TAIL INFERENCE: WHERE DOES THE TAIL BEGIN?

TILO NGUYEN AND GENNADY SAMORODNITSKY

Abstract. The quality of estimation of tail parameters, such as tail index in the univariate case, or the spectral measure in the multivariate case, depends crucially on the part of the sample included in the estimation. A simple approach involving sequential statistical testing is proposed in order to choose this part of the sample. This method can be used both in the univariate and multivariate cases. It is computationally efficient, and can be easily automated. No visual inspection of the data is required. We establish consistency of the Hill estimator when used in conjunction with the proposed method, as well describe its asymptotic fluctuations. We compare our method to existing methods in univariate and multivariate tail estimation, and use it to analyze Danish fire insurance data.

1. Introduction

Let F be a univariate distribution (function). F is said to have a regularly varying right tail of index $\alpha > 0$ if the tail function $\bar{F} = 1 - F$ satisfies

$$\lim_{x\to\infty} \frac{\bar{F}(tx)}{\bar{F}(x)} = t^{-\alpha} \tag{1.1}$$

for all $t > 0$. The index α measures the heaviness of the tail. The smaller is α, the heavier is the right tail of the distribution. An encyclopedic treatment of regular variation is given by Resnick (1987, 2007).

Estimating the tail index α is of crucial importance in many applications of stochastic models, and a number of estimators have been designed for that purpose. The best known estimator of the tail index is the Hill estimator, introduced by Hill (1975), and it is defined as follows. Assume that $X_{1,n} \leq X_{2,n} \leq \ldots \leq X_{n,n}$ are the order statistics from a positive sample (or from the positive part of a general sample) $X_1, \ldots, X_n$. The Hill estimator based

1991 *Mathematics Subject Classification.* Primary 62G32, 60G70.

Key words and phrases. extremes, tail estimation, Hill estimator, spectral measure, tail region .

This research was partially supported by the ARO grant W911NF-07-1-0078, NSF grant DMS-1005903 and NSA grant H98230-11-1-0154 at Cornell University.

on k upper order statistics is defined as

$$H_{k,n} := \frac{1}{k}\sum_{i=0}^{k-1} \log \frac{X_{n-i,n}}{X_{n-k,n}} \tag{1.2}$$

If the sample is an i.i.d. sample from a distribution with a regularly varying right tail with tail index α, then, under the conditions $n \to \infty$, $k \to \infty$ and $\frac{k}{n} \to 0$, the Hill estimator $H_{k,n}$ converges in probability to $\gamma = \frac{1}{\alpha}$ (see Mason (1982)). If, additionally, $k/\log\log n \to \infty$, then even almost sure convergence holds (Deheuvels et al. (1988)). The role of the condition $\frac{k}{n} \to 0$ is to ensure that only data from the tail enter into the estimation. It has also been shown that the Hill estimator remains consistent under certain deviations from the i.i.d. assumption; a summary is in Theorem 6.4.6 of Embrechts et al. (1997).

In practice, when the Hill estimator is applied to a finite sample, a problem of choosing the appropriate number k of upper order statistics to construct the estimator arises. This problem becomes particularly critical because the Hill estimator has proved to be very sensitive to the choice of k. This sensitivity is shared by other estimators of the tail index, such as the Pickands estimator (see Dekkers and de Haan (1989)) and the moment estimator (see Dekkers et al. (1989)). Sometimes visual techniques are used: the estimator, e.g. the Hill estimator, is plotted for a range of k, and then one looks for a part of the plot that looks stable. The corresponding stable value is used to estimate α, and several smoothing techniques have been introduced to assist in this visual analysis; see Resnick and Stărică (1997). Still, the procedure is difficult to automate, and even visually it is sometimes difficult to use, as the so-called "Hill horror plots" demonstrate; see Embrechts et al. (1997).

A systematic way of selecting a "good" number of upper order statistics in Hill estimator originated with Hall (1990) and is based on the assumption that distribution F satisfies a further assumption of second order regular variation (which we introduce below). Under this assumption it becomes possible to look for $k = k(n)$ that minimizes the asymptotic bias of the estimator. This method was later refined by Danielsson et al. (2001) who suggested a two-level bootstrap procedure that works under minimal a priori

available information. An alternative approach of finding this asymptotically optimal number k of upper order statistics was suggested, under slightly more restrictive assumptions, by Drees and Kaufmann (1998).

In contrast, instead of viewing the problem of selecting the number k in Hill estimator (or any related estimator) as a problem of optimizing asymptotic efficiency, we view it as the problem of deciding which part of a given sample contains reliable information on the tail of the distribution F. Put another way, we would like to know where in a sample the tail begins. Much of our motivation lies in the multivariate context: given a sample of random vectors (potentially, in a highly dimensional space) with an appropriately defined regularly varying tail we would like to test a variety of different subvectors of these vectors for tail independence. This involves repeated estimation of the so called tail measure, a time consuming procedure, which is also highly sensitive to the contamination of the tail by the center of the distribution (see Resnick (2007)). It is, therefore, desirable to have a reasonably quick way of deciding which part of the sample belongs to the tail.

Our approach is based on a simple idea which we now introduce informally. It is well known that, under the assumption (1.1) of regular variation, vague convergence of point processes holds,

$$N_n = \sum_{i=1}^{n} \delta_{X_i/a_n} \stackrel{v}{\Rightarrow} N_* \,,$$

where δ_x is a point mass at x, and (a_n) a positive sequence satisfying $\bar{F}(a_n) \sim 1/n$ as $n \to \infty$. Further, N_* is a Poisson random measure on $(0, \infty]$ with measure measure $\mu_*(x, \infty] = x^{-\alpha},\ x > 0$; see Resnick (1987). We interpret this result as saying that any upper order statistics in the sample that fall in the tail region behave like points of a Poisson random measure with a power intensity. This property can be tested statistically, and sequentially. Specifically, one can perform appropriate statistical tests on the subsamples $X_{n-k+1,n}, X_{n-k+2,n}, \ldots, X_{n,n}$ while increasing k, and terminate the procedure once the k upper order statistics stop resembling points of a Poisson random measure with a power intensity.

Interestingly, Hill himself suggested a sequential statistical procedure for choosing k in his original paper Hill (1975). He considered the case when the distribution F had an exact Pareto tail beyond an unknown threshold D. If $X_{n-k-1,n} > D$, then, under the exact Pareto tail assumption, the random variables $iV_i := i \log \frac{X_{n-i,n}}{X_{n-i-1,n}}$ for $i = 0, 1, 2 \ldots, k$ are independent exponential random variables of parameter α. On the other hand, for k too large, the behavior of $\{iV_i\}$ would exhibit discrepancies from the exponential distribution. One can sequentially use exponential goodness of fit tests on $\{iV_i : i = 1, \ldots k\}$ for increasing k, until the hypothesis of exponentiality is rejected.

One can view Hill's procedure as a differential version of our sequential procedure. It has been criticized, perhaps too harshly, by Hall and Welsh (1985) as tending to result in too large a number k of order statistics. We use Hill's procedure as one of the benchmarks against which we test our approach in Section 3.

The formal definition of the our estimation procedure is given in Section 2. As the previous discussion indicates, we only need to decide how to test for a Poisson process with a power intensity. The test we choose in this paper becomes a certain test for exponentiality, but it is possible that other tests will perform equally well or, perhaps, even better. We prove the consistency of our estimator, and present a weak limit theorem describing the deviations of the estimator from the true value of the tail exponent. In Section 3 we test the performance of our estimator on simulated univariate data, and compare it to several benchmark estimators. We investigate how our rule of deciding on the tail part of the sample performs in estimation of the tail measure on simulated bivariate data in Section 4, and on Danish fire insurance data in Section 5.

