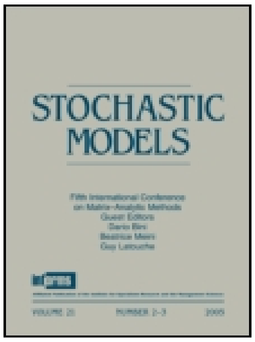

ISSN: 1532-6349 (Print) 1532-4214 (Online) Journal homepage: www.tandfonline.com/journals/lstm20

The Extremal Dependence Measure and Asymptotic Independence

Sidney Resnick

STOCHASTIC MODELS
Vol. 20, No. 2, pp. 205–227, 2004

The Extremal Dependence Measure and Asymptotic Independence

Sidney Resnick*

School of Operations Research and Industrial Engineering,
Cornell University, Ithaca, New York, USA

ABSTRACT

Extremal dependence analysis assesses the tendency of large values of components of a random vector to occur simultaneously. This kind of dependence information can be qualitatively different than what is given by correlation which averages over the total body of the joint distribution. Also, correlation may be completely inappropriate for heavy tailed data. We study the *extremal dependence measure* (EDM), a measure of the tendency of large values of components of a random vector to occur simultaneously and show consistency of an estimator of the EDM. We also show asymptotic normality of an idealized estimator in a restricted case of multivariate regular variation where scaling functions do not have to be estimated.

Key Words: Heavy tails; Multivariate regular variation; Pareto tails; Asymptotic independence; Extremal dependence.

*Correspondence: Sidney Resnick, School of Operations Research and Industrial Engineering, Rhodes Hall 214, Cornell University, Ithaca, NY 14853, USA; Fax: (607) 255-9129; E-mail: sir1@comell.edu.

DOI: 10.1081/STM-120034129 1532-6349 (Print); 1532-4214 (Online)

www.dekker.com

1. INTRODUCTION

Extremal dependence analysis assesses the tendency of large values of components of a random vector to occur simultaneously. This kind of dependence information can be qualitatively different than what is given by numerical summaries, such as correlation, which average over the total body of the joint distribution. Two examples of the type of questions that extremal dependence analysis deals with are

- Is a large movement in exchange rate returns in one currency, such as the German Mark relative to the US Dollar, likely to be accompanied by a similar large movement in another currency, such as the French Franc? (See Refs.[57,58].)
- In Internet traffic, are large file sizes likely to imply a large transmission duration? (See Refs.[5,35].)

Asymptotic independence for a bivariate vector means the probability of both components being large is of smaller order than the probability of one of them being large. If the components of the vector are non-negative and heavy tailed, bivariate data from the distribution of such a vector has a scatter plot with data tending to hug the axes because if both components are unlikely to be simultaneously big, there will be few data points in the interior of the positive quadrant.

To illustrate the point, consider the the following example from Ref.[48]. The file fm-exch1.dat included with the program *Xtremes* (cf. Ref.[40]), 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. Note the observation period is well before the introduction of the Euro.

Figure 1 gives on the left, a scatter plot of the absolute daily log returns for the French Franc against the absolute daily log returns for the German Mark. Small log absolute returns for one currency are matched by a wide range of values

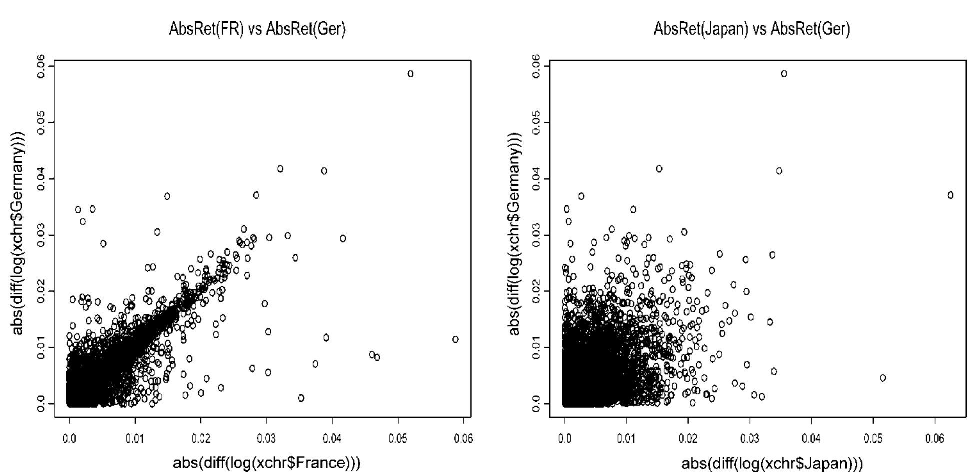

Figure 1. Scatter plots of absolute returns of (left) the French Franc against the German Mark and (right) absolute returns of the Japanese Yen against the German Mark.

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.

In the Internet traffic context, the dependence structure of large values of file size, transmission duration of the file and transmission throughput (file size divided by duration) has been analyzed in Refs.[5,35]. Correlations are emphasized in Ref.[61].

The data shown in Fig. 2 are based on HTTP responses, gathered from the University of North Carolina main link during April of 2001 in a measurement study initiated by Kevin Jaffay and Don Smith (CS, UNC). An HTTP "response" is set of packets associated with a single HTTP data transfer, and "duration" is the time between the first and last packets. Packets were gathered over 21 four hour blocks, over each of the 7 days of the week, and for "morning" (8:00AM–12:00AM), "afternoon" (1:00PM–5:00PM) and "evening" (7:30PM–11:30PM) periods on each day. The total number of HTTP flows over the four hour blocks ranged from ~1 million (weekend mornings) to ~7 million (weekday afternoons). Here we only consider Tuesday afternoon large flows, meaning thresholded data restricted to responses with more than 100 kilobytes.

The left plot in Fig. 2 corresponding to duration vs throughput rate seems to exhibit clear axis hugging meaning there is little tendency for large durations to be associated with large rates. The middle plot for file size vs rate seems to exhibit some similar tendency. The right plot of duration vs inverse rate seems to exhibit extremal dependence. WARNING: Asymptotic independence is an asymptotic distributional property requiring scales for each variable to be adjusted appropriately and merely looking at scatter plots is not adequate. More careful study is required. In Refs.[5,35], it was found that the strongest tendency towards extremal independence was in the pair (file size, rate).

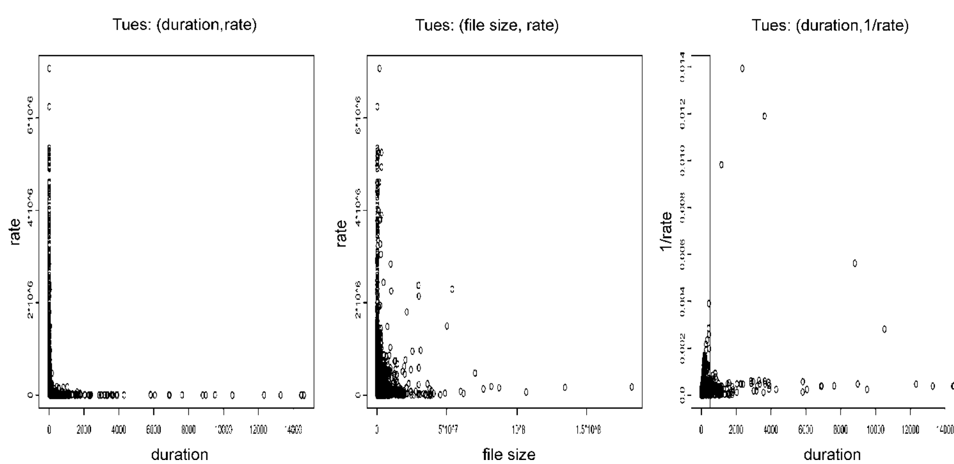

Figure 2. Scatter plots of (left) duration vs. rate, (middle) file size vs. rate and (right) duration vs. inverse rate.

