Test
DOI 10.1007/s11749-010-0204-4

ORIGINAL PAPER

Empirical likelihood confidence intervals for the endpoint of a distribution function

Deyuan Li · Liang Peng · Yongcheng Qi

Received: 24 January 2010 / Accepted: 14 June 2010

Abstract Estimating the endpoint of a distribution function is of interest in product analysis and predicting the maximum lifetime of an item. In this paper, we propose an empirical likelihood method to construct a confidence interval for the endpoint. A simulation study shows the proposed confidence interval has better coverage accuracy than the normal approximation method, and bootstrap calibration improves the accuracy.

Keywords Confidence interval · Coverage probability · Empirical likelihood method · Endpoint

Mathematics Subject Classification (2000) 62G32 · 62G15

1 Introduction

Suppose $X_1, \ldots, X_n$ are independent and identically distributed random variables with distribution function F which has a finite right endpoint θ and satisfies

$$1 - F(x) = c(\theta - x)^{\alpha} + o\big((\theta - x)^{\alpha}\big) \tag{1.1}$$

Communicated by Domingo Morales.

D. Li
Department of Statistics, School of Management, Fudan University, Shanghai, 200433, P.R. China
e-mail: deyuanli@fudan.edu.cn

L. Peng (✉)
School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332-0160, USA
e-mail: peng@math.gatech.edu

Y. Qi
Department of Mathematics and Statistics, University of Minnesota–Duluth, 1117 University Drive, Duluth, MN 55812, USA
e-mail: yqi@d.umn.edu

Published online: 03 July 2010

as $x \uparrow \theta$, where $c > 0$ and $\alpha > 0$. Note that condition (1.1) implies that F is in the domain of attraction of an extreme value distribution with a negative extreme value index; see Chap. 1 of De Haan and Ferreira (2006) for details. This condition has been employed in Hall (1982), and parameters c, α and θ specify the first-order approximation to the upper tail of the distribution function F. Based on this approximation, an approximate likelihood can be formulated and maximum likelihood estimates for parameters c, α, θ can be obtained as in Sect. 2.

Estimating the endpoint θ is of interest in production analysis and predicting the maximum lifetime of an item. The cases of $\alpha > 2$ and $0 < \alpha \le 2$ are respectively called regular case and irregular case in the literature. Estimating the endpoint for the regular case includes Athreya and Fukuchi (1997), Dekker et al. (1989), Hall (1982), Hall and Wang (1999), Li and Peng (2009a, 2009b) and Loh (1984). For the irregular case, studies include Aarssen and de Haan (1994), Falk (1994), Smith (1985, 1987), Woodroofe (1974), Zhou (2009), Peng and Qi (2009).

In this paper, we concern with the construction of confidence intervals for θ under the setup of regular case. Based on the asymptotic limit of some estimator for the endpoint, a normal approximation confidence interval can be obtained via estimating the asymptotic variance. In order to avoid estimating the asymptotic variance, bootstrap method can be employed (see Athreya and Fukuchi 1997 and Li and Peng 2009a, 2009b). Here, we investigate the possibility of applying the empirical likelihood method. Like bootstrap and jackknife methods, empirical likelihood method is a powerful nonparametric method in interval estimation and hypothesis test. Recently, empirical likelihood method has been applied to extremes, see Einmahl and Segers (2009) and Qi (2008). For overview of empirical likelihood methods, we refer to Owen (2001). Some important features of empirical likelihood method include Bartlett correction and self-normalization, i.e., without estimating the asymptotic variance explicitly. A common way to formulate the empirical likelihood function is via estimating equations as in Qin and Lawless (1994). In this paper, empirical likelihood confidence intervals for the endpoint are constructed via estimating equations. And bootstrap calibration is proposed to improve the coverage accuracy. We organize this paper as follows. Methodologies and results are given in Sect. 2. Section 3 presents some finite sample investigation and real data analysis. Proofs are put in Sect. 4.

2 Methodology and results

Let $X_{n,1} \le \cdots \le X_{n,n}$ denote the order statistics of $X_1, \ldots, X_n$. Model (1.1) is a semiparametric one since the remainder is not specified, and all the parameters of interest reveal the upper tail properties of the distribution F. It is known in the literature that only a few largest observations should be employed in the estimation. For this purpose, we define $\delta_i = I(X_i > u)$ for $i = 1, \ldots, n$, where u is a high threshold. When $\delta_i = 1$, we can approximate the distribution function of X_i by $1 - c(\theta - x)^{\alpha}$. When $\delta_i = 0$, X_i is considered to be censored so that it does not account in the estimation but the only information that the observation is below the threshold u is used in formulating a likelihood function. This results in an approximate censored likelihood function for $\{(X_i, \delta_i)\}_{i=1}^n$ as

$$\prod_{i=1}^{n}\left\{c\alpha(\theta - X_i)^{\alpha-1}\right\}^{\delta_i}\left\{1 - c(\theta - u)^{\alpha}\right\}^{1-\delta_i}. \tag{2.1}$$

Next we take the threshold $u = X_{n,n-k}$ for some k such that $k = k(n) \to \infty$ and $k/n \to 0$. Then (2.1) becomes

$$\prod_{i=1}^{k}\left\{c\alpha(\theta - X_{n,n-i+1})^{\alpha-1}\right\}\left\{1 - c(\theta - X_{n,n-k})^{\alpha}\right\}^{n-k},$$

which results in the score equations

$$\sum_{i=1}^{k}\tilde{\boldsymbol{Y}}_i(c, \alpha, \theta) = 0,$$

where $\tilde{\boldsymbol{Y}}_i(c, \alpha, \theta) = (\tilde{Y}_{i1}(c, \alpha, \theta),\ \tilde{Y}_{i2}(c, \alpha, \theta),\ \tilde{Y}_{i3}(c, \alpha, \theta))^T$ with

$$\tilde{Y}_{i1}(c, \alpha, \theta) = \frac{1}{c} - \frac{n-k}{k}\frac{(\theta - X_{n,n-k})^{\alpha}}{1 - c(\theta - X_{n,n-k})^{\alpha}},$$

$$\tilde{Y}_{i2}(c, \alpha, \theta) = \frac{1}{\alpha} + \log(\theta - X_{n,n-i+1}) - \frac{n-k}{k}\frac{c(\theta - X_{n,n-k})^{\alpha}\log(\theta - X_{n,n-k})}{1 - c(\theta - X_{n,n-k})^{\alpha}},$$

$$\tilde{Y}_{i3}(c, \alpha, \theta) = \frac{\alpha - 1}{\theta - X_{n,n-i+1}} - \frac{n-k}{k}\frac{c\alpha(\theta - X_{n,n-k})^{\alpha-1}}{1 - c(\theta - X_{n,n-k})^{\alpha}}.$$

