Biometrika (2020), **xx**, x, *pp*. 1–17

Printed in Great Britain

Supplementary Material for ”Distributed Inference for Extreme Value Index”

BY LIUJUN CHEN, DEYUAN LI

Fudan University, China

AND CHEN ZHOU

Erasmus University Rotterdam, The Netherlands

A. SIMULATION STUDY

We conduct a simulation study to demonstrate the finite sample performance of the distributed Hill estimator $\hat{\gamma}_{DH}$. 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: $F(x) = e^{-x^{-1}},\ x > 0$.
- Pareto(1) distribution: $F(x) = 1 - x^{-1},\ x > 1$.
- Absolute Cauchy distribution: the probability density function is given as $f(x) = 2\left(\pi(1+x^2)\right)^{-1},\ x > 0$.

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.

A·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 $\{X_1, \ldots, X_n\}$ 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.

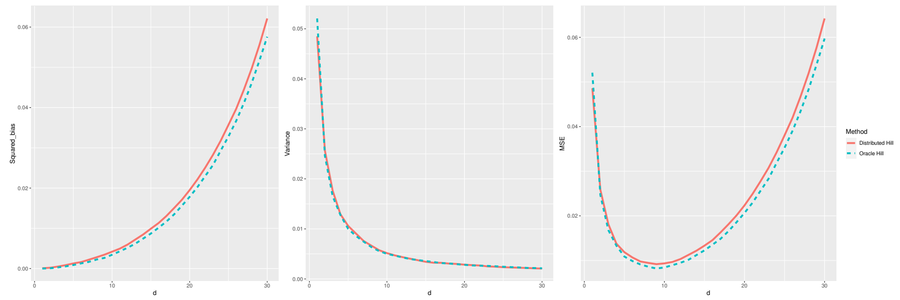
The results are presented in Figure S.1. For the Fréchet(1) distribution and the absolute Cauchy distribution, we observe a trade off between the bias and variance for the both estimators: as d increases, the bias increases while the variance decreases. For the Pareto(1) distribution, the bias is virtually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

Figure S.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 1. 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 3.

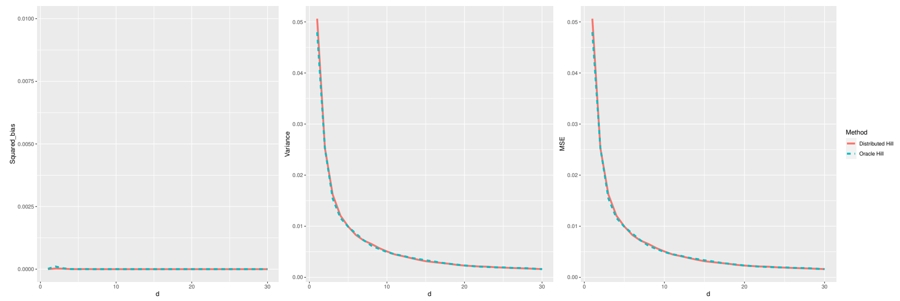
A·2. *Comparison for different level of k*

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 $k_H = 40, \ldots, 400$ 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 S.2.

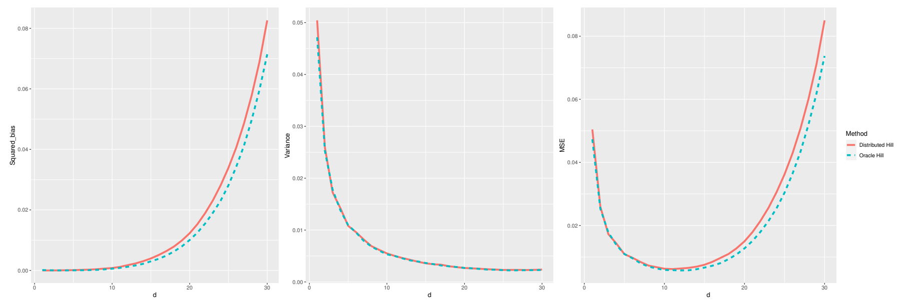
Figure S.2 shows that $\hat{\gamma}_{DH,2}$, $\hat{\gamma}_{DH,8}$ and $\hat{\gamma}_H$ have almost the same variance for all levels of k_H, which is in line with the theoretical result. For the Pareto(1) distribution, $\hat{\gamma}_{DH,2}$ and $\hat{\gamma}_{DH,8}$ behave closely to the oracle Hill 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



