HIDDEN REGULAR VARIATION
AND THE RANK TRANSFORM

JANET HEFFERNAN,* *Lancaster University*

SIDNEY RESNICK,** *Cornell University*

Abstract

Random vectors in the positive orthant whose distributions possess *hidden regular variation* are a subclass of those whose distributions are multivariate regularly varying with asymptotic independence. The concept is an elaboration of the *coefficient of tail dependence* of Ledford and Tawn. We show that the rank transform that brings unequal marginals to the standard case also preserves the hidden regular variation. We discuss applications of the results to two examples, one involving flood risk and the other Internet data.

Keywords: Heavy tail; regular variation; Pareto tail; coefficient of tail dependence; hidden regular variation; rank transform; asymptotic independence; Internet traffic; flood risk

2000 Mathematics Subject Classification: Primary 60G70
                            Secondary 62H05; 62H20; 60E05

1. Introduction

A refinement of the class of multivariate regularly varying distributions, which we call those with *hidden regular variation*, is a semiparametric subfamily of the full family of distributions possessing multivariate regular variation and asymptotic independence. Various cases of hidden regular variation have recently received considerable attention as part of the program to distinguish statistically asymptotic independence from dependence; see Campos *et al.* (2005), Coles *et al.* (1999), de Haan and de Ronde (1998), Draisma *et al.* (2004), Heffernan (2000), Ledford and Tawn (1996), (1997), Maulik and Resnick (2005), Peng (1999), Poon *et al.* (2003), Resnick (2004), and Stărică (1999). In particular, hidden regular variation is based on Ledford and Tawn's (1996), (1997) analysis of the coefficient of tail dependence.

Treatments of hidden regular variation, placing it in relation to the concepts of asymptotic independence and second-order regular variation, and giving characterizations and examples, were given in Resnick (2002) and Maulik and Resnick (2005). Here we discuss how the rank transformation yielding standard-form regular variation preserves the hidden regular variation.

1.1. Outline

In the rest of this section, we review notation (Section 1.2) and the polar coordinate transformation (Section 1.3). In Section 2, we define multivariate regular variation and hidden regular variation for heavy-tailed vectors without distributionally equal components. We review the rank method for estimating the spectral measure in Section 3, and show that the rank transform

can also be used to estimate the hidden spectral measure. In Section 4, we give some remarks on the estimation of the hidden spectral measure since the limit result of Section 3 involves a function that, typically, will not be known. In Sections 4.3 and 5, we also give a high-level view of several problems in which the existence of hidden regular variation allows for improved accuracy when estimating probabilities of extreme events. Two examples are given in Section 6, one requiring estimation of the risk of flooding and a second in which the problem is to ascertain whether the response size and transfer rate of Internet traffic exhibit asymptotic independence.

1.2. Notation

For simplicity, we will generally assume that random vectors have nonnegative components. Set

$$\mathbb{E} := [0, \infty]^d \setminus \{\mathbf{0}\}$$

so that the origin is excluded from $\mathbb{E}$. Compact subsets of $\mathbb{E}$ are compact sets of $[0, \infty]^d$ that do not intersect the origin; see the discussion in Resnick (2002). In some applications, for instance in finance, it is natural to consider the cone $[-\infty, \infty]^d \setminus \{\mathbf{0}\}$. We leave it to the reader to make the modest changes necessary to generalize to this case by considering the orthants individually.

Vectors are denoted by bold letters; capitals for random vectors and lower-case letters for nonrandom vectors. For example: $\boldsymbol{x} = (x^{(1)}, \ldots, x^{(d)}) \in \mathbb{R}^d$. Operations between vectors should be interpreted componentwise so that, for two vectors $\boldsymbol{x}$ and $\boldsymbol{z}$,

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

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

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

and so on. Also define $\mathbf{0} = (0, \ldots, 0)$, the d-dimensional zero vector. For a real number c, we write $c\boldsymbol{x} = (cx^{(1)}, \ldots, cx^{(d)})$, as usual. We denote rectangles (or higher-dimensional intervals) by

$$[\boldsymbol{a}, \boldsymbol{b}] = \{\boldsymbol{x} \in \mathbb{E} : \boldsymbol{a} \leq \boldsymbol{x} \leq \boldsymbol{b}\}.$$

Higher-dimensional rectangles with one or both endpoints open are defined analogously; for example

$$(\boldsymbol{a}, \boldsymbol{b}] = \{\boldsymbol{x} \in \mathbb{E} : \boldsymbol{a} < \boldsymbol{x} \leq \boldsymbol{b}\}.$$

Complements are taken with respect to $\mathbb{E}$, so that, for $\boldsymbol{x} > \mathbf{0}$,

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

For $i = 1, \ldots, d$, we define the basis vectors $\boldsymbol{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ so that the axes originating at $\mathbf{0}$ are given by $\mathbb{L}_i := \{t\boldsymbol{e}_i, t > 0\}$, $i = 1, \ldots, d$. We then define the cone

$$\mathbb{E}_0 = \mathbb{E} \setminus \bigcup_{i=1}^{d} \mathbb{L}_i = \{\boldsymbol{s} \in \mathbb{E} : \text{ for some } 1 \leq i < j \leq d, s^{(i)} \wedge s^{(j)} > 0\}.$$

If $d = 2$, we have $\mathbb{E}_0 = (0, \infty]^2$. The cone $\mathbb{E}_0$ consists of those points of $\mathbb{E}$ such that at most $d - 2$ coordinates are 0.

1.3. The polar coordinate transformation

It is sometimes illuminating to consider multivariate regular variation for the distribution of a random vector after a polar coordinate transformation. Suppose that $\| \cdot \| \colon \mathbb{R}^d \to [0, \infty)$ 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 $p > 0$, and the L_∞-norm: $\|\boldsymbol{x}\| = \bigvee_{i=1}^n |x^{(i)}|$. Assume that the norm has been scaled so that $\|\boldsymbol{e}_i\| = 1$ for $i = 1, \ldots, d$. Given a chosen norm $\| \cdot \|$, the points at unit distance from the origin $\boldsymbol{0}$ are

$$\aleph := \{\boldsymbol{x} \in \mathbb{E} \colon \|\boldsymbol{x}\| = 1\}.$$

For the purpose of hidden regular variation, we need to look at the subcone $\mathbb{E}_0$ of $\mathbb{E}$ and the restriction of $\aleph$ to $\mathbb{E}_0$, denoted by $\aleph_0 = \aleph \cap \mathbb{E}_0$. Recall that norms on $\mathbb{R}^d$ are all topologically equivalent, in that convergence in one norm implies convergence in another.

For a fixed norm, define the polar coordinate transformation $T \colon [0, \infty)^d \setminus \{\boldsymbol{0}\} \to (0, \infty) \times \aleph$ by

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

and the inverse transformation $T^{\leftarrow} \colon (0, \infty) \times \aleph \to [0, \infty)^d \setminus \{\boldsymbol{0}\}$ by

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

We think of $\boldsymbol{a} \in \aleph$ as defining a direction and r as telling us how far in direction $\boldsymbol{a}$ to proceed. Since we exclude $\boldsymbol{0}$ from the domain of T, both T and $T^{\leftarrow}$ are continuous bijections.

When $d = 2$, it is customary, but not obligatory, to write $T(\boldsymbol{x}) = (r, \theta)$, where $\boldsymbol{x} = (r \cos\theta, r \sin\theta)$, with $0 \leq \theta \leq \frac{1}{2}\pi$, rather than the more consistent notation

$$T(\boldsymbol{x}) = (r, (\cos\theta, \sin\theta)).$$

For a random vector $\boldsymbol{X}$ in $\mathbb{R}^d$, we sometimes write $T(\boldsymbol{X}) = (R_X, \boldsymbol{\Theta}_X)$. When there is no risk of confusion, we suppress the subscript.

2. Multivariate regular variation and hidden regular variation

We start with the definition of regular variation with unequal components.

2.1. Regular variation on $\mathbb{E}$

Suppose that $\boldsymbol{Z}$ is a d-dimensional random vector in $[0, \infty)^d$. The distribution of $\boldsymbol{Z}$ is regularly varying (with unequal components) if there exist functions $b^{(j)}(t) \uparrow \infty$, as $t \to \infty$, such that, for a Radon measure ν on $\mathbb{E}$, we have the vague convergence

$$t\, \mathrm{P}\left(\left(\frac{Z^{(j)}}{b^{(j)}(t)}, j = 1, \ldots, d \right) \in \cdot \right) = t\, \mathrm{P}\left(\frac{\boldsymbol{Z}}{\boldsymbol{b}(t)} \in \cdot \right) \xrightarrow{v} \nu \tag{1}$$

on $\mathbb{E}$ (cf. de Haan and Omey (1984) and Greenwood and Resnick (1979)). We assume that the marginal convergences satisfy

$$t\, \mathrm{P}\left(\frac{Z^{(j)}}{b^{(j)}(t)} > x \right) \to \nu_{\alpha^{(j)}}(x, \infty] := x^{-\alpha^{(j)}},$$

where $\alpha^{(j)} > 0$, $j = 1, \ldots, d$. Then $b^{(j)}(t) \in RV_{1/\alpha^{(j)}}$, the class of regularly varying functions of index $1/\alpha^{(j)}$, and we can and do assume that each $b^{(j)}(t)$ is both continuous and strictly increasing, and denote the inverse function by $b^{(j)\leftarrow}(t)$.

