

ON THE FOUNDATIONS OF MULTIVARIATE HEAVY TAIL ANALYSIS

SIDNEY RESNICK

ABSTRACT. Univariate heavy tailed analysis rests on the analytic notion of regularly varying functions. For multivariate heavy tailed analysis, reliance on functions is awkward because multivariate distribution functions are not natural objects for many purposes and are difficult to manipulate. An approach based on vague convergence of measures makes the differences between univariate and multivariate analysis evaporate. We survey the foundations of the subject and discuss statistical attempts to assess dependence.

1. INTRODUCTION

One of the many admirable aspects of Chris Heyde's research career is the effortless way he would follow research problems back and forth across the boundary between probability and statistics. In fact, it would seem that Professor Heyde's healthy outlook did not even recognize a boundary between the two subjects. To me, and no doubt to many like minded colleagues who found a comfortable home in the applied probability community, Chris Heyde's example was important to our developing tastes and scientific outlooks. Hopefully, this contribution in Chris' honor, will exhibit some of these characteristics of containing both interesting probability and useful statistical techniques.

Heavy tailed analysis has been stimulated over the last 10 year or so due to an increasing number of real world examples that seem to exhibit heavy tailed characteristics. Finance, economics, data networks, insurance, risk analysis are all subjects which have become heavy consumers of heavy tailed methods. Useful surveys by Mikosch [33] and Resnick [37] appear in a forthcoming SemStat volume or on the authors web sites.

The networking story began around 1993 with the publication of what is now known as the Bellcore study ([24, 29, 49]). Traditional queueing models had thrived on assumptions of exponentially bounded tails, Poisson inputs and assumed independence. Collected network data, studied at what was then Bellcore (now Telcordia), exhibited properties which were inconsistent with traditional queueing models. These anomalies were also found in world wide web downloads in the Boston University study ([15, 7, 8, 9, 12, 10, 11]). The unusual properties found in the data traces included:

- self-similarity (ss) and long-range dependence (LRD) of various transmission rates:
 - packet counts per unit time,
 - www bits/time.
- heavy tails of quantities such as
 - file sizes,

Sidney Resnick's research was partially supported by NSF grant DMS-97-04982 at Cornell University.
©2001 by Sidney Resnick.

- transmission rates, transmission durations,
- CPU job completion times,
- call lengths

The Bellcore study resulted in a paradigm shift and after some resistance to the presence of long range dependence, there was widespread acceptance of the statement that *packet counts per unit time exhibit self similarity and long range dependence*. Research goals then shifted from detection of the phenomena to greater understanding of the causes. Today there is fairly wide spread acceptance that a high level explanation is that heavy tailed file sizes causes long range dependence in network traffic [48, 49, 37].

In finance, ARCH and GARCH models were developed by Engle, Bollerslev and others to explain volatility or a constantly changing variance [26, 6, 5]. These models have Pareto univariate tails ([28, 47, 27]) and even exhibit multivariate regular variation [16, 45, 46]. Insurance, especially *re-insurance* is also a natural setting for heavy tailed analysis.

Our focus here is on multivariate problems and is primarily didactic. We show that with the right formulation, multivariate heavy tailed analysis can be framed so that it is indistinguishable from the univariate case. We highlight some inference problems special to the multivariate case and discuss some current developments.

Heavy tailed analysis is dependent for its analytical underpinnings on the theory of *regularly varying functions*. Such functions are, roughly speaking, functions which behave asymptotically like power functions. We start by considering real functions of a real variable. Consideration of multivariate cases and probability concepts suggests recasting definitions in terms of vague convergence of measures but we will consider this reformulation a bit later. More formally, a measurable function $U : \mathbb{R}_+ \mapsto \mathbb{R}_+$ is regularly varying at ∞ with index $\rho \in \mathbb{R}$ (written $U \in RV_\rho$) if for $x > 0$

$$(1.1) \quad \lim_{t \rightarrow \infty} \frac{U(tx)}{U(t)} = x^\rho.$$

We call ρ the *exponent of variation*. If $\rho = 0$ we say U is *slowly varying*. Slowly varying functions are generically denoted by $L(x)$. If $U \in RV_\rho$, then we write $U(x) = x^\rho L(x)$.

The canonical ρ -varying function is x^ρ . The functions $\log(1+x)$, $\log \log(e+x)$ are slowly varying, as is $\exp\{(\log x)^\alpha\}$, $0 < \alpha < 1$. In probability applications we are concerned with distributions whose tails are regularly varying. Examples are the standard Pareto, the extreme value Fréchet distribution, the stable distribution, and most familiarly, the Cauchy distribution tail with density $f(x) = (\pi(1+x^2))^{-1}$.

Distributions with regularly varying tails are called *heavy tailed distributions*. In probability theory, the property of regularly varying tails serves as exact necessary and sufficient condition for a variety of limit theorems and in statistics provides a semi-parametric family of distributions that can be reasonably fit to data.

2. MULTIVARIATE REGULAR VARIATION FOR FUNCTIONS

The obvious way to generalize (1.1) is to replace the scalar vector x by the vector \mathbf{x} . Before doing this, we review notational conventions which aid a dimensionless treatment of the theory.

2.1. **Notation review.** Vectors are denoted by bold letters, capitals for random vectors and lower case for non-random vectors. For example:

$$\mathbf{x} = (x^{(1)}, \dots, x^{(d)}) \in \mathbb{R}^d.$$

Operations between vectors *should (almost) always be interpreted componentwise* so that for two vectors \mathbf{x} and \mathbf{z}

$$\begin{aligned} \mathbf{x} < \mathbf{z} &\text{ means } x^{(i)} < z^{(i)}, \quad i = 1, \dots, d, \\ \mathbf{x} \leq \mathbf{z} &\text{ means } x^{(i)} \leq z^{(i)}, \quad i = 1, \dots, d, \\ \mathbf{x} = \mathbf{z} &\text{ means } x^{(i)} = z^{(i)}, \quad i = 1, \dots, d, \\ \mathbf{z}\mathbf{x} &= (z^{(1)}x^{(1)}, \dots, z^{(d)}x^{(d)}), \\ \mathbf{x} \vee \mathbf{z} &= (x^{(1)} \vee z^{(1)}, \dots, x^{(d)} \vee z^{(d)}), \\ \frac{\mathbf{x}}{\mathbf{z}} &= \left(\frac{x^{(1)}}{z^{(1)}}, \dots, \frac{x^{(d)}}{z^{(d)}} \right). \end{aligned}$$

Also $\mathbf{0} = (0, \dots, 0)$, $\mathbf{1} = (1, \dots, 1)$, $\mathbf{e}_i = (0, \dots, 1, \dots, 0)$ for $i = 1, \dots, d$. For a real number c , denote as usual

$$c\mathbf{x} = (cx^{(1)}, \dots, cx^{(d)}).$$

We denote rectangles by

$$[\mathbf{a}, \mathbf{b}] = \{\mathbf{x} \in \mathbb{R} : \mathbf{a} \leq \mathbf{x} \leq \mathbf{b}\}$$

so that for $\mathbf{x} > \mathbf{0}$ and $\mathbb{E} = [\mathbf{0}, \infty) \setminus \{\mathbf{0}\}$,

$$[\mathbf{0}, \mathbf{x}]^c = \mathbb{E} \setminus [\mathbf{0}, \mathbf{x}] = \{\mathbf{y} \in \mathbb{E} : \bigvee_{i=1}^d \frac{y^{(i)}}{x^{(i)}} > 1\}.$$

2.2. **Multivariate regularly varying functions.** A subset $C \subset \mathbb{R}^d$ is a *cone* if whenever $\mathbf{x} \in C$ also $t\mathbf{x} \in C$ for any $t > 0$. A function $h : C \mapsto (0, \infty)$ is monotone if it is either non-decreasing in each component or non-increasing in each component. For h non-decreasing, this is equivalent to saying that whenever $\mathbf{x}, \mathbf{y} \in C$ and $\mathbf{x} \leq \mathbf{y}$ we have $h(\mathbf{x}) \leq h(\mathbf{y})$. The natural domain for a multivariate regularly varying function is a cone.

Suppose $h(\cdot) \geq 0$ is measurable and defined on C . Suppose $\mathbf{1} = (1, \dots, 1) \in C$. Call h *multivariate regularly varying on C with limit function $\lambda(\cdot)$* if $\lambda(\mathbf{x}) > 0$ for $\mathbf{x} \in C$ and for all $\mathbf{x} \in C$ we have

$$(2.1) \quad \lim_{t \rightarrow \infty} \frac{h(t\mathbf{x})}{h(t\mathbf{1})} = \lambda(\mathbf{x}).$$

Note $\lambda(\mathbf{1}) = 1$. A simple scaling argument shows that $\lambda(\cdot)$ is homogeneous:

$$(2.2) \quad \lambda(s\mathbf{x}) = s^\rho \lambda(\mathbf{x}), \quad s > 0, \mathbf{x} \in C, \rho \in \mathbb{R}.$$

For multivariate distributions F concentrating on $[0, \infty)^d =: [\mathbf{0}, \infty)$, it is ambiguous what we mean by *distribution tail*. The usual interpretation has been to consider $1 - F(\mathbf{x})$ for

$\mathbf{x} \geq \mathbf{0}$ but $\mathbf{x} \neq \mathbf{0}$ and so it is required that

$$(2.3) \quad \lim_{t \rightarrow \infty} \frac{1 - F(t\mathbf{x})}{1 - F(t\mathbf{1})} = \lambda(\mathbf{x}).$$

It is awkward to deal with distribution functions and more natural to deal with measures.

