Distributed Inference for Extreme Value Index

Liujun Chen
Fudan University, China

Deyuan Li
Fudan University, China

Chen Zhou
Erasmus University Rotterdam, The Netherlands

Abstract

This paper investigates a divide-and-conquer (DC) algorithm for estimating the extreme value index when data are stored in multiple machines. The oracle property of such a DC algorithm based on extreme value methods is not guaranteed by the general theory of distributed inference. We propose a distributed Hill estimator and establish its asymptotic theories. We consider various cases where the number of observations involved in each machine can be either homogeneous or heterogeneous, either fixed or varying according to the total sample size. In each case, we provide sufficient, sometimes also necessary, condition, under which the oracle property holds. A simulation study shows the finite sample performance of the distributed Hill estimator.

Keywords: Extreme value index, Distributed inference, Distributed Hill estimator

1 Introduction

The availability of enormously large dataset presents a great challenge to traditional statistical methods running on a standalone computer. Such datasets are often stored in multiple machines and cannot be combined into one dataset due to either lacking of hardware or confidentiality. Analyzing such "big data" often requires a divide-and-conquer (DC) algorithm. A DC algorithm, or sometimes referred to as distributed inference, estimates a desired quantity or parameter on each machine and transmits the results to a central machine. The central machine combines all the results, often by a simple averaging, to obtain a computationally feasible estimator.

For a broader set of statistical procedures, under mild conditions, the DC algorithm possesses the oracle property: its speed of convergence and asymptotic distribution coincides with the oracle estimator when applying the same statistical procedure to the hypothetically combined dataset; see Kleiner et al. (2014) for a general discussion and Li et al. (2013) for kernel density estimation.

Nevertheless, the oracle property may not hold for some specific statistical methods, or requires additional conditions. For example, Volgushev et al. (2019) studies distributed inference in quantile regression, and shows both a necessary condition and a sufficient condition to ensure the oracle property of a standard DC algorithm for quantile regression.

Extreme value analysis focuses on statistical inference regarding the tail of a distribution. Similar to quantile regression, the oracle property of a standard DC algorithm based on extreme value methods is not guaranteed by the general theory in distributed inference. For example, considering a distribution with a finite endpoint, a natural estimator for the endpoint is the sample maxima. If an oracle sample maxima cannot be obtained, one may consider collecting the maxima from each subset of the data stored in each machine. Clearly, to obtain the oracle estimator, one needs to take the maximum of the subset maxima rather than taking average. Therefore, the standard DC algorithm based on averaging will lead to an estimator that may not possess the oracle property. Since extreme value analysis are often based on high order statistics, it requires further validation for the oracle property of the standard DC algorithm. This paper aims at providing an initial attempt in this direction.

Consider a distribution function F which belongs to the max-domain of attraction of an extreme value distribution G_γ with index $\gamma \in \mathbb{R}$, denote by $F \in D(G_\gamma)$. Mathematically, there exist constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that

$$\lim_{n\to\infty} F^n\left(a_n x + b_n\right) = G_\gamma(x) := \exp\left(-(1+\gamma x)^{-1/\gamma}\right), \tag{1.1}$$

for all $1+\gamma x > 0$. This is equivalent to the convergence of the quantile function of F: there exists a function $a(t) > 0$ such that

$$\lim_{t\to\infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma}, \quad x > 0, \tag{1.2}$$

where $U(t) = \left(\frac{1}{1-F}\right)^{\leftarrow}(t)$ and $\leftarrow$ denotes the left-continuous inverse function. In this paper, We restrict our attention to the heavy tailed case, *i.e.* $\gamma > 0$. Then the relation (1.2) simplifies to that U is a regular varying function with index $\gamma > 0$, denoted by $U \in RV(\gamma)$, *i.e.*,

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^\gamma, \quad x > 0. \tag{1.3}$$

Suppose we have independent and identically distributed (i.i.d.) observations $X_1, \ldots, X_n$ drawn from F. A key question in extreme value analysis is to estimate the extreme value index γ using the observed sample, see Chapter 3 in de Haan and Ferreira (2006). After estimating the extreme value index, both relations (1.2) and (1.3) can be used for extrapolation beyond the observed tail region, which leads to estimators for high quantiles, see Chapter 4 in de Haan and Ferreira (2006).

We assume that the i.i.d. observations $X_1, \ldots, X_n$ are stored in k machines with m observations each and $n = mk$. Further we assume that $m \to \infty$ and $k \to \infty$ as $n \to \infty$. Under the distributed inference setup, we assume that only one result can be transmitted from each machine to the central machine, due to hardware limitation or confidentiality issue. Practically, we cannot apply statistical procedures to the oracle sample, i.e. the hypothetically combined dataset $\{X_1, \ldots, X_n\}$. Instead, we shall invoke a DC algorithm to solve this problem.

In this paper, we limit our focus to the case $\gamma > 0$. For $\gamma > 0$, an efficient estimator for γ is the Hill estimator (Hill (1975)). Consider an intermediate sequence $l = l(n)$ such that $l \to \infty, l/n \to 0$ as $n \to \infty$. Then the oracle Hill estimator is defined as

$$\hat{\gamma}_H = \frac{1}{l} \sum_{i=1}^{l} \left(\log M^{(i)} - \log M^{(l+1)} \right),$$

where $M^{(1)} \geq \cdots \geq M^{(n)}$ are the order statistics of the oracle sample $\{X_1, X_2, \cdots, X_n\}$. Notice that the oracle Hill estimator involves the top $l+1$ highest order statistics, or in other words, l exceedance ratios $M^{(i)}/M^{(l+1)}$ for $i = 1, 2, \ldots, l$.

Since it is not feasible to obtain the top order statistics of the oracle sample, we follow a DC algorithm by first applying the Hill estimator at each machine, and then taking the average of the Hill estimates from all machines. Let $M_j^{(1)} \geq M_j^{(2)} \geq \cdots \geq M_j^{(m)}$ denote the order statistics within the machine j, we construct the Hill estimator using the d_j exceedance ratios,

$$\hat{\gamma}_j := \frac{1}{d_j} \sum_{i=1}^{d_j} \left(\log M_j^{(i)} - \log M_j^{(d_j+1)} \right).$$

The distributed Hill estimator is then defined as

$$\hat{\gamma}_{DH} := \frac{1}{k} \sum_{j=1}^{k} \hat{\gamma}_j = \frac{1}{k} \sum_{j=1}^{k} \frac{1}{d_j} \sum_{i=1}^{d_j} \left(\log M_j^{(i)} - \log M_j^{(d_j+1)} \right). \tag{1.4}$$

In this paper, we establish the asymptotic theory for the distributed Hill estimator $\hat{\gamma}_{DH}$ and compare its performance to the oracle Hill estimator $\hat{\gamma}_H$. We particularly focus on two issues. Firstly, the construction of the Hill estimators at each machine may be homogeneous or heterogeneous depending on whether they use the same amount of exceedance ratios. Denote the numbers of exceedance ratios used at each machine as $d_1, d_2, \ldots, d_k$. We handle both the homogeneous case where $d_1 = d_2 = \cdots = d_k = d$ and the heterogeneous case where they are different but uniformly bounded. In addition, for the homogeneous case, d can be regarded as either a fixed integer or an intermediate sequence depending on n. Secondly, for each of the aforementioned cases, we show asymptotic theories for the distributed Hill estimator and provide at least a sufficient condition for the oracle property. In some cases the sufficient condition is also necessary.

In addition, we relax the assumption that the oracle sample $X_1, \ldots, X_n$ are i.i.d., in particular, the assumption of being drawn from the same distribution. In application, observations from different sources may follow different distributions, but nevertheless share some common properties in the tail such as the extreme value index. A practical example is to analyze operational risk using operational loss data from various banks. Given that operational losses are scarce, it is ideal to combine data from various institutions in order to analyze the tail of the loss distribution. However, due to confidentiality, banks may not share such data among each other. Conducting distributed inference using extreme value analysis is therefore necessary. Nevertheless, since banks differ essentially in terms of size, business model and activities, it is inappropriate to assume that their operational losses follow the same distribution.

In line with the aforementioned example, we assume that observations on a given machine follow the same distribution, whereas across machines, the tails of distributions are only comparable in the context of heteroscedastic extremes, see Einmahl et al. (2016). Consequently, all distributions share the same extreme value index. This relaxation does not affect the main asymptotic theory for estimating the common extreme value index. The discussion on the oracle property remains valid.

The paper is organized as follows. Section 2 shows the asymptotic behaviour for the distributed Hill estimator $\hat{\gamma}_{DH}$ based on i.i.d. observations. Section 3 shows the asymptotic behaviour when the observations are not identically distributed. Section 4 provides a simulation study to illustrate the finite sample behaviour of the distributed Hill estimator. The proofs are given in Section 5.

2 Main Results: IID Observations

In this section, we show the asymptotic behaviour of the distributed Hill estimator $\hat{\gamma}_{DH}$ when the oracle sample $X_1, X_2, \ldots, X_n$ are i.i.d.. Recall that the oracle sample are stored in k machines with m observations each.

We discuss the homogeneous case where $d_1 = \cdots = d_k = d$ is a fixed integer, the heterogeneous case where d_j are uniformly bounded and the homogeneous case where $d = d(m)$ is an intermediate sequence in Section 2.1, Section 2.2 and Section 2.3, respectively.

In order to investigate the asymptotic behaviour of the distributed Hill estimator, we impose the following condition on the sequences k and m.

Condition A. $k = k(n) \to \infty$, $m = m(n) \to \infty$ and $m/\log k \to \infty$, as $n \to \infty$.

The last limit relation in **Condition A** requires that the number of observations in one machine m is not too low. Notice that the distributed Hill estimator essentially involves kd exceedance ratios: d from each machine. Comparing with an oracle Hill estimator which involves kd exceedance ratios of the oracle sample, if $k = 1$ and $m = n$, the two sets of exceedance ratios are the same. As m decreases, the difference between the two sets of exceedance ratio arises with the distributed Hill estimator involving more "non-extreme" observations from the oracle sample. Theoretically, if m is too low, the distributed Hill estimator fails to be consistent. We further remark that the condition is not too restrictive: for instance, if $m = n^a$ with any $a \in (0, 1)$, then **Condition A** holds.

2.1 Homogeneous case where $d_1 = d_2 = \cdots = d_k = d$ is a fixed integer

The following theorem shows the consistency of the distributed Hill estimator $\hat{\gamma}_{DH}$ when $d_1 = d_2 = \cdots = d_k = d \geq 1$ is a fixed integer.

Theorem 2.1. Suppose $F \in D(G_\gamma)$ with $\gamma > 0$ and **Condition A** holds. Let $d_1 = d_2 = \cdots = d_k = d$, where $d \geq 1$ is a fixed integer. Then as $n \to \infty$,

$$\hat{\gamma}_{DH} \xrightarrow{P} \gamma.$$

To obtain the asymptotic normality of the distributed Hill estimator $\hat{\gamma}_{DH}$, we need some second order condition quantifying the rate of convergence in (1.3) as follows.

Condition B. (Second Order Condition) There exists an eventually positive or negative function

$A(t) \in RV(\rho)$ with $\rho \leq 0$ and $\lim_{t\to\infty} A(t) = 0$ 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 $x > 0$ (see e.g. de Haan and Ferreira (2006), Corollary 2.3.4).

