Adapting the Hill estimator to distributed inference: dealing with the bias

Liujun Chen

Fudan University, China

Deyuan Li

Fudan University, China

Chen Zhou

Erasmus University Rotterdam, The Netherlands

Abstract

The distributed Hill estimator is a divide-and-conquer algorithm for estimating the extreme value index when data are stored in multiple machines. In applications, estimates based on the distributed Hill estimator can be sensitive to the choice of the number of the exceedance ratios used in each machine. Even with choosing the number at a low level, a high asymptotic bias may arise. We overcome this potential drawback by designing a bias correction procedure for the distributed Hill estimator, adhere to the setup of distributed inference. The asymptotically unbiased distributed estimator we obtained, on the one hand, is applicable to distributed stored data, on the other hand, inherits all known advantages of bias correction methods in extreme value statistics.

Keywords: Extreme value index, Distributed inference, Bias correction

1 Introduction

Consider a distribution function F which belongs to the maximum domain of attraction of an extreme value distribution with a positive *extreme value index* $\gamma > 0$, that is,

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^{\gamma},\ x > 0,$$

where $U(t) := F^{\leftarrow}(1 - 1/t)$ with $t > 1$, and $^{\leftarrow}$ denotes the left-continuous inverse function. Such a distribution is also called a heavy-tailed distribution, where the extreme value index governs the tail of the distribution. Estimating the extreme value index is a key step for making statistical inference on the tail behaviour of F. Various methods have been proposed to estimate the extreme value index, such as the Hill estimator (Hill, 1975) , the maximum likelihood estimator (Prescott and Walden, 1980; Drees et al., 2004; Zhou, 2009) and the moment estimator (Dekkers et al., 1989).

Conducting extreme value analysis often requires large datasets in order to select extreme observations in the tail. Such datasets may be stored in multiple machines and cannot be combined into one dataset due to data privacy issue. For example, datasets collected in industries such as banking and healthcare require high level consumer privacy and cannot be shared across different organizations. Another potential situation is that some massive datasets cannot be processed by a single computer due to internet traffic or memory constraints. *Distributed inference* refers to the statistical problem of analyzing data stored in multiple machines. It often requires a divide-and-conquer (DC) algorithm. In a DC algorithm, one calculates statistical estimators on each machine in parallel and then communicates them to a central machine. The final estimator is obtained on the central machine, often by a simple average; see, for example, Li et al. (2013) for kernel density estimation, Fan et al. (2019) for principal component analysis, Volgushev et al. (2019) for quantile regression.

In this paper, we aim at estimating the extreme value index in the distributed inference context. Assume that independent and identically distributed (i.i.d.) observations $X_1, \ldots, X_N$ drawn from F are stored in m machines with n observations on each machine, i.e. $N = mn$. In the context of distributed inference, we assume that only limited (finite) number of results can be transmited from each machine to the central machine. As a result, we cannot apply statistical procedures to the oracle sample, i.e., the hypothetically combined dataset $\{X_1, \ldots, X_N\}$.

Chen et al. (2021) proposes the distributed Hill estimator to estimate the extreme value index γ. On each machine, the Hill estimator is applied and then transmitted to the central machine. On the central machine, the average of the Hill estimates collected from the m machines are calculated. Let $M_j^{(1)} \geq \cdots \geq M_j^{(n)}$ denote the order statistics of the observations on machine j for $j = 1, \ldots, m$. Then the Hill estimator on machine j can be constructed by using the top k exceedance ratios $M_j^{(i)}/M_j^{(k+1)}, i = 1, \ldots, k$, as

$$\hat{\gamma}_{j,k} = \frac{1}{k} \sum_{i=1}^{k} \left(\log M_j^{(i)} - \log M_j^{(k+1)} \right), \quad j = 1, \ldots, m.$$

The distributed Hill estimator is defined as

$$\hat{\gamma}_{DH,k} := \frac{1}{m}\sum_{j=1}^{m}\hat{\gamma}_{j,k} = \frac{1}{m}\sum_{j=1}^{m}\frac{1}{k}\sum_{i=1}^{k}\left(\log M_j^{(i)} - \log M_j^{(k+1)}\right).$$

Chen et al. (2021) studies the asymptotic behaviour of the distributed Hill estimator and shows sufficient conditions under which the distributed Hill estimator possesses the oracle property: its speed of convergence and asymptotic distribution coincides with the oracle Hill estimator. Here, the oracle Hill estimator is the Hill estimator using the top km exceedance ratios of the oracle sample $\{X_1, \ldots, X_N\}$, i.e. $\hat{\gamma} = l^{-1}\sum_{i=1}^{l}\left(\log M^{(i)} - \log M^{(l+1)}\right)$, where $l = km$ and $M^{(1)} \geq \cdots \geq M^{(N)}$ are the order statistics of the oracle sample $\{X_1, \ldots, X_N\}$.

In applications with finite sample size, one important tuning parameter in the Hill estimator is the number of exceedance ratios l used in the estimation. For the distributed Hill estimator $\hat{\gamma}_{DH,k}$, this issue is regarding the choice of k on each machine. The choice leads to a bias-variance tradeoff: with a low level of k, the estimation variance is at a high level; by increasing the level of k, the estimation variance is reduced but the estimation bias may arise. Such a problem is more pronounced in the context of distributed inference: the distributed Hill estimator may fail to achieve the oracle property with a high level of k. In addition, recall that the effective number of exceedance ratios involved in $\hat{\gamma}_{DH,k}$ is km. As k increases by 1, the effective number of exceedance ratios will increase by m. Thus, the performance of $\hat{\gamma}_{DH,k}$ is very sensitive to the choice of k. If m is large, with even a low level of k, the asymptotic bias may be at a high level which may not be acceptable in applications.

In existing extreme value statistics literature, there are two types of solutions for selecting the number of exceedance ratios in the estimation. The first stream of literature aims at finding the optimal level that balances the asymptotic bias and variance, see e.g. Danielsson et al. (2001) and Guillou and Hall (2001). The second stream of literature corrects the bias and eventually allows for choosing a high level of the number of exceedance ratios, see e.g. Gomes et al. (2008) and de Haan et al. (2016). In applications, if the sample size is large, the bias correction method is preferred since they possess at least two advantages. First, bias correction methods allow for choosing a higher level of the number of exceedance ratios than that used for the original estimator, which results in also a lower level of varaince. Second, bias correction procedures lead to estimates that are less sensitive to the choice of the number of exceedance ratios.

In this paper, we shall adapt the distributed Hill estimator such that it is suitable for finite sample applications. More specifically, we introduce a bias correction procedure for estimating the extreme value index, without compromising the distributed inference setup. Notice that existing bias correction methods often rely on estimating a second order parameter and a second order scale function as given in (1) below. Such an estimation again requires the oracle sample which is infeasible in the context of distributed inference. Therefore, we resort to a different approach, sticking to the requirement that only limited (fixed) number of results can be transmitted from each machine to the central machine. In such a way, the resulting estimator is not only asymptotically unbiased, but also in the same spirit of a DC algorithm. We name it as "asymptotically unbiased distributed estimator" for the extreme value index. The asymptotically unbiased distributed estimator, on the one hand, is applicable to distributed stored data, on the other hand, inherits the advantages of bias correction methods in extreme value statistics.

The rest of the paper is organized as follows. Section 2 presents the idea for bias correction. Section 3 proposes a DC algorithm for estimating the second order parameter, defines the asymptotically unbiased distributed estimator for the extreme value index and shows the main theoretical results. Section 4 provides a simulation study to confirm that the asymptotically unbiased distributed estimator exhibits superior performance compared to the distributed Hill estimator. The proofs are given in the Appendix.

2 Bias Correction Methodology

To obtain the asymptotic normality of the distributed Hill estimator $\hat{\gamma}_{DH,k}$, Chen et al. (2021) assumes the following second order condition. Suppose that there exist an eventually positive or negative function A with $\lim_{t\to\infty} A(t) = 0$ and a real number $\rho \leq 0$ such that

$$\lim_{t\to\infty} \frac{\frac{U(tx)}{U(t)} - x^\gamma}{A(t)} = x^\gamma \frac{x^\rho - 1}{\rho},$$

for all $x > 0$, which is equivalent to

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

In addition, assume that as $N \to \infty$,

$$m = m(N) \to \infty, \quad n = n(N) \to \infty, \quad n/\log m \to \infty, \tag{2}$$

and k is either a fixed integer or an intermediate sequence, i.e. $k = k(N) \to \infty, k/n \to 0$. Under conditions (1) and (2), Chen et al. (2021) shows that the distributed Hill estimator possesses the following asymptotic expansion:

$$\hat{\gamma}_{DH,k} - \gamma = \frac{N_\gamma}{\sqrt{km}} + \frac{A(n/k)}{1-\rho} g(k,n,\rho) + \frac{1}{\sqrt{km}} o_P(1),$$

where N_γ is a normally distributed random variable with mean 0 and variance γ^2, and

$$g(k,n,\rho) := \left(\frac{k}{n}\right)^\rho \frac{\Gamma(n+1)\Gamma(k-\rho+1)}{\Gamma(n-\rho+1)\Gamma(k+1)}, \tag{3}$$

with Γ denoting the gamma function. By Lemma 2, we have that, if k is a fixed integer, then $g(k,n,\rho) \to k^\rho \Gamma(k-\rho+1)/\Gamma(k+1)$, as $N \to \infty$. If k is an intermediate sequence, then $g(k,n,\rho) \to 1$, as $N \to \infty$.

Since the bias term of the distributed Hill estimator is an explicit function $(1-\rho)^{-1}A(n/k)g(k,n,\rho)$, we shall estimate the bias, subtract it from the original distributed Hill estimator, which leads to the asymptotically unbiased distributed estimator.

