Robust Asymptotic Inference about Conditional Tail Properties: A Panel Data Approach*

Yulong Wang

Syracuse University, Economics Department

April 2019 [Incomplete, comments welcome]

click here for the latest version

Abstract

I consider inference about conditional tail properties such as conditional tail index and conditional extremal quantile. Most existing suggestions with extremal quantile regression as the leading approach rely on the key assumption that the tail shape of the underlying conditional distribution remains unchanged given different conditional values, implying that the conditional extremal quantile can be approximated by a location-scale model. However, this assumption holds only in limited cases such as joint normal distribution, but might be violated in many other empirically relevant cases in which the tail shape depends on the covariates. To construct robust inference with imposing only mild regularity conditions, I develop asymptotically valid confidence intervals for conditional tail properties based on panel data and extreme value theory. These intervals allow for unobserved heterogeneity and dynamic panel, and have excellent small sample coverage and length properties. To illustrate their empirical use, I study the tail risk of the U.S. stock return given stock size and the low birthweight risk given mother's net weight gain during pregnancy.

Keywords: panel data, conditional tail, extreme value theory, nonparametric confidence interval

*Yulong Wang is assistant professor of economics at Maxwell School, Syracuse University. Email: ywang402@maxwell.syr.edu. I thank Yoonseok Lee, Zhijie Xiao, Yichong Zhang and participants at the seminars at Boston College, PSU, and SMU for very helpful comments and advice. I gratefully acknowledge the financial support by the Applyby-Mosher fund.

1 Introduction

Tail risk and extreme events are important research topics in economics and finance. In many applications, the features of interest are conditional tail properties such as conditional tail index and conditional extremal quantile. This article provides a new method to construct confidence intervals for these features. The main advantage of the new method is its robustness to flexible distributional assumptions. In particular, it allows that the location, the scale, and the shape of the tail all depend on the covariate, in a nonparametric way.

Compared with unconditional tail properties, the conditional tail counterparts are much more difficult to study. This is because conditional tails depend on both marginal distributions and their joint behavior. Although the marginal ones can be assumed as Pareto, the joint has to be fully nonparametric and is hard to study given very limited tail observations. To model a covariate-dependent yet tractable tail, the seminal paper by Chernozhukov (2005) extends the Koenker and Bassett (1978) quantile regression (QR) estimator from mid-sample to tail, called the extremal quantile regression (EQR). Chernozhukov and Fernández-Val (2011) further investigate EQR to construct confidence intervals (CIs) based on subsampling.

The EQR approach assumes that the extremal conditional quantile is approximately a location-scale shift model. More specifically, suppose I have independently and identically distributed (i.i.d.) observations $\{Y_i, X_i\}$ for $i = 1, \ldots, n$ and am interested in the τ_y quantile of Y given $X = x$, denoted as $Q_{Y|X=x}(\tau_y)$. The EQR approach assumes that when τ_y is close to 1,

$$Q_{Y|X=x}(\tau_y) \sim \mu(x) + \sigma(x)(1-\tau_y)^{-\xi} \tag{1}$$

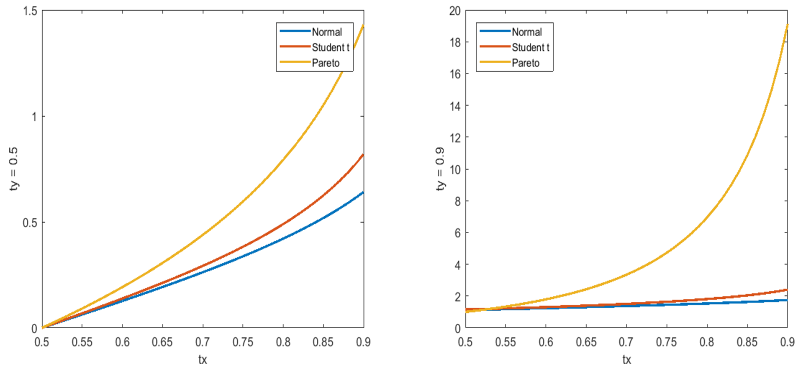
for some *parametric* functions $\mu(x)$ and $\sigma(x)$, which respectively capture the location and the scale. The element $(1-\tau_y)^{-\xi}$ can be treated as the quantile function of a standard Pareto distribution, that is,

$$\mathbb{P}(Y > y) \sim y^{-1/\xi} \tag{2}$$

where $1/\xi$ is the Pareto exponent and ξ is called the tail index. This single parameter captures the tail *shape* in the way that a larger ξ implies a heavier tail. The assumption (1) simplifies the conditional tail distribution so that the covariate X only affects the location and scale, but not the shape[1]. This is satisfied if X and Y are jointly normal but violated

[1] Wang and Li (2013) formally establish that the location-shift model assumption is equivalent to assume ξ remains constant across x.

Figure 1: Condition Quantiles of Joint Normal, Student's t, and Pareto



Note: This figure plots $Q_{Y|X=Q_X}(\tau_y)$ for $\tau_x \in [0.5, 0.9]$ and $\tau_y \in \{0.5, 0.9\}$ where Y and X are distributed as follows: (i) joint normal with zero means, unit variances, and 0.5 correlation; (ii), joint student's t with degree of freedom 3, zero mean, unit variances, and correlation 0.5; and (iii) X is standard normal and $Y|X \sim$Pa$(\xi(x))$, that is, $\mathbb{P}(Y > y|X = x) = 1 - y^{-1/\xi(x)}$ with $\xi(x) = x$.

by many other joint distributions. In contrast to mid-sample properties, such violation may lead to a substantial misspecification error in studying tail ones.

To have a better sense of the misspecification error, Figure 1 plots the conditional quantiles of Y given X equal to its τ_x unconditional quantiles, for three commonly used joint distributions. First, the blue curve depicts the standard joint normal distribution with correlation 0.5, so that $Q_{Y|X=x}(\tau_y)$ is always linear in x. Condition (1) is then satisfied. Second, the red curve depicts the joint student's t distribution with 3 degree of freedom, zero means, unit variances, and 0.5 covariance. Conditional on $X = x$, the distribution of Y is still student's t but with the mean and the variance depending on x in a highly nonlinear way. The upward slope in the tail reflects this. Condition (1) then leads to some error if $\mu(x)$ and $\sigma(x)$ are misspecified. Third, the yellow line depicts the distribution that Y given $X = x$ is Pareto with exponent $1/x$. The conditional quantile $(1 - \tau_y)^{-1/x}$ is highly nonlinear in x, and then approximating such a nonlinear tail property with the linear location-scale shift model (1) induces a large misspecification error. Such error is much more substantial in the tail than in the mid-sample. This is seen by the sharp difference in the vertical axis between the left ($\tau_y = 0.5$) and the right ($\tau_y = 0.9$) panels.

The misspecification error is not only a theoretical but also an empirical concern in some situations. First, conditional value-at-risk (VaR) is a risk measure commonly used in financial management, insurance, and actuarial science. Estimation and inference are studied by Chernozhukov and Umantsev (2001) and Engle and Manganelli (2004), among many others. Recently Adrian and Brunnermeier (2016) propose a new measure for systemic risk, Δ-CoVar, defined as the difference between two conditional VaRs. The tail shape governs the third and higher moments of the portfolio's return, which typically depend on other economic factors, say business cycle. But this is excluded by the location-scale model (1). Second, Kelly and Jiang (2014) find that extreme event risk affects asset pricing in the U.S. stock market. The distribution of stock returns is approximately Pareto in the tail with a time-varying and stock specific shape parameter. In particular, the shape parameter measures tail risk and varies across other stock characteristics such as stock size. I examine this in Section 4.1. Third, top wealth inequality is an active research question in macro finance literature (see, for example, Piketty and Saez (2003), Gabaix, Lasry, Lions, and Moll (2016), and Jones and Kim (2018)). Tail of the wealth distribution is well documented as Pareto, and the exponent is in general a function of fundamentals in general equilibrium models. For example, Beare and Toda (2017) derive a formula for the Pareto exponent and comparative statics results, and Toda (2019) applies that formula in a general equilibrium context. Finally, how infant's birthweight depends on mother's demographics and maternal behavior is an important question in health economics. See Abrevaya (2001), Koenker and Hallock (2001), and Chernozhukov and Fernández-Val (2011). I find that the left tail shape of the birthweight distribution varies across mother's net weight gain during pregnancy. See Section 4.2 for more details. Other economic problems about conditional tail properties can be found in the comprehensive review by Chernozhukov, Fernández-Val, and Kaji (2016).

To solve the misspecification issue, there have been some suggestions in the literature on relaxing (1). They all focus on estimation and can be roughly categorized into two classes. The first class maintains some parametric form but relaxes the location-shift model to allow some nonlinearity. They are more flexible but still suffer from misspecification. In particular, Wang and Tsai (2009) assume $\xi(x)$ equals to $\exp(x\theta_0)$ for some unknown parameter θ_0, and Wang and Li (2013) assume that the Box-Cox transformed Y has a linear conditional quantile in X. The second class is fully nonparametric and constructs some local smooth estimator, including, for example Beirlant, Wet, and Goegebeur (2004), Gardes, Girard, and Lekina (2010), Gardes, Guillou, and Schorgen (2012), Daouia, Gardes, and Girard (2013), and Martins-Filho, Yao, and Torero (2018). These approaches depend heavily on the richness of the data in the target neighborhood and hence require very large samples.

In this article, I focus on statistical inference instead of estimation and provide CIs of conditional tail properties that have good coverage and length properties in relatively small samples. In addition, I consider repeated cross-sectional or panel data instead of cross-sectional random samples, which is the first attempt in the extreme event literature. The main idea is very intuitive: take one particular observation from each time series and collect them into a cross-sectional sample. Suppose I have panel data of Y and X for many individuals and many time periods and am interested in some tail property of the conditional distribution of Y given $X = x$, denoted $F_{Y|X=x}$. If for every individual, there exists some time period in which X takes x, I can simply collect the associated Y's and form a cross-sectional sample from $F_{Y|X=x}$. Since this is infeasible when X is continuous, I instead collect from each individual's time series, the induced Y associated with the X that is closest to x. These induced Y's are now *approximately* stemming from $F_{Y|X=x}$, and the large (small) order statistics from them can be used for inference about the right (left) tail of $F_{Y|X=x}$. I formalize such approximation by establishing a new extreme value (EV) theory. The proof is based on the large n and large T asymptotics, where n and T denote the sample sizes in cross-sectional and time dimensions, respectively. A large T guarantees the closest X is close enough to the conditional value, and a large n provides enough observations from the tail.

Given the new EV theory, I show how existing suggestions on inference about *unconditional* tail properties can be applied using the induced Y's as input. In particular, I consider both the fixed-k asymptotic inference proposed by Müller and Wang (2017) and the methods developed by Hill (1975) and Smith (1987), which are two leading examples in the numerous increasing-k asymptotic methods. The number k denotes how many largest (smallest) observations are used to approximate the tail. The fixed-k approach is more suitable for moderate n, say 200, while the increasing-k ones have computational advantage when n is much larger.

More specifically, the fixed-k approach relies on the fact that the largest k observations for some fixed k are jointly converging to an EV distributed random vector whose density is fully determined by $\xi(x)$ (after imposing location and scale invariance). Then inference on any tail property that is a function of $\xi(x)$ amounts to determine a CI that controls size asymptotically with a random draw from the EV distribution. This approach does not aim for a consistent estimate of $\xi(x)$ but imposes the asymptotically correct size constraint for all values of $\xi(x)$ that are empirically relevant. Hence, the CIs are conservative, but the sample size can be relatively small. In contrast, the increasing-k approach first builds a consistent estimator of $\xi(x)$ and then constructs CIs of tail properties using the asymptotic normality as $k \to \infty$. I show that both the fixed-k and the increasing-k approaches can be applied in the panel data model in Sections 2.5 and 2.6. A Monte Carlo comparison is given in Section 3.

In summary, the main idea is a combination of the nearest neighbor (NN) algorithm in the time dimension and EV theory in the cross-sectional dimension. This idea does not require any parametric assumption on the joint distribution and hence is much more flexible than existing methods. A natural question is how much efficiency I lose by using only one out of T observations in each time series. It turns out that if the tail shape depends on the covariate nonlinearly, the new NN method dominates existing methods in both coverage and length when T is only moderately large, say 50. When T is very large, say 500, the new CIs are only twice longer than the kernel regression method with the optimal bandwidth. See the Monte Carlo results in Section 3 for more details.

The above idea is further generalized to allow multi-dimensional covariates. This is done by defining the NN using the L^2-norm. If a linear regression model is appropriate, the NN can be defined using the linear index. In particular, I consider the classic linear panel data regression model

$$Y_{it} = \alpha_i + \mathbf{X}_{it}^{\intercal}\beta_0 + u_{it}$$

where α_i denotes the unobserved heterogeneity and u_{it} is an error term. Given the linear regression model, I treat $\mathbf{X}_{it}^{\intercal}\beta_0$ as a scalar random variable and focus on the conditional quantile of Y given this scalar index. First, consider α_i as a random variable, which can be correlated with $\mathbf{X}_{it}$ and u_{it}. Under the condition that $(\alpha_i, \mathbf{X}_{it}^{\intercal}, u_{it})^{\intercal}$ is i.i.d. across i and strictly stationary across t, the conditional quantile is equal to $Q_{Y|\mathbf{X}^{\intercal}\beta_0=x}(\tau_y) = x + Q_{\varepsilon|\mathbf{X}^{\intercal}\beta_0=x}(\tau_y)$ where I denote ε_{it} as $\alpha_i + u_{it}$. If β_0 is given, I can directly apply the above introduced approach by taking the induced ε_{it} associated with the order statistics of $\{\mathbf{X}_{it}^{\intercal}\beta_0\}_{i=1}^{T}$ for each i. Given a consistent estimator $\hat{\beta}$ of β_0, I now pick the induced regression residuals $\hat{\varepsilon}_{it} = Y_{it} - \mathbf{X}_{it}^{\intercal}\hat{\beta}$ associated with the order statistics of $\mathbf{X}_{it}^{\intercal}\hat{\beta}$. I show that the errors resulting from both estimating β_0 and the re-ordering are negligible, and hence EV theory still holds for the largest k order statistics in the induced residuals. Then both the fixed-k and the increasing-k methods are applicable again to construct the CI of the quantile of ε_{it}. The CI for $Q_{Y|\mathbf{X}^{\intercal}\beta_0=x}(\tau_y)$ is obtained by adding back the conditional value x. Now consider α_i as a fixed but unknown constant. If I still assume $(\mathbf{X}_{it}^{\intercal}, u_{it})^{\intercal}$ is i.i.d. across i and strictly stationary across t, the conditional quantile of Y_{it} can be written as $Q_{Y_i|\mathbf{X}^{\intercal}\beta_0=x}(\tau_y) = \alpha_i + x + Q_{u|\mathbf{X}^{\intercal}\beta_0=x}(\tau_y)$, which is different across i because of α_i. Under the large T asymptotics, the constant α_i can be consistently estimated by $\bar{\varepsilon}_i$, the time average of $\hat{\varepsilon}_{it}$. Then I establish EV theory for the large order statistics of the induced $\hat{\varepsilon}_{it} - \bar{\varepsilon}_i$. This allows me to construct the CI for $Q_{u|\mathbf{X}^\intercal\beta_0=x}(\tau_y)$ and further that for $Q_{Y_i|\mathbf{X}^\intercal\beta_0=x}(\tau_y)$ by adding back x and $\bar{\varepsilon}_i$.

The rest of the paper is organized as follows. Section 2 reviews EV theory in unconditional distributions, establishes a new theory in the panel setup, and constructs the CIs for conditional tail properties. Section 3 implements an extensive Monte Carlo study, which shows that the new CIs have excellent small sample coverage and length properties for moderately large sample sizes. Section 4 applies the new intervals to the U.S. monthly stock returns and the data on baby birthweights. Finally, Section 5 concludes with proofs and omitted details collected in the Appendix.

2 Main result

2.1 Review of EV theory

Consider a random sample $Y_1, Y_2, ..., Y_n$ from some population with cumulative distribution function (CDF) F_Y. Denote the marginal quantile function as $Q_Y(\tau_y) = \inf\{y : F_Y(y) \geq \tau_y\}$ for some $\tau_y \in [0, 1]$. Let $Y_{(1)} \geq Y_{(2)} \geq \ldots \geq Y_{(n)}$ denote the order statistics, so that $Y_{(1)}$ is the sample maximum. From now on, I consider the right tail, which is without loss of generality.

The fundamental result in EV theory is developed by Fisher and Tippett (1928) and Gnedenko (1943), stating that if there exist sequences a_n and b_n such that

$$\frac{Y_{(1)} - b_n}{a_n} \xrightarrow{d} V_1 \text{ as } n \to \infty \tag{3}$$

for some nondegenerate random variable V_1, then the distribution of V_1 is, up to location and scale normalization, the generalized EV distribution with the CDF

$$G_\xi(v) = \begin{cases} \exp(-(1+\xi v)^{-1/\xi}), & 1+\xi v \geq 0, \text{ for } \xi \neq 0 \\ \exp(-e^{-v}), & v \in \mathbb{R},\ \xi = 0 \end{cases} \tag{4}$$

where $\xi \in \mathbb{R}$ is the tail index.

EV theory holds if and only if F_Y is within the domain of attraction (DOA), denoted as $F_Y \in \mathcal{D}(G_\xi)$. This is a very mild assumption as it is satisfied by most commonly used distributions. In particular, the positive ξ case covers the distributions with a Pareto-type tail such as Pareto, student's t, and F. The case with $\xi = 0$ covers the distributions with finite moments of any order. Leading examples are normal and log-normal. The case with a negative ξ covers the distributions with a bounded right end-point, that is, $Q_Y(1) < \infty$. See Chapter 1 in de Haan and Ferreira (2007) for a complete review.

Without loss of generality, assume that any location and scale normalization of V_1 is subsumed in a_n and b_n, so that the CDF of V_1 is equal to G_ξ. It is well known (see, for instance, Theorem 3.5 of Coles (2001)) that if (3) holds, then EV theory also holds jointly for the first k order statistics

$$\begin{pmatrix} \frac{Y_{(1)}-b_n}{a_n} \\ \vdots \\ \frac{Y_{(k)}-b_n}{a_n} \end{pmatrix} \xrightarrow{d} \mathbf{V} = \begin{pmatrix} V_1 \\ \vdots \\ V_k \end{pmatrix} \tag{5}$$

for any fixed k, where the joint probability density function (PDF) of $\mathbf{V}$ is given by

$$f_{\mathbf{V}|\xi}(v_1, \ldots, v_k) = G_\xi(v_k) \prod_{i=1}^{k} g_\xi(v_i)/G_\xi(v_i) \tag{6}$$

on $v_k \leq v_{k-1} \leq \ldots \leq v_1$ with $g_\xi(v) = \partial G_\xi(v)/\partial v$, and zero otherwise. Note that the constants a_n and b_n depend on ξ and are difficult to estimate. For example, a_n is n^ξ if F_Y is standard Pareto. Since a small estimation error in ξ is amplified by the n-power, inference relying on a good estimate of ξ and the scale usually requires a large k and a even larger sample size n.

The $G_\xi(v_k)$ term in (6) suggests that the largest k order statistics are not asymptotically independent, given any fixed k. In constrast, this term is negligible if k increases with n, and then $\mathbf{V}$ can be considered as independent draws from the generalized Pareto distribution (GPD) (see Section 2.6 for details).

Based on EV theory, tail properties such as extremal quantile can be expressed as known functions of ξ. The inference problem is asymptotically equivalent to the parametric one in which I have k observations drawn from the EV distribution or the GPD and aim for CIs of a function of the single parameter ξ. There have been numerous suggestions on estimation and inference along this line. Depending on whether the asymptotic embedding assumes a fixed or increasing k, I refer to them as the fixed-k or increasing-k approaches, respectively. I discuss them in Sections 2.5 and 2.6. Now, I proceed to study conditional tails and establishing a new EV theory with panel data.

2.2 One-dimentional covariate

I start with a scalar X, which is a continuous random variable with the marginal CDF and the quantile function denoted as $F_X(\cdot)$ and $Q_X(\cdot)$, respectively. The multi-dimensional case is discussed in the next subsection. My object of interest is the tail properties of the conditional distribution $F_{Y|X=x}(\cdot) = \mathbb{P}(Y \leq \cdot|X = x)$. To fix idea, I focus on the extremal conditional quantile of Y given X takes its τ_x marginal quantile, that is, $Q_{Y|X=Q_X(\tau_x)}(\tau_y)$ for some pre-specified $\tau_x \in (0, 1)$ and some τ_y close to 1.

In contrast to the unconditional case, observations from the conditional CDF are not available in a cross-sectional dataset. I overcome this issue by using panel data. Consider a balanced[2] panel dataset $\{Y_{it}, X_{it}\}_{i=1:n,t=1:T}$ that is i.i.d. across i and strictly stationary and weakly dependent across t. My approach is implemented in the following three steps.

Step 1 For each i, order $X_{i1}, \ldots, X_{iT}$ descendingly as

$$X_{i,(1)} \geq X_{i,(2)} \geq \ldots \geq X_{i,(T)}$$

so that $X_{i,(1)}$ is the maximum. Denote the induced order statistics as $Y_{i,[1]}, \ldots, Y_{i,[T]}$, that is, $Y_{i,[j]} = Y_{it}$ if $X_{i,(j)} = X_{it}$. Note that these induced values are not ranked.

Step 2 Suppose $\tau_x T$ is an integer for notational ease. Collect, for each i, the induced Y associated with $X_{i,(\tau_x T)}$ and order them descendingly as

$$Y_{(1),[\tau_x T]} \geq Y_{(2),[\tau_x T]} \geq \ldots \geq Y_{(n),[\tau_x T]}. \tag{7}$$

Step 3 Take the largest k order statistics from (7), denoted as

$$\mathbf{Y} = (Y_{(1),[\tau_x T]}, Y_{(2),[\tau_x T]}, ..., Y_{(k),[\tau_x T]}), \tag{8}$$

and use $\mathbf{Y}$ as input to apply either the fixed-k or the increasing-k approachs reviewed in Sections 2.5 and 2.6.

