Statistica Sinica Preprint No: SS-2016-0336.R4	
Title	Maximum Penalized Likelihood Estimation For The Endpoint And Exponent Of A Distribution
Manuscript ID	SS-2016-0336.R4
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202016.0336

Complete List of Authors	Fang Wang Liang Peng Yongcheng Qi and Meiping Xu
Corresponding Author	Yongcheng Qi
E-mail	yqi@d.umn.edu

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

Statistica Sinica

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*

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

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

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), and 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.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

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) and Drees, Ferreira and de Haan (2004) for the regular case, and Smith (1985, 1987), Smith and Weissman (1985), Woodroofe (1974), Zhou (2009), and 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), and 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), and 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 $X_1, \cdots, X_n$ 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 $X_1, \cdots, X_n$, and let $k = k_n$ be

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

a sequence of integers satisfying $k/n \to 0$ as $n \to \infty$. When our (2.6) 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 $\alpha > 1$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

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) and Pauli and Coles (2001)), but they are problematic for the irregular case (see Hall (1982) and 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 $X_{n,n-k+1}, \cdots, X_{n,n}$ 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 $X_{n,n-k}, \cdots, X_{n,n}$, up to a constant scale, is given by

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

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

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

(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,

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

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 $X_1, \cdots, X_n$ 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 $X_1, \cdots, X_n$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

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 $X_1, \cdots, X_n$ 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 $p(\theta, \alpha, X_{n,n-k}, \cdots, X_{n,n})$ to be a general penalization function such that

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

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

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

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

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

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's (1982) 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 (3.23) and $g(\theta)$ is defined in (2.17). 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).

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

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

$$h(\theta) := \frac{\theta - X_{n,n-k}}{\theta - X_{n,n}} + (1 - \frac{1}{\alpha_0}) \sum_{j=1}^{k-1} \frac{\theta - X_{n,n-k}}{\theta - X_{n,n-k+j}} - \frac{2}{\alpha_0} - k = 0. \tag{2.3}$$

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

$$h(X_{n,n}+) = \infty,\ h(\infty) = -\frac{k+1}{\alpha_0} < 0 \text{ and } h(\theta) \text{ is continuous,} \tag{2.4}$$

there exists at least one root to (2.3). We have that

$$h(\theta) = \frac{X_{n,n} - X_{n,n-k}}{\theta - X_{n,n}} + \frac{\alpha_0 - 1}{\alpha_0} \sum_{j=1}^{k-1} \frac{X_{n,n-k+j} - X_{n,n-k}}{\theta - X_{n,n-k+j}} - \frac{k+1}{\alpha_0} \tag{2.5}$$

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

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

$h'(\theta) < 0$ 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 (2.4) when $X_{n,n} > X_{n,n-1}$. In conclusion, there exists a unique solution to (2.3) 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 $\hat{\theta} \overset{a.s.}{\to} \theta_0$ 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

$$\lim_{t\to\infty} \frac{\frac{U(tx)-U(t)}{a(t)} - \frac{x^{\gamma_0}-1}{\gamma_0}}{A(t)} = H_{\gamma_0,\rho}(x) := \frac{1}{\rho}\Big(\frac{x^{\gamma_0+\rho}-1}{\gamma_0+\rho} - \frac{x^{\gamma_0}-1}{\gamma_0}\Big), \qquad (2.6)$$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

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 (2.6) 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

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

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

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

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

with the characteristic function $f_{\lambda,v,x}$ given by

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

Theorem 2. *Assume (2.6) holds and $k = k_n$ satisfies one of the following conditions:*

$$k \to \infty, \quad k/n \to 0, \quad k^{1/2}A(n/k) \to 0 \quad \textit{if} \quad \alpha_0 > 2; \tag{2.7}$$

$$k \to \infty, \quad k/n \to 0, \quad k^{1/2}(\log k)^{-1/2}A(n/k) \to 0 \quad \textit{if} \quad \alpha_0 = 2; \tag{2.8}$$

$$k \to \infty, \quad k/n \to 0, \quad k^{1+\gamma_0}A(n/k) \to 0 \quad \textit{if} \quad \alpha_0 \in (1, 2); \tag{2.9}$$

$$k \to \infty, \quad k/n \to 0 \quad \textit{if} \quad \alpha_0 \in (0, 1]. \tag{2.10}$$

