Extreme Quantile Estimation Based on the Tail Single-index Model

Wen Xu

Department of Statistics, Fudan University, Shanghai, China

Huixia Judy Wang

Department of Statistics, The George Washington University, Washington DC, the USA

Deyuan Li*

Department of Statistics, Fudan University, Shanghai, China

Supplementary Material

The supplementary file contains some remarks, additional simulation results, and all the technical details.

S1 Computational efficiency

Table 1 summarizes the average computing time (in seconds) of different methods for estimating the conditional quantile for one simulation data. Results show that the proposed method SIMEXQ is computationally more

*Corresponding author: deyuanli@fudan.edu.cn

efficient than the other methods except SIMQ, which is designed for central quantile analysis and does not involve the estimation of extreme value index and extrapolation.

Table 1: The average computing time (in seconds) of different methods for estimating the extreme conditional quantile in one simulation repetition.

Case	BG	ICDF	TDR	SIMQ	SIMEXQ
Case 1, $p = 1$	1.40	531.31	10.24	0.03	0.48
Case 2, $p = 4$	3.72	588.62	42.50	0.03	0.36
Case 3, $p = 4$	3.98	675.98	54.24	0.04	1.48

BG: the estimator proposed by Beirlant and Goegebeur (2004); ICDF: the inverse CDF estimator; TDR: the tail dimension reduction estimator; SIMQ: the single-index model estimator in Zhu et al. (2012) for central quantiles; SIMEXQ: the proposed extreme quantile estimator.

S2 Linear conditional mean assumption

The condition (2.3) is a linear conditional mean (LCM) assumption on the true index parameter $\boldsymbol{\beta}_0$, LCM($\boldsymbol{\beta}_0$). The LCM condition is a common assumption in the dimension reduction literature (Li, 1991; Hall et al., 1993; Zhu et al., 2012; Li, 2018). Let $\boldsymbol{\beta} \in \mathbb{R}^p$, the LCM($\boldsymbol{\beta}$) condition assumes that $E\left(X \mid \boldsymbol{\beta}^\top X\right)$ is a linear function of the random vector $\boldsymbol{\beta}^\top X$, see for instance, Li (2018). Li (1991) and Li (2018) showed that a random vector $\mathbf{X}$ satisfies LCM($\boldsymbol{\beta}$) for all $\boldsymbol{\beta} \in \mathbb{R}^p$ if and only if $\mathbf{X}$ has an elliptical distribution. Hall et al. (1993) proved that LCM($\boldsymbol{\beta}$) at a fixed $\boldsymbol{\beta}$ holds to a good approximation in single-index models when the dimension p diverges. Thus, the LCM assumption is typically regarded as mild, particularly when p is fairly large. The condition (2.3) is weaker than spherical symmetry because it requires that LCM($\boldsymbol{\beta}$) holds only at one fixed $\boldsymbol{\beta}_0$. In applications where the elliptical distribution assumption is seriously violated, we can follow the heuristic approach suggested in Chapter 7 of Li (2018) and transform each component of $\mathbf{X}$ marginally to Gaussian. The validity of this approach is built on a Gaussian copula assumption.

To study the sensitivity of the proposed method against the violation of the linear conditional mean assumption, we consider an additional simulation study: Case 2 in the main paper with $\mathbf{X}$ from a multivariate skew normal distribution. The multivariate skew normal distribution is introduced by Azzalini and Valle (1996). A p-dimensional random variable $\mathbf{Z}$ is said to have a multivariate skew normal distribution if it is continuous with density function

$$2\boldsymbol{\phi}_p(\mathbf{z};\boldsymbol{\Omega})\Phi\left(\boldsymbol{\alpha}^\top \mathbf{z}\right), \quad \mathbf{z} \in \mathbb{R}^p,$$

where $\boldsymbol{\phi}_p(\mathbf{z};\boldsymbol{\Omega})$ is the p-dimensional normal density with zero mean and correlation matrix $\boldsymbol{\Omega}$, $\Phi(\cdot)$ is the $N(0,1)$ distribution function, and $\boldsymbol{\alpha}$ is a p-dimensional vector. When $\boldsymbol{\alpha} = \mathbf{0}$, the density of $\mathbf{Z}$ reduces to $\boldsymbol{\phi}_p(\mathbf{0},\boldsymbol{\Omega})$. We set $p = 4$ and choose shape parameter $\boldsymbol{\alpha} = (2,-6,2,-6)^T$ and $\boldsymbol{\Omega} = \mathbf{I}$. Table 2 shows that even in this case when $\mathbf{X}$ follows an asymmetric distribution, the SIMEXQ method still works the best among all methods considered.

Table 2: The mean integrated squared error (standard errors) for different estimators of the extreme conditional quantiles at $\tau = 0.99$, 0.995 and 0.999 for Case 2 with $p = 4$ and with $\mathbf{X}$ following a multivariate skew normal distribution.

Method	$\tau = 0.99$	$\tau = 0.995$	$\tau = 0.999$
BG	12.24 (0.05)	8.76 (0.08)	6.76 (0.12)
ICDF	5.84 (0.07)	12.37 (0.11)	56.87 (0.17)
TDR	0.16 (0.06)	0.69 (0.09)	0.91 (0.12)
SIMQ	0.15 (0.09)	0.23 (0.15)	0.27 (0.19)
SIMEXQ	0.10 (0.03)	0.12 (0.05)	0.19 (0.07)

BG: the estimator proposed by Beirlant and Goegebeur (2004); ICDF: the inverse CDF estimator; TDR: the tail dimension reduction estimator; SIMQ: the single-index model estimator in Zhu et al. (2012) for central quantiles; SIMEXQ: the proposed extreme quantile estimator.

S3 Sensitivity with different extreme value index estimators

Our proposed method uses the Hill estimator of the extreme value index (EVI). Alternatively, we can also consider other estimators, such as the moment estimator as in Li and Wang (2019), the Pickands estimator as in Daouia et al. (2013), and the peaks over random threshold (PORT) estimator as in Santos et al. (2006). First, Theorem 1 does not include the estimation of the EVI so the result remains the same for different EVI estimators. For Theorem 2, the asymptotic properties of different EVI estimators will be similar in terms of the asymptotic normality but with different representations and variances. For example, see the asymptotic property of the moment-type estimator in Li and Wang (2019). When a different EVI estimator such as the moment-type estimator is used, we expect the asymptotic result in Theorem 3 still holds but with a different form for the term on the right hand side. To examine the numerical stability of our proposed method against different choices of EVI estimators, we include the PORT estimator in the simulation study.

The PORT estimator was introduced in Santos et al. (2006) for univariate data, and it is based on the sample of excesses over a random threshold and is invariant for changes in location and scale. We can adapt the estimator to the regression setup by taking the sample of excesses as $\{Y_i - \hat{Q}_\tau(Y|\mathbf{X})\}_+$, that is, the positive exceedances above the τ-th conditional quantile. If the tail behavior of Y is invariant of $\mathbf{X}$, then we can use this sample of excesses to estimate the common extreme value index; see the similar idea as in Chernozhukov and Du (2006) for linear extreme quantile regression. However, in cases when the EVI varies with $\mathbf{X}$, we cannot directly apply the idea to obtain the PORT estimator of $\gamma(\mathbf{X})$; and how to adapt the PORT estimator in this setting will require further investigation.

In our simulation study, Case 2 has a common extreme value index. Therefore, we include another variant of our proposed estimator, referred to as SIMEXQ$_P$, which replaces the Hill estimator in Step 2 with the PORT estimator, in Case 2 for comparison. For the PORT estimator, we choose the threshold level as $\tau = 0.9$. Tables 3 and 4 summarize the mean integrated squared error (standard errors) for different estimators of the extreme conditional quantiles at $\tau = 0.99, 0.995$ and 0.999 and $\gamma(\mathbf{x})$ for Case 2 with $p = 4$, respectively. The results show that the procedures based on the Hill and PORT estimators perform similarly, suggesting that our proposed method is stable with different extreme value index estimators.

Table 3: The mean integrated squared error (standard errors) for different estimators of the extreme conditional quantiles at $\tau = 0.99, 0.995$ and 0.999 in Case 2 with $p = 4$.

Method	$\tau = 0.99$	$\tau = 0.995$	$\tau = 0.999$
BG	12.42 (0.04)	8.37 (0.07)	5.04 (0.11)
ICDF	6.22 (0.05)	11.23 (0.08)	57.38 (0.12)
TDR	0.18 (0.04)	0.67 (0.07)	0.89 (0.09)
SIMQ	0.06 (0.09)	0.13 (0.12)	0.24 (0.15)
SIMEXQ	0.04 (0.02)	0.05 (0.03)	0.07 (0.07)
SIMEXQ$_P$	0.04 (0.03)	0.05 (0.03)	0.09 (0.07)

BG: the estimator proposed by Beirlant and Goegebeur (2004); ICDF: the inverse CDF estimator; TDR: the tail dimension reduction estimator; SIMQ: the single-index model estimator in Zhu et al. (2012) for central quantiles; SIMEXQ: the proposed extreme quantile estimator based on the Hill EVI estimator; SIMEXQ$_P$: the proposed extreme quantile estimator based on the PORT EVI estimator.