The estimation of the bias term requires estimating the second order parameter ρ and the second order scale function A in condition (1). For simplicity, we follow the bias correction literature to assume that $\rho < 0$, see e.g. de Haan et al. (2016) and Gomes and Pestana (2007). In order to obtain the asymptotic behavior of the estimator for ρ, a third order condition is often assumed. We invoke the third order condition in Alves et al. (2003) as follows. Suppose that there exist an eventually positive or negative function B with $\lim_{t\to\infty} B(t) = 0$ and a real number $\tilde{\rho} \leq 0$ such that

$$\lim_{t\to\infty} \frac{1}{B(t)} \left\{ \frac{\log U(tx) - \log U(t) - \gamma \log x}{A(t)} - \frac{x^\rho - 1}{\rho} \right\} = \frac{1}{\tilde{\rho}} \left(\frac{x^{\rho+\tilde{\rho}} - 1}{\rho + \tilde{\rho}} - \frac{x^\rho - 1}{\rho} \right). \tag{4}$$

Lastly, following Cai et al. (2012) and de Haan et al. (2016), we use a higher intermediate sequence k_ρ for estimating the second order parameter ρ. Assume that as $N \to \infty$, $k_\rho \to \infty, k_\rho/n \to 0$, and

$$\sqrt{k_\rho m} A(n/k_\rho) \to \infty, \sqrt{k_\rho m} A^2(n/k_\rho) \to \lambda_1 \in \mathbb{R}, \sqrt{k_\rho m} A(n/k_\rho) B(n/k_\rho) \to \lambda_2 \in \mathbb{R}. \tag{5}$$

Similar to de Haan et al. (2016), in the eventual asymptotically unbiased distributed estimator for the extreme value index, one can choose a high number of exceedance ratios than that used in the distributed Hill estimator. In our context, we choose a sequence k_n such that, as $N \to \infty$, $k_n/k_\rho \to 0$ and

$$\sqrt{k_n m} A(n/k_n) \to \infty, \sqrt{k_n m} A^2(n/k_n) \to 0, \sqrt{k_n m} A(n/k_n) B(n/k_n) \to 0. \tag{6}$$

Here, similar to the distributed Hill estimator, k_n can be either fixed or an intermediate sequence.

3 Main results

We first introduce the estimator for the second order parameter ρ in the distributed inference setup and study its asymptotic behavior. Then we define the asymptotically unbiased distributed estimator and show its asymptotic behavior.

3.1 Estimating the second order parameter

If the oracle sample can be used, then there are several estimators for the second order parameter ρ, see e.g. Alves et al. (2003) and Gomes et al. (2002). However, since we cannot apply a statistical procedure to the oracle sample, we need to develop a DC algorithm for estimating ρ. Consider the following statistics computed based on observations on machine j,

$$R_{j,k}^{(\alpha)} := \frac{1}{k} \sum_{i=1}^{k} \left\{ \log M_j^{(i)} - \log M_j^{(k+1)} \right\}^{\alpha}, \quad \alpha = 1, 2, 3.$$

We request that each machine sends the values $R_{j,k}^{(\alpha)}, \alpha = 1, 2, 3$ to the central machine. On the central machine, we take the average of the $R_{j,k}^{(\alpha)}$ statistics to obtain

$$R_k^{(\alpha)} = \frac{1}{m} \sum_{j=1}^{m} R_{j,k}^{(\alpha)}, \quad \alpha = 1, 2, 3.$$

Motivated by Alves et al. (2003), we define the estimator for the second order parameter ρ as

$$\widehat{\rho}_{k,\tau} := -3 \left| \frac{T_{k,\tau} - 1}{T_{k,\tau} - 3} \right|, \tag{7}$$

where

$$T_{k,\tau} := \frac{\left(R_k^{(1)}\right)^{\tau} - \left(R_k^{(2)}/2\right)^{\tau/2}}{\left(R_k^{(2)}/2\right)^{\tau/2} - \left(R_k^{(3)}/6\right)^{\tau/3}},$$

and $\tau \geq 0$ is a tuning parameter. For $\tau = 0$, $T_{\tau,k}$ is defined by continuity. In practice, it is suggested to choose $\tau \in [0, 1]$, see e.g. Gomes and Pestana (2007) and Gomes et al. (2008).

Before studying the asymptotics of $\hat{\rho}_{k,\tau}$, we first establish that for $R_k^{(\alpha)}$ in the following proposition.

Proposition 1. Assume that the distribution function F satisfies the third order condition (4) with parameters $\gamma > 0, \rho < 0$ and $\tilde{\rho} \leq 0$, and condition (2) holds. In addition, suppose that an intermediate sequence k satisfies that as $N \to \infty$, $k \to \infty, k/n \to \infty$ and $\sqrt{km}A(n/k)B(n/k) = O(1), \sqrt{km}A^2(n/k) = O(1)$. Then for suitable versions of the functions A and B, denoted as A_0 and B_0, we have that as $N \to \infty$,

(i)

$$\sqrt{km}\left(R_k^{(1)} - \gamma\right) - \gamma P^{(1)} - \frac{g(k,n,\rho)}{1-\rho}\sqrt{km}A_0(n/k) - \frac{g(k,n,\rho+\tilde{\rho})}{1-\rho-\tilde{\rho}}\sqrt{km}A_0(n/k)B_0(n/k) = o_p(1),$$

(ii)

$$\begin{aligned}
&\sqrt{km}\left(R_k^{(2)} - 2\gamma^2\right) - \gamma^2 P^{(2)} - 2\gamma\sqrt{km}A_0(n/k)\frac{g(k,n,\rho)}{\rho}\left\{\frac{1}{(1-\rho)^2} - 1\right\} \\
&\quad - \sqrt{km}A_0^2(n/k)\frac{g(k,n,2\rho)}{\rho^2}\left(\frac{1}{1-2\rho} - \frac{2}{1-\rho} + 1\right) \\
&\quad - 2\gamma\sqrt{km}A_0(n/k)B_0(n/k)\frac{g(k,n,\rho+\tilde{\rho})}{\rho+\tilde{\rho}}\left\{\frac{1}{(1-\rho-\tilde{\rho})^2} - 1\right\} = o_p(1),
\end{aligned}$$

(iii)

$$\begin{aligned}
&\sqrt{km}\left(R_k^{(3)} - 6\gamma^3\right) - \gamma^3 P^{(3)} - 6\gamma^2\sqrt{km}A_0(n/k)\frac{g(k,n,\rho)}{\rho}\left\{\frac{1}{(1-\rho)^3} - 1\right\} \\
&\quad - 3\gamma\sqrt{km}A_0^2(n/k)\frac{g(k,n,2\rho)}{\rho^2}\left\{\frac{1}{(1-2\rho)^2} - \frac{2}{(1-\rho)^2} + 1\right\} \\
&\quad - 6\gamma^2\sqrt{km}A_0(n/k)B_0(n/k)\frac{g(k,n,\rho+\tilde{\rho})}{\rho+\tilde{\rho}}\left\{\frac{1}{(1-\rho-\tilde{\rho})^3} - 1\right\} = o_P(1),
\end{aligned}$$

where $(P^{(1)}, P^{(2)}, P^{(3)})^T \sim N(\mathbf{0}, \mathbf{\Sigma})$ with

$$\mathbf{\Sigma} = \begin{pmatrix} 1 & 4 & 18 \\ 4 & 20 & 98 \\ 18 & 98 & 684 \end{pmatrix}.$$

Applying Proposition 1 leads to the asymptotic behavior of $\hat{\rho}_{k,\tau}$ as follows.

THEOREM 1. Assume that the distribution function F satisfies the third order condition (4) with parameters $\gamma > 0, \rho < 0$ and $\tilde{\rho} \leq 0$, and condition (2) holds. Suppose that the intermediate sequence k_ρ satisfies condition (5). Then as $N \to \infty$,

$$\sqrt{k_\rho m} A_0(n/k_\rho)(\hat{\rho}_{k_\rho,\tau} - \rho) = O_P(1),$$

where $\hat{\rho}_{k_\rho,\tau}$ is defined in (7).

3.2 Asymptotically unbiased distributed estimator for the extreme value index

Motived by de Haan et al. (2016), we define the following estimator as the asymptotically unbiased distributed estimator for the extreme value index:

$$\tilde{\gamma}_{k_n,k_\rho,\tau} := R_{k_n}^{(1)} - \frac{R_{k_n}^{(2)} - 2\left(R_{k_n}^{(1)}\right)^2}{2R_{k_n}^{(1)}\hat{\rho}_{k_\rho,\tau}(1-\hat{\rho}_{k_\rho,\tau})^{-1}}, \tag{8}$$

where $\tau \geq 0$ is a tuning parameter. Notice that the estimator $\tilde{\gamma}_{k_n,k_\rho,\tau}$ in (8) adheres to a DC algorithm since each machine only sends five values $\left\{R_{j,k_n}^{(1)}, R_{j,k_n}^{(2)}, R_{j,k_\rho}^{(1)}, R_{j,k_\rho}^{(2)}, R_{j,k_\rho}^{(3)}\right\}$ to the central machine.

REMARK 1. The statistic $R_{k_n}^{(1)}$ is the original distributed Hill estimator $\hat{\gamma}_{DH,k_n}$.

The following theorem shows the asymptotic normality of the asymptotically unbiased distributed estimator.

THEOREM 2. Assume that the distribution function F satisfies the third order condition (4) with parameters $\gamma > 0, \rho < 0$ and $\tilde{\rho} \leq 0$, and condition (2) holds. Suppose that k_ρ, k_n satisfy conditions

(5) and (6) respectively. Then as $N \to \infty$,

$$\sqrt{k_n m}\left(\tilde{\gamma}_{k_n,k_\rho,\tau}-\gamma\right) \xrightarrow{d} N\left[0,\gamma^2\left\{1+\left(\rho^{-1}-1\right)^2\right\}\right].$$

Remark 2. The limit distribution in Theorem 2 is the same as that of the bias corrected Hill estimator based on the oracle sample, see for example de Haan et al. (2016). In other words, the asymptotically unbiased distributed estimator achieves the oracle property.

4 Simulation Study

4.1 Comparison with the original distributed Hill estimator

In this subsection, we conduct a simulation study to demonstrate the finite sample performance of the asymptotically unbiased distributed estimator for the extreme value index. Data are simulated from three distributions: the unit Fréchet distribution, $F(x) = \exp\left(-x^{-1}\right), x > 0$; the Burr distribution, $F(x) = 1-(1+x^{1/2})^{-2}, x > 0$; and the absolute Cauchy distribution with the density function $f(x) = 2/\left\{\pi(1+x^2)\right\}, x > 0$. The first, second and third order indices of the three distributions are listed in Table 1. We generate $r = 1000$ samples with sample size $N = 10000$. The value of k_ρ is chosen to be $[n^{0.98}]$ as suggested by Cai et al. (2012), where $[x]$ denotes the largest integer less than or equal to x.

	Unit Fréchet	Burr distribution	Absolute Cauchy
γ	1	1	1
ρ	-1	$-1/2$	-2
$\tilde{\rho}$	-1	$-1/2$	-4

Table 1: The first, second and third order indicies for the distributions.

To apply the asymptotically unbiased distributed estimator, we use the following procedure:

1. On each machine j, we calculate $R^{(1)}_{j,k_n}$, $R^{(2)}_{j,k_n}$, $R^{(1)}_{j,k_\rho}$, $R^{(2)}_{j,k_\rho}$, $R^{(3)}_{j,k_\rho}$ and transmit them to the central machine.

