MAXIMUM PENALIZED LIKELIHOOD ESTIMATION FOR THE ENDPOINT AND EXPONENT OF A DISTRIBUTION

Fang Wang[1], Liang Peng[2], Yongcheng Qi[3] and Meiping Xu[4]

[1] *Capital Normal University,* [2] *Georgia State University,*

[3] *University of Minnesota Duluth, and*

[4] *Beijing Technology and Business University*

Supplementary Material

We present some further comparisons with the endpoint estimator proposed in Fraga Alves and Neves (2014) and with the moment estimator for the tail index proposed by Dekkers, Einmahl and de Haan (1989) in Section S1. All proofs of the theorems in Section 2 are given in Section S2.

S1 Further comparisons on estimators for endpoint and tail index

Per the request of an anonymous referee, we carry out the following two comparison studies: (A) comparison between our new estimator for the endpoint and the endpoint estimator proposed in Fraga Alves and Neves (2014) and (B) comparison between our new estimator for the tail index and the moment estimator in Dekkers, Einmahl and de Haan (1989). Throughout the referred equation and theorem numbers without a letter are those in the original paper.

The endpoint estimator in Fraga Alves and Neves (2014) is defined as

$$\hat{\theta}_{FAN}(2k-1) = X_{n,n} + \sum_{i=0}^{k-1} a_{ki}(X_{n,n-k} - X_{n,n-k-i}), \tag{S1.1}$$

where $a_{ki} = (\log 2)^{-1}(\log(k+i+1) - \log(k+i))$ for $0 \le i \le k-1$. We will call it FAN estimator. This estimator was originally proposed to estimate the endpoint for distributions in the Gumbel max-domain of attraction. Fraga Alves, Neves and Rosário (2017) have extended the setting to (1.1).

The moment estimator for the tail index $\gamma = -1/\alpha$ proposed by Dekkers, Einmahl and de Haan (1989) is given by

$$\hat{\gamma}_M(k) = M_{n,k}^{(1)} + 1 - \frac{1}{2}\big(1 - \frac{(M_{n,k}^{(1)})^2}{M_{n,k}^{(2)}}\big)^{-1}, \tag{S1.2}$$

where $M_{n,k}^{(j)} = \frac{1}{k}\sum_{i=1}^{k}(\log(X_{n,n-k+i}) - \log(X_{n,n-k}))^j$ for $j = 1, 2$. A natural requirement for the moment estimator $\hat{\gamma}_M(k)$ is that all the data involved in the estimation must be positive, which implies that the endpoint θ must be positive. Otherwise, one can add a positive constant to all observations to fulfill this requirement.

For empirical comparison, we will use the same setting as in Section 3.2, that is, we use both distributions defined in (3.26) and(3.27), choose the sample size $n = 500$, and repeat the experiment 1000 times. We calculate the averages and estimate the mean absolute errors (L_1 errors) of the two aforementioned estimators. The simulation results for distribution (3.27) are somewhat similar to those for distribution (3.26), and so we will report simulation results for distribution (3.26) only.

In Figures 1 and 2, we plot the averages of the estimates and their L_1 errors for the endpoint based on our new penalized likelihood method (New Estimator) and Fraga Alves and Neves's (2014) method (FAN Estimator) against the sample fraction k. We note that the FAN Estimator $\hat{\theta}_{FAN}(2k-1)$ in (S1.1) employs $2k$ upper order statistics while the New Estimator $\tilde{\theta}_N(k) = \tilde{\theta}$ given in Theorem 3 is based on $k+1$ upper order statistics. To make a fair comparison, two types of estimators are compared when the same number of observations are involved in the estimation. More precisely, we will compare $\hat{\theta}_{FAN}(k)$ and $\tilde{\theta}_N(k)$ for $k = 2p-1$, $p = 3, 4, \cdots, 102$.

We have repeated our simulation study for distribution (3.26) by selecting various values for (τ_1, τ_2). We choose $(\tau_1, \tau_2) = (0.5, 1.0)$, $(1.0, 0.5)$, $(0.5, 2.0)$, $(1.0, 2.0)$, $(0.5, 3.0)$, $(1.0, 3.0)$. For distribution (3.26), $\theta = 0$ and $\alpha = \tau_1\tau_2$. Therefore, our study covers cases of $\alpha = 0.5, 1, 1.5, 2$ and 3.

In Figures 3 and 4, we plot the averages of the estimates and their L_1 errors for the index $1/\alpha$ based on our new penalized likelihood method (New Estimator) and the moment estimator (Moment Estimator) against the sample fraction k. Since the moment estimator $\hat{\gamma}_M(k)$ defined in (S1.2) is used to estimate $\gamma = -1/\alpha$, we actually plot the estimated means and L_1 errors for $\tilde{\alpha}_N^{-1}$ given in (2.16) and $-\hat{\gamma}_M(k)$. Since the moment estimator can only be applied to positive observations, all our samples in the study are drawn from the population $20 + X$, where X is a random variable having distribution (3.26). The values of (τ_1, τ_2) selected in this study are the same as in the simulation for the endpoint. The sample fraction k is taken from 5 to 200 with an increment 5.

In conclusion, we observe from Figures 1 and 2 that the maximum penalized likelihood estimator for endpoint is very stable against the sample fraction in terms of the bias and the mean absolute error, and the FAN estimator can perform better when the upper order statistics employed in the estimation are relatively dense near the endpoint. Also we observe from Figures 3 and 4 that the maximum penalized likelihood estimator is superior to the moment estimator.

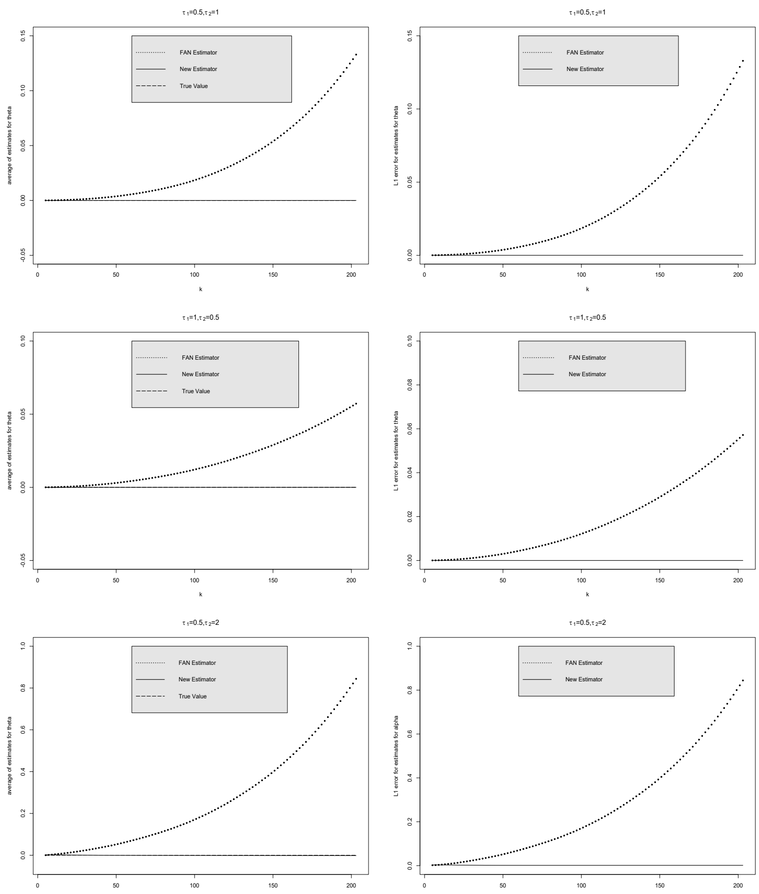

Figure 1: Estimated means (left) and estimated L_1 errors (right) for two endpoint estimators: New Estimator as the smallest solution to (2.17) and FAN Estimator defined in (S1.1). The samples are taken from distribution (3.26), where $\theta = 0$ and $\alpha = \tau_1\tau_2$

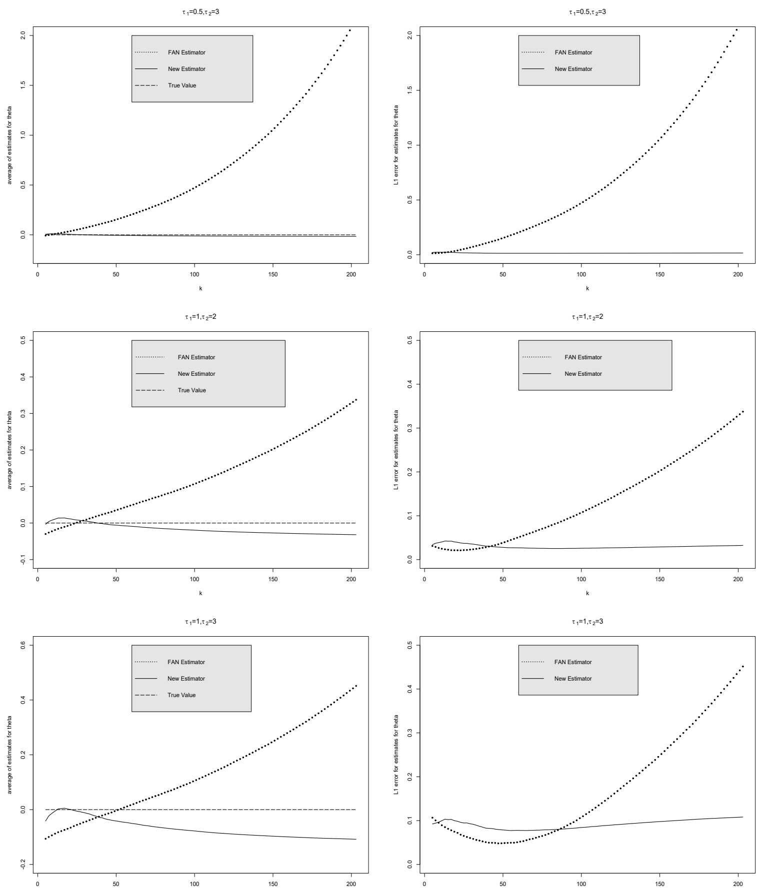

Figure 2: Estimated means (left) and estimated L_1 errors (right) for two endpoint estimators: New Estimator as the smallest solution to (2.17) and FAN Estimator defined in (S1.1). The samples are taken from distribution (3.26), where $\theta = 0$ and $\alpha = \tau_1\tau_2$

