Subsampling inference for nonparametric extremal conditional quantiles

Daisuke Kurisu*[1] and Taisuke Otsu†[2]

[1]University of Tokyo
[2]London School of Economics

October 4, 2023

Abstract

This paper proposes a subsampling inference method for extreme conditional quantiles based on a self-normalized version of a local estimator for conditional quantiles, such as the local linear quantile regression estimator. The proposed method circumvents difficulty of estimating nuisance parameters in the limiting distribution of the local estimator. A simulation study and empirical example illustrate usefulness of our subsampling inference to investigate extremal phenomena.

Keywords: quantile regression; subsampling; extreme value theory

1 Introduction

Since the seminal work of Koenker and Bassett (1978), quantile regression has been widely applied in empirical analysis. In contrast to (mean) regression analysis for conditional means of response variables given covariates, the quantile regression technique allows us to investigate conditional quantile functions for different quantiles including tail areas to study various extremal phenomena.

For linear quantile regression models, Chernozhukov (2005) developed the asymptotic theory for Koenker and Bassett's (1978) quantile regression estimator under the extremal order quantile

*Center for Spatial Information Science, The University of Tokyo, 5-1-5, Kashiwanoha, Kashiwa-shi, Chiba 277-8568, Japan. Email: daisukekurisu@csis.u-tokyo.ac.jp

†Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK. Email: t.otsu@lse.ac.uk.

asymptotics, where the quantile level converges to zero or one at the same rate as the sample size, n, by extending the extreme value theory (see, e.g., Resnick (1987) for a review). Furthermore, Chernozhukov and Fernández-Val (2011) proposed feasible inference methods for the extremal quantile regression parameters by using self-normalized statistics combined with analytical or subsampling critical values. Their inference methods are practical and much more accurate in extreme tails than the conventional inference methods based on the fixed quantile asymptotics. One major limitation of these studies on the extremal quantile regression model is that the quantile regression function must be parametrically specified.[1] Chaudhuri (1991) proposed the local polynomial quantile regression approach to estimate nonparametrically the conditional quantile function, and investigated its asymptotic properties under the conventional fixed quantile asymptotics, which is, however, inaccurate for conducting inference for the tails. The purpose of this paper is to fill this gap by developing a practical inference method for nonparametric conditional quantiles in extreme tails.

In particular, we extend the extremal order quantile asymptotics by Chernozhukov (2005) and Chernozhukov and Fernández-Val (2011) to a nonparametric setup, and consider the situation where the quantile converges to zero or one at the same rate as $n\delta_n^d$ with the sample size n, number of covariates d, and localization or bandwidth parameter δ_n for a local estimator, such as the local linear quantile regression estimator. Then we propose a subsampling inference method based on a self-normalized counterpart of the local estimator for nonparametric extremal quantiles. Our subsampling inference avoids estimation of nuisance parameters in the limiting distribution of the local estimator under the extremal quantile asymptotics. In contrast to the conventional fixed quantile asymptotics based on central limit theorems, our extremal order quantile asymptotic analysis is built upon point process theory (see, e.g., Resnick, 1987, and Embrechts, Klüppelberg and Mikosch, 1997). See also Zhang (2018) for inference on quantile treatment effects under the extremal order quantile asymptotics. The main theorem of this paper, validity of our subsampling method, covers general local estimators for conditional quantiles. In the online supplement, we verify high level conditions of this theorem by a specific example, the local linear quantile regression estimator.

We emphasize that the main focus of this paper is on inference (i.e., confidence intervals and hypothesis testing) for extreme conditional quantiles. For point estimation, we consider the

[1] In an insightful paper, Phillips (2015) characterized probabilities of quantile crossings that imply misspecification of linear quantile regression models in the context of predictive regressions. In particular, when the slope coefficient varies with the quantile levels and the regressor obeys a unit root process, the linear quantile predictive regression is inevitably misspecified with high probability. It should be noted that this misspecification problem in the population cannot be resolved by finite sample modifications of the quantile regression estimator, such as the rearrangement method in Chernozhukov, Fernández-Val and Galichon (2010).

extrapolation approach as in Daouia, Gardes and Girard (2013) is particularly suitable since it allows to use more observations from less extreme quantiles (see, also Wang, Li and He, 2012, and He, Cheng and Tong, 2016). Intuitively our point estimator uses less observations than the extrapolation approach, and in this paper the point estimator is treated merely as a centering object to conduct subsampling inference. We regard our point process approach as a complementary inference method to the extrapolation approach as in Daouia, Gardes and Girard (2013).

This paper is organized as follows. In Section 2, we present our main result, validity of subsampling inference based on the self-normalized counterpart of the local estimator for extremal conditional quantiles. In Section 3, we conduct a simulation study, and Section 4 presents an empirical illustration of our method. The proof of the main theorem is contained in the Appendix. In Section 5, we describe additional results presented in the online supplement, where we verify the high level conditions of the main theorem by a specific example, the local linear quantile regression estimator, and discuss two extensions of our subsampling inference for varying extreme value index models and varying coefficient models. Finally, Section 6 concludes.

2 Subsampling inference

Let $\{Y_i, X_i\}_{i=1}^n$ be a sample of size n from $(Y, X) \in \mathbb{R} \times \mathbb{R}^d$, and $F_Y(\cdot|\cdot)$ be the conditional distribution function of $Y|X = \cdot$. The focus of this paper is to conduct inference on the extremal (lower) quantiles $\theta_{\alpha_n}(c) = \inf\{q : F_Y(q|c) \geq \alpha_n\}$ with $\alpha_n \to 0$ as $n \to \infty$ for given $c \in \mathbb{R}^d$. The case of upper quantiles with $\alpha_n \to 1$ is investigated in the same manner.

For the linear regression quantiles (say, $\theta_{\alpha_n}(x) = x'\gamma_{\alpha_n}$), Chernozhukov and Fernández-Val (2011) considered the case of $n\alpha_n \to \tilde{k} > 0$ and proposed analytical and subsampling inference methods based on the self-normalized object

$$T_n = \frac{\sqrt{n\alpha_n}(\hat{\gamma}_{\alpha_n} - \gamma_{\alpha_n})}{\bar{X}'(\hat{\gamma}_{m\alpha_n} - \hat{\gamma}_{\alpha_n})},$$

for some $m > 1$, where $\hat{\gamma}_{\alpha_n}$ is the linear quantile regression estimator and $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$. As they argue, although the scaling constant of $(\hat{\gamma}_{\alpha_n} - \gamma_{\alpha_n})$ in the numerator is generally impossible to estimate without strong parametric assumptions, the above normalized object converges to a limiting distribution that only depends on the extreme value index of the error distribution, which allows to consistently estimate the quantiles of $c'T_n$ by analytical or subsampling methods to conduct inference on the conditional quantile $\theta_{\alpha_n}(c) = c'\gamma_{\alpha_n}$.

This paper extends the above inference approach by Chernozhukov and Fernández-Val (2011) to the situation where the researcher does not know the functional form of $\theta_{\alpha_n}(x)$. In particular, based on some local estimator $\hat{\theta}_{\alpha_n}(c)$ for $\theta_{\alpha_n}(c)$ with a localization or bandwidth parameter δ_n to select or weight the observations around $x = c$, we consider its self-normalized counterpart:

$$\Theta_n = \frac{\hat{\theta}_{\alpha_n}(c) - \theta_{\alpha_n}(c)}{\hat{\theta}_{m\alpha_n}(c) - \hat{\theta}_{\alpha_n}(c)}, \tag{2.1}$$

for some $m > 1$.

Examples of the estimator $\hat{\theta}_{\alpha_n}(c)$ include the local constant, linear, or polynomial quantile regression estimators, and the inverse of the kernel or local polynomial estimator for the conditional distribution function of $Y|X = c$ using the bandwidth δ_n. In the online supplement, we focus on the local linear quantile regression estimator as a specific example of $\hat{\theta}_{\alpha_n}(c)$ and verify high level conditions for our main theorem on validity of subsampling inference.

Chaudhuri (1991) studied asymptotic properties of the local quantile regression estimator when the quantile is fixed. Chernozhukov (1998) investigated asymptotic properties of the local quantile regression estimator under the extreme order quantile asymptotics, $n\delta_n^d\alpha_n \to 0$ as $n \to \infty$. Alternatively, motivated by Chernozhukov and Fernández-Val (2011), this paper considers the extremal order quantile asymptotics in the sense that

$$\alpha_n \to 0, \qquad n\delta_n^d\alpha_n \to k \in (0, \infty) \quad \text{as } n \to \infty. \tag{2.2}$$

In order to establish validity of subsampling inference based the self-normalized object Θ_n, a major requirement is to guarantee that

$$\Theta_n \xrightarrow{d} \Theta_\infty \text{ with a continuous limit law.} \tag{2.3}$$

Our main theorem below imposes this requirement as a high level condition (Assumption (i)). However, the limiting distributions of Θ_n or even the local quantile estimator $\hat{\theta}_{\alpha_n}(c)$ are open questions in the literature (even though the focus of this paper is not on point estimation). In Section 1 of the online supplement, we derive the limiting distribution of Θ_n for a specific example, where $\hat{\theta}_{\alpha_n}(c)$ is the local linear quantile regression estimator. In this section, we directly assume (2.3) and propose a subsampling method to estimate consistently quantiles of Θ_∞, which can be used to conduct inference on $\theta_{\alpha_n}(c)$.[2]

[2]It is known that the conventional bootstrap does not work due to the nonstandard behavior of extremal quantile regression estimators (see, e.g., Bickel and Freedman (1981, Section 6) for a proof in the classical non-regression case). In particular, the empirical bootstrap fails in our framework, which can be deduced from a general

Let q_t denote the t-th quantile of Θ_∞. The subsampling approximation for the distribution of Θ_n is obtained as follows.

(Step 1) Consider all subsets of the data $\{W_i = (Y_i, X_i)\}$ of size b. If $\{W_i\}$ is a time series, consider $B_n = n - b + 1$ subsets of size b of the form $\{W_i, W_{i+1}, \ldots, W_{i+b-1}\}$.

(Step 2) For the j-th subsample, compute a subsample analogue of Θ_n, that is

$$\hat{\Theta}_b^{(j)} = \frac{\hat{\theta}_{\alpha_b}^{(j)}(c) - \hat{\theta}_{\alpha_b}(c)}{\hat{\theta}_{m\alpha_b}^{(j)}(c) - \hat{\theta}_{\alpha_b}^{(j)}(c)}, \tag{2.4}$$

for $j = 1, \ldots, B_n$, where $\hat{\theta}_{\alpha_b}(c)$ is the α_b-th conditional quantile estimator computed using the full sample, and $\hat{\theta}_{\alpha_b}^{(j)}(c)$ is the α_b-th conditional quantile estimator computed using the j-th subsample and bandwidth $\delta_b = (k/b\alpha_b)^{1/d}$ with $k = n\delta_n^d\alpha_n$. We take α_b such that $\alpha_b/\alpha_n \to \infty$ as $n \to \infty$ (i.e., α_b satisfies the intermediate order quantile asymptotics (Ichimura, Otsu and Altonji, 2019)).

(Step2) Obtain $\hat{q}_t$ as the sample t-th quantile of $\{\hat{\Theta}_b^{(j)}\}_{j=1}^{B_n}$.

Let $\mathbb{B}$ denote some fixed closed ball around c. For any positive sequences $\{c_{1n}\}$ and $\{c_{2n}\}$, $c_{1n} \sim c_{2n}$ means $c_{1n}/c_{2n} \to 1$ as $n \to \infty$. The main result of this paper, the asymptotic validity of our subsampling inference, is obtained as follows.

Theorem 1. *Assume that:*

(i) *(2.2) and (2.3) hold true.*

(ii) *As $n \to \infty$, it holds $b \to \infty$, $b/n \to 0$, $\delta_n \to 0$, $\delta_b \to 0$, $\alpha_b \to 0$, and $\alpha_b/\alpha_n \to \infty$.*

(iii) *There exist a distribution function F_{U_*} with Pareto-type tails of extreme value index $\xi \neq 0$ and measurable function φ such that $F_{Y-\varphi(X)}(z|x) \sim \Gamma(x)F_{U_*}(z)$, as $z \downarrow F_{U_*}^{-1}(0)$, uniformly over $x \in \mathbb{B}$ for some positive continuous function $\Gamma(x)$. Furthermore, $\hat{\theta}_{\alpha_b}(c)$ based on $\hat{\theta}_{\alpha_n}(c)$ satisfies*

$$F_{U_*}^{-1}(1/b\delta_b^d)\{\hat{\theta}_{\alpha_b}(c) - \theta_{\alpha_b}(c)\} \xrightarrow{p} 0.$$

Then as $n \to \infty$,

$$\hat{q}_t \xrightarrow{p} q_t \quad \text{for } t \in (0,1).$$

theory on weak convergence of point processes and inconsistency of the conventional bootstrap for heavy-tailed data (see, Resnick, 2007, Section 6).

Assumption (i) is a high level condition on the normalized object Θ_n. See Section 1 in the online supplement for primitive conditions and derivation of the limiting distribution Θ_∞ for the case of the local linear quantile regression estimator. Assumption (ii) contains mild conditions for b (subsample size), (α_n, α_b) (quantiles), and (δ_n, δ_b) (bandwidths). Assumption (iii) is typically satisfied for the location-scale model $Y = \varphi(X) + \Gamma(x)^\xi U_*$. The error term U_* is in the minimum domain of attraction of the extreme value distribution with shape parameter ξ called the extreme value index. See Section 1.1 of the online supplement for a detail. The last condition is on the estimator $\hat{\theta}_{\alpha_b}(c)$ at the intermediate order quantile α_b, which is imposed to control the approximation error for Θ_b by $\hat{\Theta}_b^{(j)}$.

To implement our subsampling inference, we need to choose: (a) size of subsamples b, (b) constant m for normalization, (c) quantile α_b, and (d) bandwidths (δ_n, δ_b) to compute $\hat{\Theta}_b^{(j)}$. For (a), b may be chosen by applying the methods in Politis, Romano and Wolf (1999, Chapter 9) and Bertail *et al.* (2004). In practice, a smaller number B_n of randomly chosen subsets can be used, provided that $B_n \to \infty$ (see, Section 2.5 of Politis, Romano and Wolf, 1999). For (b)-(d), we suggest the following procedure.

1. Choose α_n based on researcher's interest.

2. Choose δ_n by some cross validation method adapted to local estimators for conditional quantiles (e.g., Takeuchi *et al.,* 2006).

3. Based on b, (1), and (2), set $k = n\delta_n^d\alpha_n$, $\alpha_b = n\alpha_n/b$, $\delta_b = (k/b\alpha_b)^{1/d} = \delta_n$, and $m = (d+1)/k + 1 + p$ for a spacing parameter $p > 0$.

For α_b, one may introduce a finite sample adjustment $\alpha_b = \min\{n\alpha_n/b, 0.2\}$ as in Chernozhukov and Fernández-Val (2011). The spacing parameter is set as $p = 0.1$ in our simulation study. Our preliminary simulation suggests that the results are similar for different values of p. Note that given the requirement $k = n\delta_n^d\alpha_n = b\delta_b^d\alpha_b$ in the construction of (2.4), once we choose b, α_n, and δ_n (and n) as in the above procedure, the bandwidth δ_b is determined as $\delta_b = \delta_n$. Although such a choice of δ_b may be suboptimal for estimating $\hat{\theta}_{\alpha_b}^{(j)}(c)$, it guarantees the validity of subsampling inference.

We note that our main theorem applies to general local quantile estimators for $\hat{\theta}_{\alpha_n}(c)$. For the numerical illustrations below, we employ the local linear quantile regression estimator

$$(\hat{\theta}_{\alpha_n}(c), \hat{\beta}_{\alpha_n}(c)) = \arg\min_{\theta,\beta} \sum_{i=1}^{n} K(\delta_n^{-1}(X_i - c))\rho_{\alpha_n}(Y_i - \theta - \delta_n^{-1}(X_i - c)'\beta), \tag{2.5}$$

where K is a kernel function, δ_n is the bandwidth, and $\rho_\alpha(v) = v(\alpha - \mathbb{I}\{v \leq 0\})$. In Section

1 of the online supplement, we verify that this estimator satisfies the assumptions of the main theorem under the primitive conditions below. See Section 1 of the online supplement for detailed discussions and verifications.

Proposition 1. *For the local linear estimator in (2.5), suppose Assumptions 1-3 below and Assumption (ii) in the main theorem hold. Then Assumptions (i) and (iii) in the main theorem are satisfied.*

Let $D_u f(c) = \partial f(c)/\partial c_u$ for $u = 1, \ldots, d$ and $f : \mathbb{R}^d \to \mathbb{R}$, and $\mathbb{B} \subset \mathbb{R}^d$ be some fixed closed ball around c.

Assumption 1.