3. REGULAR VARIATION FORMULATED IN TERMS OF MEASURES.

It is possible to reformulate the regular variation of distribution tails in terms of vague convergence of families of measures which give a dimensionless and elegant treatment. This is facilitated by considering random elements in various complete separable metric spaces as discussed in Billingsley [3, 4]. The spaces we need are summarized in the following Table 1.

Metric space \mathbb{S}
\mathbb{R}, \mathbb{R}^d , Euclidean spaces
$C[0, \infty)$, the space of real valued, continuous functions on $[0, \infty)$
$D[0, \infty)$, the space of real valued, right continuous functions on $[0, \infty)$ with finite left limits existing on $(0, \infty)$
$M_p(\mathbb{E})$, the space of point measures on a nice state space \mathbb{E}
$M_+(\mathbb{E})$, the space of Radon measures on a nice state space \mathbb{E}
$\mathcal{K}(\mathbb{S})$, the compact subsets of \mathbb{S} .

TABLE 1. Various metric spaces.

Examples of state spaces \mathbb{E} that are useful for us are $(0, \infty]$, $[0, \infty] \setminus \{0\}$, $[-\infty, \infty] \setminus \{0\}$. All these examples are compact sets punctured by removal of one point which is an example of the *one-point un-compactification*. This unconventional topology is necessitated by the fact that we consider regular variation at ∞ . For instance in the case $d = 1$ we need a way for sets of the form (x, ∞) to be considered relatively compact. More later in Subsection 3.3.

3.1. Basic convergences. We now give two criteria for convergence. One gives necessary and sufficient conditions for empirical measures of scaled observations to converge to a Poisson random measure limit and the other discusses convergence to a constant limit measure. The first is the basis for manipulating iid random variables with regularly varying tails by means of the Poisson transform and the second is the basis for consistency of estimators of heavy tailed parameters. Here, and in what follows, $PRM(\mu)$ means *Poisson random measure with mean measure μ* and \mathbb{E} is a nice metric space which for us means a finite dimensional Euclidean space.

Theorem 1 (Basic convergence). *Suppose that for each $n \geq 1$ we have $\{X_{n,j}, j \geq 1\}$ is a sequence of iid random elements of $(\mathbb{E}, \mathcal{E})$. Let ξ be $PRM(\mu)$ on $M_p(\mathbb{E})$.*

(i) *We have*

$$(3.1) \quad \sum_{j=1}^n \epsilon_{X_{n,j}} \Rightarrow \xi = PRM(\mu)$$

on $M_p(\mathbb{E})$, iff in $M_+(\mathbb{E})$

$$(3.2) \quad n\mathbb{P}[X_{n,1} \in \cdot] = \mathbf{E} \left(\sum_{j=1}^n \epsilon_{X_{n,j}}(\cdot) \right) \xrightarrow{v} \mu.$$

(ii) Suppose additionally that $0 < a_n \uparrow \infty$. Then for a measure $\mu \in M_+(\mathbb{E})$ we have

$$(3.3) \quad \frac{1}{a_n} \sum_{j=1}^n \epsilon_{X_{n,j}} \Rightarrow \mu$$

on $M_+(\mathbb{E})$ iff in $M_+(\mathbb{E})$

$$(3.4) \quad \frac{n}{a_n} \mathbb{P}[X_{n,1} \in \cdot] = \mathbf{E} \left(\frac{1}{a_n} \sum_{j=1}^n \epsilon_{X_{n,j}}(\cdot) \right) \xrightarrow{v} \mu.$$

The proofs are readily handled using Laplace functionals. See [40, 41].

3.2. The polar coordinate transformation. To state the equivalent forms of multivariate regular variation for the distribution of a random vector, we need the polar coordinate transformation.

Suppose $\|\cdot\| : \mathbb{R}^d \mapsto [0, \infty)$ is a *norm* on \mathbb{R}^d . The most useful norms for us are the usual Euclidean L_2 norm, the L_p norm for $p \geq 0$ and the L_∞ norm: $\|\mathbf{x}\| = \bigvee_{i=1}^n |x^{(i)}|$. Given a chosen norm $\|\cdot\|$, the unit sphere is

$$\mathfrak{N} := \{\mathbf{x} : \|\mathbf{x}\| = 1\}.$$

For the Euclidean norm, the unit sphere is really a sphere. If $d = 2$, the L_1 norm is a diamond, and the L_∞ norm is a square. Norms on \mathbb{R}^d are all topologically equivalent in that convergence in one norm implies convergence in another.

Now fix a norm. Define the polar coordinate transformation $T : \mathbb{R}^d \setminus \{\mathbf{0}\} \mapsto (0, \infty) \times \mathfrak{N}$ by

$$T(\mathbf{x}) = \left(\|\mathbf{x}\|, \frac{\mathbf{x}}{\|\mathbf{x}\|} \right) =: (r, \mathbf{a}),$$

and the inverse transformation $T^{\leftarrow} : (0, \infty) \times \mathfrak{N} \mapsto \mathbb{R}^d \setminus \{\mathbf{0}\}$ by

$$T^{\leftarrow}(r, \mathbf{a}) = r\mathbf{a}.$$

Think of $\mathbf{a} \in \mathfrak{N}$ as defining a direction and r telling how far in direction \mathbf{a} to proceed. Since we excluded $\mathbf{0}$ from the domain of T , both T and T^{\leftarrow} are continuous bijections.

When $d = 2$, it is customary, but not obligatory, to write $T(\mathbf{x}) = (r \cos \theta, r \sin \theta)$, where $0 \leq \theta \leq 2\pi$, rather than the more consistent notation $T(\mathbf{X}) = (r, (\cos \theta, r \sin \theta))$. For a random vector \mathbf{X} in \mathbb{R}^d we sometimes write $T(\mathbf{X}) = (R, \Theta)$.

3.3. Topology: The one point uncompactification. In reformulating the function theory concept of regularly varying functions into a measure theory concept, there is continual need to deal with infinite sets which are neighborhoods of ∞ . Such sets need to be regarded as “bounded” in an appropriate topology so sequences of measures of such sets can converge non-trivially. For example, when $d = 1$ tail probabilities of sets of the form (x, ∞) are considered and because vaguely converging measures only guarantee convergence of relatively compact sets, we need a topology which renders (x, ∞) as relatively compact. A convenient way to think about such issues is by means of the one point un-compactification.

Let $(\mathbb{X}, \mathcal{T})$ be a nice topological space; \mathbb{X} is the set and \mathcal{T} is the topology. For our purposes, \mathbb{X} would be a subset of Euclidean space. Consider a subset $\mathbb{D} \subset \mathbb{X}$ and define $\mathbb{X}^\# = \mathbb{X} \setminus \mathbb{D}$ and give $\mathbb{X}^\#$ the relative topology

$$\mathcal{T}^\# = \mathcal{T} \cap \mathbb{D}^c = \mathcal{T} \cap \mathbb{X}^\#.$$

So a set is open in $\mathbb{X}^\#$ if it is an open subset of \mathbb{X} intersected with $\mathbb{X}^\#$.

We need to identify the compact subsets of $\mathbb{X}^\#$.

Proposition 1. *The compact subsets of $\mathbb{X}^\#$ are*

$$\mathcal{K}(\mathbb{X}^\#) = \{K \in \mathcal{K}(\mathbb{X}) : K \cap \mathbb{D} = \emptyset\}.$$

The compact sets of $\mathbb{X}^\#$ are the original compact sets of \mathbb{X} , provided they do not intersect the piece D chopped away from \mathbb{X} to form $\mathbb{X}^\#$. Specialize this to the *one point un-compactification*, which describes what are the compact sets when a compact space is punctured by the removal of a point: Suppose \mathbb{E} is a compact set and $e \in \mathbb{E}$. Give $\mathbb{E} \setminus \{e\}$ the relative topology consisting of sets in $\mathbb{E} \setminus \{e\}$ of the form $G \setminus \{e\}$, where G is open in \mathbb{E} . The compact sets of $\mathbb{E} \setminus \{e\}$ are those compact subsets $K \subset \mathbb{E}$ such that $e \notin K$.

Consider the following special cases:

- (1) The compact sets of $[0, \infty]$ consist of any closed set. The compact subsets of $[0, \infty] \setminus \{0\}$ are closed subsets of $[0, \infty]$ bounded away from 0.
- (2) The compact sets of $[-\infty, \infty]$ consist of any closed set. The compact subsets of $[-\infty, \infty] \setminus \{0\}$ are closed subsets of $[-\infty, \infty]$ bounded away from 0. This punctured space \mathbb{E} , and the associated space of Radon measures $M_+(\mathbb{E})$ is needed for considering weak convergence of partial sums to Lévy processes.
- (3) The compact sets of $[0, \infty]$ consist of any closed set. The compact subsets of $[0, \infty] \setminus \{0\}$ are closed subsets of $[0, \infty]$ bounded away from 0. Similarly, the compacta of $[-\infty, \infty] \setminus \{0\}$ are those closed subsets of $[-\infty, \infty]$ bounded away from 0.