Like Qin and Lawless (1994), we define the empirical likelihood function as

$$\tilde{R}_n(c, \alpha, \theta) = \sup\left\{\prod_{i=1}^{k}(kp_i) : p_i \ge 0,\ \sum_{i=1}^{k}p_i = 1,\ \sum_{i=1}^{k}p_i\tilde{\boldsymbol{Y}}_i(c, \alpha, \theta) = 0\right\}.$$

Note that $\sum_{i=1}^{k} p_i\tilde{\boldsymbol{Y}}_i(c, \alpha, \theta) = 0$ is equivalent to

$$\begin{cases} \sum_{i=1}^{k} p_i\tilde{Y}_{i1}(c, \alpha, \theta) = 0, \\ \sum_{i=1}^{k} p_i\{\tilde{Y}_{i2}(c, \alpha, \theta)/\log(\theta - X_{n,n-k}) - c\tilde{Y}_{i1}(c, \alpha, \theta)\} = 0, \\ \sum_{i=1}^{k} p_i\{\frac{\theta - X_{n,n-k}}{\alpha}\tilde{Y}_{i3}(c, \alpha, \theta) - c\tilde{Y}_{i1}(c, \alpha, \theta)\} = 0, \end{cases}$$

which implies that

$$\begin{cases} \sum_{i=1}^{k} p_i\{\frac{\alpha^{-1} + \log(\theta - X_{n,n-i+1})}{\log(\theta - X_{n,n-k})} - 1\} = 0, \\ \sum_{i=1}^{k} p_i\{\frac{\alpha-1}{\alpha}\frac{\theta - X_{n,n-k}}{\theta - X_{n,n-i+1}} - 1\} = 0, \end{cases}$$

i.e.,

$$\begin{cases} \sum_{i=1}^{k} p_i\{\log(\frac{\theta - X_{n,n-i+1}}{\theta - X_{n,n-k}}) + \frac{1}{\alpha}\} = 0, \\ \sum_{i=1}^{k} p_i\{\frac{\theta - X_{n,n-k}}{\theta - X_{n,n-i+1}} - \frac{\alpha}{\alpha-1}\} = 0. \end{cases}$$

So, $\tilde{R}_n(c, \alpha, \theta)$ can be reduced to

$$R_n(\alpha, \theta) = \sup\left\{\prod_{i=1}^{k}(kp_i) : p_i \geq 0,\ \sum_{i=1}^{k} p_i = 1,\ \sum_{i=1}^{k} p_i \mathbf{Y}_i(c, \alpha, \theta) = 0\right\},$$

where $\mathbf{Y}_i(\alpha, \theta) = (Y_{i1}(\alpha, \theta), Y_{i2}(\alpha, \theta))^T$,

$$Y_{i1}(\alpha, \theta) = \log \frac{\theta - X_{n,n-i+1}}{\theta - X_{n,n-k}} + \frac{1}{\alpha} \quad \text{and} \quad Y_{i2}(\alpha, \theta) = \frac{\theta - X_{n,n-k}}{\theta - X_{n,n-i+1}} - \frac{\alpha}{\alpha - 1}.$$

By the standard Lagrange multiplier technique, we obtain the log empirical likelihood ratio as

$$l(\theta, \alpha) = -2 \log R_n(\alpha, \theta) = 2 \sum_{i=1}^{k} \log\left(1 + \boldsymbol{\lambda}^T \mathbf{Y}_i(\alpha, \theta)\right), \tag{2.2}$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^T$ satisfies

$$\sum_{i=1}^{k} \frac{\mathbf{Y}_i(\alpha, \theta)}{1 + \boldsymbol{\lambda}^T \mathbf{Y}_i(\alpha, \theta)} = 0. \tag{2.3}$$

Throughout we denote the true value of (θ, α) by (θ_0, α_0). Since we are interested in a confidence interval for θ, we consider the profile log empirical likelihood ratio $l(\theta, \hat{\alpha}(\theta))$, where $\hat{\alpha}(\theta) = \text{argmin}_{\alpha > 2} l(\theta, \alpha)$.

In order to derive the asymptotic limit of the above profile log empirical likelihood ratio, we assume the following second-order condition: there exist functions $a(t) > 0$ and $A(t) \to 0$ such that

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

where $U(t)$ is the inverse function of $F(1 - 1/t)$, $\gamma = -1/\alpha \in (-1/2, 0)$ and $\rho < 0$. We refer to De Haan and Ferreira (2006) for details on the above second-order regular variation. Under the above condition, we follow the procedure in Qin and Lawless (1994) to first show that there exists $\hat{\alpha}(\theta_0)$ with a certain rate of convergence and then we show the convergence of $l(\theta_0, \hat{\alpha}(\theta_0))$; see the following Proposition 1 and Theorem 1 for details. Note that the results in Qin and Lawless (1994) are not directly applicable to our setting since $\{(\theta - X_{n,n-i+1})/(\theta - X_{n,n-k})\}_{i=1}^k$ is not an independent sequence.

