Statistica Sinica **29** (2019), 203-224
doi:https://doi.org/10.5705/ss.202016.0336

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

Fang Wang, Liang Peng, Yongcheng Qi and Meiping Xu

Capital Normal University, Georgia State University,
University of Minnesota, Duluth and
Beijing Technology and Business University

Abstract: Consider a random sample from a regularly varying distribution function with a finite right endpoint θ and an exponent α of regular variation. The primary interest of the paper is to estimate both the endpoint and the exponent. Since the distribution is semiparametric and the endpoint and the exponent reveal asymptotic properties of the right tail for the distribution, inference can only be based on a few of the largest observations in the sample. The conventional maximum likelihood method can be used to estimate both α and θ, see e.g., Hall (1982) and Drees, Ferreira and de Haan (2004) for the regular case, $\alpha \geq 2$, and Smith (1987) and Peng and Qi (2009) for the irregular case, $\alpha \in (1,2)$. A global maximum of the likelihood function doesn't exist if one allows $\alpha \in (0,1]$, and a local maximum exists with probability tending to one only if $\alpha > 1$. We propose a penalized likelihood method to estimate both parameters. The estimators derived from this exist for all $\alpha > 0$ and any sample such that the largest two observations are distinct. We present the asymptotic distributions for the proposed maximum penalized likelihood estimators. A simulation study shows that the proposed method works very well for the irregular case, and has even better finite sample performance than the conventional maximum likelihood method for the regular case.

Key words and phrases: Endpoint, exponent, irregular case, limiting distribution, maximum likelihood.

1. Introduction

Let F be a distribution function with a finite right endpoint θ. For

$$1 - F(x) = c(\theta - x)^\alpha + o\{(\theta - x)^\alpha\} \quad \text{as } x \uparrow \theta, \tag{1.1}$$

where $c > 0$ is a constant and $\alpha > 0$ is called the exponent of F, statistical inference for θ and α has been of importance in the applications of extreme value theory; see, e.g., de Haan and Ferreira (2006); Einmahl and Magnus (2008); Einmahl and Smeets (2011). When the underlying distribution function is $F(x) = 1 - (1 - x/\theta)^\alpha$ for $x \in [0, \theta]$ and some $\alpha, \theta > 0$, it is easy to check that the Fisher information with respect to θ is finite for $\alpha > 2$ and infinite for $\alpha \leq 2$. Therefore, finding an efficient inference for the endpoint θ depends on whether $\alpha > 2$ or $\alpha \leq 2$. These are called the regular case and the irregular case, respectively, in the literature.

Taking a high threshold u_n and approximating the tail probability $1 - F(x)$ for $x \geq u_n$ by the parametric family $c(\theta - x)^\alpha$, a type of maximum likelihood (ML) method can be employed to estimate both θ and α. See, e.g., Hall (1982); Drees, Ferreira and de Haan (2004) for the regular case, and Smith (1985, 1987); Smith and Weissman (1985); Woodroofe (1974); Zhou (2009); Peng and Qi (2009) for the irregular case. For some other inference procedures for the endpoint, such as resampling, minimum distance, high order moments, Bayesian inference and others, we refer to Athreya and Fukuchi (1997); Falk (1995); Hall and Wang (1999, 2005); Loh (1984); Girard, Guillou and Stupfler (2012a,b); Beirlant, Fraga Alves and Gomes (2016); Fraga Alves and Neves (2014); Fraga Alves, Neves and Rosário (2017). Bias correction and interval estimation for the endpoint are available in Hall and Park (2002); Li and Peng (2009); Li, Peng and Xu (2011); Li, Peng and Qi (2011). Instead of assuming (1.1), Fraga Alves and Neves (2014) estimated the finite right endpoint of a distribution function by assuming that the underlying distribution is in the domain of attraction of Gumbel distribution.

Assume $X_1, \ldots, X_n$ are independent and identically distributed random variables having a distribution function F satisfying (1.1). Let $X_{n,1} \leq \cdots \leq X_{n,n}$ denote the order statistics of $X_1, \ldots, X_n$, and let $k = k_n$ be a sequence of integers satisfying $k/n \to 0$ as $n \to \infty$. When our (2.4) holds with $\rho < 0$, it is known that $X_{n,n} - \theta = O_p(n^{-1/\alpha})$. When $\alpha > 2$, an endpoint estimator based on the largest k order statistics can have a faster rate of convergence than $n^{-1/\alpha}$, especially for a larger α. Although many existing endpoint estimators work for all $\alpha > 0$, their convergence rate is usually slower than $n^{-1/\alpha}$ when $\alpha < 2$. For example, the estimators in Girard, Guillou and Stupfler (2012a,b) have the rate of convergence $n^{-1/2}p_n^{\alpha/2-1}$ for some p_n such that $np_n^{-\alpha} \to \infty$, if $\alpha < 2$, which implies that $n^{-1/2}p_n^{\alpha/2-1}/n^{-1/\alpha} \to \infty$. This is understandable since their estimators have a normal limit. Given the information that $\alpha < 2$, one can select the value of p_n as large as possible in the estimators by Girard, Guillou and Stupfler (2012a,b) such that $n^{-1/2}p_n^{\alpha/2-1}/n^{-1/\alpha} \to \infty$ at an arbitrarily slow rate. In this sense, one can argue that these estimators are essentially optimal for the irregular case. To achieve the exact rate of convergence as the maximum for the irregular case, a simple strategy suggested by Remark 4.5.5 of de Haan and Ferreira (2006) is to either use two different endpoint estimators for the regular case and the irregular case, or to employ different choices of sample fraction in the construction of an endpoint estimator. This depends on how effectively one can distinguish the regular case and the irregular case. Likelihood-based estimators via (1.1) only exist for $\alpha > 1$ and the corresponding endpoint estimators have the same rate of convergence as $X_{n,n}$ in the irregular case (see Hall (1982)). Based on exceedances and a generalized Pareto distribution, Smith (1987) estimated the endpoint separately for the regular case and the irregular case.

Likelihood-based approaches have been shown to be efficient for the regular case (see Coles and Dixon (1999); Pauli and Coles (2001)), but they are problematic for the irregular case (see Hall (1982); Smith (1987)). The problem of interest here is to find a method which is efficient as the likelihood approach in the regular case and overcomes the difficulties of the likelihood approach in the irregular case.

Treat $X_{n,n-k+1}, \ldots, X_{n,n}$ as k left-censored observations above the threshold $u_n = X_{n,n-k}$. By temporarily assuming that $1-F(x) = c(\theta-x)^\alpha$ for $u_n < x < \theta$, the censored likelihood function for $X_{n,n-k}, \ldots, X_{n,n}$, up to a constant scale, is given by

$$L(\theta, c, \alpha) = \left\{ \prod_{j=0}^{k} c\alpha(\theta - X_{n,n-k+j})^{\alpha-1} \right\} \left\{ 1 - c(\theta - X_{n,n-k})^\alpha \right\}^{n-k-1}. \qquad (1.2)$$

By maximizing the likelihood one can find ML estimators for the parameters θ, c, and α (if it is unknown). Hall (1982) derived the limiting distribution for the ML estimator for θ when $\alpha > 2$ is known, and the joint limiting distribution for the ML estimators for θ and α when $\alpha > 2$ is unknown. The limiting distribution for the ML estimator of θ was also obtained in Hall (1982) when $1 < \alpha < 2$ is known and $k \geq 2$ is fixed rather than divergent.

If $\alpha \in (0, 1)$ is known, the ML estimator for θ is $X_{n,n}$, at which the likelihood function $L(\theta, c, \alpha)$ is infinite. Hence, it is biased and always underestimates θ. On the other hand, when $\alpha > 0$ is unknown, the endpoint θ is the only parameter that can be estimated and the ML estimator for θ is $X_{n,n}$, since $L(X_{n,n}, c, \alpha)$ is infinite for any $\alpha \in (0, 1)$. The ML estimator of θ is also $X_{n,n}$ if $\alpha = 1$. Thus, when $\alpha > 0$ is unknown, jointly estimating θ and α by the maximum likelihood estimation in Hall (1982) is impossible unless we impose the constraint $\alpha > 1$.

We seek a method that avoids using the maximum observation as an estimator for the endpoint θ, can estimate θ and α simultaneously for all $\alpha > 0$ at the same rate of convergence as the maximum for estimating θ in the irregular case. We propose a penalized likelihood method to achieve these goals so as to improve the inference in Hall (1982). After showing that the corresponding score equations exist a solution for any given sample and $k \geq 2$ (as long as the largest two observations are distinct), we derive the limiting distribution for the new endpoint estimator when $\alpha > 0$ is known, and the joint limiting distribution for the new estimators of θ and α when $\alpha > 0$ is unknown. In particular, we show that the limiting distribution for this estimator of α is normal for all $\alpha > 0$ and, for the new estimator of θ, that the limiting distribution is normal if $\alpha \geq 2$ and non-normal if $\alpha < 2$.

The rest of the paper is organized as follows. Section 2 presents the penalized likelihood approach and the main asymptotic results of the paper. In Section 3 some simulation studies are reported that compared the performance of the new estimators with the maximum likelihood estimators in Hall (1982), with the high-order moments estimator for the endpoint by Girard, Guillou and Stupfler (2012b). Some discussion on these estimators is given as well. 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) can be found in Section S1 of the Supplement Materials. In Section 4, data sets on the men's and women's 100 meters dash are analyzed, and results from our likelihood method are compared with those using the moment method. More details on the data application are available in Section S2 of the Supplement. Proofs are given in Section S3 of the Supplement.

2. Methodologies and Main Results

Throughout we assume our observations $X_1, \ldots, X_n$ are independent and identically distributed random variables with distribution function F satisfying (1.1). Let $X_{n,1} \leq \cdots \leq X_{n,n}$ denote the order statistics of $X_1, \ldots, X_n$ with $k = k_n$ such that $k/n \to 0$ as $n \to \infty$. If we directly maximize the censored likelihood function $L(\theta, c, \alpha)$ at (1.2), the resulting estimator for θ is $X_{n,n}$ when $\alpha \in (0, 1]$. This underestimates the endpoint, and α is not estimable when $\alpha \in (0, 1)$. Moreover, given the sample $X_1, \ldots, X_n$ and $k \geq 2$, the score equations with respect to $L(\theta, c, \alpha)$ may have no solution even for $\alpha > 1$.