2.2. Standard form of regular variation

By a change of variables, regular variation with unequal components can be converted into regular variation for which the marginal distributions of each $Z^{(j)}$ are tail equivalent. The relation (1) is equivalent to (cf., for example, Resnick (1987, p. 277))

$$t\, \mathrm{P}\!\left(\left(\frac{b^{(j)\leftarrow}(Z^{(j)})}{t}, j = 1, \dots, d\right) \in \cdot\right) \xrightarrow{v} v_{\text{standard}}(\cdot) \tag{2}$$

on $\mathbb{E}$, where v_{standard} satisfies the homogeneity condition

$$v_{\text{standard}}(t\cdot) = t^{-1} v_{\text{standard}}(\cdot) \tag{3}$$

on $\mathbb{E}$. Relation (2) is the *standard* form of regular variation, where each component of the random vector is normalized by the same function. The measures v_{standard} and v are related by

$$v([\mathbf{0}, \mathbf{x}]^c) = v_{\text{standard}}([\mathbf{0}, \mathbf{x}^{\boldsymbol{\alpha}}]^c), \qquad \mathbf{x} \in \mathbb{E}, \tag{4}$$

where we extend our vector conventions such that

$$\mathbf{x}^{\boldsymbol{\alpha}} = ((x^{(1)})^{\alpha^{(1)}}, \dots, (x^{(d)})^{\alpha^{(d)}}).$$

After transforming to polar coordinates, the homogeneity property (3) becomes

$$v_{\text{standard}}\left(\left\{\mathbf{x} \in \mathbb{E} \colon \|\mathbf{x}\| > r, \frac{\mathbf{x}}{\|\mathbf{x}\|} \in \cdot\right\}\right) = c r^{-1} S(\cdot),$$

where $c > 0$ and S is a probability measure on Borel subsets of $\aleph$.

2.3. Hidden regular variation

We say that the distribution of $\mathbf{Z}$ has hidden regular variation if, in addition to (1) or (2), we have the following property after transforming to the standard case. There must exist a function $b_0(t) \in RV_{1/\alpha_0}$ with $b_0(t) \uparrow \infty$, $\alpha_0 \geq 1$, and

$$\lim_{t \to \infty} \frac{t}{b_0(t)} = \infty, \tag{5}$$

such that, on $\mathbb{E}_0$,

$$t\, \mathrm{P}\!\left(\left(\frac{b^{(j)\leftarrow}(Z^{(j)})}{b_0(t)}, j = 1, \dots, d\right) \in \cdot\right) \xrightarrow{v} v_0 \tag{6}$$

for some Radon measure v_0 on $\mathbb{E}_0$. Note that (6) is equivalent to

$$t\, \mathrm{P}\!\left(\left(\frac{Z^{(j)}}{b^{(j)}(b_0(t))}, j = 1, \dots, d\right) \in \cdot\right) \xrightarrow{v} \tilde{v}_0 \tag{7}$$

on $\mathbb{E}_0$, where v_0 and $\tilde{v}_0$ are related by

$$\tilde{v}_0((\mathbf{x}, \boldsymbol{\infty}]) = v_0((\mathbf{x}^{\boldsymbol{\alpha}}, \boldsymbol{\infty}]), \qquad \mathbf{x} \in \mathbb{E}_0. \tag{8}$$

The measure v_0 is also homogeneous on $\mathbb{E}_0$, i.e.

$$v_0(t\cdot) = t^{-\alpha_0} v_0(\cdot),$$

but ν_0 can be either finite or infinite on E_0 (Maulik and Resnick (2005), Resnick (2002)) and, when we transform ν_0 to polar coordinates, we obtain

$$\nu_0\left(\left\{\boldsymbol{x} \in \mathbb{E}_0 : \|\boldsymbol{x}\| > r, \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|} \in \cdot\right\}\right) = r^{-\alpha_0} S_0(\cdot),$$

where S_0 is a Radon measure on Borel subsets of

$$\aleph_0 = \aleph \cap \mathbb{E}_0.$$

Since the region

$$\aleph_{\text{inv}} := \left\{\boldsymbol{x} \in \mathbb{E}_0 : \bigwedge_{j=1}^{d} x^{(j)} \geq 1\right\}$$

is a compact subset of $\mathbb{E}_0$ and, hence, will always have finite hidden measure, we can (and do) always choose $b_0(t)$ so that

$$\nu_0(\aleph_{\text{inv}}) = 1.$$

Recall that the presence of hidden regular variation implies that the vector $\boldsymbol{Z}$ possesses *asymptotic independence* (Resnick (2002)), which, in our case, is equivalent to

$$\nu_{\text{standard}}(\mathbb{E}_0) = 0.$$

This means that the probability of two components of the vector being simultaneously large is negligible compared to the probability of only one component being large. The motivation (Ledford and Tawn (1996), (1997)) behind the concept of hidden regular variation is to create a tractable subclass of the distributions possessing asymptotic independence, which would allow for statistical analysis.

2.4. Comparison with Ledford and Tawn

Ledford and Tawn (1996) proposed the following model for the asymptotic behaviour of the joint survivor function of standardized variables $(Z^{(1)}, Z^{(2)})$:

$$P(Z^{(1)} > r, Z^{(2)} > r) \sim \mathcal{L}(r) r^{-1/\eta} \quad \text{as } r \to \infty.$$

This holds for a wide range of bivariate distributions with standard Pareto margins. Here, $\mathcal{L}$ is a slowly varying function, i.e. $\mathcal{L}(tr)/\mathcal{L}(r) \to 1$ as $r \to \infty$ for all fixed $t > 0$. The parameter $\eta \in [0, 1]$ was termed the *coefficient of tail dependence* by Ledford and Tawn, as it controls the rate of decay of $P(Z^{(1)} > r, Z^{(2)} > r)$ as $r \to \infty$. If we have positively associated variables, then $\frac{1}{2} \leq \eta \leq 1$; when $Z^{(1)}$ and $Z^{(2)}$ are independent, $\eta = \frac{1}{2}$; and if $0 < \eta < \frac{1}{2}$ then the variables are negatively dependent. The variables are termed asymptotically dependent if $\eta = 1$ and $\mathcal{L}(r) \not\to 0$ as $r \to \infty$, and asymptotically independent otherwise.

In a later paper, Ledford and Tawn (1997) refined this joint tail assumption in order to model behaviour off the diagonal. They assumed, for z_1 and z_2 simultaneously large, that

$$P(Z^{(1)} > z_1, Z^{(2)} > z_2) \sim \mathcal{L}_1(z_1, z_2) z_1^{-c_1} z_2^{-c_2} + \mathcal{L}_2(z_1, z_2) z_1^{-(c_1+d_1)} z_2^{-(c_2+d_2)} + \cdots, \quad (9)$$

where $c_1 + c_2 = 1/\eta$, $d_j \geq 0$, $\mathcal{L}_j(z_1, z_2)$ are bivariate slowly varying functions with limit functions g_j, and $\mathcal{L}_2(tz_1, tz_2) = o(\mathcal{L}_1(tz_1, tz_2))$ as $t \to \infty$ if $d_1 = d_2 = 0$. Since

$g_j(cz_1, cz_2) = g_j(z_1, z_2)$ for all $c > 0$ and $(z_1, z_2) \in \mathbb{R}_+^2$, Ledford and Tawn defined the *ray dependence function*

$$g_*^{(j)}(w) := g_*^{(j)}\left(\frac{z_1}{z_1 + z_2}\right) \equiv g_j(z_1, z_2).$$

(Note that the map

$$z \mapsto \left(\frac{z_1}{z_1 + z_2}, \frac{z_2}{z_1 + z_2}\right)$$

corresponds to our polar coordinate transform using the L_1-norm.) This function provides an alternative diagnostic to S_0 for the strength of the dependence between $Z^{(1)}$ and $Z^{(2)}$ within the class of distributions with a given value of η. Ledford and Tawn (1997) derived the connections between their ray dependence function $g_*(w) := g_*^{(1)}$ and the hidden spectral measure S_0. For $(Z^{(1)}, Z^{(2)})$ as above, with $g_*(w)$ differentiable for all $w \in (0, 1)$ and a sequence $b_0(n)$ satisfying $n\mathcal{L}_1(b_0(n), b_0(n)) \sim b_0(n)^{1/\eta}$, Ledford and Tawn defined the inhomogeneous point process

$$\mathcal{P}_n^{(1)} = \{(Z_i^{(1)}/b_0(n), Z_i^{(2)}/b_0(n)), i = 1, \ldots, n\},$$

where $(Z_i^{(1)}, Z_i^{(2)})$, $1 \le i \le n$, are independent, identically distributed (i.i.d.) copies of $(Z^{(1)}, Z^{(2)})$. Theorem 1 of Ledford and Tawn (1997) states that $\mathcal{P}_n^{(1)} \to \mathcal{P}^{(1)}$ weakly as $n \to \infty$ on $\mathbb{E}_0$, where $\mathcal{P}^{(1)}$ is an inhomogeneous Poisson process with point intensity

$$\lambda(r, w) = \lambda_0(w)r^{-(1+1/\eta)}$$

in polar coordinates, corresponding to radial and angular components

$$R = (Z^{(1)} + Z^{(2)})/b_0(n) \quad \text{and} \quad W = Z^{(1)}/(Z^{(1)} + Z^{(2)}),$$

where

$$\lambda_0(w) = \frac{c_1 c_2 g_*(w) + w(1 - w)g_*'(w)(2w - 1 + c_1 - c_2) - g_*''(w)w^2(1 - w)^2}{w^{1+c_1}(1 - w)^{1+c_2}}. \quad (10)$$

Thus stated, λ_0 is the point intensity associated with our spectral measure S_0, and our parameter α_0 corresponds to $1/\eta$. The Ledford and Tawn Poisson convergence would be a consequence of Proposition 3.21 of Resnick (1987, p. 154), if (9) were interpreted as vague convergence of measures.