2. On the central machine, we take the average of the $R^{(1)}_{j,k_n}$, $R^{(2)}_{j,k_n}$, $R^{(1)}_{j,k_\rho}$, $R^{(2)}_{j,k_\rho}$, $R^{(3)}_{j,k_\rho}$ statistics collected from the m machines to obtain $R^{(1)}_{k_n}$, $R^{(2)}_{k_n}$, $R^{(1)}_{k_\rho}$, $R^{(2)}_{k_\rho}$, $R^{(3)}_{k_\rho}$.

3. On the central machine, we estimate the second order parameter ρ by (7) with $k = k_\rho$. The value of the tuning parameter τ is set at $0, 0.5$ and 1.

4. On the central machine, we estimate the extreme value index by (8) for various values of k_n, using $\hat{\rho}_{k_\rho,\tau}$.

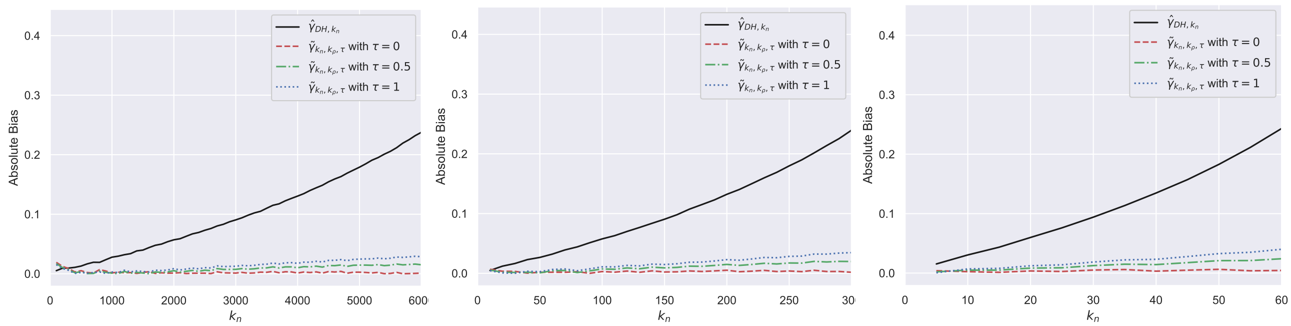
We assume that the $N = 10000$ observations are stored in $m = 1, 20, 100$ machines with $n = N/m$ observations each. Note that the case $m = 1$ corresponds to applying the statistical procedure to the oracle sample directly. The corresponding estimator is therefore the oracle estimator. Figures 1-3 show the absolute bias against various levels of k_n for the three distributions, respectively.

We have three main observations from these figures. First, the asymptotically unbiased distributed estimator $\tilde{\gamma}_{k_n,k_\rho,\tau}$ has superior performance compared to the original distributed Hill estimator $\hat{\gamma}_{DH,k_n}$. As k_n increases, the bias of the distributed Hill estimator increases, while the asymptotically unbiased distributed estimator has almost zero bias except for very high level of k_n. This is in line with the asymptotic theory.

Second, the performance of the asymptotically unbiased distributed estimator is not sensitive to the variation in m. The performance across different levels of m is comparable to the case $m = 1$, i.e., the oracle property holds. When m is at a high level, the distributed Hill estimator exhibits substantial bias even when k_n is low. By contrast, the asymptotically unbiased distributed estimator corrects such a bias.

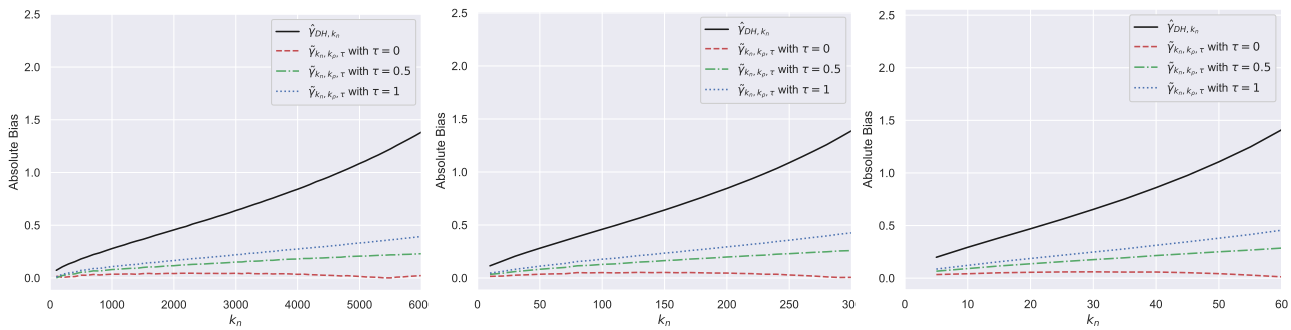
Third, the choice of τ affects the performance of the asymptotically unbiased distributed estimator. When $\rho < -1$ (Cauchy distribution), $\tau = 1$ is a better choice than $\tau = 0$. When $\rho \geq -1$ (Fréchet distribution and Burr distribution), $\tau = 0$ is a better choice than $\tau = 1$. This is in line with the findings in Alves et al. (2003).

We also plot the mean squared error (MSE) of these estimators against various levels of k_n for $m = 20$ in Figure 4. For the unit Fréchet distribution and the absolute Cauchy distribution, when k_n is low, the distributed Hill estimator has a better performance compared to the asymptotically unbiased distributed estimator. In this case, the biases for both estimators are low, which implies that the MSE is driven by the estimation variance. As k_n increases, the estimation variance of both estimators decreases towards zero. Then, the MSE is mainly driven by the bias. So, the asymptotically unbiased distributed estimator exhibits a much smaller MSE compared to the distributed Hill estimator. For the Burr distribution, the bias is high even when k_n is low since $\rho = -1/2$. Consequently, the asymptotically unbiased distributed estimator has a smaller MSE for almost all levels of k_n.



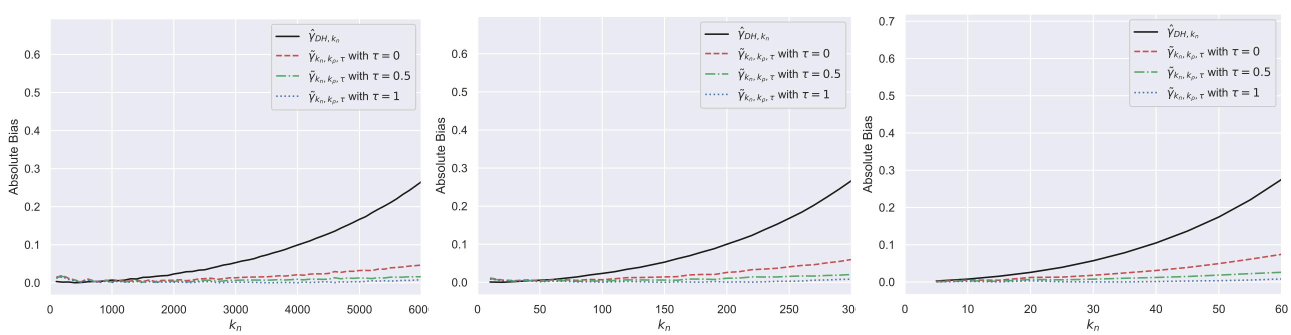
(a) m=1 (b) m=20 (c) m=100

Figure 1: Absolute bias for the unit Fréchet distribution.



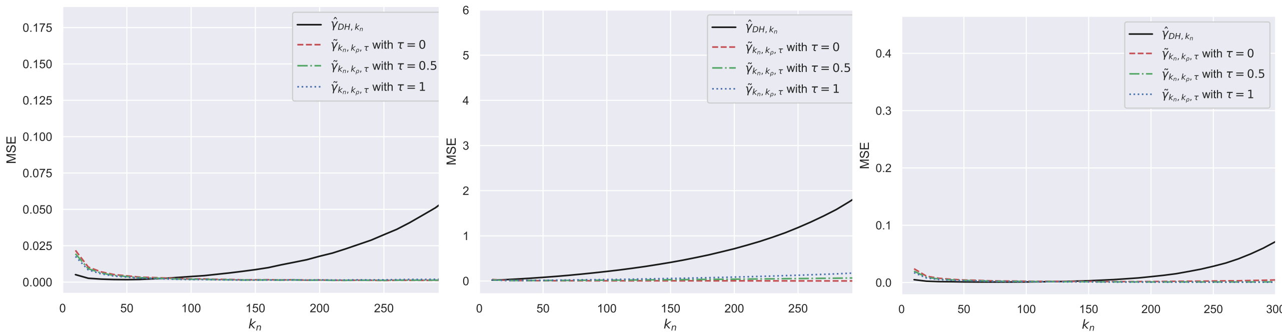
(a) m=1 (b) m=20 (c) m=100

Figure 2: Absolute bias for the Burr distribution.



(a) m=1 (b) m=20 (c) m=100

Figure 3: Absolute bias for the absolute Cauchy distribution.



(a) Unit Fréchet (b) Burr (c) Absolute Cauchy

Figure 4: MSE for the unit Fréchet, Burr and absolute Cauchy distribution for $m = 20$.

4.2 Further limitation for transmission

Recall that for the asymptotically unbiased distributed estimator, we need to transmit five statistics from each of the m machines to the central machine. If there are further limitations on the number of results that can be transmitted, such as only three, or even one statistic can be transmitted, the estimation procedure in Section 4.1 will not be applicable. In this subsection, we consider two alternative procedures for bias correction in the distributed inference setup with fewer number of transmissions.

Firstly, we consider a bias correction procedure if only three statistics can be transmitted. We can estimate the second order parameter ρ on each machine and transmit the estimates for ρ to the central machine. The detailed procedures are given as follows:

- On each machine j, we calculate $R^{(1)}_{j,k_n}$, $R^{(2)}_{j,k_n}$, $R^{(1)}_{j,k_\rho}$, $R^{(2)}_{j,k_\rho}$, $R^{(3)}_{j,k_\rho}$.

- On each machine j, we estimate the second order parameter ρ by

$$\hat{\rho}_{j,k_\rho,\tau} := -3\left|\frac{T_{j,k_\rho,\tau}-1}{T_{j,k_\rho,\tau}-3}\right|, \tag{9}$$

with

$$T_{j,k_\rho,\tau} := \frac{\left(R^{(1)}_{j,k_\rho}\right)^{\tau}-\left(R^{(2)}_{j,k_\rho}/2\right)^{\tau/2}}{\left(R^{(2)}_{j,k_\rho}/2\right)^{\tau/2}-\left(R^{(3)}_{j,k_\rho}/6\right)^{\tau/3}},$$

and transmit $\hat{\rho}_{j,k_\rho,\tau}, R^{(1)}_{j,k_n}, R^{(2)}_{j,k_n}$ to the central machine.