3.4. Multivariate regularly varying tail probabilities. We now state various equivalences which can define multivariate regularly varying tail probabilities. To deal with multivariate regular variation of tail probabilities, we have to consider a punctured space with a one-point un-compactification such as $[0, \infty] \setminus \{0\}$. Equivalences in terms of polar coordinates are then problematic since the polar coordinate transformation is not defined on the lines through ∞ , so some sort of restriction argument is necessary. A different treatment of the polar coordinate transformation is given in [1, 2, 33].

For what follows we restrict attention to the case of random vectors with non-negative components. Suppose $\{\mathbf{Z}_n, n \geq 1\}$ are iid random elements of \mathbb{R}_+^d with common distribution $F(\cdot)$. We set $\mathbb{E} = [\mathbf{0}, \infty] \setminus \{\mathbf{0}\}$. Recall $[\mathbf{0}, \mathbf{x}]^c = \mathbb{E} \setminus [\mathbf{0}, \mathbf{x}]$. We also define the measure on $(0, \infty]$

$$\nu_\alpha(x, \infty] = x^{-\alpha}, \quad x > 0, \alpha > 0.$$

Set $\mathfrak{N}_+ = \mathfrak{N} \cap \mathbb{E}$. Vague convergence of measures is denoted by \xrightarrow{v} .

Theorem 2 (Multivariate regularly varying tail probabilities). *The following statements are equivalent. (In each, we understand the phrase Radon measure to mean a not identically zero Radon measure.)*

- (1) *There exists a Radon measure ν on \mathbb{E} such that*

$$\lim_{t \rightarrow \infty} \frac{1 - F(t\mathbf{x})}{1 - F(t\mathbf{1})} = \lim_{t \rightarrow \infty} \frac{\mathbb{P}[\frac{\mathbf{Z}_1}{t} \in [\mathbf{0}, \mathbf{x}]^c]}{\mathbb{P}[\frac{\mathbf{Z}_1}{t} \in [\mathbf{0}, \mathbf{1}]^c]} = c\nu([\mathbf{0}, \mathbf{x}]^c),$$

for some $c > 0$ and for all points $\mathbf{x} \in [\mathbf{0}, \infty] \setminus \{\mathbf{0}\}$ which are continuity points of $\nu([\mathbf{0}, \cdot]^c)$.

- (2) *There exists a function $b(t) \rightarrow \infty$ and a Radon measure ν on \mathbb{E} such that in $M_+(\mathbb{E})$*

$$t\mathbb{P}\left[\frac{\mathbf{Z}_1}{b(t)} \in \cdot\right] \xrightarrow{v} \nu, \quad t \rightarrow \infty.$$

- (3) *There exists a sequence $b_n \rightarrow \infty$ and a Radon measure ν on \mathbb{E} such that in $M_+(\mathbb{E})$*

$$n\mathbb{P}\left[\frac{\mathbf{Z}_1}{b_n} \in \cdot\right] \xrightarrow{v} \nu, \quad t \rightarrow \infty.$$

- (4) *There exists a probability measure $S(\cdot)$ on \mathfrak{N}_+ and a function $b(t) \rightarrow \infty$ such that for $(R_1, \Theta_1) = \left(\|\mathbf{Z}_1\|, \frac{\mathbf{Z}_1}{\|\mathbf{Z}_1\|}\right)$ we have*

$$t\mathbb{P}\left[\left(\frac{R_1}{b(t)}, \Theta_1\right) \in \cdot\right] \xrightarrow{v} c\nu_\alpha \times S$$

in $M_+((0, \infty] \times \mathfrak{N}_+)$, where $c > 0$.

- (5) *There exists a probability measure $S(\cdot)$ on \mathfrak{N}_+ and a sequence $b_n \rightarrow \infty$ such that for $(R_1, \Theta_1) = \left(\|\mathbf{Z}_1\|, \frac{\mathbf{Z}_1}{\|\mathbf{Z}_1\|}\right)$ we have*

$$n\mathbb{P}\left[\left(\frac{R_1}{b_n}, \Theta_1\right) \in \cdot\right] \xrightarrow{v} c\nu_\alpha \times S$$

in $M_+((0, \infty] \times \mathfrak{N}_+)$, where $c > 0$.

- (6) *There exists $b_n \rightarrow \infty$ such that in $M_p(\mathbb{E})$*

$$\sum_{i=1}^n \epsilon_{\mathbf{Z}_i/b_n} \Rightarrow PRM(\nu).$$

(7) *There exists a sequence $b_n \rightarrow \infty$ such that in $M_p((0, \infty] \times \mathfrak{N}_+)$*

$$\sum_{i=1}^n \epsilon_{(R_i/b_n, \Theta_i)} \Rightarrow PRM(c\nu_\alpha \times S).$$

These conditions imply that for any sequence $k = k(n) \rightarrow \infty$ such that $n/k \rightarrow \infty$ we have

(8) *In $M_+(\mathbb{E})$*

$$(3.5) \quad \frac{1}{k} \sum_{i=1}^n \epsilon_{\mathbf{Z}_i/b(\frac{n}{k})} \Rightarrow \nu$$

and 8 is equivalent to any of 1–7, provided $k(\cdot)$ satisfies $k(n) \sim k(n+1)$. Similar statements to (3.5) can be made in terms of polar coordinates.

Before discussing elements of the proof of Theorem 2, we need to understand how vague convergence of a sequence of measures is affected by a continuous transformation of the state space.

3.4.1. Preservation of weak convergence under mappings: Consider two state spaces \mathbb{E}_1 and \mathbb{E}_2 with a mapping $T : \mathbb{E}_1 \mapsto \mathbb{E}_2$ taking one into the other. A measure $\mu \in M_+(\mathbb{E}_1)$ has an image $\hat{T}(\mu) \in M_+(\mathbb{E}_2)$ given by the map

$$\hat{T}(\mu) = \mu \circ T^{-1}.$$

If T is a continuous point transformation, is $\hat{T} : M_+(\mathbb{E}_1) \mapsto M_+(\mathbb{E}_2)$ continuous? Note, if $m \in M_p(\mathbb{E})$ is a point measure of the form $\sum_i \epsilon_{x_i}$, then

$$\hat{T}(m) = m \circ T^{-1} = \sum_i \epsilon_{T(x_i)}.$$

Continuity of T does not guarantee continuity of \hat{T} without a condition. See [41] for the simple proof of the following.

Proposition 2. *Suppose $T : \mathbb{E}_1 \mapsto \mathbb{E}_2$ is a continuous function such that*

$$(3.6) \quad T^{-1}(K_2) \in \mathcal{K}(\mathbb{E}_1), \quad \forall K_2 \in \mathcal{K}(\mathbb{E}_2).$$

Then if $\mu_n \xrightarrow{v} \mu_0$ in $M_+(\mathbb{E}_1)$, we have that

$$\hat{T}\mu_n = \mu_n \circ T^{-1} \xrightarrow{v} \mu_0 \circ T^{-1} = \hat{T}\mu_0.$$

3.4.2. Elements of the proof of Theorem 2. The equivalence between mean measure convergence in $\{1, 2, 3\}$ and convergence to a Poisson random measure in 1 is given by the basic convergences Theorem 1 (i). Similarly, Theorem 1 (ii) shows mean measure convergence is equivalent to (3.5) under the stated conditions.

To understand how to go from cartesian to polar coordinates and back we outline the proof of 5→3. We break this into steps.

STEP 1. RESTRICT THE SPACE TO THE NATURAL DOMAIN OF THE POLAR COORDINATE TRANSFORMATION. By applying a restriction functional, the vague convergence statement in item 5 implies

$$(3.7) \quad n\mathbb{P}\left[\left(\frac{R_1}{b_n}, \Theta\right) \in \cdot\right] \xrightarrow{v} c\nu_\alpha \times S, \quad \text{in } M_+((0, \infty) \times \mathfrak{N}_+).$$

STEP 2. Recall the polar coordinate transform was denoted by T . Apply T^\leftarrow to (3.7). Recall

$$T^\leftarrow : (0, \infty) \times \mathfrak{N}_+ \mapsto [\mathbf{0}, \infty) \setminus \{\mathbf{0}\}, \quad T^\leftarrow(r, \boldsymbol{\theta}) = r\boldsymbol{\theta}.$$

Let $K \in \mathcal{K}([\mathbf{0}, \infty) \setminus \{\mathbf{0}\})$. Then for some small $\delta > 0$ and large $M > 0$ we have

$$K \subset \{\mathbf{x} : \delta \leq \|\mathbf{x}\| \leq M\}.$$

Then

$$\begin{aligned} (T^\leftarrow)^{-1}(K) &= \{(r, \boldsymbol{\theta}) : r\boldsymbol{\theta} \in K\} \\ &\subset \{(r, \boldsymbol{\theta}) : \delta \leq r \leq M\} \in \mathcal{K}((0, \infty) \times \mathfrak{N}_+). \end{aligned}$$

Now $(T^\leftarrow)^{-1}(K)$ is closed and contained in a compact set so is compact.

Using Proposition 2, we apply T^\leftarrow to (3.7) and obtain

$$(3.8) \quad n\mathbb{P}\left[\frac{R_1 \Theta_1}{b_n} \in \cdot\right] = n\mathbb{P}\left[\frac{\mathbf{Z}_1}{b_n} \in \cdot\right] \xrightarrow{v} c\nu_\alpha \times S \circ (T^\leftarrow)^{-1} =: \nu(\cdot) \quad \text{in } M_+([\mathbf{0}, \infty) \setminus \{\mathbf{0}\})$$

where we called the limit ν .