A recent consideration of the coefficient of tail dependence in the context of elliptical distributions was given by Hult and Lindskog (2002).

3. Rank methods for estimating the spectral and hidden spectral measures

We now review and extend a method, based on ranks and discussed by Huang (1992), for estimating the spectral measure S. This rank method overcomes the statistical problems inherent in multivariate methods that first require one to estimate tail indices $\alpha^{(j)}$, $j = 1, \ldots, d$. Assume that we wish to estimate the spectral measure S and the hidden spectral measure S_0 on the basis of a sample $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ of size n.

3.1. Review of diagnostics, based on ranks, for estimating the spectral measure *S*

We continue to suppose that we have multivariate regular variation (without tail-equivalent marginal distributions) given by (1). It follows from Resnick (1986) (see also Resnick (1987, Exercise 3.5.7, p. 161)) that the empirical measures converge, i.e.

$$v_n(\cdot) := \frac{1}{k} \sum_{i=1}^{n} \varepsilon_{(Z_i^{(1)}/b^{(1)}(n/k),\ldots,Z_i^{(d)}/b^{(d)}(n/k))}(\cdot) \xrightarrow{P} v(\cdot) \tag{11}$$

in $M_+(\mathbb{E})$, the space of Radon measures on $\mathbb{E}$. Here, $k \equiv k(n)$ is a function of n satisfying $k(n) \to \infty$ but $k/n \to 0$, so that k/n is a vanishing proportion of the sample size, and '$\xrightarrow{P}$' denotes convergence in probability. Furthermore, ε_x is the probability measure that puts all mass at x, i.e.

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

For the sample $(Z_1^{(j)}, \ldots, Z_n^{(j)})$ of jth components, let

$$Z_{(1)}^{(j)} \geq \cdots \geq Z_{(n)}^{(j)}$$

be the order statistics in decreasing order, starting with the largest. Taking the marginal convergences in (11) and inverting yields, for each $j = 1, \ldots, d$,

$$\frac{Z_{(\lceil kt^{(j)} \rceil)}^{(j)}}{b^{(j)}(n/k)} \xrightarrow{P} (t^{(j)})^{-1/\alpha^{(j)}}$$

in $D(0, \infty]$, the space of right-continuous functions on $(0, \infty]$ with finite left-hand limits. By $\lceil \cdot \rceil$ we denote the integer-part function. Since convergence is to a constant limit, we may append this to (11) to obtain (Billingsley (1968, p. 27))

$$\left(v_n(\cdot), \left(\frac{Z_{(\lceil kt^{(j)} \rceil)}^{(j)}}{b^{(j)}(n/k)}, j = 1, \ldots, d \right) \right) \Rightarrow (v, ((t^{(j)})^{-1/\alpha^{(j)}}; j = 1, \ldots, d)) \tag{12}$$

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

Recalling (4), we use continuous mapping (Billingsley (1968, p. 30)) to convert (12) into

$$\left(v_n([\mathbf{0}, \mathbf{x}]^c), \left(\frac{Z_{(\lceil kt^{(j)} \rceil)}^{(j)}}{b^{(j)}(n/k)}, j = 1, \ldots, d \right) \right) \Rightarrow (v([\mathbf{0}, \mathbf{x}]^c), ((t^{(j)})^{-1/\alpha^{(j)}}, j = 1, \ldots, d)), \tag{13}$$

and then apply the almost surely continuous map

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

to (13), to obtain

$$v_n\left(\left[\mathbf{0}, \left(\frac{Z_{(\lceil kt^{(j)} \rceil)}^{(j)}}{b^{(j)}(n/k)}, j = 1, \ldots, d \right) \right]^c \right) \Rightarrow v([\mathbf{0}, \mathbf{t}^{-1/\alpha}]^c). \tag{14}$$

Unpacking the left-hand side of (14), we have

$$
\nu_n\left(\left[\mathbf{0}, \left(\frac{Z^{(j)}_{(\lceil kt^{(j)}\rceil)}}{b^{(j)}(n/k)}, j = 1, \ldots, d\right)\right]^c\right) = \frac{1}{k}\sum_{i=1}^{n}\mathbf{1}_{\{Z_i^{(j)}/b^{(j)}(n/k) \leq Z^{(j)}_{(\lceil kt^{(j)}\rceil)}/b^{(j)}(n/k), j=1,\ldots,d\}^c}
$$

$$
= \frac{1}{k}\sum_{i=1}^{n}\mathbf{1}_{\{Z_i^{(j)} \leq Z^{(j)}_{(\lceil kt^{(j)}\rceil)}, j=1,\ldots,d\}^c}. \tag{15}
$$

For $j = 1, \ldots, d$, define the antiranks

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

to be the number of jth components greater than or equal to $Z_i^{(j)}$. We can rephrase the final expression in (15) as

$$
\frac{1}{k}\sum_{i=1}^{n}\mathbf{1}_{\{r_i^{(j)} \geq kt^{(j)}, j=1,\ldots,d\}^c}
$$

and then make the change of variable $s \mapsto t^{-1}$ to obtain

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

or

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

or

$$
\hat{\nu}_{\text{standard},n} =: \frac{1}{k}\sum_{i=1}^{n}\varepsilon_{(k/r_i^{(j)}, j=1,\ldots,d)} \Rightarrow \nu_{\text{standard}} \tag{16}
$$

in $M_+(\mathbb{E})$, where we have used (8). A polar coordinate transformation of the points

$$
\left\{\left(\frac{k}{r_i^{(j)}}, j = 1, \ldots, d\right), i = 1, \ldots, n\right\}
$$

allows us to estimate the spectral measure S. Suppose that the polar coordinates of $(1/r_i^{(j)}, j = 1, \ldots, d)$ are (R_i, Θ_i). Then the empirical measure of Θs corresponding to exceedances in the R variable, that is

$$
\frac{\sum_{i=1}^{n}\mathbf{1}_{\{kR_i \geq 1\}}\varepsilon_{\Theta_i}}{\sum_{i=1}^{n}\mathbf{1}_{\{kR_i \geq 1\}}} = \frac{\sum_{i=1}^{n}\mathbf{1}_{\{R_i \geq k^{-1}\}}\varepsilon_{\Theta_i}}{\sum_{i=1}^{n}\mathbf{1}_{\{R_i \geq k^{-1}\}}} \Rightarrow S, \tag{17}
$$

is a consistent estimator of S.

3.2. Finding the hidden spectral measure with the rank transform

Now suppose that both regular variation, represented by (1), and hidden regular variation, represented by (6), hold. How can we use the rank transform to estimate the hidden measure S_0?

To find the hidden angular measure, we expect to have to use points

$$\left\{ \left(\frac{k}{r_i^{(j)}}, j = 1, \ldots, d \right), i = 1, \ldots, n \right\}$$

with a threshold at a lower level than in (17). Since $b_0(t)/t \to 0$ (see (5)), it seems plausible to use the points

$$\left\{ \left(\frac{k/r_i^{(j)}}{b_0(n/k)/(n/k)}, j = 1, \ldots, d \right), i = 1, \ldots, n \right\}$$

$$= \left\{ \left(\frac{n/r_i^{(j)}}{b_0(n/k)}, j = 1, \ldots, d \right), i = 1, \ldots, n \right\}.$$

This scheme yields the hidden measure.

Proposition 1. *Assume that $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ is an i.i.d. sample from a distribution on $[0, \infty)^d$ that possesses both regular and hidden regular variation, so that (1) and (6) hold. Then we have*

$$\frac{1}{k} \sum_{i=1}^{n} \varepsilon_{((n/r_i^{(j)})/b_0(n/k), j=1,\ldots,d)} \Rightarrow \nu_0 \tag{18}$$

in $M_+(\mathbb{E}_0)$, where ν_0 is as given in (6).

