Extremes (2008) 11:281–302
DOI 10.1007/s10687-008-0058-2

A two-step estimator of the extreme value index

Chen Zhou

Received: 2 April 2007 / Revised: 30 January 2008 /
Accepted: 5 February 2008 / Published online: 30 March 2008

Abstract In this paper, we build a two-step estimator $\hat{\gamma}_{\text{STEP}}$, which satisfies $\sqrt{k}(\hat{\gamma}_{\text{STEP}} - \hat{\gamma}_{ML}) \stackrel{P}{\to} 0$, where $\hat{\gamma}_{ML}$ is the well-known maximum likelihood estimator of the extreme value index. Since the two-step estimator $\hat{\gamma}_{\text{STEP}}$ can be calculated easily as a function of the observations, it is much simpler to use in practice. By properly choosing the first step estimator, such as the *Pickands estimator*, we can even get a shift and scale invariant estimator with the above property.

Keywords Extreme value index · Maximum likelihood · Shift and scale invariant estimator

AMS 2000 Subject Classification 62G05

1 Introduction

Let $X_1, X_2, \cdots$ be independent and identically distributed (i.i.d.) random variables from a distribution function F. Suppose F is in the domain of attraction of an extreme value distribution, i.e. there exist constants $a_n > 0$ and b_n, such that,

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G_\gamma(x),$$

The author thanks Laurens de Haan for motivating this work and giving helpful comments. The author also thanks two anonymous referees for their useful comments.

C. Zhou (✉)
Tinbergen Institute H09-32, Erasmus University Rotterdam,
P.O. Box 1738, 3000DR Rotterdam, The Netherlands
e-mail: zhou@few.eur.nl

for all $1+\gamma x>0$, where $G_\gamma(x)=\exp(-(1+\gamma x)^{-1/\gamma})$ is the corresponding extreme value distribution and $\gamma\in\mathbb{R}$ is called the *extreme value index* (see Gnedenko 1943). Commonly, it is denoted as $F\in D(G_\gamma)$.

There are a few characterizations of the necessary and sufficient condition for a distribution function F belonging to the domain of attraction. One of them is via the "excess distribution function" as in Balkema and de Haan (1974). Denote the excess distribution function as

$$F_t(x):=P(X-t\le x|X>t)=\frac{F(t+x)-F(t)}{1-F(t)}.$$

Then $F\in D(G_\gamma)$ is equivalent to

$$\lim_{t\to x^*}F_t(x\sigma(t))=H_\gamma(x):=1-(1+\gamma x)^{-1/\gamma},$$

for all $1+\gamma x>0$, where $\sigma(t)$ is a positive function and x^* is the right endpoint of F, i.e. $x^*=\sup\{x|F(x)<1\}$. The distribution function H_γ is the so-called *generalized Pareto distribution (GPD) function*. Intuitively, the distribution function F belongs to the domain of attraction if and only if the excesses above a high threshold are asymptotically generalized Pareto distributed.

This characterization creates several possible ways to deal with a major issue in Extreme Value Theory: estimating the extreme value index γ.

Denote $X_{n,1}\le\cdots\le X_{n,n}$ as the order statistics of $X_1,X_2,\cdots,X_n$. For a suitable sequence such that $k_n\to\infty$, $k_n/n\to 0$ as $n\to\infty$, the k_nth upper order statistic $X_{n,n-k_n}$ may take the place of the "high threshold". Then $X_{n,n}-X_{n,n-k_n},\cdots X_{n,n-k_n+1}-X_{n,n-k_n}$ can be recognized as the order statistics of the empirical excesses above the high threshold. Thus, together as a new sample, they are asymptotically generalized Pareto distributed. In the rest of the paper, without declaration, we briefly use k instead of k_n.

Theoretically, the 1/2 and 3/4-quantiles of the GPD can be calculated as $(2^\gamma-1)/\gamma$ and $(4^\gamma-1)/\gamma$; empirically, they can be estimated as $X_{n,n-[k/2]}-X_{n,n-k}$ and $X_{n,n-[k/4]}-X_{n,n-k}$ respectively. This creates the quantile estimator, suggested by Pickands (1975), as follows.

$$\hat{\gamma}_P=\frac{1}{\log 2}\log\frac{X_{n,n-[k/4]}-X_{n,n-[k/2]}}{X_{n,n-[k/2]}-X_{n,n-k}}.$$

When $\gamma>0$, the function $\sigma(t)$ can be chosen as $\sigma(t)=\gamma t$ and $x^*=+\infty$. Thus, the condition on the excess distribution function can be rewritten as

$$\lim_{t\to+\infty}P\left(\frac{X}{t}\le x|X>t\right)=1-x^{-1/\gamma}.$$

Therefore, similar to the above intuition, by taking $X_{n,n-k}$ as the "high threshold", we get that, as $n\to\infty$, the excess ratios $X_{n,n}/X_{n,n-k},\cdots X_{n,n-k+1}/X_{n,n-k}$ form a sample of order statistics from a Pareto distribution. By fitting the Pareto distribution with the maximum likelihood procedure, Hill (1975) suggested the so-called Hill estimator as

$$\hat{\gamma}_H=\frac{1}{k}\sum_{i=0}^{k-1}\log X_{n,n-i}-\log X_{n,n-k}.$$

The Hill estimator is only applied for positive γ. In order to deal with a general $\gamma \in \mathbb{R}$, Dekkers et al. (1989) introduced the moment estimator,

$$\hat{\gamma}_M = \hat{\gamma}_H + 1 - \frac{1}{2}\left(1 - \frac{\hat{\gamma}_H^2}{M_n^{(2)}}\right)^{-1},$$

where

$$M_n^{(2)} = \frac{1}{k}\sum_{i=0}^{k-1}(\log X_{n,n-i} - \log X_{n,n-k})^2.$$

An alternative way to have a general estimator was suggested by Beirlant et al. (1996) as the UH estimator. By denoting $UH_{n,n-i} = X_{n,n-i}\hat{\gamma}_H$ for $i = 0, 1, \cdots, k$, the estimator

$$\hat{\gamma}_{UH} = \frac{1}{k}\sum_{i=0}^{k-1}\log UH_{n,n-i} - \log UH_{n,n-k}$$

is valid for all $\gamma \in \mathbb{R}$.

Although $\hat{\gamma}_H$ and $\hat{\gamma}_M$ perform reasonably well for γ positive and $\gamma \in \mathbb{R}$ respectively, they both have the disadvantage that they are not shift invariant. The estimator $\hat{\gamma}_P$ is a shift and scale invariant estimator, but according to de Haan and Peng (1998), it does not perform as well as the other two in most cases.

A shift invariant estimator needs to be constructed from the excesses instead of the excess ratios. Hosking and Wallis (1987) proposed the probability-weighted moment (PWM) estimator by assigning different weights to the excesses. It is defined as

$$\hat{\gamma}_{PWM} = \frac{P_n - 4R_n}{P_n - 2R_n},$$

where

$$P_n = \frac{1}{k}\sum_{i=0}^{k-1}X_{n,n-i} - X_{n,n-k},$$

and

$$R_n = \frac{1}{k}\sum_{i=0}^{k-1}\frac{i}{k}(X_{n,n-i} - X_{n,n-k}).$$

In order to have consistency, the PWM estimator can only be applied for $\gamma < 1$. To obtain the asymptotic normality, γ should be further restricted as $\gamma < 1/2$. Note that the PWM estimator is shift and scale invariant.