The main result of this article is summarized in Theorem 1, which states that $\mathbf{Y}$ satisfies a similar convergence as in (5). The key idea is heuristically captured by the following derivation. For each i and any y,

$$\begin{aligned} & \mathbb{P}\left(Y_{i,[\tau_x T]} \leq y\right) \\ = \; & \mathbb{E}_{X_{i,(\tau_x T)}}\left[\mathbb{P}\left(Y_{i,[\tau_x T]} \leq y | X_{i,(\tau_x T)}\right)\right] \\ = \; & \mathbb{E}_{X_{i,(\tau_x T)}}\left[F_{Y|X=X_{i,(\tau_x T)}}(y)\right] \text{ (by strict stationarity)} \end{aligned}$$

[2] This is only for notational ease. The new approach is valid as long as T is large for all i.

$$
\begin{aligned}
&= F_{Y|X=Q_X(\tau_x)}(y) + \mathbb{E}_{X_{i,(\tau_x T)}}\left[\left.\frac{\partial F_{Y|X=x}(y)}{\partial x}\right|_{x=\dot{x}} \left(X_{i,(\tau_x T)} - Q_X(\tau_x)\right) \right] \text{ (by Taylor Expansion)} \\
&\to F_{Y|X=Q_X(\tau_x)}(y) \text{ as } T \to \infty
\end{aligned}
$$

where $\dot{x}$ is between $X_{i,(\tau_x T)}$ and $Q_X(\tau_x)$. The first equation is by the definition of conditional expectation. The second one is established as Theorem 2.1 in Yang (1977) in the i.i.d. case. It also holds under much more general dependence conditions as long as strict stationarity is maintained. The last equation is valid if the conditional CDF is smooth. The final convergence holds if the empirical quantile $X_{i,(\tau_x T)}$ is consistent to the true quantile $Q_X(\tau_x)$ and if the CDF is smooth again with bounded derivatives.

The above derivation states that the collection of the induced order statistics Y associated with a particular empirical quantile of X can be treated as stemming from the true conditional CDF $F_{Y|X=Q_X(\tau_x)}$ asymptotically. Thus the largest (cross-sectional) order statistics $\mathbf{Y}$ can be treated as draws from the tail of $F_{Y|X=Q_X(\tau_x)}$. Given a Pareto-type tail assumption, the problem reduces back to its unconditional analogue and hence existing suggestions on unconditional tail problems become applicable. Note that the normalizing constants a_n and b_n now depends on $\xi(x)$ evaluated at $x = Q_X(\tau_x)$.

A formal establishment requires the following conditions.

Condition 1.1 $(Y_{i1}, X_{i1}), \ldots, (Y_{iT}, X_{iT})$ are i.i.d. across i, and (Y_{it}, X_{it}) for $t = 1, \ldots, T$ is strictly stationary and weakly dependent across t such that $\mathbb{E}[|X_{i,(\tau_x T)} - Q_X(\tau_x)|] = O\left(T^{-1/2}\right)$.

Condition 1.1 requires the data to be independent across individual but weakly dependent across t such that the empirical quantile of X in each time series is root-T consistent to its true counterpart. This is satisfied for many weakly dependent processes. See Wu (2005) for examples.

Condition 1.2 $F_{Y|X=Q_X(\tau_x)} \in \mathcal{D}\left(G_{\xi(Q_X(\tau_x))}\right)$.

This condition requires that the tail of the underlying conditional distribution can be well approximated by a Pareto whose exponent depends on the covariate. It generalizes the conditional location-scale shift model (1) by allowing $\mu(x), \sigma(x)$, and $\xi(x)$ to be all unknown (but smooth) functions of x. To give a concrete sense how $\xi(x)$ enters the conditional quantile, I discuss the following examples.

Example 1 (Joint Normal) Suppose (Y, X) is joint normal with zero means, unit variances, and correlation ρ. Then Y given $X = x$ is normal with mean ρx, and variance $1 - \rho^2$. The conditional tail index is $\xi(x) = 0$ for all $x \in \mathbb{R}$. The conditional quantile is $Q_{Y|X=x}(\tau_y) = \rho x + \sqrt{1 - \rho^2}\Phi^{-1}(\tau_y)$, where $\Phi^{-1}(\cdot)$ is the quantile function of a standard normal. Thus the location-scale model assumption (1) is satisfied.

Example 2 (Joint Student's t) Suppose (Y, X) is jointly student's t distributed with d.f. v, zero means, unit variances, and correlation $\rho \neq 0$. Then Y given $X = x$ is student's t distributed with d.f. $v + 1$, mean ρx, and variance $(1 - \rho^2)(v + x^2)/(v + 1)$. The conditional tail index is $\xi(x) = 1/(v + 1)$ for all $x \in \mathbb{R}$.[3] The conditional quantile is $Q_{Y|X=x}(\tau_y) = \rho x + \sqrt{(1 - \rho^2)(v + x^2)/(v + 1)}Q_{t(v)}(\tau_y)$, where $Q_{t(v)}(\cdot)$ is the quantile function of the standard student's t distribution with d.f. v. This specification satisfies the location-scale shift model (1) but the scale function is highly nonlinear in x.

Example 3 (Conditional Pareto) Suppose X is half-normal with positive support and Y given $X = x$ is the Pareto distribution such that $\mathbb{P}(Y \leq y|X = x) = 1 - (y + 1)^{-1/x}$ for $y \geq 0$ and any $x > 0$. Then the conditional tail index is $\xi(x) = x$ and the conditional quantile is $Q_{Y|X=x}(\tau_y) = -1 + (1 - \tau_y)^{-x}$, which violates the location-scale shift model (1).

Let y_0 denote the end-point of the conditional CDF, that is, $y_0 = Q_{Y|X=x}(1) \leq \infty$. The next assumption is a high level regularity condition on the tail of the conditional CDF, whose preliminary conditions are discussed in the Appendix.

Condition 1.3 $f_{Y|X=x}(y)$ is uniformly bounded and continuously differentiable in x and y. In addition, for any fixed y and $u_n = a_n y + b_n \to y_0$ and any open ball $B_{\delta_T}(x)$ centered at x with radius $\delta_T = O(T^{-1/2})$, $\lim_{u_n \to y_0} \sup_{x \in B_\delta(Q_X(\tau_x))} T^{-1/2} \left| \frac{\frac{\partial}{\partial x} F_{Y|X=x}(u_n)}{1 - F_{Y|X=x}(u_n)} \right| = 0$ and $\lim_{u_n \to y_0} \sup_{x \in B_\delta(Q_X(\tau_x))} T^{-1/2} \left| \frac{\partial f_{Y|X=x}(u_n)}{f_{Y|X=Q_x(\tau_x)}(u_n)} \right| = 0$ as $n \to \infty$ and $T \to \infty$.

Condition 1.3 requires that the derivatives of the conditional CDF and PDF are smooth and decay quickly. This is a mild condition again, which is satisfied by the above examples by straightforward calculation. I give details in the Appendix.

Condition 1.4 $n \to \infty$, $T \to \infty$, and $T/n \to \lambda$ for some $\lambda \in (0, \infty)$.

[3] See Ding (2016) for the exact expression for the PDF.

Condition 1.4 requires both n and T are large. A large n guarantees that the error due to the Pareto tail approximation is negligible, and a large T controls the error due to replacing the true quantile $Q_X(\tau_x)$ with the sample analogue $X_{i,[\tau_x T]}$. The parameter λ can be any non-zero constant, and hence T can be much less than n.

Given the above conditions, I establish the following theorem that formalizes the idea of treating the induced Y's as draws from the true conditional distribution.

Theorem 1 *Under Conditions 1.1-1.4, there exist constants a_n and b_n depending on $Q_X(\tau_x)$ such that*

$$\frac{\mathbf{Y} - b_n}{a_n} \xrightarrow{d} \mathbf{V} \tag{9}$$

where $\mathbf{Y}$ is defined in (8) and $\mathbf{V}$ is jointly EV distributed with PDF (6) and $\xi = \xi(Q_X(\tau_x))$.

2.3 Multi-dimensional covariate and stochastic unobserved heterogeneity

This section generalizes Theorem 1 to the multi-dimensional covariate case. To that end, I impose the following panel data model

$$Y_{it} = \alpha_i + \mathbf{X}_{it}^{\intercal}\beta_0 + u_{it} \tag{10}$$

where β_0 denotes some pseudo structural parameter that leads to the best least squares approximation, α_i the unobserved heterogeneity, and u_{it} the error term. I first consider α_i as a random variable that can be correlated with u_{it} and $\mathbf{X}_{it}$. This case is commonly referred as the fixed effects model in microeconometric analysis (cf. Chapter 10 of Wooldridge (2002)). The case where α_i is treated as fixed constants is studied in the next section. The covariate $\mathbf{X}_{it}$ may include lag dependent variable so that my approach can be used for forecasting.

To apply the previous idea, I treat $X(\beta_0) \equiv \mathbf{X}^{\intercal}\beta_0$ as a scalar random variable and consider the tail properties of $F_{Y|X(\beta_0)=x}$. This is in principle more restrictive than directly considering $F_{Y|\mathbf{X}=\mathbf{x}}$. But given the linear regression setup (10), it is natural to assume that $\mathbf{X}$ determines the location, the scale, and the shape of Y through $X(\beta_0)$.[4] Denote $Q_{X(\beta_0)}(\tau_x)$ as the τ_x unconditional quantile of $X(\beta_0)$ and $Q_{Y|X(\beta_0)=Q_{X(\beta_0)}(\tau_x)}(\tau_y)$ the target conditional quantile.

[4] Such linear regression model is not necessary. To avoid this, we can define the nearest neighbor according to the L^2-norm.

If β_0 were known, I can order $X(\beta_0)_{it}$ for $t = 1, \ldots, T$ and pick the induced order statistics of Y_{it} associated with $X(\beta_0)_{i,(\tau_x T)}$. Otherwise, a feasible implementation is to replace β_0 with some consistent estimator $\hat{\beta}$, which has been suggested by many papers. Leading examples are Anderson and Hsiao (1982), Arellano and Bond (1991), Bai (2009), and Moon and Weidner (2015) among many others.

More specifically, the previous approach is implemented as follows. Define $\varepsilon_{it} = \alpha_i + u_{it}$ and $\hat{\varepsilon}_{it} = Y_{it} - \mathbf{X}_{it}^{\intercal}\hat{\beta}$. The linear model (10) implies that

$$Q_{Y|X(\beta_0)=Q_{X(\beta_0)}(\tau_x)}(\tau_y) = Q_{X(\beta_0)}(\tau_x) + Q_{\varepsilon|X(\beta_0)=Q_{X(\beta_0)}(\tau_x)}(\tau_y).$$

First, for each i, order $\hat{\varepsilon}_{i1}, \ldots, \hat{\varepsilon}_{iT}$ according to the values of $X(\hat{\beta})_{i1}, \ldots, X(\hat{\beta})_{iT}$. Denote the induced order statistics in the same way as in the previous section: $\hat{\varepsilon}_{i,[1]}, \ldots, \hat{\varepsilon}_{i,[T]}$. Second, collect $\hat{\varepsilon}_{i,[\tau_x T]}$ for $i = 1, ..., n$ and order them descendingly. Third, take the largest k order statistics, denoted

$$\hat{\mathbf{Y}} = \hat{\varepsilon}_{(1),[\tau_x T]}, \hat{\varepsilon}_{(2),[\tau_x T]}, ..., \hat{\varepsilon}_{(k),[\tau_x T]}, \tag{11}$$

and apply the fixed-k or increasing-k CIs using $\hat{\mathbf{Y}}$ as input to construct the CI for $Q_{\varepsilon|X(\beta_0)=Q_{X(\beta_0)}(\tau_x)}(\tau_y)$. Finally, add back a consistent estimator $\hat{Q}_{X(\hat{\beta})}(\tau_x)$ of the true unconditional quantile $Q_{X(\beta_0)}(\tau_x)$. One natural choice is the τ_x empirical quantile of the pooled $\{X(\hat{\beta})_{it}\}_{i=1,t=1}^{n,T}$.

To establish a similar convergence of $\hat{\mathbf{Y}}$ as Theorem 1, I impose the following conditions.

Condition 2.1 $(\alpha_i, u_{i1}, \mathbf{X}_{i1}), \ldots, (\alpha_i, u_{iT}, \mathbf{X}_{iT})$ are i.i.d. across i, and $(u_{it}, \mathbf{X}_{it})$ for $t = 1, \ldots, T$ is strictly stationary and weakly dependent across t such that $\mathbb{E}[\,||\mathbf{X}_{it}||^2\,] < \infty$ and $\sup_{\beta \in \mathcal{B}} \mathbb{E}[\,|X(\beta)_{i,(\tau_x T)} - Q_{X(\beta)}(\tau_x)|] = O(T^{-1/2})$, where $\mathcal{B}$ denotes the space of β, which is a compact subset of $\mathbb{R}^{\dim(\beta)}$ excluding $\mathbf{0}$.

Condition 2.2 $F_{\varepsilon|X(\beta_0)=Q_{X(\beta_0)}(\tau_x)} \in \mathcal{D}(G_{\xi(Q_{X(\beta_0)}(\tau_x),\beta_0)})$ for $\xi(Q_{X(\beta_0)}(\tau_x), \beta_0) \geq -1/2$.

Condition 2.3 $f_{\varepsilon|X(\beta)=x}(y)$ is continuous in β and uniformly bounded and continuously differentiable in x and y. In addition, for any fixed y and $u_n = a_n y + b_n \to y_0$ and some open ball $B_{\delta_T}(x)$ centered at x with radius $\delta_T = O(T^{-1/2})$, $\lim_{u_n \to y_0} \sup_{x \in B_{\delta_T}(Q_{X(\beta_0)}(\tau_x))} T^{-1/2} \left| \frac{\partial F_{\varepsilon|X(\beta_0)=x}(u_n)/\partial x}{1 - F_{\varepsilon|X(\beta_0)=x}(u_n)} \right| = 0$, $\lim_{u_n \to y_0} \sup_{x \in B_{\delta_T}(Q_{X(\beta_0)}(\tau_x))} T^{-1/2} \left| \frac{\partial f_{\varepsilon|X(\beta_0)=x}(u_n)/\partial x}{f_{\varepsilon|X(\beta_0)=Q_{X(\beta_0)}(\tau_x)}(u_n)} \right| = 0$, and $\lim_{u_n \to y_0} \sup_{\beta \in B_{v_{nT}}(\beta_0)} \sup_{x \in B_{\delta_T}(Q_{X(\beta_0)}(\tau_x))} T^{-1/2} \left| \frac{\partial F_{\varepsilon|X(\beta)=x}(u_n)/\partial \beta}{1 - F_{\varepsilon|X(\beta_0)=x}(u_n)} \right| = 0$ for some $v_{nT} = O((nT)^{-1/2})$, as $n \to \infty$ and $T \to \infty$.

Condition 2.4 $n \to \infty$, $T \to \infty$, and $T/n \to \lambda$ for some $\lambda \in (0, \infty)$.

Condition 2.5 $\hat{\beta} - \beta_0 = O_p\left((nT)^{-1/2}\right)$ as $n \to \infty$ and $T \to \infty$.

Condition 2 is similar to Condition 1. In particular, Condition 2.1 requires that the data are i.i.d. across i and strictly stationary across t, and the empirical quantile of $X(\beta)$ is uniformly root-T consistent. This uniform convergence is implied by the standard empirical process argument (for example, Wu (2004)). I provide details in the Appendix. Note that the point zero is excluded to make sure $X(\beta)$ is a non-degenerate scalar random variable. Condition 2.2 requires that the tail of the target conditional CDF is in the domain of attraction. Note that both β and the covariate $X(\beta)$ affect the conditional CDF. This additional error from estimating β_0 (and replacing ε_{it} by $\hat{\varepsilon}_{it}$) makes it a non-trivial generalization from Theorem 1 to Theorem 2. The condition that the tail index is not less than $-1/2$ is used to show the estimation error in $\hat{\varepsilon}$ is negligible in the tail. This rules out the distributions with a bounded support that decays faster than the triangular distribution. Condition 2.3 requires the derivatives of the conditional CDF and PDF w.r.t. x and β are both uniformly small in the tail. This is also a mild assumption that is satisfied by commonly used distributions, including the three examples discussed before. Preliminary conditions are again postponed to the Appendix for readability (see Lemma 1). Condition 2.4 is identical to 1.4. Condition 2.5 requires that the estimator of β is consistent and converges fast enough.

Given Condition 2, I establish the following theorem, which generalizes the one-dimensional covariate case to the multi-dimensional one.

Theorem 2 *Under Conditions 2.1-2.5, there exist sequences of constants a_n and b_n depending on $Q_{X(\beta_0)}(\tau_x)$ such that*

$$\frac{\hat{\mathbf{Y}} - b_n}{a_n} \xrightarrow{d} \mathbf{V}$$

where $\hat{\mathbf{Y}}$ is defined in (11) and $\mathbf{V}$ is jointly EV distributed with PDF (6) and $\xi = \xi(Q_{X(\beta_0)}(\tau_x), \beta_0)$.

2.4 Multi-dimensional covariate and constant unobserved heterogeneity

Now we consider α_i as a fixed but unknown constant. Following the previous subsection, the conditional quantile of Y is now written as

$$Q_{Y_i|X(\beta_0)=Q_{X(\beta_0)}(\tau_x)}(\tau_y) = \alpha_i + Q_{X(\beta_0)}(\tau_x) + Q_{u|X(\beta_0)=Q_{X(\beta_0)}(\tau_x)}(\tau_y)$$

where the subscript i indicates that Y_i is no longer identically distributed due to α_i. Define $\bar{u}_i = T^{-1}\sum_{t=1}^{T} u_{it}$, $\bar{\varepsilon}_i = T^{-1}\sum_{t=1}^{T} \hat{\varepsilon}_{it}$, and $\bar{\mathbf{X}}_i = T^{-1}\sum_{t=1}^{T} \mathbf{X}_{it}$. Then under the assumption that uniformly in i, $\bar{u}_i$ converges to zero and $\bar{\mathbf{X}}_i$ is $O_p(1)$, we have

$$\begin{aligned} \bar{\varepsilon}_i &= \alpha_i + \bar{u}_i - \bar{\mathbf{X}}_i^{\intercal}(\hat{\beta} - \beta_0) \\ &\xrightarrow{p} \alpha_i \end{aligned}$$

as $T \to \infty$. Consider

$$\begin{aligned} \tilde{\varepsilon}_{it} &= \hat{\varepsilon}_{it} - \bar{\varepsilon}_i \\ &= u_{it} - \bar{u}_i - (\mathbf{X}_{it} - \bar{\mathbf{X}}_i)^{\intercal}(\hat{\beta} - \beta_0) \end{aligned}$$

as the previous Y_{it} and order them according to $\mathbf{X}_{it}^{\intercal}\hat{\beta}$. Collect the largest k of the induced nearest neighbors as

$$\tilde{\mathbf{Y}} = \{\tilde{\varepsilon}_{(1),[\tau_x T]}, ..., \tilde{\varepsilon}_{(k),[\tau_x T]}\}. \tag{12}$$

Then under the following regularity condition, I establish the desired EV theory.

Condition 2.1' α_i is a finite constant for all i. $(u_{i1}, \mathbf{X}_{i1}), \ldots, (u_{iT}, \mathbf{X}_{iT})$ are i.i.d. across i, and $(u_{it}, \mathbf{X}_{it})$ for $t = 1, \ldots, T$ is strictly stationary and weakly dependent across t such that $\mathbb{E}[\,||\mathbf{X}_{it}||^2\,] < \infty$, $\sup_i |\bar{u}_i| = O_p(T^{-1/2})$, $\sup_i \left|\left|\bar{\mathbf{X}}_i\right|\right| = O_p(1)$, and $\sup_{\beta\in\mathcal{B}} \mathbb{E}[\,\left|X(\beta)_{i,(\tau_x T)} - Q_{X(\beta)}(\tau_x)\right|] = O(T^{-1/2})$, where $\mathcal{B}$ denotes the space of β, which is a compact subset of $\mathbb{R}^{\dim(\beta)}$ excluding $\mathbf{0}$.

Theorem 3 *Under Conditions 2.1' and 2.2-2.5 with* $\xi(Q_{X(\beta_0)}(\tau_x), \beta_0)) > -1/2$, *there exist sequences of constants* a_n *and* b_n *depending on* $Q_{X(\beta_0)}(\tau_x)$ *such that*

$$\frac{\tilde{\mathbf{Y}} - b_n}{a_n} \xrightarrow{d} \mathbf{V}$$

where $\tilde{\mathbf{Y}}$ *is defined in (12) and* $\mathbf{V}$ *is jointly EV distributed with PDF (6) and* $\xi = \xi(Q_{X(\beta_0)}(\tau_x), \beta_0)$.

Condition 2.1' is similar to Condition 2.1. The key difference is that α_i now is treated as a fixed constant and that $\bar{u}_i$ has to be uniformly close to zero. In particular, the stochastic boundedness of $\bar{u}_i$ is a strong assumption since it rules out that u_{it} has infinite variance. This is necessary here because the estimation error of α_i needs to be controlled. If the true conditional tail index $\xi(Q_{X(\beta_0)}(\tau_x), \beta_0))$ is non-negative, the rate of $\sup_i |\bar{u}_i|$ can be relaxed into $o_p(1)$.

Given Theorem 3, the CIs of $Q_{u|X^\intercal\beta_0=Q_{X^\intercal\beta_0}(\tau_x)}(\tau_y)$ can be constructed by using $\tilde{\mathbf{Y}}$ as input. Then the CI for the quantile of Y_i is obtained by adding back $\bar{\varepsilon}_i$ and $\hat{Q}_{X^\intercal\hat{\beta}}(\tau_x)$. The continuous mapping theorem and the consistency of $\bar{\varepsilon}_i$ and $\hat{Q}_{X^\intercal\hat{\beta}}(\tau_x)$ guarantee the asymptotically correct coverage.

2.5 Apply the fixed-k asymptotic inference

Given Theorems 1 and 2, this section applies the fixed-k approach proposed by Müller and Wang (2017) for inference about conditional tail properties. This approach is originally developed for unconditional tail properties only. The major advantage is the excellent performance in coverage probability and length given relatively small sample sizes, say $n = 200$.

