Reinforced Tail Quantile Regression

Xuan Leng[*] Yi He[†] Yongmiao Hong [‡] Yizhou Zhang[§]

This version: January 24, 2026

Abstract. Capturing tail phenomena with quantile regression (QR) typically requires the quantile level to vary with the sample size. Consequently, the effective number of observations contributing to estimation shrinks rapidly, inducing substantial variance inflation in the conventional QR estimator. Exploiting the widely observed power-law behavior of heavy-tailed distributions, this paper proposes a new estimation method for QR parameters in the tail region. The proposed approach incorporates additional tail information and substantially reinforces estimation precision relative to conventional QR. The framework accommodates both heavy-tailed and light-tailed outcomes within linear QR models. We establish the asymptotic normality of the proposed estimator over the entire tail region. Building on an orthogonal decomposition of the estimation error, we develop a multiplier double bootstrap procedure that enables valid inference. Simulation studies demonstrate that our estimator produces markedly narrower confidence intervals than QR estimator. We apply the method to estimate the marginal effect of education on upper-tail incomes.

Keywords: Quantile regressions, Power laws, Extreme value statistics, Heterogeneous data, Returns to education

[*]Xiamen University, China.

[†]University of Amsterdam, Netherlands.

[‡]University of Chinese Academy of Sciences; Chinese Academy of Sciences, Beijing 100190, China. Email: yh20@cornell.edu.

[§]Kyoto University, Japan.

1 Introduction

Koenker and Bassett (1978) introduce quantile regression (QR), a natural generalization of sample quantiles from the location model to linear model. By minimizing a weighted sum of absolute residuals, the QR estimator exhibits robustness properties under non-Gaussian errors such as heavy-tailed distributions. Since then, QR has become a widely used tool for analyzing heterogeneous covariate effects in economics, finance, and health studies. For a prespecified and fixed quantile level $\tau \in (0,1)$, the asymptotic normality of the τ-QR estimator has been studied under suitable regularity conditions; see, e.g., Koenker (2005, Section 4.3) and references therein. The estimation precision is positively associated with the conditional density of the outcome variable evaluated at its τ-th conditional quantile.

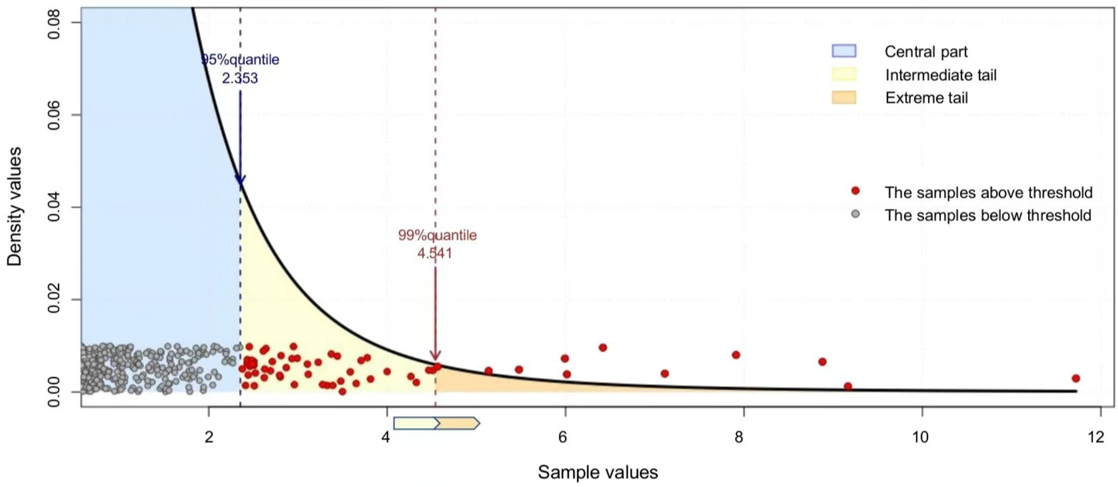
In risk analysis, investment, insurance, and policy evaluation, empirical interest often focuses on the tail region of the conditional distribution. For example, Engle and Manganelli (2004), Hong et al. (2009), and Adrian and Brunnermeier (2016) examine extreme or systemic risk in financial markets through tail conditional quantiles; Xiao (2014) highlights the particular value of right-tail information in asset return distributions for investors and portfolio managers; Wang and Li (2013) investigate the effects of chronic conditions on the upper quantiles of medical costs; and Sasaki and Wang (2025) evaluate policy interventions aimed at populations with extremely high net savings. Capturing such tail phenomena typically requires the target quantile level to grow with the sample size. As τ approaches 1, the conditional density evaluated at the τ-th conditional quantile becomes arbitrarily small. As a consequence, the effective number of observations contributing to QR estimation shrinks rapidly.

By integrating extreme value theory with QR, Chernozhukov (2005) rigorously establishes the asymptotic theory of QR estimators for quantile levels approaching 1 or 0. The resulting asymptotic behavior depends critically on the effective sample size and distinguishes between two tail regimes. As illustrated in Figure 1, which depicts the upper tail of a heavy-tailed distribution, the intermediate tail represents a transition regime in which the number of outcome observations remains large. In contrast, the extreme tail is characterized by data scarcity. In this regime, the τ-QR estimator loses asymptotic normality and under heavy-tailed outcomes its estimation error diverges, rendering the conventional QR estimator unreliable. To address inference in the extreme tail, Chernozhukov and Fernández-Val (2011) develop a subsampling-based procedure for the QR estimator. Subsequent studies have extended the extremal QR framework of Chernozhukov (2005) to settings involving censored outcomes (Altonji et al., 2012), treatment effects (Zhang, 2018; Deuber et al., 2024), endogenous selection (D'Haultfœuille et al., 2018), and time-series models (Zhang, 2021; Daouia et al., 2023), among others. Nevertheless, relatively little work has focused on improving the estimation accuracy of QR parameters in the tail region. The QR estimator may suffer from variance inflation and numerical instability in tails, as noted by He et al. (2023). As emphasized by Chernozhukov et al. (2017), "the τ-QR estimator can be very inaccurate" when τ is very close to 1 or 0.

This paper proposes a novel method for estimating τ-QR parameters when the quantile level τ approaches 1 or 0. Without loss of generality, we focus on the upper-tail. By approximating the vanishing tail probability in the population τ-QR score using an empirical power law representation, the proposed estimator effectively exploits information from the entire tail region above a threshold. As a result, the effective number of observations contributing to estimation does not shrink as τ approaches 1. The proposed method substantially reinforces estimation precision in the tail region relative to the conventional QR. Specifically, under suitable regularity conditions, the proposed estimator attains the same convergence rate in the extreme-tail regime as in the intermediate-tail regime.

Semiparametric Pareto approximation from extreme value theory (e.g., de Haan

Figure 1: Sparsity and integration in the extreme tail



Notes: The upper tail of the $t(3)$ *distribution with the 95% quantile taken as the threshold. The tail region above this threshold contains 58 observations (red circles), of which 47 belong to the intermediate tail, while the remaining 11 constitute the extreme tail.*

and Ferreira (2006) and Resnick (1987)) provides an important, if not the only, statistical method for extrapolation beyond the observed data. For a heavy-tailed random variable with quantile function Q, the tail can be extrapolated via

$$Q(\tau) \approx \left(\frac{1-\tau_0}{1-\tau}\right)^{\gamma} Q(\tau_0), \tag{1}$$

where τ corresponds to an extreme quantile level, while τ_0 lies in the intermadiate tail regime. The exponent $\gamma > 0$ is the extreme value index, which characterizes tail heaviness. Weissman (1978) uses this approximation to extrapolate intermediate sample quantile into the extreme tail with fitted γ. More recently, Wang et al. (2012) and Hou et al. (2024) extend this approach to the QR setting. Specifically, they estimate the conditional quantile at an intermediate level τ_0 using QR and extrapolate it to an extreme level τ according to the Pareto approximation. The rationale behind the approximation rests on the empirical power law behavior widely observed in real world data.

Unlike prior studies that apply Pareto approximation to estimate high conditional quantiles, our focus is on QR parameters for their economic and policy implications in the tail region. The proposed estimator is obtained by solving a convex and smooth optimization problem constructed via a convolution with an integrated Pareto distribution. Our estimator remains asymptotically normal in both the intermediate and extreme tail regimes. Moreover, owing to the Pareto-based convolution, the precision matrix of our estimator no longer involves the conditional density of the outcome evaluated at its τ-th conditional quantile, which vanishes as τ approaches 1. Leveraging recent advances in extreme value theory for heterogeneous data by Einmahl et al. (2016) and Einmahl and He (2023), we estimate the extreme value index of the underlying Pareto distribution directly from heterogeneous outcomes rather than from conditional quantiles. This yields a covariate-free approach that remains robust to the choice and dimensionality of covariates. We accommodate both heavy-tailed and light-tailed outcomes in linear QR models; for light-tailed outcomes, an exponential transformation is employed to induce power law behavior.

