2011-03-01

Estimating heavy-tail exponents through max self-similarity

Stilian A Stoev, George Michailidis, Murad S Taqqu. 2011. "Estimating Heavy-Tail Exponents Through Max Self-Similarity." IEEE TRANSACTIONS ON INFORMATION THEORY, Volume 57, Issue 3, pp. 1615 - 1636 (22). https://doi.org/10.1109/TIT.2010.2103751

https://hdl.handle.net/2144/37414

Estimating heavy–tail exponents through max self–similarity

Stilian A. Stoev*

University of Michigan, Ann Arbor

George Michailidis

University of Michigan, Ann Arbor

Murad S. Taqqu

Boston University

Abstract: In this paper, a novel approach to the problem of estimating the heavy–tail exponent $\alpha > 0$ of a distribution is proposed. It is based on the fact that block–maxima of size m of the independent and identically distributed data scale at a rate of $m^{1/\alpha}$. This scaling rate can be captured well by the *max–spectrum* plot of the data that leads to regression based estimators. Consistency and asymptotic normality of these estimators is established under mild conditions on the behavior of the tail of the distribution. The results are obtained by establishing bounds on the rate of convergence of moment–type functionals of heavy–tailed maxima. Such bounds often yield exact rates of convergence and are of independent interest. Practical issues on the automatic selection of tuning parameters for the estimators and corresponding confidence intervals are also addressed. Extensive numerical simulations show that the proposed method proves competitive for both small and large sample sizes and for a large range of tail exponents. The method is shown to be more robust than the classical Hill plot and is illustrated on two data sets of insurance claims and natural gas field sizes.

AMS 2000 subject classifications: Primary 62G32, 62G20, 62G05; secondary 62P30, 62P05.

Keywords and phrases: heavy–tail exponent, max self–similarity, max–spectrum, Hill plot, block–maxima, Fréchet distribution, moments of maxima.

1. Introduction

Heavy–tailed distributions arise in many diverse scientific areas: insurance claims, high–speed network traffic, hydrology, the topological structure of the World Wide Web and of social networks, linguistics, just to name a few (see e.g. Adler et al. (1998), McNeil (1997), Resnick (1997*b*), Faloutsos et al. (1999), Adamic and Huberman (2000, 2002), Zipf (1932, 1949), Tsonis et al. (1997)). Highly optimized physical systems also exhibit heavy–tailed behavior, as

* *Corresponding author:* sstoev@umich.edu, Department of Statistics, University of Michigan, 439 West Hall, 1085 South University, Ann Arbor, MI 48109-1107, U.S.A. phone: (734) 763-3519, fax: (734) 763-4676

discussed in Carlson and Doyle (1999).

A real valued random variable X with cumulative distribution function (c.d.f.) $F(x) = \mathbb{P}\{X \le x\},\ x \in \mathbb{R}$ is said to have (right) heavy tail if,

$$\mathbb{P}\{X > x\} = 1 - F(x) = L(x)x^{-\alpha}, \quad \text{as } x \to \infty \tag{1.1}$$

for some $\alpha > 0$, where $L(x) > 0$ is a slowly varying function. The *tail exponent* $\alpha > 0$ controls the rate of decay of F and hence characterizes its tail behavior. The problem of estimating the tail exponent has attracted a lot of attention in the literature since it poses numerous theoretical, as well as, practical challenges (de Haan et al. (2000) and de Sousa and Michailidis (2004)). Most approaches focus on the scaling behavior of the largest order statistics $X(1;N) \ge X(2;N) \ge \cdots \ge X(N;N)$ obtained from an independent and identically distributed (i.i.d.) sample $X(1), \dots, X(N)$ from F. Typical examples include Hill's estimator (1975), its numerous variations (Kratz and Resnick (1996), Resnick and Stărică (1997)), and the kernel–based estimators of Csörgő et al. (1985) (see also Feuerverger and Hall (1999)). For example, the Hill estimator, which is one of the most widely used estimators in practice, can be written as

$$\widehat{\alpha}_H(k) = \Big(\frac{1}{k}\sum_{i=1}^{k} i(\ln X(i;N) - \ln X(i+1;N))\Big)^{-1} =: \Big(\frac{1}{k}\sum_{i=1}^{k} Y_i\Big)^{-1}, \tag{1.2}$$

where $Y_i := i(\ln X(i;N) - \ln X(i+1;N))$. As shown in Weissman (1978), assumption (1.1) implies that for all fixed k's, the vector $\{Y_i\}_{i=1}^k$ converges in distribution to a vector of independent exponentially distributed variables with mean $1/\alpha$. Therefore, when both N and k are large, the statistic $\widehat{\alpha}_H(k)$ in (1.2) behaves like the sample mean of a sample of independent exponential variables. This suggests that the estimator $\widehat{\alpha}_H(k)$ is consistent (Mason (1982)), and under some additional conditions on the tail behavior of F, asymptotically normal (Hall (1982)). In practice, one relies on plotting $\widehat{\alpha}_H(k)$ as a function of the order statistics k (Hill plot) and then selecting an appropriate value for k (see example in Figure 1). In the case of the Pareto distribution ($F(x) = 1 - (x/\sigma_0)^{-\alpha},\ x \ge \sigma_0,\ \sigma_0 > 0$), the Hill estimator is also a conditional maximum likelihood estimator. However, when deviations from this ideal case occur, it exhibits substantial bias and the resulting plot can be misleading (see examples and discussion in de Haan et al. (2000) and de Sousa and Michailidis (2004) and references therein). These shortcomings were addressed in a series of papers that introduced modifications of the original Hill estimator and the resulting Hill plot. The kernel–type estimators introduced by Csörgő et al. (1985) extend the Hill estimator, by introducing non–uniform weights in (1.2) (see also Groeneboom *et al.* (2003)). Namely, given a non–negative and non–increasing kernel function $K(x),\ x > 0$, one considers

$$\widehat{\alpha}_{K,\lambda,N} := \Big(\frac{1}{N}\sum_{i=1}^{N} K(i/\lambda N)Y_i\Big)^{-1}\int_0^{1/\lambda} K(x)dx, \tag{1.3}$$

for some $\lambda > 0$. The Hill estimator can be recovered as a special choice of the function K. Observe also that the threshold parameter k in (1.2) is no longer present. The choice of the kernel function and the bandwidth parameter $\lambda > 0$, however, remain an important and difficult problem for the kernel estimators, similar to the choice of k for the Hill estimator. One practical disadvantage of kernel–type estimators is that no analogue of the Hill plot exists. Therefore, one cannot readily judge how reliable the resulting numerical estimates are.

Other important and popular estimators include the Pickands estimator (see, Pickands (1975) and Dekkers and de Haan (1989)) and de Haan's moment type estimator (see Dekkers et al. (1989)). Resnick and Stărică (1997) introduced a modified and smoothed version of the Hill plot and showed that it performs better in practice when the data depart from the Pareto model (see also de Haan et al. (2000)). The consistency of estimators based on this alternative Hill plot is also established for dependent data (see, Resnick and Stărică (1995)).

In this study, we propose a novel method for estimating the tail index α. It relies on the concept of *max self-similarity*. We focus on the case when the slowly varying function in (1.1) is asymptotically constant and consider block–wise maxima of i.i.d. random variables $X(1), X(2), \ldots$ with c.d.f. F. Block–maxima of block sizes m, scale at a rate of $m^{1/\alpha}$, as $m \to \infty$. Therefore, we can obtain an estimate of α, by focusing on a sequence of growing, dyadic block sizes $m = 2^j,\ 1 \le j \le \log_2 N,\ j \in \mathbb{N}$, and estimating the mean of logarithms of block–maxima (log–block–maxima). This is achieved by examining the *max–spectrum plot* of the data, defined as means of log–block–maxima as a function of the logarithm of the block–size. The slope of the max–spectrum plot for large block–sizes yields an estimate of $1/\alpha$ (see Figure 1 below).

When the $X(i)$'s come from a Fréchet distribution, then their block–maxima have the same Fréchet distribution, rescaled by $m^{1/\alpha}$, where m denotes the block size. Thus, in practice, the max–spectrum plot is essentially linear (Figure 2). One can view i.i.d. Fréchet sequences as *max self–similar* with self–similarity parameter $1/\alpha$ (Definition 2.1). Due to this exact max self–similarity property, our estimation framework works best for Fréchet data. On the other hand, the Hill–type estimators work best for Pareto data. This also shows the fundamental difference between the two approaches. In many important applications the Hill plot is rather volatile. The max spectrum turns out to be more robust to outliers in the data or to deviations from its corresponding ideal Fréchet model than the Hill plot. In Section 5.3, we examine two data sets: *(i)* $2,167$ insurance claims due to fire losses in Denmark and *(ii)* volumes of natural gas reserves in 406 Oil rich provinces. In both cases, the max self–similarity estimators yield values consistent with previous detailed studies of these data sets (see McNeil (1997) and de Sousa and Michailidis (2004), respectively). These values depart from values that one obtains directly from the Hill plots. In fact, in case *(ii)*, due to the peculiar discrete nature of the data set the Hill plot has a saw tooth shape and it is particularly hard to interpret, whereas the max spectrum plot appears to yield a reliable estimate.

The remainder of the paper is structured as follows. In Section 2, we introduce the max–spectrum plot and the self–similarity estimators of the heavy–tail exponent α and establish their basic properties in the ideal Fréchet setting. Some useful results on rates for moment–type functionals of heavy–tailed maxima are presented in Section 3. These results are used to prove the consistency and asymptotic normality of the max self–similarity estimators in Section 4. In Section 5, the performance of the new estimators is examined through a simulation study. The max self–similarity estimators are then shown to work well in the context of two challenging real data examples where the classical Hill plot is rather volatile and is hard to interpret.

2. Max self–similarity and tail exponent estimators

In this section, we introduce some notation and recall some basic definitions used in the remainder of the paper. We then introduce estimators of the heavy–tail exponents based on max self–similarity and discuss their basic properties in the ideal Fréchet case.

2.1. Definition and basic properties

We focus on the case where the slowly varying function L in (1.1) is trivial, that is, when

$$\mathbb{P}\{X > x\} = 1 - F(x) \sim \sigma_0^\alpha x^{-\alpha}, \quad \text{as } x \to \infty, \tag{2.1}$$

with $\sigma_0 > 0$ and where $\sim$ means that the ratio of the left–hand side (l.h.s.) to the right–hand side (r.h.s.) in (2.1) tends to 1, as $x \to \infty$. For simplicity, we further assume that the $X(i)$'s are almost surely positive ($F(0) = 0$). We address the general case where the $X(i)$'s can take negative values in Section 4 (see, Proposition 4.3).

We begin with some useful definitions: for an i.i.d. sample $X(i),\ i \in \mathbb{N} := \{1, 2, \ldots\}$ from F, consider the sequence of block–maxima

$$X_m(k) := \max_{1\le i\le m} X(m(k-1)+i) \equiv \bigvee_{i=1}^{m} X(m(k-1)+i), \quad k = 1, 2, \ldots,$$

with $m \in \mathbb{N}$, where $X_m(k)$ is the greatest observation in the $k-$th block. The Fisher–Tippett–Gnedenko Theorem (see e.g. Proposition 0.3 in Resnick (1987)) then implies that, as $m \to \infty$, $m^{-1/\alpha}X_m(k)$ converges in distribution to a random variable Z with an $\alpha-$Fréchet distribution. More precisely,

$$\mathbb{P}\{Z \le x\} = \exp\{-\sigma_0^\alpha x^{-\alpha}\}, \quad x > 0, \tag{2.2}$$

where $\sigma_0 > 0$, called the *scale coefficient* of Z, is as in (2.1). In fact, as $m \to \infty$, we have

$$\Big\{\frac{1}{m^{1/\alpha}} X_m(k)\Big\}_{k\in\mathbb{N}} \stackrel{d}{\longrightarrow} \Big\{Z(k)\Big\}_{k\in\mathbb{N}}, \tag{2.3}$$

where the $Z(k)$'s are independent copies of Z and where $\stackrel{d}{\to}$ denotes convergence of the finite–dimensional distributions. Thus, for large values of m, the normalized block–maxima behave like a sequence of i.i.d. $\alpha-$Fréchet variables. In fact, when the $X(k)$'s are $\alpha-$Fréchet, (2.3) holds with equality for all $m \in \mathbb{N}$ (see Relation (7.3) in the Appendix). The sequence of i.i.d. $\alpha-$Fréchet $X(k)$'s is thus *max self–similar* in the sense of the following definition.

Definition 2.1 A sequence of random variables $X = \{X(k)\}_{k\in\mathbb{N}}$ (defined on the same probability space) is said to be max self–similar with self–similarity parameter $H > 0$, if for any $m > 0,\ m \in \mathbb{N}$,

$$\Big\{\bigvee_{i=1}^{m} X(m(k-1)+i)\Big\}_{k\in\mathbb{N}} \stackrel{d}{=} \Big\{m^H X(k)\Big\}_{k\in\mathbb{N}}, \tag{2.4}$$

where $=^d$ denotes equality of the finite–dimensional distributions.

If the $X(k)$'s are i.i.d. but not Fréchet, then Relation (2.3) indicates that (2.4) holds asymptotically, as $m \to \infty$, with $H = 1/\alpha$. Thus, any sequence of i.i.d. heavy–tailed variables can be regarded as *asymptotically max self–similar* with self–similarity parameter $H = 1/\alpha$. This feature suggests that an estimator of H and therefore α can be obtained by focusing on the scaling of the block–maxima of growing block sizes. Crovella and Taqqu (1999) used a similar idea based on the scaling of block–wise sums to estimate a heavy–tail exponent α when $\alpha \in (0, 2)$.

Given an i.i.d. sample $X(1), \dots, X(N)$ from F, we consider

$$D(j,k) := \max_{1\le i\le 2^j} X(2^j(k-1)+i) = \bigvee_{i=1}^{2^j} X(2^j(k-1)+i), \quad k = 1, 2, \dots, N_j, \tag{2.5}$$

for all $j = 1, 2, \dots, [\log_2 N]$, where $N_j := [N/2^j]$ and $[x]$ denotes the largest integer not greater than $x \in \mathbb{R}$. By analogy to the discrete wavelet transform, we refer to the parameter j as the *scale* and to k as the *location* parameter. We consider dyadic block–sizes for algorithmic and computational convenience (for more details, see Stoev et al. (2006)).

Observe that for any fixed j, the block–maxima $D(j, k)$ are independent in k since they involve maxima over non–overlapping blocks of the $X(i)$'s. Moreover, as argued above, when the $X(i)$'s follow an $\alpha-$Fréchet distribution,

$$\{D(j,k)\}_{k\in\mathbb{N}} \stackrel{d}{=} \{2^{j/\alpha} D(0,k)\}_{k\in\mathbb{N}} = \{2^{j/\alpha} X(k)\}_{k\in\mathbb{N}}, \tag{2.6}$$

for any scale $j \in \mathbb{N}$. Introduce the statistics

$$Y_j := \frac{1}{N_j} \sum_{k=1}^{N_j} \log_2 D(j,k), \quad j = 1, 2, \dots, [\log_2(N)] \tag{2.7}$$

and observe that by the Law of Large Numbers, the Y_j's are consistent and unbiased estimators of the expectations $\mathbb{E} \log_2 D(j,1)$, provided that these are finite. (Corollary 3.1 below establishes that $\mathbb{E}|\log_2 D(j,1)|$ are finite under general conditions on the c.d.f. $F(x)$.) In view of the asymptotic max self–similarity (2.3) of X, relationship (2.6) holds approximately for large scales j, and in fact,

$$\mathbb{E}Y_j = \mathbb{E} \log_2 D(j,1) \simeq j/\alpha + C, \tag{2.8}$$

with $C = C(\sigma_0, \alpha) = \mathbb{E} \log_2 \sigma_0 Z$, where Z is an $\alpha-$Fréchet variable with unit coefficient as in (2.2) above. Here $\simeq$ means that the difference between the l.h.s. and the r.h.s. tends to zero.

In practice, one can look at the *max–spectrum plot* of the statistics Y_j's versus j (see Figure 1 below). In view of (2.8) it is expected that for large j's the slope coefficient of a linear fit of the Y_j's against j's would yield an estimate of $H = 1/\alpha$. Further, observe that the log–linear scaling relation in (2.8) becomes more precise, the larger the scale j (block–size 2^j) and holds exactly for all scales $j = 1, \dots, [\log_2(N)]$, when the $X(k)$'s come from an $\alpha-$Fréchet distribution (see (2.6)).

Thus, given a range of scales $1 \le j_1 \le j \le j_2 \le [\log_2(N)]$, we define the following regression–based estimators of $H = 1/\alpha$ and α

$$\widehat{H}_w(j_1, j_2) := \sum_{j=j_1}^{j_2} w_j Y_j, \quad \text{and} \quad \widehat{\alpha}_w(j_1, j_2) := 1/\widehat{H}_w(j_1, j_2), \tag{2.9}$$

where the weights w_j are chosen so that

$$\sum_{j=j_1}^{j_2} w_j = 0 \quad \text{and} \quad \sum_{j=j_1}^{j_2} j w_j = 1. \tag{2.10}$$

It is easy to see that the linear estimators $\widehat{H}_w$ in (2.9) with weights as in (2.10) are least squares estimators in a linear regression model. In the rest of the paper, the estimators $\widehat{H}_w$ and $\widehat{\alpha}_w$ in (2.9) are referred to as *max self–similarity* estimators.

Remark *(Computational complexity)*

The proposed estimators exhibit a significant computational advantage over Hill–type or kernel–based estimators. Given a sample of size N one can compute the max–spectrum $Y_j,\ 1 \le j \le [\log_2 N]$, with Y_j as in (2.7) by using $\mathcal{O}(N)$ operations since $\mathcal{O}(N/2^j)$ pair–wise maxima and sums are computed, for $j = 1, \ldots, [\log_2 N]$, and therefore $\mathcal{O}\Big(\sum_{j=1}^{[\log_2 N]} [N/2^j]\Big) = \mathcal{O}(N)$ operations are done. On the other hand, methods involving order statistics require sorting the sample which results in $\mathcal{O}(N \log_2(N))$ operations.

We now illustrate the nature of the max-spectrum plot and the resulting estimator using an example of Internet topology data. The data describe the degree of connectivity between autonomous systems (AS - networks under a single administrative authority) on the Internet for the year 2002 and is provided by the National Laboratory for Applied Network Research. The information has been used to characterize the topology of the Internet (see, e.g. Faloutsos et al. (1999) and Chen et al. (2002)). The size of the data set is 13,579 and each observation gives the number of connections of an AS to peer AS. The histogram of the data (in log-scale) shows that the vast majority of the AS are connected to very few peer systems, but there are a few AS that are directly connected to over 10% of their peer systems. The max–spectrum indicates a value for the tail index of about 1.5. The Hill estimator for $k = 80$ (where the Hill plot seems to stabilize) suggests a value of 1.43.