Proposition 1 *Assume* (2.4) *holds and* $k = k(n)$ *satisfies*

$$k/\log n \to \infty, \qquad k/n \to 0, \qquad \sqrt{k}A(n/k) \to 0 \quad \text{as } n \to \infty. \tag{2.5}$$

Then, with probability one, $l(\theta_0, \alpha)$ *attains its minimum value at some point* $\hat{\alpha}(\theta_0)$ *in the interior of the ball* $|\alpha - \alpha_0| \leq k^{-\delta_0}$ *with* $\delta_0 = \max\{\frac{1}{3}, \frac{1}{\alpha_0} + \frac{2^{-1} + \alpha_0^{-1}}{2}\}$, *and* $\hat{\alpha}(\theta_0)$

and $\hat{\boldsymbol{\lambda}} = \boldsymbol{\lambda}(\theta_0, \hat{\alpha}(\theta_0))$ *satisfy*

$$Q_{1n}\big(\hat{\alpha}(\theta_0), \hat{\boldsymbol{\lambda}}\big) = 0 \quad \textit{and} \quad Q_{2n}\big(\hat{\alpha}(\theta_0), \hat{\boldsymbol{\lambda}}\big) = 0,$$

where

$$Q_{1n}(\alpha, \boldsymbol{\lambda}) = \frac{1}{k}\sum_{i=1}^{k} \frac{\mathbf{Y}_i(\alpha, \theta_0)}{1 + \boldsymbol{\lambda}^T \mathbf{Y}_i(\alpha, \theta_0)}$$

and

$$Q_{2n}(\alpha, \boldsymbol{\lambda}) = \frac{1}{k}\sum_{i=1}^{k} \frac{1}{1 + \boldsymbol{\lambda}^T \mathbf{Y}_i(\alpha, \theta_0)} \left\{ \frac{d}{d\alpha} \mathbf{Y}_i(\alpha, \theta_0) \right\}^T \boldsymbol{\lambda}.$$

Theorem 1 *Under the conditions in Proposition 1, we have* $l(\theta_0, \hat{\alpha}(\theta_0)) \stackrel{d}{\to} \chi^2(1)$ *as* $n \to \infty$.

Based on the above theorem, a confidence interval for θ_0 with level d is

$$I_d^E = \big\{\bar{\theta} : l\big(\bar{\theta}, \hat{\alpha}(\bar{\theta})\big) \le \chi^2_{d,1}\big\},$$

where $\chi^2_{d,1}$ is the dth quantile of $\chi^2(1)$.

Remark 1 The above proposed empirical likelihood method does not apply to the irregular case since the asymptotic limit of the maximum likelihood estimator for the endpoint is a normal distribution when $0 < \alpha < 2$ (see Woodroofe 1974; Peng and Qi 2009). Therefore, Wilks theorem cannot be expected for the empirical likelihood method based on score equations.

Selection of k is of importance. Theoretically, one needs to choose k to minimize the coverage error. Unfortunately, how to derive the coverage expansion remains unknown. On the other hand, in general k is much smaller than the sample size n in analyzing extremes, which implies that calibration for the proposed empirical likelihood method is quite practically useful. Indeed the simulation in next section shows that the calibration makes the choice of k much less sensitive. We refer to Owen (2001) for details on calibration of empirical likelihood method. Here we propose the following bootstrap calibration method.

Draw a random sample of size n from $X_1, \ldots, X_n$ with replacement. Based on this resampling, we calculate the empirical likelihood ratio in (2.2) with θ being replaced by $\hat{\theta}$, where $\hat{\theta}$ is the solution to

$$\frac{1}{k+1}\sum_{i=1}^{k} \frac{\theta - X_{n,n-k}}{\theta - X_{n,n-i+1}} \left\{ \frac{1}{k+1}\sum_{i=1}^{k} \log \frac{\theta - X_{n,n-i+1}}{\theta - X_{n,n-k}} + 1 \right\} = 1. \tag{2.6}$$

Let us denote this bootstrapped empirical likelihood ratio as $l^*(\hat{\theta}, \alpha)$.

Note that (2.6) is the score equation for θ after parameters c and α are canceled out, and thus $\hat{\theta}$ is the maximum likelihood estimate for θ in Hall (1982). Moreover,

$\hat{\theta}$ and the maximum empirical likelihood estimate, $\arg\min_\theta l(\theta, \hat{\alpha}(\theta))$, are the same since the empirical likelihood function is formulated via score equations.

Next we profile the bootstrapped empirical likelihood ratio to obtain $\hat{\alpha}^*(\theta) = \arg\min_{\alpha\geq 2} l^*(\hat{\theta}, \alpha)$. Hence we obtain the bootstrapped profile empirical likelihood ratio $l^*(\hat{\theta}, \hat{\alpha}^*(\hat{\theta}))$. Repeat the above procedure M times so that we obtain M bootstrapped profile empirical likelihood ratios. Let d^* denote the $[Md]$th largest value of these M bootstrapped empirical likelihood ratios. Then a bootstrap calibration confidence interval is

$$I_d^{BC} = \{\bar{\theta} : l(\bar{\theta}, \hat{\alpha}(\bar{\theta})) \leq d^*\}.$$

3 Simulation and data analysis

In this section, we first investigate the finite sample behavior of the proposed empirical likelihood confidence interval I_d^E and the bootstrap calibration confidence interval I_d^{BC} in terms of coverage accuracy, and compare with the normal approximation confidence interval and percentile-t bootstrap confidence interval given below.

Define

$$\hat{\gamma} = \frac{1}{k+1}\sum_{i=1}^{k}\log\frac{\hat{\theta} - X_{n,n-i+1}}{\hat{\theta} - X_{n,n-k}}.$$

Then $\hat{\gamma}$ and $\hat{\theta}$ defined in Sect. 2 are estimators for $\gamma = -1/\alpha$ and θ proposed by Hall (1982). Put

$$\hat{\sigma} = X_{n,n-k}\left\{\frac{1}{k+1}\sum_{i=1}^{k}\log\frac{X_{n,n-i+1}}{X_{n,n-k}}\right\}\{1-\hat{\gamma}\}\hat{\gamma}^{-2}\{1+\hat{\gamma}\}\{1+2\hat{\gamma}\}^{1/2}.$$

Then, it follows from Li and Peng (2009a, 2009b) that

$$\sqrt{k}\frac{\hat{\theta} - \theta}{\hat{\sigma}} \xrightarrow{d} N(0, 1)$$

under conditions of Proposition 1. Further Li and Peng (2009a, 2009b) showed that bootstrap method based on the pivotal statistic $T_n = \sqrt{k}\{\hat{\theta} - \theta\}/\hat{\sigma}$ is consistent. Therefore, we can construct the normal approximation confidence interval and percentile-t bootstrap confidence interval with level d based on the pivotal statistic T_n. Let us denote them as I_d^N and I_d^B.