(i) $\{Y_i, X_i\}_{i=1}^n$ *is a sample from* $(Y, X) \in \mathbb{R} \times \mathbb{R}^d$*. The random variable* X *has the density function* f_X *that is positive and continuous on* $\mathbb{B}$*.*

(ii) There exist a random variable U_* *with distribution function* F_{U_*} *and a measurable function* $\varphi : \mathbb{B} \to \mathbb{R}$ *such that the conditional distribution function* $F_U(z|x)$ *of* $U = Y - \varphi(X)$ *given* $X = x$ *satisfies that* $F_U(z|x)/F_{U_*}(z) \sim \Gamma(x)$*, as* $z \downarrow F_{U_*}^{-1}(0)$*, uniformly over* $x \in \mathbb{B}$ *for some positive continuous function* $\Gamma(x)$ *on* $\mathbb{B}$*. The quantile function* $F_{U_*}^{-1}$ *of* U_* *has end-points* $F_{U_*}^{-1}(0) = 0$ *or* $F_{U_*}^{-1}(0) = -\infty$*. The distribution function* $F_{U_*}(z)$ *exhibits Pareto-type tails with extreme value index* $\xi \in \mathbb{R}$*, i.e.,*

(1) as $z \downarrow F_{U_*}^{-1}(0) = 0$ *or* $-\infty$*,* $F_{U_*}(z + va(z)) \sim e^v F_{U_*}(z)$ *for all* $v \in \mathbb{R}$ *when* $\xi = 0$*,*

(2) as $z \downarrow F_{U_*}^{-1}(0) = -\infty$*,* $F_{U_*}(vz) \sim v^{-1/\xi} F_{U_*}(z)$ *for all* $v > 0$ *when* $\xi > 0$*,*

(3) as $z \downarrow F_{U_*}^{-1}(0) = 0$*,* $F_{U_*}(vz) \sim v^{-1/\xi} F_{U_*}(z)$ *for all* $v > 0$ *when* $\xi < 0$*,*

where $a(z) = \int_{F_{U_*}^{-1}(0)}^{z} F_{U_*}(v)dv/F_{U_*}(z)$ *for* $z > F_{U_*}^{-1}(0)$*.*

(iii) Let δ_n *be a sequence of positive constants with* $\delta_n \to 0$ *as* $n \to \infty$*. Assume that* $n\delta_n^d \alpha_n \to k \in (0, \infty)$ *and* $\mathfrak{a}_n \delta_n^{1+\gamma} \to 0$ *as* $n \to \infty$*, where* γ *is defined in Assumption 1 (iv) below, and*

(1) $\mathfrak{a}_n = 1/a(F_{U_*}^{-1}(1/n\delta_n^d))$ *when* $\xi = 0$*,*

(2) $\mathfrak{a}_n = -1/F_{U_*}^{-1}(1/n\delta_n^d)$ *when* $\xi > 0$*,*

(3) $\mathfrak{a}_n = 1/F_{U_*}^{-1}(1/n\delta_n^d)$ *when* $\xi < 0$*.*

Furthermore, we define $\mathfrak{b}_n = \begin{cases} F_{U_*}^{-1}(1/n\delta_n^d) & \text{for } \xi = 0 \\ 0 & \text{for } \xi \neq 0 \end{cases}$*.*

(iv) *For each $u = 1, \ldots, d$, $D_u\varphi(x)$ exists at each $x \in \mathbb{B}$, and there exist constants $C \in (0, \infty)$ and $\gamma \in (0, 1]$ such that $D_u\varphi(x)$ is γ-Hölder continuous on $\mathbb{B}$, i.e., at each $x \in \mathbb{B}$, $|D_u\varphi(x) - D_u\varphi(c)| \leq C\|x - c\|^{\gamma}$.*

(v) *For all n large enough, $D_u\theta_{\alpha_n}(x)$ exists and is continuous at each $x \in \mathbb{B}$ and $u = 1, \ldots, d$, and $\sup_{x\in\mathbb{B}_n} \mathfrak{a}_n|\theta_{\alpha_n}(x) - \theta_{\alpha_n}(c) - (x - c)'\partial\theta_{\alpha_n}(c)/\partial x| \to 0$ as $n \to \infty$.*

Assumption 2. *The sequence $\{U_i, X_i\}_{i=1}^n$ with $U_i = Y_i - \varphi(X_i)$ defined in Assumption 1 (ii) forms a stationary and strongly mixing process with a geometric mixing rate, that is, for some $C_1 > 0$,*

$$\sup_i \sup_{A\in\mathcal{A}_i, B\in\mathcal{B}_{i+m}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| \exp(C_1 m) \to 0 \quad \text{as } m \to \infty,$$

where $\mathcal{A}_i = \sigma(U_i, X_i, U_{i-1}, X_{i-1}, \cdots)$ and $\mathcal{B}_i = \sigma(U_i, X_i, U_{i+1}, X_{i+1}, \cdots)$. Moreover, the sequence satisfies a condition that curbs clustering of extreme events in the following sense: $\mathbb{P}(U_i \leq M, U_{i+m} \leq M|\mathcal{A}_i) \leq C_2\mathbb{P}(U_i \leq M|\mathcal{A}_i)^2$ for all $M \in [s, \bar{M}]$, uniformly for all $m \geq 1$ with some constants $C_2 > 0$ and $\bar{M} > s$.

Assumption 3.

(i) *Let $w = (w_1, \ldots, w_d)' \in \mathbb{R}^d$. The kernel function K is a bounded positive Lipschitz function with support $[-1, 1]^d$ and second order, that is*

$$\int_{\mathbb{R}^d} K(w)dw = 1, \quad \int_{\mathbb{R}^d} K(w)w_u dw = 0 \text{ for } u = 1, \ldots, d.$$

(ii) *$\int_{\mathbb{R}^d} K(w)\tilde{w}\tilde{w}'dw$ is positive definite, where $\tilde{w} = (1, w_1, \ldots, w_d)' \in \mathbb{R}^{d+1}$.*

3 Simulation

In this section, we present simulation results to evaluate the finite sample performance of the proposed subsampling method. We consider the following location-scale model:

$$Y_i = 0.5 \sin(X_i) + \sqrt{2.5 + 0.5X_i^2}U_{*,i}, \tag{3.1}$$

for $i = 1, \ldots, n$, where $\{X_i\}$ are i.i.d. uniform random variables on $[-1, 0]$, and $\{U_{*,i}\}$ are i.i.d. random variables following either (i) t distribution with 3 or 30 degree of freedom, or (ii) Weibull distribution with the shape parameter 3 or 30. Note that these two cases corresponds to (i) $\xi = 1/3$ or $1/30$ and (ii) $\xi = -1/3$ or $-1/30$, respectively. When $\xi = 1/30$ or $-1/30$, U_* has a light-tailed distribution.

We compute $\hat{\theta}_{\alpha_n}(c)$ at $c = -0.5$ by using the local linear quantile regression estimator in (2.5) with the biweight kernel $K(w) = \frac{15}{16}(1-w^2)^2\mathbb{I}\{|w| \leq 1\}$. To estimate the quantile q_t of Θ_∞ in (2.3) based on the subsampling method, we consider $B_n = n - b + 1$ subsets of size b of the form $\{(Y_i, X_i), (Y_{i+1}, X_{i+1}), \ldots, (Y_{i+b-1}, X_{i+b-1})\}$. To illustrate the proposed subsample based inference on $\theta_{\alpha_n}(c)$, we see the finite sample properties of the following $100(1-t)\%$ confidence intervals ($t \in (0, 1/2)$) for the model (3.1) with Student's t and Weibull noises:

$$C_{1-t}(\alpha_n) = [\hat{\theta}_{\alpha_n}(c) - \hat{q}_{1-t/2}\{\hat{\theta}_{m\alpha_n}(c) - \hat{\theta}_{\alpha_n}(c)\}, \hat{\theta}_{\alpha_n}(c) - \hat{q}_{t/2}\{\hat{\theta}_{m\alpha_n}(c) - \hat{\theta}_{\alpha_n}(c)\}].$$

Table 1 presents empirical coverage probabilities of 90% ($t = 0.1$) and 95% ($t = 0.05$) confidence intervals $C_{1-t}(\alpha_n)$. We consider two cases for the sample size $n \in \{2000, 5000\}$ and set $b = 200$ (for $n = 2000$) and $b = 500$ (for $n = 5000$). For each Monte Carlo replication, we select the bandwidth δ_n by using leave-one-out cross validation (LOOCV) as explained in Remark 6 of the online supplement. We set $k = n\delta_n\alpha_n$, $B_n = n-b+1$, $\alpha_b = n\alpha_n/b$, and $m = 2/k+1.1$. The number of Monte Carlo repetitions is 250. The numbers in the parentheses are means of bandwidths selected by using LOOCV. We find that the simulated coverage probabilities of confidence intervals $C_{1-t}(\alpha_n)$ have similar performance in every case and they are reasonably close to the nominal coverage probabilities.

		Model	$t(3)$		$t(30)$		Weibull$(3,1)$		Weibull$(30,1)$	
n	α_n	Nominal	0.90	0.95	0.90	0.95	0.90	0.95	0.90	0.95
2000	0.01		0.848	0.920	0.860	0.928	0.856	0.928	0.876	0.936
			(0.198)		(0.197)		(0.197)		(0.196)	
	0.005		0.856	0.928	0.852	0.924	0.872	0.932	0.864	0.932
			(0.223)		(0.221)		(0.223)		(0.222)	
5000	0.01		0.876	0.948	0.860	0.920	0.860	0.924	0.872	0.940
			(0.191)		(0.195)		(0.197)		(0.164)	
	0.005		0.864	0.940	0.868	0.932	0.884	0.948	0.852	0.936
			(0.218)		(0.215)		(0.219)		(0.182)	

Table 1: Empirical coverage probabilities of $C_{1-t}(\alpha_n)$ for $\theta_{\alpha_n}(c) = F_Y^{-1}(\alpha_n|c)$ at $c = -0.5$. We set $b = 200$ for $n = 2000$ and $b = 500$ for $n = 5000$. The numbers in the parentheses are means of bandwidths selected by using LOOCV.

Table 2 presents empirical coverage probabilities of 90% ($t = 0.1$) and 95% ($t = 0.05$) confidence intervals $C_{1-t}(\alpha_n)$ with $n = 2000$ and $b \in \{80, 120, 160, 200, 300, 400, 500\}$. We also use the biweight kernel and set $k = n\delta_n\alpha_n$, $B_n = n-b+1$, $\alpha_b = n\alpha_n/b$, and $m = 2/k+1.1$. To compute confidence intervals, we use LOOCV to select δ_n. The number of Monte Carlo repetitions is 250. We find that the empirical coverage probabilities are reasonably close to the nominal ones when $1/25 \leq b/n \leq 1/10$. This motivates us to use $b = [n/10]$ as a practical choice of subsample size, which is employed in real data analysis in the next section. Note that our choices of $b \in \{80, 120, 160, 200, 300, 400, 500\}$ correspond to $\alpha_b \in \{1/4, 1/6, 1/8, 1/10, 1/15, 1/20, 1/25\}$ when $\alpha_n = 0.01$, respectively. The empirical coverage probabilities are less sensitive even for somewhat larger values of α_b.

		Model	$t(3)$		$t(30)$		Weibull(3, 1)		Weibull(30, 1)	
b	α_n	Nominal	0.90	0.95	0.90	0.95	0.90	0.95	0.90	0.95
80	0.01		0.868	0.932	0.872	0.936	0.868	0.936	0.864	0.936
	0.005		0.864	0.928	0.876	0.940	0.876	0.932	0.852	0.936
120	0.01		0.848	0.920	0.852	0.920	0.844	0.924	0.864	0.928
	0.005		0.868	0.924	0.876	0.940	0.872	0.940	0.872	0.936
160	0.01		0.852	0.924	0.864	0.924	0.868	0.932	0.864	0.924
	0.005		0.844	0.920	0.860	0.928	0.864	0.924	0.868	0.940
200	0.01		0.848	0.920	0.860	0.928	0.856	0.928	0.876	0.936
	0.005		0.856	0.928	0.852	0.924	0.872	0.932	0.864	0.932
300	0.01		0.852	0.916	0.852	0.920	0.848	0.916	0.856	0.912
	0.005		0.844	0.908	0.848	0.908	0.852	0.912	0.848	0.908
400	0.01		0.836	0.896	0.812	0.848	0.844	0.904	0.812	0.872
	0.005		0.816	0.872	0.804	0.856	0.812	0.856	0.820	0.876
500	0.01		0.796	0.860	0.752	0.808	0.792	0.868	0.800	0.852
	0.005		0.780	0.852	0.728	0.784	0.780	0.822	0.792	0.840

Table 2: Empirical coverage probabilities of $C_{1-t}(\alpha_n)$ for $\theta_{\alpha_n}(c) = F_Y^{-1}(\alpha_n|c)$ at $c = -0.5$ with $n = 2000$ and $b \in \{80, 120, 160, 200, 300, 400, 500\}$.

3.1 Comparison with other methods

We compare finite sample properties of confidence intervals based on (i) our subsampling method, (ii) normal approximation, and (iii) the extrapolation approach developed in Daouia, Gardes and Girard (2013). When the quantile level α_n is considered as fixed (i.e. $\alpha_n = \alpha \in (0, 1)$), we can also apply normal approximation of $\hat{\theta}_\alpha(c)$ to construct confidence intervals. From Fan, Hu and Truong (1994, Theorem 3), we can construct $100(1-t)\%$ confidence intervals based on normal approximation of $\hat{\theta}_\alpha(c)$ for fixed $\alpha \in (0, 1)$ as follows:

$$C^N_{1-t}(\alpha) = \left[\hat{\theta}_\alpha(c) - z_{1-t/2}\sqrt{\frac{\hat{\tau}^2(c)}{n\delta_n}}, \hat{\theta}_\alpha(c) - z_{t/2}\sqrt{\frac{\hat{\tau}^2(c)}{n\delta_n}}\right],$$

where z_t is the t-th quantile of the standard normal distribution and $\hat{\tau}^2(c)$ is an estimator of the asymptotic variance of $\hat{\theta}_\alpha(c)$ given by

$$\tau^2(c) = \frac{\alpha(1-\alpha)\int K^2(w)dw}{f_X(c)g_Y^2(\theta_\alpha(c)|c)}.$$

Here, f_X is the density of X and $g_Y(\cdot|c)$ is the conditional density of Y given $X = c$. To estimate f_X, we use kernel smoothing with the Epanechnikov kernel and bandwidth selected by using LOOCV. For the estimation of $g_Y(\cdot|c)$, we use the method proposed in Bashtannyu and Hyndman (2001). We also compute $\hat{\theta}_\alpha(c)$ in the same way as our method and the bandwidth is selected by using LOOCV.

Furthermore, in our simulation study, we consider an infeasible version of Daouia, Gardes and Girard's (2013) extrapolation-based estimator in eq. (1.13) of the online supplement, where we set $\hat{\xi}(c) = \xi$ and $\hat{a}(c) = (\theta_{\alpha_n}(c) - \theta_{\tilde{\alpha}_n}(c))/K_\xi(\tilde{\alpha}_n/\alpha_n)$. In other words, the second term in eq. (1.13) of the online supplement does not involve any preliminary estimation as in Daouia, Gardes and Girard (2013). In this case, as in Daouia, Gardes and Girard (2013, Theorem 1), one can construct $100(1-t)\%$ confidence intervals of $\theta_{\alpha_n}(c)$ as follows:

$$C^E_{1-t}(\alpha_n) = \left[\hat{\theta}_{\tilde{\alpha}_n}(c) + B_n(c) - z_{1-t/2}\sqrt{\frac{\hat{v}^2(c)}{n\delta_n}}, \hat{\theta}_{\tilde{\alpha}_n}(c) + B_n(c) - z_{t/2}\sqrt{\frac{\hat{v}^2(c)}{n\delta_n}}\right],$$

where $B_n(c) = \theta_{\alpha_n}(c) - \theta_{\tilde{\alpha}_n}(c)$ and $\hat{v}^2(c)$ is an estimator of the asymptotic variance of $\hat{\theta}_{\tilde{\alpha}_n}(c)$ given by

$$v^2(c) = \frac{\tilde{\alpha}_n \int K^2(w)dw}{f_X(c) g_Y^2(\theta_{\tilde{\alpha}_n}(c)|c)}.$$

We set $\tilde{\alpha}_n = n\alpha_n/b$ (b is the subsample size used in the computation of $C_{1-t}(\alpha_n)$) and the bandwidth δ_n is selected by using LOOCV. For the estimation of f_X, we use kernel smoothing with the Epanechnikov kernel and bandwidth selected by using LOOCV. For the estimation of $g_Y(\cdot|c)$, we use the method proposed in Bashtannyu and Hyndman (2001).

Table 3 presents empirical coverage probabilities of 90% ($t = 0.1$) and 95% ($t = 0.05$) confidence intervals $C_{1-t}(\alpha_n)$, $C^N_{1-t}(\alpha_n)$, and $C^E_{1-t}(\alpha_n)$ with $n = 2000$ and $\alpha_n \in \{0.01, 0.005\}$. Although we do not report here, the results are similar for the case of $n = 5000$. To compute the confidence interval $C_{1-t}(\alpha_n)$, we use LOOCV to select δ_n and set $k = n\delta_n\alpha_n$, $b = n/10$, $B_n = n - b + 1$, $\alpha_b = n\alpha_n/b$, and $m = 2/k + 1.1$. We also use the local linear quantile regression estimator in (2.5) with the biweight kernel to compute $\hat{\theta}_{\alpha_n}(c)$ and $\hat{\theta}_{\tilde{\alpha}_n}(c)$. The number of Monte Carlo repetitions is 250. We find that the normal approximation confidence interval $C^N_{1-t}(\alpha_n)$ exhibits severe size distortions particularly for the $t(3)$ and $t(30)$ distributions. This result clearly endorses usefulness of the asymptotic approximation based on the extreme value theory for tail areas as advocated in this paper. We also find that the confidence interval $C^E_{1-t}(\alpha_n)$ based on the infeasible estimator $\hat{\theta}_{\tilde{\alpha}_n}(c) + K_\xi(\tilde{\alpha}_n/\alpha_n)\hat{a}(c)$ (where the second term does not involve preliminary estimation) also exhibits size distortions. This result indicates that the normal approximation for $\hat{\theta}_{\tilde{\alpha}_n}(c)$ under the intermediate quantile asymptotics may not work well for inference in tail areas even after the bias correction by the second term $K_\xi(\tilde{\alpha}_n/\alpha_n)\hat{a}(c)$.

		Model	$t(3)$		$t(30)$		Weibull$(3,1)$		Weibull$(30,1)$	
n	α_n	Nominal	0.90	0.95	0.90	0.95	0.90	0.95	0.90	0.95
2000	0.01	$C_{1-t}(\alpha_n)$	0.848	0.920	0.860	0.928	0.856	0.928	0.876	0.936
		$C_{1-t}^N(\alpha_n)$	0.044	0.052	0.204	0.240	0.716	0.776	0.656	0.748
		$C_{1-t}^E(\alpha_n)$	0.436	0.528	0.512	0.608	0.784	0.860	0.808	0.864
	0.005	$C_{1-t}(\alpha_n)$	0.856	0.928	0.852	0.924	0.872	0.932	0.864	0.932
		$C_{1-t}^N(\alpha_n)$	0.032	0.040	0.128	0.152	0.628	0.684	0.668	0.708
		$C_{1-t}^E(\alpha_n)$	0.256	0.304	0.408	0.496	0.768	0.816	0.740	0.812

Table 3: Empirical coverage probabilities of $C_{1-t}(\alpha_n)$, $C_{1-t}^N(\alpha_n)$, and $C_{1-t}^E(\alpha_n)$ for $\theta_{\alpha_n}(c) = F_Y^{-1}(\alpha_n|c)$ at $c = -0.5$.

4 Real data illustration

We apply our methodology to conduct inference on the extremal quantiles of the GBP-AUD exchange rate $\{R_i\}_{i=1}^{n+1}$ observed every 3 hours from March 22nd, 2006 to August 30th, 2008 ($n = 5053$) provided by the Dukascopy Bank. Before we apply our method, we transform $\{R_i\}$ as $Y_i = 100 \times (\log(R_{i+1}) - \log(R_i))$ for $i = 1, \ldots, 5053$, and consider an AR(1)-type structure $(Y_i, X_i) = (Y_i, Y_{i-1})$. Figure 4.1 depicts the transformed GBP-AUD exchange rate $\{Y_i\}_{i=1}^n$. We also use the local linear quantile regression estimator in (2.5) with the biweight kernel and set $\alpha_n \in \{0.01, 0.005\}$, $\delta_n = 0.103$ (for $\alpha_n = 0.01$), 0.115 (for $\alpha_n = 0.005$) which are selected by the rule-of-thumb proposed in Yu and Jones (1998), $k = n\delta_n\alpha_n$, $b = [n/10] = 505$, $B_n = n - b + 1$, $\alpha_b = n\alpha_n/b$, and $m = 2/k + 1.1$. Table 4 presents estimated values of the extremal conditional quantiles $\hat{\theta}_{\alpha_n}(c)$ at $c = 0$ and confidence intervals $C_{1-t}(\alpha_n)$. We can see that our confidence intervals for the extreme quantiles $\theta_{0.01}(c)$ and $\theta_{0.005}(c)$ are reasonably informative based on the plot in Figure 4.1.