Then we have

$$n^{-\gamma_0}k^{1/2+\gamma_0}c^{-\gamma_0}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, (1 + 2\gamma_0)) \quad \textit{if} \quad \alpha_0 > 2; \tag{2.11}$$

$$(n \log k)^{1/2}c(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, 1) \quad \textit{if} \quad \alpha_0 = 2; \tag{2.12}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

$$n^{-\gamma_0}c^{-\gamma_0}(\hat{\theta}-\theta_0) \xrightarrow{d} \Lambda_{-\gamma_0} \quad \textit{if} \quad \alpha_0 \in (0,2), \quad \alpha_0 \neq 1; \tag{2.13}$$

$$n^{-\gamma_0}c^{-\gamma_0}(\hat{\theta}-\theta_0) \xrightarrow{d} 1-Z \quad \textit{if} \quad \alpha_0 = 1, \tag{2.14}$$

where Z is a standard exponential random variable.

Remark 1. (**a**) From (2.5), $\hat{\theta} = X_{n,n} + (k+1)^{-1}(X_{n,n} - X_{n,n-k})$ 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) and Robson and Whitlock (1964). The two estimators for θ in Miller (1964) and 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

$$\sigma^2_{Miller}(n) := E(\hat{\theta}_{Miller} - \theta)^2 = \frac{2\theta^2(n^2-n+1)}{n^2(n+1)(n+2)}, \tag{2.15}$$

$$\sigma^2_{RW}(n) := E(\hat{\theta}_{RW} - \theta)^2 = \frac{2\theta^2}{(n+1)(n+2)},$$

$$\sigma^2_{N}(n,k) := E(\hat{\theta} - \theta)^2 = \frac{k+2}{k+1}\frac{\theta^2}{(n+1)(n+2)}.$$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

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)); (2.15) 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 (2.8)–(2.10) are weaker than (2.7). The condition (2.10) imposes the weakest condition on k, and the second-order convergence rate A is not involved, although the second-order regular variation condition (2.6) is assumed. Some intuitive explanations are as follows. We have $X_{n,n} - \theta_0 = O_p(n^{-1/\alpha})$ under (2.6) 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.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

(**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 the estimator $(\tilde{\theta}, \tilde{c}, \tilde{\alpha})$ of (θ, c, α). By solving score equations, we have $\tilde{c} = ((k+1)/n)(\tilde{\theta} - X_{n,n-k})^{-\tilde{\alpha}}$,

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

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

$$\begin{aligned} g(\theta) \;:=\; & \sum_{j=1}^{k}\Big(\frac{\theta - X_{n,n-k}}{\theta - X_{n,n-k+j}} - 1\Big) \\ & -\frac{1}{k}\Big(\sum_{j=1}^{k}\log\frac{\theta - X_{n,n-k}}{\theta - X_{n,n-k+j}}\Big)\Big(2 + \sum_{j=1}^{k-1}\frac{\theta - X_{n,n-k}}{\theta - X_{n,n-k+j}}\Big) = 0 \end{aligned} \tag{2.17}$$

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

$$\alpha_n^{-1} = \frac{1}{k}\sum_{j=1}^{k} \log \frac{\theta - X_{n,n-k}}{\theta - X_{n,n-k+j}}, \tag{2.18}$$

see, e.g., Falk (1995). The estimator of α^{-1} given by (2.16) is coincident

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

with (2.18) 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 (2.17) 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* (2.6) *holds and*

$$k \to \infty, \quad k/n \to 0, \quad k^{1/2}A(n/k) \to 0 \quad \textit{as} \quad n \to \infty. \tag{2.19}$$

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

$$\big(n^{-\gamma_0}k^{1/2+\gamma_0}c^{-\gamma_0}(\tilde{\theta} - \theta_0), k^{1/2}(\tilde{\alpha}^{-1} - \alpha_0^{-1})\big) \xrightarrow{d} N(0, \Sigma), \tag{2.20}$$

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

$$k^{1/2}(\tilde{\alpha}^{-1} - \alpha_0^{-1}) \xrightarrow{d} N(0, \gamma_0^2), \tag{2.21}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

$\tilde{\theta}$ 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 (2.19) is required this time for all cases; it is needed only for (2.21).

