NONPARAMETRIC INFERENCE FOR EXTREMAL CONDITIONAL QUANTILES

DAISUKE KURISU AND TAISUKE OTSU

Abstract. This paper studies asymptotic properties of the local linear quantile estimator under the extremal order quantile asymptotics, and develops a practical inference method for conditional quantiles in extreme tail areas. By using a point process technique, the asymptotic distribution of the local linear quantile estimator is derived as a minimizer of certain functional of a Poisson point process that involves nuisance parameters. To circumvent difficulty of estimating those nuisance parameters, we propose a subsampling inference method for conditional extreme quantiles based on a self-normalized version of the local linear estimator. A simulation study illustrates usefulness of our subsampling inference to investigate extremal phenomena.

Keywords: quantile regression; extreme value theory; point process; subsampling

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 than the conventional inference methods based on the fixed quantile asymptotics in extreme tails.

One major limitation of these studies on the extremal quantile regression model is that the quantile regression function must be parametrically specified. For point estimation, 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

Kurisu acknowledges financial support by JSPS KAKENHI Grant Numbers JP17H02513, JP19K20881 and JP20K13468.

tails. The purpose of this paper is to fill this gap by (i) studying asymptotic properties of the local linear quantile estimator under the extremal order quantile asymptotics, and (ii) developing a practical inference method for 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, bandwidth δ_n for local linear regression, and number of covariates d. Then we derive the limiting distribution of the local linear quantile regression estimator and its self-normalized version under this asymptotic framework. We find that the limiting distribution of the the local linear estimator involves nuisance parameters: the density of covariates, extreme value index for an auxiliary error term, and an approximately multiplicative component in this error term. Particularly, it is not trivial how to estimate the multiplicative component, so inference based on the estimated limiting distribution is difficult in the nonparametric setup. Therefore, based on the self-normalized statistic, we propose a subsampling inference method which completely avoids estimation of the nuisance parameters in the limiting distribution. This result can be considered as an extension of subsampling inference by Chernozhukov and Fernández-Val (2011) to the nonparametric setup.

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. Since seminal works by Resnick (1986, 1987) on mathematical characterizations of heavy-tailed phenomena, asymptotic theory of point processes has been one of the important tools for statistical analysis of extremal events in the literature of statistical extreme value theory. See Embrechts, Klüppelberg and Mikosch (1997) and Resnick (2007) for reviews on applications of point process theory to insurance, finance, and data networks. We also refer to Davis, Mikosch and Pfaffel (2016), Heiny and Mikosch (2017, 2019) as recent contributions to heavy-tailed (and high-dimensional) time series analysis, and Zhang (2018) on inference for quantile treatment effects under the extremal order quantile asymptotics. Main results in this paper are built upon weak convergence of a point process to a Poisson point process, which enables us to develop nonparametric inference of extremal conditional quantiles.

Another approach, which we call the extrapolation approach, to estimate extremal (conditional) quantiles has been studied in the literature. Dekkers and de Haan (1989) proposed estimators of "very" extreme α_n-quantiles by estimating "less" extreme $\widetilde{\alpha}_n$-quantile such that $\alpha_n < \widetilde{\alpha}_n \to 0$ (or $\alpha_n > \widetilde{\alpha}_n \to 1$) as $n \to \infty$ and extrapolating the estimator to estimate α_n-quantile. Wang, Li and He (2012) and He, Cheng and Tong (2016) proposed closely related estimators for linear quantile regression and Daouia, Gardes and Girard (2013) investigated a similar extrapolation estimator for extremal conditional quantiles based on kernel smoothing. We note that these extrapolation approaches allow the case of $n\delta_n^d\alpha_n \to 0$, which is excluded in our analysis. However, since their idea is based on the extrapolation of the quantile function depending on the extreme value index ξ, which is a parameter to describe heavy-tailedness, they need to estimate ξ and set a less extreme quantile level $\widetilde{\alpha}_n$ to implement their estimator. On the other hand, computation of our estimator does not need to set $\widetilde{\alpha}_n$ or estimate any nuisance parameters including ξ. This can be achieved by (i) considering a self-normalized version of the local linear estimator and (ii) developing a subsampling method for inference of extremal conditional quantiles. Therefore, our point process approach can be seen as an alternative to the extrapolation approach.

This paper is organized as follows. In Section 2, we introduce the setup and local linear quantile regression estimator (Section 2.1), derive the asymptotic distribution of the estimator under the extremal order quantile regression (Section 2.2), and propose a subsampling inference method (Section 2.3). Section 3 discusses extensions of our results for the case where the extreme value index varies with covariates (Section 3.1) and varying coefficient models (Section 3.2). In Section 4, we conduct a simulation study. All proofs are contained in Appendix.

Notation. Hereafter we use the following notation. For random variables $(Y, X) \in \mathbb{R} \times \mathbb{R}^d$, let $F_Y(y|x_r)$ be the conditional distribution function of Y given $X = x_r = (x_{r,1}, \dots x_{r,d})' \in \mathbb{R}^d$ and $\theta_\alpha(x_r) = \inf_{y \in \mathbb{R}}\{y : F_Y(y|x_r) > \alpha\}$ be the α-th conditional quantile function at x_r. For any function $f : \mathbb{R}^d \to \mathbb{R}$, let $D_u f(x) = \partial f(x)/\partial x_u$ for $u = 1, \dots, d$. $\mathbb{B}(x_r) \subset \mathbb{R}^d$ denotes some fixed closed ball around x_r. For any positive sequences a_n, 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. Let $\|\cdot\|$ be the Euclidean norm.

2. Main results

2.1. **Estimator and setup.** We first define the local nonparametric quantile regression estimator for the conditional quantile function $\theta_\alpha(x_r)$ at x_r. We tentatively assume that the quantile point $\alpha \in (0, 1)$ is fixed. Let $K : \mathbb{R}^d \to \mathbb{R}$ be a kernel function. Let $\beta = (\beta_0, \dots, \beta_d)'$ and $x(x_r, \delta_n) = (1, (x_1 - x_{r,1})/\delta_n, \dots, (x_d - x_{r,d})/\delta_n)'$ for $x = (x_1, \dots, x_d)' \in \mathbb{R}^d$. Based on a sample $\{Y_j, X_j\}_{j=1}^n$, we consider the following (weighted) quantile regression estimator:

$$\widehat{\beta}_n^{(\alpha)} = \arg\min_{\beta \in \mathbb{R}^{d+1}} \sum_{j=1}^n K\left(\frac{X_j - x_r}{\delta_n}\right) \rho_\alpha(Y_j - X_j(x_r, \delta_n)'\beta), \tag{2.1}$$

where $\rho_\alpha(v) = (\alpha - \mathbb{I}\{v \leq 0\})v$ is called the check function, $\mathbb{I}\{\cdot\}$ is the indicator function and δ_n be a sequence of bandwidth parameters satisfying $\delta_n \to 0$ as $n \to \infty$. This minimization problem can be considered as a natural extension of the local linear (mean) regression problem to the quantile regression setup (see, e.g., Section 5.5.2 of Fan and Gijbels, 1996). Chaudhuri (1991) set the kernel function K as the uniform kernel. Chernozhukov (2005, p. 808) discussed localized linear regression without the kernel component. For nonparametric estimation of the conditional quantile $\theta_\alpha(x_r)$, we typically need either smoothing or growing series approximations, and this paper focuses on the smoothing approach.

Based on $\widehat{\beta}_n^{(\alpha)} = (\widehat{\beta}_{n,0}^{(\alpha)}, \ldots, \widehat{\beta}_{n,d}^{(\alpha)})'$ obtained above, our estimator for the α-th conditional quantile $\theta_\alpha(x_r)$ of $Y|X = x_r$ is given by its intercept, i.e., $\widehat{\theta}_\alpha(x_r) = \widehat{\beta}_{n,0}^{(\alpha)}$. Chaudhuri (1991) considered the case of the uniform kernel function and explored the asymptotic properties of the local and global nonparametric quantile regression estimators, respectively, when the quantile $\alpha \in (0,1)$ is fixed. Chernozhukov (1998) investigated the asymptotic properties of those estimators under the extreme order quantile asymptotics, where the quantile α depends on the sample size and satisfies $\alpha = \alpha_n \to 0$ and $n\delta_n^d\alpha_n \to 0$ as $n \to \infty$. Alternatively, this paper considers the extremal order quantile asymptotics, where the quantile α_n satisfies $\alpha_n \to 0$ and

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

Define $\beta_n^{(\alpha_n)} = (\beta_{n,0}^{(\alpha_n)}, \ldots, \beta_{n,d}^{(\alpha_n)})' = (\theta_{\alpha_n}(x_r), \delta_n D_1\theta_{\alpha_n}(x_r), \ldots \delta_n D_d\theta_{\alpha_n}(x_r))'$ and $\mathbb{N}(x_r, \delta_n) = \prod_{j=1}^d [x_{r,j} - \delta_n, x_{r,j} + \delta_n]$. We impose the following conditions.

