Econometric Theory, **0**, 2023, 1–15.
doi:10.1017/S0266466623000336

SUBSAMPLING INFERENCE FOR NONPARAMETRIC EXTREMAL CONDITIONAL QUANTILES

Daisuke Kurisu
The University of Tokyo

Taisuke Otsu
London School of Economics

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.

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 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

Address correspondence to Taisuke Otsu, Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, UK; e-mail: t.otsu@lse.ac.uk

[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

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; 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 Supplementary Material, 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 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; 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 Supplementary Material, 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.

cannot be resolved by finite-sample modifications of the quantile regression estimator, such as the rearrangement method in Chernozhukov, Fernández-Val, and Galichon (2010).

https://doi.org/10.1017/S0266466623000336 Published online by Cambridge University Press

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 let $F_Y(\cdot|\cdot)$ 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 Supplementary Material, 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

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

https://doi.org/10.1017/S0266466623000336 Published online by Cambridge University Press

In order to establish validity of subsampling inference based on 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 Supplementary Material, 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]

Let q_t denote the tth 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 jth subsample, compute a subsample analog 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_bth conditional quantile estimator computed using the full sample, and $\hat{\theta}_{\alpha_b}^{(j)}(c)$ is the α_bth conditional quantile estimator computed using the jth 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 tth 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 $\alpha_b/\alpha_n \to \infty$.*

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

(iii) *There exist a distribution function F_{U_*} with Pareto-type tails of extreme value index $\xi \neq 0$ and a 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 $n \to \infty$,

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

Assumption (i) is a high-level condition on the normalized object Θ_n. See Section 1 of the Supplementary Material 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 Supplementary Material 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, Chap. 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.

https://doi.org/10.1017/S0266466623000336 Published online by Cambridge University Press

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

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

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 Supplementary Material, we verify that this estimator satisfies the assumptions of the main theorem under the primitive conditions below. See Section 1 of the Supplementary Material 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 let $\mathbb{B} \subset \mathbb{R}^d$ 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 $\mathfrak{b}_n = \begin{cases} F_{U_*}^{-1}(1/n\delta_n^d), & \text{for } \xi = 0, \\ 0, & \text{for } \xi \neq 0. \end{cases}$

https://doi.org/10.1017/S0266466623000336 Published online by Cambridge University Press

(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 $x \in \mathbb{B}$, $|D_u\varphi(x) - D_u\varphi(c)| \le C\|x-c\|^\gamma$.

(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 $\mathcal{A}_i = \sigma(U_i, X_i, U_{i-1}, X_{i-1}, \ldots)$ and $\mathcal{B}_i = \sigma(U_i, X_i, U_{i+1}, X_{i+1}, \ldots)$. 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 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,

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

(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 correspond 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 $K(w) = \frac{15}{16}(1-w^2)^2\mathbb{I}\{|w| \le 1\}$. 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 ($t \in (0,1/2)$)

https://doi.org/10.1017/S0266466623000336 Published online by Cambridge University Press

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 $n = 2,000$ and $b = 500$ for $n = 5,000$. The numbers in the parentheses are means of bandwidths selected by using LOOCV.

		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
2,000	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)	
5,000	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)	

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 $n \in \{2,000, 5,000\}$ and set $b = 200$ (for $n = 2,000$) and $b = 500$ (for $n = 5,000$). 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 Supplementary Material. 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.

Table 2 presents empirical coverage probabilities of 90% ($t = 0.1$) and 95% ($t = 0.05$) confidence intervals $C_{1-t}(\alpha_n)$ with $n = 2,000$ 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 $1/25 \le b/n \le 1/10$. 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.

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 $n = 2,000$ and $b \in \{80, 120, 160, 200, 300, 400, 500\}$.

		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

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, Th m. 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 tth 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

https://doi.org/10.1017/S0266466623000336 Published online by Cambridge University Press

TABLE 3. Empirical coverage probabilities of $C_{1-t}(\alpha_n)$, $C^N_{1-t}(\alpha_n)$, and $C^E_{1-t}(\alpha_n)$ for $\theta_{\alpha_n}(c) = F_Y^{-1}(\alpha_n|c)$ at $c = -0.5$.