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

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 (2.7) implies conditions (2.8) and (2.9), 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 (2.7) can be employed for Theorems 2 and 3.

First, we show that there always exists a sequence of integers $\{\bar{k}_n\}$ satisfying (2.7). To see that $\max_{1 \leq k \leq \bar{k}_n} \sqrt{k}|A(n/k)| \to 0$ as $n \to \infty$, let

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

$B(t) = \sup_{s\geq t}|A(s)|$. 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

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

A choice of k satisfying (2.7) can be obtained via estimating the second order regular variation parameter ρ when (2.6) holds with some $\rho < 0$. Since $|A(t)|$ is regularly varying with exponent ρ, we can apply Potter's bound and prove that (2.7) holds for any sequence of positive integers $k = k_n$ with $k_n \sim cn^{\beta}$ for $c > 0$, and $\beta \in (0, \frac{-2\rho}{1-2\rho})$. 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 (2.16) and (2.17) as $\tilde{\alpha}(k)$ and $\tilde{\theta}(k)$. When (2.6) 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

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

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 (3.26) 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 1000 each 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 $1 \leq k \leq 999$, and all estimates are between -1.944×10^{-07} and 1.747×10^{-07}. Therefore, choosing any large k results in a satisfactory estimate.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

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 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) and 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

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

solution of

$$\sum_{j=1}^{k}\left(\frac{\theta - X_{n,n-k}}{\theta - X_{n,n-k+j}} - 1\right) - \frac{k+1}{\alpha_0 - 1} = 0, \tag{3.22}$$

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

$$m(\theta) := \frac{k+1}{\sum_{j=1}^{k} \log \frac{\theta - X_{n,n-k}}{\theta - X_{n,n-k+j}}} - \frac{k+1}{\sum_{j=1}^{k} \frac{X_{n,n-k+j} - X_{n,n-k}}{\theta - X_{n,n-k+j}}} - 1 = 0 \tag{3.23}$$

and the estimator for α^{-1} is

$$\tilde{\alpha}_H^{-1} = \frac{1}{k+1}\sum_{j=1}^{k} \log \frac{\tilde{\theta}_H - X_{n,n-k}}{\tilde{\theta}_H - X_{n,n-k+j}}. \tag{3.24}$$

To make a fair comparison, we chose the solution of (3.23) 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 (3.25).

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

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

Hill estimator,

$$\tilde{\alpha}_{NH}^{-1} = \frac{1}{k}\sum_{j=1}^{k-1} \log \frac{X_{n,n} - X_{n,n-k}}{X_{n,n} - X_{n,n-k+j}} \tag{3.25}$$

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 model as if θ were known (see, e.g., Falk (1995)): (2.21) 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 $N = 1000$ random samples of size n with n set at 100, 200, 500 and 1000, 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}.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

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 (3.25), 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 are 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 report 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,

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

so one could employ 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 $n = 1000$ 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 $N = 1000$ estimates and 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 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.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

