tau_k|x)} \\ &\quad + \frac{1}{k-1} \sum_{j=1}^{k-1} \left\{ \frac{\hat{q}_Y(\tau_j|x) - q_Y(\tau_j|x)}{q_Y(\tau_j|x)} - \frac{\hat{q}_Y(\tau_k|x) - q_Y(\tau_k|x)}{q_Y(\tau_k|x)} \right\} (1 + o_P(1)) \\ &\equiv D_{1n} + D_{2n}. \end{aligned}$$

We then note that D_{1n} is not random variable. Similar to the proof of Theorem 2.3 of Wang et al. (2012) [44], we have as $k \rightarrow \infty$,

$$D_{1n} = \gamma + O(k^{-1/2}) = \gamma + o(k^{-m/(2m+1)})$$

for $m \geq 1$. Next, we consider D_{2n} . Under the conditions for Theorem 3.3, using the result of Theorems 3.1–3.2 and the property of B -spline basis, we have

$$\frac{\tilde{q}_Y(\tau_j|x) - q_Y(\tau_j|x)}{q_Y(\tau_j|x)} = \frac{b_\lambda(\tau_j|x) + \boldsymbol{\nu}(\tau_j|x)^T \mathbf{W}}{q_Y(\tau_j|x)} (1 + o_P(1)),$$

where $\mathbf{W} \sim N_{K+p}(\mathbf{0}, I)$. Therefore, D_{2n} can be evaluated as

$$D_{2n} = \{b(k|x) + \mathbf{v}(k|x)^T \mathbf{W}\} (1 + o_P(1))$$

That is,

$$\hat{\gamma}(x) = \gamma + b(k|x) + \mathbf{v}(k|x)^T \mathbf{W} + o(k^{-m/(2m+1)}),$$

where $b(k|x) = O(k^{-m/(2m+1)})$ and $\mathbf{v}(k|x) = O(k^{-m/(2m+1)})$. Consequently, we get

$$\frac{\hat{\gamma}(x) - \gamma - b(k|x)}{\sqrt{\mathbf{v}(k|x)^T \mathbf{v}(k|x)}} \xrightarrow{D} N(0, 1)$$

and $E[\{\hat{\gamma}(x) - \gamma\}^2] = O(k^{-2m/(2m+1)})$. For the common index version $\hat{\gamma}^C$, similar to above, the straightforward calculation yields that

$$\hat{\gamma}^C = \frac{1}{n} \sum_{i=1}^n \hat{\gamma}(x_i) = \gamma + E[b(k|X)] + E[\mathbf{v}(k|X)]^T \mathbf{W} + o(k^{-m/(2m+1)}).$$

This completes the proof. □

Proof of Theorem 3.4: First, the second order condition for $U_Y(1/(1-\tau)|x) = q_Y(\tau|x)$ yields that

$$\frac{q_Y(\tau_I|x)}{q_Y(\tau_E|x)} = \left(\frac{U_Y(1/(1-\tau_E)|x)}{U_Y(1/(1-\tau_I)|x)} \right)^{-1} = \left(\frac{1-\tau_I}{1-\tau_E} \right)^{-\gamma} \{1 + o(k^{-m/(2m+1)})\}.$$

Furthermore, the result of Theorem 3.3 indicates that

$$\begin{aligned} \left(\frac{1-\tau_I}{1-\tau_E} \right)^{\hat{\gamma}(x)-\gamma} &= \exp \left[(\hat{\gamma}(x) - \gamma) \log \left(\frac{1-\tau_I}{1-\tau_E} \right) \right] \\ &= 1 + (\hat{\gamma}(x) - \gamma) \log \left(\frac{1-\tau_I}{1-\tau_E} \right) (1 + o_P(1)) \\ &= 1 + \{b(k|x) + \mathbf{v}(k|x)^T \mathbf{W}\} \log \left(\frac{1-\tau_I}{1-\tau_E} \right) (1 + o_P(1)). \end{aligned}$$