As the name suggests, the fixed-k approach relies on the asymptotic embedding that the largest k observations from an i.i.d. sample converge in distribution to the joint EV distributed vector $\mathbf{V}$. If the tail property under investigation is also asymptotically equivalent to some known function of ξ, I essentially end up with a straightforward parametric problem: construct CIs for a function of ξ with one draw from $\mathbf{V}$ whose PDF is characterized only by ξ.

Suppose I have the dataset as in Section 2.2 and aim for a $1-\alpha$ CI for the conditional extremal quantile $Q_{Y|X=Q_X(\tau_x)}(\tau_y)$ for τ_y close to 1. To be precise, I rewrite τ_y as $1-h/n$ for some $h > 0$ as similarly considered in Chernozhukov (2005) and Chernozhukov and Fernández-Val (2011). This setup means that the extremal quantile is of the same order of the sample maximum from n random draws from the true conditional CDF $F_{Y|X=Q_X(\tau_x)}$. Such an extreme quantile is too far in the tail for the normal approximation to perform well.

In the one-dimentional covariate case, the effective data becomes $\mathbf{Y}$ as in (8) and the objective is to construct a confidence set $S(\mathbf{Y}) \subset \mathbb{R}$ such that $\mathbb{P}(Q_{Y|X=Q_X(\tau_x)}(\tau_y) \in S(\mathbf{Y})) \geq 1-\alpha$, at least as $n \to \infty$ and $T \to \infty$. In particular, EV theory suggests that

$$\frac{Q_{Y|X=Q_X(\tau_x)}(1-h/n) - b_n}{a_n} \to q(\xi, h) \equiv \begin{cases} \frac{h^{-\xi}-1}{\xi} & \text{if } \xi \neq 0 \\ -\log(h) & \text{if } \xi = 0 \end{cases}$$

where I suppress $Q_X(\tau_x)$ in ξ for notational ease. Note that $q(\xi, h)$ is the $\exp(h)$ quantile of V_1. The normalizing constants a_n and b_n implicitly depend on ξ and hence are unknown. Since they are shared by both $\mathbf{Y}$ and $Q_{Y|X=Q_X(\tau_x)}(1-h/n)$, I can impose location and scale equivariance on the CI to cancel them out. Specifically, I impose that for any constants $a > 0$ and b, $S(a\mathbf{Y} + b) = aS(\mathbf{Y}) + b$, where $aS(\mathbf{Y}) + b = \{y : (y-b)/a \in S(\mathbf{Y})\}$. Under this equivariance constraint, I can write

$$\begin{aligned} & \mathbb{P}(Q_{Y|X=Q_X(\tau_x)}(1-h/n) \in S(\mathbf{Y})) \\ = \; & \mathbb{P}\left(\frac{Q_{Y|X=Q_X(\tau_x)}(1-h/n) - Y_{(k),[\tau_x T]}}{Y_{(1),[\tau_x T]} - Y_{(k),[\tau_x T]}} \in S\left(\frac{\mathbf{Y} - Y_{(k),[\tau_x T]}}{Y_{(1),[\tau_x T]} - Y_{(k),[\tau_x T]}}\right)\right) \\ \to \; & \mathbb{P}\left(Y^s \in S(\mathbf{V}^s)\right) \end{aligned}$$

where I introduce the self-normalized statistics

$$\begin{aligned} Y^s &= \frac{q(\xi,h) - V_k}{V_1 - V_k} \text{ and} \\ \mathbf{V}^s &= \left(\frac{V_1 - V_k}{V_1 - V_k}, \frac{V_2 - V_k}{V_1 - V_k}, ..., \frac{V_k - V_k}{V_1 - V_k}\right). \end{aligned}$$

The densities of Y^s and $\mathbf{V}^s$ now depend solely on ξ and can be numerically computed.

Given only a finite number of observations, a consistent estimation of ξ is out of the question. Instead of imposing the correct size control for the true ξ, I impose it for all the values of ξ that are empirically relevant. In this sense the fixed-k approach is conservative but more robust to misspecification, especially when the sample size is not large enough to support a precise estimation of ξ. Let $\Xi \subset \mathbb{R}$ be the set of tail indices for which I impose the asymptotically correct coverage[5]. The asymptotic problem then is to construct a location and scale equivariant S that satisfies

$$\mathbb{P}\left(Y^s \in S(\mathbf{V}^s)\right) \geq 1 - \alpha \text{ for all } \xi \in \Xi, \tag{13}$$

since any S that satisfies (13) also satisfies $\liminf_{n\to\infty, T\to\infty} \mathbb{P}(Q_{Y|X=Q_X(\tau_x)}(1-h/n) \in S(\mathbf{Y})) \geq 1-\alpha$ under (9) and the continuous mapping theorem. Among all solutions to this problem, I choose the optimal one that minimizes the weighted average expected length criterion

$$\int \mathbb{E}[\text{lgth}(S(\mathbf{V}))]dW(\xi), \tag{14}$$

where W is a positive measure with support on Ξ, and $\text{lgth}(A) = \int \mathbf{1}[y \in A]dy$ for any Borel set $A \subset \mathbb{R}$. The equivariance of S further implies $\mathbb{E}[\text{lgth}(S(\mathbf{V}))] = \mathbb{E}[(V_1 - V_k)\,\text{lgth}(S(\mathbf{V}^s))]$. Thus the program of minimizing (14) subject to (13) among all equivariant set estimators S asymptotically becomes

$$\begin{aligned} & \min_{S(\cdot)} \textstyle\int_{\Xi} \mathbb{E}[(V_1 - V_k)\,\text{lgth}(S(\mathbf{V}^s))]dW(\xi) \\ & \text{s.t. } \mathbb{P}(Y^s \in S(\mathbf{V}^s)) \geq 1 - \alpha \text{ for all } \xi \in \Xi. \end{aligned} \tag{15}$$

[5] I use $\Xi = [-1/2, 1/2]$ in later applications which covers all the distributions with finite variance. This can be easily extended.

Note that the above expectation and probability are w.r.t. the distribution of Y^s and $\mathbf{V}^s$. This distribution depends on $\xi(x)$ evaluated at $Q_X(\tau_x)$. Solution to problem (15) is numerically calculated with the corresponding MATLAB program provided on the author's website. The computation cost is only several seconds using a modern PC.

In addition to the extremal conditional quantile, I also construct a fixed-k CI for the conditional tail index $\xi(Q_X(\tau_x))$. The unconditional case is studied by Wang (2018), which is now generalized to the conditional case. In particular, consider the testing problem

$$\begin{aligned} H_0 &: \xi(Q_X(\tau_x)) = \xi_0 \text{ against} \\ H_1 &: \xi(Q_X(\tau_x)) \in \Xi \backslash \{\xi_0\}. \end{aligned}$$

With some weighting function W again, the likelihood ratio test is constructed as

$$\varphi(\mathbf{v}^s) = \mathbf{1}\left[\frac{\int_\Xi f_{\mathbf{V}^s|\xi}(\mathbf{v}^s) dW(\xi)}{f_{\mathbf{V}^s|\xi_0}(\mathbf{v}^s)} > \text{cv}_{(\alpha;\xi_0)}\right] \tag{16}$$

where $\mathbf{1}[\cdot]$ is the indicator function, and $\text{cv}_{(\alpha;\xi_0)}$ denotes the critical value that depends on the significance level α and the null value ξ_0. The CI for ξ is then obtained by inverting this test. Theorem 1 and the continuous mapping theorem again yield that the test (16) and the corresponding CI are asymptotically valid.

The CIs (15) and (16) are also applicable in the multivariate $\mathbf{X}$ case as implied by Theorem 2. Other tail related quantities such as the conditional tail expectation are also covered as long as they can be expressed as functions of the conditional tail index.

2.6 Apply the increasing-k asymptotic inference

If the cross-sectional sample size n is large enough, I can also switch to the increasing-k approaches that are originally developed for unconditional tail properties (see de Haan and Ferreira (2007) for an overview). In particular, I consider two popular estimators developed respectively by Hill (1975) and Smith (1987), and show that they are applicable using $\mathbf{Y}$ as input if k is large. Application of other methods is also possible but requires a method-specific consideration. I leave this for future research.

In the unconditional case, Pickands (1975) states that if $F_Y \in \mathcal{D}(G_\xi)$, the generalized Pareto distribution is a good approximation of the tail of F_Y, in the sense that

$$\lim_{u \to Q_Y(1)} \sup_{0<y<Q_Y(1)-u} \left| \frac{F_Y(y+u) - F_Y(u)}{1 - F_Y(u)} - F_{\text{GP}}(y; \xi, \sigma) \right| = 0$$

where

$$F_{\mathrm{GP}}(y;\xi,\sigma)=\begin{cases}1-(1+\xi y/\sigma)^{-1/\xi} & \text{if } \xi\neq 0\\ 1-\exp(y/\sigma) & \text{if } \xi=0.\end{cases}$$

Choose $Y_{(k),[\tau_x T]}$ as the cutoff u_n, Hill's estimator can be constructed in the panel framework as

$$\hat{\xi}_H=\frac{1}{k-1}\sum_{i=1}^{k-1}(\log Y_{(i),[\tau_x T]}-\log Y_{(k),[\tau_x T]}).$$

Second, Smith (1987) suggests fitting the differences between the largest k observations and the cutoff $Y_{(k),[\tau_x T]}$ to $F_{\mathrm{GP}}(y;\xi,\sigma)$ and constructing the maximum likelihood estimator for ξ and σ. In particular, this estimator can be implemented as

$$(\hat{\xi}_{ML},\hat{\sigma}_{ML})=\arg\max_{\xi\in\Xi,\sigma\in\mathbb{R}^+}\sum_{i=1}^{k-1}\log(f_{\mathrm{GP}}(Y_{(i),[\tau_x T]}-Y_{(k),[\tau_x T]};\xi,\sigma)) \tag{17}$$

where $f_{\mathrm{GP}}(y;\xi,\sigma)=\partial F_{\mathrm{GP}}(y;\xi,\sigma)/\partial y$.

Both Hill's and Smith's estimators are root-k consistent and asymptotically normal, provided that $k\to\infty$ and $k/n\to 0$. Since the Hill's estimator is only defined for positive indices, I restrict $\xi(x)$ to be positive. For notational ease, I present the proof for the scalar X case and write $\gamma(x)=1/\xi(x)$ and $\tilde{\gamma}(x)=1/\tilde{\xi}(x)$. The following additional conditions are imposed.

Condition 3.1 $1-F_{Y|X=x}(y)=c(x)y^{-\gamma(x)}(1+d(x)y^{-\tilde{\gamma}(x)}+r(x,y)y^{-\tilde{\gamma}(x)})$ uniformly as $y\to\infty$ where $c(\cdot)$, $d(\cdot)$, $\gamma(\cdot)$, and $\tilde{\gamma}(\cdot)$ are continuously differentiable and uniformly bounded between 0 and ∞, and $r(x,y)$ is continuously differentiable with bounded derivatives w.r.t. both x and y and satisfies $\limsup_{y\to\infty}\sup_{x\in B_{T^{-1/2}}(Q_X(\tau_x))}|r(x,y)|=0$.

Condition 3.2 $\sqrt{k}n^{-1/(\tilde{\xi}(Q_X(\tau_x))/\xi(Q_X(\tau_x))+2)}\to\mu$ for some constant $\mu\in\mathbb{R}$.

Condition 3.1 is a second order condition on the GPD approximation. In particular, the conditional CDF is approximated by a GPD in the first order, and the parameter $\tilde{\xi}$ governs the approximation bias. This condition is commonly assumed to study unconditional tail problems (see, for example, Hall (1982), Smith (1987), and Chernozhukov (2005)). Condition 3.2 specifies the choice of the tail cutoff that leads to a non-degenerate asymptotic bias in the tail index estimator (cf. eq. (3.3) in Smith (1987)). This is seen in the following theorem.

Theorem 4 *Suppose Conditions 1.1, 1.4, and 3 hold with* $\xi(Q_X(\tau_x)) > 0$. *If* $k \to \infty$, $k/n \to 0$, *then*

$$\sqrt{k}(\hat{\xi}_H - \xi) \xrightarrow{d} \mathcal{N}\left(\mu_H, \xi^2\right),$$

and

$$\sqrt{k}(\hat{\xi}_{ML} - \xi) \xrightarrow{d} \mathcal{N}\left(\mu_{ML}, (1+\xi)^2\right)$$

where

$$\begin{aligned} \mu_H &= -\mu\xi \\ \mu_{ML} &= -\frac{\mu(1+\xi)\xi(1-\tilde{\gamma})}{1+\xi-\xi\tilde{\gamma}} \end{aligned}$$

and $(\xi, \mu, \tilde{\gamma})$ *are evaluated at* $x = Q_X(\tau_x)$.

Theorem 4 derives the asymptotic distributions of the Hill's and the Smith's estimators, which are identical to that shown in Goldie and Smith (1987) and Smith (1987). The bias terms are difficult to estimate since they involve the second order parameters $d(x)$ and $\tilde{\gamma}(x)$. A feasible CI for ξ is then constructed by choosing k of a smaller order than specified in Condition 3.2 so that the bias is asymptotically zero (cf. Theorem 2 in Hall (1982)). This is similar to the undersmoothing choice of the bandwidth in kernel estimation. The CIs are obtained accordingly by plugging in the index estimate for the variance.

3 Monte Carlo results

In this section I run Monte Carlo experiments to examine the small sample performance of the new approach. Section 3.1 considers the simple panel data $\{Y_{it}, X_{it}\}$ without any fixed effect. In Section 3.2, I compare the efficiency of the new approach with a kernel estimator, which essentially uses more than one nearest neighbor observations. In Sections 3.3 and 3.4, I impose the linear regression setup (10) with classic individual random effects and fixed effects, respectively. Section 3.5 considers the CIs for the one time period forecast using dynamic panel data with fixed effects. Finally in Section 3.6, I present the results on inference about the conditional tail index, using both the fixed-k and the increasing-k methods.

3.1 Extremal conditional quantile with i.i.d. data

I continue to consider the three examples in Section 2.2 as the data generating processes (DGPs). In all experiments, data are i.i.d. across i. The dependence structure across t is as follows.

1. **Joint Normal** $X_{it} = \rho X_{it-1} + u_{it}$ with $u_{it} \sim^{iid} \mathcal{N}(0, 1-\rho^2)$ and $X_{i1} \sim \mathcal{N}(0,1)$. $Y_{it} = r_{xy} X_{it} + \sqrt{1-r_{xy}^2} v_{it}$ where $v_{it} \sim^{iid} \mathcal{N}(0,1)$ and independent of u_{it}. Set $\rho = 0.5$ and $r_{xy} = 0.5$.

2. **Joint Student's t** (X_{it}, Y_{it}) is i.i.d. across t and distributed as $t_v(\mu, \Sigma)$ with $v = 3$, $\mu = [0,0]^\intercal$, and $\Sigma = [1, 0.5; 0.5, 1]$.

3. **Conditional Pareto** $X_{it} = \rho X_{it-1} + u_{it}$ with $u_{it} \sim^{iid} \mathcal{N}(0, (1-\rho^2))$ and $X_{i1} \sim \mathcal{N}(0,1)$. $Y_{it}|X_{it} = x \sim \text{Pa}(\xi(x))$, that is, $\mathbb{P}(Y_{it} \leq y | X_{it} = x) = 1 - y^{-1/\xi(x)}$ for $y \geq 1$ where $\xi(x) = x - Q_X(\tau_x) + 0.5$.

Conditional on $X = Q_X(\tau_x)$, the tail indices for the above three DGPs are 0, 1/4, and 1/2, respectively. I construct CIs for $Q_{Y|X=Q_X(\tau_x)}(1-h/n)$ with $\tau_x = 0.5$ and 0.95 and $h = 1$ and 5. The sample sizes n and T are either 200 or 500, with smaller combinations exercised in later sections.

I compare three approaches: (i) the fixed-k approach (fixed-k) introduced in Section 2.2, (ii) quantile regression (QR), and (iii) bootstrapping the empirical quantile (Boot). More specifically, I produce the fixed-k CI using $k = 20$ in most cases if not specially noted. The space of ξ is restricted as $[-1/2, 1/2]$ in this and the following two sections where I target quantile. For the QR approach, I run a quantile regression of Y_{it} on X_{it} and a constant at the τ_y quantile for each i . The conditional quantile is estimated at $\hat{\beta}_{0i} + \hat{Q}_X(\tau_x)\hat{\beta}_{1i}$ where $\hat{Q}_X(\tau_x)$ is the empirical quantile of the pooled $X's$ and $\hat{\beta}_{0i}$ and $\hat{\beta}_{1i}$ are the coefficient estimates using the i-th individual's observations. The CI is simply the 2.5% and 97.5% quantiles of these n estimates. The bootstrap CI is based on bootstrapping the empirical τ_y quantile in $\{Y_{i,[\tau_x T]}\}_{i=1}^n$. The bootstrap size is 200.

Tables 1-3 depict the coverage probabilities (Cov) and the average lengths (Lgth) of the above three methods based on 500 simulation draws. The fixed-k approach performs very well in both coverage and length in all specifications. Regarding the QR method, since the conditional quantile is a linear function of X in the first DGP but not in the other two, the CIs based on QR perform well in the first DGP but deliver substantial undercoverage

Table 1: Finite sample performance of inference about extremal conditional quantile, no model specification

n	200 (97.5% quantile)				500 (99% quantile)			
T	200		500		200		500	
	Cov	Lgth	Cov	Lgth	Cov	Lgth	Cov	Lgth
				Joint Normal				
fixed-k	0.97	0.63	0.96	0.66	0.95	0.56	0.96	0.56
QR	1.00	0.63	1.00	0.41	1.00	0.89	1.00	0.56
Boot	0.97	0.64	0.91	0.61	0.88	0.58	0.95	0.55
				Joint Student's t				
fixed-k	0.96	1.35	0.96	1.47	0.95	1.62	0.94	1.63
QR	0.95	2.20	0.00	1.31	1.00	4.76	0.01	2.87
Boot	0.91	1.36	0.95	1.34	0.89	1.51	0.94	1.68
				Conditional Pareto				
fixed-k	0.96	7.65	0.97	7.14	0.98	15.8	0.97	11.6
QR	0.00	$>10^3$	0.00	$>10^3$	0.00	$>10^3$	0.00	$>10^3$
Boot	0.93	8.30	0.93	7.80	0.94	15.3	0.90	12.7

Note: Entries are coverages and lengths of the CIs for $Q_{Y|X=Q_X(0.5)}(1-5/n)$. See the main text for the description of the three approaches and the data generating processes. Confidence level is 5%. Based on 500 simulation draws.

and longer length in the other two DGPs due to misspecification. The bootstrap approach is robust to misspecification but requires an asymptotic normal approximation, which performs well only in the mid-sample. This is why the bootstrap intervals exhibit more undercoverage for $h = 1$ than 5.

One comment about the choice of k. A larger k leads to more tail observations and hence shorter confidence intervals, but is subject to a larger approximation bias due to including too many mid-sample data. This bias and variance trade off indicates that the choice of k is difficult, especially when n is only moderate. It is actually impossible to choose a uniformly best k allowing the underlying CDF to be flexible (see Theorem 1 of Müller and Wang (2017)). The CDFs in my Monte Carlo are all well behaved so that the k as large as 40% of the sample size performs well. This is seen in Table 3, which reports the numbers for $k = 20$ and 50. See also Table 8 in Section 3.5.

3.2 Comparison to kernel smoothing with i.i.d. data

The new approach takes only the nearest neighbor in each time series, which raises the question of efficiency loss. I answer this question by comparing the fixed-k approach with the kernel smoothing methods proposed by Gardes, Girard, and Lekina (2010). In particular, I first pool the panel data into a cross-sectional sample. Suppose the object of interest is still $Q_{Y|X=Q_X(\tau_x)}(\tau_y)$. I follow Gardes, Girard, and Lekina (2010) to pick the bin $B_{\hat{Q}_X(\tau_x)}(b_{nT})$

Table 2: Finite sample performance of inference about extremal conditional quantile, no model specification

n	200 (97.5% quantile)				500 (99% quantile)			
T	200		500		200		500	
	Cov	Lgth	Cov	Lgth	Cov	Lgth	Cov	Lgth
	Joint Normal							
fixed-k	0.96	0.65	0.96	0.65	0.95	0.57	0.94	0.57
QR	1.00	1.28	1.00	0.80	1.00	1.81	1.00	1.13
BEQ	0.93	0.63	0.92	0.64	0.92	0.55	0.91	0.56
	Joint Student's t							
fixed-k	0.97	2.42	0.95	2.36	0.97	2.88	0.97	2.77
QR	1.00	3.53	1.00	2.26	1.00	6.23	1.00	3.94
BEQ	0.94	2.31	0.93	2.25	0.93	2.83	0.95	2.72
	Conditional Pareto							
fixed-k	0.95	9.30	0.97	7.45	0.84	16.3	0.94	12.5
QR	0.00	$>10^3$	0.00	$>10^3$	0.00	$>10^3$	0.00	$>10^3$
BEQ	0.95	12.5	0.96	8.91	0.80	26.8	0.93	15.2

Note: Entries are coverages and lengths of the CIs for $Q_{Y|X=Q_X(0.95)}(1-5/n)$. See the main text for the description of the three approaches and the data generating processes. Confidence level is 5%. Based on 500 simulation draws.