Assumption 2.1.

(i) $\{Y_j, X_j\}_{j=1}^n$ *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* $\mathbb{B}(x_r)$*.*

(ii) *There exist a random variable* U_* *with distribution function* F_{U_*} *and a measurable function* $\varphi : \mathbb{B}(x_r) \to \mathbb{R}$ *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* $x \in \mathbb{B}(x_r)$ *for some positive continuous function* $\Gamma(x)$ *on* $\mathbb{B}(x_r)$*. 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* $a_n\delta_n^{1+\gamma} \to 0$ *as* $n \to \infty$*, where*
   (1) $a_n = 1/a(F_{U_*}^{-1}(1/n\delta_n^d))$ *and* $b_n = F_{U_*}^{-1}(1/n\delta_n^d)$ *when* $\xi = 0$*,*
   (2) $a_n = -1/F_{U_*}^{-1}(1/n\delta_n^d)$ *and* $b_n = 0$ *when* $\xi > 0$*,*
   (3) $a_n = 1/F_{U_*}^{-1}(1/n\delta_n^d)$ *and* $b_n = 0$ *when* $\xi < 0$*.*

(iv) *For each* $u = 1, \ldots, d$*,* $D_u\varphi(x)$ *exists at each* $x \in \mathbb{B}(x_r)$*, and there exist constants* $c_1 \in (0,\infty)$ *and* $\gamma \in (0,1]$ *such that* $D_u\varphi(x)$ *is* γ*-Hölder continuous on* $\mathbb{B}(x_r)$*, i.e., at each* $x \in \mathbb{B}(x_r)$*,* $|D_u\varphi(x) - D_u\varphi(x_r)| \le c_1\|x - x_r\|^\gamma$ *.*

(v) *For all n large enough, $D_u\theta_{\alpha_n}(x)$ exists and is continuous at each $x \in \mathbb{B}(x_r)$ and $u = 1, \dots, d$, and* $\sup_{x\in\mathbb{N}(x_r,\delta_n)} a_n|\theta_{\alpha_n}(x_r) - x(x_r,\delta_n)'\beta_n^{(\alpha_n)}| \to 0$ *as $n \to \infty$.*

Assumption 2.1 (i) is on smoothness of the density f_X for X. Assumption 2.1 (ii) introduces an auxiliary function φ, which 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 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 2.1 (ii) also requires that for any $x_1, x_2 \in \mathbb{B}(x_r)$, $z \mapsto F_U(z|x_1)$ 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 2.1 (ii) is satisfied for location-scale models. See also the comments after Assumption 2.2 below. The absolute value of ξ measures the 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 2.1 (iii) is concerned with the canonical normalization of $\widehat{\theta}_{\alpha_n}(x_r) - \theta_{\alpha_n}(x_r)$. 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 $a_n = \lambda^{-1}$ and $b_n = \lambda^{-1}(\log 2 - \log(n\delta_n^d))$. 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 $a_n = ((n\delta_n^d)^\xi - 1)^{-1}$. 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 $a_n \sim \beta^{-1}(n\delta_n^d)^{-\xi}$.

Assumption 2.1 (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 $x = x_r$ yields

$$\varphi(x) = x(x_r,\delta_n)'\beta_{\varphi,n} + R_\varphi(x,\delta_n), \quad \theta_{\alpha_n}(x) = x(x_r,\delta_n)'\beta_n^{(\alpha_n)} + R(x,\delta_n), \tag{2.3}$$

where $\beta_{\varphi,n} = (\beta_{\varphi,n,0}, \dots, \beta_{\varphi,n,d})' = (\varphi(x_r), \delta_n D_1\varphi(x_r), \dots, \delta_n D_d\varphi(x_r))'$ and Assumption 2.1 (iv) guarantees

$$\sup_{x\in\mathbb{N}(x_r,\delta_n)} |R_\varphi(x,\delta_n)| = O(\delta_n^{1+\gamma}). \tag{2.4}$$

Assumption 2.1 (v) says that the remainder of the Taylor expansion of $\theta_{\alpha_n}(x)$ around $x = x_r$ should be smaller order than a_n^{-1}, i.e.,

$$\sup_{x\in\mathbb{N}(x_r,\delta_n)} a_n|R(x,\delta_n)| = o(1). \tag{2.5}$$

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

We also assume the following dependence structure on $\{U_j, X_j\}$.

Assumption 2.2. *The sequence* $\{W_j\}_{j=1}^n$ *with* $W_j = (U_j, X_j)$ *and* $U_j = Y_i - \varphi(X_j)$ *defined in Assumption 2.1 (ii) forms a stationary and strongly mixing process with a geometric mixing rate, that is, for some* $C_1 > 0$,

$$\sup_j \sup_{A \in \mathcal{A}_j, B \in \mathcal{B}_{j+m}} |P(A \cap B) - P(A)P(B)| \exp(C_1 m) \to 0 \quad \text{as } m \to \infty,$$

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

Assumption 2.2 includes the case that the sequence of variables $\{U_j, X_j\}_{j=1}^n$, or equivalently $\{Y_j, X_j\}_{j=1}^n$, is a sequence of i.i.d. random variables. The mixing assumption on $\{U_j, X_j\}_{j=1}^n$ is equivalent to the one on $\{Y_j, X_j\}_{j=1}^n$. The non-clustering assumption is used to apply Meyer's (1973) theorem in (A.4) to establish the weak convergence of the point process (2.8) defined below.

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

$$Y_j = \varphi(X_j) + \gamma(X_j)U_{*,j}. \tag{2.6}$$

In this case, Assumption 2.1 (ii) is satisfied with $\Gamma(x) = \gamma(x)^{1/\xi}$. Also, Assumption 2.1 (v) is satisfied if $D_u\gamma(x)$ exists and $D_u\gamma(x)$ is γ-Hölder continuous at each $x \in \mathbb{B}(x_r)$ and $u = 1, \ldots, d$ since

$$a_n\theta_{\alpha_n}(x) = a_n\varphi(x) - \text{sgn}(\xi) \cdot \gamma(x)\frac{F_{U_*}^{-1}(\alpha_n)}{F_{U_*}^{-1}(1/n\delta_n^d)} \sim a_n\varphi(x) - \text{sgn}(\xi) \cdot k^{-\xi}\gamma(x), \tag{2.7}$$

as $n \to \infty$.

We note that Assumptions 2.1 and 2.2 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 3.

We impose the following conditions for the kernel function.

Assumption 2.3.

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

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 $D_j \in \{1, \ldots, M\}$, our estimator for the α-th coniditonal quantile of $Y|X = x_r, D = m$ can be obtained as in (2.1) by replacing the kernel component "$K\left(\frac{X_j - x_r}{\delta_n}\right)$" with "$K\left(\frac{X_j - x_r}{\delta_n}\right)\mathbb{I}\{D_j = m\}$".

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

2.2. **Asymptotic distribution of estimator.** Let $U_{n,j} = U_j + R_\varphi(X_j, \delta_n) - b_n$. Define $\mathbb{E} = \mathbb{E}_\infty \times \mathbb{R}^d$, where

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

As a preparation for the asymptotic analysis on the conditional quantile estimator $\widehat{\theta}_{\alpha_n}(x_r) = \widehat{\beta}^{(\alpha)}_{n,0}$, we consider the following point process

$$\widehat{N}(\cdot) = \sum_{j=1}^{n} \mathbb{I}\left\{(a_n U_{n,j}, (X_j - x_r)/\delta_n) \in \cdot\right\}, \tag{2.8}$$

as a random element of the metric space $M_p(\mathbb{E})$ of point processes defined on the measurable space $(\mathbb{E}, \sigma(\mathbb{E}))$, where $\sigma(\mathbb{E})$ is the σ-algebra generated by the open sets of $\mathbb{E}$, and the metric space $M_p(\mathbb{E})$ 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 $\xi \neq 0$, $a_n U_{n,j}$ may not be in $\mathbb{E}_\infty$ due to the term $R_\varphi(X_j, \delta_n)$ and therefore we need to restrict the state space of $a_n U_{n,j}$ on $\mathbb{E}_\infty$ in general. However, such a restriction on the state space does not cause any technical problem since $a_n|R_\varphi(x, \delta_n)| = O(a_n \delta_n^{1+\gamma}) = o(1)$ uniformly over $x \in \mathbb{N}(x_r, \delta_n)$ under Assumption 2.1 (iv), and this implies that the restriction is asymptotically negligible.