- On the central machine, we take the average of the $\hat{\rho}_{j,k_\rho,\tau}, R^{(1)}_{j,k_n}, R^{(2)}_{j,k_n}$ statistics to obtain

$$\tilde{\rho}_{k_\rho,\tau} = \frac{1}{m}\sum_{j=1}^{m} \hat{\rho}_{j,k_\rho,\tau},\ \ R^{(1)}_{k_n} = \frac{1}{m}\sum_{j=1}^{m} R^{(1)}_{j,k_n},\ \ R^{(2)}_{k_n} = \frac{1}{m}\sum_{j=1}^{m} R^{(2)}_{j,k_n}.$$

- On the central machine, we estimate the extreme value index by

$$\tilde{\gamma}^{(2)}_{k_n,k_\rho,\tau} := R^{(1)}_{k_n} - \frac{R^{(2)}_{k_n} - 2\left(R^{(1)}_{k_n}\right)^2}{2R^{(1)}_{k_n}\tilde{\rho}_{k_\rho,\tau}(1-\tilde{\rho}_{k_\rho,\tau})^{-1}}.$$

Secondly, we consider a bias correction procedure if only one statistic can be transmitted. We can conduct bias correction on each machine and transmit the estimates using the bias-corrected Hill estimator to the central machine. Then we take the average of these estimates on the central machine. In this procedure, each machine only sends one statistic to the central machine. The detailed procedures are as follows:

- On each machine j, we calculate $R^{(1)}_{j,k_n}$, $R^{(2)}_{j,k_n}$, $R^{(1)}_{j,k_\rho}$, $R^{(2)}_{j,k_\rho}$, $R^{(3)}_{j,k_\rho}$ and estimate the second order parameter ρ by (9).

- On each machine j, we estimate the extreme value index by

$$\tilde{\gamma}_{j,k_n,k_\rho,\tau} := R^{(1)}_{j,k_n} - \frac{R^{(2)}_{j,k_n} - 2\left(R^{(1)}_{j,k_n}\right)^2}{2R^{(1)}_{j,k_n}\hat{\rho}_{j,k_\rho,\tau}(1-\hat{\rho}_{j,k_\rho,\tau})^{-1}},$$

  and transmit the estimates $\tilde{\gamma}_{j,k_n,k_\rho,\tau}$ to the central machine.

- On the central machine, we take the average of these estimates by

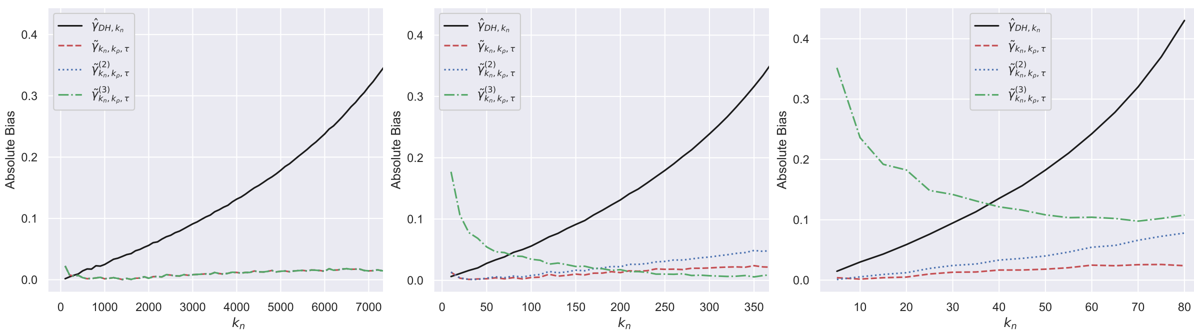
$$\tilde{\gamma}^{(3)}_{k_n,k_\rho,\tau} := \frac{1}{m}\sum_{j=1}^{m} \tilde{\gamma}_{j,k_n,k_\rho,\tau}.$$

The asymptotic theories of these two estimators $\tilde{\gamma}^{(2)}_{k_n,k_\rho,\tau}$ and $\tilde{\gamma}^{(3)}_{k_n,k_\rho,\tau}$ are left for further study. We only provide a finite sample comparison between the proposed estimator and these two estimators.

In this comparison, we fix $\tau = 0.5$. Figure 5 shows the absolute bias for the unit Fréchet distribution. The figures for the Burr distribution and the absolute Cauchy distribution have similar patterns and are therefore omitted. We observe that all three estimators $\tilde{\gamma}_{k_n,k_\rho,\tau}$, $\tilde{\gamma}^{(2)}_{k_n,k_\rho,\tau}$

and $\tilde{\gamma}^{(3)}_{k_n,k_\rho,\tau}$ perform better than the original distributed Hill estimator. In addition, $\tilde{\gamma}_{k_n,k_\rho,\tau}$ and $\tilde{\gamma}^{(2)}_{k_n,k_\rho,\tau}$ have similar performance for all three levels of m with $\tilde{\gamma}_{k_n,k_\rho,\tau}$ performing slightly better for the unit Fréchet distribution and $\tilde{\gamma}^{(2)}_{k_n,k_\rho,\tau}$ performing slightly better for the absolute Cauchy distribution.

The performance of $\tilde{\gamma}^{(3)}_{k_n,k_\rho,\tau}$ is unstable when m is at a high level. In this case, n is at a low level. Therefore, conducting bias correction on each machine is suboptimal since the bias correction procedure requires a relatively large sample size.



(a) m=1 (b) m=20 (c) m=100

Figure 5: Absolute bias for the unit Fréchet distribution.

A Proofs

A.1 Preliminary

Lemma 1. Let $Y, Y_1, \ldots, Y_n$ be i.i.d. Pareto (1) random variables with distribution function $1 - 1/y,\ y \geq 1$. Let $Y^{(1)} \geq \cdots \geq Y^{(n)}$ be the order statistics of $\{Y_1, \ldots, Y_n\}$. Let f be a function such that $\mathrm{Var}\{f(Y)\} < \infty$. Then for any $k \geq 1$,

$$\frac{1}{k}\sum_{i=1}^{k} f\left(\frac{Y^{(i)}}{Y^{(k+1)}}\right) \stackrel{d}{=} \frac{1}{k}\sum_{i=1}^{k} f(Y_i^*),$$

where $Y_1^*, Y_2^*, \ldots, Y_k^*$ are i.i.d. Pareto (1) random variables. Moreover,

$$\sqrt{k}\left\{\frac{1}{k}\sum_{i=1}^{k} f\left(\frac{Y^{(i)}}{Y^{(k+1)}}\right) - \mathbb{E}f(Y)\right\}$$

is independent of $Y^{(k+1)}$ and asymptotically normally distributed with mean zero and variance $\operatorname{Var}\{f(Y)\}$ as $n \to \infty$, provided that $k = k(n) \to \infty$ and $k/n \to 0$.

Proof of Lemma 1. This Lemma follows directly from Lemma 3.2.3 in de Haan and Ferreira (2006) with the fact that $\log Y$ follows a standard exponential distribution. $\square$

LEMMA 2. *Let $Y_1, \ldots, Y_n$ be i.i.d. Pareto (1) random variables and $Y^{(1)} \geq \cdots \geq Y^{(n)}$ be the order statistics of $\{Y_1, \ldots, Y_n\}$. Then for any $\rho < 0$,*

$$\mathbb{E}\left\{\left(\frac{k}{n}Y^{(k+1)}\right)^{\rho}\right\} = g(k, n, \rho),$$

where $g(k, n, \rho)$ is defined in (3). Moreover, if k is a fixed integer, then $g(k, n, \rho) \to k^{\rho}\Gamma(k - \rho + 1)/\Gamma(k+1)$ as $n \to \infty$. If k is an intermediate sequence, i.e. $k \to \infty, k/n \to 0$ as $n \to \infty$, then,

$$g(k, n, \rho) = 1 + \frac{1}{2}(\rho^2 - \rho)k^{-1} - \frac{1}{2}(\rho^2 - \rho)(n - \rho)^{-1} + O(k^{-2}).$$

Proof of Lemma 2.

$$\begin{aligned}
\mathbb{E}\left\{\left(\frac{k}{n}Y^{(k+1)}\right)^{\rho}\right\} &= \frac{n!}{(n-k-1)!k!}\int_1^{\infty}\left(1 - \frac{1}{y}\right)^{n-k-1}\left(\frac{1}{y}\right)^{k+2}\left(\frac{k}{n}y\right)^{\rho} dy \\
&= \left(\frac{k}{n}\right)^{\rho}\frac{n!}{(n-k-1)!k!}\int_1^{\infty}\left(1 - \frac{1}{y}\right)^{n-k-1}\left(\frac{1}{y}\right)^{k+2-\rho} dy \\
&= \left(\frac{k}{n}\right)^{\rho}\frac{\Gamma(n+1)\Gamma(k-\rho+1)}{\Gamma(n-\rho+1)\Gamma(k+1)} \\
&= g(k, n, \rho).
\end{aligned}$$

We first handle the case when k is a fixed integer. By the Stirling's formula,

$$\Gamma(x) = \sqrt{2\pi(x-1)}\left\{e^{-1}(x-1)\right\}^{x-1}\left\{1 + (x-1)^{-1}/12 + O(1/x^2)\right\}$$

as $x \to \infty$, we have that, as $n \to \infty$,

$$\Gamma(n+1) \sim (2\pi n)^{1/2}\left(\frac{n}{e}\right)^{n}, \quad \Gamma(n - \rho + 1) \sim \{2\pi(n-\rho)\}^{1/2}\left(\frac{n-\rho}{\rho}\right)^{n-\rho},$$

which leads to

$$g(k, n, \rho) \to k^{\rho}\frac{\Gamma(k - \rho + 1)}{\Gamma(k+1)}.$$

Next, we handle the case when k is an intermediate sequence. By the Stirling's formula, we have that, as $n \to \infty$,

$$g(k,n,\rho) = \left(1-\frac{\rho}{k}\right)^{k-\rho+1/2}\left(1+\frac{\rho}{n-\rho}\right)^{n-\rho+1/2}\frac{1+n^{-1}/12+O(n^{-2})}{1+(n-\rho)^{-1}/12+O(n^{-2})}\frac{1+(k-\rho)^{-1}/12+O(k^{-2})}{1+k^{-1}/12+O(k^{-2})}$$
$$= \left(1-\frac{\rho}{k}\right)^{k-\rho+1/2}\left(1+\frac{\rho}{n-\rho}\right)^{n-\rho+1/2}\left\{1+O(n^{-2})\right\}\left\{1+O(k^{-2})\right\}.$$

By the Taylor's formula and some direct calculation, we obtain that, as $n \to \infty$,

$$\left(1-\frac{\rho}{k}\right)^{k-\rho+1/2} = e^{-\rho}\left\{1+\frac{1}{2}(\rho^2-\rho)k^{-1}+O(k^{-2})\right\},$$

and

$$\left(1+\frac{\rho}{n-\rho}\right)^{n-\rho+1/2} = e^{\rho}\left\{1-\frac{1}{2}(\rho^2-\rho)(n-\rho)^{-1}+O(n^{-2})\right\}.$$

It follows that, as $n \to \infty$,

$$g(k,n,\rho) = 1+\frac{1}{2}(\rho^2-\rho)k^{-1}-\frac{1}{2}(\rho^2-\rho)(n-\rho)^{-1}+O(k^{-2}).$$

$\square$

Lemma 3. Let $Y_1,\ldots,Y_n$ be i.i.d. Pareto (1) random variables and $Y^{(1)} \geq \cdots \geq Y^{(n)}$ be the order statistics of $\{Y_1,\ldots,Y_n\}$. Define for $\rho < 0$,

$$Z_k = \frac{1}{k}\sum_{i=1}^{k}\frac{\left(Y^{(i)}/Y^{(k+1)}\right)^{\rho}-1}{\rho}.$$

Then, the following results hold.