2. The estimation procedure

Recall from the previous section that we would like to test sequentially the upper order statistics for resembling points of a Poisson process with a power intensity. We use the following property of a Poisson process: if

$V_1 > V_2 > \ldots > V_k$ are the largest points of a Poisson process on $(0, \infty)$ with the mean measure $\mu_*(x, \infty) = x^{-\alpha}$, $x > 0$, then $\{V_i/V_k,\, i = 1, \ldots, k-1\}$, considered as a set, forms an i.i.d. sample from the Pareto distribution with the tail $x^{-\alpha}$, $x > 1$, and taking the logarithms transforms this sample into an i.i.d. sample of exponential random variable with the mean $\gamma = 1/\alpha$. Accordingly, our procedure for deciding on the number k of the upper order statistics to use in the Hill estimator consists of sequentially testing the samples $\{\log \frac{X_{n-i,n}}{X_{n-k,n}} : i = 0, 1 \ldots k-1\}$ for the null hypothesis of exponential distribution. Our choice of sample fraction k used in the tail estimation is then $N_n - 1$, where N_n is the smallest k such that the test described above rejects the null hypothesis of exponentiality.

In order to implement this procedure one has to choose a test of exponentiality. Once this has been done, the only remaining choice is that of the significance level of the test. Such choices are needed in all procedures to select the number of the order statistics to use (recall the subsample size in the bootstrap procedure of Danielsson et al. (2001), or the threshold sequence of Drees and Kaufmann (1998)). We suggest a canonical way of choosing this significance level that appears to work reasonably well in the situations we have tried.

To test for exponentiality we choose the moment statistic

$$Q_{k,n} = \frac{\sqrt{k}}{2} \left(\frac{\frac{1}{k} \sum_{i=0}^{k-1} \left(\log \frac{X_{n-i,n}}{X_{n-k,n}} \right)^2}{\left(\frac{1}{k} \sum_{i=0}^{k-1} \log \frac{X_{n-i,n}}{X_{n-k,n}} \right)^2} - 2 \right);$$

its large sample distribution under the null hypothesis assumption of exponentiality is the standard normal distribution (Dahiya and Gurland (1972)). One could try to implement a sequential testing procedure by choosing a critical value ω (perhaps, a 99% quantile with respect to the limiting standard normal distribution) and use $N_n^* - 1$ upper order statistics, where

$$N_n^* := \inf\{k : 1 \le k \le n,\, |Q_{k,n}| \ge \omega\} \tag{2.1}$$

The problem with this implementation is that N_n^* stays tight as the sample size increases, and this contradicts the obvious requirement that to get any averaging effect we need to take more and more order statistics into account.

Therefore, the critical value needs to increase with the sample size n. We achieve this by selecting an increasing sequence $\theta_n \uparrow \infty$; this is the degree of freedom we mentioned above. On the other hand, in order to avoid taking into account too many order statistics, we choose to make it easier to reject the null hypothesis for larger k. It turns out that a good way to put all of this together is to set

$$N_n := \inf\left\{k : 1 \le k \le n,\, |Q_{k,n}| \ge \omega\sqrt{\frac{\theta_n}{k}}\right\}. \tag{2.2}$$

We will see in Theorem 1 below that, under a suitable growth condition on θ_n, this definition of N_n makes it, roughly, proportional to θ_n.

In order to state our first result, we introduce formally the notion of second order regular variation. Let $U = \left(\frac{1}{1-F}\right)^{\leftarrow}$ be the generalized inverse function. We assume that there exists $\rho < 0$ and a function A regularly varying at infinity with exponent ρ such that

$$\lim_{r\to\infty} \frac{\frac{U(rx)}{U(r)} - x^\gamma}{A(r)} = x^\gamma \frac{x^\rho - 1}{\rho} \tag{2.3}$$

(recall that $\gamma = 1/\alpha$). We also assume, without loss of generality, that A is continuous and $|A|$ is eventually decreasing.

Theorem 1. *Let $\omega > 0$ and (θ_n) an increasing sequence such that $\theta_n \uparrow \infty$ and $\theta_n = \mathrm{o}\left(n^{\frac{2|\rho|}{1+2|\rho|}}\right)$ as $n \to \infty$. Then $\dfrac{N_n}{\theta_n} \Rightarrow \tau_\omega$, where τ_ω is the first time a standard Brownian motion hits $\pm\omega$.*

To prove the theorem, first we need the following lemma, which is a functional version of Lemma 3.5.5 in de Haan and Ferreira (2006). In this lemma we work with spaces of the type $D[0,\infty)$, $D^2[0,\infty)$, $D[\delta,\infty)$ and $D^2[\delta,\infty)$ for $\delta > 0$. The latter spaces are, of course, only notationally different from the former spaces. We endow the D^2 spaces with the (strong) J_1 topology. See Whitt (2002) for details.

Lemma 2.1. *For $n \geq 1$ define*

$$M^j_{\theta_n,n}(t) = \begin{cases} 0 & \text{if } 0 \leq t < \dfrac{1}{\theta_n} \\ \dfrac{1}{\lfloor \theta_n t \rfloor} \sum_{i=0}^{\lfloor \theta_n t \rfloor - 1} \left(\log \dfrac{X_{n-i,n}}{X_{n-\lfloor \theta_n t \rfloor,n}} \right)^j & \text{if } \dfrac{1}{\theta_n} \leq t \leq \dfrac{n}{\theta_n} \\ \dfrac{1}{n} \sum_{i=0}^{n-1} \left(\log \dfrac{X_{n-i,n}}{X_{1,n}} \right)^j & \text{if } t > \dfrac{n}{\theta_n}, \end{cases}$$

$j = 1, 2$. *Then*

$$\sqrt{\theta_n} t \begin{pmatrix} \dfrac{M^1_{\theta_n,n}(t)}{\gamma} - 1 \\ \dfrac{M^2_{\theta_n,n}(t)}{\gamma^2} - 2 \end{pmatrix} \Rightarrow \begin{bmatrix} W_1(t) \\ W_2(t) \end{bmatrix}$$

in $D^2[0,\infty)$, where $\big((W_1(t), W_2(t)),\, t \geq 0\big)$ is a two-dimensional zero mean Brownian motion with covariance matrix

$$\begin{bmatrix} 1 & 4 \\ 4 & 20 \end{bmatrix}.$$

Proof. Fix $0 < \epsilon < |\rho|$. As in Lemma 3.5.5 in de Haan and Ferreira (2006), there are i.i.d. Pareto(1) random variables $Y_1, Y_2, \ldots, Y_n$ (i.e. $P(Y_1 > y) = 1/y$ for $y \geq 1$), a function $A_0 \sim A$ and $r_0 > 0$ with the property that

$$\begin{aligned} & \log(\frac{Y_{n-i,n}}{Y_{n-k,n}}) + A_0(Y_{n-k,n}) \frac{1}{\rho} \left(\left(\frac{Y_{n-i,n}}{Y_{n-k,n}} \right)^\rho - 1 \right) \\ & -\epsilon |A_0(Y_{n-k,n})| \frac{1}{\rho} \left(\frac{Y_{n-i,n}}{Y_{n-k,n}} \right)^{\rho+\epsilon} \\ \leq & \frac{1}{\gamma} \log \frac{X_{n-i,n}}{X_{n-k,n}} \leq \\ & \log(\frac{Y_{n-i,n}}{Y_{n-k,n}}) + A_0(Y_{n-k,n}) \frac{1}{\rho} \left(\left(\frac{Y_{n-i,n}}{Y_{n-k,n}} \right)^\rho - 1 \right) \\ & +\epsilon |A_0(Y_{n-k,n})| \left(\frac{Y_{n-i,n}}{Y_{n-k,n}} \right)^{\rho+\epsilon} \end{aligned} \tag{2.4}$$

if $Y_{n-k,n} > r_0$; note that the latter condition is satisfied for large n if $k = O(\theta_n) = \mathrm{o}(n)$ under the assumptions of the lemma.