Table 4: The mean integrated squared error (standard errors) of different estimators of $\gamma(\mathbf{x})$ in Case 2 with $p = 4$.

BG	ICDF	TDR	SIMEXQ	SIMEXQ$_P$
0.25 (0.09)	0.17 (0.07)	0.34 (0.12)	0.11 (0.07)	0.10 (0.07)

BG: the estimator proposed by Beirlant and Goegebeur (2004); ICDF: the inverse CDF estimator; TDR: the tail dimension reduction estimator; SIMEXQ: the proposed extreme quantile estimator based on the Hill EVI estimator; SIMEXQ$_P$: the proposed extreme quantile estimator based on the PORT EVI estimator.

S4 Assumption in the rule of thumb for the bandwidth selection

To determine a simple rule of thumb for the bandwidth, we assume that the second derivatives of the quantile function $G''_\tau(z)$ and the conditional mean function $G''(z)$ are similar. Even though $G(z)$ itself might vary considerably with z in terms of curvature, the second derivatives of $G(z)$ and $G_\tau(z)$ may still be similiar in some situations. For example, in the location-shift model $Y_i = G(z_i) + \varepsilon_i$, where $z_i = \mathbf{x}_i^T \boldsymbol{\beta}_0$, and ε_i's are identically distributed errors with mean zero and CDF F_ε, the mean function is $G(z)$, and τ-th quantile function is $G_\tau(z) = G(z) + F_\varepsilon^{-1}(\tau)$; thus in this case $G''(z) = G''_\tau(z)$ for any z. The assumption also holds in heteroscedastic cases if the conditional quantiles of ε do not depend on z or are linear in z. One example is: $Y_i = G(z_i) + (1 + az_i)\varepsilon_i$, where a is a constant satisfying $1 + az_i > 0$ for any z_i in the support, and ε_i are identically distributed errors. The assumption has also been used for bandwidth selection in Yu and Jones (1998), Zhu et al. (2012) and etc.

In the simulation study, Case 2 is a heteroscedastic model with the error term $z\varepsilon$, therefore the assumption of $G''(z) = G''_\tau(z)$ holds for any z except $z = 0$ and for any $0 < \tau < 1$. This assumption does not hold in

Cases 1 and 3. However, our simulation results show that the bandwidth chosen based on the rule of thumb still performs well even when the assumption is violated. On the other hand, directly estimating $G''_\tau(z)$ and $G''(z)$ would involve nonparametric estimation that depends on some other tuning parameters, and thus is difficult to apply in practice. Therefore, we would still recommend using the assumption $G''(z) = G''_\tau(z)$ to simplify the bandwidth selection procedure.

S5 Proof of Proposition 2.1

It suffices to prove that, for any constant $\mathrm{u} \in \mathbb{R}$, there exists a constant κ such that $\mathcal{L}_\tau(\mathrm{u}, \boldsymbol{\beta}) \geq \mathcal{L}_\tau(\mathrm{u}, \kappa\boldsymbol{\beta}_0)$, $\tau \in (\tau_c, 1)$. Specifically,

$$\begin{aligned}\mathcal{L}_\tau(\mathrm{u}, \boldsymbol{\beta}) &= E\left[E\left\{\rho_\tau\left(Y - \mathrm{u} - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)|\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}, Y\right\}\right] \\ &\geq E\left[\rho_\tau\left\{E\left(Y|\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}, Y\right) - \mathrm{u} - E\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}|\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}, Y\right)\right\}\right] \\ &= E\left[\rho_\tau\left\{Y - \mathrm{u} - E\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}|\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}, Y\right)\right\}\right] \\ &= E\left\{\rho_\tau\left(Y - \mathrm{u} - \kappa\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}\right)\right\}.\end{aligned}$$

The first equality follows from the iterative law of conditional expectation; the first inequality follows from Jensen's inequality and the convexity of ρ_τ; the second equality is true with the conditional expectation property, and the last equality holds true by invoking the linearity condition (2.3). This completes the proof. □

S6 Proof of Proposition 2.2

Quadratic Approximation Lemma. Suppose $\mathcal{L}_n(\boldsymbol{\delta})$ is convex and can be represented as $\boldsymbol{\delta}^T\mathbf{B}\boldsymbol{\delta}/2 + u_n^T\boldsymbol{\delta} + \mathbf{a}_n + R_n(\boldsymbol{\delta})$, where $\mathbf{B}$ is symmetric and positive definite, u_n is stochastically bounded, $\mathbf{a}_n$ is arbitrary, and $R_n(\boldsymbol{\delta})$ goes to zero in probability for each $\boldsymbol{\delta}$. Then $\hat{\boldsymbol{\delta}}$, the minimizer of $\mathcal{L}_n(\boldsymbol{\delta})$, is only $o_p(1)$ away from $-\mathbf{B}^{-1}u_n$. If u_n converges in distribution to u, then $\hat{\boldsymbol{\delta}}$ converges in distribution to $-\mathbf{B}^{-1}u$.

We will apply **Quadratic Approximation Lemma**. The main steps are similar to those in the proof of Theorem 2 in Zhu et al. (2012). But we correct some errors.

Let $\alpha_n = n^{-1/2}$, $b = n^{1/2}(\hat{u}_{\tau_0} - u_{\tau_0})$, and $\mathbf{a} = n^{1/2}(\hat{\boldsymbol{\beta}}_{\tau_0} - \boldsymbol{\beta}_{\tau_0})$. Write

$$\mathcal{L}_{\tau n}(\hat{u}_{\tau_0}, \hat{\boldsymbol{\beta}}_{0\tau}) = \mathcal{L}_{\tau n}(u_{\tau_0} + \alpha_n b, \boldsymbol{\beta}_{\tau_0} + \alpha_n \mathbf{a}).$$

We will expand $\mathcal{L}_{\tau n}(\hat{u}_{\tau_0}, \hat{\boldsymbol{\beta}}_{\tau_0})$ around $\mathcal{L}_{\tau n}(u_{\tau_0}, \boldsymbol{\beta}_{\tau_0})$.

By applying the identity

$$\rho_\tau(x-y)-\rho_\tau(x)=y\{I(x\le 0)-\tau\}+\int_0^y\{I(x\le t)-I(x\le 0)\}dt, \quad \text{(S6.1)}$$

it follows that

$$\begin{aligned}
&\mathcal{L}_{\tau n}(u_{\tau_0}+\alpha_n b,\, \boldsymbol{\beta}_{\tau_0}+\alpha_n\mathbf{a})-\mathcal{L}_{\tau n}(u_{\tau_0},\, \boldsymbol{\beta}_{\tau_0})\\
&=\frac{1}{n}\sum_{i=1}^n\rho_\tau(Y_i-u_{\tau_0}-\alpha_n b-\mathbf{X}_i^T\boldsymbol{\beta}_{\tau_0}-\alpha_n\mathbf{X}_i^T\mathbf{a})-\frac{1}{n}\sum_{i=1}^n\rho_\tau(Y_i-u_{\tau_0}-\mathbf{X}_i^T\boldsymbol{\beta}_{\tau_0})\\
&=\frac{1}{n}\sum_{i=1}^n\alpha_n(\mathbf{X}_i^T\mathbf{a}+b)\{I(Y_i-\mathbf{X}_i^T\boldsymbol{\beta}_{\tau_0}\le u_{\tau_0})-\tau_0\}\\
&\quad+\frac{1}{n}\sum_{i=1}^n\int_0^{\alpha_n(\mathbf{X}_i^T\mathbf{a}+b)}\{I(Y_i-\mathbf{X}_i^T\boldsymbol{\beta}_{\tau_0}\le u_{\tau_0}+t)-I(Y_i-\mathbf{X}_i^T\boldsymbol{\beta}_{\tau_0}\le u_{\tau_0})\}dt.
\end{aligned}$$

