

On the tail index of a heavy tailed distribution

Yongcheng Qi

Received: 12 January 2007 / Revised: 4 February 2008 / Published online: 15 April 2008
© The Institute of Statistical Mathematics, Tokyo 2008

Abstract This paper proposes some new estimators for the tail index of a heavy tailed distribution when only a few largest values are observed within blocks. These estimators are proved to be asymptotically normal under suitable conditions, and their Edgeworth expansions are obtained. Empirical likelihood method is also employed to construct confidence intervals for the tail index. The comparison for the confidence intervals based on the normal approximation and the empirical likelihood method is made in terms of coverage probability and length of the confidence intervals. The simulation study shows that the empirical likelihood method outperforms the normal approximation method.

Keywords Confidence interval · Coverage probability · Empirical likelihood · Tail index estimation · Edgeworth expansion

1 Introduction

In many fields such as meteorology, hydrology, climatology, environmental science, telecommunications, insurance and finance, heavy tailed distributions are recommended to model the data (see e.g. [Embrechts et al. 1997](#)), and hence the problem of estimation of the tail index of a heavy tailed distribution has been paid much attention in recent years. Various estimates for the tail index have been proposed in the literature,

The research was supported by NSF grant DMS 0604176.

Y. Qi (✉)
Department of Mathematics and Statistics,
University of Minnesota Duluth, SCC 140, 1117 University Drive,
Duluth, MN 55812, USA
e-mail: yqi@d.umn.edu

see, e.g., [Hill \(1975\)](#), [Pickands \(1975\)](#), and [Dekkers et al. \(1989\)](#). These estimators are based on a few of upper order statistics.

Recently, [Davydov et al. \(2000\)](#) and [Paulauskas \(2003\)](#) proposed a new estimator for the tail index. In their approach, observations are divided into several blocks and the estimator of the tail index is constructed from the ratios of the first largest and second largest terms within blocks. A similar idea was also used by [Segers and Teugels \(2000\)](#) who studied the limiting distribution of Galton's ratio computed from each of the blocks in the entire sample and developed a parallel procedure to test whether the underlying distribution is from the extremal domain of attraction of the Gumbel distribution.

Since Davydov–Paulauskas–Rackauska's (DPR) approach doesn't use all the upper order statistics, it may not be as efficient as Hill estimator when it is used to estimate the tail index. Even so, the simulation study in [Paulauskas \(2003\)](#) shows that the large sample performance of the estimator is good. A recent work by [Gadeikis and Paulauskas \(2005\)](#) used a similar method to test and estimate the change-point in the tail index. When the data can be divided into several blocks but only a few of largest observations within blocks are available for analysis, DPR's approach is feasible for estimation of the tail index. In fact, sometimes only several largest observations within blocks are available for analysis. For example, for financial data, it is very common that only the information on a few yearly largest claims or losses is reported to the public. In many Olympic games, only a few best players are allowed to participate, and thus only scores for these people are observed and these observations can be considered as the largest observations within each block (or game). In these situations, DPR's approach has its advantages over others, since none of the aforementioned methods is applicable.

One of the main purposes of the paper is to propose a new class of estimators for the tail index under a setup similar to DPR's. By using the same information as in DPR's approach, our new estimators are more efficient than DPR's in the sense that our new estimators have a smaller asymptotic variance. Hence, the confidence intervals based on our new estimators are shorter than the ones based on DPR's. This paper also applies empirical likelihood method to construct the confidence intervals for the tail index.

The empirical likelihood was introduced by [Owen \(1988, 1990\)](#) for the mean vector of independent and identically distributed observations, and it has been extended to a wide range of applications including, among others, linear models ([Owen 1991](#)), generalized linear models ([Kolaczyk 1994](#)), quantile estimation ([Chen and Hall 1993](#)), generalized estimating equations ([Qin and Lawless 1994](#)), partial linear models ([Wang and Jing 2003](#)), errors-in-covariables models ([Wang and Rao 2002](#)), censored linear regression models ([Qin and Jing 2001a](#)), censored partial linear models ([Qin and Jing 2001b](#); [Wang and Li 2002](#)), Cox regression models ([Qin and Jing 2001c](#)), additive risk models ([Lu and Qi 2004](#)), nonparametric density estimation ([Chen 1996](#)) and nonparametric regression models ([Chen and Qin 2000](#)). The empirical likelihood method possesses some advantages over other methods like the normal approximation method. It allows the use of likelihood methods without having to pick a parametric family for the data. It produces confidence regions whose shape and orientation are determined entirely by the data.

Lu and Peng (2002) applied both the empirical likelihood method and the parametric method to obtain confidence intervals for the tail index of a heavy-tailed distribution, and compared their performance with the normal approximation method based on Hill estimator (Hill 1975). The simulation study in Lu and Peng (2002) indicates that the empirical likelihood method and the parametric method are comparable, and both result in better coverage accuracy than the normal approximation method. Peng and Qi (2006b) studied the confidence intervals for high quantiles of a heavy-tailed distribution. They applied three methods, including normal approximation method, likelihood ratio method and data tilting method (similar to the empirical likelihood method). From their simulation study, the data tilting method is favorable.

The paper is organized as follows. In Sect. 2 we first introduce our new estimators for the tail index and obtain their limiting distribution, then we derive Edgeworth expansion for the new estimators in order to assess the accuracy of the coverage probability for confidence intervals based on the asymptotic normality of our estimators. We also employ the empirical likelihood method to construct the confidence intervals for the tail index. In Sect. 3, we conduct a simulation study to compare the confidence intervals based on the normal approximation of our new estimators and the empirical likelihood method. Finally, we give all the proofs in the Appendix.

2 Estimator of the tail index and its Edgeworth expansion

Throughout let X_1, \dots, X_n be independent and identically distributed (iid) random variables with a distribution function F satisfying

$$1 - F(x) = x^{-1/\gamma} L(x) \quad \text{for } x > 0, \tag{1}$$

where $\gamma > 0$ is an unknown parameter and $1/\gamma$ is called the tail index of the distribution function F , and L is slowly varying function satisfying

$$\lim_{t \rightarrow \infty} \frac{L(tx)}{L(t)} = 1.$$

We are interested in estimating γ and constructing confidence interval for γ .

Let $X_{n,1} \geq \dots \geq X_{n,n}$ denote the order statistics based on X_1, \dots, X_n . The well-known Hill estimator for γ is defined as

$$\widehat{\gamma}_H(k_n) = \frac{1}{k_n} \sum_{i=1}^{k_n} \log X_{n,i} - \log X_{n,k_n+1},$$

where k_n satisfies

$$k_n \rightarrow \infty \quad \text{and} \quad \frac{k_n}{n} \rightarrow 0 \quad \text{as } n \rightarrow \infty. \tag{2}$$

See Hill (1975). Like most of the estimators for the tail index in the literature, the Hill estimator uses only a few upper order statistics.

An alternative estimator was recently proposed by Davydov et al. (2000) and Paulauskas (2003). The procedure is as follows. First, divide the sample X_1, \dots, X_n into k_n blocks (or groups), V_1, \dots, V_{k_n} , and each block contains $m = m_n = [n/k_n]$ observations, where $[x]$ denotes the integer part of $x > 0$. To be more specific, $V_i = \{X_{(i-1)m+1}, \dots, X_{im}\}$ for $1 \leq i \leq k_n$. Let $X_{m,1}^{(i)} \geq \dots \geq X_{m,m}^{(i)}$ denote the order statistics of the m observations in the i -th block. Set

$$S_{k_n} = \sum_{i=1}^{k_n} \frac{X_{m,2}^{(i)}}{X_{m,1}^{(i)}}$$

and define

$$\widehat{\gamma}_{\text{DPR}}(k_n) = S_{k_n}^{-1}(k_n - S_{k_n})$$

an estimator of γ . Under the condition

$$1 - F(x) = cx^{-1/\gamma} + dx^{-\beta} + o(x^{-\beta}) \quad \text{as } x \rightarrow \infty, \tag{3}$$

where $0 < \gamma^{-1} < \beta \leq \infty$, it is proved that

$$k_n^{1/2}(\widehat{\gamma}_{\text{DPR}}(k_n) - \gamma) \xrightarrow{d} N\left(0, \frac{\gamma^2(1 + \gamma)^2}{(1 + 2\gamma)}\right) \tag{4}$$

if $k_n = o(n^{2(\beta\gamma-1)/(2\beta\gamma-1)})$, where \xrightarrow{d} denotes convergence in distribution.