Meanwhile, we obtain

$$\frac{\hat{q}_Y(\tau_I|x)}{q_Y(\tau_I|x)} = 1 + \frac{b_\lambda(\tau_I|x)}{q_Y(\tau_I|x)} + \frac{\boldsymbol{\nu}(\tau_I|x)^T}{q_Y(\tau_I|x)} \mathbf{W} + o_P(k^{-m/(2m+1)}),$$

where \mathbf{W} is that given in the proof of Theorem 3.3. Using above, we have

$$\begin{aligned} \frac{\hat{q}_Y(\tau_E|x)}{q_Y(\tau_E|x)} &= \left(\frac{1-\tau_I}{1-\tau_E} \right)^{\hat{\gamma}(x)} \frac{\hat{q}_Y(\tau_I|x)}{q_Y(\tau_E|x)} \\ &= \left(\frac{1-\tau_I}{1-\tau_E} \right)^{\hat{\gamma}(x)} \frac{\hat{q}_Y(\tau_I|x)}{q_Y(\tau_I|x)} \frac{q_Y(\tau_I|x)}{q_Y(\tau_E|x)} \\ &= \left(\frac{1-\tau_I}{1-\tau_E} \right)^{\hat{\gamma}(x)} \frac{\hat{q}_Y(\tau_I|x)}{q_Y(\tau_I|x)} (1 + o(k^{-m/(2m+1)})) \\ &= \left\{ 1 + (b(k|x) + \mathbf{v}(k|x)^T \mathbf{W}) \log \left(\frac{1-\tau_I}{1-\tau_E} \right) (1 + o_P(1)) \right\} \\ &\quad \times \left[1 + \left\{ \frac{b_\lambda(\tau_I|x)}{q_Y(\tau_I|x)} + \frac{\boldsymbol{\nu}(\tau_I|x)^T}{q_Y(\tau_I|x)} \mathbf{W} \right\} (1 + o_P(1)) \right] (1 + o(k^{-m/(2m+1)})) \\ &= 1 + \left\{ \log \left(\frac{1-\tau_I}{1-\tau_E} \right) \mathbf{v}(k|x) + \frac{\boldsymbol{\nu}(\tau_I|x)}{q_Y(\tau_I|x)} \right\}^T \mathbf{W} \\ &\quad + b(k|x) \log \left(\frac{1-\tau_I}{1-\tau_E} \right) + \frac{b_\lambda(\tau_I|x)}{q_Y(\tau_I|x)} \\ &\quad + o \left(k^{-m/(2m+1)} \log \left(\frac{1-\tau_I}{1-\tau_E} \right) \right) + o_P(\{n(1-\tau_I)\}^{-m/(2m+1)}). \end{aligned}$$

Consequently, we obtain

$$\frac{\frac{\hat{q}_Y(\tau_E|x)}{q_Y(\tau_E|x)} - 1 - \text{bias}(\tau_E|x)}{s(\tau_E|x)} \xrightarrow{D} N(0, 1),$$

where

$$(B.5) \quad \text{bias}(\tau_E|x) = b(k|x) \log \left(\frac{1-\tau_I}{1-\tau_E} \right) + \frac{b_\lambda(\tau_I|x)}{q_Y(\tau_I|x)}$$

and

$$(B.6) \quad s(\tau_E|x) = \left\| \log \left(\frac{1-\tau_I}{1-\tau_E} \right) \mathbf{v}(k|x) + q_Y(\tau_I|x)^{-1} \boldsymbol{\nu}(\tau_I|x) \right\|.$$

Here, for a vector \mathbf{a} , $\|\mathbf{a}\|$ means the ℓ_2 -norm of \mathbf{a} . Furthermore, we get

$$E \left[\left\{ \frac{\hat{q}_Y(\tau_E|x)}{q_Y(\tau_E|x)} - 1 \right\}^2 \right] = O \left(\max \left\{ k^{-\frac{2m}{2m+1}} \log^2 \left(\frac{1 - \tau_I}{1 - \tau_E} \right), \{(1 - \tau_I)n\}^{-\frac{2m}{2m+1}} \right\} \right).$$