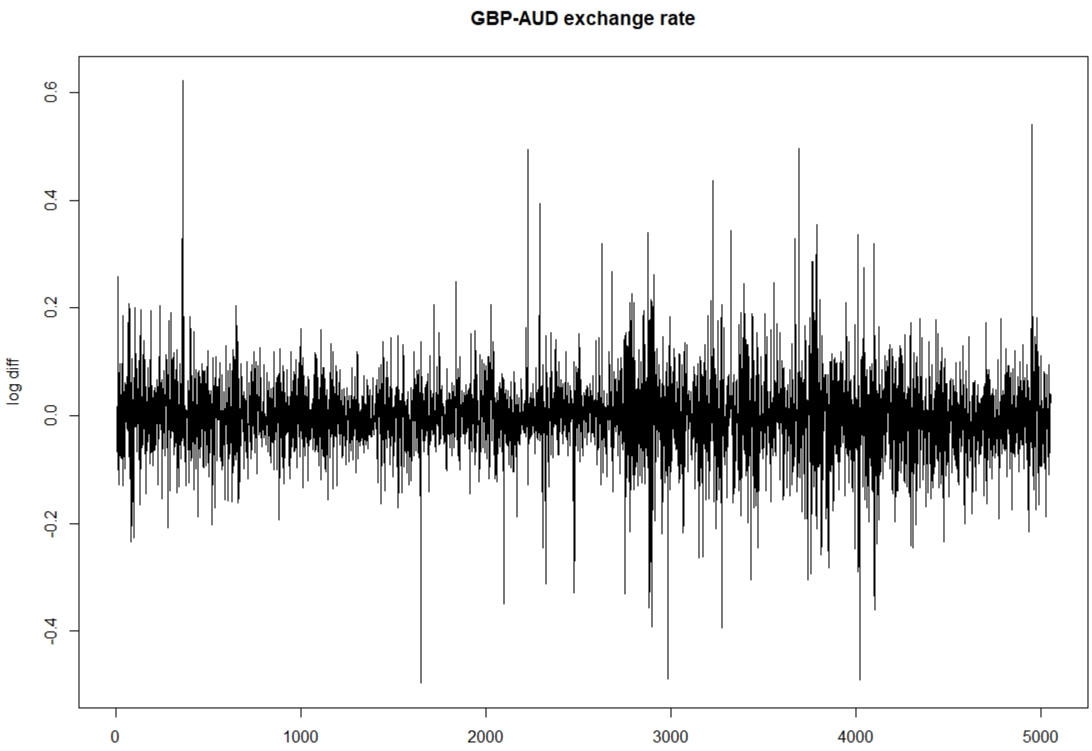

Figure 4.1: Plots of the transformed GBP-AUD exchange rate $\{Y_i\}_{i=1}^{5053}$.

α_n	$\hat{\theta}_{\alpha_n}(c)$	$C_{0.90}(\alpha_n)$	$C_{0.95}(\alpha_n)$
0.01	-0.185	[-0.265, -0.139]	[-0.274, -0.137]
0.005	-0.243	[-0.347, -0.146]	[-0.363, -0.138]

Table 4: Estimated values of $\theta_{\alpha_n}(c)$ at $c = 0$ and confidence intervals $C_{1-t}(\alpha_n)$.

5 Additional results in online supplement

A major technical challenge is to establish the weak convergence of the normalized object Θ_n in (2.3) under the extremal order quantile asymptotics (2.2). This is a key condition (Assumption (i)) to establish the validity of our subsampling inference in the main theorem. Furthermore, although the focus of this paper is inference (i.e., hypothesis testing and interval estimation) on $\theta_{\alpha_n}(c)$, it is of independent interest what is the convergence rate and limiting distribution of the local estimator $\hat{\theta}_{\alpha_n}(c)$ under the extremal order quantile asymptotics. For point estimation of $\theta_{\alpha_n}(c)$, we consider the extrapolation approach as in Daouia, Gardes and Girard (2013) is particularly suitable since it allows to use more observations from less extreme quantiles.

In Section 1 of the online supplement, we focus on the local linear quantile regression estimator as a specific example of $\hat{\theta}_{\alpha_n}(c)$, provide primitive conditions to satisfy the assumptions in our main theorem, and derive the limiting distributions of the point estimator $\hat{\theta}_{\alpha_n}(c)$ and its self-normalized counterpart Θ_n. In particular, we extend the extremal order quantile asymptotics by Chernozhukov (2005) and Chernozhukov and Fernández-Val (2011) to a nonparametric setup, and consider the situation where the quantile converges to zero or one at the same rate as $n\delta_n^d$ as in (2.2). In contrast to the conventional fixed quantile asymptotics based on central limit theorems, our extremal order quantile asymptotic analysis is built upon point process theory.

Although theoretical developments are similar, there are at least two important directions to extend our subsampling inference method. In Section 2 of the online supplement, we present extensions of our main result to (a) the case where the extreme value index ξ of the error term distribution may vary with covariates (Section 2.1), and (b) varying coefficient extremal quantile regression models $Y = X'\beta(Z) + \gamma(X, Z)V_*$ for an unknown function $\beta(\cdot)$ of covariates Z, and error term V_* in the domain of minimum attraction (Section 2.2).

These additional results are also new in the literature, and we also provide detailed comments on the assumptions and theorems in the online supplement.

6 Conclusion

This paper studies inference for nonparametric extreme conditional quantiles. We propose a subsampling inference method based on a self-normalized counterpart of a nonparametric conditional quantile estimator. An attractive feature of our method is that it avoids estimation of nuisance parameters in the limiting distribution of the quantile estimator under the extremal quantile asymptotics. We establish asymptotic validity of the proposed method, and illustrate its finite sample performance by a simulation study and empirical example. It is interesting to extend the proposed method to other econometric problems associated with quantiles, such as the quantile treatment effect analysis and quantile instrumental variable regression.

A Proof of Theorem

Let $\tilde{A}_b^{(j)} = \frac{1}{\hat{\theta}_{m\alpha_b}^{(j)}(c) - \hat{\theta}_{\alpha_b}^{(j)}(c)}$, $\Theta_b^{(j)} = \tilde{A}_b^{(j)}(\hat{\theta}_{\alpha_b}^{(j)}(c) - \theta_{\alpha_b}(c))$, and $A_b = -\text{sgn}(\xi) \cdot 1/F_{U_*}^{-1}(1/(b\delta_b^d))$. Define

$$
\begin{aligned}
\hat{G}_n(x) &= \frac{1}{B_n}\sum_{j=1}^{B_n} \mathbb{I}\{\hat{\Theta}_b^{(j)} \le x\} = \frac{1}{B_n}\sum_{j=1}^{B_n} \mathbb{I}\{\Theta_b^{(j)} + \tilde{A}_b^{(j)}(\theta_{\alpha_b}(c) - \hat{\theta}_{\alpha_b}(c)) \le x\}, \\
\tilde{G}_n(x;\Delta) &= \frac{1}{B_n}\sum_{j=1}^{B_n} \mathbb{I}\{\Theta_b^{(j)} + (\tilde{A}_b^{(j)}/A_b)\Delta \le x\}.
\end{aligned}
$$

Then

$$
\mathbb{I}\{\Theta_b^{(j)} \le x - \tilde{A}_b^{(j)} w_n/A_b\} \le \mathbb{I}\{\hat{\Theta}_b^{(j)} \le x\} \le \mathbb{I}\{\Theta_b^{(j)} \le x + \tilde{A}_b^{(j)} w_n/A_b\},
$$

for all $j = 1, \dots, B_n$, where $w_n = |A_b(\theta_{\alpha_b}(c) - \hat{\theta}_{\alpha_b}(c))|$.

Since $w_n = o_p(1)$ by Assumption (iii), there exists a sequence $\epsilon_n \downarrow 0$ as $n \to \infty$ such that the following event occurs with probability approaching one:

$$
\begin{aligned}
\Omega_n = \Big\{ & \mathbb{I}\{\Theta_b^{(j)} \le x - \tilde{A}_b^{(j)}\epsilon_n/A_b\} \le \mathbb{I}\{\Theta_b^{(j)} \le x - \tilde{A}_b^{(j)} w_n/A_b\} \le \mathbb{I}\{\hat{\Theta}_b^{(j)} \le x\} \\
& \le \mathbb{I}\{\Theta_b^{(j)} \le x + \tilde{A}_b^{(j)} w_n/A_b\} \le \mathbb{I}\{\Theta_b^{(j)} \le x + \tilde{A}_b^{(j)}\epsilon_n/A_b\} \text{ for all } j = 1, \dots, B_n \Big\}.
\end{aligned}
$$

On Ω_n, it holds

$$
\tilde{G}_n(x;\epsilon_n) \le \hat{G}_n(x) \le \tilde{G}_n(x;-\epsilon_n). \tag{A.1}
$$

We next show that at the continuity points of $G(x) = \mathbb{P}(\Theta_\infty \le x)$, it holds $\tilde{G}_n(x;\pm\epsilon_n) \xrightarrow{p} G(x)$. Non-replacement sampling implies

$$
\mathbb{E}[\tilde{G}_n(x;\epsilon_n)] = \mathbb{P}(\Theta_b - \tilde{A}_b^{(j)}\epsilon_n/A_b \le x),
$$

and at the continuity points of $G(x)$,

$$
\lim_{n\to\infty} \mathbb{E}[\tilde{G}_b^{(j)}(x;\epsilon_n)] = \lim_{b\to\infty} \mathbb{P}(\Theta_b - \tilde{A}_b^{(j)}\epsilon_n/A_b \le x) = G(x),
$$

since $\Theta_b \xrightarrow{d} \Theta_\infty$ (by Assumption (i)) and $\tilde{A}_b^{(j)}\epsilon_n/A_b = O_p(1)\cdot\epsilon_n = o_p(1)$. Since $\tilde{G}_n(x;\epsilon_n)$ is a U-statistic of degree b, the law of large numbers for U-statistics in Politis, Romano and Wolf (1999) implies $\text{Var}(\tilde{G}_n(x;\epsilon_n)) = o(1)$. This shows that $\tilde{G}_n(x;\epsilon_n) \xrightarrow{p} G(x)$. Likewise, we obtain $\tilde{G}_n(x;-\epsilon_n) \xrightarrow{p} G(x)$.

Finally, since $\mathbb{P}(\Omega_n) \to 1$, (A.1) yields $\hat{G}_n(x) \xrightarrow{p} G(x)$ for each $x \in \mathbb{R}$. Since convergence of distribution functions at continuity points implies convergence of quantile functions at the continuity points, the continuous mapping theorem yields $\hat{q}_t = \hat{G}_n^{-1}(t) \xrightarrow{p} G^{-1}(t) = q_t$, provided $G^{-1}(t)$ is a continuity point of $G(x)$.

References

[1] Bashtannyu, D. M. and R. J. Hyndman (2001) Bandwidth selection for kernel conditional density estimation, *Computational Statistics & Data Analysis*, 36, 279-298.

[2] Bertail, P., Haefke, C., Politis, D. N. and H. White (2004) Subsampling the distribution of diverging statistics with applications to finance, *Journal of Econometrics*, 120, 295-326.

[3] Bickel, P. J. and D. A. Freedman (1981) Some asymptotic theory for the bootstrap, *Annals of Statistics*, 9, 1196-1217.

[4] Chaudhuri, P. (1991) Nonparametric estimates of regression quantiles and their local Bahadur representation, *Annals of Statistics*, 19, 760-777.

[5] Chernozhukov, V. (1998) Nonparametric extreme regression quantiles, Working paper.

[6] Chernozhukov, V. (2005) Extremal quantile regression, *Annals of Statistics*, 33, 806-839.

[7] Chernozhukov, V. and I. Fernández-Val (2011) Inference for extremal conditional quantile models, with an application to market and birthweight risks, *Review of Economic Studies*, 78, 559-589.

[8] Chernozhukov, V., Fernández-Val, I. and A. Galichon (2010) Quantile and probability curves without crossing, *Econometrica*, 78, 1093-1125.

[9] Daouia, A., Gardes, L. and S. Girard (2013) On kernel smoothing for extremal quantile regression, *Bernoulli*, 19, 2557-2589.

[10] Embrechts, P., Klüppelberg, C. and T. Mikosch (1997) *Modeling Extremal Events for Insurance and Finance*, Springer, Berlin.

[11] Fan, J., Hu, T.-C. and Y. K.Truong (1994) Robust non-parametric function estimation, *Scandinavian Journal of Statistics*, 21, 433-446.

[12] He, F., Cheng, Y. and T. Tong (2016) Estimation of extreme conditional quantiles through an extrapolation of intermediate regression quantiles, *Statistics & Probability Letters*, 113, 30-37.

[13] Ichimura, H., Otsu, T. and J. Altonji (2019) Nonparametric intermediate order regression quantiles. Working paper.

[14] Koenker, R. and G. Bassett (1978) Regression quantiles, *Econometrica*, 46, 33-50.

[15] Phillips, P. C. B. (2015) Halbert White Jr. memorial JFEC lecture: Pitfalls and possibilities in predictive regression, *Journal of Financial Econometrics*, 13, 521-555.

[16] Politis, D. N., Romano, J. P. and M. Wolf (1999) *Subsampling*, Springer, New York.

[17] Resnick, S. I. (1987) *Extreme Values, Regular Variation, and Point Process*, Springer, New York.

[18] Resnick, S. I. (2007) *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, Springer, New York.

[19] Takeuchi, I., Le, Q. V., Sears, T. and A. J. Smola (2006) Nonparametric quantile regression, *Journal of Machine Learning Research*, 7, 1231–1264.

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

[21] Yu, K. and M. C. Jones (1998) Local linear quantile regression, *Journal of the American Statistical Association*, 93, 228-237.

[22] Zhang, Y. (2018) Extremal quantile treatment effects, *Annals of Statistics*, 46, 3707-3740.

Online supplement for "Subsampling inference for nonparametric extremal conditional quantiles"

Daisuke Kurisu*[1] and Taisuke Otsu†[2]

[1]University of Tokyo
[2]London School of Economics

October 4, 2023

Abstract

In this supplement, we focus on the local linear quantile regression estimator as an example of $\hat{\theta}_{\alpha_n}(c)$. Section 1 derives the limiting distribution of the normalized object Θ_n and shows that $w_n := \mathrm{sgn}(\xi)[1/F_{U^*}^{-1}(1/(b\delta_b^d))](\theta_{\alpha_b}(c) - \hat{\theta}_b(c)) = o_p(1)$, which verifies Assumptions (i) and (iii) in the main paper. In Section 1.1, we present primitive conditions for the derivation and provide detailed comments. Section 1.2 presents the limiting distribution of Θ_n and the local linear quantile regression estimator. In Section 1.3, we show that $w_n = o_p(1)$ for the local linear quantile regression estimator. Section 2 discusses extensions of our results for the case where the extreme value index varies with covariates (Section 2.1) and varying coefficient models (Section 2.2). All proofs of this supplement are contained in the Appendix.

Throughout this supplement, we focus on the local linear quantile regression estimator:

$$(\hat{\theta}_{\alpha_n}(c), \hat{\beta}_{\alpha_n}(c)) = \arg\min_{\theta,\beta} \sum_{i=1}^{n} K(\delta_n^{-1}(X_i - c))\rho_{\alpha_n}(Y_i - \theta - \delta_n^{-1}(X_i - c)'\beta),$$

where K is a kernel function (see Assumption 3 in this supplement for details) and $\rho_\alpha(v) = v(\alpha - \mathrm{I}\{v \leq 0\})$.

*Center for Spatial Information Science, The University of Tokyo, 5-1-5, Kashiwanoha, Kashiwa-shi, Chiba 277-8568, Japan. Email: daisukekurisu@csis.u-tokyo.ac.jp

†Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK. Email: t.otsu@lse.ac.uk.

Notation.

Hereafter we use the following notation. For random variables $(Y, X) \in \mathbb{R}\times\mathbb{R}^d$, let $F_Y(y|c)$ be the conditional distribution function of Y given $X = c = (c_1, \ldots, c_d)' \in \mathbb{R}^d$ and $\theta_\alpha(c) = \inf_{y\in\mathbb{R}}\{y : F_Y(y|c) > \alpha\}$ be the α-th conditional quantile function at c. For any function $f : \mathbb{R}^d \to \mathbb{R}$, let $D_u f(c) = \partial f(c)/\partial c_u$ for $u = 1, \ldots, d$. $\mathbb{B} \subset \mathbb{R}^d$ denotes some fixed closed ball around c, and $\mathbb{B}_n = \prod_{j=1}^{d}[c_j - \delta_n, c_j + \delta_n]$. For any positive sequences a_n and b_n, we write $a_n \lesssim b_n$ if there is a constant $C > 0$ independent of n such that $a_n \leq C b_n$ for all n, and $a_n \sim b_n$ means that $a_n/b_n \to 1$ as $n \to \infty$. For any $a \in \mathbb{R}$, define $\mathrm{sgn}(a) = 1$ if $a > 0$ and $\mathrm{sgn}(a) = -1$ if $a \leq 0$. We use the notations $\stackrel{d}{\to}$ and $\stackrel{p}{\to}$ as convergence in distribution and in probability, respectively. For $a \in \mathbb{R}$, let $[a]$ be the integer part of a. Let $\|\cdot\|$ be the Euclidean norm.

1 Proof of Proposition 1 in the main text (verification of Assumptions (i) and (iii))

In this section, we prepare some auxiliary results (Section 1.1) and verify (2.3) in the main paper (Section 1.2) and $w_n := \mathrm{sgn}(\xi)[1/F_{U^*}^{-1}(1/(b\delta_b^d))](\theta_{\alpha_b}(c) - \hat{\theta}_b(c)) = o_p(1)$ (Section 1.3).

1.1 Assumptions and comments

We impose the following conditions.

Assumption 1.

(i) $\{Y_i, X_i\}_{i=1}^n$ *is a sample from* $(Y, X) \in \mathbb{R}\times\mathbb{R}^d$. *The random variable* X *has the density function* f_X *that is positive and continuous on* $\mathbb{B}$.