2.2. The ideal Fréchet case

We start by assuming that $X(1), \ldots, X(N)$ is an i.i.d. sample of $\alpha-$Fréchet variables with scale coefficient $\sigma_0 > 0$ and study the behavior of $\widehat{H}_w(j_1, j_2)$ in this setting.

Consider the regression problem

$$Y_j = j/\alpha + C + \epsilon_j, \quad j_1 \le j \le j_2 \tag{2.11}$$

where

$$C = C(\sigma_0, \alpha) = \mathbb{E} \log_2(\sigma_0 Z) = \log_2(\sigma_0) + \mathbb{E} \log_2(Z) \tag{2.12}$$

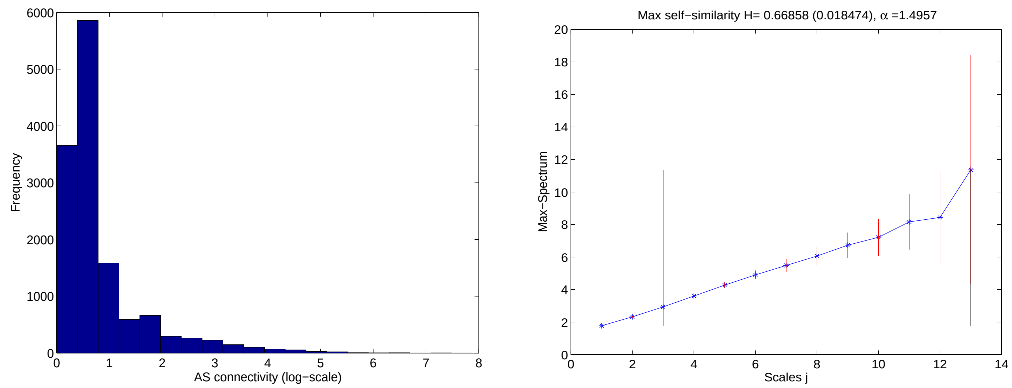

Fig 1. *Left panel: histogram (log–scale) of AS connectivities. Right panel: max–spectrum plot for the AS connectivity data. The large vertical lines indicate the range of* j*'s where a linear fit was used to estimate the heavy–tail index* α*. The shorter vertical lines are* 95% *confidence intervals for the* $\mathbb{E}Y_j$*'s. The reciprocal of the slope yields an estimate of* $\widehat{\alpha}_w(3,13) = 1.4957$*. This range was selected automatically with tunning level* $p = 0.1$*, discussed in Section 5.2.*

for an $\alpha-$Fréchet Z random variable with unit scale coefficient, and where $1 \le j_1 \le j_2 \le [\log_2 N]$. In view of (2.6), we have that the errors ϵ_j have zero means. They are, however, dependent in j due to the corresponding dependence of the Y_j statistics in (2.7). Moreover, the number of $D(j,k)$'s at a scale j in (2.7) is $N_j = [N/2^j]$ and therefore, the variances of the ϵ_j's grow exponentially in j. This implies that the minimal variance unbiased estimators of the parameters of interest $\theta = (H, C)^t$ that are linear in Y_j are obtained through *generalized least squares* (GLS). They are given by

$$\widehat{\theta}_\Sigma = \begin{pmatrix} \widehat{H}_\Sigma \\ \widehat{C}_\Sigma \end{pmatrix} = (A^t \Sigma^{-1} A)^{-1} A^t \Sigma^{-1} Y, \tag{2.13}$$

where $A = (a\, b)$ with $a^t = (j_1, \dots, j_2)$ and $b^t = (1, \dots, 1)$, and $\Sigma = (\mathrm{Cov}(Y_i, Y_j))_{i,j=j_1}^{j_2}$ is the covariance matrix of the vector $Y = \{Y_j\}_{j=j_1}^{j_2}$. An explicit expression of the matrix $\Sigma = \Sigma_\alpha(j_1, j_2; N)$ is given next.

Proposition 2.1 *Let* $Y = \{Y_j\}_{j=j_1}^{j_2}$ *be as in* (2.7)*, where the underlying distribution of the* $X(k)$*'s is* $\alpha-$*Fréchet with scale coefficient* $\sigma_0 > 0$*. Then, for all* $j_1 \le i \le j \le j_2$*,*

$$\mathbb{E}Y_j = j/\alpha + C(\sigma_0, \alpha),$$

and

$$\mathrm{Cov}(Y_i, Y_j) = \Sigma_\alpha(j_1, j_2; N)_{ij} = \frac{2^{j-i}}{\alpha^2 N_i}\psi(|i-j|), \quad N_i = [N/2^i], \tag{2.14}$$

where

$$\psi(a) := \mathrm{Cov}(\log_2(Z_1), \log_2(Z_1 \vee (2^a - 1)Z_2)), \;\; a \ge 0, \tag{2.15}$$

and where Z_1 and Z_2 are independent 1−Fréchet variables with unit scale coefficients.

Proof: Let $j_1 \le i < j \le j_2$ and observe that $N_i = 2^{j-i}N_j + R$, where $0 \le R < 2^{j-i}, \; R \in \mathbb{N}$. In view of (2.7),

$$\begin{aligned} \mathrm{Cov}(Y_i, Y_j) &= \frac{1}{N_i N_j}\sum_{k_1=1}^{N_i}\sum_{k_2=1}^{N_j} \mathrm{Cov}(\log_2 D(i,k_1), \log_2 D(j,k_2)) \\ &= \frac{1}{N_i N_j}\sum_{k_1=1}^{N_j}\sum_{\ell=1}^{2^{j-i}}\sum_{k_2=1}^{N_j} \mathrm{Cov}(\log_2 D(i,(k_1-1)2^{j-i}+\ell), \log_2 D(j,k_2)) \\ &\quad + \frac{1}{N_i N_j}\sum_{\ell=1}^{R}\sum_{k_2=1}^{N_j} \mathrm{Cov}(\log_2 D(i, N_j 2^{j-i}+\ell), \log_2 D(j,k_2)), \end{aligned} \tag{2.16}$$

where the last relation follows from expressing the sum $\sum_{k_1=1}^{N_i}$ as a double sum $\sum_{k_1=1}^{N_j}\sum_{\ell=1}^{2^{j-i}}$ plus the remainder term $\sum_{\ell=1}^{R}\sum_{k_2=1}^{N_j}$. Observe that in view of (2.5), we have that the terms $\mathrm{Cov}(\log_2 D(i,(k_1-1)2^{j-i}+\ell), \log_2 D(j,k_2)), \; 1 \le \ell < 2^{j-i}$ are non–zero only if $k_1 = k_2$ since otherwise the terms $D(i,(k_1-1)2^{j-i}+\ell)$ and $\log_2 D(j,k_2)$ involve maxima of non–overlapping sets of $X(k)$'s. Note moreover that

$$D(j,k_2) = D(i,(k_2-1)2^{j-i}+1) \vee \cdots \vee D(i, k_2 2^{j-i}), \tag{2.17}$$

where the $D(i,k)$'s are i.i.d. $\alpha-$Fréchet variables with scale coefficient $2^{i/\alpha}\sigma_0$ (see (7.3) below). Therefore, for all $k = 1, \ldots, N_j$ and $\ell = 1, \ldots, 2^{j-i}$,

$$(D(i,(k-1)2^{j-i}+\ell), D(j,k)) \stackrel{d}{=} (2^{i/\alpha}Z', 2^{i/\alpha}Z' \vee (2^{j/\alpha} - 2^{i/\alpha})Z''),$$

where Z' and Z'' are independent $\alpha-$Fréchet variables with scale coefficients $\sigma_0 > 0$. Observe that $Z' = \sigma_0 Z_1^{1/\alpha}$, where Z_1 is 1−Fréchet with unit scale coefficient. Hence, for all $k_1 = k_2 =$

$1, \ldots, N_j$ and $\ell = 1, \ldots, 2^{j-i}$, we have

$$\begin{aligned} &\mathrm{Cov}(\log_2 D(i,(k_1-1)2^{j-i}+\ell), \log_2 D(j,k_2)) \\ &\quad = \mathrm{Cov}\Big(\log_2(2^{i/\alpha}\sigma_0 Z_1^{1/\alpha}), \log_2(2^{i/\alpha}\sigma_0 Z_1^{1/\alpha} \vee (2^{j/\alpha}-2^{i/\alpha})\sigma_0 Z_2)\Big) \\ &\quad = \mathrm{Cov}\Big(\log_2(Z_1^{1/\alpha}), \log_2(Z_1^{1/\alpha} \vee (2^{(j-i)/\alpha}-1)Z_2^{1/\alpha})\Big) = \frac{1}{\alpha^2}\psi(|i-j|). \quad (2.18) \end{aligned}$$

The last two relations follow from the facts that $\log_2(2^{i/\alpha}\sigma_0 Z_1^{1/\alpha})$ equals $\log_2(2^{i/\alpha}\sigma_0) + \alpha^{-1}\log_2(Z_1)$ and since $\mathrm{Cov}(\xi + a, \eta + b) = \mathrm{Cov}(\xi, \eta)$, for any constants a and b and random variables ξ and η with finite variance.

Note that the covariances in the remainder term in (2.16) vanish since $D(i, N_j 2^{j-i} + \ell),\ \ell = 1, \ldots, 2^{j-i}$ are independent of $X(i),\ i = 1, \ldots, N_j 2^j$. Thus, by using Relation (2.18), we obtain (2.14). □

Remarks

1. Observe that the covariance matrix Σ does not depend on the scale coefficient σ_0, which is due to the fact that the Y_j's are obtained through a logarithmic transformation of the $X(k)$'s.
2. Observe that for all $1 \le j_1 < j_2 \le [\log_2 N]$ and $\alpha > 0$, we have by (2.14) that

$$\Sigma_\alpha(j_1, j_2; N) = \frac{1}{\alpha^2}\Sigma_1(j_1, j_2; N),$$

where $\Sigma_1(j_1, j_2; N)$ corresponds to the covariance matrix of $Y = \{Y_j\}_{j=j_1}^{j_2}$ from a 1−Fréchet sample.

That is, the unknown parameter α appears only in the factor $1/\alpha^2$ of the covariance matrix and thus the GLS estimators $\widehat{H}_\Sigma$ and $\widehat{C}_\Sigma$ do not depend on α. Indeed, if one multiplies Σ by a factor ϕ, the resulting estimates are not affected, since the formula (2.13) involves the product of ϕ and its inverse.

This invariance property shows that the GLS estimators can be computed *exactly*, without using plug–in approximations for the unknown parameter α involved in the matrix Σ. Table 7.1 in the Appendix contains values of $\psi(i)$ for $i = 0, 1 \ldots, 19$, obtained through

Monte Carlo simulations. This is sufficient to handle sample sizes of up to $2^{20} = 1,048,576$ observations.

3. Finally, $\Sigma_\alpha(j_1, j_2; N)$ is invertible, which follows from the fact that the joint distribution of the Y_j's has a density with respect to the Lebesgue measure.

In view of the above remarks, we have that

Corollary 2.1 *The minimum variance unbiased estimators for H and C in the regression model* (2.11)*, linear in Y_j, are given by* (2.13)*. Moreover, the covariance matrix of $\widehat{\theta}_\Sigma$ is*

$$\Sigma_{(\widehat{H}_\Sigma, \widehat{C}_\Sigma)} = (A^t \Sigma_\alpha^{-1}(j_1, j_2; N)A)^{-1} = \frac{1}{\alpha^2}(A^t \Sigma_1^{-1}(j_1, j_2; N)A)^{-1},$$

where $\Sigma_1(j_1, j_2; N)$ is the covariance matrix of the Y_j statistics based on $1-$Fréchet data.

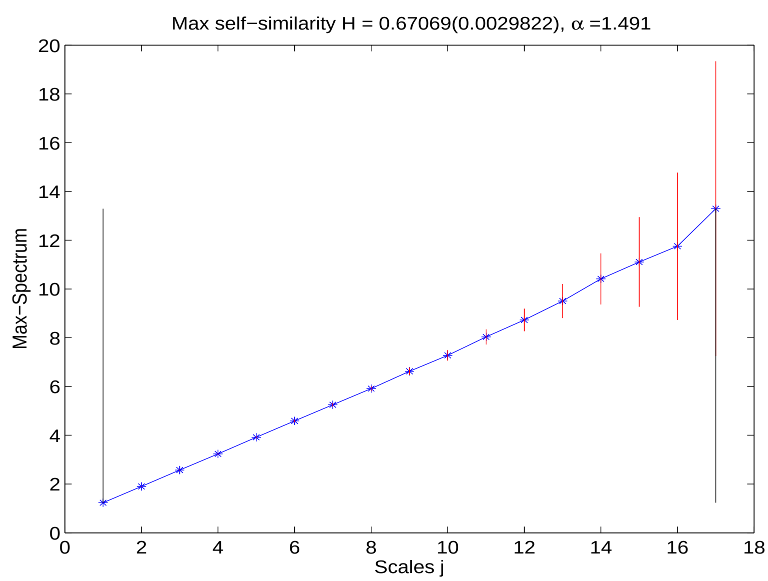

Fig 2. *Displayed is an example the max–spectrum of an i.i.d. $\alpha-$Fréchet sample of size $N = 2^{17} = 131,072$ with $\alpha = 1.5$. Observe that the max–spectrum is perfectly linear in j. The vertical intervals around every Y_j point indicate 95% confidence intervals for the mean of Y_j based on normal approximation. Observe that these confidence intervals grow with the scale j. GLS regression based on all scales $1 \le j \le 17$ was used to obtain an estimate $\widehat{\alpha} = 1.491$. The estimated standard deviation of the slope $\widehat{H} = 0.67$ is indicated in parentheses: $\widehat{\sigma}_H = 0.00298$. This last estimate is based on the asymptotic variance of $\widehat{H}$ (see Proposition 4.2).*

In Figure 2, the max–spectrum of a sample from a Fréchet distribution with $N = 2^{17}$ observations is shown. As expected, the max–spectrum is essentially linear in j and the slope yields a very good estimate of $1/\alpha$. The asymptotic properties of estimators based on the max–spectrum of general heavy–tailed samples are established in Section 4. In practice, when the sample is not Fréchet, the max–spectrum is linear in j only on a range of the largest scales j. The problem of choosing the "best" range of scales to estimate α is very important in practice and is briefly addressed in Section 5.2.

3. Rates for moment–type functionals of heavy–tailed maxima

In this section, we establish some results for moment–type functionals obtained from maxima of heavy–tailed data. They prove useful in establishing the consistency and asymptotic normality of the max self–similar estimators under general conditions, but are also of independent interest since they yield *exact rates* of convergence in many cases.

Let $X(1), X(2), \dots,$ be i.i.d. random variables with c.d.f.

$$F(x) = \exp\{-\sigma^\alpha(x)x^{-\alpha}\}, \quad x > 0, \tag{3.1}$$

where $\alpha > 0$, and where the function $\sigma(x) > 0$ is such that

$$\sigma(x) \longrightarrow \sigma_0 > 0, \quad \text{as } x \to \infty.$$

Here, we let the function $\sigma(x)$ take values in the extended half–line $(0, \infty]$, that is, $\sigma(x)$ can take the value ∞, in which case $F(x)$ becomes $e^{-\infty} = 0$ (see the Examples below). Such a representation always exists if the c.d.f. F belongs to the normal domain of attraction of an $\alpha-$Fréchet distribution, that is, if

$$M_n := \frac{1}{n^{1/\alpha}} \bigvee_{1\le i\le n} X(i) \stackrel{d}{\longrightarrow} Z, \tag{3.2}$$

where $G(x) := \mathbb{P}\{Z \le x\} = \exp\{-\sigma_0^\alpha x^{-\alpha}\},\ x > 0$, for some $\sigma_0 > 0$. For simplicity, we suppose that the $X(i)$'s are positive, almost surely, that is $F(0) = 0$. The case when the $X(i)$'s can take negative values is addressed in Section 4 below.

Our goal here is to establish bounds on the rate of convergence of $\mathbb{E}f(M_n)$ to $\mathbb{E}f(Z)$, as $n \to \infty$, for an absolutely continuous function $f : (0, \infty) \to \mathbb{R}$. We do so under general conditions on the asymptotic tail behavior of the c.d.f. $F(x)$.

In what follows, the next two conditions on the c.d.f. $F(x)$ are needed:

Condition 3.1 *For some $\beta > 0$ and $C_1 > 0$,*

$$|\sigma^\alpha(x) - \sigma_0^\alpha| \le C_1 x^{-\beta}, \quad \textit{for all sufficiently large } x > 0. \tag{3.3}$$

and

Condition 3.2 *We have $F(0) = 0$ and for some $C_2 > 0$,*

$$\sigma^\alpha(x) \ge C_2 \min\{1, x^\gamma\},\ x > 0, \quad \textit{for some} \ \ \gamma \in (0, \alpha). \tag{3.4}$$

In the examples below, we show that the Conditions 3.1 and 3.2 hold in many cases of practical interest. The second condition concerns the behavior of $F(x)$ for small x, and ensures that $\mathbb{E}(X^p 1_{\{X\le 1\}}) < \infty$, for any $p \in \mathbb{R}$. This condition always holds, for example, if the $X(i)$'s are bounded away from zero, almost surely. The case of arbitrary $X(i)$'s which can possibly take negative values is addressed in Section 4.

The following result provides an upper bound on $|\mathbb{E}f(M_n) - \mathbb{E}f(Z)|$ under the above conditions for general class of absolutely continuous functions f. Namely, we shall suppose that $f(x) = f(x_0) + \int_{x_0}^x f'(u)du,\ x > 0$, for some (any) $x_0 \in (0,\infty)$, with f' being a locally integrable function.