Draw 1000 random samples with size $n = 1000$ from the random variable $\theta - 1/Z$, where $P(Z \leq z) = 1 - (1 + z^{\tau_1})^{-\tau_2}$ for $z > 0$. For calculating the bootstrap calibration confidence interval I_d^{BC} and the percentile-t bootstrap confidence interval I_d^B, we draw 200 resamplings with size n for each sample. We consider the cases $(\theta, \tau_1, \tau_2) = (0, 8, 4/8)$ and $(0, 40, 4/40)$. Note that the case $(\tau_1, \tau_2) = (40, 8/40)$ gives a much faster rate of convergence for function A defined in (2.4) than the case $(\tau_1, \tau_2) = (8, 4/8)$, which means that a larger k can be employed for the case $(\tau_1, \tau_2) = (40, 4/40)$ than the other. In Fig. 1, we plot the coverage probabilities against $k = 50, 60, \ldots, 200$ for these four confidence intervals with levels 0.9, 0.95.

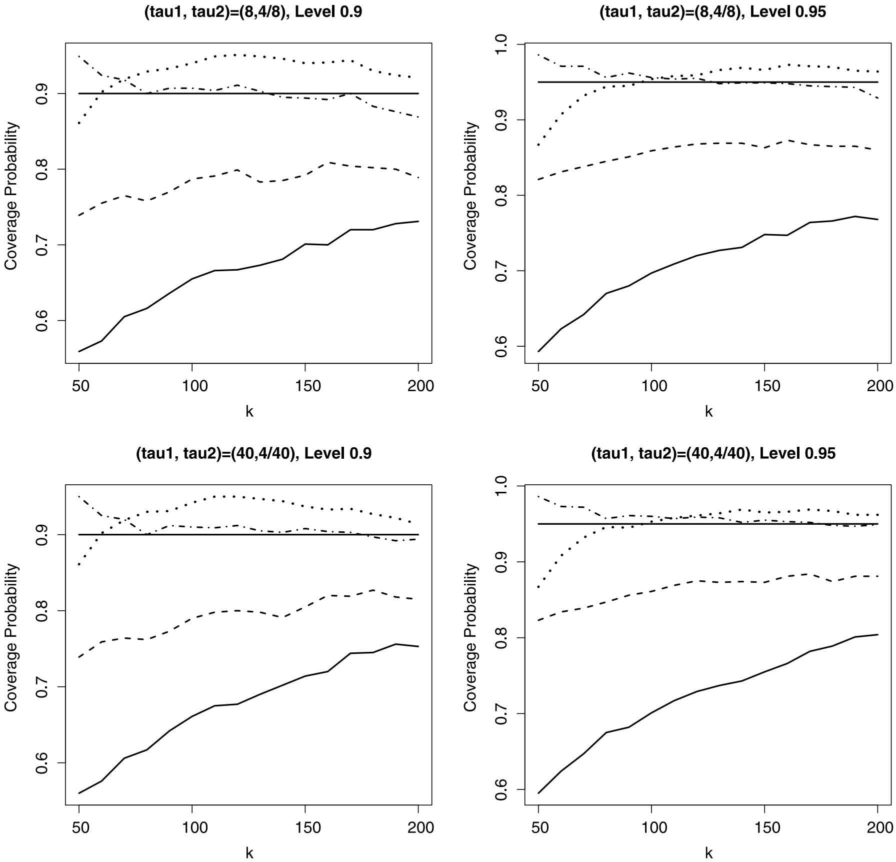

Fig. 1 Empirical coverage probabilities for the four confidence intervals are plotted against $k = 50, \ldots, 200$. *Straight solid line*, another *solid line*, *dashed line*, *dotted line* and *dotted-dashed line* represent the level and intervals I_d^N, I_d^E, I_d^B and I_d^{BC}, respectively

From Fig. 1 we found that the proposed empirical likelihood confidence interval is more accurate than the normal approximation confidence interval, and both are less accurate but less computationally intensive, than the proposed bootstrap calibration confidence interval and the percentile-t bootstrap confidence interval. The proposed bootstrap calibration confidence interval is most accurate for most cases and least sensitive to the choice of the sample fraction k. Moreover, confidence intervals for the case $(\tau_1, \tau_2) = (40, 8/40)$ are more accurate than those for the case $(\tau_1, \tau_2) = (8, 4/8)$ when a large k is employed.

Next we apply the above intervals to a real data set, which consists of the total lifespan (in days) of 10391 residents born in the Netherlands in the years 1877–1881, still alive on January 1, 1971, and who died as resident of the Netherlands. More detailed analysis on this data set can be found in Chaps. 3.7 and 4.6 of De Haan and Ferreira (2006). Here we focus on the intervals I_d^B and I_d^{BC} since the above results show that these two intervals are much more accurate than the other two.

Table 1 Intervals $I^B_{0.95}$ and $I^{BC}_{0.95}$ are calculated for the lifespan data set with $k = 100, 150, \ldots, 400$

	$I^B_{0.95}$	$I^{BC}_{0.95}$
$k = 100$	(111.84, 113.49)	(111.65, 121.61)
$k = 150$	(112.02, 117.81)	(111.92, 123.69)
$k = 200$	(112.42, 117.35)	(112.18, 126.02)
$k = 250$	(112.11, 118.36)	(112.13, 125.54)
$k = 300$	(112.29, 119.06)	(112.27, 124.85)
$k = 350$	(112.34, 119.11)	(112.27, 127.59)
$k = 400$	(112.59, 119.81)	(112.31, 128.61)

First we change the lifespan in days to years by dividing by 365. Then we calculate these two intervals by drawing 1000 bootstrap samples. To approximate I^{BC}_d by an interval, we increase and decrease $\bar{\theta}$ from $\hat{\theta}$ by step 0.01 until $l(\bar{\theta}, \hat{\alpha}(\bar{\theta}))$. Table 1 reports these two intervals with level 95% for $k = 100, 150, \ldots, 400$. From this table, we observe that the length of the interval I^{BC}_d is larger than that of I^B_d, and I^{BC}_d is much skewed to the right.

4 Proofs

Let $V_1, \ldots, V_n$ 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 $V_1, \ldots, V_n$. Consider another independent sequence of i.i.d. random variables $V^*_1, \ldots, V^*_k$ with distribution function $1 - 1/x$, and denote $V^*_{k,1} \leq \cdots \leq V^*_{k,k}$ as their order statistics. It is well known that

$$\{V_{n,n-i+1}/V_{n,n-k}\}_{i=1}^k \stackrel{d}{=} \{V^*_{k,k-i+1}\}_{i=1}^k. \tag{4.1}$$

The following lemma comes from Lemma 5.2 of Draisma et al. (1999).