		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
2,000	0.01	$C_{1-t}(\alpha_n)$	0.848	0.920	0.860	0.928	0.856	0.928	0.876	0.936
		$C^N_{1-t}(\alpha_n)$	0.044	0.052	0.204	0.240	0.716	0.776	0.656	0.748
		$C^E_{1-t}(\alpha_n)$	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
		$C^N_{1-t}(\alpha_n)$	0.032	0.040	0.128	0.152	0.628	0.684	0.668	0.708
		$C^E_{1-t}(\alpha_n)$	0.256	0.304	0.408	0.496	0.768	0.816	0.740	0.812

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 equation (1.13) of the Supplementary Material, 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 equation (1.13) of the Supplementary Material does not involve any preliminary estimation as in Daouia, Gardes, and Girard (2013). In this case, as in Daouia, Gardes, and Girard (2013, Th m. 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 $n = 2{,}000$ and $\alpha_n \in \{0.01, 0.005\}$. Although we do not report here, the results are similar for the case of $n = 5{,}000$. 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 $K_\xi(\tilde{\alpha}_n/\alpha_n)\hat{a}(c)$.

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 22,

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

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]

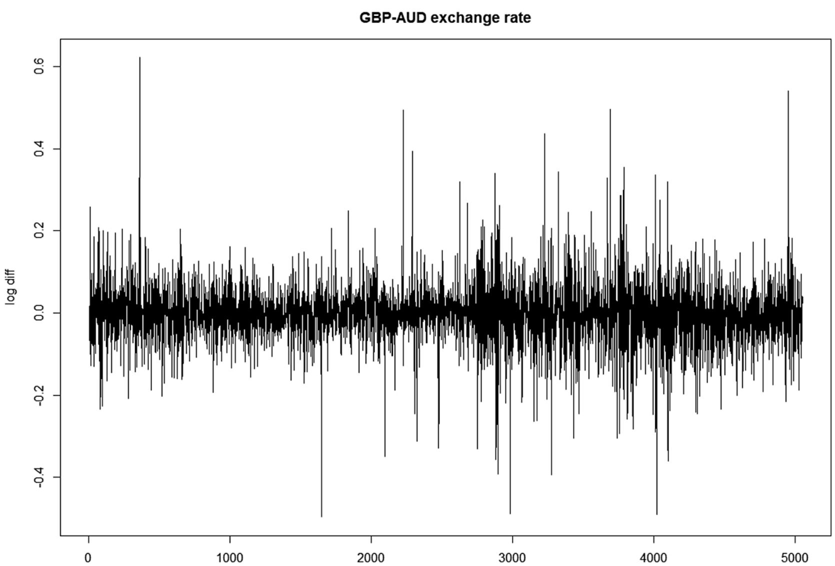

FIGURE 1. Plots of the transformed GBP-AUD exchange rate $\{Y_i\}_{i=1}^{5,053}$.

https://doi.org/10.1017/S0266466623000336 Published online by Cambridge University Press

2006 to August 30, 2008 ($n = 5,053$) 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 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 1.

5. ADDITIONAL RESULTS IN THE SUPPLEMENTARY MATERIAL

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 Supplementary Material, 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 Supplementary Material, 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).

https://doi.org/10.1017/S0266466623000336 Published online by Cambridge University Press

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

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.

