MULTIVARIATE TAIL ESTIMATION WITH APPLICATION TO ANALYSIS OF COVAR

TILO NGUYEN AND GENNADY SAMORODNITSKY

Abstract. The quality of estimation of multivariate tails depends significantly on the portion of the sample included in the estimation. A simple approach involving sequential statistical testing is proposed in order to select which observations should be used for estimation of the tail and spectral measures. We prove that the estimator is consistent. We test the proposed method on simulated data, and subsequently apply it to analyze CoVaR for stock and index returns.

1. Introduction

It is one of the stylized facts about financial returns that instruments related by industry, or by geography, tend to show highly dependent extreme movements. Quantifying this extreme dependence is, clearly, important, but statistical estimation is tricky.

1991 Mathematics Subject Classification. Primary 62G32, 60G70.
Key words and phrases. extremes, tail estimation, tail measure, spectral measure, CoVaR, tail region.

This research was partially supported by the ARO grants W911NF-07-1-0078 and W911NF-12-10385, NSF grant DMS-1005903 and NSA grant H98230-11-1-0154 at Cornell University.

Figure 1. Daily log return
Figure 1 shows a scatter plot of the daily log returns on the stock of Bank of America and JP Morgan over a period of 4362 days from January 1996 to June 2012, next to a scatter plot of daily log returns on the British FTSE 100 index and the French CAC 40 index, from January 1995 to April 2012. It is obvious to the “naked eye” that the two indices have higher dependence in the extreme movements than the two stocks do. The difficulty with making this precise via statistical estimation is two-fold. First of all, there are simply not that many extreme observations. Secondly, it is not easy to decide, in a given sample, what observations “belong to the tail region”, and should be used to estimate the dependence of the extremes.

The main tools in describing such dependence are the tail and spectral measures of the returns. Their statistical estimation has been studied in detail, see e.g. de Haan (1985), de Haan and Resnick (1993), Einmahl et al. (1993), Einmahl et al. (2001), Einmahl and Segers (2009). All known methods are subject to the difficulties described above. We propose a method that allows a user to systematically decide what part of the sample to use in tail estimation. The method is fast and easy to automate.

The rest of the paper is organized as follows. In Section 2, we summarize the notions of regular variation, both univariate and multivariate. In Section 3, we describe how one can use the method of ranks to estimate the tail and spectral measures. Our method for deciding on the “tail part of the sample” is introduced in Section 4, where we also prove that our approach leads to a consistent estimator. In Section 5 we test our approach of estimating the spectral measure on simulated data, and compare it with the “rule of thumb” approach that has been commonly used. Finally, in Section 6 we analyze the recently introduced conditional risk measure, CoVaR, for stock and index returns.

2. Regular variation: A summary

Regular variation provides us with a framework that enables us to model multivariate extreme events. We now proceed with a quick summary of the background theory.
Let $F$ be a univariate distribution function. $F$ is said to have a regularly varying right tail of index $\alpha > 0$ if the tail probability function $\bar{F} = 1 - F$ satisfies

\[
\lim_{x \to \infty} \frac{\bar{F}(tx)}{\bar{F}(x)} = t^{-\alpha}
\]

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

The concept of regular variation can be extended to the multivariate case. Let $Z$ be a $d$-dimensional random vector with nonnegative components. Denote $\mathbb{E} = [0, \infty]^d \setminus \{0\}$. We say that $Z$ is multivariate regularly varying if there exist a function $b(t) \to \infty$ and a nontrivial (i.e. nonzero) Radon measure $\nu$ on $\mathbb{E}$, vanishing on the set of infinite points (i.e. points with at least one infinite coordinate), such that

\[
t P \left[ \frac{Z}{b(t)} \in \cdot \right] \vto \nu, \quad t \to \infty,
\]

vaguely in the space of nonnegative Radon measures on $\mathbb{E}$; see Resnick (2007). The measure $\nu$ is usually called the tail measure. The index $\alpha$ of the regular variation of the tail in this definition is hidden in the index of regular variation of the function $b$. It also appears in the scaling property the limit measure $\nu$ (the so-called tail measure) must possess: $\nu(c\cdot) = c^{-\alpha} \nu(\cdot)$ for every $c > 0$.

Equivalently, $Z$ is multivariate regularly varying with index $\alpha > 0$ if there exists a probability measure $S$ on $\mathbb{S}_+ = \mathbb{S}^{d-1} \cap [0, \infty)^d$ such that for all $x > 0$

\[
t P \left( \|Z\| > b(t)x, \frac{Z}{\|Z\|} \in \cdot \right) \Rightarrow cx^{-\alpha} S(\cdot), \quad t \to \infty,
\]

weakly on the unit sphere $\mathbb{S}^{d-1}$; see Resnick (2007). The probability measure $S$ is called the spectral measure of the distribution. One can choose an arbitrary norm on $\mathbb{R}^d$; this will determine the unit sphere and, thus, the spectral measure. The constant $c$ depends on the choice of the normalizing function $b$. 


Clearly, if the same normalizing function \( b \) is used in both definitions, then the tail and spectral measures are related by

\[
\nu \left( ||Z|| > x, \frac{Z}{||Z||} \in \cdot \right) = cx^{-\alpha} S(\cdot), \ x > 0.
\]

If the tail measure (or the spectral measure) is concentrated on the axes, it is common to speak of asymptotic (or tail) independence between the components of the random vector \( Z \), in the sense of very low likelihood that more than one component takes a large value at the same time. Otherwise, the components of the random vector \( Z \) are said to be tail dependent. A special example of tail independence is the case of a random vector with i.i.d. regularly varying marginal distributions.

We should note that these definitions are really useful if all the marginal tails of the random vector \( Z = (Z^{(1)}, \ldots, Z^{(d)}) \) are equivalent, i.e. the limit as \( z \to \infty \) of the ratio \( P(|Z^{(i)}| > z)/P(|Z^{(j)}| > z) \) exists for all \( i, j \), and is finite. This means, in particular, that all marginal tails are regularly varying with the same index \( \alpha \). In practice this is often an unreasonable assumption. There exists a framework of multivariate regular variation that allows different marginal tail indices. Let \( Z \) still be a \( d \)-dimensional random vector with nonnegative components. Specifically, we assume that there exist \( d \) normalizing functions \( b_i(t) \to \infty \), \( i = 1, \ldots, d \), such that both

\[
t P \left( \left( \frac{Z^{(i)}}{b_i(t)} \right) \in \cdot \right) \to \nu_{\alpha_i}, \ t \to \infty
\]

vaguely on \([0, \infty)\) for each \( i = 1, \ldots, d \) and

\[
t P \left( \left( \frac{Z^{(1)}}{b^{(1)}(t)}, \frac{Z^{(2)}}{b^{(2)}(t)}, \ldots, \frac{Z^{(d)}}{b^{(d)}(t)} \right) \in \cdot \right) \to \nu, \ t \to \infty
\]