Let $\epsilon=Y-\mathbf{X}^T\boldsymbol{\beta}_{\tau_0}$, and denote $F_\epsilon(t|\mathbf{X})$ and $f_\epsilon(\cdot|\mathbf{X})$ as the conditional CDF and conditional density function of ϵ given $\mathbf{X}$, respectively. By Taylor expansion, it follows that

$$\begin{aligned}
&E\{\mathcal{L}_{\tau n}(u_{\tau_0}+\alpha_n b,\, \boldsymbol{\beta}_{\tau_0}+\alpha_n\mathbf{a})-\mathcal{L}_{\tau n}(u_{\tau_0},\, \boldsymbol{\beta}_{\tau_0})|\mathbf{X}_1,\ldots,\mathbf{X}_n)\}\\
&=\frac{1}{n}\sum_{i=1}^n\left[\alpha_n(\mathbf{X}_i^T\mathbf{a}+b)\{F_\epsilon(u_{\tau_0}|\mathbf{X}_i)-\tau_0\}+\int_{u_{\tau_0}}^{u_{\tau_0}+\alpha_n(\mathbf{X}_i^T\mathbf{a}+b)}\{F_\epsilon(t|\mathbf{X}_i)-F_\epsilon(0|\mathbf{X}_i)\}dt\right]\\
&=\frac{\alpha_n}{n}\sum_{i=1}^n\left[(\mathbf{X}_i^T\mathbf{a}+b)\{F_\epsilon(u_{\tau_0}|\mathbf{X}_i)-\tau_0\}+\frac{\alpha_n}{2}f_\epsilon(u_{\tau_0}|\mathbf{X}_i)(\mathbf{X}_i^T\mathbf{a}+b)^2\right]+o_p(\alpha_n^2).
\end{aligned}$$

Following standard arguments, we obtain that

$$
\begin{aligned}
R_n = &\mathcal{L}_{\tau n}(u_{\tau_0} + \alpha_n b,\, \boldsymbol{\beta}_{\tau_0} + \alpha_n \mathbf{a}) - \mathcal{L}_{\tau n}(u_{\tau_0},\, \boldsymbol{\beta}_{\tau_0}) \\
&- E\{\mathcal{L}_{\tau n}(u_{\tau_0} + \alpha_n b,\, \boldsymbol{\beta}_{\tau_0} + \alpha_n \mathbf{a}) - \mathcal{L}_{\tau n}(u_{\tau_0},\, \boldsymbol{\beta}_{\tau_0}) | \mathbf{X}_1, ..., \mathbf{X}_n\} \\
= &\, o_p(n^{-1}).
\end{aligned}
$$

So,

$$
\begin{aligned}
&\mathcal{L}_{\tau n}(u_{\tau_0} + \alpha_n b,\, \boldsymbol{\beta}_{\tau_0} + \alpha_n \mathbf{a}) \\
&= \mathcal{L}_{\tau n}(u_{\tau_0},\, \boldsymbol{\beta}_{\tau_0}) + E\{\mathcal{L}_{\tau n}(u_{\tau_0} + \alpha_n b,\, \boldsymbol{\beta}_{\tau_0} + \alpha_n \mathbf{a}) - \mathcal{L}_{\tau n}(u_{\tau_0},\, \boldsymbol{\beta}_{\tau_0}) | \mathbf{X}_1, ..., \mathbf{X}_n\} + R_n \\
&= \mathcal{L}_{\tau n}(u_{\tau_0},\, \boldsymbol{\beta}_{\tau_0}) + \frac{\alpha_n}{n} \sum_{i=1}^{n} \Big[(\mathbf{X}_i^T \mathbf{a} + b)\{F_\epsilon(u_{\tau_0}|\mathbf{X}_i) - \tau_0\} + \frac{\alpha_n}{2} f_\epsilon(u_{\tau_0}|\mathbf{X}_i)(\mathbf{X}_i^T \mathbf{a} + b)^2 \Big] \\
&\quad + o_p(n^{-1}) + R_n \\
&= \mathcal{L}_{\tau n}(u_{\tau_0},\, \boldsymbol{\beta}_{\tau_0}) + \frac{\alpha_n}{n} \sum_{i=1}^{n} \begin{pmatrix} 1 \\ \mathbf{X}_i \end{pmatrix}^T \{F_\epsilon(u_{\tau_0}|\mathbf{X}_i) - \tau_0\} \begin{pmatrix} b \\ \mathbf{a} \end{pmatrix} \\
&\quad + \begin{pmatrix} b \\ \mathbf{a} \end{pmatrix}^T \frac{\alpha_n^2}{2n} \begin{pmatrix} \sum_{i=1}^n f_\epsilon(u_{\tau_0}|\mathbf{X}_i) & \sum_{i=1}^n \mathbf{X}_i^T f_\epsilon(u_{\tau_0}|\mathbf{X}_i) \\ \sum_{i=1}^n \mathbf{X}_i f_\epsilon(u_{\tau_0}|\mathbf{X}_i) & \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^T f_\epsilon(u_{\tau_0}|\mathbf{X}_i) \end{pmatrix} \begin{pmatrix} b \\ \mathbf{a} \end{pmatrix} + o_p(n^{-1}).
\end{aligned}
$$

Applying **Quadratic Approximation Lemma**, we obtain that

$$
\begin{pmatrix} b \\ \mathbf{a} \end{pmatrix} = -\tilde{\mathbf{B}}_n^{-1} \frac{\alpha_n}{n} \sum_{i=1}^{n} \begin{pmatrix} 1 \\ \mathbf{X}_i \end{pmatrix} \big(F_\epsilon(u_{\tau_0}|\mathbf{X}_i) - \tau_0 \big),
$$

where

$$\tilde{\mathbf{B}}_n = \alpha_n^2 \begin{pmatrix} n^{-1}\sum_{i=1}^n f_\epsilon(u_{\tau_0}|\mathbf{X}_i) & n^{-1}\sum_{i=1}^n \mathbf{X}_i^T f_\epsilon(u_{\tau_0}|\mathbf{X}_i) \\ n^{-1}\sum_{i=1}^n \mathbf{X}_i f_\epsilon(u_{\tau_0}|\mathbf{X}_i) & n^{-1}\sum_{i=1}^n \mathbf{X}_i\mathbf{X}_i^T f_\epsilon(u_{\tau_0}|\mathbf{X}_i) \end{pmatrix} =: \alpha_n^2 \mathbf{B}_n.$$

Hence

$$\begin{pmatrix} n^{1/2}(\hat{u}_{\tau_0} - u_{\tau_0}) \\ n^{1/2}(\hat{\boldsymbol{\beta}}_{\tau_0} - \boldsymbol{\beta}_{\tau_0}) \end{pmatrix} = -\alpha_n^{-1}\mathbf{B}_n^{-1} \begin{pmatrix} n^{-1}\sum_{i=1}^n (F_\epsilon(u_{\tau_0}|\mathbf{X}_i) - \tau_0) \\ n^{-1}\sum_{i=1}^n \mathbf{X}_i(F_\epsilon(u_{\tau_0}|\mathbf{X}_i) - \tau_0) \end{pmatrix} + o_p(1) \tag{S6.2}$$

and by the CLT,

$$\mathbf{B}_n \xrightarrow{p} \begin{pmatrix} E(f_\epsilon(u_{\tau_0}|\mathbf{X})) & E(\mathbf{X}^T f_\epsilon(u_{\tau_0}|\mathbf{X})) \\ E(\mathbf{X} f_\epsilon(u_{\tau_0}|\mathbf{X})) & E(\mathbf{X}\mathbf{X}^T f_\epsilon(u_{\tau_0}|\mathbf{X})) \end{pmatrix} = \mathbf{B}.$$

In order to study the limitation of (S6.2), we first study the mean and variance of

$$\begin{pmatrix} F_\epsilon(u_{\tau_0}|\mathbf{X}) - \tau_0 \\ X(F_\epsilon(u_{\tau_0}|\mathbf{X}) - \tau_0) \end{pmatrix}.$$

In fact, $(u_{\tau_0}, \boldsymbol{\beta}_{\tau_0})$ solves $E[\{\tau_0 - I(Y \le u + \mathbf{X}^T\boldsymbol{\beta})\}(1, \mathbf{X}^T)^T] = \mathbf{0}$. Then it follows that

$$E[I(Y \le u_{\tau_0} + \mathbf{X}^T\boldsymbol{\beta}_{\tau_0}) - \tau_0] = 0, \quad E[\mathbf{X}(I(Y \le u_{\tau_0} + \mathbf{X}^T\boldsymbol{\beta}_{\tau_0}) - \tau_0)] = \mathbf{0}.$$

Hence, we have $E[E\{I(Y \leq u_{\tau_0} + \mathbf{X}^T\boldsymbol{\beta}_{\tau_0})|\mathbf{X}\} - \tau_0] = 0$,

$$E\big(\mathbf{X}[E\{I(Y \leq u_{\tau_0} + \mathbf{X}^T\boldsymbol{\beta}_{\tau_0})|\mathbf{X}\} - \tau_0]\big) = \mathbf{0},$$

which implies that

$$E\begin{pmatrix} F_\epsilon(u_{\tau_0}|\mathbf{X}) - \tau_0 \\ \mathbf{X}(F_\epsilon(u_{\tau_0}|\mathbf{X}) - \tau_0) \end{pmatrix} = \begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix}.$$

Let

$$\mathbf{V} = Var\begin{pmatrix} F_\epsilon(u_{\tau_0}|\mathbf{X}) - \tau_0 \\ \mathbf{X}(F_\epsilon(u_{\tau_0}|\mathbf{X}) - \tau_0) \end{pmatrix}.$$