Let $r \geq 1$ be an integer. In this paper we assume that the $r + 1$ largest random variables within the k_n blocks are observed, that is, only the observations $\{X_{m,j}^{(i)} : j = 1, \dots, r + 1, i = 1, \dots, k_n\}$ are available for inference. We propose to estimate γ by

$$\widehat{\gamma}_N(k_n) = \frac{1}{k_n r} \sum_{i=1}^{k_n} \sum_{j=1}^r (\log X_{m,j}^{(i)} - \log X_{m,r+1}^{(i)}), \tag{5}$$

where k_n satisfies condition (2).

For convenience we assume that all random variables are bounded below by 1. Otherwise we can use truncation technique and substitute $X_{m,j}^{(i)}$ by $\max(1, X_{m,j}^{(i)})$. This is very natural since the tail index is only related to the upper tail of the distribution F and should be estimated from those large observations.

To make an inference about γ , a condition stronger than (1) is required. Throughout this paper, we assume that there exists a function $A(t) \rightarrow 0$ as $t \rightarrow \infty$ such that

$$\lim_{t \rightarrow \infty} \frac{U(tx)/U(t) - x^\gamma}{A(t)} = x^\gamma \frac{x^\rho - 1}{\rho} \tag{6}$$

for all $x > 0$, where $U(x)$ is the inverse function of $\frac{1}{1-F(x)}$ and $\rho < 0$. This condition is more general than (3). In fact, if (3) holds, then one can verify that (6) holds with $A(t) = -\gamma(\beta\gamma - 1)dc^{-\beta\gamma}t^{1-\beta\gamma}$ and $\rho = 1 - \beta\gamma$.

We first present the asymptotic normality of our estimator.

Theorem 1 *Assume (6) holds. If*

$$k_n \rightarrow \infty, \quad \frac{k_n}{n} \rightarrow 0 \text{ and } k_n^{1/2} A(n/k_n) \rightarrow \delta \in (-\infty, \infty) \text{ as } n \rightarrow \infty, \quad (7)$$

then

$$k_n^{1/2}(\widehat{\gamma}_N(k_n) - \gamma) \xrightarrow{d} N\left(\delta b_r, \frac{\gamma^2}{r}\right), \quad (8)$$

where $b_r = \frac{1}{r\rho} \left(\sum_{j=1}^r \frac{\Gamma(j-\rho)}{(j-1)!} - \frac{\Gamma(r+1-\rho)}{(r-1)!}\right)$ and $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t} dt$ is the Gamma function.

Corollary 1 *In Theorem 1, if (7) holds with $\delta = 0$, then we have*

$$k_n^{1/2}(\widehat{\gamma}_N(k_n) - \gamma) \xrightarrow{d} N\left(0, \frac{\gamma^2}{r}\right).$$

When $r = 1$, that is, only the two largest observations are available within each block, we have the following corollary.

Corollary 2 *Assume that $r = 1$. Under the conditions of Theorem 1, we have*

$$k_n^{1/2}(\widehat{\gamma}_N(k_n) - \gamma) \xrightarrow{d} N(\delta\Gamma(1 - \rho), \gamma^2). \quad (9)$$

Remark 1 By using arguments similar to those in the proof of Theorem 1 we can show that

$$k_n^{1/2} \left(\frac{S_{k_n}}{k_n} - \frac{1}{1 + \gamma}\right) \xrightarrow{d} N\left(\frac{-\delta\Gamma(2 - \rho)}{(1 + \gamma)(1 + \gamma - \rho)}, \frac{\gamma^2}{(1 + \gamma)^2(1 + 2\gamma)}\right)$$

under the conditions of Theorem 1. Hence we have

$$k_n^{1/2}(\widehat{\gamma}_{DRP}(k_n) - \gamma) \xrightarrow{d} N\left(\frac{\delta(1 + \gamma)^2\Gamma(2 - \rho)}{(1 + \gamma - \rho)}, \frac{\gamma^2(1 + \gamma)^2}{(1 + 2\gamma)}\right). \quad (10)$$

This yields (4) when $\delta = 0$. Comparison of Eqs. (10) and (9) indicates that the asymptotic variance of our estimator is smaller than that of DPR’s estimator for all $\gamma > 0$ and the bias term for our new estimator is smaller than that of DPR’s estimator as well when $\delta \neq 0$.

Remark 2 It is well-known that for the Hill estimator it is held under the conditions of Theorem 1 that

$$k_n^{1/2}(\widehat{\gamma}_H(k_n) - \gamma) \xrightarrow{d} N(\delta(1 - \rho)^{-1}, \gamma^2).$$

See, e.g., [de Haan and Peng \(1998\)](#). In general, it is difficult to compare our estimator and the Hill estimator since the Hill estimator may not be applicable in case of incomplete data. When all observations are available, by assuming the same k_n in the Hill estimator $\widehat{\gamma}_H(k_n)$ and our estimator $\widehat{\gamma}_N(k_n)$, we see that the two estimators have the same asymptotic variances when $r = 1$. Hence, theoretically, the two estimators have similar large sample properties when the bias terms are asymptotically negligible when $r = 1$. To compare the small sample properties for the Hill estimator, the DPR estimator and our new estimator (when $r = 1$), we conducted a simulation study. Based on 10,000 random samples of size $n = 50$ from Fréchet(1) distribution (to be defined in Sect. 4) with $k_n = 10$ and $m_n = 5$, the mean squared errors for the Hill estimator, our new estimator and the DPR estimator are 0.114, 0.121 and 0.317, respectively. One can conclude that our new estimator and the Hill estimator are comparable and both are better than the DPR estimator in terms of mean squared error.

When $r > 1$, our new estimator has a smaller asymptotic variance than the Hill estimator. This phenomenon is due to the fact that different information is used for the two estimators. The conclusion can be different if we consider the bias terms and use the optimal asymptotic mean squared errors as a criterion. In order to calculate the optimal asymptotic mean squared errors for the estimators, let introduce a new function s which is positive, decreasing and regularly varying with index $2\rho - 1$ and satisfying

$$A^2(t) \sim \int_t^\infty s(u)du \quad \text{as } t \rightarrow \infty.$$

For the existence of such a function s see Proposition 1.7.12 of [Geluk and de Haan \(1987\)](#). Let s^- denote the inverse function of s . Let $k_n^{(H)}$ and $k_n^{(N)}$ denote the two values of k_n when the Hill estimator $\widehat{\gamma}_H(k_n)$ and our estimator $\widehat{\gamma}_N(k_n)$ achieve their optimal mean squared errors respectively. For convenience, we let $r = 1$ in our new estimator. From [de Haan and Peng \(1998\)](#) we have

$$k_n^{(H)} \sim \frac{n}{s^- \left(\frac{\gamma^2(1-\rho)^2}{n} \right)}.$$

Following the proof of Theorem 2 in [de Haan and Peng \(1998\)](#) we have

$$k_n^{(N)} \sim \frac{n}{s^- \left(\frac{\gamma^2}{n(\Gamma(1-\rho))^2} \right)}.$$

We can verify that under (3)

$$k_n^{(H)} \sim \left[(2d^2)^{-1} c^{2\beta\gamma} \beta^2 \gamma^2 (\beta\gamma - 1)^{-3} \right]^{\frac{1}{2\beta\gamma-1}} n^{\frac{2\beta\gamma-2}{2\beta\gamma-1}}$$

and

$$k_n^{(N)} \sim \left[(2d^2)^{-1} c^{2\beta\gamma} (\beta\gamma - 1)^{-3} \Gamma^{-2}(\beta\gamma) \right]^{\frac{1}{2\beta\gamma-1}} n^{\frac{2\beta\gamma-2}{2\beta\gamma-1}}. \tag{11}$$

It is easy to see that $k_n^{(H)}$ and $k_n^{(N)}$ are of the same order but $k_n^{(H)} > k_n^{(N)}$, and the optimal mean squared error for the Hill estimator is smaller. In this sense, the Hill estimator has a better performance than our new estimator when $r = 1$. On the other hand, for our new estimator, increasing the value of r decreases the asymptotic variance of the estimator, and we expect the optimal asymptotic mean squared error decreases with r . Meanwhile, increasing the value of r costs an increase of the asymptotic bias for the estimator if the bias is not negligible. Therefore, one has to be cautious in selecting the value of r in practice since increasing r doesn't implies a decrease in the asymptotic mean squared error when m_n is too small and r is too large.