STEP 3. EXTEND TO THE BOUNDARIES: Let $f \in C_K^+(\mathbb{E})$. It suffices to show

$$(3.9) \quad nEf\left(\frac{\mathbf{Z}_1}{b_n}\right) \rightarrow \nu(f).$$

This is done using a smooth truncation, appealing to Step 2, and then letting the truncation level float to infinity. \square

4. AMPLIFICATION.

Here are some additional remarks to flesh out the theory.

4.1. **What if the natural state space is not the positive orthant?** What if \mathbb{E} is the closed cone $[-\infty, \infty) \setminus \{\mathbf{0}\}$? Item 1 of Theorem 2 in terms of multivariate regular variation of the distribution function tail no longer has an easy analogue, except in $d = 1$, but there is no trouble extending items 2-7 of Theorem 2 to the more general cone, provided \mathfrak{N}_+ is replaced by $\mathfrak{N} \cap \mathbb{E}$.

If $d = 1$, then

$$(4.1) \quad n\mathbb{P}\left[\frac{Z_1}{b_n} \in \cdot\right] \xrightarrow{v} \nu, \quad \text{on } M_+([-\infty, \infty) \setminus \{0\})$$

is the basic condition. This means for $x > 0$,

$$(4.2) \quad n\mathbb{P}\left[\frac{Z_1}{b_n} > x\right] \rightarrow \nu(x, \infty], \quad (n \rightarrow \infty)$$

so

$$\nu(x, \infty] = c_+ x^{-\alpha}, \quad c_+ \geq 0.$$

Also, for $x > 0$

$$(4.3) \quad n\mathbb{P}\left[\frac{Z_1}{b_n} < -x\right] = n\mathbb{P}\left[\frac{-Z_1}{b_n} > x\right] \rightarrow \nu[-\infty, -x), \quad (n \rightarrow \infty)$$

and we get

$$\nu[-\infty, -x) = c_- x^{-\alpha}, \quad c_- \geq 0.$$

The α for the right tail must be the same as for the left tail since the same b_n works for both tails and b_n relates to α through the fact that $b_n = b(n)$ where the function $b(\cdot) \in RV_{1/\alpha}$. So

$$\nu(dx) = c_+ \alpha x^{-\alpha-1} dx 1_{(0, \infty)}(x) + c_- \alpha |x|^{-\alpha-1} dx 1_{(-\infty, 0)}(x).$$

For this $d = 1$ case, $\aleph = \{-1, 1\}$ and

$$S(\{1\}) = \frac{c_+}{c_+ + c_-}, \quad S(\{-1\}) = \frac{c_-}{c_+ + c_-}.$$

4.2. I hate multivariate distribution functions and love densities. What's in this for me? Glad you asked. Most multivariate distributions are specified by densities, not distribution functions so it would be nice to have workable criteria for when regular variation of the density implies the equivalent conditions of Theorem 2. When $d = 1$, regular variation of the density implies the distribution tail is a regularly varying function because of Karamata's theorem. In higher dimensions, some regularity is needed.

Roughly speaking, multivariate regular variation knits together one-dimensional regular variation along rays but does not control what happens as we hop from ray to ray. Imposing a uniformity condition as we move across rays overcomes this difficulty. The following is given in [21] and discussed in [41]. Related work is in [20, 17, 18]. We return to the assumption that $\mathbb{E} = [\mathbf{0}, \infty] \setminus \{\mathbf{0}\}$.

Theorem 3. *Suppose F is a probability distribution on \mathbb{E} with density F' which is regularly varying with the limit function $\lambda(\cdot)$ on $[\mathbf{0}, \infty) \setminus \{\mathbf{0}\}$. That is, we suppose for some regularly varying function $V(t) \in RV_\rho$, $\rho < 0$, we have for $\mathbf{x} \in [\mathbf{0}, \infty) \setminus \{\mathbf{0}\}$*

$$(4.4) \quad \lim_{t \rightarrow \infty} \frac{F'(t\mathbf{x})}{t^{-d}V(t)} = \lambda(\mathbf{x}) > 0.$$

Necessarily λ satisfies $\lambda(t\mathbf{x}) = t^\rho \lambda(\mathbf{x})$, for $\mathbf{x} \in [\mathbf{0}, \infty) \setminus \{\mathbf{0}\}$. Further suppose that λ is bounded on \aleph_+ and that the following uniformity condition holds

$$(4.5) \quad \lim_{t \rightarrow \infty} \sup_{\mathbf{x} \in \aleph_+} \left| \frac{F'(t\mathbf{x})}{t^{-d}V(t)} - \lambda(\mathbf{x}) \right| = 0.$$

It then follows that for any $\delta > 0$

$$(4.6) \quad \lim_{t \rightarrow \infty} \sup_{\|\mathbf{x}\| > \delta} \left| \frac{F'(t\mathbf{x})}{t^{-d}V(t)} - \lambda(\mathbf{x}) \right| = 0.$$

Furthermore, $\lambda(\cdot)$ is integrable on $[\mathbf{0}, \mathbf{x}]^c$ for $\mathbf{x} > \mathbf{0}$ and $1 - F$ is a regularly varying function on $(\mathbf{0}, \infty)$ which takes the form

$$(4.7) \quad \lim_{t \rightarrow \infty} \frac{1 - F(t\mathbf{x})}{V(t)} = \int_{[\mathbf{0}, \mathbf{x}]^c} \lambda(\mathbf{y}) d\mathbf{y}.$$

4.3. My bivariate data looks heavy tailed but the α 's do not look the same for each component. What do I do? Glad you asked. The phrasing of the regular variation condition so far in Theorem 2 assumes tail equivalence among the components in the following sense. The requirement that

$$\mathbb{P}\left[\frac{\mathbf{Z}_1}{b_n} \in \cdot\right] \xrightarrow{v} \nu, \quad \text{in } M_+(\mathbb{E}),$$

implies that

$$n\mathbb{P}\left[\frac{Z_1^{(i)}}{b_n} \in \cdot\right] \xrightarrow{v} c_i \nu_\alpha, \quad \text{in } M_+((0, \infty]),$$

for $c_i \geq 0$ and $i = 1, \dots, d$. Note we have not ruled out the possibility that for some (but not all) i , c_i could be zero. For those components with $c_i > 0$, the α 's are the same because $b_n = b(n)$ and $b(\cdot) \in RV_{1/\alpha}$. The marginal convergences with the same scaling function b_n , in turn imply that for $1 \leq i < j \leq d$

$$\lim_{x \rightarrow \infty} \frac{\mathbb{P}[Z_1^{(i)} > x]}{\mathbb{P}[Z_1^{(j)} > x]} = \frac{c_i}{c_j}.$$

While the theory is most elegantly developed using the single normalization of Theorem 2, in practice this is not adequate and sensitivity to tails with different weights is needed.

Theorem 4. Suppose $\{\mathbf{Z}_n, n \geq 1\}$ are iid non-negative random vectors and $\mathbb{E} = [\mathbf{0}, \infty] \setminus \{\mathbf{0}\}$. Suppose there exist sequences $\{b_{(i)}(n), n \geq 1, 1 \leq i \leq d\}$ satisfying

$$\lim_{n \rightarrow \infty} b_{(i)}(n) = \infty, \quad i = 1, \dots, d$$

and such that the following are true:

(i) MARGINAL CONDITION: For each $i = 1, \dots, d$

$$(4.8) \quad n\mathbb{P}\left[\frac{Z_1^{(i)}}{b_{(i)}(n)} \in \cdot\right] \xrightarrow{v} \nu_{\alpha_i}, \quad \alpha_i > 0,$$

in $M_+((0, \infty])$, and

(ii) GLOBAL CONDITION: There exists a measure ν on Borel subsets of \mathbb{E} such that

$$(4.9) \quad n\mathbb{P}\left[\frac{\mathbf{Z}_1}{(b_{(1)}(n), \dots, b_{(d)}(n))} \in \cdot\right] \xrightarrow{v} \nu,$$

in $M_+(\mathbb{E})$.

Let $\bar{F}_{(i)}(x) = \mathbb{P}[Z_1^{(i)} > x]$ be the i th marginal distribution tail and define

$$U_{(i)}(x) = \left(\frac{1}{\bar{F}_{(i)}(\cdot)} \right)^{\leftarrow}(x), \quad x > 1.$$

Then we have

(i) STANDARD GLOBAL CONVERGENCE:

$$(4.10) \quad nF_*(n\cdot) := n\mathbb{P}\left[\left(\frac{U_{(i)}^{\leftarrow}(Z_1^{(i)})}{n}, i = 1, \dots, n\right) \in \cdot\right] \xrightarrow{v} \nu_*, \quad \text{in } M_+(\mathbb{E}),$$

where

$$(4.11) \quad \nu_*(t\cdot) = t^{-1}\nu_*(\cdot),$$

on Borel subsets of \mathbb{E} , and

(ii) STANDARD MARGINAL CONVERGENCE:

$$(4.12) \quad n\mathbb{P}\left[\frac{U_{(i)}^{\leftarrow}(Z_1^{(i)})}{n} > x\right] \rightarrow x^{-1}, \quad x > 0.$$

The marginal condition (4.8) rules out tails that are not heavy. The global condition (4.9) describes dependence among the components. The standard case is where we have tail equivalent marginal tails, each of which is regularly varying with index -1 and normalization by the same constant $b_*(n) = n$ is adequate. See [41, Chapter 5].