(i) For fixed k, $\mathbb{E}(Z_k^a) < \infty$, for $a = 1,2,3,4$. Moreover, $\mathbb{E}\left(Z_k^2\right) - \left\{\mathbb{E}\left(Z_k\right)\right\}^2 > 0$.

(ii) For intermediate k, i.e., $k = k(n) \to \infty, k/n \to 0$ as $n \to \infty$, then, for $a = 1,2,3,4$,

$$\mathbb{E}\left(Z_k^a\right) = \frac{1}{(1-\rho)^a}\left\{1+\frac{a(a-1)}{2(1-2\rho)}\frac{1}{k}+O(k^{-2})\right\}.$$

Proof of Lemma 3. By Lemma 1, we have that,

$$Z_k \stackrel{d}{=} \frac{1}{k}\sum_{i=1}^{k}\frac{(Y_i^*)^\rho - 1}{\rho},$$

where $Y_1^*, \ldots, Y_k^*$ are i.i.d. Pareto (1) random variables. Denote $T_i = \{(Y_i^*)^\rho - 1\}/\rho$, for $i = 1, \ldots, k$ and $Z_k = k^{-1}\sum_{i=1}^k T_i$. Then, $T_i, i = 1, \ldots, k$ follows the generalized Pareto distribution with the cumulative distribution function $F(t) = 1 - (1+\rho t)^{-1/\rho}$. Thus, we have that for $a = 1, 2, 3, 4$,

$$\mathbb{E}(T_i^a) = \frac{a!}{(1-a\rho)\cdots(1-\rho)}.$$

First, we handle the case when k is fixed. The result is obvious since kZ_k is a finite sum of i.i.d. generalized Pareto random variables with shape parameter $\rho < 0$.

Next, we handle the case when k is an intermediate sequence. For $a = 1$, we have that, $E(Z_k) = E(T_i) = (1-\rho)^{-1}$.

For $a = 2$, we have that,

$$\begin{aligned}
\mathbb{E}\left(Z_k^2\right) &= \frac{1}{k^2}\left\{\sum_{i=1}^{k} E\left(T_i^2\right) + \sum_{i\neq j}\mathbb{E}\left(T_i\right)\mathbb{E}\left(T_j\right)\right\} \\
&= \frac{1}{k^2}\left[kE\left(T_i^2\right) + k(k-1)\left\{\mathbb{E}\left(T_i\right)\right\}^2\right] \\
&= \frac{1}{(1-\rho)^2} + \frac{1}{k}\frac{1}{(1-2\rho)(1-\rho)^2}.
\end{aligned}$$

For $a = 3$, we have that

$$\begin{aligned}
\mathbb{E}\left(Z_k^3\right) &= \frac{1}{k^2}\left\{\sum_{i=1}^{k}\mathbb{E}\left(T_i^3\right) + \sum_{i=j\neq l}\mathbb{E}\left(T_iT_j\right)\mathbb{E}\left(T_l\right) + \sum_{i\neq j\neq l}\mathbb{E}\left(T_i\right)\mathbb{E}\left(T_j\right)\mathbb{E}\left(T_l\right)\right\} \\
&= \frac{1}{k^3}\left[k\mathbb{E}\left(T_i^3\right) + 3k(k-1)\mathbb{E}\left(T_i^2\right)\mathbb{E}\left(T_i\right) + k(k-1)(k-2)\left\{\mathbb{E}(T_i)\right\}^3\right] \\
&= \frac{1}{(1-\rho)^3} + \frac{1}{k}\frac{3}{(1-2\rho)(1-\rho)^3} + O(k^{-2}).
\end{aligned}$$

The term $\mathbb{E}\left(Z_k^4\right)$ can be handled in a similar way as that for handling $\mathbb{E}\left(Z_k^3\right)$. □

LEMMA 4. Assume that the distribution function F satisfies the third order condition (4). Then there exist two functions $A_0(t) \sim A(t)$ and $B_0(t) = O\{B(t)\}$ as $t \to \infty$, such that for any $\delta > 0$, there exists a $t_0 = t_0(\delta) > 0$, for all $t \geq t_0$ and $tx \geq t_0$,

$$\left| \frac{\frac{\log U(tx) - \log U(t) - \gamma \log x}{A_0(t)} - \frac{x^\rho - 1}{\rho}}{B_0(t)} - \frac{x^{\rho+\tilde{\rho}} - 1}{\rho + \tilde{\rho}} \right| \leq \delta x^{\rho+\tilde{\rho}} \max(x^\delta, x^{-\delta}).$$

Proof of Lemma 4. This lemma follows from applying Theorem B.3.10 in de Haan and Ferreira (2006) to the function $f(t) := \log U(t) - \gamma \log t$. $\square$

A.2 Proofs for Section 3

Recall that $U = \{1/(1-F)\}^{\leftarrow}$. Then $X \stackrel{d}{=} U(Y)$, where Y follows the Pareto (1) distribution. Since we have i.i.d. observations $\{X_1, \ldots, X_N\}$, we can write $X_i \stackrel{d}{=} U(Y_i)$, where $\{Y_1, \ldots, Y_N\}$ is a random sample of Y. Recall that the N observations are stored in m machines with n observations each. For machine j, let $Y_j^{(1)} \geq \cdots \geq Y_j^{(n)}$ denote the order statistics of the n Pareto (1) distributed variables corresponding to the n observations in this machine. Then $M_j^{(i)} \stackrel{d}{=} U(Y_j^{(i)}), i = 1, \ldots, n, j = 1, \ldots, m$.

Proof of Proposition 1. We intend to replace t and tx in Lemma 4 by n/k and $Y_j^{(i)}, i = 1, \ldots, k+1, j = 1, \ldots, m$, respectively. For this purpose, we introduce the set

$$\mathcal{F}_{t_0} := \left\{ Y_j^{(k+1)} \geq t_0, \ \textit{for all}\ 1 \leq j \leq m \right\}.$$

By Lemma S.2 in the supplementary material of Chen et al. (2021), we have that for any $t_0 > 1$, if condition (2) holds, then $\lim_{N\to\infty} \mathbb{P}(\mathcal{F}_{t_0}) = 1$. Then, we can apply the intended replacement to get that, as $N \to \infty$,

$$\begin{aligned} \log U(Y_j^{(i)}) - \log U(n/k) = & -\gamma \log\left(kY_j^{(i)}/n\right) - A_0(n/k) \left\{ \left(kY_j^{(i)}/n\right)^\rho - 1 \right\} / \rho \\ & + A_0(n/k) B_0(n/k) \left\{ \left(kY_j^{(i)}/n\right)^{\rho+\tilde{\rho}} - 1 \right\} / (\rho + \tilde{\rho}) \\ & + o_P(1) A_0(n/k) B_0(n/k) \left(kY_j^{(i)}/n\right)^{\rho+\tilde{\rho}\pm\delta}, \end{aligned} \tag{10}$$

where the $o_P(1)$ term is uniform for all $1 \leq i \leq k+1$ and $1 \leq j \leq m$. By applying (10) twice for a general i and $i = k+1$ and the inequality $x^{\rho\pm\delta}/y^{\rho\pm\delta} \leq (x/y)^{\rho\pm\delta}$ for any $x, y > 0$, we get that as

$N \to \infty$,

$$\begin{aligned}
&\log U\left(Y_j^{(i)}\right) - \log U\left(Y_j^{(k+1)}\right) \\
&= \gamma\left(\log Y_j^{(i)} - \log Y_j^{(k+1)}\right) \\
&\quad + A_0(n/k)\left(kY_j^{(k+1)}/n\right)^{\rho}\left\{\left(Y_j^{(i)}/Y_j^{(k+1)}\right)^{\rho} - 1\right\}/\rho \\
&\quad + A_0(n/k)B_0(n/k)\left(kY_j^{(k+1)}/n\right)^{\rho+\tilde{\rho}}\left\{\left(Y_j^{(i)}/Y_j^{(k+1)}\right)^{\rho+\tilde{\rho}} - 1\right\}/\left(\rho+\tilde{\rho}\right) \\
&\quad + o_P(1)A_0(n/k)B_0(n/k)\left(kY_j^{(k+1)}/n\right)^{\rho+\tilde{\rho}\pm\delta}\left\{\left(Y_j^{(i)}/Y_j^{(k+1)}\right)^{\rho+\tilde{\rho}\pm\delta} + 1\right\}.
\end{aligned} \tag{11}$$

By taking the average across i and j, we obtain that

$$\begin{aligned}
&\sqrt{km}\left(R_k^{(1)} - \gamma\right) \\
&= \gamma\sqrt{km}\frac{1}{m}\frac{1}{k}\sum_{j=1}^{m}\sum_{i=1}^{k}\left\{\log\left(Y_j^{(i)}/Y_j^{(k+1)}\right) - \gamma\right\} \\
&\quad + \sqrt{km}A_0(n/k)\frac{1}{m}\sum_{j=1}^{m}\left(kY_j^{(k+1)}/n\right)^{\rho}\rho^{-1}\frac{1}{k}\sum_{i=1}^{k}\left\{\left(Y_j^{(i)}/Y_j^{(k+1)}\right)^{\rho} - 1\right\} \\
&\quad + \sqrt{km}A_0(n/k)B_0(n/k)\frac{1}{m}\sum_{j=1}^{m}\left(kY_j^{(k+1)}/n\right)^{\rho+\tilde{\rho}}\left(\rho+\tilde{\rho}\right)^{-1}\frac{1}{k}\sum_{i=1}^{k}\left\{\left(Y_j^{(i)}/Y_j^{(k+1)}\right)^{\rho+\tilde{\rho}} - 1\right\} \\
&\quad + o_P(1)\sqrt{km}A_0(n/k)B_0(n/k)\frac{1}{m}\sum_{j=1}^{m}\left(kY_j^{(k+1)}/n\right)^{\rho+\tilde{\rho}\pm\delta}\frac{1}{k}\sum_{i=1}^{k}\left\{\left(Y_j^{(i)}/Y_j^{(k+1)}\right)^{\rho+\tilde{\rho}\pm\delta} + 1\right\} \\
&=: I_1 + I_2 + I_3 + I_4.
\end{aligned}$$

Firstly, we handle I_1. By Lemma 1, we have that,

$$I_1 \overset{d}{=} \gamma\sqrt{km}\left(\frac{1}{km}\sum_{j=1}^{m}\sum_{i=1}^{k}\log Y_i^{j,*} - 1\right),$$

where $Y_i^{j,*}, i = 1, \ldots, k, j = 1, \ldots, m$ are independent and identically distributed Pareto (1) random variables. The central limit theorem yields that as $N \to \infty$, $I_1 \xrightarrow{d} \gamma P^{(1)}$, where $P^{(1)} \sim N(0, 1)$.