In Sec. 2, we review the notion of multivariate regular variation which underpins the theory of multivariate heavy tailed analysis. We review asymptotic independence and asymptotic full dependence and the statistical goal is to detect these situations from data. We also extend these ideas to stochastic processes and define a *regularly varying process*, which is an abuse of language since, of course, it is the distribution of the process with the regular variation property. Section 3 restricts attention to vectors of dimension 2 and defines the *extremal dependence measure* (EDM) which is standardized to have the look and feel of correlation. When the EDM is 0, we have asymptotic independence and when the EDM is 1, asymptotic full dependence is present. The EDM can also be used as a diagnostic for regularly varying processes and indicates independence properties between large values which are separated by sufficient time lags. The EDM was used in Ref.[5] as an exploratory data analysis tool and our goal is to begin the study of its mathematical properties.

Section 4 discusses estimators for parametric and semi-parametric quantities related to multivariate heavy tailed analysis and in particular suggests an estimator of the EDM. Consistency is shown for this estimator. Section 5 considers asymptotic normality of an idealized estimator of the EDM in the restricted case that the scaling functions do not have to be estimated.

In practice, heavy tailed vectors rarely, if ever, have component random variables with the same regular variation indices. None-the-less, this paper often restricts attention to the *standard case*, which assumes each component random variable is tail equivalent to the others and each component variable has a distribution tail which is regularly varying with index -1. This assumption makes the probabilistic analysis and exposition clearer but begs the statistical question of how to transform a non-standard case to standard. There are two suggested methods to accomplish this involving either power transformations or ranks (see Refs.[5,17,27,30–32,35,39,41,47,48,56,59]) but the supremacy of either method in practice is not yet completely clear and proving asymptotic normality of the EDM without the standard case assumption is more difficult. So we have assumed a relatively easy case and more work remains to be done. We hope to address this soon.

Of course, there are other papers which attempt to assess extremal dependence. Each has its own assumptions (explicit or implicit), parameters and their estimators, and approach. Ledford and Tawn[31–33] define the *coefficient of tail dependence* η which is related to *second order regular variation*[17] and *hidden regular variation*[47]; see also Ref.[55]. Coles et al.[6] define two measures of dependence $(\chi, \bar{\chi})$ and such measures are reviewed in Ref.[29]. Also, Ledford and Tawn[34] investigate extremal dependence structure in stationary time series. Our EDM measure joins a growing list of attempts to assess extremal dependence. Although similar in spirit, our EDM differs from what is used in these papers in that it is based only on assuming multivariate regular variation of the observation vector and is basically a distribution moment instead of a the measure of a region in the positive quadrant.

2. STANDARD CASE MULTIVARIATE REGULAR VARIATION

Before getting to the meat and potatoes, here is a list of notational conventions that will make exposition easier.

2.1. Vector Notation Review

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

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

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

$$\boldsymbol{x} < \boldsymbol{z} \text{ means } x^{(i)} < z^{(i)}, \quad i = 1, \ldots, d, \qquad \boldsymbol{x} \le \boldsymbol{z} \text{ means } x^{(i)} \le z^{(i)}, \quad i = 1, \ldots, d,$$

$$\boldsymbol{x} = \boldsymbol{z} \text{ means } x^{(i)} = z^{(i)}, \quad i = 1, \ldots, d, \qquad \boldsymbol{z}\boldsymbol{x} = (z^{(1)}x^{(1)}, \ldots, z^{(d)}x^{(d)}),$$

$$\boldsymbol{x} \vee \boldsymbol{z} = (x^{(1)} \vee z^{(1)}, \ldots, x^{(d)} \vee z^{(d)}), \qquad \frac{\boldsymbol{x}}{\boldsymbol{z}} = \left(\frac{x^{(1)}}{z^{(1)}}, \ldots, \frac{x^{(d)}}{z^{(d)}}\right)$$

$$\boldsymbol{x}^{\boldsymbol{z}} = \left((x^{(1)})^{z^{(1)}}, \ldots, (x^{(d)})^{z^{(d)}}\right) \qquad c\boldsymbol{x} = (cx^{(1)}, \ldots, cx^{(d)}),$$

for a real number c. Also $\mathbf{0} = (0, \ldots, 0)$, $\mathbf{1} = (1, \ldots, 1)$, $\boldsymbol{e}_i = (0, \ldots, 1, \ldots, 0)$ for $i = 1, \ldots, d$.

We denote rectangles by

$$[\boldsymbol{a}, \boldsymbol{b}] = \{\boldsymbol{x} \in \mathbb{R} : \boldsymbol{a} \le \boldsymbol{x} \le \boldsymbol{b}\}$$

so that for $\boldsymbol{x} > 0$ and $\mathbb{E} = [\mathbf{0}, \infty]\backslash\{\mathbf{0}\}$,

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

2.2. Multivariate Regularly Varying Functions

A subset $C \subset \mathbb{R}^d$ is a *cone* if whenever $\boldsymbol{x} \in C$ also $t\boldsymbol{x} \in C$ 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 $\boldsymbol{x}, \boldsymbol{y} \in C$ and $\boldsymbol{x} \le \boldsymbol{y}$ we have $h(\boldsymbol{x}) \le h(\boldsymbol{y})$. The natural domain for a multivariate regularly varying function is a cone.

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

$$\lim_{t\to\infty} \frac{h(t\boldsymbol{x})}{h(t\mathbf{1})} = \lambda(\boldsymbol{x}). \tag{2.1}$$

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

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

See Refs.[1,2,18,19,22,23,36,46].

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

$$\lim_{t\to\infty} \frac{1-F(t\boldsymbol{x})}{1-F(t\mathbf{1})} = \lambda(\boldsymbol{x}). \tag{2.3}$$

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

2.3. Multivariate Regularly Varying Tail Probabilities

There are various equivalences which define multivariate regularly varying tail probabilities. We restrict attention to the case of random vectors with non-negative components. Suppose $\{\boldsymbol{Z}_n, n \geq 1\}$ are iid random elements of $\mathbb{R}_+^d$ with common distribution $F(\cdot)$. Recall $\mathbb{E} = [\mathbf{0},\boldsymbol{\infty}]\backslash\{\mathbf{0}\}$ and $[\mathbf{0},\boldsymbol{x}]^c = \mathbb{E}\backslash[\mathbf{0},\boldsymbol{x}]$. Set $M_+(\mathbb{E})$ to be the space of positive Radon measures on $\mathbb{E}$ and $M_p(\mathbb{E})$ is the space of Radon point measures on $\mathbb{E}$. Define the measure on $(0,\infty]$

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

Vague convergence of measures is denoted by $\xrightarrow{v}$.

Fix a norm $\|\cdot\|$ and with respect to this norm define the unit sphere

$$\aleph = \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\| = 1\}.$$

Set $\aleph_+ = \aleph \cap \mathbb{E}$. Define the polar coordinate transformation $T: \mathbb{R}^d\backslash\{\mathbf{0}\} \mapsto (0,\infty) \times \aleph$ by

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

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

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

Think of $\boldsymbol{a} \in \aleph$ as defining a direction and r telling how far in direction $\boldsymbol{a}$ 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 $T(\boldsymbol{x}) = (r\cos\theta, r\sin\theta)$, where $0 \leq \theta \leq 2\pi$, which is the usual polar coordinate transformation. However, the more consistent notation would be $T(\boldsymbol{x}) = \big(r, (\cos\theta, \sin\theta)\big)$. When $d=2$, for a measure S on the Borel subsets of $\aleph_+$ we write

$$S_1(a,b] = S_1\left\{\theta \in (a,b] \subset \left[0,\frac{\pi}{2}\right]\right\} = S\{(\cos\theta,\sin\theta) : \theta \in (a,b]\}. \tag{2.4}$$

For a random vector $\boldsymbol{X}$ in $\mathbb{R}^d$ we sometimes write $T(\boldsymbol{X}) = (R,\boldsymbol{\Theta})$.

To deal with multivariate regular variation of tail probabilities, we have to consider a punctured space with a one-point un-compactification such as $[\mathbf{0}, \boldsymbol{\infty}]\backslash\{\mathbf{0}\}$. Equivalences in terms of polar coordinates are then problematic since the polar coordinate transformation is not defined on the lines through $\boldsymbol{\infty}$, so some sort of restriction argument is necessary. A different treatment of the polar coordinate transformation is given in Refs.[1,2,37]. For versions of the following see Refs.[46,48]. We use the notation

$$\epsilon_x(A) = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \notin A, \end{cases}$$

for the point measure with all mass at x and also denote *vague convergence* by "$\stackrel{v}{\to}$".