(ii) There exist a random variable U_* *with distribution function* F_{U_*} *and a measurable function* $\varphi : \mathbb{B} \to \mathbb{R}$ *such that the conditional distribution function* $F_U(z|x)$ *of* $U = Y - \varphi(X)$ *given* $X = x$ *satisfies that* $F_U(z|x)/F_{U_*}(z) \sim \Gamma(x)$, *as* $z \downarrow F_{U_*}^{-1}(0)$, *uniformly over* $x \in \mathbb{B}$ *for some positive continuous function* $\Gamma(x)$ *on* $\mathbb{B}$. *The quantile function* $F_{U_*}^{-1}$ *of* U_* *has end-points* $F_{U_*}^{-1}(0) = 0$ *or* $F_{U_*}^{-1}(0) = -\infty$. *The distribution function* $F_{U_*}(z)$ *exhibits Pareto-type tails with extreme value index* $\xi \in \mathbb{R}$, *i.e.,*

(1) as $z \downarrow F_{U_*}^{-1}(0) = 0$ *or* $-\infty$, $F_{U_*}(z + va(z)) \sim e^v F_{U_*}(z)$ *for all* $v \in \mathbb{R}$ *when* $\xi = 0$,

(2) as $z \downarrow F_{U_*}^{-1}(0) = -\infty$, $F_{U_*}(vz) \sim v^{-1/\xi}F_{U_*}(z)$ *for all* $v > 0$ *when* $\xi > 0$,

(3) as $z \downarrow F_{U_*}^{-1}(0) = 0$, $F_{U_*}(vz) \sim v^{-1/\xi}F_{U_*}(z)$ *for all* $v > 0$ *when* $\xi < 0$,

where $a(z) = \int_{F_{U_*}^{-1}(0)}^{z} F_{U_*}(v)dv/F_{U_*}(z)$ *for* $z > F_{U_*}^{-1}(0)$.

(iii) Let δ_n *be a sequence of positive constants with* $\delta_n \to 0$ *as* $n \to \infty$. *Assume that* $n\delta_n^d \alpha_n \to k \in (0,\infty)$ *and* $\mathfrak{a}_n \delta_n^{1+\gamma} \to 0$ *as* $n \to \infty$, *where* γ *is defined in Assumption 1 (iv) below, and*

(1) $\mathfrak{a}_n = 1/a(F_{U_*}^{-1}(1/n\delta_n^d))$ *when* $\xi = 0$,

(2) $\mathfrak{a}_n = -1/F_{U_*}^{-1}(1/n\delta_n^d)$ *when* $\xi > 0$,

(3) $\mathfrak{a}_n = 1/F_{U_*}^{-1}(1/n\delta_n^d)$ *when* $\xi < 0$.

Furthermore, we define $\mathfrak{b}_n = \begin{cases} F_{U_*}^{-1}(1/n\delta_n^d) & \text{for } \xi = 0 \\ 0 & \text{for } \xi \neq 0 \end{cases}$.

(iv) For each $u = 1, \ldots, d$, $D_u\varphi(x)$ *exists at each* $x \in \mathbb{B}$, *and there exist constants* $C \in (0,\infty)$ *and* $\gamma \in (0,1]$ *such that* $D_u\varphi(x)$ *is* γ*-Hölder continuous on* $\mathbb{B}$, *i.e., at each* $x \in \mathbb{B}$, $|D_u\varphi(x) - D_u\varphi(c)| \leq C\|x-c\|^\gamma$.

(v) For all n large enough, $D_u\theta_{\alpha_n}(x)$ *exists and is continuous at each* $x \in \mathbb{B}$ *and* $u = 1, \ldots, d$, *and* $\sup_{x\in\mathbb{B}_n} \mathfrak{a}_n|\theta_{\alpha_n}(x) - \theta_{\alpha_n}(c) - (x-c)'\partial\theta_{\alpha_n}(c)/\partial x| \to 0$ *as* $n \to \infty$.

Assumption 1 (ii) is a key condition, which involves auxiliary objects $\varphi(x)$, $\Gamma(x)$, and U_*. Intuitively, the function $\varphi(x)$ can be considered as a general notion of the 'boundary' of the conditional distribution $Y|X = x$, and the conditional distribution of the error term $U|X = x$ is approximated by a multiplicative form $\Gamma(x)F_{U_*}(\cdot)$ so that U_* and $\Gamma(x)$ may be interpreted as an idiosyncratic shock and skedastic function in heteroskedastic errors, respectively. Under this assumption, the quantile function $\theta_{\alpha_n}(x)$ can be approximately decomposed into the function $\varphi(x)$ and remaining term, i.e.,

$$\theta_{\alpha_n}(x) \approx \varphi(x) + F_{U_*}^{-1}(\alpha_n/\Gamma(x)) \approx \varphi(x) + \Gamma(x)^\xi F_{U_*}^{-1}(\alpha_n). \tag{1.1}$$

Based on this decomposition and Taylor expansions of $\theta_{\alpha_n}(x)$ and $\varphi(x)$ by using Assumption 1 (iv)-(v), in Theorem 1 below, we derive the limiting distribution of

$$(\hat{\theta}_{\alpha_n}(c), \hat{\beta}_{\alpha_n}(c)) = \arg\min_{\theta,\beta} \sum_{i=1}^{n} K(\delta_n^{-1}(X_i - c))\rho_{\alpha_n}(Y_i - \theta - \delta_n^{-1}(X_i - c)'\beta), \tag{1.2}$$

centered around the expansion coefficients for $\varphi(x)$ and the second term in (1.1)[1].

[1]Since we employ the conventional local linear quantile regression estimator, the quantile crossing problem also occurs to our estimator (i.e., $\hat{\theta}_\alpha(c)$ may not be increasing in α in finite samples). In our context with $\alpha = \alpha_n \to 0$, the sequence of the estimators $\{\hat{\theta}_{\alpha_n}(c)\}$ may not be decreasing even though this feature does not affect our asymptotic analysis. One way to circumvent the quantile crossing is to rearrange the quantile regression estimator as in Chernozhukov, Fernández-Val and Galichon (2010) (i.e, estimate $\theta_\alpha(c)$ by the α-th quantile of

More precisely, the auxiliary function φ is considered as (1) the boundary function for the case when Y has a finite lower end-point, $F_Y^{-1}(0|x) = \theta_0(x) = \lim_{\alpha_n \downarrow 0} \theta_{\alpha_n}(x) = \lim_{\alpha_n \downarrow 0}(\varphi(x) + F_U^{-1}(\alpha_n|x)) = \varphi(x) > -\infty$, or (2) the location function of Y given X for the unbounded support case, $F_Y^{-1}(0|x) = \theta_0(x) = \lim_{\alpha_n \downarrow 0}(\varphi(x) + F_U^{-1}(\alpha_n|x)) = -\infty$, and the condition on φ restricts the shape of the conditional distribution $F_U(\cdot|x)$ of $U = Y - \varphi(X)$ given $X = x$. In particular, we assume that $F_U(\cdot|x)$ is approximated by a multiplicative form $\Gamma(x)F_{U_*}(\cdot)$, and that F_{U_*} has a tail of type 1, 2, and 3 when $\xi = 0$, $\xi > 0$, and $\xi < 0$, respectively (see Resnick, 1987, for details on these types). Assumption 1 (ii) also requires that for any $x_1, x_2 \in \mathbb{B}$, $z \mapsto F_U(z|x_1)$ and $z \mapsto F_U(z|x_2)$ are tail equivalent up to a constant. This condition is motivated by the closure of the domain of minimum attraction under tail equivalence (see Proposition 1.19 of Resnick, 1987). Typically, Assumption 1 (ii) is satisfied for location-scale models. See also the comments after Assumption 2 below. The absolute value of ξ measures heavy-tailedness of the distribution. Distributions with $\xi = 0$ include normal and exponential. Distributions with $\xi > 0$ include stable, Pareto, and Student's t. Distributions with $\xi < 0$ include uniform, exponential, and Weibull.

Assumption 1 (iii) is concerned with the canonical normalization of $\hat{\theta}_{\alpha_n}(c) - \theta_{\alpha_n}(c)$. For example, for Case (1), if U_* follows the Laplace distribution $F_{U_*}(z) = 2^{-1}e^{-\lambda|z|}\mathbb{I}\{z < 0\} + (1 - 2^{-1}e^{-\lambda|z|})\mathbb{I}\{z \geq 0\}$ for some $\lambda > 0$, then we have $a(z) = \lambda^{-1}$ and $F_{U_*}^{-1}(\tau) = \lambda^{-1}\log(2\tau)$ (as $\tau \downarrow 0$) implying $\mathfrak{a}_n = \lambda^{-1}$ and $\mathfrak{b}_n = \lambda^{-1}(\log 2 - \log(n\delta_n^d))$. For Case (2), if U_* follows the Pareto distribution $F_{U_*}(z) = (1 + |z|)^{-1/\xi}\mathbb{I}\{z \leq 0\}$ for some $\xi > 0$, then we have $F_{U_*}^{-1}(\tau) = 1 - \tau^{-\xi}$ implying $\mathfrak{a}_n = ((n\delta_n^d)^\xi - 1)^{-1}$. For Case (3), if U_* follows the Weibull distribution $F_{U_*}(z) = (1 - e^{-(z/\beta)^{-1/\xi}})\mathbb{I}\{z \geq 0\}$ for some $\xi < 0$ and $\beta > 0$, then we have $F_{U_*}^{-1}(\tau) = \beta\{-\log(1-\tau)\}^{-\xi} \sim \beta\tau^{-\xi}$ (as $\tau \downarrow 0$) implying $\mathfrak{a}_n \sim \beta^{-1}(n\delta_n^d)^{-\xi}$.

Assumption 1 (iv) and (v) are concerned with smoothness of the conditional quantile function θ_{α_n} and auxiliary function φ. A Taylor expansion of φ around $x = c$ yields

$$\begin{aligned} \varphi(x) &= \varphi(c) + (x-c)'\frac{\partial \varphi(c)}{\partial x} + R_\varphi(x, \delta_n), \\ \theta_{\alpha_n}(x) &= \theta_{\alpha_n}(c) + (x-c)'\frac{\partial \theta_{\alpha_n}(c)}{\partial x} + R(x, \delta_n), \end{aligned} \tag{1.3}$$

$\hat{\theta}_U(c)$ with $U \sim$ Uniform$[0, 1]$) even though its theoretical analysis for the extremal case is beyond the scope of this paper. Furthermore, it should be noted that such a rearrangement method for the linear quantile regression is a finite sample modification and does not resolve misspecification problems of the linear model in the population. Indeed Phillips (2005) characterized probabilities of quantile crossings implying misspecification of linear quantile regression models in the context of predictive regressions, and argued that the linear quantile predictive regression may be inevitably misspecified with high probability. Although formal analysis for predictive regressions is beyond the scope, Phillips' (2005) analysis also endorses importance of nonparametric methods to investigate conditional quantiles.

and Assumption 1 (iv) guarantees

$$\sup_{x\in\mathbb{B}_n} |R_\varphi(x,\delta_n)| = O(\delta_n^{1+\gamma}). \tag{1.4}$$

Assumption 1 (v) says that the remainder of the Taylor expansion of $\theta_{\alpha_n}(x)$ around $x = c$ should be smaller order than $\mathfrak{a}_n^{-1}$, i.e.,

$$\sup_{x\in\mathbb{B}_n} \mathfrak{a}_n|R(x,\delta_n)| = o(1). \tag{1.5}$$

As shown below, this condition is satisfied for location-scale models under certain smoothness conditions.[2]

We also assume the following dependence structure on $\{U_i, X_i\}$.

Assumption 2. *The sequence $\{W_i\}_{i=1}^n$ with $W_i = (U_i, X_i)$ and $U_i = Y_i - \varphi(X_i)$ defined in Assumption 1 (ii) forms a stationary and strongly mixing process with a geometric mixing rate, that is, for some $C_1 > 0$,*

$$\sup_i \sup_{A\in\mathcal{A}_i, B\in\mathcal{B}_{i+m}} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)| \exp(C_1 m) \to 0 \quad \text{as } m\to\infty,$$

where $\mathcal{A}_i = \sigma(W_i, W_{i-1}, \cdots)$ and $\mathcal{B}_i = \sigma(W_i, W_{i+1}, \cdots)$. Moreover, the sequence satisfies a condition that curbs clustering of extreme events in the following sense: $\mathbb{P}(U_i \le M, U_{i+m} \le M|\mathcal{A}_i) \le C_2\mathbb{P}(U_i \le M|\mathcal{A}_i)^2$ for all $M \in [s, \bar{M}]$, uniformly for all $m \ge 1$ with some constants $C_2 > 0$ and $\bar{M} > s$.

Assumption 2 includes the case that the sequence of variables $\{U_i, X_i\}_{i=1}^n$, or equivalently $\{Y_i, X_i\}_{i=1}^n$, is a sequence of i.i.d. random variables. The mixing assumption on $\{U_i, X_i\}_{i=1}^n$ is equivalent to the one on $\{Y_i, X_i\}_{i=1}^n$. The non-clustering assumption is used to apply Meyer's (1973) theorem in (A.4) to establish the weak convergence of the point process (1.7) defined below.

We now provide an example satisfying our assumptions. Let $\{U_{*,i}\}$ be a sequence of i.i.d. random variables and $\{Y_i, X_i\}$ are observations. Letting $\xi \ne 0$, consider the following location-scale model

$$Y_i = \varphi(X_i) + \gamma(X_i)U_{*,i}. \tag{1.6}$$

[2]When $F_Y(y|x)$ does not have a finite end-point, the remainder $R(x,\delta_n)$ may diverge as $\alpha_n \downarrow 0$ in some cases. However, in such cases, the definition of $\mathfrak{a}_n = -1/F_{U_*}^{-1}(1/n\delta_n^d)$ implies $\mathfrak{a}_n \downarrow 0$ as $n\to\infty$ so that the condition in (1.5) can be still satisfied. On the other hand, the condition in (1.5) becomes more stringent for δ_n when $\mathfrak{a}_n \to \infty$. For example, when U_* follows the Weibull distribution $F_{U_*}(z) = (1 - e^{-(z/\beta)^{-1/\xi}})\mathbb{I}\{z \ge 0\}$ for some $\xi < 0$ and $\beta > 0$, then we have $\mathfrak{a}_n \sim \beta^{-1}(n\delta_n^d)^{-\xi}$. Additionally, consider the location-scale model in (1.6) below with scale function $\gamma(x)$ such that $D_u\gamma(x)$ exists and $D_u\gamma(x)$ is γ-Hölder continuous at each $x \in \mathbb{B}$ and $u = 1,\dots,d$. In this case, we have $\sup_{x\in\mathbb{B}_n}|R(x,\delta_n)| = O(\delta_n^{1+\gamma})$. Therefore, the condition (1.5) is satisfied when $\delta_n = o(n^{\xi/(1+\gamma-d\xi)})$ (note that $\xi < 0$).

In this case, Assumption 1 (ii) is satisfied with $\Gamma(x) = \gamma(x)^{1/\xi}$. Also, Assumption 1 (v) is satisfied if $D_u\gamma(x)$ exists and $D_u\gamma(x)$ is γ-Hölder continuous at each $x \in \mathbb{B}$ and $u = 1, \ldots, d$.[3]

We note that Assumptions 1 and 2 could be relaxed in certain directions for some of the results stated below, but we decided to state a single set of sufficient assumptions for all the results in this section. We will extend results in this section later in Section 2.

We impose the following conditions for the kernel function.

Assumption 3.

(i) Let $w = (w_1, \ldots, w_d)' \in \mathbb{R}^d$. The kernel function K is a bounded positive Lipschitz function with support $[-1, 1]^d$ and second order, that is

$$\int_{\mathbb{R}^d} K(w)dw = 1, \quad \int_{\mathbb{R}^d} K(w)w_u dw = 0 \text{ for } u = 1, \ldots, d.$$

(ii) $\int_{\mathbb{R}^d} K(w)\tilde{w}\tilde{w}'dw$ is positive definite, where $\tilde{w} = (1, w_1, \ldots, w_d)' \in \mathbb{R}^{d+1}$.

These assumptions are standard in the literature and satisfied by popular kernel functions, such as the uniform and biweight kernels. If one wishes to incorporate a discrete covariate, say $D_i \in \{1, \ldots, M\}$, our estimator for the α-th coniditonal quantile of $Y|X = c, D = m$ can be obtained as in (1.2) by replacing the kernel component "$K(\delta_n^{-1}(X_i - c))$" with "$K(\delta_n^{-1}(X_i - c))\mathbb{I}\{D_i = m\}$".

In the next section, we derive the asymptotic distribution of our local linear quantile regression estimator.

[3]To see this, a Taylor expansion of $\gamma(x)$ around $x = c$ yields $\gamma(x) = \gamma(c) + (x - c)'\frac{\partial\gamma(c)}{\partial x} + R_\gamma(x, \delta_n)$, where $\sup_{x\in\mathbb{B}_n} |R_\gamma(x, \delta_n)| = O(\delta_n^{1+\gamma})$. Thus, by noting $\theta_{\alpha_n}(c) = \varphi(c) + F_{U_*}^{-1}(\alpha_n)\gamma(c)$ and $\frac{\partial\theta_{\alpha_n}(c)}{\partial x} = \frac{\partial\varphi(c)}{\partial x} + F_{U_*}^{-1}(\alpha_n)\frac{\partial\gamma(c)}{\partial x}$, (1.3)-(1.5) imply

$$\begin{aligned}
& \sup_{x\in\mathbb{B}_n} \mathfrak{a}_n \left|\theta_{\alpha_n}(x) - \theta_{\alpha_n}(c) - (x - c)'\frac{\partial\theta_{\alpha_n}(c)}{\partial x}\right| \\
= & \sup_{x\in\mathbb{B}_n} \mathfrak{a}_n \left|\theta_{\alpha_n}(x) - \{\varphi(c) + F_{U_*}^{-1}(\alpha_n)\gamma(c)\} - (x - c)'\left\{\frac{\partial\varphi(c)}{\partial x} + F_{U_*}^{-1}(\alpha_n)\frac{\partial\gamma(c)}{\partial x}\right\}\right| \\
\leq & \sup_{x\in\mathbb{B}_n} \mathfrak{a}_n|R_\varphi(x, \delta_n)| + \frac{F_{U_*}^{-1}(\alpha_n)}{F_{U_*}^{-1}(1/n\delta_n^d)} \sup_{x\in\mathbb{B}_n} |R_\gamma(x, \delta_n)| \to 0.
\end{aligned}$$

1.2 Asymptotic distribution of estimator

Let $U_{n,i} = U_i + R_\varphi(X_i, \delta_n) - \mathfrak{b}_n$. Define $\mathbb{S} = \mathbb{S}_\infty \times \mathbb{R}^d$, where

$$\mathbb{S}_\infty = \begin{cases} [-\infty, \infty) & \text{if } \xi = 0, \\ [-\infty, 0) & \text{if } \xi > 0, \\ [0, \infty) & \text{if } \xi < 0. \end{cases}$$

As a preparation for the asymptotic analysis on the conditional quantile estimator $\hat{\theta}_{\alpha_n}(c)$, we consider the following point process

$$\hat{N}(\cdot) = \sum_{i=1}^{n} \mathbb{I}\{(\mathfrak{a}_n U_{n,i}, \delta_n^{-1}(X_i - c)) \in \cdot\}, \tag{1.7}$$

as a random element of the metric space $M_p(\mathbb{S})$ of point processes defined on the measurable space $(\mathbb{S}, \sigma(\mathbb{S}))$, where $\sigma(\mathbb{S})$ is the σ-algebra generated by the open sets of $\mathbb{S}$, and the metric space $M_p(\mathbb{S})$ is equipped with the metric induced by the topology of vague convergence (see Resnick, 1987, for details on the theory of point process). In finite samples, if $\xi \neq 0$, $\mathfrak{a}_n U_{n,i}$ may not be in $\mathbb{S}_\infty$ due to the term $R_\varphi(X_i, \delta_n)$ and therefore we need to restrict the state space of $\mathfrak{a}_n U_{n,i}$ on $\mathbb{S}_\infty$ in general. However, such a restriction on the state space does not cause any technical problem since $\mathfrak{a}_n |R_\varphi(x, \delta_n)| = O(\mathfrak{a}_n \delta_n^{1+\gamma}) = o(1)$ uniformly over $x \in \mathbb{B}_n$ under Assumption 1 (iv), and this implies that the restriction is asymptotically negligible.