For I_2, write $\delta_{j,n} = \left(kY_j^{(k+1)}/n\right)^{\rho}(k\rho)^{-1}\sum_{i=1}^{k}\left\{\left(Y_j^{(i)}/Y_j^{(k+1)}\right)^{\rho} - 1\right\}$. Then we have that $I_2 = \sqrt{km}A_0(n/k)m^{-1}\sum_{j=1}^{m}\delta_{j,n}$, where $\delta_{j,n}, j = 1, \ldots, m$ are i.i.d. random variables.

We are going to show that, as $N \to \infty$,

$$\sqrt{km}\left\{\frac{1}{m}\sum_{j=1}^{m}\delta_{j,n} - \mathbb{E}\left(\delta_{j,n}\right)\right\} = O_P(1). \tag{12}$$

If k is fixed, (12) follows directly from Lemma 3 (i) and the Lyapunov central limit theorem for triangular array.

Next, we handle the case when k is an intermediate sequence. In this case, in order to apply the Lyapunov central limit theorem with 4-th moment, we need to calculate $\operatorname{Var}\left(\delta_{j,n}\right)$ and $\mathbb{E}\left[\left\{\delta_{j,n} - \mathbb{E}\left(\delta_{j,n}\right)\right\}^4\right]$. Denote $m_n^{(a)} := \mathbb{E}\left\{\left(\delta_{j,n}\right)^a\right\},\ a = 1,2,3,4$. By Lemma 1, we have that,

$$m_n^{(a)} = g(k,n,a\rho)\mathbb{E}\left[\left\{\frac{1}{k}\sum_{i=1}^{k}\frac{\left(Y_j^{(i)}/Y_j^{(k+1)}\right)^\rho - 1}{\rho}\right\}^a\right].$$

First, we calculate $\operatorname{Var}\left(\delta_{j,n}\right)$. By Lemma 3, we have that,

$$\begin{aligned}
\operatorname{Var}(\delta_{j,n}) &= m_n^{(2)} - \left(m_n^{(1)}\right)^2 \\
&= g(k,n,2\rho)\left\{\frac{1}{(1-\rho)^2} + \frac{1}{k}\frac{1}{(1-2\rho)(1-\rho)^2} + O(k^{-2})\right\} - \{g(k,n,\rho)\}^2\left\{\frac{1}{(1-\rho)^2} + O(k^{-2})\right\} \\
&= \frac{1}{k}g(k,n,2\rho)\frac{1}{(1-2\rho)(1-\rho)^2} + \left[g(k,n,2\rho) - \{g(k,n,\rho)\}^2\right]\frac{1}{(1-\rho)^2} + O(k^{-2}),
\end{aligned}$$

here in the last step, we used the fact that as $n \to \infty$, $g(k,n,\rho) \to 1$ and $g(k,n,2\rho) \to 1$. By Lemma 2, we have that, as $n \to \infty$,

$$g(k,n,2\rho) - \{g(k,n,\rho)\}^2 = 1 + \frac{1}{2}\left(4\rho^2 - 2\rho\right)\frac{1}{k} + o(1/k) - \left\{1 + \frac{1}{2}\left(\rho^2 - \rho\right)\frac{1}{k} + o(k^{-1})\right\}^2 = \frac{1}{k}\rho^2 + o(k^{-1}).$$

Hence, as $n \to \infty$, $\operatorname{Var}\left(\delta_{j,n}\right) = k^{-1}(1-\rho)^{-2}\left((1-2\rho)^{-1} + \rho^2\right) + o(k^{-1})$.

Next, we calculate $\mathbb{E}\left[\left\{\delta_{j,n} - \mathbb{E}\left(\delta_{j,n}\right)\right\}^4\right]$. By Lemma 2 and Lemma 3, we have that, for $a = 3,4$, as $N \to \infty$,

$$\begin{aligned}
m_n^{(a)} &= (1-\rho)^{-a}\left\{1 + \frac{1}{2}\frac{1}{k}\frac{a(a-1)}{1-2\rho} + O(k^{-2})\right\}\left\{1 + \frac{1}{2}(a^2\rho^2 - a\rho)k^{-1} - \frac{1}{2}(a^2\rho^2 - a\rho)(n-a\rho)^{-1} + O(k^{-2})\right\} \\
&= (1-\rho)^{-a}\left\{1 + k^{-1}\frac{1}{2}\frac{a(a-1)}{1-2\rho} + \frac{1}{2}(a^2\rho^2 - a\rho)k^{-1} - \frac{1}{2}(a^2\rho^2 - a\rho)(n-a\rho)^{-1} + O(k^{-2})\right\}.
\end{aligned}$$

Note that,

$$\mathbb{E}\left[\{(\delta_{j,n} - \mathbb{E}(\delta_{j,n})\}^4\right] = m_n^{(4)} - 4m_n^{(3)}m_n^{(1)} + 6m_n^{(2)}\left(m_n^{(1)}\right)^2 - 3\left(m_n^{(1)}\right)^4.$$

By some direct calculation, all terms of order k^{-1} and n^{-1} are cancelled out. Thus, as $N \to \infty$, $\mathbb{E}\left[\{(\delta_{j,n} - \mathbb{E}(\delta_{j,n})\}^4\right] = O(k^{-2})$. Combining $\mathrm{Var}(\delta_{j,n})$ and $\mathbb{E}\left[\{\delta_{j,n} - \mathbb{E}(\delta_{j,n})\}^4\right]$, we conclude that the sequences $\{\delta_{j,n}\}_{j=1}^m$ satisfy the Lyapunov's condition. Then, (12) follows by the central limit theorem.

Applying (12), we obtain that, as $N \to \infty$,

$$I_2 = \sqrt{km}A_0(n/k)\left\{\mathbb{E}(\delta_{j,n}) + O_P(1/\sqrt{km})\right\} = \frac{g(k,n,\rho)}{1-\rho}\sqrt{km}A_0(n/k) + o_P(1).$$

For I_3, by using the weak law of large numbers for triangular array, we have that, as $N \to \infty$,

$$\begin{aligned}
I_3 &= \frac{\sqrt{km}A_0(n/k)B_0(n/k)}{1-\rho-\tilde{\rho}}\mathbb{E}\left\{\left(kY_1^{(k+1)}/n\right)^{\rho+\tilde{\rho}}\right\}\{1+o_P(1)\} \\
&= \sqrt{km}A_0(n/k)B_0(n/k)\frac{g(k,n,\rho+\tilde{\rho})}{1-\rho-\tilde{\rho}} + o_P(1),
\end{aligned}$$

where the last equality follows by the condition $\sqrt{km}A(n/k)B(n/k) = O(1)$.

For I_4, by similar arguments as for I_3, we obtain that, as $N \to \infty$, $I_4 \xrightarrow{P} 0$. Combining I_1, I_2, I_3 and I_4, we have proved (i).

Next, we handle $R_k^{(2)}$. By (11), we obtain that, as $N \to \infty$,

$$
\begin{aligned}
&\sqrt{km}\left(R_k^{(2)} - 2\gamma^2\right) \\
&= \gamma^2 \frac{1}{mk} \sum_{j=1}^{m} \sum_{i=1}^{k} \left\{ \log^2 \left(Y_j^{(i)} / Y_j^{(k+1)} \right) - 2 \right\} \\
&\quad + 2\gamma\sqrt{km} A_0(n/k) \frac{1}{km} \sum_{j=1}^{m} \left(k Y_j^{(k+1)} / n \right)^{\rho} \sum_{i=1}^{k} \log \left(Y_j^{(i)} / Y_j^{(k+1)} \right) \left\{ \left(Y_j^{(i)} / Y_j^{(k+1)} \right)^{\rho} - 1 \right\} / \rho \\
&\quad + \sqrt{km} A_0^2(n/k) \frac{1}{km} \sum_{j=1}^{m} \left(k Y_j^{(k+1)} / n \right)^{2\rho} \sum_{i=1}^{k} \left\{ \left(Y_j^{(i)} / Y_j^{(k+1)} \right)^{\rho} - 1 \right\}^2 / \rho^2 \\
&\quad + 2\gamma\sqrt{km} A_0(n/k) B_0(n/k) \frac{1}{km} \sum_{j=1}^{m} \left(k Y_j^{(k+1)} / n \right)^{\rho+\tilde{\rho}} \sum_{i=1}^{k} \log \left(Y_j^{(i)} / Y_j^{(k+1)} \right) \frac{\left(Y_j^{(i)} / Y_j^{(k+1)} \right)^{\rho+\tilde{\rho}} - 1}{\rho + \tilde{\rho}} \\
&\quad + o_P(1) \\
&:= I_5 + I_6 + I_7 + I_8 + o_P(1).
\end{aligned}
$$

For I_5, by Lemma 1, we have that

$$
I_5 \stackrel{d}{=} \gamma^2 \sqrt{km} \left\{ \frac{1}{km} \sum_{j=1}^{m} \sum_{i=1}^{k} \left(\log Y_i^{j,*} \right)^2 - 2 \right\}.
$$

The central limit theorem yields that as $N \to \infty$, $I_5 \xrightarrow{d} \gamma^2 P^{(2)}$, where $P^{(2)}$ is distributed as $N(0, 20)$. In addition, the covariance of $P^{(1)}$ and $P^{(2)}$ is equal to the covariance of $\log Y_i^{j,*}$ and $\left(\log Y_i^{j,*}\right)^2$, where $Y_i^{j,*}$ follows the Pareto (1) distribution. Hence, $\mathrm{Cov}(P^{(1)}, P^{(2)}) = 4$.