Here we add a penalization multiplier to $L(\theta, c, \alpha)$ such that the penalized likelihood function is always bounded, and the corresponding score equations always exist, and have a solution for any given sample and k, as long as the largest two observations are distinct. Take $p(\theta, \alpha, X_{n,n-k}, \ldots, X_{n,n})$ to be a general penalization function such that

$$L_1(\theta, c, \alpha) = L(\theta, c, \alpha) p(\theta, \alpha, X_{n,n-k}, \ldots, X_{n,n})$$

is bounded globally. Since $L(\theta, c, \alpha)$ is unbounded as $\theta \to X_{n,n}$, we need $p(\theta, \alpha, X_{n,n-k}, \ldots, X_{n,n}) \to 0$ as $\theta \to X_{n,n}$. A simple choice then is

$$p(\theta, \alpha, X_{n,n-k}, \ldots, X_{n,n}) = \frac{\theta - X_{n,n}}{\alpha(\theta - X_{n,n-k})},$$

where the numerator ensures that the penalization goes to zero as $\theta \to X_{n,n}$, but the denominator slows the convergence to avoid over-penalization, and the involved α is to ensure that the corresponding score equations always have a solution. Using this penalization, the penalized likelihood function is

$$\begin{aligned} L_1(\theta, c, \alpha) &= c^{k+1} \alpha^k (\theta - X_{n,n})^{\alpha} (\theta - X_{n,n-k})^{\alpha-2} \\ &\quad \times \left\{ \prod_{j=1}^{k-1} (\theta - X_{n,n-k+j})^{\alpha-1} \right\} \left\{ 1 - c(\theta - X_{n,n-k})^{\alpha} \right\}^{n-k-1} \end{aligned}$$

for $\theta > X_{n,n}$, and zero otherwise. The maximum penalized likelihood estimators are obtained by maximizing this likelihood function. When both α and θ are unknown, Hall (1982)'s estimator and the maximum penalized likelihood estimator for θ are defined as the smallest solutions to $m(\theta) = 0$ and $g(\theta) = 0$, respectively, where $m(\theta)$ is defined in (3.2) and $g(\theta)$ is defined in (2.15). In a simulation study, we had plotted functions $m(\theta)$ and $g(\theta)$ against θ for some samples drawn from the reverse Gamma distribution with true $\theta = 0$ and $n = 200$, which clearly shows that maximum likelihood estimate in Hall (1982) may not exist, but that the proposed maximum penalized likelihood estimate always exists.

We consider the cases of known α and unknown α separately. When α is assumed to be known, we focus on the endpoint estimation. When α is unknown, we estimate θ and α jointly. Throughout we let (α_0, θ_0) denote the true value of (α, θ).

2.1. Estimating θ with known α

Suppose the parameter $\alpha = \alpha_0 > 0$ is known and we are interested in estimating θ. We maximize L_1 with respect to c and θ, and denote the estimators of c and θ as $\hat{c}$ and $\hat{\theta}$, respectively. By differentiating the log-likelihood function $\log L_1$ with respect to θ and c, we have $\hat{c} = \{(k+1)/n\}(\hat{\theta} - X_{n,n-k})^{-\alpha_0}$, and $\hat{\theta}$ is the solution to

$$h(\theta) := \frac{\theta - X_{n,n-k}}{\theta - X_{n,n}} + \left(1 - \frac{1}{\alpha_0}\right) \sum_{j=1}^{k-1} \frac{\theta - X_{n,n-k}}{\theta - X_{n,n-k+j}} - \frac{2}{\alpha_0} - k = 0. \tag{2.1}$$

Assume that $X_{n,n} > X_{n,n-1}$. Since

$$h(X_{n,n}+) = \infty,\ h(\infty) = -\frac{k+1}{\alpha_0} < 0 \text{ and } h(\theta) \text{ is continuous,} \tag{2.2}$$

there exists at least one root to (2.1). We have that

$$h(\theta) = \frac{X_{n,n} - X_{n,n-k}}{\theta - X_{n,n}} + \frac{\alpha_0 - 1}{\alpha_0} \sum_{j=1}^{k-1} \frac{X_{n,n-k+j} - X_{n,n-k}}{\theta - X_{n,n-k+j}} - \frac{k+1}{\alpha_0} \tag{2.3}$$

is strictly decreasing in $\theta \in (X_{n,n}, \infty)$ when $\alpha_0 \geq 1$. Therefore the estimator $\hat{\theta}$ is unique if $\alpha_0 \geq 1$ and $X_{n,n} > X_{n,n-1}$. If $\alpha_0 \in (0,1)$, then

$$h'(\theta)(\theta - X_{n,n})^2 = \frac{1-\alpha_0}{\alpha_0} \sum_{j=1}^{k-1} \frac{(X_{n,n-k+j} - X_{n,n-k})(\theta - X_{n,n})^2}{(\theta - X_{n,n-k+j})^2} - X_{n,n} - X_{n,n-k}$$

is increasing in θ if $X_{n,n} > X_{n,n-1}$, which implies that the equation $h'(\theta) = 0$ has at most one root in $(X_{n,n}, \infty)$. As $h'(X_{n,n}+) = -\infty$, we conclude that i) $h'(\theta) < 0$ for all $\theta > X_{n,n}$, or ii) there exists a unique $\theta^* > X_{n,n}$ such that $h'(\theta) < 0$ for $\theta \in (X_{n,n}, \theta^*)$, $h'(\theta^*) = 0$ and $h'(\theta) > 0$ for $\theta > \theta^*$, or iii) there exists a unique $\theta^* > X_{n,n}$ such that $h'(\theta) < 0$ for $\theta \in (X_{n,n}, \theta^*) \cup (\theta^*, \infty)$ and $h'(\theta^*) = 0$. Thus, $h(\theta)$ is either i) a decreasing function on $(X_{n,n}, \infty)$, or ii) a decreasing function on $(X_{n,n}, \theta^*)$ and an increasing function on (θ^*, ∞), or iii) a decreasing function on $(X_{n,n}, \theta^*) \cup (\theta^*, \infty)$, which implies that there exists a unique estimator $\hat{\theta}$ for $\alpha_0 \in (0,1)$ by using (2.2) when $X_{n,n} > X_{n,n-1}$. In conclusion, there exists a unique solution to (2.1) for all $\alpha > 0$, any $k \geq 2$, when $X_{n,n} > X_{n,n-1}$.

We show that the estimator $\hat{\theta}$ is strongly consistent under some general conditions.

Theorem 1. *Assume that F has a finite right endpoint θ and is continuous in a neighborhood of θ. If $k \geq 2$ and $k/n \to 0$ as $n \to \infty$, then $\hat{\theta} \xrightarrow{a.s.} \theta_0$ as $n \to \infty$.*

Consistency does not require $k \to \infty$ as $n \to \infty$. In order to derive the asymptotic distribution for the proposed endpoint estimator, we need a second order regular variation condition to control the asymptotic bias of the proposed estimator. Suppose there exist functions $a(t) > 0$ and $A(t) \to 0$ such that

$$\lim_{t\to\infty} \frac{\{U(tx) - U(t)\}/a(t) - (x^{\gamma_0} - 1)/\gamma_0}{A(t)} = H_{\gamma_0,\rho}(x) := \frac{1}{\rho}\Big(\frac{x^{\gamma_0+\rho} - 1}{\gamma_0 + \rho} - \frac{x^{\gamma_0} - 1}{\gamma_0}\Big), \tag{2.4}$$

where $U(t)$ is the inverse function of $1/(1-F)$, $\gamma_0 = -1/\alpha_0 < 0$, and $\rho \leq 0$. Here $H_{\gamma_0,0}(x)$ is defined as $\lim_{\rho\uparrow 0} H_{\gamma_0,\rho}(x)$. When (2.4) holds, $|A(t)|$ is a regularly varying function with exponent ρ and (1.1) holds with $c = [\lim_{t\to\infty}\{\theta_0-$

$U(t)\}t^{-\gamma_0}]^{1/\gamma_0}$; see Lemma 4 in the Supplement for an explicit expression of U.

It is expected that the asymptotic distribution of the endpoint estimator is quite different for the case $\alpha > 2$ and the case $\alpha < 2$. A typical technique in handling the irregular case $\alpha < 2$ is via conditional characteristic functions as in Woodroofe (1974). Our analyses are more complicated since the new endpoint estimator is valid for all $\alpha > 0$ instead of $\alpha > 1$ as in Woodroofe (1974).

Let

$$\varphi_x = \begin{cases} (-x)^{-1}, & \text{if } x < 0, \\ \infty, & \text{if } x \geq 0, \end{cases}$$

$$H_{\lambda,x}(y) = \begin{cases} \displaystyle\int_0^{\varphi_x^{1/\lambda}} G_{\lambda,v,x}\left\{\frac{1}{1-\lambda}\left(y-\frac{v^\lambda}{1+v^\lambda x}\right)\right\} v^{-2}\exp(-v^{-1})dv, & \lambda \in \left(\frac{1}{2},1\right), \\ \displaystyle\int_0^{\varphi_x^{1/\lambda}} \left[1-G_{\lambda,v,x}\left\{\frac{1}{1-\lambda}\left(y-\frac{v^\lambda}{1+v^\lambda x}\right)\right\}\right] v^{-2}\exp(-v^{-1})dv, & \lambda > 1, \end{cases}$$

and write $\Lambda_\lambda(x) = H_{\lambda,x}(0)$ for $x \in \mathbb{R}$, where $G_{\lambda,v,x}$ is a distribution function with the characteristic function $f_{\lambda,v,x}$ given by

$$f_{\lambda,v,x}(t) = \begin{cases} \exp\left(\displaystyle\int_0^v \left\{\exp\left(it\frac{y^\lambda}{1+y^\lambda x}\right) - 1 - it\frac{y^\lambda}{1+y^\lambda x}\right\} y^{-2}dy \right. \\ \qquad \left. -it\left(\displaystyle\int_0^v \frac{y^{2\lambda-2}x}{1+y^\lambda x}dy + \frac{v^{\lambda-1}}{1-\lambda}\right)\right), & \lambda \in \left(\frac{1}{2},1\right) \\ \exp\left(\displaystyle\int_0^v \left\{\exp\left(it\frac{y^\lambda}{1+y^\lambda x}\right) - 1\right\} y^{-2}dy\right), & \lambda > 1. \end{cases}$$