Table 3: Finite sample performance of inference about extremal conditional quantile, no model specification

n	200 (99.5% quantile)				500 (99.8% quantile)			
T	200		500		200		500	
	Cov	Lgth	Cov	Lgth	Cov	Lgth	Cov	Lgth
	Joint Normal							
fixed-k(k=20)	0.95	1.82	0.96	1.83	0.97	1.69	0.96	1.70
QR	1.00	1.19	1.00	0.75	1.00	1.18	1.00	1.07
BEQ	0.63	0.62	0.64	0.59	0.64	0.57	0.65	0.59
	Joint Student's t							
fixed-k(k=20)	0.96	4.71	0.96	4.69	0.96	5.62	0.97	5.61
fixed-k(k=50)	0.94	3.91	0.92	3.90	0.95	4.85	0.92	4.73
QR	1.00	8.51	0.68	5.51	1.00	8.47	1.00	11.5
BEQ	0.62	2.01	0.60	2.02	0.63	2.57	0.61	2.56
	Conditional Pareto							
fixed-k(k=20)	0.98	27.6	0.98	26.1	0.94	48.1	0.97	40.5
QR	0.00	$>10^3$	0.00	$>10^3$	0.00	$>10^3$	0.00	$>10^3$
BEQ	0.71	25.9	0.63	30.4	0.78	76.2	0.77	43.9

Note: Entries are coverages and lengths of the CIs for $Q_{Y|X=Q_X(0.5)}(1-1/n)$. See the main text for the description of the three approaches and the data generating processes. Confidence level is 5%. Based on 500 simulation draws.

centered at $\hat{Q}_X(\tau_x)$ with a bandwidth b_{nT}. Since there is no theoretical justification for the optimal choice of b_{nT}, I take the rule-of-thumb choice $c(nT)^{-1/5}$ with different values of the constant c. Now a certain choice of b_{nT} leads to a certain collection of $Y's$ whose paired $X's$ are in the bin $B_{\hat{Q}_X(\tau_x)}(b_{nT})$. Order these induced $Y's$ descendingly into $\{Y_{(1)} \geq Y_{(2)} \geq ... \geq Y_{(m)}\}$ where m denotes the local sample size determined by the bandwidth. It is approximately nTb_n in the kernel smoothing and is n in the nearest neighbor.

Given the induced Y, the conditional quantile is estimated as $\hat{Q}_{Y|X=Q_X(\tau_x)}(\tau_y) = Y_{(\tau_y m)}$. Gardes, Girard, and Lekina (2010) show that under $m(1-\tau_y) \to \infty$ and some other regularity conditions,

$$\sqrt{m(1-\tau_y)}\left(\frac{\hat{Q}_{Y|X=Q_X(\tau_x)}(\tau_y)}{Q_{Y|X=Q_X(\tau_x)}(\tau_y)} - 1\right) \xrightarrow{d} \mathcal{N}(0, 1/\xi_0^2)$$

where ξ_0 depends $Q_X(\tau_x)$. Then the CI of $Q_{Y|X=Q_X(\tau_x)}(\tau_y)$ is constructed by the delta method and plugging in some consistent estimator of ξ_0. One choice they propose is the Hill-type estimator

$$1/\hat{\xi} = \frac{1}{k-1}\sum_{i=1}^{k-1} i \log(Y_{(i)}/Y_{(i+1)}) \tag{18}$$

for some choice of $k < m$.

For comparison, I implement the fixed-k approach by using the panel data and the above kernel estimator by pooling the data. In particular, I implement the conditional Pareto DGP in the previous experiment with $n = 200$ and T ranging from 25 to 500. For the fixed-k CI, I set $k = 50$. For the kernel method, I implement $c \in \{0.1, 0.25, 0.5, 1, 2\}$ and set k (in the Hill-type index estimator (18)) as the largest integer less than $m/4$.

Table 4 presents the coverage and the length of both methods. Several interesting observations are made. First, the kernel approach is sensitive to the choice of the bandwidth. In particular, a correct coverage relies on a narrow window of the bandwidth choice. A larger choice can lead to a substantial undercoverage since the smoothing bias dominates quickly in the tail. Second, when T is only moderately large (say 25 and 50), the fixed-k CIs are much shorter than the kernel one given both of them have good coverage properties. This is because the fully nonparametric method ignores the Pareto tail information, which is utilized by the fixed-k method. Third, when T is very large, say 500, choosing only the nearest neighbor does incur an efficiency loss as we compare the length between the fixed-k and the kernel CIs. But such loss is approximately in a factor of 2 or 3 instead of $T^{1/2}$. This means a general covariate dependent tail is very difficult to estimate in a fully nonparametric way.

It is also worth mentioning that the conditional tail index estimator (18) is not invariant

Table 4: Finite sample performance of inference about extremal conditional quantile, comparison with kernel method

T	25		50		100		200		500	
	Cov	Lgth	Cov	Lgth	Cov	Lgth	Cov	Lgth	Cov	Lgth
fixed-k	0.97	24.3	0.97	21.1	0.97	19.8	0.97	16.8	0.98	15.3
NP(c=0.1)	0.86	59.6	0.91	50.1	0.89	30.0	0.94	24.4	0.93	14.6
NP(c=0.25)	0.96	50.5	0.94	33.1	0.96	19.0	0.93	13.3	0.96	9.15
NP(c=0.5)	0.95	27.6	0.93	17.1	0.94	12.7	0.93	9.28	0.95	6.36
NP(c=1)	0.95	20.7	0.93	13.9	0.90	9.84	0.89	7.14	0.88	4.77
NP(c=2)	0.58	24.3	0.37	15.6	0.24	10.1	0.14	6.86	0.11	4.18

Note: Entries are coverages and lengths of the CIs for $Q_{Y|X=Q_X(0.5)}(1-1/n)$ under the conditional Pareto DGP with no shifting. See the main text for the description of the two approaches and details of the DGP. Confidence level is 5%. Based on 500 simulation draws.

Table 5: Finite sample performance of inference about extremal conditional quantile, comparison with kernel method

T	25		50		100		200		500	
	Cov	Lgth	Cov	Lgth	Cov	Lgth	Cov	Lgth	Cov	Lgth
fixed-k	0.97	24.3	0.97	21.1	0.97	19.8	0.97	16.8	0.98	15.3
NP(c=0.1)	0.78	57.8	0.88	42.4	0.83	23.2	0.89	20.0	0.87	11.7
NP(c=0.25)	0.88	37.2	0.87	21.4	0.90	15.8	0.90	11.2	0.90	7.27
NP(c=0.5)	0.89	22.0	0.87	14.0	0.87	10.3	0.88	7.45	0.85	5.14
NP(c=1)	0.85	17.1	0.80	11.4	0.83	8.00	0.83	5.75	0.79	3.85
NP(c=2)	0.34	20.0	0.21	12.2	0.11	7.99	0.08	5.43	0.04	3.38

Note: Entries are coverages and lengths of the CIs for $Q_{Y|X=Q_X(0.5)}(1-1/n)$ under the conditional Pareto DGP with Y shifted by 1. See the main text for the description of the two approaches and details of the DGP. Confidence level is 5%. Based on 500 simulation draws.

to location shift but the fixed-k approach is. Such location invariance also contributes to the efficiency as depicted in Table 5. In particular, I use the same DGP as above except that Y is shifted by adding 1. This shifting does not affect the fixed-k CIs but does impede the performance of the conditional tail index estimator (18). Consequently, the kernel CIs based on plugging in (18) exhibit undercoverage. Some unreported results show that if (18) is replaced with some location invariant estimator, the kernel CIs are substantially longer and comparable with the fixed-k ones.

As a final remark of this subsection, I also implement the standard kernel weighted quantile regression method designed for the mid-sample quantiles (cf. Chapter 10 of Li and Racine (2007)). Given a large T, the target $1-1/n$ conditional quantile is relatively in the mid-sample, and hence the confidence interval based on asymptotic normality might work. However, unreported Monte Carlo simulations show that this method works only if

T is substantially, say 5 times, larger than n. In our setups, it is strictly dominated by the method proposed by Gardes, Girard, and Lekina (2010).

3.3 Extremal conditional quantile with stochastic unobserved effects

In this section, I impose the linear regression model $Y_{it} = \alpha_i + X_{it}\beta_0 + u_{it}$ and assume data are i.i.d. across i. For time dependence, I set $\alpha_i = T^{-1}\sum_{t=1}^{T} X_{it}$ and $X_{it} = \rho X_{it-1} + e_{it}$ with $e_{it} \sim^{iid} \mathcal{N}(0, (1-\rho^2))$ and $X_{i0} \sim \mathcal{N}(0,1)$. The distribution of u_{it} conditional on $X_{it} = x$ is as follows.

1. **Conditional Normal** $u_{it}|X_{it} = x \sim \mathcal{N}(0, 1 + x^2)$.

2. **Conditional Student's t** $u_{it}|X_{it} = x \sim t(2 + |x|)$.

3. **Conditional Pareto** $u_{it}|X_{it} = x \sim \pm\mathrm{Pa}(\xi(x))$, that is, $\mathbb{P}(u_{it} \le y|X_{it} = x) = 1/2 + (1 - (1+y)^{-1/\xi(x)})/2$ for $y \ge 0$, and $\mathbb{P}(u_{it} \le y|X_{it} = x) = (-y+1)^{-1/\xi(x)}/2$ for $y \le 0$ where $\xi(x) = x - Q_X(\tau_x) + 0.5$.

I use the same three approaches as in the Section 3.1 to construct CIs for the extremal conditional quantile $Q_{Y_{it}|X_{it}=Q_X(\tau_x)}(\tau_y) = Q_{\alpha_i+u_{it}|X_{it}=Q_X(\tau_x)}(\tau_y) + Q_X(\tau_x)\beta_0$. This specification assumes that β_0 is not included for the conditional value. I relax this in the following subsection. In particular, the fixed-k approach is conducted in two ways: with or without using any least squares estimator of $\hat{\beta}$ in (10). For the former (fixed-k w. LS), I first estimate β_0 using the standard demean estimator $\hat{\beta}$ and back out $\hat{\varepsilon}_{it} = Y_{it} - X_{it}\hat{\beta}$ and then implement the steps in Section 2.3 to construct the CIs for the conditional quantile of ε_{it}. The CIs for $Q_{Y_{it}|X_{it}=Q_X(\tau_x)}(\tau_y)$ is obtained by adding back $\hat{Q}_X(\tau_x)\hat{\beta}$ where $\hat{Q}_X(\tau_x)$ is the sample τ_x quantile in the pooled X's. For the one ignoring the fixed effects structure (fixed-k w/o LS), I directly use Y_{it} and X_{it} and apply Steps 1-3 in Section 2.2.

Tables 6 present the results for $n \in \{100, 200\}$ and $T \in \{25, 50, 200, 500\}$. Several interesting observations can be found. First, the error in the conditional t and conditional Pareto models does not have a finite variance when τ_x is 0.5, and hence the fixed effects estimator of β behaves poorly. This leads to a poor performance of the fixed-k approach if the linear regression model is utilized. This problem can be solved by using the least absolute deviation (LAD) estimator, which is implemented in the next table. In comparison, the fixed-k CIs without using the linear regression model always perform well given a large

Table 6: Finite sample performance of inference about extremal conditional quantile, non-dynamic model with random effects

n	200 (99.5% quantile)				100 (99% quantile)			
T	200		500		25		50	
	Cov	Lgth	Cov	Lgth	Cov	Lgth	Cov	Lgth
				Conditional Normal				
fixed-k w. LS	0.94	2.23	0.93	2.13	0.80	2.85	0.88	2.34
fixed-k w/o LS	0.93	2.22	0.92	2.14	0.53	3.17	0.76	2.62
QR	0.00	3.04	0.00	2.19	1.00	3.10	1.00	2.96
Boot	0.73	0.77	0.67	0.71	0.73	1.11	0.81	0.92
				Conditional Student's t				
fixed-k w. LS	0.95	15.3	0.94	15.5	0.93	10.0	0.96	10.4
fixed-k w/o LS	0.95	15.3	0.94	15.5	0.91	10.0	0.93	10.3
QR	1.00	26.5	1.00	7.98	0.99	11.0	1.00	15.0
Boot	0.56	11.5	0.60	11.0	0.47	6.32	0.51	6.01
				Conditional Pareto				
fixed-k w. LS	0.00	16.2	0.00	5.90	0.02	59.6	0.01	58.1
fixed-k w/o LS	0.97	18.8	0.97	16.9	0.95	16.0	0.96	15.4
QR	0.00	$>10^3$	0.00	$>10^3$	1.00	$>10^3$	1.00	$>10^3$
Boot	0.71	16.2	0.67	16.8	0.76	313	0.78	31.4

Note: Entries are coverages and lengths of the CIs for $Q_{Y|X=Q_X(0.1)}(1-1/n)$. See the main text for the description of different approaches and the data generating processes. Confidence level is 5%. Based on 500 simulation draws.

enough sample size. Second, the QR approach still suffers undercoverage in all three specifications since the normal and the student's t DGPs have nonlinear heteroskedasticity and the conditional Pareto DGP violates the constant tail shape condition. Finally, the bootstrap method performs poorly if the extremal quantile under investigation is too far in the tail.

3.4 Extremal conditional quantile with constant unobserved effects

Now I consider the same linear regression model (10) as in the previous experiment but with fixed effects. In particular, I set $\alpha_i = i$ and use the same DGP for the conditional Normal and the conditional Student's t distribution as before. Regarding the conditional Pareto, the condition that $\bar{u}_i \xrightarrow{p} 0$ is violated since X_{it} may occasionally take large values. I modify this by setting $\xi(x) = \max\{\min\{x - Q_X(\tau_x), 0.25\}, -0.25\} + 0.5$. This assumption restricts the effect of X_{it} on the tail shape but still allows for a high nonlinearity. Since the LS estimator behaves poorly in the conditional Pareto DGP due to heavy tail errors, I implement the LAD estimator for β instead. Table 7 present the results the same parameter combination and the same three CIs as before. The fixed-k CI with the nearest neighbor again delivers excellent coverage and length performance.

Table 7: Finite sample performance of inference about extremal conditional quantile, non-dynamic model with fixed effect

n	200 (99.5% quantile)				100 (99% quantile)			
T	200		500		25		50	
	Cov	Lgth	Cov	Lgth	Cov	Lgth	Cov	Lgth
				Joint Normal				
fixed-k with LS	0.96	2.25	0.99	2.17	0.81	2.59	0.88	2.40
QR	0.05	2.92	0.00	2.16	1.00	2.46	0.97	2.57
BEQ	0.68	0.76	0.62	0.71	0.66	0.84	0.69	0.81
				Conditional Student's t				
fixed-k with LS	0.94	14.9	0.92	16.3	0.92	10.0	0.96	10.0
QR	1.00	25.3	1.00	7.87	0.99	10.3	1.00	14.0
BEQ	0.53	8.00	0.60	9.33	0.54	5.78	0.53	6.77
				Conditional Pareto				
fixed-k with LAD	0.95	21.1	0.95	19.2	0.96	16.5	0.96	17.8
QR	1.00	212	0.98	53.7	0.99	42.7	0.99	65.2
BEQ	0.73	27.6	0.72	22.3	0.79	20.2	0.73	26.7

Note: Entries are coverages and lengths of the CIs for $Q_{Y_i|X=Q_X(0.5)}(1-1/n)$. See the main text for the description of different approaches and the data generating processes. Confidence level is 5%. Based on 500 simulation draws.

3.5 Forecast with dynamic panel

I now consider the following dynamic panel with random effects for forecasting purpose

$$Y_{it} = \alpha_i + Y_{i,t-1}\rho_0 + X_{it}\beta_0 + u_{it}.$$

This model is also implemented for Monte Carlo simulation by Galvao Jr. (2011), whose target is the mid-sample instead of the extremal quantile. I choose $\rho_0 = 0.5$, $\beta_0 = 1$, and $\alpha_i = T^{-1}\sum_{t=1}^{T} X_{it}$ where X_{it} is stationary AR(1) as in the previous subsection. The conditional distribution u_{it} given $X_{it} = x$ is as follows.

1. **Conditional Normal** $u_{it}|X_{it} = x \sim \mathcal{N}(0, 1 + (x\beta_0)^2)$.

2. **Conditional Student's t** $u_{it}|X_{it} = x \sim t(2 + |x\beta_0|)$.

3. **Conditional Pareto** $u_{it}|X_{it} = x \sim \pm\mathrm{Pa}(\xi(x\beta_0))$, that is, $\mathbb{P}(u_{it} \le y|X_{it} = x) = 1/2 + (1 - (1+y)^{-1/\xi(x\beta_0)})/2$ for $y \ge 0$, and $\mathbb{P}(u_{it} \le y|X_{it} = x) = (-y+1)^{-1/\xi(x\beta_0)}/2$ for $y \le 0$ where $\xi(x) = x - Q_X(\tau_x) + 0.5$.

These specifications assume that the tail of the error term u_{it} depends on $x_{it}\beta_0$ only but not the lag dependent variable. Otherwise it will be very complicated to maintain the strict stationarity. Given this specification, the one period ahead conditional extremal quantile is then $Q_{Y_{i,t+1}|Y_{it}=y,X_{it}=x}(\tau_y) = y\rho_0 + x\beta_0 + Q_{\varepsilon_{i,t+1}|X_{it}\beta_0=x\beta_0}(\tau_y)$ where ε_{it} again denotes $\alpha_i + u_{it}$. I construct the fixed-k CIs and the QR estimates for the quantiles with $\tau_y = 1 - h/n, y = Q_Y(\tau_x)$, and $x = Q_X(\tau_x)$ for $h = 1, n \in \{100, 200\}$ and $\tau_x = 0.5$.

More specifically, the fixed-k approach is implemented as follows. First, I construct the instrument variable estimator $(\hat{\rho}, \hat{\beta})$ proposed by Arellano and Bond (1991) using $Y_{i,t-2}$ as the instrument, and back out $\hat{\varepsilon}_{it} = Y_{it} - Y_{i,t-1}\hat{\rho} - X_{it}\hat{\beta}$. Then I follow the steps described in Section 2.3 to construct CIs for $Q_{\varepsilon_{i,t+1}|X_{it}\beta_0=Q_X(\tau_x)\beta_0}(\tau_y)$, using the induced $\hat{\varepsilon}$ associated with the rank statistics of $X_{it}\hat{\beta}$. Thus the estimation error in $\hat{\beta}$ is included in the ranking. Finally, I add back $\hat{Q}_Y(\tau_x)\hat{\rho} + \hat{Q}_X(\tau_x)\hat{\beta}$ for the CIs of $Q_{Y_{i,t+1}|Y_{it}=y,X_{it}=x}(\tau_y)$, where $\hat{Q}_Y(\tau_x)$ and $\hat{Q}_X(\tau_x)$ are the empirical quantiles of Y and X using pooled data. There is no other existing suggestion on this problem. For comparison purpose, I modify the instrument variable quantile regression (Modified IVQR) to the panel data case, which is originally developed by Chernozhukov and Hansen (2005) for cross-sectional data. Specifically, I first construct, for each i, the IVQR estimator using T observations to obtain the conditional quantile estimate $\hat{Q}_{Y_{i,t+1}|Y_{it}=Q_Y(\tau_x),X_{it}=Q_X(\tau_x)}(\tau_y)_i = \hat{\alpha}_i(\tau_y) + \hat{Q}_Y(\tau_x)\hat{\rho}(\tau_y)_i + \hat{Q}_X(\tau_x)\hat{\beta}(\tau_y)_i$. The subscript i indicates that this estimate is conducted for each i. Second, I simply take the 2.5% and 97.5% quantiles in $\{\hat{Q}_{Y_{i,t+1}|Y_{it}=Q_Y(\tau_x),X_{it}=Q_X(\tau_x)}(\tau_y)_i\}_{i=1}^n$ as the CI.

Table 8 presents the results with $T \in \{25, 50, 200, 500\}$. Similarly as in previous experiments, the fixed-k approach works well with large enough sample sizes and worse as the sample size gets smaller. The estimation error in $\hat{\beta}$ also contributes to the approximation error. Choosing a smaller k alleviates the undercoverage at the cost of a longer interval. The quantile regression approach is again subject to misspecification, which is due to the nonlinearity in the tail shape and the complicate heteroskedasticity.

3.6 Conditional tail index

The last experiment examines the CIs of the conditional tail index. I consider the following three DGPs.

1. **Joint Student's t** (X_{it}, Y_{it}) is i.i.d. across i and t and is distributed as $t_v(\mu, \Sigma)$ with $v = 2$, $\mu = [0, 0]^\intercal$, and $\Sigma = [1, 0.5; 0.5, 1]$.

2. **Conditional Pareto** $X_{it} = \rho X_{it-1} + u_{it}$ with $u_{it} \sim \mathcal{N}(0, (1 - \rho^2))$ and i.i.d. across i and t, and $X_{i1} \sim \mathcal{N}(0, 1)$. $Y_{it}|X_{it} = x \sim \text{Pa}(1/\xi(x))$, that is, $\mathbb{P}(Y_{it} \le y|X_{it} = x) = 1 - y^{-1/\xi(x)}$ for $y \ge 1$. Set $\xi(x) = x - Q_X(\tau_x) + 0.5$.

Table 8: Finite sample performance of inference about one period ahead extremal conditional quantile, dynamic model with random effects

n	200 (99.5% quantile)				100 (99% quantile)			
T	200		500		25		50	
	Cov	Lgth	Cov	Lgth	Cov	Lgth	Cov	Lgth
				Conditional Normal				
fixed-k(k=10)	0.94	2.32	0.98	2.44	0.79	3.14	0.86	2.85
fixed-k(k=20)	0.96	2.20	0.95	2.21	0.73	2.69	0.82	2.52
Modified IVQR	0.00	2.53	0.00	2.54	1.00	5.48	1.00	3.46
				Conditional Student's t				
fixed-k(k=10)	0.94	19.5	0.94	18.6	0.96	12.5	0.94	13.6
fixed-k(k=20)	0.96	14.3	0.94	15.3	0.89	10.3	0.94	10.7
Modified IVQR	1.00	24.4	1.00	8.87	0.98	13.8	1.00	14.0
				Conditional Pareto				
fixed-k(k=10)	0.93	28.0	0.95	24.0	0.86	36.0	0.86	31.7
fixed-k(k=20)	0.96	20.6	0.94	17.2	0.94	10.4	0.92	18.1
Modified IVQR	0.00	$>10^3$	0.00	$>10^3$	0.99	15.1	1.00	9.90

Note: Entries are coverages and lengths of the CIs for $Q_{Y_{i,t+1}|Y_{it}=Q_Y(\tau_x),X_{it}=Q_X(\tau_x)}(\tau_y)$ with $\tau_x = 0.5$ and $\tau_y = 0.995$ and 0.99. See the main text for the description of two approaches and the data generating processes. Confidence level is 5%. Based on 500 simulation draws.

3. Independent F $X_{it} = \rho X_{it-1} + u_{it}$ with $u_{it} \sim \mathcal{N}(0, (1-\rho^2))$ and $X_{i1} \sim \mathcal{N}(0,1)$. Y_{it} is F(4,4) and independent of X_{it}. Y_{it} and u_{it} are both i.i.d. across i and t.