Similar to the idea of the Hill estimator, Smith (1987) applied the maximum likelihood procedure to fit the GPD with a general γ, which leads to the maximum likelihood estimator of the extreme value index. The likelihood equations are as follows:

$$\sum_{i=1}^{k} \frac{1}{\gamma^2} \log\left(1 + \frac{\gamma}{\sigma}(X_{n,n-i+1} - X_{n,n-k})\right)$$
$$- \left(\frac{1}{\gamma} + 1\right) \frac{(1/\sigma)(X_{n,n-i+1} - X_{n,n-k})}{1 + (\gamma/\sigma)(X_{n,n-i+1} - X_{n,n-k})} = 0,$$
$$\sum_{i=1}^{k} \left(\frac{1}{\gamma} + 1\right) \frac{(\gamma/\sigma)(X_{n,n-i+1} - X_{n,n-k})}{1 + (\gamma/\sigma)(X_{n,n-i+1} - X_{n,n-k})} = k, \tag{1}$$

(the equations for $\gamma = 0$ should be interpreted as the limit when $\gamma \to 0$). For $\gamma \neq 0$, they can be simplified to

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

When $\gamma > -1/2$, the maximum likelihood estimators for the extreme value index and the scale, $\hat{\gamma}_{ML}$ and $\hat{\sigma}_{ML}$, are obtained by solving these equations.

In order to obtain the asymptotic normality for most of the estimators of the extreme value index, further restrictive condition on F is required. de Haan and Stadtmüller (1996) proposed the generalized second order condition as follows. Denote $F^{\leftarrow}$ as the generalized inverse of F. Assume that there exist measurable, locally bounded functions $a, \Phi : (0, 1) \to (0, \infty)$ and $\Psi : (0, \infty) \to \mathbb{R}$, such that for all $x > 0$

$$\lim_{t \downarrow 0} \frac{(F^{\leftarrow}(1 - tx) - F^{\leftarrow}(1 - t))/a(t) - (x^{-\gamma} - 1)/\gamma}{\Phi(t)} = \Psi(x). \tag{2}$$

According to de Haan and Stadtmüller (1996), $|\Phi|$ is $-\rho$-varying at 0 for some $\rho \leq 0$, and

$$\Psi(x) = \begin{cases} (x^{-(\gamma+\rho)} - 1)/(\gamma + \rho), & \rho < 0 \\ -x^{-\gamma} \log(x)/\gamma, & \gamma \neq 0, \rho = 0 \\ \log^2(x), & \gamma = \rho = 0. \end{cases}$$

Under the generalized second order condition, for $\gamma > -1/2$, Drees et al. (2004) proved asymptotic normality of the maximum likelihood estimator by assuming that the sequence k_n satisfies

$$\Phi(k_n/n) = O(k_n^{-1/2}), \tag{3}$$

as $n \to \infty$. The asymptotic normality is a direct consequence of the following theorem (Theorem 2.1 in Drees et al. 2004).

Theorem 1.1 *Assume condition (2) holds for some $\gamma > -1/2$, and the sequence k_n satisfies Eq. 3. Then the system of likelihood equations (1) has a sequence of solutions $(\hat{\gamma}_n, \hat{\sigma}_n)$ that verifies*

$$k^{1/2}(\hat{\gamma}_n - \gamma) - \frac{(\gamma+1)^2}{\gamma} k^{1/2} \Phi\left(\frac{k}{n}\right) \int_0^1 (t^\gamma - (2\gamma+1)t^{2\gamma})\Psi(t)dt$$

$$\xrightarrow{d} \frac{(\gamma+1)^2}{\gamma} \int_0^1 (t^\gamma - (2\gamma+1)t^{2\gamma})(W(1) - t^{-(\gamma+1)}W(t))dt, \tag{4}$$

$$k^{1/2}\left(\frac{\hat{\sigma}_n}{a(k/n)} - 1\right) - \frac{(\gamma+1)^2}{\gamma} k^{1/2} \Phi\left(\frac{k}{n}\right) \int_0^1 ((\gamma+1)(2\gamma+1)t^{2\gamma} - t^\gamma)\Psi(t)dt$$

$$\xrightarrow{d} \frac{(\gamma+1)^2}{\gamma} \int_0^1 ((\gamma+1)(2\gamma+1)t^{2\gamma} - t^\gamma)(W(1) - t^{-(\gamma+1)}W(t))dt, \tag{5}$$

as $n \to \infty$, and the convergence holds jointly with the same standard Brownian motion W. For $\gamma = 0$ these equations should be interpreted as their limits when $\gamma \to 0$.

From this theorem, Eq. 4 can be rewritten as

$$k^{1/2}(\hat{\gamma}_{ML} - \gamma) = \frac{(\gamma+1)^2}{\gamma} \int_0^1 \left(t^\gamma - (2\gamma+1)t^{2\gamma}\right) L_n(t)dt + o_p(1), \tag{6}$$

where

$$L_n(t) = W_n(1) - t^{-(\gamma+1)}W_n(t) + k^{1/2}\tilde{\Phi}\left(\frac{k}{n}\right)\Psi(t), \tag{7}$$

$$W_n(t) = k^{-1/2}W(kt),$$

$\tilde{\Phi}(k/n) \sim \Phi(k/n)$ as $n \to \infty$ and W is a standard Brownian motion which implies that W_n is also a standard Brownian motion. Then the integral of the two parts $W_n(1) - t^{-(\gamma+1)}W_n(t)$ and $k^{1/2}\tilde{\Phi}\left(\frac{k}{n}\right)\Psi(t)$ lead to a mean-zero normal distribution and the asymptotic bias respectively, which completes the proof of the asymptotic normality. Notice that the asymptotic bias depends on the second order parameter ρ and the asymptotic variance can be calculated as shown in Remark 2.1 and Corollary 2.1 in Drees et al. (2004).

It is clear that the maximum likelihood estimator is shift and scale invariant. Meanwhile, it performs well for $\gamma > -1/2$. But it still has a disadvantage: there is no explicit formula for this estimator. It is always given by solving the likelihood equations, but there is even no guarantee for the existence of a solution. The existence was stated in Drees et al. (2004) but there is no proof of that statement in the paper. The numerical way to find a solution of these equations had been discussed in Grimshaw (1993).

An alternative way to deal with this problem is to find an approximate solution for the likelihood equations, i.e. an explicit estimator such that the difference between the maximum likelihood estimator and the alternative estimator is approximately 0. As an example, Theorem 2.2 and Remark 2.4 in Drees et al. (2004) proved that, when $\gamma = 0$, with the generalized second order condition and assumption on the sequence k as in Eq. 3, we have that

$$k^{1/2}(\hat{\gamma}_* - \hat{\gamma}_{ML}) \xrightarrow{P} 0,$$

where

$$\hat{\gamma}_* = 1 - \frac{1}{2}\left(1 - \frac{(m_n^{(1)})^2}{m_n^{(2)}}\right)^{-1},$$

and

$$m_n^{(j)} = \frac{1}{k}\sum_{i=1}^{k}(X_{n,n-i+1} - X_{n,n-k})^j, \quad j = 1, 2.$$

In this case, $\hat{\gamma}_*$, a shift and scale invariant estimator with explicit formula, is close enough to the maximum likelihood estimator. But this is only for a special case $\gamma = 0$. Can we find such kind of estimator in general case? In this paper, a two-step estimator is established which gives a positive answer to this question. The idea is similar to the PWM estimator which is based on the weighted sum of the excesses. However, in the two-step estimator, the weights are determined in prior according to a pre-estimation of the extreme value index. This is similar to the UH estimator where the extreme value index is pre-estimated by the Hill estimator.

In Section 2, Theorem 2.2 shows that, the two-step estimator is close enough to the maximum likelihood estimator. By suitable choice in the first step, we may get a shift and scale invariant estimator. Simulations are given in Section 3. Section 4 concludes the paper.

2 Result and Proof

We start with stating the following theorem in Drees (1998).