THEOREM 2.2. Suppose $F \in D(G_\gamma)$ with $\gamma > 0$ and **Condition A** and **B** hold. Let $d_1 = d_2 = \cdots = d_k = d$, where $d \geq 1$ is a fixed integer. If $\sqrt{kd}A(m/d) = O(1)$ as $n \to \infty$, then

$$\sqrt{kd}\,(\hat{\gamma}_{DH} - \gamma - A(m/d)g(d,m,\rho)) \xrightarrow{d} N(0, \gamma^2),$$

where

$$g(d,m,\rho) = \frac{1}{1-\rho}\left(\frac{m}{d}\right)^{-\rho} \frac{\Gamma(m+1)\Gamma(d-\rho+1)}{\Gamma(m-\rho+1)\Gamma(d+1)}.$$

REMARK 2.1. If **Condition B** holds and $\sqrt{kd}A(m/d) = O(1)$ as $n \to \infty$, then there exists a constant $l < 1$ such that $kd = O(n^l)$ as $n \to \infty$, which implies $m/\log k \to \infty$ required in **Condition A**.

To investigate the oracle property of the distributed Hill estimator, we compare its asymptotic behavior to that of the oracle Hill estimator using kd exceedance ratios, denoted as $\hat{\gamma}_H$. As in Theorem 3.2.5 in de Haan and Ferreira (2006), for the oracle Hill estimator, assume that $\sqrt{kd}A(n/(kd)) = \sqrt{kd}A(m/d) \to \lambda \in \mathbb{R}$ as $n \to \infty$. Then it possesses asymptotic normality as follows: as $n \to \infty$,

$$\sqrt{kd}(\hat{\gamma}_H - \gamma) \xrightarrow{d} N\left(\frac{\lambda}{1-\rho}, \gamma^2\right).$$

Under the same condition, we have that as $n \to \infty$,

$$g(d,m,\rho) \to \frac{d^\rho}{1-\rho}\frac{\Gamma(d-\rho+1)}{\Gamma(d+1)},$$

see (5.6) in the proof of Theorem 2.2. Therefore, the asymptotic normality for the distributed Hill estimator can be expressed as

$$\sqrt{kd}(\hat{\gamma}_{DH} - \gamma) \xrightarrow{d} N\left(\lambda\frac{d^\rho}{1-\rho}\frac{\Gamma(d-\rho+1)}{\Gamma(d+1)}, \gamma^2\right),$$

as $n \to \infty$.

Notice that the two estimators share the same asymptotic variance. Obviously, if $\lambda = 0$ or $\rho = 0$, the two estimators share the same asymptotic bias. In other words, the distributed Hill estimator achieves the oracle property if $\lambda = 0$ or $\rho = 0$. The following corollary shows that $\lambda = 0$ or $\rho = 0$ is not only sufficient but also necessary for the oracle property.

COROLLARY 2.1. Under the same conditions as in Theorem 2.2, if $\sqrt{kd}A(m/d) \to \lambda$ as $n \to \infty$, then the distributed Hill estimator possess oracle property if and only if $\lambda = 0$ or $\rho = 0$.

2.2 Heterogeneous case where d_j are uniformly bounded

Next, we handle the heterogeneous case with different $d_1, d_2, \ldots, d_k$. We assume that $\{d_j\}_{j=1}^k, k \in \mathbb{N}$ are uniformly bounded positive integer series, i.e.,

$$d_{max} = \sup_{k \in \mathbb{N}} \max_{1 \le j \le k} d_j < \infty.$$

The asymptotic behaviour of the distributed Hill estimator is presented in the following theorems.

THEOREM 2.3. Suppose $F \in D(G_\gamma)$ with $\gamma > 0$ and **Condition A** holds. Let $d_1, d_2, \ldots, d_k$ be uniformly bounded positive integers. Then as $n \to \infty$,

$$\hat{\gamma}_{DH} \xrightarrow{P} \gamma.$$

THEOREM 2.4. Suppose $F \in D(G_\gamma)$ with $\gamma > 0$ and **Condition A** and **B** hold. Let $d_1, d_2, \ldots, d_k$ be uniformly bounded positive integers. If $\sqrt{k\bar{d}}A(m/\bar{d}) = O(1)$ as $n \to \infty$, then

$$\sqrt{k\bar{d}}\left(\hat{\gamma}_{DH} - \gamma - A(m/\bar{d})\frac{1}{k}\sum_{j=1}^{k}\left(\frac{\bar{d}}{d_j}\right)^{\rho} g(d_j, m, \rho)\right) \xrightarrow{d} N(0, \gamma^2),$$

where $\bar{d} = k^{-1}\sum_{j=1}^k d_j$.

We further investigate the oracle property of the distributed Hill estimator in this case. First, we derive the asymptotic behavior of the oracle Hill estimator using $\sum_{j=1}^k d_j = k\bar{d}$ exceedance ratios, denoted as $\hat{\gamma}_H$. Similar to Section 2.1, with assuming that $\sqrt{k\bar{d}}A(n/(k\bar{d})) = \sqrt{k\bar{d}}A(m/\bar{d}) \to \lambda \in \mathbb{R}$ as $n \to \infty$, the oracle Hill estimator possesses the following asymptotic normality,

$$\sqrt{k\bar{d}}(\hat{\gamma}_H - \gamma) \xrightarrow{d} N\left(\frac{\lambda}{1-\rho}, \gamma^2\right).$$

Under the same condition, we calculate the asymptotic bias for the distributed Hill estimator using the Stirling's formula. Since the positive integers $\{d_j\}_{j=1}^k$ are uniformly bounded, as $n \to \infty$,

$$g(d_j, m, \rho) \to \frac{d_j^\rho}{1-\rho} \frac{\Gamma(d_j - \rho + 1)}{\Gamma(d_j + 1)},$$

holds uniformly for all $1 \le j \le k$. It follows that as $n \to \infty$,

$$\frac{1}{k} \sum_{j=1}^k \left(\frac{\bar{d}}{d_j} \right)^\rho g(d_j, m, \rho) \sim \frac{\bar{d}^\rho}{1-\rho} \frac{1}{k} \sum_{j=1}^k \frac{\Gamma(d_j - \rho + 1)}{\Gamma(d_j + 1)}.$$

We conclude again that the distributed Hill estimator achieves the oracle property when $\lambda = 0$ or $\rho = 0$. Nevertheless, due to the complex structure of the asymptotic bias, it is not guaranteed that this condition is also necessary.

2.3 Homogeneous case where $d = d(m)$ is an intermediate sequence

Lastly, we handle the case when the number of exceedance ratios used in each machine is an intermediate sequence tending to infinity. Assume that $d_1 = d_2 = \cdots = d_k = d$ and $d = d(m)$ is an intermediate sequence, i.e.,

$$d = d(m) \to \infty, d/m \to 0$$

as $n \to \infty$. The following theorems show that the asymptotic theories for the distributed Hill estimator still hold.

Theorem 2.5. Suppose $F \in D(G_\gamma)$ with $\gamma > 0$ and **Condition A** holds. Let $d_1 = d_2 = \cdots = d_k = d,\ d = d(m) \to \infty$ and $d/m \to 0$ as $n \to \infty$, then

$$\hat{\gamma}_{DH} \xrightarrow{P} \gamma.$$

Theorem 2.6. Suppose $F \in D(G_\gamma)$ with $\gamma > 0$ and **Condition A** and **B** hold. Let $d_1 = d_2 = \cdots = d_k = d,\ d = d(m) \to \infty$ and $d/m \to 0$ as $n \to \infty$. If $\sqrt{kd} A(m/d) = O(1)$ as $n \to \infty$, then

$$\sqrt{kd} \left(\hat{\gamma}_{DH} - \gamma - A(m/d) g(d, m, \rho) \right) \xrightarrow{d} N(0, \gamma^2).$$

Remark 2.2. We remark that the results in Theorem 2.5 and 2.6 hold under a weaker version of

Condition A. In particular, the requirements $k \to \infty$ and $m \to \infty$ as $n \to \infty$ are not needed. Recall that the distributed Hill estimator essentially uses kd exceedance ratios in the estimation. To ensure the asymptotic theories, we need $kd \to \infty$ as $n \to \infty$. Since $d \to \infty$, as $n \to \infty$, it is not necessary to further require $k \to \infty$. In addition, since $m \geq d$, $m \to \infty$ as $n \to \infty$ holds automatically.

We further investigate the oracle property of the distributed Hill estimator in this case. First, similar to Section 2.1, with assuming that $\sqrt{kd}A(n/(kd)) = \sqrt{kd}A(m/d) \to \lambda \in \mathbb{R}$ as $n \to \infty$, the oracle Hill estimator possesses the following asymptotic normality,

$$\sqrt{kd}(\hat{\gamma}_H - \gamma) \xrightarrow{d} N\left(\frac{\lambda}{1-\rho}, \gamma^2\right).$$

Under the same condition, since as $n \to \infty$,

$$g(d, m, \rho) \to \frac{1}{1-\rho},$$

we get that

$$\sqrt{kd}(\hat{\gamma}_{DH} - \gamma) \xrightarrow{d} N\left(\frac{\lambda}{1-\rho}, \gamma^2\right).$$

Hence, we conclude that the distributed Hill estimator always achieves the oracle property.

Comparing the three scenarios in Section 2.1-2.3, we conclude that whether the distributed Hill estimator achieves the oracle property depends on the theoretical setup for the the number of exceedance ratios used in each machine d. If d is homogeneous across machines and can be regarded as an intermediate sequence, the oracle property always holds true. If d is homogeneous across machines but is at a low level, i.e. regarded as a fixed integer, the oracle property holds true if and only if $\lambda = 0$ or $\rho = 0$. Notice that $\lambda = 0$ corresponds to zero asymptotic bias for the oracle Hill estimator. Finally, if $d_1, d_2, \cdots, d_k$ vary across machines but are uniformly bounded, the condition $\lambda = 0$ or $\rho = 0$ is only sufficient and may not be necessary.

Overall, based on the asymptotic theories, the DC algorithm can be applied to the Hill estimator without additional condition, if one uses sizable amount of exceedance ratios in each machine. We shall verify this theoretical finding by a simulation study when the sample size is finite.

3 Non Identically Distributed Case

We further investigate the case where observations on different machines follow different distributions. Recall again that the oracle sample $\{X_1, X_2, \ldots, X_n\}$ are stored in k machines with m observations each. Let X_j^i denote ith observation in machine j. Assume all observations are independent, but only observations on the same machine follow the same distribution. Observations across machines are not necessarily identically distributed. Denote the common distribution function of $\{X_j^1, X_j^2, \ldots, X_j^m\}$ as $F_{k,j}$ for $j = 1, 2, \ldots, k$. We assume that the tails of $\{F_{k,j}\}_{j=1}^k$ are comparable as follows.

Condition C. There exists a continuous distribution function F such that

$$\lim_{x\to\infty} \frac{1 - F_{k,j}(x)}{1 - F(x)} = c_{k,j}, \tag{3.1}$$

uniformly for all $1 \leq j \leq k$ and all $k \in \mathbb{N}$ with $c_{k,j}$ uniformly bounded away from 0 and ∞.

Note that **Condition C** describes a flexible non-parametric model that allows for different scales in the tails, which is similar to the model in Einmahl et al. (2016) for heteroscedastic extremes.

Define

$$U_{k,j}(t) = \left(\frac{1}{1 - F_{k,j}}\right)^{\leftarrow}(t).$$

Again, we assume $F \in D(G_\gamma)$ with $\gamma > 0$. It is straightforward to derive that (3.1) is equivalent to

$$\lim_{t\to\infty} \frac{U_{k,j}(t)}{U(t)} = c_{k,j}^{\gamma} \tag{3.2}$$

uniformly for all $1 \leq j \leq k$ and all $k \in \mathbb{N}$.

To obtain the asymptotic normality, we need some second-order condition quantifying the speed of convergence in (3.2) as follows.

Condition D. There exists an eventually positive or negative function $A_1(t) \in RV(\tilde{\rho})$ with index $\tilde{\rho} \leq 0$ and $\lim_{t\to\infty} A_1(t) = 0$ such that

$$\sup_{k\in\mathbb{N}} \max_{1\leq j\leq k} \left| \frac{U_{k,j}(t)}{U(t)} - c_{k,j}^{\gamma} \right| = O\left(A_1(t)\right), \tag{3.3}$$

as $t \to \infty$.