vaguely on \( \mathbb{E} \), where \( \nu_{\alpha_i} \) is a measure on the Borel sets \([0, \infty)\) with the density \( c_i \alpha_i x^{-(1+\alpha_i)}, \ x > 0, \ \alpha_i, \ c_i > 0, \ i = 1, \ldots, d \) with respect to the Lebesgue measure. Moreover, \( \nu \) is a nontrivial Radon measure on \( \mathbb{E} \), vanishing on the set of infinite points, i.e. points which have \( \infty \) in at least one coordinate; it is also called the tail measure. The \( i \)th normalizing function \( b_i \) is regularly varying with index \( 1/\alpha_i \), \( i = 1, \ldots, d \). Since the indices of regular variation may be different in different directions, the normalizing functions may be different as well; this is the nonstandard version of regular variation. It can
be converted into the standard version of regular variation, with equivalent
tails, as follows. Let $F_i, i = 1, \ldots, d$ be the marginal distribution functions.
Define the generalized inverse function by
\begin{equation}
(2.5) \quad u_i(x) = \left(\frac{1}{1 - F_i}\right)^\leftarrow (x), \ i = 1, \ldots, d.
\end{equation}
Then
\begin{equation}
(2.6) \quad t \mathbb{P}\left[\left(\frac{u_i^\leftarrow (Z^{(i)})}{t}, i = 1, \ldots, d\right) \in \cdot\right] \overset{v}{\to} \nu_+(\cdot)
\end{equation}
vaguely on $\mathbb{E}$, while for each $i = 1, \ldots, d$,
\begin{equation}
(2.7) \quad t \mathbb{P}\left[\frac{u_i^\leftarrow (Z^{(i)})}{t} > x\right] \to x^{-1}, \ x > 0.
\end{equation}
Here $\nu_+$ is a nontrivial Radon measure on $\mathbb{E}$, vanishing on the set of infinite
points, with the scaling property
\begin{equation}
(2.8) \quad \nu_+(\cdot) = c^{-1}\nu_+(\cdot), \ c > 0.
\end{equation}
The measures $\nu$ and $\nu_+$ are related by
\begin{equation*}
\nu([0, x^{1/\alpha}]^c) = \nu_+([0, x]^c),
\end{equation*}
where $x^{1/\alpha}$ is defined as the vector with coordinates $x_i^{1/\alpha}, i = 1, \ldots, d$.

Therefore, the transformation (2.5) achieves both the standard global reg-
ular variation and the standard Pareto index 1 marginal distributions.

3. Estimation of the Tail Measure

Estimating the tail measure $\nu$ (or $\nu_+$) is of crucial importance in many
applications of stochastic models, and a number of estimators have been
designed for that purpose. The method chosen for this paper is the ranks
method that automatically standardizes the problem without requiring one
to estimate first the marginal quantile functions $u$ in (2.5). This method also
bypass the need for estimation of each of the marginal tail indices.

Suppose that conditions (2.3) and (2.4) for global and marginal regular
variation hold. Given a $d$-dimensional sample of size $n$, for each $j = 1, \ldots, d$,
we define
\begin{equation*}
r_{ij}^j = \sum_{m=1}^{n} 1_{[Z_m^{(j)} > Z_i^{(j)}]}, \ 1 \leq i \leq n,
\end{equation*}
which is the sequence of ranks. If $k = k(n)$ is an intermediate sequence, i.e. if $k \to \infty$ and $\frac{k}{n} \to 0$ as $n \to \infty$ then weak consistency of the empirical estimator

\[
\frac{1}{k} \sum_{i=1}^{n} \epsilon \left( \frac{k}{r_{i}^{(j)}, j=1, \ldots, d} \right) \Rightarrow \nu^*_\star
\]

in the vague topology on the space of nonnegative Radon measures on $\mathbb{E}$ holds; see pp. 311-312 in Resnick (2007). Note that the condition $\frac{k}{n} \to 0$ ensures that only “tail observations” affect the estimator. The second order behavior of this estimator is difficult; in the two-dimensional case, and under additional regularity assumptions, asymptotic normality was established in Einmahl et al. (2001).

To estimate the spectral measure, we first apply the polar transformation on the rank data, by

\[
(R_{i,k}, \theta_{i,k}) = \left( \frac{k}{r_{i}^{(j)}, j=1, \ldots, d}, \frac{\arctan \left( \frac{k}{r_{i}^{(2)}}, \frac{k}{r_{i}^{(1)}} \right)}{\pi}, i = 1, \ldots, n. \right)
\]

In this paper, we use the $l^\infty$ norm, and the polar transformation for nonnegative bivariate data can be written more explicitly in the form

\[
(R_{i,k}, \theta_{i,k}) = \left( \frac{k}{r_{i}^{(j)}, j=1, 2}, \frac{\arctan \left( \frac{k}{r_{i}^{(2)}}, \frac{k}{r_{i}^{(1)}} \right)}{\pi}, i = 1, \ldots, n. \right)
\]

When we apply the transformation (3.2) to the points of a point process, we obtain a continuous mapping from the space of point processes to $(0, \infty) \times S_+$, so we can use (3.1) and continuous mapping theorem. Recalling the representation (2.2) of the tail measure (applied to $\nu^*_\star$ in this case), we obtain

\[
\frac{1}{k} \sum_{i=1}^{n} 1_{\left( R_{i,k} > 1, \theta_{i,k} \in \cdot \right)} \Rightarrow cS_*(\cdot)
\]

weakly on the unit sphere, where $S_*$ is the spectral measure corresponding to $\nu^*_\star$ and $c$ is a normalizing constant so that $S_*$ is a probability measure.
This then gives us the following consistent estimator for $S_*$:

$$S_{k,n}^*(\cdot) := \sum_{i=1}^{n} \frac{1(\mathbf{R}_{i,k}>1, \theta_{i,k} \in \cdot)}{\sum_{i=1}^{n} 1(\mathbf{R}_{i,k}>1)} \Rightarrow S_*(\cdot).$$

(3.3)

Note that here and in the sequel we identify the angular component of a non-zero vector with the projection of the vector on the unit sphere (the $l^\infty$ unit sphere in this paper.)

4. The choice of $k$

In practice, the estimator $S_{k,n}^*$ in (3.3) shares with other “tail related” estimators the sensitivity to the choice of $k$. A decision on what $k$ to use can be viewed as a decision on what part of the sample corresponds to “tail observations”. For estimators of the index of univariate regular variation, $k$ is the number of upper order statistics to use in the estimator, and several visual techniques have been suggested to pick out that number, including the most popular choice, the Hill plot; see Section 4.4 in Resnick (2007). Other methods to optimize the asymptotic behaviour of the tail estimators have been suggested in e.g. Danielsson et al. (2001) and Drees and Kaufmann (1998).