The normal distribution is replaced with the independent F(4,4), so that the true conditional tail index is 0.5 in all three designs. I set $\Xi = [0,1]$ to avoid the true ξ being on the boundary. Table 9 reports the coverage and length of the fixed-k CI (fixed-k) based on inverting (16) and that based on the Hill's estimator (Hill) and the maximum likelihood estimator (MLE) as described in Section 2.6. I set $\tau_x = 0.95$, $\rho = 0.5$ and $n = T = 1000$, and choose $k \in \{20, 50, 100, 200\}$. This is to make sure k is relatively small to n but still large enough for the increasing-k asymptotic approximation. As expected from the theoretical derivation, the fixed-k CIs deliver excellent coverage and length when k is small. As k grows, the MLE gradually performs better. A heuristic rule-of-thumb for choosing the MLE instead of the fixed-k CI is based on whether k is over 100 or not, provided n is substantially larger. The Hill's estimator is not location invariant and thus heavily relies on the Pareto tail approximation. This is why it has short length and precise coverage when the underlying distribution is exactly Pareto, but has large undercoverage when the DGP is student's t or F.

Table 9: Small sample performance for inference about the conditional tail index

k	20		50		100		200	
	Cov	Lgth	Cov	Lgth	Cov	Lgth	Cov	Lgth
	Conditional Student's t							
fixed-k	0.97	0.76	0.96	0.69	0.92	0.53	0.81	0.37
Hill	0.88	0.41	0.83	0.26	0.86	0.18	0.96	0.14
MLE	0.76	0.80	0.86	0.69	0.86	0.55	0.75	0.38
	Conditional Pareto							
fixed-k	0.95	0.75	0.94	0.68	0.90	0.54	0.88	0.39
Hill	0.96	0.48	0.95	0.30	0.93	0.21	0.94	0.14
MLE	0.81	1.00	0.90	0.79	0.94	0.59	0.95	0.42
	Independent F							
fixed-k	0.96	0.75	0.94	0.69	0.92	0.54	0.94	0.39
Hill	0.97	0.49	0.94	0.31	0.63	0.24	0.02	0.19
MLE	0.77	0.79	0.90	0.70	0.92	0.56	0.94	0.41

Note: Entries are coverages and lengths of CIs on the tail index of the underlying conditional distribution, based on the largest k order statistics. See the main text for a description of the two types of CIs and the DGPs. Confidence level is 5%. Based on 500 Monte Carlo simulations.

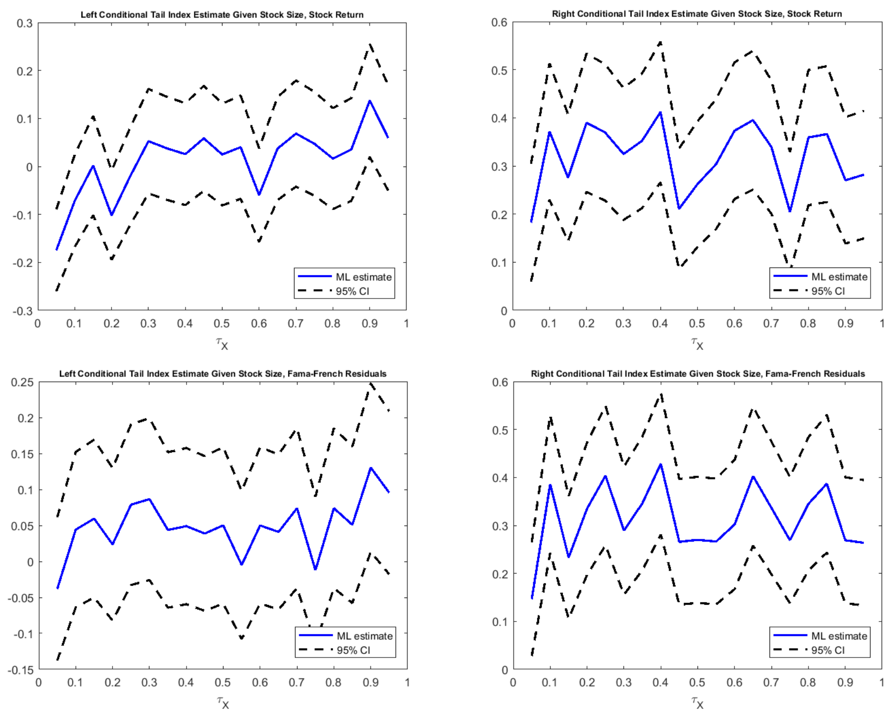
4 Empirical applications

4.1 Tail risk in stock returns

Tail risk in stock returns has been an important topic in finance. See Backus, Chernov, and Martin (2011) and Bollerslev and Todorov (2011) among many others. Due to limited observations, tail properties are usually difficult to study using time series data only. Motivated by this, Kelly and Jiang (2014) apply panel data on stock returns and assume the left tail of the i-th stock at period t has a time varying tail index $\xi_{it} = \lambda_t / a_i$. The λ_t term captures the dynamics that are shared by all assets and a_i measures the stock specific tail risk. I relax such ratio structure by considering a covariate dependent tail index. In particular, I illustrate with the stock size as the covariate.

I follow convention to use monthly returns of NYSE/AMEX/NASDAQ stocks with share codes 10 and 11. To obtain a large T, I only use the stocks that are traded for more than 120 months. This leads to an unbalanced panel dataset with $n = 1744$ and T ranging from 121 to 1104. Given such a large n, I apply the maximum likelihood estimator (17) with $k = 349$ (20% of n) and the corresponding CIs based on its asymptotic normality. Top panels in Figure 2 plot the estimated left and right conditional tail indices of stock returns given stock size equals to its τ_x unconditional quantile. Lower panels plot the same estimates and CIs based on the residuals of the Fama-French three-factor regression. The results suggest that large stocks tend to exhibit heavier left tails, but such relation is weak for right tail. This is coherent with Chen, Hong, and Stein (2001) who use linear regression to find such pattern.

Figure 2: Plot of QR estimate and fixed-k CIs of conditional tail Index of stock return or Fama-French residual given stock size



Note: This figure plots the QR estimates (solid line) and 95% fixed-k CIs (dash line) of the left and right conditional tail indices of the stock returns (upper row) or the residuals from the Fama-French 3 factors regression (lower row). See main context for details of these two approaches. Data are available from http://www.crsp.com.

In addition, I also examine the extremal conditional quantiles. Figure 3 plots the QR estimates and the fixed-k CIs of the τ_y conditional quantiles of stock returns and the Fama-French residuals, conditional on the τ_x quantile of the stock size. In particular, I present the results for $\tau_x \in \{0.05, 0.5, 0.95\}$ and $\tau_y = h/n$ and $1-h/n$ for $\{1, 2, ..., 10\}$. The QR estimate is based on regressing either stock return or the Fama-French residual on a constant and the stock size. The fixed-k CIs are based on $k = 100$. This is set for good coverage but possibly conservative length. The figure shows that the QR estimates are outside the fixed-k CIs for the left tail, but not the right one, indicating that the left conditional extremal quantiles are highly nonlinear but the right ones are close to be linear.

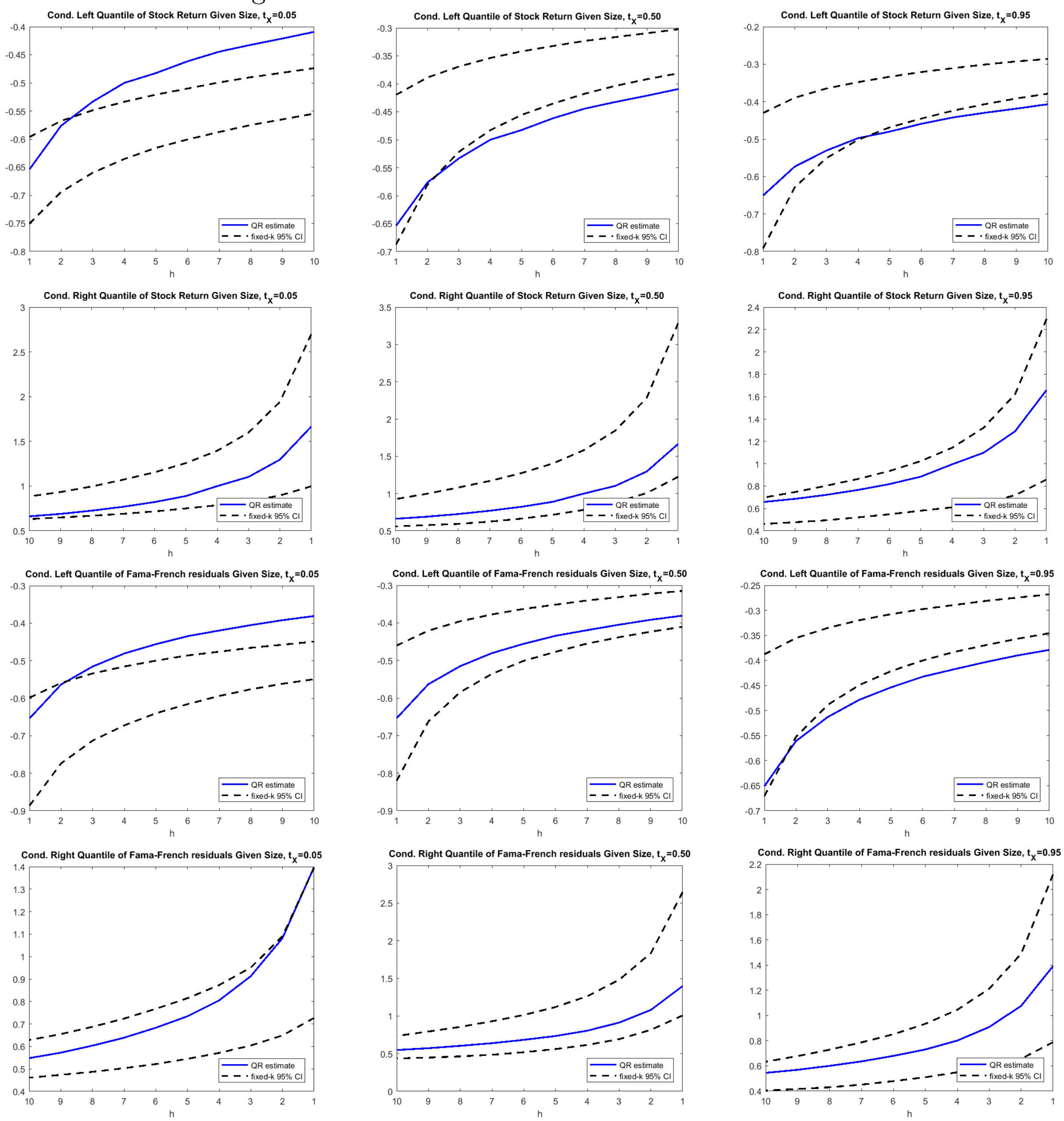
4.2 Low Infant Birthweight and Mother' Characteristic

The second example examines the effect of mother' demographic characteristics on infant birthweight. Chernozhukov and Fernández-Val (2011) study this problem by applying the EQR approach to extremely low birthweight and find substantially different results than the previous literature that uses mid-sample QR (see Abrevaya (2001) and Koenker and Hallock (2001)). I use the same dataset as Chernozhukov and Fernández-Val (2011) and complement their study by examining the effect of mother's net weight gain. This variable is excluded by them to avoid endogeneity but is intuitively correlated with infant's birthweight.

I first consider extremely low quantiles of infant birthweight conditional on mother's net weight gain equal to 7kg, 27kg, and 49kg, corresponding to its 5%, 50%, and 95% unconditional quantiles, respectively. Figure 4 plots the QR regression estimate (solid line) and the 95% fixed-k CIs for τ_y ranging from 0.002 to 0.02. In particular, there are 198377 observations which are all used to run a quantile regression of birthweight on a constant and mother's net weight gain. Regarding the fixed-k approach, I randomly decompose data into a panel with $n = T = 445$ and apply the procedure in Section 2.2 with $k = 100$. I repeat this random decomposition by 100 times and plot the average of the boundary points of the fixed-k CIs. The QR estimates are inside the fixed-k CIs (but very close to the upper bound) when mother's net weight gain is either very low or high but outside otherwise. This suggests that the location-shift model condition might be violated here, and mother's net weight gain has a large effect on extremely low quantile of infant birthweight. This is coherent with the QR results based on mid-sample quantiles (see Abrevaya (2001)).

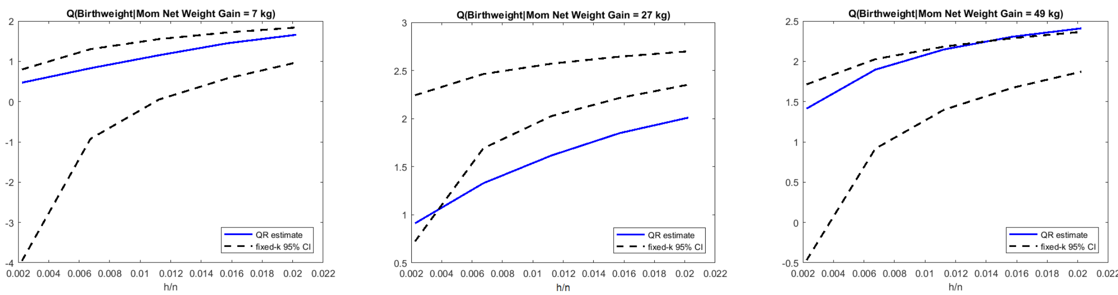
Next, I examine whether an extremal quantile regression is sufficient to capture the tail shape. To this end, I first run a 5% quantile regression of infant's birthweight on the same controls as in Chernozhukov and Fernández-Val (2011). Then I collect the residuals and

Figure 3: Plot of QR estimate and fixed-k CIs of conditional quantile of stock return and Fama-French residual given stock size



Note: This figure plots the QR estimates (solid line) and 95% fixed-k CIs (dash line) of the right and left conditional quantiles of the stock returns (upper two rows) or residuals from the Fama-French 3 factors regression (lower two rows). See main context for details of these two approaches. Data are available from http://www.crsp.com.

Figure 4: Plot of QR estimate and fixed-k CIs of Conditional Quantile of Birthweight Given Mother's Net Weight Gain



Note: The solid line plots the fitted value in the quantile regression of birthweight on mother's net weight gain and a constant at levels from 0.002 to 0.02. The dash line plots the fixed-k confidence intervals using the same input based on $k = 100$. See main context for details of these two approaches. Data are available from Chernozhukov and Fernández-Val (2011).

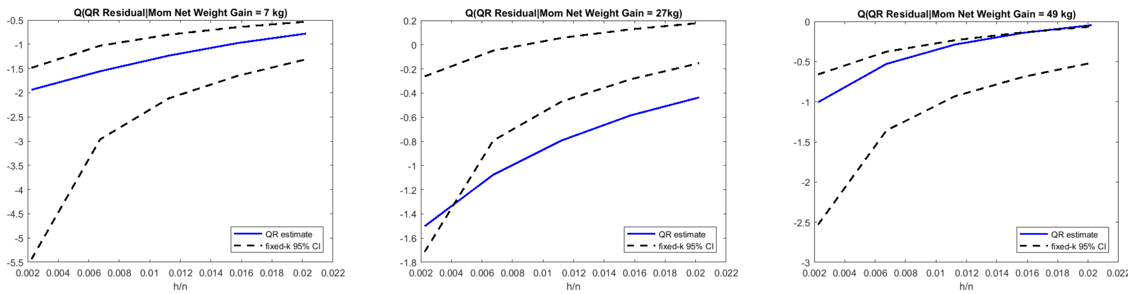
conduct the previous exercise. Figure 5 plots the fixed-k CIs and the QR estimates of the conditional quantile of the residuals given mother's net weight gain. The pattern is the same as in Figures 4, and hence reinforces my previous finding on the nonlinearity of the extremal conditional quantile.

5 Concluding Remarks

This paper develops a new framework on inference about conditional tail properties using panel data. The key insight is that the induced order statistics in each time series can be treated as approximately stemming from the true conditional distribution, and the large order statistics among these induced values can then be used to study the conditional tail. By focusing on the induced order statistics, I essentially reduce the conditional tail problem into an unconditional one, so that existing approaches about unconditional tail properties become applicable. Monte Carlo simulations show that the new method delivers excellent small sample performance in terms of coverage probability and length.

The new method is substantially more flexible than quantile regression because the latter assumes that the extremal conditional quantile is a parametric location-shift model, which is an empirical concern in some applications. If a linear regression model is imposed, the new method is easily combined with any existing consistent estimator of the structural parameter and allows arbitrary unobserved heterogeneity. If a large cross-sectional sample is available, the econometrician can first randomly decompose the sample into a panel/repeated cross-

Figure 5: Plot of QR estimate and fixed-k CIs of Conditional Quantile of QR Residuals Given Mother's Net Weight Gain



Note: The solid line plots the fitted value in the quantile regression of a previous extremal quantile regression residuals on mother's net weight gain and a constant at levels from 0.002 to 0.02. The dash line plots the fixed-k confidence intervals using the same input based on $k = 100$. See main context for details of these two approaches. Data are available at Chernozhukov and Fernández-Val (2011).

sectional dataset and then apply the new method.

A Appendix

A.1 Omitted Details and Preliminary Conditions in Section 2

This section provides more details about Conditions 1 and 2. I focus on the latter since it covers Condition 1 by treating β_0 as a known parameter.

First, Condition 2.1 assumes that

$$\sup_{\beta \in \mathcal{B}} \mathbb{E}[\left|X(\beta)_{(\tau_x T)} - Q_{X(\beta)}(\tau_x)\right|] = O(T^{-1/2}) \tag{19}$$

where I suppress the subscript i in this subsection for notational ease. This holds under the following sufficient (but not necessary) regularity conditions.

Condition A.1 $(\mathbf{X}_t)_{t \in \mathbb{Z}}$ is strictly stationary with $\mathbb{E}[||\mathbf{X}_t||^4] < \infty$ and ϕ-mixing with the mixing coefficient satisfying $\sum_{t=1}^{\infty} \phi(t)^{1/2} < \infty$.

Condition A.2 $f_{X(\beta)}(x)$ is uniformly continuously differentiable in β and x and bounded from 0 and ∞.

Condition A.1 imposes ϕ-mixing, which can be replaced with much more general forms of weak dependence. See Wu (2004) for an overview. Condition A.2 requires that the joint density of $\mathbf{X}$ is smooth and uniformly bounded. Define $g(\mathbf{x}; \beta, r) = \mathbf{1}[\mathbf{x}^\intercal \beta \leq r]$ for $\mathbf{x} \in \mathbb{R}^{\dim(\beta)}$ and $(\beta, r) \in \mathcal{B} \times \mathbb{R}$.

Also define $P_T g(\beta,r) = T^{-1}\sum_{t=1}^{T} g(\mathbf{X}_t;\beta,r)$ and $Pg(\beta,r) = \mathbb{E}[g(\mathbf{X}_t;\beta,r)]$. Then the definition of the empirical quantile yields

$$\begin{aligned} O(T^{-1/2}) &= T^{-1/2}\sum_{t=1}^{T}\left(\mathbf{1}\left[\mathbf{X}_t^{\intercal}\beta \le X(\beta)_{(\tau_x T)}\right] - \tau_x\right) \\ &= T^{1/2}(P_T - P)g(\beta, X(\beta)_{(\tau_x T)}) + T^{1/2}\left(Pg(\beta, X(\beta)_{(\tau_x T)}) - Pg(\beta, Q_{X(\beta)}(\tau_x))\right) \\ &= T^{1/2}(P_T - P)g(\beta, X(\beta)_{(\tau_x T)}) + f_{X(\beta)}(\dot{X})T^{1/2}(X(\beta)_{(\tau_x T)} - Q_{X(\beta)}(\tau_x)) \end{aligned}$$

where $\dot{X}$ is between $X(\beta)_{(\tau_x T)}$ and $Q_{X(\beta)}(\tau_x)$. Condition A.2 guarantees that $f_{X(\beta)}$ is uniformly $O(1)$ and bounded below from 0, say $\underline{f}$. Then

$$T^{1/2}\mathbb{E}[\left|X(\beta)_{(\tau_x T)} - Q_{X(\beta)}(\tau_x)\right|] \le \underline{f}T^{1/2}\mathbb{E}[\left|(P_T - P)g(\beta, X(\beta)_{(\tau_x T)})\right|] + O(T^{-1/2}).$$

So it suffices to show that

$$\sup_{\beta\in\mathcal{B}, r\in\mathbb{R}} \mathbb{E}[T^{1/2}\left|(P_T - P)g(\beta,r)\right|] = O(1). \tag{20}$$

To this end, apply Corollary 14.5 in Davidson (1994) to obtain

$$\begin{aligned} &\mathbb{E}[\left|(P_T - P)g(\beta,r)\right|^2] \\ \le\ & n^{-1}\left(\mathbb{E}[g(\beta,r)^2] - (\mathbb{E}[g(\beta,r)])^2\right) + n^{-2}\sum_{i\neq j} Cov[\mathbf{1}[\mathbf{X}_i^{\intercal}\beta \le r], \mathbf{1}[\mathbf{X}_j^{\intercal}\beta \le r]] \\ \le\ & n^{-1} + n^{-1}4\sum_{t=1}^{\infty}\phi_t^{1/2}\mathbb{E}[g(\beta,r)^2] \le Cn^{-1}. \end{aligned}$$

The above bound is uniform in β and r and then (20) follows from Cauchy-Schwarz inequality. So (19) is established.

Next, I provide preliminary conditions for Condition 2.3. The following conditions are sufficient. Recall that y_0 denotes the right end-point $\sup\{y, F_{\varepsilon|X(\beta_0)}(y) < 1\}$. The notation is simpler if I write with the following notation: $\gamma(\cdot) = 1/\xi(\cdot)$, g_i denotes the the partial derivative of a generic function $g(\cdot,\cdot)$ w.r.t. the i-th element, and g_{ij} the i,j-th cross derivative.