Hence for fixed $T > 0$, eventually (i.e. for n large) we have

$$\begin{aligned}\frac{\lfloor \theta_n t \rfloor}{\sqrt{\theta_n}} \left(\frac{M^1_{\theta_n,n}(t)}{\gamma} - 1 \right) &\leq \frac{1}{\sqrt{\theta_n}} \sum_{i=0}^{\lfloor \theta_n t \rfloor - 1} \left(\log \frac{Y_{n-i,n}}{Y_{n-\lfloor \theta_n t \rfloor,n}} - 1 \right) \\ &+ \frac{1}{\rho} \sqrt{\theta_n} A_0(Y_{n-\lfloor \theta_n t \rfloor,n}) \frac{1}{\theta_n} \sum_{i=0}^{\lfloor \theta_n t \rfloor - 1} \left(\left(\frac{Y_{n-i,n}}{Y_{n-\lfloor \theta_n t \rfloor,n}} \right)^{\rho} - 1 \right) \\ &+ \epsilon \sqrt{\theta_n} |A_0(Y_{n-\lfloor \theta_n t \rfloor,n})| \frac{1}{\theta_n} \sum_{i=0}^{\lfloor \theta_n t \rfloor - 1} \left(\left(\frac{Y_{n-i,n}}{Y_{n-\lfloor \theta_n t \rfloor,n}} \right)^{\rho+\epsilon} \right)\end{aligned} \tag{2.5}$$

for every $t \in [0,T]$. Note that all the terms in the right hand side of (2.5) (which we interpret as 0 for $0 \leq t < 1/\theta_n$) are in $D[0,T]$. Let us denote the second and the third terms by $W^{(2)}_n(t)$ and $W^{(3)}_n(t)$, correspondingly. We start with showing that

$$\sup_{0 \leq t \leq T} |W^{(2)}_n(t)| \to 0 \ \text{ in probability as } n \to \infty. \tag{2.6}$$

Since $\dfrac{A_0(Y_{n-\lfloor \theta_n t \rfloor,n})}{A(Y_{n-\lfloor \theta_n t \rfloor,n})} \to 1$ a.s uniformly in $t \in [0,T]$, we may and will replace A_0 by A in this calculation. Further, as $|A|$ is eventually decreasing, for n large enough we have

$$|A(Y_{n-\lfloor \theta_n t \rfloor,n})| \leq |A(Y_{n-\lfloor \theta_n T \rfloor,n})|$$

for all relevant t. Furthermore, by Corollary 2.2.2 in de Haan and Ferreira (2006),

$$\frac{\lfloor \theta_n T \rfloor}{n} Y_{n-\lfloor \theta_n T \rfloor} \xrightarrow{P} 1\,,$$

and, since A is regularly varying,

$$\frac{A(Y_{n-\lfloor \theta_n T \rfloor})}{A(\frac{n}{\lfloor \theta_n T \rfloor})} \xrightarrow{P} 1\,.$$

Putting everything together, we see that, in order to prove (2.6), it is enough to prove that

$$\lim_{n \to \infty} \mathbb{P} \left(\sup_{0 \leq t \leq T} \left| \sqrt{\theta_n} A(\frac{n}{\lfloor \theta_n T \rfloor}) \frac{1}{\theta_n} \sum_{i=0}^{\lfloor \theta_n t \rfloor - 1} \left(\left(\frac{Y_{n-i,n}}{Y_{n-\lfloor \theta_n t \rfloor,n}} \right)^{\rho} - 1 \right) \right| > \zeta \right) = 0\,.$$

Let $\mu_\rho = 1/(1-\rho) = EY_1^\rho$. For large n we have

$$\mathbb{P} \left(\sup_{0 \leq t \leq T} \left| \sqrt{\theta_n} A(\frac{n}{\lfloor \theta_n T \rfloor}) \frac{1}{\theta_n} \sum_{i=0}^{\lfloor \theta_n t \rfloor - 1} \left(\left(\frac{Y_{n-i,n}}{Y_{n-\lfloor \theta_n t \rfloor,n}} \right)^{\rho} - 1 \right) \right| > \zeta \right)$$

$$\leq \mathbb{P}\left(\sup_{0\leq t\leq T}\left|\sqrt{\theta_n}A(\frac{n}{\lfloor\theta_n T\rfloor})\frac{1}{\theta_n}\sum_{i=0}^{\lfloor\theta_n t\rfloor-1}\left(\left(\frac{Y_{n-i,n}}{Y_{n-\lfloor\theta_n t\rfloor,n}}\right)^{\rho}-\mu_\rho\right)\right|>\zeta/2\right),$$

since

$$\sqrt{\theta_n}A(\frac{n}{\lfloor\theta_n T\rfloor})\to 0$$

as $n\to\infty$, by the growth assumption on the sequence (θ_n). The process above is a step function with jumps at multiples of $\frac{1}{\theta_n}$, hence largest value of the process is achieved at one of these steps. Therefore, the above probability does not exceed

$$\sum_{j=1}^{\lfloor T\theta_n\rfloor}\mathbb{P}\left(\left|\sqrt{\theta_n}A(\frac{n}{\lfloor\theta_n T\rfloor})\frac{1}{\theta_n}\sum_{i=0}^{j-1}\left(\left(\frac{Y_{n-i,n}}{Y_{n-j,n}}\right)^{\rho}-\mu_\rho\right)\right|>\zeta/2\right)$$

$$\leq \sum_{j=1}^{\lfloor T\theta_n\rfloor}\left(\sqrt{\theta_n}A(\frac{n}{\lfloor\theta_n T\rfloor})\right)^2\frac{4}{\zeta^2\theta_n^2}E\left[\sum_{i=0}^{j-1}\left(\left(\frac{Y_{n-i,n}}{Y_{n-j,n}}\right)^{\rho}-\mu_\rho\right)\right]^2.$$

For each fixed j, by the Renyi representation, $\left\{\frac{Y_{n-i,n}}{Y_{n-j,n}}\right\}_i \stackrel{d}{=} \left\{Y^*_{j-i,j}\right\}_i$ where $Y^*_0, Y^*_1, \ldots, Y^*_{j-1}$ are, once again, i.i.d. Pareto(1) random variables. Therefore,

$$E\left[\sum_{i=0}^{j-1}\left(\left(\frac{Y_{n-i,n}}{Y_{n-j,n}}\right)^{\rho}-\mu_\rho\right)\right]^2 = j\mathrm{Var}(Y^{*\rho}).$$

By the growth assumption of the sequence (θ_n) we conclude that

$$\sum_{j=1}^{\lfloor T\theta_n\rfloor}\left(\sqrt{\theta_n}A(\frac{n}{\lfloor\theta_n T\rfloor})\right)^2\frac{4}{\zeta^2\theta_n^2}E\left[\sum_{i=0}^{j-1}\left(\left(\frac{Y_{n-i,n}}{Y_{n-j,n}}\right)^{\rho}-\mu_\rho\right)\right]^2$$

$$\leq \left(\sqrt{\theta_n}A(\frac{n}{\lfloor\theta_n T\rfloor})\right)^2\frac{4}{\zeta^2}\mathrm{Var}(Y^{*\rho})\frac{1}{\theta_n^2}\sum_{j=1}^{\lfloor T\theta_n\rfloor} j\to 0$$

as $n\to\infty$. This proves (2.6). In the same way we can show that

$$\sup_{\delta\leq t\leq T}|W_n^{(3)}(t)|\to 0 \text{ in probability as } n\to\infty. \tag{2.7}$$

Applying the corresponding lower bounds, we see that

$$\frac{\lfloor\theta_n t\rfloor}{\sqrt{\theta_n}}\left(\frac{M^1_{\theta_n,n}(t)}{\gamma}-1\right)-\frac{1}{\sqrt{\theta_n}}\sum_{i=0}^{\lfloor\theta_n t\rfloor-1}\left(\log\frac{Y_{n-i,n}}{Y_{n-\lfloor\theta_n t\rfloor,n}}-1\right)\to 0 \tag{2.8}$$

$t\in[\delta,\infty)$, uniformly on compact intervals, in probability. Next, we recall that $\log Y_1$ is a standard exponential random variable, so that the differences