Lemma 1 *Let f be a measurable function. Suppose there exist a real parameter α_1 and functions $a_1(t) > 0$ and $A_1(t) \to 0$ such that for all $x > 0$*

$$\lim_{t\to\infty} \frac{\frac{f(tx)-f(t)}{a_1(t)} - \frac{x^{\alpha_1}-1}{\alpha_1}}{A_1(t)} = H_1(x) = \frac{1}{\beta_1}\left\{\frac{x^{\alpha_1+\beta_1}-1}{\alpha_1+\beta_1} - \frac{x^{\alpha_1}-1}{\alpha_1}\right\},$$

where $\beta_1 \leq 0$. Then for any $\epsilon > 0$ there exists $t_0 > 0$ such that for all $t \geq t_0, tx \geq t_0$,

$$\left|\frac{\frac{f(tx)-f(t)}{a_1(t)} - \frac{x^{\alpha_1}-1}{\alpha_1}}{A_1(t)} - H_1(x)\right| \leq \epsilon\left\{1 + x^{\alpha_1} + 2x^{\alpha_1+\beta_1}e^{\epsilon|\log x|}\right\}.$$

Define

$$\bar{\mathbf{Y}}_n(\alpha) = \frac{1}{k}\sum_{i=1}^k \mathbf{Y}_i(\alpha, \theta_0)$$

and

$$S_n(\alpha) = \frac{1}{k}\sum_{i=1}^{k} \mathbf{Y}_i(\alpha, \theta_0)\mathbf{Y}_i^T(\alpha, \theta_0).$$

Lemma 2 *Under the conditions of Proposition 1,*

$$\sup_{\alpha>2} \frac{1}{k}\sum_{i=1}^{k} \left\|\mathbf{Y}_i(\alpha, \theta_0)\right\|^2 = O_p(1) \tag{4.2}$$

and

$$\sup_{\alpha>2} \max_{1\le i\le k} \left\|\mathbf{Y}_i(\alpha, \theta_0)\right\| = o_p\big(k^{\delta}\big) \tag{4.3}$$

for any $\delta \in (1/\alpha_0, 1/2]$.

Proof Since $X_{n,n-i+1} \overset{d}{=} U(V_{n,n-i+1})$, we can write

$$\begin{aligned}\frac{\theta_0 - X_{n,n-i+1}}{a(V_{n,n-k})} &\overset{d}{=} \left\{\frac{\theta_0 - U(V_{n,n-k})}{a(V_{n,n-k})} + \frac{1}{\gamma_0}\right\} - \frac{(V_{n,n-i+1}/V_{n,n-k})^{\gamma_0}}{\gamma_0} \\ &\quad - \left\{\frac{U(\frac{V_{n-i+1}}{V_{n,n-k}} V_{n,n-k}) - U(V_{n,n-k})}{a(V_{n,n-k})} - \frac{(V_{n,n-i+1}/V_{n,n-k})^{\gamma_0} - 1}{\gamma_0}\right\} \\ &= I_1 - I_2 - I_3\end{aligned} \tag{4.4}$$

for $i = 1, \ldots, k$. It follows from (2.4) that

$$\frac{\frac{\theta_0 - U(t)}{a(t)} + \frac{1}{\gamma_0}}{A(t)} \to H_{\gamma_0,\rho}(\infty) \tag{4.5}$$

(see the proof of Lemma 4.2 in Ferreira et al. 2003), which implies that

$$\frac{\frac{\theta_0 - X_{n,n-k}}{a(V_{n,n-k})} + \frac{1}{\gamma_0}}{A(n/k)} \to H_{\gamma_0,\rho}(\infty) \quad \text{a.s.} \tag{4.6}$$

since $\frac{k}{n} V_{n,n-k} \to 1$ a.s. By (2.4), (4.6), Lemma 1, (2.5) and $\delta \le 1/2$, we can show that

$$I_1 = O\big(A(n/k)\big) = o\big(k^{-\delta}\big) \quad \text{and} \quad I_3 = O\big(A(n/k)\big) = o\big(k^{-\delta}\big) \quad \text{a.s.} \tag{4.7}$$

By (4.1) and $\gamma_0 = -1/\alpha_0 < -\delta$, we have

$$\min_{1\le i\le k} k^{\delta}(V_{n,n-i+1}/V_{n,n-k})^{\gamma_0} = k^{\delta}(V_{n,n}/V_{n,n-k})^{\gamma_0} \overset{d}{=} k^{\delta}(V_{k,k}^*)^{\gamma_0} \to \infty \quad \text{a.s.} \tag{4.8}$$

It follows from (4.4), (4.7) and (4.8) that

$$\min_{1\le i\le k} k^{\delta} \frac{\theta_0 - X_{n,n-i+1}}{a(V_{n,n-k})} \to \infty \quad \text{a.s.} \tag{4.9}$$

By (4.6) and (4.9), we have

$$\sup_{\alpha>2}\max_{1\le i\le k}\left|Y_{ij}(\alpha,\theta_0)\right| = o\left(k^{\delta}\right) \quad \text{a.s.} \tag{4.10}$$

for $j = 1, 2$, which implies (4.3). If follows from (4.1) that

$$\frac{1}{k}\sum_{i=1}^{k}(V_{n,n-i+1}/V_{n,n-k})^{\gamma_0} \overset{d}{=} \frac{1}{k}\sum_{i=1}^{k}(V_i^*)^{\gamma_0} \to \frac{1}{1-\gamma_0} \quad \text{a.s.} \tag{4.11}$$

Hence, (4.2) follows from (4.4), (4.6), (2.4), Lemma 1, (2.5) and (4.11). □

Lemma 3 *Under the conditions in Proposition 1, we have*

$$\sqrt{k}\bar{\mathbf{Y}}_n(\alpha_0) \overset{d}{\to} N(0, V), \tag{4.12}$$

$$\left\|\bar{\mathbf{Y}}_n(\alpha) - \bar{\mathbf{Y}}_n(\alpha_0)\right\| = O\left(k^{-1/2}\log\log k + |\alpha-\alpha_0|\right) \quad a.s., \tag{4.13}$$

$$\bar{Y}_n(\alpha_0) = O\left(k^{-1/2}\log\log k\right) \quad a.s. \tag{4.14}$$

and

$$\sup_{|\alpha-\alpha_0|\le\delta_n}\left\|S_n(\alpha) - V\right\| \to 0 \quad a.s. \tag{4.15}$$

for any sequence $\delta_n \to 0$, *where*

$$V = \begin{pmatrix} \frac{1}{\alpha_0^2} & -\frac{1}{(\alpha_0-1)^2} \\ -\frac{1}{(\alpha_0-1)^2} & \frac{\alpha_0}{(\alpha_0-1)^2(\alpha_0-2)} \end{pmatrix}.$$

