Biometrika (2020), **xx**, x, *pp*. 1–8
Printed in Great Britain

Distributed Inference for Extreme Value Index

BY LIUJUN CHEN, DEYUAN LI
Fudan University, China

AND CHEN ZHOU
Erasmus University Rotterdam, The Netherlands

SUMMARY

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.

Some key words: 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. 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. However, due to confidentiality, banks may not share such data among each other. Another example is to analyze insurance claims from various insurance companies. To protect the privacy of their customers, insurance firms cannot share even a small subset of the data, or even a single observation, to externals. Alternatively, banks and insurance firms may share a statistical result conducted based on their own data such that individual operational loss or individual customer data cannot be re-identified from the shared information.

Abstracting from the practical examples, analyzing "big data" stored in distributed machines 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 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, because extreme value analysis are often based on high order statistics. Therefore, it requires further validation for the oracle property of the DC algorithm. This paper provides a first 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 $1 + \gamma x > 0$. In this paper, we restrict our attention to the heavy tailed case, $i.e.$ $\gamma > 0$. Denote $U(t) = (1/(1-F))^{\leftarrow}(t)$ where $\leftarrow$ denotes the left-continuous inverse function. Then the domain of attraction condition simplifies to $U \in RV(\gamma)$, $i.e.$,

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

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, see Chapter 3 in de Haan and Ferreira (2006). After estimating the extreme value index, the relation (1.2) 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. Practically, we cannot apply statistical procedures to the oracle sample, i.e. the hypothetically combined dataset $\{X_1, \ldots, X_n\}$.

For $\gamma > 0$, an efficient estimator for the extreme value index γ is the Hill estimator (Hill (1975)). Consider an intermediate sequence $l = l(n)$ such that $l \to \infty, l/n \to 0$ as $n \to \infty$. The oracle Hill estimator is defined as $\hat{\gamma}_H = l^{-1} \sum_{i=1}^{l} \left(\log M^{(i)} - \log M^{(l+1)}\right)$, 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, top l exceedance ratios $M^{(i)}/M^{(l+1)}$ for $i = 1, 2, \ldots, l$.

The distributed inference setup makes it infeasible to obtain the top order statistics of the oracle sample. Following a DC algorithm, we first apply the Hill estimator at each machine, and then take 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. Define the Hill estimator on the machine j as $\hat{\gamma}_j := {d_j}^{-1} \sum_{i=1}^{d_j} \left(\log M_j^{(i)} - \log M_j^{(d_j+1)}\right)$. This estimator uses the top d_j exceedance ratios. The distributed Hill estimator is then defined as

$$\hat{\gamma}_{DH} := \frac{1}{k}\sum_{i=1}^{k} \hat{\gamma}_j = \frac{1}{k}\sum_{i=1}^{k} \frac{1}{d_j}\sum_{i=1}^{d_j} \left(\log M_j^{(i)} - \log M_j^{(d_j+1)}\right). \tag{1.3}$$

In this paper, we establish the asymptotic theory for the distributed Hill estimator $\hat{\gamma}_{DH}$ and compare it to the oracle Hill estimator $\hat{\gamma}_H$. We particularly focus on two issues. Firstly, the numbers of exceedance ratios used in each machine may be homogeneous or heterogeneous. 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. Continuing with the operational loss example, 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 this 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 handles non-identically distributed observations. Section 4 provides the proof of Theorem 1. Preliminary lemmas used for the proof are gathered in Section B of the supplementary material. The proofs for other results in the paper follow similar steps and can be found in Section C of the supplementary material. In addition, we verify the theoretical finding by an extensive simulation study for finite sample size; see Section A of the supplementary material.

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: the distributed Hill estimator involves more “non-extreme” observations from the oracle sample. If m is too low, it may even fail to be consistent. We further remark that the condition is not too restrictive: 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*

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 (1.2) 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 1. *Suppose $F \in D(G_\gamma)$ with $\gamma > 0$ and Conditions 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

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

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: $\sqrt{kd}(\hat{\gamma}_H - \gamma) \xrightarrow{d} N\left(\lambda/(1-\rho), \gamma^2\right)$, as $n \to \infty$. Under the same condition, we have that $g(d,m,\rho) \to d^\rho(1-\rho)^{-1}\Gamma(d-\rho+1)/\Gamma(d+1)$, as $n \to \infty$, see (4.4) in the proof of Theorem 1. Hence, 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 1. *Under the same conditions as in Theorem 1, 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*

Assume that $\{d_j\}_{j=1}^k, k \in \mathbb{N}$ are uniformly bounded positive integer series, i.e., $d_{max} = \sup_{k\in\mathbb{N}} \max_{1\leq j\leq k} d_j < \infty$. The asymptotic normality of the distributed Hill estimator is presented in the following theorem.

THEOREM 2. *Suppose $F \in D(G_\gamma)$ with $\gamma > 0$ and Conditions 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*

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

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, $\sqrt{k\bar{d}}(\hat{\gamma}_H - \gamma) \xrightarrow{d} N\left(\lambda/(1-\rho), \gamma^2\right)$. 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 $n \to \infty$, $g(d_j, m, \rho) \to d_j^\rho(1-\rho)^{-1}\Gamma(d_j - \rho + 1)/\Gamma(d_j + 1)$ holds uniformly for all $1 \leq j \leq k$. It follows that as $n \to \infty$,

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

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*

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 theorem shows the asymptotic normality of the distributed Hill estimator.