(a) Fréchet(1) Distribution



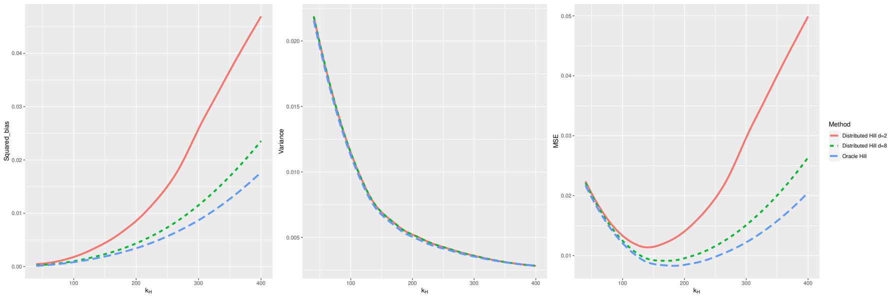
(b) Pareto(1) Distribution



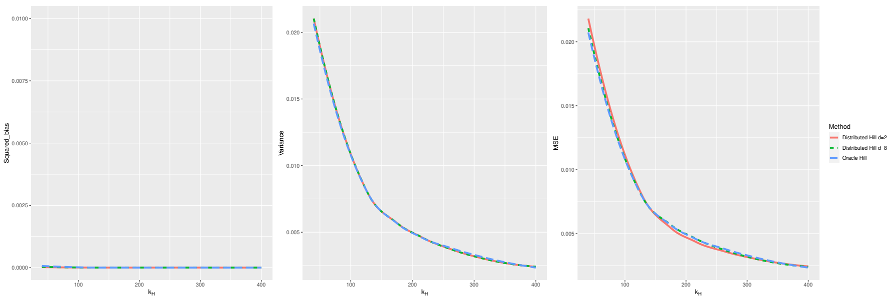
(c) Absolute Cauchy Distribution

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

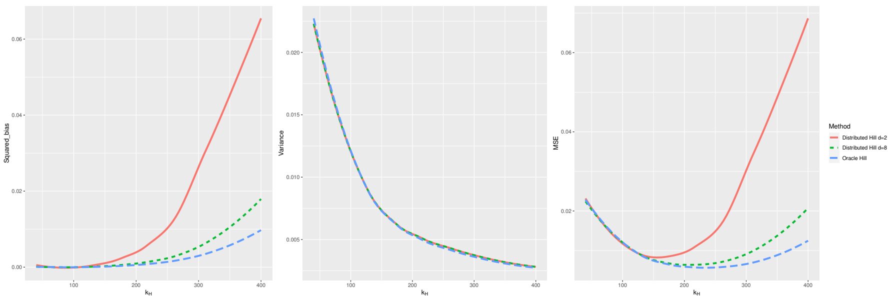
1. As the level of k_H increases, the performance of $\gamma_{DH,8}$ is closer to the performance of the oracle Hill estimator compared to that of $\gamma_{DH,2}$. This is in line with the theoretical result since $d = 2$ can be regarded as fixed integers while $d = 8$ can be regarded as an intermediate sequences as the level of k_H increases.



(a) Fréchet(1) Distribution



(b) Pareto(1) Distribution



(c) Absolute Cauchy Distribution

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

A·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) is equivalent to the convergence in sample maxima:

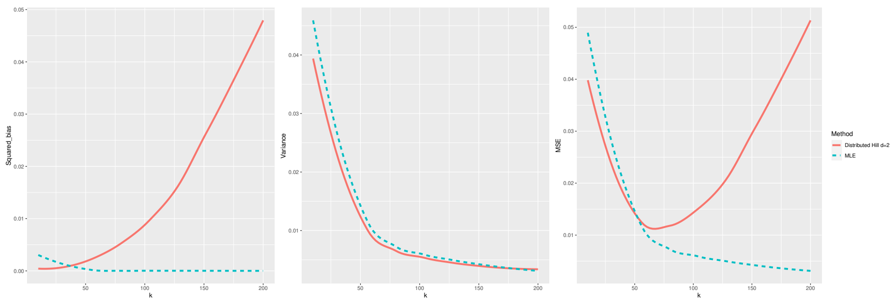
$$\frac{\max\{X_1, \ldots, X_n\} - b_n}{a_n} \to G_\gamma$$

in distribution, where $a_n > 0$ and $b_n \in \mathbb{R}$ are the same series as in (1). Suppose we collect the maxima from each machine. They can be regarded as i.i.d. observations from the distribution of $\max\{X_1, \ldots, X_m\}$. 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 divide-and-conquer 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.