Building on the established asymptotic normality and an orthogonal decomposition of the estimation error, we develop a multiplier double bootstrap procedure that enables valid inference. Recent work by He et al. (2023) and Galvao et al. (2024) investigates random-weight bootstrap methods within the QR framework. Our numerical analysis, including simulation studies and an application to returns to education at high income levels, demonstrates appealing performance of the proposed estimation and inference methods. In particular, the proposed estimator produces markedly narrower confidence intervals than conventional QR when the effective sample size is small.

The remainder of the paper is organized as follows. Section 2 introduces the proposed estimator for QR parameters in the tail region and establishes its asymptotic normality. Section 3 proposes a bootstrap procedure with random weights and proves its asymptotic validity. Simulation studies in Section 4 demonstrate the advantages of the proposed method relative to conventional QR. Section 5 applies the method to an income dataset. Section 6 concludes.

2 New estimator for linear QR

Let Y_i be a scalar response and $\boldsymbol{X}_i = (X_{i1}, \dots, X_{id})^\top$ a d-dimensional vector of covariates for unit i, where d is fixed. Denote by $Q_{Y_i}(\tau \mid \boldsymbol{x}_i)$ the conditional τ-quantile of Y_i given $\boldsymbol{X}_i = \boldsymbol{x}_i$ for $\tau \in (0, 1)$. A commonly used specification assumes that the conditional quantile function is linear in the covariates:

$$Q_{Y_i}(\tau \mid \boldsymbol{x}_i) = \alpha_\tau + \boldsymbol{x}_i^\top \boldsymbol{\beta}_\tau, \tag{2}$$

where $\alpha_\tau \in \mathbb{R}$ is the quantile-specific intercept and $\boldsymbol{\beta}_\tau = (\beta_1(\tau), \dots, \beta_d(\tau))^\top \in \mathbb{R}^d$ is the vector of quantile-specific slope coefficients. Koenker and Xiao (2002) propose tests for linear QR models. Let $\rho_\tau(u) = u\{\tau - \mathbb{1}(u < 0)\}$ denote the check loss function. The τ-QR coefficients of model (2) minimize the expected check loss:

$$(\alpha_\tau, \boldsymbol{\beta}_\tau) = \arg\min_{\alpha, \boldsymbol{\beta}} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\big\{\rho_\tau(Y_i - \alpha - \boldsymbol{x}_i^\top \boldsymbol{\beta})\big\}, \tag{3}$$

where $\mathbb{E}$ denotes expectation conditional on the covariates and n is the sample size. Suppose that $\{Y_i, \boldsymbol{X}_i\}_{i=1}^n$ are independent over i. The τ-QR estimator is then obtained by minimizing the corresponding empirical counterpart,

$$(\widehat{\alpha}_{\text{qr}}(\tau), \widehat{\boldsymbol{\beta}}_{\text{qr}}(\tau)) = \arg\min_{\alpha, \boldsymbol{\beta}} \frac{1}{n} \sum_{i=1}^{n} \rho_\tau(Y_i - \alpha - \boldsymbol{x}_i^\top \boldsymbol{\beta}). \tag{4}$$

For a comprehensive overview of QR methodology under a fixed quantile level τ, see Koenker (2005). In this paper, we focus on precise estimation of the QR parameters in the tail region where $\tau = \tau_n \uparrow 1$ as $n \to \infty$. We distinguish between two tail regimes: an intermediate tail, where $(1 - \tau)n \to \infty$, and an extreme tail, where

$(1-\tau)n \to C < \infty$. If the lower tail ($\tau = \tau_n \downarrow 0$) is of interest, one can equivalently consider $-Y_i$ instead of Y_i.

Suppose that the conditional distribution of Y_i given $\boldsymbol{x}_i$ is heavy-tailed. We order the observations $Y_1, \ldots, Y_n$ as $Y_{1,n} \leq \ldots \leq Y_{n,n}$, and take the $(k+1)$-th largest observation $Y_{n-k,n}$ as the threshold t_k. Here k denotes the number of the observations exceeding threshold t_k. Then rewrite the population objective function in (3) as

$$(\alpha_\tau, \boldsymbol{\beta}_\tau) = \arg\min_{\alpha,\boldsymbol{\beta}} \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\big\{\rho_\tau\big((Y_i - t_k) - (\alpha + \boldsymbol{x}_i^\top \boldsymbol{\beta} - t_k)\big)\big\}. \tag{5}$$

A fundamental result in extreme value statistics states that the exceedances over a high threshold, $\{Y_{n-k+i} - t_k\}_{i=1}^{k}$, asymptotically follow a generalized Pareto distribution under mild regularity conditions, known as the peak-over-threshold approach (Pickands III, 1975). Motivated by this property, we propose a Pareto-convolution-based loss to approximate the expected check loss for τ in the tail region. Specifically, we modify the standard check loss ρ_τ and define

$$\rho_{\tau,D}(u) = u\{\mathbb{1}(u > 0)D - (1-\tau)\}, \quad D = \mathbb{1}\{Y > t_k\}.$$

This tail-adapted check loss incorporates the exceedance indicator D into ρ_τ. When Y exceeds the threshold, $\rho_{\tau,D}$ coincides with ρ_τ. Otherwise, $\rho_{\tau,D}$ becomes asymptotically negligible as $\tau \to 1$. Thus, the modified check loss effectively downweights the non-tail observations. Based on (5), we define the Pareto-convolution-based loss as

$$\ell_{\tau,D}(v) = \int_0^\infty \{\rho_{\tau,D}(u - v) - \rho_{\tau,D}(u)\}\, dG_{\widehat{\gamma}_H, t_k}(u), \tag{6}$$

G_{γ,t_k} denotes the Pareto distribution function

$$G_{\gamma,t_k}(u) = 1 - \left(1 + \frac{u}{t_k}\right)^{-1/\gamma}, \quad u > 0,$$

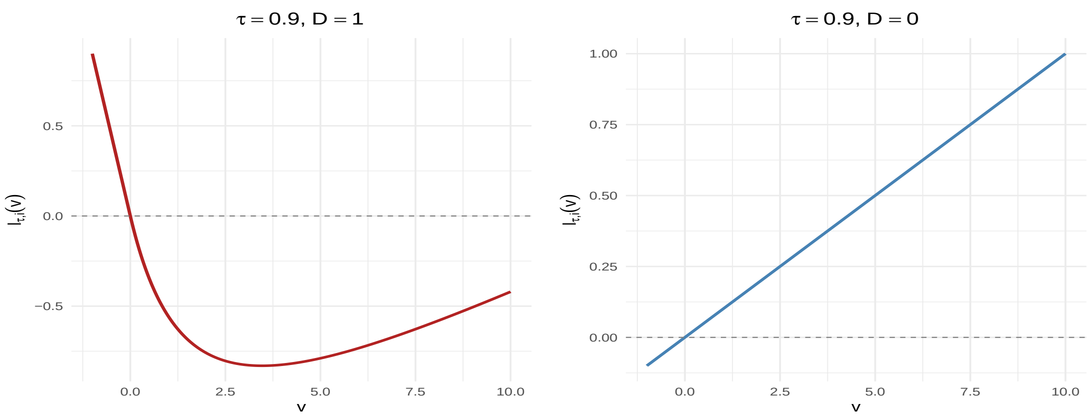
and $\widehat{\gamma}_H$ is the Hill estimator (Hill, 1975) for the shape parameter $\gamma > 0$,

$$\widehat{\gamma}_H = \frac{1}{k}\sum_{i=1}^{k}(\log Y_{n-k+i,n} - \log t_k). \tag{7}$$

Here γ corresponds to the extreme value index of the conditional outcome distributions. Under the linear QR model (2), Proposition 2.1 of Wang and Li (2013) shows that γ is invariant with respect to the covariates.

Figure 2 illustrates the shape of the convolution-based loss function $\ell_{\tau,D}(v)$ which is both convex and smooth. Note that the term $\rho_{\tau,D}(u-v)$ in the integrand of $\ell_{\tau,D}(v)$ corresponds to the check loss $\rho_\tau((Y_i - t_k) - (\alpha + \boldsymbol{x}_i^\top \boldsymbol{\beta} - t_k))$ in (5), while the integrator $dG_{\widehat{\gamma}_H,t_k}$ plays the role of the expectation $\mathbb{E}$. The additional subtraction of $\rho_{\tau,D}(u)$ ensures that the integral in (6) is well defined for any $\widehat{\gamma}_H > 0$. Recently, the asymptotic properties of the Hill estimator for γ under the independent but non-identically distributed univariate samples have been studied by Einmahl et al. (2016) and Einmahl and He (2023).

Figure 2: Visualization of function $\ell_{\tau,D}$ by setting $t_k = 3$, $\widehat{\gamma}_H = 1/3$ and $\tau = 0.9$.