Proof By (4.4), (2.4), Lemma 1, (2.5), (4.8) and Taylor expansion, we can show that

$$\max_{1\le i\le k}\left|\frac{a(V_{n,n-k})}{\theta_0 - X_{n,n-i+1}} - \frac{-\gamma_0}{(V_{n,n-i+1}/V_{n,n-k})^{\gamma_0}}\right| = o\left(1/\sqrt{k}\right) \quad \text{a.s.} \tag{4.16}$$

Since $\min_{1\le i\le k}\frac{-\gamma_0}{(V_{n,n-i+1}/V_{n,n-k})^{\gamma_0}} \ge -\gamma_0$, (4.16) implies that

$$\max_{1\le i\le k}\left|\log\frac{a(V_{n,n-k})}{\theta_0 - X_{n,n-i+1}} - \log\frac{-\gamma_0}{(V_{n,n-i+1}/V_{n,n-k})^{\gamma_0}}\right| = o\left(1/\sqrt{k}\right) \quad \text{a.s.} \tag{4.17}$$

It follows from (4.6) and (2.5) that

$$\begin{cases} \frac{\theta_0 - X_{n,n-k}}{a(V_{n,n-k})} + \frac{1}{\gamma_0} = o(1/\sqrt{k}) & \text{a.s.} \\ \log\frac{\theta_0 - X_{n,n-k}}{a(V_{n,n-k})} - \log\frac{1}{-\gamma_0} = o(1/\sqrt{k}) & \text{a.s.} \end{cases} \tag{4.18}$$

By (4.16)–(4.18) and (4.1), we have

$$\begin{cases} \frac{1}{k}\sum_{i=1}^{k} Y_{i1}(\alpha,\theta_0) \overset{d}{=} \{\frac{1}{k}\sum_{i=1}^{k}\log(V_i^*)^{\gamma_0}+\frac{1}{\alpha_0}\}+\{\frac{1}{\alpha}-\frac{1}{\alpha_0}\}+o(1/\sqrt{k}) & \text{a.s.,} \\ \frac{1}{k}\sum_{i=1}^{k} Y_{i2}(\alpha,\theta_0) \overset{d}{=} \{\frac{1}{k}\sum_{i=1}^{k}(V_i^*)^{-\gamma_0}-\frac{\alpha_0}{\alpha_0-1}\}-\{\frac{\alpha}{\alpha-1}-\frac{\alpha_0}{\alpha_0-1}\}+o(1/\sqrt{k}) & \text{a.s.} \end{cases} \tag{4.19}$$

Hence, (4.12), (4.13) and (4.14) follow from (4.19) and the fact that

$$\frac{1}{k}\sum_{i=1}^{k}(V_i^*)^{-\gamma_0}-\frac{\alpha_0}{\alpha_0-1}=O\big(k^{-1/2}\log\log k\big) \quad \text{a.s.,}$$

$$\frac{1}{k}\sum_{i=1}^{k}\log(V_i^*)^{\gamma_0}+\frac{1}{\alpha_0}=O\big(k^{-1/2}\log\log k\big) \quad \text{a.s.}$$

and

$$\sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{k}(V_i^*)^{-\gamma_0}-\frac{\alpha_0}{\alpha_0-1},\ \frac{1}{k}\sum_{i=1}^{k}\log(V_i^*)^{\gamma_0}+\frac{1}{\alpha_0}\right)\overset{d}{\to}N(0,V).$$

Similarly we can show (4.15). □

Proof of Proposition 1 By setting

$$\boldsymbol{\mu}(\alpha)=\left(\frac{1}{\alpha_0}-\frac{1}{\alpha},\ \frac{\alpha}{\alpha-1}-\frac{\alpha_0}{\alpha_0-1}\right)^T,$$

we have

$$\mathbf{Y}_i(\alpha,\theta)=\mathbf{Y}_i(\alpha_0,\theta)-\boldsymbol{\mu}(\alpha).$$

Put

$$g(\boldsymbol{\lambda},\alpha)=\frac{1}{k}\sum_{i=1}^{k}\frac{\mathbf{Y}_i(\alpha,\theta_0)}{1+\boldsymbol{\lambda}^T\mathbf{Y}_i(\alpha,\theta_0)} \tag{4.20}$$

and $D_n(\alpha)=\max_{1\le i\le k}\|\mathbf{Y}_i(\alpha,\theta_0)\|$. Let $\eta=\boldsymbol{\lambda}/\|\boldsymbol{\lambda}\|$, i.e., $\boldsymbol{\lambda}=\|\boldsymbol{\lambda}\|\eta$. It follows from (2.3) that

$$0=\eta^T g(\boldsymbol{\lambda},\alpha)=\frac{1}{k}\sum_{i=1}^{k}\left(\frac{\eta^T\mathbf{Y}_i(\alpha,\theta_0)}{1+\boldsymbol{\lambda}^T\mathbf{Y}_i(\alpha,\theta_0)}-\eta^T\mathbf{Y}_i(\alpha,\theta_0)\right)+\eta^T\bar{\mathbf{Y}}_n(\alpha),$$

which implies that

$$\|\boldsymbol{\lambda}\|\eta^T\tilde{S}_n(\alpha)\eta=\eta^T\bar{\mathbf{Y}}_n(\alpha),$$

where

$$\tilde{S}_n(\alpha)=\frac{1}{k}\sum_{i=1}^{k}\frac{\mathbf{Y}_i(\alpha,\theta_0)\mathbf{Y}_i^T(\alpha,\theta_0)}{1+\|\boldsymbol{\lambda}\|\eta^T\mathbf{Y}_i(\alpha,\theta_0)}.$$