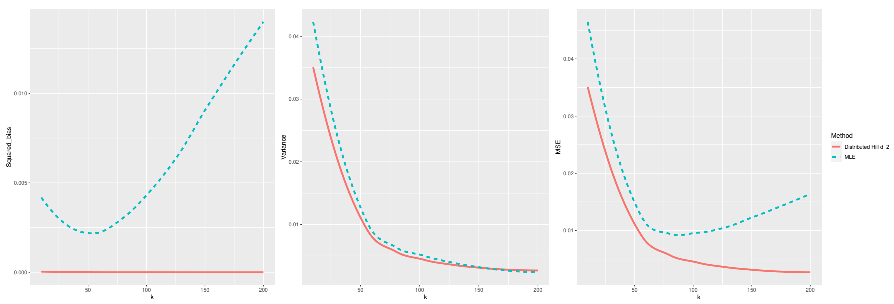
For the calculation of $\hat{\gamma}_M$, we have three technical choices. First, since we restrict our attention to the heavy-tailed case, we fit the block maxima to the generalized Fréchet distribution (with scale and shape parameter) instead of the generalized extreme value distribution; see Bücher and Segers (2018). Secondly, we use the left-truncated block maxima $M_j^{(1)} \vee c,\ j = 1, \ldots, k$, for some small positive constant c, as suggested by Bücher and Segers (2018) . Asymptotically, such left-truncation does not affect the limit behaviour, since $\mathrm{pr}(X_j^{(i)} > c) \to 1$ in probability. 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 S.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

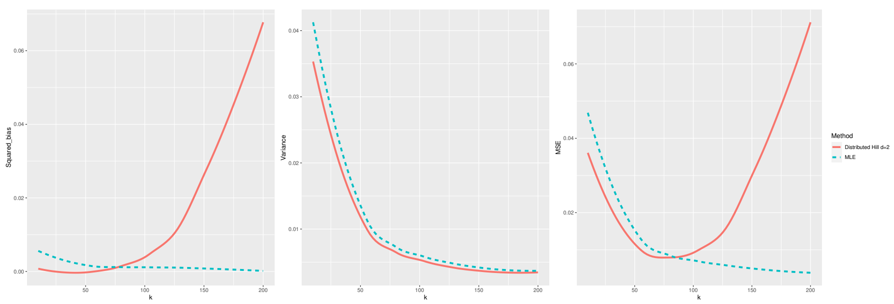
Fréchet(1) and the absolute Cauchy distribution, $\hat{\gamma}_{DH}$ outperforms $\hat{\gamma}_M$ when k is small. As k increases, the bias of $\hat{\gamma}_{DH}$ increases for the Fréchet(1) and the absolute Cauchy distribution. Due to the max-stability of the Fréchet(1) distribution, $\hat{\gamma}_M$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



(a) Fréchet(1) Distribution



(b) Pareto(1) Distribution



(c) Absolute Cauchy Distribution

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

B. Real Data Analysis

We employ a dataset containing car insurance claims in five states of the United States during January and February 2011. [1] We consider five hypothetical insurance companies, one in each state, where each company monopolies all the car insurance in its own state. Due to business privacy, companies cannot share their data to others but they are willing to share their statistical results. To analyze the risk of having extreme claims, the companies, named after the states they operate in, will collaborate in a divide-and-conquer algorithm to estimate the common extreme value index of the total claim amount.

The dataset contains 9134 observations, with 2601, 798, 3150, 1703 and 882 observations for Iowa, Kansas, Missouri, Nebraska and Oklahoma, repsectively. The histograms of the total claim amount based on the combined data and the data for each state are presented in Figure S.4. The histograms across different states are similar to each other. In addition, we observe that the distribution of total claim amount is heavy tailed.

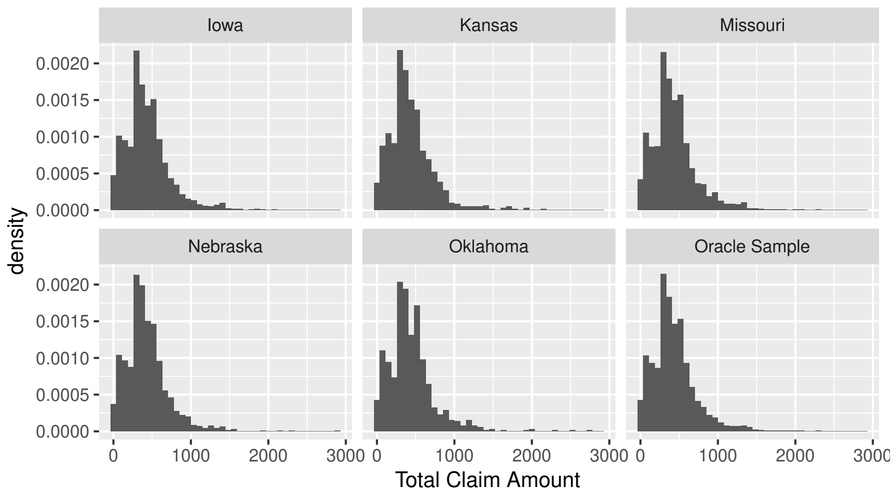


Fig. S.4: Histogram of the total claim amount for each state.