Proof. The proof mimics the scheme followed when using ranks to estimate ν or S. Observe, for $\mathbf{x} \in \mathbb{E}_0$, that

$$\frac{1}{k} \sum_{i=1}^{n} \varepsilon_{((n/r_i^{(j)})/b_0(n/k), j=1,\ldots,d)}([\mathbf{x}, \infty])$$

$$= \frac{1}{k} \sum_{i=1}^{n} \mathbf{1}_{\{n(x^{(j)})^{-1} b_0^{-1}(n/k) \geq r_i^{(j)}, j=1,\ldots,d\}}$$

$$= \frac{1}{k} \sum_{i=1}^{n} \mathbf{1}_{\{Z_i^{(j)}/b^{(j)}(b_0(n/k)) \geq Z_{(\lceil n(x^{(j)} b_0(n/k))^{-1} \rceil)}^{(j)}/b^{(j)}(b_0(n/k)), j=1,\ldots,d\}}.$$

We claim (see below for the proof) that, for each $j = 1, \ldots, d$,

$$\frac{Z_{(\lceil n(x^{(j)} b_0(n/k))^{-1} \rceil)}^{(j)}}{b^{(j)}(b_0(n/k))} \xrightarrow{\text{P}} (x^{(j)})^{1/\alpha^{(j)}}. \tag{19}$$

Using this to scale the convergence in (7), we obtain

$$\frac{1}{k} \sum_{i=1}^{n} \varepsilon_{((n/r_i^{(j)})/b_0(n/k), j=1,\ldots,d)}([\mathbf{x}, \infty]) \Rightarrow \tilde{\nu}_0([\mathbf{x}^{1/\alpha}, \infty]) = \nu_0([\mathbf{x}, \infty]),$$

where we have used (8) in the last equality. This suffices to prove the result, given the claim.

Lemma 1. *Assume that* $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ *is an i.i.d. sample from a distribution on* $[0, \infty)^d$ *that possesses both regular and hidden regular variation, so that (1) and (6) hold. Then (19) holds in* $D[0, \infty)$ *for each* $j = 1, \ldots, d$.

Proof. We have, for each $j = 1, \ldots, d$,

$$\frac{b_0(n/k)}{n} \sum_{i=1}^n \varepsilon_{Z_i^{(j)}/b^{(j)}(b_0(n/k))} \Rightarrow \nu_{\alpha^{(j)}}$$

in $M_+(0, \infty]$, using Resnick (1986) or Resnick (1987, Exercise 3.5.7, p. 161). In particular,

$$\frac{b_0(n/k)}{n} \sum_{i=1}^n \varepsilon_{Z_i^{(j)}/b^{(j)}(b_0(n/k))}((x^{-1}, \infty]) \Rightarrow x^{\alpha^{(j)}}$$

in $D[0, \infty)$. This is a sequence of nondecreasing functions converging to a continuous limit and, so, the inverse functions converge as well. This yields the statement of the lemma.

If we convert (18) to polar coordinates in order to estimate S_0, we obtain the analogue of (17). As in (17), let $\mathbf{\Theta}_i$ be the polar angular coordinate of $(1/r_i^{(j)}, 1 \leq j \leq d)$ and let R_i be the norm of $(1/r_i^{(j)}, 1 \leq j \leq d)$. Then, assuming that S_0 is finite (otherwise one has to restrict $\mathbf{\Theta}_i$ to a compact subset of $\aleph_0$), we have

$$\frac{\sum_{i=1}^n \mathbf{1}_{\{R_i \geq n^{-1} b_0(n/k)\}} \varepsilon_{\mathbf{\Theta}_i}}{\sum_{i=1}^n \mathbf{1}_{\{R_i \geq n^{-1} b_0(n/k)\}}} \Rightarrow S_0 \tag{20}$$

in $M_+(\mathbb{E}_0)$. Since $b_0(n/k)$ is unknown for statistical purposes, it must be estimated before we can regard (20) as a suitable estimate of S_0.

4. Why estimation of the hidden measure matters

Hidden regular variation is designed to produce a model subclass of the multivariate regularly varying distributions possessing asymptotic independence. This model subclass is better suited to estimating the very small probabilities of the random vector falling in jointly remote regions well beyond the range of the observed data; see de Haan and de Ronde (1998) and Ledford and Tawn (1996), (1997). In this section, we review how hidden regular variation can help in the estimation of small probabilities.

For example, consider the following problems.

4.1. Estimate the probability of noncompliance

Suppose that the vector $\mathbf{Z} = (Z^{(1)}, \ldots, Z^{(d)})$ represents concentrations of a specific pollutant at d locations. (Alternatively, $\mathbf{Z}$ could represent concentrations of different pollutants at a single site.) Environmental agencies set standards by insisting that critical levels $\mathbf{t}_0 = (t_0^{(1)}, \ldots, t_0^{(d)})$ not be exceeded at any of the d sites; that is, that $\mathbf{Z} \leq \mathbf{t}_0$. Noncompliance is represented by the event

$$\{\text{noncompliance}\} = \{\mathbf{Z} \leq \mathbf{t}_0\}^c = \bigcup_{j=1}^d \{Z^{(j)} > t_0^{(j)}\}.$$

Noncompliance results in a fine or withdrawal of government support; it has various economic and political implications, none of which is desirable. How do we estimate the probability of noncompliance?

We assume only that (1) and (6) hold and, for simplicity, that $d = 2$. We observe that the probability of noncompliance is

$$P\left(\bigcup_{j=1}^{d}\{Z^{(j)} > t_0^{(j)}\}\right) = \sum_{j=1}^{2} P(Z^{(j)} > t_0^{(j)}) - P(Z^{(1)} > t_0^{(1)}, Z^{(2)} > t_0^{(2)}). \quad (21)$$

Assuming that only asymptotic independence holds, we would be inclined (if unacquainted with the work of Ledford and Tawn (1996), (1997)) to neglect the joint probability on the right-hand side since it is negligible compared to the univariate probabilities. However, as the following outline shows, the assumption of hidden regular variation allows for reasonable estimation of the joint probability.

Assume that we have an i.i.d. random sample $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$. For the univariate probabilities, we have

$$\begin{aligned}
\sum_{j=1}^{2} P(Z^{(j)} > t_0^{(j)}) &= \sum_{j=1}^{2} P\left(\frac{Z^{(j)}}{b^{(j)}(n/k)} > \frac{t_0^{(j)}}{b^{(j)}(n/k)}\right) \\
&\approx \frac{k}{n} \nu\left(\left[\mathbf{0}, \left(\frac{t_0^{(1)}}{b^{(1)}(n/k)}, \frac{t_0^{(2)}}{b^{(2)}(n/k)}\right)\right]^{c}\right) \\
&= \frac{k}{n}\left(\left(\frac{t_0^{(1)}}{b^{(1)}(n/k)}\right)^{-\alpha^{(1)}} + \left(\frac{t_0^{(2)}}{b^{(2)}(n/k)}\right)^{-\alpha^{(2)}}\right) \\
&\approx \frac{k}{n}\left(\left(\frac{t_0^{(1)}}{\hat{b}^{(1)}(n/k)}\right)^{-\hat{\alpha}^{(1)}} + \left(\frac{t_0^{(2)}}{\hat{b}^{(2)}(n/k)}\right)^{-\hat{\alpha}^{(2)}}\right),
\end{aligned}$$

where

$$\hat{b}^{(j)}(n/k) = Z_{(k)}^{(j)},$$

the kth largest of the jth components of the data, and $\hat{\alpha}^{(j)}$ is an estimate of $\alpha^{(j)}$ obtained from the one-dimensional sample of jth components. For example, $\hat{\alpha}^{(j)}$ could be the Hill estimator or the maximum likelihood estimator.

For the multivariate tail probability on the right-hand side of (21), we estimate

$$P(Z^{(1)} > t_0^{(1)}, Z^{(2)} > t_0^{(2)}) \quad (22)$$

$$\begin{aligned}
&= P\left(\frac{Z^{(j)}}{b^{(j)}(b_0(n/k))} > \frac{t_0^{(j)}}{b^{(j)}(b_0(n/k))}, j = 1, 2\right) \\
&\approx \frac{k}{n} \tilde{\nu}_0\left(\left(\left(\frac{t_0^{(1)}}{b^{(1)}(b_0(n/k))}, \frac{t_0^{(2)}}{b^{(2)}(b_0(n/k))}\right), \infty\right]\right) \quad \text{(from (7))} \\
&= \frac{k}{n} \nu_0\left(\left(\left(\left(\frac{t_0^{(1)}}{b^{(1)}(b_0(n/k))}\right)^{\alpha^{(1)}}, \left(\frac{t_0^{(2)}}{b^{(2)}(b_0(n/k))}\right)^{\alpha^{(2)}}\right), \infty\right]\right) \quad \text{(from (8))} \\
&\approx \frac{k}{n} \hat{\nu}_0\left(\left(\left(\left(\frac{t_0^{(1)}}{\widehat{b^{(1)} \circ b_0}(n/k)}\right)^{\hat{\alpha}^{(1)}}, \left(\frac{t_0^{(2)}}{\widehat{b^{(2)} \circ b_0}(n/k)}\right)^{\hat{\alpha}^{(2)}}\right), \infty\right]\right). \quad (23)
\end{aligned}$$

Here,

$$\hat{\nu}_0 := \frac{1}{k} \sum_{i=1}^{n} \varepsilon_{((1/r_i^{(j)})/m_{(k)}, 1 \le j \le d)}, \quad (24)$$

where $m_{(k)}$ is defined as follows: set

$$m_i = \bigwedge_{j=1}^{d} \frac{1}{r_i^{(j)}}, \qquad i = 1, \ldots, n, \tag{25}$$

and suppose that

$$m_{(1)} \geq m_{(2)} \geq \cdots \geq m_{(n)} \tag{26}$$

is the ordering of $m_1, \ldots, m_n$, with the largest first. We will relate $\hat{v}_0$ to v_0 in Proposition 2, below. For our estimate of $\widehat{b^{(j)} \circ b_0}(n/k)$, $j = 1, 2$, we use

$$\widehat{b^{(j)} \circ b_0}(n/k) = Z_{(\lceil 1/m_{(k)} \rceil)}^{(j)}. \tag{27}$$

This choice is justified in Section 4.3, below.