Under the heteroscedastic extremes setup, Einmahl et al. (2016) shows that one may pool all observations together and apply the Hill estimator to estimate the common extreme value index γ. In other words, hypothetically, one could apply the Hill estimator to the oracle example while discarding the fact that they are not from the same distribution. We define such an estimator as the oracle Hill estimator.

In the distributed inference setup, we can only apply the Hill estimator at each machine and then average the estimates. The two theorems below show that, the consistency and asymptotic normality of the distributed Hill estimator remain valid.

THEOREM 3.1. Suppose $F \in D(G_\gamma)$ with $\gamma > 0$ and **Condition A** and **C** hold. Then as $n \to \infty$,

$$\hat{\gamma}_{DH} \xrightarrow{P} \gamma.$$

THEOREM 3.2. Suppose $F \in D(G_\gamma)$ with $\gamma > 0$ and **Conditions A-D** hold. Let $d_1 = d_2 = \cdots = d_k = d$, where $d \geq 1$ is a fixed integer. If $\sqrt{kd}A(m/d) = O(1)$ and $\sqrt{kd}A_1(m/d) \to 0$ as $n \to \infty$, then

$$\sqrt{kd}\,(\hat{\gamma}_{DH} - \gamma - A(m/d)g(d, m, \rho)) \xrightarrow{d} N(0, \gamma^2).$$

REMARK 3.1. Following the same lines of proof as that for Theorem 2.4 and Theorem 2.6, Theorem 3.2 can be extended to the heterogeneous case where d_j are uniformly bounded and the homogeneous case where $d = d(m)$ is an intermediate sequence respectively.

Einmahl et al. (2016) shows that if the oracle Hill estimator $\hat{\gamma}_H$ uses kd exceedance ratios with kd satisfying $\sqrt{kd}A(n/(kd)) \to 0$ as $n \to \infty$, the asymptotic normality holds: as $n \to \infty$,

$$\sqrt{kd}\,(\hat{\gamma}_H - \gamma) \xrightarrow{d} N(0, \gamma^2).$$

Compared to the result in Theorem 3.2, we conclude that the distributed Hill estimator $\hat{\gamma}_{DH}$ possesses the oracle property under the same condition. Note that Einmahl et al. (2016) did not handle the case $\sqrt{kd}A(n/(kd)) = O(1)$ as $n \to \infty$, while we can handle this case for the distributed Hill estimator.

4 Simulation

We conduct a simulation study to demonstrate the finite sample performance of the distributed Hill estimators $\hat{\gamma}_{DH}$. We consider three distributions: the Fréchet(1) distribution, the Pareto(1) distribution and the absolute Cauchy distribution. All three distributions belong to the max-domain of attraction of an extreme value distribution with $\gamma = 1$.

- Fréchet(1) distribution:
$$F(x) = e^{-x^{-1}}, \quad x > 0.$$

- Pareto(1) distribution:
$$F(x) = 1 - x^{-1}, \quad x > 0.$$

- Absolute Cauchy distribution: the probability density function is given as
$$f(x) = \frac{2}{\pi(1 + x^2)}, \quad x > 0.$$

We generate samples from all of three distributions with sample size $n = 1000$. Based on $r = 1000$ Monte Carlo repetitions, we obtain the finite sample bias, variance and mean squared error (MSE) for all considered estimators.

4.1 Comparison for different level of d

First, we vary the level of d in the distributed Hill estimator to verify the theoretical results on the oracle property. The oracle sample $\{X_1, X_2, \dots, X_n\}$ contains $n = 1000$ observations stored in k machines with m observations each. We fix $k = 20$ and $m = 50$ and compare the finite sample performance of the distributed Hill estimator with that of the oracle Hill estimator for different values of d. Recall that total number of exceedance ratios involved in the oracle Hill estimator is kd.

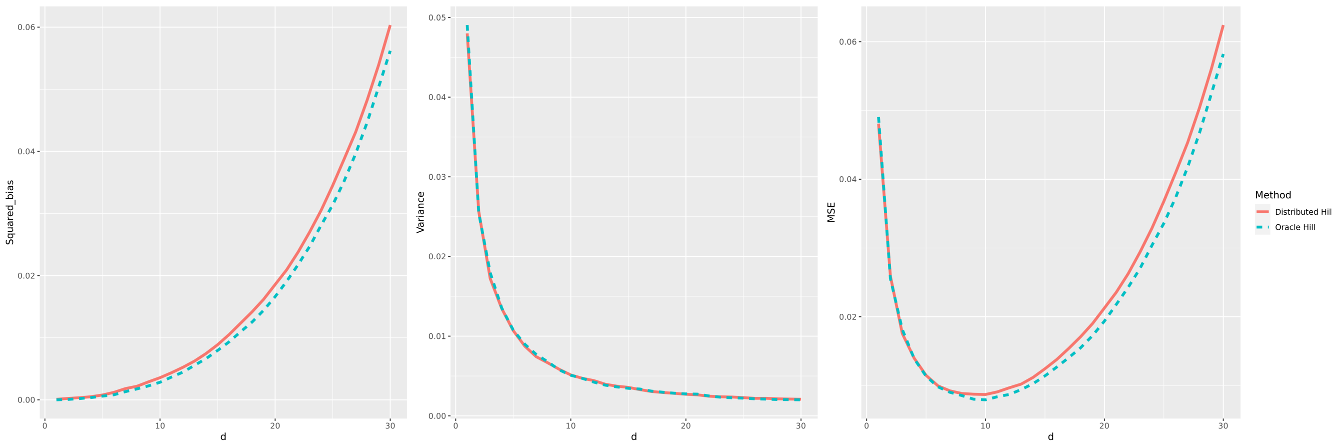
The results are presented in Figure 1. For the Fréchet(1) distribution and the absolute Cauchy distribution, we observe a trade off between the bias and variance for the both estimators: as d increases, the bias increases while the variance decreases. For the Pareto(1) distribution, the bias is virtually zero for all levels of d.

Figure 1 shows that $\hat{\gamma}_H$ and $\hat{\gamma}_{DH}$ have almost the same variance for all levels of d. When d is low, the number of exceedance ratios used in both estimators kd is low, and hence the bias for the oracle Hill estimator is close to zero. This in line with the situation $\lambda = 0$ in Theorem 2.2. For this case, we do not observe sizeable difference between the biases of $\hat{\gamma}_H$ and $\hat{\gamma}_{DH}$. Consequently, the difference in MSE is also negligible. As d increases, the biases of the distributed Hill estimator and the oracle Hill estimator become visible for the Fréchet(1) and the absolute Cauchy distribution. Nevertheless, the biases for these two estimators are still comparable. This is in line with the situation when d is an intermediate sequence as in Theorem 2.6.

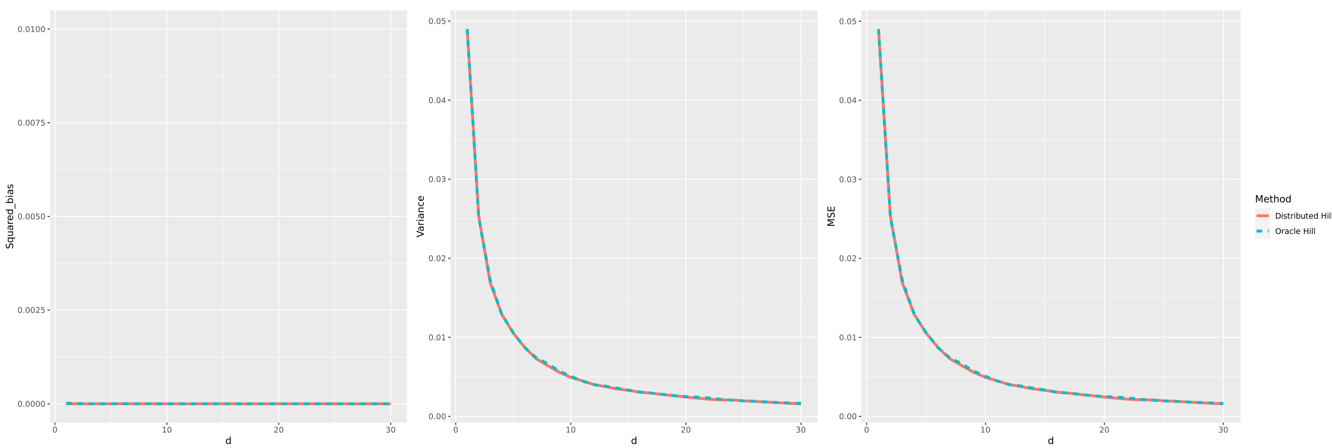
4.2 Comparison for different level of k

Next, we compare the finite sample performances of the distributed Hill estimator $\hat{\gamma}_{DH}$ and the oracle Hill estimator $\hat{\gamma}_H$ for various levels of k. For the oracle estimator, we use $k_H = 40, \cdots, 400$ exceedance ratios. Then for each d, we construct the distributed Hill estimator with $k = k_H/d$ number of machines. We fix d at two levels: a low level ($d = 2$) and a relatively high level ($d = 8$) and denote the corresponding estimator as $\hat{\gamma}_{DH,2}$ and $\hat{\gamma}_{DH,8}$ respectively. The results are presented in Figure 2.

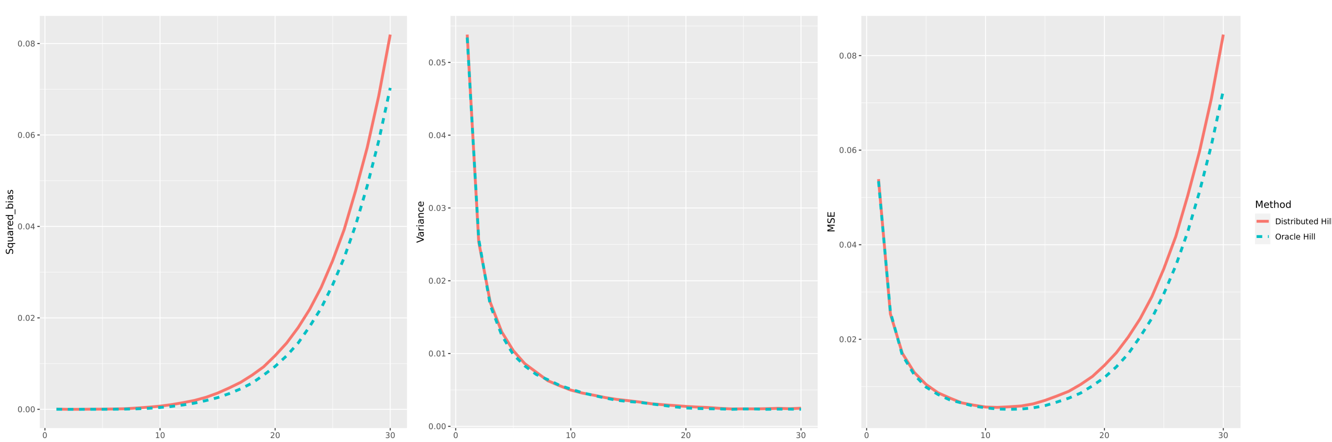
Figure 2 shows that $\hat{\gamma}_{DH,2}$, $\hat{\gamma}_{DH,8}$ and $\hat{\gamma}_H$ have almost the same variance for all levels of k_H, which is in line with the theoretical result. For the Pareto(1) distribution, $\hat{\gamma}_{DH,2}$ and $\hat{\gamma}_{DH,8}$ behave closely to the oracle Hill estimator $\hat{\gamma}_H$. Note that the Pareto(1) distribution corresponds to $\lambda = 0$ for all levels of k_H. For the Fréchet(1) distribution and the absolute Cauchy distribution, the oracle property holds for $\hat{\gamma}_{DH,2}$ and $\hat{\gamma}_{DH,8}$ when the number of exceedance ratios k_H is low. This is line with the situation $\lambda = 0$ in Theorem 2.2. As the level of k_H increases, the performance of $\gamma_{DH,8}$ is closer to the performance of the oracle Hill estimator compared to that of $\gamma_{DH,2}$. This is in line with the theoretical result since $d = 2$ can be regarded as fixed integers while $d = 8$ can be regarded as an intermediate sequences as the level of k_H increases.