[1] The dataset is available at

`http://dyzz9obi78pm5.cloudfront.net/app/image/id/560ec66d32131c9409f2ba54/n/Auto_Insurance_Claims_Sample.csv`

To ensure that data stored at the five companies are comparable, we sample 700 observations from each state as the hypothetical data possessed by each company. To validate our assumption that different datasets possessed by different companies share the same extreme value index, we plot the estimated extreme value indices for each company against various levels of d using the Hill estimator in Figure S.5. The estimated extreme value indices are close across the five companies.

Next, we average the estimates obtained from the five companies to get a divide-and-conquer estimate using the distributed Hill estimator. The result is plotted in Figure S.6, the solid black line. To compare the distributed Hill estimator with the oracle Hill estimator, we also plot the estimates using the oracle Hill estimator as well as its 95% confidence intervals. We observe that the point estimate based on the distributed Hill estimator is close to that using the oracle Hill estimator for almost all level of d. In addition, the distributed Hill estimates always fall within the 95% confidence intervals.

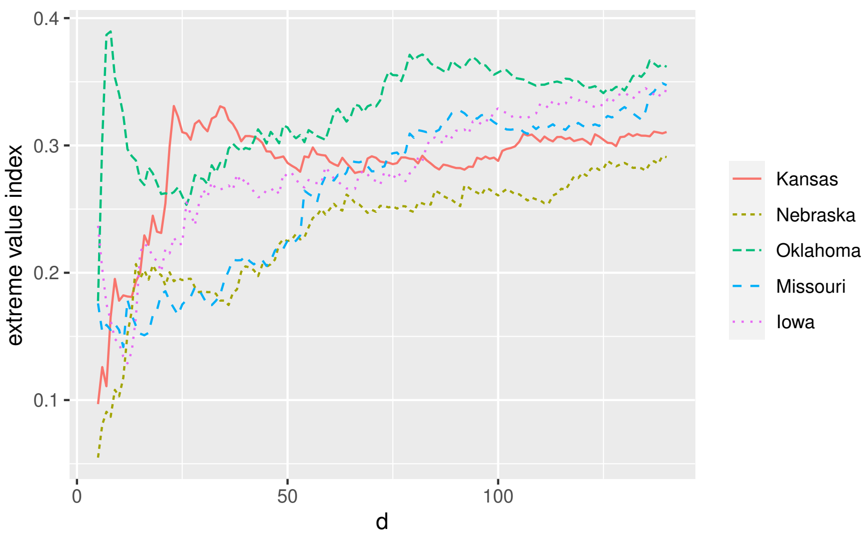


Fig. S.5: Estimation for the extreme value index of total claim amount for each state.

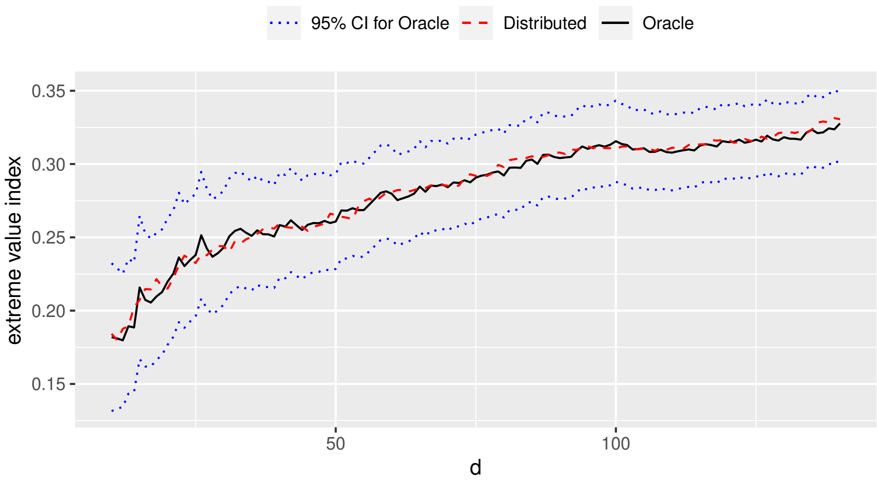


Fig. S.6: The point estimation and the 95% confidence interval of the distributed Hill estimator and the oracle Hill estimator for γ of total claim amount.