5. ESTIMATION

We now survey several estimation problems. After initiating the section with discussion of the Hill estimator, we emphasize multivariate problems.

5.1. The Hill estimator. Suppose $d = 1$ and consider the problem of estimating α , the index of regular variation, based on a sample $\{Z_1, \dots, Z_n\}$ of non-negative iid random variables when the tail probability

$$(5.1) \quad \bar{F}(x) = P[Z_1 > x] \sim x^{-\alpha}L(x), \quad x \rightarrow \infty.$$

A popular estimator is the Hill estimator which is close in spirit to a peaks over threshold maximum likelihood estimator. For $1 \leq i \leq n$, write $Z_{(i)}$ for the i th largest value of Z_1, Z_2, \dots, Z_n . and then Hill's estimator based on k upper order statistics is defined as:

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

The quantile function is

$$(5.3) \quad b(t) := \left(\frac{1}{1-F}\right)^{\leftarrow}(t) = F^{\leftarrow}\left(1 - \frac{1}{t}\right), \quad t > 1.$$

The random measure

$$(5.4) \quad \nu_n =: \frac{1}{k} \sum_{i=1}^n \epsilon_{Z_i/b(n/k)} \Rightarrow \nu_\alpha(\cdot)$$

is a random element of $M_+(0, \infty]$ and is a consistent estimator (in the vague topology) of the measure $\nu_\alpha \in M_+(0, \infty]$ defined by $\nu_\alpha(x, \infty] = x^{-\alpha}$, provided $n \rightarrow \infty$, and $k/n \rightarrow 0$. (See Theorem 2, (8).) Because $b(\cdot)$ is unknown, $b(n/k)$ will be estimated by a consistent estimator, $\hat{b}(n/k)$ to be specified, and we will denote

$$(5.5) \quad \hat{\nu}_n =: \frac{1}{k} \sum_{i=1}^n \epsilon_{Z_i/\hat{b}(n/k)}.$$

We now have ingredients in place to quickly show

$$H_{k,n} \xrightarrow{P} \frac{1}{\alpha}, \quad n \rightarrow \infty, \quad k \rightarrow \infty, \quad k/n \rightarrow 0.$$

The verification of consistency proceeds by a series of steps.

STEP 1. Consistency of the empirical measure given in (5.4) implies

$$(5.6) \quad \frac{Z_{(\lceil kt \rceil)}}{b(n/k)} \xrightarrow{P} t^{-1/\alpha}, \quad \text{in } D(0, \infty],$$

where $\lceil kt \rceil$ is the smallest integer greater than or equal to kt . Assertion (5.6) is proved via the continuous mapping theorem by applying two almost surely continuous maps to (5.4). First apply the map from $M_+((0, \infty]) \mapsto D[0, \infty)$ defined by

$$\mu \mapsto \mu(t^{-1}, \infty], \quad t \geq 0,$$

and then the map which takes a non-decreasing function in $D[0, \infty)$ into its inverse.

STEP 2. Next show $\hat{\nu}_n \xrightarrow{P} \nu_\alpha$, as $n \rightarrow \infty$, $k \rightarrow \infty$ and $k/n \rightarrow 0$. From (5.4) and Step 1, we have

$$(5.7) \quad \left(\nu_n, \frac{Z^{(k)}}{b(\frac{n}{k})} \right) \Rightarrow (\nu_\alpha, 1)$$

in $M_+(0, \infty] \times (0, \infty)$. Since

$$\hat{\nu}_n(\cdot) = \nu_n \left(\frac{Z^{(k)}}{b(\frac{n}{k})} \cdot \right) = \tilde{T} \left(\nu_n, \frac{Z^{(k)}}{b(\frac{n}{k})} \right),$$

the conclusion will follow by the continuous mapping theorem, applied using the scaling operator

$$\tilde{T} : M_+(E) \times (0, \infty) \mapsto M_+(E)$$

defined by

$$\tilde{T}(\mu, x)(A) = \mu(xA).$$

STEP 3. Finish by showing

$$H_{k,n} = \int_1^\infty \hat{\nu}_n(x, \infty] x^{-1} dx \xrightarrow{P} \int_1^\infty \nu_\alpha(x, \infty] x^{-1} dx = \frac{1}{\alpha}.$$

□

Asymptotic normality is usually considered under a second order condition which controls the bias $E(H_{k,n}) - \alpha^{-1}$ and yields asymptotic variance $1/\alpha^2$:

$$\sqrt{k}(H_{k,n} - \alpha^{-1}) \Rightarrow N(0, \alpha^{-2}).$$

The condition is uncheckable but probably a reasonable thing to assume. There are many problems associated with the Hill estimator: lack of location invariance, choice of k , effect of the slowly varying component. See [39, 42, 36, 35, 43, 38, 37, 23, 22, 19, 30, 31, 32, 13, 14].

5.2. Multivariate modeling: Dependence among extreme events. Now consider estimation problems in the multivariate heavy tailed case. For such problems, one can estimate something about the one-dimensional marginal distributions relatively easily using a combination of QQ-plotting techniques and various estimators such as Hill's. It is more difficult to say something informative about the dependence structure.

Bivariate time series of financial returns often show little dependence when returns are small. However, when returns are larger and more extreme in absolute value, then there is more pronounced dependence between the components.

The file `fm-exch1.dat` included with the program *Xtremes* (cf [34]), gives daily spot exchange rates of the currencies of France, Germany, Japan, Switzerland and the UK against the US dollar over a period of 6041 days from January 1971 to February 1994.

Figure 1 gives on the left, a scatter plot of the absolute log daily returns for the French Franc against the absolute log daily returns for the German Deutchmark. Small log absolute returns for one currency are matched by a wide range of values for the other currency. Visually, however, dependence increases as the size of the absolute returns for the pair increases. The pattern varies, however, between different exchange rate processes. The dependence among large daily absolute returns between Japan and Germany (right) is much less pronounced than between France and Germany.

One of the ways to measure dependence components in bivariate data is to compute correlations (assuming they exist). However, correlation is a crude summary of dependence which is most informative only between jointly normal variables. It is a meat cleaver which does not distinguish between the dependence between large values and the dependence between small values. The correlation between the original exchange rate data for Germany and France is 0.579. The correlation between Japan and Germany is higher, namely 0.882. The large dependence between Germany and Japan as measured by correlation is not reflected in the scatterplot indicated in the right side of Figure 1 which indicates less dependence between extremes. The smaller correlation between France and Germany does not indicate the stronger dependence shown in the left plot of Figure 1 for the large absolute values of log returns.

Examine the tails of the squared log returns for France and Germany individually. Based on a combination of QQ plots and Hill plots, we conclude that squared log returns of France and Germany are each heavy tailed with

$$\alpha_{\text{germany}^2} = 1.98, \quad \alpha_{\text{france}^2} = 1.75.$$

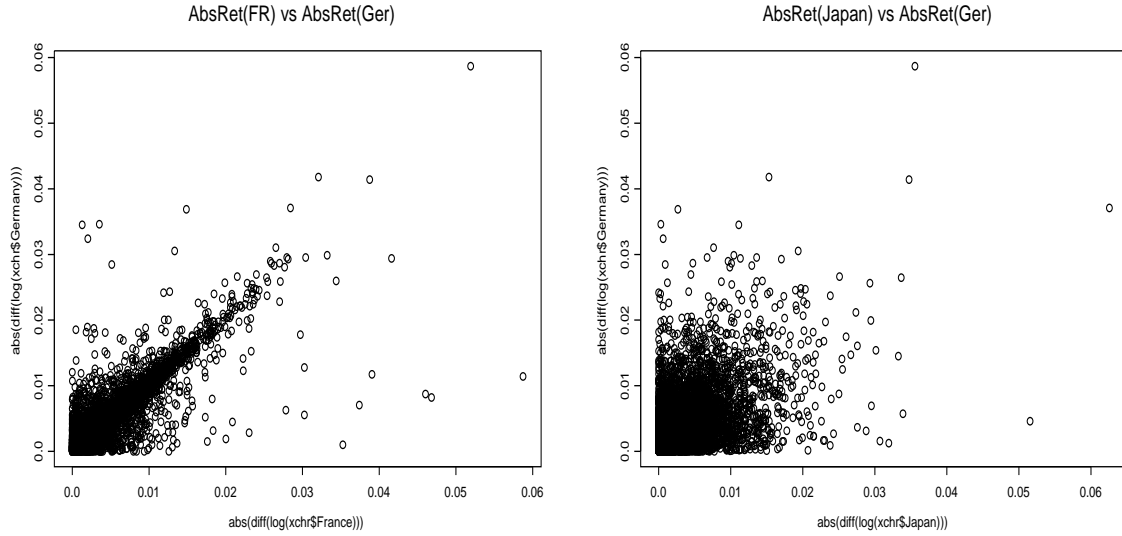


FIGURE 1. Scatter plots of absolute returns of (left) the French Franc against the German Deutschmark and (right) absolute returns of the Japanese Yen against the German Deutschmark.

Figure 2 gives the QQ-plots, which match empirical quantiles of the log transformed data with theoretical quantiles of the exponential distribution.