Remark 3 It is possible to consider the situation when the numbers of random variables within the blocks are different and the numbers of the observations available for inference are also different. Assume there are k_n blocks of observations, $V_i, 1 \leq i \leq k_n$, and the i -th block V_i contains m_i observations and only the $r_i + 1$ largest order statistics $X_{m_i, j}^{(i)}, j = 1, \dots, r_i + 1$ are observed. Also assume that all random variables are independent and identically distributed with a heavy-tailed distribution (1). We define the estimator of γ as

$$\hat{\gamma}_N^*(k_n) = \frac{1}{\sum_{i=1}^{k_n} r_i} \sum_{i=1}^{k_n} \sum_{j=1}^{r_i} \left(\log X_{m_i, j}^{(i)} - \log X_{m_i, r_i+1}^{(i)} \right)$$

Then we can prove the asymptotic normality of $\hat{\gamma}_N^*(k_n)$. We present the result as the following Theorem 2 whose proof is similar to that of Theorem 1 and will be omitted.

Theorem 2 *Under the setup in Remark 3, if (6) holds and*

$$k_n \rightarrow \infty, \quad \frac{k_n}{n} \rightarrow 0 \quad \text{and} \quad \left(\sum_{i=1}^{k_n} r_i \right)^{1/2} A(q_n) \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

where $q_n = \min_{1 \leq i \leq k_n} (m_i / r_i) \rightarrow \infty$ as $n \rightarrow \infty$, then

$$\left(\sum_{i=1}^{k_n} r_i \right)^{1/2} (\hat{\gamma}_N^*(k_n) - \gamma) \xrightarrow{d} N(0, \gamma^2).$$

In order to demonstrate how the selection of k_n impacts the convergence rate of the normal approximation for our estimator $\widehat{\gamma}_N(k_n)$, we derive its Edgeworth expansion. Denote the standard normal density function and its cumulative distribution function by ϕ and Φ , respectively; that is

$$\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2} \text{ and } \Phi(x) = \int_{-\infty}^x \phi(t)dt \text{ for } x \in R.$$

Theorem 3 Assume (6) holds. If (2) holds and $\delta_n := k_n^{1/2} A(n/k_n) \rightarrow 0$ as $n \rightarrow \infty$, then

$$\begin{aligned} &P\left(\frac{k_n^{1/2}(\widehat{\gamma}_N(k_n) - \gamma)}{r^{-1/2}\gamma} \leq x\right) \\ &= \Phi(x) + \frac{1}{3(rk_n)^{1/2}}(1 - x^2)\phi(x) + a(\gamma, \rho)A(n/k_n)x\phi(x) \\ &\quad + c_n\phi(x) - \frac{b_r^2\delta_n^2}{r\gamma^2}x\phi(x) + o\left(\frac{1}{k_n^{1/2}} + A(n/k_n) + \delta_n^2\right) \end{aligned}$$

uniformly in x , where

$$\begin{aligned} a(\gamma, \rho) = &-\frac{1}{r\gamma\rho} \left(\sum_{j=1}^r \frac{\Gamma(j - \rho)}{(j - 1)!} \left(j - 1 - r + \sum_{t=j}^r \frac{t}{t - \rho} \right) \right. \\ &\left. + \frac{(r - 1)\Gamma(r + 1 - \rho)}{(r - 1)!} \right) \end{aligned}$$

and

$$c_n := \frac{k_n^{1/2} \left(r\gamma - E\left(\sum_{j=1}^r (\log X_{m,j}^{(1)} - \log X_{m,r+1}^{(1)})\right) \right)}{r^{1/2}\gamma} \sim -\frac{r^{1/2}b_r\delta_n}{\gamma}$$

as $n \rightarrow \infty$.

Remark 4 A $100(1 - \alpha)\%$ confidence interval for γ based on the normal approximation of $\widehat{\gamma}_N(k_n)$ is defined by

$$I_N(1 - \alpha) = \left\{ \gamma > 0 : \left| \frac{k_n^{1/2}(\widehat{\gamma}_N(k_n) - \gamma)}{r^{-1/2}\gamma} \right| < z_{\alpha/2} \right\},$$

that is,

$$I_N(1 - \alpha) = \left(\frac{\widehat{\gamma}_N(k_n)}{1 + z_{\alpha/2}(rk_n)^{-1/2}}, \frac{\widehat{\gamma}_N(k_n)}{1 - z_{\alpha/2}(rk_n)^{-1/2}} \right), \tag{12}$$

where $z_{\alpha/2}$ is the critical value of the standard normal distribution at level $\alpha/2$; that is, $1 - \Phi(z_{\alpha/2}) = \alpha/2$. According to Corollary 1, this confidence interval has an asymptotically correct coverage probability, that is, $P(\gamma_0 \in I_N(1 - \alpha)) \rightarrow 1 - \alpha$ as $n \rightarrow \infty$, where γ_0 is the true value of the parameter γ .

Remark 5 According to Theorem 3, the convergence rate of the coverage probability to the nominal level $1 - \alpha$ depends on the choice of k_n . It is easy to see that

$$P(\gamma_0 \in I_N(1 - \alpha)) - (1 - \alpha) = 2a(\gamma, \rho)z_{\alpha/2}\phi(z_{\alpha/2})A\left(\frac{n}{k_n}\right) - \frac{2b_r^2 z_{\alpha/2}\phi(z_{\alpha/2})}{r\gamma^2}\delta_n^2 + o\left(\frac{1}{k_n^{1/2}} + A\left(\frac{n}{k_n}\right) + \delta_n^2\right).$$

From Corollary 1, in order to achieve a faster convergence rate to γ (in terms of the asymptotic variance) for the estimator $\widehat{\gamma}_N(k_n)$, one should try to select a larger k_n . However, with the increase of k_n within certain range, the convergence rate of $\delta_n = k_n^{1/2}A(n/k_n)$ to zero is getting slower. From the above equation, we see that the leading error term for the convergence rate of the coverage probability for large k_n is δ_n^2 , up to a constant scale. This rate can be arbitrarily slow.

3 Empirical likelihood method

Set $Z_j^{(i)} = j(\log X_{m,j}^{(i)} - \log X_{m,j+1}^{(i)})$ for $j = 1, \dots, r$ and $i = 1, \dots, k_n$. As we will see from the proofs in the appendix, the limiting distribution for $Z_j^{(i)}$'s as n tends to infinity is exponential with mean γ . Thus, the mean of $Z_j^{(i)}$ is approximately equal to γ . We notice that the random variables $Z_j^{(i)}$'s are independent between blocks and the r random variables within each block are asymptotically independent. Therefore, the rk_n random variables $Z_j^{(i)}$'s are approximately iid with mean γ . We apply Owen's empirical likelihood method (Owen 1990) to construct the confidence interval or to test the hypothesis for the tail index γ .