C. Preliminary Lemmas

This section contains preliminary lemmas used for the proof.

Lemma S.1. *Let $Z^{(1)} \geq \cdots \geq Z^{(m)}$ denote the order statistics of the i.i.d. Pareto(1) distributed random variables $\{Z_1, \ldots, Z_m\}$. 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 S.1. Note that $\{E_i = \log Z_i, i = 1, 2, \ldots, m\}$ forms a random sample from the standard exponential distribution. Denote $E^{(1)} \geq \cdots \geq E^{(m)}$ as the order statistics of $\{E_1, \ldots, E_m\}$. By Rényi (1953) we have

$$\left\{E^{(m)}, E^{(m-1)}, \ldots, E^{(1)}\right\} = \left\{\frac{E_m^*}{m}, \frac{E_m^*}{m} + \frac{E_{m-1}^*}{m-1}, \ldots, \frac{E_m^*}{m} + \frac{E_{m-1}^*}{m-1} + \cdots + \frac{E_1^*}{1}\right\}$$

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

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

in distribution, which yields Lemma S.1. □

Lemma S.2. *Let $Z^{(1)} \geq \cdots \geq Z^{(m)}$ denote the order statistics of the i.i.d. Pareto(1) distributed random variables $\{Z_1, \ldots, Z_m\}$. 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 $t_0 > 1$, $\lim_{n\to\infty} \left\{pr\left(Z^{(d+1)} \geq t_0\right)\right\}^k = 1$.*

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

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

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

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

By Hoeffding's inequality, it follows that

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

Combining with $m/\log k \to \infty$ and $d/m \to 0$, we obtain that as $n \to \infty$, $k \log \mathrm{pr}\left(Z^{(d+1)} \geq t_0\right) \to 0$, which yields Lemma S.2. $\square$

LEMMA S.3. *Let $Z^{(1)} \geq \cdots \geq Z^{(m)}$ denote the order statistics of the i.i.d. Pareto(1) distributed random variables $\{Z_1, \ldots, Z_m\}$. Then for any $\rho \leq 0$,*

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

Furthermore, if d is a fixed integer, then as $m \to \infty$,

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

Proof Lemma S.3.

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

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

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

which leads to

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

D. PROOFS OF ASYMPTOTIC NORMALITY

This section contains proofs of Theorems 1-4 and Corollary 1.

Proof of Theorem 1 when $\rho = 0$. If $\rho = 0$, (9) is equivalent to

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

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

$$\begin{aligned}
(kd)^{1/2}(\hat{\gamma}_{DH} - \gamma) &= \gamma(kd)^{1/2}\left\{\frac{1}{kd}\sum_{j=1}^{k}\sum_{i=1}^{d}\log\frac{Z_j^{(i)}}{Z_j^{(d+1)}} - 1\right\} \\
&+ (kd)^{1/2}A_0(m/d)\frac{1}{k}\sum_{j=1}^{k}\frac{1}{d}\sum_{i=1}^{d}\log\left(Z_j^{(i)}/Z_j^{(d+1)}\right) \\
&+ o_P(1)(kd)^{1/2}A_0(m/d)\frac{1}{k}\sum_{j=1}^{k}\left[\left(\frac{dZ_j^{(d+1)}}{m}\right)^{\pm\delta}\frac{1}{d}\sum_{i=1}^{d}\left\{\left(\frac{Z_j^{(i)}}{Z_j^{(d+1)}}\right)^{\delta} + 1\right\}\right] \\
&=: I_1 + I_2 + I_3.
\end{aligned}$$

As $n \to \infty$, we can show that $I_1 \to N(0, \gamma^2)$ in distribution, $I_2 = (kd)^{1/2}A_0(m/d)\{1 + o_P(1)\}$, and $I_3 = o_P(1)(kd)^{1/2}A_0(m/d)$, which yields the statement in Theorem 1. □

Proof of Corollary 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:

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

Denote $q = [-\rho]$. If $-\rho = q$, then $q > 0$, (S.1) 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

$$\frac{\Gamma(x+1)}{\Gamma(x+s)} > x^{1-s} \tag{S.2}$$

for any $x>0$ and $s\in(0,1)$. We apply (S.2) with $x=d+\eta$ and $s=1-\eta$ to obtain that

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

Proof of Theorem 2. We only show the proof for $\rho<0$. The proof for $\rho=0$ is similar.