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

Proposition 2.1 (Weak convergence of $\widehat{N}$)**.** *Under Assumptions 2.1-2.2,* $\widehat{N} \xrightarrow{d} N$ *in* $M_p(\mathbb{E})$*, where* N *is a Poisson point process in* $M_p(\mathbb{E})$ *with mean measure*

$$m(du, dw) = \begin{cases} \Gamma(x_r) f_X(x_r) e^u du dw & \text{if } \xi = 0, \\ \Gamma(x_r) f_X(x_r) \frac{1}{\xi}(-u)^{-1/\xi - 1} du dw & \text{if } \xi > 0, \\ -\Gamma(x_r) f_X(x_r) \frac{1}{\xi} u^{-1/\xi - 1} du dw & \text{if } \xi < 0. \end{cases} \tag{2.9}$$

Remark 2.1. Proposition 2.1 can be established by asymptotic theory of point process and the weak convergence $\widehat{N} \xrightarrow{d} N$ enables us to develop statistical inference on extreme order conditional quantiles. It also should be noted that the limit distribution of $\widehat{\theta}_{\alpha_n}(x_r)$ is not normal (see Theorem

2.2). 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 define $\Delta_n(k) = a_n(\widehat{\beta}_n^{(\alpha_n)} - \beta_{\varphi,n} - b_n\mathbf{e}_1)$, where $\mathbf{e}_1 = (1, 0, \ldots, 0)' \in \mathbb{R}^{d+1}$. Recall that $\beta_{\varphi,n}$ is the coefficient vector in the Taylor expansion (2.3). By using the Poisson point process N in the above proposition, the asymptotic distribution of $\Delta_n(k)$ is obtained as follows.

Theorem 2.1 (Asymptotic distribution of $\widehat{\beta}_n^{(\alpha_n)}$)**.** *Under Assumptions 2.1-2.3, it holds $\Delta_n(k) \stackrel{d}{\to} \Delta_\infty(k)$ provided $\Delta_\infty(k)$ is defined as a random vector in $\mathbb{R}^{d+1}$ which uniquely minimizes the objective function*

$$\begin{aligned} Q_\infty(\Delta, k) &= \operatorname{sgn}(\xi) \cdot k f_X(x_r) \int_{[-1,1]^d} K(w)\widetilde{w}' dw \Delta - \int_{\mathbb{E}} K(w) \min\{u - \widetilde{w}'\Delta, 0\} dN(u, w) \\ &= \operatorname{sgn}(\xi) \cdot k f_X(x_r) \int_{[-1,1]^d} K(w)\widetilde{w}' dw \Delta - \sum_{i=1}^{\infty} K(\mathcal{W}_i) \min\{\mathcal{J}_i - \widetilde{\mathcal{W}}_i'\Delta, 0\}, \quad (2.10) \end{aligned}$$

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

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

Remark 2.2. Theorem 2.1 implies that the limiting distribution may be approximated by

$$\arg\min_{\Delta \in \mathbb{R}^{d+1}} \left\{ \operatorname{sgn}(\xi) \cdot \frac{k f_X(x_r)}{S} \sum_{i=1}^{S} K(\mathcal{W}_i)\widetilde{\mathcal{W}}_i{}'\Delta - \sum_{i=1}^{S} K(\mathcal{W}_i) \min\{\mathcal{J}_i - \widetilde{\mathcal{W}}_i{}'\Delta, 0\} \right\}, \qquad (2.11)$$

for large values of S. In particular when $\xi < 0$, (2.11) is equivalent to

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

and we can simulate the asymptotic distribution of $\Delta_n(k)$ from the weighted quantile regression. However, this simulation requires knowledge of unknown objects, ξ, $f_X(x_r)$, and $\Gamma(x_r)$. For example when $\{Y_j, X_j\}$ is an i.i.d. sample, Daouia, Gardes and Girard (2013) proposes a Pickands type estimator of ξ, which also can be applied to the varying extreme value index where ξ may depend on x_r and the authors show 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 3. The density $f_X(x_r)$ may be estimated by the kernel estimator, for example. On the other hand, it is not clear how to estimate $\Gamma(x_r)$ (defined in Assumption 2.1 (ii)) to implement the simulation based on (2.11). Therefore, we do not pursue such an analytical approach for inference of the conditional quantile $\theta_{\alpha_n}(x_r)$ and we instead propose a subsampling method in the next section which completely avoids estimation of the nuisance components ξ, $f_X(x_r)$, and $\Gamma(x_r)$.

Define $\Delta_\infty(k) = (\Delta_{\infty,0}(k), \ldots, \Delta_{\infty,d}(k))'$. Based on Theorem 2.2, the asymptotic distribution of $\widehat{\theta}_{\alpha_n}(x_r)$ is obtained as follows.

Theorem 2.2 (Asymptotic distribution of $\widehat{\theta}_{\alpha_n}(x_r)$)**.** *Under Assumptions 2.1-2.3, we have that*

$$a_n(\widehat{\theta}_{\alpha_n}(x_r) - \theta_{\alpha_n}(x_r)) \xrightarrow{d} \Delta_{\infty,0}(k) + g(x_r;\xi), \tag{2.12}$$

and

$$\begin{aligned}\Theta_n(k,m) \quad &:= \quad \frac{\widehat{\theta}_{\alpha_n}(x_r) - \theta_{\alpha_n}(x_r)}{\widehat{\theta}_{m\alpha_n}(x_r) - \widehat{\theta}_{\alpha_n}(x_r)} \\ &\xrightarrow{d} \quad \frac{\Delta_{\infty,0}(k) + g(x_r;\xi)}{\Delta_{\infty,0}(mk) - \Delta_{\infty,0}(k)} =: \Theta_\infty(k,m),\end{aligned} \tag{2.13}$$

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 2.3. Theorem 2.2 implies that $\widehat{\theta}_{\alpha_n}(x_r) - \theta_{\alpha_n}(x_r) = O_p(1/a_n)$, where a_n is defined in Assumption 2.1 (iii). Since a_n is unknown in general, we cannot use (2.12) to provide some practical inference tools for $\theta_{\alpha_n}(x_r)$. On the other hand, the weak convergence result in (2.13) is useful for inference on $\theta_{\alpha_n}(x_r)$ since we can compute $\Theta_n(k,m)$, which is a randomly self-normalized version of $\widehat{\theta}_{\alpha_n}(x_r) - \theta_{\alpha_n}(x_r)$, without the knowledge of canonical normalization a_n.

Remark 2.4 (Uniqueness of $\Delta_\infty(k)$ and continuity of $G(x) = P(\Theta_\infty(k,m) \le x)$)**.** Uniqueness of $\Delta_\infty(k)$ is necessary to apply the convexity lemma (Geyer, 1996, and Knight, 1999) to show the weak convergence of $\Delta_n(k)$. Furthermore, we need the continuity of $G(x)$ to show the asymptotic validity of the subsampling method proposed in the next subsection. We can show the uniqueness of $\Delta_\infty(k)$ and continuity of $G(x)$ if $\int_{\mathbb{R}^d} K(w)\widetilde{w}\widetilde{w}'dw$ 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 2.3 (ii), we can show that (a) $\Delta_\infty(k)$ 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 $\Theta_\infty(k,m)$ is a proper random variable with a continuous distribution function.

Remark 2.5 (Choice of the bandwidth δ_n)**.** To implement our point estimator $\widehat{\theta}_{\alpha_n}(x_r)$ in (2.1), 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 $\sum_{j=1}^n K\left(\frac{X_j-x_r}{\delta_n}\right)\rho_{\alpha_n}(Y_j - X_j(x_r,\delta_n)'\widehat{\beta}_{n,-j}^{(\alpha_n)})$ with respect to δ_n, where $\widehat{\beta}_{n,-j}^{(\alpha_n)}$ is obtained as in (2.1) by deleting the j-th observation (Y_j, X_j). However, its theoretical analysis is beyond the scope of this paper.

2.3. **Subsampling inference.** In this subsection we propose a subsampling method for constructing confidence intervals of the conditional quantile $\theta_{\alpha_n}(x_r)$. It should be noted that our method does not need to estimate nuisance parameters ξ, $f_X(x_r)$, and $\Gamma(x_r)$ in the limiting distribution of the point estimator $\widehat{\theta}_{\alpha_n}(x_r)$. To obtain the asymptotic validity of the subsampling method, we assume the following additional conditions on quantile density functions.

Assumption 2.4. *The conditional quantile density function $\partial F_U^{-1}(\tau|x)/\partial\tau$ exists and satisfies the tail equivalence relationship*

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

uniformly over $x\in\mathbb{B}(x_r)$, where $\partial F_{U_*}^{-1}(\tau)/\partial\tau$ is regularly varying at 0 with exponent $\xi+1$, i.e., $\partial F_{U_*}^{-1}(\tau)/\partial\tau \sim L(\tau)\cdot\tau^{-\xi-1}$ for a slowly-varying function $L(\tau)$. Furthermore, assume $\lim_{\tau\downarrow 0}\left|\frac{\partial F_{U_*}^{-1}(\tau)/\partial\tau}{\tau^{-1}F_{U_*}^{-1}(\tau)}\right| \in (0,\infty)$.