Let $p = (p_1^{(1)}, \dots, p_r^{(1)}, \dots, p_1^{(k_n)}, \dots, p_r^{(k_n)})$ be a probability vector satisfying $\sum_{i=1}^{k_n} \sum_{j=1}^r p_j^{(i)} = 1$ and all $p_j^{(i)} \geq 0$. Then the empirical likelihood, evaluated at true value γ_0 for the tail index γ , is defined by

$$EL(\gamma_0) = \sup \left\{ \prod_{i=1}^{k_n} \prod_{j=1}^r p_j^{(i)} : \sum_{i=1}^{k_n} \sum_{j=1}^r p_j^{(i)} Z_j^{(i)} = \gamma_0 \right\}.$$

Then, by the method of Lagrange multipliers, we can easily get

$$p_j^{(i)} = \frac{1}{rk_n} \{1 + \lambda(Z_j^{(i)} - \gamma_0)\}^{-1}, \quad j = 1, \dots, r, \quad i = 1, \dots, k_n,$$

where λ is the solution to the equation

$$\frac{1}{rk_n} \sum_{i=1}^{k_n} \sum_{j=1}^r \frac{Z_j^{(i)} - \gamma_0}{1 + \lambda(Z_j^{(i)} - \gamma_0)} = 0. \tag{13}$$

On the other hand, $\prod_{i=1}^{k_n} \prod_{j=1}^r p_j^{(i)}$, subject to $\sum_{i=1}^{k_n} \sum_{j=1}^r p_j^{(i)} = 1$, attains its maximum $(rk_n)^{-rk_n}$ at $p_j^{(i)} = (rk_n)^{-1}$. So we define the empirical likelihood ratio at γ_0 by

$$l(\gamma_0) = \prod_{i=1}^{k_n} \prod_{j=1}^r (rk_n p_j^{(i)}) = \prod_{i=1}^{k_n} \prod_{j=1}^r \{1 + \lambda(Z_j^{(i)} - \gamma_0)\}^{-1},$$

and the corresponding empirical log-likelihood ratio statistic is defined as

$$\mathcal{L}(\gamma_0) = -2 \log l(\gamma_0) = 2 \sum_{i=1}^{k_n} \sum_{j=1}^r \log\{1 + \lambda(Z_j^{(i)} - \gamma_0)\},$$

where λ is the solution of (13).

The following theorem gives the asymptotic distribution of $\mathcal{L}(\gamma_0)$.

Theorem 4 *Under the conditions of Corollary 1 we have*

$$\mathcal{L}(\gamma_0) \xrightarrow{d} \chi_1^2,$$

where χ_1^2 denotes a chi-squared random variable with one degree of freedom, and γ_0 is the true value of the tail index γ .

According to the theorem, a $100(1 - \alpha)\%$ confidence interval for γ based on the empirical likelihood ratio statistic is determined by

$$I_E(1 - \alpha) = \{\gamma > 0 : \mathcal{L}(\gamma) < c(\alpha)\},$$

where $c(\alpha)$ is the α level critical value of a chi-squared distribution with one degree of freedom.

The asymptotic χ^2 calibrated empirical likelihood-based confidence intervals (or regions) have an undercoverage problem when the sample size is small; that is, these confidence intervals have a lower coverage probability than the nominal level. See, e.g., Owen (1988), Hall and La Scala (1990), Qin and Lawless (1994) and Tsao (2004). As pointed out by Tsao (2004), the distribution of the empirical likelihood ratio statistic has an atom at infinity, and the atom can be substantial if the sample size is not large. This causes the undercoverage. Tsao (2004) proposed a new method of calibration for the empirical likelihood-based confidence region for means. This new confidence region is computed by approximating the quantiles of the empirical likelihood ratio

statistic by an E-distribution. The E-distribution is defined as the distribution of the empirical likelihood ratio statistic for a normal mean. This method is easy to implement and significantly improves the coverage probabilities for small samples.

The undercoverage problem exists as well for the empirical likelihood-based confidence interval for the tail index since the number (rk_n) of observations employed in the inference of the tail index is very limited. Set $r = 1$. Consider a Fréchet distribution, $\text{Fréchet}(\alpha_0)$ (to be defined in Sect. 4), where $\alpha_0 > 0$. This is a heavy tailed distribution satisfying (3) with $\gamma = 1/\alpha_0$, $\beta = 2\alpha_0$, $c = 1$ and $d = -1/2$. From (11) we have $k_n^{(N)} \sim 72^{-1/3}n^{2/3}$. Note that the normal approximation based confidence interval (12) is valid only if $k_n = o(k_n^{(N)})$. However, even for a sample size as large as $n = 1,000$, since $72^{-1/3}1,000^{2/3} \doteq 24$, the theoretic range of k_n is very small. Therefore, the undercoverage is not unusual when the empirical likelihood method is used in the inference of the tail index. This stimulates us to pursue some new method of calibration that makes the empirical likelihood method more accurate.

Motivated by the work of Tsao (2004), Peng and Qi (2006a) proposed a similar calibration method for constructing confidence intervals for the tail index of a heavy-tailed distribution. Assume that $v \geq 2$ be an integer. Let E_1, \dots, E_v be *i.i.d.* random variables with unit exponential distribution. Peng and Qi (2006a) defined the following empirical likelihood ratio statistic

$$\text{ELR}(v) = 2 \sum_{i=1}^v \log(1 + \lambda'(E_i - 1)),$$

where λ' is the solution to the equation

$$\sum_{i=1}^v \frac{E_i - 1}{1 + \lambda'(E_i - 1)} = 0.$$

Let $c(v, \alpha)$ be the upper α -level critical value of the distribution of $\text{ELR}(v)$. After using this new critical value, the coverage probability for the confidence interval based on the empirical likelihood method has been greatly improved; see Peng and Qi (2006a). By using these new critical values, we can construct a $100(1 - \alpha)\%$ confidence interval for γ based on the empirical likelihood ratio statistic as follows:

$$I_E^*(1 - \alpha) = \{\gamma > 0 : \mathcal{L}(\gamma) < c(rk_n, \alpha)\}.$$

Through simulation Peng and Qi (2006a) estimated the $c(v, \alpha)$ for $10 \leq v \leq 200$. The results for $10 \leq v \leq 29$ are published in Table 1 in Peng and Qi (2006a) when $\alpha = 0.01, 0.05$ and 0.10 . For $30 \leq v \leq 200$, these critical values are fitted by regression lines

$$c(v, 0.01) = 6.6349 - \frac{4.56941}{\sqrt{v}} + \frac{98.98899}{v},$$

Table 1 Coverage probabilities of the two confidence intervals

k_n	Fréchet(1) $I_E^*(0.95)$	Fréchet(1) $I_N(0.95)$	Burr(0.5,1) $I_E^*(0.95)$	Burr(0.5,1) $I_N(0.95)$	Burr(1,0.5) $I_E^*(0.95)$	Burr(1,0.5) $I_N(0.95)$
10	0.9531	0.9550	0.9508	0.9540	0.9486	0.9555
15	0.9530	0.9509	0.9525	0.9500	0.9505	0.9513
20	0.9525	0.9501	0.9558	0.9478	0.9507	0.9527
25	0.9528	0.9507	0.9543	0.9461	0.9518	0.9533
30	0.9510	0.9484	0.9497	0.9434	0.9497	0.9512
35	0.9539	0.9488	0.9520	0.9412	0.9515	0.9528
40	0.9540	0.9474	0.9513	0.9373	0.9511	0.9516
45	0.9522	0.9437	0.9458	0.9305	0.9520	0.9525
50	0.9490	0.9405	0.9412	0.9271	0.9490	0.9483
55	0.9457	0.9377	0.9347	0.9170	0.9485	0.9471
60	0.9456	0.9368	0.9248	0.9043	0.9499	0.9488
65	0.9428	0.9339	0.9201	0.8999	0.9462	0.9460
70	0.9397	0.9316	0.9056	0.8858	0.9469	0.9456
75	0.9389	0.9286	0.8943	0.8736	0.9506	0.9505
80	0.9354	0.9239	0.8808	0.8617	0.9501	0.9475
85	0.9296	0.9191	0.8551	0.8350	0.9494	0.9472
90	0.9280	0.9183	0.8472	0.8291	0.9475	0.9468
95	0.9238	0.9108	0.8152	0.7964	0.9477	0.9466
100	0.9211	0.9075	0.8080	0.7886	0.9500	0.9458

$$c(v, 0.05) = 3.8415 - \frac{1.12486}{\sqrt{v}} + \frac{32.90613}{v}$$

and

$$c(v, 0.10) = 2.7055 - \frac{0.51269}{\sqrt{v}} + \frac{18.14242}{v},$$

where all three intercepts are the 1%, 5% and 10% critical values from a chi-squared distribution with one degree of freedom.