The following result plays an important role to investigate the asymptotic properties of $\hat{\theta}_{\alpha_n}(c)$.

Proposition 1 (Weak convergence of $\hat{N}$)**.** *Under Assumptions 1-2, $\hat{N} \xrightarrow{d} N$ in $M_p(\mathbb{S})$, where N is a Poisson point process in $M_p(\mathbb{S})$ with mean measure*

$$m(du, dw) = \begin{cases} \Gamma(c) f_X(c) e^u du dw & \text{if } \xi = 0, \\ \Gamma(c) f_X(c) \frac{1}{\xi} (-u)^{-1/\xi - 1} du dw & \text{if } \xi > 0, \\ -\Gamma(c) f_X(c) \frac{1}{\xi} u^{-1/\xi - 1} du dw & \text{if } \xi < 0. \end{cases} \tag{1.8}$$

Remark 1. Proposition 1 can be established by asymptotic theory of point process and the weak convergence $\hat{N} \xrightarrow{d} N$ enables us to develop statistical inference on extreme order conditional quantiles. It should be noted that the limit distribution of $\hat{\theta}_{\alpha_n}(c)$ is not normal (see Theorem 2). Therefore, our analysis is quite different from the extrapolation approach, in which extremal order conditional quantiles are estimated by extrapolations of estimators for intermediate order quantiles ($n\delta_n^d \alpha_n \to \infty$ as $n \to \infty$), investigated by e.g. Wang, Li and He (2012) and Daouia, Gardes and Girard (2013) for linear regression and kernel smoothing, respectively.

Now we study asymptotic properties of the quantile regression estimator $\hat{\theta}_{\alpha_n}(c)$. To this end, we first characterize the limiting behavior of the coefficient estimator $(\hat{\theta}_{\alpha_n}(c), \hat{\beta}_{\alpha_n}(c))$. In particular, we consider the normalized object

$$\Delta_n = \mathfrak{a}_n \begin{pmatrix} \hat{\theta}_{\alpha_n}(c) - \varphi(c) - \mathfrak{b}_n \\ \hat{\beta}_{\alpha_n}(c) - \delta_n \frac{\partial \varphi(c)}{\partial x} \end{pmatrix}.$$

The object Δ_n is centered around $\left(\varphi(c) + \mathfrak{b}_n, \delta_n \frac{\partial \varphi(c)}{\partial x'}\right)$ instead of the coefficients $\left(\theta_{\alpha_n}(c), \delta_n \frac{\partial \theta_{\alpha_n}(c)}{\partial x'}\right)$ to cover all the cases (1)-(3) in Assumption 1. For the cases (2)-(3), we have $\mathfrak{b}_n = 0$. Also $\left(\theta_{\alpha_n}(c), \delta_n \frac{\partial \theta_{\alpha_n}(c)}{\partial x'}\right)$ involves a bias component as illustrated in the location-scale example in (1.6) implying $\left(\theta_{\alpha_n}(c), \delta_n \frac{\partial \theta_{\alpha_n}(c)}{\partial x'}\right) = \left(\varphi(c), \delta_n \frac{\partial \varphi(c)}{\partial x'}\right) + F_{U_*}^{-1}(\alpha_n) \left(\gamma(c), \delta_n \frac{\partial \gamma(c)}{\partial x'}\right)$.

By using the Poisson point process N in Proposition 1, the asymptotic distribution of Δ_n is obtained as follows.

Theorem 1 (Asymptotic distribution of Δ_n). *Under Assumptions 1-3, it holds $\Delta_n \xrightarrow{d} \Delta_\infty(k)$ provided $\Delta_\infty(k)$ is defined as a random vector in $\mathbb{R}^{d+1}$ which uniquely minimizes the objective function*

$$\begin{aligned} Q_\infty(\Delta, k) &= -k f_X(c) \left\{ \int_{[-1,1]^d} K(w) \tilde{w} dw \right\}' \Delta - \int_{\mathbb{S}} K(w) \min\{u - \tilde{w}'\Delta, 0\} dN(u, w) \\ &= -k f_X(c) \left\{ \int_{[-1,1]^d} K(w) \tilde{w} dw \right\}' \Delta - \sum_{i=1}^{\infty} K(\mathcal{W}_i) \min\{\mathcal{J}_i - \tilde{\mathcal{W}}_i'\Delta, 0\}, \qquad (1.9) \end{aligned}$$

with respect to $\Delta \in \mathcal{Q}$ where $\mathcal{Q} = \mathbb{R}^{d+1}$ for $\xi \leq 0$ and $\mathcal{Q} = \{a \in \mathbb{R}^{d+1} : \max_{w \in [-1,1]^d} \tilde{w}'a \leq 0\}$ for $\xi > 0$,

$$\begin{aligned} \mathcal{J}_i &= \begin{cases} \log\left(\frac{\mathcal{G}_i}{2^d \Gamma(c) f_X(c)}\right) & \text{if } \xi = 0, \\ -\text{sgn}(\xi)\left(\frac{\mathcal{G}_i}{2^d \Gamma(c) f_X(c)}\right)^{-\xi} & \text{if } \xi \neq 0, \end{cases} \\ \mathcal{G}_i &= \sum_{j=1}^{i} \eta_j, \\ \{\eta_j\} &= \text{i.i.d. sequence of Exp}(1) \text{ random variables}, \\ \{\mathcal{W}_i\} &= \text{i.i.d. sequence of uniform random variables on } [-1,1]^d, \text{ and } \tilde{\mathcal{W}}_i = (1, \mathcal{W}_i')'. \end{aligned}$$

Remark 2. Theorem 1 implies that the limiting distribution may be approximated by

$$\arg\min_{\Delta\in\mathbb{R}^{d+1}}\left\{-\frac{kf_X(c)}{S}\sum_{i=1}^{S}K(\mathcal{W}_i)\tilde{\mathcal{W}}_i{}'\Delta-\sum_{i=1}^{S}K(\mathcal{W}_i)\min\{\mathcal{J}_i-\tilde{\mathcal{W}}_i{}'\Delta,0\}\right\}, \tag{1.10}$$

for large values of S. In particular, (1.10) is equivalent to

$$\arg\min_{\Delta\in\mathbb{R}^{d+1}}\sum_{i=1}^{S}K(\mathcal{W}_i)\rho_{\frac{kf_X(c)}{S}}(\mathcal{J}_i-\tilde{\mathcal{W}}_i{}'\Delta),$$

and we can simulate the asymptotic distribution of Δ_n from the weighted quantile regression. However, this simulation requires knowledge of the objects ξ, $f_X(c)$, and $\Gamma(c)$, which are unknown to the researcher. For example when $\{Y_i, X_i\}$ is an i.i.d. sample, Daouia, Gardes and Girard (2013) proposed a Pickands type estimator of ξ, which also can be applied to the varying extreme value index where ξ may depend on c and they showed its consistency under intermediate order asymptotics ($n\delta_n^d\alpha_n\to\infty$ as $n\to\infty$). We will discuss extensions of our results to the varying extreme value index in Section 2. The density $f_X(c)$ may be estimated by the kernel estimator, for example. On the other hand, it is not clear how to estimate $\Gamma(c)$ (defined in Assumption 1 (ii)) to implement the simulation based on (1.10). Therefore, we do not pursue such an analytical approach for inference of the conditional quantile $\theta_{\alpha_n}(c)$ and we instead consider a subsampling method which completely avoids estimation of the nuisance components ξ, $f_X(c)$, and $\Gamma(c)$.

Define $\Delta_\infty(k)=(\Delta_{\infty,0}(k),\ldots,\Delta_{\infty,d}(k))'$. Based on Theorem 1, the asymptotic distribution of $\hat{\theta}_{\alpha_n}(c)$ is obtained as follows.

Theorem 2 (Asymptotic distribution of $\hat{\theta}_{\alpha_n}(c)$ and Θ_n)**.** *Under Assumptions 1-3, we have that*

$$\mathfrak{a}_n(\hat{\theta}_{\alpha_n}(c)-\theta_{\alpha_n}(c))\xrightarrow{d}\Delta_{\infty,0}(k)+g(c;\xi), \tag{1.11}$$

and

$$\begin{aligned}\Theta_n &= \frac{\hat{\theta}_{\alpha_n}(c)-\theta_{\alpha_n}(c)}{\hat{\theta}_{m\alpha_n}(c)-\hat{\theta}_{\alpha_n}(c)}\\ &\xrightarrow{d}\frac{\Delta_{\infty,0}(k)+g(c;\xi)}{\Delta_{\infty,0}(mk)-\Delta_{\infty,0}(k)}=:\Theta_\infty,\end{aligned} \tag{1.12}$$

for any m such that $k(m-1)>d+1$, provided $\Delta_\infty(k)$ and $\Delta_\infty(mk)$ are uniquely defined random vectors in $\mathbb{R}^{d+1}$ and

$$g(x;\xi)=\begin{cases}\log(\Gamma(x)/k) & \text{if } \xi=0,\\ \operatorname{sgn}(\xi)\cdot(\Gamma(x)/k)^{\xi} & \text{if } \xi\neq 0.\end{cases}$$

Remark 3. Theorem 2 implies that $\hat{\theta}_{\alpha_n}(c) - \theta_{\alpha_n}(c) = O_p(1/\mathfrak{a}_n)$, where $\mathfrak{a}_n$ is defined in Assumption 1 (iii). We note that $\mathfrak{a}_n \downarrow 0$ for the case of $\xi > 0$ and $\mathfrak{a}_n \to \infty$ for the case of $\xi < 0$. Since $\mathfrak{a}_n$ is unknown in general, we cannot use (1.11) to provide practical inference tools for $\theta_{\alpha_n}(c)$. On the other hand, the weak convergence result in (1.12) is useful for inference on $\theta_{\alpha_n}(c)$ since we can compute Θ_n, which is a randomly self-normalized version of $\hat{\theta}_{\alpha_n}(c) - \theta_{\alpha_n}(c)$, without the knowledge of canonical normalization $\mathfrak{a}_n$.

Remark 4 (Comparison with Daouia, Gardes and Girard (2013))**.** We now compare the point estimator $\hat{\theta}_{\alpha_n}(c)$ with the extrapolation-based approach. Daouia, Gardes and Girard (2013) studied kernel smoothing for estimating extremal conditional quantiles by using the relation

$$\frac{\theta_{t\alpha}(c) - \theta_{\alpha}(c)}{a(\theta_{\alpha}(c)|c)} - K_{\xi(c)}(1/t) \to 0,$$

for all $t > 0$ as $\alpha \to 0$ under Assumption (A.1) in their paper. Here, $a(\cdot|c)$ is an auxiliary function defined in Daouia, Gardes and Girard (2013), $\xi(c)$ is the extreme value index of $F_Y(y|c)$, and $K_{\xi}(u) = \int_1^u v^{\xi-1} dv$. Based on this result, one can construct an estimator of $\theta_{\alpha_n}(c)$ by

$$\hat{\theta}^{E}_{\alpha_n}(c) = \hat{\theta}_{\tilde{\alpha}_n}(c) + K_{\hat{\xi}(c)}(\tilde{\alpha}_n/\alpha_n)\hat{a}(c), \tag{1.13}$$

where $\hat{\xi}(c)$ and $\hat{a}(c)$ are estimators of $\xi(c)$ and $a(\theta_{\tilde{\alpha}_n}(c)|c)$ respectively, $\tilde{\alpha}_n$ is an intermediate quantile level such that $\tilde{\alpha}_n \to 0$ and $n\delta_n^d\tilde{\alpha}_n \to \infty$ as $n \to \infty$, and $\hat{\theta}_{\tilde{\alpha}_n}(c)$ is the intermediate quantile regression estimator defined as

$$\hat{\theta}_{\tilde{\alpha}_n} = \inf_{y\in\mathbb{R}}\{y : \hat{F}_Y(y|c) > \tilde{\alpha}_n\}, \quad \hat{F}_Y(y|c) = \frac{\sum_{i=1}^n K(\delta_n^{-1}(X_i - c))\mathbb{I}\{Y_i \le y\}}{\sum_{i=1}^n K(\delta_n^{-1}(X_i - c))}.$$

Intuitively the estimator $\hat{\theta}^{E}_{\alpha_n}(c)$ uses sample information from less extreme observations to estimate intermediate quantiles at $\tilde{\alpha}_n$, which can yield desirable risk properties as a point estimator. Indeed Daouia, Gardes and Girard (2013) carefully studied the estimation method of the second term in (1.13) and investigated the asymptotic properties of $\hat{\theta}^{E}_{\alpha_n}(c)$. Compared to $\hat{\theta}^{E}_{\alpha_n}(c)$, our point estimator $\hat{\theta}_{\alpha_n}(c)$ uses less sample information and the convergence rate tends to be slower. Rather our focus is on inference (i.e., confidence interval and hypothesis testing) based on the point process theory instead of central limit theorems, and the result in Theorem 2 should be understood as a building block for subsampling inference.

Remark 5 (Uniqueness of $\Delta_\infty(k)$ and continuity of $G(x) = \mathbb{P}(\Theta_\infty \le x)$)**.** Uniqueness of $\Delta_\infty(k)$ is necessary to apply the convexity lemma (Geyer, 1996, and Knight, 1999) to show the weak convergence of Δ_n. Furthermore, we need the continuity of $G(x)$ to show the asymptotic validity of our subsampling method. We can show the uniqueness of $\Delta_\infty(k)$ and continuity of $G(x)$ if $\int_{\mathbb{R}^d} K(w)\tilde{w}\tilde{w}'dw$ is positive definite. Indeed, since $Q_\infty(\Delta, k)$ is convex in Δ and $\mathcal{W}$ is the uniform random variable on $[-1,1]^d$, Chernozhukov (2005, Condition PJ) is satisfied. Therefore, we can show the tightness of $\Delta_\infty(k)$ similarly to the proof of Chernozhukov (2005, Lemma 9.7). Taking the tightness of $\Delta_\infty(k)$ as given and under Assumption 3 (ii), we can show that (a) $\Delta_\infty(k)$ is uniquely determined almost surely, (b) $\Delta_{\infty,0}(mk) - \Delta_{\infty,0}(k) > 0$ almost surely, and (c) $\Delta_{\infty,0}(k)$ has the continuous distribution function by a similar argument to the proof of Chernozhukov and Fernández-Val (2011, Lemma E1). Therefore, (b) and (c) imply that Θ_∞ is a proper random variable with a continuous distribution function.

Remark 6 (Choice of the bandwidth δ_n)**.** To implement our point estimator $\hat{\theta}_{\alpha_n}(c)$ in (1.2), we need to choose the bandwidth δ_n. One data-driven approach is to adapt cross validation to the local quantile regression as in Takeuchi *et al.* (2006). For example, the leave-one-out cross validation minimizes $\sum_{i=1}^n K(\delta_n^{-1}(X_i - c))\rho_{\alpha_n}(Y_i - \hat{\theta}_{\alpha_n}^{(-i)}(c) - \delta_n^{-1}(X_i - c)'\hat{\beta}_{\alpha_n}^{(-i)}(c))$ with respect to δ_n, where $(\hat{\theta}_{\alpha_n}^{(-i)}(c), \hat{\beta}_{\alpha_n}^{(-i)}(c))$ is obtained as in (1.2) by deleting the i-th observation (Y_i, X_i). However, its theoretical analysis is beyond the scope of this paper.

1.3 Verification of the condition on $\hat{\theta}_{\alpha_b}(c)$ in Assumption (iii) of the main paper for the local linear quantile regression estimator

Define $A_b = -\mathrm{sgn}(\xi) \cdot 1/F_{U^*}^{-1}(1/(b\delta_b^d))$ and $w_n = |A_b(\theta_{\alpha_b}(c) - \hat{\theta}_{\alpha_b}(c))|$. Note that $\hat{\theta}_{\alpha_b}(c)$ is the intermediate order conditional quantile computed using the full sample of size n since $\alpha_b n\delta_n^d = k_n\alpha_b/\alpha_n \to \infty$ as $n \to \infty$. To show $w_n = o_p(1)$, we apply the results in Ichimura, Otsu and Altonji (2019). Now we assume the following conditions:

(a) $\frac{\partial F_U^{-1}(\tau|x)}{\partial\tau} \sim \frac{\partial F_{U_*}^{-1}(\tau/\Gamma(x))}{\partial\tau}$ as $\tau \downarrow 0$ uniformly over $x \in \mathbb{B}$. $\frac{\partial F_{U_*}^{-1}(\tau)}{\partial\tau}$ is regularly varying at 0 with exponent $\xi + 1$ for some $\xi \neq 0$, and $\lim_{\tau\downarrow 0}\left|\frac{\partial F_{U_*}^{-1}(\tau)/\partial\tau}{\tau^{-1}F_{U_*}^{-1}(\tau)}\right| \in (0,\infty)$.

(b) Conditions C1, F1, and R1 (when $\xi < 0$) or Conditions C2, F2, and R2 (when $\xi > 0$) hold by replacing α_n (in their notation) with α_b.