Although the rank method of estimating v_0 obviates the need to compute the αs, the α estimates are needed for estimating probabilities of remote events.

The calculation leading to (23) will be used in Section 6.2 to assess the risk of flooding. The data analysed there are precipitation in successive hours. For these bivariate data, simultaneous exceedance of a threshold represents the greatest risk of flooding due to rainfall.

4.2. Estimate the probability of a failure region

In certain water resource problems a failure region of the form

$$A := \{z \in \mathbb{E} \colon f(z, t_0) > 0\}$$

is determined by a function f and is required to estimate $P(\mathbf{Z} \in A)$ (Bruun and Tawn (1998), de Haan and de Ronde (1998)). For example, the failure probability could be of the form

$$P(a^{(1)} Z^{(1)} + a^{(2)} Z^{(2)} > t_0),$$

where $Z^{(1)}$ represents still water level, $Z^{(2)}$ represents wave height, and $a^{(j)} > 0$, $j = 1, 2$.

As before, assume that the estimation of the failure probability is based on an i.i.d. sample $\mathbf{Z}, \mathbf{Z}_1, \ldots, \mathbf{Z}_n$ and that there is hidden regular variation. Then,

$$P(a^{(1)} Z^{(1)} + a^{(2)} Z^{(2)} > t_0) = P(a^{(1)} Z^{(1)} + a^{(2)} Z^{(2)} > t_0, \mathbf{Z} \in \mathbb{E}_0)$$
$$+ P(a^{(1)} Z^{(1)} + a^{(2)} Z^{(2)} > t_0, \mathbf{Z} \in \mathbb{E} \setminus \mathbb{E}_0).$$

The second probability on the right-hand side is approximately

$$\frac{k}{n} \sum_{j=1}^{2} P\left(a^{(j)} \frac{Z^{(j)}}{b^{(j)}(n/k)} > \frac{t_0}{b^{(j)}(n/k)}\right) \approx \frac{k}{n} \sum_{j=1}^{2} \left(\frac{t_0}{a^{(j)} b^{(j)}(n/k)}\right)^{-\alpha^{(j)}}$$
$$\approx \frac{k}{n} \sum_{j=1}^{2} \left(\frac{t_0}{a^{(j)} Z_{(k)}^{(j)}}\right)^{-\hat{\alpha}^{(j)}}.$$

The probability on $\mathbb{E}_0$ is approximately

$$\frac{k}{n} \tilde{v}_0 \left\{ z \in \mathbb{E}_0 \colon \sum_{j=1}^{2} a^{(j)} b^{(j)} (b_0(n/k)) z^{(j)} > t_0 \right\}$$
$$\approx \frac{k}{n} \hat{\tilde{v}}_0 \left\{ z \in \mathbb{E}_0 \colon \sum_{j=1}^{2} a^{(j)} Z_{(\lceil 1/m_{(k)} \rceil)}^{(j)} z^{(j)} > t_0 \right\},$$

where $\hat{\bar{v}}_0(x, \infty] = \hat{v}_0(x^{\hat{\alpha}}, \infty]$, using the estimator in (27). Recall that $\hat{v}_0$ is given in (24) (it is also used in (29)) and that $\hat{\alpha} = (\hat{\alpha}^{(1)}, \hat{\alpha}^{(2)})$ can be estimated using the Hill estimator or the maximum likelihood estimator applied to the marginal one-dimensional samples.

4.3. Justification of the estimate of $b^{(j)} \circ b_0(n/k)$

In Section 4.1, (27) suggests that we estimate $b^{(j)} \circ b_0(n/k)$ by $Z^{(j)}_{(\lceil 1/m_{(k)} \rceil)}$. To see why this is an appropriate estimator of the composition, note that, from (30) (below, with $s = 1$) and (19), we have

$$\left(\frac{Z^{(j)}_{(\lceil n(x b_0(n/k))^{-1} \rceil)}}{b^{(j)}(b_0(n/k))}, \frac{n}{b_0(n/k)} m_{(k)} \right) \xrightarrow{P} (x^{1/\alpha^{(j)}}, 1) \tag{28}$$

in $D(0, \infty) \times [0, \infty)$. Therefore, by scaling using the map $(x(\cdot), t) \mapsto x(t\cdot)$, we find from (28) that

$$\frac{Z^{(j)}_{(\lceil n((nm_{(k)}/b_0(n/k))b_0(n/k))^{-1} \rceil)}}{b^{(j)}(b_0(n/k))} = \frac{Z^{(j)}_{(\lceil 1/m_{(k)} \rceil)}}{b^{(j)}(b_0(n/k))} \xrightarrow{P} 1.$$

Hence, we make the choice shown in (27).

5. Estimation of the hidden measure

For this section, recall that we write

$$\aleph_{\mathrm{inv}} := \left\{ x \in \mathbb{E}_0 : \bigwedge_{j=1}^{d} x^{(j)} \geq 1 \right\}$$

for the set of vectors all of whose components are at least 1. We continue to use the notation of (25) and (26). We have the following result.

Proposition 2. *Assume that* $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ *is an i.i.d. sample from a distribution on* $[0, \infty)^d$ *that possesses both regular and hidden regular variation, so that (1) and (6) hold, and continue to assume that* $v_0(\aleph_{\mathrm{inv}}) = 1$. *Then we have*

$$\hat{v}_0 := \frac{1}{k} \sum_{i=1}^{n} \varepsilon_{((1/r_i^{(j)})/m_{(k)}, 1 \leq j \leq d)} \Rightarrow v_0 \tag{29}$$

in $M_+(\mathbb{E}_0)$.

Thus, we have removed the unknown $b_0(n/k)$ and replaced it by a random variable.

Proof of Proposition 2. In $D[0, \infty)$ we have, from Proposition 1 and continuous mapping, that

$$\eta_n(t) := \frac{1}{k} \sum_{i=1}^{k} \varepsilon_{(n/b_0(n/k)) \bigwedge_{j=1}^{d} 1/r_i^{(j)}} ((t^{-1}, \infty])$$

$$\Rightarrow v_0\left(\left\{ x : \bigwedge_{j=1}^{d} x^{(j)} \geq t^{-1} \right\} \right) = t^{\alpha_0} v_0(\aleph_{\mathrm{inv}}) = t^{\alpha_0} =: \eta_\infty(t).$$

Therefore, in $D[0, \infty)$, we also find that the inverse processes converge:

$$\eta_n^{\leftarrow}(s) \Rightarrow \eta_\infty^{\leftarrow}(s) = s^{1/\alpha_0}.$$

Unpack the left-hand side. We have

$$
\begin{aligned}
\eta_n^{\leftarrow}(s) &= \inf\{u \colon \eta_n(u) \geq s\} \\
&= \inf\left\{u \colon \sum_{i=1}^n \varepsilon_{n/b_0(n/k)} m_i((u^{-1}, \infty]) \geq ks\right\} \\
&= \left(\sup\left\{v \colon \sum_{i=1}^n \varepsilon_{n/b_0(n/k)} m_i((v, \infty]) \geq ks\right\}\right)^{-1} \\
&= \frac{b_0(n/k)}{n}\left(\sup\left\{w \colon \sum_{i=1}^n \varepsilon_{m_i}((w, \infty]) \geq ks\right\}\right)^{-1} \\
&= \frac{b_0(n/k)}{n} m_{(\lceil ks \rceil)}^{-1}.
\end{aligned}
$$

Therefore, we see that

$$
\frac{n}{b_0(n/k)} m_{(\lceil ks \rceil)} \Rightarrow s^{-1/\alpha_0} \tag{30}
$$

in $D(0, \infty]$.

The rest is a scaling argument: we couple (30) with (18) and compose to obtain

$$
\begin{aligned}
\frac{1}{k}\sum_{i=1}^n \varepsilon_{((n/r_i^{(j)})/b_0(n/k),\, j=1,\ldots,d)}&\left(\left[\frac{n}{b_0(n/k)} m_{(k)} \boldsymbol{x}, \boldsymbol{\infty}\right]\right) \\
&= \frac{1}{k}\sum_{i=1}^n \varepsilon_{((1/r_i^{(j)})/m_{(k)},\, j=1,\ldots,d)}([\boldsymbol{x}, \boldsymbol{\infty}]) \Rightarrow v_0((\boldsymbol{x}, \boldsymbol{\infty}])
\end{aligned}
$$

in $D(\mathbb{E}_0)$, as required.