Theorem 2.1 *Given Eq. 2 with $\gamma > -1/2$ and Eq. 3, one can find a probability space and define on that space a Brownian Motion W and a sequence of stochastic processes Q_n such that*

(1) *For each n, $(Q_n(t))_{t\in[0,1]} \overset{d}{=} (X_{n,n-[kt]})_{t\in[0,1]}$;*

(2) *There exist functions* $\tilde{a}(k/n) = a(k/n)(1 + o(\Phi(k/n)))$ *and* $\tilde{\Phi}(k/n) \sim \Phi(k/n)$ *such that, for all* $\varepsilon > 0$,

$$\sup_{t\in[0,1]} t^{\gamma+1/2+\varepsilon} \left| \frac{Q_n(t) - F^{\leftarrow}(1-k/n)}{\tilde{a}(k/n)} - \left(\frac{t^{-\gamma}-1}{\gamma} - t^{-(\gamma+1)} \frac{W(kt)}{k} + \tilde{\Phi}\left(\frac{k}{n}\right)\Psi(t) \right) \right| = o_p(k^{-1/2}) + o_p\left(\tilde{\Phi}\left(\frac{k}{n}\right)\right), \tag{8}$$

as $n \to \infty$.

The following notation is introduced in order to shorten the proof in the rest of the paper.

$$Y_n(t) = k^{1/2}\left(\frac{Q_n(t) - Q_n(1)}{\tilde{a}(k/n)} - \frac{t^{-\gamma}-1}{\gamma} \right). \tag{9}$$

When $\gamma = 0$, $\frac{t^{-\gamma}-1}{\gamma}$ should be read as $-\log t$.

With the notations in Eqs. 7 and 9, a direct consequence of Theorem 2.1 is the following lemma.

Lemma 2.1 *Suppose Eqs. 2 and 3 hold. Then for all* $\varepsilon > 0$,

$$Y_n(t) = L_n(t) + o_p(1)t^{-(\gamma+1/2+\varepsilon)}, \tag{10}$$

as $n \to \infty$, *where the* o_p*−term is uniform for* $t \in [0, 1]$.

Our purpose is to find an estimator which is close enough to the maximum likelihood estimator. Hence, it should have the same asymptotic structure as the right side of Eq. 6. In order to do so, we should connect $L_n(t)$ with the observations. From Lemma 2.1, intuitively, we can substitute $L_n(t)$ by $Y_n(t)$, which is partially based on the observations.

There are still two remaining difficulties. First, the asymptotic structure in Eq. 6 is an integral of the product of $L_n(t)$ and $t^{\gamma} - (2\gamma+1)t^{2\gamma}$. To replace $L_n(t)$ by $Y_n(t)$, we have to study the functional approximation between them, i.e. whether the asymptotic structure is close to the integral of the product of $Y_n(t)$ and such kind of function. Secondly, there is still the parameter γ unknown. We solve this problem by using a first step estimator of γ, and show that it is still close enough.

To deal with the first difficulty, we study the weighted integral of the process $Y_n(t)$ on [0, 1]. For the weight function, we focus on (pseudo) power functions. Suppose a continuous function $f : (0, 1] \to \mathbb{R}$ satisfies

$$|f(t)| = O(t^{\gamma-\delta}) \qquad \text{when } t \to 0^+, \tag{11}$$

for some $0 < \delta < 1/2$. Then, we can choose a positive ε such that $\varepsilon + \delta < 1/2$. By applying Eq. 10 for this ε, we get

$$\int_0^1 f(t)(Y_n(t) - L_n(t))dt = o_p(1).$$

By checking

$$\int_0^1 f(t)\Psi(t)dt < \infty$$

for $f(t)$ satisfying condition (11), it is insured that,

$$\int_0^1 f(t)L_n(t)dt$$

is bounded in probability as $n \to \infty$. Hence we have the following corollary.

Corollary 2.1 *With the same conditions as in Theorem 2.1, given a continuous function* $f : (0, 1] \to \mathbb{R}$ *satisfying Eq. 11 for some* $0 < \delta < 1/2$*, we have that*

$$k^{1/2}\left(\int_0^1 f(t)\frac{Q_n(t) - Q_n(1)}{\tilde{a}(k/n)}dt - \int_0^1 f(t)\frac{t^{-\gamma} - 1}{\gamma}dt\right) = \int_0^1 f(t)L_n(t)dt + o_p(1).$$

Next let us consider (for $\gamma > -1/2$) a continuous function $g : (0, 1] \to \mathbb{R}$ satisfying

$$|g(t)| = O(t^{2\gamma-\delta}) \qquad \text{when } t \to 0^+, \tag{12}$$

For some $0 < \delta < (\gamma \wedge 0) + 1/2$. We can find positive numbers ε and δ such that $2\varepsilon + \delta < (\gamma \wedge 0) + 1/2$. We write

$$\begin{aligned}
&k^{1/2}\left(\int_0^1 g(t)\left(\frac{Q_n(t) - Q_n(1)}{\tilde{a}(k/n)}\right)^2 dt - \int_0^1 g(t)\left(\frac{t^{-\gamma} - 1}{\gamma}\right)^2 dt\right) \\
&= \left(\int_0^{k^{-1}} + \int_{k^{-1}}^1\right) g(t)\left(k^{-1/2}Y_n(t) + 2\frac{t^{-\gamma} - 1}{\gamma}\right) Y_n(t)dt \\
&= I_1 + I_2.
\end{aligned}$$

Because $k^{-1/2}Y_n(t) = o_p(t^{-(\gamma+\varepsilon)})$ uniformly for all $t \in [k^{-1}, 1]$, and $\int_0^1 g(t)t^{-(2\gamma+1/2+2\varepsilon)}dt$ is finite, by applying Eq. 10 for this ε, we get

$$\begin{aligned}
I_2 &= \int_{k^{-1}}^1 g(t)\left(2\frac{t^{-\gamma} - 1}{\gamma} + o_p\left(t^{-(\gamma+\varepsilon)}\right)\right)(L_n(t) + o_p(1)t^{-(\gamma+1/2+\varepsilon)})dt \\
&= \int_{k^{-1}}^1 2g(t)\frac{t^{-\gamma} - 1}{\gamma}L_n(t)dt + o_p(1) \\
&= \int_0^1 2g(t)\frac{t^{-\gamma} - 1}{\gamma}L_n(t)dt + o_p(1).
\end{aligned}$$

The last equality comes from that the final integration is bounded in probability as $n \to \infty$.

For the rest part, I_1, it is going to be proved that $I_1 = o_p(1)$. On the interval $[0, k^{-1})$, for any $0 < \eta < 1$

$$\left(\frac{Q_n(t) - Q_n(1)}{\tilde{a}(k/n)}\right)^2 = \left(\frac{Q_n(\eta k^{-1}) - Q_n(1)}{\tilde{a}(k/n)}\right)^2 = o_p(k^{2(\gamma+\varepsilon)}).$$

Note that

$$\delta < 2\gamma + 1 \Rightarrow \int_0^{k^{-1}} g(t)dt = O(k^{-(2\gamma-\delta+1)}).$$

Finally, we have that

$$k^{1/2}\int_0^{k^{-1}} g(t)\left(\frac{Q_n(t) - Q_n(1)}{\tilde{a}(k/n)}\right)^2 dt = o_p(k^{2\varepsilon+\delta-1/2}) = o_p(1).$$

Meanwhile,

$$k^{1/2}\int_0^{k^{-1}} g(t)\left(\frac{t^{-\gamma}-1}{\gamma}\right)^2 dt = O(k^{\delta-1/2}) = o_p(1),$$

which completes the proof of $I_1 = o_p(1)$.

This conclusion is rewritten as the following corollary.