Then, Theorems 1 and 3 in Ichimura, Otsu and Altonji (2019) yield

$$\theta_{\alpha_b}(c) - \hat{\theta}_{\alpha_b}(c) = O_p\left(\sqrt{\frac{\alpha_b}{n\delta_n^d\phi_b^2}}\right),$$

where

$$\begin{aligned}\phi_b &= f_Y(\theta_{\alpha_b}(c)|c) = f_{\theta_0(c)+U}(\theta_0(c) + F_U^{-1}(\alpha_b|c)|c) = f_U(F_U^{-1}(\alpha_b|c)|c) \\ &= \frac{1}{\partial F_U^{-1}(\tau|c)/\partial\tau\big|_{\tau=\alpha_b}} \sim \frac{1}{\partial F_{U_*}^{-1}(\tau/\Gamma(c))/\partial\tau\big|_{\tau=\alpha_b}} \sim \frac{\alpha_b/\Gamma(c)}{L_0 F_{U_*}^{-1}(\alpha_b/\Gamma(c))},\end{aligned}$$

for $L_0 = \lim_{\tau\downarrow 0} \frac{\partial F_{U_*}^{-1}(\tau)/\partial\tau}{\tau^{-1}F_{U_*}^{-1}(\tau)} \in (0,\infty)$, and the wave relations follow from Assumption (ii) of the main paper. This implies

$$\theta_{\alpha_b}(c) - \hat{\theta}_{\alpha_b}(c) = O_p\left(\frac{\Gamma(c)F_{U_*}^{-1}(\alpha_b/\Gamma(c))}{\sqrt{\alpha_b n\delta_n^d}}\right),$$

and we obtain

$$\begin{aligned}w_n &= \left|\frac{\text{sgn}(\xi)}{F_{U_*}^{-1}(1/b\delta_b^d)}\right| O_p\left(\frac{\Gamma(c)F_{U_*}^{-1}(\alpha_b/\Gamma(c))}{\sqrt{\alpha_b n\delta_n^d}}\right) \\ &= O_p\left(k^{-\xi}\Gamma(c)^{\xi+1}\sqrt{\frac{1}{\alpha_b n\delta_n^d}}\right) = O_p\left(\sqrt{\frac{\alpha_n}{\alpha_b}}\right) = o_p(1),\end{aligned}$$

since $k_n = n\delta_n^d\alpha_n (= b\delta_b^d\alpha_b) \to k \in (0,\infty)$ and $\alpha_n/\alpha_b \to 0$ as $b, n \to \infty$.

2 Extension

In this section, we discuss two extensions of the results in Section 1. In particular, we extend our results to (i) the case where the extreme value index of U_* may vary with covariates (Section 2.1) and (ii) varying coefficient extremal quantile regression models (Section 2.2).

2.1 Varying extreme value index

In this section we extend our results to the case where the extreme value index of U_* may vary with covariates, that is, the distribution of U_* depends on $X = x$ through the extreme value index $\xi(x) \neq 0$. Before we state our results, we provide an example to motivate such an extension.

Example 1. Suppose that X is half-normal with negative support and Y given $X = x$ is the negative Pareto distribution such that $F_Y(y|x) = (1+|y|)^{-1/|x|}$ for $y \leq 0$ and $x < 0$. Then the conditional quantile is $\theta_\tau(x) = 1 - \tau^{-|x|} = 1 - \tau^x$. In this case, we cannot apply Theorem 1 in the supplement to estimate $\theta_{\alpha_n}(c)$ $(c < 0)$ since the conditional tail index is $\xi(x) = |x| > 0$ is not constant.[4]

[4]See Daouia, Gardes and Girard (2013, Section 4) for further examples of varying extreme value index models.

To allow dependence of ξ on $X = x$, we impose the following assumption.

Assumption 4. *(ii') There exists a measurable function $\varphi : \mathbb{B} \to \mathbb{R}$ such that the conditional distribution function $F_U(z|x)$ of $U = Y - \varphi(X)$ given $X = x$ satisfies that $F_U(z|x)/F_{U_*}(z|x) \sim \Gamma(x)$, as $z \downarrow F_{U_*}^{-1}(0|x)$, uniformly over $x \in \mathbb{B}$ for some positive continuous function $\Gamma(x)$ on $\mathbb{B}$. The quantile function $F_{U_*}^{-1}(\cdot|x)$ of U_* given $X = x$ has end-points $F_{U_*}^{-1}(0|x) = 0$ or $F_{U_*}^{-1}(0|x) = -\infty$. The conditional distribution function $F_{U_*}(z|x)$ exhibits Pareto-type tails with extreme value index $\xi(x) \neq 0$, i.e.,*

(2) as $z \downarrow F_{U_*}^{-1}(0|x) = -\infty$, $F_{U_*}(vz|x) \sim v^{-1/\xi(x)} F_{U_*}(z|x)$ for all $v > 0$, where $\xi : \mathbb{R}^d \to [0, \infty)$ is positive and continuous on $\mathbb{B}$.

(3) as $z \downarrow F_{U_*}^{-1}(0|x) = 0$, $F_{U_*}(vz|x) \sim v^{-1/\xi(x)} F_{U_*}(z|x)$ for all $v > 0$, where $\xi : \mathbb{R}^d \to (-\infty, 0]$ is negative and continuous on $\mathbb{B}$.

(iii') Let δ_n be a sequence of positive constants with $\delta_n \to 0$ as $n \to \infty$. Assume that $n\delta_n^d \alpha_n \to k \in (0, \infty)$ and $\mathfrak{a}_n \delta_n^{1+\gamma} \to 0$ as $n \to \infty$, where

(2) $\mathfrak{a}_n = -1/F_{U_*}^{-1}(1/n\delta_n^d|c)$ and $\mathfrak{b}_n = 0$ when $\xi(c) > 0$,

(3) $\mathfrak{a}_n = 1/F_{U_*}^{-1}(1/n\delta_n^d|c)$ and $\mathfrak{b}_n = 0$ when $\xi(c) < 0$.

We call the set of Assumptions 1 (i), (iv) and (v), and Assumptions 4 (ii') and (iii') as Assumption 1'

Remark 7. In Example 1, Assumptions 4 (ii') and (iii') are satisfied with $\varphi(x) = 0$, $\Gamma(x) = 1$, and $\mathfrak{a}_n = 1/((1/n\delta_n^d)^c - 1)$. We can also check that Example 1 satisfies Assumption 1 (v). Now we additionally assume that $\delta_n(\log n)^2 \to 0$ as $n \to \infty$. This implies that $\alpha_n^{-\delta_n} \to 1$ as $n \to \infty$ since

$$\delta_n \log \alpha_n \sim \delta_n \log(k/n\delta_n^d) = \delta_n(\log k - \log n - d\log\delta_n) \to 0.$$

Define $D\theta_\tau(x) = d\theta_\tau(x)/dx = -\tau^x(\log\tau)$. We have

$$\begin{aligned} \sup_{|x-c|\le\delta_n} \mathfrak{a}_n |D\theta_{\alpha_n}(x) - D\theta_{\alpha_n}(c)| &\lesssim k^c |\log\alpha_n| \sup_{|x-c|\le\delta_n} |\alpha_n^{x-c} - 1| \lesssim k^c|\log\alpha_n|(\alpha_n^{-\delta_n} - 1) \\ &\lesssim k^c|\log\alpha_n|\alpha_n^{-\pi_n\delta_n}\delta_n|\log\alpha_n|, \text{ for } \pi_n \in (0,1) \\ &\lesssim (\log n)^2\delta_n, \end{aligned} \tag{2.1}$$

As they argue, under the von-Mises type condition, $\xi(x)$ may be characterized as

$$\lim_{y\downarrow\theta_0(x)} \frac{F_Y(y|x) d^2F_Y(y|x)/dy^2}{\{dF_Y(y|x)/dy\}^2} = \xi(x) + 1,$$

where $\theta_0(x) = \lim_{\alpha\downarrow 0}\theta_\alpha(x)$.

where the third inequality follows from the mean value theorem. Therefore, (2.1) implies $\sup_{x\in\mathbb{B}_n}\mathfrak{a}_n|R(x,\delta_n)| = o(1)$ and this also implies Assumption 1 (v). Analogously, the condition would be satisfied for a wide class of models if $\xi(x)$ and $\Gamma(x)$ are sufficiently smooth on $\mathbb{B}$. Furthermore, it is easy to check that Example 1 satisfies Assumption 5 below.

Under this assumption, our main results are extended as follows.

Theorem 3. *Suppose that Assumptions 1' and 2 hold. Then the same result of Proposition 1 when $\xi \neq 0$ holds by replacing ξ with $\xi(c)$. Additionally, suppose that Assumption 3 holds. Then the same results of Theorems 1 and 2 when $\xi \neq 0$ hold by replacing ξ with $\xi(c)$.*

For subsampling inference, we impose the following assumption.

Assumption 5. *The conditional quantile density function $\partial F_U^{-1}(\tau|x)/\partial\tau$ exists and satisfies the tail equivalence relationship*

$$\frac{\partial F_U^{-1}(\tau|x)}{\partial\tau} \sim \frac{\partial F_{U_*}^{-1}(\tau/\Gamma(x)|x)}{\partial\tau} \quad \text{as } \tau \downarrow 0,$$

uniformly over $x \in \mathbb{B}$, where $\partial F_{U_*}^{-1}(\tau|x)/\partial\tau$ is regularly varying at 0 with exponent $\xi(x)+1$ on $\mathbb{B}$. We also assume that there exists a function h such that h is continuous on $\mathbb{B}$ and $\lim_{\tau\downarrow 0}\left|\frac{\partial F_{U_*}^{-1}(\tau|x)/\partial\tau}{\tau^{-1}F_{U_*}^{-1}(\tau|x)}\right| = h(x) \in (0,\infty)$ on $\mathbb{B}$.

For the case of the location-scale model $Y = \varphi(X)+\gamma(X)U_*$ with $F_{U_*}(u|x) = (1+|u|)^{-1/\xi(x)}$ for $u \leq 0$ and some positive continuous function $\xi(x)$, we have $F_{U_*}^{-1}(\tau|x) = 1-\tau^{-\xi(x)}$ and the function $h(x)$ in the above assumption coincides with $\xi(x)$.

Under the above assumptions, the validity of our subsampling inference for the case of varying extreme value indices is established as follows. The proof requires the convergence rate of our estimator under the intermediate order quantile asymptotics, which can be obtained by adapting the argument in Ichimura, Otsu and Altonji (2019) for the case of varying extreme value indices.

Theorem 4. *Let $t \in (0,1)$. As $n \to \infty$, it holds $b \to \infty$, $b/n \to 0$, $\delta_n \to 0$, $\delta_b \to 0$, $\alpha_b \to 0$, and $\alpha_b/\alpha_n \to \infty$. Under Assumptions 1', 2, 3, 5, and condition (b) in Section 1.3 of this supplement, the same result of Theorem 1 in the main paper holds.*

2.2 Varying coefficient extremal quantile regression

We can also extend our analysis in Section 2 to varying coefficient extremal quantile regression models. Let Z be a random variable in $\mathbb{R}^{d_Z}$, and fix $c_Z \in \mathbb{R}^{d_Z}$. We consider the following varying coefficient model

$$Y = X'\beta(Z) + \gamma(X,Z)V_*, \tag{2.2}$$

where $\beta(\cdot) = (\beta_0(\cdot), \ldots, \beta_d(\cdot))'$ are unknown functions of Z, $\gamma(\cdot)$ is a scale function, and V_* is an error term that is independent of (X, Z) and is in the domain of minimum attraction with $\xi \neq 0$. This specification allows the effect of each element of X to depend on Z in a nonparametric way. As well as nesting nonparametric additive models (Hastie and Tibshirani, 1993), this varying coefficient model is also a generalization of the partially linear model (Robinson, 1988). Also in the literature of regression quantiles, many papers studied the varying coefficient model and its variants for fixed quantiles; see Lee (2003) for partially linear models, Horowitz and Lee (2005) for additive models, Honda (2004) and Kim (2007) for varying coefficient models, among others. In this subsection, we contribute to this literature by considering varying coefficient models in the context of extremal quantiles.

In this setup, the assumptions in Section 2 are adapted as follows. Let $\mathbb{B}_Z$ denote some fixed closed ball around c_Z.

Assumption 6. *(i) $\{Y_i, X_i, Z_i\}$ is a sample from $(Y, X, Z) \in \mathbb{R} \times \mathbb{R}^{d+1} \times \mathbb{R}^{d_Z}$. The random variable (X, Z) has the distribution function $F(x, z)$ with compactly supported conditional distribution function $F_X(x|z)$ for $z \in \mathbb{B}_Z$. Z has the density function $f_Z(z)$ that is positive and continuous on a neighborhood around $\mathbb{B}_Z$.*

(ii) $\mathbb{E}[XX'|Z = c_Z]$ is positive definite. Without loss of generality, let $\mathbb{E}[X|Z = c_Z] = (1, 0, \ldots, 0)'$.

Assumption 7. *(i) There exists a measurable function $\beta(\cdot) = (\beta_0(\cdot), \ldots, \beta_d(\cdot))' : \mathbb{B}_Z \to \mathbb{R}^{d+1}$ such that the conditional quantile function of $V = Y - X'\beta(Z)$ given $X = x$ and $Z = z$ satisfies that $F_V^{-1}(v|x, z)/F_{V_*}^{-1}(v) \sim \gamma(x, z)$, as $v \downarrow 0$, uniformly over $\{(x, z) : x \in S(X|z), z \in \mathbb{B}_Z\}$ for some positive continuous function $\gamma(x, z)$ on $\{(x, z) : x \in S(X|z), u \in \mathbb{B}_Z\}$, where $S(X|z)$ is the support of $F_X(x|z)$. The quantile function $F_{V_*}^{-1}$ of V_* has end-points $F_{V_*}^{-1}(0) = 0$ or $F_{V_*}^{-1}(0) = -\infty$. The distribution function $F_{V_*}(v)$ exhibits Pareto-type tails with extreme value index $\xi \neq 0$.*

(ii) Let δ_n be a sequence of positive constants with $\delta_n \to 0$ as $n \to \infty$. Assume that $n\delta_n^{d_Z}\alpha_n \to k \in (0, \infty)$ and $\mathfrak{a}_n\delta_n \to 0$ as $n \to \infty$, where

(1) $\mathfrak{a}_n = -1/F_{V_*}^{-1}(1/n\delta_n^{d_Z})$ when $\xi > 0$,

(2) $\mathfrak{a}_n = 1/F_{V_*}^{-1}(1/n\delta_n^{d_Z})$ when $\xi < 0$.

Assumption 8. *(i) $D_v\gamma(x, z) = \partial\gamma(x, z)/\partial z_v$ exists and is continuous at each $z \in \mathbb{B}_Z, x \in S(X|z)$ and for each $v = 1, \ldots, d_Z$.*

(ii) $D_v\beta_j(z)$ *exists and is* γ*-Hölder continuous at each* $z \in \mathbb{B}_Z$ *and for each* $v = 1, \ldots, d_Z$*.*

Assumption 9. *The kernel function* K *is a bounded Lipschitz function with support* $[-1,1]^{d_Z}$ *and second order.*

Under these assumptions, we consider the following point process

$$\hat{N}_1(\cdot) = \sum_{i=1}^{n} \mathbb{I}\left\{\left(\mathfrak{a}_n(V_i + X_i'(\beta(Z_i) - \beta(c_Z))), X_i, (Z_i - c_Z)/\delta_n\right) \in \cdot\right\},$$

as a random element of $M_p(\mathbb{S}_1)$, where

$$\mathbb{S}_1 = \begin{cases} [-\infty, 0) \times S(X|c_Z) \times \mathbb{R}^{d_Z} & \text{if } \xi > 0, \\ [0, \infty) \times S(X|c_Z) \times \mathbb{R}^{d_Z} & \text{if } \xi < 0. \end{cases}$$

Let $\Gamma(x,z) = \gamma(x,z)^{1/\xi}$.

Proposition 2 (Weak convergence of $\hat{N}_1$)**.** *Under Assumptions 6-9 and Assumption 2 by replacing* U_i *and* W_i *with* V_i *and* $\tilde{W}_i = (V_i, X_i', Z_i')'$*, respectively, it holds* $\hat{N}_1 \xrightarrow{d} N_1$ *in* $M_p(\mathbb{S}_1)$*, where* N_1 *is a Poisson point process in* $M_p(\mathbb{S}_1)$ *with mean measure*

$$m(dv, dx, dw) = \begin{cases} \Gamma(x, c_Z) f_Z(c_Z) \frac{1}{\xi}(-v)^{-1/\xi - 1} dv dF_X(x|c_Z) dw & \textit{if } \xi > 0, \\ -\Gamma(x, c_Z) f_Z(c_Z) \frac{1}{\xi} v^{-1/\xi - 1} dv dF_X(x|c_Z) dw & \textit{if } \xi < 0. \end{cases}$$

Now we focus on the model (2.2) and assume that $\gamma(x,z) = x'\sigma(z)$ where $\sigma(z) = (\sigma_0(z), \ldots, \sigma_d(z))'$ and $X'\sigma(z) > 0$ almost surely for $z \in \mathbb{B}_Z$. We also assume that $D_v\sigma_j(z)$ exists and is γ-Hölder continuous at each $z \in \mathbb{B}_Z$ and $v = 1, \ldots, d_Z$. In this case, the conditional quantile can be written as

$$F_Y^{-1}(\alpha_n|x, c_Z) = x'(\beta(c_Z) + \sigma(c_Z)F_{V_*}^{-1}(\alpha_n)) = x'\beta_{\alpha_n}(c_Z),$$

where $\beta_{\alpha_n}(c_Z) = \beta(c_Z) + \sigma(c_Z)F_{V_*}^{-1}(\alpha_n)$.

Based on this expression, we consider the following quantile regression problem:

$$\hat{\beta}_{\alpha_n}(c_Z) = \arg\min_{\beta \in \mathbb{R}^{d+1}} \sum_{i=1}^{n} K(\delta_n^{-1}(Z_i - c_Z))\rho_\alpha(Y_i - X_i'\beta). \tag{2.3}$$

Let $\overline{\Delta}_n = \mathfrak{a}_n(\hat{\beta}_n^{(\alpha_n)}(c_Z) - \beta(c_Z))$. The asymptotic distribution of the quantile regression estimator (2.3) for the varying coefficient model (2.2) is obtained as follows. Since the proofs of Theorems 5 and 6 are analogous to those of Theorems 1 and 2, they are omitted.