$\log Y_{n-i,n} - \log Y_{n-i-1,n}$, $i = 0, 1, \ldots, n-1$ are independent exponential random variables with the means $1/(i+1)$, $i = 0, 1, \ldots, n-1$. Therefore, denoting the ith of these exponential random variables by $E_i/(i+1)$, we see that for $k = 1, \ldots, n$,

$$\sum_{i=0}^{k-1} \log \frac{Y_{n-i,n}}{Y_{n-k,n}} = \sum_{i=0}^{k-1} E_i \, .$$

Therefore,

$$\frac{\lfloor \theta_n t \rfloor}{\sqrt{\theta_n}} \left(\frac{M^1_{\theta_n,n}(t)}{\gamma} - 1 \right) - \frac{1}{\sqrt{\theta_n}} \sum_{i=0}^{\lfloor \theta_n t \rfloor - 1} (E_i - 1) \to 0 \tag{2.9}$$

$t \in [\delta, \infty)$, uniformly on compact intervals, in probability. Squaring (2.4) and repeating the argument gives us

$$\frac{\lfloor \theta_n t \rfloor}{\sqrt{\theta_n}} \left(\frac{M^2_{\theta_n,n}(t)}{\gamma^2} - 2 \right) - \frac{1}{\sqrt{\theta_n}} \sum_{i=0}^{\lfloor \theta_n t \rfloor - 1} \big((E_i^2 - 2) \to 0, \tag{2.10}$$

$t \in [\delta, \infty)$, uniformly on compact intervals, in probability.

By Theorem 12.6.1 and Remark 12.6.2 in Whitt (2002), the statement of the lemma will follow once we check that

$$\frac{1}{\sqrt{\theta_n}} \begin{pmatrix} \sum_{i=0}^{\lfloor \theta_n t \rfloor - 1} (E_i - 1) \\ \sum_{i=0}^{\lfloor \theta_n t \rfloor - 1} ((E_i^2 - 2) \end{pmatrix} \Rightarrow \begin{bmatrix} W_1(t) \\ W_2(t) \end{bmatrix}$$

in $D^2[\delta, \infty)$, where $\big((W_1(t), W_2(t)), \, t \geq \delta \big)$ is a two-dimensional zero mean Brownian motion covariance matrix

$$\begin{bmatrix} 1 & 4 \\ 4 & 20 \end{bmatrix}.$$

This is, however, an immediate consequence of the multivariate version of Donsker's theorem; see e.g. Theorem 4.3.5 in Whitt (2002). □

Proof of theorem. We start by showing that for any $\delta > 0$,

$$\sqrt{t} Q^*_n(t) \Rightarrow \frac{1}{2}(W_2(t) - 2W_1(t)) \tag{2.11}$$

in $D[\delta, \infty)$, where $\big((W_1(t), W_2(t)), \, t \geq 0 \big)$ is the two-dimensional Brownian motion of Lemma 2.1, and

$$Q^*_n(t) = \begin{cases} Q_{\lfloor \theta_n t \rfloor, n} & \text{if } \ \delta \leq t \leq \dfrac{n}{\theta_n}, \\ Q_{n,n} & \text{if } \ t > \dfrac{n}{\theta_n} \, . \end{cases}$$

By Theorem 16.7 in Billingsley (1999), we have to prove convergence in $D[\delta, T]$ for each $\delta < T < \infty$. Straightforward algebra shows that

$$\sqrt{t}Q_n^*(t) = \frac{\sqrt{\theta_n}t}{2}\frac{\big(M^2_{\theta_n,n}(t)/\gamma^2 - 2\big) - 4\big(M^1_{\theta_n,n}(t)/\gamma - 1\big)}{\big(M^1_{\theta_n,n}(t)\big)^2/\gamma^2}$$

$$-\sqrt{\theta_n}t\frac{\big(M^1_{\theta_n,n}(t)/\gamma - 1\big)^2}{\big(M^1_{\theta_n,n}(t)\big)^2/\gamma^2} := V_n^{(1)}(t) - V_n^{(2)}(t)\,,$$

$\delta \le t \le T$, while for $0 \le t < 1/n$ we define both $V_n^{(1)}(t) = 0$ and $V_n^{(2)}(t) = 0$. Call

$$D_n^{(1)}(t) = \frac{\sqrt{\theta_n}t}{2}\Big[\big(M^2_{\theta_n,n}(t)/\gamma^2 - 2\big) - 4\big(M^1_{\theta_n,n}(t)/\gamma - 1\big)\Big],\ \delta \le t \le T.$$

Since the limiting process in Lemma 2.1 is continuous, the weak convergence holds also in the uniform topology (on $[\delta, T]$), and addition is continuous in this topology. By Lemma 2.1 and continuous mapping theorem we conclude that

$$D_n^{(1)}(t) \Rightarrow \frac{1}{2}(W_2(t) - 4W_1(t))$$

in $D[\delta, T]$. Furthermore, by Lemma 2.1,

$$M^1_{\theta_n,n}(t)/\gamma \to 1$$

uniformly on $[\delta, T]$ in probability. By Theorem 3.1 in Billingsley (1999) we conclude that

$$V_n^{(1)}(t) \Rightarrow \frac{1}{2}(W_2(t) - 4W_1(t))$$

in $D[\delta, T]$ as well. Similarly,

$$V_n^{(2)}(t) \to 0$$

uniformly on $[\delta, T]$ in probability, so that we may apply Theorem 3.1 in Billingsley (1999) once again and obtain (2.11).

Fix $x > 0$, and write

$$P(N_n \le \theta_n x) = P\left(|Q_{k,n}| \ge \omega\sqrt{\frac{\theta_n}{k}} \text{ for some } 1 \le k \le \theta_n x\right).$$

Therefore, for $0 < \delta < x$ we have

$$P\left(|Q_{k,n}| \ge \omega\sqrt{\frac{\theta_n}{k}} \text{ for some } \theta_n\delta \le k \le \theta_n x\right) \le P(N_n \le \theta_n x)$$

$$\leq P\left(|Q_{k,n}| \geq \omega\sqrt{\frac{\theta_n}{k}} \text{ for some } \theta_n\delta \leq k \leq \theta_n x\right)$$
$$+P\left(|Q_{k,n}| \geq \omega\sqrt{\frac{\theta_n}{k}} \text{ for some } 1 \leq k \leq \theta_n\delta\right).$$

We will show that for any $0 < \delta < x$

(2.12)

$$\lim_{n\to\infty} P\left(|Q_{k,n}| \geq \omega\sqrt{\frac{\theta_n}{k}} \text{ for some } \theta_n\delta \leq k \leq \theta_n x\right) = P\big(\sup_{\delta\leq t\leq x} |B(t)| \geq \omega\big)\,,$$

where $\big(B(t),\, t \geq 0\big)$ is a standard Brownian motion, while

$$\lim_{\delta\to 0} \limsup_{n\to\infty} P\left(|Q_{k,n}| \geq \omega\sqrt{\frac{\theta_n}{k}} \text{ for some } 1 \leq k \leq \theta_n\delta\right) = 0\,. \tag{2.13}$$

It will follow from the above relations that

$$P(N_n \leq \theta_n x) \to P\big(\sup_{0\leq t\leq x} |B(t)| \geq \omega\big) = P(\tau_\omega \leq x),$$

which is what we need for the statement of the theorem.

Observe that for any $\delta > 0$,

$$P\left(|Q_{k,n}| \geq \omega\sqrt{\frac{\theta_n}{k}} \text{ for some } \theta_n\delta \leq k \leq \theta_n x\right)$$
$$= P\left(|\sqrt{t}Q_n^*(t)| \geq \sqrt{t}\omega\sqrt{\frac{\theta_n}{[\theta_n t]}} \text{ for some } \delta \leq t \leq x\right)$$
$$= P\left(|\sqrt{t}Q_n^*(t)| \geq \omega(1+o(1)) \text{ for some } \delta \leq t \leq x\right)$$

(with the same $o(1)$ for all relevant t). Now (2.12) follows from (2.11) and the continuity of the supremum distribution of the Brownian motion.