(a) Fréchet(1) Distribution

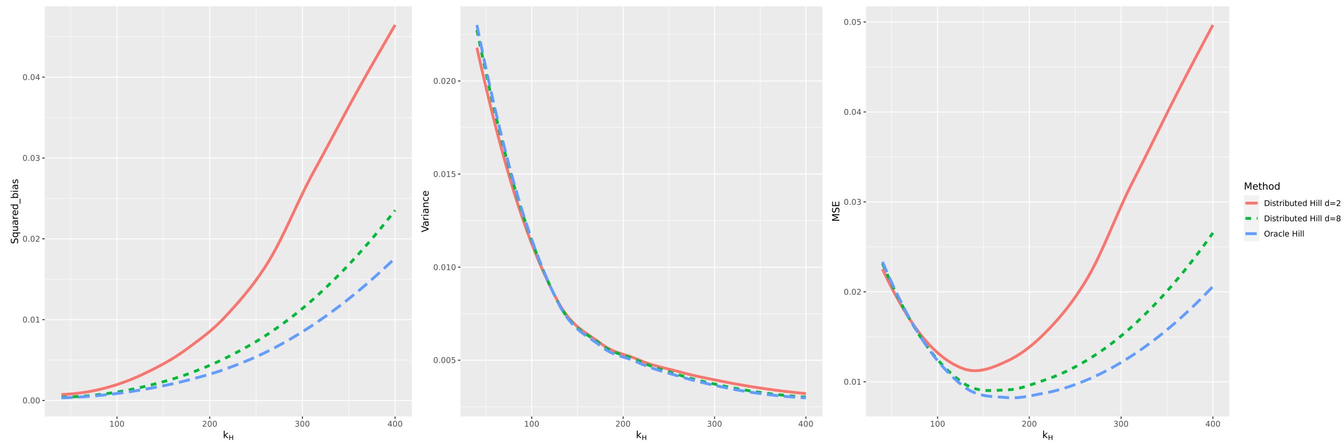


(b) Pareto(1) Distribution

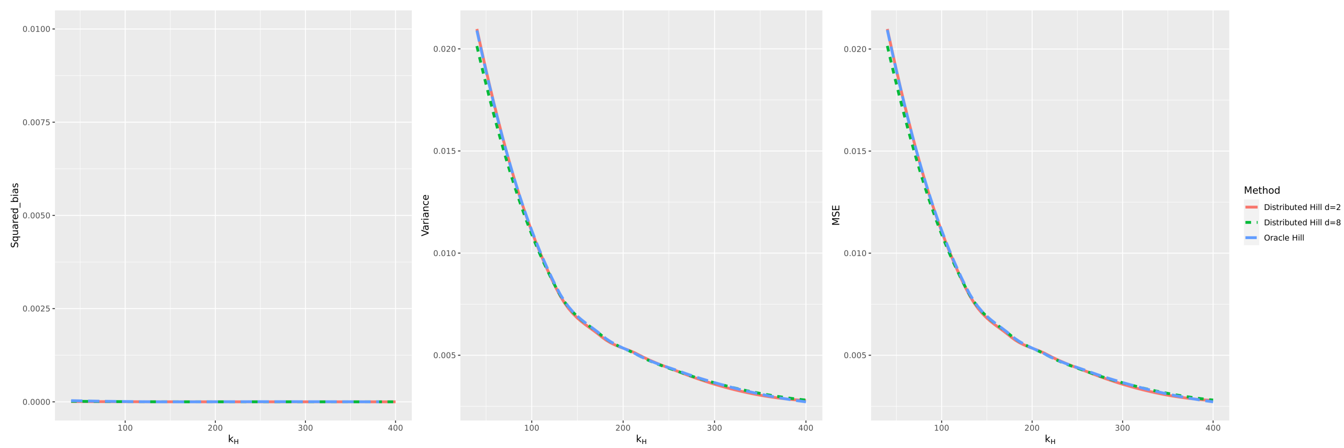


(c) Absolute Cauchy Distribution

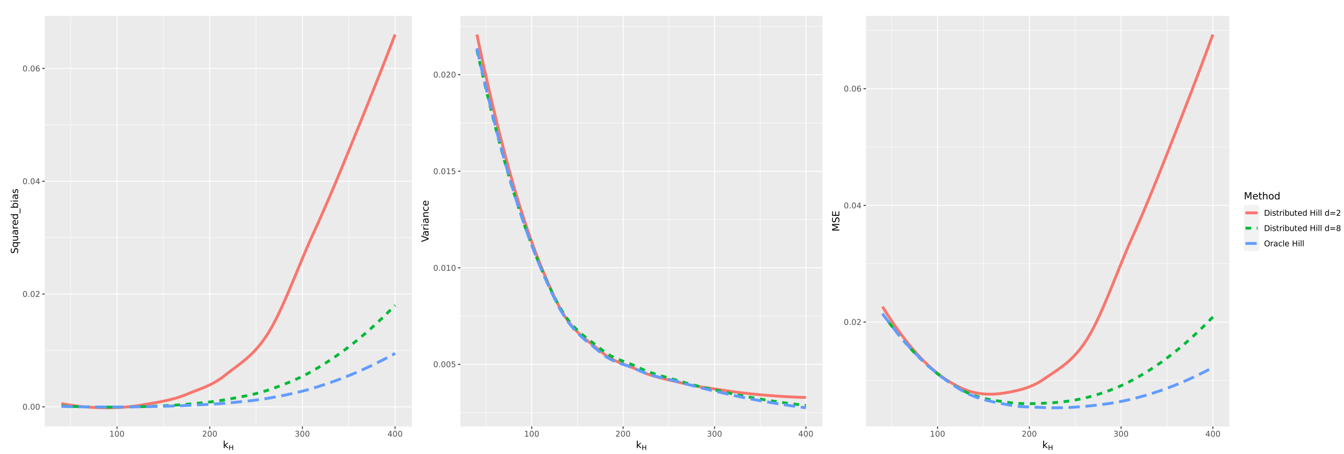
Figure 1: Finite sample performance for the distributed Hill estimator and the oracle Hill estimator for different levels of d.



(a) Fréchet(1) Distribution



(b) Pareto(1) Distribution



(c) Absolute Cauchy Distribution

Figure 2: Finite sample performance for the distributed Hill estimator and the oracle Hill estimator for different levels of k.

4.3 Comparison with the block maxima approach

The distributed inference setup fits into the so-called "block maxima" approach in extreme value statistics. Note that condition (1.1) is equivalent to the convergence in sample maxima:

$$\frac{\max\{X_1, \ldots, X_n\} - b_n}{a_n} \xrightarrow{d} G_\gamma,$$

where $a_n > 0$ and $b_n \in \mathbb{R}$ are the same series as in (1.1). Suppose we collect the maxima from each machine. They can be regarded as i.i.d. observations from the distribution of $\max\{X_1, X_2, \ldots, X_m\}$. Thus, after linear normalization using a_m and b_m, machine-wise maxima should follow approximately the extreme value distribution G_γ. One may transmit the observed machine-wise maxima to the central machine and fit them to a generalized extreme value distribution (GEV) to obtain an estimator of γ. This is the so-called "block maxima" approach, see e.g. Bücher and Segers (2018) and Dombry and Ferreira (2019). However, the block maxima approach is not a DC algorithm: the treatment of all machine-wise maxima is based on available information of the underlying statistical problem, rather than using a simple average. We are therefore interested in comparing the finite sample performance of the distributed Hill estimator with the maximum likelihood estimator (MLE) using the block maxima approach, denote by $\hat{\gamma}_M$. We fix $d = 2$ in this simulation for the distributed Hill estimator. The block size is always set to m.

For the calculation of $\hat{\gamma}_M$, we have three technical choices. First, since we restrict our attention to the heavy-tailed case, we fit the block maxima to the generalized Fréchet distribution (with scale and shape parameter) instead of the generalized extreme value distribution; see Bücher and Segers (2018). Secondly, we use the left-truncated block maxima $M_j^{(1)} \vee c$, $j = 1, 2, \ldots, k$, for some small positive constant c, as suggested by Bücher and Segers (2018) . Asymptotically, such left-truncation does not affect the limit behaviour, since $\mathbb{P}(M_j^{(1)} > c) \to 1$ as block size $m \to \infty$. In this simulation study, we set $c = 0.1$. Lastly, $\hat{\gamma}_M$ does not admit an analytical form and we obtain it by numerical optimization. However, the numerical algorithm may not converge in some situations. We omit these failures and only use the successful repetitions to calculate the bias, variance and MSE. The failure occurs more frequently when k is high. Nevertheless, the number of failure cases is only a small proportion (less than 5%) of the total repetitions in this simulation.

Figure 3 shows that $\hat{\gamma}_{DH}$ has a lower variance for all distributions compared to $\hat{\gamma}_M$. This is in line with the asymptotic theory: the asymptotic variance for $\hat{\gamma}_{DH}$ and $\hat{\gamma}_M$ is $\gamma^2/2$ and $(6/\pi^2)\gamma^2$

respectively. For the Pareto(1) distribution, $\hat{\gamma}_{DH}$ outperforms $\hat{\gamma}_M$ for all levels of k as the bias is virtually zero. For the Fréchet(1) and the absolute Cauchy distribution, $\hat{\gamma}_{DH}$ outperforms $\hat{\gamma}_M$ when k is small. As k increases, the bias of $\hat{\gamma}_{DH}$ increases for the Fréchet(1) and the absolute Cauchy distribution. Due to the max-stability of the Fréchet(1) distribution, $\hat{\gamma}_M$ does not suffer from an asymptotic bias. Eventually, the difference in biases dominates the difference in variance.

5 Proofs

5.1 Preliminary

Recall that $U = (1/(1-F))^{\leftarrow}$. Then $X \stackrel{d}{=} U(Z)$ where Z follows the Pareto(1) distribution with distribution function $1 - 1/z, z \geq 1$. Since we have i.i.d. observations $\{X_1, X_2, \ldots, X_n\}$, we can write $X_i \stackrel{d}{=} U(Z_i)$ where $\{Z_1, Z_2, \ldots, Z_n\}$ is a random sample of Z. The n observations are stored in k machines with m observations each. For each machine j, let $Z_j^{(1)} \geq Z_j^{(2)} \geq \cdots \geq Z_j^{(m)}$ denote the order statistics of the observations in this machine. We have the following lemmas regarding these order statistics.

