

# LIVING ON THE MULTIDIMENSIONAL EDGE: SEEKING HIDDEN RISKS USING REGULAR VARIATION

BIKRAMJIT DAS,\* *ETH Zürich*  
ABHIMANYU MITRA \*\* \*\*\*\* AND  
SIDNEY RESNICK,\*\* \*\*\*\* *Cornell University*

## Abstract

Multivariate regular variation plays a role in assessing tail risk in diverse applications such as finance, telecommunications, insurance, and environmental science. The classical theory, being based on an asymptotic model, sometimes leads to inaccurate and useless estimates of probabilities of joint tail regions. This problem can be partly ameliorated by using *hidden regular variation* (see Resnick (2002) and Mitra and Resnick (2011)). We offer a more flexible definition of hidden regular variation that provides improved risk estimates for a larger class of tail risk regions.

*Keywords:* Regular variation; spectral measure; asymptotic independence; risk set; vague convergence

2010 Mathematics Subject Classification: Primary 60F99; 62G32  
Secondary 60G70

## 1. Introduction

Daily we observe environmental, technological, and financial phenomena possessing inherent risks. There are financial risks from large investment losses; environmental risks from health hazards resulting from high concentrations of atmospheric pollutants; and hydrological risks from river floods. Risk analysis requires estimation of tail probabilities that provide measures of such risks. The mathematical framework of multivariate regular variation provides tools to compute tail probabilities associated with such risks; see [3], [11], and [21, Chapter 6]. These tools have limitations which we begin to address in this paper.

Consider a nonnegative random vector  $\mathbf{Z} = (Z_1, Z_2, \dots, Z_d)$  called a *risk vector*. The distribution of  $\mathbf{Z}$  has multivariate regular variation if there exist a function  $b(t) \uparrow \infty$  and a nonnegative, nondegenerate Radon measure  $\mu(\cdot)$  on  $\mathbb{E} = [0, \infty]^d \setminus \{(0, 0, \dots, 0)\}$  such that

$$t \mathbf{P} \left[ \frac{\mathbf{Z}}{b(t)} \in \cdot \right] \xrightarrow{v} \mu(\cdot) \quad \text{as } t \rightarrow \infty, \quad (1.1)$$

where ‘ $\xrightarrow{v}$ ’ denotes vague convergence in  $\mathbb{M}_+(\mathbb{E})$ , the set of Radon measures on  $\mathbb{E}$  [21, p. 172]. Note that (1.1) effectively assumes tail equivalence of the marginal components

Received 27 September 2011; revision received 19 June 2012.

\* Postal address: RiskLab, Department of Mathematics, ETH Zürich, Rämistrasse 101, 8092 Zürich, Switzerland.  
Email address: bikram@math.ethz.ch

\*\* Postal address: School of Operations Research and Information Engineering, Cornell University, Ithaca, NY 14853, USA.

\*\*\* Email address: am492@cornell.edu

\*\*\*\* Email address: sir1@cornell.edu

[21, Section 6.5.6], so while (1.1) is valuable as a theoretical foundation, it must be modified for applications.

The asymptotic relation (1.1) allows the limit measure  $\mu(\cdot)$  to be used for approximating tail probabilities. For example, approximation of the probability of the event  $\{Z_i > x_i$  for some  $i\}$  for large thresholds  $x_i$ ,  $i = 1, 2, \dots, d$ , requires the ability to compute the quantity  $\mu(\{(z_1, \dots, z_d) \in \mathbb{E} : z_i > w_i \text{ for some } i\})$  for  $w_i > 0$ ,  $i = 1, 2, \dots, d$ . Such approximations of tail probabilities are sensitive to degeneracies in the limit measure  $\mu(\cdot)$ . For example, when *asymptotic independence* is present as in Gaussian copula models, the limit measure  $\mu(\cdot)$  in (1.1) concentrates on the coordinate axes  $\mathbb{L}_i := \{\mathbf{x} \in \mathbb{R}^d : x_j = 0 \text{ for all } j \neq i\}$ ,  $i = 1, \dots, d$ , and  $\mu(\{(z_1, z_2, \dots, z_d) \in \mathbb{E} : z_i > w_1, z_j > w_2\}) = 0$  for any  $1 \leq i < j \leq d$  and  $w_1, w_2 > 0$ . Consequently, for large  $x_1$  and  $x_2$ , we approximate the joint tail probability  $P[Z_i > x_1, Z_j > x_2] \approx 0$  and conclude that risk contagion is absent. This conclusion may be naive and building on the seminal concept of *coefficient of tail dependence* [13], [14], *hidden regular variation* (HRV) [20] was introduced as a refinement of the approximation; see [7], [15], and [17].

HRV offers some advantages, but also has weaknesses. The existing definition provides insight into only the presence of a restricted class of degeneracies in the limit measure  $\mu(\cdot)$  in (1.1); namely when  $\mu(\cdot)$  concentrates either on the coordinate axes, or the coordinate planes or similar coordinate hyperplanes in higher dimensions. However, other degeneracies in  $\mu(\cdot)$  are possible; for example,  $\mu(\cdot)$  may concentrate on the diagonal  $\{(z_1, z_2, \dots, z_d) \in \mathbb{E} : z_1 = z_2 = \dots = z_d\}$ , a condition called asymptotic full dependence. To deal with situations where  $\mu$  puts zero mass on large portions of the state space, in Section 3 we define HRV on cones. (We define set  $\mathbb{C} \subset \mathbb{R}^d$  to be a cone if  $\mathbf{x} \in \mathbb{C}$  implies that  $t\mathbf{x} \in \mathbb{C}$  for  $t > 0$ .) The basic idea is to seek a lower-order regular variation property on the complement of the support of  $\mu$  and our claim is that the complement of the support offers the possibility of further tail modeling.