In our previous work (Nguyen and Samorodnitsky, 2012), we proposed a sequential testing method for deciding “where the tail begins” in which the significance level increases with the sample size. Our approach is based on a simple idea. Let $(X_j)$ be a nonnegative i.i.d. sequence with a distribution $F$. It is well known that, under the assumption (2.1) of regular variation, we have the following weak convergence

$$N_n = \sum_{i=1}^{n} \delta_{X_i/c_n} \Rightarrow N_*$$

in the vague topology on the space of nonnegative Radon measures on $(0, \infty]$ (see e.g. Proposition 4.26 in Resnick (1987)). Here $\delta_x$ is the point mass at $x$, and $(c_n)$ a positive sequence satisfying $\bar{F}(c_n) \sim 1/n$ as $n \to \infty$. Further, $N_*$ is a Poisson random measure on $(0, \infty]$ with mean measure $\mu_*(x, \infty] = x^{-\alpha}$, $x > 0$. Accordingly, our procedure for deciding on the number $k$ of the upper order statistics to use in the tail estimation consists of sequentially
testing the samples \( \{ \log(X_{n-i,n}/X_{n-k,n}), i = 0, 1 \ldots k - 1 \} \), with increasing value of \( k \), for the null hypothesis of exponential distribution. Our choice of \( k \) used in 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.

To test for exponentiality we used the statistic

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

under the null hypothesis it converges to the standard normal distribution (Dahiya and Gurland (1972)). Our choice of the significance level of the test resulted in

\[
N_n := \inf \left\{ k : 1 \leq k \leq n, |Q_{k,n}| \geq \omega \sqrt{\frac{\lambda_n}{k}} \right\},
\]

where \( \omega > 0 \) and \( (\lambda_n) \) is an increasing to infinity sequence such that \( \lambda_n = o \left( \frac{2(\log n)}{n^{1+2|\rho|}} \right) \). Here \( \rho \) is the parameter of the assumed second order regular variation; we state this assumption explicitly below. The role of parameter \( \omega \) is limited in comparison with that of the sequence \( (\lambda_n) \) (and it can be, in principle, absorbed into the latter). Explicitly, parameter \( \omega \) appears in the limiting result \( \frac{N_n}{\lambda_n} \Rightarrow \tau_\omega \), where \( \tau_\omega \) is the first time a standard Brownian motion hits \( \pm \omega \). For practical purposes, one can fix \( \omega \) and then find an appropriate \( \lambda_n \). This is the procedure we follow in the examples later in the paper.

The choice of \( k \) described above (stop the first time the exponentiality hypothesis is rejected) is only one of several possibilities; another possibility is to stop the last time the exponentiality hypothesis is accepted. We use the choice of \( k \) for which the asymptotic theory in Nguyen and Samorodnitsky (2012) applies.

For estimating the spectral measure in the multivariate regular variation context, one can adapt the one-dimensional rule of thumb and use that part of the sample whose radial component is in the top 5% among all the observations. The only existing systematic method to choose \( k \) appears to be the “Stărică plots”, introduced in Stărică (1999). It is a visual method, based on the fact that the tail measure \( \nu \) has the scaling property (2.8). This
property is checked empirically for a range of scales, and a “good” choice of $k$ (of the upper radial order statistics) is the one for which the estimated tail measure seems to be close to satisfying the scaling property. This checking is performed graphically, and a “good” choice of $k$ is the one where the plot stays as much as possible around the value equal to 1; see p. 314 in Resnick (2007). One would then use the observations of the norm larger than the $k$th radial order statistic in estimating the spectral measure.

In this work we extend our sequential testing procedure to the multivariate framework of estimating the spectral measure. Specifically, we start by applying the testing procedure to each set of marginal observations. For the $j$th marginal, $j = 1, \ldots, d$ we calculate the first rejection time, $N^{(j)}$, via

$$N_n^{(j)} := \inf \left\{ k : 1 \leq k \leq n, |Q_{k,n}| \geq \omega_j \sqrt{\frac{\lambda_n^{(j)}}{k}} \right\},$$

using the data for that marginal, with some $\omega_j > 0$ and some sequence $(\lambda_n^{(j)})$. Any choice of $(\omega_j)$ is possible; again, we can fix them and find appropriate $(\lambda_n^{(j)})$. We set our choice of $k$ to be

$$N_n = \bigwedge_{j=1}^d N_n^{(j)}.$$

Taking the minimum in (4.2) means that our procedure is, generally, conservative about deciding on the “tail part” of the data. Other choices of $k$ are possible, with the largest rejection point, $N_n = \bigvee_{j=1}^d N_n^{(j)}$, being an obvious choice. When tested on simulated data, the latter choice of $k$ often appeared to be “too generous” with deciding which observations are “in the tails”.

The consistency of the resulting estimator of the tail measure is proven in the following theorem. In order to state our result, we introduce formally the notion of second order regular variation. For a univariate regular varying distribution $F$ with index $\alpha$, let $u$ be the generalized inverse function in (2.5). We assume that there exists $\rho < 0$ and a function $G$ regularly varying at infinity with exponent $\rho$ such that for all $x > 0$

$$\lim_{r \to \infty} \frac{\frac{u(rx)}{u(r)} - x^{1/\alpha}}{G(r)} = x^{1/\alpha} x^\rho - 1,$$
It is important to note that our consistency result does require the second-order regular variation assumption, unlike most of the previous approaches, in which only the (first-order) regular variation is used. When a non-random number $k$ of upper order statistics is used in estimation, the second-order regular variation is often used to obtain asymptotic normality of the estimator; see e.g. de Haan and Fereira (2006). With a random choice of $k$ as in our procedure, we needed to use the second-order regular variation even to obtain consistency of the estimator, even in the one-dimensional case; see Nguyen and Samorodnitsky (2012). In fact, the behaviour of the random variable $N_n$ above seems to undergo a phase transition if the condition $\lambda_n = o\left(\frac{2|\rho|}{n^{1+2|\rho|}}\right)$ breaks down. It remains to be understood what the behaviour of our estimator is if this happens (or if the assumption of the second-order regular variation does not hold at all). However, as we will see in the sequel, the choice of the sequences $(\lambda_n^{(j)})$ we advocate (as leading to a good performance) is logarithmic and, hence, does not require knowing the value of the exponent of the second-order regular variation.

**Theorem 1.** Assume that the marginal and the joint regular variation conditions (2.3) and (2.4) hold. Assume, further, that the $j$th marginal distribution satisfies the second order regular variation condition with exponent $\rho_j < 0$, $j = 1, \ldots, d$. Let $\omega_j > 0$ and $(\lambda_n^{(j)})$ be an increasing to infinity sequence such that $\lambda_n^{(j)} = o\left(\frac{2|\rho|}{n^{1+2|\rho_j|}}\right)$, $j = 1, \ldots, d$, and let $N_n$ be defined by (4.2). Then

$$\frac{1}{N_n} \sum_{i=1}^{n} \epsilon \left(\frac{N_n}{r_i^{(j)}}\right)_{j=1,\ldots,d} \Rightarrow \nu_*$$

weakly in the vague topology.