In order to show (2.13), we start with showing that, for any $\delta > 0$, both

$$\inf_{1\leq k\leq\theta_n\delta} M^1_{k,n} \text{ is stochastically bounded away from 0, and} \tag{2.14}$$
$$\sup_{1\leq k\leq\theta_n\delta} M^1_{k,n} \text{ is stochastically bounded away from infinity.}$$

To see this, we recall from the proof of Lemma 2.1 that

$$\Big(M^1_{k,n},\, 1 \leq k \leq \theta_n\delta\Big) = \Big(M^{1,Y}_{k,n}\gamma + W_{k,n},\, 1 \leq k \leq \theta_n\delta\Big)\,,$$

where $(M^{1,Y}_{k,n})$ is constructed using the standard Pareto random variables, while

$$\sup_{1\leq k\leq\theta_n\delta} |W_{k,n}| \to 0 \text{ in probability.}$$

Therefore, (2.14) will follow once we check it for the standard Pareto random variables. However, we have seen that, in the latter case,

$$\Big(M_{k,n}^{1,Y},\, 1 \le k \le \theta_n\delta\Big) \stackrel{d}{=} \Big(\frac{1}{k}\sum_{i=1}^{k} E_i,\, 1 \le k \le \theta_n\delta\Big)\,,$$

where $E_1, E_2, \dots$ are i.i.d. standard exponential random variables, and, hence, (2.14) follows from the law of large numbers.

We continue by writing

$$P\left(|Q_{k,n}| \ge \omega\sqrt{\frac{\theta_n}{k}} \text{ for some } 1 \le k \le \theta_n\delta\right)$$

$$= P\left(k\left|\frac{M_{k,n}^2 - 2(M_{k,n}^1)^2}{2(M_{k,n}^1)^2}\right| > \omega\sqrt{\theta_n} \text{ for some } 1 \le k \le \theta_n\delta\right)$$

and, using (2.14), we see that, in order to prove (2.13), it is enough to prove that

$$\lim_{\delta\to 0}\limsup_{n\to\infty} P\left(|R_{k,n}| \ge \omega\sqrt{\theta_n} \text{ for some } 1 \le k \le \theta_n\delta\right) = 0\,, \tag{2.15}$$

where

$$R_{k,n} = k\frac{M_{k,n}^2 - 2(M_{k,n}^1)^2}{2(M_{k,n}^1)^2},\ k = 1, 2, \dots, n\,.$$

A straightforward algebra allows us to write the above probability as

$$P\left(\frac{k}{2}\left|\Big(\frac{M_{k,n}^2}{\gamma^2} - 2\Big) - 4\Big(\frac{M_{k,n}^1}{\gamma} - 1\Big)\right.\right.$$

$$\left.\left. -2\Big(\frac{M_{k,n}^1}{\gamma} - 1\Big)^2\right| > \omega\sqrt{\theta_n} \text{ for some } 1 \le k \le \theta_n\delta\right)\,.$$

By (2.14), we can write now

$$P\left(|R_{k,n}| \ge \omega\sqrt{\theta_n} \text{ for some } 1 \le k \le \theta_n\delta\right)$$

$$\le P\Big(\frac{k}{2}\left(\Big|\frac{M_{k,n}^2}{\gamma^2} - 2\Big| + 4\Big|\frac{M_{k,n}^1}{\gamma} - 1\Big|\right)(1 + K_n) \ge \omega\sqrt{\theta_n}$$

$$\text{for some } 1 \le k \le \theta_n\delta\Big)\,,$$

where (K_n) is a tight family of nonnegative random variables. Since by Lemma 2.1,

$$P\Big(\frac{k}{2}\left(\Big|\frac{M_{k,n}^2}{\gamma^2} - 2\Big| + 4\Big|\frac{M_{k,n}^1}{\gamma} - 1\Big|\right) \ge \omega\sqrt{\theta_n} \text{ for some } 1 \le k \le \theta_n\delta\Big)$$

$$= P\Big(\sup_{0\le t\le \delta} \sqrt{\theta_n} t \Big(\Big| \frac{M^2_{[\theta_n t],n}}{\gamma^2} - 2 \Big| + 4 \Big| \frac{M^1_{[\theta_n t],n}}{\gamma} - 1 \Big| \Big) > \omega \Big)$$

$$\to P(\sup_{0\le t\le \delta} |B(t)| > \omega) \,,$$

where (B) is some Brownian motion, the claim (2.13) follows. This completes the proof of the theorem. □

The following theorem, which is the main theorem of this section, shows that using the Hill estimator with the random number of upper order statistics given by (2.2) is, indeed, a consistent estimator of $\gamma = 1/\alpha$. We also derive a weak limit result for the deviations of the estimator from $1/\gamma$. It shows that these deviations are of the order $1/\sqrt{\theta_n}$, which is expected, since by Theorem 1, the number N_n of the order statistics in the Hill estimator is of the order θ_n.

Theorem 2. *Let $\theta_n = \mathrm{o}\left(n^{\frac{2|\rho|}{1+2|\rho|}} \right)$ as $n \to \infty$, and let N_n be given by (2.2). The Hill estimator based on N_n upper order statistics is consistent, i.e.*

$$H_{N_n,n} = \frac{1}{N_n} \sum_{i=0}^{N_n-1} \log \frac{X_{n-1,n}}{X_{n-N_n,n}} \xrightarrow{P} \gamma$$

as $n \to \infty$. Furthermore,

$$\sqrt{\theta_n} \left(\frac{H_{N_n,n}}{\gamma} - 1 \right) \Rightarrow \frac{G}{(\tau_\omega)^{1/2}} \,, \tag{2.16}$$

where G is a standard normal random variable independent of the first hitting time τ_ω from Theorem 1.

Proof. The idea is to use a random stopping technique in a weak convergence context. The formulation we will use is the one given in Theorem 2.2.1 in Silvestrov (2004). If, for each n, $(X_n(t),\, t \ge 0)$ is a càdlàg process, and τ_n is a nonnegative random variable, such that for all $0 \le a < b < \infty$,

$$\begin{aligned} &(\tau_n,\, \sup_{t\in[a,b)} X_n(t)) \Rightarrow (\tau_0,\, \sup_{t\in[a,b)} X_0(t)) \\ &(\tau_n,\, \inf_{t\in[a,b)} X_n(t)) \Rightarrow (\tau_0,\, \inf_{t\in[a,b)} X_0(t)) \end{aligned} \tag{2.17}$$

for some continuous process $(X_0(t),\, t \ge 0)$ and a nonnegative random variable τ_0, then $X_n(\tau_n) \Rightarrow X_0(\tau_0)$.

It is, clearly, enough to prove the weak convergence (2.16), as the consistence of the estimator would then follow automatically. Note that (2.16) is equivalent to

$$\sqrt{\theta_n}\frac{1}{N_n}\sum_{i=0}^{N_n-1}\Big(\frac{1}{\gamma}\log\frac{X_{n-1,n}}{X_{n-N_n,n}}-1\Big)\Rightarrow\frac{G}{(\tau_\omega)^{1/2}}\,. \tag{2.18}$$

We will prove that, for any $\delta>0$,

$$\sqrt{\theta_n}\frac{1}{N_n\vee\theta_n\delta}\sum_{i=0}^{[N_n\vee\theta_n\delta]-1}\Big(\frac{1}{\gamma}\log\frac{X_{n-1,n}}{X_{n-[N_n\vee\theta_n\delta],n}}-1\Big)\Rightarrow\frac{G}{(\tau_\omega\vee\delta)^{1/2}}\,. \tag{2.19}$$

The claim (2.18) would then follows from (2.13).