Theorem 1 (Multivariate Regularly Varying Tail Probabilities)**.** *Suppose* $\{\boldsymbol{Z}_m, m \geq 1\}$ *are iid* $\mathbb{R}_+^d$*-valued random vectors with common distribution* F*. The following statements are equivalent.* (*In each, we understand the phrase Radon measure to mean a not identically zero Radon measure. Also, repeated use of the symbols* ν, $b(\cdot)$, $\{b_n\}$ *from statement to statement, does not require these objects to be exactly the same in different statements. See Remark* 1 *after Theorem* 1.)

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

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

for all points $\boldsymbol{x} \in [\mathbf{0}, \boldsymbol{\infty})\backslash\{\mathbf{0}\}$ *which are continuity points of* $\nu([\mathbf{0}, \cdot]^c)$.

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

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

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

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

(4) *There exists a probability measure* $S(\cdot)$ *on* $\aleph_+$ *and a function* $b(t) \to \infty$ *such that for* $(R_1, \boldsymbol{\Theta}_1) = \left(\|\boldsymbol{Z}_1\|, \frac{\boldsymbol{Z}_1}{\|\boldsymbol{Z}_1\|}\right)$ *we have*

$$t\mathbb{P}\left[\left(\frac{R_1}{b(t)}, \boldsymbol{\Theta}_1\right) \in \cdot\right] \stackrel{v}{\to} c\nu_\alpha \times S$$

in $M_+\big((0, \infty] \times \aleph_+\big)$, *where* $c > 0$.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

(5) *There exists a probability measure* $S(\cdot)$ *on* $\aleph_+$ *and a sequence* $b_n \to \infty$ *such that for* $(R_1, \boldsymbol{\Theta}_1) = \left(\|\boldsymbol{Z}_1\|, \dfrac{\boldsymbol{Z}_1}{\|\boldsymbol{Z}_1\|} \right)$ *we have*

$$n\mathbb{P}\left[\left(\frac{R_1}{b_n}, \boldsymbol{\Theta}_1\right) \in \cdot\right] \xrightarrow{v} c\nu_\alpha \times S$$

in $M_+\big((0, \infty] \times \aleph_+\big)$, *where* $c > 0$.

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

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

where $PRM(\nu)$ *is a Poisson random measure with mean measure* ν.

(7) *There exists a sequence* $b_n \to \infty$ *such that in* $M_p((0, \infty] \times \aleph_+)$

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

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

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

$$\nu_n := \frac{1}{k} \sum_{i=1}^{n} \epsilon_{\boldsymbol{Z}_i/b\left(\frac{n}{k}\right)} \Rightarrow \nu \tag{2.5}$$

and (8) *is equivalent to any of* (1)–(7), *provided* $k(\cdot)$ *satisfies* $k(n) \sim k(n+1)$; *that is,* $\lim_{n\to\infty} k(n)/k(n+1) = 1$. *Similar statements to* (2.5) *can be made in terms of polar coordinates.*

Remark 1. (1) Normalization of all components by the same function means that marginal distributions are tail equivalent; that is, in Refs.[44,46]

$$\lim_{x\to\infty} \frac{P[Z_1^{(i)} > x]}{P[Z_1^{(j)} > x]} =: r_{ij} \in [0, \infty],$$

for $1 \le i, j \le d$. To avoid cases where some marginal tails are heavier than others, corresponding to $r_{ij} = 0$ or ∞ for some (i, j), we usually assume all components $\{Z_1^{(i)}, 1 \le i \le d\}$ are identically distributed.

(2) When marginals are the same and $b(t) = t$ or $b_n = n$ we are in the *standard case*[21,46] and all marginal distributions are tail equivalent to a standard Pareto distribution with $\alpha = 1$. In general, the possible choices of $b(t)$ include

(i) $b(t) = \left(\dfrac{1}{1 - F_{(1)}}\right)^{\leftarrow}(t)$ where $F_{(1)}(x) = P[Z_1^{(1)} \le x]$ is the one-dimensional marginal distribution.

(ii) $b(t) = \left(\frac{1}{1-F_R}\right)^{\leftarrow}(t)$ where $F_R(x) = P[R_1 \leq x]$ is the distribution of $\|\boldsymbol{Z}_1\|$. Note this choice of $b(\cdot)$ depends on the choice of norm $\|\cdot\|$.

Different choices of $b(\cdot)$ may introduce different constants c in the limit statements.

(3) The parameter χ discussed in Refs.[6,29] is, apart from a multiplicative constant, $\nu(\mathbf{1}, \boldsymbol{\infty}]$.

Definition 1. A non-negative stationary stochastic sequence $\boldsymbol{X} = \{X^{(n)}, n \geq 0\}$ is called a regularly varying process if for every $d \geq 0$, there exists a Radon measure $\nu_{0,\ldots,d}(\cdot)$ on $[0,\infty]^{d+1}\backslash\{\mathbf{0}\}$ such that

$$nP[b_n^{-1}(X^{(0)}, \ldots, X^{(d)}) \in \cdot] \xrightarrow{v} \nu_{0,\ldots,d}(\cdot), \tag{2.6}$$

where we assume b_n is chosen to satisfy

$$nP[X^{(1)} > b_n x] \to x^{-\alpha}, \quad x > 0.$$

An analogous definition can be made in continuous time. Examples of continuous time regularly varying processes are stationary stable processes[53,54] and stationary max-stable processes.[15,16,20] In discrete time there are moving and max-moving averages with heavy tailed innovations as well as autoregressions and ARMA's. Stationary ARCH and GARCH processes are also regularly varying processes.[2,12,37,38]

2.4. Asymptotic Independence

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

(1) $\nu(\mathbb{E}^0) = 0$ so that ν concentrates on the axes where

$$\mathbb{E}^0 = \{\boldsymbol{s} \in \mathbb{E} : \text{For some } 1 \leq i < j \leq d, s^{(i)} \wedge s^{(j)} > 0\};$$

OR

(2) S concentrates on $\{\boldsymbol{e}_i, i = 1, \ldots, d\}$, where recall

$$\boldsymbol{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)$$

is the ith basis vector.

If $d = 2$, then ν concentrates on the horizontal and vertical axes and S puts all mass on the two basis vectors on the two axes.

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

(1) ν concentrates on $\left\{t\frac{\mathbf{1}}{\|\mathbf{1}\|} : t > 0\right\}$, the diagonal line,

or

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

When $d = 2$ and components of the random vector are tail equivalent, asymptotic independence means

$$\mathbb{P}\left[Z_1^{(1)} > b_n | Z_1^{(2)} > b_n\right] = \frac{\mathbb{P}\left[Z_1^{(1)} > b_n, Z_1^{(2)} > b_n\right]}{\mathbb{P}\left[Z_1^{(2)} > b_n\right]} \to v((\mathbf{1}, \boldsymbol{\infty})) = 0.$$

Hence the name, asymptotic independence. The extreme value background is discussed in Ref.[46], page 290.

The goal is to detect asymptotic independence statistically.

3. THE EXTREMAL DEPENDENCE MEASURE

The extremal dependence measure is a crude measure of dependence between large values of the various components of a random vector. It is convenient to suppose $d = 2$; higher dimensional analogues are possible but would be cumbersome and, in any event, a d-dimensional random vector possesses asymptotic independence iff all component pairs are asymptotically independent (see page 296 of Ref.[46]).

Suppose $\boldsymbol{Z} = (Z^{(1)}, Z^{(2)})$ is a bivariate heavy tailed vector whose distribution F satisfies the conditions of Theorem 1. For $d = 2$, assume the parameterization on $[0, \pi/2]$ given by (2.4) and the angular measure S is replaced by S_1. Define

$$\upsilon := \int_0^{\pi/2} \left(\theta - \frac{\pi}{4}\right)^2 S_1(d\theta). \tag{3.1}$$

The distribution of $\boldsymbol{Z}$ possesses asymptotic full dependence iff $\upsilon = 0$ and asymptotic independence iff $\upsilon = (\pi/4)^2$. These facts follow easily since the integral in (3.1) is extreme when S_1 concentrates all mass on the atoms $\{0, \pi/2\}$. The extremal dependence measure is defined by

$$\rho := 1 - \frac{\upsilon}{(\pi/4)^2}, \tag{3.2}$$

and the distribution of $\boldsymbol{Z}$ possesses asymptotic independence iff $\rho = 0$ and asymptotic full dependence iff $\rho = 1$.