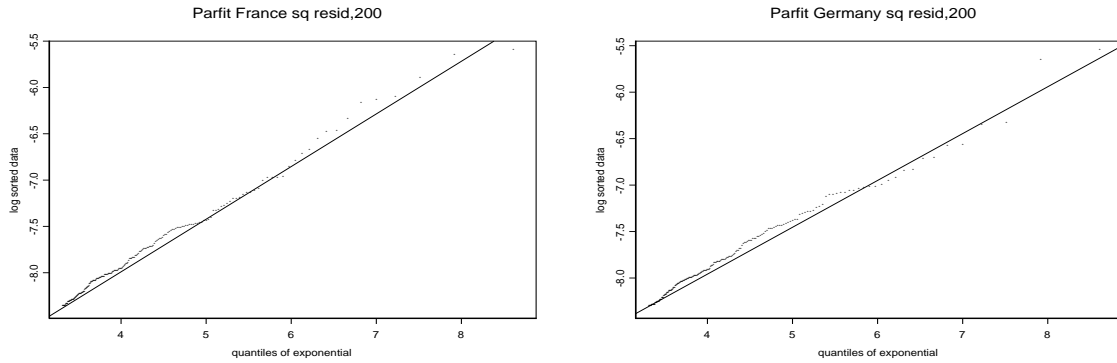


FIGURE 2. QQ plot fitting of α to squared log returns for French exchange rates (left) and German exchange rates.

5.3. How to choose k ? The Stărică plot. The achilles heel of many heavy tailed methods is the difficulty in choosing k . Suppose ν_* is a standard limit measure with the the scaling property (4.11). Suppose $\hat{\nu}_{*,n}$ is our estimator of ν_* . We pick k so that $\nu_{*,n}$ mimics the scaling. We try to use the set

$$\mathfrak{N}^> = \{\mathbf{x} \in \mathbb{E} : \|\mathbf{x}\| > 1\},$$

thinking such a set encompasses information from all directions. So for a fixed k , we graph

$$(5.8) \quad \left\{ \frac{u \hat{\nu}_{*,n}(u \aleph^>)}{\hat{\nu}_{*,n}(\aleph^>)}, u \geq 0.1 \right\},$$

which we call the Stărică plot. The idea [45] is the ratio should be roughly constant for u in a neighborhood of 1 if k is chosen wisely so that $\hat{\nu}_{*,n}$ is close to ν_* . If u is too small, we are using too many small observations from the center of the distribution which are not likely to carry accurate information about the tail.

In practice, we make the Stărică plot (5.8) for various values of k and choose the k which seems to have the plot most closely hugging the horizontal line at height 1. In the standard case, based on the non-negative iid vectors $\{\mathbf{Z}_i, 1 \leq i \leq n\}$ from a distribution $F(\mathbf{x})$ with multivariate regularly varying tail, we make the Stărică plot as follows. Estimate ν_* with the tail empirical measure $\hat{\nu}_{*,n}$. Evaluate (5.8):

$$\frac{u \hat{\nu}_{*,n}(u \aleph^>)}{\hat{\nu}_{*,n}(\aleph^>)} = \frac{u \frac{1}{k} \sum_{i=1}^n \epsilon_{\mathbf{Z}_i / \hat{b}(n/k)}(u \aleph^>)}{\frac{1}{k} \sum_{i=1}^n \epsilon_{\mathbf{Z}_i / \hat{b}(n/k)}(\aleph^>)}$$

where $\hat{b}(n/k) = \|\mathbf{Z}\|_{(k+1)}$, the k -th largest among the one-dimensional set $\{\|\mathbf{Z}_i\|, 1 \leq i \leq n\}$. Since the sum in the denominator of the ratio is k , if we set $u = \|\mathbf{Z}\|_{(j+1)} / \hat{b}(n/k)$, we get for the ratio $\frac{\|\mathbf{Z}\|_{(j+1)}}{\|\mathbf{Z}\|_{(k+1)}} \cdot \frac{j}{k}$. So, to get the Stărică plot, we graph

$$(5.9) \quad \left\{ \left(\frac{\|\mathbf{Z}\|_{(j+1)}}{\|\mathbf{Z}\|_{(k+1)}}, \frac{\|\mathbf{Z}\|_{(j+1)}}{\|\mathbf{Z}\|_{(k+1)}} \cdot \frac{j}{k} \right), 0 \leq j \leq n-1 \right\}.$$

but look closely at values of the abscissa in a neighborhood of 1. The plots will look different for various values of k and some experimentation can be done. It would be desirable to automate this procedure for choosing k and to prove the Stărică plot technique for choosing k is sensible.

Figure 3 gives 4 views of a Starica plot for 5000 independent pairs of iid observations simulated from a Pareto distribution with $\alpha = 1$ and using $k = 1000, 2000, 3000, 3500$.

5.4. Diagnostics based on ranks. To estimate the angular measure S one first must estimate ν_* which may require transformations which depend on information on the marginal tails. A method based on ranks [44, 25] overcomes some of the discomfort inherent in tail estimation since it does not require marginal tail estimates.

Continue to suppose multivariate regular variation (without tail equivalent marginal distributions) and that

$$(5.10) \quad \nu_n(\cdot) = \frac{1}{k} \sum_{i=1}^n \epsilon_{\left(\frac{\mathbf{Z}_i^{(1)}}{b_{(1)}(n/k)}, \dots, \frac{\mathbf{Z}_i^{(d)}}{b_{(d)}(n/k)} \right)}(\cdot) \xrightarrow{P} \nu$$

in $M_+(\mathbb{E})$, where $\mathbb{E} = [0, \infty] \setminus \{0\}$. From our knowledge of marginal properties we know (see (5.6)) for each j , that

$$\frac{\mathbf{Z}_{(\lceil kt^{(j)} \rceil)}^{(j)}}{b_{(j)}(n/k)} \xrightarrow{P} (t^{(j)})^{-1/\alpha_j},$$

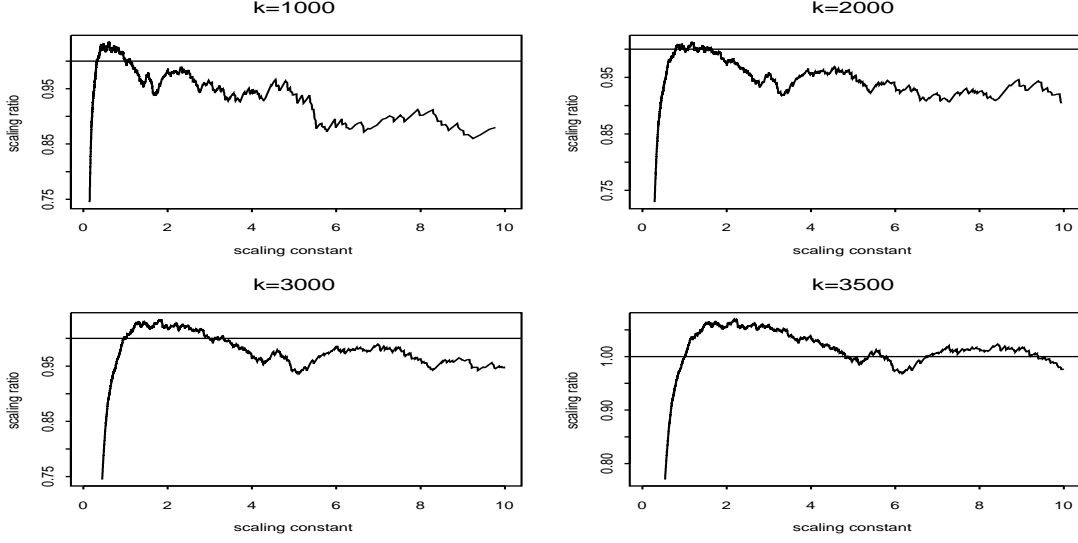


FIGURE 3. Four views of the Starica plot for 5000 simulated Pareto independent pairs, each with $\alpha = 1$.

in $D(0, \infty]$. Because convergence is to a constant limit, we may append this to (5.10) to get

$$(5.11) \quad \left(\nu_n(\cdot), \left(\frac{Z_{(\lceil kt^{(j)} \rceil)}^{(j)}}{b_{(j)}(n/k)}; j = 1, \dots, d \right) \right) \Rightarrow \left(\nu, ((t^{(j)})^{-1/\alpha_j}; j = 1, \dots, d) \right)$$

in $M_+(\mathbb{E}) \times D(0, \infty] \times \dots \times D(0, \infty]$.