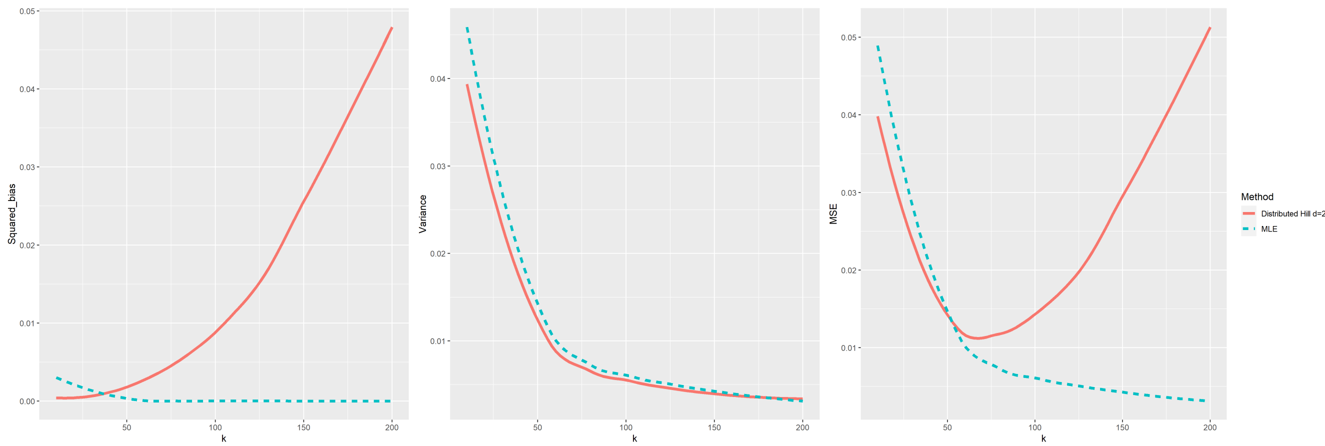
Lemma 5.1. Let $Z^{(1)} \geq Z^{(2)} \geq \cdots \geq Z^{(m)}$ denote the order statistics of the i.i.d. Pareto(1) distributed random variables $\{Z_1, Z_2, \ldots, Z_m\}$. Then

$$\log \frac{Z^{(1)}}{Z^{(2)}},\ 2\log \frac{Z^{(2)}}{Z^{(3)}},\ \ldots, (m-1)\log \frac{Z^{(m-1)}}{Z^{(m)}}$$

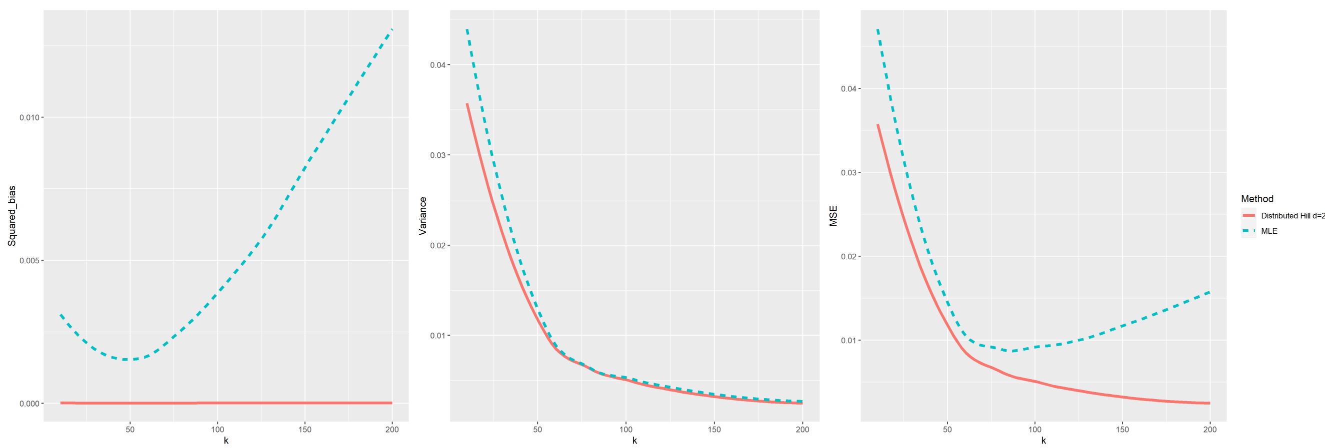
can be regarded as i.i.d. standard exponentially distributed random variables.

Proof of Lemma 5.1. Note that $\{E_i = \log Z_i, i = 1, 2, \ldots, m\}$ form a a random sample from the standard exponential distribution. Denote $E^{(1)} \geq E^{(2)} \geq \cdots \geq E^{(m)}$ as the order statistics of $\{E_1, E_2, \ldots, E_m\}$. By Rényi (1953) we have

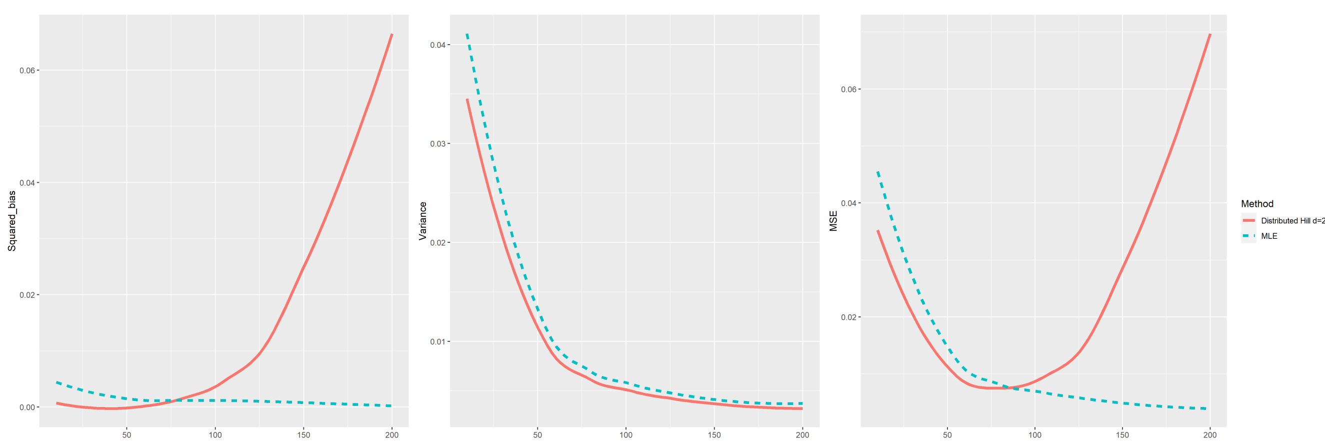
$$\begin{aligned}
&(E^{(m)}, E^{(m-1)}, \ldots, E^{(1)}) \\
&\qquad \stackrel{d}{=} \left(\frac{E_m^*}{m}, \frac{E_m^*}{m} + \frac{E_{m-1}^*}{m-1}, \cdots, \frac{E_m^*}{m} + \frac{E_{m-1}^*}{m-1} + \cdots + \frac{E_1^*}{1} \right),
\end{aligned}$$



(a) Fréchet(1) Distribution



(b) Pareto(1) Distribution



(c) Absolute Cauchy Distribution

Figure 3: Finite sample performance of the distributed Hill estimator and the maximum likelihood estimator using the block maxima approach.

where $E_1^*, \ldots, E_m^*$ are i.i.d. standard exponentially distributed random variables. This implies

$$\left(\log \frac{Z^{(1)}}{Z^{(2)}}, 2\log \frac{Z^{(2)}}{Z^{(3)}}, \ldots, (m-1)\log \frac{Z^{(m-1)}}{Z^{(m)}}\right) \stackrel{d}{=} \left(E_1^*, E_2^*, \ldots, E_{m-1}^*\right),$$

which yields Lemma 5.1. $\square$

LEMMA 5.2. Let $Z^{(1)} \geq Z^{(2)} \geq \cdots \geq Z^{(m)}$ denote the order statistics of the i.i.d. Pareto(1) distributed random variables $\{Z_1, Z_2, \ldots, Z_m\}$. Suppose **Condition A** holds. Let d be a fixed integer or an intermediate sequence $d = d(m) \to \infty, d/m \to 0$ as $n \to \infty$. Then for any $t_0 > 1$,

$$\lim_{n\to\infty} \mathbb{P}^k\Big(Z^{(d+1)} \geq t_0\Big) = 1.$$

Proof of Lemma 5.2. Let $U^{(1)} \geq U^{(2)} \geq \cdots \geq U^{(m)}$ denote the order statistics of i.i.d. uniformly distributed random variables $\{U_1, U_2, \ldots, U_m\}$. Then as $n \to \infty$,

$$\begin{aligned}\mathbb{P}\Big(Z^{(d+1)} \geq t_0\Big) &= \mathbb{P}\left(\sum_{i=1}^m I_{Z_i \geq t_0} \geq d+1\right) = \mathbb{P}\left(\sum_{i=1}^m I_{U_i \leq \frac{1}{t_0}} \geq d+1\right)\\ &= 1 - \mathbb{P}\left(\sum_{i=1}^m I_{U_i \leq \frac{1}{t_0}} < d+1\right) \to 1.\end{aligned}$$

As a result, as $n \to \infty$,

$$k \log \mathbb{P}\left(Z^{(d+1)} \geq t_0\right) = k \log\left(1 - \mathbb{P}\left(\sum_{i=1}^m I_{U_i \leq \frac{1}{t_0}} < d+1\right)\right) \sim -k\mathbb{P}\left(\sum_{i=1}^m I_{U_i \leq \frac{1}{t_0}} < d+1\right).$$

By Hoeffding's Inequality, it follows that

$$\mathbb{P}\left(\sum_{i=1}^m \left(I_{U_i \leq \frac{1}{t_0}} - \frac{1}{t_0}\right) < d+1-\frac{m}{t_0}\right) \leq 2\exp\left\{-\frac{2}{m}\left(d+1-\frac{m}{t_0}\right)^2\right\}.$$

Combining with $m/\log k \to \infty$ and $d/m \to 0$, we obtain that as $n \to \infty$,

$$k \log \mathbb{P}\left(Z^{(d+1)} \geq t_0\right) \to 0,$$

which yields Lemma 5.2. $\square$

Define

$$\mathscr{I}_{n,d,t_0} = \left\{ Z_j^{(d+1)} \geq t_0, \quad \text{for all} \quad 1 \leq j \leq k \right\}.$$

Let d be a fixed integer or an intermediate sequence $d = d(m) \to \infty, d/m \to 0$ as $n \to \infty$. By Lemma 5.2, we have for any $t_0 > 1$,

$$\lim_{n\to\infty} \mathbb{P}\left(\mathscr{I}_{n,d,t_0}\right) = 1. \tag{5.1}$$

5.2 Proofs for Section 2

Proof of Theorem 2.1. By (1.3) and Corollary 1.2.10 in de Haan and Ferreira (2006), it follows that for any $\varepsilon > 0,\ \delta > 0$, there exists a $t_0 > 0$ such that for $t \geq t_0$ and $x \geq 1$,

$$(1-\varepsilon)x^{\gamma-\delta} \leq \frac{U(tx)}{U(t)} \leq (1-\varepsilon)x^{\gamma+\delta},$$

which implies that

$$\begin{aligned} \log(1-\varepsilon) + (\gamma-\delta)\log x \leq \log U(tx) - \log U(t) \\ \leq \log(1+\varepsilon) + (\gamma+\delta)\log x. \end{aligned} \tag{5.2}$$

Recall that $M_j^{(i)} \overset{d}{=} U(Z_j^{(i)})$ where $Z_j^{(1)} \geq Z_j^{(2)} \geq \cdots \geq Z_j^{(m)}$ are the order statistics of the observations in machine j. We intend to replace t and tx in (5.2) by $Z_j^{(d+1)}$ and $Z_j^{(i)}$ for $i = 1, 2, \ldots, d$. By (5.1), we have for any $t_0 > 1$,

$$\lim_{n\to\infty} \mathbb{P}\left(\mathscr{I}_{n,d,t_0}\right) = 1.$$

On the set $\mathscr{I}_{n,d,t_0}$, applying the upper bound in (5.2) with $x = Z_j^{(i)}/Z_j^{(d+1)}$ and $t = Z_j^{(d+1)}$ for $i = 1, 2, \ldots, d$, it follows that

$$\log U(Z_j^{(i)}) - \log U(Z_j^{(d+1)}) \leq \log(1+\varepsilon) + (\gamma+\delta)\log\frac{Z_j^{(i)}}{Z_j^{(d+1)}}.$$