Similarly, for the common index estimator $\hat{q}_Y^C(\tau_E|x)$, we have

$$\begin{aligned} \frac{\hat{q}_Y^C(\tau_E|x)}{q_Y(\tau_E|x)} &= 1 + E[b(k|X)] \log \left(\frac{1 - \tau_I}{1 - \tau_E} \right) + \frac{b_\lambda(\tau_I|x)}{q_Y(\tau_I|x)} \\ &\quad + \left\{ E[\mathbf{v}(k|X)] \log \left(\frac{1 - \tau_I}{1 - \tau_E} \right) + \frac{\boldsymbol{\nu}(\tau_I|x)}{q_Y(\tau_I|x)} \right\}^T \mathbf{W} \\ &\quad + o_P \left(k^{-m/(2m+1)} \log \left(\frac{1 - \tau_I}{1 - \tau_E} \right) \right) + o_P(\{(1 - \tau_I)n\}^{-m/(2m+1)}). \end{aligned}$$

Accordingly,

$$\frac{\frac{\hat{q}_Y^C(\tau_E|x)}{q_Y(\tau_E|x)} - 1 - \text{bias}^C(x)}{s^C(\tau_E|x)} \xrightarrow{D} N(0, 1),$$

where

$$(B.7) \quad \text{bias}^C(\tau_E|x) = E[b(k|X)] \log \left(\frac{1 - \tau_I}{1 - \tau_E} \right) + \frac{b_\lambda(\tau_I|x)}{q_Y(\tau_I|x)}$$

and

$$(B.8) \quad s^C(\tau_E|x) = \left\| E[\mathbf{v}(k|X)] \log \left(\frac{1 - \tau_I}{1 - \tau_E} \right) + \frac{\boldsymbol{\nu}(\tau_I|x)}{q_Y(\tau_I|x)} \right\|.$$

Finally, we obtain the optimal rate of convergence of MISE of the common index estimator as

$$E \left[\left\{ \frac{\hat{q}_Y^C(\tau_E|x)}{q_Y(\tau_E|x)} - 1 \right\}^2 \right] = O \left(\max \left\{ k^{-\frac{2m}{2m+1}} \log^2 \left(\frac{1 - \tau_I}{1 - \tau_E} \right), \{(1 - \tau_I)n\}^{-\frac{2m}{2m+1}} \right\} \right). \quad \square$$

ACKNOWLEDGMENTS

The authors are grateful to Editor, Associate Editor and anonymous referees for their valuable comments and suggestions, which led to improvements of the paper. The research of the author was partially supported by JSPS-KAKENHI 26730019 and 18K18011.