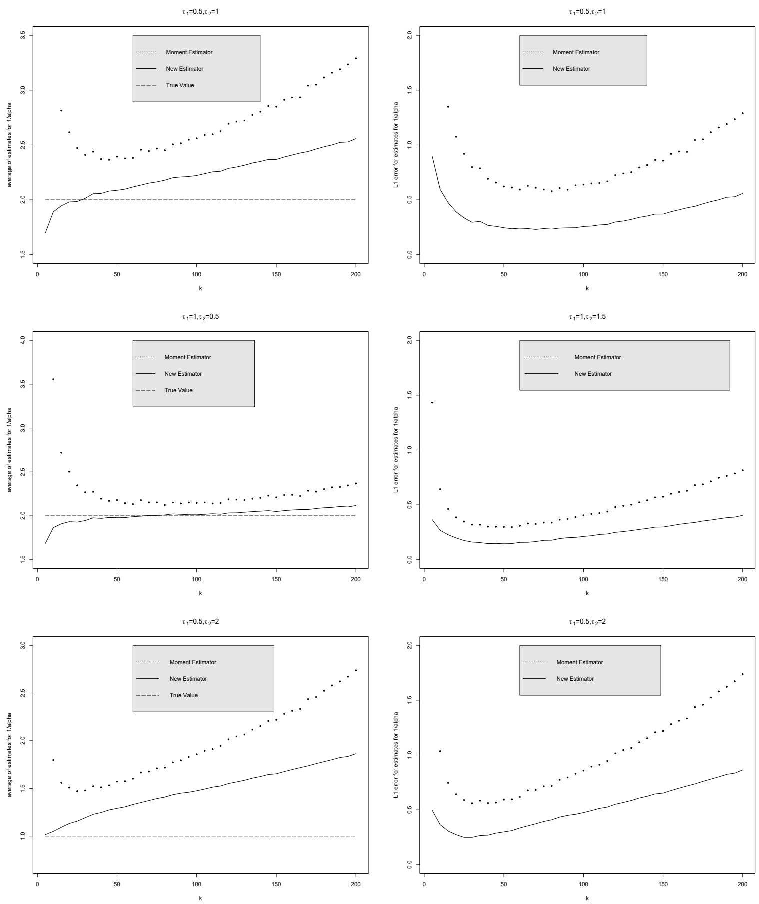

Figure 3: Estimated means (left) and estimated L_1 errors (right) for two estimators for α^{-1}: New Estimator $\tilde{\alpha}_N^{-1}$ defined in (2.16) and minus Moment Estimator $-\hat{\gamma}_M(k)$, where $\hat{\gamma}_M(k)$ is defined in (S1.2). The samples are taken from population $20 + X$, where X has distribution (3.26) and $\alpha = \tau_1\tau_2$

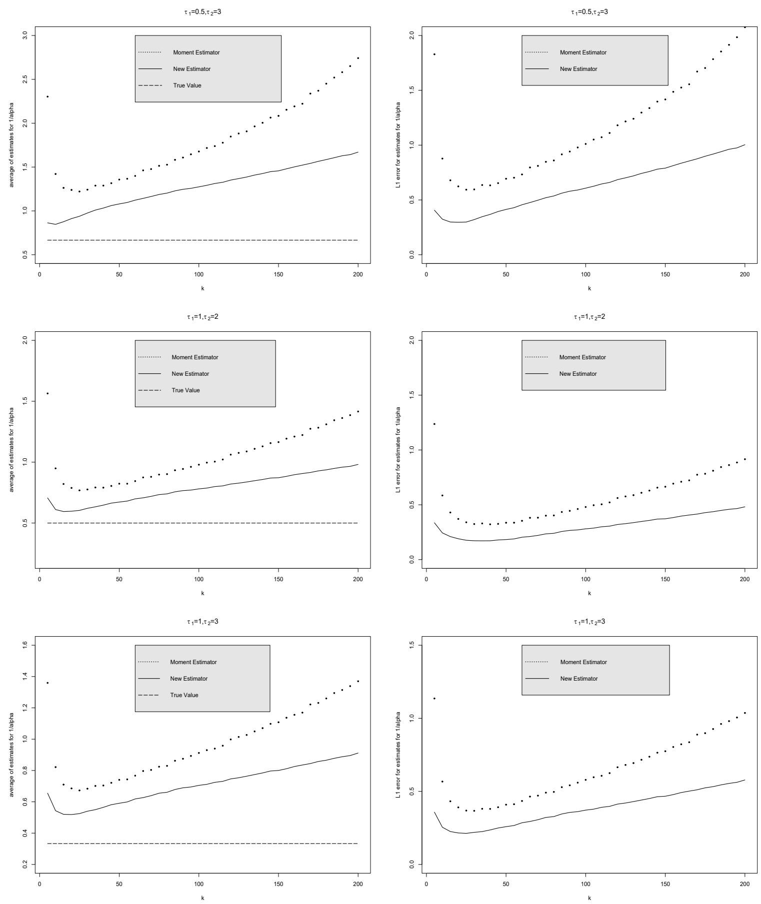

Figure 4: Estimated means (left) and estimated L_1 errors (right) for two estimators for α^{-1}: New Estimator $\tilde{\alpha}_N^{-1}$ defined in (2.16) and minus Moment Estimator $-\hat{\gamma}_M(k)$, where $\hat{\gamma}_M(k)$ is defined in (S1.2). The samples are taken from population $20+X$, where X has distribution (3.26) and $\alpha = \tau_1\tau_2$

S2 Details on data applications

Einmahl and Smeets (2011) tested extreme-value conditions for the two data sets. They applied the moment estimators for both the tail index $\gamma = -1/\alpha$ and the endpoint θ. It is important to decide the sample fraction or threshold k in the estimation, and this can be done by minimizing the so-called asymptotic mean squared errors (AMSE). They estimated γ by identifying some k-regions over which the AMSEs are relative small and stable and then used the average of all estimates of γ in these regions as the final estimate for the tail index for each event. Next, they estimated the endpoint for speed for each event by identifying k-regions and using the average of estimates for the endpoints over the regions. The two k-regions for men's 100 meters and women's 100 meters are $110-200$ and $80-210$, respectively.

First, we compare the performance of our likelihood method with the moment method. We estimate the speed endpoint and tail index for each of the two events and plot the estimates based on the likelihood method and the moment method in Figures 5 and 6, respectively. Note that the estimates for the endpoints in the moment method in Einmahl and Smeets (2011) use the same (fixed) estimates for tail index while in our study the estimates of γ depend on the sample fraction k. Therefore, our plots for moment estimates and the endpoints are different from those in Einmahl and Smeets (2011). We notice that there are similar patterns or trends for two types of estimation methods. But our likelihood estimators are more stable than the moment estimators in general.

Second, we decide a single value of sample fraction k for our likelihood estimates in the k-regions as the moment methods by Einmahl and Smeets (2011) so that we don't have to worry about violation of the extreme-value condition. For men's 100 meters, we check the k-region 110-200 and find out that both estimates for the tail index and the endpoint are highly stable when k changes from 140 to 160. We select $k = 160$ and the resulting estimates for γ and θ are -0.18 and 37.96. Based on Theorem **??**, the standard error for the endpoint estimate is 0.6837, and thus a 95% upper confidence limit is $37.95 + 1.645 \times 0.6937 = 39.09$. From formula $t = 36/s$, the estimates for the time endpoint and its 95% lower confidence limit are 9.48 and 9.21, respectively. Similarly, for women's 100 meters, we find out that both our estimates for the tail index and the endpoint are highly stable when k changes from 100 to 200 which is within the k-region $80 - 210$, and thus we are able to decide the sample fraction $k = 200$. The corresponding estimates are listed in Table 4. The results for the moment method from Einmahl and Smeets (2011) are also listed in Table 4 for comparison. The standard error of the likelihood estimate for the speed endpoint is 0.5606 for women's 100 meters.

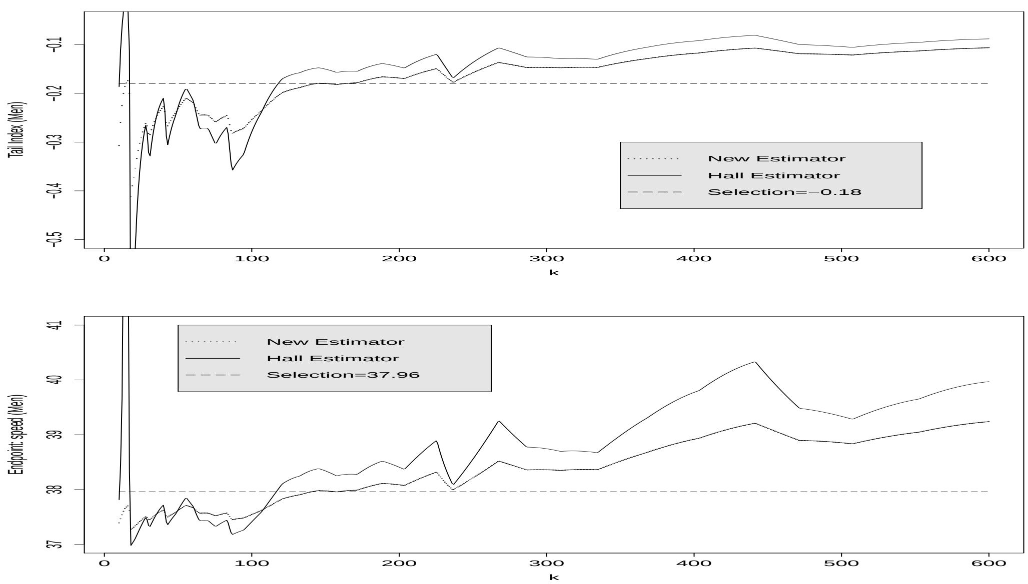

Figure 5: Our new likelihood estimates and the moment estimates for tail index $\gamma = -1/\alpha$ and the endpoint θ for speed (in km/h) for men's 100 meters.

S3 Proofs of Theorems 1, 2 and 3 in Section 2

S3.1 Some notation and lemmas

Let $V_1, \cdots, V_n$ be i.i.d. random variables with distribution function $1-1/x$ for $x \geq 1$ and $V_{n,1} \leq \cdots \leq V_{n,n}$ denote the order statistics of $V_1, \cdots, V_n$. Since $U(V_1), \cdots, U(V_n)$ are iid random variables with the distribution F, for

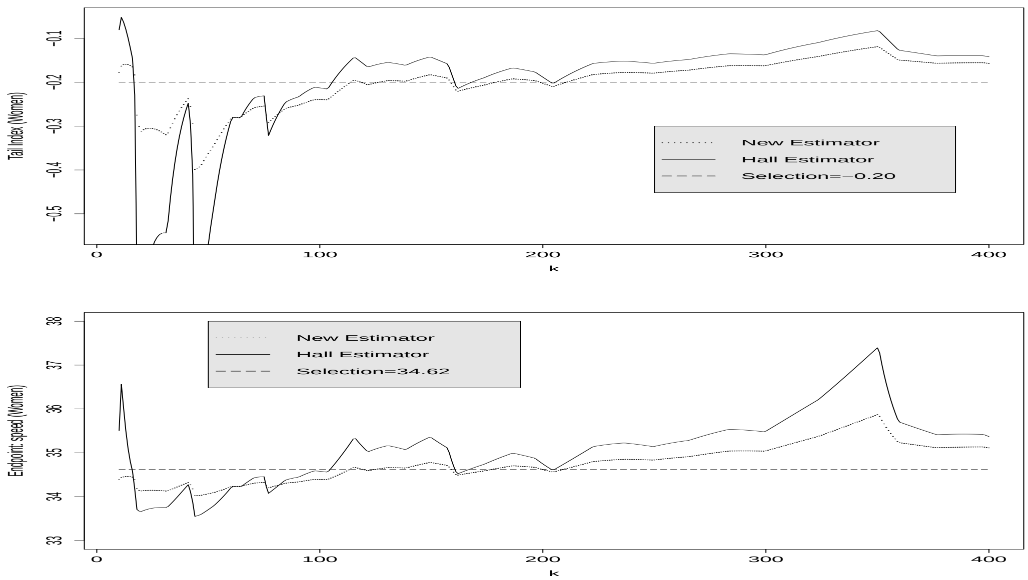

Figure 6: Our new likelihood estimates and the moment estimates for tail index $\gamma = -1/\alpha$ and the endpoint θ for speed (in km/h) for women's 100 meters.

convenience we assume $X_i = U(V_i)$ for $1 \le i \le n$ and hence $X_{n,i} = U(V_{n,i})$ for $1 \le i \le n$.