**Proof.** See Appendix

The following is an immediate corollary of Theorem 1. It provides the estimator of the spectral measure we will use in the rest of the paper.
Corollary 2. Under the assumptions of Theorem 1,

\[ S_{*N,n}(\cdot) = \frac{\sum_{i=1}^{n} 1(R_{i,N} > 1, \theta_{i,N} \in \theta(\cdot))}{\sum_{i=1}^{n} 1(R_{i,N} > 1)} \Rightarrow S_*(\cdot), \]

weakly on the unit sphere.

Note that this result is a weak convergence result. Other procedures have been shown to lead, under appropriate assumptions, to uniform convergence over certain classes of sets; see e.g. Einmahl et al. (1993) and Einmahl et al. (2001).

5. Testing the estimator on simulated data

In this section, we evaluate our estimator (4.4) of the spectral measure on simulated data. We consider 3 two-dimensional examples. The first example looks at two scenarios with asymptotic independence (so that the true spectral measure is concentrated at the extreme points of the arc), the second example looks at a case of a complete tail dependence (so that the true spectral measure is concentrated at a single point in the interior of the arc), and the last example considers a situation where the true spectral measure is absolutely continuous with respect to the Lebesgue measure on the arc. In each case we ran simulations with sample sizes \( n = 1000, n = 5000 \) and \( n = 20000 \), and each simulation was performed 500 times. The results we report are the averages obtained from the 500 simulations. In all cases, we choose the numbers \( N(j), j = 1, 2 \) by (4.1) with \( \lambda_n^{(j)} = (\log n)^2 \), as recommended in Nguyen and Samorodnitsky (2012). This choice of \( \lambda_n \) was empirically observed to work well in the univariate estimation of the exponent of regular variation for a variety of models. It has an added advantage of satisfying the condition \( \lambda_n = o \left( n^{\frac{2|\rho|}{1+2|\rho|}} \right) \) regardless of the actual value of the exponent \( \rho \) in the second-order regular variation. We also choose \( \omega_j = 1.64 \) for all \( j \). This is the rejection level of the original exponentiality test (Dahiya and Gurland (1972)) with significance level .1, as in Nguyen and Samorodnitsky (2012).
Example 5.1. The generic model is $X^{(i)} = |Z + Y^{(i)}|$, $i = 1, 2$ where $Z \sim N(0, 1)$ and $Y^{(1)}$ and $Y^{(2)}$ are independent identically distributed random variables independent of $Z$, with regularly varying tails.

(a) $Y^{(i)}$, $i = 1, 2$ have a $t$ distribution with 3 degrees of freedom. In this case, the second order regular variation index is $\rho = -\frac{2}{3}$.

(b) $Y^{(i)}$, $i = 1, 2$ have a Generalized Pareto distribution with parameters $\mu = 1, \alpha = 1, \sigma = 2$. Recall that the distribution function of such a distribution is given by

$$G_{\alpha, \sigma}(y) = 1 - \left(1 + \frac{1}{\alpha} \frac{(y - \mu)}{\sigma}\right)^{-\alpha}, \quad y \geq \mu. \quad (5.1)$$

In this case, $\rho = -1/\alpha$.

Our decision on the tail part of the sample plays a role not only in the estimation of the spectral measure, but also in estimation of the parameters of the distribution. Estimating the latter is not necessary for the estimation of the spectral measure, which is our main goal. However, the interplay between estimation of the parameters and the choice of the part of the data to use for it, is instructive, and we include a short discussion of the resulting insights.

In Example 5.1 (a) we estimate the marginal tail index (coinciding with the number of the degrees of freedom) by the Hill estimator, which is defined as follows. If $X_{1,n} \leq X_{2,n} \leq \ldots \leq X_{n,n}$ are the order statistics from a positive sample, $X_1, \ldots, X_n$, then the Hill estimator based on $k$ upper order statistics is

$$H_{k,n} := \frac{1}{k} \sum_{i=0}^{k-1} \log \frac{X_{n-i,n}}{X_{n-k,n}}. \quad (5.2)$$

If the observations are i.i.d., with a regularly varying right tail with tail index $\alpha$, then, with $n \to \infty$, $k \to \infty$ and $\frac{k}{n} \to 0$, the Hill estimator $H_{k,n}$ converges in probability to $\gamma = \frac{1}{\alpha}$; see Mason (1982)), who also shows, under certain conditions of $k = k_n$, equivalence between consistency of the Hill estimator and regular variation. We follow Nguyen and Samorodnitsky (2012) in our choice of $k$ in estimating the $j$th marginal tail index. We choose it to be $N_n^{(j)}$ as above, $j = 1, 2$. 


In the case of the generalized Pareto distribution of Example 5.1 (b) we use the $N_n^{(j)}$th largest order statistic of $X_j$ as the location parameter $\mu^{(j)}$, $j = 1, 2$. The reason for this is that a generalized Pareto distribution is used as a model for the “excess tails” once a certain threshold has been crossed, for data in which the “real distribution” is not known; this is the so called POT (peaks over threshold) method; see e.g. Embrechts et al. (1997). Choosing the threshold is often difficult, as it is supposed to represent a point beyond which the “tail behaviour” of the distribution is observed. This fits well into our approach, where $N_n^{(j)}$th largest order statistic represents the estimated “beginning of the tails”. Therefore, instead of using the maximum likelihood estimation to estimate all the parameters of the known distribution in this example, we prefer to set the initial point $\mu^{(j)}$ of the generalized Pareto distribution, by what the data indicates, i.e. by the $N_n^{(j)}$th largest order statistic. Once the location parameter is decided on, the remaining parameters $\alpha$ and $\sigma$ are obtained via maximum likelihood estimation.

We note that in both parts of Example 5.1 the tails are asymptotic independent, and the spectral measure splits its mass equally between the two extreme points, at the angles 0 and $\pi/2$.