Theorem 2. *Assume* (2.4) *holds and* $k = k_n$ *satisfies one of the following conditions:*

$$k \to \infty, \quad \frac{k}{n} \to 0, \quad k^{1/2}A\left(\frac{n}{k}\right) \to 0 \quad \textit{if} \quad \alpha_0 > 2; \tag{2.5}$$

$$k \to \infty, \quad \frac{k}{n} \to 0, \quad k^{1/2}(\log k)^{-1/2}A\left(\frac{n}{k}\right) \to 0 \quad \textit{if} \quad \alpha_0 = 2; \tag{2.6}$$

$$k \to \infty, \quad \frac{k}{n} \to 0, \quad k^{1+\gamma_0}A\left(\frac{n}{k}\right) \to 0 \quad \textit{if} \quad \alpha_0 \in (1,2); \tag{2.7}$$

$$k \to \infty, \quad \frac{k}{n} \to 0 \quad \textit{if} \quad \alpha_0 \in (0,1]. \tag{2.8}$$

Then we have

$$n^{-\gamma_0}k^{1/2+\gamma_0}c^{-\gamma_0}(\hat{\theta}-\theta_0) \xrightarrow{d} N(0,(1+2\gamma_0)) \quad \textit{if} \quad \alpha_0 > 2; \tag{2.9}$$

$$(n\log k)^{1/2}c(\hat{\theta}-\theta_0) \xrightarrow{d} N(0,1) \quad \textit{if} \quad \alpha_0 = 2; \tag{2.10}$$

$$n^{-\gamma_0}c^{-\gamma_0}(\hat{\theta}-\theta_0) \xrightarrow{d} \Lambda_{-\gamma_0} \quad \textit{if} \quad \alpha_0 \in (0,2), \quad \alpha_0 \neq 1; \tag{2.11}$$

$$n^{-\gamma_0}c^{-\gamma_0}(\hat{\theta}-\theta_0) \xrightarrow{d} 1 - Z \quad \textit{if} \quad \alpha_0 = 1, \tag{2.12}$$

where Z is a standard exponential random variable.

Remark 1. (**a**) From (2.3), $\hat{\theta} = X_{n,n} + (k+1)^{-1}(X_{n,n} - X_{n,n-k})$ when $\alpha_0 = 1$, and it is asymptotically unbiased in the sense that its limiting distribution has a zero mean. An anonymous referee has drawn our attention to the jackknife estimators for the endpoint in Miller (1964); Robson and Whitlock (1964). The two estimators for θ in Miller (1964); Robson and Whitlock (1964) are given, respectively, by

$$\hat{\theta}_{Miller} = X_{n,n} + \frac{n-1}{n}(X_{n,n} - X_{n,n-1}), \quad \hat{\theta}_{RW} = X_{n,n} + (X_{n,n} - X_{n,n-1}).$$

Our estimator $\hat{\theta} = X_{n,n} + (k+1)^{-1}(X_{n,n} - X_{n,n-k})$ has a similar form. For a brief comparison, let F be a uniform $(0,\theta)$ with $\theta > 0$. Then the mean squared errors for the three estimators are

$$\begin{aligned} \sigma^2_{Miller}(n) &:= E(\hat{\theta}_{Miller} - \theta)^2 = \frac{2\theta^2(n^2-n+1)}{n^2(n+1)(n+2)}, \\ \sigma^2_{RW}(n) &:= E(\hat{\theta}_{RW} - \theta)^2 = \frac{2\theta^2}{(n+1)(n+2)}, \\ \sigma^2_N(n,k) &:= E(\hat{\theta} - \theta)^2 = \frac{k+2}{k+1}\frac{\theta^2}{(n+1)(n+2)}. \end{aligned} \tag{2.13}$$

These mean squared errors can be obtained by using the formulas for the variances and covariances of order statistics from uniform distributions (see, e.g., Section 3.4 in Balakrishnan and Cohen (1991)); (2.13) is available in Miller (1964). One can see that $\sigma^2_{RW}(n) > \sigma^2_{Miller}(n) > \sigma^2_N(n,k)$ for $n \geq 4$, $k \geq 1$. Since $k \to \infty$, we have

$$\lim_{n\to\infty} \frac{\sigma^2_N(n,k)}{\sigma^2_{Miller}(n)} = \frac{1}{2} \quad \text{and} \quad \lim_{n\to\infty} \frac{\sigma^2_N(n,k)}{\sigma^2_{RW}(n)} = \frac{1}{2}.$$

(**b**) The conditions (2.6)–(2.8) are weaker than (2.5). The condition (2.8) imposes the weakest condition on k, and the second-order convergence rate A is not involved, although the second-order regular variation condition (2.4) is assumed. Some intuitive explanations are as follows. We have $X_{n,n} - \theta_0 = O_p(n^{-1/\alpha})$ under (2.4) with $\rho < 0$, which means $X_{n,n}$ is further from the endpoint for a larger α. If $\alpha > 2$, an endpoint estimator using the upper k order statistics generally has the rate of convergence $n^{-1/\alpha}k^{-1/2+1/\alpha}$, which is faster than $n^{-1/\alpha}$ when the second order approximation error is smaller. In this case, the second order approximation rate determines that k cannot be too large in order to ensure that the bias is negligible. However, when α is smaller, many observations are quite close to the endpoint. Hence, in the irregular case, the rate of convergence $n^{-1/\alpha}$ cannot be improved and so the second order approximation does not play a role in determining the asymptotic distribution, unlike in the regular case.

(**c**) It can be shown that the estimator $\hat{c} = \{(k+1)/n\}(\hat{\theta} - X_{n,n-k})^{-\alpha_0}$ for c is consistent.

2.2. Estimating θ and α jointly

When both θ and α are unknown, we can develop our new estimators of c, θ, and α via maximizing the penalized likelihood function $L_1(\theta, c, \alpha)$, obtaining the estimator $(\tilde{\theta}, \tilde{c}, \tilde{\alpha})$ of (θ, c, α). By solving score equations, we have $\tilde{c} = \{(k+1)/n\}(\tilde{\theta} - X_{n,n-k})^{-\tilde{\alpha}}$,

$$\tilde{\alpha}^{-1} = \frac{1}{k} \sum_{j=1}^{k} \log \frac{\tilde{\theta} - X_{n,n-k}}{\tilde{\theta} - X_{n,n-k+j}}, \tag{2.14}$$

and $\tilde{\theta}$ is the smallest root to the equation

$$\begin{aligned} g(\theta) := & \sum_{j=1}^{k} \Big(\frac{\theta - X_{n,n-k}}{\theta - X_{n,n-k+j}} - 1 \Big) \\ & - \frac{1}{k} \left(\sum_{j=1}^{k} \log \frac{\theta - X_{n,n-k}}{\theta - X_{n,n-k+j}} \right) \left(2 + \sum_{j=1}^{k-1} \frac{\theta - X_{n,n-k}}{\theta - X_{n,n-k+j}} \right) = 0 \end{aligned} \tag{2.15}$$

for $\theta > X_{n,n}$.

When θ is known, the best estimator for α^{-1} in a certain class of distributions is the uniform minimum variance unbiased (UMVU) estimator α_n^{-1} given by

$$\alpha_n^{-1} = \frac{1}{k} \sum_{j=1}^{k} \log \frac{\theta - X_{n,n-k}}{\theta - X_{n,n-k+j}}, \tag{2.16}$$

see, e.g., Falk (1995). The estimator of α^{-1} given by (2.14) is coincident with (2.16) if $\tilde{\theta}$ happens to be θ. Thus, if $\tilde{\theta}$ gives a good estimate for θ, $\tilde{\alpha}^{-1}$ should perform well as an estimator of α^{-1}.

If $X_{n,n-1} < X_{n,n}$, we have $g(X_{n,n}+) = \infty$. By using Taylor's expansion one can verify that $g(\theta) < 0$ if θ is large enough. Hence, it follows from the continuity of $g(\theta)$ that there exists at least one root to (2.15) for any given sample and k such that $X_{n,n-1} < X_{n,n}$. Unlike the case of known α, we cannot show that there is a unique solution when θ and α are jointly estimated.

Here are the joint limiting distributions for the estimators $\tilde{\theta}$ and $\tilde{\alpha}$.

Theorem 3. *Assume condition* (2.4) *holds and*

$$k \to \infty, \quad \frac{k}{n} \to 0, \quad k^{1/2} A\left(\frac{n}{k}\right) \to 0 \quad as \quad n \to \infty. \tag{2.17}$$

(i) *If* $\alpha_0 > 2$, *then*

$$\{n^{-\gamma_0} k^{1/2+\gamma_0} c^{-\gamma_0}(\tilde{\theta} - \theta_0), k^{1/2}(\tilde{\alpha}^{-1} - \alpha_0^{-1})\} \xrightarrow{d} N(0, \Sigma), \tag{2.18}$$

where

$$\Sigma = \begin{pmatrix} \gamma_0^{-2}(1+\gamma_0)^2(1+2\gamma_0) & (-\gamma_0)^{-1}(1+\gamma_0)(1+2\gamma_0) \\ (-\gamma_0)^{-1}(1+\gamma_0)(1+2\gamma_0) & (1+\gamma_0)^2 \end{pmatrix};$$

(ii) *If* $\alpha_0 \in (0, 2]$, *then*

$$k^{1/2}(\tilde{\alpha}^{-1} - \alpha_0^{-1}) \xrightarrow{d} N(0, \gamma_0^2), \tag{2.19}$$

$\tilde{\theta}$ *has the same limiting distribution as* $\hat{\theta}$ *given in Theorem* 2, *and* $\tilde{\alpha}^{-1}$ *and* $\tilde{\theta}$ *are asymptotically independent.*

Remark 2. (**a**) The estimator for α is always asymptotically normal, and the estimator for θ, when α is unknown, behaves as if α were known in the irregular case $\alpha \leq 2$. The condition (2.17) is required this time for all cases; it is needed only for (2.19).

(**b**) It can be shown that the estimator $\tilde{c} = \{(k+1)/n\}(\tilde{\theta} - X_{n,n-k})^{-\tilde{\alpha}}$ is consistent for c, which can be used to construct confidence intervals for θ in the regular case.

(**c**) In (2.18), $n^{-\gamma_0} k^{1/2+\gamma_0} = (n/k)^{1/\alpha_0} k^{1/2} \to \infty$.

2.3. Selection of the sample fraction

Theorems 2 and 3 provide answers to how one can select the sample fraction k so as to achieve the desired asymptotic distributions for estimators of the tail index and the endpoint. One has that condition (2.5) implies conditions (2.6) and (2.7), since both $(\log k_n)^{-1/2}$ and $k_n^{\gamma_0} = k_n^{-1/\alpha_0}$ go to zero as $n \to \infty$. Therefore a choice of k_n satisfying (2.5) can be employed for Theorems 2 and 3.