4 Simulation study

In this section we compare the performance of the confidence intervals based on the empirical likelihood ($I_E^*(1 - \alpha)$) and on the normal approximation ($I_N(1 - \alpha)$) in terms of both coverage probability and interval length with the following two types of cumulative distribution functions (cdfs): (i) the Fréchet cdf given by $F(x) = \exp(-x^{-\alpha_0})$ ($x > 0$), where $\alpha_0 > 0$ (notation: Fréchet(α_0)); (ii) the Burr cdf given by $F(x) = 1 - (1 + x^{\alpha_0})^{-\beta_0}$ ($x > 0$), where $\alpha_0 > 0, \beta_0 > 0$ (notation: Burr(α_0, β_0)). We choose $r = 1$, that is, we consider the case when only two largest observations within blocks are used for the inference.

First we drew 10,000 random samples of sample size $n = 1,000$ from the Fréchet (1), Burr(0.5,1) and Burr(1,0.5) distributions, and then computed the coverage probabilities for $I_E^*(0.95)$ and $I_N(0.95)$ for $k_n = 10, 15, \dots, 100$. These coverage probabilities are reported in Table 1.

Second, we drew 10,000 random samples of size $n = 1,000$ from the Fréchet(1), Burr(0.5,1) and Burr(1,0.5) distributions, and then computed the average lengths of the intervals $I_E^*(0.95)$ and $I_N(0.95)$ for $k_n = 10, 15, \dots, 100$; see Table 2 for the simulation results.

From Tables 1 and 2 we can conclude that the empirical likelihood method gives a better coverage accuracy than the normal approximation method, and it also generates shorter confidence intervals in general. In particular, for small values of k_n , the two methods have comparable coverage probabilities, but the empirical likelihood method yields significantly shorter confidence intervals. With the increase of k_n , both the methods produce confidence intervals whose coverage probabilities are lower than the nominal levels, but we have noticed that the coverage probabilities for the empirical likelihood method are about 2% higher. This is extremely important in practice since it is very difficult to determine the optimal value for k_n in general, and the method with certain robust properties is preferable against the choice of k_n .

Table 2 Average lengths of the two confidence intervals

k_n	Fréchet(1) $I_E^*(0.95)$	Fréchet(1) $I_N(0.95)$	Burr(.5,1) $I_E^*(0.95)$	Burr(0.5,1) $I_N(0.95)$	Burr(1,0.5) $I_E^*(0.95)$	Burr(1,0.5) $I_N(0.95)$
10	1.548	2.026	3.106	4.072	3.088	4.032
15	1.196	1.368	2.401	2.756	2.382	2.716
20	1.004	1.094	2.019	2.210	1.988	2.168
25	0.879	0.935	1.769	1.894	1.747	1.849
30	0.790	0.832	1.592	1.689	1.568	1.641
35	0.725	0.755	1.464	1.543	1.438	1.491
40	0.673	0.699	1.360	1.427	1.333	1.374
45	0.630	0.653	1.276	1.336	1.248	1.280
50	0.596	0.616	1.209	1.265	1.180	1.206
55	0.567	0.584	1.150	1.203	1.119	1.142
60	0.543	0.559	1.104	1.155	1.071	1.090
65	0.518	0.535	1.056	1.108	1.022	1.042
70	0.498	0.514	1.016	1.067	0.981	0.999
75	0.482	0.497	0.984	1.034	0.948	0.964
80	0.465	0.480	0.953	1.004	0.915	0.931
85	0.453	0.467	0.929	0.979	0.888	0.902
90	0.438	0.452	0.898	0.949	0.860	0.875
95	0.427	0.442	0.878	0.932	0.837	0.852
100	0.415	0.430	0.854	0.906	0.813	0.829

Acknowledgments The author would like to thank the referees for their valuable comments and suggestions that have led to significant improvements on the paper.

Appendix

Proof of Theorem 1. Condition (6) implies that $A(t)$ is a regularly varying function with index ρ . For detail see de Haan and Stadtmüller (1996). Then $A(t) = t^\rho \eta(t)$, where $\eta(t)$ is a slowly varying function. Without loss of generality, assume $|\eta(t)|$ is bounded away from 0 and ∞ on every compact subset of $[0, \infty)$. Then by applying Potter’s bounds to $\eta(t)$ we have for every $\delta > 0$ there exists a constant $c(\delta) > 0$ such that

$$\left| \frac{A(x)}{A(y)} \right| \leq c(\delta) \max \left(\left(\frac{x}{y} \right)^{\rho+\delta}, \left(\frac{x}{y} \right)^{\rho-\delta} \right) \text{ for all } x, y > 0. \tag{14}$$

See, e.g., Theorem 1.5.6 in Bingham et al. (1987).

Note that (6) is equivalent to

$$\lim_{t \rightarrow \infty} \frac{\log U(tx) - \log U(t) - \gamma \log x}{A(t)} = \frac{x^\rho - 1}{\rho}. \tag{15}$$

Applying Potter’s bounds to A we have

$$\left| \frac{\log U(tx) - \log U(t) - \gamma \log x}{A(t)} \right| \leq \frac{2}{|\rho|}$$

holds for all $x \geq 1$ and $t \geq t_0$, where $t_0 > 0$ is a constant. See also Lemma 1 in Drees (1998). Let $h(x) = \log U(x) - \gamma \log x$. Then (15) implies that for each $x > 0$,

$$\lim_{t \rightarrow \infty} \frac{h(tx) - h(t)}{A(t)} = \frac{x^\rho - 1}{\rho},$$

from which we conclude that $A(t) \sim \rho h(t)$ as $t \rightarrow \infty$. Therefore, $h(x)$ is a regularly varying function with index ρ . This implies $h(x) \rightarrow 0$ as $x \rightarrow \infty$ and

$$\left| \frac{\log U(tx) - \log U(t) - \gamma \log x}{A(t)} \right| = \left| \frac{h(tx) - h(t)}{A(t)} \right| \text{ is bounded for } x \geq 1, \\ 1 \leq t \leq t_0.$$

Since $\rho < 0$ we conclude that

$$\varepsilon(t, x) := \frac{\log U(tx) - \log U(t) - \gamma \log x}{A(t)} - \frac{x^\rho - 1}{\rho} \\ \text{is bounded for } x \geq 1, t \geq 1. \tag{16}$$

Assume $\{E_j, j \geq 1\}$ are iid random variables with a unit exponential distribution. It is easy to see that $U(e^{E_1})$ and X_1 have the same distribution function. Hence, $\{U(e^{E_j}), j \geq 1\}$ are iid random variables with the distribution F . Without loss of generality, we assume that $X_j = U(e^{E_j})$ for $j \geq 1$. Let $E_{m,1}^{(i)} \geq \dots \geq E_{m,m}^{(i)}$ be the order statistics of $E_{(i-1)m+1}, \dots, E_{im}$. Then $X_{m,j}^{(i)} = U(e^{E_{m,j}^{(i)}})$ for $j = 1, \dots, m$ and $i = 1, \dots, k_n$.

Set $J_m^{(i)} = E_{m,r+1}^{(i)}, K_j^{(i)} = E_{m,j}^{(i)} - E_{m,r+1}^{(i)}$ and $I_j^{(i)} = j(E_{m,j}^{(i)} - E_{m,j+1}^{(i)})$ for $j = 1, \dots, r$ and $i = 1, \dots, k_n$. In particular, set $K_{r+1}^{(i)} = 0$. Then we have $K_j^{(i)} = \sum_{t=j}^r I_t^{(i)}/t$. Furthermore, $\{I_j^{(i)}, 1 \leq j \leq r, i \geq 1\}$ are iid with a unit exponential distribution.