The measure ν in (4.9) is related to ν_* by the relationship

$$(5.12) \quad \nu([\mathbf{0}, \mathbf{x}^{1/\alpha}]^c) =: \nu_*([\mathbf{0}, \mathbf{x}]^c).$$

Convert (5.11) into

$$(5.13) \quad \left(\nu_n([\mathbf{0}, \mathbf{x}]^c), \left(\frac{Z_{(\lceil kt^{(j)} \rceil)}^{(j)}}{b_{(j)}(n/k)}; j = 1, \dots, d \right) \right) \Rightarrow \left(\nu([\mathbf{0}, \mathbf{x}]^c), ((t^{(j)})^{-1/\alpha_j}; j = 1, \dots, d) \right)$$

and then apply the continuous map

$$(\nu([\mathbf{0}, \mathbf{x}]^c), \mathbf{t}) \mapsto \nu([\mathbf{0}, \mathbf{t}]^c)$$

to (5.13) to get

$$(5.14) \quad \nu_n \left(\left[\mathbf{0}, \left(\frac{Z_{(\lceil kt^{(j)} \rceil)}^{(j)}}{b_{(j)}(n/k)}; j = 1, \dots, d \right) \right]^c \right) \Rightarrow \nu([\mathbf{0}, \mathbf{t}^{-1/\alpha}]^c).$$

Unpacking the left side of (5.14), we get

$$(5.15) \quad \frac{1}{k} \sum_{i=1}^n \mathbf{1}_{\left[\frac{Z_i^{(j)}}{b_{(j)}(n/k)} \leq \frac{Z_{(\lceil kt^{(j)} \rceil)}^{(j)}}{b_{(j)}(n/k)}; j=1, \dots, d \right]^c} = \frac{1}{k} \sum_{i=1}^n \mathbf{1}_{[Z_i^{(j)} \leq Z_{(\lceil kt^{(j)} \rceil)}^{(j)}; j=1, \dots, d]^c}.$$

Define the (anti)-ranks for $j = 1, \dots, d$

$$r_i^{(j)} = \sum_{l=1}^n 1_{[Z_l^{(j)} \geq Z_i^{(j)}]}$$

to be the number of j -th components bigger than $Z_i^{(j)}$. Rephrase (5.15) as

$$\frac{1}{k} \sum_{i=1}^n 1_{[r_i^{(j)} \geq kt^{(j)}; j=1, \dots, d]^c}.$$

Change variables $\mathbf{s} \mapsto \mathbf{t}^{-1}$ to get

$$\frac{1}{k} \sum_{i=1}^n 1_{[r_i^{(j)} \geq k(s^{(j)})^{-1}; j=1, \dots, d]^c} \Rightarrow \nu([\mathbf{0}, \mathbf{s}^{1/\alpha}]^c)$$

or in $M_+(\mathbb{E})$

$$(5.16) \quad \frac{1}{k} \sum_{i=1}^n \epsilon_{\left(\frac{k}{r_i^{(j)}}; j=1, \dots, d\right)} \Rightarrow \nu_*.$$

To estimate S , apply the polar coordinate transformation

$$T\left(\frac{k}{r_i^{(j)}; j=1, \dots, d}\right) = (R_{i,k}, \Theta_{i,k})$$

and then

$$\frac{1}{k} \sum_{i=1}^n \epsilon_{(R_{i,k}, \Theta_{i,k})} \Rightarrow c\nu_{-1} \times S$$

and the estimator of S is

$$(5.17) \quad \hat{S}_n(\cdot) = \frac{\sum_{i=1}^n \epsilon_{(R_{i,k}, \Theta_{i,k})} \left((1, \infty] \times \cdot \right)}{\sum_{i=1}^n \epsilon_{R_{i,k}}((1, \infty])} = \frac{\sum_{i=1}^n 1_{[R_{i,k} > 1]} \epsilon_{\Theta_{i,k}}(\cdot)}{\sum_{i=1}^n \epsilon_{R_{i,k}}((1, \infty])} \Rightarrow S.$$

5.5. Asymptotic independence. Suppose $\{\mathbf{Z}_n, n \geq 1\}$ are iid and satisfy the conditions of Theorem 4. The distribution F of \mathbf{Z}_1 possesses the property of *asymptotic independence* if

- (1) $\nu_*((\mathbf{0}, \infty)) = 0$ so that ν_* concentrates on the axes;

OR

- (2) S concentrates on $\{\mathbf{e}_i, i = 1, \dots, d\}$.

The distribution F of \mathbf{Z}_1 possesses the property of *asymptotic dependence* if

- (1) ν_* concentrates on $\{t \frac{\mathbf{1}}{\|\mathbf{1}\|} : t > 0\}$, the diagonal line,

or

- (2) S concentrates on $\{\frac{\mathbf{1}}{\|\mathbf{1}\|}\}$.

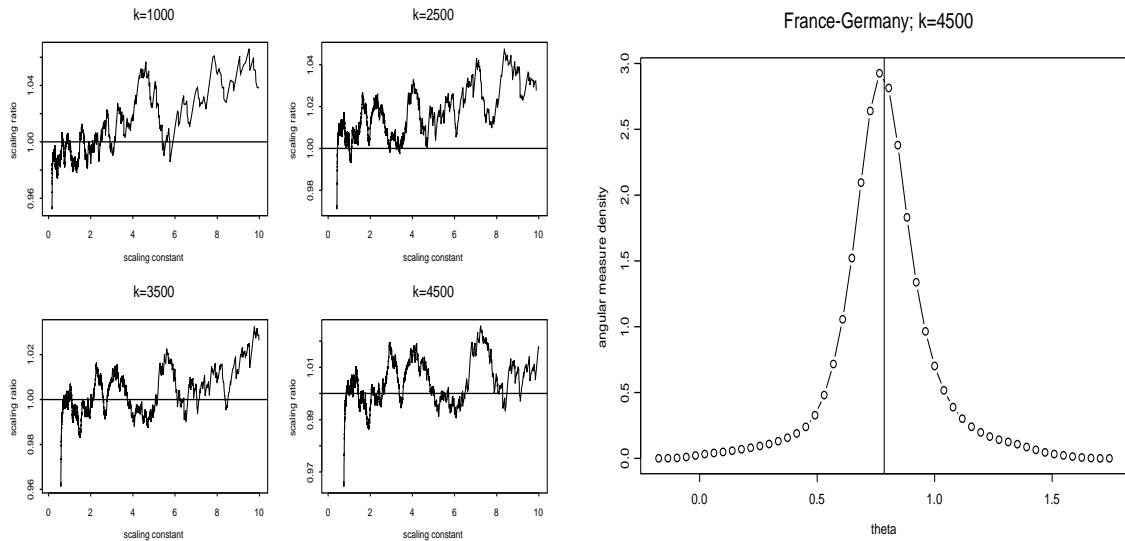


FIGURE 4. Returns from the German mark vs French frank: Starica plot (left) and spectral density estimation.

If $d = 2$ and components of the random vector are tail equivalent, then asymptotic independence corresponds to

$$\mathbb{P}[Z_1^{(1)} > b_n | Z_1^{(2)} > b_n] = \frac{\mathbb{P}[Z_1^{(1)} > b_n, Z_1^{(2)} > b_n]}{\mathbb{P}[Z_1^{(2)} > b_n]} \rightarrow \nu((\mathbf{1}, \infty)) = 0.$$

Hence the name, asymptotic independence. The extreme value background is discussed [41, page 290].

It appears from the scatterplots in Figure 1, that exchange rate returns for (Japan, Germany) might be modelled as asymptotically independent whereas (Germany, France) might be from an asymptotically dependent model.

5.6. Estimating the angular measure. The interpretation of (5.17) is that the empirical measure of those Θ 's whose radius vector is greater than 1 is approximately proportional to S . Apart from normalization of the plot, if we consider the points

$$\{\Theta_i^{(n)} : R_i^{(n)} > 1\}$$

and make a density plot, we should get an estimate of the density of $S(\cdot)$. A notable mode in the density at $\pi/4$ reveals a tendency toward dependence. Modes at 0 and $\pi/2$ show a tendency toward independence, or at least asymptotic independence. Of course, we do not know that S has a density and we could proceed by making plots for the distribution function $S[0, \theta]$, $0 \leq \theta \leq \pi/2$. However, often, density estimate plots are striking and show qualitative differences effectively.

Figure 4 shows the spectral density estimate for the German and French exchange rate returns using ranks. Based on the Starica plot applied to the points in (5.16), we used

$k = 4500$. Note the mode at approximately $\pi/4$, which is indicated by the vertical line. This bears out the promise of the plot in Figure 1 where it seemed that large values seemed to be highly dependent.

Contrast this with the returns from the German mark jointly with those of the Japanese yen. Large values are much less dependent and in fact appear to exhibit asymptotic independence. Figure 5 shows the Starica plot which helps in choosing $k = 1000$ and on the right there is the density estimate which shows a clear tendency towards having 2 modes at 0 and $\pi/2$.

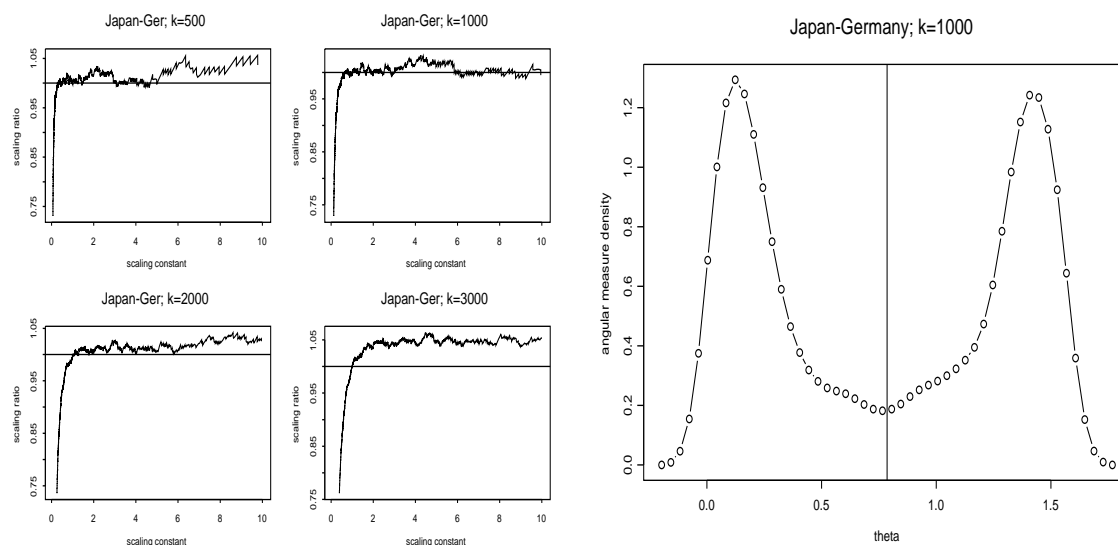


FIGURE 5. Returns from the German mark vs Japanese yen: Starica plot (left) and spectral density estimation.