First, we show that there always exists a sequence of integers $\{\bar{k}_n\}$ satisfying (2.5). To see that $\max_{1 \leq k \leq \bar{k}_n} \sqrt{k}|A(n/k)| \to 0$ as $n \to \infty$, let $B(t) = \sup_{s \geq t} |A(s)|$. Since $A(t) \to 0$ as $t \to \infty$ regardless of ρ, $B(t)$ is non-increasing and vanishes at infinity. If we define $\bar{k}_n$ as the integer part of $\min\{\sqrt{n}, B^{-1}(\sqrt{n})\}$, then $\bar{k}_n \to \infty$ and $\bar{k}_n/n \to 0$ as $n \to \infty$, and $\sqrt{\bar{k}_n}|A(n/\bar{k}_n)| \leq \{B(\sqrt{n})\}^{1/2} \to 0$ as $n \to \infty$. Then

$$\max_{1 \leq k \leq \bar{k}_n} \sqrt{k}\left|A\left(\frac{n}{k}\right)\right| \leq \max_{1 \leq k \leq \bar{k}_n} \sqrt{k}\left|B\left(\frac{n}{k}\right)\right| \leq \{B(\sqrt{n})\}^{1/2} \to 0 \text{ as } n \to \infty.$$

A choice of k satisfying (2.5) can be obtained via estimating the second order regular variation parameter ρ when (2.4) holds with some $\rho < 0$. Since $|A(t)|$ is regularly varying with exponent ρ, we can apply Potter's bound and prove that (2.5) holds for any sequence of positive integers $k = k_n$ with $k_n \sim cn^{\beta}$ for $c > 0$, and $\beta \in (0, -2\rho/(1-2)\rho)$. For estimating ρ, we refer to Gomes, de Haan and Peng (2002).

A plot of the estimator against the sample fraction can be helpful in determining a sample fraction that can be used for inference. To construct confidence intervals or test some hypotheses, one looks for a sample fraction that results in an estimator with a negligible bias. Denote the estimators of α and θ given in (2.14) and (2.15) as $\tilde{\alpha}(k)$ and $\tilde{\theta}(k)$. When (2.4) holds, both estimators may fluctuate wildly when the values of k are small, and are relatively stable in a range of the sample fraction k from small to relatively large. The existence of such relatively stable ranges is implied by the asymptotic bias of the estimators. Hence for each estimator, one can observe a turning point for k, followed by an upward or a downward trend. We will examine several examples of this.

We consider some distribution functions given in (3.5) with parameters $\tau_1, \tau_2 > 0$. These distributions are related to the Burr distributions. The exponent of such a distribution with parameters τ_1 and τ_2 is $\alpha = \tau_1\tau_2$, and its endpoint is $\theta = 0$. We generated a random sample of size 1,000 each from the distribution with $(\tau_1, \tau_2) = (1, 2)$, or $(1, 1)$, or $(1, 0.5)$. The corresponding plots are given in Figure 1. The dashed lines in these plots are the true values of α and θ.

For the distribution with $(\tau_1, \tau_2) = (1, 2)$, both plots suggest the use of $k = 63$, and the corresponding estimates for α and θ are 1.8344 and 0.014626, respectively. For the distribution with $(\tau_1, \tau_2) = (1, 1)$, both plots suggest the use of $k = 183$, and the corresponding estimates for α and θ are 0.9731 and 0.0004992, respectively. For the distribution with $(\tau_1, \tau_2) = (1, 0.5)$, the plot for the estimates of α suggests the use of $k = 183$ with an estimate 0.5212 for α. The estimates for θ have no significant difference in the full range $1 \le k \le 999$, and all estimates are between -1.944×10^{-07} and 1.747×10^{-07}. Therefore, choosing any large k results in a satisfactory estimate.

3. Simulation Study and Further Discussions

Our comparison study consists of three parts. In the first part, we compare the performance of our likelihood method with Hall's conventional likelihood method. We consider the biases and mean squared errors for estimators for

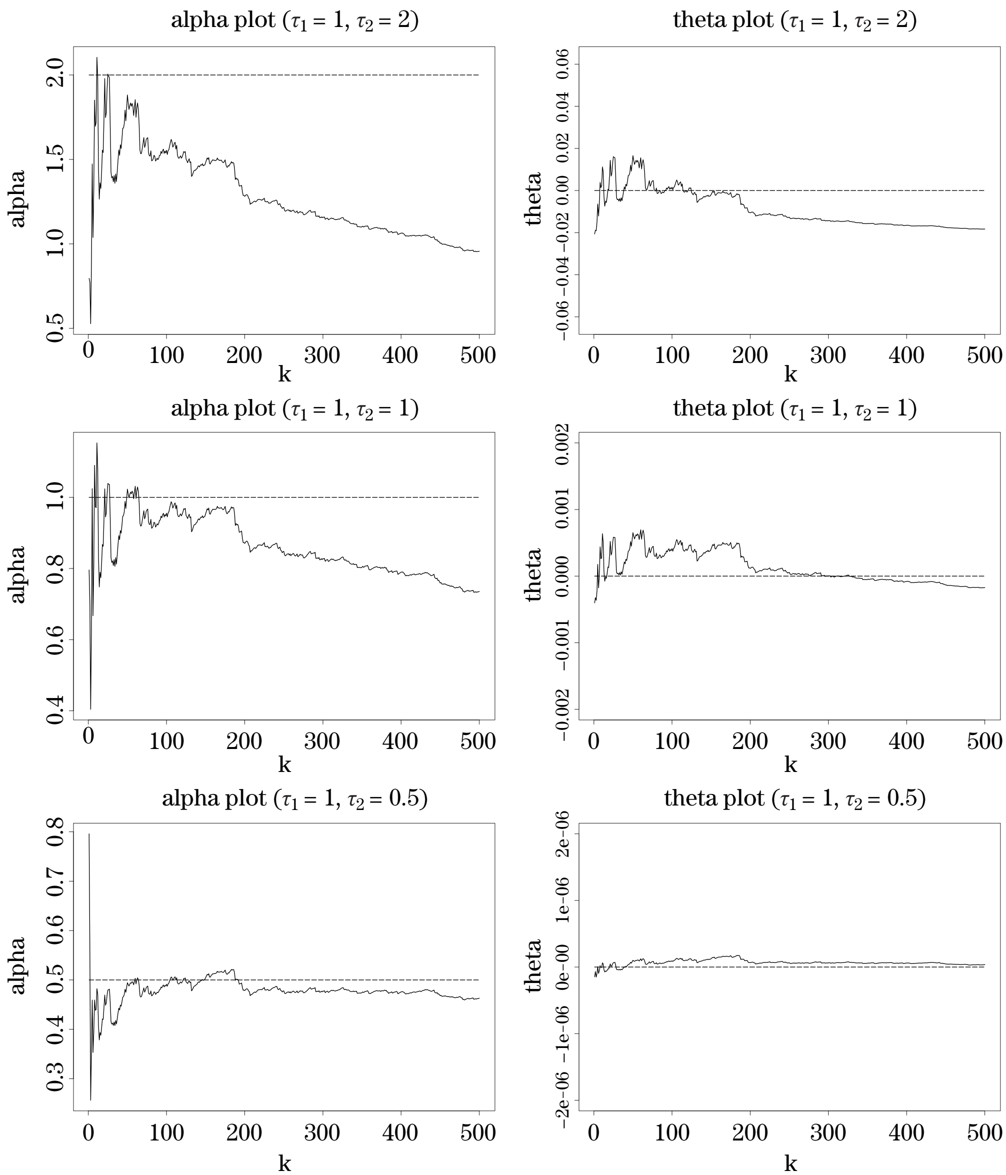

Figure 1. Plots of estimates for α and θ based a random sample of size 1,000 from distribution (3.5) with parameters with $(\tau_1, \tau_2) = (1, 2)$, $(1, 1)$, $(1, 0.5)$, respectively.

both the endpoint and the exponent of the distribution. In the second part, we compare the performance of the endpoint estimators based on our likelihood method with the high-order moments method proposed in Girard, Guillou and Stupfler (2012b). In the third part, we compare the new estimators with the estimators in Fraga Alves and Neves (2014); Dekkers, Einmahl and de Haan (1989).

We use $\tilde{\theta}_N$ and $\tilde{\alpha}_N^{-1}$ to denote our estimators $\tilde{\theta}$ and $\tilde{\alpha}^{-1}$ defined in Section 2.2.

3.1. Comparisons with the conventional likelihood method

Here the comparisons are with the conventional ML estimators proposed in Hall (1982), and the negative Hill estimator (see, e.g., Falk (1995) or Section 3.6.2 in de Haan and Ferreira (2006)).

When $\alpha = \alpha_0 \geq 2$ is known, Hall's ML estimator for θ is the unique solution of

$$\sum_{j=1}^{k}\left(\frac{\theta - X_{n,n-k}}{\theta - X_{n,n-k+j}} - 1\right) - \frac{k+1}{\alpha_0 - 1} = 0, \tag{3.1}$$

say $\hat{\theta}_H$. When $\alpha = \alpha_0 \in (1,2)$, Hall (1982) defined the estimator of θ by using a linear combination of a fixed number of largest order statistics. Theoretically, this could be extended to the case $\alpha_0 \in (1,2)$. This works only when $\alpha_0 > 1$. When $\alpha_0 \leq 1$, the conventional ML estimator for θ is simply $X_{n,n}$.

If $\alpha \geq 2$ is unknown, it follows from Hall (1982) that Hall's estimator for θ, denoted as $\tilde{\theta}_H$, is the smallest solution of

$$\begin{aligned} m(\theta) := {} & \frac{k+1}{\sum_{j=1}^{k} \log((\theta - X_{n,n-k})/(\theta - X_{n,n-k+j}))} \\ & - \frac{k+1}{\sum_{j=1}^{k} (X_{n,n-k+j} - X_{n,n-k})/(\theta - X_{n,n-k+j})} - 1 = 0 \end{aligned} \tag{3.2}$$

and the estimator for α^{-1} is

$$\tilde{\alpha}_H^{-1} = \frac{1}{k+1}\sum_{j=1}^{k} \log \frac{\tilde{\theta}_H - X_{n,n-k}}{\tilde{\theta}_H - X_{n,n-k+j}}. \tag{3.3}$$

To make a fair comparison, we chose the solution of (3.2) closest to the true value of θ, and if there is no root at all, we took $\tilde{\theta}_H = X_{n,n}$, as Hall (1982) suggested, and the estimator for α as the negative Hill estimator in (3.4).