From (16) we have

$$\begin{aligned} \log(X_{m,j}^{(1)}) - \log(X_{m,r+1}^{(1)}) &= \log U(e^{J_m^{(1)}} e^{K_j^{(1)}}) - \log U(e^{J_m^{(1)}}) \\ &= \gamma K_j^{(1)} + A(e^{J_m^{(1)}}) \\ &\quad \times \frac{e^{\rho K_j^{(1)}} - 1}{\rho} + A(e^{J_m^{(1)}}) \varepsilon(J_m^{(1)}, e^{K_j^{(1)}}). \end{aligned} \tag{17}$$

We also need the following properties:

Fact 1. $J_m^{(1)}$ and $\{K_j^{(1)}, j = 1, \dots, r\}$ are independent;

Fact 2. $P(e^{J_m^{(1)}} \leq y) = \frac{m!}{(m-r-1)!r!} \int_0^{1-y^{-1}} t^{m-r-1}(1-t)^r dt$ for $y \geq 1$;

Fact 3. $\{(e^{J_m^{(1)}}/m)^t, m \geq r+1\}$ are uniformly integrable for each $t < 0$, and

$$\frac{e^{J_m^{(1)}}}{m} \xrightarrow{d} G(y) = \frac{1}{r!} \int_{1/y}^\infty e^{-t} t^r dt \quad \text{for } y > 0, \quad \text{as } n \rightarrow \infty,$$

where $G(y)$ has a density function $g(y) = y^{-r-2} \exp\{-y^{-1}\}/r!$ for $y > 0$.

Fact 1 is well-known in the literature. Fact 2 can be derived from equation (2.4.16) in Balakrishnan and Cohen (1991), and Fact 3 can be concluded from Fact 2. The details are omitted here.

From Fact 2 we get

$$\frac{A(e^{J_m^{(1)}})}{A(m)} \xrightarrow{d} J^\rho,$$

where J is a random variable with the distribution G . By applying (14) with $\delta = -\rho/2$ we have

$$\left| \frac{A(e^{J_m^{(1)}})}{A(m)} \right| \leq c \left(\frac{|\rho|}{2} \right) \left(\left(\frac{e^{J_m^{(1)}}}{m} \right)^{-1.5\rho} + \left(\frac{e^{J_m^{(1)}}}{m_n} \right)^{-0.5\rho} \right).$$

Then by using the dominated convergence theorem we have

$$\begin{aligned} \frac{E\left(A\left(e^{J_m^{(1)}}\right)\right)}{A(m)} &= E\left(\frac{A\left(e^{J_m^{(1)}}\right)}{A(m)}\right) \rightarrow E\left(J^\rho\right) = \frac{1}{r!} \int_0^\infty y^\rho g(y) dy \\ &= \frac{1}{r!} \int_0^\infty y^{r-\rho} e^{-y} dy = \frac{\Gamma(r+1-\rho)}{r!}. \end{aligned} \tag{18}$$

Similarly we have

$$\frac{E\left(A^2\left(e^{J_m^{(1)}}\right)\right)}{A^2(m)} \rightarrow \frac{\Gamma(r+1-2\rho)}{r!}. \tag{19}$$

Since $e^{J_m^{(1)}} \xrightarrow{P} \infty$ as $n \rightarrow \infty$, from (15) we conclude that $\varepsilon\left(e^{J_m^{(1)}}, e^{K_j^{(1)}}\right) \xrightarrow{P} 0$, where \xrightarrow{P} denotes the convergence in probability. By using Schwartz inequality, (16) and the dominated convergence theorem we get

$$\begin{aligned} k_n^{1/2} E\left|\varepsilon\left(e^{J_m^{(1)}}, e^{K_j^{(1)}}\right) A\left(e^{J_m^{(1)}}\right)\right| &\leq k_n^{1/2} E^{1/2}\left(\varepsilon\left(e^{J_m^{(1)}}, e^{K_j^{(1)}}\right)\right)^2 E^{1/2}\left(A\left(e^{J_m^{(1)}}\right)\right)^2 \\ &= o\left(k_n^{1/2} A(m)\right) = o(1). \end{aligned}$$

Recall that $Z_j^{(i)} = j\left(\log X_{m,j}^{(i)} - \log X_{m,j+1}^{(i)}\right)$ as defined in the beginning of Sect. 3. From (17) we have

$$Z_j^{(i)} = \gamma I_j^{(i)} + A(m) C_j^{(i)} + A(m) D_j^{(i)}, \tag{20}$$

where

$$C_j^{(i)} = \frac{A\left(e^{J_m^{(i)}}\right) j\left(e^{\rho K_j^{(i)}} - e^{\rho K_{j+1}^{(i)}}\right)}{A(m) \rho}$$

and

$$D_j^{(i)} = \frac{A\left(e^{J_m^{(i)}}\right)}{A(m)} j\left(\varepsilon\left(e^{J_m^{(i)}}, e^{K_j^{(i)}}\right) - \varepsilon\left(e^{J_m^{(i)}}, e^{K_{j+1}^{(i)}}\right)\right).$$

We have shown that $E|D_j^{(1)}| \rightarrow 0$ as $n \rightarrow \infty$ for $j = 1, \dots, r$. Since each $C_j^{(i)}$ is a product of two independent random variables, we conclude from the above proofs that $E\left(C_j^{(1)}\right)^2$ converges to a finite constant for $j = 1, \dots, r$.

By using $K_j^{(1)} = \sum_{t=j}^r I_t^{(1)}/t$, a straightforward calculation shows that

$$\tau_j := E \left(e^{\rho K_j^{(1)}} \right) = \prod_{t=j}^r \frac{t}{t - \rho} = \frac{r!}{\Gamma(r + 1 - \rho)} \frac{\Gamma(j - \rho)}{(j - 1)!}$$

if $1 \leq j \leq r$, and $\tau_j = 1$ if $j = r + 1$. Hence, from (18) we have

$$\begin{aligned} \beta_m &:= \sum_{j=1}^r E(C_j^{(1)}) \\ &= \frac{E \left(A \left(e^{J_m^{(1)}} \right) \right)}{A(m)} \sum_{j=1}^r j(\tau_j - \tau_{j+1})/\rho \\ &= \frac{E(A(e^{J_m^{(1)}}))}{A(m)} \left(\sum_{j=1}^r \tau_j - r \right) / \rho \\ &\sim \frac{1}{\rho} \left(\sum_{j=1}^r \frac{\Gamma(j - \rho)}{(j - 1)!} - \frac{\Gamma(r + 1 - \rho)}{(r - 1)!} \right) \\ &= rb_r. \end{aligned}$$