This suggests a way to circumvent the problem of the unknown function $b_0(n/k)$ in (20): we replace $n^{-1}b_0(n/k)$ by $m_{(k)}$. We can then write the analogue of (20). If v_0 is infinite, let $\aleph_0(K)$ be a convenient compact subset of $\aleph_0$. For $d = 2$, where $\aleph$ can be parametrized as $\aleph = [0, \frac{1}{2}\pi]$ and $\aleph_0 = (0, \frac{1}{2}\pi)$, we can set $\aleph_0(K) = [\delta, \frac{1}{2}\pi - \delta]$ for some small $\delta > 0$. Then, from Proposition 2, we have

$$
\frac{\sum_{i=1}^n \mathbf{1}_{\{R_i \geq m_{(k)}, \boldsymbol{\Theta}_i \in \aleph_0(K)\}}\, \varepsilon_{\boldsymbol{\Theta}_i}}{\sum_{i=1}^n \mathbf{1}_{\{R_i \geq m_{(k)}, \boldsymbol{\Theta}_i \in \aleph_0(K)\}}} \Rightarrow S_0(\cdot \cap \aleph_0(K)).
$$

If v_0 is finite, we can replace $\aleph_0(K)$ with $\aleph_0$, as was done in (20).

Thus, to summarize, we proceed as follows when estimating S_0.

1. Replace the heavy-tailed multivariate sample $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ by the n vectors $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_n$ of antiranks, where, recall,

$$
r_i^{(j)} = \sum_{l=1}^n \mathbf{1}_{\{Z_l^{(j)} \geq Z_i^{(j)}\}}, \qquad j = 1, \ldots, d, \ i = 1, \ldots, n.
$$

2. Compute the normalizing factors $m_i = \bigwedge_{j=1}^d 1/r_i^{(j)}$, $i = 1, \ldots, n$, and their order statistics

$$
m_{(1)} \geq \cdots \geq m_{(n)}.
$$

3. Compute the polar coordinates $\{(R_i, \Theta_i), i = 1, \ldots, n\}$ of $\{(1/r_i^{(j)}, j = 1, \ldots, d), i = 1, \ldots, n\}$.

4. Estimate S_0 using the Θ_i corresponding to $R_i \geq m_{(k)}$.

Thus, we have a simple method of estimating S_0. Alternatively, we can construct kernel density estimates corresponding to S_0, using the points Θ_i identified in step 4 above. This is the technique used in Section 6.

Ledford and Tawn (1997) gave an empirical estimator for their ray dependence function $g_*(w)$. However, constructing estimates of the hidden spectral density using this estimator and (10) involves an extra smoothing step, as this requires knowledge of $g_*'(w)$ and $g_*''(w)$. Empirical estimates of $g_*(w)$ will not provide information about its derivatives without some assumptions concerning the smoothness of the function.

Our results offer some advice on how to pick thresholds to estimate both S and S_0, but of course we must still choose k. The Stărică scaling device (Resnick (2003), Stărică (1999)) seems to offer some guidance, although the authors believe that nothing is presently known about its theoretical properties.

6. Two examples

This section illustrates the usefulness of our theoretical results in two contexts: Internet traffic studies and flood risk analysis.

6.1. Internet Hypertext Transfer Protocol (HTTP) response data

In our first example, we analyse HTTP response data describing Internet transmissions observed during a four-hour period from 1 p.m. to 5 p.m. on 26 April 2001 at the University of North Carolina at Chapel Hill. The datasets were obtained from the Department of Computer Science's Distributed and Real-Time Systems Group under the direction of Don Smith and Kevin Jeffay. Interest in this subject was stimulated by Steve Marron in his Mary Upson lectures at Cornell University in autumn 2001.

Internet file transfers are subject to delays and although one expects larger file transfers to encounter more delays, this is overly simplistic (Campos *et al.* (2005)). Large file transfers, while comparatively rare, comprise a significant fraction of all the bytes transferred on the Internet and, hence, are important in understanding the impact on end-user performance measures of diverse networking technologies such as routeing, congestion control, and server design. For HTTP (web browsing) responses, the joint behaviour of large values of three variables – size of response, time duration of response, and throughput (or, synonymously, rate = size/time) – can be considered. All three quantities are typically heavy tailed, but size and rate tend to be asymptotically independent; see Maulik *et al.* (2002) and Resnick (2003). Here, we consider the existence of hidden regular variation for the pair of variables (size, throughput).

The dataset consists of responses (bytes), in the stated time period, whose size is in excess of 100 000 bytes, coupled with the time required for transmission (seconds). There were 21 829 such transmissions. As opposed to the next example, on flood risk, our interest here is not specifically in estimating probabilities of rare events, such as joint threshold exceedance, but rather in understanding the underlying structure of the transmission process. This is intended to aid network engineers in their development of realistic models of Internet traffic processes, with which they simulate network behaviour. To this end, we focus on the establishment of asymptotic independence and hidden regular variation between transmission size and throughput, and then on obtaining an estimate of the hidden spectral measure characterizing the joint
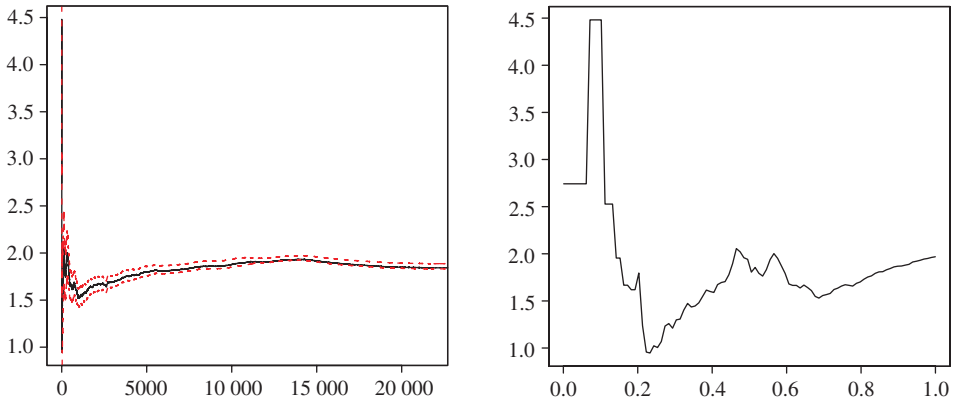
FIGURE 1: Hill plots on two scales for estimating the coefficient of tail dependence η. The left-hand plot is a Hill plot for $\alpha_0 = 1/\eta$, showing the Hill estimate of α_0 versus the number of order statistics. The right-hand plot is an altHill plot for α_0, showing the Hill estimate of α_0 based on $\lceil n^\theta \rceil$ upper order statistics, $0 \le \theta \le 1$.

tail of the distribution of these variables. This would suggest that a suitable model for (size, throughput) could be obtained from a mixture model, as in Maulik and Resnick (2005).

Although estimation of the marginal distributions is not required in the examination of the dependence structure, we first estimate the marginal tail parameters to establish the heavy-tailed behaviour of the transmission rate and size variables. We use Hill plots and alternative Hill (altHill) plots (not shown) to choose values of $k = 150$ and $k = 250$ for the size and rate variables, respectively. Estimates of tail indices for size and rate are relatively stable around the values of $\alpha^{(1)} = 1.8$ and $\alpha^{(2)} = 2.1$ in the ranges $[50, 3000]$ and $[50, 400]$ for these variables, respectively.

The next stage is to establish whether these Internet data exhibit asymptotic independence. We calculate the coefficient of tail dependence η of Ledford and Tawn (1996), (1997) using the Hill estimator for the shape parameter of the distribution of componentwise minima taken after rank transformation to standard Pareto margins. (See also Beirlant and Vandewalle (2002), Drees *et al.* (2004), and Peng (1999).) The Hill and altHill plots used to inform our choice of k for this estimation are shown in Figure 1. These plots show estimates of $\alpha_0 = 1/\eta$ constructed using different numbers of order statistics. The plots show the estimated value of η to be stable at around 0.6 for k in the range $[50, 400]$. This value is consistent with asymptotic independence and weak dependence between transmission size and rate at finite levels.

Another visual confirmation of asymptotic independence is to calculate the empirical measure (17) for different values of k. Figure 2 shows kernel density estimates corresponding to these estimates for $k = 300, 600, 900, 1200$. These plots clearly show the concentration of mass towards the endpoints of the interval $[0, 1]$ as we use fewer and fewer order statistics. This gives further compelling evidence of asymptotic independence.