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

$$
\begin{aligned}
&\left(k\bar{d}\right)^{1/2}\left(\hat{\gamma}_{DH}-\gamma\right) \\
&\quad=\gamma\left(k\bar{d}\right)^{1/2}\left(\frac{1}{k}\sum_{j=1}^{k}\frac{1}{d_j}\sum_{i=1}^{d_j}\log\frac{Z_j^{(i)}}{Z_j^{(d_j+1)}}-1\right) \\
&\qquad+\left(k\bar{d}\right)^{1/2}A_0(m/\bar{d})\frac{1}{\rho}\frac{1}{k}\sum_{j=1}^{k}\left(\frac{\bar{d}}{d_j}\right)^{\rho}\left(\frac{d_jZ_j^{(d_j+1)}}{m}\right)^{\rho}\frac{1}{d_j}\sum_{i=1}^{d_j}\left\{\left(\frac{Z_j^{(i)}}{Z_j^{(d_j+1)}}\right)^{\rho}-1\right\} \\
&\qquad+o_P(1)\left(k\bar{d}\right)^{1/2}A_0(m/\bar{d})\frac{1}{k}\sum_{j=1}^{k}\left(\frac{\bar{d}}{d_j}\right)^{\rho\pm\delta}\left(\frac{d_jZ_j^{(d_j+1)}}{m}\right)^{\rho\pm\delta}\frac{1}{d_j}\sum_{i=1}^{d_j}\left\{\left(\frac{Z_j^{(i)}}{Z_j^{(d_j+1)}}\right)^{\rho+\delta}+1\right\} \\
&\quad=:I_1+I_2+I_3.
\end{aligned}
$$

By Lemma S.1 and the central limit theorem, we have that as $n\to\infty$, $I_1\to N\left(0,\gamma^2\right)$ in distribution. For I_2, recall that d_j may take values in $\{1,\ldots,d_{max}\}$, write

$$
I_2=\left(k\bar{d}\right)^{1/2}A_0(m/\bar{d})\left[\sum_{s=1}^{d_{max}}\left(\frac{\bar{d}}{s}\right)^{\rho}\frac{c_{s,n}}{k}\frac{1}{\rho}\frac{1}{c_{s,n}}\sum_{j:d_j=s}\left(\frac{sZ_j^{(s+1)}}{m}\right)^{\rho}\frac{1}{s}\sum_{i=1}^{s}\left\{\left(\frac{Z_j^{(i)}}{Z_j^{(s+1)}}\right)^{\rho}-1\right\}\right].
$$

Denote $\mathcal{S}=\{s\in\{1,\ldots,d_{max}\}:c_{s,n}\to\infty \text{ as } n\to\infty\}$. 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 1, we have that as $n\to\infty$,

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

in probability. 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 $s \in \mathcal{S}^c$, $\left(k\bar{d}\right)^{1/2} A_0(m/\bar{d}) = O(1)$ and $\mathbb{E}\left\{\left(sZ_1^{(s+1)}/m\right)^{\rho}\right\}$ converges to a constant, we obtain that as $n \to \infty$,

$$\left(k\bar{d}\right)^{1/2} A_0(m/\bar{d})\left[\sum_{s\in\mathcal{S}^c}\left(\frac{\bar{d}}{s}\right)^{\rho}\frac{c_{s,n}}{k}\frac{1}{\rho}\frac{1}{c_{s,n}}\sum_{j:d_j=s}\left(\frac{sZ_j^{(s+1)}}{m}\right)^{\rho}\frac{1}{s}\sum_{i=1}^{s}\left\{\left(\frac{Z_j^{(i)}}{Z_j^{(s+1)}}\right)^{\rho}-1\right\}\right]=o_P(1).$$

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

$$\begin{aligned}
I_2 &= \left(k\bar{d}\right)^{1/2} A_0(m/\bar{d})\frac{1}{1-\rho}\sum_{s\in\mathcal{S}}\left(\frac{\bar{d}}{s}\right)^{\rho}\frac{c_{s,n}}{k}\mathbb{E}\left\{\left(sZ_1^{(s+1)}/m\right)^{\rho}\right\}\{1+o_P(1)\}\\
&= \left(k\bar{d}\right)^{1/2} A_0(m/\bar{d})\frac{1}{k}\sum_{d_j\in\mathcal{S}}\left(\frac{\bar{d}}{d_j}\right)^{\rho} g(d_j,m,\rho)\{1+o_P(1)\}.
\end{aligned}$$

Since $\left(k\bar{d}\right)^{1/2} A_0(m/\bar{d}) = O(1)$ and $g(d_j, m, \rho)$ converges to a constant as $n \to \infty$, we get that

$$\left(k\bar{d}\right)^{1/2} A_0(m/\bar{d})\frac{1}{k}\sum_{d_j\in\mathcal{S}^c}\left(\frac{\bar{d}}{d_j}\right)^{\rho} g(d_j,m,\rho) = o(1).$$