By taking average across i, we obtain that

$$\hat{\gamma}_j \leq \log(1+\varepsilon) + (\gamma+\delta)\frac{1}{d}\sum_{i=1}^{d}\log\frac{Z_j^{(i)}}{Z_j^{(d+1)}}.$$

By taking average across j, we obtain that

$$\begin{aligned}\hat{\gamma}_{DH} &\leq \log(1+\varepsilon) + (\gamma+\delta)\frac{1}{k}\sum_{j=1}^{k}\frac{1}{d}\sum_{i=1}^{d}\log\frac{Z_j^{(i)}}{Z_j^{(d+1)}}\\ &= \log(1+\varepsilon) + (\gamma+\delta)\frac{1}{k}\sum_{j=1}^{k}\frac{1}{d}\sum_{i=1}^{d} i\log\frac{Z_j^{(i)}}{Z_j^{(i+1)}}.\end{aligned}$$

Similarly, we have the lower bound for $\hat{\gamma}_{DH}$,

$$\hat{\gamma}_{DH} \geq \log(1-\varepsilon) + (\gamma-\delta)\frac{1}{k}\sum_{j=1}^{k}\frac{1}{d}\sum_{i=1}^{d} i\log\frac{Z_j^{(i)}}{Z_j^{(i+1)}}.$$

It now suffices to prove that as $n \to \infty$,

$$\frac{1}{k}\sum_{j=1}^{k}\frac{1}{d}\sum_{i=1}^{d} i\log\frac{Z_j^{(i)}}{Z_j^{(i+1)}} \xrightarrow{P} 1,$$

which follows by Lemma 5.1 and the law of large numbers. $\square$

Proof of Theorem 2.2. By Theorem B.2.18 in de Haan and Ferreira (2006), **Condition B** implies that there exists a function $A_0(t)$ such that $A_0(t) \sim A(t)$ as $t \to \infty$, and that for all $\varepsilon > 0,\ \delta > 0$, there exists a $t_0 > 0$ such that for $tx \geq t_0,\ t \geq t_0$,

$$\left|\frac{\log U(tx) - \log U(t) - \gamma\log x}{A_0(t)} - \frac{x^\rho - 1}{\rho}\right| \leq \varepsilon x^\rho x^{\pm\delta}, \tag{5.3}$$

where we use the notation $x^{\pm\delta} = \max\{x^\delta, x^{-\delta}\}$. In fact, A_0 is well defined for different $\gamma \in \mathbb{R}$ and $\rho \leq 0$. For more details on A_0, see page 48 in de Haan and Ferreira (2006).

Recall that $\lim_{n\to\infty}\mathbb{P}(\mathscr{I}_{n,d,t_0}) = 1$ for any $t_0 > 1$. On the set $\mathscr{I}_{n,d,t_0}$, by applying (5.3) with $t = m/d$ and $x = dZ_j^{(i)}/m$ for $i = 1, 2, \ldots, d+1$, we obtain that

$$\left|\frac{\log U(Z_j^{(i)}) - \log U(m/d) - \gamma\log(dZ_j^{(i)}/m)}{A_0(m/d)} - \frac{(dZ_j^{(i)}/m)^\rho - 1}{\rho}\right| \leq \varepsilon\left(\frac{dZ_j^{(i)}}{m}\right)^{\rho\pm\delta}. \tag{5.4}$$

By applying the inequality (5.4) twice for a general i and $i = d+1$, we get that as $n \to \infty$,

$$
\begin{aligned}
&\frac{\log U(Z_j^{(i)}) - \log U(Z_j^{(d+1)}) - \gamma(\log Z_j^{(i)} - \log Z_j^{(d+1)})}{A_0(m/d)} \\
&\quad = \frac{(dZ_j^{(i)}/m)^\rho - 1}{\rho} - \frac{(dZ_j^{(d+1)}/m)^\rho - 1}{\rho} + o_P(1)\left(\left(\frac{dZ_j^{(i)}}{m}\right)^{\rho\pm\delta} + \left(\frac{dZ_j^{(d+1)}}{m}\right)^{\rho\pm\delta}\right).
\end{aligned}
\tag{5.5}
$$

Note that $\lim_{n\to\infty} \mathbb{P}\left(\mathscr{I}_{n,d,t_0}\right) = 1$ implies that the $o_P(1)$ term is uniform for all $1 \le j \le k, 1 \le i \le d$ and all $k \in \mathbb{N}$.

We first prove the result for $\rho < 0$. Choose $\delta > 0$ such that $\rho + \delta < 0$. By taking average across i, we obtain that

$$
\begin{aligned}
\hat{\gamma}_j = \gamma &\left[\frac{1}{d}\sum_{i=1}^{d}(\log Z_j^{(i)} - \log Z_j^{(d+1)})\right] \\
&+ A_0(m/d)\frac{1}{\rho}\left(\frac{dZ_j^{(d+1)}}{m}\right)^\rho \left[\frac{1}{d}\sum_{i=1}^{d}\left(\left(\frac{Z_j^{(i)}}{Z_j^{(d+1)}}\right)^\rho - 1\right)\right] \\
&+ o_P(1)A_0(m/d)\left[\frac{1}{d}\sum_{i=1}^{d}\left(\left(\frac{dZ_j^{(i)}}{m}\right)^{\rho\pm\delta} + \left(\frac{dZ_j^{(d+1)}}{m}\right)^{\rho\pm\delta}\right)\right].
\end{aligned}
$$

By taking average across j and applying the inequality $\frac{x^{\rho\pm\delta}}{y^{\rho\pm\delta}} \le \left(\frac{x}{y}\right)^{\rho\pm\delta}$ for any $x, y > 0$, we obtain that

$$
\begin{aligned}
\sqrt{kd}(\hat{\gamma}_{DH} - \gamma) = \gamma\sqrt{kd} &\left(\frac{1}{kd}\sum_{j=1}^{k}\sum_{i=1}^{d}\log\frac{Z_j^{(i)}}{Z_j^{(d+1)}} - 1\right) \\
&+ \sqrt{kd}\frac{A_0(m/d)}{\rho}\frac{1}{k}\sum_{j=1}^{k}\left[\left(\frac{dZ_j^{(d+1)}}{m}\right)^\rho \frac{1}{d}\sum_{i=1}^{d}\left(\left(\frac{Z_j^{(i)}}{Z_j^{(d+1)}}\right)^\rho - 1\right)\right] \\
&+ o_P(1)\sqrt{kd}A_0(m/d)\frac{1}{k}\sum_{j=1}^{k}\left[\left(\frac{dZ_j^{(d+1)}}{m}\right)^{\rho\pm\delta}\frac{1}{d}\sum_{i=1}^{d}\left(\left(\frac{Z_j^{(i)}}{Z_j^{(d+1)}}\right)^{\rho+\delta} + 1\right)\right] \\
&=: I_1 + I_2 + I_3.
\end{aligned}
$$

By Lemma 5.1 and the central limit theorem, we have that as $n \to \infty$,

$$
I_1 \xrightarrow{d} N(0, \gamma^2).
$$

For I_2, write

$$I_2 = \sqrt{kd}\frac{A_0(m/d)}{\rho}\mathbb{E}\left[\left(\frac{dZ_1^{(d+1)}}{m}\right)^\rho\right]\frac{1}{k}\sum_{j=1}^{k}\frac{\left(dZ_j^{(d+1)}/m\right)^\rho}{\mathbb{E}\left[\left(dZ_1^{(d+1)}/m\right)^\rho\right]}\frac{1}{d}\sum_{i=1}^{d}\left(\left(\frac{Z_j^{(i)}}{Z_j^{(d+1)}}\right)^\rho - 1\right).$$

By the weak law of large numbers for triangular array, we have that as $n \to \infty$,

$$\frac{1}{k}\sum_{j=1}^{k}\frac{\left(dZ_j^{(d+1)}/m\right)^\rho}{\mathbb{E}\left[\left(dZ_1^{(d+1)}/m\right)^\rho\right]}\frac{1}{d}\sum_{i=1}^{d}\left(\left(\frac{Z_j^{(i)}}{Z_j^{(d+1)}}\right)^\rho - 1\right) \xrightarrow{P} \mathbb{E}\left[\frac{1}{d}\sum_{i=1}^{d}\left(\left(\frac{Z_j^{(i)}}{Z_j^{(d+1)}}\right)^\rho - 1\right)\right] = \frac{\rho}{1-\rho},$$

where the second equality follows by Lemma 5.1 and direct calculation. Hence, as $n \to \infty$,

$$I_2 = \frac{\sqrt{kd}A_0(m/d)}{1-\rho}\mathbb{E}\left[\left(\frac{dZ_1^{(d+1)}}{m}\right)^\rho\right](1 + o_P(1)).$$

For I_3, by similar arguments as for I_2, as $n \to \infty$,

$$I_3 = o_P(1)\sqrt{kd}A_0(m/d)\mathbb{E}\left[\left(\frac{dZ_1^{(d+1)}}{m}\right)^{\rho\pm\delta}\right].$$

Moreover, we calculate that

$$\begin{aligned}
\frac{1}{1-\rho}\mathbb{E}\left[\left(\frac{dZ_1^{(d+1)}}{m}\right)^\rho\right] &= \frac{1}{1-\rho}\frac{m!}{(m-d-1)!d!}\int_1^\infty\left(1-\frac{1}{z}\right)^{m-d-1}\frac{1}{z^{d+2}}\left(\frac{dz}{m}\right)^\rho dz \\
&= \frac{1}{1-\rho}\left(\frac{m}{d}\right)^{-\rho}\frac{m!}{(m-d-1)!d!}\int_1^\infty\left(1-\frac{1}{z}\right)^{m-d-1}\left(\frac{1}{z}\right)^{d+2-\rho}dz \\
&= \frac{1}{1-\rho}\left(\frac{m}{d}\right)^{-\rho}\frac{\Gamma(m+1)}{\Gamma(m-d)\Gamma(d+1)}\int_0^1(1-t)^{m-d-1}t^{d-\rho}dt \\
&= \frac{1}{1-\rho}\left(\frac{m}{d}\right)^{-\rho}\frac{\Gamma(m+1)\Gamma(d-\rho+1)}{\Gamma(m-\rho+1)\Gamma(d+1)} \\
&= g(d,m,\rho).
\end{aligned}$$

By the Stirling's formula $\Gamma(x) \sim \sqrt{2\pi(x-1)}\left(e^{-1}(x-1)\right)^{x-1}$ as $x \to \infty$, it follows that as $m \to \infty$,

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

which leads to

$$g(d,m,\rho) \to \frac{d^\rho}{1-\rho}\frac{\Gamma(d-\rho+1)}{\Gamma(d+1)}. \tag{5.6}$$

Combining with the assumption that $\sqrt{kd}A(m/d) = O(1)$, we obtain that $I_3 \xrightarrow{P} 0$ as $n \to \infty$. We conclude that as $n \to \infty$,

$$\sqrt{kd}\left(\hat{\gamma} - \gamma - A_0(m/d)g(d,m,\rho)\right) \xrightarrow{d} N(0,\gamma^2).$$

By the assumption that $\sqrt{kd}A(m/d) = O(1)$ and $\dfrac{A_0(m/d)}{A(m/d)} \to 1$ as $n \to \infty$, we can replace A_0 by A and then the statement in Theorem 2.2 follows.