Remark 2. The limit measure v is associated with *spectral functions* $f_1, f_2 \in L_1([0, 1], ds)$ which are integrable on $[0, 1]$ with respect to Lebesgue measure. See, for example, Proposition 5.11, page 268 of Ref.[46]. In terms of these functions, the integral υ in (3.1) can be expressed as

$$\upsilon = \int_0^1 \left| \arctan \frac{f_2(s)}{f_1(s)} - \arctan 1 \right|^2 ds.$$

Suppose $\{X^{(n)}, n \geq 0\}$ is a regularly varying process defined in (2.6). We define

$$\rho(l) = \text{ the EDM of } (X^{(0)}, X^{(l)}), \tag{3.3}$$

which we write in shorthand as

$$\rho(l) = \mathrm{EDM}(X^{(0)}, X^{(l)}).$$

Return to the case $d = 2$ and assume Theorem 1 holds. Then

$$\bigvee_{i=1}^{n} \frac{\boldsymbol{Z}_i}{b_n} \Rightarrow \left(M^{(\infty)}(1), M^{(\infty)}(2)\right)$$

where the limit is a max-stable random vector. (Recall the maximum notation "$\vee$" introduced in Subsec. 2.1.) If the EDM is 0, the components of the limit are independent. The next Theorem helps explain how the EDM controls simultaneous large values in a regularly varying process. (cf. Proposition 3.2.1, page 89 of Ref.[4].)

Theorem 2. *Suppose* $\boldsymbol{X} := \{X^{(n)}, n \geq 0\}$ *is a regularly varying sequence satisfying* (2.6); *recall we assume* $X^{(n)} \geq 0$ *for all* $n \geq 0$. *Suppose the* EDM *of* $\boldsymbol{X}$ *has the property that*

$$\rho(l) = \mathrm{EDM}(X^{(0)}, X^{(l)}) = 0, \quad \textit{for } l \geq L. \tag{3.4}$$

Let $\{\boldsymbol{X}(j), j \geq 1\}$ *be iid* $\mathbb{R}_+^{\infty}$ *valued random elements with* $\boldsymbol{X}(j) \stackrel{d}{=} \boldsymbol{X}$. *Then in* $\mathbb{R}_+^{\infty}$,

$$\frac{\bigvee_{j=1}^{n} \boldsymbol{X}(j)}{b_n} \Rightarrow \boldsymbol{M}(\infty), \tag{3.5}$$

where $\boldsymbol{M}(\infty) = (M^{(1)}(\infty), M^{(2)}(\infty), \ldots)$ *is a max-stable process with the property that if* I_1, I_2 *are two finite subsets of the non-negative integers,* $I_m \subset \mathbb{N} = \{0, 1, \ldots\}$, $m = 1, 2$ *and* $I_1 \cap I_2 = \emptyset$ *and* $\inf I_2 - \sup I_1 \geq L$ *then*

$$\{M^{(l)}(\infty), l \in I_1\} \quad \textit{and} \quad \{M^{(l)}(\infty), l \in I_2\}$$

are independent.

Proof. The limit sequence must be a stationary max-stable sequence (cf. Refs.[15,16,20,52,60]). There exist (see, for example page 268 of Ref.[46]) non-negative functions $f_j, g_m,\ j \in I_1, m \in I_2$ in $L_1([0,1], ds)$ such that (assuming, without loss of generality, that $\alpha = 1$)

$$\begin{aligned} &P[M^{(j)}(\infty) \leq x_j, j \in I_1; M^{(m)}(\infty) \leq x_m, m \in I_2] \\ &\quad = \exp\left\{-\int_0^1 \left(\bigvee_{j \in I_1} \frac{f_j(s)}{x_j} \vee \bigvee_{m \in I_2} \frac{g_m(s)}{x_m}\right) ds\right\}. \end{aligned} \tag{3.6}$$

Note that $\rho(l)=0$ for $l \geq L$ implies that for $j \in I_1,\ m \in I_2\ M^{(\infty)}(i)$ and $M^{(\infty)}(m)$ are independent and therefore,

$$f_j(s) g_m(s) = 0, \tag{3.7}$$

almost everywhere. Now write the exponent in (3.6) as

$$\int_0^1 \left(\bigvee_{j \in I_1} \frac{f_j(s)}{x_j} \vee \bigvee_{m \in I_2} \frac{g_m(s)}{x_m} \right) ds = \int_{[\vee_{j \in I_1} f_j > 0]} + \int_{[\vee_{j \in I_1} f_j = 0]}$$

and because of (3.7), this is the same as

$$\begin{aligned} &= \int_{[\vee_{j \in I_1} f_j > 0]} \bigvee_{j \in I_1} \frac{f_j(s)}{x_i}\, ds + \int_{[\vee_{j \in I_1} f_j = 0]} \bigvee_{m \in I_2} \frac{g_m(s)}{x_m}\, ds \\ &= \int_{[\vee_{j \in I_1} f_j > 0]} \bigvee_{j \in I_1} \frac{f_j(s)}{x_i}\, ds + \int_{[\vee_{m \in I_2} g_m > 0]} \bigvee_{m \in I_2} \frac{g_m(s)}{x_m}\, ds \end{aligned}$$

which is of the form

$$h_1(x_j, j \in I_1) + h_2(x_m, m \in I_2). \tag{3.8}$$

This suffices to show the independence. □

Corollary 1. *Assume the notation and conditions of Theorem* 2. *Let* $|I_j|$ *be the number of elements in* I_j. *Then in* $M_p\big([0,\infty]^{|I_1|+|I_2|} \setminus \{\mathbf{0}\}\big)$ *we have*

$$\sum_{i=1}^{n} \epsilon_{b_n^{-1}\left(X^{(j)}(i), j \in I_1; X^{(m)}(i), m \in I_2\right)} \Rightarrow N_1 + N_2 \tag{3.9}$$

where N_1 *and* N_2 *are independent PRM's of the form*

$$N_1 = \sum_k \epsilon_{(\,j^{(l)}(k),\ l \in I_1,\ \mathbf{0})},$$

$$N_2 = \sum_k \epsilon_{(\mathbf{0},\ j^{(m)}(k),\ m \in I_2)}.$$

Proof. The conclusion (3.5) of Theorem 2 implies

$$nP[b_n^{-1}(X^{(j)}(i), j \in I_1; X^{(m)}(i), m \in I_2) \in [\mathbf{0}, \boldsymbol{x}]^c] \to h_1(x_j, j \in I_1) + h_2(x_m, m \in I_2)$$

using the notation of (3.8). The stated result now follows from [Proposition 3.21, page 154 of Ref.[46]]. □

Corollary 2. *Assume the notation and conditions of Theorem* 2. *Suppose* $0 < \alpha < 2$. *Then there exists* c_n *such that in* $\mathbb{R}_+^{\infty}$

$$\sum_{i=1}^{n} \frac{\boldsymbol{X}(i)}{b_n} - nc_n \Rightarrow \boldsymbol{X}_\alpha,$$

where $\boldsymbol{X}_\alpha$ *is a stationary* α*-stable sequence such that*

$$\{X_\alpha^{(i)}, i \in I_1\} \quad \text{and} \quad \{X_\alpha^{(m)}, m \in I_2\}$$

are independent.