3.1 Comparisons with the conventional likelihood method

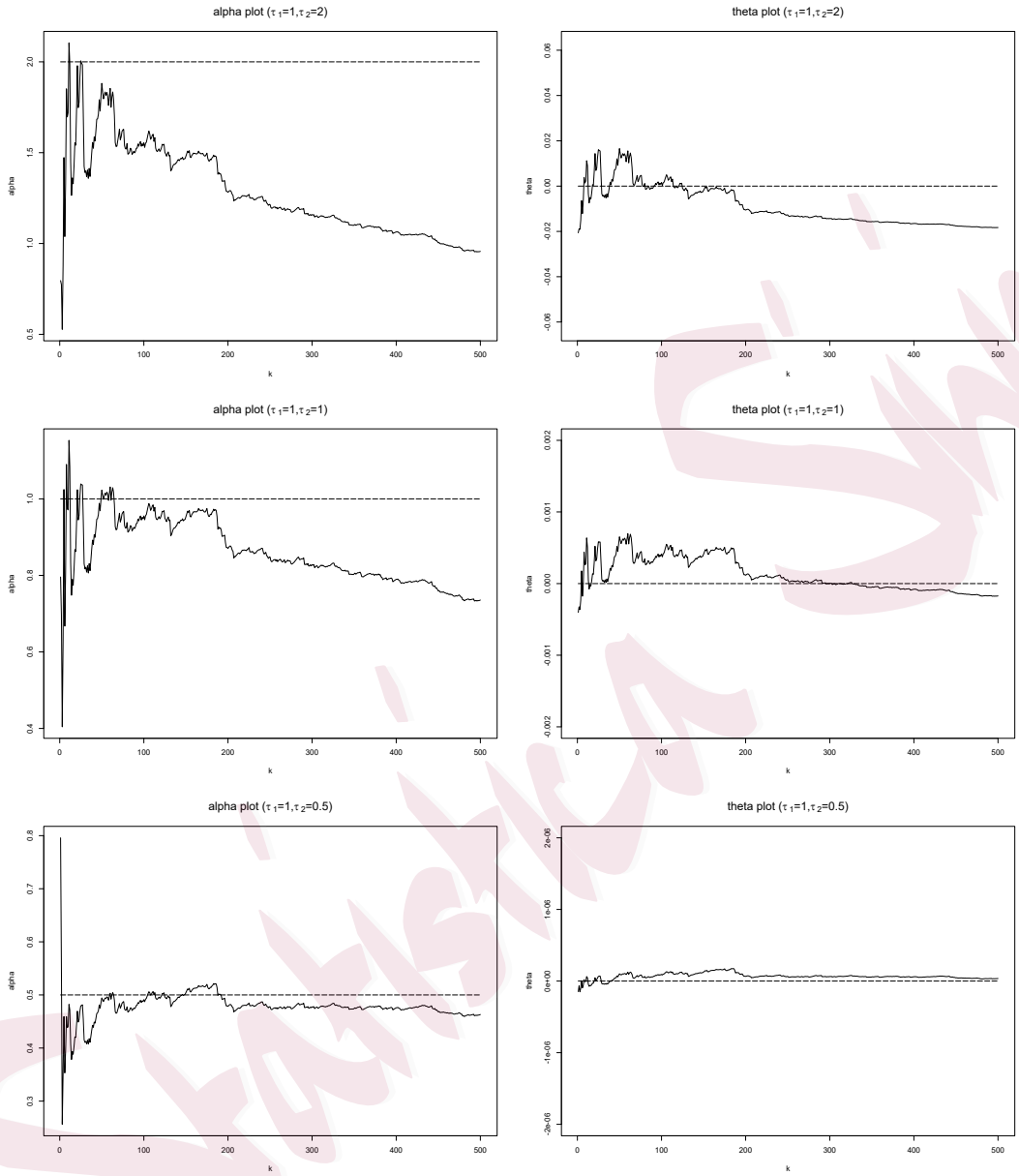

Figure 1: Plots of estimates for α and θ based a random sample of size 1000 from distribution (3.26) with parameters with $(\tau_1, \tau_2) = (1, 2)$, $(1, 1)$, $(1, 0.5)$, respectively.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

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 (3.25).