Theorem 3.1 *Let $f(x), x > 0$ be an absolutely continuous function on all compact intervals $[a, b] \subset (0,\infty)$. Let also $F_n(x) := \mathbb{P}\{M_n \le x\}$ and $G(x) = \mathbb{P}\{Z \le x\},\ x \in \mathbb{R}$, be the c.d.f.'s of the random variables M_n and Z in (3.2). Suppose that Conditions 3.1 and 3.2 hold.*

(a) *If for some $m \in \mathbb{R}$ and $\delta > 0$,*

$$x^m|f(x)| + \operatorname*{esssup}_{0<y\le x} y^m|f'(y)| \to 0,\ x \downarrow 0, \quad \textit{and} \quad x^{-\alpha}|f(x)| + x^{1+\delta} \operatorname*{esssup}_{y\ge x} y^{-\alpha}|f'(y)| \to 0,\ x \to \infty, \tag{3.5}$$

then $\mathbb{E}|f(Z)|$ and $\mathbb{E}|f(M_n)|,\ n \in \mathbb{N}$ are finite. Moreover,

$$\mathbb{E}f(M_n) - \mathbb{E}f(Z) = \int_0^\infty (G(x) - F_n(x))f'(x)dx. \tag{3.6}$$

Here esssup *denotes the essential supremum of a measurable function* g*, that is,*

$$\mathrm{esssup}_{y\in A} g(y) := \inf_{A_0\subset A,\ |A\setminus A_0|=0} \sup_{y\in A_0} g(y),$$

for any Borel set A*, where* $|A|$ *denotes the Lebesgue measure of the set* A*.*

(b) *If in addition to* (3.5)*,* $\int_1^\infty x^{-(\alpha+\beta)}|f'(x)|dx < \infty$*, then for any* $\epsilon(n) \to 0$*, such that* $n^{1/\alpha}\epsilon(n) \to \infty$*, as* $n \to \infty$*, we have*

$$|\mathbb{E}f(M_n) - \mathbb{E}f(Z)| \leq C_1 n^{-\beta/\alpha}\Big(\int_0^\infty x^{-(\alpha+\beta)}|f'(x)|e^{-cx^{-\alpha}}dx\Big) + 2\int_0^{\epsilon(n)} e^{-C_2 x^{-(\alpha-\gamma)}}|f'(x)|dx, \quad (3.7)$$

for all sufficiently large n*, where* $c \in (0, \sigma_0^\alpha)$ *can be chosen arbitrarily close to* σ_0^α*. Moreover,*

$$|\mathbb{E}f(M_n) - \mathbb{E}f(Z)| \leq C_f n^{-\beta/\alpha}, \quad (3.8)$$

for all sufficiently large n *with some* $C_f > 0$*.*

Proof: *We first prove part* **(a)***.* Let $f(x) = f(x_0) + \int_{x_0}^x f'(u)du,\ x > 0$, with $x_0 \in (0,\infty)$, where $f'(x),\ x \in (0,\infty)$ is locally integrable, and where $\int_a^b = -\int_b^a$. Let now $[a,b] \subset (0,\infty),\ x_0 \in (a,b)$ be an arbitrary interval and observe that $\int_a^b f(x)dF_n(x)$ equals

$$\int_a^{x_0} f(x)dF_n(x) + \int_{x_0}^b f(x)d(F_n(x)-1) = F_n(x_0)f(x_0) - F_n(a)f(a) - \int_a^{x_0} F_n(x)f'(x)dx$$
$$+(F_n(b)-1)f(b) - (F_n(x_0)-1)f(x_0) - \int_{x_0}^b (F_n(x)-1)f'(x)dx \quad (3.9)$$
$$= (F_n(b)-1)f(b) - F_n(a)f(a) + f(x_0)$$
$$- \int_a^{x_0} F_n(x)f'(x)dx + \int_{x_0}^b (1-F_n(x))f'(x)dx. \quad (3.10)$$

The equality in Relation (3.9) follows from Lemma 7.1.

In view of Relation (3.10), the monotone convergence theorem implies that $\mathbb{E}|f(M_n)| = \int_0^\infty |f(x)|dF_n(x)$ is finite if

$$|(F_n(b)-1)f(b)| + |F_n(a)f(a)| \longrightarrow 0, \quad \text{as } a\downarrow 0 \quad \text{and} \quad b\to\infty, \quad (3.11)$$

and if

$$\int_0^{x_0} F_n(x)|f'(x)|dx + \int_{x_0}^\infty (1-F_n(x))|f'(x)|dx < \infty. \quad (3.12)$$

Observe that by (3.1),

$$F_n(x) = F(n^{1/\alpha}x)^n = \exp\{-\sigma^\alpha(n^{1/\alpha}x)x^{-\alpha}\}, \quad x > 0.$$

Hence, in view of (3.3) we have

$$1 - F_n(x) \sim \sigma_0^\alpha x^{-\alpha}, \quad \text{as } x \to \infty, \tag{3.13}$$

since $1 - e^{-u} \sim u$, as $u \downarrow 0$. Thus, the second convergence in (3.5), implies $|(F_n(b) - 1)f(b)| \to 0,\ b \to \infty$. On the other hand, by (3.4), for $n \geq 1,\ n \in \mathbb{N}$,

$$\sigma^\alpha(n^{1/\alpha}x) \geq C_2 n^{\gamma/\alpha}x^\gamma \geq C_2 x^\gamma, \quad \text{for all } x \in (0, n^{-1/\alpha}), \tag{3.14}$$

and hence

$$F_n(x) = \exp\{-\sigma^\alpha(n^{1/\alpha}x)x^{-\alpha}\} \leq \exp\{-C_2 x^{-(\alpha-\gamma)}\}, \quad \text{for all} \ \ x \in (0, n^{-1/\alpha}). \tag{3.15}$$

Thus, since $u^p e^{-u} \to 0$, as $u \to \infty$, for any $p \in \mathbb{R}$, the first convergence in (3.5) implies that $F_n(a)f(a) \to 0$, as $a \to \infty$. We have thus shown that (3.11) holds. One can similarly show that the integrals in (3.12) are finite by the using the conditions in (3.5) on f' and Relations (3.13) and (3.14). Indeed, for almost all $x > 0$, we have

$$F_n(x)|f'(x)| \leq (\sup_{0<y\leq x} F_n(y)y^{-m})(\text{esssup}_{0<y\leq x} y^m|f'(y)|) = \mathcal{O}(x^{-|m|}\exp\{-C_2 x^{-(\alpha-\gamma)}\}) \to 0, \tag{3.16}$$

as $x \downarrow 0$ and, for almost all $x > 0$,

$$(1 - F_n(x))|f'(x)| \leq (\sup_{y\geq x}(1 - F_n(y))y^{-\alpha})(\text{esssup}_{y\geq x} y^\alpha|f'(y)|) = \mathcal{O}(x^{-(1+\delta)}), \tag{3.17}$$

as $x \to \infty$. We have thus shown that $\int_0^\infty |f(x)|dF_n(x) < \infty$ for all $n \in \mathbb{N}$. One can similarly show that $\int_0^\infty |f(x)|dG(x) < \infty$, by replacing $F_n(x)$ with $G(x)$, above, and using the fact that $G(x) = \exp\{-\sigma_0^\alpha x^{-\alpha}\},\ x > 0$ satisfies trivially Conditions 3.1 and 3.2.

Observe that (3.6) follows from the relations

$$\int_0^\infty f(x)dF_n(x) = f(x_0) - \int_0^{x_0} F_n(x)f'(x)dx + \int_{x_0}^\infty (1 - F_n(x))f'(x)dx$$

and

$$\int_0^\infty f(x)dG(x) = f(x_0) - \int_0^{x_0} G(x)f'(x)dx + \int_{x_0}^\infty (1-G(x))f'(x)dx.$$

We now turn to proving part **(b)***.* Let $\epsilon(n) \downarrow 0$ be such that $n^{1/\alpha}\epsilon(n) \to \infty$, as $n \to \infty$. By (3.6), using the triangle inequality, we get

$$\begin{aligned} |\mathbb{E}f(M_n) - \mathbb{E}f(Z)| \ &\le \ \int_0^{\epsilon(n)} G(x)|f'(x)|dx + \int_0^{\epsilon(n)} F_n(x)|f'(x)|dx \\ &\quad + \int_{\epsilon(n)}^\infty |F_n(x) - G(x)||f'(x)|dx =: I_1 + I_2 + I_3. \end{aligned}$$

We first consider the integral I_3. Since $n^{1/\alpha}\epsilon(n) \to \infty,\ n \to \infty$, in view of (3.3), for all sufficiently large n, we have

$$\begin{aligned} |F_n(x) - G(x)| \ &= \ |\sigma^\alpha(n^{1/\alpha}x) - \sigma_0^\alpha| x^{-\alpha} e^{-\theta_n(x)x^{-\alpha}} \\ &\le \ C_1 n^{-\beta/\alpha} x^{-(\alpha+\beta)} e^{-cx^{-\alpha}}, \end{aligned} \tag{3.18}$$

for all $x \in (\epsilon(n), \infty)$, where c is an arbitrary constant in $(0, \sigma_0^\alpha)$, and where $\theta_n(x)$ is between $\sigma^\alpha(n^{1/\alpha}x)$ and σ_0^α. Indeed, the first relation in (3.18) follows by the mean value theorem applied to the function $g(u) = \exp\{-ux^{-\alpha}\},\ u > 0$. The inequality in (3.18), follows from (3.3) since $n^{1/\alpha}\epsilon(n) \to \infty$ implies $\sup_{x \ge \epsilon(n)} \sigma^\alpha(n^{1/\alpha}x) \ge c,\ c \in (0, \sigma_0^\alpha)$, for all sufficiently large n.

Therefore (3.18) implies

$$I_3 \le C_1 n^{-\beta/\alpha} \int_{\epsilon(n)}^\infty x^{-(\alpha+\beta)} e^{-cx^{-\alpha}} |f'(x)|dx \le C_1 n^{-\beta/\alpha} \int_0^\infty x^{-(\alpha+\beta)} |f'(x)| e^{-cx^{-\alpha}} dx,$$

for all sufficiently large n. The last integral is finite. Indeed, by assumption $\int_1^\infty x^{-(\alpha+\beta)}|f'(x)|dx < \infty$. The integral $\int_0^1 x^{-(\alpha+\beta)}|f'(x)|e^{-cx^{-\alpha}}dx$ is finite since in view of (3.5),

$$(\mathrm{esssup}_{0 \le y \le x} y^m |f'(y)|) x^{-(\alpha+\beta+|m|)} e^{-cx^{-\alpha}} = \mathcal{O}\Big(x^{-(\alpha+\beta+|m|)} e^{-cx^{-\alpha}}\Big) = \mathcal{O}(x^p),\ x \downarrow 0, \tag{3.19}$$

for any $p > 0$.

We now consider the integral I_2. Observe that $\epsilon(n) > n^{-1/\alpha}$, eventually, and hence

$$I_2 \le \int_0^{n^{-1/\alpha}} \exp\{-C_2 x^{-(\alpha-\gamma)}\}|f'(x)|dx + \int_{n^{-1/\alpha}}^{\epsilon(n)} F_n(x)|f'(x)|dx, \tag{3.20}$$

by (3.15). Relation (3.4) implies that $\sigma^\alpha(n^{1/\alpha}x) \geq C_2$, for all $x \in (n^{-1/\alpha}, \epsilon(n))$, and hence $F_n(x) \leq \exp\{-C_2 x^{-\alpha}\} \leq \exp\{-C_2 x^{-(\alpha-\gamma)}\}$, $x \in (n^{-1/\alpha}, \epsilon(n))$. Therefore, the second integral in (3.20) can be bounded above by $\int_{n^{-1/\alpha}}^{\epsilon(n)} \exp\{-C_2 x^{-(\alpha-\gamma)}\}|f'(x)|dx$ and hence

$$I_2 \leq \int_0^{\epsilon(n)} \exp\{-C_2 x^{-(\alpha-\gamma)}\}|f'(x)|dx.$$

One can similarly bound I_1. Indeed, Relation (3.4) implies that $\sigma_0^\alpha \geq C_2$, since $\sigma^\alpha(x) \sim \sigma_0^\alpha,\ x \to \infty$. For all $0 < x < \epsilon(n) < 1$ and $\gamma \in (0, \alpha)$, we have $x^{-\alpha} \geq x^{-(\alpha-\gamma)}$, and hence we obtain

$$I_1 = \int_0^{\epsilon(n)} \exp\{-\sigma_0^\alpha x^{-\alpha}\}|f'(x)|dx \leq \int_0^{\epsilon(n)} \exp\{-C_2 x^{-(\alpha-\gamma)}\}|f'(x)|dx.$$

The last three bounds for I_1, I_2 and I_3 imply (3.7).

Now, to prove (3.8), observe that, as in (3.19), since $\alpha - \gamma > 0$, for almost all $x > 0$, we have

$$\exp\{-C_2 x^{-(\alpha-\gamma)}\}|f'(x)| \leq \mathcal{O}\Big(x^{-|m|}e^{-C_2 x^{-(\alpha-\gamma)}}\Big) = \mathcal{O}(x^p),\ x \downarrow 0, \tag{3.21}$$

for any $p > 0$. Thus, the second integral in (3.7) is of order $\mathcal{O}(\epsilon(n)^p)$, for any $p > 0$ and by setting $\epsilon(n) := n^{-\delta}$, for some $\delta \in (0, 1/\alpha)$, we obtain that (3.8) holds. This completes the proof of the theorem. □

In the following examples we show that most heavy–tailed distributions of practical interest satisfy the conditions of Theorem 3.1.

Examples:

- *(Pareto laws)* Let $F(x) = 1 - (x/\sigma_0)^{-\alpha},\ x \geq \sigma_0$, and $F(x) = 0$, $x < \sigma_0$, for some $\sigma_0 > 0$ and $\alpha > 0$. Then, Relation (3.1) holds with

$$\sigma^\alpha(x) = \infty 1_{(0,\sigma_0]}(x) - x^\alpha \ln(1 - (x/\sigma_0)^{-\alpha}) 1_{(\sigma_0,\infty)}(x),$$

that is, the function $\sigma(x)$ equals ∞ for all $x \in (0, \sigma_0]$ to account for the fact that $F(x) = 0,\ x \in (0, \sigma_0]$.

Observe that $\sigma^\alpha(x)$ *satisfies Condition 3.1 with* $\beta = \alpha$. Indeed, since $\ln(1-u) = -u + u^2/2 + \mathcal{O}(u^3),\ u \to 0$, by setting $u := (x/\sigma_0)^{-\alpha}$, we obtain

$$|\sigma^\alpha(x) - \sigma_0^\alpha| = \Big|\frac{\ln(1 - (x/\sigma_0)^{-\alpha})}{x^{-\alpha}} + \sigma_0^\alpha\Big| = \sigma_0^\alpha\Big|\frac{\ln(1-u)}{u} + 1\Big| \leq \sigma_0^\alpha u = \sigma_0^{2\alpha} x^{-\alpha}, \tag{3.22}$$

for all sufficiently large x.

One has, moreover, that

$$\sigma^{\alpha}(x) - \sigma_0^{\alpha} \sim \frac{\sigma_0^{2\alpha}}{2} x^{-\alpha}, \quad \text{as } x \to \infty. \tag{3.23}$$

(see Proposition 3.1, below).

Condition 3.2 also holds. Indeed, $\sigma(x) = \infty \geq x^{\gamma}$, for all $x \in (0, \sigma_0]$ and $\gamma \in (0, \alpha)$. To prove (3.4), it remains to show that $\sigma^{\alpha}(x) \geq C_2 > 0$, for all $x > 0$. As shown in (3.22) above $\sigma^{\alpha}(x) \to \sigma_0^{\alpha},\ x \to \infty$, where $\sigma_0 > 0$. On the other hand $\sigma^{\alpha}(x)$ is a positive, continuous function over all compact intervals of (σ_0, ∞) and $\sigma(x) \to \infty$, as $x \to \sigma_0$. This shows that $\sigma^{\alpha}(x)$ is bounded below by a positive constant.

- *(Products of Fréchet laws)* Let $F(x) = G_{\alpha_0}(x/\sigma_0) G_{\alpha_1}(x/\sigma_1)$, where $\sigma_0,\ \sigma_1 > 0$ and $0 < \alpha_0 < \alpha_1$, and where $G_{\alpha}(x) = \exp\{-x^{-\alpha}\},\ x > 0$ denotes the c.d.f. of a standard $\alpha-$Fréchet variable. Observe that the function $F(x)$ is the c.d.f. of $\max\{\sigma_0 Z_0, \sigma_1 Z_1\}$, where Z_0 and Z_1 are independent standard α_0- and α_1-Fréchet random variables, respectively. Therefore, (3.1) holds with $\alpha = \alpha_0$ and

$$\sigma^{\alpha}(x) = \sigma_0^{\alpha} + \sigma_1^{\alpha} x^{-(\alpha_1 - \alpha_0)}, \ \ x > 0. \tag{3.24}$$

Conditions 3.1 and 3.2 are readily satisfied where $\beta = \alpha_1 - \alpha_0 > 0$.

- *(Mixtures of Pareto laws)* Let

$$F(x) = p(1 - (x/\sigma_0)^{-\alpha_0}) 1_{\{x \geq \sigma_0\}} + (1-p)(1 - (x/\sigma_1)^{-\alpha_1}) 1_{\{x \geq \sigma_1\}}, \ \ 0 < \alpha_0 < \alpha_1,$$

where $p \in (0,1)$ and $\sigma_0,\ \sigma_1 > 0$.

Then, (3.1) holds with $\alpha \equiv \alpha_0$, and $\sigma^{\alpha}(x) = \infty 1_{(0,\sigma_*]}(x) - x^{\alpha} \ln(F(x)) 1_{(\sigma_*,\infty)}(x)$, where $\sigma_* := \min\{\sigma_0, \sigma_1\} > 0$.

As in the case of Pareto laws, one can show that Condition 3.1 holds with $\beta = \min\{\alpha_0, \alpha_1 - \alpha_0\}$ and, σ_0 replaced by $p\sigma_0$. In fact,

$$\sigma^{\alpha}(x) - p\sigma_0^{\alpha} \sim C_0 x^{-\beta}, \quad \text{as } x \to \infty, \tag{3.25}$$

where

$$C_0 = \begin{cases} \sigma_1^{\alpha_1}(1-p) & , \text{ if } \alpha_1 - \alpha_0 < \alpha_0 \\ \sigma_1^{\alpha_1}(1-p) + p^2\sigma_0^{2\alpha_0}/2 & , \text{ if } \alpha_1 - \alpha_0 = \alpha_0 \\ p^2\sigma_0^{2\alpha_0}/2 & , \text{ if } \alpha_1 - \alpha_0 > \alpha_0 \end{cases}$$

One can also show that Condition 3.2 holds as in the case of Pareto laws.

- Absolute values of $\alpha-$stable ($0 < \alpha < 2$) and $t-$distributed random variables X_i's, for example, also satisfy Condition 3.1. They *do not* satisfy Condition 3.2, however, since $\mathbb{E}(|X_1|^{-1}1_{\{|X|\le 1\}})$ is infinite. In Proposition 4.3 below, we address the general case where Condition 3.2 fails and in fact the case where the X_i's can take negative values.

The following result shows that the rate $n^{-\beta/\alpha}$ in (3.8) is optimal, if so is the inequality in (3.3).