Proof. Convergence of finite dimensional distributions is proven as in Ref.[42]. The sequence $\{X_\alpha^{(i)}, i \in I_p\}$ is constructed from N_p, $p = 1, 2$. Independence follows from Corollary 1. □

4. ESTIMATORS AND DIAGNOSTICS

Suppose the conditions of Theorem 1 hold. The analogue of (2.5) provides a way to estimate the spectral measure S:

$$\frac{1}{k}\sum_{i=1}^{n} \epsilon_{\left(\frac{R_i}{b(n/k)}, \Theta_i\right)} \Rightarrow c\nu_\alpha \times S. \tag{4.1}$$

as $k = k(n) \to \infty$ and $k/n \to 0$. There are (at least) three problems:

- How does one estimate $b(\cdot)$? The function $b(\cdot)$ is only determined up to an asymptotic constant. In the standard case $b(t) = t$ but if Theorem 1 is assumed to hold with general $b(\cdot)$, we can replace $b(n/k)$ with $\hat{b}(n/k)$, the kth largest order statistic of the iid sample $R_1, \ldots, R_n$. Explanations of why this works are now standard; see Refs.[17,25,30,49]. (In practice, if we have transformed a non-standard case to standard by power transformation, the order statistic normalization seems to work better than normalizing by n/k.)
- What is c? This can be scaled away either by taking ratios in (4.1) or by assuming $b(t) = \left(\frac{1}{1-F_R}\right)^{\leftarrow}(t)$, since then $tP[R_1 > b(t)] \to 1$ as $t \to \infty$.
- How does one choose k? A weakness of many extreme value methodologies is sensitivity of estimates to choice of threshold or choice of k. There are various solutions – bootstrap,[11] smoothing,[50,51] asymptotic optimality[26] – but none is completely satisfactory. There is a graphical technique due to Stărică[41,48,57] which seems to work reasonably well in the multivariate case and which requires examining scaling plots and some trial and error. This technique is exploratory and to-date, nothing is proven about it.

Specialize to $d = 2$. Let $\{\boldsymbol{Z}_j, 1 \le j \le n\}$ be iid random pairs. Using an estimator $\hat{b}\left(\frac{n}{k}\right)$, define

$$\hat{\nu}(\cdot) = \frac{1}{k}\sum_{i=1}^{n} \epsilon_{\frac{Z_i}{\hat{b}(n/k)}} \tag{4.2}$$

$$\begin{aligned}\widehat{S}_1(\cdot) &= \frac{\hat{\nu}\{\boldsymbol{x} : \|\boldsymbol{x}\| > 1, \arctan\frac{x^{(2)}}{x^{(1)}} \in \cdot\}}{\hat{\nu}\{\boldsymbol{x} : \|\boldsymbol{x}\| > 1\}} \\ &= \frac{\sum_{i=1}^{n} 1_{[\|\boldsymbol{Z}_i\|/\hat{b}(n/k) > 1]}\epsilon_{\Theta_i}(\cdot)}{\sum_{i=1}^{n} 1_{[\|\boldsymbol{Z}_i\|/\hat{b}(n/k) > 1]}}\end{aligned} \tag{4.3}$$

where $\Theta_i = \arctan\big(Z_i^{(2)}/Z_i^{(1)}\big)$. Also define

$$\hat{v} = \int_0^{\pi/2} (\theta - \pi/4)^2 \hat{S}_1(d\theta)$$
$$= \frac{\sum_{i=1}^n 1_{[\|\boldsymbol{Z}_i\|/\hat{b}(n/k)>1]}(\Theta_i - \pi/4)^2}{\sum_{i=1}^n 1_{[\|\boldsymbol{Z}_i\|/\hat{b}(n/k)>1]}} \tag{4.4}$$

and

$$\hat{\rho} = 1 - \frac{\hat{v}}{(\pi/4)^2}. \tag{4.5}$$

Provided in (4.2) that

$$\hat{v} \Rightarrow v$$

in $M_+([0,\infty]^2\backslash\{\mathbf{0}\}$, as will certainly be the case if, as assumed, the $\boldsymbol{Z}$'s are iid,[30,46,49] we get by continuous mapping that all other quantities are also consistent:

$$\widehat{S}_1 \Rightarrow S_1, \quad \hat{v} \xrightarrow{P} v, \quad \hat{\rho} \xrightarrow{P} \rho.$$

5. ASYMPTOTIC NORMALITY OF THE EDM ESTIMATOR

Continue to suppose $d = 2$ and that the conditions of Theorem 1 hold for the iid random vectors $\{\boldsymbol{Z}_i, i \geq 1\}$. Suppose the polar coordinates of $\boldsymbol{Z}_i$ are the usual (R_i, Θ_i), where $\Theta_i \in [0, \pi/2]$. Assume α is known and does not have to be estimated; for simplicity assume $\alpha = 1$. Furthermore assume that $b(\cdot)$ is known, perhaps because the standard case assumption holds. It is most convenient to assume

$$tP[R_1 > b(t)] \to 1, \quad (t \to \infty). \tag{5.1}$$

Since we assume we know $b(\cdot)$, there is no need for an estimate and we consider the quantities in (4.2)–(4.5) with $b(\cdot)$ instead of $\hat{b}(\cdot)$. We continue to use the notation $\hat{v}$ and $\hat{\rho}$ in this section with the understanding that in (4.4) and (4.5) b replaces $\hat{b}$. Define

$$N_n = \sum_{i=1}^n 1_{[R_i > b(n/k)]} \tag{5.2}$$

as the random number of exceedances. Let $\{i(l,n), l \geq 1\}$ be the random indices such that $R_{i(l,n)} > b(n/k)$ so that (see page 212 of Ref.[46]) $\{\Theta_{i(l,n)}, l \geq 1\}$ are iid, independent of $\{N_n\}$ and

$$P[\Theta_{i(1,n)} \in \cdot] = P[\Theta_1 \in \cdot | R_1 > b(n/k)] = \frac{P[R_1 > b(n/k), \Theta_1 \in \cdot]}{P[R_1 > b(n/k)]} \to S_1(\cdot), \tag{5.3}$$

so that for an iid sequence $\{\Theta_l^{(\infty)}, l \geq 1\}$ with common distribution S_1 we have

$$\{\Theta_{i(l,n)}, l \geq 1\} \Rightarrow \{\Theta_l^{(\infty)}, l \geq 1\}, \quad n \to \infty.$$

This allows us to represent $\hat{v}$ as a random sum of iid random variables. From (4.4)

$$\hat{v} = \frac{1}{N_n} \sum_{l=1}^{N_n} (\Theta_{i(l,n)} - \pi/4)^2. \tag{5.4}$$

Now we add the assumption that the distribution of $\boldsymbol{Z}_1$ possesses asymptotic independence. From asymptotic independence we observe that, as $n \to \infty$,

$$E(\Theta_{i(1,n)} - \pi/4)^2 \to \int_0^{\pi/2} (\theta - \pi/4)^2 S_1(d\theta) = \left(\frac{\pi}{4}\right)^2,$$

and

$$\operatorname{Var}\big((\Theta_{i(1,n)} - \pi/4)^2\big) \to 0, \tag{5.5}$$

the last line following from the fact that $(\Theta_1^{(\infty)} - \pi/4)^2$ is almost surely constant with respect to the two-point distribution on $\{0, \pi/2\}$, the constant being $(\pi/2 - \pi/4)^2 = (\pi/4)^2$ or $(0 - \pi/4)^2 = (\pi/4)^2$.

Theorem 3. *Suppose $\{\boldsymbol{Z}_n, n \geq 1\}$ is iid with common distribution F and that $d = 2$ and Theorem* 1 *holds as well as* (5.1). *Assume also that asymptotic independence holds so that S_1 is a two-point distribution concentrating mass on $\{0, \pi/2\}$. Finally, suppose*

$$\operatorname{Var}\big(((\Theta_1 - \pi/4)^2) | R_1 > b(n/k)\big) \neq 0, \tag{5.6}$$

for $n \geq n_0$ for some n_0. Then as $n \to \infty$, $k = k(n) \to \infty$, $k/n \to 0$, we have

$$\frac{\sqrt{k}}{\sqrt{\operatorname{Var}((\Theta_{i(1,n)} - \pi/4)^2)}} \big(\hat{v} - E((\Theta_{i(1,n)} - \pi/4)^2)\big) \Rightarrow W(1), \quad n \to \infty, \tag{5.7}$$

where $\{W(t), t \geq 0\}$ is a standard Wiener process. Consequently,

$$\frac{\sqrt{k}}{\sqrt{\operatorname{Var}((\Theta_{i(1,n)} - \pi/4)^2)}} \left(\hat{\rho} - \left[1 - \frac{E(\Theta_{i(1,n)} - \pi/4)^2}{(\pi/4)^2}\right]\right) \Rightarrow \frac{-W(1)}{(\pi/4)^2} \stackrel{d}{=} \frac{W(1)}{(\pi/4)^2} \tag{5.8}$$

Remarks. (1) It is important to note that in the asymptotic independence case, the rate of convergence is not $\sqrt{k}$ but

$$\frac{\sqrt{k}}{\sqrt{\operatorname{Var}((\Theta_{i(1,n)} - \pi/4)^2)}}$$

where the denominator is converging to zero. Without asymptotic independence, the denominator would converge to a constant and the rate of convergence would be $\sqrt{k}$. This is an essential difference between the asymptotic independence case and cases without the asymptotic independence.