When $\alpha \leq 1$, the conventional ML estimator for θ is $X_{n,n}$, but the conventional ML estimator for α does not exist. In this case, the negative Hill estimator,

$$\tilde{\alpha}_{NH}^{-1} = \frac{1}{k}\sum_{j=1}^{k-1} \log \frac{X_{n,n} - X_{n,n-k}}{X_{n,n} - X_{n,n-k+j}} \tag{3.4}$$

can serve as an estimator of α^{-1}. If $\alpha \in (0,2)$, this estimator behaves asymptotically like the UMVU estimator of α^{-1} in some ideal models as if θ were known (see, e.g., Falk (1995)): (2.19) holds for the estimator $\tilde{\alpha}_{NH}^{-1}$.

We conducted a simulation study on several distribution functions, including the reverse Gamma distributions with density function

$$f(x,\alpha,\theta) = \frac{(\theta - x)^{\alpha-1}}{\Gamma(\alpha)} \exp(-(\theta - x)), \quad x < \theta,$$

and the reverse Weibull distributions with density function

$$f(x,\alpha,\theta) = \alpha(\theta - x)^{\alpha-1} \exp(-(\theta - x)^{\alpha}), \quad x < \theta.$$

We only present the results for the reverse Gamma distributions since results are similar for others.

In the simulation we took the true value of θ to be zero and selected different values of $\alpha = 0.5, 1, 2$, and 3.

We generated $N = 1{,}000$ random samples of size n with n set at 100, 200, 500 and 1,000, and the values of k selected accordingly. For each combination of n and k, we calculated the estimates for θ and α by the different methods and then computed the biases and root mean squared errors of estimators for θ and α^{-1}.

Table 1 contains the results for the cases $\alpha = 0.5$ and 1. We only compared $\tilde{\theta}_N$ and $\tilde{\alpha}_N^{-1}$ with the negative Hill estimator $\tilde{\alpha}_{NH}^{-1}$ given by (3.4), and the endpoint estimator given by $\tilde{\theta}_M = X_{n,n}$. The column for $\tilde{\theta}_M$ has both the biases and root mean squared errors for different values of k as they are the same since the estimators $\tilde{\theta}_M$ do not depend on k. Our estimators for both θ and α^{-1} are less biased than the estimators $\tilde{\theta}_M$ and $\tilde{\alpha}_{NH}^{-1}$, with comparable root mean squared errors.

Table 2 presents the simulation results for the cases $\alpha = 2$ and 3. We reported simulation results for $\tilde{\theta}_N$ and $\tilde{\alpha}_N^{-1}$ and the estimators $\tilde{\theta}_H$ and $\tilde{\alpha}_H^{-1}$ in Hall (1982). Based on the results in Table 2, clearly our method is superior to the conventional ML method for both estimators of θ and α^{-1}. For both estimators of θ and α^{-1}, our estimators have the smallest biases; the root mean squared errors for the new estimators for θ are smaller in most cases, and the root mean squared errors for α^{-1} are the smallest among the three estimators for all cases reported in the table. The performance of the Hall estimators for α^{-1} is much worse than that of the negative Hill estimators, especially when k is small.

The choice of an optimal k is always challenging in extreme value theory, and needs more complicated justifications. The rate of convergence of the new endpoint estimator is independent of k for the irregular case, so one could employ

Table 1. Biases (upper values) and root mean-squared errors (lower values in the parentheses) of estimators of θ and α^{-1} when unknown $\alpha = 0.5$ and 1: $\tilde{\theta}_N$ and $\tilde{\alpha}_N^{-1}$ are our estimators for θ and α^{-1}, $\tilde{\theta}_M = X_{n,n}$ is the largest observation, and $\tilde{\alpha}_{NH}^{-1}$ is the negative Hill estimator as defined in (3.4).

			estimators of θ		estimators of α^{-1}	
α	n	k	$\tilde{\theta}_N$	$\tilde{\theta}_M$	$\tilde{\alpha}_N^{-1}$	$\tilde{\alpha}_{NH}^{-1}$
0.5	100	20	6.34×10^{-5}	-1.63×10^{-4}	−0.0431	−0.2779
			(4.45×10^{-4})	(3.84×10^{-4})	(0.4849)	(0.4728)
		30	1.82×10^{-5}	-1.63×10^{-4}	−0.0001	−0.1881
			(3.81×10^{-4})	(3.84×10^{-4})	(0.3862)	(0.3769)
0.5	200	20	2.29×10^{-5}	-4.24×10^{-5}	−0.0763	−0.3051
			(1.29×10^{-4})	(9.66×10^{-5})	(0.4976)	(0.4930)
		40	2.33×10^{-6}	-4.24×10^{-5}	−0.0171	−0.1701
			(8.50×10^{-5})	(9.66×10^{-5})	(0.3328)	(0.3370)
0.5	500	30	1.64×10^{-6}	-6.82×10^{-6}	−0.0710	−0.2492
			(1.80×10^{-5})	(1.69×10^{-5})	(0.3759)	(0.3980)
		60	-3.83×10^{-7}	-6.82×10^{-6}	−0.0258	−0.1399
			(1.54×10^{-5})	(1.69×10^{-5})	(0.2639)	(0.2769)
0.5	1,000	50	-2.66×10^{-8}	-1.76×10^{-6}	−0.0397	−0.1689
			(4.42×10^{-6})	(4.77×10^{-6})	(0.2854)	(0.3043)
		100	-3.20×10^{-7}	-1.76×10^{-6}	−0.0087	−0.0884
			(4.35×10^{-6})	(4.77×10^{-6})	(0.2024)	(0.2101)
1.0	100	20	5.08×10^{-3}	-1.02×10^{-2}	0.0253	0.0088
			(2.12×10^{-2})	(1.48×10^{-2})	(0.2990)	(0.2189)
		30	2.04×10^{-3}	-1.02×10^{-2}	0.0767	0.0533
			(1.62×10^{-2})	(1.48×10^{-2})	(0.2446)	(0.1893)
1.0	200	20	3.12×10^{-3}	-5.03×10^{-3}	−0.0191	−0.0246
			(1.05×10^{-2})	(7.01×10^{-3})	(0.2900)	(0.2153)
		40	9.95×10^{-4}	-5.03×10^{-3}	0.0467	0.0335
			(6.97×10^{-3})	(7.01×10^{-3})	(0.2015)	(0.1627)
1.0	500	30	9.80×10^{-4}	-2.01×10^{-3}	−0.0126	−0.0128
			(3.94×10^{-3})	(2.83×10^{-3})	(0.2362)	(0.1834)
		60	2.91×10^{-4}	-2.01×10^{-3}	0.0250	0.0193
			(2.51×10^{-3})	(2.83×10^{-3})	(0.1539)	(0.1304)
1.0	1,000	50	2.47×10^{-4}	-1.02×10^{-3}	−0.0016	−0.0040
			(1.41×10^{-3})	(1.45×10^{-3})	(0.1694)	(0.1400)
		100	6.55×10^{-5}	-1.02×10^{-3}	0.0195	0.0157
			(1.18×10^{-3})	(1.45×10^{-3})	(0.1121)	(0.0989)

a k obtained by any existing data-driven method for estimating an endpoint. Rather than choosing an optimal k, we conducted a simulation study for sample size $n = 1{,}000$ and different values for α by allowing a large range of values of k. We took all k from 10 to 100 and plot averages of the $N = 1{,}000$ estimates and

Table 2. Biases (upper values) and root mean-squared errors (lower values in the parentheses) of estimators of both θ and α^{-1} when unknown $\alpha = 2$ and 3: $\tilde{\theta}_N$ and $\tilde{\alpha}_N^{-1}$ are our estimators for θ and α^{-1}, $\tilde{\theta}_H$ and $\tilde{\alpha}_H^{-1}$ are Hall's ML estimators for θ and α^{-1}.

			estimators of θ			estimators of α^{-1}		
α	n	k	$\tilde{\theta}_N$	$\tilde{\theta}_H$	$\tilde{\theta}_M$	$\tilde{\alpha}_N^{-1}$	$\tilde{\alpha}_H^{-1}$	$\tilde{\alpha}_{NH}^{-1}$
2	100	20	0.0240	−0.0856	−0.1312	0.1292	1.0329	0.2124
			(0.1690)	(0.1741)	(0.1495)	(0.2696)	(1.3986)	(0.2797)
		30	−0.0047	−0.1031	−0.1312	0.1757	0.6758	0.2404
			(0.1348)	(0.1462)	(0.1495)	(0.2601)	(0.9265)	(0.2835)
2	200	20	0.0257	−0.0497	−0.0918	0.0783	0.8839	0.1824
			(0.1257)	(0.1526)	(0.1042)	(0.2380)	(1.2884)	(0.2524)
		40	0.0011	−0.0650	0.0918	0.1242	0.3547	0.1998
			(0.0945)	(0.0973)	(0.1042)	(0.1998)	(0.5332)	(0.2359)
2	500	30	0.0158	−0.0300	−0.0567	0.0505	0.3784	0.1565
			(0.0739)	(0.0733)	(0.0640)	(0.1827)	(0.6703)	(0.2083)
		60	0.0007	−0.0368	−0.0567	0.0868	0.1783	0.1605
			(0.0528)	(0.0551)	(0.0640)	(0.1482)	(0.2480)	0.1884
2	1,000	50	0.0070	−0.0218	−0.0405	0.0460	0.1589	0.1380
			(0.0465)	(0.0452)	(0.0458)	(0.1435)	(0.2781)	(0.1760)
		100	−0.0024	−0.0250	−0.0405	0.0773	0.1296	0.1382
			(0.0336)	(0.0374)	(0.0458)	(0.1206)	(0.1629)	(0.1589)
3	100	20	−0.0236	−0.2450	−0.3883	0.1934	0.7759	0.3047
			(0.3827)	(0.4208)	(0.4182)	(0.2955)	(1.1699)	0.3543
		30	−0.0724	−0.2768	0.3883	0.2281	0.5350	0.3220
			(0.3556)	(0.3915)	(0.4182)	(0.2954)	(0.7622)	(0.3557)
3	200	20	−0.0056	−0.1660	−0.3036	0.1499	0.6478	0.2734
			(0.3099)	(0.3588)	(0.3263)	(0.2621)	(1.0532)	(0.3255)
		40	−0.0526	−0.2018	−0.3036	0.1748	0.3157	0.2775
			(0.2654)	(0.2900)	(0.3263)	(0.2306)	(0.4342)	(0.3051)
3	500	30	0.0077	−0.1052	−0.2196	0.1049	0.2946	0.2375
			(0.2239)	(0.2409)	(0.2347)	(0.2012)	(0.5224)	(0.2771)
		60	−0.0336	−0.1287	−0.2196	0.1270	0.1954	0.2298
			(0.1703)	(0.1911)	(0.2347)	(0.1727)	(0.2387)	(0.2502)
3	1,000	50	0.0034	−0.0685	−0.1719	0.0855	0.1598	0.2079
			(0.1649)	(0.1907)	0.1833	(0.1562)	(0.2384)	(0.2347)
		100	−0.0321	−0.0955	−0.1719	0.1123	0.1570	0.1997
			(0.1277)	(0.1437)	0.1833	(0.1448)	(0.1849)	0.2147

their root mean squared errors for α^{-1} and θ in Figures 2 and 3, respectively. Here the true value for α is 3. Our estimators are superior to Hall's over the range of values selected for k in terms of biases and root mean squared errors.