Assumption 2.4 is imposed to obtain the convergence rate of the conditional quantile estimator $\widehat{\theta}_{\alpha_b}(x_r)$ with an intermediate order quantile level α_b ($\alpha_b\downarrow 0$ and $\alpha_b n\delta_n^d \sim k\alpha_b/\alpha_n\to\infty$ as $n\to\infty$ and $b=b_n\to\infty$), which is introduced to define the subsample counterpart of $\Theta_n(k,m)$ (see (2.14) below). This assumption excludes the case that U_* has exponentially light tail ($\xi=0$). Light-tailed cases may be approximated by taking the absolute value of ξ small since Assumption 2.1 (ii) implies that the distribution of U_* is in the minimum domain of attraction of the extreme value distribution with shape parameter $\xi\in\mathbb{R}$ (see Theorem 1.1.8 in de Haan and Ferreira, 2006) and the distributions are continuous in ξ including at $\xi=0$.

Let c_τ denote the τ-quantile of $\Theta_\infty(k,m)$. The subsampling approximation for the distribution of $\Theta_n(k,m)$ in (2.13) is obtained as follows.

(Step1) Consider all subsets of the data $\{W_j=(Y_j,X_j)\}$ of size b. If $\{W_j\}$ is a time series data, consider $B_n=n-b+1$ subsets of size b of the form $\{W_i,W_{i+1},\dots,W_{i+b-1}\}$. Then compute the analogue of the self-normalized statistics $\Theta_n(k,m)$, denoted by $\widehat{\Theta}_{i,b,n}$ and defined below for each i-th subsample for $1,\dots,B_n$.

(Step2) Obtain $\widehat{c}_\tau$ as the sample τ-quantile of $\{\widehat{\Theta}_{i,b,n}, i = 1, \ldots, B_n\}$.

In practice, a smaller number B_n of randomly chosen subsets can be used, provided that $B_n \to \infty$ and $n \to \infty$ (see, Section 2.5 of Politis, Romano and Wolf, 1999). Politis, Romano and Wolf (1999) and Bertail *et al.* (2004) provide rules for the choice of subsample size b. For m, we set $m = (d+1)/(n\delta_n^d \alpha_n) + 1 + p = (d+1)/k + 1 + p + o(1)$, where $p > 0$ is a spacing parameter. In each i-th subsample of size b, the analogue of $\Theta_n(k,m)$ is computed by

$$\widehat{\Theta}_{i,b,n} := \frac{\widehat{\theta}^{(i,b,n)}_{\alpha_b}(x_r) - \widehat{\theta}_{\alpha_b}(x_r)}{\widehat{\theta}^{(i,b,n)}_{m\alpha_b}(x_r) - \widehat{\theta}^{(i,b,n)}_{\alpha_b}(x_r)} = \widetilde{a}_{i,b,n}(\widehat{\theta}^{(i,b,n)}_{\alpha_b}(x_r) - \widehat{\theta}_{\alpha_b}(x_r)), \tag{2.14}$$

where $\widehat{\theta}_{\alpha_b}(x_r)$ is the α_b-th quantile regression function computed using the full sample and $\widehat{\theta}^{(i,b,n)}_{\alpha_b}(x_r) = \widehat{\beta}^{(\alpha_b)}_{i,b,0}$ is the α_b-quantile regression coefficient computed using the i-th subsample and bandwidth $\delta_b = (k/b\alpha_b)^{1/d}$, where $k = n\delta_n^d \alpha_n$. We take α_b such that $\alpha_b n \delta_n^d = k\alpha_b/\alpha_n \to \infty$ as $n \to \infty$, so that α_b satisfies the intermediate order quantile asymptotics; see Ichimura, Otsu and Altonji (2019) for detailed discussion on nonparametric intermediate order extremal quantile regression.

Let $a_b = -\text{sgn}(\xi)/F^{-1}_{U_*}(1/b\delta_b^d)$. The asymptotic validity of our subsampling inference is obtained as follows.

Theorem 2.3 (Subsampling of self-normalized statistics)**.** *Let* $\tau \in (0,1)$*. Suppose* $b \to \infty$*,* $\alpha_b \to 0$*,* $\delta_b \to 0$*,* $a_b \delta_b^{1+\gamma} \to \infty$*,* $b/n \to 0$*,* $\alpha_b/\alpha_n \to \infty$*,* $\frac{\log\log n}{n\delta_n^d \alpha_b} \to 0$ *and* $\sup_{x \in \mathbb{N}(x_r,\delta_b)} a_b |\theta_{\alpha_b}(x_r) - x(x_r,\delta_b)' \beta_b^{(\alpha_b)}| \to 0$ *as* $n \to \infty$*. Under Assumptions 2.1-2.4, it holds* $\widehat{c}_\tau \xrightarrow{p} c_\tau$*.*

Remark 2.6. To implement our subsampling inference, we need to choose the quantile point α_n, bandwidth δ_n, and subsample size b. Then other constants are obtained as

$$k = n\alpha_n \delta_n^d, \quad \alpha_b = n\alpha_n/b, \quad \delta_b = (k/b\alpha_b)^{1/d}.$$

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 quantile point α_n is chosen based on researcher's interest. The bandwidth δ_n may be selected by the cross validation method as in Remark 2.5. The size of subsamples b may be chosen by applying the methods in Politis, Romano and Wolf (1999, Chapter 9).

3. Extension

3.1. **Varying extreme value index.** In this section we extend our results for 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 3.1. 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 $y \le 0$ and any $x < 0$. Then the conditional quantile is $\theta_\tau(x) = 1 - \tau^{-|x|} = 1 - \tau^x$. In this case, we cannot apply Theorem 2.1 to estimate $\theta_{\alpha_n}(x_r)$ ($x_r < 0$) since the conditional tail index is $\xi(x) = |x| > 0$ is not constant.

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

Assumption 3.1. (ii') *There exists a measurable function $\varphi : \mathbb{B}(x_r) \to \mathbb{R}$ 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 $x \in \mathbb{B}(x_r)$ for some positive continuous function $\Gamma(x)$ on $\mathbb{B}(x_r)$. 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 $\xi(x) \ne 0$, 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 $\mathbb{B}(x_r)$.*

(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 $\mathbb{B}(x_r)$.*

(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 $a_n \delta_n^{1+\gamma} \to 0$ as $n \to \infty$, where*

(2) *$a_n = -1/F_{U_*}^{-1}(1/n\delta_n^d|x_r)$ and $b_n = 0$ when $\xi(x_r) > 0$,*

(3) *$a_n = 1/F_{U_*}^{-1}(1/n\delta_n^d|x_r)$ and $b_n = 0$ when $\xi(x_r) < 0$.*

We call the set of Assumptions 2.1 (i), (iv) and (v), and Assumptions 3.1 (ii') and (iii') as Assumption 2.1'

Remark 3.1. In Example 3.1, Assumptions 3.1 (ii') and (iii') are satisfied with $\varphi(x) = 0$, $\Gamma(x) = 1$ and $a_n = 1/((1/n\delta_n^d)^{x_r} - 1)$. We can also check that Example 3.1 satisfies Assumption 2.1 (v). Now we additionally assume that $\delta_n (\log n)^2 \to 0$ as $n \to \infty$. This implies that $\alpha_n^{-\delta_n} \downarrow 1$ 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

$$\begin{aligned} \sup_{|x-x_r| \le \delta_n} a_n |D\theta_{\alpha_n}(x) - D\theta_{\alpha_n}(x_r)| &\lesssim k^{x_r} |\log \alpha_n| \sup_{|x-x_r| \le \delta_n} |\alpha_n^{x-x_r} - 1| \lesssim k^{x_r} |\log \alpha_n| (\alpha_n^{-\delta_n} - 1) \\ &\lesssim k^{x_r} |\log \alpha_n| \alpha_n^{-\pi_n \delta_n} \delta_n |\log \alpha_n|, \text{ for } \pi_n \in (0,1) \\ &\lesssim (\log n)^2 \delta_n, \end{aligned} \tag{3.1}$$

where the third inequality follows from the mean value theorem. Therefore, (3.1) implies $\sup_{x \in \mathbb{N}(x_r, \delta_n)} a_n |R(x, \delta_n)| = o(1)$ and this also implies Assumption 2.1 (v). Analogously, the condition would be satisfied for a wide class of models if $\xi(x)$ and $\Gamma(x)$ are sufficiently smooth on $\mathbb{B}(x_r)$. Furthermore, it is easy to check that Example 3.1 satisfies Assumption 3.2 below.

Under this assumption, the results in Section 2 are extended as follows.