Condition B $\mathbf{X}_{it}$ has a compact support. $F_{\varepsilon|X(\beta)=x}(y)$ satisfies (i) if $\xi(x,\beta) > 0$,

$$1 - F_{\varepsilon|X(\beta)=x}(y) = c(x,\beta)y^{-\gamma(x,\beta)}(1 + d(x,\beta)(y)^{-\tilde{\gamma}(x,\beta)} + r(x,y,\beta)) \text{ for some } \tilde{\gamma}(x,\beta) > 0$$

where $c(\cdot) > 0$, $d(\cdot)$, $\gamma(\cdot)$, and $\tilde{\gamma}(\cdot)$ are uniformly bounded and continuously differentiable with uniformly bounded derivatives w.r.t. β and x, and $r(x,y,\beta)$ satisfies

$$\limsup_{y\to y_0} \sup_{(x,\beta)\in(B_\delta(Q_{X(\beta_0)}(\tau_x))\times B_\delta(\beta_0))\cap\{(x,\beta):\xi(x,\beta)>0\}} \left|r_2(x,y,\beta)/(\tilde{\gamma}(x,\beta)y^{-\tilde{\gamma}(x,\beta)-1})\right| \to 0,$$

$$\limsup_{y\to y_0} \sup_{(x,\beta)\in(B_\delta(Q_{X(\beta_0)}(\tau_x))\times B_\delta(\beta_0))\cap\{(x,\beta):\xi(x,\beta)>0\}} \left| r_1(x,y,\beta)/y^{-\tilde{\gamma}(x,\beta)}) \right| \;\to\; 0,$$

$$\limsup_{y\to y_0} \sup_{(x,\beta)\in(B_\delta(Q_{X(\beta_0)}(\tau_x))\times B_\delta(\beta_0))\cap\{(x,\beta):\xi(x,\beta)>0\}} \left| r_{12}(x,y,\beta)/(\tilde{\gamma}(x,\beta)y^{-\tilde{\gamma}(x,\beta)-1}) \right| \;\to\; 0;$$

(ii) if $\xi(x,\beta) < 0$,

$1-F_{\varepsilon|X(\beta)=x}(y) = c(x,\beta)(y_0-y)^{-\gamma(x,\beta)}(1+d(x,\beta)(y_0-y)^{-\tilde{\gamma}(x,\beta)}+r(x,y,\beta))$ for some $\tilde{\gamma}(x,\beta) < 0$

where $c(\cdot) > 0$, $d(\cdot)$, and $\tilde{\gamma}(\cdot)$ are uniformly bounded and continuously differentiable with uniformly bounded derivatives, and $r(x,y,\beta)$ satisfies

$$\limsup_{y\to y_0} \sup_{(x,\beta)\in(B_\delta(Q_{X(\beta_0)}(\tau_x))\times B_\delta(\beta_0))\cap\{(x,\beta):\xi(x,\beta)<0\}} \left| r_2(x,y,\beta)/(\tilde{\gamma}(x,\beta)(y_0-y)^{-\tilde{\gamma}(x,\beta)-1}) \right| \;\to\; 0,$$

$$\limsup_{y\to y_0} \sup_{(x,\beta)\in(B_\delta(Q_{X(\beta_0)}(\tau_x))\times B_\delta(\beta_0))\cap\{(x,\beta):\xi(x,\beta)<0\}} \left| r_1(x,y,\beta)/(y_0-y)^{-\tilde{\gamma}(x,\beta)}) \right| \;\to\; 0,$$

$$\limsup_{y\to y_0} \sup_{(x,\beta)\in B_\delta(Q_{X(\beta_0)}(\tau_x))\cap\mathbb{R}^-\times B_\delta(\beta_0)} \left| r_{12}(x,y,\beta)/(\tilde{\gamma}(x,\beta)(y_0-y)^{-\tilde{\gamma}(x,\beta)-1}) \right| \;\to\; 0;$$

(iii) if $\xi(x,\beta) = 0$,

$$f_{\varepsilon|X(\beta)=x}(y) = c(x,\beta)y^{\tilde{c}(x,\beta)}\exp(-d(x,\beta)\tilde{d}(y))(1+r(x,\beta,y))$$

where $c(x,\beta)$ and $d(x,\beta)$ are some continuously differential functions that are uniformly bounded between 0 and ∞, $\tilde{c}(x,\beta)$ is continuously differentiable and uniformly bounded by -1 and ∞, and $\tilde{d}(y)$ is continuously differentiable and satisfies $C_1(\log y)^2 \le \tilde{d}(y) \le C_2 y^{C_3}$ for some constants $0 \le C_1, C_2, C_3 < \infty$. The remainder $r(x,\beta,y)$ is uniformly bounded and continuously differentiable w.r.t. all three arguments with bounded derivatives, and satisfies

$$\limsup_{y\to y_0} \sup_{(x,\beta)\in(B_\delta(Q_{X(\beta_0)}(\tau_x))\times B_\delta(\beta_0))\cap\{(x,\beta):\xi(x,\beta)=0\}} |\max\{r_1(x,y,\beta), r_2(x,\beta,y), r_{12}(x,\beta,y)\}| \to 0.$$

Condition B assumes that the error of approximating the true CDF consists of the leading term $1+(d(x,\beta)y)^{-\tilde{\gamma}(x,\beta)}$ or $c(x,\beta)y^{\tilde{c}(x,\beta)}\exp(-d(x,\beta)\tilde{d}(y))$ and the remainder $r(x,y,\beta)$. The cases (i) and (ii) cover regularly varying tails, and are imposed by Smith (1982) to study unconditional problems. See also Hall (1982) and Smith (1987). The case (iii) covers slowly varying tails, including Gaussian ($\tilde{c}(x,\beta)=0$ and $\tilde{d}(y)=y^2$), lognormal ($\tilde{c}(x,\beta)=-1$ and $\tilde{d}(y)=(\log y)^2$), and the exponential family ($\tilde{c}(x,\beta)=0$ and $\tilde{d}(y)=y$). See, for example, Chapter B in de Haan and Ferreira (2007). Compared with those literature, I require a stronger version that the derivatives of $r(x,y,\beta)$ are uniformly bounded. This is to guarantee that the tail of $f_{\varepsilon|X(\beta)}$ is also uniformly bounded. The compact support of $\mathbf{X}$ is imposed to simplify the proof (cf. Wang and Li (2013)). The following lemma establishes Condition 2.3 using Conditions 2.4 and B. Its proof is collected at the very end of the article.

Lemma 1 *If Condition 2.4 and Condition B hold, then Condition 2.3 holds, i.e., for* $u_n = a_n y + b_n$ *for any* y*, (a)*

$$\lim_{u_n \to y_0} \sup_{x \in B_{\delta_T}(Q_{X(\beta_0)}(\tau_x))} T^{-1/2} \left| \frac{\partial F_{\varepsilon|X(\beta_0)=x}(u_n) / \partial x}{1 - F_{\varepsilon|X(\beta_0)=x}(u_n)} \right| = 0,$$

(b)

$$\lim_{u_n \to y_0} \sup_{x \in B_{\delta_T}(Q_{X(\beta_0)}(\tau_x))} T^{-1/2} \left| \frac{\partial f_{\varepsilon|X(\beta_0)=x}(u_n) / \partial x}{f_{\varepsilon|X(\beta_0)=Q_x(\tau_x)}(u_n)} \right| = 0,$$

and (c)

$$\lim_{u_n \to y_0} \sup_{(x,\beta) \in B_{\delta_T}(Q_{X(\beta_0)}(\tau_x)) \times B_{v_{nT}}(\beta_0)} T^{-1/2} \left| \frac{\partial F_{\varepsilon|X(\beta)=x}(u_n) / \partial \beta}{1 - F_{\varepsilon|X(\beta_0)=x}(u_n)} \right| = 0,$$

for some $v_{nT} = O((nT)^{-1/2})$ *and* $\delta_T = O(T^{-1/2})$ *as* $n \to \infty$ *and* $T \to \infty$*.*

To give a better sense of Condition B, I now show that it is satisfied by the three one-dimensional covariate examples in Section 2.2.

First consider the joint normal distribution. Condition B.(iii) is satisfied by setting $c(x) = \sqrt{2\pi(1-\rho^2)}$, $d(x) = 1$, $\tilde{d}(y) = y^2/(2(1-\rho^2))$, and $r(x,y) = \exp(2\rho x/y + \rho^2 x^2/y^2) - 1$. Second, for the conditional student's t distribution, Ding (2016) derives that the conditional PDF of Y given $X = x$ is

$$f_{Y|X=x}(y) = \frac{C}{\sigma(x)} \left(1 + \frac{(y-\rho x)^2}{(v+1)\,\sigma(x)^2}\right)^{-\frac{v+2}{2}}$$

for some constant C depending on v only and $\sigma(x) = \sqrt{(1-\rho^2)(v+x^2)/(v+1)}$. Then Condition B holds with $\gamma(x) = 1/(v+1)$, $c(x) \propto \sigma(x)^{v+1}, d(x) \propto \rho x, \tilde{\gamma}(x) = 1$, and $r(x,y) = O(y^{-2})$ for any $x \in \mathbb{R}$. Finally, for the conditional Pareto distribution, Taylor expansion yields

$$\begin{aligned} 1 - F_{Y|X=x}(y) &= y^{-1/x}(1+1/y)^{-1/x} \\ &= y^{-1/x}(1 - \frac{1}{xy} + O(\frac{1}{y^2})). \end{aligned}$$

Thus Condition B holds with $c(x) = 1, \gamma(x) = 1/x, d(x) = -1/x, \tilde{\gamma}(x) = 1$, and $r(x,y) = O(y^{-2})$ for x bounded below from 0.

A.2 Proofs of Theorems

Proof of Theorem 1

I suppress $Q_X(\tau_x)$ in $a_n(\cdot)$ and $b_n(\cdot)$ and consider $k = 1$ first. By strict stationarity across t,

$$\begin{aligned} \mathbb{P}\left(Y_{i,[\tau_x T]} \le a_n y + b_n\right) &= \mathbb{E}_{X_{i,(\tau_x T)}}\left[\mathbb{P}\left(Y_{i,[\tau_x T]} \le a_n y + b_n | X_{i,(\tau_x T)}\right)\right] \\ &= \mathbb{E}_{X_{i,(\tau_x T)}}\left[F_{Y|X=X_{i,(\tau_x T)}}(a_n y + b_n)\right]. \end{aligned} \tag{21}$$

Thus

$$\begin{aligned}
& \mathbb{P}\left(Y_{(1),[\tau_x T]} \leq a_n y + b_n\right) \\
= \; & F^n_{Y_{i,[\tau_x T]}}(a_n y + b_n) \text{ (by i.i.d. across } i) \\
= \; & F^n_{Y|X=Q_X(\tau_x)}\left(a_n y + b_n\right)\left(\frac{\mathbb{P}\left(Y_{i,[\tau_x T]} \leq a_n y + b_n\right)}{F_{Y|X=Q_X(\tau_x)}\left(a_n y + b_n\right)}\right)^n \\
= \; & F^n_{Y|X=Q_X(\tau_x)}\left(a_n y + b_n\right)\left(\frac{\mathbb{E}_{X_{i,(\tau_x T)}}\left[F_{Y|X=X_{i,(\tau_x T)}}\left(a_n y + b_n\right)\right]}{F_{Y|X=Q_X(\tau_x)}\left(a_n y + b_n\right)}\right)^n \quad \text{(by (21))} \\
= \; & F^n_{Y|X=Q_X(\tau_x)}\left(a_n y + b_n\right)\left(1 + \frac{\mathbb{E}_{X_{i,(\tau_x T)}}\left[F_{Y|X=X_{i,(\tau_x T)}}\left(a_n y + b_n\right)\right] - F_{Y|X=Q_X(\tau_x)}\left(a_n y + b_n\right)}{F_{Y|X=Q_X(\tau_x)}\left(a_n y + b_n\right)}\right)^n \\
\equiv \; & A_n\left(y\right)\left(1 + \frac{B_{n,T}\left(y\right)}{F_{Y|X=Q_X(\tau_x)}\left(a_n y + b_n\right)}\right)^n .
\end{aligned}$$

By the standard EV theory and Condition 1.2, $A_n(y) \to G_\xi(y)$ as $n \to \infty$. Regarding $B_{n,T}(y)$, Theorem 1 in Wu (2005) yields that $\mathbb{E}\left[\left|X_{i,(\tau_x T)} - Q_X(\tau_x)\right|\right] = O\left(T^{-1/2}\right)$. Then together with Condition 1.3, I have, for some $\dot{x}$ between $X_{i,(\tau_x T)}$ and $Q_X(\tau_x)$

$$\begin{aligned}
\left|B_{n,T}\left(y\right)\right| &= \mathbb{E}\left[\frac{\partial}{\partial x} F_{Y|X=x}\left(a_n y + b_n\right)|_{x=\dot{x}}\left(X_{i,(\tau_x T)} - Q_X\left(\tau_x\right)\right)\right] \\
&\leq \sup_x \left|\frac{\partial}{\partial x} F_{Y|X=x}\left(a_n y + b_n\right)\right| \mathbb{E}\left[\left|X_{i,(\tau_x T)} - Q_X\left(\tau_x\right)\right|\right] \\
&= o(T^{-1/2})
\end{aligned}$$

and hence given $a_n y + b_n \to y_0$,

$$\begin{aligned}
\left(1 + \frac{B_{n,T}\left(y\right)}{F_{Y|X=Q_X(\tau_x)}\left(a_n y + b_n\right)}\right)^n &\leq \left(\left(1 + \frac{o(T^{-1/2})}{F_{Y|X=Q_X(\tau_x)}\left(a_n y + b_n\right)}\right)^{T^{1/2}}\right)^{2C/\lambda} \\
&\leq \exp\left(2C/\lambda \times o(1)\right) \to 1.
\end{aligned}$$

The proof for $k = 1$ is then complete by the continuous mapping theorem.

Generalization to $k > 1$ is as follows. Consider $y_1 > y_2 > \cdots > y_k$. Chapter 8.4 in Arnold, Balakrishnan, and Nagaraja (1992) gives that

$$\begin{aligned}
& \mathbb{P}\left(Y_{(1),[\tau_x T]} \leq a_n y_1 + b_n, ..., Y_{(k),[\tau_x T]} \leq a_n y_k + b_n\right) \\
= \; & F^{n-k}_{Y_{i,[\tau_x T]}}(a_n y_k + b_n) \prod_{r=1}^{k}\left(n - r + 1\right) a_n f_{Y_{i,[\tau_x T]}}\left(a_n y_r + b_n\right) \text{ (by i.i.d. across } i) \\
= \; & \left[F^{n-k}_{Y|X=Q_X(\tau_x)}\left(a_n y_k + b_n\right) \prod_{r=1}^{k}\left(n - r + 1\right) a_n f_{Y|X=Q_x(\tau_x)}\left(a_n y_r + b_n\right)\right] \times
\end{aligned}$$

$$
\begin{aligned}
& \left[\left(\frac{\mathbb{P}\left(Y_{i,[\tau_x T]} \leq a_n y_k + b_n\right)}{F_{Y|X=Q_X(\tau_x)}\left(a_n y_k + b_n\right)}\right)^{n-k} \prod_{r=1}^{k} \frac{f_{Y_{i,[\tau_x T]}}\left(a_n y_r + b_n\right)}{f_{Y|X=Q_x(\tau_x)}\left(a_n y_r + b_n\right)}\right] \\
\equiv \quad & \tilde{A}_n \times \tilde{B}_{nT}.
\end{aligned}
$$

The convergence that $\tilde{A}_n \rightarrow G_{\xi}\left(y_k\right) \prod_{r=1}^{k}\{g_{\xi}\left(y_r\right)/G_{\xi}\left(y_k\right)\}$ is established by Theorem 8.4.2 in Arnold, Balakrishnan, and Nagaraja (1992). It now remains to show $\tilde{B}_{nT} \rightarrow 1$. First, $(\mathbb{P}\left(Y_{i,[\tau_x T]} \leq a_n y_k + b_n\right)/F_{Y|X=Q_X(\tau_x)}\left(a_n y_k + b_n\right))^{n-k} \rightarrow 1$ is shown by the same argument as above in the $k=1$ case. Second, for any $r \in \{1, ..., k\}$ and any v

$$
\begin{aligned}
\frac{f_{Y_{i,[\tau_x T]}}(v)}{f_{Y|X=Q_x(\tau_x)}(v)} &= \frac{\frac{\partial \mathbb{P}\left(Y_{i,[\tau_x T]} \leq v\right)}{\partial v}}{f_{Y|X=Q_x(\tau_x)}(v)} \\
&= \frac{\frac{\partial}{\partial v}\mathbb{E}_{X_{i,(\tau_x T)}}\left[F_{Y|X=X_{i,(\tau_x T)}}(v)\right]}{f_{Y|X=Q_X(\tau_x)}(v)} \\
&= \frac{\frac{\partial}{\partial v}\int F_{Y|X=x}(v)\, f_{X_{i,(\tau_x T)}}(x)\, dx}{f_{Y|X=Q_X(\tau_x)}(v)} \\
&= \frac{\int \frac{\partial}{\partial v} F_{Y|X=x}(v)\, f_{X_{i,(\tau_x T)}}(x)\, dx}{f_{Y|X=Q_X(\tau_x)}(v)} \quad \text{(by Leibniz's rule)} \\
&= \frac{\mathbb{E}_{X_{i,(\tau_x T)}}\left[f_{Y|X=X_{i,(\tau_x T)}}(v)\right]}{f_{Y|X=Q_X(\tau_x)}(v)}
\end{aligned}
$$

where applying Leibniz's rule is permitted by the assumption that $f_{Y|X=x}(v)$ is uniformly continuous in x and v. Then the intermediate value theorem yields that for some $\dot{X}$ between $X_{i,(\tau_x T)}$ and $Q_X(\tau_x)$,

$$
\begin{aligned}
& \left|\frac{f_{Y_{i,[\tau_x T]}}\left(a_n y_r + b_n\right)}{f_{Y|X=Q_X(\tau_x)}\left(a_n y_r + b_n\right)} - 1\right| \\
= \quad & \left|\mathbb{E}_{X_{i,(\tau_x T)}}\left[\frac{\partial f_{Y|X=x}\left(a_n y_r + b_n\right)}{\partial x}\Big|_{x=\dot{X}} \frac{\left(X_{i,(\tau_x T)} - Q_X(\tau_x)\right)}{f_{Y|X=Q_X(\tau_x)}\left(a_n y_r + b_n\right)}\right]\right| \\
\leq \quad & \mathbb{E}_{X_{i,(\tau_x T)}}\left[\frac{\partial f_{Y|X=x}\left(a_n y_r + b_n\right)}{\partial x}\Big|_{x=\dot{X}} \frac{\left(X_{i,(\tau_x T)} - Q_X(\tau_x)\right)}{f_{Y|X=Q_X(\tau_x)}\left(a_n y_r + b_n\right)}\right] \\
\leq \quad & \sup_{x \in B_{\delta_T}(Q_X(\tau_x))}\left|\frac{\partial f_{Y|X=x}\left(a_n y_r + b_n\right)/\partial x}{f_{Y|X=Q_X(\tau_x)}\left(a_n y_r + b_n\right)}\right| \mathbb{E}_{X_{i,(\tau_x T)}}\left[\left|X_{i,(\tau_x T)} - Q_X(\tau_x)\right|\right] \text{ for some large constant } C \\
\rightarrow \quad & 0
\end{aligned}
$$

where the last convergence follows from Conditions 1.1 and 1.3 again. ■

Proof of Theorem 2

The proof consists of 5 steps.

Step 1. For any fixed β, define $\varepsilon_{i,J([\tau_x T],\beta)}$ as the induced ε_{it} associated with the $\tau_x T$-th largest observation in $\{X(\beta)_{i1}, X(\beta)_{i2}, ..., X(\beta)_{iT}\}$, that is, $\varepsilon_{i,J([\tau_x T],\beta)} = \varepsilon_{it}$ if $X(\beta)_{i,(\tau_x T)} = X(\beta)_{it}$. I claim that as $n \to \infty$ and $T \to \infty$,

$$\frac{\{\varepsilon_{(1),J([\tau_x T],\beta_0)}, \varepsilon_{(2),J([\tau_x T],\beta_0)}, ..., \varepsilon_{(k),J([\tau_x T],\beta_0)}\} - b_n}{a_n} \xrightarrow{d} V_1, ..., V_k.$$

This is done by treating $X(\beta_0)$ as a scaler variable so that $(\varepsilon_{it}, X(\beta_0)_{it})$ satisfies Condition 1.

Step 2. I claim that for each i,

$$\sup_{\beta \in B_{\delta_T}(\beta_0)} \mathbb{E}\left[\left|X(\beta)_{i,(\tau_x T)} - X(\beta_0)_{i,(\tau_x T)}\right|\right] = O(T^{-1/2})$$

where $B_{\delta_T}(\beta_0)$ denotes an open ball centered at β_0 with radius $\delta_T = o(T^{-1/2})$.

To prove this, the triangular inequality implies

$$\begin{aligned} & \sup_{\beta \in B_{\delta_T}(\beta_0)} \mathbb{E}\left[\left|X(\beta)_{i,(\tau_x T)} - X(\beta_0)_{i,(\tau_x T)}\right|\right] \\ \leq & \sup_{\beta \in B_{\delta_T}(\beta_0)} \mathbb{E}\left[\left|X(\beta)_{i,(\tau_x T)} - Q_{X(\beta)}(\tau_x)\right|\right] + \sup_{\beta \in B_{\delta_T}(\beta_0)} \left|Q_{X(\beta)}(\tau_x) - Q_{X(\beta_0)}(\tau_x)\right| \\ & + \mathbb{E}\left[\left|X(\beta_0)_{i,(\tau_x T)} - Q_{X(\beta_0)}(\tau_x)\right|\right]. \end{aligned}$$

The first one is $O(T^{-1/2})$ given Condition 2.1. The second item is $o(T^{-1/2})$ since Condition 2.3 implies $Q_{X(\beta)}(\tau_x)$ is continuous in β. The last item is $O(T^{-1/2})$ by Theorem 1 of Wu (2005).