			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	1000	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	1000	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)

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

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	1000	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	1000	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

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

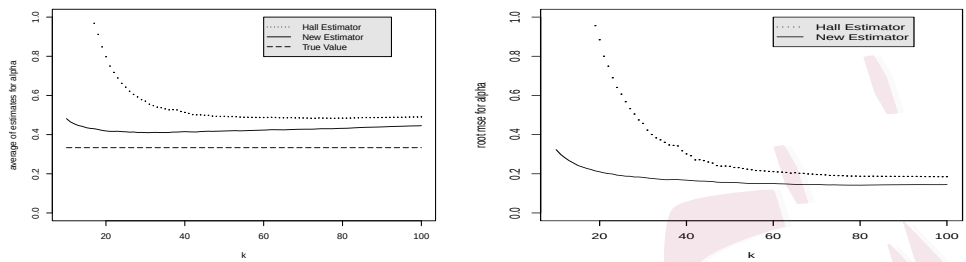

Figure 2: Estimated biases (left) and root mean-squared errors (right) of the new estimator and Hall's estimator for α^{-1} with sample size $n = 1000$.

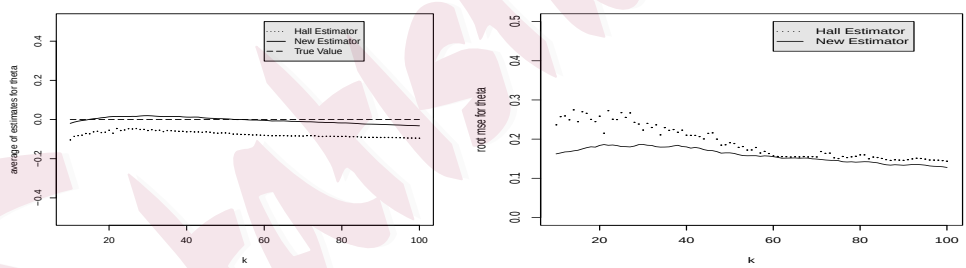

Figure 3: Estimated biases (left) and root mean-squared errors (right) of the new estimator and Hall's estimator for θ with sample size $n = 1000$.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

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 a different role.

Girard, Guillou and Stupfler (2012b) compared the performance of their estimator with the maximum value estimator ($X_{n,n}$) and the moment es-

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

timator in terms of the optimal mean absolute errors, under two types of distributions. The first one was

$$1 - F(x) = \left(1 + (-x)^{-\tau_1}\right)^{-\tau_2}, \quad x < 0 \tag{3.26}$$

with $\tau_1, \tau_2 > 0$. A random variable X with distribution (3.26) 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

$$1 - F(x) = \int_{\log(1-1/x)}^{\infty} \lambda^2 t e^{-\lambda t} dt, \quad x < 0 \tag{3.27}$$

with $\lambda > 0$. A random variable X with distribution (3.27) can be written as $X = -1/(e^Y - 1)$, where Y has a Gamma$(2, \lambda)$ distribution.

Models (3.26) and (3.27) have a right endpoint $\theta = 0$. Choose one distribution from (3.26) or (3.27). 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 (3.26) and (3.27), generated $N = 1000$ replicates of random samples of size $n = 500$ each, chose the same values for p and

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

a, and used the same choices for parameters in the two distributions. We computed our estimate for θ with choices $k \in \{5, 10, 15, \cdots, 300\}$ 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;

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

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) and 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 $y-0.005+0.01(2j-1)/(2m)$, $j = 1, \cdots, m$. 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.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

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 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

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

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

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.

Acknowledgements

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 L. de Haan (1994). On the maximal life span of humans. *Math. Popul. Stud.* 4 (4), pp. 259-281.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

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

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

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

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

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.

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

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

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

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

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

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.

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

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

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

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

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

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

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

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

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

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

Li, D., L. Peng and Y. Qi (2011). Empirical likelihood confidence intervals for the endpoint of a distribution function. *Test* 20, pp. 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, pp. 1534–1550.

Miller, R.G. (1964). A trustworthy jackknife. *Annals of Mathematical Statistics* 35, pp. 1594–1605.

Pauli, F. and S. Coles (2001). Penalized likelihood inference in extreme value analyses. *Journal of Applied Statistics* 28, pp. 547–560.

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

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

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

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

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0336

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

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

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

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

E-mail: xumeiping2006@163.com