Theorem 5 (Asymptotic distribution of $\overline{\Delta}_n$)**.** *Under Assumptions 6-9 and Assumption 2 by replacing U_i and W_i with V_i and $\tilde{W}_i = (V_i, X_i', Z_i')'$, respectively, it holds $\overline{\Delta}_n \xrightarrow{d} \overline{\Delta}_\infty(k)$ provided $\overline{\Delta}_\infty(k)$ is defined as a random vector in $\mathbb{R}^{d+1}$ which uniquely minimizes the objective function*

$$\begin{aligned}\overline{Q}_\infty(\Delta, k) &= -k\mathbb{E}[X|Z = c_Z]'\Delta - \int_{\mathbb{S}_1} K(w)\min\{v - x'\Delta, 0\}dN_1(v, x, w) \\ &= -k\mathbb{E}[X|Z = c_Z]'\Delta - \sum_{i=1}^{\infty} K(\mathcal{W}_i)\min\{\mathcal{J}_i - \mathcal{X}_i'\Delta, 0\},\end{aligned}$$

with respect to $\Delta \in \mathcal{Q}_1$, where $\mathcal{Q}_1 = \mathbb{R}^{d+1}$for $\xi \leq 0$ and $\mathcal{Q}_1 \in \{a \in \mathbb{R}^{d+1} : \max_{x \in S(X|c_Z)} x'a \leq 0\}$ for $\xi > 0$,

$$\begin{aligned}\mathcal{J}_i &= -\text{sgn}(\xi) \cdot \left(\frac{\mathcal{G}_i}{2^{d_Z}\Gamma(\mathcal{X}_i, c_Z) f_Z(c_Z)}\right)^{-\xi}, \\ \mathcal{G}_i &= \sum_{j=1}^{i} \eta_j, \\ \{\eta_j\} &= \textit{i.i.d. sequence of Exp(1) random variables}, \\ \{\mathcal{X}_{1,i}\} &= \textit{i.i.d. sequence of random variables with the distribution function } F_X(\cdot|c_Z), \\ \{\mathcal{W}_i\} &= \textit{i.i.d. sequence of uniform random variables on } [-1, 1]^{d_Z}.\end{aligned}$$

For inference, we can consider the self-normalized version of $\overline{\Delta}_n$:

$$\overline{\Theta}_n = \frac{\hat{\beta}_{\alpha_n}(c_Z) - \beta_{\alpha_n}(c_Z)}{\sum_{i=1}^n K_{n,i} X_i'(\hat{\beta}_{m\alpha_n}(c_Z) - \hat{\beta}_{\alpha_n}(c_Z)) / \sum_{i=1}^n K_{n,i}},$$

where $K_{n,i} = K(\delta_n^{-1}(Z_i - c_Z))$.

Theorem 6 (Asymptotic distribution of $\overline{\Theta}_n$)**.** *Under Assumptions 6-9 and Assumption 2 by replacing U_i and W_i with V_i and $\tilde{W}_i = (V_i, X_i', Z_i')'$, respectively, it holds*

$$\overline{\Theta}_n \xrightarrow{d} \frac{\overline{\Delta}_\infty(k) + \text{sgn}(\xi) \cdot k^{-\xi}\sigma(c_Z)}{\mathbb{E}[X|Z = c_Z]'(\overline{\Delta}_\infty(mk) - \overline{\Delta}_\infty(k))} =: \overline{\Theta}_\infty(k, m),$$

for any $k(m-1) > d+1$, provided $\overline{\Delta}_\infty(k)$ and $\overline{\Delta}_\infty(mk)$ are uniquely defined random vectors in $\mathbb{R}^{d+1}$.

Remark 8. It is possible to show the uniqueness of $\overline{\Delta}_\infty(k)$ and continuity of the distribution function of $\overline{\Theta}_\infty(k, m)$ under Assumption 6 (ii) by a similar argument to $\Delta_\infty(k)$ and $\Theta_\infty(k, m)$. It also would be possible to develop subsampling based inference for each component of $\overline{\Theta}_\infty(k, m)$, i.e., we could consistently estimate the quantile of each component of $\overline{\Theta}_\infty(k, m)$ by following the procedure in Section 2.3 and by using the analogue of $\overline{\Theta}_n$ computed from each subsample. To this end, we need to derive the convergence rate of our varying coefficient estimator under the intermediate order quantile asymptotics, which is beyond the scope of this paper.

A Proofs

A.1 Proof of Proposition 1

We first consider the case where $\{U_i, X_i\}_{i=1}^n$ is i.i.d. Let $\mathcal{E}$ be finite unions and intersections of bounded open rectangles in $\mathbb{S}$. From the definition of the mean measure m in (1.8),

$$m(S) = \begin{cases} \Gamma(c) f_X(c) \int_{(u,w)\in S} e^u du dw & \text{if } \xi = 0, \\ \Gamma(c) f_X(c) \int_{(u,w)\in S} \frac{1}{\xi}(-u)^{-1/\xi-1} du dw & \text{if } \xi > 0, \\ \Gamma(c) f_X(c) \int_{(u,w)\in S} -\frac{1}{\xi} u^{-1/\xi-1} du dw & \text{if } \xi < 0, \end{cases}$$

for $S \in \mathcal{E}$. Resnick (1987, Proposition 3.22) implies that if

$$\lim_{n\to\infty} \mathbb{E}[\hat{N}(S)] = \mathbb{E}[N(S)] = m(S), \tag{A.1}$$

$$\lim_{n\to\infty} \mathbb{P}(\hat{N}(S) = 0) = \mathbb{P}(N(S) = 0) = \exp(-m(S)), \tag{A.2}$$

for all $S \in \mathcal{E}$, then it holds $\hat{N} \xrightarrow{d} N$ in $M_p(\mathbb{S})$. Thus it is sufficient for the conclusion to show (A.1) and (A.2). Hereafter we present a proof for the case of $\xi < 0$. Proofs for other cases are similar.

First, we show (A.1). For this it is sufficient to consider E of the form $S = \cup_{j=1}^M S_j$, where $S_j = (\underline{u}_j, \bar{u}_j) \times S_j^W$ for $j = 1, \ldots, M$ are nonoverlapping and nonempty subsets of $\mathbb{S}$, and S_j^W are intersections of open bounded rectangles of $\mathbb{R}^d$. Observe that

$$\begin{aligned} \mathbb{E}[\hat{N}(S)] &= \sum_{j=1}^M \mathbb{E}[\hat{N}(S_j)] = \sum_{j=1}^M n\mathbb{E}\left[\mathbb{I}\{(\mathfrak{a}_n U_{n,i}, \delta_n^{-1}(X_i - c)) \in (\underline{u}_j, \bar{u}_j) \times S_j^W\}\right] \\ &= \sum_{j=1}^M n\mathbb{E}\left[\mathbb{E}\left[\mathbb{I}\{(\mathfrak{a}_n U_{n,i}, \delta_n^{-1}(X_i - c)) \in (\underline{u}_j, \bar{u}_j) \times S_j^W\} | X_i\right]\right] \\ &= \sum_{j=1}^M n\mathbb{E}\left[\mathbb{E}\left[\mathbb{I}\{\mathfrak{a}_n(U_i + R_\varphi(X_i, \delta_n)) \in (\underline{u}_j, \bar{u}_j)\} | X_i\right] \mathbb{I}\{\delta_n^{-1}(X_i - c) \in S_j^W\}\right] \\ &= \sum_{j=1}^M n\delta_n^d \int_{w\in S_j^W} \left(F_U\left(\frac{\bar{u}_j + o(1)}{\mathfrak{a}_n} \middle| c + \delta_n w \right) - F_U\left(\frac{\underline{u}_j + o(1)}{\mathfrak{a}_n} \middle| c + \delta_n w \right) \right) f_X(c + \delta_n w) dw \\ &=: \mathbb{I}_n, \end{aligned}$$

where the first equality follows from the definition of $\{S_j\}_{j=1}^M$ (nonoverlapping), the second equality follows from the stationarity of $\{U_i, X_i\}_{i=1}^n$, the third equality follows from the law of iterated expectation, the fourth equality follows from the definition of $U_{n,i}$ and property of conditional expectation, and the fifth equality follows from the change of variables and Assumption 1 (iv) (implying (1.4)). Also, observe that

$$
\begin{aligned}
& \frac{F_U\left(\frac{u+o(1)}{\mathfrak{a}_n}\Big| c+\delta_n w\right)}{F_{U_*}\left(\frac{u+o(1)}{\mathfrak{a}_n}\right)} \times n\delta_n^d F_{U_*}\left(\frac{u+o(1)}{\mathfrak{a}_n}\right) \\
= & \frac{F_U\left((u+o(1))F_{U_*}^{-1}(1/(n\delta_n^d))\big| c+\delta_n w\right)}{F_{U_*}\left((u+o(1))F_{U_*}^{-1}(1/(n\delta_n^d))\right)} \frac{F_{U_*}\left((u+o(1))F_{U_*}^{-1}(1/(n\delta_n^d))\right)}{F_{U_*}(F_{U_*}^{-1}(1/(n\delta_n^d)))} \\
= & (\Gamma(c+\delta_n w)+o(1)) \times (u^{-1/\xi}+o(1)) \to \Gamma(c)u^{-1/\xi}, \quad \text{as } n\to\infty,
\end{aligned}
$$

uniformly over $w \in [-1,1]^d$, where the first equality follows from Assumption 1 (iii) and the second equality follows from the tail properties of U (given X) and U_* (Assumption 1 (ii)). Therefore, (A.1) is obtained as

$$
\mathbb{I}_n = \sum_{j=1}^{M} \Gamma(c+\delta_n w)\{\bar{u}_j^{-1/\xi} - \underline{u}_j^{-1/\xi}\} \int_{w\in S_j^W} f_X(c+\delta_n w)dw + o(1) \to m(S). \qquad \text{(A.3)}
$$

Next, we show (A.2). The same argument to derive (A.3) yields $\mathbb{P}\left((\mathfrak{a}_n U_{n,i}, \delta_n^{-1}(X_i - c)) \in S\right) \sim \frac{m(S)}{n}$ for any $S \in \mathcal{E}$. Thus, an application of Meyer (1973) yields (A.2). Therefore, we obtain $\hat{N} \xrightarrow{d} N$ for the case of $\xi < 0$ with i.i.d. observations.

We can also show the same result under geometric strong mixing condition (Assumption 2) as an application of Meyer's (1973) theorem and by observing that

$$
\begin{aligned}
& n\sum_{i=2}^{[n/m]} \mathbb{P}\left((\mathfrak{a}_n U_{n,1}, \delta_n^{-1}(X_1-c)) \in S, (\mathfrak{a}_n U_{n,i}, \delta_n^{-1}(X_i-c)) \in S\right) \\
\leq\ & O\left(n[n/m]\mathbb{P}((\mathfrak{a}_n U_{n,1}, \delta_n^{-1}(X_1-c)) \in S)^2\right) = O(n[n/m]\delta_n^{2d}\alpha_n^2) = O(1/m). \qquad \text{(A.4)}
\end{aligned}
$$

A.2 Proof of Theorem 1

Step 1: Overall sketch

Let

$$
\begin{aligned}
K_{n,i} &= K(\delta_n^{-1}(X_i - c)), \qquad \tilde{X}_{n,i} = (1, \delta_n^{-1}(X_i-c)')', \qquad \iota = (\theta, \beta')', \qquad \hat{\iota}_{\alpha_n} = (\hat{\theta}_{\alpha_n}(c), \hat{\beta}_{\alpha_n}(c)')', \\
\Delta &= \mathfrak{a}_n(\iota - \iota_{\varphi,n} - \mathfrak{b}_n \mathbf{e}_1), \qquad \iota_{\varphi,n} = \left(\varphi(c), \delta_n \frac{\partial \varphi(c)}{\partial x'}\right)', \qquad \mathbf{e}_1 = (1,0,\ldots,0)' \in \mathbb{R}^{d+1}.
\end{aligned}
$$

The objective function for $\hat{\iota}_{\alpha_n}$ is written as

$$\begin{aligned}
& \sum_{i=1}^{n} K_{n,i}\rho_{\alpha_n}(Y_i - \tilde{X}_{n,i}'\iota) \\
= & \sum_{i=1}^{n} K_{n,i}[\alpha_n - \mathbb{I}\{Y_i - \tilde{X}_{n,i}'\iota \le 0\}](Y_i - \tilde{X}_{n,i}'\iota) \\
= & \mathfrak{a}_n^{-1}\sum_{i=1}^{n} K_{n,i}[\alpha_n - \mathbb{I}\{\mathfrak{a}_n(U_i + R_\varphi(X_i,\delta_n) - \mathfrak{b}_n) - \tilde{X}_{n,i}'\mathfrak{a}_n(\iota - \iota_{\varphi,n} - \mathfrak{b}_n\mathbf{e}_1) \le 0\}] \\
& \times\{\mathfrak{a}_n(U_i + R_\varphi(X_i,\delta_n) - \mathfrak{b}_n) - \tilde{X}_{n,i}'\mathfrak{a}_n(\iota - \iota_{\varphi,n} - \mathfrak{b}_n\mathbf{e}_1)\} \\
= & \mathfrak{a}_n^{-1}\sum_{i=1}^{n} K_{n,i}[\alpha_n - \mathbb{I}\{\mathfrak{a}_n U_{n,i} - \tilde{X}_{n,i}'\Delta \le 0\}]\{\mathfrak{a}_n U_{n,i} - \tilde{X}_{n,i}'\Delta\}.
\end{aligned}$$

Thus, we have $\Delta_n \in \arg\min_{\Delta\in\mathbb{R}^{d+1}} Q_n(\Delta)$, where

$$\begin{aligned}
Q_n(\Delta) &= -\alpha_n\sum_{i=1}^{n} K_{n,i}\tilde{X}_{n,i}'\Delta - \sum_{i=1}^{n} K_{n,i}\mathbb{I}\{\mathfrak{a}_n U_{n,i} \le \tilde{X}_{n,i}'\Delta\}\{\mathfrak{a}_n U_{n,i} - \tilde{X}_{n,i}'\Delta\} \\
&=: -Q_{1n}(\Delta) - Q_{2n}(\Delta).
\end{aligned}$$

We also note that subtracting $\sum_{i=1}^{n} K_{n,i}\mathbb{I}\{\mathfrak{a}_n U_{n,i} \le -\delta\}(-\delta - \mathfrak{a}_n U_{n,i})$ for some $\delta > 0$ from $Q_n(\Delta)$ does not affect optimization for Δ, and denote the new objective function:

$$\tilde{Q}_n(\Delta) := -Q_{1n}(\Delta) + \tilde{Q}_{2n}(\Delta) := -Q_{1n}(\Delta) + \sum_{i=1}^{n} K_{n,i}\ell_\delta(\mathfrak{a}_n U_{n,i}, \tilde{X}_{n,i}; \Delta),$$

where

$$\ell_\delta(u,w;\Delta) = \mathbb{I}\{u \le \tilde{w}'\Delta\}(\tilde{w}'\Delta - u) - \mathbb{I}\{u \le -\delta\}(-\delta - u).$$

Since $K(w)\ell_\delta(u,w;\Delta)$ is a sum of convex function in Δ, $\tilde{Q}_n(\Delta)$ and $Q_n(\Delta)$ are also convex in Δ. Observe that

$$-Q_{1n}(\Delta) = -\frac{k+o(1)}{n\delta_n^d}\sum_{i=1}^{n} K_{n,i}\tilde{X}_{n,i}'\Delta \xrightarrow{p} -kf_X(c)\left\{\int_{[-1,1]^d} K(w)\tilde{w}dw\right\}'\Delta,$$

as $n \to \infty$ due to the law of large numbers. Moreover, by the definition of $\hat{N}$, it holds

$$\begin{aligned}
Q_{2n}(\Delta) &= \sum_{i=1}^{n} K_{n,i}\min\{\mathfrak{a}_n U_{n,i} - \tilde{X}_{n,i}'\Delta, 0\} = \int_{\mathbb{S}} K(w)\min\{u - \tilde{w}'\Delta, 0\}d\hat{N}(u,w), \\
\tilde{Q}_{2n}(\Delta) &= \int_{\mathbb{S}} K(w)\ell_\delta(u,w;\Delta)d\hat{N}(u,w).
\end{aligned}$$

Based on these notations, the convexity lemma (Geyer, 1996, and Knight, 1999) says that if

(a) $\tilde{Q}_n$ (or Q_n): $\mathbb{R}^{d+1} \to \bar{\mathbb{R}}$ is convex and lower semicontinuous in Δ for each $n \in \mathbb{N}$,

(b) $\tilde{Q}_n$ (or Q_n) marginally converges to a limit function $\tilde{Q}_\infty : \mathbb{R}^{d+1} \to \bar{\mathbb{R}}$ defined by

$$\tilde{Q}_\infty(\Delta, k) = -kf_X(c) \left\{ \int_{[-1,1]^d} K(w)\tilde{w}dw \right\}' \Delta + \int_{\mathbb{S}} K(w)\ell_\delta(u, w; \Delta)dN(u, w), \qquad \text{(A.5)}$$

over a dense subset of $\mathbb{R}^{d+1}$,

(c) $\tilde{Q}_n$ (or Q_n) is finite over a non-empty open set $\mathbb{D}_0 \subset \mathbb{R}^{d+1}$,

(d) $\tilde{Q}_\infty$ is uniquely minimized over $\mathbb{R}^{d+1}$ at a random vector $\Delta_\infty(k)$,

then we obtain the conclusion, $\Delta_n \xrightarrow{d} \Delta_\infty(k)$.

Condition (a) is satisfied from the definitions of $Q_n(\Delta)$ and $\tilde{Q}_n(\Delta)$. Condition (d) is assumed. Condition (c) is satisfied by setting $\mathbb{D}_0$ as (i) any non-empty open bounded subset of $\mathbb{R}^{d+1}$ (for $\xi \le 0$) or (ii) any non-empty open bounded subset of $\Delta_N := \{\Delta \in \mathbb{R}^{d+1} : \max_{w\in[-1,1]^d} \tilde{w}'\Delta < 0\}$. Thus, it remains to check Condition (b) (in Step 2). Finally in Step 3, we verify the second equality in (1.9).

Step 2: Check Condition (b)

Note that $\tilde{Q}_\infty(\cdot, k)$ in (A.5) is the marginal weak limit of $\{\tilde{Q}_n(\cdot)\}$ if and only if $(\tilde{Q}_n(\Delta_j), j = 1, \ldots, L) \xrightarrow{d} (\tilde{Q}_\infty(\Delta_j, k), j = 1, \ldots, L)$ for any finite collection $\{\Delta_1, \ldots, \Delta_L\}$. Let $T : M_p(\mathbb{S}) \to \mathbb{R}^L$ be a mapping defined by

$$N \mapsto \left(\int_{\mathbb{S}} K(w)\ell_\delta(u, w; \Delta_1)dN(u, w), \ldots, \int_{\mathbb{S}} K(w)\ell_\delta(u, w; \Delta_L)dN(u, w) \right)'.$$

Also define

$$\kappa = \max_{w\in[-1,1]^d, \Delta\in\{\Delta_1,\ldots,\Delta_L\}} \tilde{w}'\Delta, \quad \kappa_0 = \max_{w\in[-1,1]^d} \tilde{w}'\Delta.$$

Based on this notation, we check Condition (b) for three cases: (i) $\xi = 0$, (ii) $\xi < 0$, and (iii) $\xi > 0$.