APPENDIX

A. PROOF OF THEOREM 1

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

$$\hat{G}_n(x) = \frac{1}{B_n}\sum_{j=1}^{B_n} \mathbb{I}\{\hat{\Theta}_b^{(j)} \le x\} = \frac{1}{B_n}\sum_{j=1}^{B_n} \mathbb{I}\{\Theta_b^{(j)} + \tilde{A}_b^{(j)}(\theta_{\alpha_b}(c) - \hat{\theta}_{\alpha_b}(c)) \le x\},$$

$$\tilde{G}_n(x;\Delta) = \frac{1}{B_n}\sum_{j=1}^{B_n} \mathbb{I}\{\Theta_b^{(j)} + (\tilde{A}_b^{(j)}/A_b)\Delta \le x\}.$$

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 $j = 1, \ldots, B_n$, 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:

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

On Ω_n, it holds

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

https://doi.org/10.1017/S0266466623000336 Published online by Cambridge University Press

We next show that at the continuity points of $G(x) = \mathbb{P}(\Theta_\infty \le x)$, it holds $\tilde{G}_n(x; \pm\epsilon_n) \stackrel{p}{\to} G(x)$. 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 $\Theta_b \stackrel{d}{\to} \Theta_\infty$ (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 $\mathrm{Var}(\tilde{G}_n(x;\epsilon_n)) = o(1)$. This shows that $\tilde{G}_n(x;\epsilon_n) \stackrel{p}{\to} G(x)$. Likewise, we obtain $\tilde{G}_n(x;-\epsilon_n) \stackrel{p}{\to} G(x)$.

Finally, since $\mathbb{P}(\Omega_n) \to 1$, (A.1) yields $\hat{G}_n(x) \stackrel{p}{\to} G(x)$ 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 $\hat{q}_t = \hat{G}_n^{-1}(t) \stackrel{p}{\to} G^{-1}(t) = q_t$, provided $G^{-1}(t)$ is a continuity point of $G(x)$.

Supplementary Material

Kurisu, D. and Otsu, T. (2023): Supplement to "Subsampling inference for nonparametric extremal conditional quantiles," Econometric Theory Supplementary Material. To view, please visit: https://doi.org/10.1017/S0266466623000336

References

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

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

Bickel, P.J. & D.A. Freedman (1981) Some asymptotic theory for the bootstrap. *Annals of Statistics* 9, 1196–1217.

Chaudhuri, P. (1991) Nonparametric estimates of regression quantiles and their local bahadur representation. *Annals of Statistics* 19, 760–777.

Chernozhukov, V. (1998) Nonparametric Extreme Regression Quantiles. Massachusetts Institute of Technology. Working paper.

Chernozhukov, V. (2005) Extremal quantile regression. *Annals of Statistics* 33, 806–839.

Chernozhukov, V. & 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.

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

Daouia, A., L. Gardes, & S. Girard (2013) On kernel smoothing for extremal quantile regression. *Bernoulli* 19, 2557–2589.

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

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

He, F., Y. Cheng, & T. Tong (2016) Estimation of extreme conditional quantiles through an extrapolation of intermediate regression quantiles. *Statistics & Probability Letters* 113, 30–37.
Ichimura, H., Otsu, T. and J. Altonji (2019) Nonparametric Intermediate Order Regression Quantiles. Yale University. Working paper.
Koenker, R. & G. Bassett (1978) Regression quantiles. *Econometrica* 46, 33–50.
Phillips, P.C.B. (2015) Halbert White Jr. memorial JFEC lecture: Pitfalls and possibilities in predictive regression. *Journal of Financial Econometrics* 13, 521–555.
Politis, D.N., J.P. Romano, & M. Wolf (1999) *Subsampling*. Springer.
Resnick, S.I. (1987) *Extreme Values, Regular Variation, and Point Process*. Springer.
Resnick, S.I. (2007) *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer.
Takeuchi, I., Q.V. Le, T. Sears, & A.J. Smola (2006) Nonparametric quantile regression. *Journal of Machine Learning Research* 7, 1231–1264.
Wang, H.J., D. Li, & X. He (2012) Estimation of high conditional quantiles for heavy-tailed distributions. *Journal of American Statistical Association* 107, 1453–1464.
Yu, K. & M.C. Jones (1998) Local linear quantile regression. *Journal of the American Statistical Association* 93, 228–237.
Zhang, Y. (2018) Extremal quantile treatment effects. *Annals of Statistics* 46, 3707–3740.