From $1+\|\boldsymbol{\lambda}\|\eta^T\mathbf{Y}_i(\alpha,\theta_0)>0$, we get that

$$\begin{aligned}\|\boldsymbol{\lambda}\|\eta^T S_n(\alpha)\eta &\le \|\boldsymbol{\lambda}\|\eta^T \tilde{S}_n(\alpha)\eta\Big(1+\max_{1\le i\le k}\|\boldsymbol{\lambda}\|\big|\eta^T\mathbf{Y}_i(\alpha,\theta_0)\big|\Big)\\ &\le \|\boldsymbol{\lambda}\|\eta^T \tilde{S}_n(\alpha)\eta\big(1+\|\boldsymbol{\lambda}\|D_n(\alpha)\big)\\ &= \eta^T\bar{\mathbf{Y}}_n(\alpha)\big(1+\|\boldsymbol{\lambda}\|D_n(\alpha)\big),\end{aligned}$$

i.e.,

$$\|\boldsymbol{\lambda}\|\big\{\eta^T S_n(\alpha)\eta - D_n(\alpha)\eta^T\bar{\mathbf{Y}}_n(\alpha)\big\}\le \eta^T\bar{\mathbf{Y}}_n(\alpha). \tag{4.21}$$

Denote the smallest eigenvalue of V as σ_1. It follows from Lemmas 2 and 3 that

$$\begin{cases}\min_\eta(\eta^T S_n(\alpha)\eta)=\sigma_1(1+o(1)) & \text{a.s.},\\ \min_\eta|D_n(\alpha)\eta^T\bar{Y}_n(\alpha)|=o(1) & \text{a.s.}\end{cases} \tag{4.22}$$

uniformly for $|\alpha-\alpha_0|\le k^{-\delta_0}$ with δ_0 given in Proposition 1. Therefore, by (4.21), (4.22) and Lemma 2,

$$\|\boldsymbol{\lambda}\|=O\big(k^{-\delta_0}\big)\quad\text{a.s.} \tag{4.23}$$

uniformly for $|\alpha-\alpha_0|\le k^{-\delta_0}$.

Note that

$$\begin{aligned}0&=\frac{1}{k}\sum_{i=1}^k\mathbf{Y}_i(\alpha,\theta_0)\bigg(1-\boldsymbol{\lambda}^T\mathbf{Y}_i(\alpha,\theta_0)+\frac{(\boldsymbol{\lambda}^T\mathbf{Y}_i(\alpha,\theta_0))^2}{1-\boldsymbol{\lambda}^T\mathbf{Y}_i(\alpha,\theta_0)}\bigg)\\ &=\bar{\mathbf{Y}}_n(\alpha)-S_n(\alpha)\boldsymbol{\lambda}+\frac{1}{k}\sum_{i=1}^k\frac{\mathbf{Y}_i(\alpha,\theta_0)(\boldsymbol{\lambda}^T\mathbf{Y}_i(\alpha,\theta_0))^2}{1-\boldsymbol{\lambda}^T\mathbf{Y}_i(\alpha,\theta_0)}.\end{aligned}$$

By Lemma 2 and (4.23), the last term is dominated by

$$\frac{1}{k}\sum_{i=1}^k\frac{\|\mathbf{Y}_i(\alpha,\theta_0)\|^3\|\boldsymbol{\lambda}\|^2}{1-\|\boldsymbol{\lambda}\|D_n(\alpha)}=O(1)D_n(\alpha)\|\boldsymbol{\lambda}\|^2=o\big(k^{\delta_0}\big)O\big(k^{-2\delta_0}\big)=o\big(k^{-\delta_0}\big)\quad\text{a.s.}$$

which implies that

$$\boldsymbol{\lambda}=S_n^{-1}(\alpha)\bar{\mathbf{Y}}_n(\alpha)+o\big(k^{-\delta_0}\big)\quad\text{a.s.} \tag{4.24}$$

uniformly in $|\alpha-\alpha_0|\le k^{-\delta_0}$.

By (4.24) and Taylor expansion, we have, almost surely, that

$$\begin{aligned}&l\big(\theta_0,\alpha_0+k^{-\delta_0}\big)\\ &=2\sum_{i=1}^k\boldsymbol{\lambda}^T\mathbf{Y}_i\big(\alpha_0+k^{-\delta_0},\theta_0\big)-\sum_{i=1}^k\big\{\boldsymbol{\lambda}^T\mathbf{Y}_i\big(\alpha_0+k^{-\delta_0},\theta_0\big)\big\}^2+o\big(kk^{-2\delta_0}\big)\\ &=k\bar{\mathbf{Y}}_n^T\big(\alpha_0+k^{-\delta_0}\big)S_n^{-1}\big(\alpha_0+k^{-\delta_0}\big)\bar{\mathbf{Y}}_n\big(\alpha_0+k^{-\delta_0}\big)+o\big(k^{1-2\delta_0}\big)\end{aligned}$$

$$
\begin{aligned}
&= k\boldsymbol{\mu}^T\big(\alpha_0 + k^{-\delta_0}\big)S_n^{-1}\big(\alpha_0 + k^{-\delta_0}\big)\boldsymbol{\mu}\big(\alpha_0 + k^{-\delta_0}\big) + o\big(k^{1-2\delta_0}\big) \\
&\geq c_1 k^{1-2\delta_0}
\end{aligned}
\tag{4.25}
$$

for some constant $c_1 > 0$. Similarly, we can show that

$$
l\big(\theta_0, \alpha_0 - k^{-\delta_0}\big) \geq c_2 k^{1-2\delta_0} \quad \text{a.s.} \tag{4.26}
$$

and

$$
l(\theta_0, \alpha_0) = k\bar{\mathbf{Y}}_n^T(\alpha_0)S_n^{-1}(\alpha_0)\bar{\mathbf{Y}}_n(\alpha_0) + o\big(k^{1-2\delta_0}\big) = o\big(k^{1-2\delta_0}\big) \quad \text{a.s.} \tag{4.27}
$$