Corollary 2.2 *With the same conditions as in Theorem* 2.1, *given a continuous function* $g : (0, 1] \to \mathbb{R}$ *satisfying Eq.* 12, *for some* $0 < \delta < 1/2 + \gamma$, *we have that*

$$k^{1/2}\left(\int_0^1 g(t)\left(\frac{Q_n(t) - Q_n(1)}{\tilde{a}(k/n)}\right)^2 dt - \int_0^1 g(t)\left(\frac{t^{-\gamma}-1}{\gamma}\right)^2 dt\right)$$
$$= \int_0^1 2g(t)\frac{t^{-\gamma}-1}{\gamma}L_n(t)dt + o_p(1).$$

In order to obtain the right side of Eq. 4, we introduce the following functions to apply Corollary 2.1 and Corollary 2.2.

$$f_1(\gamma, t) = \frac{1}{\sqrt{2\gamma+1}}t^{\gamma},$$
$$f_2(\gamma, t) = \frac{\partial f_1(\gamma, t)}{\partial \gamma},$$
$$f_3(\gamma, t) = t^{2\gamma},$$
$$f_4(\gamma, t) = \frac{\partial f_3(\gamma, t)}{\partial \gamma}.$$

Denote weighted moments with true γ as

$$h_n^{(1)}(\gamma) = \int_0^1 f_1(\gamma, t)(Q_n(t) - Q_n(1))dt,$$

$$h_n^{(2)}(\gamma) = \int_0^1 f_3(\gamma, t)(Q_n(t) - Q_n(1))^2 dt.$$

The asymptotic behavior of $h_n^{(1)}(\gamma)$ and $h_n^{(2)}(\gamma)$ can be obtained by applying Corollary 2.1 and Corollary 2.2 for f_1 and f_3 as follows

$$k^{1/2}\left(\frac{h_n^{(1)}(\gamma)}{\tilde{a}(k/n)} - \frac{1}{\sqrt{2\gamma+1}(\gamma+1)}\right) = \int_0^1 f_1(\gamma, t)L_n(t)dt + o_p(1), \quad (13)$$

$$k^{1/2}\left(\frac{h_n^{(2)}(\gamma)}{(\tilde{a}(k/n))^2} - \frac{2}{(2\gamma+1)(\gamma+1)}\right) = \int_0^1 2f_3(\gamma, t)\frac{t^{-\gamma}-1}{\gamma}L_n(t)dt + o_p(1). \quad (14)$$

They lead to the asymptotical behavior of their combination as

$$\begin{aligned} &k^{1/2}\left(\frac{(h_n^{(1)}(\gamma))^2}{h_n^{(2)}(\gamma)} - \frac{1}{2(\gamma+1)}\right) \\ &= k^{1/2}\left(\frac{h_n^{(1)}(\gamma)}{\tilde{a}(k/n)} - \frac{1}{\sqrt{2\gamma+1}(\gamma+1)}\right)\cdot 2\frac{\sqrt{2\gamma+1}}{2} \\ &\quad + k^{1/2}\left(\frac{h_n^{(2)}(\gamma)}{(\tilde{a}(k/n))^2} - \frac{2}{(2\gamma+1)(\gamma+1)}\right)\cdot(-1)\frac{2\gamma+1}{4} + o_p(1) \\ &= \sqrt{2\gamma+1}\int_0^1 f_1(\gamma, t)L_n(t)dt - \frac{2\gamma+1}{2}\int_0^1 f_3(\gamma, t)\frac{t^{-\gamma}-1}{\gamma}L_n(t)dt + o_p(1) \\ &= -\frac{1}{2\gamma}\int_0^1 \left(t^{\gamma} - (2\gamma+1)t^{2\gamma}\right)L_n(t)dt + o_p(1). \quad (15) \end{aligned}$$

Define an auxiliary random variable as

$$\varphi(\gamma) := \frac{1}{2}\frac{h_n^{(2)}(\gamma)}{(h_n^{(1)}(\gamma))^2} - 1.$$

From Eq. 15 and

$$\frac{1}{2(\varphi(\gamma)+1)} = \frac{\left(h_n^{(1)}(\gamma)\right)^2}{h_n^{(2)}(\gamma)},$$

we get the asymptotical behavior of $\varphi(\gamma)$ as

$$k^{1/2}(\varphi(\gamma) - \gamma) = \frac{(\gamma+1)^2}{\gamma}\int_0^1 \left(t^{\gamma} - (2\gamma+1)t^{2\gamma}\right)L_n(t)dt + o_p(1).$$

Compared to Eq. 6, we proved that

$$k^{1/2}(\varphi(\gamma) - \hat{\gamma}_{ML}) = o_p(1).$$

Now the only problem is that, the real parameter γ is still a part of the auxiliary random variable $\varphi(\gamma)$. We introduce a first step estimator to replace it and try to keep the asymptotic property at the same time. By rewriting the final estimator in explicit form, we define the two-step estimator as

Definition 2.1 Suppose a first step estimator of the extreme value index, $\hat{\gamma}^{(1)}$, is given, which uses the largest k order statistics. Assume the first step estimator approaches γ in speed $1/\sqrt{k}$, i.e.

$$k^{1/2}\left(\hat{\gamma}^{(1)} - \gamma\right) \xrightarrow{d} N, \tag{16}$$

where N is a random variable with a suitable distribution. For all of the suggested estimators above, it follows a normal distribution.

If $\hat{\gamma}^{(1)} > -1/2$, define the weights $w_i^{(j)}$ as

$$w_i^{(j)} = \int_{\frac{i-1}{k}}^{\frac{i}{k}} t^{j\hat{\gamma}^{(1)}} dt = \frac{1}{j\hat{\gamma}^{(1)} + 1}\left(\left(\frac{i}{k}\right)^{j\hat{\gamma}^{(1)}+1} - \left(\frac{i-1}{k}\right)^{j\hat{\gamma}^{(1)}+1}\right), \tag{17}$$

for $j = 1, 2$ and $i = 1, \cdots, k$. Then, define the weighted moments as

$$WM_n^{(j)} = \sum_{i=1}^{k} w_i^{(j)}(X_{n,n-i+1} - X_{n,n-k})^j, \quad j = 1, 2. \tag{18}$$

Finally, define the estimator

$$\hat{\gamma}_{\text{STEP}} = \frac{2\hat{\gamma}^{(1)} + 1}{2}\frac{WM_n^{(2)}}{\left(WM_n^{(1)}\right)^2} - 1, \tag{19}$$

as the *two-step estimator* of the extreme value index.

The following theorem shows that this estimator is close enough to the maximum likelihood estimator.

Theorem 2.2 *Assume Eq. 2 holds and the sequence k satisfies Eq. 3. If* $\gamma > -1/2$*, then*

$$k^{1/2}(\hat{\gamma}_{\text{STEP}} - \hat{\gamma}_{ML}) \xrightarrow{P} 0.$$

Proof of Theorem 2.2 We have already proved that the auxiliary random variable $\varphi(\gamma)$ is close enough to the maximum likelihood estimator. Since the two-step estimator is in fact $\varphi(\hat{\gamma}^{(1)})$, in order to complete the proof of the theorem, we only need to show that the difference between the two-step estimator and the auxiliary random variable is also negligible, i.e.

$$\sqrt{k}(\varphi(\gamma) - \hat{\gamma}_{\text{STEP}}) = o_p(1). \tag{20}$$

From Eq. 18 and the definition of $Q_n(t)$ in Theorem 2.1, by changing γ into its estimator $\hat{\gamma}^{(1)}$ in $h_n^{(1)}(\gamma)$ and $h_n^{(2)}(\gamma)$, we can rewrite the weighted moments in the definition as

$$\left(WM_n^{(1)}, WM_n^{(2)}\right) \stackrel{d}{=} \left(h_n^{(1)}\left(\hat{\gamma}^{(1)}\right)\sqrt{2\hat{\gamma}^{(1)}+1}, h_n^{(2)}\left(\hat{\gamma}^{(1)}\right)\right).$$

According to the definition of $\hat{\gamma}_{\mathrm{STEP}}$ in Eq. 19, it is clear that

$$\frac{1}{2(\hat{\gamma}_{\mathrm{STEP}}+1)} = \frac{1}{2\hat{\gamma}^{(1)}+1}\frac{\left(WM_n^{(1)}\right)^2}{WM_n^{(2)}} \stackrel{d}{=} \frac{\left(h_n^{(1)}\left(\hat{\gamma}^{(1)}\right)\right)^2}{h_n^{(2)}\left(\hat{\gamma}^{(1)}\right)}. \tag{21}$$

□

This is a slight change from $\varphi(\gamma)$ in sense of the following lemma.