Having established that our variables are both heavy tailed and asymptotically independent variables, we can now employ steps 1–4 given at the end of Section 5 to estimate the hidden measure. Implementation of these steps involves a choice of k, and we use the Stărică scaling device (Resnick (2003), Stărică (1999)) to aid us in making this choice. This tool is motivated by the homogeneity property (3) of the spectral measure ν_{standard}. Recall, from (16), that
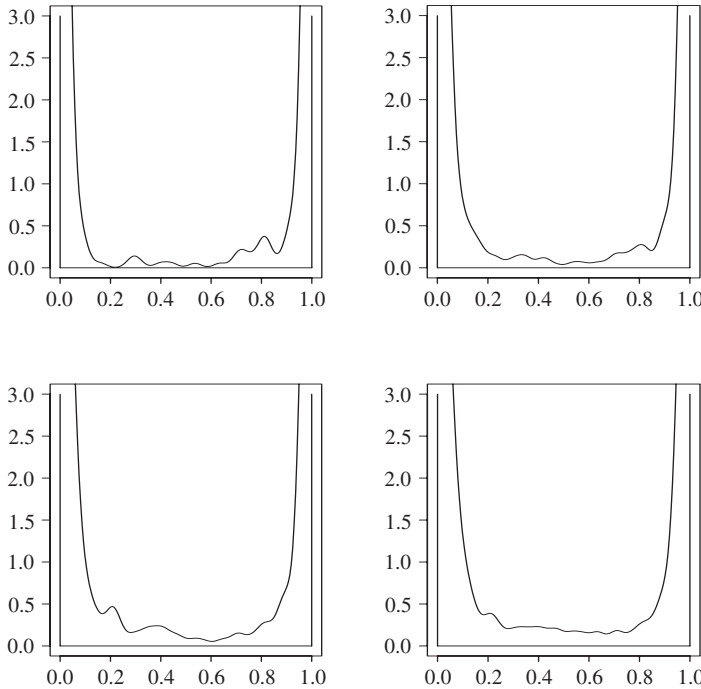
FIGURE 2: Density estimates of the spectral measure S for the Internet transmission size and rate data. The plots show spectral density estimate versus angle for norm r greater than the 300th largest norm (*top left*), r greater than the 600th largest norm (*top right*), r greater than the 900th largest norm (*bottom left*), and r greater than the 1200th largest norm (*bottom right*). (The norm is for the rank-transformed data.)

$\hat{v}_{\text{standard},n}$ is the empirical estimate of v_{standard} and contains the parameter k that must be selected. Since v_{standard} is homogeneous, Stărică argued that a good choice of k is that for which the following approximation is true, where u is in a neighbourhood of 1 and $A = \{x \in \mathbb{E}: \|x\| > 1\}$:

$$\hat{v}_{\text{standard},n}(uA) \approx u^{-1} \hat{v}_{\text{standard},n}(A).$$

For the Internet data, the value of k for which the scaling ratio

$$\hat{v}_{\text{standard},n}(uA)/u^{-1}\hat{v}_{\text{standard},n}(A) \tag{31}$$

is most stable around the value 1, for u in the interval $[0.5, 1.5]$, is $k = 1112$. The resulting value of $m_{(1112)}$ corresponds to the 0.45 quantile of the radial components. Figure 3 shows this scaling ratio, calculated using $k = 1112$ for u between 0 and 10.

Following this suggestion for the choice of k, we plot the estimated hidden measure for $k = 1000, 1150, 1200, 1250$ in Figure 4. These plots show stability of the estimated measure for these values of k. Again, since the measure may be infinite, the interval on which we estimate the measure has been bounded away from 0 and 1, and we show the kernel density estimate only on the interval $[0.1, 0.9]$. An edge correction has been applied so that the density integrates to 1 on this interval. All plots show the hidden measure to be bimodal with peaks around 0.2 and 0.85.
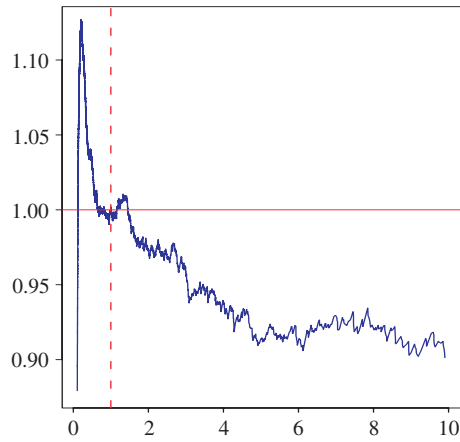
FIGURE 3: Stărică scaling plot for the Internet transmission size and rate data, showing the scaling ratio (31) versus the scaling constant u, for $k = 1112$.
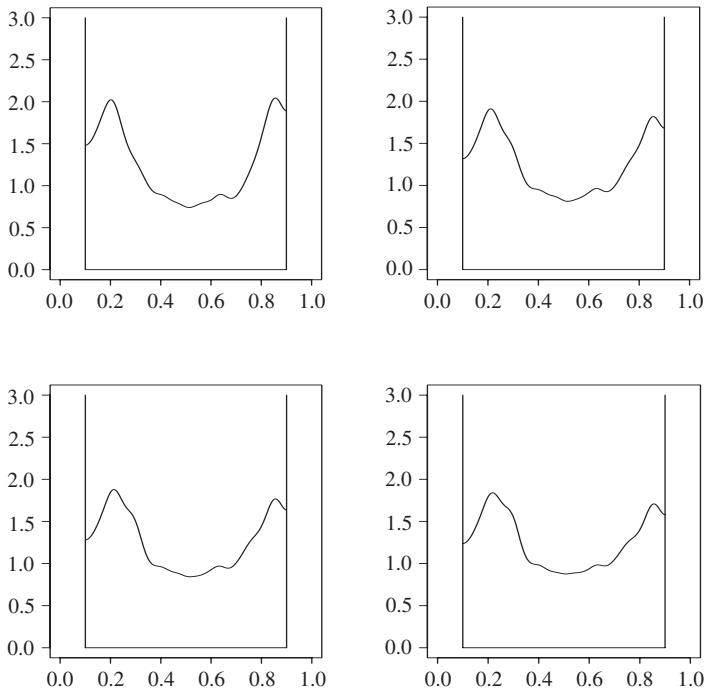

FIGURE 4: Density estimates of the hidden spectral measure S_0 for the rank-transformed Internet transmission size and rate data. The plots show density versus angle for $r > m_{(1000)}$ (*top left*), $r > m_{(1150)}$ (*top right*), $r > m_{(1200)}$ (*bottom left*), and $r > m_{(1250)}$ (*bottom right*).
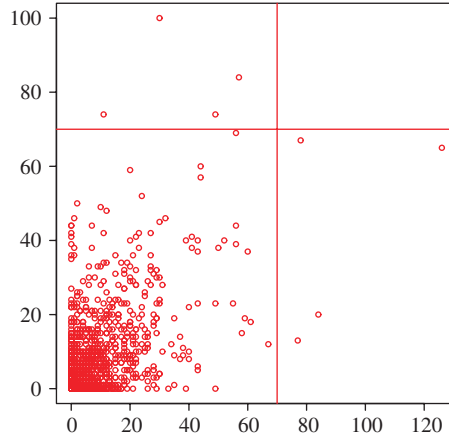
FIGURE 5: Scatter plot of one-hour rainfall measurements (in units of 0.1 mm) from Eskdalemuir, 1970–1986, showing $Z^{(2)}$ versus $Z^{(1)}$. The lines show the exceedance thresholds $t_0^{(1)} = t_0^{(2)} = 70$, which are not jointly exceeded by any pair in the dataset.

The dependence between variables is summarized by the coefficient of tail dependence, which was estimated to be approximately $\eta = 0.6$. The hidden spectral measure refines the characterization of extreme dependence by revealing the finer structure of the joint tail within the class of distributions having $\eta \approx 0.6$. Figure 4 shows the Internet response data to exhibit weak dependence within this class, as the measure places more mass towards the ends of the interval than in the centre.

This estimate of the hidden spectral measure, in combination with the estimated coefficient of tail dependence, provides network engineers with a model of the joint tail of the response distribution, from which they could construct importance samples. Simulation of data from this region may prove to be another technique useful in studying such data.

6.2. Risk from flooding in Eskdalemuir

For our next example, we illustrate the estimation of risk from flooding using hourly rainfall measurements (in units of 0.1 mm) from Eskdalemuir in the south-west of Scotland, during the years 1970–1986 inclusive. This dataset was previously considered in Nadarajah *et al.* (1998). The rainfall process exhibits both serial dependence and diurnal cycles. To avoid complications arising from these features we focus on consecutive pairs of hourly observations taken from 11:00 to 12:00 and from 12:00 to 13:00 daily. The choice of these particular pairs of consecutive hours is arbitrary but does not materially affect our analysis. This gives us a series of 6209 daily observations of consecutive hours represented by the random pair $\mathbf{Z} = (Z^{(1)}, Z^{(2)})$, which we assume to be i.i.d. Around 80% of the observations are 0, corresponding to dry hours. The data are plotted in Figure 5.