Hence, the proposition follows from (4.25)–(4.27). □

Proof of Theorem 1 Taking derivatives with respect to α and $\boldsymbol{\lambda}^T$, we have

$$
\begin{aligned}
&\frac{\partial}{\partial \alpha} Q_{1n}(\alpha, 0) = \left(-\frac{1}{\alpha^2}, \frac{1}{(\alpha-1)^2}\right)^T, \\
&\frac{\partial}{\partial \boldsymbol{\lambda}^T} Q_{1n}(\alpha, 0) = -\frac{1}{k}\sum_{i=1}^{k} \mathbf{Y}_i(\alpha, \theta_0)\mathbf{Y}_i^T(\alpha, \theta_0), \\
&\frac{\partial}{\partial \alpha} Q_{2n}(\alpha, 0) = 0, \qquad \frac{\partial}{\partial \boldsymbol{\lambda}^T} Q_{2n}(\alpha, 0) = \left(-\frac{1}{\alpha^2}, \frac{1}{(\alpha-1)^2}\right).
\end{aligned}
$$

Expanding $Q_{1n}(\hat{\alpha}(\theta_0), \hat{\boldsymbol{\lambda}})$ and $Q_{2n}(\hat{\alpha}(\theta_0), \hat{\boldsymbol{\lambda}})$ at $(\alpha_0, 0)$ and using Proposition 1, we have that

$$
0 = Q_{jn}(\alpha_0, 0) + \frac{\partial}{\partial \alpha} Q_{jn}(\alpha_0, 0)\big(\hat{\alpha}(\theta_0) - \alpha_0\big) + \frac{\partial}{\partial \boldsymbol{\lambda}^T} Q_{jn}(\alpha_0, 0)(\hat{\boldsymbol{\lambda}} - 0) + o_p(\delta_n) \tag{4.28}
$$

for $j = 1, 2$, where $\delta_n = |\hat{\alpha}(\theta_0) - \alpha_0| + \|\hat{\boldsymbol{\lambda}}\|$. Hence

$$
\big(\hat{\boldsymbol{\lambda}}, \hat{\alpha}(\theta_0) - \alpha_0\big)^T = \Delta_n^{-1}\big(-Q_{1n}^T(\alpha_0, 0) + o_p(\delta_n), o_p(\delta_n)\big)^T, \tag{4.29}
$$

where

$$
\Delta_n = \begin{pmatrix} \frac{\partial}{\partial \boldsymbol{\lambda}^T} Q_{1n}(\alpha_0, 0) & \frac{\partial}{\partial \alpha} Q_{1n}(\alpha_0, 0) \\ \frac{\partial}{\partial \boldsymbol{\lambda}^T} Q_{2n}(\alpha_0, 0) & 0 \end{pmatrix} \tag{4.30}
$$

$$
\xrightarrow{p} \begin{pmatrix} -V & (-\alpha_0^{-2}, (\alpha_0 - 1)^{-2})^T \\ (-\alpha_0^{-2}, (\alpha_0 - 1)^{-2}) & 0 \end{pmatrix}. \tag{4.31}
$$

Since $Q_{1n}(\alpha_0, 0) = O_p(k^{-1/2})$, (4.29) and (4.30) imply that

$$
\delta_n = O_p\big(k^{-1/2}\big). \tag{4.32}
$$

Hence, by (4.29)–(4.32) and Lemma 3,

$$
\begin{aligned}
l\big(\theta_0, \hat{\alpha}(\theta_0)\big) &= 2k\big(\hat{\boldsymbol{\lambda}}^T, \hat{\alpha}(\theta_0) - \theta_0\big)\big(Q_{1n}^T(\alpha_0, 0), 0\big)^T \\
&\quad + k\big(\hat{\boldsymbol{\lambda}}^T, \hat{\alpha}(\theta_0) - \theta_0\big)\Delta_n\big(\hat{\boldsymbol{\lambda}}^T, \hat{\alpha}(\theta_0) - \theta_0\big)^T + o_p\big(k^{-1/2}\big) \\
&= -k\big(Q_{1n}^T(\alpha_0, 0), 0\big)\Delta_n^{-1}\big(Q_{1n}^T(\alpha_0, 0), 0\big)^T + o_p\big(k^{-1/2}\big) \\
&= -\big(\sqrt{k}\bar{\mathbf{Y}}_n^T(\alpha_0), 0\big)\Delta_n^{-1}\big(\sqrt{k}\bar{\mathbf{T}}_n^T(\alpha, 0), 0\big)^T + o_p\big(k^{-1/2}\big) \\
&\xrightarrow{d} \chi^2(1). \qquad \square
\end{aligned}
$$

Acknowledgements We thank two reviewers for helpful comments. Li's research was partly supported by NNSFC Grant 10801038. Peng's research was supported by NSA Grant H98230-10-1-0170 and Qi's research was supported by NSA Grant H98230-10-1-0161.

References

Aarssen K, de Haan L (1994) On the maximal lifespan of humans. Math Popul Stud 4:259–281

Athreya KB, Fukuchi JI (1997) Confidence intervals for endpoints of a cdf via bootstrap. J Stat Plann Inference 58:299–320

Dekker ALM, Einmahl JHJ, de Haan L (1989) A moment estimator for the index of an extreme value distribution. Ann Stat 17:1833–1855

Draisma G, de Haan L, Peng L, Pereira TT (1999) A bootstrap-based method to achieve optimality in estimating the extreme-value index. Extremes 2:367–404

Einmahl JHJ, Segers J (2009) Maximum empirical likelihood estimation of the spectral measure of an extreme value distribution. Ann Stat 37:2953–2989

Falk M (1994) Some best estimators for distributions with finite endpoints. Statistics 27:115–125

Ferreira A, de Haan L, Peng L (2003) On optimising the estimation of high quantiles of a probability distribution. Statistics 37:403–434

De Haan L, Ferreira A (2006) Extreme value theory: an introduction. Springer, Berlin

Hall P (1982) On estimating the endpoint of a distribution. Ann Stat 10:556–568

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

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

Li Z, Peng L (2009b). Bootstrapping endpoint. Technical report

Loh WY (1984) Estimating an endpoint of a distribution with resampling methods. Ann Stat 12:1534–1550

Owen A (2001) Empirical likelihood. Chapman & Hall, London

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

Qi Y (2008) Bootstrap and empirical likelihood methods in extremes. Extremes 11:81–98

Qin J, Lawless J (1994) Empirical likelihood and general estimating equations. Ann Stat 22:300–325

Smith RL (1985) Maximum likelihood estimation in a class of nonregular cases. Biometrika 72:67–90

Smith RL (1987) Estimating tails of probability distributions. Ann Stat 15:1174–1207

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

Zhou C (2009) Existence and consistency of the maximum likelihood estimator for the extreme value index. J Multivar Anal 4:794–815