Then, by (S6.2) and central limit theory, we obtain that

$$\begin{pmatrix} n^{1/2}(\hat{u}_{\tau_0} - u_{\tau_0}) \\ n^{1/2}(\hat{\boldsymbol{\beta}}_{\tau_0} - \boldsymbol{\beta}_{\tau_0}) \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix}, \ \boldsymbol{\Sigma}\right)$$

with $\boldsymbol{\Sigma} = \mathbf{B}^{-1}\mathbf{V}\mathbf{B}^{-1}$. □

S7 Proof of Theorem 1

Recall the notations $z = \mathbf{X}_0^T\boldsymbol{\beta}_{\tau_0}, \hat{z} = \mathbf{X}_0^T\hat{\boldsymbol{\beta}}_{\tau_0}, Z_i = \mathbf{X}_i^T\boldsymbol{\beta}_{\tau_0}, \hat{Z}_i = \mathbf{X}_i^T\hat{\boldsymbol{\beta}}_{\tau_0}$. Let $a_0 = G_\tau(z) = G_\tau(\mathbf{X}_0^T\boldsymbol{\beta}_{\tau_0})$, $b_0 = G'_\tau(z)$, $\hat{K}_i = K\{(\hat{Z}_i - \hat{z})/h\}$, $K_i =$

$K\{(Z_i - z)/h\}$ and $K'((Z_i - z)/h)$ for $i = 1, 2, ..., n$.

We first show that, for $\hat{\boldsymbol{\beta}}_{\tau_0}$ which is a $\sqrt{n}$-consistent estimator of $\boldsymbol{\beta}_{\tau_0}$,

$$\frac{1}{n}\sum_{i=1}^{n}\left[\rho_\tau\{Y_i - a - b(\hat{Z}_i - \hat{z})\}\hat{K}_i - \rho_\tau\{Y_i - a - b(Z_i - z)\}K_i\right] = O_p(n^{-1/2}). \tag{S7.1}$$

Define

$$R_{11} = \frac{1}{n}\sum_{i=1}^{n}K_i\left[\rho_\tau\{Y_i - a - b(\hat{Z}_i - \hat{z})\} - \rho_\tau\{Y_i - a - b(Z_i - z)\}\right],$$

$$R_{12} = \frac{1}{n}\sum_{i=1}^{n}(\hat{K}_i - K_i)\rho_\tau\{Y_i - a - b(Z_i - z)\},$$

$$R_{13} = \frac{1}{n}\sum_{i=1}^{n}\left[\rho_\tau\{Y_i - a - b(\hat{Z}_i - \hat{z})\} - \rho_\tau\{Y_i - a - b(Z_i - z)\}\right](\hat{K}_i - K_i).$$

Then,

$$\frac{1}{n}\sum_{i=1}^{n}\left[\rho_\tau\{Y_i - a - b(\hat{Z}_i - \hat{z})\}\hat{K}_i - \rho_\tau\{Y_i - a - b(Z_i - z)\}K_i\right] = R_{11}+R_{12}+R_{13}.$$

Using the identity (S6.1), we have $|\rho_\tau(x-y) - \rho_\tau(x)| \le 2|y|$. Invoking the $\sqrt{n}$-consistency of $\hat{\boldsymbol{\beta}}_{\tau_0}$, we have

$$
\begin{aligned}
|R_{11}| &\le \frac{1}{n}\sum_{i=1}^{n} K_i \left|\rho_\tau\{Y_i - a - b(\hat{Z}_i - \hat{z})\} - \rho_\tau\{Y_i - a - b(Z_i - z)\}\right| \\
&\le 2|b| \sup_u |K(u)| \frac{1}{n}\sum_{i=1}^{n}\{|\hat{Z}_i - Z_i| + |\hat{z} - z|\} \\
&= 2|b| \sup_u |K(u)| \frac{1}{n}\sum_{i=1}^{n}\{|\mathbf{X}_i^T(\hat{\boldsymbol{\beta}}_{\tau_0} - \boldsymbol{\beta}_{\tau_0})| + |\mathbf{X}_0(\hat{\boldsymbol{\beta}}_{\tau_0} - \boldsymbol{\beta}_{\tau_0})|\} = O_p(n^{-1/2}).
\end{aligned}
$$

Next we deal with R_{12}. By Taylor expansion, we have

$$
R_{12} = (\hat{\boldsymbol{\beta}}_{\tau_0} - \boldsymbol{\beta}_{\tau_0})^T \frac{1}{nh}\sum_{i=1}^{n} K'\Big(\frac{Z_i - z}{h}\Big)(\mathbf{X}_i - \mathbf{X}_0)\rho_\tau\{Y_i - a - b(Z_i - z)\}\{1 + o_p(1)\}.
$$

Let $\mathbf{Z}_{i0} = (\mathbf{X}_i - \mathbf{X}_0)/h$. Then

$$
\begin{aligned}
&\frac{1}{nh}\sum_{i=1}^{n} K'\Big(\frac{Z_i - z}{h}\Big)(\mathbf{X}_i - \mathbf{X}_0)\rho_\tau\{Y_i - a - b(Z_i - z)\} \\
&= \frac{1}{n}\sum_{i=1}^{n} K'\big(\mathbf{Z}_{i0}^T\boldsymbol{\beta}_{\tau_0}\big)\mathbf{Z}_{i0}\rho_\tau(Y_i - a - bh\mathbf{Z}_{i0}^T\boldsymbol{\beta}_{\tau_0}) \\
&\le \frac{1}{n}\sum_{i=1}^{n} K'\big(\mathbf{Z}_{i0}^T\boldsymbol{\beta}_{\tau_0}\big)\mathbf{Z}_{i0}|Y_i - a - bh\mathbf{Z}_{i0}^T\boldsymbol{\beta}_{\tau_0}| \\
&= O_p(1).
\end{aligned}
$$

Together with the $\sqrt{n}$-consistency of $\hat{\boldsymbol{\beta}}_{\tau_0}$, we prove that $R_{12} = O_p(n^{-1/2})$.

It remains to investigate the order of R_{13}. Following similar arguments, we have $R_{13} = O_p(n^{-1/2})$. Thus, (S7.3) follows and implies

$$\frac{1}{n}\sum_{i=1}^{n}\rho_\tau\{Y_i - a - b(\hat{Z}_i - \hat{z})\}\hat{K}_i = \frac{1}{n}\sum_{i=1}^{n}\rho_\tau\{Y_i - a - b(Z_i - z)\}K_i + o_p\{(nh)^{-1/2}\}.$$

This means, the quantile regression estimate of $G_\tau(\cdot)$ based on $\{(\mathbf{X}_i^T\hat{\boldsymbol{\beta}}_{\tau_0}, Y_i) : i = 1, 2, ..., n\}$ is asymptotically as efficient as that based on $\{(\mathbf{X}_i^T\boldsymbol{\beta}_{\tau_0}, Y_i) : i = 1, 2, ..., n\}$.

The rest of the proof follows literally from Lemma 3 of Wang et al. (2012). Based on the auxiliary data points $\{(\mathbf{X}_i^T\boldsymbol{\beta}_0, Y_i), i = 1, 2, ..., n\}$, the technique for establishing the asymptotic normality involves the convexity arguments for argmin process (Kato, 2009). We only sketch the outline below.

Recall that

$$F\{U(t; z)|z\} = 1 - \frac{1}{t}.$$

By taking the derivation with respect to t, we have

$$F'\{U(t; z)\}U'(t; z) = \frac{1}{t^2}.$$

By Condition C_6, it follows that $U(t; z) = G_{1-1/t}(z)$ and $\frac{tU'(t;z)}{U(t;z)} = \gamma(z)$

uniformly for $z \in \mathcal{Z}$, and hence

$$f_Y(G_\tau(z)|z) = \frac{1-\tau}{\gamma(z)G_\tau(z)}\{1+o(1)\}. \tag{S7.2}$$

Let $\boldsymbol{\theta}(\tau) = \frac{\sqrt{n(1-\tau)h}}{\gamma(z)G_\tau(z)}\big\{a-a_0,\ h(b-b_0)\big\}^T$,

$$\begin{aligned}
&f_n(\boldsymbol{\theta},\tau)\\
&= \frac{1}{\gamma(z)G_\tau(z)}\Big[\sum_{i=1}^n \rho_\tau\Big\{Y_i - a_0 - b_0(Z_i - z) - \frac{G_\tau(z)\gamma(z)}{\sqrt{n(1-\tau)h}}\begin{pmatrix}1\\(Z_i-z)/h\end{pmatrix}^T\boldsymbol{\theta}\Big\}K_i\\
&\qquad - \sum_{i=1}^n \rho_\tau\big\{Y_i - a_0 - b_0(Z_i - z)\big\}K_i\Big]
\end{aligned}$$

and

$$g_n(\boldsymbol{\theta},\tau) = -\boldsymbol{\theta}^T\tilde{W}_n(\tau) + \frac{1}{2}\boldsymbol{\theta}^T Q(\tau)\boldsymbol{\theta},$$

where

$$\begin{aligned}
\tilde{W}_n(\tau) =& \frac{1}{\sqrt{n(1-\tau)h}}\sum_{i=1}^n \begin{pmatrix}1\\(Z_i-z)/h\end{pmatrix}[\tau - I\{Y_i \le G_\tau(Z_i)\}]K_i\\
&+ \frac{\sqrt{nh(1-\tau)}}{2G_\tau(z)\gamma(z)}\begin{pmatrix}\mu_2\\ \mu_3\end{pmatrix} h^2 G''(z) f_Z(z),
\end{aligned}$$

with $\mu_2 = \int_{-1}^{1} u^2 K(u)du$ and $\mu_3 = \int_{-1}^{1} u^3 K(u)du$ and $Q(\tau) = f_Z(z)\begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix}$.