Finally, our estimator of $(\alpha_\tau, \boldsymbol{\beta}_\tau)$ for $\tau = \tau_n \to 1$ is defined as

$$(\widehat{\alpha}(\tau), \widehat{\boldsymbol{\beta}}(\tau)) = \arg\min_{\alpha,\boldsymbol{\beta}} \frac{1}{n}\sum_{i=1}^{n} \underbrace{\ell_{\tau,D_i}(\alpha + \boldsymbol{x}_i^\top \boldsymbol{\beta} - t_k)}_{\text{convex in } (\alpha,\boldsymbol{\beta})}, \tag{8}$$

where $D_i = \mathbb{1}\{Y_i > t_k\}$ indicates whether observation Y_i exceeds the threshold. Unlike conventional QR, the proposed objective function is differentiable. The approach is related to convolution-based smoothing methods for QR, such as Fernandes et al. (2021) and He et al. (2023), which employ kernel-based nonparametric smoothers. Define the function

$$\widehat{\boldsymbol{s}}_\gamma(\alpha, \boldsymbol{\beta}) := \frac{1}{n}\sum_{i=1}^{n}(1, \boldsymbol{x}_i^\top)^\top \left\{ \left(\frac{\alpha + \boldsymbol{x}_i^\top \boldsymbol{\beta}}{t_k}\right)^{-1/\gamma} D_i - (1-\tau) \right\}.$$

The following theorem shows that $\widehat{\boldsymbol{s}}_{\widehat{\gamma}_H}()$ is the gradient of the proposed objective function with respect to parameters.

Theorem 1. *The proposed estimator from (8) satisfies* $\widehat{\boldsymbol{s}}_{\widehat{\gamma}_H}(\widehat{\alpha}(\tau), \widehat{\boldsymbol{\beta}}(\tau)) = \mathbf{0}$.

Note that the gradient of the population objective function in (3) is given by

$$\boldsymbol{s}(\alpha, \boldsymbol{\beta}) := \frac{1}{n}\sum_{i=1}^{n}(1, \boldsymbol{x}_i^\top)^\top \left\{ \mathbb{P}(Y_i > \alpha + \boldsymbol{x}_i^\top \boldsymbol{\beta} \mid \boldsymbol{x}_i) - (1-\tau) \right\},$$

which satisfies $\boldsymbol{s}(\alpha_\tau, \boldsymbol{\beta}_\tau) = \mathbf{0}$ at the true parameter values. When τ approaches 1, $\widehat{\boldsymbol{s}}_{\widehat{\gamma}_H}(\alpha_\tau, \boldsymbol{\beta}_\tau)$ provides a tractable approximation to $\boldsymbol{s}(\alpha_\tau, \boldsymbol{\beta}_\tau)$. To see this, rewrite the population gradient as

$$\boldsymbol{s}(\alpha, \boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}(1, \boldsymbol{x}_i^\top)^\top \left\{ \frac{\mathbb{P}(Y_i > \alpha + \boldsymbol{x}_i^\top \boldsymbol{\beta} \mid \boldsymbol{x}_i)}{\mathbb{P}(Y_i > t_k \mid \boldsymbol{x}_i)} \cdot \mathbb{P}(Y_i > t_k \mid \boldsymbol{x}_i) - (1-\tau) \right\}. \quad (9)$$

Since $Y_i \mid \boldsymbol{x}_i$ follows a heavy-tailed distribution with extreme value index γ, the conditional tail probabilities obey the power-law relation

$$\lim_{y\to\infty} \frac{\mathbb{P}(Y_i > sy \mid \boldsymbol{x}_i)}{\mathbb{P}(Y_i > y \mid \boldsymbol{x}_i)} = s^{-1/\gamma}, \qquad s > 0.$$

This implies

$$\frac{\mathbb{P}(Y_i > \alpha_\tau + \boldsymbol{x}_i^\top \boldsymbol{\beta}_\tau \mid \boldsymbol{x}_i)}{\mathbb{P}(Y_i > t_k \mid \boldsymbol{x}_i)} \approx \left(\frac{\alpha_\tau + \boldsymbol{x}_i^\top \boldsymbol{\beta}_\tau}{t_k}\right)^{-1/\gamma}. \quad (10)$$

Combining (9) and (10) yields

$$
\begin{aligned}
\boldsymbol{s}(\alpha_\tau, \boldsymbol{\beta}_\tau) &\approx \frac{1}{n}\sum_{i=1}^{n}(1, \boldsymbol{x}_i^\top)^\top \left\{ \left(\frac{\alpha_\tau + \boldsymbol{x}_i^\top \boldsymbol{\beta}_\tau}{t_k} \right)^{-1/\gamma} \cdot \mathbb{P}(Y_i > t_k \mid \boldsymbol{x}_i) - (1-\tau) \right\} \\
&\approx \widehat{\boldsymbol{s}}_{\widehat{\gamma}_H}(\alpha_\tau, \boldsymbol{\beta}_\tau), \qquad (11)
\end{aligned}
$$

where the final approximation replaces the unknown exceedance probability $\mathbb{P}(Y_i > t_k \mid \boldsymbol{x}_i)$ with its empirical counterpart under suitable regularity conditions.

By leveraging power-law behavior, our method avoids direct estimation of the vanishing tail probability, $\mathbb{P}(Y_i > \alpha_\tau + \boldsymbol{x}_i^\top \boldsymbol{\beta}_\tau \mid \boldsymbol{x}_i) = 1 - \tau \downarrow 0$, which appears in the population τ-QR score. This enables us to effectively use the k exceedances above a threshold to estimate the τ-QR parameters for all $\tau > 1 - k/n$, regardless of whether τ lies in the intermediate or extreme tail regime. In extreme value analysis, $k = k_n$ is typically taken as an intermediate sequence satisfying $k \to \infty$ and $k/n \to 0$. Relative to the conventional τ-QR estimator, the convolution-based approach substantially reinforces estimation precision, particularly when $k/\{n(1-\tau)\} \to \infty$. We refer to $(\widehat{\alpha}(\tau), \widehat{\boldsymbol{\beta}}(\tau))$ as the *reinforced tail quantile regression* (ReQR) estimator.

We establish the consistency of the ReQR estimator, in the sense that the estimation errors are asymptotically negligible relative to the true parameter values. The required regularity conditions are provided in Appendix A of the online supplementary material. In particular, we assume that the conditional outcome distributions are tail-equivalent and asymptotically proportional to a common "mother" heavy-tailed distribution F,

$$
\sup_{i \geq 1} \left| \frac{1 - F_i(y \mid \boldsymbol{x}_i)}{1 - F(y)} - g_{ni} \right| \to 0, \quad \text{as } y \to \infty.
$$

Here, g_{ni} captures covariate-induced heterogeneity in tail scale across outcome variables, which we refer to as skedastic factors. When $g_{ni} \equiv 1$ for all i, the conditional outcome distributions have asymptotically identical upper tails. Several illustrative examples of g_{ni} are provided in Appendix A.

Proposition 1. *Suppose that Assumptions S.1–S.4 in Appendix A hold and that* $\tau = \tau_n \uparrow 1$ *as* $n \to \infty$. *Then* $\frac{\widehat{\alpha}(\tau)-\alpha_\tau}{Q(\tau)} \xrightarrow{\mathbb{P}} 0$, *and* $\frac{\widehat{\boldsymbol{\beta}}(\tau)-\boldsymbol{\beta}_\tau}{Q(\tau)} \xrightarrow{\mathbb{P}} \mathbf{0}$, *where* $Q(\tau) := \inf\{y : F(y) \geq \tau\}$ *denotes the quantile function of* F.

Since $Q_{Y_i}(\tau \mid \boldsymbol{x}_i) = \alpha_\tau + \boldsymbol{x}_i^\top \boldsymbol{\beta}_\tau \to \infty$, it follows that either $\alpha_\tau \to \infty$ or $\|\boldsymbol{\beta}_\tau\| \to \infty$, or both. Without loss of generality, we assume that $\alpha_\tau \to \infty$. Note that α_τ is asymptotically proportional to $Q(\tau)$. Proposition 1 then implies $\frac{\widehat{\alpha}(\tau)}{\alpha_\tau} - 1 = o_{\mathbb{P}}(1)$, as $n \to \infty$. A similar consistency result holds for the slope coefficients $\boldsymbol{\beta}_\tau$. Specifically, for the location-scale shift model (Example 1 in simulation studies), Proposition 1 yields $\max_{1\leq j\leq d} \left|\frac{\widehat{\beta}_j(\tau)}{\beta_{j,\tau}} - 1\right| = o_{\mathbb{P}}(1)$, where $\widehat{\boldsymbol{\beta}}(\tau) = (\widehat{\beta}_1(\tau), \ldots, \widehat{\beta}_d(\tau))^\top$ and $\boldsymbol{\beta}_\tau = (\beta_{1,\tau}, \ldots, \beta_{d,\tau})^\top$.