Note the that expression in the left hand side of (2.19) results from a substitution of the random time

$$\tau_n=\max\left(\delta,\frac{N_n}{\theta_n}\right)$$

into the càdlàg process

$$V_n(t)=\sqrt{\theta_n}\frac{1}{\theta_n t}\sum_{i=0}^{[\theta_n t]-1}\Big(\frac{1}{\gamma}\log\frac{X_{n-1,n}}{X_{n-[\theta_n t],n}}-1\Big),\ t\ge\delta\,.$$

According to (2.17) and to self-similarity of the Brownian motion, it is enough to check that for all $\delta\le a<b<\infty$,

$$\Big(\max\left(\delta,\frac{N_n}{\theta_n}\right),\ \sup_{t\in[a,b)}V_n(t)\Big)\Rightarrow\Big(\tau_\omega\vee\delta,\ \sup_{t\in[a,b)}\frac{B(t)}{t}\Big)\,, \tag{2.20}$$

where τ_ω is, as before, the first hitting time of a standard Brownian motion, independent of another standard Brownian motion B. We also need to prove an analogous statement with suprema replaced by infima but, since the two statements can be proved in the same way, we concentrate on suprema only.

In order to prove (2.20), it is enough to show that for any $x\ge\delta$ and $y\ge0$,

$$P\Big(N_n\le\theta_n x,\ \sup_{t\in[a,b)}V_n(t)\le y\Big)\to P\Big(\tau_\omega\le x,\ \sup_{t\in[a,b)}\frac{B(t)}{t}\le y\Big)\,.$$

We have seen in an analogous situation in the proof of Theorem 1 that this statement will follow once we check that for any $0<\delta'<x$,

$$\begin{aligned}&P\Big(\theta_n\delta'\le N_n\le\theta_n x,\ \sup_{t\in[a,b)}V_n(t)\le y\Big)\\ &\to P\Big(\sup_{\delta'\le t\le x}|B_1(t)|\ge\omega,\ \sup_{t\in[a,b)}\frac{B(t)}{t}\le y\Big)\,,\end{aligned}$$

where B and B_1 are independent standard Brownian motions, which, in turn, will be implied by the statement

$$P\Big(\sup_{\delta'\le t\le x}|\sqrt{t}\,Q_n^*(t)|\ge w,\ \sup_{t\in[a,b)}V_n(t)\le y\Big) \tag{2.21}$$

$$\to P\Big(\sup_{\delta'\le t\le x}|B_1(t)|\ge \omega,\ \sup_{t\in[a,b)}\frac{B(t)}{t}\le y\Big)\,.$$

To this end note that

$$V_n(t)=\frac{1}{t}\left[\sqrt{\theta_n}\,t\left(\frac{M^1_{\theta_n,n}(t)}{\gamma}-1\right)\right],$$

and the map $\big(f(t),\,t\ge\delta\big)\to\big(f(t)/t,\,t\ge\delta\big)$ is continuous on $D[\delta,\infty)$. Therefore, the argument leading to (2.11) applies, and it gives us the joint convergence

$$\begin{bmatrix}\sqrt{t}\,Q_n^*(t),\,t\ge\delta'\\ V_n(t),\,t\ge\delta'\end{bmatrix}\Rightarrow\begin{bmatrix}\frac{1}{2}\big(W_2(t)-4W_1(t)\big),\,t\ge\delta'\\ \frac{W_1(t)}{t},\,t\ge\delta'\end{bmatrix},$$

where W_1 and W_2 are as in Lemma 2.1. It is a simple matter to compute the correlations and check that W_1 and $(W_2-4W_1)/2$ are independent standard Brownian motions. Therefore, (2.21) follows, and the proof of the theorem is complete. □

Remark 3. We have already mentioned that, according to Theorem 2, the deviations of our estimator from the true value of γ are of the order $1/\sqrt{\theta_n}$. Since, under conditions of that theorem, the rate of growth of θ_n can go all the way up to $n^{\frac{2|\rho|}{1+2|\rho|}}$, our estimator can almost achieve the optimal rate of decay of the asymptotic deviation from the true γ, given by $n^{\frac{-|\rho|}{1+2|\rho|}}$; see e.g. Danielsson et al. (2001). Since the exponent ρ of the second order regular variation (2.3) is unknown, one could, potentially, combine our method with the bootstrap technique of Danielsson et al. (2001). We do not pursue this approach. In fact, our goal is not necessarily asymptotic efficiency, since for some distributions it can take a very long time until these asymptotics become effective. Our goal is, rather, determining, for a given sample size, which (upper) part of the sample appears, statistically, to be consistent with being in the tail region. For this purpose, sequences (θ_n) that increase at a much slower rate, appear to be appropriate. In fact, as the reader will see

in the subsequent sections, we advocate using logarithmically fast increasing sequences.

3. Testing the estimator on simulated data

In this section we evaluate our procedure (2.2) for selecting the number of upper order statistics in the Hill estimator on simulated univariate data. We compare the resulting performance of the estimator with the bootstrap procedure of Danielsson et al. (2001) , the optimal sample fraction choice of Drees and Kaufmann (1998), and to the original testing procedure of Hill (1975). For the test data we choose i.i.d. samples from the Student-t distribution and from the symmetric stable distribution. We have chosen these distributions because the "usual" Hill plots are often difficult to interpret in these cases. Recall that for the Student-t distribution, the tail exponent α is equal to the number of degrees of freedom. For the Student-t distribution we considered the cases $\alpha = 1, 3, 4$, while for the stable distribution, we tested the cases $\alpha = 1$ and $\alpha = 1.7$.

The Hill estimator has been shown to be consistent not only on i.i.d. data but also under certain kinds of dependence, (see Hsing (1991), Resnick and Stărică (1995), Resnick and Stărică (1998)). This includes the class of moving average processes, and we also test our estimator on the MA(1) process Y_t=$X_t + X_{t-1}$ where X_t are i.i.d. Student-t random variables with 3 degrees of freedom. The tail index in this case is equal to 3.

We have tested our estimator with logarithmic sequences $\theta_n = \log n$ and $\theta_n = (\log n)^2$. In all cases we chose ω to be the 95th quantile of the standard normal distribution. The results of the simulation are displayed in the tables below. Table 1 and 2 present the results using sample size $n = 5000$ and $n = 50000$ of absolute values of the above distribution respectively. Each simulation was performed 250 times. The following information is displayed.

- The testing procedure suggested by Hill (1975) using the moment statistic with significant level .05.
- Our choice of the sample fraction N_n with $\theta_n = \log n$.
- Our choice of the sample fraction N_n with $\theta_n = (\log n)^2$.

- The choice of sample fraction using bootstrap method proposed by Danielsson et al. (2001) with n_1 going from 1000 to 4000 in increments of 250 for sample size 5000 and from 10000 to 35000 in increments of 2500 for sample size 50000. In both case the number of bootstrap samples is 500.
- The optimal sample fraction choice $\hat{k}_{opt}$ of Drees and Kaufmann (1998) with the initial $\tilde{\beta}_n$ based on the upper $2\sqrt{n}$ order statistics, $r_n = 2.5\tilde{\beta}_n n^{.25}$, $\psi = .7$ and $\rho_0 = 1$.

The results in Table 1 and Table 2 indicate that we consistently obtain good results with $\theta_n = (\log n)^2$. Our choice of the number of order statistics to use in the Hill estimator performs reasonably well with both moderate and large sample sizes. Its computational complexity is significantly lower than that of the competing methods. In particular, our estimator significantly outperform the other estimators in the notoriously difficult case of the stable distribution with parameter $\alpha = 1.7$.

4. Estimation of the spectral measure

It is particularly important to have a procedure to determine which part of the sample belongs to the tail region, in the multivariate context. Our approach, discussed above in the univariate context, can be conveniently applied in the multivariate context as well. In this section we explore, on simulated bivariate data, how well our approach detects the tail region.

We recall briefly the notion of multivariate regular variation; see Resnick (2007) for details. The distribution of a d-dimensional random vector X is said to be regularly varying with index α if there exists a random vector Θ taking values on the unit sphere $\mathcal{S}^{d-1}$ in $\mathbb{R}^d$ such that

$$\lim x \to \infty \frac{P(|X| > tx, \frac{X}{|X|} \in B)}{P(|X| > x)} = t^{-\alpha} P(\Theta \in B) \tag{4.1}$$

for all Θ-continuity Borel subsets of $\mathcal{S}^{d-1}$. This, clearly, implies that all marginal distributions are also regularly varying with index α, and the same is true for the length (the radial component) of the observations. The law of Θ on the unit sphere is called the (normalized) spectral measure.