Consider another independent sequence of i.i.d. random variables V_1^*, $\cdots$, V_k^* with distribution function $1 - 1/x$ for $x \ge 1$. Denote $V_{k,1}^* \le \cdots \le V_{k,k}^*$ as their order statistics. It is well known that

$$\{V_{n,n-k+j}/V_{n,n-k}\}_{j=1}^k \stackrel{d}{=} \{V_{k,j}^*\}_{j=1}^k, \tag{S3.3}$$

see Page 71 of de Haan and Ferreira (2006). That is, $\{V_{n,n-k+j}/V_{n,n-k}\}_{j=1}^k$ are distributed the same as the order statistics of a sample of size k from the distribution function $1 - 1/x$ for $x \ge 1$. In the sequel, we will simply denote $V_{n,n-k+j}/V_{n,n-k}$ by $V^*_{k,j}$ for $1 \le j \le k$.

Set $S_k(\lambda) = \sum_{j=1}^k (V^*_{k,j})^\lambda = \sum_{j=1}^k (V^*_j)^\lambda$ for $\lambda > 0$ and define for $x \in \mathbb{R}$,

$$Q_k = \sqrt{k}\Big(\frac{1}{k}\sum_{j=1}^k \log V^*_{k,j} - 1\Big),$$

$$T^{(k)}_{\lambda,x} = \sum_{j=1}^{k-1} \frac{(V^*_{k,j})^\lambda}{1+(V^*_{k,j}/k)^\lambda x} \quad \text{for } \lambda > 1/2$$

and

$$R^{(k)}_{\lambda,x} = \begin{cases} \frac{1}{k^\lambda}\Big(\frac{(V^*_{k,k})^\lambda}{1+(V^*_{k,k}/k)^\lambda x} + (1-\lambda)(T^{(k)}_{\lambda,x} - \frac{k-1}{1-\lambda})\Big) & \text{if } \lambda \in (1/2, 1), \\ \frac{1}{k^\lambda}\Big(\frac{(V^*_{k,k})^\lambda}{1+(V^*_{k,k}/k)^\lambda x} + (1-\lambda)T^{(k)}_{\lambda,x}\Big) & \text{if } \lambda > 1. \end{cases}$$

Let $\{Y_n\}$ be a sequence of random variables and $\{a_n\}$ be a sequence of positive constants. Assume $\{A_n\}$ is a sequence of measurable sets. If $P(\{|Y_n/a_n| > \varepsilon\} \cap A_n) \to 0$ for every $\varepsilon > 0$, then we say Y_n/a_n converges in probability to zero on A_n and denote it by $Y_n = o_p(a_n)$ on A_n. If $\lim_{\varepsilon\to\infty}\limsup_{n\to\infty} P(\{|Y_n/a_n| > \varepsilon\} \cap A_n) = 0$, then we say Y_n/a_n is bounded on A_n and denote it by $Y_n = O_p(a_n)$ on A_n.

The following two lemmas are very helpful and easy to prove, and the details of the proofs are omitted here.

Lemma 1. *$Y_n = o_p(a_n)$ if and only if for every $\delta \in (0,1)$ there exists a sequence of measurable sets $\{A_n\}$ with $P(A_n) \ge \delta$ for all large n such that $Y_n = o_p(a_n)$ on A_n. The same conclusion is true if $o_p(a_n)$ is replaced by $O_p(a_n)$.*

Lemma 2. *Let $\{Y_n\}$ and $\{Z_n\}$ be two sequences of random variables such that $Y_n - Z_n = o_p(1)$. If the limiting distribution of Z_n exists and is continuous at x, then $\lim_{n\to\infty} P(Y_n \le x) = \lim_{n\to\infty} P(Z_n \le x)$.*

The following lemma deals with limits of $V^*_{k,k}$, $S_k(\lambda)$ and Q_k.

Lemma 3. *(i) $V^*_{k,k}/k \xrightarrow{d} \exp(-x^{-1})$ $(x > 0)$.*

(ii) If $\lambda \in (0,1)$, then $\frac{1}{k}S_k(\lambda) \xrightarrow{p} \frac{1}{1-\lambda}$.

(iii) If $\lambda \in (0,1/2)$, then $\frac{1}{\sqrt{k}}\Big(S_k(\lambda) - \frac{k}{1-\lambda}\Big) \xrightarrow{d} N(0, \frac{\lambda^2}{(1-\lambda)^2(1-2\lambda)})$.

(iv) If $\lambda = 1/2$, then $\frac{1}{\sqrt{k\log k}}\Big(S_k(1/2) - 2k\Big) \xrightarrow{d} N(0,1)$.

(v) If $\lambda = 1$, then $\frac{S_k(\lambda)}{k\log k} \xrightarrow{p} 1$.

(vi) If $\lambda > 1$, then $\frac{S_k(\lambda)}{k^\lambda} = O_p(1)$.

(vii) $Q_k \xrightarrow{d} N(0,1)$ as $k \to \infty$. If $\lambda \in (0,1/2)$, then $(\frac{1}{\sqrt{k}}(S_k(\lambda) - \frac{k}{1-\lambda}), Q_k) \xrightarrow{d} N(0,\Sigma_1)$, where

$$\Sigma_1 = \begin{pmatrix} \lambda^2(1-\lambda)^{-2}(1-2\lambda)^{-1} & \lambda(1-\lambda)^{-2} \\ \lambda(1-\lambda)^{-2} & 1 \end{pmatrix};$$

if $\lambda = 1/2$, then $\frac{1}{\sqrt{k\log k}}(S_k(1/2) - 2k)$ and Q_k are asymptotically independent.

Proof. (i) follows from a direct calculation that for $x > 0$, $P(V^*_{k,k}/k \le x) = (1 - \frac{1}{kx})^k$ for large k such that $kx > 1$, which has a limit $\exp(-x^{-1})$ as $k \to \infty$. See, also, de Haan and Ferreira (2006). Parts (ii) to (vii) follow from the classic theory of probability (see, eg, Loève (1977)) since

$S_k(\lambda) = \sum_{j=1}^{k}(V_j^*)^\lambda$ is the sum of k i.i.d. random variables for each $\lambda > 0$. Note that the mean $E((V_1^*)^\lambda) = \frac{1}{1-\lambda}$ is finite only if $\lambda \in (0,1)$ and the variance $Var((V_1^*)^\lambda) = \frac{\lambda^2}{(1-\lambda)^2(1-2\lambda)}$ is finite if $\lambda \in (0, 1/2)$. Therefore, part (ii) is a consequence of the classic law of large numbers and part (iii) follows from the standard central limit theorem. When $\lambda \geq 1/2$, the distribution of $(V_1^*)^\lambda$ is in the domain of attraction of a $1/\lambda$-stable law. If $\lambda = 1/2$, the stable law is normal and part (iv) follows from Loève (1977), page 364. IF $\lambda > 1$, $S_k(\lambda)/k^\lambda$ converges in distribution to a $1/\lambda$-stale law and part (vi) follows immediately. If $\lambda = 1$, $(S_k(1) - k\log k)/k$ converges in distribution to a 1-stable law, which implies part (v). The first part of (vi) follows from the standard central limit theorem, and the second part follows from the multivariate central limit theorem since

$$(\frac{1}{\sqrt{k}}(S_k(\lambda) - \frac{k}{1-\lambda}), Q_k) = \frac{1}{\sqrt{k}}\sum_{j=1}^{k}\Big((V_j^*)^\lambda - \frac{1}{1-\lambda}, \log V_j^* - 1\Big)$$

and Σ_1 is the covariance matrix of $(V_1^*)^\lambda$ and $\log V_1^*$. $\square$

Lemma 4. *Under condition (2.6) there exists a regularly varying function $A_1(t) \sim A(t)$ such that*

$$\theta_0 - U(t) = bt^{\gamma_0}\Big(-\frac{1}{\gamma_0} - \frac{1}{\gamma_0 + \rho}A_1(t)\Big) \text{ for all large } t$$

where $b = \lim_{t\to\infty} t^{-\gamma_0}a(t) = c^{\gamma_0}(-\gamma_0)$, and c is given in (1.1).

Proof. From Theorem 2.3.6 of de Haan and Ferria (2006) there exists a function $A_1(t) \sim A(t)$ such that for any $\epsilon > 0$ and $\delta > 0$

$$\left|\frac{\frac{U(tx)-U(t)}{bt^{\gamma_0}} - \frac{x^{\gamma_0}-1}{\gamma_0}}{A_1(t)} - \frac{x^{\gamma_0+\rho}-1}{\gamma_0+\rho}\right| \le \epsilon x^{\gamma_0+\rho} \max(x^{\delta}, x^{-\delta})$$

for all t as $tx \ge t_0$ for some $t_0 > 0$. Since $\lim_{x\to\infty} U(x) = \theta_0$, we get the desired result by selecting $\delta < -\gamma_0$ and letting $x \to \infty$. □

In Lemmas 5, 6 and 7 below and their proofs we use e^{ix} to denote the complex number $\cos x + i \sin x$.

Lemma 5. *Let $x \in \mathbb{R}$ and $v > 0$ be any constants such that $1 + v^{\lambda}x > 0$.*

(i) Conditional on $V^*_{k,k} = kv$,

$$\hat{T}_{\lambda,x} := \frac{1}{k^{\lambda}}(T^{(k)}_{\lambda,x} - \frac{k-1}{1-\lambda}) \xrightarrow{d} G_{\lambda,v,x} \qquad \text{if } \lambda \in (\frac{1}{2}, 1)$$

and

$$\hat{T}_{\lambda,x} := \frac{1}{k^{\lambda}} T^{(k)}_{\lambda,x} \xrightarrow{d} G_{\lambda,v,x} \qquad \text{if } \lambda \in (1, \infty).$$

(ii) Conditional on $V^*_{k,k} = kv$, Q_k converges in distribution to the standard normal, and Q_k and $\hat{T}_{\lambda,x}$ are asymptotically independent for $\lambda \in (\frac{1}{2}, 1)$ and $\lambda \in (1, \infty)$.