Lemma 2.2 *Under the conditions of Theorem* 2.2, *we have*

$$k^{1/2}\left(\frac{\left(h_n^{(1)}\left(\hat{\gamma}^{(1)}\right)\right)^2}{h_n^{(2)}\left(\hat{\gamma}^{(1)}\right)} - \frac{\left(h_n^{(1)}(\gamma)\right)^2}{h_n^{(2)}(\gamma)}\right) = o_p(1). \tag{22}$$

Proof of Lemma 2.2 We start with the Taylor expansion of $f_1(\hat{\gamma}^{(1)}, t)$,

$$f_1(\hat{\gamma}^{(1)}, t) = f_1(\gamma, t) + \left(\hat{\gamma}^{(1)}-\gamma\right) f_2(\gamma, t) + \frac{\left(\hat{\gamma}^{(1)}-\gamma\right)^2}{2}\frac{\partial f_2(s,t)}{\partial s}\Big|_{s=\eta_n(t)},$$

where $\eta_n(t)$ is a random variable depending on n and t, but always between γ and $\hat{\gamma}^{(1)}$. Since $\hat{\gamma}^{(1)} \stackrel{P}{\to} \gamma$ as $n \to \infty$, we have $\eta_n(t) \stackrel{P}{\to} \gamma$ uniformly for all $t \in (0, 1]$. Then, for any $\delta > 0$,

$$\frac{\partial f_2(s,t)}{\partial s}\Big|_{s=\eta_n(t)} = O_p\left(t^{\eta_n(t)}(\log t)^2\right) = t^{\gamma-\delta}O_p(1),$$

when $n \to \infty$, the O_p-term is uniform for all $t \in (0, 1]$. So we can continue with the Taylor expansion as follows,

$$\begin{aligned} f_1\left(\hat{\gamma}^{(1)}, t\right) &= f_1(\gamma, t) + \left(\hat{\gamma}^{(1)}-\gamma\right) f_2(\gamma, t) + \frac{\left(\hat{\gamma}^{(1)}-\gamma\right)^2}{2}t^{\gamma-\delta}O_p(1) \\ &= f_1(\gamma, t) + \left(\hat{\gamma}^{(1)}-\gamma\right) f_2(\gamma, t) + \left(\hat{\gamma}^{(1)}-\gamma\right)t^{\gamma-\delta}o_p(1), \end{aligned}$$

as $n \to \infty$. In this expansion, the o_p−term is also uniform for $t \in (0, 1]$. Then, by using this expansion and applying Corollary 2.1 for f_2 satisfying Eq. 12, we get that

$$
\begin{aligned}
& k^{1/2}\left(\frac{h_n^{(1)}\left(\hat{\gamma}^{(1)}\right)-h_n^{(1)}(\gamma)}{\tilde{a}(k/n)}+\left(\hat{\gamma}^{(1)}-\gamma\right)\frac{2\gamma^2+6\gamma+3}{(\gamma+1)^2(2\gamma+1)^{3/2}}\right) \\
&= k^{1/2}\left(\int_0^1\left(f_1\left(\hat{\gamma}^{(1)},t\right)-f_1(\gamma,t)\right)\frac{Q_n(t)-Q_n(1)}{\tilde{a}(k/n)}dt\right. \\
&\qquad\left.-\left(\hat{\gamma}^{(1)}-\gamma\right)\int_0^1 f_2(\gamma,t)\frac{t^{-\gamma}-1}{\gamma}dt\right) \\
&= k^{1/2}\left\{\int_0^1\left[f_2(\gamma,t)\left(\hat{\gamma}^{(1)}-\gamma\right)+\left(\hat{\gamma}^{(1)}-\gamma\right)t^{\gamma-\delta}o_p(1)\right]\frac{Q_n(t)-Q_n(1)}{\tilde{a}(k/n)}dt\right. \\
&\qquad\left.-\left(\hat{\gamma}^{(1)}-\gamma\right)\int_0^1 f_2(\gamma,t)\frac{t^{-\gamma}-1}{\gamma}dt\right\} \qquad (23) \\
&= \left(\hat{\gamma}^{(1)}-\gamma\right)\left(\int_0^1 f_2(\gamma,t)L_n(t)dt\right)+k^{1/2}(\hat{\gamma}^{(1)}-\gamma)o_p(1)+o_p(1) \\
&= o_p(1). \qquad (24)
\end{aligned}
$$

A similar relationship between $h_n^{(2)}(\hat{\gamma}^{(1)})$ and $h_n^{(2)}(\gamma)$ is given as

$$
k^{1/2}\left(\frac{h_n^{(2)}\left(\hat{\gamma}^{(1)}\right)-h_n^{(2)}(\gamma)}{(\tilde{a}(k/n))^2}+\left(\hat{\gamma}^{(1)}-\gamma\right)\frac{8\gamma^2+24\gamma+12}{(\gamma+1)^2(2\gamma+1)^2}\right)=o_p(1). \qquad (25)
$$

From Eqs. 13 and 14, we have that as $n \to \infty$,

$$
\frac{h_n^{(1)}(\gamma)}{\tilde{a}(k/n)} \xrightarrow{P} \frac{1}{\sqrt{2\gamma+1}(\gamma+1)},
$$

and

$$
\frac{h_n^{(2)}(\gamma)}{(\tilde{a}(k/n))^2} \xrightarrow{P} \frac{2}{(2\gamma+1)(\gamma+1)}.
$$

Considering with Eqs. 24 and 25, we also have that as $n \to \infty$,

$$
\begin{aligned}
\frac{h_n^{(1)}(\hat{\gamma}^{(1)})}{\tilde{a}(k/n)} &\xrightarrow{P} \frac{1}{\sqrt{2\gamma+1}(\gamma+1)}, \\
\frac{h_n^{(2)}(\hat{\gamma}^{(1)})}{(\tilde{a}(k/n))^2} &\xrightarrow{P} \frac{2}{(2\gamma+1)(\gamma+1)}.
\end{aligned}
$$

Finally, by using Eqs. 24, 25 and the four equations above, we can calculate that

$$\begin{aligned}
&k^{1/2}\left(\frac{\left(h_n^{(1)}\left(\hat{\gamma}^{(1)}\right)\right)^2}{h_n^{(2)}\left(\hat{\gamma}^{(1)}\right)}-\frac{\left(h_n^{(1)}(\gamma)\right)^2}{h_n^{(2)}(\gamma)}\right)\\
&\quad=k^{1/2}\left(\frac{h_n^{(1)}\left(\hat{\gamma}^{(1)}\right)-h_n^{(1)}(\gamma)}{\tilde{a}(k/n)}\right)\left(\frac{h_n^{(1)}\left(\hat{\gamma}^{(1)}\right)+h_n^{(1)}(\gamma)}{\tilde{a}(k/n)}\right)\frac{1}{h_n^{(2)}\left(\hat{\gamma}^{(1)}\right)/(\tilde{a}(k/n))^2}\\
&\qquad+k^{1/2}\left(\frac{h_n^{(2)}\left(\hat{\gamma}^{(1)}\right)-h_n^{(2)}(\gamma)}{(\tilde{a}(k/n))^2}\right)\\
&\qquad\cdot(-1)\frac{\left(h_n^{(1)}(\gamma)/\tilde{a}(k/n)\right)^2}{\left(h_n^{(2)}(\gamma)/(\tilde{a}(k/n))^2\right)\left(h_n^{(2)}\left(\hat{\gamma}^{(1)}\right)/(\tilde{a}(k/n))^2\right)}\\
&\quad=k^{1/2}\left(\hat{\gamma}^{(1)}-\gamma\right)\Bigg(-\frac{2\gamma^2+6\gamma+3}{(\gamma+1)^2(2\gamma+1)^{3/2}}\cdot 2\frac{\sqrt{2\gamma+1}}{2}\\
&\qquad+\frac{8\gamma^2+24\gamma+12}{(\gamma+1)^2(2\gamma+1)^2}\frac{2\gamma+1}{4}\Bigg)+o_p(1)\\
&\quad=o_p(1).
\end{aligned}$$