Following similar arguments as in Chernozhukov (2005), we have

$$\sup_{\tau\in\mathcal{T}} |f_n(\boldsymbol{\theta},\tau) - g_n(\boldsymbol{\theta},\tau)| \xrightarrow{p} 0,$$

By Theorem 2.2 of Shorack (1979), it follows that $\tilde{W}_n(\tau)$ converges to a Gaussian process. Then, by applying Theorem 2 in Kato (2009), we have

$$\boldsymbol{\theta}(\tau) = \{Q(\tau)\}^{-1}\tilde{W}_n(\tau) + o_p(1),$$

which means

$$\begin{aligned}
&\frac{\sqrt{n(1-\tau)h}}{\gamma(z)G_\tau(z)}(a - a_0,\, h(b-b_0))^T \\
&\quad = \frac{1}{f_Z(z)}\begin{pmatrix} 1 & 0 \\ 0 & \mu_2^{-1} \end{pmatrix}\left\{\frac{1}{\sqrt{n(1-\tau)h}}\sum_{i=1}^{n}\begin{pmatrix} 1 \\ (Z_i - z)/h \end{pmatrix}[\tau - I\{Y_i \le G_\tau(Z_i)\}]K_i \right. \\
&\qquad\qquad \left. + \frac{\sqrt{nh(1-\tau)}}{2G_\tau(z)\gamma(z)}\begin{pmatrix} \mu_2 \\ \mu_3 \end{pmatrix} h^2 G''(z) f_Z(z)\right\}.
\end{aligned}$$

Since we are more interested in the estimation of $a_0 = G_\tau(z)$, taking the first row in above, we obtain

$$\frac{\{nh(1-\tau)\}^{1/2}}{\gamma(z)G_\tau(z)}\{\hat{G}_\tau(\hat{z})-G_\tau(z)-\frac{1}{2}h^2G''_\tau(z)\mu_2\}=W_n(\tau)\{1+o_p(1)\}, \tag{S7.3}$$

where $\mu_2 = \int_{-1}^{1} u^2K(u)du$, $f_Y(y|z)$ is the conditional density function of $Y|Z=z$,

$$W_n(\tau)=\frac{1}{\sqrt{nh(1-\tau)}}f_Z^{-1}(z)\sum_{i=1}^{n}[\tau-I\{Y_i\leq G_\tau(Z_i)\}]K_i, \quad \tau\in\mathcal{T},$$

which converges to a Gaussian process with mean zero and covariance

$$\mathbf{\Sigma}(\tau_t,\tau_s)=\frac{\nu_0\{\min(\tau_t,\tau_s)-\tau_t\tau_s\}}{\sqrt{(1-\tau_t)(1-\tau_s)}}f_Z^{-1}(z),$$

where $\nu_0=\int_{-1}^{1}K^2(u)du$ and $f_Z(z)$ is the density function of Z. □

S8 Proof of Theorem 2

From Theorem 1, we have that

$$\frac{\{nh(1-\tau_j)\}^{1/2}}{\gamma(z)G_{\tau_j}(z)}\left\{\hat{G}_{\tau_j}(\hat{z})-G_{\tau_j}(z)-\frac{1}{2}h^2G''_{\tau_j}(z)\mu_2\right\}=W_n(\tau_j)\{1+o_p(1)\} \tag{S8.1}$$

uniformly for $\tau_j \in \mathcal{T}$ where $\tau_j = 1 - j/n$ and $j = \lceil n^{\eta} \rceil, ..., k$, and that

$$W_n(\tau) = \{nh(1-\tau)\}^{-1/2} f_Z^{-1}(z) \Big\{ \frac{1}{nh} \sum_{i=1}^{n} [\tau - I\{Y_i \le G_\tau(Z_i)\}] K_i \Big\}, \quad \tau \in (0,1),$$

converges to a Gaussian process with mean zero and covariance

$$\mathbf{\Sigma}(\tau_t, \tau_s) = \frac{\nu_0 \min(\tau_t, \tau_s) - \tau_t \tau_s\}}{f_Z(z)\sqrt{(1-\tau_t)(1-\tau_s)}}.$$

Based on (S8.6), we have

$$\hat{G}_{\tau_j}(\hat{z}) = G_{\tau_j}(z) + \frac{1}{2} h^2 G''_{\tau_j}(z)\mu_2 + \{nh(1-\tau_j)\}^{-1/2} \gamma(z) G_{\tau_j}(z) W_n(\tau_j)\{1 + o_p(1)\}.$$

Hence, by Taylor expansion, it follows that

$$\begin{aligned}
&\log \frac{\hat{G}_{\tau_j}(\hat{z})}{\hat{G}_{\tau_k}(\hat{z})} \\
&= \log \frac{G_{\tau_j}(z)[1 + \frac{1}{2}h^2 G_{\tau_j}^{-1}(z) G''_{\tau_j}(z)\mu_2 + \{nh(1-\tau_j)\}^{-1/2}\gamma(z) W_n(\tau_j)(1 + o_p(1))]}{G_{\tau_k}(z)[1 + \frac{1}{2}h^2 G_{\tau_k}^{-1}(z) G''_{\tau_k}(z)\mu_2 + \{nh(1-\tau_k)\}^{-1/2}\gamma(z) W_n(\tau_k)\{1 + o_p(1)\}]} \\
&= \log \frac{G_{\tau_j}(z)}{G_{\tau_k}(z)} \\
&\quad + \frac{1}{2}h^2 G_{\tau_j}^{-1}(z) G''_{\tau_j}(z)\mu_2\{1 + o(1)\} + \{nh(1-\tau_j)\}^{-1/2}\gamma(z) W_n(\tau_j)\{1 + o_p(1)\} \\
&\quad - \frac{1}{2}h^2 G_{\tau_k}^{-1}(z) G''_{\tau_k}(z)\mu_2\{1 + o(1)\} - \{nh(1-\tau_k)\}^{-1/2}\gamma(z) W_n(\tau_k)\{1 + o_p(1)\}.
\end{aligned}$$

By (3.1), we have

$$\frac{U(tx;z)}{U(t;z)} = x^{\gamma(z)}\Big[1 + A(t;z)\frac{x^{\varrho(z)}-1}{\varrho(z)}\{1+o(1)\}\Big].$$

So, with $\tau_j = 1 - j/n$,

$$\begin{aligned}\frac{G_{\tau_j}(z)}{G_{\tau_k}(z)} &= \frac{U(\frac{1}{1-\tau_j})}{U(\frac{1}{1-\tau_k})} = \frac{U(\frac{n}{j})}{U(\frac{n}{k})} = \frac{U(\frac{n}{k}\cdot\frac{k}{j})}{U(\frac{n}{k})} \\ &= (k/j)^{\gamma(z)}[1 + A(n/k;z)\frac{(k/j)^{\varrho(z)}-1}{\varrho(z)}\{1+o(1)\}],\end{aligned}$$

and hence

$$\log\frac{G_{\tau_j}(z)}{G_{\tau_k}(z)} = -\gamma(z)\log(j/k) + A(n/k;z)\frac{(k/j)^{\varrho(z)}-1}{\varrho(z)}\{1+o(1)\}.$$

Now we have

$$\begin{aligned}\log\frac{\hat{G}_{\tau_j}(\hat{z})}{\hat{G}_{\tau_k}(\hat{z})} &= -\gamma(z)\log(j/k) \\ &\quad + A(n/k;z)\frac{(k/j)^{\varrho(z)}-1}{\varrho(z)}\{1+o(1)\} \\ &\quad + \frac{1}{2}h^2\mu_2\{G_{\tau_j}^{-1}(z)G''_{\tau_j}(z) - G_{\tau_k}^{-1}(z)G''_{\tau_k}(z)\}\{1+o(1)\} \\ &\quad + (nh)^{-1/2}\gamma(z)\big\{(1-\tau_j)^{-1/2}W_n(\tau_j) - (1-\tau_k)^{-1/2}W_n(\tau_k)\big\}\{1+o_p(1)\}.\end{aligned}$$

Note that

$$\frac{1}{k}\sum_{j=\lceil n^{\eta}\rceil}^{k}\log(j/k)=\int_0^1\log x dx+o(\max\{\lceil n^{\eta}\rceil,1/k\})$$
$$=-1+o(\max\{\lceil n^{\eta}\rceil,1/k\})=-1+o(k^{-1/2})$$

and that

$$\frac{1}{k}\sum_{j=\lceil n^{\eta}\rceil}^{k}\frac{(k/j)^{\varrho(z)}-1}{\varrho(z)}=\int_0^1\frac{x^{-\varrho(z)}-1}{\varrho(z)}dx+o(\max\{\lceil n^{\eta}\rceil,1/k\})=\frac{1}{1-\varrho(z)}+o(k^{-1/2}).$$

Recall that $G_\tau(z)=U\{1/(1-\tau);z\}$ and $G''_\tau(z)=\frac{\partial^2 G_\tau(z)}{\partial z^2}=\frac{\partial^2 U\{1/(1-\tau);z\}}{\partial z^2}$. In order to approximate $G^{-1}_{\tau_j}(z)G''_{\tau_j}(z)$, we have to approximate $\frac{\partial^2 U\{1/(1-\tau);z\}}{\partial z^2}$.