Building on the consistency result in Proposition 1, we establish the asymptotic normality of the ReQR slope estimator.

Theorem 2. *Suppose that Assumptions S.1–S.5 in Appendix A hold. Let* k *be an intermediate sequence and consider quantile levels* $\tau > 1 - k/n$.

(i) If $\frac{k}{n(1-\tau)} \to c \in (1, \infty)$ *as* $n \to \infty$, *we have*

$$\sqrt{k}\,\frac{\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_\tau}{Q(\tau)} \xrightarrow{\mathbb{D}} \mathcal{N}\big(\mathbf{0}, \gamma^2 \Gamma_{\boldsymbol{xx}}^{-1} \Phi_{c,\boldsymbol{xx}} \Gamma_{\boldsymbol{xx}}^{-1}\big).$$

(ii) If $\frac{k}{n(1-\tau)} \to \infty$ *and* $\frac{\sqrt{k}}{\log^2(n(1-\tau)/k)} \to \infty$ *as* $n \to \infty$, *then the following results hold.*

(a) When the conditional outcome distributions have asymptotically identical upper tails,

$$\sqrt{k}\frac{\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_\tau}{Q(\tau)} \xrightarrow{\mathbb{D}} \mathcal{N}\big(\mathbf{0}, \gamma^2 \Gamma_{\boldsymbol{xx}}^{-1}\big),$$

where $\Gamma_{\boldsymbol{xx}} = \lim_{n\to\infty} 1/n \sum_{i=1}^{n} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_n)(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_n)^\top$ *in this simple case.*

(b) When the conditional outcome distributions are asymptotically non-identical in the upper tails,

$$\frac{\sqrt{k}}{\log(n(1-\tau)/k)}\frac{\widehat{\boldsymbol{\beta}}(\tau)-\boldsymbol{\beta}_\tau}{Q(\tau)} \xrightarrow{\mathbb{D}} \mathcal{N}\big(\mathbf{0}, \gamma^2\Gamma_{\boldsymbol{xx}}^{-1}\boldsymbol{h}_{\boldsymbol{x}}\boldsymbol{h}_{\boldsymbol{x}}^\top\Gamma_{\boldsymbol{xx}}^{-1}\big).$$

The explicit expressions for the asymptotic variance matrices are provided in Assumption S.5 of Appendix A.

Theorem 2 establishes the asymptotic normality under both intermediate and extreme quantile levels. The variance matrices rely on weighted sample moments of the covariates incorporating skedastic factors and no longer involve the conditional outcome densities. Case (i) belongs to the intermediate tail regime, in which $n(1-\tau)$ is proportional to $k\to\infty$, and $n(1-\tau)<k$ for large n. In this case, the ReQR estimator converges at the rate $k^{-1/2}$, whereas the conventional QR estimator converges at the slower rate $\{n(1-\tau)\}^{-1/2}$ (Chernozhukov, 2005, Theorem 5.1 for type 2 tails). If the conditional outcome distributions have asymptotically identical upper tails, the result of case (i) coincides with that in case $(ii)(a)$. For the intermediate tail satisfying $n(1-\tau)=k^\alpha$ with $\alpha\in(0,1)$, by case $(ii)(a)$, the ReQR slope estimator converges at the rate $k^{-1/2}$, whereas the conventional QR estimator converges at the slower rate $k^{-\alpha/2}$.

Case (ii) also covers the extreme tail regime in which $n(1-\tau)\to C<\infty$. In this regime, the estimator exhibits asymptotic behavior analogous to that in the intermediate tail and retains asymptotic normality. In contrast, the conventional QR estimator in the extreme tail has a non-Gaussian limiting distribution (Chernozhukov, 2005, Theorem 4.1 for type-2 tails) and converges at a substantially slower rate.

Remark 1. The joint asymptotic normality of $\widehat{\alpha}(\tau)$ and $\widehat{\boldsymbol{\beta}}(\tau)$ is established in Theorem **??** of online supplementary material Appendix B.

2.1 ReQR estimator for Log-transformed linear QR

Many economics and finance variables, such as income, wealth, and asset returns, are highly skewed. A common practice to mitigate skewness is to apply a logarithmic transformation. We consider the following log-linear QR model:

$$Q_{\log Y_i}(\tau \mid \boldsymbol{x}_i) = \alpha_\tau + \boldsymbol{x}_i^\top \boldsymbol{\beta}_\tau, \tag{12}$$

where the conditional τ-quantile of $\log Y_i$ is specified as a linear function of covariates. By the equivariance of quantiles to monotone transformations, $Q_{\log Y_i}(\tau \mid \boldsymbol{x}_i) = \log Q_{Y_i}(\tau \mid \boldsymbol{x}_i)$. Thus, the conditional quantile of Y_i is given by

$$Q_{Y_i}(\tau \mid \boldsymbol{x}_i) = \exp(\alpha_\tau + \boldsymbol{x}_i^\top \boldsymbol{\beta}_\tau). \tag{13}$$

We extend the ReQR method developed in Section 2 to the log-transformed QR model (12) which accommodates light-tailed outcomes within linear QR framework. More generally, the proposed approach can be adapted to Box-Cox transformed QR models, as studied in Mu and He (2007) and Wang and Li (2013).

Under Assumptions A.2 and A.3 in Appendix A, the conditional distribution of Y_i given X_i is heavy-tailed with extreme value index γ. Based on (13), we define the ReQR estimator as

$$(\widetilde{\alpha}(\tau), \widetilde{\boldsymbol{\beta}}(\tau)) = \arg\min_{\alpha, \boldsymbol{\beta}} \frac{1}{n} \sum_{i=1}^{n} \ell_{\tau, D_i}\big(\exp(\alpha + \boldsymbol{x}_i^\top \boldsymbol{\beta}) - t_k\big), \tag{14}$$

where the smooth loss function $\ell_{\tau,D}$ is defined in (6). Since the exponential function is convex, the objective function in (14) is convex in $(\alpha, \boldsymbol{\beta})$. The theorem below shows the gradient of the objective function with respect to the parameters.

Theorem 3. *The ReQR estimator defined in* (14) *satisfies*

$$\frac{1}{n}\sum_{i=1}^{n}(1,\boldsymbol{x}_i^\top)^\top \widetilde{Q}_{Y_i}(\tau \mid \boldsymbol{x}_i)\left\{\left(\frac{\widetilde{Q}_{Y_i}(\tau \mid \boldsymbol{x}_i)}{t_k}\right)^{-1/\widehat{\gamma}_H} D_i - (1-\tau)\right\} = \mathbf{0},$$

where $\widetilde{Q}_{Y_i}(\tau \mid \boldsymbol{x}_i) := \exp(\widetilde{\alpha}(\tau) + \boldsymbol{x}_i^\top \widetilde{\boldsymbol{\beta}}(\tau))$ *and* $D_i = \mathbb{1}\{Y_i > t_k\}$.

Note that the population score evaluated at $(\alpha_\tau, \boldsymbol{\beta}_\tau)$ is given by

$$\frac{1}{n}\sum_{i=1}^{n}(1,\boldsymbol{x}_i^\top)^\top Q_{Y_i}(\tau \mid \boldsymbol{x}_i)\big\{\mathbb{P}(Y_i > Q_{Y_i}(\tau \mid \boldsymbol{x}_i) \mid \boldsymbol{x}_i) - (1-\tau)\big\} = \mathbf{0},$$

where $Q_{Y_i}(\tau \mid \boldsymbol{x}_i) = \exp(\alpha_\tau + \boldsymbol{x}_i^\top \boldsymbol{\beta}_\tau)$. Analogous to the analysis of Theorem 1 for the linear QR model, the feasible score in Theorem 3 approximates this population score by invoking a power-law approximation for the tail probability $\mathbb{P}(Y_i > Q_{Y_i}(\tau \mid \boldsymbol{x}_i) \mid \boldsymbol{x}_i) = 1-\tau$. The next proposition establishes the consistency of the ReQR estimator.

Proposition 2. *Suppose that Assumptions S.1–S.4 hold and that* $\tau = \tau_n \to 1$ *as* $n \to \infty$. *Then* $\widetilde{\alpha}(\tau) - \alpha_\tau \xrightarrow{\mathbb{P}} 0$ *and* $\widetilde{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_\tau \xrightarrow{\mathbb{P}} \mathbf{0}$.

Building on this consistency result, we derive the asymptotic normality of the ReQR slope estimator.

Theorem 4. *Suppose that Assumptions S.1–S.4 and Assumption S.6 hold. Let* k *be an intermediate sequence and consider quantile levels* $\tau > 1-k/n$. *If* $\frac{\sqrt{k}}{\log^2(n(1-\tau)/k)} \to \infty$ *as* $n \to \infty$, *we have*

$$\sqrt{k}\big(\widetilde{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_\tau\big) \xrightarrow{\mathbb{D}} \mathcal{N}\big(\mathbf{0}, \gamma^2 \Omega_{\boldsymbol{xx}}^{-1}\Psi_{\boldsymbol{xx}}\Omega_{\boldsymbol{xx}}^{-1}\big),$$

where the asymptotic variance matrix is defined in Assumption S.6 of Appendix A.