TABLE 1. Simulation with $n = 5000$

Method		Hill		N_n, $\theta_n = \log n$		N_n, $\theta_n = (\log n)^2$		Bootstrap		$\hat{k}_{opt}$	
Dist.	α	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE
Student(4)	4	2.7794	1.7098	4.0496	.9128	3.4568	.6510	3.6135	.6859	3.4270	.6629
Student(3)	3	2.1719	.9843	3.1460	.7434	2.7726	.3657	2.8490	.4383	2.7669	.3358
Student(1)	1	.9326	.3937	1.0738	.2203	1.0109	.0890	.9881	.0502	.9965	.0391
Stable(1.7)	1.7	2.0347	.6951	1.9401	.4586	2.0013	.3887	2.2515	.5654	2.2138	.5283
Stable(1)	1	.8965	.1683	1.0684	.2567	1.0099	.0855	.9945	.0689	.9912	.0404
MA(1)	3	2.8335	1.8239	3.8765	1.6059	3.1434	.5232	3.1365	.5647	3.0955	.3708

TABLE 2. Simulation with $n = 50000$

Method		Hill		N_n, $\theta_n = \log n$		N_n, $\theta_n = (\log n)^2$		Bootstrap		$\hat{k}_{opt}$	
Dist.	α	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE
Student(4)	4	3.2954	1.5064	4.2361	1.0055	3.7958	.4743	3.7690	.5282	3.6080	.4217
Student(3)	3	2.5280	.6734	3.1677	.7360	2.9391	.2245	2.8900	.3013	2.8490	.1839
Student(1)	1	.9499	.2182	1.0857	.2382	1.0103	.0697	.9959	.0215	.9970	.0159
Stable(1.7)	1.7	2.0894	.5303	1.8276	.3935	1.7733	.1670	2.2276	.5288	2.1057	.4079
Stable(1)	1	.9608	.1357	1.0480	.2022	1.0079	.0764	.9918	.0204	.9965	.0165
MA(1)	3	3.6480	5.4884	3.5814	1.2982	3.1893	.4743	3.1014	.2113	3.0898	.1775

The natural way to estimate the spectral measure is via

$$\hat{\mu}(B) = \frac{\sum_{i=1}^{n} 1_{(|X_i|>m, \frac{X_i}{|X_i|}\in B)}}{\sum_{i=1}^{n} 1_{(|X_i|>m)}};$$

see Resnick (2007). Here where m is some threshold, usually chosen to be one of the upper order statistics $R_{(k)}$ of the radial components of X_is. The task is to select this threshold so as to base the estimator on the maximal number of observations that fall in the tail region. One approach is to use the so-called "Stărică plots", introduced in Stărică (1999) and involving plotting the radial order statistics via $\left\{\left(\frac{R_{n-j}}{R_{n-k}}, \frac{R_{n-j}}{R_{n-k}}\frac{j}{k}\right), 1 \le j \le n\right\}$ for different values of k (see Resnick (2007)), p. 314. A "good" choice of k would result in a plot staying, roughly, around the value equal to 1, and one would use in estimating the spectral measure the observations corresponding to the radial upper order statistics from that k on. This procedure is mostly done visually. The automated procedure to pick k by minimizing some distance function does not generally seem to be reliable.

Our testing procedure for deciding which observations to use in estimating the spectral measure is as follows. We partition the unit sphere $\mathcal{S}^{d-1}$ into s subsets. If there is a sufficient number of observations with the spherical component faling in the ith subset, say, S_i, we apply (2.2) to the radial components of these observations and determine the value, say, r_i^*, so that all observations with the spherical component in S_i and radial component greater than r_i^*, are taken to be in the tail region. We then use $m = \max_i r_i^*$ in the estimation of the tail measure. In the case of a nonnegative bivariate sample, for instance, we simply use $S_i = [(i-1)\frac{\pi}{2s}, i\frac{\pi}{2s}]\}$ for $i = 1, 2, \ldots, s$.

We first simulate 5000 i.i.d. bivariate random vectors whose coordinate are independent Pareto(1) random variables. The true spectral measure puts equal mass at 0 and $\frac{\pi}{2}$, and nowhere else.

Based on our simulation, any k between 1000 and 2000 seems to produce an acceptable Stărică plot. The plot for the case $k = 1000$ is show in Figure 1. In Figure 2, we plot the cdf of the spectral measure using our method of choosing sample fraction with $s = 1$ and choices of θ_n. We also plot the cdf using the 1000th order statistic of the radius as threshold. The Stărică

method appears to take too much data into account, and our method with $\theta_n = \log n$ on the other hand takes too little data. Overall, the choice $\theta_n = (\log n)^2$ appears to perform the best.

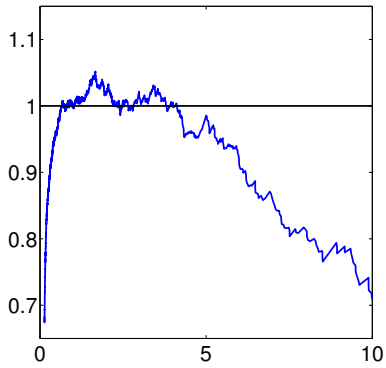

FIGURE 1. Starica plots for 5000 bivariate independent Pareto random variables using $k = 1000$

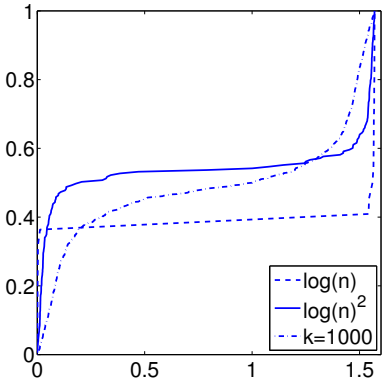

FIGURE 2. The spectral measure of 5000 independent bivariate Pareto random variables.

In order to explore what might happen of the multivariate observations are not, sequentially, i.i.d., we simulated 5000 absolute values of observations from a bivariate MA(1) model

$$\begin{pmatrix} Y_{1,i} \\ Y_{2,i} \end{pmatrix} = \begin{pmatrix} X_{1,i} + X_{1,i-1} + Z_i \\ X_{2,i} + X_{2,i-1} + Z_i \end{pmatrix},$$

where $X_{j,i}$, $j \in \{1, 2\}, i \in \{1, 2, ..., n\}$ are i.i.d. Student-t random variables with 3 degrees of freedom, Z_i are i.i.d. standard normal random variables,

and $\{X_{j,i}\}$ and $\{Z_i\}$ are independent. Once again, the true spectral measure has equal mass at 0 and $\frac{\pi}{2}$.

The Stărică plot for $k = 100$ is shown in Figure 3. In our simulations, any k from 50 to 500 produced similar Stărică plots, so it was hard to choose one of them as the best. In Figure 4, we plot the cdf of the spectral measure using our method of choosing sample fraction with $s = 1$ and two choices of θ_n. We also plot the cdf using the 100th order statistic of the radius as the threshold. The figure shows that none of the estimators performed particularly well in this situation.

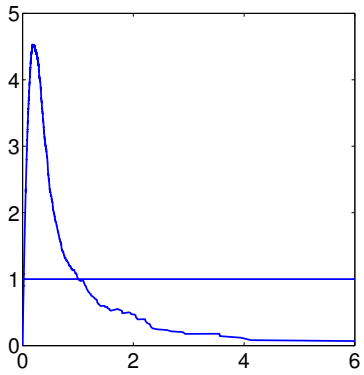

FIGURE 3. Starica plots for 5000 realizations of bivariate ARMA using k=100

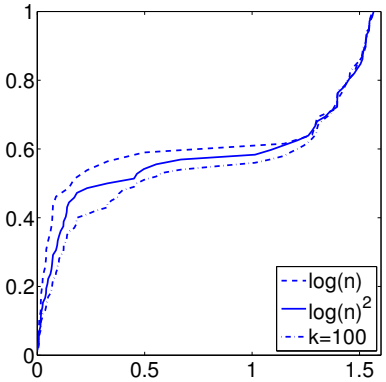

FIGURE 4. The spectral measure of 5000 realizations of bivariate ARMA.