Since $\{I_j^{(i)}, j = 1, \dots, r, i = 1, \dots, k_n\}$ are iid random variables with a unit exponential distribution, and $\{\sum_{j=1}^r C_j^{(i)} - \beta_m, 1 \leq i \leq k_n\}$ are iid random variables with mean zero and bounded variances, we have

$$\frac{\sum_{i=1}^{k_n} \sum_{j=1}^r I_j^{(i)} - k_n r}{(k_n r)^{1/2}} \xrightarrow{d} N(0, 1),$$

and

$$\sum_{i=1}^{k_n} \sum_{j=1}^r C_j^{(i)} - k_n \beta_m = O_p(k_n^{1/2}).$$

Furthermore, we have

$$\sum_{i=1}^{k_n} \sum_{j=1}^r D_j^{(i)} = o_p(k_n).$$

since $E \sum_{i=1}^{k_n} \sum_{j=1}^r |D_j^{(i)}| = k_n \sum_{j=1}^r E |D_j^{(i)}| = o(k_n)$ as $n \rightarrow \infty$. Therefore, we obtain

$$\begin{aligned}
 k_n^{1/2}(\widehat{\gamma}_N(k_n) - \gamma) &= \frac{1}{k_n^{1/2}r} \sum_{i=1}^{k_n} \sum_{j=1}^r (Z_j^{(i)} - \gamma) \\
 &= \frac{\gamma \sum_{i=1}^{k_n} \sum_{j=1}^r I_j^{(i)} - k_n r}{r^{1/2}(k_n r)^{1/2}} + \frac{A(m) \sum_{i=1}^{k_n} \sum_{j=1}^r C_j^{(i)} - k_n \beta_m}{r k_n^{1/2}} \\
 &\quad + \frac{k_n^{1/2} A(m) \beta_m}{r} + k_n^{1/2} A(m) \frac{\sum_{i=1}^{k_n} \sum_{j=1}^r D_j^{(i)}}{k_n r} \\
 &= \frac{\gamma \sum_{i=1}^{k_n} \sum_{j=1}^r I_j^{(i)} - k_n r}{r^{1/2}(k_n r)^{1/2}} + \frac{k_n^{1/2} A(m) \beta_m}{r} + o_p(1) \\
 &\xrightarrow{d} N\left(\delta b_r, \frac{\gamma^2}{r}\right)
 \end{aligned}$$

as $n \rightarrow \infty$, proving (9). We complete the proof of the theorem. □

Proof of Theorem 3. From (20) we have

$$\begin{aligned}
 \sum_{j=1}^r Z_j^{(1)} &= \gamma \sum_{j=1}^r I_j^{(1)} + A(m) \sum_{j=1}^r C_j^{(1)} + A(m) \sum_{j=1}^r D_j^{(1)} \\
 &= \gamma \sum_{j=1}^r I_j^{(1)} + A(e^{J_m^{(1)}}) \frac{\sum_{j=1}^r e^{\rho K_j^{(1)}} - r}{\rho} + A(m) \sum_{j=1}^r D_j^{(1)}. \tag{21}
 \end{aligned}$$

One can see that for any $1 \leq j \leq r$ and $d > 0$, $E |D_j^{(1)}|^d \rightarrow 0$ as $n \rightarrow \infty$. One can also check that

$$\begin{aligned}
 E \left(\sum_{j=1}^r I_j^{(1)} \right) \left(\sum_{j=1}^r e^{\rho K_j^{(1)}} \right) &= \sum_{j=1}^r \tau_j \left(j - 1 + \sum_{t=j}^r \frac{t}{t - \rho} \right) \\
 &= \frac{r!}{\Gamma(r + 1 - \rho)} \sum_{j=1}^r \frac{\Gamma(j - \rho)}{(j - 1)!} \left(j - 1 + \sum_{t=j}^r \frac{t}{t - \rho} \right).
 \end{aligned}$$

Let $\mu_{n,i} = E \left(\sum_{j=1}^r Z_j^{(1)} \right)^i$ for $i = 1, 2$. Then from the proof of Theorem 1 we have as $n \rightarrow \infty$

$$\begin{aligned}
 \mu_{n,1} &= E \left(\sum_{j=1}^r Z_j^{(1)} \right) \\
 &= r\gamma + A(m) \sum_{j=1}^r \frac{\Gamma(r + 1 - \rho)}{r!} \frac{j(\tau_j - \tau_{j+1})}{\rho} + o(A(m))
 \end{aligned}$$

$$\begin{aligned}
 &= r\gamma + \frac{A(m)}{\rho} \left(\sum_{j=1}^r \frac{\Gamma(j - \rho)}{(j - 1)!} - \frac{\Gamma(r + 1 - \rho)}{(r - 1)!} \right) + o(A(m)) \\
 &= r\gamma + r b_r A(m) + o(A(m))
 \end{aligned} \tag{22}$$

and

$$\begin{aligned}
 \mu_{n,2} &= \frac{2\gamma A(m)}{\rho} \left(\sum_{j=1}^r \frac{\Gamma(j - \rho)}{(j - 1)!} \left(j - 1 + \sum_{t=j}^r \frac{t}{t - \rho} \right) - \frac{\Gamma(r + 1 - \rho)}{(r - 1)!} \right) \\
 &\quad + r(r + 1)\gamma^2 + o(A(m)).
 \end{aligned}$$

Set

$$\sigma_n^2 = E \left(\sum_{j=1}^r Z_j^{(1)} - \mu_{n,1} \right)^2 \quad \text{and} \quad v_n = E \left(\sum_{j=1}^r Z_j^{(1)} - \mu_{n,1} \right)^3.$$

Then

$$\begin{aligned}
 \sigma_n^2 &= r\gamma^2 + \frac{2\gamma A(m)}{\rho} \left(\sum_{j=1}^r \frac{\Gamma(j - \rho)}{(j - 1)!} \left(j - 1 - r + \sum_{t=j}^r \frac{t}{t - \rho} \right) \right. \\
 &\quad \left. + \frac{(r - 1)\Gamma(r + 1 - \rho)}{(r - 1)!} \right) + o(A(m))
 \end{aligned} \tag{23}$$

and

$$v_n = 2r\gamma^3 + o(1).$$

Since $\left\{ \sum_{j=1}^r Z_j^{(i)}, 1 \leq i \leq k_n \right\}$ are iid random variables with bounded moments. By following the proof of Theorem 5.22 in [Petrov \(1995\)](#), we have

$$\begin{aligned}
 &P \left(\frac{k_n^{1/2} (r\widehat{\gamma}_N(k_n) - \mu_{n,1})}{\sigma_n} \leq x \right) \\
 &= \Phi(x) + \frac{1}{6k_n^{1/2}} \frac{v_n}{\sigma_n^3} (1 - x^2)\phi(x) + o(k_n^{-1}) \\
 &= \Phi(x) + \frac{1}{3(rk_n)^{1/2}} (1 - x^2)\phi(x) + o(k_n^{-1/2})
 \end{aligned}$$

uniformly in x , which implies

$$\begin{aligned} P\left(\frac{k_n^{1/2}(\widehat{\gamma}_N(k_n) - \gamma)}{r^{-1/2}\gamma} \leq x\right) - \Phi(x) &= P\left(\frac{k_n^{1/2}(r\widehat{\gamma}_N(k_n) - \mu_{n,1})}{\sigma_n} \leq x\right) \\ &\leq \frac{k_n^{1/2}(r\gamma - \mu_{n,1})}{\sigma_n} + \frac{r^{1/2}\gamma}{\sigma_n}x - \Phi(x) \\ &= \Phi\left(\frac{k_n^{1/2}(r\gamma - \mu_{n,1})}{\sigma_n} + \frac{r^{1/2}\gamma}{\sigma_n}x\right) - \Phi(x) \\ &\quad + \frac{1}{3(rk_n)^{1/2}}(1 - x^2)\phi(x) + o(k_n^{-1/2}). \end{aligned}$$

By Taylor’s expansion, Theorem 3 follows from (22), (23) and the above equation. The details are omitted. This completes the proof. \square

Proof of Theorem 4. From (20) we conclude that

$$\max_{1 \leq j \leq r} \max_{1 \leq i \leq k_n} |Z_j^{(i)} - \gamma_0| = o_p(k_n^{1/2})$$

and

$$s_n^2 := \frac{1}{rk_n} \sum_{i=1}^{k_n} \sum_{j=1}^r (Z_j^{(i)} - \gamma_0)^2 \xrightarrow{p} \gamma_0^2.$$