Our general format for reporting the results of estimation is as follows. We report the numbers $N_n^{(j)}$, $j = 1, 2$, averaged over all the runs, as well as their minimum, $N_n$, also averaged over all the runs. To measure how close the estimated spectral measure is to the true one, we calculate and report the integrated squared bias and variance of the estimator, given, respectively, by $\int_0^{\pi/2} (S_\ast([0, \theta]) - E(\hat{S}_\ast([0, \theta])))^2\, d\theta$ and $\int_0^{\pi/2} \text{Var}(\hat{S}_\ast([0, \theta]))\, d\theta$. Here $S_\ast([0, \theta])$ is the mass assigned by the true spectral measure $S$ to the part of the $l^\infty$ unit sphere whose angle, in the polar coordinates, is in the interval $[0, \theta]$, $0 \leq \theta \leq \pi/2$. Similarly, $\hat{S}_\ast([0, \theta])$ is the mass assigned by the estimated spectral measure to the same set; these are the “cumulative distribution functions” indexed by the angle. The expected value and variance in the integrals are replaced by sample mean and sample variance. We compare the calculated biases and variances obtained using our approach to the “rule of thumb” choice of the part of the sample whose radial component
is in the top 5% among all the observations. The alternative “Stărică plot” is
not useful in this situation; it cannot distinguish between different $k$ in the
relevant range.

The estimation results for the $t$ distribution of Example 5.1 (a) are re-
ported in Table 1 and Table 2. We present estimates of the spectral measure
from 20 simulation in Figure 2.

The estimation results for the Generalized Pareto distribution of Example
5.1 (b) are similarly reported in Table 3, Table 4, Table 5 and Figure 3.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\alpha^{(1)}$</th>
<th>$\alpha^{(2)}$</th>
<th>$N^{(1)}$</th>
<th>$N^{(2)}$</th>
<th>$N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>2.9906</td>
<td>2.9819</td>
<td>111.814</td>
<td>110.33</td>
<td>82.336</td>
</tr>
<tr>
<td>5000</td>
<td>3.1411</td>
<td>3.1174</td>
<td>204.822</td>
<td>210.2520</td>
<td>131.44</td>
</tr>
<tr>
<td>20000</td>
<td>3.1011</td>
<td>3.1217</td>
<td>330.7920</td>
<td>299.7080</td>
<td>195.9760</td>
</tr>
</tbody>
</table>

Table 1. Simulation results for Example 5.1 (a)

<table>
<thead>
<tr>
<th>$n$</th>
<th>Adaptive bias</th>
<th>Adaptive variance</th>
<th>5% bias</th>
<th>5% variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>.0337</td>
<td>.0026</td>
<td>.0246</td>
<td>4.784 * 10^{-4}</td>
</tr>
<tr>
<td>5000</td>
<td>.0127</td>
<td>.0016</td>
<td>.0243</td>
<td>9.836 * 10^{-5}</td>
</tr>
<tr>
<td>20000</td>
<td>.0065</td>
<td>6.519 * 10^{-4}</td>
<td>.0243</td>
<td>2.433 * 10^{-5}</td>
</tr>
</tbody>
</table>

Table 2. Simulation results for Example 5.1 (a)
Table 3. Simulation results for Example 5.1 (b)

<table>
<thead>
<tr>
<th>n</th>
<th>$N^{(1)}$</th>
<th>$N^{(2)}$</th>
<th>$N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>153.922</td>
<td>148.026</td>
<td>96.972</td>
</tr>
<tr>
<td>5000</td>
<td>242.882</td>
<td>226.010</td>
<td>138.098</td>
</tr>
<tr>
<td>20000</td>
<td>315.744</td>
<td>31310.996</td>
<td>192.712</td>
</tr>
</tbody>
</table>

Table 4. Simulation results for Example 5.1 (b)

<table>
<thead>
<tr>
<th>n</th>
<th>Adaptive bias</th>
<th>Adaptive variance</th>
<th>5% bias</th>
<th>5% variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>.0328</td>
<td>.0037</td>
<td>.0217</td>
<td>3.886 * 10^{-4}</td>
</tr>
<tr>
<td>5000</td>
<td>.0119</td>
<td>.0015</td>
<td>.0214</td>
<td>7.889 * 10^{-5}</td>
</tr>
<tr>
<td>20000</td>
<td>.0062</td>
<td>5.282 * 10^{-4}</td>
<td>.0214</td>
<td>2.026 * 10^{-5}</td>
</tr>
</tbody>
</table>

Table 5. Simulation results for Example 5.1 (b)

<table>
<thead>
<tr>
<th>n</th>
<th>$\alpha^{(1)}$</th>
<th>$\alpha^{(2)}$</th>
<th>$\sigma^{(1)}$</th>
<th>$\sigma^{(2)}$</th>
<th>$\mu^{(1)}$</th>
<th>$\mu^{(2)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>.9948</td>
<td>.9992</td>
<td>20.4228</td>
<td>22.2245</td>
<td>19.6233</td>
<td>22.5595</td>
</tr>
<tr>
<td>5000</td>
<td>1.0209</td>
<td>1.0042</td>
<td>70.3615</td>
<td>76.6433</td>
<td>77.2565</td>
<td>79.6734</td>
</tr>
<tr>
<td>20000</td>
<td>1.0172</td>
<td>1.0198</td>
<td>218.7182</td>
<td>210.5060</td>
<td>237.6335</td>
<td>231.3167</td>
</tr>
</tbody>
</table>

Figure 3. Estimated spectral measures from 20 simulations of Example 5.1 (b) with $n = 5000$ and $n = 20000$.

Observe that, while the “5% rule of thumb” results in a lower bias and variance than our estimator for the smallest sample size $n = 1000$, the situation is reversed for larger sample sizes. By construction, the bias of the 5% rule stays constant as sample size increases. On the other hand, our adaptive
method quickly picks up the emerging tails and improves the quality of the estimation as the sample size increases.

As far as the estimation of other parameters is concerned, we see that we are estimating the tail exponent $\alpha$ well even for small sample sizes. However, we are missing the true values of the location parameter $\mu$ and the scale parameter $\sigma$ of the Generalized Pareto distribution of Example 5.1 (b) by a wide margin. The reason, as we explained above, is that in deciding the location parameter we pretend that we do not not know the parametric form of the model and set the location parameter by the appropriate order statistic, similar to the POT method. This results in inflation of the location parameter. Recall that, in a Generalized Pareto distributed with parameters $(\mu, \alpha, \sigma)$, given that $X > \mu + \delta$ for some $\delta > 0$, $X$ has again a Generalized Pareto distribution, this time with parameters $(\mu + \delta, \alpha, \sigma + \delta/\alpha)$. Hence inflating the location parameter leads to overestimating the scale parameter as well. Note, however, that the tail exponent $\alpha$ is still adequately estimated.