Case (i) $\xi = 0$. Note that the map $(u, w) \mapsto K(w)\ell_\delta(u, w; \Delta)$ is continuous on $\mathbb{S} = [-\infty, \infty) \times \mathbb{R}^d$ and vanishes outside the compact set $[-\infty, \max(\kappa, -\delta)] \times [-1, 1]^d$ with $\kappa < \infty$. Then since $M_p(\mathbb{S})$ is equipped with the vague topology, this implies that $T : M_p(\mathbb{S}) \to \mathbb{R}^L$ is continuous, and the continuous mapping theorem combined with $\hat{N} \xrightarrow{d} N$ (Proposition 1) yields Condition (b).

Case (ii) $\xi < 0$. Note that the map $(u, w) \mapsto K(w) \min\{u - \tilde{w}'\Delta, 0\}$ is continuous on $\mathbb{S} = [0, \infty) \times \mathbb{R}^d$ and vanishes outside the compact set $[0, \max(\kappa, 0)] \times [-1, 1]^d$ with $\kappa < \infty$. Then

$T : M_p(\mathbb{S}) \to \mathbb{R}^L$ is continuous, and the continuous mapping theorem combined with $\hat{N} \xrightarrow{d} N$ (Proposition 1) yields Condition (b).

Case (iii) $\xi > 0$. Let $\Delta_P := \{\Delta \in \mathbb{R}^{d+1} : \max_{w\in[-1,1]^d} \tilde{w}'\Delta > 0\}$. Since $\Delta_N \cup \Delta_P$ is dense in $\mathbb{R}^{d+1}$, it is enough to show that $\tilde{Q}_n(\Delta) \xrightarrow{d} \tilde{Q}_\infty(\Delta, k)$ for each $\Delta \in \Delta_N$, and $\tilde{Q}_n(\Delta) \xrightarrow{p} +\infty$ with $\tilde{Q}_\infty(\Delta, k) = +\infty$ for each $\Delta \in \Delta_P$.

(I) Pick any $\Delta \in \Delta_N$. The map $(u, w) \mapsto K(w)\ell_\delta(u, w; \Delta)$ is continuous on $\mathbb{S} = [-\infty, 0) \times \mathbb{R}^d$ and vanishes outside the set $S = [-\infty, \max(\kappa, -\delta)] \times [-1, 1]^d$, where $\kappa < 0$ on Δ_N. Note that S is compact since $\kappa < 0$ if $\Delta \in \Delta_N$. Thus, the continuous mapping theorem combined with $\hat{N} \xrightarrow{d} N$ (Proposition 1) yields $\tilde{Q}_n(\Delta) \xrightarrow{d} \tilde{Q}_\infty(\Delta, k)$.

(II) Now pick $\Delta \in \Delta_P$. Note that $\ell_\delta(u, w; \Delta) = \min\{\tilde{w}'\Delta - u, 0\} \geq 0$ for any $u \geq -\delta$. Hence, for any $u \geq -\delta$ and $\epsilon > 0$, it holds

$$\ell_\delta(u, w; \Delta) = \mathbb{I}\{-\delta \leq u \leq \tilde{w}'\Delta\}(\tilde{w}'\Delta - u) \geq \mathbb{I}\{-\delta \leq u \leq 0, \tilde{w}'\Delta \geq \epsilon\}\epsilon. \tag{A.6}$$

This implies

$$\tilde{Q}_n(\Delta) \geq -Q_{1n}(\Delta) + V_{1,n} + V_{2,n},$$

where

$$\begin{aligned} V_{1,n} &:= \sum_{i=1}^n K(\delta_n^{-1}(X_i - c))\ell_\delta(\mathfrak{a}_n U_{n,i}, \tilde{X}_{n,i}; \Delta)\mathbb{I}\{\mathfrak{a}_n U_{n,i} \leq -\delta\}, \\ V_{2,n} &:= \sum_{i=1}^n K(\delta_n^{-1}(X_i - c))\mathbb{I}\{-\delta/\mathfrak{a}_n \leq U_{n,i} \leq 0, \tilde{X}'_{n,i}\Delta \geq \epsilon\}\epsilon. \end{aligned}$$

Observe that $V_{1,n} = O_p(1)$ by the argument in (I). For $V_{2,n}$, note that for each $\epsilon > 0$,

$$\begin{aligned} & \mathbb{P}(-\delta/\mathfrak{a}_n \leq U_{n,1} \leq 0, \tilde{X}'_{n,1}\Delta \geq \epsilon, \delta_n^{-1}(X_1 - c) \in [-1, 1]^d) \\ = & \int \mathbb{I}\left\{-\delta/\mathfrak{a}_n \leq u + R_\varphi(x, \delta_n) \leq 0, (1, \delta_n^{-1}(x - c)')\Delta \geq \epsilon, \delta_n^{-1}(x - c) \in [-1, 1]^d\right\} dF_U(u|x)f_X(x)dx \\ = & \delta_n^d \int \mathbb{I}\{-\delta/\mathfrak{a}_n \leq u + R_\varphi(c + \delta_n w, \delta_n) \leq 0\} \\ & \times\mathbb{I}\{\tilde{w}'\Delta \geq \epsilon, w \in [-1, 1]^d\}dF_U(u|c + \delta_n w)f_X(c + \delta_n w)dw \\ \gtrsim & \delta_n^d \int \mathbb{I}\{-\delta/\mathfrak{a}_n + \delta_n^{1+\gamma} \leq u \leq -\delta_n^{1+\gamma}, \tilde{w}'\Delta \geq \epsilon, w \in [-1, 1]^d\}dF_U(u|c + \delta_n w)dw \\ \gtrsim & \delta_n^d, \end{aligned}$$

where the second equality follows from the change of variables, the first inequality follows from (1.4) and $\inf_{x\in\mathbb{B}} f_X(x) > 0$ (by Assumption 1 (i) and (iv)), and the second inequality follows from $\inf_{x\in\mathbb{B}} \mathbb{P}(U \leq 0|X = x) > 0$ (by Assumption 1 (ii)). Therefore, $V_{2,n} \gtrsim O_p(n\delta_n^d) \xrightarrow{p} +\infty$ in $\bar{\mathbb{R}}$. Combining these results, we obtain $\tilde{Q}_n(\Delta) \xrightarrow{p} +\infty$ for any $\Delta \in \Delta_P$. Therefore, Condition (b) is satisfied when $\xi > 0$.

Step 3: Alternative representation of $Q_\infty(\Delta, k)$ (2nd equality in (1.9))

From Resnick (1987, Proposition 3.8), the point process defined by $\{\mathcal{G}_i, \mathcal{W}_i\}$ corresponds to the Poisson point process with mean measure $\tilde{m}(du, dw) = du \times 2^{-d}dw$ on

$$\tilde{\mathbb{S}} = \begin{cases} [-\infty, \infty) \times [-1, 1]^d & \text{if } \xi = 0, \\ [-\infty, 0) \times [-1, 1]^d & \text{if } \xi > 0, \\ [0, \infty) \times [-1, 1]^d & \text{if } \xi < 0. \end{cases}$$

Now consider the mapping $J : \tilde{\mathbb{S}} \to \tilde{\mathbb{S}}$ defined by

$$(u, w) \mapsto \begin{cases} \left(\log\left(\frac{u}{2^d\Gamma(c)f_X(c)}\right), w\right) & \text{if } \xi = 0, \\ \left(-\left(\frac{u}{2^d\Gamma(c)f_X(c)}\right)^{-\xi}, w\right) & \text{if } \xi > 0, \\ \left(\left(\frac{u}{2^d\Gamma(c)f_X(c)}\right)^{-\xi}, w\right) & \text{if } \xi < 0. \end{cases}$$

Then from Resnick (1987, Proposition 3.7), the point process defined by $\{J(\mathcal{G}_i, \mathcal{W}_i)\}$ corresponds to the Poisson point process with mean measure

$$\tilde{m}(J^{-1}(du, dw)) = \begin{cases} 2^d\Gamma(c)f_X(c) \times e^u du \times 2^{-d}dw & \text{if } \xi = 0, \\ 2^d\Gamma(c)f_X(c) \times \left(\frac{1}{\xi}(-u)^{-1/\xi-1}\right) du \times 2^{-d}dw & \text{if } \xi > 0, \\ 2^d\Gamma(c)f_X(c) \times \left(-\frac{1}{\xi}u^{-1/\xi-1}\right) du \times 2^{-d}dw & \text{if } \xi < 0. \end{cases}$$

This implies that $\tilde{m}(J^{-1}(\cdot)) = m(\cdot)$ on $\sigma(\tilde{\mathbb{S}})$. Recall that the kernel function K is compactly supported on $[-1, 1]^d$. Then we can restrict the state space $\mathbb{S}$ of N on $\tilde{\mathbb{S}}$. Therefore, $Q_\infty(\Delta, k)$ can be represented as

$$Q_\infty(\Delta, k) = -kf_X(c)\left\{\int_{[-1,1]^d} K(w)\tilde{w}dw\right\}' \Delta - \sum_{i=1}^{\infty} K(\mathcal{W}_i) \min\{\mathcal{J}_i - \tilde{\mathcal{W}}_i'\Delta, 0\}.$$

A.3 Proof of Theorem 2

A.3.1 Proof of (1.11)

Note that $\theta_{\alpha_n}(x) = F_Y^{-1}(\alpha_n|x) = \varphi(x) + F_U^{-1}(\alpha_n|x)$ by Assumption 1 (ii). When $\xi \neq 0$, Assumption 1 (ii)-(iii) imply

$$\begin{aligned}
&\mathfrak{a}_n(\theta_{\alpha_n}(c) - \varphi(c)) = \mathfrak{a}_n F_U^{-1}(\alpha_n|c) \\
= \quad & -\mathrm{sgn}(\xi)\cdot(\Gamma(c)^{\xi} + o(1))\frac{F_{U_*}^{-1}(\alpha_n)}{F_{U_*}^{-1}(1/(n\delta_n^d))} \to -\mathrm{sgn}(\xi)\cdot k^{-\xi}\Gamma(c)^{\xi}. \qquad (A.7)
\end{aligned}$$

When $\xi = 0$, we can similarly show that

$$\mathfrak{a}_n(\theta_{\alpha_n}(c) - \varphi(c) - \mathfrak{b}_n) \to -\log\Gamma(c) + \log k. \qquad (A.8)$$

Indeed, similarly to Step 1 in the proof of Chernozhukov (2005, Lemma 9.1), we can show that for $m \in (0,1)\cup(1,\infty)$,

$$\frac{F_U^{-1}(\alpha_n|c) - F_{U_*}^{-1}(\alpha_n)}{F_{U_*}^{-1}(m\alpha_n) - F_{U_*}^{-1}(\alpha_n)} \to \frac{\log(1/\Gamma(c))}{\log m}. \qquad (A.9)$$

Furthermore, the following result is well known in extreme value theory (cf. de Haan (1984) or Chapters 1 and 2 in Resnick (1987)): When $\xi = 0$, for $m, \ell \in (0,\infty)$,

$$\frac{F_{U_*}^{-1}(\ell m\tau) - F_{U_*}^{-1}(\ell\tau)}{a(F_{U_*}^{-1}(\tau))} \to \log m, \quad \text{as } \tau \downarrow 0, \qquad (A.10)$$

where $a(\cdot)$ is the auxiliary function defined in Assumption 1 (ii) (see also Lemma 9.2 (iv) and the proof of Chernozhukov (2005, Lemma 9.1)). Therefore, (A.9) and (A.10) yield (A.8) as follows:

$$\begin{aligned}
&\mathfrak{a}_n(\theta_{\alpha_n}(c) - \varphi(c) - \mathfrak{b}_n) = \frac{F_U^{-1}(\alpha_n|c) - F_{U_*}^{-1}(1/n\delta_n^d)}{a(F_{U_*}^{-1}(1/n\delta_n^d))} \\
\sim \quad & \frac{F_U^{-1}(k/n\delta_n^d|c) - F_{U_*}^{-1}(k/n\delta_n^d)}{a(F_{U_*}^{-1}(1/n\delta_n^d))} + \frac{F_{U_*}^{-1}(\alpha_n) - F_{U_*}^{-1}(1/n\delta_n^d)}{a(F_{U_*}^{-1}(1/n\delta_n^d))} \\
\sim \quad & \frac{F_{U_*}^{-1}(ek/n\delta_n) - F_{U_*}^{-1}(k/n\delta_n^d)}{a(F_{U_*}^{-1}(1/n\delta_n^d))}\cdot\frac{\log(1/\Gamma(c))}{\log e} + \frac{F_{U_*}^{-1}(k/n\delta_n^d) - F_{U_*}^{-1}(1/n\delta_n^d)}{a(F_{U_*}^{-1}(1/n\delta_n^d))} \\
\to \quad & \log e\cdot\frac{-\log\Gamma(c)}{\log e} + \log k = -\log\Gamma(c) + \log k. \qquad (A.11)
\end{aligned}$$

Theorem 1 in the supplement implies

$$\Delta_n \xrightarrow{d} \Delta_\infty(k). \qquad (A.12)$$

Therefore, (1.11) is obtained as

$$\begin{aligned} & \mathfrak{a}_n(\hat{\theta}_{\alpha_n}(c) - \theta_{\alpha_n}(c)) \\ = \; & \mathfrak{a}_n(\hat{\theta}_{\alpha_n}(c) - \varphi(c) - \mathfrak{b}_n) - \mathfrak{a}_n(\theta_{\alpha_n}(c) - \varphi(c) - \mathfrak{b}_n) \\ \stackrel{d}{\to} \; & \Delta_{\infty,0}(k) + g(c;\xi), \end{aligned} \tag{A.13}$$

where the convergence of the first term follows from (A.12) and the convergence of the second term follows from (A.7) and (A.8). Therefore, we obtain the conclusion.

A.3.2 Proof of (1.12)

Define Δ_n^m and $Q_n^m(\Delta)$ by replacing α_n with $m\alpha_n$ in Δ_n and $Q_n(\Delta)$, respectively. A similar argument to the proof of Theorem 1 in the supplement yields the weak convergence of

$$(\Delta_n^m, \Delta_n) \in \arg \min_{(\Delta^{m\prime}, \Delta^{\prime})^{\prime} \in \mathbb{R}^{2(d+1)}} \{Q_n^m(\Delta^m) + Q_n(\Delta)\},$$

to the limiting distribution

$$(\Delta_\infty(mk), \Delta_\infty(k)) = \arg \min_{(\Delta^{m\prime}, \Delta^{\prime})^{\prime} \in \mathbb{R}^{2(d+1)}} \{Q_\infty(\Delta^m, mk) + Q_\infty(\Delta, k)\}. \tag{A.14}$$

Here the random vectors $\Delta_\infty(mk)$ and $\Delta_\infty(k)$ are uniquely determined since the objective function $Q_n^m(\Delta^m) + Q_n(\Delta)$ is a sum of objective functions in the proof of Theorem 1. From (A.14), the continuous mapping theorem yields

$$\begin{aligned} & \mathfrak{a}_n(\hat{\theta}_{m\alpha_n}(c) - \hat{\theta}_{\alpha_n}(c)) \\ = \; & \mathfrak{a}_n\{(\hat{\theta}_{m\alpha_n}(c) - \varphi(c) - \mathfrak{b}_n) - (\hat{\theta}_{\alpha_n}(c) - \varphi(c) - \mathfrak{b}_n)\} \\ \stackrel{d}{\to} \; & \Delta_{\infty,0}(mk) - \Delta_{\infty,0}(k). \end{aligned} \tag{A.15}$$

By (A.13) and (A.15), we obtain the conclusion as

$$\Theta_n = \frac{\mathfrak{a}_n(\hat{\theta}_{\alpha_n}(c) - \theta_{\alpha_n}(c))}{\mathfrak{a}_n(\hat{\theta}_{m\alpha_n}(c) - \hat{\theta}_{\alpha_n}(c))} \stackrel{d}{\to} \frac{\Delta_{\infty,0}(k) + g(c;\xi)}{\Delta_{\infty,0}(mk) - \Delta_{\infty,0}(k)}.$$

A.4 Proof of Theorem 3

The proof is analogous to the ones for Theorems 1 (in the supplement) and 2.

A.5 Proof of Theorem 4

The proof is analogous to the one for Theorem 1 in the main paper.

References

[1] Chernozhukov, V. (2005) Extremal quantile regression, *Annals of Statistics*, 33, 806-839.

[2] Chernozhukov, V. and I. Fernández-Val (2011) Inference for extremal conditional quantile models, with an application to market and birthweight risks, *Review of Economic Studies*, 78, 559-589.

[3] Chernozhukov, V., Fernández-Val, I. and A. Galichon (2010) Quantile and probability curves without crossing, *Econometrica*, 78, 1093-1125.

[4] Daouia, A., Gardes, L. and S. Girard (2013) On kernel smoothing for extremal quantile regression, *Bernoulli*, 19, 2557-2589.

[5] de Haan, L. (1984) Slow variation and characterization of domains of attraction, in *Statistical Extremes and Applications* (I. Tiago de Oliveira, ed.) 31-48. Reidel, Dordrecht.

[6] Geyer, C. J. (1996) On the asymptotics of convex stochastic optimization, Working paper.

[7] Hastie, T. and R. Tibshirani (1993) Varying-coefficient models, *Journal of the Royal Statistical Society*, B 55, 757-779.

[8] Honda, T. (2004) Quantile regression in varying coefficient models, *Journal of Statistical Planning and Inference*, 121, 113-125.

[9] Horowitz, J. L. and S. Lee (2005) Nonparametric estimation of an additive quantile regression model, *Journal of the American Statistical Association*, 100, 1238-1249.

[10] Ichimura, H., Otsu, T. and J. Altonji (2019) Nonparametric intermediate order regression quantiles. Working paper.

[11] Kim, M.-O. (2007) Quantile regression with varying coefficients, *Annals of Statistics*, 35, 92-108.

[12] Knight, K. (1999) Epi-convergence and stochastic equisemicontinuity. Working paper.

[13] Lee, S. (2003) Efficient semiparametric estimation of a partially linear quantile regression model, *Econometric Theory*, 19, 1-31.

[14] Meyer, R. M. (1973) A Poisson-type limit theorem for mixing sequence of dependent “rare” events, *Annals of Probability*, 1, 480-483.

[15] Phillips, P. C. B. (2015) Halbert White Jr. memorial JFEC lecture: Pitfalls and possibilities in predictive regression, *Journal of Financial Econometrics*, 13, 521-555.

[16] Resnick, S. I. (1987) *Extreme Values, Regular Variation, and Point Process*, Springer, New York.

[17] Robinson, P. M. (1988) Root-N-consistent semiparametric regression, *Econometrica*, 56, 931-954.

[18] Takeuchi, I., Le, Q. V., Sears, T. and A. J. Smola (2006) Nonparametric quantile regression, Journal of Machine Learning Research, 7, 1231–1264.

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