Our estimators are competitive in that they can be applied directly without requiring any prior information on the parameters. They have satisfactory large

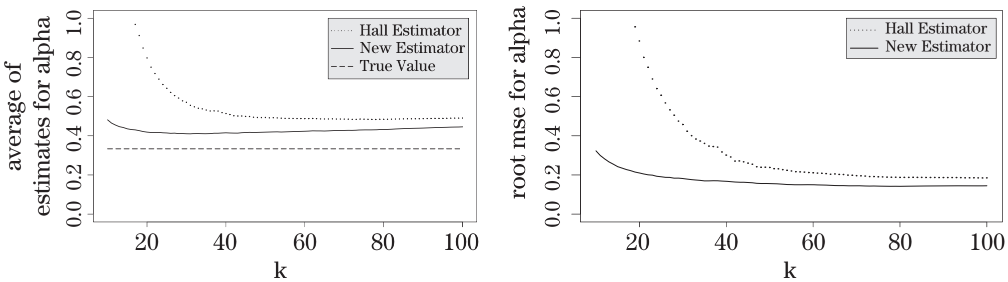

Figure 2. Estimated biases (left) and root mean-squared errors (right) of the new estimator and Hall's estimator for α^{-1} with sample size $n = 1{,}000$.

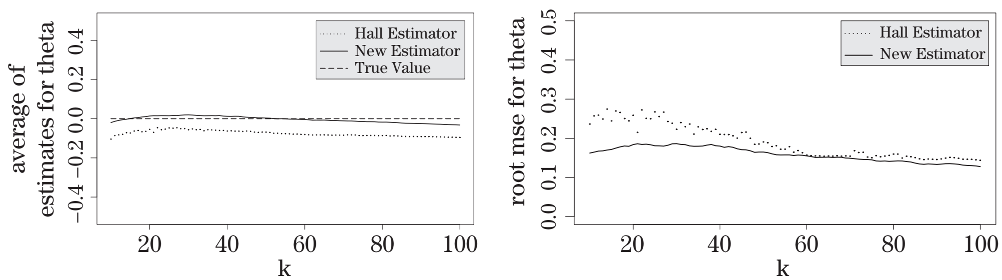

Figure 3. Estimated biases (left) and root mean-squared errors (right) of the new estimator and Hall's estimator for θ with sample size $n = 1{,}000$.

sample properties as well as very good small sample performance. Since the asymptotic distribution for the estimator of the endpoint is nonnormal for certain values of α, a simple unified interval estimate would be provided by a subsample bootstrap method. Further research is needed for constructing an efficient unified interval estimation procedure for the endpoint.

3.2. Comparisons with high-order moments method

Girard, Guillou and Stupfler (2012b) proposed a high-order moments estimator for endpoint θ based on the empirical moment-generating function

$$\mu(p) = \frac{1}{n}\sum_{j=1}^{n} e^{pX_j}, \quad p > 0.$$

The high-order moments estimator for θ is then

$$\Theta_n = \frac{1}{a}\left\{\log\frac{\mu(p_n)}{\mu(p_n+1)} - \log\frac{\mu((a+1)p_n)}{\mu((a+1)(p_n+1))}\right\},$$

where $a > 0$ is a fixed constant and p_n is a sequence of constants such that $p_n \to \infty$ as $n \to \infty$. Under certain conditions involving the underlying distribution F

and p_n, they showed that Θ_n is asymptotically normal. For our estimators, as well as for many other estimators such as moment estimators (see, Aarssen and de Haan (1994)), the parameter k represents the proportion of the sample that is used in the estimation. The high-order moments estimator uses all data points, and parameters p and a may be related to weights of the data points used in the estimation. In general, it seems not easy to compare the performance of different estimation methods at specific levels of their tuning parameters when the tuning parameters in different methods have different roles.

Girard, Guillou and Stupfler (2012b) compared the performance of their estimator with the maximum value estimator ($X_{n,n}$) and the moment estimator in terms of the optimal mean absolute errors, under two types of distributions. The first one was

$$1 - F(x) = \left\{1 + (-x)^{-\tau_1}\right\}^{-\tau_2}, \quad x < 0 \tag{3.5}$$

with $\tau_1, \tau_2 > 0$. A random variable X with distribution (3.5) can be written as $X = -1/Y$, where Y has a Burr$(1, \tau_1, \tau_2)$ type III distribution.

The second distribution employed was

$$1 - F(x) = \int_{\log(1-1/x)}^{\infty} \lambda^2 t e^{-\lambda t} dt, \quad x < 0 \tag{3.6}$$

with $\lambda > 0$. A random variable X with distribution (3.6) can be written as $X = -1/(e^Y - 1)$, where Y has a Gamma$(2, \lambda)$ distribution.

Models (3.5) and (3.6) have a right endpoint $\theta = 0$. Choose one distribution from (3.5) or (3.6). Their simulations compared the high-order moments estimator, the maximum value estimator and the moment estimator for θ. From their Table 1, Girard, Guillou and Stupfler (2012b) asserted that the high-order moments estimator outperforms over other two estimators in all cases.

We compared our likelihood estimator with the high-order moments estimator using the setups of Girard, Guillou and Stupfler (2012b). We took distributions from (3.5) and (3.6), generated $N = 1{,}000$ replicates of random samples of size $n = 500$ each, chose the same values for p and a, and used the same choices for parameters in the two distributions. We computed our estimate for θ with choices $k \in \{5, 10, 15, \ldots, 300\}$ and then estimated the corresponding optimal mean absolute error. Our simulation results are reported in Table 3, showing that the estimated optimal mean absolute errors for the our estimator are smaller than those for the high-order moments estimator.

Table 3. Comparisons of the endpoint estimators based on high-order moments (HOM) method and maximum penalized likelihood (MPL) method in terms of optimal mean absolute errors.

Distribution	$-1/\text{Burr}(1,\tau_1,\tau_2)$		Distribution	$-1/(\exp(\text{Gamma}(2,\lambda))-1)$	
Parameters \methods	HOM	MPL	Parameters \methods	HOM	MPL
$(\tau_1,\tau_2)=(1,1)$	$1.48\cdot 10^{-3}$	$1.40\cdot 10^{-3}$	$\lambda=1$	$1.68\cdot 10^{-4}$	$1.57\cdot 10^{-4}$
$(\tau_1,\tau_2)=(5/6,6/5)$	$1.50\cdot 10^{-3}$	$1.42\cdot 10^{-3}$	$\lambda=5/4$	$7.94\cdot 10^{-4}$	$7.47\cdot 10^{-4}$
$(\tau_1,\tau_2)=(2/3,3/2)$	$1.55\cdot 10^{-3}$	$1.47\cdot 10^{-3}$	$\lambda=5/3$	$3.87\cdot 10^{-3}$	$3.60\cdot 10^{-3}$
$(\tau_1,\tau_2)=(1/2,2)$	$1.72\cdot 10^{-3}$	$1.63\cdot 10^{-3}$	$\lambda=5/2$	$2.03\cdot 10^{-2}$	$1.83\cdot 10^{-2}$

4. Data Applications

We analyzed two data sets: the fastest personal times of 100-meters for men and women recorded from January 1, 1991 to June 19, 2008. The aim is to predict the ultimate world records for these two events. The current Men's record is 9.58 seconds, run by Usain Bolt at the 2009 World Championships; the Women's record is 10.49 seconds, run by Florence Griffith-Joyner at the 1988 Olympic Trials. These records are not included in the data sets because they were not set in this time period.

The two datasets have been studied in Einmahl and Smeets (2011) by using the moment estimators proposed in Dekkers, Einmahl and de Haan (1989). The dataset for men's 100 meters consists of 762 best personal times ranging from 9.72 to 10.30 (seconds), while the dataset for women's 100 meters has 479 data points ranging from 10.65 to 11.38 (seconds).

Times for the two events are available in hundredths of seconds and thus there are many ties in the data sets. A smoothed method was used as in Einmahl and Magnus (2008); Einmahl and Smeets (2011); given $m(m \geq 2)$ athletes with equal personal best time y (in seconds), smooth them equally over the interval $(y-0.005, y+0.005)$ by the m data points $y-0.005+0.01(2j-1)/(2m)$, $j=1,\ldots,m$. We calculated speeds in kilometers per hour and analyzed those. Our estimates as well as the results from the moment method in Einmahl and Smeets (2011) are listed in Table 4.

We compared results from the two different estimation methods. For men's 100 meters, our estimate gives an estimated ultimate men's world record 9.48 seconds, 0.10 seconds lower than the current world record 9.58 seconds, while the moment method provides an estimate of 9.51 seconds. Both methods yield the same 95% lower confidence limit 9.21 seconds. For the women's 100 meters, our method gives an estimate 10.40 seconds, 0.09 seconds lower than the

Table 4. Ultimate world records in speed (km/h) and time (seconds).

Events	Current World Record	Estimation Method	Tail Index	Endpoint (speed)	Endpoint (time)	95% Lower Limit (time)
100-m men	9.58	Moment	−0.19	37.85	9.51	9.21
		Likelihood	−0.18	37.95	9.48	9.21
100-m women	10.49	Moment	−0.18	34.85	10.33	9.88
		Likelihood	−0.20	34.62	10.40	10.13

current world record. The moment method yields a much lower estimate 10.33 seconds, 0.16 seconds lower than the current world record, a much bigger room for improvement. For the 95% lower confidence limit for women's 100 meters, our method gives 10.12 seconds, while the moment method has a much smaller estimate 9.88 seconds. We can further calculate a 99% upper confidence limit for the speed endpoint, $10.40 + 0.5606 \times 2.326 = 35.92$ (kilometers per hour) and thus 99% lower confidence limit of $360/35.92 = 10.02$ seconds for the time endpoint. If we think the 99% lower confidence limit as a possible true endpoint then, by comparing it with the current world record 10.49 seconds established almost thirty years ago, we may well expect that it will be a long way for female athletes to achieve a personal best time within 10.00 seconds, a time shorter than the 99% lower confidence limit for women's 100 meters ultimate world record.