The lemma has been proved. □

Lemma 2.2 shows that

$$\sqrt{k}\left(\frac{1}{2(\varphi(\gamma)+1)}-\frac{1}{2(\hat{\gamma}_{\mathrm{STEP}}+1)}\right)=o_p(1),$$

which implies Eq. 20 as a direct consequence. Hence we complete the proof of Theorem 2.2.

Remark 2.1 From the definition of the two-step estimator, it is clear that, if the first step estimator is shift and scale invariant, the final estimator should be the same. So we can choose the Pickands' estimator $\hat{\gamma}_P$ mentioned in Section 1 as the first step estimator. Although the Pickands' estimator itself does not perform very well in most of the cases, after the two-step procedure, it will be close enough to the maximum likelihood estimator.

Remark 2.2 Obviously, we can also use the final two-step estimator as the first step estimator, and iterate the same procedure once more. It results in a three-step estimator. If the first step estimator is shift and scale invariant, so is the final three-step one. Simulations in Section 3 will show that the three-step estimator is even more accurate.

Remark 2.3 The weighted moments $WM_n^{(j)}$ $(j = 1, 2)$ can be represented in another way as

$$WM_n^{(1)} = \frac{1}{\hat{\gamma}^{(1)} + 1} \sum_{i=1}^{k} \left(\frac{i}{k}\right)^{\hat{\gamma}^{(1)}+1} (X_{n,n-i+1} - X_{n,n-i}),$$

$$WM_n^{(2)} = \frac{1}{2\hat{\gamma}^{(1)} + 1} \sum_{i=1}^{k} \left(\frac{i}{k}\right)^{2\hat{\gamma}^{(1)}+1} (X_{n,n-i+1} - X_{n,n-i}) \times \left(X_{n,n-i+1} + X_{n,n-i} - 2X_{n,n-k}\right).$$

3 Simulations

Simulations have been done for three cases: γ positive, negative and $\gamma = 0$. We also try to simulate for both large and small sample size.

For large sample size simulation, a sample with sample size 10,000 from a certain distribution is generated. In case $\gamma > 0$, we choose the Cauchy distribution which has a positive extreme value index $\gamma = 1$ and a second order index $\rho = -2$. In case $\gamma = 0$, we choose standard Normal distribution. Both the extreme value index γ and the second order index ρ are equal to 0. In case $\gamma < 0$, we choose the Reversed Burr distribution. Such a distribution function is given as

$$F(x) = 1 - \left(\frac{4}{4 + x^{-2}}\right)^2, \qquad x < 0.$$

It belongs to the domain of attraction of the extreme value distribution with extreme value index $\gamma = -1/4$ and $\rho = -1/2$.

We choose the Pickands estimator as the first step estimator, calculate the two-step estimator and the maximum likelihood estimator. The three-step estimators described in Remark 2.2 are also demonstrated in the figures. In order to study the sensitivity of the first step estimator, we also use the moment estimator as the first step estimator for the same simulated samples. For γ positive, they are presented separately in Figs. 1 and 2. For $\gamma = 0$ and γ negative, the results are shown in Figs. 3–6.

From these figures we observe that, the two-step estimator is close enough to the maximum likelihood estimator. Hence, it can be a good substitute of the maximum likelihood estimator with explicit formula. Furthermore, the three-step estimator is closer, i.e. it will be better to iterate the procedure for more steps. With the moment estimator as the first step estimator, the performance of the two-step is improved. Hence, it will be helpful to choose an accurate first step estimator, even if not location invariant.

Secondly, we turn to small sample size. We generate 500 samples with sample size 1,000 each, calculate the maximum likelihood estimator and two-step estimator in each sample, and take the average of the estimators among the samples. We also calculate the mean squared error (MSE) for both

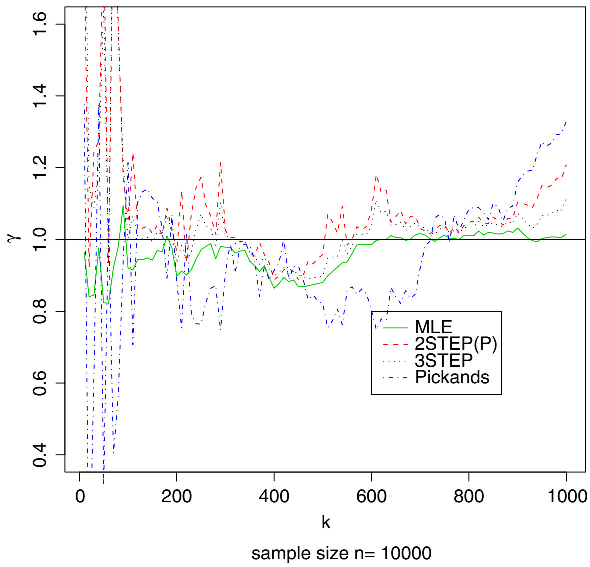


Fig. 1 Cauchy ($\gamma = 1, \rho = -2$)

the maximum likelihood estimator and the two-step estimator. Denote the calculated estimator as $\hat{\gamma}_i$ for sample i, where $1 \leq i \leq 500$. Then the mean squared error is defined as follows

$$MSE = \frac{1}{500} \sum_{i=1}^{500} (\hat{\gamma}_i - \gamma)^2,$$

where γ is the known extreme value index.

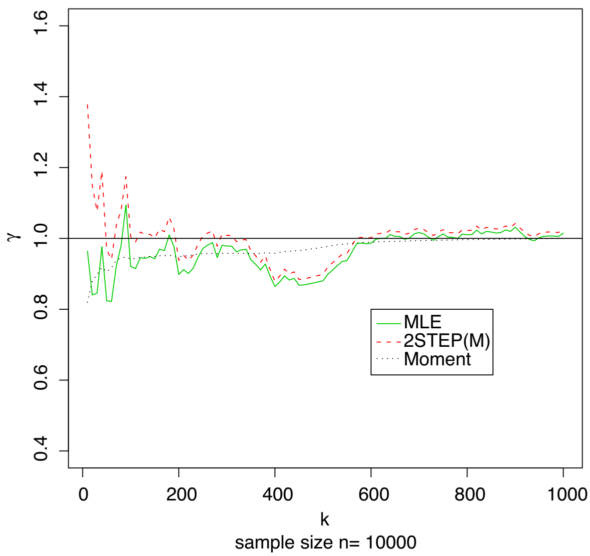
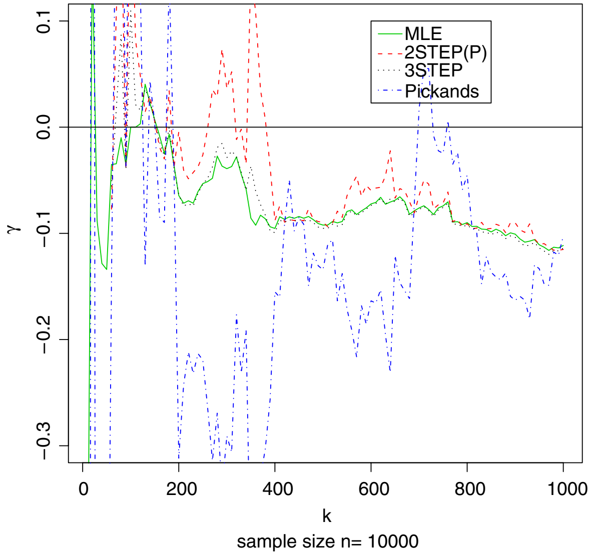