Theorem 4 establishes the asymptotic normality of the ReQR estimator for the log-linear QR model (12), which allows $\log Y_i$ to be light-tailed. Relative to the corresponding result for the linear QR model (2), the asymptotic behavior of the log-linear QR is more transparent. In particular, the slope estimator exhibits the same asymptotic distribution in both the intermediate- and extreme-tail regimes. The resulting improvement in estimation accuracy in the tail region constitutes a key advantage of the reinforced approach over the conventional QR.

Remark 2. The joint asymptotic normality of $\widetilde{\alpha}(\tau)$ and $\widetilde{\boldsymbol{\beta}}(\tau)$ is established in Theorem **??** of online supplementary material Appendix B.

3 Inference: Multiplier double bootstrap

Since the asymptotic variances established in Theorems 2 and 4 depend on the unknown skedastic factors g_{ni}, direct plug-in estimation is challenging. Building on the asymptotic normality of the ReQR estimator in both the intermediate and extreme tails, together with an orthogonal decomposition of the estimation error, we develop a multiplier double bootstrap procedure that delivers valid inference over the entire tail region.

Models (2) and (13) can be written in a unified form as $Q_{Y_i}(\tau \mid \boldsymbol{x}_i) = \Lambda\big(\alpha_\tau + \boldsymbol{x}_i^\top \boldsymbol{\beta}_\tau\big)$, where the link function $\Lambda(\cdot)$ is defined by

$$\Lambda(x) = \begin{cases} x, & \text{linear QR model (2)}, \\ \exp(x), & \text{log-linear QR model (12)}. \end{cases}$$

Accordingly, the ReQR estimator can be defined uniformly as

$$(\widehat{\alpha}_\Lambda(\tau), \widehat{\boldsymbol{\beta}}_\Lambda(\tau)) = \arg\min_{\alpha, \boldsymbol{\beta}} \frac{1}{n} \sum_{i=1}^{n} \ell_{\tau, D_i}\big(\Lambda(\alpha + \boldsymbol{x}_i^\top \boldsymbol{\beta}) - t_k\big).$$

By Theorems 1 and 3, the ReQR estimator solves a feasible score equation that approximates the population score evaluated at the true parameter values. The approximation involves two layers: (i) a power-law approximation to the conditional tail probability, implemented via the Hill estimator for the extreme value index γ; and (ii) an empirical approximation to the intermediate exceedance probability. Lemma **??** in Appendix E of online supplementary material shows that the estimation uncertainties arising from these two layers are asymptotically *independent*. This result motivates our multiplier double boostatrap procedure for constructing confidence intervals for the ReQR estimator.

Step 1 Draw two mutually independent i.i.d. sequences of nonnegative random weights, denoted $\{\omega_i^{(b)}\}_{i=1}^n$ and $\{\xi_i^{(b)}\}_{i=1}^n$, also independent of the data. The weights $\omega_i^{(b)}$ have mean 1 and variance 1/2, while $\xi_i^{(b)}$ have mean 1 and variance 1.

Step 2 Using the weights $\omega_i^{(b)}$, define the randomly weighted Hill estimator as $\widehat{\gamma}_H^b = \frac{1}{k}\sum_{i=1}^n \omega_i^{(b)} (\log Y_i - \log t_k) \mathbb{1}\{Y_i > t_k\}$, and construct the corresponding loss function $\ell_{\tau,D}^b(v) = \int_0^\infty \{\rho_{\tau,D}(u-v) - \rho_{\tau,D}(u)\}\, dG_{\widehat{\gamma}_H^b, t_k}(u)$.

Step 3 Using the weights $\xi_i^{(b)}$, obtain the bootstrap ReQR estimator

$$(\widehat{\alpha}_\Lambda^b(\tau), \widehat{\boldsymbol{\beta}}_\Lambda^b(\tau)) = \underset{\alpha,\boldsymbol{\beta}}{\operatorname{argmin}} \frac{1}{n}\sum_{i=1}^n \xi_i^{(b)} \ell_{\tau,D_i}^b \left(\Lambda(\alpha + \boldsymbol{x}_i^\top \boldsymbol{\beta}) - t_k\right).$$

Step 4 Repeat Steps 1–3 to generate a pool of bootstrap estimators $\{\widehat{\boldsymbol{\beta}}_\Lambda^b(\tau)\}_{b=1}^B$. Let $\widehat{\boldsymbol{\beta}}_\Lambda^b(\tau) = (\widehat{\beta}_{\Lambda,1}^b(\tau), \ldots, \widehat{\beta}_{\Lambda,d}^b(\tau))^\top$ and $\widehat{\boldsymbol{\beta}}_\Lambda(\tau) = (\widehat{\beta}_{\Lambda,1}(\tau), \ldots, \widehat{\beta}_{\Lambda,d}(\tau))^\top$. Define $\delta_j^b := |\widehat{\beta}_{\Lambda,j}^b(\tau) - \widehat{\beta}_{\Lambda,j}(\tau)|$. Order $\delta_j^1, \ldots, \delta_j^B$ as $\delta_j^{(1)} \le \ldots \le \delta_j^{(B)}$. For a confidence level $a \in (0,1)$, we construct the two-sided bootstrap confidence intervals $\mathrm{CI}_{\widehat{\beta}_{\Lambda,j}}(\tau, a) = \left(\widehat{\beta}_{\Lambda,j}(\tau) - \delta_j^{([Ba])}, \widehat{\beta}_{\Lambda,j}(\tau) + \delta_j^{([Ba])}\right)$. for $j = 1, \ldots, d$.

The weight sequence $\{\omega_i^{(b)}\}_{i=1}^n$ employed in Step 2 is designed to replicate the asymptotic behavior of the Hill estimator $\widehat{\gamma}_H$. In Step 3 we perturb the objective function using an additional sequence of weights $\{\xi_i^{(b)}\}_{i=1}^n$, following the standard multiplier bootstrap approach; see, for example, Jin et al. (2001) and Chatterjee and Bose

(2005). Specifically, we draw $\xi_i^{(b)}$ from a standard exponential distribution and construct $\omega_i^{(b)} = (\widetilde{\xi}_i^{(b)} - 1)/\sqrt{2} + 1$ where $\widetilde{\xi}_i^{(b)}$ is an independent copy of $\xi_i^{(b)}$, ensuring that $\omega_i^{(b)}$ has mean 1 and variance 1/2. We use the ReQR estimator as the initial value when computing the bootstrap ReQR estimator. Our bootstrap procedure approximates the distribution of $\widehat{\boldsymbol{\beta}}_\Lambda(\tau) - \boldsymbol{\beta}_\tau$ by the conditional distribution of $\widehat{\boldsymbol{\beta}}_\Lambda^b(\tau) - \widehat{\boldsymbol{\beta}}_\Lambda(\tau)$ given the sample. The theoretical validity of this inference method is established in the theorems of online supplementary material Appendix C.

4 Simulation

In this section, we evaluate the estimation accuracy of the proposed ReQR estimators at high quantile levels and compare their performance with that of conventional QR estimators. We also examine empirical coverage probabilities and widths of the confidence intervals constructed using the proposed bootstrap procedure, and compare them with those obtained from the subsampling-based inference method for conventional QR developed by Chernozhukov and Fernández-Val (2011). Finally, we assess prediction errors for high conditional quantiles by comparing the ReQR approach with both conventional QR and the extrapolation method proposed by Wang et al. (2012). We generate data from the following three models:

Model 1 (Location-scale shift model):

$$Y_i = \alpha + \boldsymbol{x}_i^\top \boldsymbol{\beta} + (\boldsymbol{x}_i^\top \boldsymbol{\kappa} + 1)\varepsilon_i, \quad i = 1, \ldots, n.$$

The covariate vector is drawn from a bivariate uniform distribution on $[0,1]^2$ with covariance matrix $(0.7^{|m-l|})_{1 \le m,l \le 2}$. The error ε_i is independently generated from a Student $t(3)$ distribution. We set $\alpha = 2$, $\boldsymbol{\beta} = (3,1)^\top$ and $\boldsymbol{\kappa} = (0.5,1)^\top$. For a given quantile level τ, the corresponding QR parameters are $\alpha_\tau = \alpha + Q_{t(3)}(\tau)$ and $\boldsymbol{\beta}_\tau = \boldsymbol{\beta} + \boldsymbol{\kappa} Q_{t(3)}(\tau)$, where $Q_{t(3)}(\tau)$ denotes the τ-quantile of the $t(3)$ distribution.