(2) If (5.6) fails then there exists a sequence n_p, $p \geq 1$ such that

$$P\left[\left|\Theta_{i(1,n_p)} - \frac{\pi}{4}\right| = \sqrt{c_{n_p}}\right] = 1,$$

where $c_n \to (\pi/4)^2$. This could not happen if, for instance, the underlying distribution F had a density.

Proof. From (5.2) and the Law of Large Numbers or the analogue of (2.5)

$$\frac{N_n}{k} \xrightarrow{P} 1. \tag{5.9}$$

From the functional central limit theorem for triangular arrays

$$\frac{1}{\sqrt{k}\sqrt{\mathrm{Var}((\Theta_{i(1,n)} - \pi/4)^2)}} \sum_{l=1}^{[kt]} \left((\Theta_{i(l,n)} - \pi/4)^2 - E((\Theta_{i(l,n)} - \pi/4)^2)\right) \Rightarrow W(t), \tag{5.10}$$

in $D[0, \infty)$. Because $\{\Theta_{i(l,n)}, l \geq 1\}$ are independent of $\{N_n\}$, the statements (5.9) and (5.10) can be combined into a joint convergence statement. Apply the almost surely continuous map $(f(\cdot), c) \mapsto f(c)$ from $D[0, \infty) \times [0, \infty) \mapsto [0, \infty)$ and we get

$$\frac{1}{\sqrt{k}\sqrt{\mathrm{Var}\left((\Theta_{i(1,n)} - \pi/4)^2\right)}} \sum_{l=1}^{N_n} \left((\Theta_{i(l,n)} - \pi/4)^2 - E((\Theta_{i(l,n)} - \pi/4)^2)\right) \Rightarrow W(1),$$

and since

$$N_n \hat{v} = \sum_{l=1}^{N_n} (\Theta_{i(l,n)} - \pi/4)^2$$

we conclude

$$\frac{N_n}{\sqrt{k}\sqrt{\mathrm{Var}((\Theta_{i(1,n)} - \pi/4)^2)}} \left(\hat{v} - E((\Theta_{i(l,n)} - \pi/4)^2)\right) \Rightarrow W(1).$$

Since $N_n/k \xrightarrow{P} 1$, the result follows. □

Of course, we would prefer the centering in (5.7) to be $v = (\pi/4)^2$ and the centering in (5.8) to be zero but then we would have one sided random variables converging to the two-sided normal limit. Any change of centering would require a second order regular variation condition which would be difficult to check in practice. (See Ref.[24] for background discussion.) Attempts to derive a second order condition in polar coordinates from a second order condition in Cartesian coordinates are discussed in Ref.[47]. We will not pursue this here.

Obviously, for statistical usage, we need, at least, an approximate solution to equations of the form

$$0.05 = P[\hat{\rho} > x_{0.05}].$$

Using (5.6) we have with the notation $P[W(1) > N^{\leftarrow}(0.95)] = 0.05$ that under the hypothesis of asymptotic independence

$$x_{0.05} \approx \frac{N^{\leftarrow}(0.95)}{(\pi/4)^2 \frac{\sqrt{k}}{\sqrt{\mathrm{Var}((\Theta_{i(1,n)}-\pi/4)^2)}}} + 1 - \frac{E((\Theta_{i(1,n)} - \pi/4)^2)}{(\pi/4)^2}. \tag{5.11}$$

One would like to be able to replace $E((\Theta_{i(1,n)} - \pi/4)^2)$ by the plug-in-estimator consisting of the sample average $(1/N_n)\sum_{l=1}^{N_n}(\Theta_{i(l,n)} - \pi/4)^2$ and this requires analysis of the difference between the two terms. Similarly, one would like to be able to replace $\mathrm{Var}((\Theta_{i(1,n)} - \pi/4)^2)$ by the plug-in-estimator consisting of $\widehat{\mathrm{Var}}$, the sample variance of $\{(\Theta_{i(l,n)} - \pi/4)^2, 1 \le l \le N_n\}$ and justifying this requires

$$\frac{\widehat{\mathrm{Var}}}{\mathrm{Var}((\Theta_{i(1,n)} - \pi/4)^2)} \xrightarrow{P} 1,$$

as $n \to \infty$.

6. EXAMPLES AND CONCLUDING REMARKS

Theorems 2 and 3 indicate that one ought to be able to treat a regularly varying sequence as a time series and test for asymptotic independence beyond some lag. We have experimented with doing this and the results of the experimentation are promising. The following points require further investigation and justification. In general, we do not know $b(n/k)$ and so we estimate $b(n/k)$ by the kth largest order statistic of the norms. This requires choice of k and experimentation with various values. In (5.11) the mean and variance are replaced by the sample versions and this needs justification. Finally, in the time series context, if we suspect asymptotic independence beyond lag L of the time series $\{X_j, 1 \le j \le n\}$, then we analyze pairs $\{(X_j, X_{j+L}), 1 \le j \le n - L\}$ and it is unlikely that for different values of j all pairs are independent as assumed in Theorem 3.

Undaunted by these difficulties, here are the results of three experiments. In addition to these, note that in Ref.[5] the EDM has been used to analyze UNC Internet data mentioned in the introduction.

6.1. Moving Averages

We constructed a moving average of order 6 using equal weights applied to 100,000 standard Pareto random variables with $\alpha = 1$. This is a case where beyond lag 6, variables are independent as well as asymptotically independent.

On the left of Fig. 3 is the EDM plot $(l, \mathrm{EDM}(X^{(0)}, X^{(l)}), 1 \le l \le 25)$ of MA as a function of lag using $k = 1000$. The plot certainly captures the asymptotic independence beyond lag 6. The horizontal line is given by (5.11). For comparison,

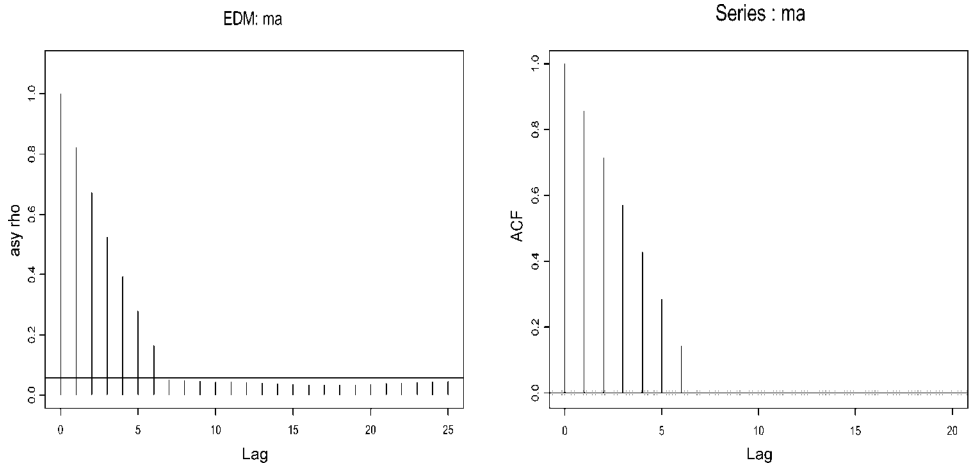

Figure 3. EDM plot of MA as a function of the lag (left) and acf plot of MA (right).

the sample correlation plot of MA is given on the right of Fig. 3. Moving average processes are one of the few classes of regularly varying processes for which the acf plot is informative. Even though the mathematical correlation does not exist, the sample correlation gives reliable indication of independence. See Refs.[13,14,43].