Proposition 3.1 *Assume that F is as in* (3.1) *and satisfies Conditions 3.1 and 3.2 above, and let f be as in Theorem 3.1* **(b)**. *Suppose, in addition, that $\sigma^\alpha(x) - \sigma_0^\alpha \sim C_1 x^{-\beta}$, as $x \to \infty$, for some $C_1 \neq 0$. Then*

$$n^{-\beta/\alpha}(\mathbb{E}f(M_n) - \mathbb{E}f(Z)) \longrightarrow C_1 \int_0^\infty x^{-(\alpha+\beta)} f'(x) e^{-\sigma_0^\alpha x^{-\alpha}} dx, \quad \text{as } n \to \infty. \tag{3.26}$$

Proof: Let as in Theorem 3.1, $\epsilon(n) \to 0$ be such that $n^{1/\alpha}\epsilon(n) \to \infty$, as $n \to \infty$. The triangle inequality applied to Relation (3.6) implies

$$\Big|\mathbb{E}f(M_n) - \mathbb{E}f(Z) - \int_{\epsilon(n)}^\infty (G(x) - F_n(x)) f'(x) dx\Big| \le \int_0^{\epsilon(n)} G(x)|f'(x)|dx + \int_0^{\epsilon(n)} F_n(x)|f'(x)|dx. \tag{3.27}$$

As in the proof of Theorem 3.1 one can show that the integrals in the right–hand side of the last expression are of order $o(n^{-\beta/\alpha})$, as $n \to \infty$, if $\epsilon(n) := n^{-\delta}$, $\delta \in (0, 1/\alpha)$ (see (3.21)).

To establish (3.26) we will now examine the order of the integral in the left–hand side of (3.27). Observe that

$$\frac{\sigma^\alpha(n^{1/\alpha}x) - \sigma_0^\alpha}{n^{-\beta/\alpha}} \longrightarrow C_1 x^{-\beta}, \tag{3.28}$$

as $n \to \infty$, for all $x > 0$. Hence (as in Theorem 3.1), in view of (3.1) and (3.28), the mean value theorem implies

$$n^{\beta/\alpha}(G(x) - F_n(x))f'(x) \longrightarrow C_1 x^{-(\alpha+\beta)} f'(x) e^{-\sigma_0^\alpha x^{-\alpha}},$$

as $n \to \infty$, for any $x \in (\epsilon(n), \infty)$ and hence for any $x > 0$ ($\epsilon(n) \to 0$, $n \to \infty$). As in the proof of Theorem 3.1, one can show that the left–hand side of the last expression is bounded above in absolute value by an integrable function. Therefore, the dominated convergence theorem implies that $n^{\beta/\alpha} \int_0^\infty (G(x) - F_n(x)) f'(x) dx$ converges to the integral in (3.26), as $n \to \infty$. □

The next result, which follows directly from Theorem 3.1 is used in Section 4.

Corollary 3.1 *Assume that F is as in* (3.1) *and satisfies Conditions 3.1 and 3.2 above. Then $\mathbb{E}|\ln(M_n)|^p < \infty$ for all $n \in \mathbb{N}$ and $p > 0$. Moreover, for any $p > 0$ and $k \in \mathbb{N}$, we have*

$$\Big|\mathbb{E}|\ln(M_n)|^p - \mathbb{E}|\ln(Z)|^p\Big| = \mathcal{O}(n^{-\beta/\alpha}) \quad \textit{and} \quad \Big|\mathbb{E}\ln(M_n)^k - \mathbb{E}\ln(Z)^k\Big| = \mathcal{O}(n^{-\beta/\alpha}),$$

as $n \to \infty$, where M_n and Z are as in Theorem 3.1.

In Section 4, one encounters covariance functionals of maxima over blocks of heavy–tailed variables, that is, bivariate moment–type functionals arise. The following result establishes rates of convergence for such functionals in the special case of logarithms.

Corollary 3.2 *Suppose that F is as in* (3.1) *and satisfies Conditions 3.1 and 3.2. Let $X(1), \ldots, X(n)$ and $Y(1), \ldots, Y(m)$, $n, m \in \mathbb{N}$ be i.i.d. random variables with c.d.f. $F(x)$. Consider the normalized maxima*

$$M_n^X := \frac{1}{n^{1/\alpha}} \bigvee_{1 \le i \le n} X(i) \quad \textit{and} \quad M_m^Y := \frac{1}{m^{1/\alpha}} \bigvee_{1 \le i \le m} Y(i), \ n, m \in \mathbb{N}.$$

Then, for any $a > 0$, as $n,\ m \to \infty$, we have that

$$\mathbb{E}\ln(M_n^X)\ln(M_n^X \vee a M_m^Y) - \mathbb{E}\ln(Z_X)\ln(Z_X \vee a Z_Y) = \mathcal{O}(n^{-\beta/\alpha} + m^{-\beta/\alpha}), \tag{3.29}$$

where Z_X and Z_Y are independent $\alpha-$Fréchet random variables with scale coefficients σ_0.

Corollary 3.2 was stated in generality which allows us to have different number of $X(i)$'s and $Y(i)$'s (n and m, respectively) in the maxima M_n^X and M_m^Y. This flexibility is needed for the proof of Proposition 4.1 below.

PROOF OF COROLLARY 3.1: Let $f(x) = |\ln(x)|^p,\ p > 0,\ x > 0$. Observe that $f(x) = \int_1^x f'(u)du$, where $f'(x) = p|\ln(x)|^{p-1}/x$ for $x \geq 1$ and $f'(x) = -p|\ln(x)|^{p-1}/x$, for $0 < x \leq 1$. One can verify that the conditions in (3.5) are fulfilled and therefore, Theorem 3.1 implies the result. The argument in the case when $f(x) = (\ln(x))^k,\ k \in \mathbb{N}$ is similar. □

PROOF OF COROLLARY 3.2: By Corollary 3.1, the expected values in (3.29) exist since $\mathbb{E}|\ln(M_n^X)|^p < \infty,\ \forall p > 0$ and since $a \vee b \leq a + b$ for any $a,\ b \geq 0$. Observe that by independence and Fubini's theorem,

$$\mathbb{E}\ln(M_n^X)\ln(M_n^X \vee aM_m^Y) = \int_0^\infty \Big(\int_0^\infty f(x,y)dF_n(x)\Big)dF_m(y),$$

and

$$\mathbb{E}\ln(Z_X)\ln(Z_X \vee aZ_Y) = \int_0^\infty \Big(\int_0^\infty f(x,y)dG(x)\Big)dG(y),$$

where $f(x,y) = \ln(x)\ln(x \vee ay),\ x,y,a > 0,\ F_n(x) := F(n^{1/\alpha}x)^n$ is the c.d.f. of M_n^X (and M_n^Y), and where $G(x) = \exp\{-\sigma_0^\alpha x^{-\alpha}\},\ x > 0$. Now, by adding and subtracting the term $\int_0^\infty(\int_0^\infty f(x,y)dG(x))dF_m(y)$, applying Fubini's theorem and then the triangle inequality, we obtain that the left–hand side of (3.29) is bounded above in absolute value by

$$\int_0^\infty \Big| \int_0^\infty f(x,y)dF_n(x) - \int_0^\infty f(x,y)dG(x)\Big| dF_m(y)$$
$$+ \int_0^\infty \Big| \int_0^\infty f(x,y)dF_m(y) - \int_0^\infty f(x,y)dG(y)\Big| dG(x) =: I_1 + I_2.$$

Focus next on the term I_1. Let $g(y) := \int_0^\infty f(x,y)(dG(x) - dF_n(x)), y > 0$. Observe that for each $y > 0,\ y \neq x/a,\ f(x,y)$ is differentiable in x since

$$f(x,y) = \begin{cases} \ln(x)\ln(ay) & ,\ 0 < x < ay \\ \ln(x)^2 & ,\ ay \leq x \end{cases}$$

In fact,

$$|f'_x(x,y)| \leq 2|\ln(x)|/x + |\ln(ay)|/x, \quad x > 0,\ y > 0.$$

Thus, Theorem 3.1 **(b)**, applied to the inner integral $g(y)$ in I_1 implies

$$|g(y)| \le n^{-\beta/\alpha}(C' + C''|\ln(y)|), \tag{3.30}$$

for all sufficiently large n, where the constants $C' > 0$ and $C'' > 0$ do not depend on y (This follows from Relation (3.7) by taking $\epsilon(n) := n^{-\delta},\ \delta \in (0, 1/\alpha)$ and observing that the second integral therein is negligible with respect to the term $(1 + |\ln(y)|)n^{-\beta/\alpha}$.)

Note now that the function $|\ln(y)|$ satisfies the assumptions of Theorem 3.1 **(b)** and hence $\int_0^\infty |\ln(y)|dF_m(y) \to \int_0^\infty |\ln(y)|dG(y)$, as $m \to \infty$. Therefore, the inequality (3.30) implies that $I_1 = \mathcal{O}(n^{-\beta/\alpha})$, as $n \to \infty$. One can similarly show that $I_2 = \mathcal{O}(m^{-\beta/\alpha}),\ m \to \infty$. $\Box$

4. Asymptotic properties of the max self–similarity estimators

We establish here the consistency and asymptotic normality of the estimators defined in (2.9), above. In fact, we prove joint asymptotic normality of the max self–similarity estimators of the tail exponent α and the scale coefficient σ_0. These results rely on the behavior of moment–type functionals of heavy–tailed maxima established in Section 3.

The general case where the $X(i)$'s may be 0 or even take negative values is addressed at the end of this section.

Let the Y_j's be defined as in (2.7), where now N denotes the sample size of available $X(i)$'s, $1 \le j \le [\log_2 N]$ and where $N_j := [N/2^j]$. As noted above, the larger the scales j, the more precise the asymptotic relation (2.8). Therefore, to obtain consistent estimates for the parameter $H = 1/\alpha$ one should focus on a range of scales which grows as the sample size increases. We therefore *fix* a range $j_1 \le j \le j_2,\ j_1, j_2 \in \mathbb{N}$ and focus on the vectors

$$Y_r := \{Y_{j+r}\}_{j=j_1}^{j_2},$$

with $r \in \mathbb{N},\ j_2 + r \le [\log_2 N]$ where the parameter $r = r(N)$ grows with the sample size.

The following result shows that the mean and the covariance matrix of the vector Y_r are asymptotically equivalent to the mean and and the covariance matrix in the case where the $X(i)$'s are $\alpha-$Fréchet (see Proposition 2.1).

Proposition 4.1 *Suppose that the c.d.f. F has the representation* (3.1) *and satisfies Conditions 3.1 and 3.2, above.*

Then,

$$\Big|\mathbb{E}Y_{j+r} - \mu_r(j)\Big| = \mathcal{O}\Big(1/2^{r\beta/\alpha}\Big), \quad \text{as } r \to \infty, \tag{4.1}$$

and for any fixed $j_1 \le i \le j \le j_2,\ i, j \in \mathbb{N}$, we have

$$\Big|N_{j_2+r}\mathrm{Cov}(Y_{i+r}, Y_{j+r}) - \alpha^{-2}\Sigma_1(i,j)\Big| = \mathcal{O}\Big(1/2^{r\beta/\alpha}\Big) + \mathcal{O}\Big(2^r/N\Big), \quad \text{as } r \to \infty. \tag{4.2}$$

Here

$$\mu_r(j) := (j+r)/\alpha + C(\sigma_0, \alpha) \quad \text{and} \quad \Sigma_1(i,j) = 2^{j-j_2}\psi(|i-j|), \tag{4.3}$$

where the function ψ is defined in (2.15) *and where $C(\sigma_0, \alpha)$ is as in* (2.12).

Proof: Observe that by (2.7), we have $\mathbb{E}Y_{j+r} = \mathbb{E}\log_2(D(j+r,1)) = \mathbb{E}\log_2\Big(\bigvee_{i=1}^{2^{j+r}} X(i)\Big)$. Therefore,

$$\begin{aligned} \mathbb{E}Y_{j+r} - (j+r)/\alpha - \mathbb{E}\log_2(\sigma_0 Z) &= \mathbb{E}\log_2\Big(\frac{1}{2^{(j+r)/\alpha}}\bigvee_{i=1}^{2^{j+r}} X(i)\Big) - \mathbb{E}\log_2(\sigma_0 Z) \\ &= \mathbb{E}\log_2(M_n) - \mathbb{E}\log_2(\sigma_0 Z), \end{aligned} \tag{4.4}$$

where $M_n := n^{-1/\alpha}\bigvee_{i=1}^n X(i)$ and where $n := 2^{(j+r)}$. Corollary 3.1 implies that the right–hand side of (4.4) is of order $\mathcal{O}(n^{-\beta/\alpha}) = \mathcal{O}(2^{-(j+r)\beta/\alpha}) = \mathcal{O}(2^{-r\beta/\alpha})$, as $r \to \infty$, which in turn implies (4.1).

We now focus on proving (4.2). Let $i < j$ and recall that $N_{j+r} = [N/2^{j+r}]$, and $N_{i+r} = [N/2^{i+r}]$. We also have that

$$D(j+r, k) = \bigvee_{r=1}^{2^{j-i}} D(i+r, 2^{j-i}(k-1)+i), \quad \text{for all } k = 1, \ldots, N_{j+r}. \tag{4.5}$$

Note that $2^{j-i}N_{j+r} \le N_{i+r}$ and therefore as in the proof of Proposition 2.1 above, we get

$$\begin{aligned} \mathrm{Cov}(Y_{j+r}, Y_{i+r}) &= \frac{1}{N_{j+r}N_{i+r}} \sum_{k_1=1}^{N_{j+r}} \sum_{k_2=1}^{N_{i+r}} \mathrm{Cov}(\log_2 D(j+r, k_1), \log_2 D(i+r, k_2)) \\ &= \frac{1}{N_{j+r}N_{i+r}} \sum_{k_1=1}^{N_{j+r}} \sum_{\ell=1}^{2^{j-i}} \mathrm{Cov}\Big(\log_2 D(j+r, k_1), \log_2 D(i+r, 2^{j-i}(k_1-1)+\ell)\Big). \end{aligned}$$

The second sum in the last expression involves only terms $D(i+r, 2^{j-i}(k_1-1)+\ell)$, for $\ell = 1, \dots, 2^{j-i}$ since in view of (4.5), the independence of the $D(i+r,k)$'s implies that $\mathrm{Cov}(D(j+r,k_1), D(i+r,k_2)) = 0$, for all k_2 outside the range $2^{j-i}(k_1-1)+\ell,\ \ell = 1,\dots,2^{j-i}$.

Now, by using the stationarity of the $D(i+r,k)$'s and Relation (4.5) again, we obtain from the last relation that

$$\begin{aligned}\mathrm{Cov}(Y_{j+r}, Y_{i+r}) &= \frac{2^{j-i}}{N_{i+r}} \mathrm{Cov}\Big(\log_2 \Big(\bigvee_{\ell=1}^{2^{j-i}} D(i+r,\ell)\Big), \log_2 D(i+r,1)\Big) \\ &= \frac{2^{j-i}}{N_{i+r}} \mathrm{Cov}\Big(\log_2 \Big(M_n' \vee (2^{j-i}-1)^{1/\alpha} M_m''\Big), \log_2 (M_n')\Big), \end{aligned} \tag{4.6}$$

where $n := 2^{i+r}$ and $m := (2^{j-i}-1)n$ with $M_n' := n^{-1/\alpha} D(i+r,1) = n^{-1/\alpha} \bigvee_{\ell=1}^n X(\ell)$, and

$$M_m'' := m^{-1/\alpha} \bigvee_{\ell=2}^{2^{j-i}} D(i+r,\ell) = m^{-1/\alpha} \bigvee_{\ell=n+1}^{n+m} X(\ell) \stackrel{d}{=} M_m'.$$

Observe that the normalized maxima M_n' and M_m'' are independent since they involve maxima of disjoint sets of $X(r)$'s. Thus, by combining the results of Corollaries 3.1 and 3.2, we obtain that

$$\mathrm{Cov}\Big(\log_2 (M_n' \vee (2^{j-i}-1)^{1/\alpha} M_m''), \log_2 (M_n')\Big) - \alpha^{-2}\psi(|i-j|) = \mathcal{O}\Big(1/2^{r\beta/\alpha}\Big), \quad r \to \infty, \tag{4.7}$$

where ψ is as in (2.15). Now, note that $N_{i+r} = 2^{j_2-i} N_{j_2+r} + q$, where $q < 2^{j_2-i},\ q \in \mathbb{N}$. This follows from the facts that $N_{i+r} = [N/2^{i+r}]$, $i = j_1, \dots, j_2$ and $i \le j_2$. Thus

$$\frac{N_{j_2+r}}{N_{i+r}} - 2^{i-j_2} = \mathcal{O}(1/N_r) = \mathcal{O}(2^r/N). \tag{4.8}$$

Now, by applying Relations (4.7) and (4.8), to (4.6), we obtain (4.2). This completes the proof of the proposition. □

The following theorem is the main result of the section. It establishes the uniform convergence of the vector Y_r to a normal vector and provides bounds on its rate of convergence. The asymptotic normality of the estimators defined in (2.13) is then an immediate consequence of this result (see Corollary 4.1 below).

Theorem 4.1 *Suppose that the c.d.f. F has the representation* (3.1) *and satisfies Conditions 3.1 and 3.2, above. Let $\theta = \{\theta_j\}_{j=j_1}^{j_2} \in \mathbb{R}^m \setminus \{\mathbf{0}\},\ m = j_2 - j_1 + 1$ be an arbitrary fixed, non–zero vector and consider the linear combination $(\theta, Y_r) := \sum_{j=j_1}^{j_2} \theta_j Y_{j+r}$.*

Then,

$$\sup_{x\in\mathbb{R}} \Big| \mathbb{P}\{\sqrt{N_{j_2+r}}\Big((\theta, Y_r) - (\theta, \mu_r)\Big) \le x\} - \Phi(x/\sigma_\theta) \Big| \le C_\theta \Big(1/2^{r\beta/\alpha} + r2^{r/2}/\sqrt{N} \Big), \tag{4.9}$$

where Φ stands for the standard Normal c.d.f. and where $C_\theta > 0$ does not depend on N. Here $N_j = [N/2^j]$ denotes the number of coefficients $D(j,k)$ available on scale j, $(\theta, \mu_r) := \sum_{j=j_1}^{j_2} \theta_j \mu_r(j)$ and

$$\sigma_\theta^2 = \alpha^{-2}(\theta, \Sigma_1 \theta) := \alpha^{-2} \sum_{i,j=j_1}^{j_2} \theta_i \Sigma_1(i,j) \theta_j > 0. \tag{4.10}$$

Proof: Since $N_i = [N/2^i],\ i = 1, \ldots, [\log_2 N]$, for all $j = j_1, \ldots, j_2$, and $r \in \mathbb{N},\ r \le [\log_2 N] - j_2$, we have $N_{j+r} = 2^{j_2-j} N_{j_2+r} + q_j$, where $0 \le q_j < 2^{j_2-j},\ q_j \in \mathbb{N}$. Thus, for all $j = j_1, \ldots, j_2$,

$$\begin{aligned} Y_{j+r} &= \frac{1}{N_{j+r}} \sum_{k=1}^{N_{j_2+r}} \sum_{i=1}^{2^{j_2-j}} \log_2 D(j+r, 2^{j_2-j}(k-1)+i) + \frac{1}{N_{j+r}} \sum_{i=1}^{q_j} \log_2 D(j+r, 2^{j_2-j} N_{j_2+r} + i) \\ &=: \frac{1}{N_{j_2+r}} \sum_{k=1}^{N_{j_2+r}} y_{j+r}(k) + R_j, \end{aligned} \tag{4.11}$$

where $y_{j+r}(k) := N_{j_2+r} N_{j+r}^{-1} \sum_{i=1}^{2^{j_2-j}} \log_2 D(j+r, 2^{j_2-j}(k-1)+i)$.