Example 5.2. This is an example of a bivariate law with a complete tail dependence. Let $Y$ be Pareto(2) distributed, independent of two i.i.d. standard normal random variables $Z^{(1)}$ and $Z^{(2)}$. Let $X^{(1)} = Z^{(1)} + Y$ and $X^{(2)} = Z^{(2)} + Y^2$. The true spectral measure in this case has mass only at a single point, $\pi/4$. We present the results in Table 6, Table 7 and Figure 4. The lessons we learn here are similar to the lessons learned in Example 5.1; for larger sample sizes our algorithm estimates the spectral measure more precisely than the 5% rule does. Again, the reason here is that the bias of the 5% method remain constant as sample size increases, while our algorithm detects emerging tails for larger sample sizes.
### Table 7. Simulation results for Example 5.2

<table>
<thead>
<tr>
<th>n</th>
<th>Adaptive bias</th>
<th>Adaptive variance</th>
<th>5% bias</th>
<th>5% variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>.0384</td>
<td>.0031</td>
<td>.0248</td>
<td>.0014</td>
</tr>
<tr>
<td>5000</td>
<td>.0184</td>
<td>.0012</td>
<td>.0261</td>
<td>2.821 * 10^{-4}</td>
</tr>
<tr>
<td>20000</td>
<td>.0099</td>
<td>5.907 * 10^{-4}</td>
<td>.0261</td>
<td>7.525 * 10^{-5}</td>
</tr>
</tbody>
</table>

**Example 5.3.** In this example the true spectral measure is absolutely continuous with respect to the Lebesgue measure on the $l^\infty$ unit sphere. We consider two cases.

(a) $X^{(i)} = Y^{(i)}$, $i = 1, 2$ with $(Y^{(1)}, Y^{(2)})$ having the bivariate Cauchy distribution restricted to $(0, \infty)^2$. This distribution has the density given by

$$f(y^{(1)}, y^{(2)}) = \frac{2}{\pi} \left(1 + (y^{(1)})^2 + (y^{(2)})^2\right)^{-\frac{3}{2}}.$$  

(b) $X^{(i)} = Y^{(i)} + Z^{(i)}$, $i = 1, 2$, with $(Y^{(1)}, Y^{(2)})$ having the bivariate Cauchy distribution restricted to $(0, \infty)^2$, independent of i.i.d. Pareto(2) random variables $Z^{(1)}$ and $Z^{(2)}$.

Note that, in both cases, the “c.d.f.” of the true spectral measure on the unit sphere with respect to the $l_p$ norm is:

$$S_\ast([0, \theta]) = \frac{1}{\sqrt{2}} \int_0^\theta ||\sin(\psi), \cos(\psi)||_p d\psi,$$

$0 \leq \theta \leq \pi/2$; it is, therefore, absolutely continuous with respect to the Lebesgue measure for any $1 \leq p \leq \infty$. It is, actually uniform (i.e. coincides,
The results for part (a) of the example are reported in Table 8, Table 9 and Figure 5, and for part (b) of the example in Table 10, Table 11 and Figure 6.

We see that, as expected, it is harder to estimate the spectral measure in the “contaminated” case (b). Nevertheless, our approach works reasonably well. It is instructive to observe that in the pure bivariate Cauchy case of part (a) the 5% rule works very well in all situations. This is because the entire up to a multiplicative constant, with the Lebesgue measure) if $p = 2$. Recall that we are using the $l^\infty$ norm for our estimation. The (marginal) exponent of the second-order regular variation in this example is $\rho = -2$.

The results for part (a) of the example are reported in Table 8, Table 9 and Figure 5, and for part (b) of the example in Table 10, Table 11 and Figure 6.

We see that, as expected, it is harder to estimate the spectral measure in the “contaminated” case (b). Nevertheless, our approach works reasonably well. It is instructive to observe that in the pure bivariate Cauchy case of part (a) the 5% rule works very well in all situations. This is because the entire
distribution is rotationally invariant and, hence, in this special case all parts of the sample provide exactly the same information, and one does not need to detect the tail part of the sample. In contrast, in the “contaminated” case of part (b), it is important to know where the tails begin, and our approach, again, is superior to the 5% rule, apart from the smallest sample size situation.

It is sometimes of interest to estimate the actual density of the spectral measure, as opposed to the CDF. We use a kernel density estimation procedure with the values of $\theta_{i,N}$ in (4.4) such that $R_{i,N} > 1$ as the input. Since the spectral measure is concentrated on the interval $[0, \pi/2]$, we use the beta
kernel of Chen (1999) (with bandwidth $b = 0.1$) that takes into account the proximity of an observation to the endpoints of the interval. The results of estimating the spectral density for the contaminated Cauchy example are presented in Figure 7. Since the density, as a derivative of the CDF, is more sensitive to the noise in the data than the latter, it is not surprising that only at relatively large sample sizes we obtain a reasonable fit. Even here the true spectral density is estimated much better by our procedure than using the 5\% rule.

6. Analyzing CoVaR

In this section we apply our approach of estimating multivariate tails to the problem of evaluating the conditional risk measure CoVaR. It is based on the widely used risk measure of Value-at-Risk.

Recall that for $0 < q < 1$, and close to 1, $VaR_q^{(j)}$ is defined as the $q$-quantile of the loss distribution $F^{(j)}$ of the $j$th asset. To be precise,

$$VaR_q^{(j)} = F^{(j)}\leftarrow(q)$$

(here and in the sequel we will assume, for simplicity of notation, that the appropriate distributions are continuous). In light of the recent financial crisis, efforts have been made to capture better the spillover effects between financial institutions. This gave rise to the conditional risk measure CoVaR defined by

$$P(X^{(j)} \geq CoVaR_q^{(j)|i}(C(X^{(i)}))) = 1 - q.$$  \hspace{1cm} \text{(6.1)}

or

$$CoVaR_q^{(j|i)} = VaR_q(X^{(j)}|C(X^{(i)})) .$$

Here $C(X^{(i)})$ is some event describing a measure of distress of institution $i$. In this section we will consider events of the type $X^i \geq VaR_r^{(i)}$ for some $0 < r < 1$, close to 1. We are also interested in estimating the natural versions of CoVaR in the case of more than one institution being in distress.

The original definition of CoVaR proposed by Adrian and Brunnermeier (2011) used conditioning on $X^i = VaR_r^{(i)}$. Improvements to this definition were suggested by Klyman (2011). The alternative version of CoVaR to
Figure 7. 20 kernel estimations for the density function of contaminated Cauchy tail with (1) \( n=5000 \) using our adaptive method, (2) \( n=20000 \) using our adaptive method, (3) \( n=20000 \) using the 5\% rule.
involving conditioning on \( X^i \geq VaR_r(i) \) was proposed by Girardi and Ergun (2012). In this paper, as well as in Mainik and Schaanning (2012), it was shown this conditioning leads to more consistent behaviour of CoVaR as a risk measure. Other systemic risk measures were proposed by Zhou (2010); the estimation procedure there also uses the multivariate extreme value theory.