Note that, for $U(t;z)$ satisfying the second order condition (3.4) with $\varrho(z)<0$, we can write it as

$$U(t;z)=c(z)t^{\gamma(z)}+d(z)t^{\gamma(z)+\varrho(z)}\{1+o(1)\}, \tag{S8.2}$$

for some $c(z)>0$ and $d(z)\in\mathbb{R}$. Assume $c(z),d(z),\gamma(z)$ and $\varrho(z)$ are continuous functions with second derivative functions. Then

$$\frac{\partial U(t;z)}{\partial z}=c'(z)t^{\gamma(z)}+c(z)t^{\gamma(z)}(\log t)\gamma'(z)$$
$$+\left(d'(z)t^{\gamma(z)+\varrho(z)}+d(z)t^{\gamma(z)+\varrho(z)}(\log t)(\gamma'(z)+\varrho'(z))\right)\{1+o(1)\}.$$

Now we distinguish three cases: (1) $\gamma'(z)\neq 0$; (2) $\gamma'(z)=0$ but $\varrho'(z)\neq 0$;

(3) $\gamma^{'}(z) = 0$ and $\varrho^{'}(z) = 0$.

For the first case $\gamma^{'}(z) \neq 0$, we can easily have

$$\frac{\partial^2 U(t;z)}{\partial z^2} \sim c(z) t^{\gamma(z)} (\log t)^2 \{\gamma^{'}(z)\}^2$$

and that, with $\tau_j = 1 - j/n$ for $j = \lceil n^\eta \rceil, ..., k$,

$$G_{\tau_j}^{-1}(z) G_{\tau_j}^{''}(z) \sim \{\log(n/j)\}^2 \{\gamma^{'}(z)\}^2.$$

Hence

$$\begin{aligned} G_{\tau_j}^{-1}(z) G_{\tau_j}^{''}(z) - G_{\tau_k}^{-1}(z) G_{\tau_j}^{''}(z) &\sim \{\gamma^{'}(z)\}^2 \{(\log(n/j))^2 - \{\log(n/k)\}^2\} \\ &= \{\gamma^{'}(z)\}^2 \{2\log(n/k) + \log(k/j)\} \log(k/j) \end{aligned}$$

and thus

$$\begin{aligned} \frac{1}{k} \sum_{j=\lceil n^\eta \rceil}^{k} \{G_{\tau_j}^{-1}(z) G_{\tau_j}^{''}(z) - G_{\tau_k}^{-1}(z) G_{\tau_k}^{''}(z)\} &\sim 2\log(n/k) \{\gamma^{'}(z)\}^2 \int_0^1 -\log x dx \\ &= 2\log(n/k) \{\gamma^{'}(z)\}^2. \end{aligned}$$

For the second case $\gamma^{'}(z) = 0$ but $\varrho^{'}(z) \neq 0$, we denote $\gamma(z) = \gamma > 0$. Then we have

$$\frac{\partial^2 U(t;z)}{\partial z^2} = c^{''}(z) t^\gamma + d(z) t^{\gamma + \varrho(z)} (\log t)^2 \{\varrho^{'}(z)\}^2 \{1 + o(1)\}$$

and that

$$G_{\tau_j}^{-1}(z)G_{\tau_j}^{''}(z) = \frac{c^{''}(z)}{c(z)} + \frac{d(z)}{c(z)}(n/j)^{\varrho(z)}\{\log(n/j)\}^2\{\varrho^{'}(z)\}^2\{1+o(1)\}.$$

Hence

$$\begin{aligned}
&G_{\tau_j}^{-1}(z)G_{\tau_j}^{''}(z) - G_{\tau_k}^{-1}(z)G_{\tau_k}^{''}(z)\\
&\sim \frac{d(z)\{\varrho^{'}(z)\}^2}{c(z)}\{(n/j)^{\varrho(z)}\{\log(n/j)\}^2 - (n/k)^{\varrho(z)}\{\log(n/k)\}^2\}\\
&\sim \frac{d(z)\{\varrho^{'}(z)\}^2}{c(z)}\Big[(\frac{n}{k})^{\varrho(z)}(\frac{k}{j})^{\varrho(z)}\Big\{\log(n/k)+\log(k/j)\Big\}^2 - (\frac{n}{k})^{\varrho(z)}\{\log(n/k)\}^2\Big]\\
&\sim \frac{d(z)\{\varrho^{'}(z)\}^2}{c(z)}(\frac{n}{k})^{\varrho(z)}\{\log(n/k)\}^2\Big\{(\frac{k}{j})^{\varrho(z)} - 1\Big\}
\end{aligned}$$

and thus

$$\begin{aligned}
&\frac{1}{k}\sum_{j=\lceil n^\eta\rceil}^{k}\{G_{\tau_j}^{-1}(z)G_{\tau_j}^{''}(z) - G_{\tau_k}^{-1}(z)G_{\tau_k}^{''}(z)\}\\
&\sim \frac{d(z)\{\varrho^{'}(z)\}^2}{c(z)}(\frac{n}{k})^{\varrho(z)}\{\log(n/k)\}^2\int_0^1(x^{-\varrho(z)}-1)dx\\
&= \frac{d(z)\{\varrho^{'}(z)\}^2}{c(z)}(\frac{n}{k})^{\varrho(z)}\{\log(n/k)\}^2\frac{\varrho(z)}{1-\varrho(z)}.
\end{aligned}$$

For the third case $\gamma^{'}(z) = 0$ and $\varrho^{'}(z) = 0$, we denote $\gamma(z) = \gamma > 0$ and $\varrho(z) = \varrho < 0$. Without loss of generality, we assume $c^{''}(z) \neq 0$ or $d^{''}(z) \neq 0$. Similarly, we have

$$\frac{\partial^2 U(t;z)}{\partial z^2} = c^{''}(z)t^\gamma + d^{''}(z)t^{\gamma+\varrho}\{1+o(1)\}$$

and that

$$G_{\tau_j}^{-1}(z)G_{\tau_j}''(z) = \frac{c''(z)}{c(z)} - \Big\{\frac{c''(z)d(z)}{c^2(z)} - \frac{d''(z)}{c(z)}\Big\}\big(\frac{n}{j}\big)^{\varrho}\{1+o(1)\}.$$

Hence

$$G_{\tau_j}^{-1}(z)G_{\tau_j}''(z) - G_{\tau_k}^{-1}(z)G_{\tau_k}''(z) \sim -\Big\{\frac{c''(z)d(z)}{c^2(z)} - \frac{d''(z)}{c(z)}\Big\}\Big[\big(\frac{n}{j}\big)^{\varrho} - \big(\frac{n}{k}\big)^{\varrho}\Big]$$

and thus

$$\begin{aligned}
&\frac{1}{k}\sum_{j=\lceil n^\eta\rceil}^{k}\{G_{\tau_j}^{-1}(z)G_{\tau_j}''(z) - G_{\tau_k}^{-1}(z)G_{\tau_k}''(z)\}\\
&\sim -\Big\{\frac{c''(z)d(z)}{c^2(z)} - \frac{d''(z)}{c(z)}\Big\}\big(\frac{n}{k}\big)^{\varrho}\int_0^1 (x^{-\varrho}-1)dx\\
&= -\Big\{\frac{c''(z)d(z)}{c^2(z)} - \frac{d''(z)}{c(z)}\Big\}\big(\frac{n}{k}\big)^{\varrho}\frac{\varrho}{1-\varrho}.
\end{aligned}$$