Model 2 (Location shift model):

$$Y_i = \alpha + \boldsymbol{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \ldots, n.$$

When $\boldsymbol{\kappa} = \mathbf{0}$, Model 1 reduces to Model 2 in which $\alpha_\tau = \alpha + Q_{t(3)}(\tau)$ and $\boldsymbol{\beta}_\tau \equiv \boldsymbol{\beta}$, so that the slope coefficients remain constant across quantile levels. The data are generated under the same setup as in Model 1.

Model 3 (Log-transformed location shift model):

$$Y_i = \exp\left(\alpha + \boldsymbol{x}_i^\top \boldsymbol{\beta} + 0.5\varepsilon_i\right), \quad \varepsilon_i \sim \text{Exp}(1), \quad i = 1, \ldots, n.$$

We set $\alpha = 1$ and $\boldsymbol{\beta} = (1, 1)^\top$, and generate the covariates from the same distribution as in Model 1. The errors follow a standard exponential distribution, implying that the outcome variable Y_i exhibits a power-law upper tail with extreme value index $\gamma = 0.5$. In contrast $\log Y_i$ is light-tailed. The conditional quantile function of Y_i given $\boldsymbol{x}_i$ takes the form $Q_{Y_i}(\tau \mid \boldsymbol{x}_i) = \exp\left(\alpha_\tau + \boldsymbol{x}_i^\top \boldsymbol{\beta}_\tau\right)$, where $\alpha_\tau = \alpha + 0.5Q_{\exp}(\tau)$, $\boldsymbol{\beta}_\tau \equiv \boldsymbol{\beta}$, and $Q_{\exp}(\tau)$ denotes the τ-quantile of the standard exponential distribution. Model 3 is calibrated to match key characteristics of the empirical dataset.

For each model, we generate $S = 3000$ random samples with size $n = 1000$ and 2000, respectively. The quantile level is set to $\tau = 0.99$, corresponding to the *effective* sample size of 10 and 20, respectively. The number of tail observations k is fixed at 100, which determines both the threshold t_k and Hill estimator $\widehat{\gamma}_H$. To guide the choice of k, one may use the log-log plot, as illustrated in the left panel of Figure 4 in Section 5, where an appropriate value of k is associated with an approximately linear pattern. Alternatively, k can be selected from the first stable region of the Hill plot, shown in the right panel of Figure 4, which reports Hill estimates over a range of k values.

To evaluate estimation performance, we compute the standard deviation and root mean squared error (RMSE) of both the ReQR and the conventional QR estimators for each component of the slope vector. Table **??** in Appendix D of online supplementary material reports the average standard deviation and RMSE across 3000 simulated samples. Overall, the bias of both estimators is negligible, while the ReQR estimator consistently achieves substantially smaller RMSE. The improvement is particularly pronounced when $n = 1000$. Figure 3 further illustrates estimation accuracy through boxplots of the slope estimates obtained from the two methods. The estimation precision of the ReQR estimator is approximately doubled relative to the conventional QR estimator which may experience significant variance inflation in the extreme tail. The choice of k governs a bias–variance trade-off. In practice, we recommend selecting a relatively large value of k to reduce variance, while ensuring that the induced bias remains well controlled.

We further evaluate the finite-sample performance of the proposed bootstrap inference procedure. Table 1 reports the average confidence interval widths and empirical coverage rates at the 95% confidence level over 3000 simulated samples, using 500 bootstrap repetitions for each sample. For comparison, we also report the results from the subsampling procedure of Chernozhukov et al. (2010) for conventional QR, with the subsample size set to $n/5$ and the spacing parameter specified as $20/(\tau n) + 1$. The rule of thumb condition $n(1-\tau)/(d+1) \leq 20$ recommended by Chernozhukov et al. (2010) is satisfied. As the sample size increases, the coverage rates of both methods converge to the nominal level of 95%, and the corresponding confidence interval widths decrease. Notably, the confidence intervals constructed using ReQR are approximately one half as wide as those based on conventional QR across all cases, which is consistent with the standard deviation results reported in Table **??**. These findings highlight the validity of the proposed inference procedure and the superior efficiency of the ReQR estimator in the extreme tail.

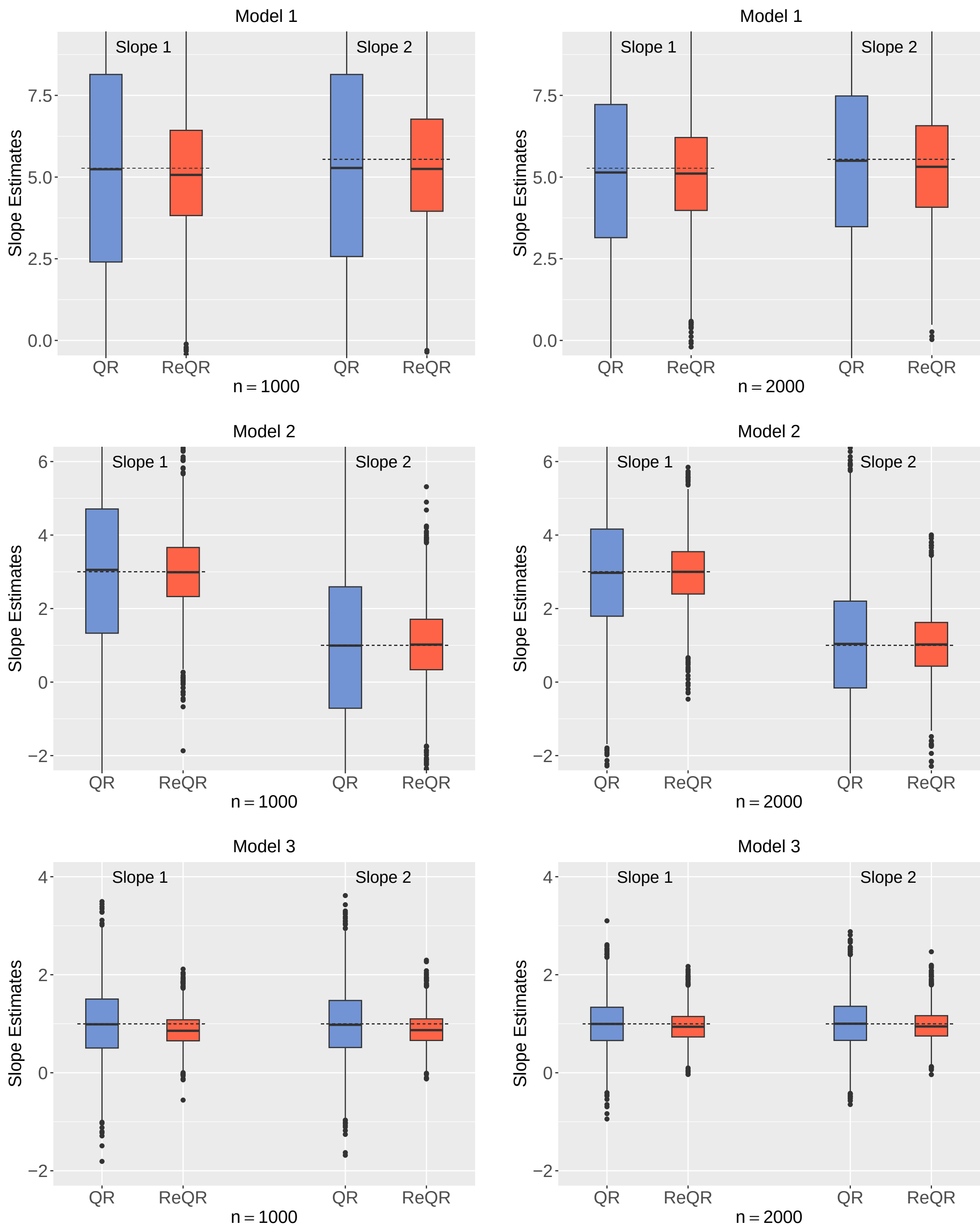
Finally, we investigate prediction accuracy for high conditional quantiles. We obtain the conditional quantile estimates by plugging the ReQR estimators into the QR models (2) and (13). We compare ReQR with the approach of Wang et al. (2012), which extrapolates intermediate regression quantiles via a power law, as described in (1). Table **??** in Appendix D of online supplementary material reports prediction accracy measured by the average absolute percentage error (APE) over 3000 simulated samples for Models 1-3. For a given covariate vector $\boldsymbol{x}_0$, the APE is defined as $1/S\sum_{s=1}^{S}\left|\frac{\widehat{Q}_Y^{(s)}(\tau|\boldsymbol{x}_0)}{Q_Y^{(s)}(\tau|\boldsymbol{x}_0)}-1\right|$. The column labeled "QR" corresponds to the plug-in estimator based on conventional QR. while "ExQR" refers to the extrapolation method of Wang et al. (2012). In general, both ExQR and ReQR outperform conventional QR by achieving smaller APE. The performance of ReQR is comparable to that of ExQR and slightly superior when the sample size becomes large ($n = 2000$). We further consider a combined method, labeled "CQR", which first uses ReQR to estimate an intermediate conditional quantile at a selected level τ_0, and then applies the ExQR approach to extrapolate from the τ_0-conditional quantile to the target τ-conditional quantile. The results for this combined approach are reported in the last column. Overall, CQR attains the lowest APE among all methods when the quantile level is very close to 1 ($\tau = 0.995$).