Interest in extreme rainfall is generally motivated by a desire to understand and protect against rainfall that could result in flooding. Four factors contribute to the likelihood of flooding following a rain event: rainfall intensity and duration (meteorological), and ground saturation and rainfall-catchment response (hydrological). We focus on the meteorological factors. A historical study of extreme rainfall events in the UK (Hand (2002)) clearly shows that the rainfall events with the highest hourly precipitation rates in the UK are the shortest events, typically lasting less than one hour and exclusively resulting from convective rain. This type
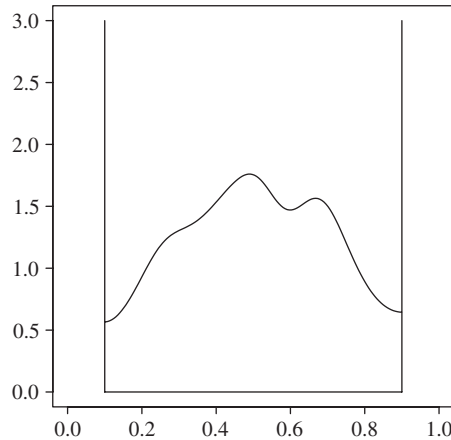
FIGURE 6: Density estimate of the hidden spectral measure S_0, showing density versus angle for $R > m_{(210)}$.

of precipitation is typically intense, of short duration, and is often accompanied by thunder and lightning. This finding suggests that rainfall amounts may be asymptotically independent from one hour to the next, as the intensity of such precipitation events is rarely sustained for long enough to make the rainfall counts extreme in both hours. Our aim in this example is to estimate the probability of flooding, i.e. rainfall in excess of a threshold t_0 in consecutive hours, for large values of t_0.

Diagnostic plots (not shown) such as the Hill, altHill, and quantile–quantile (QQ) plots (Beirlant *et al.* (1996), Drees *et al.* (2000), Kratz and Resnick (1996), Resnick (2003), Resnick and Stărică (1997)) suggest that $k = 300$ is appropriate for estimation of the marginal parameters $\alpha^{(1)}$ and $\alpha^{(2)}$. These are estimated, using the Hill estimator, to be 1.66 and 1.57 for the periods 11:00–12:00 and 12:00–13:00, respectively. We checked for asymptotic independence by estimating S, using the estimator in (17), for a variety of values of k. Plots of this empirical measure (not shown) show the movement of the mass to the endpoints of its support as k decreases. We confirmed this finding by estimating the coefficient of tail dependence of Ledford and Tawn (1996), (1997) using the Hill estimator for the shape parameter of the distribution of componentwise minima taken after rank transformation. The estimated value of this parameter is stable at around 0.83 for k in the range [10, 500]. This value is consistent with asymptotic independence, but with reasonably strong dependence at finite levels.

We choose the value of k to use for the dependence estimation (29) using the Stărică scaling plot (Resnick (2003), Stărică (1999)). This device suggests a value of $k = 210$ for this pair of variables. Using the above values of $(\hat{\alpha}^{(1)}, \hat{\alpha}^{(2)}) = (1.66, 1.57)$ and $k = 210$, we find the estimates of the probability of joint exceedance of $t_0^{(1)} = t_0^{(2)} = 40, 50, 60, 70$ to be 6.0×10^{-5}, 3.2×10^{-5}, 2.2×10^{-5}, and 1.6×10^{-5}, respectively. For such values of $t_0^{(1)}$ and $t_0^{(2)}$, empirical exceedance rates using the observed counts of joint exceedances are based on too few data to be reliable. In particular, for $t_0^{(1)} = t_0^{(2)} = 70$ (indicated by the lines plotted in Figure 5), there are no observed joint exceedances of this threshold. Our estimate of $\hat{v}_0$ uses a greater proportion of the data and then scales by k/n, according to (23).

An additional benefit of our proposed estimation scheme is the ability to visualize the hidden spectral measure S_0. We again follow steps 1–4, given at the end of Section 5, to construct points that we treat as a sample from the hidden angular distribution. Figure 6 shows the

kernel density estimate of the angular components constructed using the L_1-norm, and radial components above $m_{(210)}$. Since the hidden spectral measure may be infinite, we have bounded the support of our estimate away from the ends of the interval $[0, 1]$, and show the density estimate only on the interval $[0.1, 0.9]$. An edge correction has been applied to ensure that the density integrates to 1 on this interval. This density estimate is relatively stable for k in the range $[100, 300]$. Note that $m_{(210)}$ corresponds to the 0.84 quantile of the distribution of radial components.

In contrast to the estimated measure shown in Figure 4, Figure 6 shows the hidden spectral measure for the rainfall data to be roughly unimodal and approximately symmetric. This indicates a relatively strong degree of dependence between the variables, within the class of distributions having $\eta \approx 0.8$. This, taken with the greater value of the coefficient of tail dependence than that estimated in the Internet response example, shows the rainfall data to exhibit materially stronger dependence, although both datasets exhibit asymptotic independence.

Acknowledgements

References

BEIRLANT, J. AND VANDEWALLE, B. (2002). Some comments on the estimation of a dependence index in bivariate extreme value statistics. *Statist. Prob. Lett.* **60,** 265–278.

BEIRLANT, J., VYNCKIER, P. AND TEUGELS, J. (1996). Tail index estimation, Pareto quantile plots, and regression diagnostics. *J. Amer. Statist. Assoc.* **91,** 1659–1667.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. John Wiley, New York.

BRUUN, J. T. AND TAWN, J. A. (1998). Comparison of approaches for estimating the probability of coastal flooding. *J. R. Statist. Soc. C* **47,** 405–423.

CAMPOS, F. H. *et al.* (2005). Extremal dependence: Internet traffic applications. *Stoch. Models* **21,** 1–35.

COLES, S. G., HEFFERNAN, J. E. AND TAWN, J. A. (1999). Dependence measures for extreme value analyses. *Extremes* **2,** 339–365.

DE HAAN, L. AND DE RONDE, J. (1998). Sea and wind: multivariate extremes at work. *Extremes* **1,** 7–46.

DE HAAN, L. AND OMEY, E. (1984). Integrals and derivatives of regularly varying functions in $\mathbb{R}^d$ and domains of attraction of stable distributions. II. *Stoch. Process. Appl.* **16,** 157–170.

DRAISMA, G., DREES, H., FERREIRA, A. AND DE HAAN, L. (2004). Bivariate tail estimation: dependence in asymptotic independence. *Bernoulli* **10,** 251–280.

DREES, H., DE HAAN, L. AND RESNICK, S. I. (2000). How to make a Hill plot. *Ann. Statist.* **28,** 254–274.

DREES, H., FERREIRA, A. AND DE HAAN, L. (2004). On maximum likelihood estimation of the extreme value index. *Ann. Appl. Prob.* **14,** 1179–1201.

GREENWOOD, P. AND RESNICK, S. (1979). A bivariate stable characterization and domains of attraction. *J. Multivariate Anal.* **9,** 206–221.

HAND, W. H. (2002). Numerical weather prediction. A historical study of rainfall events in the 20th century. Forecast. Tech. Rep. 384, UK Met Office. Available at http://www.metoffice.com/research/nwp/publications/papers/technical_reports/2002/FRTR384/FRTR384.pdf.

HEFFERNAN, J. E. (2000). A directory of coefficients of tail dependence. *Extremes* **3,** 279–290.

HUANG, X. (1992). Statistics of bivariate extreme values. Doctoral Thesis, Erasmus University Rotterdam.

HULT, H. AND LINDSKOG, F. (2002). Multivariate extremes, aggregation and dependence in elliptical distributions. *Adv. Appl. Prob.* **34,** 587–608.

KRATZ, M. AND RESNICK, S. I. (1996). The QQ-estimator and heavy tails. *Commun. Statist. Stoch. Models* **12,** 699–724.

LEDFORD, A. W. AND TAWN, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika* **83,** 169–187.

LEDFORD, A. W. AND TAWN, J. A. (1997). Modelling dependence within joint tail regions. *J. R. Statist. Soc. B* **59,** 475–499.

MAULIK, K. AND RESNICK, S. I. (2005). Characterizations and examples of hidden regular variation. *Extremes* **7,** 31–67.

MAULIK, K., RESNICK, S. I. AND ROOTZÉN, H. (2002). Asymptotic independence and a network traffic model. *J. Appl. Prob.* **39,** 671–699.

NADARAJAH, S., ANDERSON, C. W. AND TAWN, J. A. (1998). Ordered multivariate extremes. *J. R. Statist. Soc. B* **60,** 473–496.

PENG, L. (1999). Estimation of the coefficient of tail dependence in bivariate extremes. *Statist. Prob. Lett.* **43,** 399–409.

POON, S.-H., ROCKINGER, M. AND TAWN, J. (2003). Modelling extreme-value dependence in international stock markets. *Statistica Sinica* **13,** 929–953.

RESNICK, S. I. (1986). Point processes, regular variation and weak convergence. *Adv. Appl. Prob.* **18,** 66–138.

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

RESNICK, S. (2002). Hidden regular variation, second order regular variation and asymptotic independence. *Extremes* **5,** 303–336.

RESNICK, S. (2003). Modeling data networks. In *Exteme Values in Finance, Telecommunications, and the Environment*, eds B. Finkenstadt and H. Rootzen, Chapman and Hall, London, pp. 287–372.

RESNICK, S. (2004). The extremal dependence measure and asymptotic independence. *Stoch. Models* **20,** 205–227.

RESNICK, S. AND STĂRICĂ, C. (1997). Smoothing the Hill estimator. *Adv. Appl. Prob.* **29,** 271–293.

STĂRICĂ, C. (1999). Multivariate extremes for models with constant conditional correlations. *J. Empirical Finance* **6,** 515–553.