Theorem 3.1. *Suppose that Assumptions 2.1' and 2.2 hold. Then the same result of Proposition 2.1 when $\xi \neq 0$ holds by replacing ξ with $\xi(x_r)$. Additionally, suppose that Assumption 2.3 holds. Then the same results of Theorems 2.1 and 2.2 when $\xi \neq 0$ hold by replacing ξ with $\xi(x_r)$.*

To extend Theorem 2.3 for subsampling inference, we impose the following assumption.

Assumption 3.2. *The conditional quantile density function $\partial F_U^{-1}(\tau|x)/\partial\tau$ exists and satisfies the tail equivalence relationship*

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

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

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 3.2. *Let $\tau \in (0,1)$. Suppose $b \to \infty$, $\alpha_b \to 0$, $\delta_b \to 0$, $a_b\delta_b^{1+\gamma} \to 0$, $b/n \to 0$, $\alpha_b/\alpha_n \to \infty$, $\frac{\log\log n}{n\delta_n^d\alpha_b} \to 0$ and $\sup_{x\in\mathbb{N}(x_r,\delta_b)} a_b|\theta_{\alpha_b}(x_r) - x(x_r,\delta_b)'\beta_b^{(\alpha_b)}| \to 0$ as $n \to \infty$. Under Assumptions 2.1', 2.2, 2.3, and 3.2, the same result of Theorem 2.3 holds.*

3.2. **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 $z_r \in \mathbb{R}^{d_Z}$. We consider the following varying coefficient model

$$Y = X'\beta(Z) + \gamma(X,Z)V_*, \tag{3.2}$$

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 V_* 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) to the extremal quantile regression context.

In this setup, the assumptions in Section 2 are adapted as follows.

Assumption 3.3. (i) *$\{Y_j, X_j, Z_j\}$ 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 $z \in \mathbb{B}(z_r)$. Z has the density function $f_Z(z)$ that is positive and continuous on $\mathbb{B}(z_r)$.*

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

Assumption 3.4. (i) *There exists a measurable function $\beta(\cdot) = (\beta_0(\cdot), \ldots, \beta_d(\cdot))' : \mathbb{B}(z_r) \to \mathbb{R}^{d+1}$ 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 $\{(x, z) : x \in S(X|z), z \in \mathbb{B}(z_r)\}$ for some positive continuous function $\gamma(x, z)$ on $\{(x, z) : x \in S(X|z), u \in \mathbb{B}(z_r)\}$, 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 $a_n\delta_n \to 0$ as $n \to \infty$, where*

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

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

Assumption 3.5. (i) *$D_v\gamma(x, z) = \partial\gamma(x, z)/\partial z_v$ exists and is continuous at each $z \in \mathbb{B}(z_r), x \in S(X|z)$ and for each $v = 1, \ldots, d_Z$.*

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

Assumption 3.6. *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

$$\widehat{N}_1(\cdot) = \sum_{j=1}^{n} \mathbb{I}\left\{\left(a_n(V_j + X_j'(\beta(Z_j) - \beta(z_r))), X_j, (Z_j - z_r)/\delta_n\right) \in \cdot\right\}$$

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

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

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

Proposition 3.1 (Weak convergence of $\widehat{N}_1$). *Under Assumptions 3.3-3.6 and Assumption 2.2 by replacing U_j and W_j with V_j and $W_j = (V_j, X_j', Z_j')'$, respectively, it holds $\widehat{N}_1 \xrightarrow{d} N_1$ in $M_p(\mathbb{E}_1)$, where N_1 is a Poisson point process in $M_p(\mathbb{E}_1)$ with mean measure*

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

Now we focus on the model (3.2) 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 $z \in \mathbb{B}(z_r)$. We also assume that $D_v\sigma_j(z)$ exists and is γ-Hölder continuous at each $z \in \mathbb{B}(z_r)$ and $v = 1, \ldots, d_Z$. In this case, the conditional quantile can be written as

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

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

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

$$\widehat{\beta}_n^{(\alpha)} = \arg\min_{\beta \in \mathbb{R}^{d+1}} \sum_{j=1}^{n} K\left(\frac{Z_j - z_r}{\delta_n}\right) \rho_\alpha(Y_j - X_j'\beta). \tag{3.3}$$

Let $\overline{\Delta}_n(k) = a_n(\widehat{\beta}_n^{(\alpha_n)} - \beta(z_r))$. The asymptotic distribution of the quantile regression estimator (3.3) for the varying coefficient model (3.2) is obtained as follows. Since the proofs of Theorems 3.3 and 3.4 are analogous to those of Theorems 2.1 and 2.2, they are omitted.

Theorem 3.3 (Asymptotic distribution of $\widehat{\beta}_n^{(\alpha_n)}$). *Under Assumptions 3.3-3.6 and Assumption 2.2 by replacing U_j and W_j with V_j and $\widetilde{W}_j = (V_j, X_j', Z_j')'$, respectively, it holds $\overline{\Delta}_n(k) \xrightarrow{d} \overline{\Delta}_\infty(k)$ provided $\overline{\Delta}_\infty(k)$ is defined as a random vector in $\mathbb{R}^{d+1}$ which uniquely minimizes the objective function*

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

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

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

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

$$\overline{\Theta}_n(k,m) = \frac{\widehat{\beta}_n^{(\alpha_n)} - \beta^{(\alpha_n)}(z_r)}{\sum_{j=1}^n K_{n,j} X_j'(\widehat{\beta}_n^{(m\alpha_n)} - \widehat{\beta}_n^{(\alpha_n)})/\sum_{j=1}^n K_{n,j}},$$

where $K_{n,j} = K((Z_j - z_r)/\delta_n)$.

Theorem 3.4 (Asymptotic distribution of $\overline{\Theta}_n(k,m)$)**.** *Under Assumptions 3.3-3.6 and Assumption 2.2 by replacing U_j and W_j with V_j and $W_j = (V_j, X_j', Z_j')'$, respectively, it holds*

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

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 3.2. 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 3.3 (ii) by similar argument to $\Delta_\infty(k)$ and $\Theta_\infty(k,m)$. See also Remark 2.4. 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 $\overline{\Theta}_n(k,m)$ 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.

4. 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:

$$Y_j = 0.5\sin(X_j) + \sqrt{2.5 + 0.5X_j^2}\,U_{*,j}, \tag{4.1}$$

for $j = 1, \ldots, n$, where $n = 2000$, $\{X_j\}$ are i.i.d. uniform random variables on $[-1, 0]$, and $\{U_{*,j}\}$ 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 light-tailed distribution.

We compute $\widehat{\theta}_{\alpha_n}(x_r) = \widehat{\beta}_{n,0}^{(\alpha_n)}$ at $x_r = -0.5$ by using the biweight kernel $K(w) = \frac{15}{16}(1 - w^2)^2\mathbb{I}\{|w| \le 1\}$. To estimate the quantile c_τ of $\Theta_\infty(k,m)$ in (2.13) based on the subsampling method, we consider $B = n - b + 1$ subsets of size $b = 200$ 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 $\theta_{\alpha_n}(x_r)$, we see the finite-sample properties of the following $100(\tau_1 - \tau_2)\%$ confidence

intervals for the model (4.1) with Student's t noise and Weibull noise when $(\tau_1, \tau_2) = (0.91, 0.01)$ and $(0.96, 0.01)$:

$$\theta_{\alpha_n}(x_r) \in [\widehat{\theta}_{\alpha_n}(x_r) - \widehat{c}_{\tau_1}\mathcal{A}_n, \widehat{\theta}_{\alpha_n}(x_r) - \widehat{c}_{\tau_2}\mathcal{A}_n],$$

where $\mathcal{A}_n = \widehat{\beta}_{n,0}^{(m\alpha_n)} - \widehat{\beta}_{n,0}^{(\alpha_n)}$. We consider the following choices for $k(= n\delta_n\alpha_n = b\delta_b\alpha_b)$, the quantile levels α_n and α_b, and m: $(k, \alpha_n) = (2, 0.01)$ or $(1, 0.005)$, $\alpha_b = n\alpha_n/b$, and $m = 2/k+1.1$. The number of Monte Carlo repetitions is 250.

Table 1 presents empirical coverage probabilities of 90% and 95% confidence intervals. We find that the proposed subsampling method works well even for very low quantile $\alpha_n = 0.005$ where confidence intervals based on normal approximations with fixed quantile levels would not work. We also note that the subsampling method can be applicable to "nearly light-tailed" case (i.e., $U_* \sim t(30)$ or Weibull$(30, 1)$). Overall, the simulated coverage probabilities of confidence intervals based on $\widehat{\theta}_{\alpha_n}$ have similar performance and they are reasonably close to the nominal coverage probabilities, although in some cases there are some rooms for improvement.