Proof. (i) Conditional on $V^*_{k,k} = kv$, the vector $(V^*_{k,1}, \cdots, V^*_{k,k-1})$ has the same joint distribution as that of the order statistics from $k-1$ iid random variables $Y_1(v), \cdots, Y_{k-1}(v)$ with a distribution function $F_{k,v}$ given by

$$F_{k,v}(y) = \frac{1 - y^{-1}}{1 - (kv)^{-1}} \quad \text{for} \quad 1 < y < kv.$$

Therefore, for each fixed $x \in \mathbb{R}$ and $v > 0$ such that $1 + v^\lambda x > 0$ we have that

$$P(T^{(k)}_{\lambda,x} \le s | V^*_{k,k} = kv) = P(\sum_{j=1}^{k-1} \frac{Y_j^\lambda(v)}{1 + (Y_j(v)/k)^\lambda x} \le s) \quad \text{for } s \in \mathbb{R}.$$

Set $Z_j = k^{-\lambda}(\frac{Y_j^\lambda(v)}{1+(Y_j(v)/k)^\lambda x} - \frac{1}{1-\lambda})$. Then we have

$$G^{(k)}_{\lambda,v,x}(s) := P(\frac{1}{k^\lambda}(T^{(k)}_{\lambda,x} - \frac{k-1}{1-\lambda}) \le s | V^*_{k,k} = kv) = P(\sum_{j=1}^{k-1} Z_j \le s).$$

We can check that

$$\begin{aligned}
& \delta_n(t) := E(e^{itZ_j}) - 1 \\
= \ & \frac{1}{1-(kv)^{-1}} \int_1^{kv} \big(\exp\{it(\frac{(y/k)^\lambda}{1+(y/k)^\lambda x} - k^{-\lambda}(1-\lambda)^{-1})\} - 1\big) y^{-2} dy \\
= \ & \frac{1}{k(1-(kv)^{-1})} \int_{1/k}^{v} \big(\exp\{it(\frac{y^\lambda}{1+y^\lambda x} - k^{-\lambda}(1-\lambda)^{-1})\} - 1\big) y^{-2} dy \\
= \ & \frac{1}{k(1-(kv)^{-1})} \int_{1/k}^{v} \big(\exp\{it\frac{y^\lambda}{1+y^\lambda x}\} \exp\{-itk^{-\lambda}(1-\lambda)^{-1}\} - 1\big) y^{-2} dy \\
= \ & \frac{1}{k(1-(kv)^{-1})} \int_{1/k}^{v} \big(\exp\{it\frac{y^\lambda}{1+y^\lambda x}\}(1 - itk^{-\lambda}(1-\lambda)^{-1}) - 1\big) y^{-2} dy + o(\frac{1}{k}) \\
= \ & \frac{1}{k(1-(kv)^{-1})} \int_{1/k}^{v} \big(\exp\{it\frac{y^\lambda}{1+y^\lambda x}\}(1 - itk^{-\lambda}(1-\lambda)^{-1})) - 1\big) y^{-2} dy + o(\frac{1}{k}) \\
= \ & \frac{1}{k(1-(kv)^{-1})} \int_{1/k}^{v} \big(\exp\{it\frac{y^\lambda}{1+y^\lambda x}\} - 1 - it\frac{y^\lambda}{1+y^\lambda x}\big)(1 - itk^{-\lambda}(1-\lambda)^{-1}) y^{-2} dy \\
& + \frac{1}{k(1-(kv)^{-1})} \int_{1/k}^{v} ((1 + it\frac{y^\lambda}{1+y^\lambda x})(1 - itk^{-\lambda}(1-\lambda)^{-1}) - 1\big) y^{-2} dy + o(\frac{1}{k}).
\end{aligned}$$

Some further manipulations show that

$$\begin{aligned}
\delta_n(t) \ = \ & \frac{1}{k} \int_0^v \big(\exp\{it\frac{y^\lambda}{1+y^\lambda x}\} - 1 - it\frac{y^\lambda}{1+y^\lambda x}\big) y^{-2} dy \\
& - \frac{it}{k}\Big(\int_0^v \frac{y^{2\lambda-2} x}{1+y^\lambda x} dy + \frac{v^{\lambda-1}}{1-\lambda}\Big) + o(\frac{1}{k}).
\end{aligned}$$

Note that the conditional characteristic function of $\sum_{j=1}^{k-1} Z_j$ is $(1 + \delta_n(t))^k$.

Thus

$$(1+\delta_n(t))^k \to f_{\lambda,v,x}(t).$$

Similarly, the case for $\lambda > 1$ can be verified.

(ii) The proof is standard by showing the convergence of the characteristic functions $E(e^{it_1 Q_k}|V^*_{k,k}=kv) \to e^{-t_1^2/2}$ and

$$E(e^{it_1 Q_k}e^{it_2\hat{T}_{\lambda,x}}|V^*_{k,k}=kv) \to e^{-t_1^2/2}f_{\lambda,v,x}(t_2)$$

for (t_1,t_2) in a neighborhood of $(0,0)$. The details are omitted here. $\square$

The following two lemmas consider the limiting distributions of $R^{(k)}_{\lambda,x}$ and Q_k.

Lemma 6. *Let* $\lambda \in (\frac{1}{2},1)$ *or* $\lambda \in (1,\infty)$.

(i) If $x \geq 0$, *then*

$$R^{(k)}_{\lambda,x} \xrightarrow{d} H_{\lambda,x};$$

(ii) If $x<0$, *then conditional on* $1+(V^*_{k,k}/k)^\lambda x > 0$,

$$R^{(k)}_{\lambda,x} \xrightarrow{d} \exp\{(-x)^{1/\lambda}\}H_{\lambda,x}.$$

Proof. Note that

$$R^{(k)}_{\lambda,x} = \frac{(V^*_{k,k}/k)^\lambda}{1+(V^*_{k,k}/k)^\lambda x} + (1-\lambda)\hat{T}_{\lambda,x},$$

where $\hat{T}_{\lambda,x}$ is defined in Lemma 5. We have shown in Lemma 5 that for any

$x \in \mathbb{R}$ and $v > 0$ such that $1 + v^\lambda x > 0$

$$f^{(k)}_{\lambda,v,x}(t) := E(e^{it\hat{T}_{\lambda,x}} | V^*_{k,k} = kv) \to f_{\lambda,v,x}(t) \tag{S3.4}$$

where $f_{\lambda,v,x}$ is the characteristic function of $G_{\lambda,v,x}$. Since $f^{(k)}_{\lambda,v,x}(t)$ is not defined when $kv \in (0,1]$, for convenience, we set $f^{(k)}_{\lambda,v,x}(t) = f_{\lambda,v,x}(t)$ when $kv \in (0,1]$.

Denote $\ell_k(v) := v^{-2}(1-(kv)^{-1})^k I(kv > 1)$, i.e., the density function of $V^*_{k,k}$. Set $\ell(v) = v^{-2}\exp(-v^{-1})I(v > 0)$, which is the density function of the distribution function $\exp(1 - v^{-1})$, $v > 0$. We can easily verify that $\int_0^\infty |\ell_k(v) - \ell(v)| dv \to 0$ as $k \to \infty$. In view of the dominated convergence theorem and (S3.4) we have

$$\int_0^\infty |f^{(k)}_{\lambda,v,x}((1-\lambda)t) - f_{\lambda,v,x}((1-\lambda)t)| \ell(v) dv \to 0.$$

When $x > 0$, the constraint $1 + v^\lambda x > 0$ is trivial and thus

$$E(e^{itR^{(k)}_{\lambda,x}}) = E\big(E(e^{itR^{(k)}_{\lambda,x}} | V^*_{k,k}/k)\big) = \int_0^\infty \exp(it(\frac{v^\lambda}{1+v^\lambda})) f^{(k)}_{\lambda,v,x}((1-\lambda)t)\ell_k(v)dv,$$

from which we have as $k \to \infty$

$$
\begin{aligned}
& |E(e^{itR^{(k)}_{\lambda,x}}) - \int_0^\infty \exp(it(\frac{v^\lambda}{1+v^\lambda}))f_{\lambda,v,x}((1-\lambda)t)\ell(v)dv| \\
\le \; & |\int_0^\infty \exp(it(\frac{v^\lambda}{1+v^\lambda}))f^{(k)}_{\lambda,v,x}((1-\lambda)t)\ell_k(v)dv \\
& \quad - \int_0^\infty \exp(it(\frac{v^\lambda}{1+v^\lambda}))f^{(k)}_{\lambda,v,x}((1-\lambda)t)\ell(v)dv| \\
& + \int_0^\infty \exp(it(\frac{v^\lambda}{1+v^\lambda}))f^{(k)}_{\lambda,v,x}((1-\lambda)t)\ell(v)dv \\
& \quad - \int_0^\infty \exp(it(\frac{v^\lambda}{1+v^\lambda}))f_{\lambda,v,x}((1-\lambda)t)\ell(v)dv| \\
\le \; & \int_0^\infty |\ell_k(v)-\ell(v)|dv + \int_0^\infty |f^{(k)}_{\lambda,v,x}((1-\lambda)t) - f_{\lambda,v,x}((1-\lambda)t)|\ell(v)dv \\
\to \; & 0.
\end{aligned}
$$

It is easily seen that $\int_0^\infty \exp(it(\frac{v^\lambda}{1+v^\lambda}))f_{\lambda,v,x}((1-\lambda)t)\ell(v)dv$ is the characteristic function of the distribution $H_{\lambda,x}$. This proves part (i) of the lemma.

When $x<0$, the natural constraint $1+(V^*_{k,k}/k)^\lambda x>0$ is equivalent to $V^*_{k,k}/k \in (0,\varphi_x)$. Therefore, we have

$$
\begin{aligned}
& E\big(e^{itR^{(k)}_{\lambda,x}}|1+(V^*_{k,k}/k)^\lambda x>0\big) \\
= \; & \tfrac{1}{P(1+(V^*_{k,k}/k)^\lambda x>0)} \int_0^{\varphi_x} \exp(it(\tfrac{v^\lambda}{1+v^\lambda}))f^{(k)}_{\lambda,v,x}((1-\lambda)t)\ell_k(v)dv.
\end{aligned}
$$

From Lemma 3 (i) we get $P(1+(V^*_{k,k}/k)^\lambda x>0) = P(V^*_{k,k}/k<\varphi_x) \to \exp(-(-x)^{1/\lambda})$. Similar to the proof for part (i), we have as $k\to\infty$

$$
\begin{aligned}
& \int_0^{\varphi_x} \exp(it(\frac{v^\lambda}{1+v^\lambda}))f^{(k)}_{\lambda,v,x}((1-\lambda)t)\ell_k(v)dv \\
\to \; & \int_0^{\varphi_x} \exp(it(\frac{v^\lambda}{1+v^\lambda}))f_{\lambda,v,x}((1-\lambda)t)\ell(v)dv.
\end{aligned}
$$

Hence, we get

$$
\begin{aligned}
& E\big(e^{itR_{\lambda,x}^{(k)}}|1+(V_{k,k}^{*}/k)^{\lambda}x>0\big) \\
\rightarrow\ & \exp((-x)^{1/\lambda})\int_0^{\varphi_x}\exp(it(\frac{v^{\lambda}}{1+v^{\lambda}}))f_{\lambda,v,x}((1-\lambda)t)\ell(v)dv.
\end{aligned}
$$

The limiting function is the characteristic function of the distribution

$$
\exp\{(-x)^{1/\lambda}\}H_{\lambda,x}(y)
$$

which is the conditional distribution of $V^{\lambda}(1+V^{\lambda}x)+(1-\lambda)Z_{\lambda,x}$ given $V<\varphi_x^{1/\lambda}$, where $Z_{\lambda,x}$ and V are two random variables such that V has a distribution $\exp(-v^{-1})$, $v>0$, and the conditional distribution of $Z_{\lambda,x}$ given $V=v$ is $G_{\lambda,v,x}$ defined in Section 2. This completes the proof of the lemma. □

Lemma 7. *Let $\lambda\in(\frac{1}{2},1)$ or $\lambda\in(1,\infty)$.*

(i) If $x\geq 0$, then $R_{\lambda,x}^{(k)}$ and Q_k are asymptotically independent.