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

$$I_2 = \left(k\bar{d}\right)^{1/2} A_0(m/\bar{d})\frac{1}{k}\sum_{j=1}^{k}\left(\frac{\bar{d}}{d_j}\right)^{\rho} g(d_j,m,\rho)\{1+o_P(1)\}.$$

Lastly, I_3 can be handled in a similar way as that in the proof of Theorem 1. By the assumption that $(kd)^{1/2}A(m/d) = O(1)$ 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 2 follows. □

Proof of Theorem 3. 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 Lemma S.2, we have $\lim_{n\to\infty} \mathrm{pr}(\mathscr{I}_{n,d,t_0}) = 1$ for any $t_0 > 1$. Then by applying (7) 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 1, we obtain that as $n \to \infty$,

$$
\begin{aligned}
(kd)^{1/2}(\hat{\gamma}_{DH} - \gamma) = \gamma (kd)^{1/2} & \left(\frac{1}{kd} \sum_{j=1}^{k} \sum_{i=1}^{d} \log \frac{Z_j^{(i)}}{Z_j^{(d+1)}} - 1 \right) \\
& + (kd)^{1/2} \frac{A_0(m/d)}{\rho} \frac{1}{k} \sum_{j=1}^{k} \left[\left(\frac{dZ_j^{(d+1)}}{m} \right)^{\rho} \frac{1}{d} \sum_{i=1}^{d} \left\{ \left(\frac{Z_j^{(i)}}{Z_j^{(d+1)}} \right)^{\rho} - 1 \right\} \right] \\
& + o_P(1)(kd)^{1/2} A_0(m/d) \frac{1}{k} \sum_{j=1}^{k} \left[\left(\frac{dZ_j^{(d+1)}}{m} \right)^{\rho \pm \delta} \frac{1}{d} \sum_{i=1}^{d} \left\{ \left(\frac{Z_j^{(i)}}{Z_j^{(d+1)}} \right)^{\rho + \delta} + 1 \right\} \right] \\
& =: I_1 + I_2 + I_3.
\end{aligned}
$$

By Lemma S.1 and the central limit theorem, we have that as $n \to \infty$, $I_1 \to N(0, \gamma^2)$ in distribution. The law of large numbers for triangular array implies that as $n \to \infty$,

$$
\begin{aligned}
I_2 &= (kd)^{1/2} \frac{A_0(m/d)}{\rho} \mathbb{E} \left\{ \left(\frac{dZ_1^{(d+1)}}{m} \right)^{\rho} \right\} \frac{1}{k} \sum_{j=1}^{k} \left[\frac{\left(dZ_j^{(d+1)}/m \right)^{\rho}}{\mathbb{E} \left\{ \left(dZ_1^{(d+1)}/m \right)^{\rho} \right\}} \frac{1}{d} \sum_{i=1}^{d} \left\{ \left(\frac{Z_j^{(i)}}{Z_j^{(d+1)}} \right)^{\rho} - 1 \right\} \right] \\
&= (kd)^{1/2} \frac{A_0(m/d)}{\rho} \mathbb{E} \left\{ \left(\frac{dZ_1^{(d+1)}}{m} \right)^{\rho} \right\} \frac{\rho}{1-\rho} \{1 + o_P(1)\} \\
&= (kd)^{1/2} A_0(m/d) \frac{1}{1-\rho} \mathbb{E} \left\{ \left(\frac{dZ_1^{(d+1)}}{m} \right)^{\rho} \right\} \{1 + o_P(1)\}.
\end{aligned}
$$

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

$$
\begin{aligned}
\mathbb{E} \left\{ \left(\frac{dZ_1^{(d+1)}}{m} \right)^{\rho} \right\} &= \left(\frac{m}{d} \right)^{-\rho} \frac{\Gamma(m+1)\Gamma(d-\rho+1)}{\Gamma(m-\rho+1)\Gamma(d+1)} \\
&\sim \left(\frac{m}{d} \right)^{-\rho} \frac{\{m(d-\rho)\}^{1/2}}{\{d(m-\rho)\}^{1/2}} \frac{m^m (d-\rho)^{d-\rho}}{d^d (m-\rho)^{m-\rho}} \\
&\sim \left(\frac{m}{m-\rho} \right)^{m-\rho} \left(\frac{d-\rho}{d} \right)^{d} \\
&= \left(1 + \frac{\rho}{m-\rho} \right)^{m-\rho} \left(1 + \frac{-\rho}{d} \right)^{d} \sim 1.
\end{aligned}
$$