The rest of the proof is very standard. Following the same lines in the proof in Sect. 11.2 in Owen (2001) we have

$$\lambda = O_p(k_n^{-1/2}),$$

where λ is the solution to (13). By setting $W_j^{(i)} = Z_j^{(i)} - \gamma_0$ we get

$$\left| \frac{\lambda^2}{rk_n} \sum_{i=1}^{k_n} \sum_{j=1}^r \frac{(W_j^{(i)})^3}{1 + \lambda W_j^{(i)}} \right| = O_p\left(\lambda^2 s_n^2 \max_{1 \leq j \leq r} \max_{1 \leq i \leq k_n} |W_j^{(i)}|\right) = o_p(k_n^{-1/2})$$

and then using (13) we obtain

$$\begin{aligned} 0 &= \frac{1}{rk_n} \sum_{i=1}^{k_n} \sum_{j=1}^r \frac{W_j^{(i)}}{1 + \lambda W_j^{(i)}} \\ &= \frac{1}{rk_n} \sum_{i=1}^{k_n} \sum_{j=1}^r W_j^{(i)} \left(1 - \lambda W_j^{(i)} + \frac{\lambda^2 (W_j^{(i)})^2}{1 + \lambda W_j^{(i)}}\right) \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{rk_n} \sum_{i=1}^{k_n} \sum_{j=1}^r W_j^{(i)} - \frac{\lambda}{rk_n} \sum_{i=1}^{k_n} \sum_{j=1}^r \left(W_j^{(i)}\right)^2 + \frac{\lambda^2}{rk_n} \sum_{i=1}^{k_n} \sum_{j=1}^r \frac{\left(W_j^{(i)}\right)^3}{1 + \lambda W_j^{(i)}} \\
 &= \widehat{\gamma}_N(k_n) - \gamma_0 - \lambda s_n^2 + o_p(k_n^{-1/2}).
 \end{aligned}$$

Therefore, we conclude that

$$\lambda = \frac{\widehat{\gamma}_N(k_n) - \gamma_0}{s_n^2} + o_p(k^{-1/2}).$$

Furthermore, since $\max_{1 \leq j \leq r} \max_{1 \leq i \leq k_n} \left| \lambda W_j^{(i)} \right| \xrightarrow{p} 0$ as $n \rightarrow \infty$, we have

$$\begin{aligned}
 R_n &:= \sum_{i=1}^{k_n} \sum_{j=1}^r \left(\log \left(1 + \lambda W_j^{(i)} \right) - \lambda W_j^{(i)} + \frac{1}{2} \left(\lambda W_j^{(i)} \right)^2 \right) \\
 &= O_p \left(\sum_{i=1}^{k_n} \sum_{j=1}^r \left| \lambda W_j^{(i)} \right|^3 \right) = O_p \left(\left| \lambda \right|^3 k_n s_n^2 \max_{1 \leq j \leq r} \max_{1 \leq i \leq k_n} \left| W_j^{(i)} \right| \right) = o_p(1).
 \end{aligned}$$

Hence, from Corollary 1

$$\begin{aligned}
 \mathcal{L}(\gamma_0) &= 2 \sum_{i=1}^{k_n} \sum_{j=1}^r \lambda W_j^{(i)} - \lambda^2 \sum_{i=1}^{k_n} \sum_{j=1}^r \left(W_j^{(i)}\right)^2 + R_n \\
 &= \frac{\gamma_0^2}{s_n^2} \frac{rk_n \left(\widehat{\gamma}_N(k_n) - \gamma_0\right)^2}{\gamma_0^2} + o_p(1) \\
 &\xrightarrow{d} \chi_1^2.
 \end{aligned}$$

This completes the proof of Theorem 4. □

References

Balakrishnan, N., Cohen, A. C. (1991). *Order statistics and inference*. Boston: Academic Press.

Bingham, N. H., Goldie, C. M., Teugels, J. L. (1987). *Regular variation*. New York: Cambridge University.

Chen, S. (1996). Empirical likelihood confidence intervals for nonparametric density estimation. *Biometrika*, 83, 329–341.

Chen, S., Hall, P. (1993). Smoothed empirical likelihood confidence intervals for quantiles. *The Annals of Statistics*, 21, 1166–1181.

Chen, S., Qin, Y.S. (2000). Empirical likelihood confidence intervals for local linear smoothers. *Biometrika*, 87, 946–953.

Davydov, Yu., Paulauskas, V., Račkauskas, A. (2000). More on P -stable convex sets in Banach spaces. *Journal of Theoretical Probability*, 13, 39–64.

De Haan, L., Peng, L. (1998). Comparison of tail index estimators. *Statistica Neerlandica*, 32, 60–70.

De Haan, L., Stadtmüller, U. (1996). Generalized regular variation of second order. *Journal of the Australian Mathematical Society (Series A)*, 61, 381–395.

- Dekkers, A. L. M., Einmahl, J. H. J., de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics*, 17, 1833–1855.
- Drees, H. (1998). On smooth statistical tail functionals. *Scandinavian Journal of Statistics*, 25, 187–210.
- Embrechts, P., Klüppelberg, C., Mikosch, T. (1997). *Modelling extremal events for insurance and finance*. Berlin: Springer.
- Gadeikis, K., Paulauskas, V. (2005) On the estimation of a changepoint in a tail index. *Lithuanian Mathematical Journal*, 45, 272–283.
- Geluk, J. L., de Haan, L. (1987). *Regular variation, extensions and Tauberian theorems*. CWI Tract, 40. Amsterdam: Centre for Mathematics and Computer Science.
- Hall, P., La Scala, L. (1990). Methodology and algorithms of empirical likelihood. *International Statistical Review*, 58, 109–127.
- Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3, 1163–1174.
- Kolaczyk, E. D. (1994). Empirical likelihood for generalized linear models. *Statistica Sinica*, 4, 199–218.
- Lu, J., Peng, L. (2002). Likelihood based confidence intervals for the tail index. *Extremes*, 5(4), 337–352.
- Lu, X., Qi, Y. (2004). Empirical likelihood for the additive risk model. *Probability and Mathematical Statistics*, 24, 419–431.
- Owen, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75, 237–249.
- Owen, A. (1990). Empirical likelihood regions. *The Annals of Statistics*, 18, 90–120.
- Owen, A. (1991). Empirical likelihood for linear models. *The Annals of Statistics*, 19, 1725–1747.
- Owen, A. (2001). *Empirical Likelihood*. London: Chapman and Hall.
- Paulauskas, V. (2003). A new estimator for a tail index. *Acta Applicandae Mathematicae*, 79, 55–67.
- Peng, L., Qi, Y. (2004). Estimating the first- and second-order parameters of a heavy-tailed distribution. *Australian & New Zealand Journal of Statistics*, 46, 305–312.
- Peng, L., Qi, Y. (2006a). A new calibration method of constructing empirical likelihood-based confidence intervals for the tail index. *Australian & New Zealand Journal of Statistics*, 48, 59–66.
- Peng, L., Qi, Y. (2006b). Confidence regions for high quantiles of a heavy-tailed distribution. *The Annals of Statistics*, 34 (4), 1964–1986.
- Petrov, V. V. (1995). *Limit theorems of probability theory: sequences of independent random variables*. Oxford: Clarendon Press.
- Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 3, 119–131.
- Qin, G., Jing, B. Y. (2001a). Empirical likelihood for censored linear regression. *Scandinavian Journal of Statistics*, 28, 661–673.
- Qin, G., Jing, B. Y. (2001b). Censored partial linear models and empirical likelihood. *Journal of Multivariate Analysis*, 78, 37–61.
- Qin, G., Jing, B. Y. (2001c). Empirical likelihood for Cox regression model under random censorship. *Communications in Statistics Simulation and Computation*, 30, 79–90.
- Qin, J., Lawless, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics*, 22, 300–325.
- Segers, J., Teugels, J. (2000). Testing the Gumbel hypothesis by the Galton's ratio. *Extremes*, 3(3), 291–303.
- Tsao, M. (2004). A new method of calibration for the empirical loglikelihood ratio. *Statistics & Probability Letters*, 68, 305–314.
- Wang, Q. H., Jing, B. Y. (2003). Empirical likelihood for partial linear models. *Annals of the Institute of Statistical Mathematics*, 55, 585–595.
- Wang, Q. H., Li, G. (2002). Empirical likelihood semiparametric regression analysis under random censorship. *Journal of Multivariate Analysis*, 83, 469–486.
- Wang, Q. H., Rao, J. N. K. (2002). Empirical likelihood-based inference in linear errors-in-covariables models with validation data. *Biometrika*, 89, 345–358.