(ii) If $x<0$, then conditional on $1+(V_{k,k}^{*}/k)^{\lambda}x>0$, $R_{\lambda,x}^{(k)}$ and Q_k are asymptotically independent.

Proof. We will sketch the proof for part (i) only. The proof for part (ii) is similar. From Lemma 5 we have

$$
f_{\lambda,v,x}^{(k)}(t,s):=E(e^{it\hat{T}_{\lambda,x}+isQ_k}|V_{k,k}^{*}=kv)\rightarrow f_{\lambda,v,x}(t)\exp(-\frac{s^2}{2}),
$$

which is parallel to (S3.4) in the proof of Lemma 6. Note that $\exp(-\frac{s^2}{2})$ is the characteristic function of the standard normal and is free of v. The rest of the proof follows the exactly same lines as those in the proof of Lemma 6. We then obtain that

$$|E(e^{it\hat{T}_{\lambda,x}+isQ_k})-\Big(\int_0^\infty \exp(it(\frac{v^\lambda}{1+v^\lambda}))f_{\lambda,v,x}((1-\lambda)t)\ell(v)dv\Big)\exp(-\frac{s^2}{2})| \to 0$$

as $k\to\infty$, which implies the asymptotic independence in part (i). $\square$

Before proving our theorems, we derive some useful inequalities. It follows from Lemma 4 that there exists a $C>0$ such that for all large t

$$|\frac{\theta_0-U(tx)}{\theta_0-U(t)}-x^{\gamma_0}|\le Cx^{\gamma_0}A_1(t) \quad \text{for all } x\ge 1.$$

Write

$$\delta(t,x)=(\frac{\theta_0-U(tx)}{\theta_0-U(t)}-x^{\gamma_0})/x^{\gamma_0}.$$

Then $|\delta(t,x)|\le CA_1(t)$ uniformly in $x\ge 1$ for all large t, and

$$\frac{U(tx)-U(t)}{\theta_0-U(t)}=1-x^{\gamma_0}(1+\delta(t,x)).$$

Now for each j, $1\le j\le k$, plug in $t=V_{n,n-k}$ and $x=\frac{V_{n,n-k+j}}{V_{n,n-k}}$ in the above equation we have

$$\frac{X_{n,n-k+j}-X_{n,n-k}}{\theta_0-X_{n,n-k}}=1-\Big(\frac{V_{n,n-k+j}}{V_{n,n-k}}\Big)^{\gamma_0}(1+\varepsilon_{n,j})=1-(V^*_{k,j})^{\gamma_0}(1+\varepsilon_{n,j}), \tag{S3.5}$$

where $\varepsilon_{n,j}=\delta(V_{n,n-k},\frac{V_{n,n-k+j}}{V_{n,n-k}})$. Since $A_1(t)$ is regularly varying with exponent ρ and $kV_{n,n-k}/n\to 1$ in probability, we get $A_1(V_{n,n-k})/A_1(n/k)\to 1$

in probability, and thus we have

$$\varepsilon_n := \max_{1\le j\le k} |\varepsilon_{n,j}| = O_p(A(n/k)).$$

For every $\theta > X_{n,n}$, define

$$\tau = \frac{\theta - X_{n,n-k}}{\theta_0 - X_{n,n-k}} \tag{S3.6}$$

and thus $\theta = X_{n,n-k} + \tau(\theta_0 - X_{n,n-k})$ for $\tau > \frac{X_{n,n}-X_{n,n-k}}{\theta_0 - X_{n,n-k}}$. Then we can write

$$\begin{aligned}
\frac{\theta - X_{n,n-k+j}}{\theta - X_{n,n-k}} &= 1 - \frac{X_{n,n-k+j} - X_{n,n-k}}{\theta - X_{n,n-k}} \\
&= 1 - \frac{X_{n,n-k+j} - X_{n,n-k}}{\tau(\theta_0 - X_{n,n-k})} \\
&= \frac{(V^*_{k,j})^{\gamma_0}(1 + (\tau - 1)(V^*_{k,j})^{-\gamma_0} + \varepsilon_{n,j})}{\tau}. \qquad \text{(S3.7)}
\end{aligned}$$

For each given $\delta \in (0, 1)$ define

$$A_n = \{1 + (\tau - 1)(V^*_{k,k})^{-\gamma_0} > \delta\} \cap \{\varepsilon_n < \delta/2\}$$

and

$$B_n = \{1 + (\tau - 1)(V^*_{k,k})^{-\gamma_0} > \delta\} \cap \{(\tau - 1)(V^*_{k,k})^{-\gamma_0} < \frac{1}{\delta}\} \cap \{\varepsilon_n < \delta/3\}.$$

Define $\beta_{n,j}$ and $\xi_{n,j}$ such that

$$\frac{\theta - X_{n,n-k}}{\theta - X_{n,n-k+j}} = \tau(V^*_{k,j})^{-\gamma_0} - (\tau - 1)(V^*_{k,j})^{-2\gamma_0} + \beta_{n,j} \tag{S3.8}$$

and

$$\frac{\theta - X_{n,n-k}}{\theta - X_{n,n-k+j}} = \frac{\tau(V^*_{k,j})^{-\gamma_0}}{1 + (\tau - 1)(V^*_{k,j})^{-\gamma_0}}(1 + \xi_{n,j}). \tag{S3.9}$$

Then, from (S3.7) we can show for all large n,

$$|\beta_{n,j}| \leq (\tau-1)^2(V^*_{k,j})^{-2\gamma_0} + \frac{5\tau}{\delta}\{(\tau-1)^2(V^*_{k,j})^{-3\gamma_0} + (V^*_{k,j})^{-\gamma_0}\varepsilon_n\} \quad \text{(S3.10)}$$

uniformly in $1 \leq j \leq k$ and τ on A_n and

$$\max_{1\leq j\leq n} |\xi_{n,j}| \leq \frac{2}{\delta}\varepsilon_n \quad \text{uniformly in } \tau \text{ on } B_n.$$

S3.2 Proof of Theorem 1

As we have known, there exists a unique solution to $h(\theta) = 0$ as defined in (2.3) on $\{X_{n,n} > X_{n,n-1}\}$. Since F is continuous in a neighborhood of θ and $X_{n,n-k} \to \theta$ almost surely, with probability one, $X_{n,n} = X_{n,n-1}$ can occur only finitely many times (in n). Set $A = \{X_{n,n} > X_{n,n-1} \text{ ultimately}\}$. Then $P(A) = 1$. Set $B = \{\hat{\theta} > \theta + \varepsilon \text{ infinitely often}\}$. If the statement in the theorem is false, then $P(B) > 0$ for some $\varepsilon > 0$, and hence $P(A \cap B) > 0$. We have from (2.5) that infinitely often in $A \cap B$

$$\begin{aligned}
1 &\leq \frac{\alpha_0}{k+1}\frac{X_{n,n}-X_{n,n-k}}{\hat{\theta}-X_{n,n}} + \frac{|\alpha_0-1|}{k+1}\sum_{j=1}^{k-1}\frac{X_{n,n-k+j}-X_{n,n-k}}{\hat{\theta}-X_{n,n-k+j}} \\
&\leq \frac{\alpha_0}{k+1}\frac{X_{n,n}-X_{n,n-k}}{\varepsilon} + \frac{|\alpha_0-1|}{k+1}\sum_{j=1}^{k-1}\frac{X_{n,n-k+j}-X_{n,n-k}}{\varepsilon} \\
&\leq \frac{2\alpha_0+1}{\varepsilon}(\theta - X_{n,n-k}) \\
&< 1,
\end{aligned}$$

which yields a contradiction. This completes the proof. □

S3.3 Proof of Theorem 2

Define

$$h_1(\tau) = h(X_{n,n-k} + \tau(\theta_0 - X_{n,n-k}))$$

and denote $\hat{\tau}$ as the solution to equation $h_1(\tau) = 0$. Then it is readily seen that

$$\hat{\theta} = X_{n,n-k} + \hat{\tau}(\theta_0 - X_{n,n-k}), \tag{S3.11}$$

or equivalently,

$$\hat{\theta} - \theta_0 = (\hat{\tau} - 1)(\theta_0 - X_{n,n-k}). \tag{S3.12}$$

Since $k = k_n \to \infty$, we have under condition (2.6) that $P(X_{n,n} > X_{n,n-k}) \to 1$ as $n \to \infty$. Thus, with probability tending to one, the ML estimator $\hat{\theta}$ is unique, and hence $\hat{\tau}$ is also the unique solution to $h_1(\tau) = 0$. It follows from Lemma 4 that

$$(\theta_0 - X_{n,n-k})/(n/k)^{\gamma_0} \xrightarrow{p} b/(-\gamma_0) = c^{\gamma_0}. \tag{S3.13}$$

We will aim at the limiting distribution of $\hat{\tau} - 1$ since the limiting distribution for $\hat{\theta} - \theta_0$ follows immediately from (S3.12) and (S3.13).

It is easy to see that for any sequence $\{\tau_n\}$, on $\{X_{n,n} > X_{n,n-k}\}$, $\hat{\tau} \le \tau_n$ if and only if $h_1(\tau_n) \le 0$ and $\tau_n > \frac{X_{n,n} - X_{n,n-k}}{\theta_0 - X_{n,n-k}}$, which implies

$$P(\hat{\tau} \le \tau_n) = P(h_1(\tau_n) \le 0,\ \frac{X_{n,n} - X_{n,n-k}}{\theta_0 - X_{n,n-k}} < \tau_n) + o(1). \tag{S3.14}$$

It follows from Lemma 3 and equation (S3.5) that

$$k^{-\gamma_0}(\frac{X_{n,n}-X_{n,n-k}}{\theta_0-X_{n,n-k}}-1)=-(\frac{V_{k,k}^*}{k})^{\gamma_0}(1+o_p(1))\xrightarrow{d}1-\exp(-(\max(0,-x))^{-\frac{1}{\gamma_0}}). \tag{S3.15}$$

Equations (S3.14) and (S3.15) play very important role in getting the limiting distributions of $\hat{\tau}$.

We will consider four cases: $\alpha_0>2$, $\alpha_0=2$, $\alpha_0\in(0,2)$, $\alpha_0\neq 1$, and $\alpha_0=1$.