Step 3. I claim that as $n \to \infty$ and $T \to \infty$,

$$\frac{\{\varepsilon_{(1),J([\tau_x T],\hat{\beta})}, \varepsilon_{(2),J([\tau_x T],\hat{\beta})}, ..., \varepsilon_{(k),J([\tau_x T],\hat{\beta})}\} - b_n}{a_n} \xrightarrow{d} V_1, ..., V_k.$$

I present the proof for $\varepsilon_{(1),J([\tau_x T],\hat{\beta})}$ only to save space. Generalization to k is similar and omitted.

Since $\hat{\beta} = \beta_0 + O_p\left((nT)^{-1/2}\right)$, it suffices to consider $\hat{\beta} \in B_{v(nT)^{-1/2}}(\beta_0)$ for some open ball centered at β_0 with radius $v(nT)^{-1/2}$ for $v \in \mathbb{R}^+$. Denote $g_n(\beta) = \mathbb{P}\left(\varepsilon_{(1),J([\tau_x T],\beta)} \leq a_n y + b_n\right)$ and $g(\beta_0) = G_{\xi\left(Q_{X(\beta_0)}(\tau_x)\right)}(y)$. Then the triangular inequality yields that $\left|g_n\left(\hat{\beta}\right) - g(\beta_0)\right| \leq |g_n(\beta_0) - g(\beta_0)| + \sup_{\beta \in B_{v(nT)^{-1/2}}(\beta_0)} |g_n(\beta) - g_n(\beta_0)|$. The first item is $o(1)$ as established in step 1. Regarding the second item, I suppress $(a_n y + b_n)$ to save space and write

$$\sup_{\beta \in B_{v(nT)^{-1/2}}(\beta_0)} |g_n(\beta) - g_n(\beta_0)|$$

$$\begin{aligned}
&\leq \left(\mathbb{E}_{X(\beta_0)_{i,(\tau_x T)}}\left[F_{\varepsilon|X(\beta_0)=X(\beta_0)_{i,(\tau_x T)}}\right]\right)^n \times \\
&\quad \sup_{\beta\in B_{v(nT)^{-1/2}}(\beta_0)}\left|\left(\frac{\mathbb{E}_{X(\beta)_{i,(\tau_x T)}}\left[F_{\varepsilon|X(\beta)=X(\beta)_{i,(\tau_x T)}}\right]}{\mathbb{E}_{X(\beta_0)_{i,(\tau_x T)}}\left[F_{\varepsilon|X(\beta_0)=X(\beta_0)_{i,(\tau_x T)}}\right]}\right)^n-1\right| \\
&= O(1)\sup_{\beta\in B_{v(nT)^{-1/2}}(\beta_0)}\left|\left(\frac{\mathbb{E}_{X(\beta)_{i,(\tau_x T)}}\left[F_{\varepsilon|X(\beta_0)=X(\beta)_{i,(\tau_x T)}}+\frac{\partial}{\partial\beta^\intercal}F_{\varepsilon|X(\dot\beta)=X(\beta)_{i,(\tau_x T)}}(\beta-\beta_0)\right]}{\mathbb{E}_{X(\beta_0)_{i,(\tau_x T)}}\left[F_{\varepsilon|X(\beta_0)=X(\beta_0)_{i,(\tau_x T)}}\right]}\right)^n-1\right| \\
&= o(1)
\end{aligned}$$

where $\dot\beta = r\beta+(1-r)\beta_0$ for some $r\in[0,1]$ and the last line holds if I establish

$$\sup_{\beta\in B_{v(nT)^{-1/2}}(\beta_0)}\frac{\mathbb{E}_{X(\beta)_{i,(\tau_x T)}}\left[F_{\varepsilon|X(\beta_0)=X(\beta)_{i,(\tau_x T)}}\right]}{\mathbb{E}_{X(\beta_0)_{i,(\tau_x T)}}\left[F_{\varepsilon|X(\beta_0)=X(\beta_0)_{i,(\tau_x T)}}\right]} = 1+o(n^{-1}) \tag{22}$$

and

$$\sup_{\beta\in B_{v(nT)^{-1/2}}(\beta_0)}\frac{\left|\mathbb{E}\left[\frac{\partial}{\partial\beta^\intercal}F_{\varepsilon|X(\dot\beta)=X(\beta)_{i,(\tau_x T)}}\right](\beta-\beta_0)\right|}{\mathbb{E}_{X(\beta_0)_{i,(\tau_x T)}}\left[F_{\varepsilon|X(\beta_0)=X(\beta_0)_{i,(\tau_x T)}}\right]} = o(n^{-1}). \tag{23}$$

I now prove (22). By Taylor expansion, I have for any β

$$\begin{aligned}
&\frac{\mathbb{E}_{X(\beta)_{i,(\tau_x T)}}\left[F_{\varepsilon|X(\beta_0)=X(\beta)_{i,(\tau_x T)}}\right]}{\mathbb{E}_{X(\beta_0)_{i,(\tau_x T)}}\left[F_{\varepsilon|X(\beta_0)=X(\beta_0)_{i,(\tau_x T)}}\right]} \\
&= 1+\frac{\mathbb{E}[F_{\varepsilon|X(\beta_0)=X(\beta)_{i,(\tau_x T)}}-F_{\varepsilon|X(\beta_0)=X(\beta_0)_{i,(\tau_x T)}}]}{\mathbb{E}_{X(\beta_0)_{i,(\tau_x T)}}\left[F_{\varepsilon|X(\beta_0)=X(\beta_0)_{i,(\tau_x T)}}\right]} \\
&= 1+\frac{\mathbb{E}[\frac{\partial}{\partial x}F_{\varepsilon|X(\beta_0)=x}|_{x=\dot X}(X(\beta)_{i,(\tau_x T)}-X(\beta_0)_{i,(\tau_x T)})]}{\mathbb{E}_{X(\beta_0)_{i,(\tau_x T)}}\left[F_{\varepsilon|X(\beta_0)=X(\beta_0)_{i,(\tau_x T)}}\right]}
\end{aligned} \tag{24}$$

where $\dot X$ is between $X(\beta)_{i,(\tau_x T)} - X(\beta_0)_{i,(\tau_x T)}$. First, Step 2 yields that $\sup_{\beta\in B_{v(nT)^{-1/2}}(\beta_0)}\mathbb{E}[\,|X(\beta)_{i,(\tau_x T)}-X(\beta_0)_{i,(\tau_x T)}|] = O(T^{-1/2})$. Second, Condition 2.3 and the fact that $1-F_{\varepsilon|X(\beta_0)=Q_{X(\beta_0)}(\tau_x)}(a_n y+b_n) = O(n^{-1})$ imply $\sup_{x\in B_{T^{-1/2}}(Q_{X(\beta_0)}(\tau_x))}|\frac{\partial}{\partial x}F_{\varepsilon|X(\beta_0)=x}(a_n y+b_n)| = o(n^{-1/2})$. Hence the second item in (24) is $o(n^{-1})$ given Condition 2.4, and (22) follows.

To prove (23), I have

$$\begin{aligned}
&\sup_{\beta\in B_{v(nT)^{-1/2}}(\beta_0)}\frac{\mathbb{E}\left[\frac{\partial}{\partial\beta^\intercal}F_{\varepsilon|X(\dot\beta)=X(\beta)_{i,(\tau_x T)}}(\beta-\beta_0)\right]}{\mathbb{E}_{X(\beta_0)_{i,(\tau_x T)}}\left[F_{\varepsilon|X(\beta_0)=X(\beta_0)_{i,(\tau_x T)}}\right]} \\
&\leq \sup_{\beta\in B_{v(nT)^{-1/2}}(\beta_0)}\frac{\sup_{x\in B_{T^{-1/2}}(Q_{X(\beta_0)}(\tau_x))}\max_{j\in\{1,\ldots,\dim(\beta)\}}\mathbb{E}\left[\left|\frac{\partial}{\partial\beta_j}F_{\varepsilon|X(\beta)=x}\right|\right]}{\mathbb{E}_{X(\beta_0)_{i,(\tau_x T)}}\left[F_{\varepsilon|X(\beta_0)=X(\beta_0)_{i,(\tau_x T)}}\right]}\times O((nT)^{-1/2})
\end{aligned}$$

$$\leq \quad O(n^{-1} \times (nT)^{-1/2}) = o(n^{-1})$$

where the last inequality again follows from Condition 2.3 and the fact that $1 - F_{\varepsilon|X(\beta_0)=Q_{X(\beta_0)}(\tau_x)}(a_n y + b_n) = O(n^{-1})$.

Step 4. Let $I = (I_1, \ldots, I_k) \in \{1, \ldots, n\}^k$ be the k random (cross-sectional) indices based on observing $\{\varepsilon_{i,J([\tau_x T],\hat{\beta})}\}_{i=1}^n$ such that $\varepsilon_{(j),J([\tau_x T],\hat{\beta})} = \varepsilon_{I_j,J([\tau_x T],\hat{\beta})}$, $j = 1, \ldots, k$, and let $\hat{I}$ be the corresponding indices based on $\{\hat{\varepsilon}_{i,J([\tau_x T],\hat{\beta})}\}_{i=1}^n$ such that $\hat{\varepsilon}_{(j),J([\tau_x T],\hat{\beta})} = \hat{\varepsilon}_{I_j,J([\tau_x T],\hat{\beta})}$, $j = 1, \ldots, k$. I claim that $I - \hat{I} \xrightarrow{p} 0$ and prove this by contradiction.

Suppose the claim is false, then step 3 implies that $\sup_i \left|\hat{\varepsilon}_{i,J([\tau_x T],\hat{\beta})} - \varepsilon_{i,J([\tau_x T],\hat{\beta})}\right|$ is not $o_p(a_n)$. But this contradicts that

$$\begin{aligned}
& a_n^{-1} \sup_i \left|\hat{\varepsilon}_{i,J([\tau_x T],\hat{\beta})} - \varepsilon_{i,J([\tau_x T],\hat{\beta})}\right| \qquad (25)\\
= \;& a_n^{-1} \sup_i \left|\mathbf{X}^{\intercal}_{i,J([\tau_x T],\hat{\beta})}(\hat{\beta} - \beta_0)\right| \\
\leq \;& a_n^{-1} \sup_i \left|\mathbf{X}^{\intercal}_{i,J([\tau_x T],\hat{\beta})}\right| \times \left|\hat{\beta} - \beta_0\right| \\
\leq \;& a_n^{-1} \times o_p\left(n^{1/2}\right) \times O_p\left((nT)^{-1/2}\right) \\
= \;& o_p\left(a_n^{-1} T^{-1/2}\right) \\
= \;& o_p(1)
\end{aligned}$$

where the last equation follows from Condition 2.4 and $\xi \geq -1/2$.

Step 5. Combine Steps 1-4 to obtain

$$\begin{aligned}
& \frac{\{\hat{\varepsilon}_{(1),[\tau_x T]}, \hat{\varepsilon}_{(2),[\tau_x T]}, \ldots, \hat{\varepsilon}_{(k),[\tau_x T]}\} - b_n}{a_n} \\
= \;& \frac{\{\hat{\varepsilon}_{\hat{I}_1,J([\tau_x T],\hat{\beta})}, \hat{\varepsilon}_{\hat{I}_2,J([\tau_x T],\hat{\beta})}, \ldots, \hat{\varepsilon}_{\hat{I}_k,J([\tau_x T],\hat{\beta})}\} - b_n}{a_n} \quad \text{(by definition of } J([\tau_x T], \hat{\beta})) \\
= \;& \frac{\{\hat{\varepsilon}_{I_1,J([\tau_x T],\hat{\beta})}, \hat{\varepsilon}_{I_2,J([\tau_x T],\hat{\beta})}, \ldots, \hat{\varepsilon}_{I_k,J([\tau_x T],\hat{\beta})}\} - b_n}{a_n} + o_p(1) \quad \text{(by step 4)} \\
= \;& \frac{\{\varepsilon_{I_1,J([\tau_x T],\hat{\beta})}, \varepsilon_{I_2,J([\tau_x T],\hat{\beta})}, \ldots, \varepsilon_{I_k,J([\tau_x T],\hat{\beta})}\} - b_n}{a_n} + o_p(1) \quad \text{(by (25))} \\
& \xrightarrow{d} \mathbf{V} \qquad \text{(by step 3).}
\end{aligned}$$

■

Proof of Theorem 3

The argument is very similar to the proof of Theorem 2, and hence I only highlight the difference. In particular, Steps 1 and 2 are identical. Step 3 is now

$$\frac{\{u_{(1),J\left([\tau_x T],\hat{\beta}\right)},u_{(2),J\left([\tau_x T],\hat{\beta}\right)},...,u_{(k),J\left([\tau_x T],\hat{\beta}\right)}\}-b_n}{a_n}\xrightarrow{d}\{V_1,...,V_k\}.$$

For Step 4, let $I=(I_1,\ldots,I_k)\in\{1,\ldots,n\}^k$ be the k random (cross-sectional) indices based on observing $\{u_{i,J([\tau_x T],\hat{\beta})}\}_{i=1}^n$ such that $u_{(j),J([\tau_x T],\hat{\beta})}=u_{I_j,J([\tau_x T],\hat{\beta})}$, $j=1,\ldots,k$, and let $\hat{I}$ be the corresponding indices based on $\{\tilde{\varepsilon}_{i,J([\tau_x T],\hat{\beta})}\}_{i=1}^n$ such that $\tilde{\varepsilon}_{(j),J([\tau_x T],\hat{\beta})}=\tilde{\varepsilon}_{I_j,J([\tau_x T],\hat{\beta})}$, $j=1,\ldots,k$. Recall that

$$\begin{aligned}\tilde{\varepsilon}_{it} &= \hat{\varepsilon}_{it}-\bar{\varepsilon}_i\\ &= u_{it}-\bar{u}_i-(\mathbf{X}_{it}-\bar{\mathbf{X}}_i)^\intercal(\hat{\beta}-\beta_0).\end{aligned}$$

I claim that $I-\hat{I}\xrightarrow{p}0$ and prove this by contradiction.

Suppose the claim is false, then step 3 implies that $\sup_i\left|\tilde{\varepsilon}_{i,J([\tau_x T],\hat{\beta})}-u_{i,J([\tau_x T],\hat{\beta})}\right|$ is not $o_p(a_n)$. But this contradicts that

$$\begin{aligned}
& a_n^{-1}\sup_i\left|\tilde{\varepsilon}_{i,J([\tau_x T],\hat{\beta})}-u_{i,J([\tau_x T],\hat{\beta})}\right| \qquad (26)\\
= {} & a_n^{-1}\sup_i\left|\bar{u}_i+\left(\mathbf{X}_{i,J([\tau_x T],\hat{\beta})}-\bar{\mathbf{X}}_i\right)^\intercal(\hat{\beta}-\beta_0)\right|\\
\le {} & a_n^{-1}\sup_i|\bar{u}_i|+a_n^{-1}\left(\sup_i\left|\left|\mathbf{X}_{i,J([\tau_x T],\hat{\beta})}\right|\right|+\sup_i\left|\left|\bar{\mathbf{X}}_i\right|\right|\right)\times\left|\left|\hat{\beta}-\beta_0\right|\right|\\
\le {} & a_n^{-1}\times O_p\left(T^{-1/2}\right)+a_n^{-1}\times(o_p(n^{1/2})+O_p(1))O_p\left((nT)^{-1/2}\right)\\
= {} & o_p(1)
\end{aligned}$$

where the last equation follows from Condition 2.4 and $\xi>-1/2$.

Step 5 now follows similarly with (26).

Proof of Theorem 4

I first derive the asymptotic distribution of the Hill's estimator. I suppress $Q_X(\tau_x)$ in c, d, ξ, and $\tilde{\xi}$ for notational ease. The notation is easier if I write $\gamma=1/\xi$ and $\tilde{\gamma}=1/\tilde{\xi}$. By Theorem 2 of Hall (1982), it suffices to show that

$$1-\mathbb{P}\left(Y_{i,[\tau_x T]}\le u\right)=cu^{-\gamma}(1+du^{-\tilde{\gamma}}+o(u^{-\tilde{\gamma}}))\text{ as }u\to y_0.$$

Condition 3 states that $F_{Y|X=Q_X(\tau_x)}$ satisfies the above expansion. So it suffices to show that as $u \to y_0$

$$\left|\mathbb{P}\left(Y_{i,[\tau_x T]} \le u\right) - F_{Y|X=Q_X(\tau_x)}(u)\right| = o(u^{-\gamma}(1 + du^{-\tilde{\gamma}}).$$

To this end, I first use Condition 3 to obtain

$$\begin{aligned}
& \sup_{x \in B_{\delta_T}(Q_X(\tau_x))} \left|\frac{\partial F_{Y|X=x}(u)/\partial y}{1 - F_{Y|X=x}(u)}\right| \\
\le & \sup_{x \in B_{\delta_T}(Q_X(\tau_x))} \left|\frac{\partial c(x)/\partial x}{c(x)} + \log(u)\frac{\partial \gamma(x)}{\partial x} + \frac{\partial d(x)}{\partial x} u^{-\tilde{\gamma}(x)}\right. \\
& \left. -d(x)u^{-\tilde{\gamma}(x)}\tilde{\gamma}(x)\log(u) + \frac{\partial r(x,y)}{\partial x}\right| \\
= & \ O(\log u) \qquad (27)
\end{aligned}$$

and

$$\sup_{x \in B_{\delta_T}(Q_X(\tau_x))} \left|1 - F_{Y|X=x}(u)\right| = O(u^{-\gamma}(1 + du^{-\tilde{\gamma}})). \qquad (28)$$

Second, the fact that $u = An^{(\gamma+2\tilde{\gamma})}$ implies that

$$\begin{aligned}
\log(u)n^{-1/2} &= O(n^{-1/2}\log n) \\
&= o(n^{-\tilde{\gamma}/(\gamma+2\tilde{\gamma})}) \\
&= o(u^{-\tilde{\gamma}}). \qquad (29)
\end{aligned}$$

Then using Conditions 1.1 and (27)-(29) gets that

$$\begin{aligned}
& \left|\mathbb{P}\left(Y_{i,[\tau_x T]} \le u\right) - F_{Y|X=Q_X(\tau_x)}(u)\right| \\
= & \left|\mathbb{E}_{X_{i,(\tau_x T)}}\left[F_{Y|X=X_{i,(\tau_x T)}}(u) - F_{Y|X=Q_X(\tau_x)}(u)\right]\right| \\
\le & \sup_{x \in B_{\delta_T}(Q_X(\tau_x))} \left|\frac{\partial F_{Y|X=x}(u)/\partial y}{1 - F_{Y|X=x}(u)}\right| \left[\left|X_{i,(\tau_x T)} - Q_X(\tau_x)\right|\right] \sup_{x \in B_{\delta_T}(Q_X(\tau_x))} \left|1 - F_{Y|X=x}(u)\right| \mathbb{E} \\
= & \ O(\log u) \times O(T^{-1/2}) \times O(u^{-\gamma}(1 + du^{-\tilde{\gamma}})). \\
= & \ o(u^{-\gamma}(1 + du^{-\tilde{\gamma}})).
\end{aligned}$$

Now I derive the asymptotic distribution of the Smith's estimator. By the exact argument in Smith (1987), it suffices to show $\mathbb{P}\left(Y_{i,[\tau_x T]} \le y\right)$ satisfies his SR2 condition. Since Condition 3 implies that $F_{Y|X=Q_X(\tau_x)}$ satisfies this condition with $\phi(u) = u^{-\tilde{\gamma}}$, it then remains to bound

$$\left|\frac{1 - \mathbb{P}\left(Y_{i,[\tau_x T]} \le yu\right)}{1 - \mathbb{P}\left(Y_{i,[\tau_x T]} \le u\right)} - \frac{1 - F_{Y|X=Q_X(\tau_x)}(yu)}{1 - F_{Y|X=Q_X(\tau_x)}(u)}\right|$$

$$\leq \left|\frac{\mathbb{P}\left(Y_{i,[\tau_x T]} \leq yu\right) - F_{Y|X=Q_X(\tau_x)}(yu)}{1-\mathbb{P}\left(Y_{i,[\tau_x T]} \leq u\right)}\right| + \left|\frac{\mathbb{P}\left(Y_{i,[\tau_x T]} \leq u\right) - F_{Y|X=Q_X(\tau_x)}(u)}{1-\mathbb{P}\left(Y_{i,[\tau_x T]} \leq u\right)}\right| \times \left|\frac{1-F_{Y|X=Q_X(\tau_x)}(yu)}{1-F_{Y|X=Q_X(\tau_x)}(u)}\right|.$$

I show the first item is uniform $o(\phi(u))$. The second one follows identically since Condition 3 implies $(1-F_{Y|X=Q_X(\tau_x)}(yu))/(1-F_{Y|X=Q_X(\tau_x)}(u)) = 1+O(\phi(u))$. Then apply Taylor expansion, Condition 1.1 and (27)-(29) to obtain that

$$\begin{aligned}
&\left|\frac{\mathbb{P}\left(Y_{i,[\tau_x T]} \leq yu\right) - F_{Y|X=Q_X(\tau_x)}(yu)}{1-\mathbb{P}\left(Y_{i,[\tau_x T]} \leq u\right)}\right| \\
= &\left|\frac{\mathbb{E}_{X_{i,(\tau_x T)}}\left[F_{Y|X=X_{i,(\tau_x T)}}(yu) - F_{Y|X=Q_X(\tau_x)}(yu)\right]}{1-\mathbb{E}_{X_{i,(\tau_x T)}}\left[F_{Y|X=X_{i,(\tau_x T)}}(u)\right]}\right| \\
\leq &\sup_{y>0, x\in B_{\delta_T}(Q_X(\tau_x))}\left|\frac{\frac{\partial}{\partial x}F_{Y|X=x}(yu)}{1-F_{Y|X=Q_X(\tau_x)}(u)+O(T^{-1/2})}\right| \mathbb{E}\left[\left|X_{i,(\tau_x T)} - Q_X(\tau_x)\right|\right] \\
= &\ O(\log(u)\times T^{-1/2}) \\
= &\ o(\phi(u)).
\end{aligned}$$