For I_6, we write $I_6 = 2\sqrt{km} A_0(n/k) m^{-1} \sum_{j=1}^{m} \eta_{j,n}$, where

$$
\eta_{j,n} = \left(k Y_j^{(k+1)} / n \right)^{\rho} (k\rho)^{-1} \sum_{i=1}^{k} \log \left(Y_j^{(i)} / Y_j^{(k+1)} \right) \left\{ \left(Y_j^{(i)} / Y_j^{(k+1)} \right)^{\rho} - 1 \right\}
$$

are i.i.d. random variables for $j = 1, 2, \ldots, m$. We can verify the Lyapunov's condition for the series $\{\eta_{j,n}\}_{j=1}^{m}$ following similar steps as those for $\{\delta_{j,n}\}_{j=1}^{m}$. Then by applying the central limit theorem and Lemma 2, we obtain that

$$
I_6 = 2\gamma\sqrt{km} A_0(n/k) g(k, n, \rho) \frac{1}{\rho} \left\{ \frac{1}{(1-\rho)^2} - 1 \right\} + o_P(1).
$$

By the weak law of large numbers for triangular array, we have that

$$I_7 = \sqrt{km} A_0^2(n/k) \frac{g(k,n,2\rho)}{\rho^2} \left\{ \frac{1}{1-2\rho} - \frac{2}{1-\rho} + 1 \right\} + o_P(1),$$

and

$$I_8 = 2\gamma\sqrt{km} A_0(n/k) B_0(n/k) \frac{g(k,n,\rho+\tilde{\rho})}{\rho+\tilde{\rho}} \left\{ \frac{1}{(1-\rho-\tilde{\rho})^2} - 1 \right\} + o_P(1).$$

Combining the results for I_5, I_6, I_7 and I_8, we have proved (ii).

Finally, we handle $R_k^{(3)}$. Also, by (11), we have that

$$\begin{aligned}
&\sqrt{km}\left(R_k^{(3)} - 6\gamma^3\right) \\
&= \gamma^3 \frac{1}{mk} \sum_{j=1}^{m} \sum_{i=1}^{k} \left\{ \log^3\left(Y_j^{(i)}/Y_j^{(k+1)}\right) - 6 \right\} \\
&\quad + 3\gamma^2 \sqrt{km} A_0(n/k) \frac{1}{km} \sum_{j=1}^{m} \left(kY_j^{(k+1)}/n\right)^{\rho} \sum_{i=1}^{k} \left\{ \log\left(Y_j^{(i)}/Y_j^{(k+1)}\right) \right\}^2 \frac{\left(Y_j^{(i)}/Y_j^{(k+1)}\right)^{\rho} - 1}{\rho} \\
&\quad + 3\gamma \sqrt{km} A_0^2(n/k) \frac{1}{km} \sum_{j=1}^{m} \left(kY_j^{(k+1)}/n\right)^{2\rho} \sum_{i=1}^{k} \log\left(Y_j^{(i)}/Y_j^{(k+1)}\right) \left\{ \frac{\left(Y_j^{(i)}/Y_j^{(k+1)}\right)^{\rho} - 1}{\rho} \right\}^2 \\
&\quad + 3\gamma^2 \sqrt{km} A_0(n/k) B_0(n/k) \frac{1}{km} \sum_{j=1}^{m} \left(kY_j^{(k+1)}/n\right)^{\rho+\tilde{\rho}} \sum_{i=1}^{k} \left\{ \log\left(Y_j^{(i)}/Y_j^{(k+1)}\right) \right\}^2 \frac{\left(Y_j^{(i)}/Y_j^{(k+1)}\right)^{\rho+\tilde{\rho}} - 1}{\rho+\tilde{\rho}} \\
&\quad + o_P(1) \\
&:= I_9 + I_{10} + I_{11} + I_{12} + o_P(1).
\end{aligned}$$

By similar steps as for handling the four items I_5, I_6, I_7 and I_8, we can show that $I_9 \xrightarrow{d} \gamma^3 P^{(3)}$, where $P^{(3)} \sim N(0, 684)$ and $\mathrm{Cov}(P^{(1)}, P^{(3)}) = 18, \mathrm{Cov}(P^{(2)}, P^{(3)}) = 98$. And

$$\begin{aligned}
I_{10} &= 6\gamma^2 \sqrt{km} A_0(n/k) \frac{g(k,n,\rho)}{\rho} \left\{ \frac{1}{(1-\rho)^3} - 1 \right\} + o_P(1), \\
I_{11} &= 3\gamma \sqrt{km} A_0^2(n/k) \frac{g(k,n,2\rho)}{\rho^2} \left\{ \frac{1}{(1-2\rho)^2} - \frac{2}{(1-\rho)^2} + 1 \right\} + o_P(1), \\
I_{12} &= 6\gamma^2 \sqrt{km} A_0(n/k) B_0(n/k) \frac{g(k,n,\rho+\tilde{\rho})}{\rho+\tilde{\rho}} \left\{ \frac{1}{(1-\rho-\tilde{\rho})^3} - 1 \right\} + o_P(1),
\end{aligned}$$

which yields (iii). □

Proof of Theorem 1. Applying Proposition 1 with $k = k_\rho$, we have that, as $N \to \infty$,

$$
\begin{aligned}
R_{k_\rho}^{(1)} &= \gamma + \frac{\gamma}{\sqrt{k_\rho m}} P^{(1)} + \frac{g(k_\rho, n, \rho)}{1-\rho} A_0(n/k_\rho) + \frac{g(k_\rho, n, \rho + \tilde{\rho})}{1-\rho-\tilde{\rho}} A_0(n/k_\rho) B_0(n/k_\rho) + \frac{1}{\sqrt{k_\rho m}} o_P(1), \\
R_{k_\rho}^{(2)} &= 2\gamma^2 + \frac{\gamma^2}{\sqrt{k_\rho m}} P^{(2)} + 2\gamma A_0(n/k_\rho) \frac{g(k_\rho, n, \rho)}{\rho} \left\{ \frac{1}{(1-\rho)^2} - 1 \right\} \\
&\quad + A_0^2(n/k_\rho) \frac{g(k_\rho, n, 2\rho)}{\rho^2} \left(\frac{1}{1-2\rho} - \frac{2}{1-\rho} + 1 \right) \\
&\quad + 2\gamma A_0(n/k_\rho) B_0(n/k_\rho) \frac{g(k_\rho, n, \rho + \tilde{\rho})}{\rho + \tilde{\rho}} \left\{ \frac{1}{(1-\rho-\tilde{\rho})^2} - 1 \right\} + \frac{1}{\sqrt{k_\rho m}} o_P(1), \\
R_{k_\rho}^{(3)} &= 6\gamma^3 + \frac{\gamma^3}{\sqrt{k_\rho m}} P^{(3)} + 6\gamma A_0(n/k_\rho) \frac{g(k_\rho, n, \rho)}{\rho} \left\{ \frac{1}{(1-\rho)^3} - 1 \right\} \\
&\quad + 3A_0^2(n/k_\rho) \frac{g(k_\rho, n, 2\rho)}{\rho^2} \left\{ \frac{1}{(1-2\rho)^2} - \frac{2}{(1-\rho)^2} + 1 \right\} \\
&\quad + 6\gamma A_0(n/k_\rho) B_0(n/k_\rho) \frac{g(k_\rho, n, \rho + \tilde{\rho})}{\rho + \tilde{\rho}} \left\{ \frac{1}{(1-\rho-\tilde{\rho})^3} - 1 \right\} + \frac{1}{\sqrt{k_\rho m}} o_P(1).
\end{aligned}
$$

As a consequence, we have that, as $N \to \infty$,

$$
\begin{aligned}
\left(R_{k_\rho}^{(1)} \right)^\tau &= \gamma^\tau \left\{ 1 + \frac{\tau}{\sqrt{k_\rho m}} P^{(1)} + \frac{\tau}{\gamma} \frac{g(k_\rho, n, \rho)}{1-\rho} A_0(n/k_\rho) + \frac{\tau}{\gamma} \frac{g(k_\rho, n, \rho + \tilde{\rho})}{1-\rho-\tilde{\rho}} A_0(n/k_\rho) B_0(n/k_\rho) \right\} \\
&\quad + \frac{1}{\sqrt{k_\rho m}} o_P(1), \\
\left(R_{k_\rho}^{(2)}/2 \right)^{\tau/2} &= \gamma^\tau \left[1 + \frac{\tau}{\sqrt{k_\rho m}} P^{(2)} + \frac{\tau}{2\gamma} A_0(n/k_\rho) \frac{g(k_\rho, n, \rho)}{\rho} \left\{ \frac{1}{(1-\rho)^2} - 1 \right\} \right. \\
&\quad + \frac{\tau}{4\gamma} A_0^2(n/k_\rho) \frac{g(k_\rho, n, 2\rho)}{\rho^2} \left(\frac{1}{1-2\rho} - \frac{2}{1-\rho} + 1 \right) \\
&\quad \left. + \frac{\tau}{2\gamma} A_0(n/k_\rho) B_0(n/k_\rho) \frac{g(k_\rho, n, \rho + \tilde{\rho})}{\rho + \tilde{\rho}} \left\{ \frac{1}{(1-\rho-\tilde{\rho})^2} - 1 \right\} \right] + \frac{1}{\sqrt{k_\rho m}} o_P(1), \\
\left(R_{k_\rho}^{(3)}/6 \right)^{\tau/3} &= \gamma^\tau \left[1 + \frac{\tau}{\sqrt{k_\rho m}} P^{(3)} + \frac{\tau}{3\gamma} A_0(n/k_\rho) \frac{g(k_\rho, n, \rho)}{\rho} \left\{ \frac{1}{(1-\rho)^3} - 1 \right\} \right. \\
&\quad + \frac{\tau}{6\gamma} A_0^2(n/k_\rho) \frac{g(k_\rho, n, 2\rho)}{\rho^2} \left\{ \frac{1}{(1-2\rho)^2} - \frac{2}{(1-\rho)^2} + 1 \right\} \\
&\quad \left. + \frac{\tau}{3\gamma} A_0(n/k_\rho) B_0(n/k_\rho) \frac{g(k_\rho, n, \rho + \tilde{\rho})}{\rho + \tilde{\rho}} \left\{ \frac{1}{(1-\rho-\tilde{\rho})^3} - 1 \right\} \right] + \frac{1}{\sqrt{k_\rho m}} o_P(1).
\end{aligned}
$$