Consider the last part: approximation to $(1-\tau_j)^{-1/2}W_n(\tau_j)$. Note that, by Theorem 1, $\sqrt{1-\tau}\{f_Z(z)/\nu_0\}^{1/2}W_n(\tau)$, $\tau \in (0,1)$ converges to a Gaussian process with mean zero and covariance function $\min(\tau_t,\tau_s)-\tau_s\tau_t$, for $\tau_t,\tau_s \in (0,1)$. Thus, there exists a sequence of Brownian bridges $\{B_n(t) : t \in (0,1)\}_{n\geq 1}$ such that

$$\sqrt{1-\tau}\sqrt{\frac{f_Z(z)}{\nu_0}}W_n(\tau) = B_n(\tau)\{1+o_p(1)\}, \quad \tau \in (0,1)$$

and hence

$$(1-\tau)^{-1/2}W_n(\tau) = (1-\tau)^{-1}\sqrt{\frac{\nu_0}{f_Z(z)}}B_n(\tau)\{1+o_p(1)\}.$$

On the other hand, there exist a sequence Brownian motions $\{\tilde{W}_n(t) : t \in (0,1)\}_{n\geq 1}$ such that

$$B_n(t) = \tilde{W}_n(t) - t\tilde{W}_n(1).$$

By the fact $B_n(t) \overset{d}{=} B_n(1-t)$, it follows that

$$B_n(\tau_j) = B_n\big(1-\frac{k}{n}\frac{j}{k}\big) \overset{d}{=} B_n\big(\frac{k}{n}\frac{j}{k}\big) = \tilde{W}_n\big(\frac{k}{n}\frac{j}{k}\big)\{1+o_p(1)\} \overset{d}{=} \big(\frac{k}{n}\big)^{1/2}\tilde{W}_n\big(\frac{j}{k}\big)\{1+o_p(1)\}.$$

Then

$$\begin{aligned}
&(1-\tau_j)^{-1/2}W_n(\tau_j) - (1-\tau_k)^{-1/2}W_n(\tau_k) \\
\overset{d}{=}& \sqrt{\frac{\nu_0}{f_Z(z)}}\Big\{(1-\tau_j)^{-1}B_n(\tau_j) - (1-\tau_k)^{-1}B_n(\tau_k)\Big\}\{1+o_p(1)\} \\
\overset{d}{=}& \sqrt{\frac{n}{k}}\sqrt{\frac{\nu_0}{f_Z(z)}}\Big\{\big(\frac{j}{k}\big)^{-1}\tilde{W}_n\big(\frac{j}{k}\big) - \tilde{W}_n(1)\Big\}\{1+o_p(1)\}.
\end{aligned}$$

Thus

$$\frac{1}{k}\sum_{j=\lceil n^\eta\rceil}^{k}\Big\{(1-\tau_i)^{-1/2}W_n(\tau_j)-(1-\tau_k)^{-1/2}W_n(\tau_k)\Big\}$$
$$\stackrel{d}{=}\sqrt{\frac{n}{k}}\sqrt{\frac{\nu_0}{f_Z(z)}}\int_0^1\{x^{-1}\tilde{W}_n(x)-\tilde{W}_n(1)\}dx\,\{1+o_p(1)\}.$$

Without loss of generality, we replace "$\stackrel{d}{=}$" by "$=$" above. Therefore,

$$\begin{aligned}
\hat{\gamma}(z) &= \frac{1}{k}\sum_{i=\lceil n^\eta\rceil}^{k}\log\frac{\hat{G}_{\tau_j}(\hat{z})}{\hat{G}_{\tau_k}(\hat{z})}\\
&= \frac{1}{k}\sum_{i=\lceil n^\eta\rceil}^{k}\{-\gamma(z)\log(j/k)\}\\
&\quad + \frac{1}{k}\sum_{i=\lceil n^\eta\rceil}^{k}A(n/k;z)\frac{(k/j)^{\varrho(z)}-1}{\varrho(z)}\{1+o(1)\}\\
&\quad + \frac{1}{k}\sum_{i=\lceil n^\eta\rceil}^{k}\frac{1}{2}h^2\mu_2\{G_{\tau_j}^{-1}(z)G''_{\tau_j}(z)-G_{\tau_k}^{-1}(z)G''_{\tau_k}(z)\}\{1+o(1)\}\\
&\quad + \frac{1}{k}\sum_{i=\lceil n^\eta\rceil}^{k}(nh)^{-1/2}\gamma(z)\Big\{(1-\tau_j)^{-1/2}W_n(\tau_j)-(1-\tau_k)^{-1/2}W_n(\tau_k)\Big\}\{1+o_p(1)\}\\
&=: I_{1n}(z)+I_{2n}(z)+I_{3n}(z)+I_{4n}(z).
\end{aligned}$$

Obviously, $I_{1n}(z)=\gamma(z)\{1+o(k^{-1/2})\}$,

$$(I_{2n}(z)=A(n/k;z)\{\frac{1}{1-\varrho(z)}+o(k^{-1/2})\}\{1+o(1)\},$$

$$I_{3n}(z) = \begin{cases} h^2\mu_2 \log(n/k)(\gamma'(z))^2\{1+o(k^{-1/2})\}\{1+o(1)\} \\ \qquad\qquad \text{for } \gamma'(z) \neq 0, \\ h^2\mu_2 \dfrac{d(z)\{\varrho^{'}(z)\}^2}{2c(z)}(\dfrac{n}{k})^{\varrho(z)}(\log(n/k))^2\dfrac{\varrho(z)}{1-\varrho(z)}\{1+o(k^{-1/2})\}\{1+o(1)\} \\ \qquad\qquad \text{for } \gamma'(z) = 0,\ \varrho'(z) \neq 0, \\ -\dfrac{1}{2}h^2\mu_2\Big\{\dfrac{c^{''}(z)d(z)}{c^2(z)} - \dfrac{d^{''}(z)}{c(z)}\Big\}\big(\dfrac{n}{k}\big)^{\varrho}\dfrac{\varrho}{1-\varrho}\{1+o(k^{-1/2})\}\{1+o(1)\} \\ \qquad\qquad \text{for } \gamma'(z) = \varrho'(z) = 0, \end{cases}$$

and

$$I_{4n}(z) = \gamma(z)(kh)^{-1/2}\sqrt{\frac{\nu_0}{f_Z(z)}}\int_0^1 \{x^{-1}\tilde{W}_n(x) - \tilde{W}_n(1)\}dx\{1+o_p(1)\}.$$

By taking $I_{2n}(z)$ and $I_{3n}(z)$ as the bias, it follows that

$$\begin{aligned}&(kh)^{1/2}\Big\{\hat{\gamma}(z) - \gamma(z) - I_{2n}(z) - I_{3n}(z)\Big\} \\ &= \gamma(z)\sqrt{\frac{\nu_0}{f_Z(z)}}\int_0^1 \{x^{-1}\tilde{W}_n(x) - \tilde{W}_n(1)\}dx\{1+o_p(1)\}.\end{aligned}$$

By the assumption $(kh)^{1/2}A(n/k;z) \to \lambda_2 \in \mathbb{R}$ and $(kh)^{1/2}h^2\log(n/k) \to \lambda_1 \in \mathbb{R}$, it follows that

$$\begin{aligned}&(kh)^{1/2}\Big\{\hat{\gamma}(z) - \gamma(z) - \frac{A(n/k;z)}{1-\varrho(z)} - \tilde{I}_{3n}(z)\Big\} \\ &= \gamma(z)\sqrt{\frac{\nu_0}{f_Z(z)}}\int_0^1 \{x^{-1}\tilde{W}_n(x) - \tilde{W}_n(1)\}dx\{1+o_p(1)\},\end{aligned}$$

where

$$\tilde{I}_{3n}(z) = \begin{cases} h^2\mu_2 \log(n/k)(\gamma'(z))^2 & , \quad \gamma'(z) \neq 0, \\ \frac{1}{2}h^2\mu_2 \frac{d(z)\{\varrho'(z)\}^2}{c(z)}(\frac{n}{k})^{\varrho(z)}\{\log(n/k)\}^2 \frac{\varrho(z)}{1-\varrho(z)} & , \quad \gamma'(z) = 0, \ \varrho'(z) \neq 0, \\ -\frac{1}{2}h^2\mu_2\Big\{\frac{c''(z)d(z)}{c^2(z)} - \frac{d''(z)}{c(z)}\Big\}(\frac{n}{k})^{\varrho}\frac{\varrho}{1-\varrho} & , \quad \gamma'(z) = \varrho'(z) = 0. \end{cases} \tag{S8.3}$$

□

S9 Proof of Theorem 3

Note that, with notation $\tau_k = 1 - k/n$ and $p_n = 1 - \tau^*$,

$$\frac{\hat{G}_{\tau^*}(\hat{z})}{G_{\tau^*}(z)} = \Big(\frac{1-\tau_k}{1-\tau^*}\Big)^{\hat{\gamma}(\hat{z})} \frac{\hat{G}_{\tau_k}(\hat{z})}{G_{\tau_k}(z)} \Big\{\frac{G_{\tau^*}(z)}{G_{\tau_k}(z)}\Big\}^{-1} = \Big(\frac{k}{np_n}\Big)^{\hat{\gamma}(\hat{z})} \frac{\hat{G}_{\tau_k}(\hat{z})}{G_{\tau_k}(z)} \Big\{\frac{U(\frac{1}{1-\tau^*})}{U(\frac{1}{1-\tau_k})}\Big\}^{-1}.$$

We will approximate the item $(k/(np_n))^{\hat{\gamma}(\hat{z})}$ by applying Theorem 2, approximate the item $\frac{\hat{G}_{\tau_k}(\hat{z})}{G_{\tau_k}(z)}$ by applying Theorem 1, and approximate the item $\frac{U(1/(1-\tau^*))}{U(1/(1-\tau_k))}$ by applying the second order condition (3.4).