Model	$t(3)$		$t(30)$		Weibull$(3, 1)$		Weibull$(30, 1)$	
Nominal	0.90	0.95	0.90	0.95	0.90	0.95	0.90	0.95
$(k, \alpha_n) = (2, 0.01)$	0.872	0.932	0.856	0.920	0.852	0.916	0.848	0.912
$(k, \alpha_n) = (1, 0.005)$	0.852	0.924	0.868	0.928	0.892	0.940	0.876	0.936

TABLE 1. Empirical coverage probabilities of $\theta_{\alpha_n}(x_r) = F_Y^{-1}(\alpha_n|x_r)$ at $x_r = -0.5$

Appendix A. Proofs

A.1. **Proof of Proposition 2.1.** We first consider the case where $\{U_j, X_j\}_{j=1}^n$ is i.i.d. Let $\mathcal{E}$ be finite unions and intersections of bounded open rectangles in $\mathbb{E}$. From the definition of the mean measure m in (2.9),

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

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

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

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

for all $E \in \mathcal{E}$, then it holds $\widehat{N} \stackrel{d}{\to} N$ in $M_p(\mathbb{E})$. 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 $E = \cup_{i=1}^M E_i$, where $E_i = (\underline{u}_i, \bar{u}_i) \times E_i^W$ for $i = 1, \dots, M$ are nonoverlapping and nonempty subsets of $\mathbb{E}$, and E_i^W are intersections of open bounded rectangles of $\mathbb{R}^d$. Observe that

$$\begin{aligned} E[\widehat{N}(E)] &= \sum_{i=1}^M E[\widehat{N}(E_i)] = \sum_{i=1}^M nE\left[\mathbb{I}\left\{\left(a_n U_{n,j}, \frac{X_j - x_r}{\delta_n}\right) \in (\underline{u}_i, \bar{u}_i) \times E_i^W\right\}\right] \\ &= \sum_{i=1}^M nE\left[E\left[\mathbb{I}\left\{\left(a_n U_{n,j}, \frac{X_j - x_r}{\delta_n}\right) \in (\underline{u}_i, \bar{u}_i) \times E_i^W\right\}\middle| X_j\right]\right] \\ &= \sum_{i=1}^M nE\left[E\left[\mathbb{I}\left\{a_n(U_j + R_\varphi(X_j, \delta_n)) \in (\underline{u}_i, \bar{u}_i)\right\}\middle| X_j\right]\mathbb{I}\left\{\frac{X_j - x_r}{\delta_n} \in E_i^W\right\}\right] \\ &= \sum_{i=1}^M n\delta_n^d \int_{w\in E_i^W} \left(F_U\left(\frac{\bar{u}_i + o(1)}{a_n}\middle| x_r + \delta_n w\right) - F_U\left(\frac{\underline{u}_i + o(1)}{a_n}\middle| x_r + \delta_n w\right)\right) f_X(x_r + \delta_n w) dw \\ &=: \mathbb{I}_n, \end{aligned}$$

where the first equality follows from the definition of $\{E_i\}_{i=1}^M$ (nonoverlapping), the second equality follows from the stationarity of $\{U_j, X_j\}_{j=1}^n$, the third equality follows from the law of iterated expectation, the fourth equality follows from the definition of $U_{n,j}$ and property of conditional expectation, and the fifth equality follows from the change of variable and Assumption 2.1 (iv) (implying (2.4)). Also, observe that

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

uniformly over $w \in [-1,1]^d$, where the first equality follows from Assumption 2.1 (iii) and the second equality follows from the tail properties of U (given X) and U_* (Assumption 2.1 (ii)). Therefore, (A.1) is obtained as

$$
\mathbb{I}_n \quad = \quad \sum_{i=1}^{M} \Gamma(x_r+\delta_n w)\{\bar{u}_i^{-1/\xi} - \underline{u}_i^{-1/\xi}\} \int_{w\in E_i^W} f_X(x_r+\delta_n w)dw + o(1) \to m(E). \quad \text{(A.3)}
$$

Next, we show (A.2). The same argument to derive (A.3) yields $P\left(\left(a_n U_{n,j}, \frac{X_j-x_r}{\delta_n}\right) \in E\right) \sim \frac{m(E)}{n}$ for any $E \in \mathcal{E}$. Thus, an application of Meyer (1973) yields (A.2). Therefore, we obtain $\widehat{N} \xrightarrow{d} N$ for the case of $\xi < 0$ with i.i.d. observations.

We can also show the same result under geometric strong mixing condition (Assumption 2.2) as an application of Meyer's (1973) theorem and by observing that

$$
\begin{aligned}
& n\sum_{j=2}^{[n/m]} P\left((a_n U_{n,1}, (X_1-x_r)/\delta_n) \in E, (a_n U_{n,j}, (X_j-x_r)/\delta_n) \in E\right) \\
\leq \quad & O\left(n[n/m]P((a_n U_{n,1}, (X_1-x_r)/\delta_n) \in E)^2\right) = O(n[n/m]\delta_n^{2d}\alpha_n^2) = O(1/m). \quad \text{(A.4)}
\end{aligned}
$$

A.2. **Proof of Theorem 2.1.**

Step 1: Overall sketch. Let $K_{n,j} = K\left(\frac{X_j-x_r}{\delta_n}\right)$ and $\Delta_n = a_n(\beta - \beta_{\varphi,n} - b_n\mathbf{e}_1)$. The objective function for $\widehat{\beta}_n^{(\alpha_n)}$ is written as

$$
\begin{aligned}
& \sum_{j=1}^{n} K_{n,j}\rho_{\alpha_n}(Y_j - X_j(x_r,\delta_n)'\beta) \\
= \quad & \sum_{j=1}^{n} K_{n,j}\left[\alpha_n - \mathbb{I}\left\{Y_j - X_j(x_r,\delta_n)'\beta \leq 0\right\}\right](Y_j - X_j(x_r,\delta_n)'\beta) \\
= \quad & a_n^{-1}\sum_{j=1}^{n} K_{n,j}\left[\alpha_n - \mathbb{I}\left\{a_n(U_j + R_\varphi(X_j,\delta_n) - b_n) - X_j(x_r,\delta_n)'a_n(\beta-\beta_{\varphi,n}-b_n\mathbf{e}_1) \leq 0\right\}\right] \\
& \times\left\{a_n(U_j + R_\varphi(X_j,\delta_n) - b_n) - X_j(x_r,\delta_n)'a_n(\beta-\beta_{\varphi,n}-b_n\mathbf{e}_1)\right\} \\
= \quad & a_n^{-1}\sum_{j=1}^{n} K_{n,j}\left[\alpha_n - \mathbb{I}\left\{a_n U_{n,j} - X_j(x_r,\delta_n)'\Delta_n \leq 0\right\}\right]\{a_n U_{j,n} - X_j(x_r,\delta_n)'\Delta_n\}.
\end{aligned}
$$

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

$$\begin{aligned} Q_n(\Delta,k) &= -\alpha_n \sum_{j=1}^n K_{n,j}X_j(x_r,\delta_n)'\Delta - \sum_{j=1}^n K_{n,j}\mathbb{I}\left\{a_nU_{n,j} \leq X_j(x_r,\delta_n)'\Delta\right\}\left\{a_nU_{n,j} - X_j(x_r,\delta_n)'\Delta\right\} \\ &=: -Q_{1n}(\Delta,k) - Q_{2n}(\Delta,k). \end{aligned}$$

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

$$\widetilde{Q}_n(\Delta,k) := -Q_{1n}(\Delta,k) + \widetilde{Q}_{2n}(\Delta,k) := -Q_{1n}(\Delta,k) + \sum_{j=1}^n K_{n,j}\ell_\delta(a_nU_{n,j}, X_j(x_r,\delta_n);\Delta),$$

where

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

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

$$-Q_{1n}(\Delta,k) = \mathrm{sgn}(\xi)\cdot\frac{k+o(1)}{n\delta_n^d}\sum_{j=1}^n K_{n,j}X_j(x_r,\delta_n)'\Delta \xrightarrow{p} \mathrm{sgn}(\xi)\cdot kf_X(x_r)\int_{[-1,1]^d} K(w)\widetilde{w}'dw\Delta,$$

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

$$\begin{aligned} Q_{2n}(\Delta,k) &= \sum_{j=1}^n K_{n,j}\min\{a_nU_{n,j} - X_j(x_r,\delta_n)'\Delta, 0\} = \int_{\mathbb{E}} K(w)\min\{u - \widetilde{w}'\Delta, 0\}d\widehat{N}(u,w), \\ \widetilde{Q}_{2n}(\Delta,k) &= \int_{\mathbb{E}} K(w)\ell_\delta(u,w;\Delta)d\widehat{N}(u,w). \end{aligned}$$

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

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

$$\widetilde{Q}_\infty(\Delta,k) = \mathrm{sgn}(\xi)\cdot kf_X(x_r)\int_{\mathbb{R}^d} K(w)\widetilde{w}'dw\Delta + \int_{\mathbb{E}} K(w)\ell_\delta(u,w;\Delta)dN(u,w), \tag{A.5}$$

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

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

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

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

Condition (a) is satisfied from the definitions of $Q_n(\Delta,k)$ and $\widetilde{Q}_n(\Delta,k)$. 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 $\xi \leq 0$) or (ii) any non-empty open bounded subset of $\Delta_N := \{\Delta \in \mathbb{R}^{d+1} : \max_{w\in[-1,1]^d}\widetilde{w}'\Delta < 0\}$. Thus, it remains to check Condition (b) (in Step 2). Finally in Step 3, we verify the second equality in (2.10).