THEOREM 3. *Suppose $F \in D(G_\gamma)$ with $\gamma > 0$ and Conditions 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*

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

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, $\sqrt{kd}(\hat{\gamma}_H - \gamma) \xrightarrow{d} N\left(\lambda/(1-\rho), \gamma^2\right)$. Under the same condition, since as $n \to \infty$, $g(d, m, \rho) \to 1/(1-\rho)$, we get that $\sqrt{kd}(\hat{\gamma}_{DH} - \gamma) \xrightarrow{d} N\left(\lambda/(1-\rho), \gamma^2\right)$. 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. To summarize, the DC algorithm can be applied to the Hill estimator without additional condition, if one uses sizable amount of exceedance ratios in each machine.

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 $\{X_j^1, X_j^2, \dots, X_j^m\}$ as $F_{k,j}$ for $j = 1, 2, \dots, k$. 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 $1 \le j \le k$ and all $k \in \mathbb{N}$ with $c_{k,j}$ uniformly bounded away from 0 and ∞.

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

Define $U_{k,j}(t) = (1/(1 - F_{k,j}))^{\leftarrow}(t)$. 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 $1 \le j \le k$ 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 $\tilde{\rho} \le 0$ and $\lim_{t\to\infty} A_1(t) = 0$ such that as $t \to \infty$,

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

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. By contrast, in the distributed inference setup, we first apply the Hill estimator at each machine and then average the estimates. The theorem below shows the asymptotic normality of the distributed Hill estimator.

THEOREM 4. *Suppose $F \in D(G_\gamma)$ with $\gamma > 0$ and Conditions A-D hold. Let $d_1 = d_2 = \cdots = d_k = d$, where $d \ge 1$ 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*

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

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 $n \to \infty$, $\sqrt{kd}\left(\hat{\gamma}_H - \gamma\right) \xrightarrow{d} N(0, \gamma^2)$. Compared to the result in Theorem 4, 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. PROOF OF THEOREM 1

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

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 $tx \geq t_0,\ t \geq t_0$,

$$\left| \frac{\log U(tx) - \log U(t) - \gamma \log x}{A_0(t)} - \frac{x^\rho - 1}{\rho} \right| \leq \varepsilon x^\rho x^{\pm\delta}, \tag{4.1}$$

where $x^{\pm\delta} = \max\{x^\delta, x^{-\delta}\}$. For more details on A_0, see page 48 in de Haan and Ferreira (2006).

Define $\mathscr{I}_{n,d,t_0} = \left\{ Z_j^{(d+1)} \geq t_0,\ \text{for all } 1 \leq j \leq k \right\}$. Lemma B.2 in the supplementary material implies that for any $t_0 > 1$, $\lim_{n\to\infty} \mathbb{P}\left(\mathscr{I}_{n,d,t_0}\right) = 1$. Consequently, on the set $\mathscr{I}_{n,d,t_0}$, we can replace t and tx in (4.1) by m/d and $Z_j^{(i)}$ for $i = 1, 2, \ldots, d+1$ and obtain that

$$\left| \frac{\log U(Z_j^{(i)}) - \log U(m/d) - \gamma \log(dZ_j^{(i)}/m)}{A_0(m/d)} - \frac{(dZ_j^{(i)}/m)^\rho - 1}{\rho} \right| \leq \varepsilon \left(\frac{dZ_j^{(i)}}{m} \right)^{\rho\pm\delta}. \tag{4.2}$$

By applying the inequality (4.2) twice for a general i and $i = d+1$, we get that as $n \to \infty$,

$$\begin{aligned} &\frac{\log U(Z_j^{(i)}) - \log U(Z_j^{(d+1)}) - \gamma(\log Z_j^{(i)} - \log Z_j^{(d+1)})}{A_0(m/d)} \\ &\quad = \frac{(dZ_j^{(i)}/m)^\rho - 1}{\rho} - \frac{(dZ_j^{(d+1)}/m)^\rho - 1}{\rho} + o_P(1) \left(\left(\frac{dZ_j^{(i)}}{m} \right)^{\rho\pm\delta} + \left(\frac{dZ_j^{(d+1)}}{m} \right)^{\rho\pm\delta} \right). \end{aligned} \tag{4.3}$$

Here, the $o_P(1)$ term is uniform for all $1 \leq j \leq k, 1 \leq i \leq d$ and all $k \in \mathbb{N}$.

We only give the proof for $\rho < 0$. The proof for $\rho = 0$ is similar and is given in the supplementary material. Choose $\delta > 0$ such that $\rho + \delta < 0$. By taking average across i and j and applying the inequality $x^{\rho\pm\delta}/y^{\rho\pm\delta} \leq (x/y)^{\rho\pm\delta}$ for any $x, y > 0$, we obtain that

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

By Lemma B.1 in the supplementary material and the central limit theorem, we have that as $n \to \infty$, $I_1 \stackrel{d}{\to} N(0, \gamma^2)$. 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 from Lemma B.1 in the supplementary material 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 $n \to \infty$,

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

By Lemma B.3 in the supplementary material, we have that as $m \to \infty$,

$$\frac{1}{1-\rho}\mathbb{E}\left[\left(\frac{dZ_1^{(d+1)}}{m}\right)^{\rho}\right] = g(d,m,\rho) \to \frac{d^{\rho}}{1-\rho}\frac{\Gamma(d-\rho+1)}{\Gamma(d+1)} \tag{4.4}$$

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 $A_0(m/d)(A(m/d))^{-1} \to 1$ as $n \to \infty$, we can replace A_0 by A and then the statement in Theorem 1 follows.

□

REFERENCES

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

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.

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