REFERENCES

- [1] BARROW, D.L. and SMITH, P.W. (1978). Asymptotic properties of best $L_2[0, 1]$ approximation by spline with variable knots, *Quart. Appl. Math.*, **36**, 293–304.
- [2] BEIRLANT, J. and GOEGBEUR, Y. (2004). Local polynomial maximum likelihood estimation for Pareto-type distributions, *J. Multivar. Anal.*, **89**, 97–118.
- [3] BEIRLANT, J.; GOEGBEUR, Y.; SEGERS, J. and TEUGELS, J. (2004a). *Statistics of Extremes: Theory and Applications*, John Wiley & Sons.
- [4] BEIRLANT, J.; DE WET, T. and GOEGBEUR, Y. (2004b). Nonparametric estimation of extreme conditional quantiles, *J. Statist. Comput. Simul.*, **74**, 567–580.
- [5] CARDOT, H. (2000). Nonparametric estimation of smoothed principal components analysis of sampled noisy functions, *J. Nonparam. Statist.*, **12**, 503–538.
- [6] CHERNOZHUKOV, V. (2005). Extremal quantile regression, *Ann. Statist.*, **33**, 806–839.
- [7] CHERNOZHUKOV, V. and FERNÁNDEZ-VAL (2011). Inference for extremal conditional quantile models, with an application to market and birthweight risks, *The Review of Economic Studies*, **78**, 559–589.
- [8] CHENG, Y.; DE GOOIJER, J.G. and ZEROM, D. (2011). Efficient estimation of an additive quantile regression model, *Scandinavian Journal of Statistics*, **38**, 46–62.
- [9] CLAESKENS, G.; KRIVOBOKOVA, T. and OPSOMER, J.D. (2009). Asymptotic properties of penalized spline estimators, *Biometrika*, **96**, 529–544.
- [10] COLES, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*, Springer-Verlag.
- [11] DAOUIA, A.; GARDES, L. and GIRARD, S. (2013). On kernel smoothing for extremal quantile regression, *Bernoulli*, **19**, 2557–2589.
- [12] DAOUIA, A.; GARDES, L.; GIRARD, S. and LEKINA, A. (2011). Kernel estimators of extreme level curves, *Test*, **20**, 311–333.
- [13] DE HAAN, L. and FERREIRA, A. (2006). *Extreme Value Theory: An Introduction*, New York, Springer-Verlag.
- [14] DEMKO, S. (1977). Inverses of band matrices and local convergence of spline projections, *SIAM. J. Numer. Anal.*, **14**, 616–619.
- [15] DURRIEU, G.; GRAMA, I.; PHAM, Q.K. and TRICOT, J.M. (2015). Nonparametric adaptive estimation of conditional probabilities of rare events and extreme quantiles, *Extremes*, **18**, 437–478.
- [16] EILERS, P.H.C. and MARX, B.D. (1996). Flexible smoothing with B -splines and penalties, *Statist. Sci.*, **11**, 89–121.
- [17] EINMAHL, J.H.J.; DE HAAN, L. and ZHOU, C. (2016). Statistics of heteroscedastic extremes, *J. Roy. Statist. Soc. Ser. B*, **78**, 31–51.
- [18] EL METHNI, J.; GARDES, L. and GIRARD, S. (2014). Non-parametric estimation of extreme risk measures from conditional heavy-tailed distributions, *Scand. J. Statist.*, **41**, 988–1012.
- [19] GARDES, L. and GIRARD, S. (2010). Conditional extremes from heavy-tailed distributions: An application to the estimation of extreme rainfall return levels, *Extremes*, **13**, 177–204.
- [20] GARDES, L.; GIRARD, S. and LEKINA, A. (2010). Functional nonparametric estimation of conditional extreme quantiles, *Journal of Multivariate Analysis*, **101**, 419–433.
- [21] GIRARD, S. and LOUHICHI, S. (2015). On the strong consistency of the kernel estimator of extreme conditional quantiles, *Functional Statistics and Applications*, 59–77.