We approximate the marginal quantiles \( VaR_r(i), i = 1, \ldots, d \) using the Generalized Pareto model (5.1) of Example 5.1 (b). As before, we set the location parameter to be the \( N_i \)th order statistic in (4.1), and the remaining two parameters by the maximum likelihood estimation. Consequently, we estimate, using (2.6), the probabilities in (6.1) for \( t \) of the order \( u_j(1/(1-r)) \) (where \( u_j \) is the \( j \)th marginal generalized inverse function) as follows.

\[
P(Y(j) > t \mid Y(i) > VaR_r(i), i = 1, \ldots, d, i \neq j) = P\left( \frac{1 - r}{1 - F(j)(Y(j))} > \frac{1 - r}{1 - F(i)(Y(i))} \right) > 1, i = 1, \ldots, d, i \neq j
\]

\[
\approx \nu_s(y \in \mathbb{R}^d : y(j) > 1 - F(j)(t), y(i) > 1, i = 1, \ldots, d, i \neq j) / \nu_s(y \in \mathbb{R}^d : y(i) > 1, i = 1, \ldots, d, i \neq j). \tag{6.2}
\]

When using the estimator (6.2) in practice, one needs to deal with the following issue. Since we would like to estimate high conditional quantiles, we will need to use the estimator for \( t \) such that \( (1 - r)/(1 - F(j)(t)) \) is very large. In such cases the sample evaluation of the value of the tail measure \( \nu_s \) in the numerator of the ratio in the right hand side of (6.2), described in Theorem 1, may be based on a very small number of observations. We have chosen to resolve this difficulty by using the scaling property (2.8) of \( \nu_s \): for every \( L > 0 \) we can rewrite the ratio in the right hand side of (6.2) as

\[
\approx \nu_s(y \in \mathbb{R}^d : y(j) > 1, i = 1, \ldots, d, i \neq j) / \nu_s(y \in \mathbb{R}^d : y(i) > 1, i = 1, \ldots, d, i \neq j). \tag{6.3}
\]
By choosing an appropriately large $L$, the difficulty described above is reduced. Another difficulty arises, however. Notice that an appropriate choice of $L$ is, by the nature of the problem, $t$-dependent. We are estimating a function of $t$ that is, clearly, monotone, and so is the expression in (6.3). However, we are using empirical estimates of $\nu_*$ both in the numerator and the denominator of (6.3). With a $t$-dependent $L$, the monotonicity may be occasionally violated. We have found that a choice of $L$, that seems to reduce the frequency and the effect of such violations, is

$$L_t = \left\| \left( \frac{1 - r}{1 - F(t)} \right)^{-1}, \ldots, 1 \right\|_2.$$  

We demonstrate our procedure on two examples, that show markedly different behavior of $\text{CoVaR}$: daily net returns of shares of 4 major US financial institutions and daily net returns of 4 major European indices.

**Example 6.1. Daily returns of 4 major US financial companies**

We analyze the daily net returns on the stock of Bank of America, JP Morgan, Morgan Stanley and Citigroup, from January 2, 1996 to June 30, 2012. The data set contains 4169 observations.

As described above, we concentrate on the respective losses $Y^{(i)}$, $i = 1, \ldots, 4$ of these 4 stocks. We use as the distress events $C_i$ the exceedance events $Y^{(i)} > \text{VaR}^{(i)}_r$, $i = 2, 3, 4$ with $r = .95$ and $r = .99$, and estimate the conditional quantiles of the losses on the stock of Bank of America. We specifically concentrate on the change in these conditional quantiles as we condition on more and more distress events for other companies. That is, we will estimate the conditional probabilities $P(Y^{(1)} > t|C_2)$, $P(Y^{(1)} > t|C_2, C_3)$ and $P(Y^{(1)} > t|C_2, C_3, C_4)$, as well as the resulting conditional quantiles. For comparison, we estimate unconditional quantiles as well.

The results are presented in Figure 8. Note that including every additional distress event adds mass to the conditional tail of the loss distribution on the stock of Bank of America and affects, correspondingly, the resulting value of $\text{CoVaR}$.

Some estimated quantiles for losses of Bank of America are reported in Table 12.
Quantile | Unconditional | Given $C_2$ | Given $C_2, C_3$ | Given $C_2, C_3, C_4$
---|---|---|---|---
90% | .026 | .116 | .157 | .174
95% | .039 | .157 | .218 | .238
99% | .090 | .292 | .380 | .411

Table 12. Estimated quantiles for losses of Bank of America, unconditional and conditioned on $Y^{(i)} > VaR_{r}(Y^{(i)})$.

Example 6.2. Daily returns of 4 major European indices

We analyze the daily returns of the following indices: British FTSE 100, French CAC 40, Deutsche Börse AG and Spain’s IBEX 35, from January 2, 1995 to April 27, 2012. The data set contains 4402 observations.

Once again, we concentrate on the respective losses $Y^{(i)} = -X^{(i)}, i = 1, \ldots, 4$. Using the same type of distress events as in the previous examples, we study the conditional loss tail of the return on British FTSE 100 and the corresponding CoVaR. The results are displayed in Figure 9 and Table 13. An obvious difference between the results obtained in this example and the previous example is that no significant movement in the conditional loss probabilities is noticeable after the first conditioning. This is due to the fact that there is strong tail dependence between the losses on French CAC 40, Deutsche Börse AG and Spain’s IBEX 35. This results in the tail measure that concentrated in a small part of the four-dimensional space, and in the conditional loss curves that barely move.
### Table 13. Estimated quantiles for losses of FTSE 100 unconditioned and conditioned on $Y^{(i)} > VaR_{r}(Y^{(i)})$

<table>
<thead>
<tr>
<th>Quantile</th>
<th>Unconditional</th>
<th>Cond. on $Y^{(2)}$</th>
<th>on $Y^{(2)}, Y^{(3)}$</th>
<th>on $Y^{(2)}, Y^{(3)}, Y^{(4)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>90%</td>
<td>.013</td>
<td>.043</td>
<td>.049</td>
<td>.050</td>
</tr>
<tr>
<td>95%</td>
<td>.019</td>
<td>.053</td>
<td>.059</td>
<td>.060</td>
</tr>
<tr>
<td>99%</td>
<td>.035</td>
<td>.078</td>
<td>.086</td>
<td>.087</td>
</tr>
</tbody>
</table>

**Figure 9.** European indices: Unconditional and conditional $P(Y^{(1)} > t | C_i)$ where $C_i$ is the event $Y^{(i)} > VaR_{r}(Y^{(i)})$ for $r = .95$ and $r = .99$

### Appendix A. Proof of Theorem 1

**Proof.** It is enough to show that for each $j = 1, \ldots, d$ and $t > 0$,

$$X^{(j)}_{[N_n t]} \frac{b^{(j)}(\frac{n}{N_n})}{b^{(j)}(\sum_{m=1}^{d} Y_{[N_n t]}^{(m)})} \to t^{-1/\alpha_j} \quad \text{in probability,}$$

because then the argument on pp. 311-312 in Resnick (2007) can be used to establish the claim of the theorem. We start with checking that for every $j, m = 1, \ldots, d$,

$$X^{(j)}_{[N_n^{(m)} t]} \frac{b^{(j)}(\frac{n}{N_n^{(m)}})}{b^{(j)}(\sum_{m=1}^{d} Y_{[N_n^{(m)} t]}^{(m)})} \to t^{-1/\alpha_j} \quad \text{in probability.}$$