Step 2: Check Condition (b). Note that $\widetilde{Q}_\infty(\cdot,k)$ in (A.5) is the marginal weak limit of $\{\widetilde{Q}_n(\cdot,k)\}$ if and only if $(\widetilde{Q}_n(\Delta_j,k), j=1,\dots,L) \overset{d}{\to} (\widetilde{Q}_\infty(\Delta_j,k), j=1,\dots,L)$ for any finite collection $\{\Delta_1,\dots,\Delta_L\}$. Let $T: M_p(\mathbb{E}) \to \mathbb{R}^L$ be a mapping defined by

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

Also define

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

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 $\mathbb{E}=[-\infty,\infty)\times\mathbb{R}^d$ and vanishes outside the compact set $[-\infty,\max(\kappa,-\delta)]\times[-1,1]^d$ with $\kappa<\infty$. Then since $M_p(\mathbb{E})$ is equipped with the vague topology, this implies that $T: M_p(\mathbb{E})\to\mathbb{R}^L$ is continuous, and the continuous mapping theorem combined with $\widehat{N}\overset{d}{\to}N$ (Proposition 2.1) yields Condition (b).

Case (ii) $\xi<0$. Note that the map $(u,w)\mapsto K(w)\min\{u-\widetilde{w}'\Delta,0\}$ is continuous on $\mathbb{E}=[0,\infty)\times\mathbb{R}^d$ and vanishes outside the compact set $[0,\max(\kappa,0)]\times[-1,1]^d$ with $\kappa<\infty$. Then $T: M_p(\mathbb{E})\to\mathbb{R}^L$ is continuous, and the continuous mapping theorem combined with $\widehat{N}\overset{d}{\to}N$ (Proposition 2.1) yields Condition (b).

Case (iii) $\xi>0$. Let $\Delta_P := \{\Delta\in\mathbb{R}^{d+1} : \max_{w\in[-1,1]^d}\widetilde{w}'\Delta>0\}$. Since $\Delta_N\cup\Delta_P$ is dense in $\mathbb{R}^{d+1}$, it is enough to show that $\widetilde{Q}_n(\Delta,k)\overset{d}{\to}\widetilde{Q}_\infty(\Delta,k)$ for each $\Delta\in\Delta_N$, and $\widetilde{Q}_n(\Delta,k)\overset{p}{\to}+\infty$ with $\widetilde{Q}_\infty(\Delta,k)=+\infty$ 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 $\mathbb{E}=[-\infty,0)\times\mathbb{R}^d$ 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 $\widehat{N}\overset{d}{\to}N$ (Proposition 2.1) yields $\widetilde{Q}_n(\Delta,k)\overset{d}{\to}\widetilde{Q}_\infty(\Delta,k)$.

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

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

This implies

$$\widetilde{Q}_n(\Delta,k)\ge -Q_{1n}(\Delta,k)+V_{1,n}+V_{2,n},$$

where

$$\begin{aligned} V_{1,n} &:= \sum_{j=1}^n K((X_j-x_r)/\delta_n)\ell_\delta(a_nU_{n,j},X_j(x_r,\delta_n);\Delta)\mathbb{I}\{a_nU_{n,j}\le-\delta\},\\ V_{2,n} &:= \sum_{j=1}^n K((X_j-x_r)/\delta_n)\mathbb{I}\{-\delta/a_n\le U_{n,j}\le 0, X_j(x_r,\delta_n)'\Delta\ge\epsilon\}\epsilon. \end{aligned}$$

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

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

where the second equality follows from the change of variables, the first inequality follows from (2.4) and $\inf_{x\in\mathbb{B}(x_r)} f_X(x) > 0$ (by Assumption 2.1 (i) and (iv)), and the second inequality follows from $\inf_{x\in\mathbb{B}(x_r)} P(U \leq 0|X = x) > 0$ (by Assumption 2.1 (ii)). Therefore, $V_{2,n} \gtrsim O_p(n\delta_n^d) \xrightarrow{p} +\infty$ in $\bar{\mathbb{R}}$. Combining these results, we obtain $\widetilde{Q}_n(\Delta, k) \xrightarrow{p} +\infty$ 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 (2.10)). 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 $\widetilde{m}(du, dw) = du \times 2^{-d}dw$ on

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

Now consider the mapping $J : \widetilde{\mathbb{E}} \to \widetilde{\mathbb{E}}$ defined by

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

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

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

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

can be represented as

$$Q_\infty(\Delta, k) = \mathrm{sgn}(\xi) \cdot k f_X(x_r) \int_{\mathbb{R}^d} K(w) \widetilde{w}' dw \Delta - \sum_{i=1}^{\infty} K(\mathcal{W}_i) \min\{\mathcal{J}_i - \widetilde{\mathcal{W}}_i' \Delta, 0\}.$$

A.3. **Proof of Theorem 2.2.**

A.3.1. *Proof of (2.12).* Note that $\theta_{\alpha_n}(x) = F_Y^{-1}(\alpha_n|x) = \varphi(x) + F_U^{-1}(\alpha_n|x)$ by Assumption 2.1 (ii). When $\xi \neq 0$, Assumption 2.1 (ii)-(iii) imply

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

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

$$a_n(\theta_{\alpha_n}(x_r) - \varphi(x_r) - b_n) \to -\log \Gamma(x_r) + \log k, \tag{A.8}$$

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

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

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 $m, \ell \in (0, \infty)$,

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

where $a(\cdot)$ is the auxiliary function defined in Assumption 2.1 (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:

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

Recall that a Taylor expansion of φ and θ_{α_n} around $x = x_r$ yields $\varphi(x) = x(x_r, \delta_n)'\beta_{\varphi,n} + R_\varphi(x, \delta_n)$ and $\theta_{\alpha_n}(x) = x(x_r, \delta_n)'\beta_n^{(\alpha_n)} + R(x, \delta_n)$. Also Theorem 2.1 implies

$$\Delta_n(k) = a_n(\widehat{\beta}_n^{(\alpha_n)} - \beta_{\varphi,n} - b_n) \xrightarrow{d} \Delta_\infty(k). \tag{A.12}$$

Based on these preparations, we now show (2.12). Note that $\beta_{n,0}^{(\alpha_n)} = \theta_{\alpha_n}(x_r)$ and $\beta_{\varphi,n,0} = \varphi(x_r)$. Thus, we have

$$\begin{aligned} & a_n(\widehat{\beta}_{n,0}^{(\alpha_n)} - \beta_{n,0}^{(\alpha_n)}) \\ = \ & a_n(\widehat{\beta}_{n,0}^{(\alpha_n)} - \beta_{\varphi,n,0} - b_n) - a(\theta_{\alpha_n}(x_r) - \varphi(x_r) - b_n) \\ \stackrel{d}{\to} \ & \Delta_{\infty,0}(k) + g(x_r;\xi), \end{aligned} \tag{A.13}$$

where the convergence of the first term follows from (A.12), the convergence of the second term follows from (A.7) and (A.8), and the convergence of the third term (to zero) follows from (2.4) and (2.5). Therefore, we obtain the conclusion.

A.3.2. *Proof of (2.13).* A similar argument to the proof of Theorem 2.1 yields the weak convergence of

$$(\Delta_n(mk), \Delta_n(k)) \in \arg \min_{(\Delta_1', \Delta_2')' \in \mathbb{R}^{2(d+1)}} \{Q_n(\Delta_1, mk) + Q_n(\Delta_2, k)\},$$

to the limiting distribution

$$(\Delta_\infty(mk), \Delta_\infty(k)) = \arg \min_{(\Delta_1', \Delta_2')' \in \mathbb{R}^{2(d+1)}} \{Q_\infty(\Delta_1, mk) + Q_\infty(\Delta_2, k)\}. \tag{A.14}$$

Here the random vectors $\Delta_\infty(mk)$ and $\Delta_\infty(k)$ are uniquely determined since the objective function $Q_n(\Delta_1, mk) + Q_n(\Delta_2, k)$ is a sum of objective functions in the proof of Theorem 2.1. Observe that

$$\begin{aligned} & a_n(\widehat{\beta}_{n,0}^{(m\alpha_n)} - \widehat{\beta}_{n,0}^{(\alpha_n)}) \\ = \ & a_n\{(\widehat{\beta}_{n,0}^{(m\alpha_n)} - \beta_{\varphi,n,0} - b_n) - (\widehat{\beta}_{n,0}^{(\alpha_n)} - \beta_{\varphi,n,0} - b_n)\} =: A_n. \end{aligned}$$

Combine this with (A.14), the continuous mapping theorem yields

$$A_n \stackrel{d}{\to} \Delta_{\infty,0}(mk) - \Delta_{\infty,0}(k). \tag{A.15}$$

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

$$\Theta_n(k,m) = \frac{a_n(\widehat{\theta}_{\alpha_n}(x_r) - \theta_{\alpha_n}(x_r))}{a_n(\widehat{\theta}_{m\alpha_n}(x_r) - \widehat{\theta}_{\alpha_n}(x_r))} \stackrel{d}{\to} \frac{\Delta_{\infty,0}(k) + g(x_r;\xi)}{\Delta_{\infty,0}(mk) - \Delta_{\infty,0}(k)}.$$

A.4. **Proof of Theorem 2.3.** Let $G_n(x) := P(\Theta_n(k,m) \le x)$ and $G(x) := P(\Theta_\infty(k,m) \le x) = \lim_{n\to\infty} G_n(x)$. We will show Theorem 2.3 in three steps.