Case 1: $\alpha_0>2$. For $x\in\mathbb{R}$ define $\tau_n=\tau_n(x)=1+\frac{x}{\sqrt{k}}$. For any $\delta>0$, we have that $P(A_n)\to 1$ as $n\to\infty$. It follows from (S3.8) and Lemma 3 that on A_n

$$\begin{aligned}
& |h_1(\tau_n)+\frac{1+\gamma_0}{\gamma_0}\Big((S_k(-\gamma_0)-\frac{k}{1+\gamma_0})-(S_k(-\gamma_0)-S_k(-2\gamma_0))\frac{x}{\sqrt{k}}\Big)| \\
\le\ & O(\frac{1}{k})(S_k(-2\gamma_0)+S_k(-3\gamma_0))+O(1)S_k(-\gamma_0)\varepsilon_n \\
& \quad +O_p(1)((V_{k,k}^*)^{-\gamma_0}+\frac{(V_{k,k}^*)^{-3\gamma_0}}{k}) \\
\le\ & O_p(kA(n/k)+k^{-\gamma_0}).
\end{aligned}$$

We have used the fact that $S_n(-3\gamma_0)\le(V_{k,k}^*)^{-\gamma_0}S_k(-2\gamma_0)$. Set $Y_n=h_1(\tau_n)/\sqrt{k}$ and $Z_n=\frac{1+\gamma_0}{-\gamma_0}\Big(\frac{1}{\sqrt{k}}(S_k(-\gamma_0)-\frac{k}{1+\gamma_0})+(S_k(-\gamma_0)-S_k(-2\gamma_0))\frac{x}{k}\Big)$. It follows that $Y_n-Z_n=o_p(1)$ under condition (2.7) and

$$Z_n\xrightarrow{d}N(\frac{-x}{1+2\gamma_0},\frac{1}{1+2\gamma_0})$$

from Lemma 3. Then we obtain from Lemma 2 that

$$\lim_{n\to\infty} P(h_1(\tau_n) \le 0) = \Phi(\frac{x}{\sqrt{1+2\gamma_0}}).$$

Since (S3.15) implies $P(\frac{X_{n,n}-X_{n,n-k}}{\theta_0-X_{n,n-k}} < \tau_n) \to 1$, we get from (S3.14) that $P(\hat{\tau} \le \tau_n(x)) \to \Phi(\frac{x}{\sqrt{1+2\gamma_0}})$ for all $x \in \mathbb{R}$, that is,

$$\sqrt{k}(\hat{\tau} - 1) \xrightarrow{d} N(0, 1+2\gamma_0),$$

which together with (S3.12) and (S3.13) yields (2.11).

Case 2: $\alpha_0 = 2$. We can show (2.12) similarly to **Case 1** by setting $\tau_n = \tau_n(x) = 1 + \frac{x}{\sqrt{k\log k}}$. The details are omitted here.

Case 3: $\alpha_0 \in (0,2)$**,** $\alpha_0 \ne 1$. Set $\tau_n = \tau_n(x) = 1 + k^{\gamma_0}x$. We consider two cases: $x \ge 0$ and $x < 0$.

Case 3.1: $x \ge 0$. It follows from Lemma 3 (i) that for any $\varepsilon > 0$, there exists a $\delta > 0$ such that $P(B_n) > 1 - \varepsilon$ for all large n. We have from Lemma 5 that $T^{(k)}_{-\gamma_0,x} = O_p(k)$ if $\alpha_0 \in (1,2)$ and $T^{(k)}_{-\gamma_0,x} = O_p(k^{-\gamma_0})$ if $\alpha_0 \in (0,1)$. Therefore, it follows from Lemma 1 and equation (S3.9) that for $\alpha_0 \in (1,2)$

$$\begin{aligned}
& k^{\gamma_0} h_1(\tau_n) \\
= \;& k^{\gamma_0}\big(\frac{(V^*_{k,k})^{-\gamma_0}}{1+(V^*_{k,k}/k)^{-\gamma_0}x} + (1+\gamma_0)(T^{(k)}_{-\gamma_0,x} - \frac{k-1}{1+\gamma_0})\big) + O_p(k^{1+\gamma_0})\varepsilon_n \\
= \;& k^{\gamma_0}\big(\frac{(V^*_{k,k})^{-\gamma_0}}{1+(V^*_{k,k}/k)^{-\gamma_0}x} + (1+\gamma_0)(T^{(k)}_{-\gamma_0,x} - \frac{k-1}{1+\gamma_0})\big) \qquad \text{(S3.16)} \\
& + O_p(k^{1+\gamma_0}A(n/k)),
\end{aligned}$$

which converges in distribution to $H_{-\gamma_0,x}$ in view of Lemma 5. Since $G_{-\gamma_0,v,x}(y)$ is continuous in y, it can be verified that $H_{-\gamma_0,x}(y)$ is continuous in y as well. The constraint $\frac{X_{n,n}-X_{n,n-k}}{\theta_0-X_{n,n-k}} < \tau_n$ is fulfilled automatically since $\frac{X_{n,n}-X_{n,n-k}}{\theta_0-X_{n,n-k}} < 1$. Therefore, we have from Lemma 2 and (S3.14) that

$$\lim_{n\to\infty} P(\hat{\tau} \le \tau_n) = \lim_{n\to\infty} P(k^{\gamma_0} h_1(\tau_n) \le 0) = H_{-\gamma_0,x}(0) = \Lambda_{-\gamma_0}(x) \quad \text{(S3.17)}$$

when $\alpha_0 \in (1,2)$. For $\alpha_0 \in (0,1)$ we have

$$\begin{aligned} k^{\gamma_0} h_1(\tau_n) &= k^{\gamma_0}\big(\frac{(V^*_{k,k})^{-\gamma_0}}{1+(V^*_{k,k}/k)^{-\gamma_0}x} + (1+\gamma_0)T^{(k)}_{-\gamma_0,x}\big) + O_p(\varepsilon_n) \\ &= k^{\gamma_0}\big(\frac{(V^*_{k,k})^{-\gamma_0}}{1+(V^*_{k,k}/k)^{-\gamma_0}x} + (1+\gamma_0)T^{(k)}_{-\gamma_0,x}\big) + O_p(A(n/k)). \end{aligned}$$

Similarly, by using Lemma 5 we obtain (S3.17) for $x \ge 0$.

Case 3.2: $x < 0$. The proof for $x < 0$ with $\alpha_0 \in (0,2)$ and $\alpha_0 \ne 1$ is a little bit complicated since we have to take into account of the constraint $\frac{X_{n,n}-X_{n,n-k}}{\theta_0-X_{n,n-k}} < \tau_n$. We only consider the case $x < 0$ and $\alpha_0 \in (1,2)$ since proof for $\alpha_0 \in (0,1)$ is similar.

From (S3.5) with $j = k$ and Lemma 3 (i) we have for $y < 0$

$$\begin{aligned} \lim_{n\to\infty} P(\frac{X_{n,n}-X_{n,n-k}}{\theta_0-X_{n,n-k}} < \tau_n(y)) &= \lim_{n\to\infty} P(k^{\gamma_0}(V^*_{k,k})^{-\gamma_0} < (-y)^{-1}) \\ &= \exp(-(-y)^{-1/\gamma_0}), \quad \text{(S3.18)} \end{aligned}$$

which is a continuous distribution function. Moreover, it follows from

(S3.15) that

$$\lim_{n\to\infty} E|I(\frac{X_{n,n}-X_{n,n-k}}{\theta_0-X_{n,n-k}}<\tau_n(y))-I(k^{\gamma_0}(V_{k,k}^*)^{-\gamma_0}<(-y)^{-1})|=0, \quad \text{(S3.19)}$$

where $I(A)$ denotes the indicator function of the event A. For any given small $\varepsilon>0$, if $\delta>0$ is small enough, we have that

$$\begin{aligned}
& E|I(k^{\gamma_0}(V_{k,k}^*)^{-\gamma_0}<(-x)^{-1})-I(k^{\gamma_0}(V_{k,k}^*)^{-\gamma_0}<(-x/(1-\delta))^{-1})| \\
= \;& P(k^{\gamma_0}(V_{k,k}^*)^{-\gamma_0}<(-x)^{-1})-P(k^{\gamma_0}(V_{k,k}^*)^{-\gamma_0}<(-x/(1-\delta))^{-1}) \\
\to \;& \exp(-(-x)^{-1/\gamma_0})-\exp(-(-x/(1-\delta))^{-1/\gamma_0}) \\
< \;& \varepsilon/2,
\end{aligned}$$

which implies that for all large k,

$$E|I(k^{\gamma_0}(V_{k,k}^*)^{-\gamma_0}<(-x)^{-1})-I(k^{\gamma_0}(V_{k,k}^*)^{-\gamma_0}<(-x/(1-\delta))^{-1})|<\varepsilon. \quad \text{(S3.20)}$$

Since $\{(V_{k,k}^*)^{-\gamma_0}<(-x/(1-\delta))^{-1}\}=\{1+(\tau_n(x)-1)V_{k,k}^*>\delta\}$, we have

$$E|I((V_{k,k}^*)^{-\gamma_0}<(-x/(1-\delta))^{-1})-I(B_n)|\to 0 \qquad \text{as } n\to\infty. \quad \text{(S3.21)}$$

Then it follows from approximation (S3.9) that (S3.16) holds on B_n. Since $\delta>0$ can be arbitrarily small, by using (S3.19) with $y=x$, (S3.20) and

(S3.21) we can show that

$$\begin{aligned}
& \lim_{n\to\infty} P(h_1(\tau_n) \le 0 | \frac{X_{n,n} - X_{n,n-k}}{\theta_0 - X_{n,n-k}} < \tau_n) \\
= & \lim_{n\to\infty} P(k^{\gamma_0}\big(\frac{(V^*_{k,k})^{-\gamma_0}}{1+(V^*_{k,k}/k)^{-\gamma_0}x} + (1+\gamma_0)(T^{(k)}_{-\gamma_0,x} - \frac{k-1}{1+\gamma_0})\big) \le 0 \\
& \quad |1 + (\frac{V^*_{k,k}}{k})^{-\gamma_0}x > 0) \\
= & \exp\{(-x)^{-1/\gamma_0}\} H_{-\gamma_0,x}(0),
\end{aligned}$$

where the last step follows from Lemma 6(ii). Once again we have (S3.17) by using (S3.14) and (S3.18) with $y = x$. Hence (2.13) follows from (S3.17) and (S3.13).

Case 4: $\alpha_0 = 1$**.** The case $\alpha_0 = 1$ can be verified directly since there is a close form solutuon $\hat{\theta} = X_{n,n} + (k+1)^{-1}(X_{n,n} - X_{n,n-k})$ as in Remark 1 in Section 2. Then, it follows from (S3.13) and (S3.15) that

$$\begin{aligned}
nc(\hat{\theta} - \theta_0) &= k\big(\frac{X_{n,n} - X_{n,n-k}}{\theta_0 - X_{n,n-k}} - 1\big)(1 + \frac{1}{k+1})\frac{c(\theta_0 - X_{n,n-k})}{(n/k)^{-1}} \\
&\quad + \frac{k}{k+1}\frac{c(\theta_0 - X_{n,n-k})}{(n/k)^{-1}} \\
&= k\big(\frac{X_{n,n} - X_{n,n-k}}{\theta_0 - X_{n,n-k}} - 1\big)(1 + o_p(1)) + 1 + o_p(1) \\
&\xrightarrow{d} 1 - Z
\end{aligned}$$

since the distribution function on the right-hand side of (S3.15) is the same as that of $-Z$, where Z is the standard exponential random variable. This completes the proof of Theorem 2. $\square$

S3.4 Proof of Theorem 3

Our approach in the proof is first to identify that the estimator $\tilde{\theta}$ falls within a small neighborhood of θ_0 and then to use some expansions to get the asymptotic distributions for both $\tilde{\theta}$ and $\tilde{\alpha}^{-1}$. The proof is very lengthy. We will consider three cases: $\alpha_0 > 2$, $\alpha_0 = 2$, and $\alpha_0 \in (0, 2)$.

Case 1: $\alpha_0 > 2$. The idea for the proof is somewhat similar to that of Theorem 6 in Hall (1982). We will split the proof into several steps.

Step 1. Some preparations.