- [22] GOEGBEUR, Y.; GUILLOU, A. and OSMANN, M. (2017). A local moment type estimator for an extreme quantile in regression with random covariates, *Communications in Statistics – Theory and Methods*, **46**, 319–343.
- [23] HASTIE, T. and TIBSHIRANI, R. (1990). *Generalized Additive Models*, Chapman and Hall.
- [24] HE, F.; CHENG, Y. and TONG, T. (2016). Estimation of extreme conditional quantiles through an extrapolation of intermediate regression quantiles, *Statistics and Probability Letters*, **113**, 30–37.
- [25] HILL, B.M. (1975). A simple general approach to inference about the tail of a distribution, *Ann. Statist.*, **13**, 331–341.
- [26] HOROWITZ, J. and LEE, S. (2005). Nonparametric estimation of an additive quantile regression model, *J. Amer. Statist. Assoc.*, **100**, 1238–1249.
- [27] KIM, S.J.; KOH, K.; BOYD, S. and GORINEVSKY, D. (2009). ℓ_1 trend filtering, *SIAM Review*, **51**, 339–360.
- [28] KNIGHT, K. (1998). Limiting distributions for L_1 regression estimators under general conditions, *Ann. Statist.*, **26**, 755–770.
- [29] KOENKER, R. (2005). *Quantile Regression*, Cambridge University Press, Cambridge.
- [30] KOENKER, R. (2011). Additive models for quantile regression: Model selection and confidence band-aids, *Brazilian Journal of Probability and Statistics*, **25**, 239–262.
- [31] KOENKER, R. and BASSETT, G. (1978). Regression quantiles, *Econometrica*, **46**, 33–50.
- [32] KOENKER, R.; NG, P. and PORTNOY, S. (1994). Quantile smoothing splines, *Biometrika*, **81**, 673–680.
- [33] LEE, Y.K.; MAMMEN, E. and PARK, B.U. (2010). Backfitting and smooth backfitting for additive quantile models, *Ann. Statist.*, **38**, 2857–2883.
- [34] LIANG, X.; ZOU, T.; GUO, B.; LI, S.; ZHANG, H.; ZHANG, S.; HUANG, H. and CHEN, S.X. (2015). Assessing Beijing’s PM2.5 pollution: severity, weather impact, APEC and winter heating, *Proceedings of the Royal Society A*, **471**.
- [35] LU, Z. and YU, K. (2004). Local linear additive quantile regression, *Scand. J. Statist.*, **31**, 333–346.
- [36] NYCHKA, D.; GRAY, G.; HAALAND, P.; MARTIN, D. and O’CONNELL, M. (1995). A nonparametric regression approach to syringe grading for quality improvement, *J. Amer. Statist. Assoc.*, **90**, 1171–1178.
- [37] O’SULLIVAN, F. (1986). A statistical perspective on ill-posed inverse problems (with discussion), *Statist. Sci.*, **1**, 505–27.
- [38] POLLARD, D. (1991). Asymptotics for least absolute deviation regression estimators, *Econometric Theory*, **7**, 186–199.
- [39] PRATESI, M.; RANALLI, M.G. and SALVATI, N. (2009). Nonparametric M-quantile regression using penalised splines, *J. Nonparam. Statist.*, **21**, 287–304.
- [40] REISS, P.T. and HUANG, L. (2012). Smoothness selection for penalized quantile regression splines, *The International Journal of Biostatistics*, **8.1**.
- [41] RUPPERT, D. (2002). Selecting the number of knots for penalized splines, *J. Comp. Graph. Statist.*, **11**, 735–757.
- [42] SADHANALA, V. and TIBSHIRANI, R.J. (2017). Additive models with trend filtering, *arXiv*, arXiv:1702.05037v2.
- [43] TIBSHIRANI, R.J. (2014). Adaptive piecewise polynomial estimation via trend filtering, *Ann. Statist.*, **42**, 285–323.
- [44] WANG, H.J.; LI, D. and HE, X. (2012). Estimation of high dimensional conditional quantiles for heavy-tailed distributions, *J. Amer. Statist. Assoc.*, **107**, 1453–1464.

- [45] WEISSMAN, I. (1978). Estimation of parameters and large quantiles based on the k largest observations, *J. Amer. Statist. Assoc.*, **73**, 812–815.
- [46] YOSHIDA, T. (2013). Asymptotics for penalized spline estimators in quantile regression, *Communications in Statistics – Theory and Methods*, DOI:10.1080/03610926.2013.765477.
- [47] YUAN, M. (2006). GACV for quantile smoothing splines, *Computational Statistics & Data Analysis*, **50**, 813–829.
- [48] ZHOU, S.; SHEN, X. and WOLFE, D.A. (1998). Local asymptotics for regression splines and confidence regions, *Ann. Statist.*, **26**, 1760–1782.