Therefore,

$$(\theta, Y_r) = \frac{1}{N_{j_2+r}} \sum_{k=1}^{N_{j_2+r}} \xi_r(k) + (\theta, R), \tag{4.12}$$

where $\xi_r(k) := (\theta, y_r(k)),\ k = 1, \ldots, N_{j_2+r}$, with $y_r(k) = \{y_{j+r}(k)\}_{j=j_1}^{j_2}$ and $R = \{R_j\}_{j=j_1}^{j_2}$.

Observe that the random vectors $y_r(k),\ k = 1, \ldots, N_{j_2+r}$ are i.i.d. and independent from the remainder term (θ, R). Indeed, this follows from the fact that the $X(i)$'s are i.i.d. and because for any $j = j_1, \ldots, j_2$, the random variable $y_{j+r}(k)$ depends only on the $X(i)$'s with indices $2^{j_2+r}(k-1)+1 \le i \le 2^{j_2+r}k,\ k = 1, \ldots, N_{j_2+r}$, and R_j depends on the $X(i)$'s with indices $2^{j_2+r} N_{j_2+r} + 1 \le i \le N$.

Thus, to prove (4.9), we proceed in two steps. First, we apply the Central Limit Theorem to the first term on the right–hand side (r.h.s.) of (4.12). Then, we will argue that the remainder term therein can be neglected.

Step 1. Note that the $\xi_r(k)$'s are i.i.d. but their distributions depend on N and hence the ordinary C.L.T. does not apply. The Berry–Esseen bound, however, (see e.g. Theorem V.2.4 in Petrov (1995)) implies that

$$\sup_{x\in\mathbb{R}} \Big| Q_{N,r}(x) - \Phi(x) \Big| \le A \frac{\mathbb{E}|\xi_r(1) - \mathbb{E}\xi_r(1)|^3}{\sigma_{\xi_r}^3} \frac{1}{\sqrt{N_{j_2+r}}}, \tag{4.13}$$

where

$$Q_{N,r}(x) := \mathbb{P}\Big\{ \frac{1}{\sigma_{\xi_r}\sqrt{N_{j_2+r}}} \sum_{k=1}^{N_{j_2+r}} (\xi_r(k) - \mathbb{E}\xi_r(k)) \le x \Big\},$$

$\Phi(x)$ denotes the standard Normal c.d.f., and where $A > 0$ is an absolute constant. This is so, provided that the variance $\sigma_{\xi_r}^2 := \mathrm{Var}(\xi_r(1))$ and the third moment $\mathbb{E}|\xi_r(1)|^3$ of the $\xi_r(k)$'s are finite.

Observe first that, by (4.12) and by the independence of the $\xi_r(k)$'s from R,

$$\sigma_{\xi_r}^2 = N_{j_2+r}\Big(\mathrm{Var}(\theta, Y_r) - \mathrm{Var}(\theta, R)\Big) = \sigma_\theta^2 + \mathcal{O}(1/2^{r\beta/\alpha}) + \mathcal{O}(2^r/N), \tag{4.14}$$

where σ_θ is as in (4.10). Indeed, this follows from Proposition 4.1 above, provided that $\mathrm{Var}(\theta, R)$ is negligible. In view of (4.11), however, since $0 \le q_j < 2^j \le 2^{j_2},\ j = j_1, \dots, j_2$,

$$\begin{aligned} \mathrm{Var}(\theta, R) &\le \frac{m^2 2^{j_2}}{N_{j_2+r}} \sum_{j=j_1}^{j_2} \mathrm{Var}(\log_2 D(j+r,1)) \\ &= \frac{m^2 2^{j_2}}{N_{j_2+r}} \sum_{j=j_1}^{j_2} \mathrm{Var}(\log_2(2^{-(j+r)/\alpha} D(j+r,1))), \end{aligned} \tag{4.15}$$

where $m = j_2 - j_1 + 1$. In the last relation, we used the inequality $\mathrm{Var}(\eta_1 + \cdots + \eta_m) \le m^2(\mathrm{Var}(\eta_1) + \cdots + \mathrm{Var}(\eta_m)),\ m \in \mathbb{N}$ and the fact that

$$\mathrm{Var}(\log_2 D(j+r,1)) = \mathrm{Var}(\log_2(2^{-(j+r)/\alpha} D(j+r,1))).$$

In view of (2.5), however, by Corollary 3.1 below, the variances on the r.h.s. of (4.15) are bounded, as $r \to \infty$. This implies that $\mathrm{Var}(\theta, R) = \mathcal{O}(2^r/N)$, which completes the proof of (4.14).

We now focus on bounding the term $\mathbb{E}|\xi_r(1) - \mathbb{E}\xi_r(1)|^3$ in (4.13). The inequality

$$\Big|\sum_{i=1}^m x_i\Big|^p \le m^{0\vee(p-1)}\sum_{i=1}^m |x_i|^p,\ m\in\mathbb{N},\quad \text{valid for all } p,\ x_i\in\mathbb{R},\ i=1,\dots,m, \tag{4.16}$$

implies

$$\begin{aligned}\mathbb{E}|\xi_r(1)-\mathbb{E}\xi_r(1)|^3 &\le m^2\sum_{j=j_1}^{j_2}|\theta_j|^3\mathbb{E}|y_{j+r}(1)-\mathbb{E}y_{j+r}(1)|^3\\ &\le m^2\sum_{j=j_1}^{j_2}|\theta_j|^3\mathbb{E}\Big|\frac{1}{2^{j_2-j}}\sum_{i=1}^{2^{j_2-j}}\log_2 D(j+r,i)-\mathbb{E}\log_2 D(j+r,1)\Big|^3\\ &\le m^2\sum_{j=j_1}^{j_2}\frac{|\theta_j|^3}{2^{j_2-j}}\sum_{i=1}^{2^{j_2-j}}\mathbb{E}|\log_2 D(j+r,i)-\mathbb{E}\log_2 D(j+r,1)|^3\\ &= m^2\sum_{j=j_1}^{j_2}|\theta_j|^3\mathbb{E}|\log_2 D(j+r,1)-\mathbb{E}\log_2 D(j+r,1)|^3, \end{aligned}\tag{4.17}$$

where $m = j_2 - j_1 + 1$ and where the last bound follows from the Jensen's inequality. As in (4.15) above, we have that $\log_2 D(j+r,1) - \mathbb{E}\log_2 D(j+r,1)$ equals

$$\log_2(2^{-(j+r)/\alpha}D(j+r,1)) - \mathbb{E}\log_2(2^{-(j+r)/\alpha}D(j+r,1)),$$

Therefore, by using inequality (4.16), we get that the r.h.s. of (4.15) is bounded above by

$$4m^2\sum_{j=j_1}^{j_2}|\theta_j|^3\Big(\mathbb{E}|\log_2(2^{-(j+r)/\alpha}D(j+r,1))|^3 + (\mathbb{E}|\log_2(2^{-(j+r)/\alpha}D(j+r,1)|)^3\Big).$$

The last term is bounded, as $r\to\infty$, in view of (2.5) and Corollary 3.1.

We have thus far shown that (4.13) holds with the r.h.s. being of order $\mathcal{O}(1/\sqrt{N_r})$, uniformly in r, that is,

$$\sup_{x\in\mathbb{R}}\Big|Q_{N,r}(x)-\Phi(x)\Big| \le C_\theta/\sqrt{N_r} = \mathcal{O}\Big(2^{r/2}/\sqrt{N}\Big). \tag{4.18}$$

We will now use this fact to prove (4.9).

Step 2. By (4.12), the probability in (4.9) equals

$$\mathbb{E}Q_{N,r}\Big(x/\sigma_{\xi_r} - \sqrt{N_{j_2+r}}((\theta,R)+\mathbb{E}\xi_r(1)-(\theta,\mu_r))/\sigma_{\xi_r}\Big) =: \mathbb{E}Q_{N,r}\Big(x/\sigma_{\xi_r}-\Delta_{N,r}\Big). \tag{4.19}$$

Indeed, this follows from the independence of the $\xi_r(k)$'s and the remainder term R.

Now, by applying the triangle inequality, we obtain that the l.h.s. of (4.9) is bounded above by:

$$\sup_{x\in\mathbb{R}} \mathbb{E}\Big|Q_{N,r}(x/\sigma_{\xi_r} - \Delta_{N,r}) - \Phi(x/\sigma_{\xi_r} - \Delta_{N,r})\Big| + \sup_{x\in\mathbb{R}} \mathbb{E}\Big|\Phi(x/\sigma_{\xi_r} - \Delta_{N,r}) - \Phi(x/\sigma_{\xi_r})\Big|$$
$$+ \sup_{x\in\mathbb{R}} |\Phi(x/\sigma_{\xi_r}) - \Phi(x/\sigma_\theta)| =: A_1 + A_2 + A_3. \tag{4.20}$$

In view of (4.18), we have that

$$A_1 \le \sup_{x\in\mathbb{R}} |Q_{N,r}(x) - \Phi(x)| = \mathcal{O}\Big(2^{r/2}/\sqrt{N}\Big), \tag{4.21}$$

as $N \to \infty$ and $N/2^r \to \infty$.

Now, focus on the term A_2 in (4.20). By using the mean value theorem, for any $a < b,\ a, b \in \mathbb{R}$, we have that $|\Phi(a) - \Phi(b)| \le |a - b|/\sqrt{2\pi}$. Therefore (see (4.19)),

$$A_2 \le \frac{1}{\sqrt{2\pi}} \mathbb{E}|\Delta_{N,r}| \le \frac{\sqrt{N_{j_2+r}}}{\sqrt{2\pi}\sigma_{\xi_r}} \Big(\mathbb{E}|(\theta, R)| + \mathbb{E}|\xi_r(1) - (\theta, \mu_r)|\Big). \tag{4.22}$$

As argued above, in view of (4.11), we obtain by the triangle inequality, that

$$\begin{aligned} \mathbb{E}|(\theta, R)| &\le \frac{\text{const}}{N_{j_2+r}} \sum_{j=j_1}^{j_2} \mathbb{E}|\log_2 D(j + r, 1)| \\ &\le \frac{\text{const}}{N_{j_2+r}} \sum_{j=j_1}^{j_2} \mathbb{E}|\log_2(2^{-(j+r)/\alpha} D(j + r, 1))| + \text{const}\,\frac{r}{N_{j_2+r}} = \mathcal{O}(r/N_r). \end{aligned} \tag{4.23}$$

The last relation follows by adding and subtracting the term $(j + r)/\alpha$, and by applying Corollary 3.1 to the terms $\mathbb{E}|\log_2(2^{-(j+r)/\alpha} D(j + r, 1))|$.

By (4.11), $\mathbb{E}\xi_r(1) = \mathbb{E}(\theta, Y_r) - \mathbb{E}(\theta, R)$ and thus by applying the triangle inequality, Proposition 4.1 and Relation (4.23), to the second term in the r.h.s. of (4.22), we obtain

$$A_2 \le \text{const} \sqrt{N_r}\Big(r/N_r + 1/2^{r\beta/\alpha}\Big) = \mathcal{O}\Big(r2^{r/2}/N\Big) + \mathcal{O}\Big(1/2^{r\beta/\alpha}\Big). \tag{4.24}$$

Here, we also used the fact that $\sigma_{\xi_r} \to \sigma_\theta,\ \sigma_\theta > 0$, as $r \to \infty$ (see (4.14) above).

Consider now the term A_3 in (4.20). As above, by using the mean value theorem, we obtain

$$\begin{aligned} A_3 &\le \text{const}\,|1/\sigma_\theta - 1/\sigma_{\xi_r}| = \text{const}\,\frac{|\sigma_\theta - \sigma_{\xi_r}|}{\sigma_\theta \sigma_{\xi_r}} \\ &= \mathcal{O}(1/2^{r\beta/\alpha}) + \mathcal{O}(2^r/N), \end{aligned} \tag{4.25}$$

as $r \to \infty$ and $N/2^r \to \infty$, where the last inequality follows form Relation (4.14) above and the fact that $\sigma_\theta^2 - \sigma_{\xi_r}^2 = (\sigma_\theta - \sigma_{\xi_r})(\sigma_\theta + \sigma_{\xi_r})$.

Now, by combining the bounds in Relations (4.20), (4.21), (4.24) and (4.25), we obtain (4.9). This completes the proof of the theorem. □

Let now the scales $j_1 \le j_2$ be fixed and let $r = r(N) \in \mathbb{N},\ r + j_2 \le [\log_2 N]$. Theorem 4.1 shows that one can obtain consistent and asymptotically normal estimators of H and $C = C(\sigma_0, \alpha)$, as in the ideal Fréchet case (2.13). Indeed, let $A = (a\ b)$ be as in (2.13) and define $\widehat{\theta}_{\Sigma_1} = (\widehat{H}_{\Sigma_1},\ \widehat{C}_{\Sigma_1})$ as in (2.13) and $\alpha^{-2}\Sigma_1$ being the asymptotic covariance matrix in Proposition 4.1.

By using (2.13), one can show that

$$\widehat{H} := \widehat{H}_{\Sigma_1} = \sum_{j=j_1}^{j_2} w_j Y_{j+r} \quad \text{and} \quad \widehat{C} := \widehat{C}_{\Sigma_1} = \sum_{j=j_1}^{j_2} v_j Y_{j+r} - r\widehat{H}_{\Sigma_1}, \tag{4.26}$$

where the w_j's and the v_j's are *fixed* weights such that

$$\sum_{j=j_1}^{j_2} j w_j = \sum_{j=j_1}^{j_2} v_j = 1 \quad \text{and} \quad \sum_{j=j_1}^{j_2} w_j = \sum_{j=j_1}^{j_2} j v_j = 0. \tag{4.27}$$

The following result establishes the asymptotic normality of these estimators.

Proposition 4.2 *Assume the conditions of Theorem 4.1 hold. If $r = r(N) \in \mathbb{N}$ is such that $r2^r/N + 1/2^{r\beta/\alpha} \to 0$, as $N \to \infty$, then for the estimators defined in (4.26), we have*

$$\sqrt{N_{j_2+r}}(\widehat{H} - H) \stackrel{d}{\longrightarrow} \mathcal{N}(0, H^2 c_w) \quad \textit{and} \quad \sqrt{N_{j_2+r}/r}(\widehat{C} - C) \stackrel{d}{\longrightarrow} \mathcal{N}(0, H^2 c_w), \tag{4.28}$$

as $N \to \infty$, where $c_w = \sum_{i,j=j_1}^{j_2} w_i w_j \Sigma_1(i,j)$ and where $C = C(\sigma_0, \alpha)$ is as in (2.12). Moreover,

$$\lim_{N\to\infty} N_{j_2+r}\mathrm{Var}(\widehat{H}) = \lim_{N\to\infty} r^{-1} N_{j_2+r}\mathrm{Var}(\widehat{C}) = H^2 c_w.$$

Proof: The first convergence in (4.28) follows directly from Theorem 4.1 by setting $\theta_j := w_j,\ j = j_1, \ldots, j_2$. Indeed, since $\mu_r(j) = (j + r)/\alpha + C$, Relation (4.27) implies that

$$(\theta, \mu_r) = \sum_{j=j_1}^{j_2} w_j((j + r)/\alpha + C) = 1/\alpha \equiv H.$$

Thus, for $\widehat{H} = (\theta, Y_r) = \sum_{j=j_1}^{j_2} w_j Y_{j+r}$, by Relation (4.9), we obtain that

$$\sup_{x\in\mathbb{R}} |\mathbb{P}\{\sqrt{N_{j_2+r}}(\widehat{H} - H) \le x\} - \Phi(x/\sigma_w)| \longrightarrow 0,$$

as $N \to \infty$. This implies the asymptotic normality of $\widehat{H}$ in (4.28), where in view of (4.10) $\sigma_w^2 = H^2(w, \Sigma_1 w) = H^2 \sum_{i,j=j_1}^{j_2} w_i w_j \Sigma_1(i,j)$.

We now focus on the estimator $\widehat{C}$. By setting $\theta_j := v_j,\ j = j_1, \dots, j_2$, we get by using (4.27) that

$$(\theta, \mu_r) = \sum_{j=j_1}^{j_2} ((j+r)/\alpha + C)v_j = r/\alpha + C.$$

On the other hand, in view of (4.26),

$$(\theta, Y_r) = \sum_{j=j_1}^{j_2} v_j Y_{j+r} = \widehat{C} + r\widehat{H}$$

and thus

$$\widehat{C} - C = (\theta, Y_r) - (\theta, \mu_r) - r(\widehat{H} - H). \tag{4.29}$$

We have already shown that the term $(\widehat{H} - H)$ above is asymptotically normal and by Theorem 4.1 the term $(\theta, Y_r) - (\theta, \mu_r)$ in (4.29) is also asymptotically normal. Since $r = r(N) \to \infty$, the second term in the r.h.s. of (4.29) dominates in the limit. This implies that second convergence in (4.28).

To complete the proof, observe that by Proposition 4.1, $N_{j_2+r}\mathrm{Var}(\widehat{H}) \to \sigma_w^2 = H^2 c_w$, as $N \to \infty$. We now consider the variance of $\widehat{C} - C$ in (4.29), and apply the inequality

$$\mathrm{Var}(\xi) - 2(\mathrm{Var}(\xi)\mathrm{Var}(\eta))^{1/2} + \mathrm{Var}(\eta) \le \mathrm{Var}(\xi - \eta) \le \mathrm{Var}(\xi) + 2(\mathrm{Var}(\xi)\mathrm{Var}(\eta))^{1/2} + \mathrm{Var}(\eta)$$

with $\xi := (\theta, Y_r) - (\theta, \mu_r)$ and $\eta := r(\widehat{H} - H)$. Since $\mathrm{Var}(\eta)$ dominates $\mathrm{Var}(\xi)$, in the limit, we obtain that $r^{-1} N_{j_2+r}\mathrm{Var}(\widehat{C}) \to \sigma_w^2 = H^2 c_w$, as $N \to \infty$. □

Corollary 4.1 *Assume the conditions of Theorem 4.1 hold. Define the estimators*

$$\widehat{\alpha} := 1/\widehat{H} \quad \textit{and} \quad \widehat{\sigma}_0 := 2^{\widehat{C} - (\mathbb{E}\log_2 Z)/\widehat{\alpha}},$$

where Z is a $1-$Fréchet random variable with unit scale coefficient. Then with $r = r(N)$ as in Proposition 4.2, we have

$$\sqrt{N_{j_2+r}}(\widehat{\alpha} - \alpha) \stackrel{d}{\longrightarrow} \mathcal{N}(0, \alpha^2 c_w) \quad \textit{and} \quad \sqrt{N_{j_2+r}/r}(\widehat{\sigma}_0 - \sigma_0) \stackrel{d}{\longrightarrow} \mathcal{N}(0, (\ln 2)^2 \sigma_0^2 \alpha^{-2} c_w). \tag{4.30}$$

This result follows from Proposition 4.2 by an application of the Delta-method.