Let $\{\theta_n\}$ be any sequence of random variables such that

$$n^{-\gamma_0}(\theta_n - \theta_0) = o_p(1). \tag{S3.22}$$

Define

$$\tau_n = \frac{\theta_n - X_{n,n-k}}{\theta_0 - X_{n,n-k}}.$$

Then it follows from (S3.13) that

$$k^{-\gamma_0}(\tau_n - 1) = \frac{n^{-\gamma_0}(\theta_n - \theta_0)}{(n/k)^{-\gamma_0}(\theta_0 - X_{n,n-k})} = \frac{n^{-\gamma_0}}{c^{\gamma_0}}(\theta_n - \theta_0)(1 + o_p(1)) = o_p(1). \tag{S3.23}$$

Since $n^{-\gamma_0}(\theta_0 - X_{n,n})$ converges in distribution to a positive and continuous random variable, we conclude that $P(\theta_n > X_{n,n}) \to 1$.

For any $\delta \in (0, 1)$, $P(A_n) \to 1$ as $n \to \infty$. By virtue of (S3.8) we have

$$\frac{\theta_n - X_{n,n-k}}{\theta_n - X_{n,n-k+j}} = (V_{k,j}^*)^{-\gamma_0}\big(1 + (\tau_n - 1)(1 - (V_{k,j}^*)^{-\gamma_0}) + (V_{k,j}^*)^{\gamma_0}\beta_{n,j}\big)$$

for $1 \le j \le k$.

From (S3.10) we have

$$\max_{1\le j\le k} (V_{k,j}^*)^{\gamma_0}|\beta_{n,j}| = o_p(1)$$

and thus

$$\begin{aligned}\log \frac{\theta_n - X_{n,n-k}}{\theta_n - X_{n,n-k+j}} &= -\gamma_0 \log V_{k,j}^* + (\tau_n - 1)(1-(V_{k,j}^*)^{-\gamma_0}) \\ &\quad + \big((V_{k,j}^*)^{\gamma_0}\beta_{n,j} + (\tau_n-1)^2(V_{k,j}^*)^{-2\gamma_0}\big)O_p(1),\end{aligned}$$

where $O_p(1)$ terms are uniform in j. Therefore we get that

$$\begin{aligned}&\frac{1}{k}\sum_{j=1}^{k}\log\frac{\theta_n - X_{n,n-k}}{\theta_n - X_{n,n-k+j}} \\ =\ & \frac{-\gamma_0}{k}\sum_{j=1}^{k}\log V_{k,j}^* + (\tau_n-1)(1-\frac{1}{k}S_k(-\gamma_0)) \\ & +O_p(1)\frac{1}{k}\sum_{j=1}^{k}(V_{k,j}^*)^{\gamma_0}|\beta_{n,j}| + (\tau_n-1)^2\frac{S_k(-2\gamma_0)}{k}O_p(1) \\ =\ & \frac{-\gamma_0}{k}\sum_{j=1}^{k}\log V_{k,j}^* + (\tau_n-1)(1-\frac{1}{k}S_k(-\gamma_0)) + O_p\big((\tau_n-1)^2 + A(n/k)\big),\end{aligned}$$

where the last step follows from Lemma 3 and (S3.10). Hence we conclude that

$$\begin{aligned}&\frac{1}{k}\sum_{j=1}^{k}\log\frac{\theta_n - X_{n,n-k}}{\theta_n - X_{n,n-k+j}} \\ =\ & \frac{-\gamma_0}{k}\sum_{j=1}^{k}\log V_{k,j}^* + (\tau_n-1)(1-\frac{1}{k}S_k(-\gamma_0)) + O_p\big((\tau_n-1)^2 + A(n/k) + \frac{1}{k}\big).\end{aligned} \tag{S3.24}$$

In a similar manner we obtain that

$$\begin{aligned}
&\sum_{j=1}^{k}\frac{\theta_n - X_{n,n-k}}{\theta_n - X_{n,n-k+j}}\\
= \;& S_k(-\gamma_0) + (\tau_n - 1)(S_n(-\gamma_0) - S_k(-2\gamma_0)) + O_p\big((\tau_n-1)^2 k^{1-\gamma_0} + kA(n/k)\big).
\end{aligned}$$

Since $\frac{\theta_n - X_{n,n-k}}{\theta_n - X_{n,n}} = O_p(k^{-\gamma_0})$, we have

$$\sum_{j=1}^{k-1}\frac{\theta_n - X_{n,n-k}}{\theta_n - X_{n,n-k+j}} = \sum_{j=1}^{k}\frac{\theta_n - X_{n,n-k}}{\theta_n - X_{n,n-k+j}} + O_p(k^{-\gamma_0}).$$

With some tedious calculations we obtain

$$\begin{aligned}
g(\theta_n) \;=\;& \big(S_k(-\gamma_0) - \frac{k}{1+\gamma_0}\big)(1+\gamma_0) + \frac{k\gamma_0}{1+\gamma_0}\big(\frac{1}{k}\sum_{j=1}^{k}\log V_{k,j}^{*} - 1\big)\\
&+\frac{\gamma_0^3}{(1+\gamma_0)^2(1+2\gamma_0)}k(\tau_n-1)(1+o_p(1)) \qquad \text{(S3.25)}\\
&+(\tau_n-1)^2 O_p(k^{1-\gamma_0}) + O_p\big(kA(n/k) + k^{-\gamma_0}\big).
\end{aligned}$$

From Lemma 3 we get

$$\frac{g(\theta_n)}{\sqrt{k}} = \frac{\gamma_0^3}{(1+\gamma_0)^2(1+2\gamma_0)}\sqrt{k}(\tau_n-1)(1+o_p(1)) + O_p(1). \qquad \text{(S3.26)}$$

Step 2. We show $n^{-\gamma_0}(\tilde{\theta} - \theta_0) \xrightarrow{p} 0$ as $n \to \infty$, that is,

$$P(n^{-\gamma_0}(\tilde{\theta} - \theta_0) > v) \to 0 \quad \text{for all } v > 0 \qquad \text{(S3.27)}$$

and

$$P(n^{-\gamma_0}(\tilde{\theta} - \theta_0) < -v) \to 0 \quad \text{for all } v > 0. \qquad \text{(S3.28)}$$

We will show (S3.27) here. The proof for (S3.28) is tedious and will be given in *Step 4*.

By setting $\theta_n = \theta_0 \pm n^{\gamma_0}/\log\log k$ in (S3.26) we have from (S3.23) that $g(\theta_n)/\sqrt{k} \xrightarrow{p} \mp\infty$, which implies that with probability tending to one, there exists a root $\theta \in (\theta_0 - n^{\gamma_0}/\log\log k, \theta_0 + n^{\gamma_0}/\log\log k)$ to the equation $g(\theta) = 0$. Since $\tilde{\theta}$ is defined to be the smallest solution to $g(\theta) = 0$ we have $P(n^{-\gamma_0}(\tilde{\theta} - \theta_0) > v) \to 0$ for all $v > 0$.

Step 3. Proof of (2.20).

Note that (S3.22) holds with $\theta_n = \tilde{\theta}$. Then it follows from (S3.25) and (S3.24) that

$$\begin{aligned}
&\sqrt{k}(\tilde{\tau} - 1) \\
= \;& \frac{(1+\gamma_0)^2(1+2\gamma_0)}{\gamma_0^3\sqrt{k}}\Big(\big(S_k(-\gamma_0) - \frac{k}{1+\gamma_0}\big)(1+\gamma_0) \\
&\quad + \frac{\gamma_0}{1+\gamma_0}\big(\sum_{j=1}^{k}\log V_{k,j}^* - k\big)\Big) + o_p(1)
\end{aligned}$$

and

$$\sqrt{k}(\tilde{\alpha}^{-1} - \alpha_0^{-1}) = \frac{-\gamma_0\sqrt{k}(\tilde{\tau} - 1)}{1+\gamma_0} - \frac{\gamma_0}{\sqrt{k}}\big(\sum_{j=1}^{k}\log V_{k,j}^* - k\big) + o_p(1).$$

Hence (2.20) follows from Lemma 3 (vii), (S3.12) and (S3.13).

Step 4: Proof of (S3.28).

We will expand $g(\theta)$ uniformly for $X_{n,n} < \theta < \theta_0$ or equivalently for $\frac{X_{n,n} - X_{n,n-k}}{\theta_0 - X_{n,n-k}} < \tau < 1$ via (S3.6). From (S3.5), this latter constraint is equivalent to $\{-1 - \varepsilon_{n,n} < (\tau - 1)(V_{k,k}^*)^{-\gamma_0} < 0\} =: C_n$.

Since $P(V^*_{k,k}/V^*_{k,k-1} > x) = 1/x$ for $x > 1$, by setting $\delta_1 = (2/(2-\varepsilon))^{-\gamma_0}$ (> 1) we have $P((V^*_{k,k})^{-\gamma_0}/(V^*_{k,k-1})^{-\gamma_0} > \delta_1) = 1 - \varepsilon/2$ for every $\varepsilon \in (0, 1)$. Hence, on $\{(V^*_{k,k})^{-\gamma_0}/(V^*_{k,k-1})^{-\gamma_0} > \delta_1\} \cap \{\varepsilon_n < (\delta_1 - 1)/2\}$,

$$-\frac{\delta_1 + 1}{2\delta_1} < (\tau - 1)(V^*_{k,k-1})^{-\gamma_0} < 0$$

holds uniformly for all $\tau \in C_n$, that is, $1 + (\tau - 1)(V^*_{k,k-1})^{-\gamma_0} > \delta$ holds uniformly on $\tau \in C_n$, where $\delta = (\delta_1 - 1)/(2\delta_1)$. Therefore, on $D_n = \{(V^*_{k,k})^{-\gamma_0}/(V^*_{k,k-1})^{-\gamma_0} > \delta_1\} \cap \{\varepsilon_n < \delta/3\}$, we have $1+(\tau-1)(V^*_{k,k-1})^{-\gamma_0} > \delta$, and thus by redefining B_n as $\{1 + (\tau - 1)(V^*_{k,k-1})^{-\gamma_0} > \delta\} \cap \{\varepsilon_n < \delta/3\}$ we have expansion (S3.9) for $1 \le j \le k - 1$ with $\max_{\tau \in C_n} \max_{1 \le j \le k-1} |\xi_{n,j}| \le \frac{2}{\delta}\varepsilon_n$ uniformly on B_n. So we have on D_n ($\subseteq B_n$)

$$\begin{aligned} g(\theta) &= K_n + J_{k,\tau}(1 + O_p(\varepsilon_n))(1 - W_{k,\tau} + O_p(\varepsilon_n)) - k \\ &= K_n + J_{k,\tau}(1 - W_{k,\tau}) - k + J_{k,\tau}(1 + W_{k,\tau})O_p(\varepsilon_n) \end{aligned}$$

uniformly on $\tau \in C_n$, where

$$J_{k,\tau} = 2 + \sum_{j=1}^{k-1} \frac{\tau(V^*_{k,j})^{-\gamma_0}}{1 + (\tau - 1)(V^*_{k,j})^{-\gamma_0}}, \quad W_{k,\tau} = \frac{1}{k}\sum_{j=1}^{k-1} \log \frac{\tau(V^*_{k,j})^{-\gamma_0}}{1 + (\tau - 1)(V^*_{k,j})^{-\gamma_0}}$$

and

$$K_n = \frac{\theta - X_{n,n-k}}{\theta - X_{n,n}} - \frac{J_{k,\tau}}{k}(\log \frac{\theta - X_{n,n-k}}{\theta - X_{n,n}})(1 + O_p(\varepsilon_n)).$$