It follows that, as $N \to \infty$,

$$\gamma^{-\tau}\left\{\left(R_{k_\rho}^{(1)}\right)^{\tau}-\left(R_{k_\rho}^{(2)}/2\right)^{\tau/2}\right\}=\frac{\tau}{\sqrt{k_\rho m}}\left(P^{(1)}-P^{(2)}\right)+\frac{\tau}{\gamma}g(k_\rho,n,\rho)A_0(n/k_\rho)\frac{-\rho}{2(1-\rho)^2}$$
$$+A_0^2\left(n/k_\rho\right)O(1)+A_0\left(n/k_\rho\right)B_0\left(n/k_\rho\right)O(1)+\frac{1}{\sqrt{k_\rho m}}o_P(1),$$

and

$$\gamma^{-\tau}\left\{\left(R_{k_\rho}^{(2)}/2\right)^{\tau/2}-\left(R_{k_\rho}^{(2)}/6\right)^{\tau/3}\right\}=\frac{\tau}{\sqrt{k_\rho m}}\left(P^{(2)}-P^{(3)}\right)+\frac{\tau}{\gamma}g(k_\rho,n,\rho)A_0(n/k_\rho)\frac{\rho(\rho-3)}{6(1-\rho)^3}$$
$$+A_0^2\left(n/k_\rho\right)O(1)+A_0\left(n/k_\rho\right)B_0\left(n/k_\rho\right)O(1)+\frac{1}{\sqrt{k_\rho m}}o_P(1).$$

By the condition (5), the dominating terms in the two expressions above are

$$\frac{\tau}{\gamma}g(k_\rho,n,\rho)A_0(n/k_\rho)\frac{-\rho}{2(1-\rho)^2}\quad\text{and}\quad\frac{\tau}{\gamma}g(k_\rho,n,\rho)A_0(n/k_\rho)\frac{\rho(\rho-3)}{6(1-\rho)^3},$$

respectively. Therefore, as $N \to \infty$,

$$T_{k_\rho,\tau}=3\frac{\rho-1}{\rho-3}\left\{1+\frac{\gamma}{\sqrt{k_\rho m}}\frac{2(1-\rho)^2}{-\rho A_0(n/k_\rho)}\left(P^{(1)}-P^{(2)}\right)-\frac{\gamma}{\sqrt{k_\rho m}A_0(n/k_\rho)}\frac{6(1-\rho)^3}{\rho^2-3\rho}\left(P^{(2)}-P^{(3)}\right)\right\}$$
$$+O_P\left\{A_0\left(n/k_\rho\right)\right\}+O_P\left\{B_0\left(n/k_\rho\right)\right\}+\frac{1}{\sqrt{k_\rho m}A_0(n/k_\rho)}o_P(1).$$

It follows that as $N \to \infty$,

$$\sqrt{k_\rho m}A_0(n/k_\rho)\left(T_{k_\rho,\tau}-3\frac{\rho-1}{\rho-3}\right)=-\gamma\frac{2(1-\rho)^2}{\rho}\left(P^{(1)}-P^{(2)}\right)-\gamma\frac{6(1-\rho)^3}{\rho^2-3\rho}\left(P^{(2)}-P^{(3)}\right)+O_P(1).$$

Theorem 1 is thus proved by applying the Cramér's delta method. □

Proof of Theorem 2. By Proposition 1, as $N \to \infty$, $R_{k_n}^{(1)}$ has the following asymptotic expansion:

$$\sqrt{k_n m}\left(R_{k_n}^{(1)}-\gamma\right)-\gamma P^{(1)}-\frac{g(k_n,n,\rho)}{1-\rho}\sqrt{k_n m}A_0(n/k_n)=o_P(1),$$

which leads to

$$\sqrt{k_n m}\left\{\left(R_{k_n}^{(1)}\right)^2-\gamma^2\right\}-2\gamma^2P^{(1)}-2\gamma\frac{g(k_n,n,\rho)}{1-\rho}\sqrt{k_n m}A_0(n/k_n)=o_P(1).$$

Together with the asymptotic expansion of $R_{k_n}^{(2)}$, we have that, as $N \to \infty$,

$$\sqrt{k_n m}\left\{R_{k_n}^{(2)} - 2\left(R_{k_n}^{(1)}\right)^2\right\} - \gamma^2\left(P^{(2)} - 4P^{(1)}\right) - \sqrt{k_n m}A_0(n/k_n)g(k_n, n, \rho)\frac{2\gamma\rho}{(1-\rho)^2} = o_P(1).$$

Thus, as $N \to \infty$,

$$\begin{aligned}
&\sqrt{k_n m}\left(\tilde{\gamma}_{k_n,k_\rho,\tau} - \gamma\right) \\
&\quad = \sqrt{k_n m}\left(R_{k_n}^{(1)} - \gamma\right) - \frac{1}{2R_{k_n}^{(1)}\hat{\rho}_{k_\rho,\tau}(1-\hat{\rho}_{k_\rho,\tau})^{-1}}\sqrt{k_n m}\left\{R_{k_n}^{(2)} - 2\left(R_{k_n}^{(1)}\right)^2\right\} \\
&\quad = \gamma P^{(1)} + \sqrt{k_n m}A_0(n/k_n)\frac{g(k_n, n, \rho)}{1-\rho} + o_P(1) \\
&\qquad - \frac{1}{2R_{k_n}^{(1)}\hat{\rho}_{k_\rho,\tau}(1-\hat{\rho}_{k_\rho,\tau})^{-1}}\left\{\gamma^2\left(P^{(2)} - 4P^{(1)}\right) + \sqrt{k_n m}A_0(n/k_n)g(k_n, n, \rho)\frac{2\gamma\rho}{(1-\rho)^2} + o_P(1)\right\} \\
&\quad = \gamma P^{(1)} - \frac{\gamma^2(1-\hat{\rho}_{k_\rho,\tau})}{R_{k_n}^{(1)}\hat{\rho}_{k_\rho,\tau}}\left(P^{(2)}/2 - 2P^{(1)}\right) \\
&\qquad + \sqrt{k_n m}A_0(n/k_n)\frac{\rho}{(1-\rho)^2}g(k_n, n, \rho)\left(\frac{1-\rho}{\rho} - \frac{1-\hat{\rho}_{k_\rho,\tau}}{\hat{\rho}_{k_\rho,\tau}}\right) + o_P(1).
\end{aligned}$$

The relation $k_n/k_\rho \to 0$ implies that $A(n/k_n)/A(n/k_\rho) \to 0$ as $N \to \infty$. Thus, by Theorem 1, we have that, as $N \to \infty$,

$$\sqrt{k_n m}A_0(n/k_n)\frac{\rho}{(1-\rho)^2}g(k_n, n, \rho)\left(\frac{1-\rho}{\rho} - \frac{1-\hat{\rho}_{k_\rho,\tau}}{\hat{\rho}_{k_\rho,\tau}}\right) = o_P(1).$$

Together with the consistency of $\hat{\rho}_{k_\rho,\tau}$ and $R_{k_n}^{(1)}$, we have that, as $N \to \infty$,

$$\sqrt{k_n m}\left(\tilde{\gamma}_{k_n,k_\rho,\tau} - \gamma\right) \xrightarrow{P} \frac{\gamma}{\rho}\left\{P^{(2)}(\rho-1)/2 + P^{(1)}(2-\rho)\right\}.$$

Combining with Proposition 1, we obtain that, as $N \to \infty$,

$$\sqrt{k_n m}\left(\tilde{\gamma}_{k_n,k_\rho,\tau} - \gamma\right) \xrightarrow{d} N\left[0, \gamma^2\left\{1 + \left(\rho^{-1} - 1\right)^2\right\}\right].$$

□

Acknowledgement

Liujun Chen and Deyuan Li's research is partially supported by the National Nature Science Foundations of China grants 11971115 and 71661137005.

References

Alves, M. F., Gomes, M. I., and de Haan, L. (2003). A new class of semi-parametric estimators of the second order parameter. *Portugaliae Mathematica*, 60(2):193–214.

Cai, J. J., de Haan, L., and Zhou, C. (2012). Bias correction in extreme value statistics with index around zero. *Extremes*, 16(2):173–201.

Chen, L., Li, D., and Zhou, C. (2021). Distributed inference for extreme value index. *Biometrika*. to appear, https://doi.org/10.1093/biomet/asab001.

Danielsson, J., de Haan, L., Peng, L., and de Vries, C. G. (2001). Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate Analysis*, 76(2):226–248.

de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer Science & Business Media.

de Haan, L., Mercadier, C., and Zhou, C. (2016). Adapting extreme value statistics to financial time series: dealing with bias and serial dependence. *Finance and Stochastics*, 20(2):321–354.

Dekkers, A. L., Einmahl, J. H., and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *Annals of Statistics*, 17(4):1833–1855.

Drees, H., Ferreira, A., and de Haan, L. (2004). On maximum likelihood estimation of the extreme value index. *Annals of Applied Probability*, 14(3):1179–1201.

Fan, J., Wang, D., Wang, K., and Zhu, Z. (2019). Distributed estimation of principal eigenspaces. *Annals of Statistics*, 47(6):3009–3031.

Gomes, M. I., de Haan, L., and Peng, L. (2002). Semi-parametric estimation of the second order parameter in statistics of extremes. *Extremes*, 4(5):387–414.

Gomes, M. I., de Haan, L., and Rodrigues, L. H. (2008). Tail index estimation for heavy-tailed models: accommodation of bias in weighted log-excesses. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):31–52.

Gomes, M. I. and Pestana, D. (2007). A simple second-order reduced bias' tail index estimator. *Journal of Statistical Computation and Simulation*, 77(6):487–502.

Guillou, A. and Hall, P. (2001). A diagnostic for selecting the threshold in extreme value analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):293–305.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3(5):1163–1174.

Li, R., Lin, D. K., and Li, B. (2013). Statistical inference in massive data sets. *Applied Stochastic Models in Business and Industry*, 29(5):399–409.

Prescott, P. and Walden, A. (1980). Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika*, 67(3):723–724.

Volgushev, S., Chao, S.-K., and Cheng, G. (2019). Distributed inference for quantile regression processes. *Annals of Statistics*, 47(3):1634–1662.

Zhou, C. (2009). Existence and consistency of the maximum likelihood estimator for the extreme value index. *Journal of Multivariate Analysis*, 100(4):794–815.