Most heavy–tailed distributions used in applications satisfy Condition 3.1, but some do not satisfy Condition 3.2. Indeed, (3.4) implies that $\mathbb{E}|X|^p 1_{\{X\le 1\}} < \infty$, for all $p \in \mathbb{R}$, which is rather stringent. Nevertheless, the results of Proposition 4.2 and Corollary 4.1 continue to hold even if Condition 3.2 is not satisfied and even if the $X(i)$'s can take negative values. This is so, because block–maxima become strictly positive as the block–size grows. We make this more precise in Proposition 4.3 below.

Now, for convenience, introduce a special value $*$ and suppose that our statistics take values in the extended real line $\mathbb{R}^* := \mathbb{R} \cup \{*\}$. If a statistic is not well–defined (because it involves $\log_2 x$ for $x \le 0$, for example), we assign to it the special value $*$. The set $\{*\} \subset \mathbb{R}^*$ is considered as both closed and open in the topology of $\mathbb{R}^*$ and the topology of $\mathbb{R} \subset \mathbb{R}^*$ is the same as that of the real line. Therefore, the statistics Y_j in (2.7) and the estimators $\widehat{H}$ and $\widehat{C}$ in (4.26), become proper random variables which can sometimes take the value $*$ if some of the $X(i)$'s are negative.

The following result shows that, asymptotically, the estimators $\widehat{H}$ and $\widehat{C}$ become real–valued with probability one, provided that $\ln(N)/2^{r(N)} \to 0$, as $N \to \infty$.

Proposition 4.3 *Suppose that the c.d.f. F has the representation* (3.1) *and satisfies Condition 3.1, where $F(0)$ is not necessarily zero. Let also $r = r(N) \in \mathbb{N}$, $\widehat{H}$ and $\widehat{C}$ be as in* (4.26)*. If $\ln(N)/2^{r(N)} \longrightarrow 0,\ N \to \infty$, then*

$$\mathbb{P}(\{\widehat{H} = *\}) + \mathbb{P}(\{\widehat{C} = *\}) \longrightarrow 0, \quad \text{as } N \to \infty. \tag{4.31}$$

If in addition $r2^r/N + 1/2^{r\beta/\alpha} \to 0$, as $N \to \infty$, then the convergences (4.28) *and* (4.30) *continue to hold.*

Proof: Let $X(i),\ i \in \mathbb{N}$ be i.i.d. with c.d.f. F and let $x_0 > 0$ be arbitrary. Define the truncated variables $\widetilde{X}(i) := X(i)1_{\{X(i)>x_0\}} + x_0 1_{\{X(i)\le x_0\}},\ i \in \mathbb{N}$ and observe that they are i.i.d. with c.d.f. $\widetilde{F}(x) := F(x),\ x \ge x_0$ and $\widetilde{F}(x) = 0,\ x < x_0$. Thus, $\widetilde{F}(x)$ has a representation as in (3.1) with the function $\sigma^\alpha(x)$ replaced by

$$\widetilde{\sigma}^\alpha(x) = \infty 1_{(-\infty,x_0)}(x) + \sigma^\alpha(x) 1_{[x_0,\infty)}(x),$$

where $\sigma^\alpha(x)$ is the function involved in the corresponding representation of $F(x)$.

Consider the statistics $\widetilde{D}(j,k)$ and $\widetilde{Y}_j$ defined as in (2.5) and (2.7) with $X(i)$'s replaced by $\widetilde{X}(i)$'s. Let also $\widetilde{H}$ and $\widetilde{C}$ be the corresponding statistics defined as in (4.26) with Y_j's replaced by $\widetilde{Y}_j$'s. Observe that $\widetilde{F}$ satisfies Condition 3.1 and also trivially Condition 3.2 since $x_0 > 0$ and $\widetilde{\sigma}^\alpha(x) = \infty$ for all $x \in (0, x_0)$. Therefore, the results of Proposition 4.2 apply to the statistics $\widetilde{H}$ and $\widetilde{C}$. We will now show that the statistics $\widehat{H}$ and $\widehat{C}$, which may not be always real–valued random variables (i.e. can take the special value $*$) coincide with the statistics $\widetilde{H}$ and $\widetilde{C}$, eventually.

Let $1 \le j_0 \le \log_2 N,\ j \in \mathbb{N}$. Observe that the event

$$\mathcal{C}_{j_0} := \{\widetilde{D}(j_0,k) = D(j_0,k),\ k = 1,\dots,N_{j_0}\}$$

implies the events $\mathcal{C}_j = \{\widetilde{D}(j,k) = D(j,k),\ k = 1,\dots,N_j\}$, for all $j_0 \le j \le \log_2 N$ and in particular the events $\{\widetilde{Y}_j = Y_j\},\ j \ge j_0$. Thus, the statistics $\widetilde{H}$ and $\widehat{H}$ (and $\widetilde{C}$ and $\widehat{C}$, respectively) coincide on the event $\mathcal{C}_{j_1+r}$. Thus, to complete the proof of the proposition, it is sufficient to show that $\mathbb{P}(\mathcal{C}_{j_1+r}) \to 1$, as $N \to \infty$.

Let $j_0 := j_1 + r$ and observe that by independence,

$$\mathbb{P}(\mathcal{C}_{j_0}) = \mathbb{P}\{\widetilde{D}(j_0,1) = D(j_0,1)\}^{N_{j_0}} = \Big(1 - F(x_0)^{2^{j_0}}\Big)^{N_{j_0}}.$$

In view of Condition 3.1, $p_0 := F(x_0) < 1$ and hence

$$\ln \mathbb{P}(\mathcal{C}_{j_0}) = N_{j_0} \ln(1 - p_0^{2^j}) = -\frac{N}{2^{j_0}} p_0^{2^{j_0}} (1 + o(1)), \quad \text{as } j_0 \to \infty.$$

Since $p_0 < 1$, the first convergence in (4.31) implies that $N p_0^{2^{j_1+r(N)}} \to 0$, as $N \to \infty$, and hence $\mathbb{P}(\mathcal{C}_{j_1+r(N)}) \to 1$, as $N \to \infty$. We have thus shown that (4.28) holds. Relation (4.30) follows from (4.28) by using the Delta–method. $\square$

Remarks:

1. Observe that in view of (2.13), $\widehat{H}_{\Sigma_1} = \widehat{H}_{\phi\Sigma_1}$ and $\widehat{C}_{\Sigma_1} = \widehat{C}_{\phi\Sigma_1}$, for any $\phi > 0$. That is, one can compute, in practice, the generalized least squares estimators $\widehat{H}$ and $\widehat{C}$ without having to use a plug–in estimator for α in (4.2) (see also the Remarks in Section 2.2).
2. The constants c_w appearing in Proposition 4.2 and Corollary 4.1 are given in Table 7.2 below. We now comment on the optimal rate in these asymptotic results.

   Proposition 4.1 indicates that the bias of the estimator $\widehat{H}$ in (4.28) is of order $\mathcal{O}(1/2^{r\beta/\alpha})$. On the other hand, the standard error of $\widehat{H}$ is of order $\mathcal{O}(2^r/N)$. By balancing these orders, we obtain that

   $$2^r = 2^{r(N)} \propto N^{\alpha/(2\beta+\alpha)}$$

   yields the *optimal order* of the mean squared error (m.s.e.) $\mathbb{E}(\widehat{H} - H)^2$, and a corresponding rate of convergence

   $$2^{r/2}/\sqrt{N} = \mathcal{O}(1/N^{\beta/(2\alpha+\beta)})$$

   to the limit distribution of $\widehat{H}$ in (4.28).

   Hall (1982) (see Theorem 2 therein) obtained the same optimal order of convergence for the Hill–type estimators under the following semi–parametric assumptions on the tail of F:

   $$1 - F(x) = c_1 x^{-\alpha}(1 + c_2 x^{-\beta} + o(x^{-\beta})), \quad \text{as } x \to \infty,\ \alpha, \beta > 0. \tag{4.32}$$

   A Taylor expansion shows that this tail behavior corresponds to Condition 3.1 above in the case when $0 < \beta \le \alpha$. *Note that in Hall (1982) the parameter r corresponds to $N/2^r$ in our case.*

   Observe that Theorems 1 and 2 in Hall (1982) involve also asymptotic normality results for the scale parameter c_1 in (4.32). These results are similar to those about $\widehat{C}$ in Proposition 4.2. Note in particular the presence of the logarithmic in N factor $r = r(N)$.
3. The optimal rate in the previous remark may not be improved, in general. Indeed, by Proposition 3.1 the rate of the bias is exact if $\sigma^\alpha(x) - \sigma_0^\alpha \sim c_1 x^{-\beta},\ x \to \infty,\ c_1 \neq 0$. This is typically the case in practice (see the Examples above). Relation (4.2) also implies that the order of the variance of $\widehat{H}$ is precisely $\mathcal{O}(1/\sqrt{N_r})$, and cannot be improved.

Furthermore, the rate in the Berry–Esseen bound may not be improved, in general (see e.g. Ch. V.2 in Petrov (1995)). Thus, the result of Theorem 4.1 is optimal in our setting.

4. Consider the case of optimal m.s.e. of $\widehat{H}$, that is, $2^r \propto N^{\alpha/(2\beta+\alpha)}$. Observe that the r.h.s. in (4.9) is up to the logarithmic in N factor of $r(N)$ of the same order as the root–m.s.e. $(\mathbb{E}(\widehat{H}-H)^2)^{1/2}$. This indicates that the precision (in terms of coverage probability) of the confidence intervals for H based on the asymptotic distribution for $\widehat{H}$ will be of order at least $\mathcal{O}(1/N^{\beta'/(2\alpha+\beta)})$ for any $\beta' \in (0,\beta)$.
5. Even though the estimators $\widehat{\alpha}$ and $\widehat{\sigma}_0$ in Corollary 4.1 are asymptotically normal, it is not a good idea to use their asymptotic distributions to construct confidence intervals for α and σ_0. Indeed, for simplicity consider the ideal Fréchet case. In this case, the estimator $\widehat{H}$ is unbiased and hence the estimator $\widehat{\alpha} = 1/\widehat{H}$ is *biased*. Moreover, since the variance of the random variable $1/X$, where X has Normal distribution is infinite, we expect that $\mathrm{Var}(\widehat{\alpha})$ does not converge to the asymptotic variance of $\widehat{\alpha}$ in (4.28). In our experience, the distribution of $\widehat{\alpha}$ tends to be skewed in practice. Therefore, one can get better confidence interval estimates for α by using *inversion* from the corresponding confidence intervals for H. For example, $((\widehat{H}+z_p\widehat{H}\sqrt{c_w}/\sqrt{N_{j_2+r}})^{-1}, (\widehat{H}-z_p\widehat{H}\sqrt{c_w}/\sqrt{N_{j_2+r}})^{-1})$ is an asymptotically correct $100(1-p)\%$ confidence interval for α, where $z_p := \Phi^{-1}(1-p/2),\ p \in (0,1)$. As indicated in the previous remark the error in the coverage probability of this interval is of order $\mathcal{O}(1/N^{\beta'/(2\alpha+\beta)})$ for any $\beta' \in (0,\beta)$, if m.s.e.–optimal r's are chosen.

5. Performance evaluation and data analysis

5.1. Typical models: small and large sample properties

We study the performance of the max self–similarity estimators when the data are heavy–tailed but deviate from the ideal Fréchet case. Specifically, given a sample of size $N = 2^n,\ n \in \mathbb{N}$, the GLS estimators $\widehat{H} = \widehat{H}(j_1, j_2)$ and $\widehat{\alpha} = \widehat{\alpha}(j_1, j_2) = 1/\widehat{H}$ are computed for a range of scales $j_1 \le j \le j_2$. We choose here $j_2 = n$ as the maximal available scale and focus on optimal j_1's in the sense of mean squared error. Namely, we let

$$j_1^{opt} := \underset{j_1,\, 1\le j_1\le j_2}{\mathrm{Argmin}}\ \mathbb{E}(\widehat{H}(j_1,j_2) - H)^2, \tag{5.1}$$

where the last expectation is computed from samples of independent realizations of the estimators $\widehat{H}$.

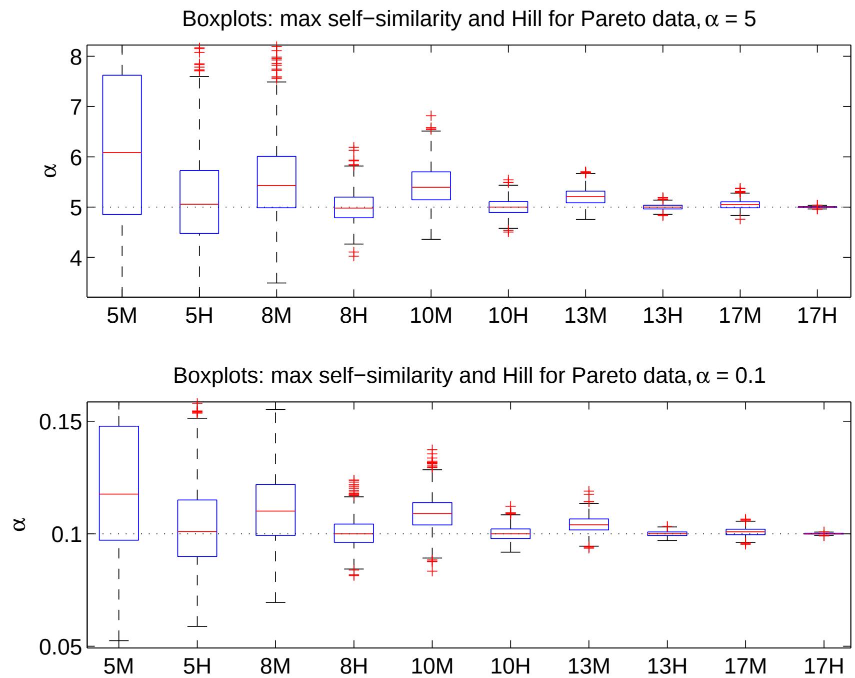

Fig 3. *Boxplots of* $1,000$ *independent realizations of max self–similarity and Hill estimators for different sample sizes from Pareto distributions with* $\alpha = 5$ *(top panel) and* $\alpha = 0.1$ *(bottom panel) are shown. The labels* nM *and* nH *correspond to sample size* 2^n *of max self–similarity and Hill estimators, respectively. The Hill estimators were computed by using* (1.2) *with* $k = 2^n - 1$*, and the max self–similarity estimators are based on a range of scales* $j_1 \le j \le j_2 = n$*, where* j_1 *was chosen to minimize the mean squared error.*

We first compare the max self–similarity estimators to the classical Hill estimator over Pareto data with unit scale, i.e. with c.d.f. $F(x) = 1 - x^{-\alpha},\ x \ge 1$. In this case, the Hill estimator corresponds to the maximum likelihood estimator. Figure 3 indicates that, as expected, the Hill estimator outperforms the max self–similarity estimator. However, as seen from the box–plots, the max self–similarity estimator works relatively well for small, moderate and large samples and essentially keeps up with the Hill estimators. In fact, as the sample size grows the max self–similarity estimator improves almost at the same rate as the Hill estimator. Here the max self–similarity estimator was computed by using the range of scales $j_1^{opt} \le j \le j_2$, where $j_2 = \log_2 N$ and j_1^{opt} is as in (5.1).

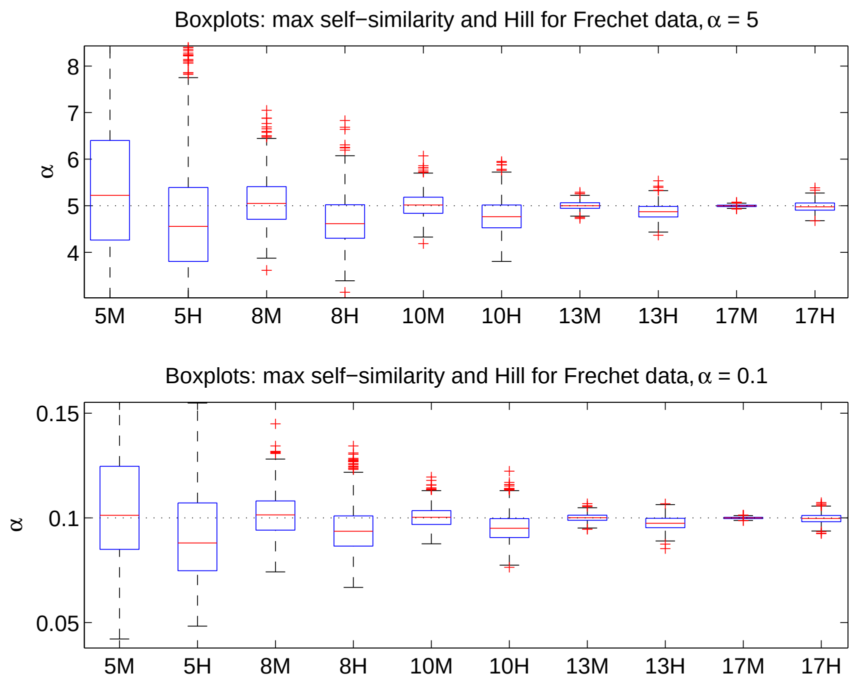

Fig 4. *Boxplots of* 1,000 *independent realizations of max self–similarity and Hill estimators for different sample sizes from Fréchet distributions with* $\alpha = 5$ *(top panel) and* $\alpha = 0.1$ *(bottom panel) are shown. The labels* nM *and* nH *correspond to sample size* 2^n *of max self–similarity and Hill estimators, respectively. The Hill estimator were computed by using an optimal value for* k *in* (1.2)*, which yields the smallest mean squared error. The max self–similarity estimators were computed from the entire range of scales* j*.*

In Figure 4, we compare the performance of the max self–similarity and the Hill estimators for Fréchet data. The parameter k in (1.2) of the Hill estimator was chosen to minimize the mean squared error of the statistics $1/\widehat{\alpha}_H(k)$, by analogy with (5.1). Now, the entire range of scales $j_1 = 1 \le j_2 = \log_2 N$ was used to compute the max self–similarity estimators. Observe that as compared to the case of Pareto data (see Figure 3), now the roles of the two estimators are reversed. As expected, the max self–similarity estimator works best in the Fréchet setting and dominates the Hill estimator. In fact, the method of choosing the parameter k here is unusually favorable to the Hill estimator since it is not based on examining and determining a range where the Hill plot is constant. It is well known that in practice, the Hill plot is quite volatile and the resulting choice of k based on this plot would yield far more biased estimators than the ones shown in Figure 4.