Table 1: Confidence Interval Widths and Coverage Rates

		Widths				Coverage Rates			
	$\tau = 0.99$	$n = 1000$		$n = 2000$		$n = 1000$		$n = 2000$	
		QR	ReQR	QR	ReQR	QR	ReQR	QR	ReQR
Model 1	$\widehat{\beta}_1(\tau)$	19	**8**	14	**7**	0.93	0.96	0.94	0.96
	$\widehat{\beta}_2(\tau)$	19	**9**	15	**8**	0.93	0.94	0.94	0.95
Model 2	$\widehat{\beta}_1(\tau)$	11	**4**	8	**3**	0.92	0.96	0.93	0.95
	$\widehat{\beta}_2(\tau)$	11	**4**	8	**3**	0.92	0.96	0.94	0.95
Model 3	$\widehat{\beta}_1(\tau)$	3	**1**	3	**1**	0.93	0.92	0.95	0.94
	$\widehat{\beta}_2(\tau)$	3	**1**	3	**1**	0.92	0.92	0.94	0.95

Notes: Widths and coverage rates at the 95% confidence level.

Figure 3: Boxplots of slope estimates from 3000 simulations of Models 1–3



Notes: The horizontal dashed lines represent the true parameter values for each case. The quantile level $\tau = 0.99$. *Boxplots are based on 3000 simulated samples.*

5 Extremely high returns to education

Beginning with the seminal work of Mincer (1974), the estimation of returns to education has been a central and enduring topic in labor economics. The core hypothesis that education constitutes an investment that enhances individual productivity and, consequently, earnings, has been tested and refined across a wide range of economic contexts. Using urban Chinese data, studies such as Fleisher and Wang (2004) and Zhang et al. (2005) find a rising trend in returns to education over the period 1988-2001. Li et al. (2012) further show that both vocational and college education yield high returns in China, comparable to those observed in the United States. While early work primarily focused on estimating average returns, Buchinsky (1994) employed quantile regression to examine heterogeneity in returns to education across the wage distribution.

In this section, we apply the proposed ReQR method to examine the marginal effect of education on upper-tail monthly income. The analysis is based on a survey conducted by the National Bureau of Statistics of China in June and July 2002. The sample comprises 2,412 individuals drawn from five major Chinese cities: Chengdu, Chongqing, Harbin, Hefei, and Wuhan. Average monthly income is CNY 912, with observations ranging from a minimum of CNY 30 to a maximum of CNY 30,000. To assess upper-tail behavior of the income distribution, the left panel of Figure 4 presents a log–log plot of the 100 largest income observations, where sample ranks (in descending order) are plotted against income values on logarithmic scales. The pronounced linear pattern provides strong evidence of heavy-tailed behavior. Guided by this feature, we set $k = 100$, corresponding to the 95th sample percentile (CNY 2,043) as the threshold t_k. The Hill estimator of the extreme value index, $\widehat{\gamma}_H = 0.49$ (right panel of Figure 4), further indicates that the income distribution may lack finite second moments.

The sample has an average of 11.4 years of schooling per individual. Based on educational attainment, we classify individuals into five groups: elementary school (≤ 6 years), middle school ($(6, 9]$ years), high school ($(9, 12]$ years), college ($(12, 16]$ years), and graduate (> 16 years). Figure 5 displays the proportion of high-income individuals in each group, where "high-income" is defined as income exceeding the 95th sample percentile. The proportion of high-income individuals increases monotonically with years of schooling, indicating a positive association between educational attainment and high-income outcomes.

We estimate the log-linear QR model (12) for individual income Y_i. Years of schooling is the primary covariate of interest, and we include employment status, age, age squared, and gender as control variables to account for potential confounding factors. The left panel of Figure 6 displays the conventional QR estimates of the education effect across quantile levels $\tau \in [0.50, 0.99]$. The 95% confidence intervals are constructed by the rank-score method of Koenker and Machado (1999), which is robust to heteroscedasticity in the error terms. The estimated effects are significantly positive for $\tau < 0.98$. To examine the effect of education on the extreme upper-tail income, we apply the subsampling-based inference procedure of Chernozhukov and Fernández-Val (2011), as shown in the right panel. For $\tau \in [0.980, 0.999]$, the QR estimates become statistically insignificant and are accompanied by wide confidence intervals. However, this is inconsistent with the descriptive evidence presented in Figure 5.

As τ increases from 0.980 to 0.999, the effective sample size for the QR estimator declines sharply, from 48 observations to as few as 3. To circumvent the data sparsity in this region, we apply the proposed ReQR approach. In contrast to the QR estimates, the ReQR estimates indicate significantly positive returns to education, as illustrated in Figure 7. Our results reveal that for the monthly income levels between CNY 3,000 and CNY 6,000, an additional year of schooling increases individual income by about 4%–5% when $k = 100$. To assess the robustness of our findings with respect to the choice of k, we vary k from 100 to 200 in increments of 10. Across the selected values of k, the estimated effects exhibit a consistent pattern. Overall, the marginal returns to education are significantly positive across the income distribution, but are more pronounced in the central region than in the upper tail.

6 Conclusion

Leveraging the widely observed power-law behavior of heavy-tailed distributions, this paper proposes an efficient method for estimating QR parameters in the tail region. By approximating the vanishing tail probability in the population score, the proposed method substantially reinforces estimation precision when the effective tail sample size is small. We establish the asymptotic normality of the ReQR estimator over the entire tail region and develop a multiplier double bootstrap procedure to conduct inference on both intermediate and extreme regression quantiles. Extending the ReQR framework to settings with temporal dependence remains an important and challenging topic for future research.

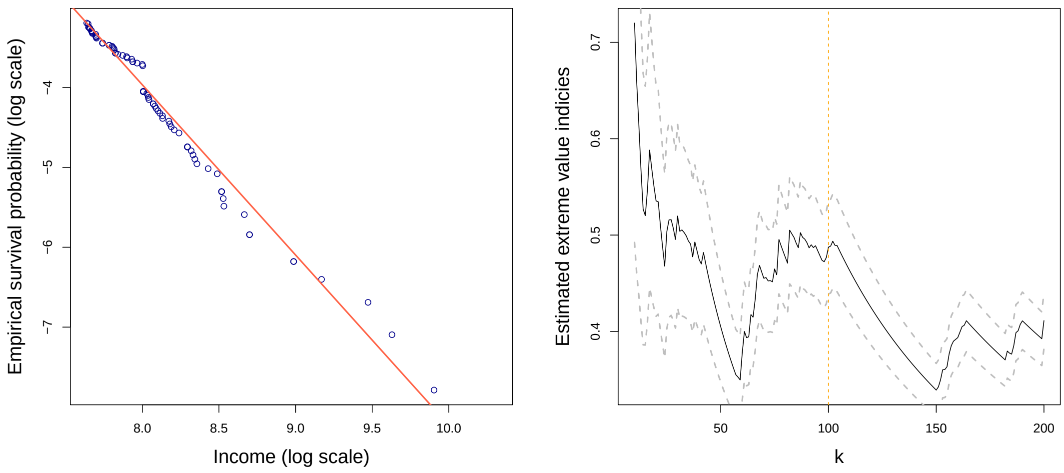
Acknowledgement

Leng's research is supported by National Natural Science Foundation of China (NSFC) (No. 72573136). Hong acknowledges the support from NSFC's Basic Scientific Center Project (No. 71988101).

References

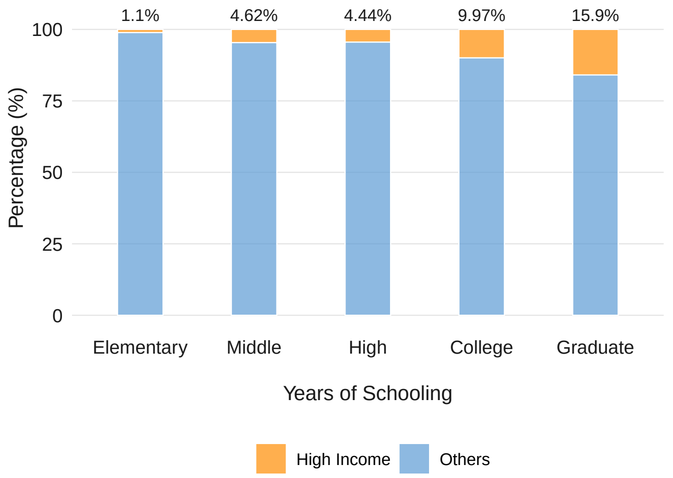
Adrian, T. and M. K. Brunnermeier (2016). CoVaR. *American Economic Review 106*, 1705–41.