Supplementary Materials

We have conducted some further simulation study to compare our new estimators 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). The comparison results can be found in Section S1 of the Supplement. Some details on the data application can be found in Section S2 of the Supplement. The proofs of the theorems in Section 2 are available in Section S3 of the Supplement.

Acknowledgment

We thank two reviewers and an associate editor for their helpful comments. Peng's research was supported by the Simons Foundation, Qi's research was supported by NSF grant DMS-1005345, and Wang's research was supported by NSFC Grant No. 11671021, NSFC Grant No. 11471222 and Foundation of Beijing Education Bureau Grant No. 201510028002.

References

Aarssen, K. and de Haan, L. (1994). On the maximal life span of humans. *Math. Popul. Stud.* **4**(4), 259–281.

Athreya, K. B. and Fukuchi, J. I. (1997). Confidence intervals for endpoints of a cdf via bootstrap. *J. Statist. Plann. Inference* **58**, 299–320.

Balakrishnan, N. and Cohen, A. C. (1991). *Order Statistics and Inference: Estimation Methods.* Academic Press.

Beirlant, J., Fraga Alves, M. I. and Gomes, M. I. (2016). Tail fitting for truncated and non-truncated Pareto-type distributions. *Extremes* **19**, 429–462.

Coles, S. G. and Dixon, M. J. (1999). Likelihood-Based Inference for Extreme Value Models. *Extremes* **2**, 5–23.

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

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

Drees, H., Ferreira, A. and de Haan, L. (2004). On maximum likelihood estimation of the extreme value index. *Ann. Appl. Probab.* **14**, 1179–1201.

Einmahl, J. and Magnus, J. (2008). Records in athletics through extreme-value theory. *J. Amer. Statist. Assoc.* **103**, 1382–1391.

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

Falk, M. (1995). Some best parameter estimates for distributions with finite endpoint. *Statistics* **27**, 115–125.

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

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

Girard, S., Guillou, A. and Stupfler, G. (2012a). Estimating an endpoint with high-order moments. *Test* **21**, 697–729.

Girard, S., Guillou, A. and Stupfler, G. (2012b). Estimating an endpoint with high order moments in the Weibull domain of attraction. *Statist. Probab. Lett.* **82**, 2136–2144.

Gomes, M., de Haan, L. and Peng, L. (2002). Semiparametric estimation of the second order parameter in statistics of extremes. *Extremes* **5**, 387–414.

Hall, P. (1982). On estimating the endpoint of a distribution. *Ann. Statist.* **10**, 556–568.

Hall, P. and Park, B. U. (2002). New methods for bias correction at endpoints and boundaries. *Ann. Statist.* **30**, 1460–1479.

Hall, P. and Wang, J. Z. (1999). Estimating the end-point of a probability distribution using minimum-distance methods. *Bernoulli* **5**, 177–189.

Hall, P. and Wang, J. Z. (2005). Bayesian likelihood methods for estimating the end point of a distribution. *J. R. Stat. Soc. Ser B Stat. Methodol.* **67**, 717–729.

Li, D. and Peng, L. (2009). Does bias reduction with external estimator of second order parameter work for endpoint? *J. Statist. Plann. Inference* **139**, 1937–1952.

Li, D., Peng, L. and Xu, X. (2011). Bias reduction for endpoint estimation. *Extremes* **14**, 393–412.

Li, D., Peng, L. and Qi, Y. (2011). Empirical likelihood confidence intervals for the endpoint of a distribution function. *Test* **20**, 353–366.

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

Loh, W. Y. (1984). Estimating an endpoint of a distribution with resampling methods. *Ann. Statist.* **12**, 1534–1550.

Miller, R. G. (1964). A trustworthy jackknife. *Ann. Math. Stat.* **35**, 1594–1605.

Pauli, F. and Coles, S. (2001). Penalized likelihood inference in extreme value analyses. *J. Appl. Stat.* **28**, 547–560.

Peng, L. and Qi, Y. (2009). Maximum likelihood estimation of extreme value index for irregular cases. *J. Statist. Plann. Inference* **139**, 3361–3376.

Robson, D. S. and Whitlock, J. H. (1964). Estimation of a truncation point. *Biometrika* **51**, 33–39.

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika* **72**, 67–90.

Smith, R. L. (1987). Estimating tails of probability distributions. *Ann. Statist.* **15**, 1174–1207.

Smith, R. L. and Weissman, I. (1985). Maximum likelihood estimation of the lower tail of a probability distribution. *J. Roy. Statist. Soc. Ser. B Stat. Methodol.* **47**, 285–298.

Woodroofe, M. (1974). Maximum likelihood estimation of translation parameter of truncated distribution II. *Ann. Statist.* **2**, 474–488.

Zhou, C. (2009). Existence and consistency of the maximum likelihood estimator for the extreme value index. *J. Multivariate Anal.* **100**, 794–815.

School of Mathematical Sciences, Capital Normal University, Beijing 100037, China.

E-mail: fang72_wang@cnu.edu.cn

Department of Risk Management and Insurance, Georgia State University, Atlanta, GA 30303, USA.

E-mail: lpeng@gsu.edu

Department of Mathematics and Statistics, University of Minnesota Duluth, 1117 University Drive, Duluth, MN 55812, USA.

E-mail: yqi@d.umn.edu

School of Science, Beijing Technology and Business University, Beijing 100048, China.

E-mail: xumeiping2006@163.com

(Received July 2016; accepted July 2017)

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.

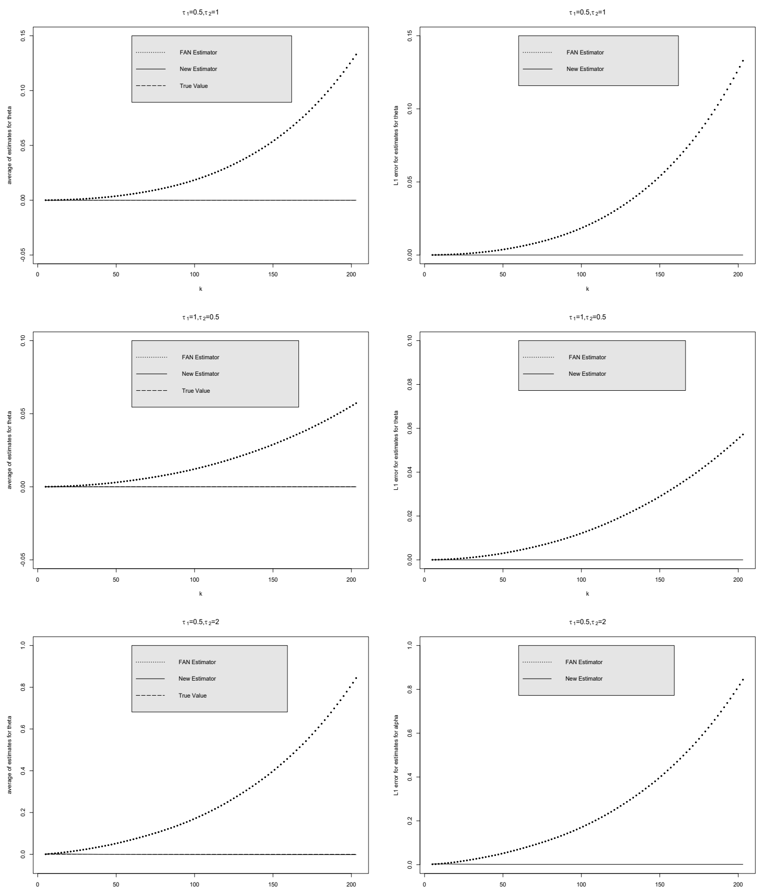

Figure 1: Estimated means (left) and estimated L_1 errors (right) for two endpoint estimators: New Estimator as the smallest solution to (2.17) and FAN Estimator defined in (S1.1). The samples are taken from distribution (3.26), where $\theta = 0$ and $\alpha = \tau_1 \tau_2$

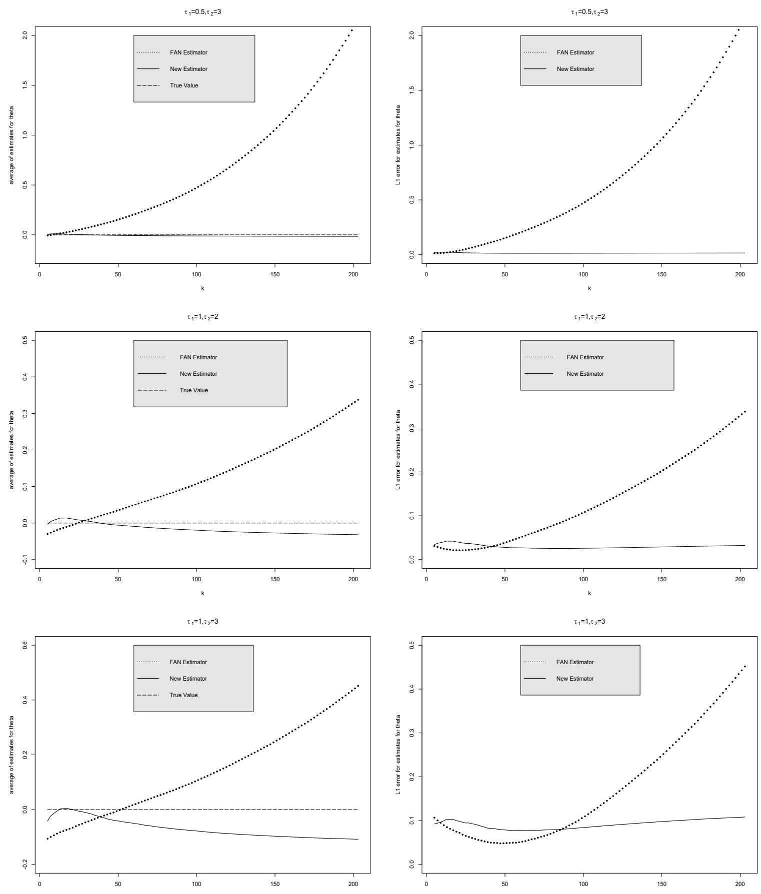

Figure 2: Estimated means (left) and estimated L_1 errors (right) for two endpoint estimators: New Estimator as the smallest solution to (2.17) and FAN Estimator defined in (S1.1). The samples are taken from distribution (3.26), where $\theta = 0$ and $\alpha = \tau_1 \tau_2$