Fig. 2 Cauchy ($\gamma = 1, \rho = -2$)

Fig. 3 Normal ($\gamma = 0, \rho = 0$)



Because the Pickands estimator does not perform very well for small sample size, we use the moment estimator as the first step estimator. In these simulations, the three-step estimator is ignored. For γ positive, we change to the Pareto distribution with $\gamma = 1/2$, i.e. the distribution function is $F(x) = 1 - 1/x^2$. In this case $\rho = -\infty$. Together with the large sample size simulation study, all these simulation studies cover the entire range of all possible ρ, i.e. $\rho \in [-\infty, 0]$.

Fig. 4 Normal ($\gamma = 0, \rho = 0$)

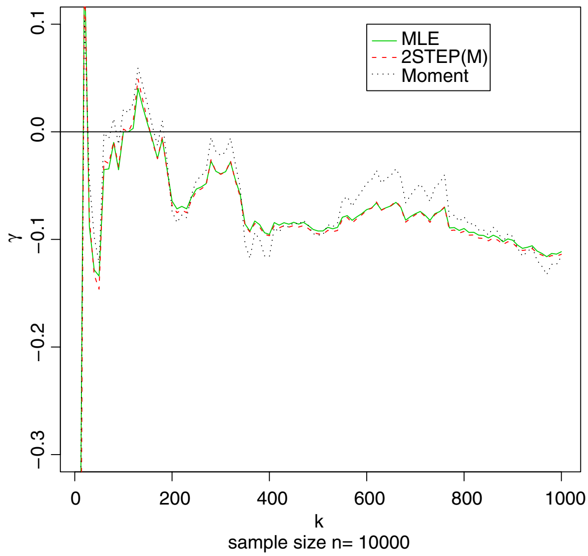
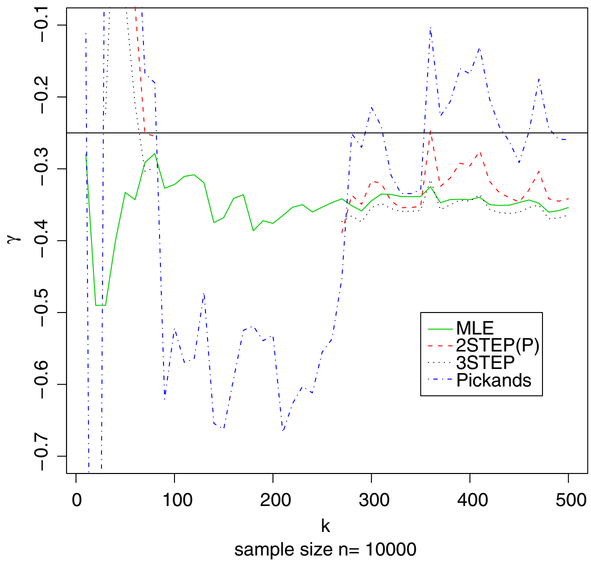


Fig. 5 Reversed Burr ($\gamma = -1/4$, $\rho = -1/2$)



The averaged estimations for γ positive are shown in Fig. 7 with its corresponding MSE pictures in Fig. 8. Figures 9, 10, 11 and 12 present the results for $\gamma = 0$ and γ negative.

From the multi-sample simulations, we again observe that the two-step estimator is close enough to the maximum likelihood estimator, while the MSE is in a comparable level.

Fig. 6 Reversed Burr ($\gamma = -1/4$, $\rho = -1/2$)

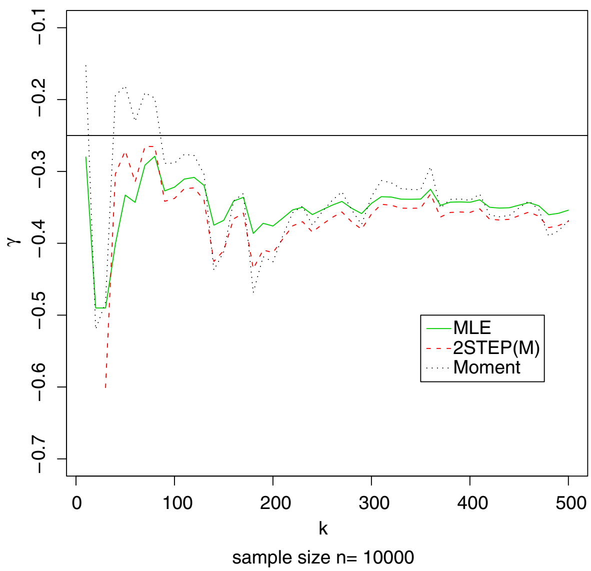
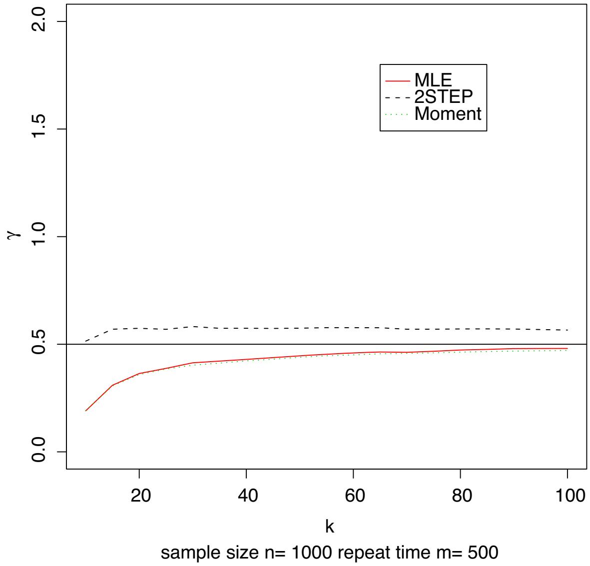


Fig. 7 Pareto ($\gamma = 1/2, \rho = -\infty$)



Furthermore, we also make simulations for even smaller sample size, for example, 100. The results are no longer comparable with the maximum likelihood estimator. From the simulation study, we recommend the two-step estimator for relatively larger sample size, for instance, at least 1,000.