Figure 4: Empirical power law of income data and Hill estimator



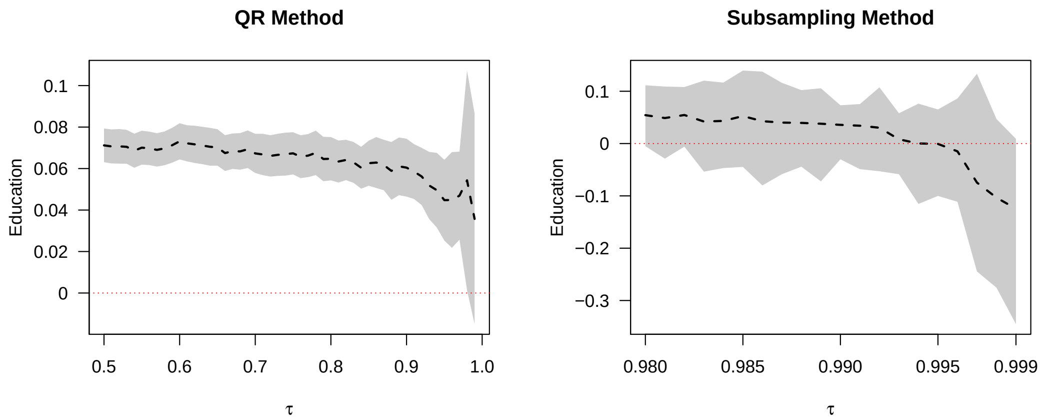
Notes: The left panel presents a log-log plot of the 100 largest income observations, where the sample ranks are plotted against income values. The right panel displays the Hill plot for the extreme value index, with the dashed vertical line indicating the selected value $k = 100$*, corresponding to a Hill estimate of* $\widehat{\gamma}_H = 0.49$.

Figure 5: Proportion of high-income individuals across groups



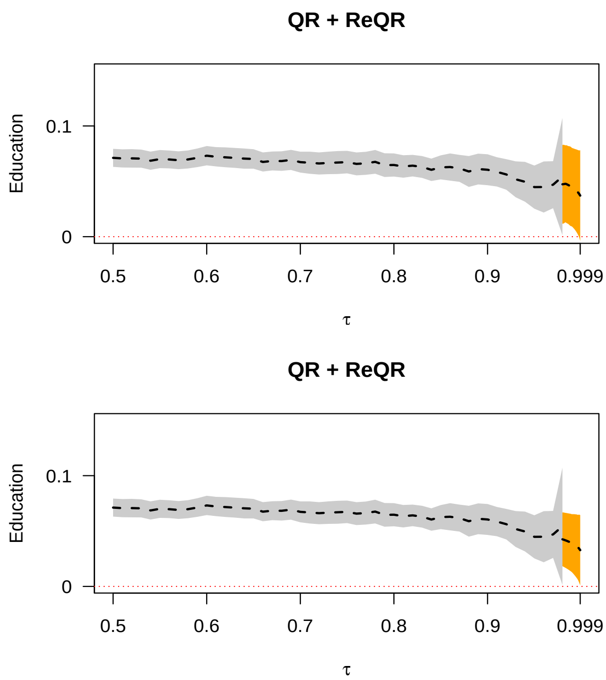
Notes: Among 2412 individuals, 91 with elementary education, 671 with middle school education, 945 with high school education, 592 with college education, and 113 with graduate education. The orange segments of the histograms represent the proportion of high-income individuals.

Figure 6: Effects of education on incomes



Left panel: QR estimates for $\tau \in [0.50, 0.99]$ with 95% confidence intervals constructed using the rank-score method of Koenker and Machado (1999). Right panel: QR estimates for $\tau \in [0.980, 0.999]$ with 95% confidence intervals constructed using the subsampling approach of Chernozhukov and Fernández-Val (2011).

Figure 7: Returns to education



Notes: Estimated education effects with 95% confidence intervals. Results for $\tau \in [0.500, 0.980)$ are obtained using conventional QR and those for $\tau \in [0.980, 0.999]$ are obtained using ReQR. $k = 100$ for the upper panel, and $k = 150$ for the lower panel.

Altonji, J. G., H. Ichimura, and T. Otsu (2012). Estimating derivatives in nonseparable models with limited dependent variables. *Econometrica 80*(4), 1701–1719.

Buchinsky, M. (1994). Changes in the US wage structure 1963-1987: Application of quantile regression. *Econometrica*, 405–458.

Chatterjee, S. and A. Bose (2005). Generalized bootstrap for estimating equations. *The Annals of Statistics 33*(1), 414–436.

Chernozhukov, V. (2005). Extremal quantile regression. *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*(2), 559–589.

Chernozhukov, V., I. Fernández-Val, and T. Kaji (2017). Extremal quantile regression: An overview. *Handbook of Quantile Regression*, 333–362.

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

Daouia, A., G. Stupfler, and A. Usseglio-Carleve (2023). Inference for extremal regression with dependent heavy-tailed data. *The Annals of Statistics 51*(5), 2040–2066.

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

Deuber, D., J. Li, S. Engelke, and M. H. Maathuis (2024). Estimation and inference of extremal quantile treatment effects for heavy-tailed distributions. *Journal of the American Statistical Association 119*(547), 2206–2216.

D'Haultfœuille, X., A. Maurel, and Y. Zhang (2018). Extremal quantile regressions for selection models and the black–white wage gap. *Journal of Econometrics 203*(1), 129–142.

Einmahl, J. H., L. de Haan, and C. Zhou (2016). Statistics of heteroscedastic extremes. *Journal of the Royal Statistical Society Series B 78*(1), 31–51.

Einmahl, J. H. and Y. He (2023). Extreme value inference for heterogeneous power law data. *The Annals of Statistics 51*(3), 1331–1356.

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.

Fernandes, M., E. Guerre, and E. Horta (2021). Smoothing quantile regressions. *Journal of Business & Economic Statistics 39*(1), 338–357.

Fleisher, B. M. and X. Wang (2004). Skill differentials, return to schooling, and market segmentation in a transition economy: the case of Mainland China. *Journal of Development Economics 73*(1), 315–328.

Galvao, A. F., T. Parker, and Z. Xiao (2024). Bootstrap inference for panel data quantile regression. *Journal of Business & Economic Statistics 42*(2), 628–639.

He, X., X. Pan, K. M. Tan, and W.-X. Zhou (2023). Smoothed quantile regression with large-scale inference. *Journal of Econometrics 232*(2), 367–388.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics 3*(5), 1163–1174.

Hong, Y., Y. Liu, and S. Wang (2009). Granger causality in risk and detection of extreme risk spillover between financial markets. *Journal of Econometrics 150*(2), 271–287.

Hou, Y., X. Leng, L. Peng, and Y. Zhou (2024). Panel quantile regression for extreme risk. *Journal of Econometrics 240*(1), 105674.

Jin, Z., Z. Ying, and L. Wei (2001). A simple resampling method by perturbing the minimand. *Biometrika 88*(2), 381–390.

Koenker, R. (2005). *Quantile Regression*. Cambridge, UK: Cambridge University Press.

Koenker, R. and G. Bassett (1978). Regression quantiles. *Econometrica 46*(1), 33–50.

Koenker, R. and J. A. Machado (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association 94*(448), 1296–1310.

Koenker, R. and Z. Xiao (2002). Inference on the quantile regression process. *Econometrica 70*(4), 1583–1612.

Li, H., P. W. Liu, and J. Zhang (2012). Estimating returns to education using twins in urban China. *Journal of Development Economics 97*(2), 494–504.

Mincer, J. (1974). *Schooling, Experience, and Earnings*. Cambridge, MA: National Bureau of Economic Research.

Mu, Y. and X. He (2007). Power transformation toward a linear regression quantile. *Journal of the American Statistical Association 102*(477), 269–279.

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

Resnick, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes*. New York, US: Springer.

Sasaki, Y. and Y. Wang (2025). Extreme quantile treatment effects under endogeneity. *Journal of Business & Economic Statistics*.

Wang, H. J. and D. Li (2013). Estimation of extreme conditional quantiles through power transformation. *Journal of the American Statistical Association 108*, 1062–1074.

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

Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association 73*(364), 812–815.

Xiao, Z. (2014). Right-tail information in financial markets. *Econometric Theory 30*(1), 94–126.

Zhang, J., Y. Zhao, A. Park, and X. Song (2005). Economic returns to schooling in urban China, 1988 to 2001. *Journal of Comparative Economics 33*(4), 730–752.

Zhang, T. (2021). High quantile regression for tail dependent time series. *Biometrika 108*(1), 113–126.

Zhang, Y. (2018). Extremal quantile treatment effects. *The Annals of Statistics 46*(6B), 3707–3740.