Step1: Let $\Theta_{i,b,n} := \widetilde{a}_{i,b,n}(\widehat{\theta}_{\alpha_b}^{(i,b,n)}(x_r) - \theta_{\alpha_b}(x_r))$ and define

$$\begin{aligned} \widehat{G}_{b,n}(x) &:= \frac{1}{B_n}\sum_{i=1}^{B_n} \mathbb{I}\{\widehat{\Theta}_{i,b,n} \le x\} = \frac{1}{B_n}\sum_{i=1}^{B_n} \mathbb{I}\{\Theta_{i,b,n} + \widetilde{a}_{i,b,n}(\theta_{\alpha_b}(x_r) - \widehat{\theta}_{\alpha_b}(x_r)) \le x\}, \\ \widetilde{G}_{b,n}(x;\Delta) &:= \frac{1}{B_n}\sum_{i=1}^{B_n} \mathbb{I}\{\Theta_{i,b,n} + (\widetilde{a}_{i,b,n}/a_b)\Delta \le x\}, \end{aligned}$$

where $a_b = 1/F_{U'}^{-1}(1/(b\delta_b^d))$. Then

$$\mathbb{I}\{\Theta_{i,b,n} \leq x - \widetilde{a}_{i,b,n} w_n/a_b\} \leq \mathbb{I}\{\widehat{\Theta}_{i,b,n} \leq x\} \leq \mathbb{I}\{\Theta_{i,b,n} \leq x + \widetilde{a}_{i,b,n} w_n/a_b\},$$

for all $i = 1, \ldots, B_n$, where $w_n = |a_b(\theta_{\alpha_b}(x_r) - \widehat{\theta}_{\alpha_b}(x_r))|$ and $a_b = -\text{sgn}(\xi) \cdot 1/F_{U'}^{-1}(1/(b\delta_b^d))$.

We can show $w_n = o_p(1)$ as follows. Note that $\widehat{\theta}_{\alpha_b}(x_r)$ is the intermediate order conditional quantile computed using the full sample of size n and since $\alpha_b n \delta_n^d = k_n \alpha_b/\alpha_n \to \infty$ as $n \to \infty$. Thus, Ichimura, Otsu and Altonji (2019, Theorem 3) yields

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

where

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

for $L_0 := \lim_{\tau\downarrow 0} \frac{\partial F_{U_*}^{-1}(\tau)/\partial\tau}{\tau^{-1}F_{U_*}^{-1}(\tau)}$ with $|L_0| \in (0, \infty)$, and the wave relations follow from Assumption 2.4. This implies

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

and we obtain

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

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 $b, n \to \infty$.

Given that $w_n = o_p(1)$, for some sequence of constants $\epsilon_n \downarrow 0$ as $n \to \infty$ the following event occurs with probability approaching one:

$$\begin{aligned} \Omega_n := \Big\{&\mathbb{I}\{\Theta_{i,b,n} \leq x - \widetilde{a}_{i,b,n}\epsilon_n/a_b\} \leq \mathbb{I}\{\Theta_{i,b,n} \leq x - \widetilde{a}_{i,b,n} w_n/a_b\} \leq \mathbb{I}\{\widehat{\Theta}_{i,b,n} \leq x\} \\ &\leq \mathbb{I}\{\Theta_{i,b,n} \leq x + \widetilde{a}_{i,b,n} w_n/a_b\} \leq \mathbb{I}\{\Theta_{i,b,n} \leq x + \widetilde{a}_{i,b,n}\epsilon_n/a_b\} \text{ for all } i = 1, \ldots, B_n\Big\}. \end{aligned}$$

On Ω_n, it holds

$$\widetilde{G}_{b,n}(x; \epsilon_n) \leq \widehat{G}_{b,n}(x) \leq \widetilde{G}_{b,n}(x; -\epsilon_n). \tag{A.16}$$

Step 2: In this step we show that at the continuity points of $G(x)$, $\widetilde{G}_{b,n}(x; \pm\epsilon_n) \stackrel{p}{\to} G(x)$. Non-replacement sampling implies

$$E[\widetilde{G}_{b,n}(x; \epsilon_n)] = P(\Theta_b(k, m) - \widetilde{a}_{i,b,n}\epsilon_n/a_b \leq x),$$

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

$$\lim_{n\to\infty} E[\widetilde{G}_{b,n}(x;\epsilon_n)] = \lim_{b\to\infty} P(\Theta_b(k,m) - \widetilde{a}_{i,b,n}\epsilon_n/a_b \le x) = P(\Theta_\infty(k,m) \le x) = G(x),$$

since $\Theta_b(k,m) \stackrel{d}{\to} \Theta_\infty(k,m)$ and $\widetilde{a}_{i,b,n}\epsilon_n/a_b = O_p(1)\cdot\epsilon_n = o_p(1)$. Since $\widetilde{G}_{b,n}(x;\epsilon_n)$ is a U-statistics of degree b, the law of large numbers for U-statistics in Politis, Romano and Wolf (1999) implies $\mathrm{Var}(\widetilde{G}_{b,n}(x;\epsilon_n)) = o(1)$. This shows that $\widetilde{G}_{b,n}(x;\epsilon_n) \stackrel{p}{\to} G(x)$. Likewise, we obtain $\widetilde{G}_{b,n}(x;-\epsilon_n) \stackrel{p}{\to} G(x)$.

Step 3: Since $P(\Omega_n) \to 1$, (A.16) yields $\widehat{G}_{b,n}(x) \stackrel{p}{\to} G(x)$ for each $x \in \mathbb{R}$. Since convergence of distribution functions at continuity points implies convergence of quantile functions at continuity points, the continuous mapping theorem yields $\widehat{c}_\tau = \widehat{G}_{b,n}^{-1}(\tau) \stackrel{p}{\to} G^{-1}(\tau) = c_\tau$, provided $G^{-1}(\tau)$ is a continuity point of $G(x)$.

A.5. **Proof of Theorem 3.1.** The proof is analogous to the ones for Theorems 2.1 and 2.2.

A.6. **Proof of Theorem 3.2.** The proof is analogous to the one for Theorems 2.3.

References

[1] 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.

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

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

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

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

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

[7] Davis, R. A., Mikosch, T. and O. Pfaffel (2016) Asymptotic theory for the sample covariance matrix of a heavy-tailed multivariate time series, *Stochastic Processes and their Applications*, 126, 767-799.

[8] 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.

[9] de Haan, L. and A. Ferreira (2006) *Extreme Value Theory: An Introduction*, Springer, New York.

[10] Dekkers, A. and L. de Haan (1989) On the estimation of the extreme-value index and large quantile estimation, *Annals of Statistics*, 17, 1795-1832.

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

[12] Fan, J. and I. Gijbels (1996) *Local Polynomial Modeling and Its Applications*, Chapman & Hall.

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

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

[15] 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.

[16] Heiny, J. and T. Mikosch (2017) Eigenvalues and eigenvectors of heavy-tailed sample covariance matrices with general growth rates: the iid case, *Stochastic Processes and their Applications*, 127, 2179-2242.

[17] Heiny, J. and T. Mikosch (2019) The eigenstructure of the sample covariance matrices of high-dimensional stochastic volatility models with heavy tails, *Bernoulli*, 25, 2949-2981.

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

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

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

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

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

[23] Resnick, S. I. (1986) Point processes, regular variation and weak convergence, *Advances in Applied Probability*, 18, 66-138.

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

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

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

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

[28] 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.

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

Department of Industrial Engineering and Economics, School of Engineering, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552, Japan.

Email address: kurisu.d.aa@m.titech.ac.jp

Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK.

Email address: t.otsu@lse.ac.uk