Fig. 8 MSE: Pareto ($\gamma = 1/2, \rho = -\infty$)

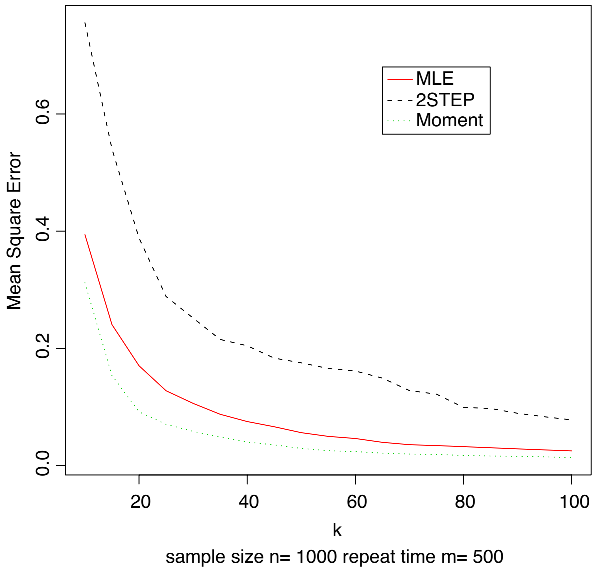


Fig. 9 Normal ($\gamma = 0, \rho = 0$)

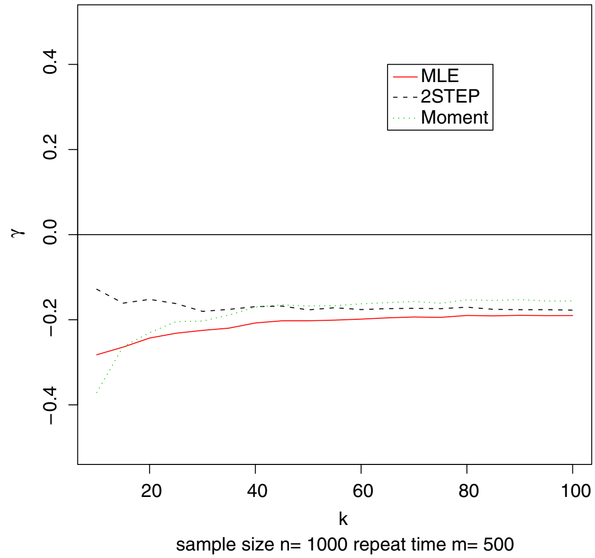


Fig. 10 MSE: Normal ($\gamma = 0, \rho = 0$)

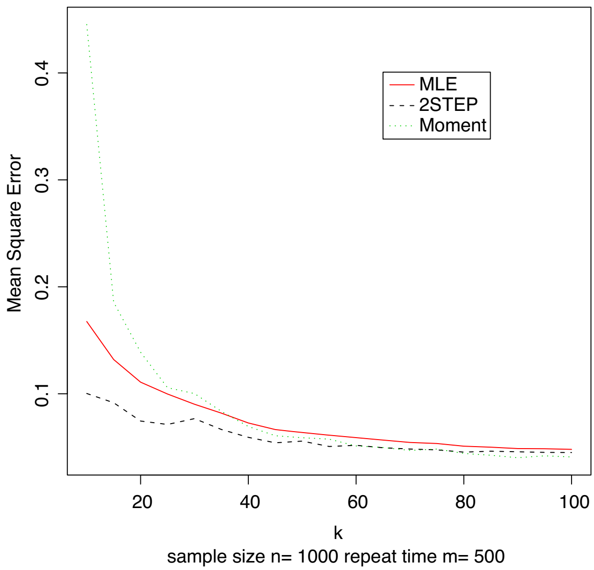


Fig. 11 R-Burr ($\gamma = -1/4$, $\rho = -1/2$)

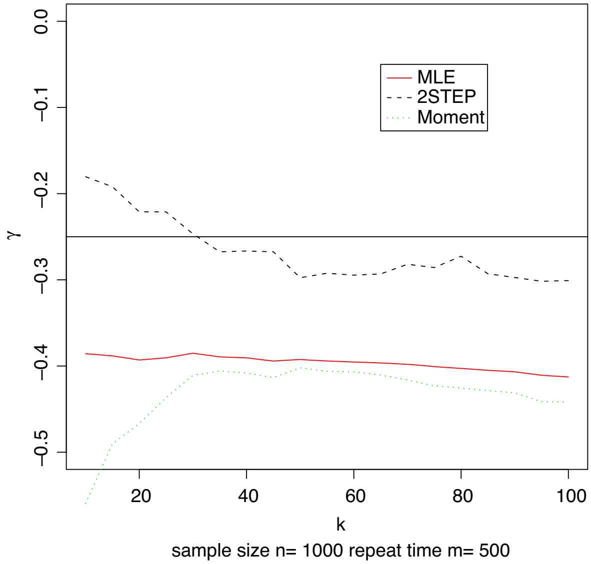
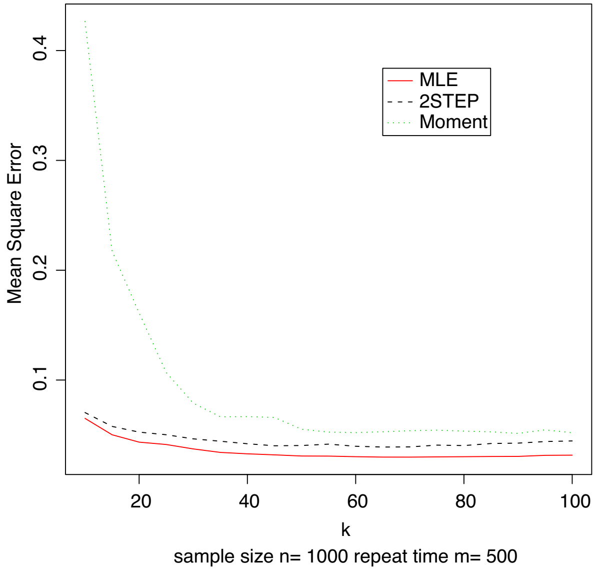


Fig. 12 MSE: R-Burr($\gamma = -1/4$, $\rho = -1/2$)



4 Conclusion

In the literature of estimating the extreme value index, a variety of estimators are proposed. A good estimator should have the following properties:

1) Performing a smaller estimation error;
2) Satisfying theoretical properties such as shift and scale invariant;
3) Easy to calculate.

Most of the explicit estimators do not satisfy the shift invariant property or perform a relatively worse estimation, while the maximum likelihood estimator is shift and scale invariant and provides a reasonably nice performance. However, it is not explicitly given.

In this paper, we propose an explicit two-step estimator which is close enough to the maximum likelihood estimator. Therefore it has the same asymptotic behavior. Furthermore, by a suitable choice of the first step estimator, it is shift and scale invariant. By iteration, we can get three-step or even more step estimators which performs better according to an extensive simulation study.

References

Balkema, A.A., de Haan, L.: Residual life time at great age. Ann. Probab. **2**, 792–804 (1974)
Beirlant, J., Vynckier, P., Teugels, J.L.: Tail index estimation, pareto quantile plots, and regression diagnostics. J. Am. Stat. Assoc. **91**, 1659–1667 (1996)
de Haan, L., Peng, L.: Comparison of tail index estimators. Stat. Neerl. **52**(1), 60–70 (1998)
de Haan, L., Stadtmüller, U.: Generalized regular variation of second order. J. Aust. Math. Soc. Ser. A **61**, 381–395 (1996)
Dekkers, A.L.M., Einmahl, J.H.J., de Haan, L.: A moment estimator for the index of an extreme-value distribution. Ann. Stat. **17**, 1833–1855 (1989)
Drees, H.: On smooth statistical tail functionals. Scand. J. Statist. **25**, 187–210 (1998)
Drees, H., Ferreira, A., de Haan, L.: On maximum likelihood estimation of the extreme value index. Ann. Appl. Probab. **14**, 1179–1201 (2004)
Gnedenko, B.: Sur la distribution limite du terme maximum d'une série aléatoire. Ann. Math. **44**, 423–453 (1943)
Grimshaw, S.D.: Computing maximum likelihood estimates for the generalized pareto distribution. Technometrics **35**, 185–191 (1993)
Hill, B.M.: A simple general approach to inference about the tail of a distribution. Ann. Stat. **3**, 1163–1174 (1975)
Hosking, J., Wallis, J.: Parameter and quantile estimation for the generalized pareto distribution. Technometrics **29**, 339–349 (1987)
Pickands III, J.: Statistical inference using extreme order statistics. Ann. Stat. **3**, 119–131 (1975)
Smith, R.L.: Estimating tails of probability distributions. Ann. Stat. **15**, 1174–1207 (1987)