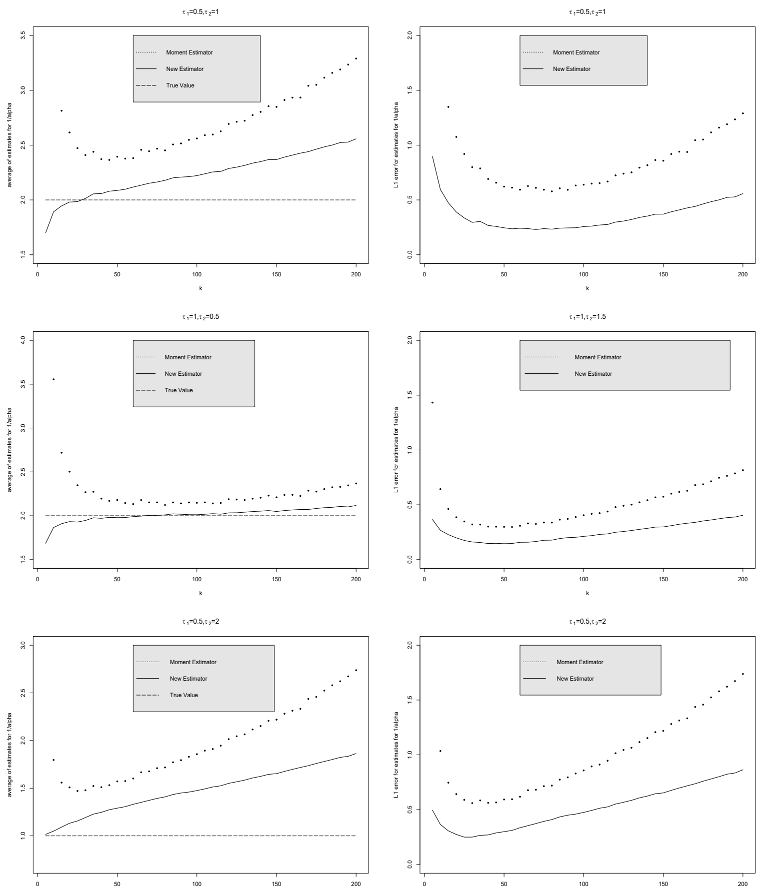

Figure 3: Estimated means (left) and estimated L_1 errors (right) for two estimators for α^{-1}: New Estimator $\tilde{\alpha}_N^{-1}$ defined in (2.16) and minus Moment Estimator $-\hat{\gamma}_M(k)$, where $\hat{\gamma}_M(k)$ is defined in (S1.2). The samples are taken from population $20+X$, where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

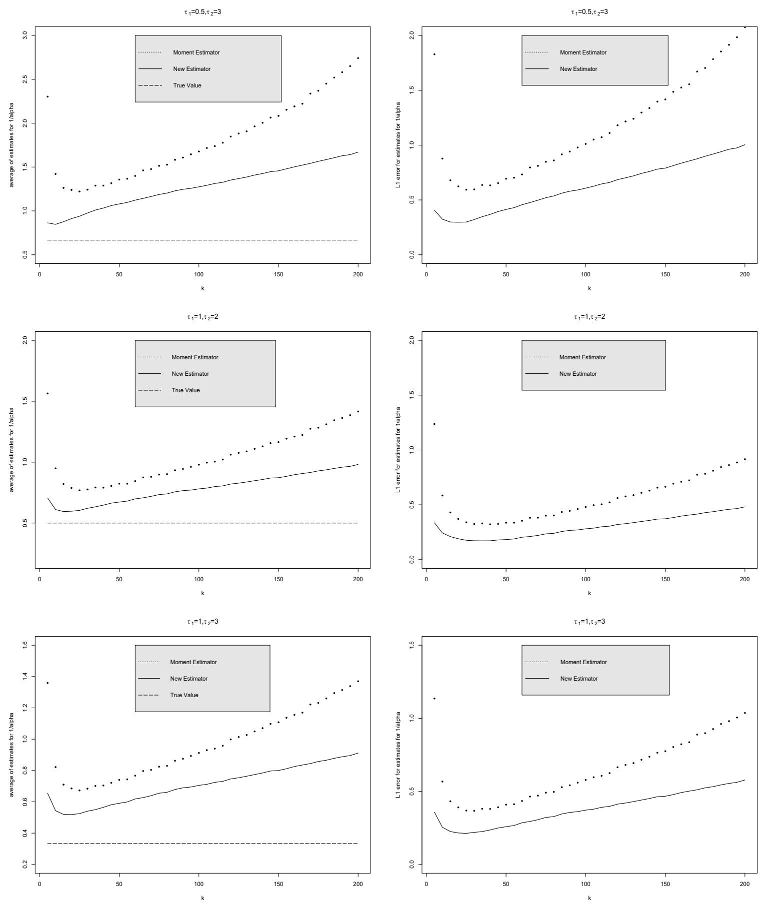

Figure 4: Estimated means (left) and estimated L_1 errors (right) for two estimators for α^{-1}: New Estimator $\tilde{\alpha}_N^{-1}$ defined in (2.16) and minus Moment Estimator $-\hat{\gamma}_M(k)$, where $\hat{\gamma}_M(k)$ is defined in (S1.2). The samples are taken from population $20+X$, where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

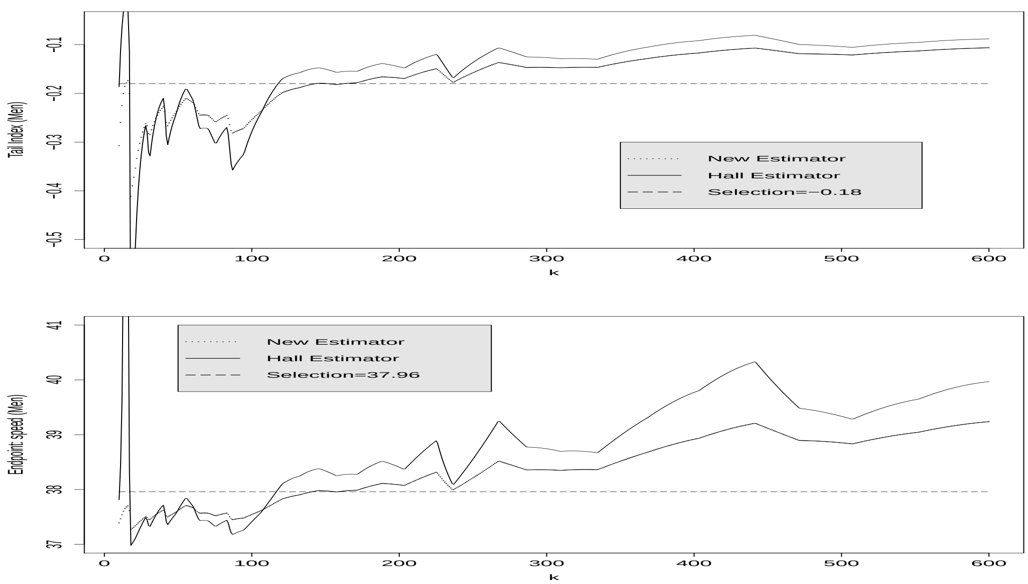

Figure 5: Our new likelihood estimates and the moment estimates for tail index $\gamma = -1/\alpha$ and the endpoint θ for speed (in km/h) for men's 100 meters.

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

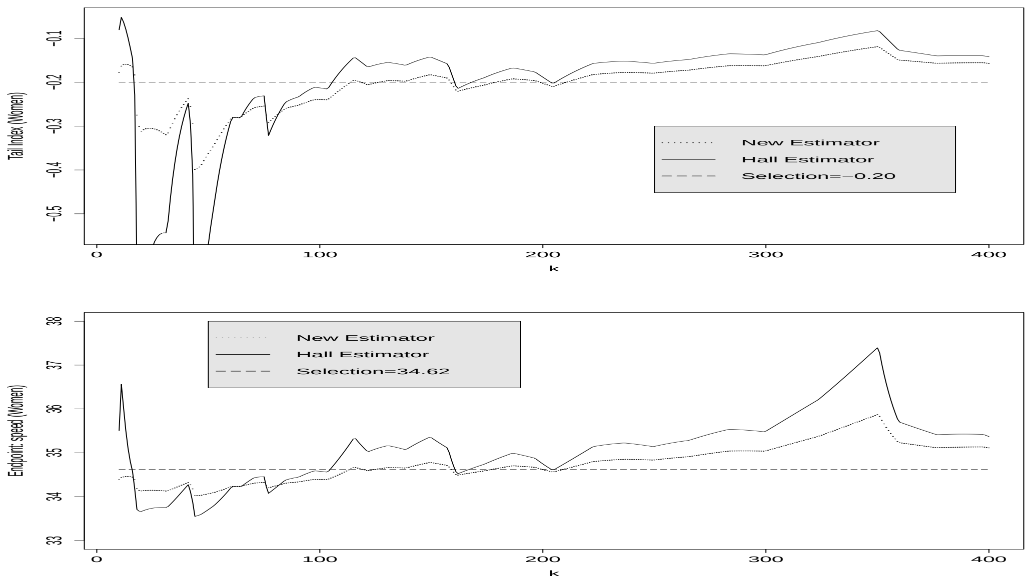

Figure 6: Our new likelihood estimates and the moment estimates for tail index $\gamma = -1/\alpha$ and the endpoint θ for speed (in km/h) for women's 100 meters.

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

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

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 $\log V_1^*$. □

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 $V^*_{k,k} = kv$,

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

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

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

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

$$G_{\lambda,v,x}^{(k)}(s) := P(\frac{1}{k^\lambda}(T_{\lambda,x}^{(k)} - \frac{k-1}{1-\lambda}) \leq s | V_{k,k}^* = kv) = P(\sum_{j=1}^{k-1} Z_j \leq s).$$

We can check that

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

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 $k \to \infty$

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

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

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

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 $R^{(k)}_{\lambda,x}$ and Q_k are asymptotically independent.

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

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

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

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

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. $\square$

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

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

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 $y < 0$

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

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).

<u>**Case 4:** $\alpha_0 = 1$.</u> 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

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

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

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

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

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

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 $v > 0$

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

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.