5. Tail analysis of the Danish fire insurance data

In this section we explore the bivariate spectral measure analysis for a three-dimensional Danish fire insurance data set. The data consists of recorded damage to building, damage to content and loss of profits from fires in which the total loss exceed 1 million Danish Kroner (mDKK), from 1980 to 2002. This data set has 6870 recorded fires and accounts for 90% of the Danish fire insurance market.

We preprocessed the data as follows. We first used the Danish consumer price index to adjust all the losses to the 2011 values. Next, we omitted all incidents in which the total loss was below 1 mDKK in 1980 numbers which left us with a data set of size 2867. We applied the Pareto transform to the marginals, $x' = \dfrac{1}{1-\hat{F}(x)}$, using the corresponding marginal empirical c.d.f. This was done in order to convert the data to the same tail index 1 for each marginal.

Displayed in Figure 5 are the scatter plots of the processed data. Here X denotes the damage to content, Y denotes the loss of profit and Z denotes the damage to building.

Figure 6 shows the estimated spectral measures with $s = 1$ and different choices θ_n. In all plots, the damage to building displays marked independence in the extremes of both loss of profit and loss of content. As expected, the estimated spectral measure does not show extremal independence between the damage to content and loss of profit (but there a large mass seems to be concentrated in the neighborhood of 0).

Acknowledgments

We are grateful to Mette Havning for kindly providing us with the Danish fire insurance data, and to Pernille Rytgaard for the explanations provided with the data.

References

P. Billingsley (1999): *Convergence of Probability Measures*. Wiley, New York, 2nd edition.

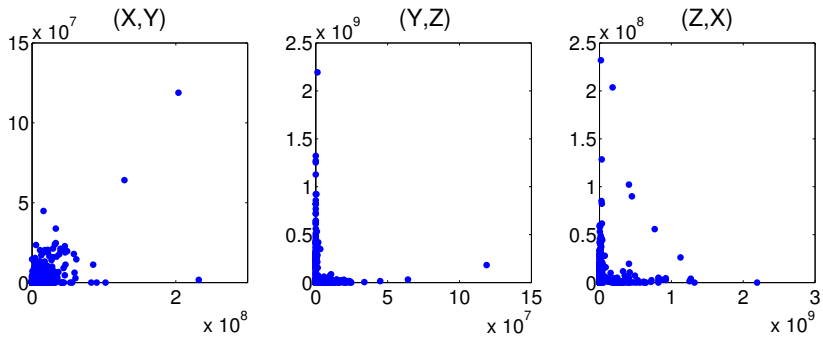

FIGURE 5. Scatterplot of the bivariate data after Pareto transform

R. DAHIYA and J. GURLAND (1972): Goodness of fit tests for the gamma and exponential distributions. *Technometrics* 14:791–801.

J. DANIELSSON, L. DE HAAN, L. PENG and C. DE VRIES (2001): Using a Bootstrap Method to Choose the Sample Fraction in Tail Index Estimation* 1. *Journal of Multivariate analysis* 76:226–248.

L. DE HAAN and A. FERREIRA (2006): *Extreme Value Theory: An Introduction*. Springer, New York.

P. DEHEUVELS, E. HÄUSLER and D. MASON (1988): Almost sure convergence of the Hill estimator. *Mathematical Proceedings of Cambridge Philosophical Society* 104:371–381.

A. DEKKERS and L. DE HAAN (1989): On the estimation of the extreme-value index and large quantile estimation. *Annals of Statistics* 17:1795–1832.

A. DEKKERS, J. EINMAHL and L. DE HAAN (1989): A moment estimator for the index of an extreme-value distribution. *Annals of Statistics* 17:1833–1855.

H. DREES and E. KAUFMANN (1998): Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Processes and their Applications* 75:149–172.

P. EMBRECHTS, C. KLÜPPELBERG and T. MIKOSCH (1997): *Modelling*

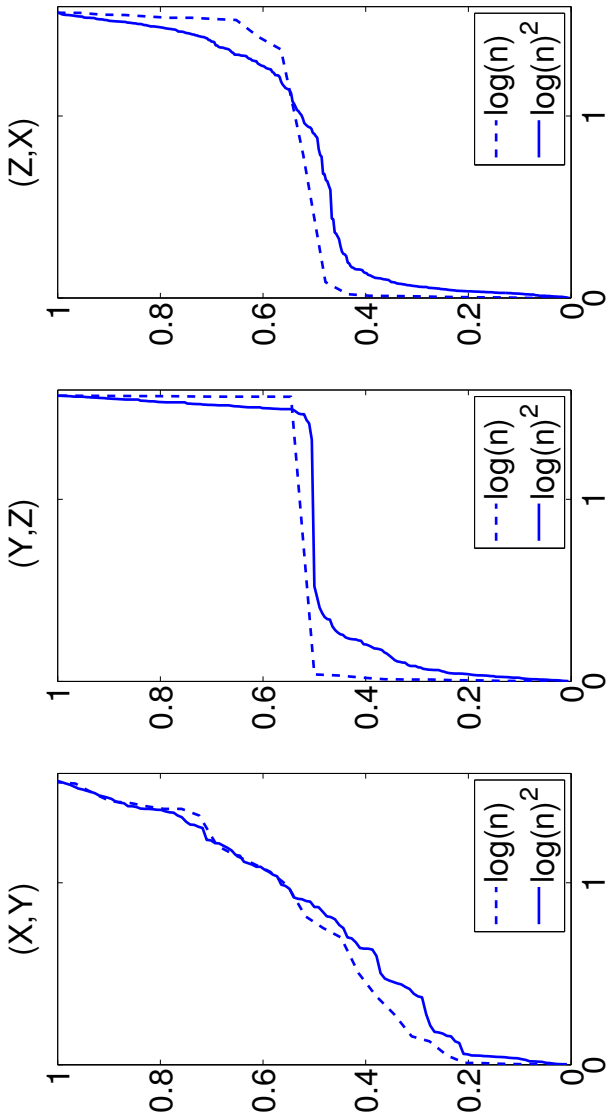

FIGURE 6. The spectral cdf in the 3 cases of bivariate data with different choices of s and θ_n.

Extremal Events for Insurance and Finance. Springer-Verlag, Berlin.

P. Hall (1990): Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *Journal of Multivariate analysis* 32:177–203.

P. Hall and A. Welsh (1985): Adaptive estimates of parameters of regular variation. *The Annals of Statistics* 13:331–341.

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

T. Hsing (1991): On tail index estimation using dependent data. *The Annals of Statistics* 19:1547–1569.

D. Mason (1982): Laws of large numbers for sums of extreme values. *Annals of Probability* 10:756–764.

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

S. Resnick (2007): *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer, New York.

S. Resnick and C. Stărică (1998): Tail index estimation for dependent data. *Annals of Applied Probability* 8:1156–1183.

S. Resnick and C. Stărică (1995): Consistency of Hill estimator for dependent data. *Journal of Applied Probability* 32:139–167.

S. Resnick and C. Stărică (1997): Smoothing the Hill estimator. *Advances in Applied Probability* 20:271–293.

D. Silvestrov (2004): *Limit theorems for randomly stopped stochastic processes*. Springer, London.

C. Stărică (1999): Multivariate extremes for models with constant conditional correlations. *Journal of Empirical Finance* 6:515–553.

W. Whitt (2002): *Stochastic-Process Limits. An Introduction to Stochastic-Process Limits and Their Applications to Queues*. Springer, New York.

Center for Applied Math, Cornell University, Ithaca, NY 14853
E-mail address: dhn43@cornell.edu

School of Operations Research and Information Engineering, and Department of Statistical Science, Cornell University, Ithaca, NY 14853
E-mail address: gs18@cornell.edu