Note that for all $\tau \in C_n$

$$\begin{aligned} J_{k,\tau} &= 2+\sum_{j=1}^{k-1}(V_{k,j}^*)^{-\gamma_0}+(1-\tau)\sum_{j=1}^{k-1}\frac{(V_{k,j}^*)^{-2\gamma_0}-(V_{k,j}^*)^{-\gamma_0}}{1+(\tau-1)(V_{k,j}^*)^{-\gamma_0}} \\ &\geq \sum_{j=1}^{k-1}(V_{k,j}^*)^{-\gamma_0}+(1-\tau)\sum_{j=1}^{k-1}((V_{k,j}^*)^{-2\gamma_0}-(V_{k,j}^*)^{-\gamma_0}) \\ &= \frac{k}{1+\gamma_0}+O_p(\sqrt{k})+(1-\tau)k\Big(\frac{1}{1+2\gamma_0}-\frac{1}{1+\gamma_0}+o_p(1)\Big) \end{aligned}$$

from Lemma 3. Meanwhile, we have

$$\begin{aligned} W_{k,\tau} &= \frac{-\gamma_0}{k}\sum_{j=1}^{k-1}\log V_{k,j}^*+\frac{1}{k}\sum_{j=1}^{k-1}\log(1+(1-\tau)\frac{(V_{k,j}^*)^{-\gamma_0}-1}{1+(\tau-1)(V_{k,j}^*)^{-\gamma_0}}) \\ &\leq \frac{-\gamma_0}{k}\sum_{j=1}^{k-1}\log V_{k,j}^*+\frac{1-\tau}{k}\sum_{j=1}^{k-1}\frac{(V_{k,j}^*)^{-\gamma_0}-1}{1+(\tau-1)(V_{k,j}^*)^{-\gamma_0}} \\ &\leq \frac{-\gamma_0}{k}\sum_{j=1}^{k}\log V_{k,j}^*+\frac{1-\tau}{k}\sum_{j=1}^{k-1}\frac{(V_{k,j}^*)^{-\gamma_0}-1}{1-(1+\delta/3)V_{k,k}^{\gamma_0}(V_{k,j}^*)^{-\gamma_0}}. \end{aligned}$$

It follows from Lemma 3 (vii) that

$$\sum_{j=1}^{k}\log V_{k,j}^* = k+O_p(\sqrt{k}).$$

Following those arguments in the proof of Lemma 5 and considering the conditional distribution given on $V_{k,k}^*$ we can show that on D_n

$$\frac{1}{k}\sum_{j=1}^{k-1}\frac{(V_{k,j}^*)^{-\gamma_0}-1}{1-(1+\delta/3)V_{k,k}^{\gamma_0}(V_{k,j}^*)^{-\gamma_0}} = (\frac{1}{1+\gamma_0}-1)(1+o_p(1)),$$

which coupled with the above estimates implies that

$$J_{k,r}(1-W_{k,r})-k \geq \frac{|\gamma_0|^3}{(1+\gamma_0)^2(1+2\gamma_0)}(1-\tau)k(1+o_p(1))+O_p(k^{1/2}).$$

We also notice that on D_n

$$J_{k,\tau} = 2 + O_p(\sum_{j=1}^{k-1} (V_{k,j}^*)^{-\gamma_0}) = O_p(k)$$

holds uniformly on $\tau \in C_n$, which implies $K_n \xrightarrow{p} \infty$ uniformly on $\tau \in C_n$. Therefore, we have from the above equations that on D_n

$$\frac{g(\theta)}{\sqrt{k}} \geq \frac{|\gamma_0|^3}{(1+\gamma_0)^2(1+2\gamma_0)}(1-\tau)\sqrt{k}(1+o_p(1)) + O_p(1) \tag{S3.29}$$

holds uniformly for $\tau \in C_n$. Since $P(D_n) > 1 - \varepsilon$ for all large n and any given $\varepsilon > 0$, we conclude from Lemma 1 that (S3.29) holds uniformly on C_n, and thus for every $v > 0$

$$\begin{aligned} \min_{X_{n,n}<\theta<\theta_0 - n^{\gamma_0} v} \frac{g(\theta)}{\sqrt{k}} &\geq O_p(1) + \frac{|\gamma_0|^3}{(1+\gamma_0)^2(1+2\gamma_0)} \frac{n^{\gamma_0} v k^{1/2}}{\theta_0 - X_{n,n-k}}(1+o_p(1)) \\ &= O_p(1) + \frac{|\gamma_0|^3}{(1+\gamma_0)^2(1+2\gamma_0)} c^{-\gamma_0} v k^{1/2+\gamma_0}(1+o_p(1)) \\ &\xrightarrow{p} \infty \end{aligned}$$

from (S3.13), which implies (S3.28).

Case 2: $\alpha_0 = 2$. The proof is similar to **Case 1**, and the details are omitted here.

Case 3: $\alpha_0 \in (0, 2)$. A different approach from the case $\alpha_0 \geq 2$ is needed in this case. We will approximate the function g defined in (2.17) by the function h defined in (2.3). Define the lower bound

$$h_L(\theta) = h(\theta) - a_n$$

and the upper bound

$$h_U(\theta) = h(\theta) + a_n,$$

where $\{a_n\}$ is a sequence of constants given by

$$a_n = \begin{cases} k^{1/2}(\log k)^2, & \text{if } \alpha_0 \in [1,2), \\ k^{-\gamma_0-1/2}(\log k)^2, & \text{if } \alpha_0 \in (0,1). \end{cases}$$

Then $a_n/k^{-\gamma_0} \to 0$ as $n \to \infty$.

Let θ_L and θ_U be the solutions to $h_L(\theta) = 0$ and $h_U(\theta) = 0$, respectively. If such solutions are not unique, θ_L and θ_U should be interpreted as the smallest ones.

For $\alpha_0 \in [1,2)$, we have $a_n/n \to 0$ as $n \to \infty$, and both h_L and h_U are decreasing functions of θ for $\theta > X_{n,n}$. Therefore, the solutions to $h_L(\theta) = 0$ and $h_U(\theta) = 0$ exist and are unique.

We continue to use the notation in the proof of Theorem 2. For $\alpha_0 \in (0,1)$, let $\tau_n = \tau_n(x) = 1 + k^{\gamma_0}x$, and define $\theta_n = \theta_n(x) = X_{n,n-k} + \tau_n(x)(\theta_0 - X_{n,n-k})$. Note that $h(\theta_n) = h_1(\tau_n)$, where h_1 is defined in the beginning of Section S3.3. It is readily seen that $k^{\gamma_0}h_U(\theta_n)$, $k^{\gamma_0}h_L(\theta_n)$, and $k^{\gamma_0}h(\theta_n)$ have the same limiting distribution function. From (S3.17), for every $\varepsilon > 0$, we can choose an $x > 0$ such that $P(k^{\gamma_0}h_L(\theta_n(x)) < 0) > 1 - \varepsilon$ and $P(k^{\gamma_0}h_U(\theta_n(x)) < 0) > 1 - \varepsilon$ for all large n. This ensures that $P(k^{\gamma_0}h_L(\theta_n(x_n)) < 0) \to 1$ and $P(k^{\gamma_0}h_U(\theta_n(x_n)) < 0) \to 1$ for some sequence of constants $\{x_n\}$ with $\lim_{n\to\infty} x_n = \infty$. Since $h'_L(\theta) = h'_U(\theta) = h'(\theta)$, we conclude, by using the same arguments in Section 2.1, that with probability tending to one, the solutions to $h_L(\theta) = 0$ and $h_U(\theta) = 0$ exist and are unique in the interval $(X_{n,n}, \theta_n(x_n))$.

From the proof of Theorem 2 we conclude that the limiting distributions for θ_L, θ_U and $\hat{\theta}$ are the same. Note that $\theta_L < \theta_U$. By using (S3.9) we can show that

$$\frac{1}{k}\sum_{j=1}^{k} \log \frac{\theta - X_{n,n-k}}{\theta - X_{n,n-k+j}} = \frac{-\gamma_0}{k}\sum_{j=1}^{k} \log V^*_{k,j} + O_p(|\tau - 1|) = -\gamma_0 + O_p(k^{-1/2}) \tag{S3.30}$$

uniformly on $\theta \in [\theta_L, \theta_U]$. Similarly, from (S3.8) and Lemma 3 we have

$$\sum_{j=1}^{k-1} \frac{\theta - X_{n,n-k}}{\theta - X_{n,n-k+j}} = \begin{cases} O_p(k), & \text{if } \alpha_0 \in (1,2); \\ O_p(k \log k), & \text{if } \alpha_0 = 1; \\ O_p(k^{-\gamma_0}), & \text{if } \alpha_0 \in (0,1) \end{cases}$$

uniformly on $\theta \in [\theta_L, \theta_U]$. It is easily seen that with probability tending to one,

$$h_L(\theta) \le g(\theta) \le h_U(\theta)$$

holds uniformly for $\theta \in [\theta_L, \theta_U]$. Therefore, there exists a root to the equation $g(\theta) = 0$ in the interval $[\theta_L, \theta_U]$ with probability tending to one, and we conclude that $P(n^{-\gamma_0}(\tilde{\theta} - \hat{\theta}) > v) \to 0$ for $v > 0$. Similar to the proof of (S3.28) we can show $P(n^{-\gamma_0}(\tilde{\theta} - \hat{\theta}) < -v) \to 0$ for $v > 0$. As a result we obtain that

$$n^{-\gamma_0}(\tilde{\theta} - \hat{\theta}) \xrightarrow{p} 0,$$

which implies that $\tilde{\theta}$ and $\hat{\theta}$ have the same limiting distributions.

For $\tilde{\alpha}^{-1}$, using a similar expansion to (S3.30) we have

$$\tilde{\alpha}^{-1} = \frac{-\gamma_0}{k} \sum_{j=1}^{k} \log V_{k,j}^{*} + o(k^{-1/2}),$$

which together with Lemma 7 yields (2.21). The asymptotic independence of $\tilde{\theta}$ and $\tilde{\alpha}^{-1}$ follows from the asymptotic independence of $\hat{\theta}$ and Q_k, which can be verified from Lemma 7 and the proof of Theorem 2. This completes the proof. $\square$

Bibliography

De Haan, L. and A. Ferreira (2006). *Extreme Value Theory: An Introduction.* Springer.

Dekkers, A.L.M, J.H.J. Einmahl and L. de Haan (1989). A moment estimator for the index of an extreme-value distribution. *Ann. Statist.* 17, pp. 1833 - 1855.

Einmahl, J. and S. Smeets (2011). Ultimate 100-m world records through extreme-value theory. *Stat. Neerl.* 65, pp. 32–42.

Fraga Alves, I. and C. Neves (2014). Estimation of the finite right endpoint in the Gumbel domain. *Statist. Sinica* 24, pp. 1811–1835.

Fraga Alves, I., C. Neves and P. Rosário (2017). A general estimator for the right endpoint with an application to supercentenarian womens records. Extremes. 20, pp. 199–237.

Loève, M. (1977). *Probability Theory I*, 4th Edition. Springer, New York.