6.2. Mixtures

We next constructed a regularly varying sequence called *notindep* that was dependent but possessed asymptotic independence. We did this by taking 20,000 iid standard Pareto observations with $\alpha = 1$ and multiplying the whole collection by a single independent Pareto observation with $\alpha = 2$. By a generalization[37,45] of Breiman's Theorem,[3] this produces a regularly varying sequence whose marginal distribution has $\alpha = 1$ and it is easy to check that the sequence is asymptotically independent.

Figure 4 displays the EDM plot on the left for *notindep*. The small values of $\{\rho(l), 1 \le l \le 20\}$ show strong tendency towards asymptotic independence. The ACF-plot on the right is expected but note that this plot fails to capture the lack of independence since the ACF thinks the data is uncorrelated (even though correlations do not exist).

6.3. BU Data

The Boston University data is extensively documented and studied (see Refs.[7–10,28,35]) and is available at `www.acm.org/sigcomm/ITA/`, the Internet Traffic Archive (ITA) web site. This data used here is processed from the original 1995 Boston University data and consists of 4161 file sizes (F) and transmission rates (R) inferred from the time required for downloading a file and the file size. The data, thus, consists of bivariate pairs (F, R) and a scatter plot is shown in Fig. 5.

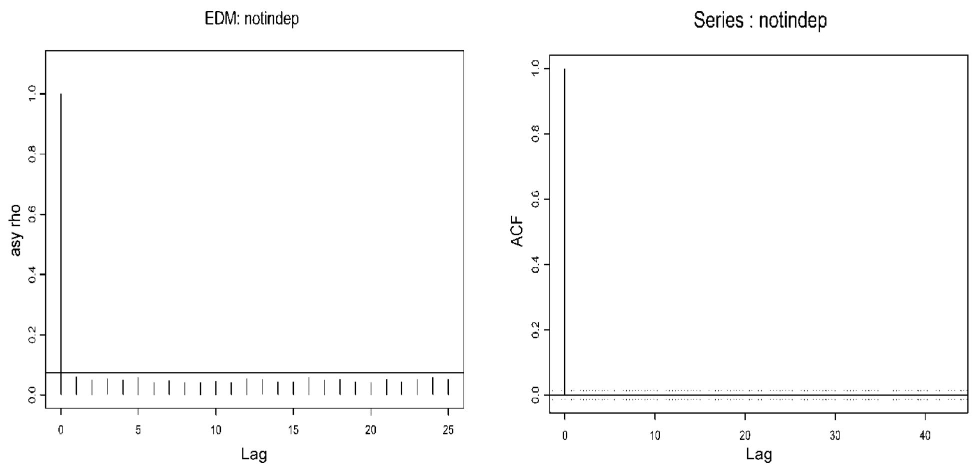

Figure 4. EDM plot of *notindep* as a function of the lag (left) and acf plot of *notindep* (right).

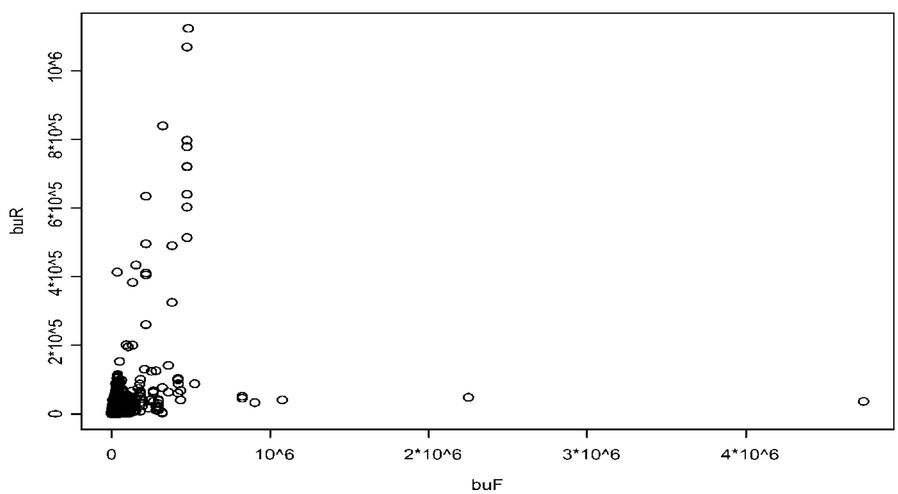

Figure 5. Scatter plot of buF vs. buR.

Each marginal distribution is heavy tailed and a combination of Hill and QQ plotting estimates the alphas as (1.157, 1.138) for F and R, respectively. Then we raise each component to its α-power to transform to the standard case where each marginal is regularly varying with $\alpha = 1$. We then compute $\hat{\rho}$ and the quantile $x_{0.05}$ in (5.11) and find with $k = 1000$ that

$$\hat{\rho} = 0.579 \quad \text{and} \quad x_{0.05} = 0.614.$$

and thus, despite the relatively large value of $\hat{\rho}$, this analysis presents no evidence at the 0.05 level against the hypothesis of asymptotic independence. Experimenting with a range of k-values produces the same conclusion.

ACKNOWLEDGMENTS

Krishanu Maulik made many helpful comments on an early draft of this paper which corrected some sloppy errors and led to several improvements. The research was partially supported by NSF grant DMS-0303493 and a grant from the Mathematical Sciences Program of NSA, grant MSPF-02G-183 at Cornell University.

REFERENCES

1. Basrak, B. The Sample Autocorrelation Function of Non-Linear Time Series. Ph.D. thesis, Rijksuniversiteit Groningen, Groningen, Netherlands, 2000.
2. Basrak, B.; Davis, R.A.; Mikosch, T. A characterization of multivariate regular variation. Ann. Appl. Probab. **2002**, *12* (3), 908–920.
3. Breiman, L. On some limit theorems similar to the arc-sin law. Theory Probab. Appl. **1965**, *10*, 323–331.
4. Brockwell, P.J.; Davis, R.A. *Time Series: Theory and Methods*, 2nd Ed.; Springer-Verlag: New York, 1991.
5. Campos, F.H.; Marron, J.S.; Resnick, S.; Jaffay, K. Extremal dependence: Internet traffic applications. Web available at http://www.orie.cornell.edu/~sid, 2002.
6. Coles, S.; Heffernan, J.; Tawn, J. Dependence measures for extreme value analyses. Extremes **1999**, *2* (4), 339–365.
7. Crovella, M.; Bestavros, A. Explaining world wide web traffic self-similarity. Preprint available as TR-95-015 from {crovella,best}@cs.bu.edu, 1995.
8. Crovella, M.; Bestavros, A. Self-similarity in world wide web traffic: Evidence and possible causes. In *Proceedings of the ACM SIGMETRICS '96 International Conference on Measurement and Modeling of Computer Systems*, 1996; Vol. 24, 160–169.
9. Crovella, M.; Bestavros, A.; Taqqu, M.S. Heavy-tailed probability distributions in the world wide web. In *A Practical Guide to Heavy Tails: Statistical Techniques for Analysing Heavy Tailed Distributions*; Taqqu, M.S., Adler, R., Feldman, R., Eds.; Birkhäuser: Boston, 1999.
10. Cunha, C.; Bestavros, A.; Crovella, M. Characteristics of www client-based traces. Preprint available as BU-CS-95-010 from {crovella,best}@cs.bu.edu, 1995.
11. Danielsson, J.; de Haan, L.; Peng, L.; de Vries, C.G. Using a bootstrap method to choose the sample fraction in tail index estimation. J. Multivariate Anal. **2001**, *76* (2), 226–248.
12. Davis, R.A.; Mikosch, T. The sample autocorrelations of heavy-tailed processes with applications to ARCH. Ann. Stat. **1998**, *26* (5), 2049–2080.
13. Davis, R.A.; Resnick, S.I. Limit theory for moving averages of random variables with regularly varying tail probabilities. Ann. Probab. **1985**, *13* (1), 179–195.
14. Davis, R.A.; Resnick, S.I. More limit theory for the sample correlation function of moving averages. Stochastic Process. Appl. **1985**, *20*, 257–279.