REFERENCES

- [1] B. Basrak. *The sample autocorrelation function of non-linear time series*. PhD thesis, Rijksuniversiteit Groningen, Groningen, Netherlands, 2000.
- [2] B. Basrak, R. Davis, and T. Mikosch. A characterization of multivariate regular variation. Preprint, 2000.
- [3] P. Billingsley. *Convergence of Probability Measures*. Wiley, New York, 1968.
- [4] Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons Inc., New York, second edition, 1999. A Wiley-Interscience Publication.
- [5] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327, 1986.
- [6] T. Bollerslev, R.Y. Chow, and K.F. Kroner. ARCH modeling in Finance: a review of the theory and empirical evidence. *Journal of Econometrics*, 52:5–59, 1992.
- [7] M. Crovella and A. Bestavros. Explaining world wide web traffic self-similarity. Preprint available as TR-95-015 from {crovella,best}@cs.bu.edu, 1995.
- [8] M. Crovella and A. Bestavros. Self-similarity in world wide web traffic: evidence and possible causes. In *Proceedings of the 1996 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, volume 24, pages 160–169, 1996.

- [9] M. Crovella and A. Bestavros. Self-similarity in world wide web traffic: evidence and possible causes. *Performance Evaluation Review*, 24:160–169, 1996.
- [10] M. Crovella and A. Bestavros. Self-similarity in world wide web traffic: evidence and possible causes. *IEEE/ACM Transactions on Networking*, 5(6):835–846, 1997.
- [11] M. Crovella, A. Bestavros, and M. Taqqu. Heavy-tailed probability distributions in the world wide web. In Murad S. Taqq Robert Adler, Raisa Feldman, editor, *A PRACTICAL GUIDE TO HEAVY TAILS: Statistical Techniques for Analysing Heavy Tailed Distributions*. Birkhäuser, Boston, 1999.
- [12] M. Crovella, G. Kim, and K. Park. On the relationship between file sizes, transport protocols, and self-similar network traffic. In *Proceedings of the Fourth International Conference on Network Protocols (ICNP'96)*, pages 171–180, 1996.
- [13] S. Csörgő, P. Deheuvels, and D. Mason. Kernel estimates for the tail index of a distribution. *Ann. Statist.*, 13:1050–1077, 1985.
- [14] S. Csorgo and D. Mason. Central limit theorems for sums of extreme values. *Math. Proc. Camb. Phil. Soc.*, 98:547–558, 1985.
- [15] Cunha, A. Besavros, and M. Crovella. Characteristics of www client-based traces. Preprint available as BU-CS-95-010 from {crovella,best}@cs.bu.edu, 1995.
- [16] Richard A. Davis and Thomas Mikosch. The sample autocorrelations of heavy-tailed processes with applications to ARCH. *Ann. Stat.*, 26(5):2049–2080, 1998.
- [17] L. de Haan and E. Omey. Integrals and derivatives of regularly varying functions in \mathbf{r}^d and domains of attraction of stable distributions. II. *Stochastic Process. Appl.*, 16(2):157–170, 1984.
- [18] L. de Haan, E. Omey, and S. Resnick. Domains of attraction and regular variation in \mathbf{r}^d . *J. Multivariate Anal.*, 14(1):17–33, 1984.
- [19] L. de Haan and S. Resnick. On asymptotic normality of the hill estimator. *Stochastic Models*, 14:849–867, 1998.
- [20] L. de Haan and S. I. Resnick. Derivatives of regularly varying functions in \mathbf{r}^d and domains of attraction of stable distributions. *Stochastic Process. Appl.*, 8(3):349–355, 1978/79.
- [21] L. de Haan and S.I. Resnick. On regular variation of probability densities. *Stochastic Processes and their Applications*, 25:83–95, 1987.
- [22] H. Drees. *Estimating the Extreme Value Index*. PhD thesis, University of Cologne, Germany, 1998. Habilitationsschrift, Mathematics Faculty.
- [23] H. Drees, L. de Haan, and S. Resnick. How to make a Hill plot. *Annals of Statistics*, 28(1):254–274, 2000.
- [24] Diane E. Duffy, Allen A. McIntosh, Mark Rosenstein, and Walter Willinger. Analyzing telecommunications traffic data from working common channel signaling subnetworks. In *PCmpScSt25*, pages 156–165, 1993.
- [25] J. Einmahl, L. de Haan, and V. Piterbarg. Nonparametric estimation of the spectral measure of an extreme value distribution. *Annals of Statistics*, 29(5):1401–1423, 2001.
- [26] R.F. Engle. Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica*, 50:987–1008, 1982.
- [27] C.M. Goldie. Implicit renewal theory and tails of solutions of random equations. *Ann. Appl. Probab.*, 1(1):126–166, 1991.
- [28] H. Kesten. Random difference equations and renewal theory for products of random matrices. *Acta Math.*, 131:207–248, 1973.
- [29] Will E. Leland, Murad S. Taqqu, Walter Willinger, and Daniel V. Wilson. Statistical analysis of high time-resolution ethernet Lan traffic measurements. In *Proceedings of the 25th Symposium on the Interface between Statistics and Computer Science*, pages 146– 155, 1993.
- [30] D. Mason. Laws of large numbers for sums of extreme values. *Ann. Probability*, 10:754–764, 1982.
- [31] D. Mason. A strong invariance theorem for the tail empirical process. *Ann. Inst. Henri Poincaré*, 24:491–506, 1988.

- [32] D. Mason and T. Turova. Weak convergence of the hill estimator process. In J. Galambos, J. Lechner, and E. Simiu, editors, *Extreme Value Theory and Applications*. Kluwer Academic Publishers, Dordrecht, Holland, 1994.
- [33] T. Mikosch. Modeling dependence and tails of financial time series. In *SemStat: Seminaire Europeen de Statistique, Extreme Value Theory and Applications*. Chapman-Hall, London, 2001. 75 pages.
- [34] R.-D. Reiss and M. Thomas. *Statistical Analysis of Extreme Values*. Birkhäuser Verlag, Basel, second edition, 2001. From insurance, finance, hydrology and other fields, With 1 Windows CD-ROM.
- [35] S. Resnick. Discussion of the danish data on large fire insurance losses. *Astin Bulletin*, 27:139–151, 1997.
- [36] S. Resnick. Heavy tail modeling and teletraffic data. *Ann. Statist.*, 25:1805–1869, 1997.
- [37] S. Resnick. Modeling data networks. In *SemStat: Seminaire Europeen de Statistique, Extreme Value Theory and Applications*. Chapman-Hall, London, 2001. 75 pages.
- [38] S. Resnick and C. Stărică. Asymptotic behavior of Hill’s estimator for autoregressive data. *Stochastic Models*, 13:703–723, 1997.
- [39] S. Resnick and C. Stărică. Smoothing the Hill estimator. *Adv. Appl. Probab.*, 29:271–293, 1997.
- [40] S.I. Resnick. Point processes, regular variation and weak convergence. *Adv. Appl. Probab.*, 18:66–138, 1986.
- [41] S.I. Resnick. *Extreme Values, Regular Variation and Point Processes*. Springer-Verlag, New York, 1987.
- [42] Sidney Resnick and Cătălin Stărică. Consistency of Hill’s estimator for dependent data. *J. Appl. Probab.*, 32(1):139–167, 1995.
- [43] Sidney Resnick and Catalin Stărică. Tail index estimation for dependent data. *Ann. Appl. Probab.*, 8(4):1156–1183, 1998.
- [44] Ashoke Kumar Sinha. *Estimating Failure Probability when Failure is Rare: Multidimensional Case*. PhD thesis, Tinbergen Institute, Erasmus University Rotterdam, 1997. PhD Thesis # 165, Promotor: Prof.dr. L.F.M. de Haan, published by Thela Thesis, Amsterdam.
- [45] C. Starica. Multivariate extremes for models with constant conditional correlations. *J. Empirical Finance*, 6:515–553, 1999.
- [46] C. Starica. Multivariate extremes for models with constant conditional correlations. In P. Embrechts, editor, *Extremes and Integrated Risk Management*. Risk Books, London, 2000.
- [47] W. Vervaat. On a stochastic difference equation and a representation of non-negative infinitely divisible random variables. *Adv. in Appl. Probab.*, 11:750–783, 1979.
- [48] W. Willinger, M.S. Taqqu, M. Leland, and D. Wilson. Self-similarity in high-speed packet traffic: analysis and modelling of ethernet traffic measurements. *Statistical Science*, 10:67–85, 1995.
- [49] W. Willinger, M.S. Taqqu, M. Leland, and D. Wilson. Self-similarity through high variability: statistical analysis of ethernet lan traffic at the source level. *Computer Communications Review*, 25:100–113, 1995. Proceedings of the ACM/SIGCOMM’95, Cambridge, MA.

SIDNEY RESNICK, SCHOOL OF OPERATIONS RESEARCH AND INDUSTRIAL ENGINEERING, CORNELL UNIVERSITY, ITHACA, NY 14853

WWW.ORIE.CORNELL.EDU/~SID

E-mail address: sid@orie.cornell.edu