If $\rho = 0$, (5.5) is equivalent to

$$\begin{aligned}
&\frac{\log U(Z_j^{(i)}) - \log U(Z_j^{(d+1)}) - \gamma(\log Z_j^{(i)} - \log Z_j^{(d+1)})}{A_0(m/d)} \\
&\qquad = \log Z_j^{(i)} - \log Z_j^{(d+1)} + o_P(1)\left[\left(\frac{dZ_j^{(i)}}{m}\right)^{\pm\delta} + \left(\frac{dZ_j^{(d+1)}}{m}\right)^{\pm\delta}\right],
\end{aligned}$$

as $n \to \infty$, where the $o_P(1)$ term is uniform for all $1 \le j \le k, 1 \le i \le d$ and all $k \in \mathbb{N}$. By using similar arguments, we obtain that

$$\begin{aligned}
\sqrt{kd}(\hat{\gamma}_{DH} - \gamma) &= \gamma\sqrt{kd}\left(\frac{1}{kd}\sum_{j=1}^{k}\sum_{i=1}^{d}\log\frac{Z_j^{(i)}}{Z_j^{(d+1)}} - 1\right) \\
&\quad + \sqrt{kd}A_0(m/d)\frac{1}{k}\sum_{j=1}^{k}\frac{1}{d}\sum_{i=1}^{d}\log\left(Z_j^{(i)}/Z_j^{(d+1)}\right) \\
&\quad + o_P(1)\sqrt{kd}A_0(m/d)\frac{1}{k}\sum_{j=1}^{k}\left[\left(\frac{dZ_j^{(d+1)}}{m}\right)^{\pm\delta}\frac{1}{d}\sum_{i=1}^{d}\left(\left(\frac{Z_j^{(i)}}{Z_j^{(d+1)}}\right)^{\delta} + 1\right)\right] \\
&=: I_1' + I_2' + I_3'.
\end{aligned}$$

As $n \to \infty$, we can show that

$$\begin{aligned}
&I_1' \xrightarrow{d} N(0,\gamma^2), \\
&I_2' = \sqrt{kd}A_0(m/d)(1 + o_P(1)), \\
&I_3' = o_P(1)\sqrt{kd}A_0(m/d),
\end{aligned}$$

which yields the statement in Theorem 2.2. $\square$

Proof of Corollary 2.1. The sufficiency is obvious. We only prove the necessity by proving a stronger result: if $\lambda \neq 0$ and $\rho \neq 0$, the distributed Hill estimator bears a higher bias than the oracle Hill estimator in the following sense:

$$\frac{d^{\rho}}{1-\rho}\frac{\Gamma(d-\rho+1)}{\Gamma(d+1)} > \frac{1}{1-\rho}. \tag{5.7}$$

Denote $q = [-\rho]$. If $-\rho = q$, then $q > 0$, (5.7) follows from $d^{-q}\Gamma(d+1+q)/\Gamma(d+1) > 1$. If $-\rho > q$, then denote $\eta = -\rho - q \in (0,1)$. By the Gautschi's inequality, we have

$$\frac{\Gamma(x+1)}{\Gamma(x+s)} > x^{1-s} \tag{5.8}$$

for any $x > 0$ and $s \in (0,1)$. We apply (5.8) with $x = d+\eta$ and $s = 1-\eta$ to obtain that

$$\begin{aligned}
d^{\rho}\frac{\Gamma(d-\rho+1)}{\Gamma(d+1)} &= d^{\rho}\frac{\Gamma(d+1+q+\eta)}{\Gamma(d+1)} \\
&\geq d^{\rho}(d+1+\eta)^{q}\frac{\Gamma(d+1+\eta)}{\Gamma(d+1)} \\
&> d^{\rho}(d+1+\eta)^{q}(d+\eta)^{1-(1-\eta)} \\
&> d^{\rho+q+\eta} \\
&= 1.
\end{aligned}$$

$\square$

Proof of Theorem 2.3. The proof follows similar steps as that for Theorem 2.1 and is thus omitted. $\square$

Proof of Theorem 2.4. We only show the proof for $\rho < 0$. The proof for $\rho = 0$ is similar.

Let $c_{s,n} = \sum_{j=1}^{k} I_{d_j = s}$ for $s = 1, 2, \ldots, d_{max}$. Obviously, $\sum_{s=1}^{d_{max}} c_{s,n} = k$. Combining the assumption $d_{max} < \infty$ and (5.1), we have $\mathbb{P}(\mathscr{I}_{n,d_{max},t_0}) \to 1$ for any $t_0 > 1$ as $n \to \infty$. Similar to the proof of Theorem 2.2, we have that for any $\delta > 0$, as $n \to \infty$,

$$\begin{aligned}
\sqrt{k\bar{d}}(\hat{\gamma}_{DH} - \gamma) &= \gamma\sqrt{k\bar{d}}\left(\frac{1}{k}\sum_{j=1}^{k}\frac{1}{d_j}\sum_{i=1}^{d_j}\log\frac{Z_j^{(i)}}{Z_j^{(d_j+1)}} - 1\right) \\
&\quad + \sqrt{k\bar{d}}A_0(m/\bar{d})\left[\frac{1}{\rho}\frac{1}{k}\sum_{j=1}^{k}\left(\frac{\bar{d}}{d_j}\right)^{\rho}\left(\frac{d_j Z_j^{(d_j+1)}}{m}\right)^{\rho}\frac{1}{d_j}\sum_{i=1}^{d_j}\left(\left(\frac{Z_j^{(i)}}{Z_j^{(d_j+1)}}\right)^{\rho} - 1\right)\right] \\
&\quad + o_P(1)\sqrt{k\bar{d}}A_0(m/\bar{d})\frac{1}{k}\sum_{j=1}^{k}\left(\frac{\bar{d}}{d_j}\right)^{\rho\pm\delta}\left(\frac{d_j Z_j^{(d_j+1)}}{m}\right)^{\rho\pm\delta}\frac{1}{d_j}\sum_{i=1}^{d_j}\left(\left(\frac{Z_j^{(i)}}{Z_j^{(d_j+1)}}\right)^{\rho+\delta} + 1\right) \\
&=: I_1 + I_2 + I_3.
\end{aligned}$$

By Lemma 5.1 and the central limit theorem, we have that as $n \to \infty$,

$$I_1 \xrightarrow{d} N\left(0, \gamma^2\right).$$

For I_2, recall that d_j may take values in $\{1, 2, \ldots, d_{max}\}$, write

$$I_2 = \sqrt{k\bar{d}}A_0(m/\bar{d})\left[\sum_{s=1}^{d_{max}}\left(\frac{\bar{d}}{s}\right)^{\rho}\frac{c_{s,n}}{k}\frac{1}{\rho}\frac{1}{c_{s,n}}\sum_{j:d_j=s}\left(\frac{sZ_j^{(s+1)}}{m}\right)^{\rho}\frac{1}{s}\sum_{i=1}^{s}\left(\left(\frac{Z_j^{(i)}}{Z_j^{(s+1)}}\right)^{\rho} - 1\right)\right].$$

Denote $\mathcal{S} = \{s \in \{1, 2, \ldots, d_{max}\} : c_{s,n} \to \infty \quad \text{as} \quad n \to \infty\}$. Since $d_{max} < \infty$ and $\sum_{s=1}^{d_{max}} c_{s,n} = k$, $\mathcal{S}$ is not an empty set. In addition, $\sum_{s\in\mathcal{S}} c_{s,n}/k \to 1$ as $n \to \infty$. For any $s \in \mathcal{S}$, by the proof of Theorem 2.2, we have that as $n \to \infty$,

$$\frac{1}{\rho}\frac{1}{c_{s,n}}\sum_{j:d_j=s}\frac{\left(sZ_j^{(s+1)}/m\right)^{\rho}}{\mathbb{E}\left[\left(sZ_1^{(s+1)}/m\right)^{\rho}\right]}\frac{1}{s}\sum_{i=1}^{s}\left(\left(\frac{Z_j^{(i)}}{Z_j^{(s+1)}}\right)^{\rho} - 1\right) \xrightarrow{P} \frac{1}{1-\rho}.$$

For any $s \in \mathcal{S}^c$, as $n \to \infty$,

$$\frac{1}{\rho}\frac{1}{c_{s,n}}\sum_{j:d_j=s}\frac{\left(sZ_j^{(s+1)}/m\right)^{\rho}}{\mathbb{E}\left[\left(sZ_1^{(s+1)}/m\right)^{\rho}\right]}\frac{1}{s}\sum_{i=1}^{s}\left(\left(\frac{Z_j^{(i)}}{Z_j^{(s+1)}}\right)^{\rho} - 1\right) = O_P(1).$$

Combining the facts that as $n \to \infty$, $c_{s,n}/k \to 0$ for any $s \in \mathcal{S}^c$, $\sqrt{k\bar{d}}A_0(m/\bar{d}) = O(1)$ and

$\mathbb{E}\left[\left(sZ_1^{(s+1)}/m\right)^\rho\right]$ converges to a constant, we obtain that as $n \to \infty$,

$$\sqrt{k\bar{d}}A_0(m/\bar{d})\left[\sum_{s\in\mathcal{S}^c}\left(\frac{\bar{d}}{s}\right)^\rho \frac{c_{s,n}}{k}\frac{1}{\rho}\frac{1}{c_{s,n}}\sum_{j:d_j=s}\left(\frac{sZ_j^{(s+1)}}{m}\right)^\rho \frac{1}{s}\sum_{i=1}^{s}\left(\left(\frac{Z_j^{(i)}}{Z_j^{(s+1)}}\right)^\rho - 1\right)\right] = o_P(1).$$

Together with $\lim_{n\to\infty}\sum_{s\in\mathcal{S}} c_{s,n}/k = 1$, we obtain that as $n \to \infty$,

$$\begin{aligned} I_2 &= \sqrt{k\bar{d}}A_0(m/\bar{d})\frac{1}{1-\rho}\sum_{s\in\mathcal{S}}\left(\frac{\bar{d}}{s}\right)^\rho \frac{c_{s,n}}{k}\mathbb{E}\left[\left(sZ_1^{(s+1)}/m\right)^\rho\right](1+o_P(1)) \\ &= \sqrt{k\bar{d}}A_0(m/\bar{d})\frac{1}{k}\sum_{d_j\in\mathcal{S}}\left(\frac{\bar{d}}{d_j}\right)^\rho g(d_j,m,\rho)(1+o_P(1)). \end{aligned} \tag{5.9}$$

Since $\sqrt{k\bar{d}}A_0(m/\bar{d}) = O(1)$ and $g(d_j, m, \rho)$ converges to a constant as $n \to \infty$, we get that

$$\sqrt{k\bar{d}}A_0(m/\bar{d})\frac{1}{k}\sum_{d_j\in\mathcal{S}^c}\left(\frac{\bar{d}}{d_j}\right)^\rho g(d_j,m,\rho) = o(1).$$

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

$$I_2 = \sqrt{k\bar{d}}A_0(m/\bar{d})\frac{1}{k}\sum_{j=1}^{k}\left(\frac{\bar{d}}{d_j}\right)^\rho g(d_j,m,\rho)(1+o_P(1))$$

Lastly, I_3 can be handled in a similar way as that in the proof of Theorem 2.3. By the assumption that $\sqrt{kd}A(m/d) = O(1)$ and $\dfrac{A_0(m/d)}{A(m/d)} \to 1$ as $n \to \infty$, we can replace A_0 by A and then the statement in Theorem 2.4 follows. □

Proof of Theorem 2.5. The proof follows similar steps as that for Theorem 2.1 and is thus omitted. □

Proof of Theorem 2.6. We only show the proof for $\rho < 0$. The proof for $\rho = 0$ is similar.