15. Davis, R.A.; Resnick, S.I. Prediction of stationary max-stable processes. Ann. Appl. Probab. **1993**, *3* (2), 497–525.
16. de Haan, L. A spectral representation for max-stable processes. Ann. Probab. **1984**, *12*, 1194–1204.
17. de Haan, L.; de Ronde, J. Sea and wind: Multivariate extremes at work. Extremes **1998**, *1* (1), 7–46.
18. de Haan, L.; Omey, E. Integrals and derivatives of regularly varying functions in $\mathbb{R}^d$ and domains of attraction of stable distributions II. Stochastic Process. Appl. **1984**, *16* (2), 157–170.
19. de Haan, L.; Omey, E.; Resnick, S. Domains of attraction and regular variation in $\mathbb{R}^d$. J. Multivariate Anal. **1984**, *14* (1), 17–33.
20. de Haan, L.; Pickands, J. III. Stationary min-stable stochastic processes. Probab. Theory Relat. Fields **1986**, *72* (4), 477–492.
21. de Haan, L.; Resnick, S.I. Limit theory for multivariate sample extremes. Z. Wahrscheinlichkeitstheorie **1977**, *40*, 317–337.
22. de Haan, L.; Resnick, S.I. Derivatives of regularly varying functions in $\mathbb{R}^d$ and domains of attraction of stable distributions. Stochastic Process. Appl. **1978/1979**, *8* (3), 349–355.
23. de Haan, L.; Resnick, S.I. On regular variation of probability densities. Stochastic Process. Appl. **1987**, *25*, 83–95.
24. de Haan, L.; Resnick, S.I. On asymptotic normality of the Hill estimator. Stochastic Models **1998**, *14*, 849–867.
25. Dekkers, A.L.M.; de Haan, L. On the estimation of the extreme-value index and large quantile estimation. Ann. Stat. **1989**, *17*, 1795–1832.
26. Dekkers, A.L.M.; de Haan, L. Optimal choice of sample fraction in extreme-value estimation. J. Multivariate Anal. **1993**, *47* (2), 173–195.
27. Einmahl, J.; de Haan, L.; Piterbarg, V. Nonparametric estimation of the spectral measure of an extreme value distribution. Ann. Stat. **2001**, *29* (5), 1401–1423.
28. Guerin, C.A.; Nyberg, H.; Perrin, O.; Resnick, S.; Rootzen, H.; Stărică, C. Empirical testing of the infinite source poisson data traffic model. Stochastic Models **2003**, *19* (2), 151–200.
29. Heffernan, J.E. A directory of coefficients of tail dependence. Extremes **2000**, *3* (3), 279–290 **(2001)**.
30. Huang, X. Statistics of Bivariate Extreme Values. Ph.D. thesis, Tinbergen Institute Research Series 22, Erasmus University Rotterdam, Postbus 1735, 3000DR, Rotterdam, The Netherlands, 1992.
31. Ledford, A.W.; Tawn, J.A. Statistics for near independence in multivariate extreme values. Biometrika **1996**, *83* (1), 169–187.
32. Ledford, A.W.; Tawn, J.A. Modelling dependence within joint tail regions. J. Roy. Stat. Soc. Ser. B **1997**, *59* (2), 475–499.
33. Ledford, A.W.; Tawn, J.A. Concomitant tail behaviour for extremes. Adv. Appl. Probab. **1998**, *30* (1), 197–215.
34. Ledford, A.W.; Tawn, J.A. Diagnostics for dependence within time series extremes. J. R. Stat. Soc. B **2003**, *65* (2), 521–543.
35. Maulik, K.; Resnick, S.I.; Rootzén, H. Asymptotic independence and a network traffic model. J. Appl. Probab. **2002**, *39* (4), 671–699.
36. Meerschaert, M.; Scheffler, H.P. *Limit Distributions for Sums of Independent Random Vectors*; John Wiley & Sons Inc.: New York, 2001.

37. Mikosch, T. Modeling dependence and tails of financial time series. In *SemStat: Seminaire Europeen de Statistique, Exteme Values in Finance, Telecommunications, and the Environment*; Finkenstadt, B., Rootzen, H., Eds.; Chapman-Hall: London, 2003, 185–286.
38. Mikosch, T.; Stărică, C. Long range dependence effects and ARCH modeling. In *Theory and applications of long-range dependence*; Doukhan, P., Oppenheim, G., Taqqu, M.S., Eds.; Birkhäuser Boston Inc.: Boston, MA, 2003, pages xii+719.
39. Poon, S.-H.; Rockinger, M.; Tawn, J. New extreme-value dependence measures and finance applications. University of Strathclyde, 100 Cathedral Street, Glasgow G40LN, UK, February 2001.
40. Reiss, R.-D.; Thomas, M. *Statistical Analysis of Extreme Values*, 2nd Ed.; Birkhäuser Verlag: Basel, 2001.
41. Resnick, S. Modeling data networks. In *SemStat: Seminaire Europeen de Statistique, Exteme Values in Finance, Telecommunications, and the Environment*; Finkenstadt, B., Rootzen, H., Eds.; Chapman-Hall: London, 2003, 287–372.
42. Resnick, S.; Greenwood, P. A bivariate stable characterization and domains of attraction. J. Multivariate Anal. **1979**, *9*, 206–221.
43. Resnick, S.; Samorodnitsky, G.; Xue, F. How misleading can sample acf's of stable ma's be? (Very!). Ann. Appl. Probab. **1999**, *9* (3), 797–817.
44. Resnick, S.I. Tail equivalence and its applications. J. Appl. Probab. **1971**, *8*, 136–156.
45. Resnick, S.I. Point processes, regular variation and weak convergence. Adv. Appl. Probab. **1986**, *18*, 66–138.
46. Resnick, S.I. *Extreme Values, Regular Variation and Point Processes*; Springer-Verlag: New York, 1987.
47. Resnick, S.I. Hidden regular variation, second order regular variation and asymptotic independence. Extremes **2002**, *5* (4), 303–336.
48. Resnick, S.I. On the foundation of multivariate heavy tail analysis. Submitted for C.C. Heyde festschrift; Web available at http://www.orie.cornell.edu/~sid, 2002.
49. Resnick, S.I.; Stărică, C. Consistency of Hill's estimator for dependent data. J. Appl. Probab. **1995**, *32* (1), 139–167.
50. Resnick, S.I.; Stărică, C. Smoothing the moment estimator of the extreme value parameter. Extremes **1998**, *1* (3).
51. Resnick, S.I.; Stărică, C. Smoothing the Hill estimator. Adv. Appl. Probab. **1997**, *29*, 271–293.
52. Resnick, S.I.; Roy, R. Superextremal processes, max-stability and dynamic continuous choice. Ann. Appl. Probab. **1994**, *4* (3), 791–811.
53. Rosiński, J. On the structure of stationary stable processes. Ann. Probab. **1995**, *23*, 1163–1187.
54. Samorodnitsky, G.; Taqqu, M. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance. Stochastic Modeling*; Chapman & Hall: New York, NY, 1994, xviii, 632.
55. Schlather, M. Examples for the coefficient of tail dependence and the domain of attraction of a bivariate extreme value distribution. Stat. Probab. Lett. **2001**, *53* (3), 325–329.

56. Ashoke Kumar Sinha. Estimating Failure Probability when Failure is Rare: Multidimensional Case. Ph.D. thesis, Tinbergen Institute, Erasmus University Rotterdam, 1997. Ph.D. Thesis # 165, Promotor: Prof. Dr. L.F.M. de Haan, published by Thela Thesis, Amsterdam.
57. Stărică, C. Multivariate extremes for models with constant conditional correlations. J. Empirical Finance **1999**, *6*, 515–553.
58. Stărică, C. Multivariate extremes for models with constant conditional correlations. In *Extremes and Integrated Risk Management*; Embrechts, P., Ed.; Risk Books: London, 2000.
59. Tawn, J. Bivariate extreme value theory: Models and estimation. Biometrica **1988**, *75*, 397–415.
60. Vervaat, W. Random upper semicontinuous functions and extremal processes. In *Probability and Lattices*; Centrum voor Wiskunde en Informatica (CWI). CWI Tracts: Amsterdam, 1997, Vol. 110, 1–56.
61. Zhang, Y.; Breslau, L.; Paxson, V.; Shenker, S. On the characteristics and origins of internet flow rates, ACM Sigcom 2002 Conference, Pittsburgh, Pa; August 19–23, 2002.

Received September 9, 2002
Accepted October 8, 2003