■

Proof of Lemma 1

The proof is different for $\xi(Q_{X(\beta_0)}(\tau_x), \beta_0) >, =,$ or < 0. I first consider the positive case. For (a), $\inf_{x\in B_{\delta_T}(Q_{X(\beta_0)}(\tau_x))} \xi(x,\beta_0) > 0$ if T is large enough. This is feasible given the continuity of $\xi(\cdot)$. After applying triangular inequality and the smoothness and boundedness of $c(\cdot)$, $d(\cdot)$, and $\gamma_1(\cdot)$, I have

$$\begin{aligned}
&\sup_{x\in B_{\delta_T}(Q_{X(\beta_0)}(\tau_x))}\left|\frac{\partial F_{\varepsilon|X(\beta_0)=x}(u_n)/\partial x}{1-F_{\varepsilon|X(\beta_0)=x}(u_n)}\right| \\
\leq &\sup_{x\in B_{\delta_T}(Q_{X(\beta_0)}(\tau_x))}\left|\frac{c_1(x,\beta_0)}{c(x,\beta_0)} + \log(u_n)\gamma_1(x,\beta_0) + d_1(x,\beta_0)(u_n)^{-\tilde{\gamma}(x,\beta_0)}\right. \\
&\left. -d(x,\beta_0)u_n^{-\tilde{\gamma}(x,\beta_0)}\tilde{\gamma}(x,\beta_0)\log(u_n) + r_1(x,u_n,\beta_0)\right| \\
= &\ O(\log(u_n)) \\
= &\ O(\log(n)) \text{ (by } a_n = Q_{\varepsilon|X(\beta_0)}(1-n^{-1})) \\
= &\ o(T^{1/2}) \text{ (by Condition 2.4).}
\end{aligned}$$

For (b), Condition A.3 implies that

$$f_{\varepsilon|X(\beta_0)=Q_x(\tau_x)}(y) = -c(x,\beta_0)\gamma(x,\beta_0)(y)^{-\gamma(x,\beta_0)-1}(1+d(x,\beta_0)(y)^{-\tilde{\gamma}(x,\beta_0)} + r(x,y,\beta_0))$$

$$+c(x,\beta_0)(y)^{-\gamma(x,\beta_0)}(-d(x,\beta_0)y^{-\tilde{\gamma}(x,\beta_0)-1}\tilde{\gamma}(x,\beta_0)+r_2(x,y,\beta_0)).$$

A similar argument as above yields

$$\sup_{x\in B_{\delta_T}(Q_{X(\beta_0)}(\tau_x))}\left|\frac{\partial f_{\varepsilon|X(\beta_0)=x}(u_n)/\partial x}{f_{\varepsilon|X(\beta_0)=Q_{X(\beta_0)}(\tau_x)}(u_n)}\right|\leq O(\log(u_n))=o(T^{1/2}).$$

For (c), for any $j\in\{1,...,\dim(\beta)\}$, define

$$\begin{aligned}\left|\frac{\partial F_{\varepsilon|X(\beta)=x}(u_n)/\partial\beta_j}{1-F_{\varepsilon|X(\beta_0)=x}(u_n)}\right| &= \left|\frac{\partial F_{\varepsilon|X(\beta)=x}(u_n)/\partial\beta_j}{1-F_{\varepsilon|X(\beta)=x}(u_n)}\right|\times\left|\frac{1-F_{\varepsilon|X(\beta)=x}(u_n)}{1-F_{\varepsilon|X(\beta_0)=x}(u_n)}\right| \\ &\equiv A_{nT}\times B_{nT}.\end{aligned}\tag{30}$$

Regarding A_{nT}, a similar argument as in (a) and (b) yields

$$\begin{aligned}&\sup_{(x,\beta)\in B_{\delta_T}(Q_{X(\beta_0)}(\tau_x))\times B_{v_{nT}}(\beta_0)}|A_{nT}| \\ \leq&\sup_{(x,\beta)\in B_{\delta_T}(Q_{X(\beta_0)}(\tau_x))\times B_{v_{nT}}(\beta_0)}\left|\frac{c_1(x,\beta)}{c(x,\beta)}+\log(u_n)\gamma_1(x,\beta)+d_1(x,\beta)(u_n)^{-\tilde{\gamma}(x,\beta)}\right. \\ &\left.-d(x,\beta)u_n^{-\tilde{\gamma}(x,\beta)}\tilde{\gamma}(x,\beta)\log(u_n)+r_1(x,u_n,\beta)\right| \\ =&\ O(\log(u_n))=o(T^{1/2}).\end{aligned}$$

Regarding B_{nT}, the continuous differentiability of γ and the fact that $u_n\sim Q_{X(\beta_0)=x}(1-1/n)=O(n^{\xi(Q_{X(\beta_0)}(\tau_x))})$ yield

$$\begin{aligned}&\sup_{(x,\beta)\in B_{\delta_T}(Q_{X(\beta_0)}(\tau_x))\times B_{v_{nT}}(\beta_0)}\left|\frac{(u_n)^{-\gamma(x,\beta)}}{(u_n)^{-\gamma(x,\beta_0)}}\right| \\ \leq&\sup_{(x,\beta)\in B_{\delta_T}(Q_{X(\beta_0)}(\tau_x))\times B_{v_{nT}}(\beta_0)}\left|n^{-\xi(Q_{X(\beta_0)}(\tau_x))(\gamma_2(x,\dot{\beta})(\beta-\beta_0)}\right| \\ \leq&\ \exp(Cv_{nT}\log n)\to 1.\end{aligned}$$

Thus

$$\begin{aligned}&\sup_{(x,\beta)\in B_{\delta_T}(Q_{X(\beta_0)}(\tau_x))\times B_{v_{nT}}(\beta_0)}|B_{nT}| \\ =&\sup_{(x,\beta)\in B_{\delta_T}(Q_{X(\beta_0)}(\tau_x))\times B_{v_{nT}}(\beta_0)}\left|\frac{c(x,\beta)(u_n)^{-\gamma(x,\beta)}(1+d(x,\beta)(u_n)^{-\tilde{\gamma}(x,\beta)}+r(x,u_n,\beta))}{c(x,\beta_0)(u_n)^{-\gamma(x,\beta_0)}(1+d(x,\beta_0)(u_n)^{-\tilde{\gamma}(x,\beta_0)}+r(x,u_n,\beta_0))}\right| \\ =&\ O(1),\end{aligned}$$

and hence (c) is established. The proof for $\xi(Q_{X(\beta_0)}(\tau_x),\beta_0)<0$ is identical to the positive case if y is replaced with y_0-y, and δ is small enough so that $\sup_{x\in B_\delta(Q_{X(\beta_0)}(\tau_x))}\xi(x)<0$.

Now it remains to prove (a)-(c) for $\xi(Q_{X(\beta_0)}(\tau_x),\beta_0)=0$. Note that $u_n \sim Q_{X(\beta_0)}(1-1/n)$, which is at most of the order $\exp(\Phi^{-1}(1-1/n)) = \exp(\sqrt{2\log n})$ by the condition $C_1(\log y)^2 \leq \tilde{d}(y) \leq C_2 y^{C_3}$.

For (a),

$$\begin{aligned}
&\sup_{x\in B_{\delta_T}(Q_{X(\beta_0)}(\tau_x))}\left|\frac{\partial F_{\varepsilon|X(\beta_0)=x}(u_n)/\partial x}{1-F_{\varepsilon|X(\beta_0)=x}(u_n)}\right| \\
\leq\ & \max\left\{\sup_{x\in B_{\delta_T}(Q_{X(\beta_0)}(\tau_x))\cap\{x:\xi(x,\beta_0)<0\}}\left|\frac{\partial F_{\varepsilon|X(\beta_0)=x}(u_n)/\partial x}{1-F_{\varepsilon|X(\beta_0)=x}(u_n)}\right|, \sup_{x\in B_{\delta}(Q_{X(\beta_0)}(\tau_x))\cap\{x:\xi(x,\beta_0)=0\}}\left|\frac{\partial F_{\varepsilon|X(\beta_0)=x}(u_n)/\partial x}{1-F_{\varepsilon|X(\beta_0)=x}(u_n)}\right|\right. \\
&\left.\sup_{x\in B_{\delta_T}(Q_{X(\beta_0)}(\tau_x))\cap\{x:\xi(x,\beta_0)>0\}}\left|\frac{\partial F_{\varepsilon|X(\beta_0)=x}(u_n)/\partial x}{1-F_{\varepsilon|X(\beta_0)=x}(u_n)}\right|\right\}.
\end{aligned}$$

By the same argument as $\xi>0$, the first and the third terms in the parenthesis are both $O(\log u_n)$, which is at most $O(\sqrt{\log n}) = o(T^{1/2})$. For the second term, apply Leibniz's rule and Condition B.(iii) to obtain

$$\begin{aligned}
&\sup_{x\in B_{\delta_T}(Q_{X(\beta_0)}(\tau_x))\cap\{x:\xi(x,\beta_0)=0\}}\left|\frac{\partial F_{\varepsilon|X(\beta_0)=x}(u_n)/\partial x}{1-F_{\varepsilon|X(\beta_0)=x}(u_n)}\right| \\
\leq\ & \sup_{x\in B_{\delta_T}(Q_{X(\beta_0)}(\tau_x))\cap\{x:\xi(x,\beta_0)=0\}} Cn\int_{u_n}^{y_0} y^{C_3} f_{\varepsilon|X(\beta_0)=x}(y)dy \\
\leq\ & Cn\int_{u_n}^{y_0} y^{C_3+\bar{C}_T}\exp(-\underline{D}_T(\log y)^2)dy \qquad (31)\\
=\ & Cn\int_{\log u_n}^{y_0}\exp(-\underline{D}_T s^2 + (C_3+\bar{C}_T+1)s)ds \text{ (by change of variables)} \\
=\ & O(1)
\end{aligned}$$

where I denote $\bar{C}_T = \sup_{x\in B_{\delta_T}(Q_{X(\beta_0)}(\tau_x))}\tilde{c}(x,\beta_0) < \infty$ and $\underline{D}_T = \inf_{x\in B_{\delta_T}(Q_{X(\beta_0)}(\tau_x))} d(x,\beta_0) > 0$, and the last equation follows from that u_n is at most of the order $\exp(\sqrt{2\log n})$ and the fact that $\Phi^{-1}(1-1/n) = O(\sqrt{\log(n)})$.

For (b),

$$\begin{aligned}
&\sup_{x\in B_{\delta_T}(Q_{X(\beta_0)}(\tau_x))}\left|\frac{\partial f_{\varepsilon|X(\beta_0)=x}(u_n)/\partial x}{f_{\varepsilon|X(\beta_0)=Q_{X(\beta_0)}(\tau_x)}(u_n)}\right| \leq \max\left\{\sup_{x\in B_{\delta_T}(Q_{X(\beta_0)}(\tau_x))\cap\{x:\xi(x,\beta_0)<0\}}\left|\frac{\partial f_{\varepsilon|X(\beta_0)=x}(u_n)/\partial x}{f_{\varepsilon|X(\beta_0)=Q_{X(\beta_0)}(\tau_x)}(u_n)}\right|\right. \\
&\left.\sup_{x\in B_{\delta_T}(Q_{X(\beta_0)}(\tau_x))\cap\{x:\xi(x,\beta_0)=0\}}\left|\frac{\partial f_{\varepsilon|X(\beta_0)=x}(u_n)/\partial x}{f_{\varepsilon|X(\beta_0)=Q_{X(\beta_0)}(\tau_x)}(u_n)}\right|, \sup_{x\in B_{\delta_T}(Q_{X(\beta_0)}(\tau_x))\cap\{x:\xi(x,\beta_0)>0\}}\left|\frac{\partial f_{\varepsilon|X(\beta_0)=x}(u_n)/\partial x}{f_{\varepsilon|X(\beta_0)=Q_{X(\beta_0)}(\tau_x)}(u_n)}\right|\right\}.
\end{aligned}$$

A similar argument as in (a) yields that the first ant the third items are $o(T^{1/2})$. The middle term is bounded by

$$\sup_{x\in B_{\delta_T}(Q_{X(\beta_0)}(\tau_x))}\left|\left(\frac{c_1(x,\beta)}{c(x,\beta)}+\frac{1}{u_n}\tilde{c}_1(x,\beta)+d_1(x,\beta)\tilde{d}(u_n)+\frac{r_1(x,\beta,u_n)}{1+r(x,\beta,u_n)}\right)\right|$$

$$\leq \quad O(u_n^{C_3}) = o(T^{1/2}).$$

For (c), a similar argument as in $\xi > 0$ and that in (a) yield for $j \in \{1, ..., \dim(\beta)\}$

$$\begin{aligned}
& \sup_{(x,\beta)\in B_{\delta_T}(Q_{X(\beta_0)}(\tau_x))\times B_{v_{nT}}(\beta_0)} \left| \frac{\partial F_{\varepsilon|X(\beta)=x}(u_n)/\partial\beta_j}{1-F_{\varepsilon|X(\beta)=x}(u_n)} \right| \\
\leq \quad & o(T^{1/2}) + \sup_{(x,\beta)\in (B_{\delta_T}(Q_{X(\beta_0)}(\tau_x))\times B_{v_{nT}}(\beta_0))\cap\{(x,\beta):\xi(x,\beta)=0\}} \left| \frac{\partial F_{\varepsilon|X(\beta)=x}(u_n)/\partial\beta_j}{1-F_{\varepsilon|X(\beta)=x}(u_n)} \right| \\
\leq \quad & o(T^{1/2}) + \sup_{(x,\beta)\in (B_{\delta_T}(Q_{X(\beta_0)}(\tau_x))\times B_{v_{nT}}(\beta_0))\cap\{(x,\beta):\xi(x,\beta)=0\}} \left| \frac{\int_{u_n}^{y_0} \partial f_{\varepsilon|X(\beta)=x}(y)/\partial\beta_j dy}{1-F_{\varepsilon|X(\beta)=x}(u_n)} \right|.
\end{aligned}$$

where the last line follows from Leibniz's rule. Then Condition B.(iii) and (31) imply

$$\begin{aligned}
& \sup_{(x,\beta)\in (B_{\delta_T}(Q_{X(\beta_0)}(\tau_x))\times B_{v_{nT}}(\beta_0))\cap\{(x,\beta):\xi(x,\beta)=0\}} \left| \frac{\int_{u_n}^{y_0} \partial f_{\varepsilon|X(\beta)=x}(y)/\partial\beta_j dy}{1-F_{\varepsilon|X(\beta)=x}(u_n)} \right| \\
\leq \quad & \sup_{(x,\beta)\in (B_{\delta_T}(Q_{X(\beta_0)}(\tau_x))\times B_{v_{nT}}(\beta_0))} n \left| \int_{u_n}^{y_0} f_{\varepsilon|X(\beta)=x}(y) \begin{pmatrix} \frac{c_2(x,\beta)}{c(x,\beta)} + \frac{1}{y}\tilde{c}_2(x,\beta) \\ +d_2(x,\beta)\tilde{d}(y) + \frac{r_2(x,\beta,y)}{1+r(x,\beta,y)} \end{pmatrix} dy \right| \\
\leq \quad & \sup_{(x,\beta)\in B_{\delta_T}(Q_{X(\beta_0)}(\tau_x))\times B_{v_{nT}}(\beta_0)} Cn \int_{u_n}^{y_0} y^{C_3} f_{\varepsilon|X(\beta)=x}(y) dy \\
= \quad & o(T^{1/2}).
\end{aligned}$$

The proof is then complete. ■

References

Abrevaya, J. (2001): "The effects of demographics and maternal Behavior on the distribution of birth outcomes," *Empirical Economics*, 26, 247–257.

Adrian, T., and M. K. Brunnermeier (2016): "CoVaR," *American Economic Review*, 106(7), 1705–1741.

Anderson, T. W., and C. Hsiao (1982): "Formulation and estimation of dynamic models using panel data," *Journal of Econometrics*, 18, 47–82.

Arellano, M., and S. Bond (1991): "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," *Review of Economic Studies*, 58, 277–279.

Arnold, B. C., N. Balakrishnan, and H. H. N. Nagaraja (1992): *A First Course in Order Statistics*. Siam.

Backus, D., M. Chernov, and I. Martin (2011): "Disasters implied by equity index options," *The Journal of Finance*, 66(6), 1969–2012.

Bai, J. (2009): "Panel data models with interactive fixed Effects," *Econometrica*, 77(4), 1229–1279.

Beare, B., and A. A. Toda (2017): "Geometrically Stopped Markovian Random Growth Processes and Pareto Tails," *arXiv:1712.01431*.

Beirlant, J., T. Wet, and Y. Goegebeur (2004): "Nonparametric Estimation of Extreme Conditional Quantiles," *Journal of Statistical Computation and Simulation*, 74, 567–580.

Bollerslev, T., and V. Todorov (2011): "Tails, fears, and risk premia," *The Journal of Finance*, 66(6), 2165–2211.

Chen, J., H. Hong, and J. C. Stein (2001): "Forecasting crashes: trading volume, past returns, and conditional skewness in stock prices," *Journal of Financial Economics*, 61, 345–381.

Chernozhukov, V. (2005): "Extremal Quantile Autoregression," *The Annals of Statistics*, 33(2), 806–839.

Chernozhukov, V., and I. Fernández-Val (2011): "Inference for extremal conditional quantile models, with an application to market and birthweight Risks," *The Review of Economic Studies*, 78, 559–589.

Chernozhukov, V., I. Fernández-Val, and T. Kaji (2016): "Extremal Quantile Regression: An Overview," *arXiv: 1612.06850*.

Chernozhukov, V., and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73(1), 245–261.

Chernozhukov, V., and L. Umantsev (2001): "Conditional value-at-risk: Aspects of modeling and estimation," *Empirical Economics*, 26(1), 271–293.

Coles, S. (2001): *An Introduction to Statistical Modeling of Extreme Values*. Springer, London.

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

Davidson, J. (1994): *Stochastic Limit Theory.* Oxford University Press, New York.

de Haan, L., and A. Ferreira (2007): *Extreme Value Theory: An Introduction.* Springer Science and Business Media, New York.

Ding, P. (2016): "On the Conditional Distribution of the Multivariate t Distribution," *The American Statistician*, 70, 293–295.

Engle, R. F., and S. Manganelli (2004): "CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles," *Journal of Business & Economic Statistics*, 22(4), 367–381.

Fisher, R. A., and L. H. C. Tippett (1928): *"Limiting forms of the frequency distribution of the largest or smallest member of a sample." Mathematical Proceedings of the Cambridge Philosophical Society.* Cambridge University Press.

Gabaix, X., J. Lasry, P. Lions, and B. Moll (2016): "The Dynamics of Inequality," *Econometrica*, 85(6), 2071–2111.

Galvao Jr., A. F. (2011): "Quantile regression for dynamic panel data with fixed effects," *Journal of Econometrics*, 164, 142–157.

Gardes, L., S. Girard, and A. Lekina (2010): "Functional Nonparametric Estimation of Conditional Extreme Quantiles," *Journal of Multivariate Analysis*, 101, 419–433.

Gardes, L., A. Guillou, and A. Schorgen (2012): "Estimating the Conditional Tail Index by Integrating a Kernel Conditional Quantile Estimator," *Journal of Statistical Planning and Inference*, 142, 1586–1598.

Gnedenko, B. V. (1943): "Sur la distribution limite du terme maximum d'une serie aléatoire," *Annals of Mathematics*, 44, 423–453.

Goldie, C. M., and R. L. Smith (1987): "Slow Variation with remainder: A survey of the theory and its applications," *Quarterly Journal of Mathematics*, 38, 45–71.

Hall, P. (1982): "On Some Simple Estimates of an Exponent of Regular Variation," *Journal of Royal Statistic Society, Series B.*

Hill, B. M. (1975): "A Simple General Approach to Inference about the Tail of a Distribution," *Annals of Statistics*, 3(5), 1163–1174.

Jones, C. I., and J. Kim (2018): "A Schumpeterian Model of Top Income Inequality," *Journal of Political Economy*, 126(5), 1785–1826.

Kelly, B., and H. Jiang (2014): "Tail Risk and Asset Prices," *The Review of Financial Studies*, 27(10), 2841–2871.

Koenker, R., and G. S. Bassett (1978): "Regression Quantiles," *Econometrica*, 46, 33–50.

Koenker, R., and K. Hallock (2001): "Quantile Regression: An introduction," *Journal of Economic Perspectives*, 15, 143–156.

Li, Q., and J. S. Racine (2007): *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

Martins-Filho, C., F. Yao, and M. Torero (2018): "Nonparametric Estimation of Conditional Valuat-at-Risk and Expected Shortfall Based on Extreme Value Theory," *Econometric Theory*, 34, 23–67.

Moon, H. R., and M. Weidner (2015): "Linear regression for panel with unknown number of factors as interactive fixed effects," *Econometrica*, 84(4), 1543–1579.

Müller, U. K., and Y. Wang (2017): "Fixed-k Asymptotic Inference about Tail Properties," *the Journal of the American Statistical Association*, 112, 1134–1143.

Pickands, III, J. (1975): "Statistical inference using extreme order statistics," *Annals of Statistics*, 3(1), 119–131.

Piketty, T., and E. Saez (2003): "Income inquality in the United States, 1913-1998," *The Quarterly Journal of Economics*, 118(1), 1–41.

Smith, R. L. (1982): "Uniform rates of convergence in Extreme-Value theory," *Advances in Applied Probability*, 13(3), 600–622.

——— (1987): "Estimating Tails of Probability Distributions," *Annals of Statistics*, 15, 1174–1207.

Toda, A. A. (2019): "Wealth Distribution with Random Discount Factors," *Journal of Monetary Economics*, 131(3), 571–592.

WANG, H., AND D. LI (2013): "Estimation of Extreme Conditional Quantiles Through Power Transformation," *Journal of the American Statistical Association*, 108(503), 1062–1074.

WANG, H., AND C. L. TSAI (2009): "Tail Index Regression," *Journal of the American Statistical Association*, 104, 1233–1240.

WANG, Y. (2018): "Unbiased Estimation of Tail Properties in Small Samples with Complete, Censored, or Truncated Data," *Working Paper*.

WOOLDRIDGE, J. M. (2002): *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, Cambridge, Massachusetts.

WU, W. B. (2004): "Empirical processes of dependent random variables," *arXiv preprint math/0412267*.

——— (2005): "On the Bahadur Representation of Sample Quantiles for Dependent Sequences," *Annuals of Statistics*, 33(4), 1934–1963.

YANG, S. S. (1977): "General distribution theory of the concomitants of order Statistics," *The Annals of Statistics*, 5(5), 996–1002.