By Theorem 1, it follows that

$$\frac{\hat{G}_{\tau_k}(\hat{z})}{G_{\tau_k}(z)} = 1 + \frac{\frac{1}{2}h^2 G''_{\tau_k}(z)\mu_2 + (nh)^{-1/2} f_Y^{-1}(G_{\tau_k}(z)|z) W_n(\tau_k)\{1 + o_p(1)\}}{G_{\tau_k}(z)}.$$

By the second order condition (3.4), it follows that

$$\begin{aligned}
&\Big\{\frac{U(\frac{1}{1-\tau^*})}{U(\frac{1}{1-\tau_k})}\Big\}^{-1} \\
&=\Big\{\Big(\frac{1-\tau^*}{1-\tau_k}\Big)^{\gamma(z)}\Big[1+A\big(\frac{1}{1-\tau_k};z\big)\frac{(\frac{1-\tau_k}{1-\tau^*})^{\varrho(z)}-1}{\varrho(z)}+o\big\{A(\frac{1}{1-\tau_k};z)\big\}\Big]\Big\}^{-1}.
\end{aligned}$$

Then

$$\frac{\hat{G}_{\tau^*}(\hat{z})}{G_{\tau^*}(z)}=\Big(\frac{k}{np_n}\Big)^{\hat{\gamma}(\hat{z})-\gamma(z)}\times\frac{\hat{G}_{\tau_k}(\hat{z})}{G_{\tau_k}(z)}\times\Big[1-A(\frac{n}{k};z)\frac{(\frac{k}{np_n})^{\varrho(z)}-1}{\varrho(z)}+o\{A(\frac{n}{k};z)\}\Big].$$

Then, under the assumption $(kh)^{-1/2}\log(k/(np_n))\to 0$, we have $(\hat{\gamma}(\hat{z})-\gamma(z))\log\{k/(np_n)\}\to 0$ and

$$\begin{aligned}
(\frac{k}{np_n})^{\hat{\gamma}(\hat{z})-\gamma(z)}&=e^{\{\hat{\gamma}(\hat{z})-\gamma(z)\}\log\{k/(np_n)\}}\\
&=1+\{\hat{\gamma}(\hat{z})-\gamma(z)\}\log\{k/(np_n)\}\{1+o_p(1)\}.
\end{aligned}$$

Then, we write

$$\begin{aligned}\frac{\hat{G}_{\tau^*}(\hat{z})}{G_{\tau^*}(z)} &= \big[1 + \{\hat{\gamma}(\hat{z}) - \gamma(z)\}\log(k/np_n)\big] \\ &\quad \times \big[1 + \frac{\frac{1}{2}h^2 G''_{\tau_k}(z)\mu_2 + (nh)^{-1/2} f_Y^{-1}\{G_{\tau_k}(z)|z\}W_n(\tau_k)\{1+o_p(1)\}}{G_{\tau_k}(z)}\big] \\ &\quad \times [1 - A(\frac{n}{k};z)\frac{(\frac{k}{np_n})^{\varrho(z)} - 1}{\varrho(z)} + o\{A(\frac{n}{k};z)\}] \\ &= 1 + \Big[\{\hat{\gamma}(\hat{z}) - \gamma(z)\}\log(k/np_n) \\ &\qquad + \frac{1}{2}h^2 G^{-1}_{\tau_k}(z) G''_{\tau_k}(z)\mu_2 - A(\frac{n}{k};z)\frac{(\frac{k}{np_n})^{\varrho(z)} - 1}{\varrho(z)} \\ &\qquad + (nh)^{-1/2} f_Y^{-1}\{G_{\tau_k}(z)|z\} G^{-1}_{\tau_k}(z) W_n(\tau_k)\Big]\{1 + o_p(1)\}.\end{aligned}$$

Let $N_n(z) = \gamma(z)\sqrt{\frac{\nu_0}{f_Z(z)}}\int_0^1\{x^{-1}\tilde{W}_n(x) - \tilde{W}_n(1)\}dx$. Then by Theorem 2,

$$\begin{aligned}&(\hat{\gamma}(\hat{z}) - \gamma(z))\log(k/np_n) \\ &= (kh)^{-1/2}\log\{k/(np_n)\}\big\{N_n(z) + I_{3n}(z) + \frac{A(n/k;z)}{1-\varrho(z)}\big\}\{1 + o_p(1)\}.\end{aligned}$$

Recall that $(nh)^{-1/2} f_Y^{-1}\{G_{\tau_k}(z)|z\} G^{-1}_{\tau_k}(z) W_n(\tau_k) = O_p((kh)^{-1/2})$ (see the proof of Theorem 2). By our assumption $k/(np_n) \to \infty$ and $(kh)^{-1/2}\log\{k/(np_n)\} \to$

0, we have

$$\frac{\hat{G}_{\tau^*}(\hat{z})}{G_{\tau^*}(z)} = 1 + \Big[(kh)^{-1/2}\log\{k/(np_n)\}N_n(z) + \frac{1}{2}h^2 G_{\tau_k}^{-1}(z)G''_{\tau_k}(z)\mu_2 - A(\frac{n}{k};z)\frac{(\frac{k}{np_n})^{\varrho(z)}-1}{\varrho(z)}\Big] \times \{1+o_p(1)\}.$$

Thus

$$\frac{(kh)^{1/2}}{\log\{k/(np_n)\}}\Big\{\frac{\hat{G}_{\tau^*}(z)}{G_{\tau^*}(z)} - 1 - \frac{1}{2}h^2 G_{\tau_k}^{-1}(z)G''_{\tau_k}(z)\mu_2 + A(\frac{n}{k};z)\frac{(\frac{k}{np_n})^{\varrho(z)}-1}{\varrho(z)}\Big\} = N_n(z)\{1+o_p(1)\},$$

which completes the proof. □

Bibliography

Azzalini, A. and Valle, A, D. (1996). The multivariate skew-normal distribution *Biometrika* **83**, 715 − 726.

Beirlant, J. and Goegebeur, Y. (2004). Local polynomial maximum likelihood estimation for Pareto-type distributions. *J. Multiariate Statist.* **89**, 97 − 118.

Chernozhukov, V. (2005). Extremal quantile regression. *Ann. Statist.* **33**, 806 − 839.

Chernozhukov, V. and Du, S. (2006). Extremal quantiles and value-at-risk *https://papers.ssrn.com/sol3/papers.cfm?abstract_id=956433.*

Daouia, A., Gardes, L., Girard, S. and Lekina, A. (2011). Kernel estimators of extreme level curves. *Test* **20**, 311 − 333.

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

Hall, P. and Li, K. (1993). On almost linearity of low dimensional projections from high dimensional data. *Ann. Statist.* **93**, 867 − 889.

Kato, K. (2009). Asymptotics for argmin processes: Convexity arguments. *J. Multiariate Statist.* **100**, 1816 − 1829.

Santos, P. A., Alves, M. I. and Gomes, M. I. (2006). Peaks over random threshold methodology for tail index and high quantile estimation. *REV-STAT* **4**, 227 − 247.

Li, D. and Wang, H. (2019). Extreme quantile estimation for autoregressive models. *J. Business Econom. Statist.* **37**, 661 − 670.

Li, K. (1991). Sliced inverse regression for dimension reduction. *J. Amer. Statist. Assoc.* **86**, 316 − 327.

Li, B. (2018). Sufficient dimension reduction: Methods and applications with R. *CRC Press.*

Shorack, G. R. (1979). The weighted empirical process of row independent random variables with arbitrary distribution functions. *Statist. Neerlandica* **33**, 169 − 189.

Wang, H., Li, D. and He, X. (2012). Estimation of high conditional quantiles for heavy-tailed distributions. *J. Amer. Statist. Assoc.* **107**, 1453 − 1464.

Yu, K. and Jones, M. C. (1998) Local linear quantile regression. *J. Amer. Statist. Assoc.* **93**, 228 − 237.

Zhu, L., Huang, M. and Li, R. (2012). Semiparametric quantile regression with high-dimensional covariates. *Statist. Sinica* **22**, 1379 − 1424.