We now examine the max self–similarity estimators in more detail when the data are drawn from a stable and a $t-$distribution. Tables **??** and **??** below, indicate that the estimators $\widehat{H}_{opt} := \widehat{H}(j_1^{opt}, j_2)$ work well in practice for a variety of sample sizes and parameter values. Their performance is particularly good in the stable context. The performance in the case of $t-$distributions is comparable with the stable cases when the heavy–tail exponent α is not large. Notice that α corresponds to the degrees of freedom of the $t-$distribution and therefore as α grows, the $t-$distribution gets closer to the Normal distribution. Although it it still heavy tailed, most of the body of the distribution is not and therefore the quality of the tail estimators deteriorates.

Table **??** indicates that the max self-similarity estimator outperforms the Hill estimator for stable distributions with $\alpha \leq 1$ and that the two estimators are comparable for $1 < \alpha < 2$. The Hill estimator is slightly better than or comparable to the max self-similarity one for the t-distributions with low α's and slightly worse or comparable for moderate and large α's (Table **??**).

The MSE–optimal choice of the parameter k is unrealistically favorable to the Hill estimator. In practice, these choices of k typically do not correspond to constant regions in the Hill plot. On the other hand the MSE-optimal values of j_1 usually correspond to the knee in the max–spectrum plot, which can be identified in practice (either visually or automatically). These observations suggest that in reality the max self–similarity estimators are more reliable and accurate than estimators based on the Hill plot.

5.2. On the selection of the scales j_1 and j_2

In the ideal case of α-Frechet data, the max–spectrum plot of Y_j is almost perfectly linear in j (see Figure 2). However, most real data sets deviate from the ideal case and thus the max–spectrum becomes linear only over a range of relatively large scales j. The selection of an

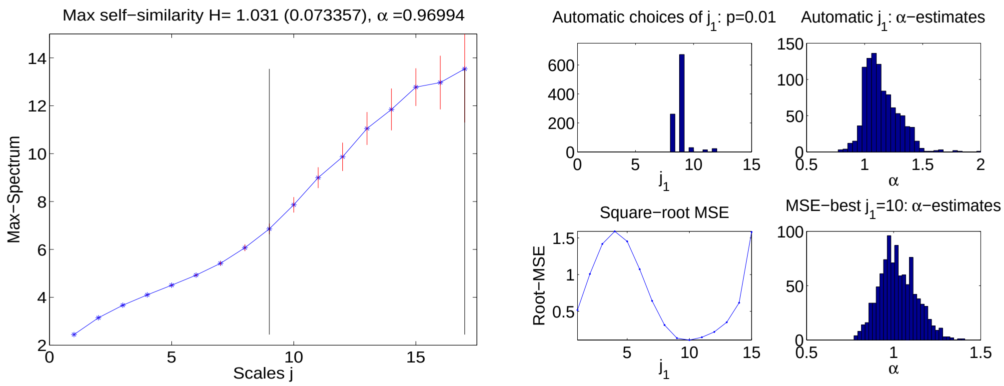

Fig 5. *Mixtures of $\alpha-$Fréchet (10%) and Exponential of mean 5 (90%) were simulated. The heavy–tail exponent is $\alpha = 1$ and the sample sizes are $N = 2^{17} = 131,072$. Left panel: max–spectrum of a typical sample. Right panel: 1,000 independent replications of the GLS max self–similarity estimators were obtained, where automatic selection for the parameter j_1 was used with $p = 0.01$ and $b = 4$. The top–right graph shows a histogram of the resulting selections of j_1. The bottom–right graph shows the root–mean squared error of the estimators $\widehat{H} = 1/\widehat{\alpha}$. The top–left and top–right plots shows histograms of the $\widehat{\alpha}$ estimates obtained by using automatically selected j_1's and with $j_1 = j_1^{opt} = 10$, respectively.*

appropriate range of scales $j_1 \le j \le j_2$, where the max self–similarity estimators are computed, becomes an important practical problem. Because of (2.8), one can always choose $j_2 = [\log_2 N]$ to be the largest available scale and the scale j_1 can be chosen by visual inspection, a strategy that work fairly well in practice. Nevertheless, we also propose an *automatic procedure* for choosing the scale j_1, which turns out to also work well in practice. It relies on the following simplifying assumptions:

Assumption 1. The vector $Y_j,\ j = 1,\dots,j_2$ follows a *multivariate Normal distribution.*

Assumption 2. The covariance matrix $\Sigma_\alpha(1, j_2; N) = \alpha^{-1}\Sigma_1(1, j_2; N)$ of the vector $Y = \{Y_j\}_{j=1}^{j_2}$ is given by (2.14).

These assumptions are valid asymptotically, provided that $N_{j_2} \to \infty$ (Theorem 4.1 and Proposition 4.1). Since the N_j's grow exponentially fast as j decreases, choosing j_2 as the largest available scale $[\log_2 N]$ is not critical in practice. Let now $\widehat{H}(j_1, j_2)$ denote the GLS estimate of $H = 1/\alpha$, computed over the range of scales $j_1 \le j \le j_2$ as in (2.13) (see also (4.26)).

Algorithm

Tunning parameters:

Pick a relatively small *significance threshold* $p \in (0,1)$ (e.g. $p = 0.1$ or 0.01) and an integer b called *back–start parameter* (e.g. $b = 3$ or 4 for moderate sample sizes). Set $j_2 := [\log_2 N]$ and $j_1 := \max\{1, j_2 - b\}$.

Step 1. If $j_1 = 1$ then stop, else calculate $\widehat{H}_{\rm new} = \widehat{H}(j_1 - 1, j_2)$ and $\widehat{H}_{\rm old} = \widehat{H}(j_1, j_2)$.

Step 2. Let $w_{\rm new}$ and $w_{\rm old}$ be vectors of weights as in (4.26), such that $\widehat{H}_{\rm new} = (w_{\rm new}, Y)$ and $\widehat{H}_{\rm old} = (w_{\rm old}, Y)$, where $Y = \{Y_j\}_{j=1}^{j_2} \in \mathbb{R}^{j_2}$ and where the vectors $w_{\rm new},\ w_{\rm old} \in \mathbb{R}^{j_2}$ are appropriately padded with zeros. Consider the quantity:

$$S_1 := \Big((w_{\rm new} - w_{\rm old}), \Sigma_1(1, j_2; N)(w_{\rm new} - w_{\rm old}) \Big)^{1/2}$$

Now, consider the approximate $(1-p)-$level confidence interval for $\mathbb{E}(\widehat{H}_{new} - \widehat{H}_{old})$:

$$\Big(\widehat{H}_{new} - \widehat{H}_{old} - z_{p/2} \widehat{H}_{\rm old} S_1, \widehat{H}_{new} - \widehat{H}_{old} + z_{p/2} \widehat{H}_{\rm old} S_1 \Big),$$

where $z_{p/2} = \Phi^{-1}(1 - p/2)$ is a $(1-p/2)-$th quantile of the standard Normal distribution.

Step 3. If zero is contained in the confidence interval computed in *Step 2*, then set $j_1 := j_1 - 1$ and go to *Step 1* otherwise stop and report the selected j_1 and $\hat{\alpha} := 1/\widehat{H}_{\rm old}$.

The choice of tunning parameters p and b and the validity of the above simplifying assumptions is addressed in Stoev et al. (2006). In Figure 5, we briefly demonstrate the performance of the above automatic selection procedure for a mixture of an Exponential and an $\alpha-$Fréchet distributions. Samples of size $N = 2^{17} = 131,072$ were generated and a level $p = 0.01$ and back–start parameter $b = 4$ employed. The left panel indicates the presence of a “knee” in the max–spectrum plot in one such mixture sample. The automatic selection procedure identified well the location of the knee by selecting $j_1 = 9$ and the resulting estimate $\widehat{\alpha} = 0.97$ is rather close to the nominal value of $\alpha = 1$. In the right panel, we demonstrate the performance of the automatic selection procedure by using $1,000$ independent replications of the mixture samples.

The histogram of the automatic choices for j_1 (left panel) indicates that most of the times values close to the MSE–optimal one $j_1^{opt} = 10$ were chosen. The histogram of the resulting estimates of the heavy–tail exponent (top–right graph in the left panel) is similar to the histogram corresponding to the MSE–optimal choice of j_1 (bottom–right in the left plot). The slight bias in the histogram on the top–right is due to the fact that often slightly lower than the MSE–optimal values of j_1 were chosen by the automatic procedure. More extensive analysis of this procedure is presented in Stoev et al. (2006).

5.3. Data analysis

We first discuss a popular insurance data set of $2,167$ fire losses in Denmark from 1980 to 1990. This data set has been studied extensively, see e.g. McNeil (1997), Resnick (1997*a*), Lu and Peng (2002) and Peng and Qi (2004).

Figure 6 displays the data, its corresponding Hill plot (bottom left) and its max–spectrum (bottom right). The max–spectrum yields an estimate $\widehat{\alpha} = 1.66$ obtained with an automatic selection of the scale j_1 by using a tunning parameter $p = 0.01$ (see Section 5.2), and the Hill plot yields an estimate $\widehat{\alpha}_H(k) = 1.39$ for $k = 1,000$. This discrepancy between the two methods is interesting since they yield comparable results in many typical models (see Section 5.1, above). To explore further the significance of this difference, we resort to calculating confidence intervals.

A particular advantage of the max–spectrum type estimators is that one can naturally obtain the following two types of confidence intervals for the parameters H and $\alpha = 1/H$: *(i)* based on the asymptotic normal distribution (see Proposition 4.2) and *(ii)* based on a *permutation bootstrapping* procedure. We will only briefly describe the procedure for obtaining permutation bootstrap confidence intervals. Its theoretical analysis is outside the scope of the present paper.

Permutation bootstrap confidence intervals

Given an i.i.d. sample $X(1), \dots, X(N)$, generate M independent random permutations $\pi_i : \{1, \dots, N\} \to \{1, \dots, N\}$, $i = 1, \dots, M$. Then, construct the *permuted samples*

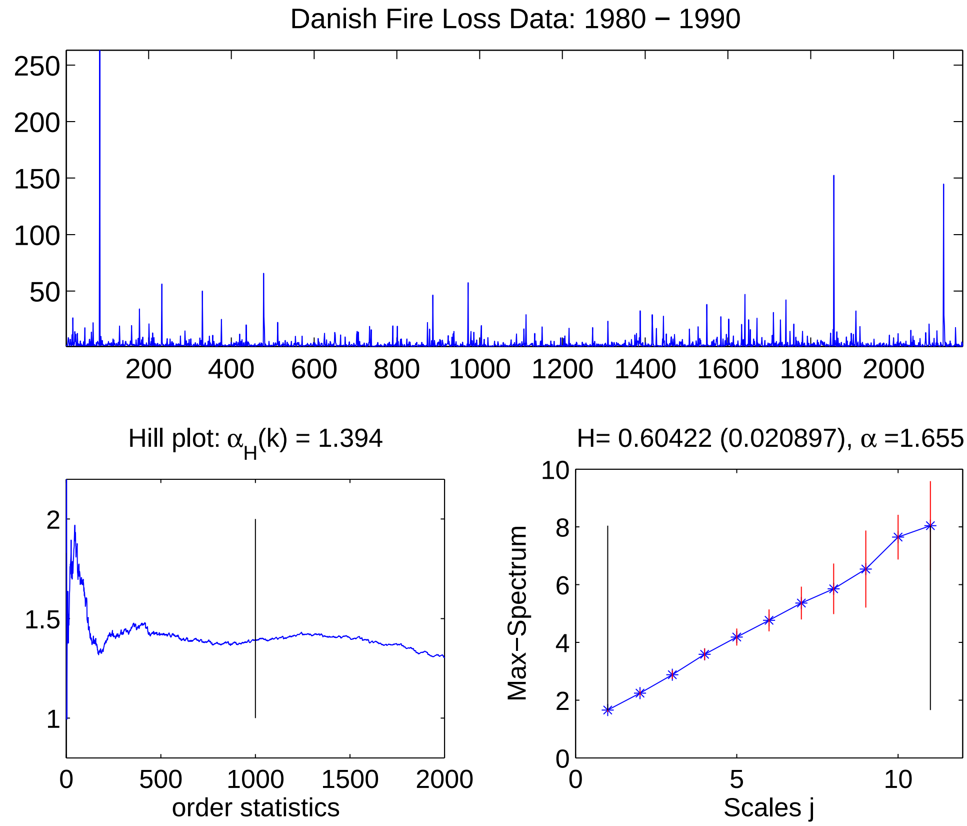

Fig 6. *Top panel: time series of insurance losses due to fire in Denmark from 1980 to 1990 losses (in million Danish krones). Bottom left panel: the Hill plot of the fire loss data set. Bottom right: the max–spectrum of the data. Note that the Hill estimate is $\widehat{\alpha}_H(k) = 1.39$, with $k = 1,000$ and the max self–similarity estimate is $\widehat{\alpha} = 1.66$.*

$\widetilde{X}_i(1), \dots, \widetilde{X}_i(N),\ i = 1, \dots, M$, where $\widetilde{X}_i(k) = X(\pi_i(k)),\ k = 1, \dots, N$. Fix a range of scales $j_1 < j_2 \le \log_2 N$ and for each $i = 1, \dots, M$, compute the GLS max self–similarity estimator $\widehat{H}_i = \widehat{H}_i(j_1, j_2)$, from the permuted sample $\widetilde{X}_i(1), \dots, \widetilde{X}_i(N)$. We will refer to the sample $\widehat{H}_i,\ i = 1, \dots, M$ as to the *permutation bootstrap* sample of the estimator $\widehat{H} = \widehat{H}(j_1, j_2)$, based on the original data set $X(1), \dots, X(N)$.

Observe that the statistics $\widehat{H}_i,\ i = 1, \dots, M$ are mutually dependent, since they are based on the original sample $X(1), \dots, X(N)$. However, since the $X(k)$'s are i.i.d. and the permutations π_i's are independent, we have that $\widehat{H}_i =^d \widehat{H}(j_1, j_2)$, for all $i = 1, \dots, M$. One has moreover that the sequence $\widehat{H}_i,\ i = 1, \dots, M$ is *exchangeable*. This suggests using the permutation bootstrap

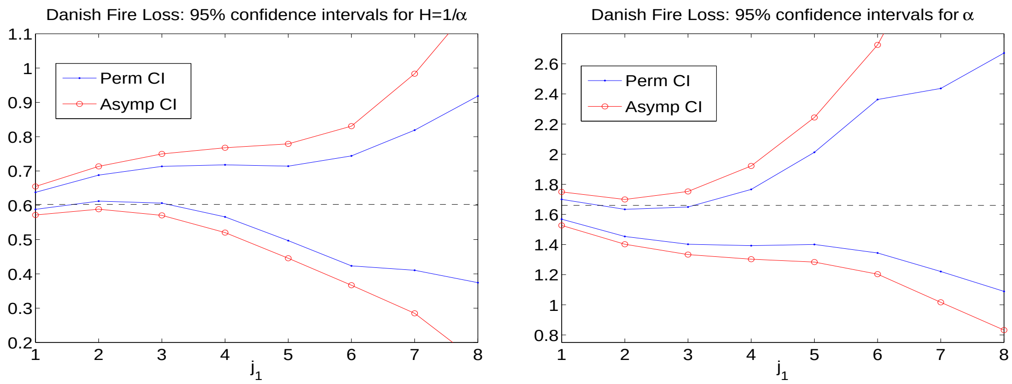

FIG 7. *Left panel: 95% confidence intervals for $H = 1/\alpha$ based on: (i) permutation–bootstrap from 10,000 independent permutations and (ii) asymptotic distribution for the max self–similarity estimators. Right panel: 95% confidence intervals for $\alpha = 1/H$ obtained by inverting the confidence intervals in the left panel. The horizontal lines indicate the estimated value of $\widehat{H} = 0.6$ and $\widehat{\alpha} = 1/\widehat{H} = 1.66$ for H and α, respectively, obtained with the max self–similarity estimator in Figure 6.*

sample $\widehat{H}_1, \ldots, \widehat{H}_M$ as a proxy to the sampling distribution of $\widehat{H}$. We thus propose to use the empirical confidence interval based on the permutation bootstrap sample as a confidence interval for H. Corresponding bootstrap confidence intervals for $\alpha = 1/H$ are obtained through the inversion method.

Experience with several simulation experiments suggests the following conjecture.

Conjecture 5.1 *Let $\widehat{H}_i$, $i = 1, \ldots, M$ be a permutation bootstrap sample of the estimator $\widehat{H}(j_1, j_2)$. Consider the scales j_1, j_2 and the permutation sample size M as functions of the sample size N, which tend to infinity as $N \to \infty$.*

Under certain conditions on the rates of growth of j_1, j_2 and M, the empirical distribution of the permutation bootstrap sample $\widehat{H}_i$, $i = 1, \ldots, M$ yields asymptotically consistent confidence intervals for $\widehat{H}$.

Figure 7 displays 95% confidence intervals for H (left panel) and $\alpha = 1/H$ (right panel) for the Danish fire loss data. Different scales j_1 were used and j_2 was chosen as the largest available scale 11. The permutation confidence intervals (denoted by dots) are obtained from $M = 10,000$ random permutations and the asymptotic confidence intervals (denoted by circles) are obtained from the asymptotic variance in Proposition 4.2 where the unknown value of H was replaced by $\widehat{H}$. To be able to compare the two types of intervals, we centered the asymptotic confidence intervals at the means of the permutation bootstrap samples $\widehat{H}_i,\ i = 1, \ldots, M$. Observe that although the two procedures for constructing confidence intervals are different, they yield very similar results. The permutation bootstrap intervals are always slightly more narrow than the asymptotic ones. As Figure 6 indicates, the use of scales $j_1 = 1$ and $j_2 = 11$ is acceptable. The resulting permutation and asymptotic confidence intervals for H are: $[0.5880, 0.6361]$ and $[0.5710, 0.6540]$, respectively. They are consistent with, but considerably tighter than the likelihood–based intervals in Figure 8 of Lu and Peng (2002) for the same data set. This can be contributed to the fact that the max–spectrum estimators and the Hill–type estimators are based on different principles. The performance of the permutation bootstrap and asymptotic confidence intervals is addressed in more detail in Stoev et al. (2006).

The second data set to be analyzed in this section consists of the volumes in trillion cubic feet of the 406 largest natural gas world provinces. The data were obtained from Table 1 in (n.d.). The study of the patterns in such data will help in the development of future natural gas resources leading to better assessments of the reserve growth potential of the world's provinces. The max self–similarity estimator, obtained from a typical randomly permuted sample is $\widehat{\alpha} = 1.284$ (Figure 8). Observe that the Hill plot shown in the bottom–left panel of Figure 8 is very volatile and appears to stabilize in a narrow range around $k = 60$, where the resulting estimator is $\widehat{\alpha}_H(60) = 0.826$. Notice that the integer nature of the observations makes the Hill plot exhibit a saw-tooth like pattern and hence difficult to obtain a good estimate for α. Due to the discrepancy between the two methods, obtaining confidence intervals becomes particularly pertinent.