For I_3, Lemma S.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 $n \to \infty$,

$$
\begin{aligned}
I_3 &= o_P(1)(kd)^{1/2} A_0(m/d) \mathbb{E} \left\{ \left(\frac{dZ_1^{(d+1)}}{m} \right)^{\rho \pm \delta} \right\} \\
&= o_P(1)(kd)^{1/2} A_0(m/d) \mathbb{E} \left\{ \left(\frac{dZ_1^{(d+1)}}{m} \right)^{\rho + \delta} + \left(\frac{dZ_1^{(d+1)}}{m} \right)^{\rho - \delta} \right\}.
\end{aligned}
$$

Combining with the assumption that $(kd)^{1/2}A(m/d) = O(1)$, we obtain that $I_3 \to 0$ in probability as $n \to \infty$. By the assumption that $(kd)^{1/2}A(m/d) = O(1)$ and $A_0(m/d)\left\{A(m/d)\right\}^{-1} \to 1$ as $n \to \infty$, we can replace A_0 by A and then the statement in Theorem 3 follows. $\square$

Proof of Theorem 4. By Condition D, we have that as $t \to \infty$,

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

uniformly for all $1 \le j \le k$ and all $k \in \mathbb{N}$. Since $c_{k,j}$ are uniformly bounded for all $1 \le j \le k$ 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 exist a $M > 0$ and a $t_1 > 0$ such that for all $tx \ge t_1$,

$$\left|\frac{\log U_{k,j}(tx) - \log U(tx) - \gamma \log(c_{k,j})}{A_1(tx)}\right| \le M, \tag{S.3}$$

for all $1 \le j \le k$. Since A_1 is a regular varying function with index $\tilde{\rho}$, for any $\varepsilon > 0, \delta > 0$, there exists a $t_2 \ge 0$ such that for $t \ge t_2, tx \ge t_2$,

$$\left|\frac{A_1(tx)}{A_1(t)} - x^{\tilde{\rho}}\right| \le \varepsilon x^{\tilde{\rho}\pm\delta}. \tag{S.4}$$

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 $tx \ge t_3,\ t \ge t_3$,

$$\left|\frac{\log U(tx) - \log U(t) - \gamma \log x}{A_0(t)} - \frac{x^{\rho} - 1}{\rho}\right| \le \varepsilon x^{\rho} x^{\pm\delta}. \tag{S.5}$$

Without loss of generality, we assume that $A_0(t)$ is eventually positive. Combining (S.3), (S.4) and (S.5), denote $t_0 = \max\{t_1, t_2, t_3\}$, we have that for all $tx \geq t_0, t \geq t_0$,

$$\begin{aligned} & M(x^{\tilde{\rho}} - \varepsilon x^{\tilde{\rho}\pm\delta})|A_1(t)| - \varepsilon A_0(t) x^{\rho\pm\delta} \\ & \leq \log U_{k,j}(tx) - \log U(t) - \gamma \log x - \gamma \log c_{k,j} - A_0(t)\frac{x^\rho - 1}{\rho} \\ & \leq M(x^{\tilde{\rho}} + \varepsilon x^{\tilde{\rho}\pm\delta})|A_1(t)| + \varepsilon A_0(t) x^{\rho\pm\delta}. \end{aligned} \tag{S.6}$$

Recall that $\lim_{n\to\infty} \mathrm{pr}(\mathscr{I}_{n,d,t_0}) = 1$ for any $t_0 > 1$. On the set $\mathscr{I}_{n,d,t_0}$, we can apply (S.6) 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 1. Eventually, the statement in Theorem 4 can be proved, provided that as $n \to \infty$,

$$\begin{aligned} & (kd)^{1/2}\, |A_1(m/d)| \\ & \times \frac{1}{k}\sum_{j=1}^{k}\frac{1}{d}\sum_{i=1}^{d}\left\{\left(\frac{dZ_j^{(i)}}{m}\right)^{\tilde{\rho}} + \varepsilon\left(\frac{dZ_j^{(i)}}{m}\right)^{\tilde{\rho}\pm\delta} + \left(\frac{dZ_j^{(d+1)}}{m}\right)^{\tilde{\rho}} + \varepsilon\left(\frac{dZ_j^{(d+1)}}{m}\right)^{\tilde{\rho}\pm\delta}\right\} \to 0 \end{aligned} \tag{S.7}$$

in probability. Proving (S.7) follows similar steps as handing I_3 in the proof of Theorem 1, and the assumption that $(kd)^{1/2}A_1(m/d) \to 0$ 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.

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