Fix $j, m$ and let $0 < a < A < \infty$. We verify first that

$$X^{(j)}_{[\lambda_n^{(m)} u t]} \frac{a}{b^{(j)}(\lambda_n^{(m)} u)} \quad (a \leq u \leq A) \to (t^{-1/\alpha_j} e(u), a \leq u \leq A)$$

in probability in the Skorohod $J_{1}$ topology on $D[a, A]$. Here $e$ is a function on $[a, A]$ equal identically to 1.
Indeed, let $0 < \varepsilon < t^{-1/\alpha_j}$, and choose $K$ so large that
\[
\frac{u_i}{u_{i-1}} < (1 + \varepsilon t^{1/\alpha_j})^{\alpha_j}
\]
for each $i = 1, \ldots, k$, with $u_i = a + (A - a)i/K$, $i = 0, 1, \ldots, K$. Then
\[
P \left( \sup_{a \leq u \leq A} \left| \frac{X^{(j)}_{\lfloor n(\lambda_{n}^{(m)} u) \rfloor}}{b(j)(\frac{n}{\lambda_{n}^{(m)} u})} - t^{-1/\alpha_j} \right| > \varepsilon \right)
\]
\[
\leq P \left( X^{(j)}_{\lfloor \lambda_{n}^{(m)} u \rfloor} > (t^{-1/\alpha_j} + \varepsilon) b(j)(\frac{n}{\lambda_{n}^{(m)} u}) \right. 
\]
for some $a \leq u \leq A)
\]
\[
+ P \left( X^{(j)}_{\lfloor \lambda_{n}^{(m)} u \rfloor} < (t^{-1/\alpha_j} - \varepsilon) b(j)(\frac{n}{\lambda_{n}^{(m)} u}) \right. 
\]
for some $a \leq u \leq A)
\]
\[
:= a_n + b_n.
\]
By monotonicity,
\[
a_n \leq \sum_{i=1}^{K} P \left( X^{(j)}_{\lfloor \lambda_{n}^{(m)} u_{i-1} \rfloor} > (t^{-1/\alpha_j} + \varepsilon) b(j)(\frac{n}{\lambda_{n}^{(m)} u_{i-1}}) \right)
\]
and the $i$th term in the finite sum can be written as
\[
P \left( \frac{X^{(j)}_{\lfloor \lambda_{n}^{(m)} u_{i-1} \rfloor}}{b(j)(\frac{n}{\lambda_{n}^{(m)} u_{i-1}})} > (t^{-1/\alpha_j} + \varepsilon) \frac{b(j)(\frac{n}{\lambda_{n}^{(m)} u_{i-1}})}{b(j)(\frac{n}{\lambda_{n}^{(m)} u_{i-1}})} \right) \to 0
\]
as $n \to \infty$, because by (4.18) in Resnick (2007), the random variable in the left hand side of the inequality converges in probability to $t^{-1/\alpha_j}$, while the expression in the right hand side of the inequality converges, by the regular variation, to
\[
(t^{-1/\alpha_j} + \varepsilon) \left( \frac{u_{i-1}}{u_i} \right)^{1/\alpha_j} \to t^{-1/\alpha_j}
\]
by the choice of $K$. Hence $a_n \to 0$ and, similarly, $b_n \to 0$ as $n \to \infty$, which establishes (A.3).

Let $\varphi : (0, \infty) \to [a, A]$ be defined by
\[
\varphi(x) = \begin{cases} 
  x, & a \leq x \leq A \\
  a, & 0 < x < a \\
  A, & x > A.
\end{cases}
\]
Since $\varphi$ is a continuous function, and $N_{n}^{(m)}/\lambda_{n}^{(m)} \Rightarrow \tau_{\omega_{m}}$ by Theorem 1 in Nguyen and Samorodnitsky (2012) (recall that $\tau_{\omega_{m}}$ is the first time a standard Brownian motion hits $\pm \omega_{m}$), we conclude that $\varphi(N_{n}^{(m)}/\lambda_{n}^{(m)}) \Rightarrow
\( \varphi(\tau_{\omega_m}) \). Since the convergence in (A.3) is to a deterministic limit, it follows by Theorem 4.4 in Billingsley (1968) that

\[
\left[ \left( \frac{X^{(j)}[\lambda_n^{(m)}u]}{b^{(j)}(\frac{n}{\lambda_n^{(m)}}u)}, a \leq u \leq A \right), \varphi(\frac{N_n^{(m)}}{\lambda_n^{(m)}}) \right] \\
\Rightarrow \left[ (t^{-1/\alpha_j}e(u), a \leq u \leq A), \varphi(\tau_{\omega_m}) \right]
\]

weakly in \( D[a, A] \times [a, A] \). Since the map \((x, t) \to x(t)\) on \( D[a, A] \times [a, A] \) is continuous at every point of its domain with a continuous first coordinate, we conclude that

\[
\frac{X^{(j)}[\lambda_n^{(m)}\varphi(\frac{N_n^{(m)}}{\lambda_n^{(m)}})\varphi(\tau_{\omega_m})t]}{b^{(j)}(\frac{n}{\lambda_n^{(m)}}\varphi(\frac{N_n^{(m)}}{\lambda_n^{(m)}}))} \to t^{-1/\alpha_j}
\]

in probability. Since

\[
P\left( \frac{X^{(j)}[\lambda_n^{(m)}]}{b^{(j)}(\frac{n}{\lambda_n^{(m)}})} \neq \frac{X^{(j)}[\lambda_n^{(m)}\varphi(\frac{N_n^{(m)}}{\lambda_n^{(m)}})\varphi(\tau_{\omega_m})t]}{b^{(j)}(\frac{n}{\lambda_n^{(m)}}\varphi(\frac{N_n^{(m)}}{\lambda_n^{(m)}}))} \right) \leq P\left( \frac{N_n^{(m)}}{\lambda_n^{(m)}} \notin [a, A] \right)
\]

\[
\Rightarrow P(\tau_{\omega_m} \notin [a, A])
\]

which can be made arbitrarily small by taking \( a \) small and \( A \) large, (A.2) follows.

The connection, through subsequences, between convergence in probability and a.s. convergence shows that (A.2) implies (A.1), so that the proof of the theorem is complete. \( \square \)

**Acknowledgments**

We are grateful to Zari Rachev for the suggestion of using CoVaR to describe the tails of high dimensional financial data and for providing us with the data on the European indices. We are also indebted to two anonymous referees whose unusually careful and detailed reading of the manuscript greatly helped to improve the paper.


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