Permutation bootstrap and asymptotic confidence intervals for the max self–similarity estimators for $H = 1/\alpha$ and α are presented in Figure 9. As in Figure 9, the asymptotic confidence intervals are slightly wider than the ones based on the permutation bootstrap. Observe that, contrary to the case of fire loss data in Figure 7, the locations of the confidence intervals for the gas data set stabilize only at scales $j \geq 4$. This indicates that the value $\widehat{\alpha} = 1.284$, obtained

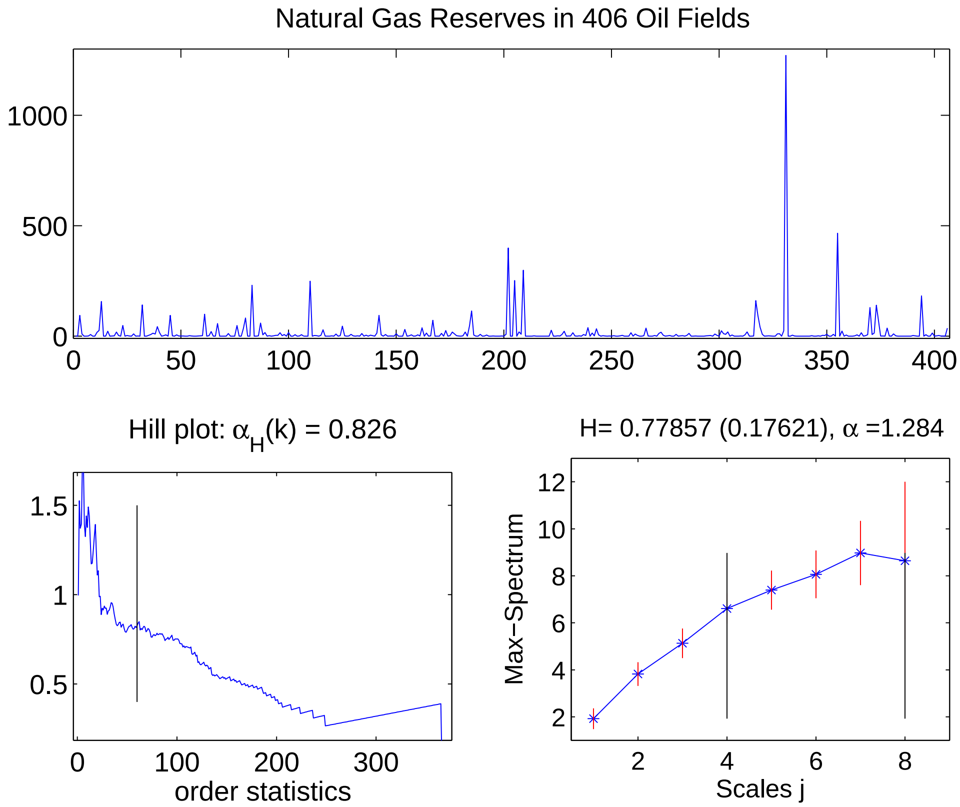

Fig 8. *Top panel: randomly permuted sample of volumes natural gas reserves (in trillion cubic feet) found in* 406 *provinces. Bottom left panel: the Hill plot of the data set. Bottom right panel: the max–spectrum of the data. Note that the Hill estimate is* $\widehat{\alpha}_H(k) = 0.826$*, with* $k = 60$ *and the max self–similarity estimate is* $\widehat{\alpha} = 1.284$.

from the range of scales $j_1 = 4$ and j_2 in Figure 8 is credible. The fact that the resulting Hill estimate $\widehat{\alpha}_H(60) = 0.826$ is less than 1 appears to be not statistically significant, according to the confidence intervals in Figure 9, which is in line with the findings in de Sousa and Michailidis (2004). This last fact and the volatility of the Hill plot suggest that the max self–similarity estimators can be viewed as more reliable in this setting.

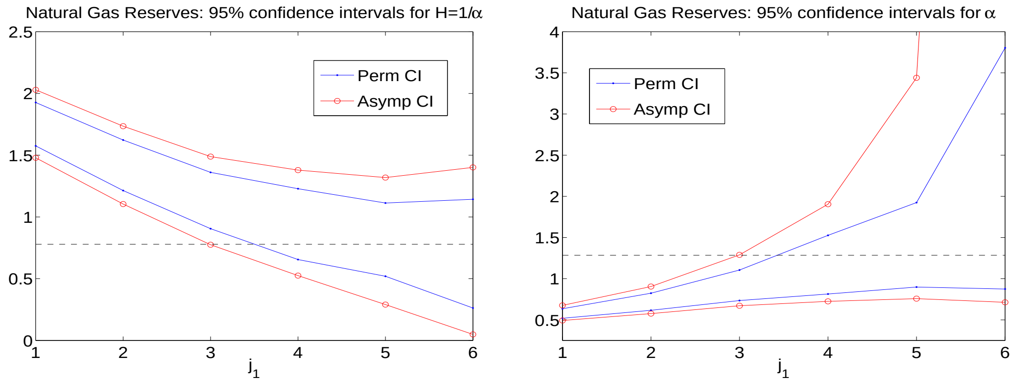

Fig 9. *Left panel:* 95% *confidence intervals for* $H = 1/\alpha$ *based on: (i) permutation–bootstrap from* 10,000 *independent permutations and (ii) asymptotic distribution for the max self–similarity estimators. Right panel:* 95% *confidence intervals for* $\alpha = 1/H$ *obtained by inverting the confidence intervals in the left panel. The horizontal lines indicate the estimated value of* $\widehat{H} = 0.78$ *and* $\widehat{\alpha} = 1/\widehat{H} = 1.28$ *for* H *and* α*, respectively, obtained with the max self–similarity estimator in Figure 8.*

6. Concluding remarks

In this paper, a new estimator for the tail exponent of a distribution was introduced and its asymptotic properties established. The estimator is based on block–maxima of the data and can be visualized through a new graphical device called the max–spectrum plot. Numerical work shows that compared to the widely used Hill estimator, the max self–similarity estimator performs competitively in the case of the Pareto distribution and it outperforms the Hill estimators in the cases of the stable, Fréchet and certain t-distributions. In practice, the max–spectrum plot is less volatile than the classical Hill plot. Thus, the max self–similarity estimator can be used in situations where the Hill plot fails or when it is hard to interpret. Finally, the fact that the estimator is based on block maxima makes it particularly suitable for time series data, a topic discussed in a companion paper Stoev et al. (2006).

Acknowledgments: The authors would like to thank Professor Robert Keener for suggesting the proof of Lemma 7.1 and for many useful suggestions. We also thank Kamal Hamidieh for stimulating discussions on the automatic selection procedure of scales. The work was supported in part by a grant from the Horace H. Rackham School of Graduate Studies at the University of Michigan (SS), and by NSF grants CCR-0325571, DMS-0505535 (GM) and DMS-0505747

(MT).

7. Appendix: auxiliary results and tables

7.1. Auxiliary results

We briefly review some properties of the $\alpha-$Fréchet distributions used above.

Definition 7.1 A random variable Z is said to have an $\alpha-$Fréchet distribution, if

$$\mathbb{P}\{Z \le x\} = \begin{cases} \exp\{-\sigma^\alpha x^{-\alpha}\} & , \ x > 0 \\ 0 & , \ x \le 0, \end{cases} \tag{7.1}$$

with $\sigma > 0$. The parameter σ is referred to as the scale coefficient of Z. The random variable Z is said to be standard $\alpha-$Fréchet if $\sigma = 1$.

Let Z be an $\alpha-$Fréchet variable with scale coefficient $\sigma > 0$. The next properties follow directly from Relation (7.1).

Properties

1. *(scale family)* For all $c > 0$, the random variable cZ is $\alpha-$Fréchet and has scale coefficient $c\sigma$.
2. *(heavy tails)* The Taylor expansion of the exponential around the origin implies that
$$\mathbb{P}\{Z > x\} = 1 - e^{-\sigma^\alpha x^{-\alpha}} \sim \sigma^\alpha x^{-\alpha}, \quad \text{as } x \to \infty. \tag{7.2}$$
3. *(moments)* In view of (7.2), for all $p > 0$,
$$\mathbb{E}Z^p < \infty \quad \text{if and only if} \quad p < \alpha.$$
One has moreover, that $\mathbb{E}Z^p = \sigma^p \Gamma(1-p/\alpha),\ p \in (0,\alpha)$, with $\Gamma(x) = \int_0^\infty u^{x-1}e^{-u}du,\ x > 0$.
4. *(log–moments)* For all $p > 0$, the moments $\mathbb{E}|\ln Z|^p$ are finite. This follows from the fact that $\xi := \alpha \ln(Z/\sigma)$ has the Gumbel distribution, i.e. $\mathbb{P}\{\xi \le x\} = \exp\{-e^{-x}\},\ x \in \mathbb{R}$. See also Corollary 3.1 below.

5. *(power transformations)* For any $p > 0$, the random variable Z^p is $\alpha/p-$Fréchet with scale coefficient σ^p. Consequently, if Z_1 is a standard $1-$Fréchet variable, then

$$Z := Z_1^{1/\alpha}$$

is standard $\alpha-$Fréchet, for all $\alpha > 0$.

The $\alpha-$Fréchet distributions are also *max–stable* in the following sense.

Definition 7.2 A random variable Z is said to be max–stable, if for all $a, b > 0$ there exist $c > 0,\ d \in \mathbb{R}$, such that

$$\max\{aZ', bZ''\} \stackrel{d}{=} cZ + d,$$

where Z' and Z'' are independent copies of Z and where $=^d$ means equality in distribution.

In particular, by (7.1), one gets that if $Z(1), \dots, Z(n),\ n \in \mathbb{N}$ are i.i.d. $\alpha-$Fréchet, then

$$Z(1) \vee \cdots \vee Z(n) \stackrel{d}{=} n^{1/\alpha} Z(1). \tag{7.3}$$

This last relation shows that a sequence of i.i.d. $\alpha-$Fréchet variables is also max self–similar with parameter $H = 1/\alpha$ (see Definition 2.1 above). Relation (7.3) served as the main motivation to define the max self–similarity estimators in Section 2 above.

The class of max–stable distributions in the sense of Definition 7.2 above includes, in addition to the Fréchet, only the classes of negative Fréchet and the Gumbel laws. These three classes of distributions are the only distributions arising in the limit of maxima of i.i.d. variables under appropriate normalization (see e.g. Proposition 0.3 in Resnick (1987) and also Leadbetter, Lindgren and Rootzén (1983)).

The following integration by parts formula is used in the proof of Theorem 3.1.

Lemma 7.1 *Let $f : [a, b] \to \mathbb{R},\ a, b \in \mathbb{R}$ be an absolutely continuous function, that is, $f(x) = f(a) + \int_a^x f'(u)du$, for some Lebesgue integrable $f'(x),\ x \in [a, b]$. Then, for any c.d.f. $G(x)$, we have*

$$\int_a^b f(x)dG(x) = f(b)G(b) - f(a)G(a) - \int_a^b G(x)f'(x)dx. \tag{7.4}$$

Proof: Since $f(x) = f(a) + \int_a^b f'(u)1_{[a,x)}(u)du$, we have that

$$\int_a^b f(x)dG(x) = f(a)G(b) - f(a)G(a) + \int_a^b \Big(\int_a^b f'(u)1_{[a,x)}(u)du \Big) dG(x).$$

An application of Fubini's theorem yields

$$\begin{aligned} & f(a)G(b) - f(a)G(a) + \int_a^b f'(u)(G(b) - G(u))du \\ & \quad = f(a)G(b) - f(a)G(a) + (f(b) - f(a))G(b) - \int_a^b f'(u)G(u)du. \end{aligned}$$

Observe that the right–hand sides of the last expression and Relation (7.4) coincide. □

7.2. Tables

$\psi(i)$	$i+0$	$i+1$	$i+2$	$i+3$	$i+4$
$i=0$	3.423696	2.211864	1.387207	0.846734	0.504666
$i=5$	0.294581	0.168963	0.095563	0.053288	0.029470
$i=10$	0.016072	0.008755	0.004756	0.002552	0.001405
$i=15$	0.000709	0.000335	0.000175	0.000097	0.000032

Table 7.1

We present here numerical approximations of the values $\psi(i)$, $i=0,1,\ldots,19$ involved in the expression of the covariance matrices $\Sigma_\alpha(j_1,j_2;N)$ in (2.14) (see also (2.15)). We used Monte Carlo simulations with 10,000,000 independent pairs of 1−Fréchet variables. To reduce the variance of the estimates we used "bagging". That is, the Monte Carlo simulations were repeated independently 1,000 times and then the resulting means were taken as the final estimates reported in the table above.

j	2	3	4	5	6	7	8	9	10	11
$\sqrt{c_w(j)}$	1.417	0.802	0.515	0.346	0.238	0.166	0.116	0.082	0.058	0.041
$j+=10$	0.029	0.020	0.014	0.010	0.007	0.005	0.004	0.003	0.002	0.001
$\sqrt{2^j c_w(j)}$	2.834	2.267	2.060	1.960	1.905	1.875	1.857	1.847	1.841	1.837
$j+=10$	1.835	1.834	1.834	1.833	1.833	1.833	1.833	1.833	1.833	1.833

Table 7.2

We present here numerical estimates of the constants c_w involved in the asymptotic variances in Proposition 4.2 above. Here, we use $j_1=1$, for simplicity, and display 20 different values corresponding to $j_2=j=2,\ldots,21$. For convenience, we present $\sqrt{c_w}$ together with $\sqrt{2^{j_2}c_w}$ where the latter constant is useful if one normalizes in (4.28) by using $\sqrt{N_r}$ instead of $\sqrt{N_{j_2+r}}$.

References

Adamic, L. & Huberman, B. (2000), 'The nature of markets in the world wide web', *Quarterly Journal of Electronic Commerce* **1**, 5–12.

Adamic, L. & Huberman, B. (2002), 'Zipf's power law and the Internet', *Glottometrics* **3**, 143–150.

Adler, R., Feldman, R. & Taqqu, M. S., eds (1998), *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, Birkhäuser, Boston.

Carlson, J. M. & Doyle, J. (1999), 'Highly optimized tolerance: a mechanism for power laws in designed systems', *Physical Review E* **60**(2), 1412–1427.

Chen, Q., Chang, H., Govindan, R., Jamin, S., Shenker, S. & Willinger, W. (2002), The origin of power laws in Internet topologies revisited, INFOCOM, IEEE.

Crovella, M. E. & Taqqu, M. S. (1999), 'Estimating the heavy tail index from scaling properties', *Methodology and Computing in Applied Probability* **1**, 55–79.

Csörgő, S., Deheuvels, P. & Mason, D. (1985), 'Kernel estimates of the tail index of a distribution', *Annals of Statistics* **13**(3), 1050–1077.

de Haan, L., Drees, H. & Resnick, S. (2000), 'How to make a Hill plot', *Annals of Statistics* **28**(1), 254–274.

de Sousa, B. & Michailidis, G. (2004), 'A diagnostic plot for estimating the tail index of a distribution', *Journal of Computational and Graphical Statistics* **13**(4), 974–995.

Dekkers, A. & de Haan, L. (1989), 'On the estimation of the extreme–value index and large quantile estimation', *Ann. Statist.* **17**(4), 1795–1832.

Dekkers, A., Einmahl, J. & de Haan, L. (1989), 'A moment estimator for the index of an extreme–value distribution', *Ann. Statist.* **17**(4), 1833–1855.

Faloutsos, M., Faloutsos, P. & Faloutsos, C. (1999), On power–law relationships of the Internet topology, *in* 'SIGCOMM', pp. 251–262.

Feuerverger, A. & Hall, P. (1999), 'Estimating a tail exponent by modeling departure from a Pareto distribution', *Ann. Statist.* **27**(2), 760–781.

Groeneboom, P., Lopuhaä, H. & de Wolf, P. (2003), 'Kernel–type estimators for the extreme value index', *Annals of Statistics* **31**(6), 1956–1995.

Hall, P. (1982), 'On some simple estimates of an exponent of regular variation', *J. Roy. Stat. Assoc.* **44**, 37–42. Series B.

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

`http://greenwood.cr.usgs.gov/energy/WorldEnergy/OF97-463` (n.d.), U.S. Department of the Interior Geological Survey.

Kratz, M. & Resnick, S. I. (1996), 'The qq-estimator and heavy tails', *Stochastic Models* **12**, 699–724.

Leadbetter, M. R., Lindgren, G. & Rootzén, H. (1983), *Extremes and Related Properties of Random Sequences and Processes*, Springer-Verlag, New York.

Lu, J.-C. & Peng, L. (2002), 'Likelihood based confidence intervals for the tail index', *Extremes* **5**(4), 337–352 (2003).

Mason, D. (1982), 'Laws of large numbers for sums of extreme values', *Annals of Probability* **10**, 754–764.

McNeil, A. (1997), Estimating the tails of loss severity distributions using extreme value theory, *in* 'ASTIN Bulletin', Vol. 27, pp. 117–137.

Peng, L. & Qi, Y. (2004), 'Estimating the first– and second–order parameters of a heavy-tailed distribution', *Aust. N. Z. J. Stat.* **46**(2), 305–312.

Petrov, V. V. (1995), *Limit Theorems of Probability Theory*, Oxford University Press, Oxford.

Pickands, J. (1975), 'Statistical inference using extreme order statistics', *Ann. Statist.* **3**, 119–131.

Resnick, S. (1997*a*), Discussion of the Danish data on large fire insurance losses, *in* 'ASTIN Bulletin', Vol. 27, pp. 139–151.

Resnick, S. & Stărică, C. (1995), 'Consistency of Hill's estimator for dependent data', *Journal of Applied Probability* **32**, 139–167.

Resnick, S. & Stărică, C. (1997), 'Smoothing the Hill estimator', *Adv. in Appl. Probab.* **29**(1), 271–293.

Resnick, S. I. (1987), *Extreme Values, Regular Variation and Point Processes*, Springer-Verlag, New York.

Resnick, S. I. (1997*b*), 'Heavy tail modeling and teletraffic data', *The Annals of Statistics* **25**, 1805–1869. With discussions and rejoinder.

Stoev, S., Michailidis, G., Hamidieh, K. & Taqqu, M. (2006), On the estimation of the heavy–tail exponent in time series using the max–spectrum, Preprint.

Tsonis, A., Schultz, C. & Tsonis, P. (1997), 'Zipf's law and the structure and evolution of languages', *Complexity* **2**(5), 12–13.

Weissman, I. (1978), 'Estimation of parameters and large quantiles based on the k largest observations', *Journal of the American Statistical Association* **73**, 812–815.

Zipf, G. (1932), *Selective Studies and the Principle of Relative Frequency in Language*, Harvard

University Press.

Zipf, G. (1949), *Human Behavior and the Principle of Least Effort*, Addison–Wesley.