In this setting, $d = d(m) \to \infty, d/m \to 0$ as $n \to \infty$. By (5.1), we have $\lim_{n\to\infty}\mathbb{P}(\mathscr{I}_{n,d,t_0}) = 1$ for any $t_0 > 1$. Then by applying (5.2) with $t = m/d$ and $x = dZ_j^{(i)}/m, i = 1, 2, \ldots, d+1$ and using similar arguments as in the proof of Theorem 2.2, we obtain that as $n \to \infty$,

$$
\begin{aligned}
\sqrt{kd}(\hat{\gamma}_{DH} - \gamma) &= \gamma\sqrt{kd}\left(\frac{1}{kd}\sum_{j=1}^{k}\sum_{i=1}^{d}\log\frac{Z_j^{(i)}}{Z_j^{(d+1)}} - 1\right) \\
&\quad + \sqrt{kd}\frac{A_0(m/d)}{\rho}\frac{1}{k}\sum_{j=1}^{k}\left[\left(\frac{dZ_j^{(d+1)}}{m}\right)^{\rho}\frac{1}{d}\sum_{i=1}^{d}\left(\left(\frac{Z_j^{(i)}}{Z_j^{(d+1)}}\right)^{\rho} - 1\right)\right] \\
&\quad + o_P(1)\sqrt{kd}A_0(m/d)\frac{1}{k}\sum_{j=1}^{k}\left[\left(\frac{dZ_j^{(d+1)}}{m}\right)^{\rho\pm\delta}\frac{1}{d}\sum_{i=1}^{d}\left(\left(\frac{Z_j^{(i)}}{Z_j^{(d+1)}}\right)^{\rho+\delta} + 1\right)\right] \\
&=: I_1 + I_2 + I_3.
\end{aligned}
$$

By Lemma 5.1 and the central limit theorem, we have that as $n \to \infty$,

$$
I_1 \xrightarrow{d} N(0, \gamma^2).
$$

The law of large numbers for triangular array implies that as $n \to \infty$,

$$
\begin{aligned}
I_2 &= \sqrt{kd}\frac{A_0(m/d)}{\rho}\mathbb{E}\left[\left(\frac{dZ_1^{(d+1)}}{m}\right)^{\rho}\right]\frac{1}{k}\sum_{j=1}^{k}\left[\frac{\left(dZ_j^{(d+1)}/m\right)^{\rho}}{\mathbb{E}\left[\left(dZ_1^{(d+1)}/m\right)^{\rho}\right]}\frac{1}{d}\sum_{i=1}^{d}\left(\left(\frac{Z_j^{(i)}}{Z_j^{(d+1)}}\right)^{\rho} - 1\right)\right] \\
&= \sqrt{kd}\frac{A_0(m/d)}{\rho}\mathbb{E}\left[\left(\frac{dZ_1^{(d+1)}}{m}\right)^{\rho}\right]\frac{\rho}{1-\rho}(1 + o_P(1)) \\
&= \sqrt{kd}A_0(m/d)\frac{1}{1-\rho}\mathbb{E}\left[\left(\frac{dZ_1^{(d+1)}}{m}\right)^{\rho}\right](1 + o_P(1))
\end{aligned}
$$

In this setting $d \to \infty$ as $n \to \infty$. By the Stirling's formula, it follows that as $n \to \infty$,

$$
\begin{aligned}
\mathbb{E}\left[\left(\frac{dZ_1^{(d+1)}}{m}\right)^{\rho}\right] &= \left(\frac{m}{d}\right)^{-\rho}\frac{\Gamma(m+1)\Gamma(d-\rho+1)}{\Gamma(m-\rho+1)\Gamma(d+1)} \\
&\sim \left(\frac{m}{d}\right)^{-\rho}\frac{\sqrt{m(d-\rho)}}{\sqrt{d(m-\rho)}}\frac{m^m(d-\rho)^{d-\rho}}{d^d(m-\rho)^{m-\rho}} \\
&\sim \left(\frac{m}{m-\rho}\right)^{m-\rho}\left(\frac{d-\rho}{d}\right)^{d} \\
&= \left(1 + \frac{\rho}{m-\rho}\right)^{m-\rho}\left(1 + \frac{-\rho}{d}\right)^{d} \\
&\sim 1.
\end{aligned}
$$

For I_3, Lemma 5.1 implies that $Z_j^{(d+1)}$ is independent with $Z_j^{(i)}/Z_j^{(d+1)}, i = 1, 2, \ldots, d$. Choose δ such that $\rho + \delta < 0$. It follows that as $n \to \infty$,

$$
\begin{aligned}
I_3 &= o_P(1)\sqrt{kd}A_0(m/d)\mathbb{E}\left[\left(\frac{dZ_1^{(d+1)}}{m}\right)^{\rho\pm\delta}\right] \\
&= o_P(1)\sqrt{kd}A_0(m/d)\mathbb{E}\left[\left(\frac{dZ_1^{(d+1)}}{m}\right)^{\rho+\delta} + \left(\frac{dZ_1^{(d+1)}}{m}\right)^{\rho-\delta}\right]
\end{aligned}
$$

Combining with the assumption that $\sqrt{kd}A(m/d) = O(1)$, we obtain that $I_3 \xrightarrow{P} 0$ as $n \to \infty$. By the assumption that $\sqrt{kd}A(m/d) = O(1)$ and $\dfrac{A_0(m/d)}{A(m/d)} \to 1$ as $n \to \infty$, we can replace A_0 by A and then the statement in Theorem 2.6 follows. $\square$

5.3 Proofs for Section 3

Proof of Theorem 3.1. By (3.2), we have that for any $\varepsilon > 0$, there exists a t_1 such that for any $t \geq t_1$,

$$
\sup_{k\in\mathbb{N}} \max_{1\leq j\leq k} \left| \frac{U_{k,j}(t)}{U(t)} - c_{k,j}^{\gamma} \right| < \varepsilon. \tag{5.10}
$$

Since $U \in RV(\gamma)$, we have that for any $\varepsilon > 0, \delta > 0$, there exists a $t_2 > 0$ such that for $t \geq t_2$ and $x \geq 1$,

$$
(1-\varepsilon)x^{\gamma-\delta} \leq \frac{U(tx)}{U(t)} \leq (1+\varepsilon)x^{\gamma+\delta}. \tag{5.11}
$$

Combining (5.10) and (5.11), denote $t_0 = \max\{t_1, t_2\}$, we have that for any $t \geq t_0$ and $x \geq 1$,

$$
\frac{c_{k,j}^{\gamma} - \varepsilon}{c_{k,j}^{\gamma} + \varepsilon}(1-\varepsilon)x^{\gamma-\delta} \leq \frac{U_{k,j}(tx)}{U_{k,j}(t)} \leq \frac{c_{k,j}^{\gamma} + \varepsilon}{c_{k,j}^{\gamma} - \varepsilon}(1+\varepsilon)x^{\gamma+\delta},
$$

for all $1 \leq j \leq k$. The rest of proof follows similar steps as that for Theorem 2.1 and is thus omitted. $\square$

Proof of Theorem 3.2. By **Condition D**,we have that as $t \to \infty$,

$$
U_{k,j}(t) = U(t)\left(c_{k,j}^{\gamma} + O(A_1(t))\right),
$$

uniformly for all $1 \leq j \leq k$ and all $k \in \mathbb{N}$. Since $c_{k,j}$ are uniformly bounded for all $1 \leq j \leq k$ and all $k \in \mathbb{N}$, we get that

$$\begin{aligned}\log U_{k,j}(t) &= \log U(t) + \log\left(c_{k,j}^{\gamma} + A_1(t)O(1)\right)\\ &= \log U(t) + \gamma \log c_{k,j} + A_1(t)O(1),\end{aligned}$$

where the $O(1)$ term is uniform for all $1 \le j \le k$ and all $k \in \mathbb{N}$ as $t \to \infty$. This implies that there exits a $M > 0$ and a $t_1 > 0$ such that for all $tx \ge t_1$,

$$\left|\frac{\log U_{k,j}(tx) - \log U(tx) - \gamma \log(c_{k,j})}{A_1(tx)}\right| \le M, \tag{5.12}$$

for all $1 \le j \le k$. Since $A_1 \in RV(\tilde{\rho})$, for any $\varepsilon > 0, \delta > 0$, there exits a $t_2 \ge 0$ such that for $t \ge t_2, tx \ge t_2$,

$$\left|\frac{A_1(tx)}{A_1(t)} - x^{\tilde{\rho}}\right| \le \varepsilon x^{\tilde{\rho} \pm \delta} \tag{5.13}$$

By Theorem B.2.18 in de Haan and Ferreira (2006), **Condition B** implies that there exists a function $A_0(t)$ such that $A_0(t) \sim A(t)$ as $t \to \infty$, and that for all $\varepsilon > 0,\ \delta > 0$, there exists a $t_3 > 0$ such that for $tx \ge t_3,\ t \ge t_3$,

$$\left|\frac{\log U(tx) - \log U(t) - \gamma \log x}{A_0(t)} - \frac{x^{\rho} - 1}{\rho}\right| \le \varepsilon x^{\rho} x^{\pm\delta}. \tag{5.14}$$

Without loss of generality, we assume that $A_0(t)$ is eventually positive. Combining (5.12), (5.13) and (5.14), denote $t_0 = \max\{t_1, t_2, t_3\}$, we have that for all $tx \ge t_0, t \ge t_0$,

$$\begin{aligned}M(x^{\tilde{\rho}} - \varepsilon x^{\tilde{\rho}\pm\delta})|A_1(t)| - \varepsilon A_0(t) x^{\rho\pm\delta} &\le \log U_{k,j}(tx) - \log U(t) - \gamma \log x - \gamma \log c_{k,j} - A_0(t)\frac{x^{\rho} - 1}{\rho}\\ &\le M(x^{\tilde{\rho}} + \varepsilon x^{\tilde{\rho}\pm\delta})|A_1(t)| + \varepsilon A_0(t) x^{\rho\pm\delta}.\end{aligned} \tag{5.15}$$

Recall that $\lim_{n\to\infty} \mathbb{P}(\mathscr{I}_{n,d,t_0}) = 1$ for any $t_0 > 1$. On the set $\mathscr{I}_{n,d,t_0}$, we can apply (5.15) with $t = m/d$ and $x = dZ_j^{(i)}/m$ for $i = 1, 2, \ldots, d+1$, and use similar arguments as in the proof of Theorem 2.2. Eventually, the statement in Theorem 3.2 can be proved, provided that as $n \to \infty$,

$$\sqrt{kd}\,|A_1(m/d)|\,\frac{1}{k}\sum_{j=1}^{k}\frac{1}{d}\sum_{i=1}^{d}\left(\left(\frac{dZ_j^{(i)}}{m}\right)^{\tilde{\rho}} + \varepsilon\left(\frac{dZ_j^{(i)}}{m}\right)^{\tilde{\rho}\pm\delta} + \left(\frac{dZ_j^{(d+1)}}{m}\right)^{\tilde{\rho}} + \varepsilon\left(\frac{dZ_j^{(d+1)}}{m}\right)^{\tilde{\rho}\pm\delta}\right) \xrightarrow{P} 0. \tag{5.16}$$

Proving (5.16) follows similar steps as handing I_3 in the proof of Theorem 2.2, and the assumption that $\sqrt{kd}A_1(m/d) \to 0$ as $n \to \infty$.

$\square$

References

Bücher, A. and Segers, J. (2018). Maximum likelihood estimation for the Fréchet distribution based on block maxima extracted from a time series. *Bernoulli*, 24(2):1427–1462.

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

Dombry, C. and Ferreira, A. (2019). Maximum likelihood estimators based on the block maxima method. *Bernoulli*, 25(3):1690–1723.

Einmahl, J. H., de Haan, L., and Zhou, C. (2016). Statistics of heteroscedastic extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(1):31–51.

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

Kleiner, A., Talwalkar, A., Sarkar, P., and Jordan, M. I. (2014). A scalable bootstrap for massive data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(4):795–816.

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.

Rényi, A. (1953). On the theory of order statistics. *Acta Mathematica Hungarica*, 4(3):191–231.

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