In practice, different risk assessment problems require calculating tail probabilities for different kinds of events. Hidden regular variation, as originally defined, may be a natural choice for some calculations but not for others. For example, suppose that  $(Z_1, Z_2) \in \mathbb{R}_+^2$  is a risk vector and we must calculate risk probabilities of the form  $P[|Z_1 - Z_2| > w]$  for large thresholds  $w > 0$ . If we use multivariate regular variation when the limit measure  $\mu(\cdot)$  in (1.1) is concentrated on the diagonal  $\{(z_1, z_2) \in \mathbb{E} : z_1 = z_2\}$ , the tail probability  $P[|Z_1 - Z_2| > w]$  must be approximated as 0 for large thresholds  $w > 0$ . The existing notion of HRV designed to help when  $\mu$  concentrates on the axes offers no assistance in this case. In Example 5.2 we illustrate how a more general theory overcomes this difficulty. Along with other examples in Section 5, we see that there is a need for HRV on general cones.

The conditional extreme value (CEV) model [4], [8], [9] provides one alternative approach to multivariate extreme value modeling. In standard form, the CEV model can also be formulated as regular variation on a particular cone in  $\mathbb{E}$  and this is discussed in Section 4. In Section 4 we also consider nonstandard regular variation from the point of view of regular variation on a sequence of cones. Nonstandard regular variation is essential in practice since in applications we cannot assume tail equivalence of all marginals in a multivariate model, as is done in (1.1).

In Section 6 we discuss how to fit our model of HRV on cones to data as well as estimation techniques of tail probabilities using our model. We have adapted ideas previously used in multivariate heavy tail analysis; this discussion is not comprehensive, merely a feasibility display. In particular, we have not performed data analyses. We close our discussion with concluding remarks in Section 7 and give some deferred results and proofs in Appendix A.

Our definition of HRV relies on a notion of convergence of measures called  $\mathbb{M}^*$ -convergence that is similar to the  $\mathbb{M}_0$ -convergence of [10]. We review  $\mathbb{M}^*$ -convergence in Section 2 and also discuss reasons to abandon the standard practice of defining regular variation through vague convergence on a compactification of  $\mathbb{R}^d$ .

**1.1. Notation**

We briefly discuss some frequently used notation and concepts.

1.1.1. *Vectors, norms, and topology.* Bold letters are used to denote vectors, with capital letters reserved for random vectors and small letters for nonrandom vectors, e.g.  $\mathbf{x} = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$ . We also define  $\mathbf{0} = (0, 0, \dots, 0)$ ,  $\mathbf{1} = (1, 1, \dots, 1)$ , and  $\infty = (\infty, \infty, \dots, \infty)$ . Operations on and between vectors are understood componentwise. For example, for vectors  $\mathbf{x}$  and  $\mathbf{z}$ ,  $\mathbf{x} \leq \mathbf{z}$  means that  $x_i \leq z_i$ ,  $i = 1, \dots, d$ . For a set  $A \subset [0, \infty)^d$  and  $\mathbf{x} \in A$ , we use  $[\mathbf{0}, \mathbf{x}]^c$  to mean  $[\mathbf{0}, \mathbf{x}]^c = A \setminus [\mathbf{0}, \mathbf{x}] = \{\mathbf{y} \in A : \bigvee_{i=1}^d y_i/x_i > 1\}$ . When we use the notation  $[\mathbf{0}, \mathbf{x}]^c$ , the set  $A$  should be clear from the context.

For the  $i$ th largest component of  $\mathbf{x}$ , we write  $x_{(i)}$ , that is,  $x_{(1)} \geq x_{(2)} \geq \dots \geq x_{(d)}$ . Thus, a subscript  $i$  denotes the  $i$ th component of a vector, whereas a subscript  $(i)$  denotes the  $i$ th largest component in the vector.

Operations with  $\infty$  are understood using the conventions

$$\begin{aligned} \infty + \infty &= \infty, & \infty - \infty &= 0, & \infty + x &= \infty - x = \infty & \text{for } x \in \mathbb{R}, \\ 0 \cdot \infty &= 0, & x \cdot \infty &= \infty & \text{for } x > 0, & x \cdot \infty &= -\infty & \text{for } x < 0. \end{aligned}$$

Fix a norm on  $\mathbb{R}^d$ , and denote the norm of  $\mathbf{x}$  as  $\|\mathbf{x}\|$ . Let  $\text{dist}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$  be the metric induced by the norm and, as usual, for  $A \subset \mathbb{R}^d$ , set  $\text{dist}(\mathbf{x}, A) = \inf_{\mathbf{y} \in A} \text{dist}(\mathbf{x}, \mathbf{y})$ . When attention is focused on the set  $\mathbb{C}$ , and  $A \subset \mathbb{C}$ , the  $\delta$ -dilation or swelling of  $A$  in  $\mathbb{C}$  is  $A^\delta := \{\mathbf{x} \in \mathbb{C} : \text{dist}(\mathbf{x}, A) < \delta\}$ . The topology on  $\mathbb{R}^d$  is the usual norm topology referred to as the Euclidean topology and the topology on a subset of  $\mathbb{R}^d$  is the relative topology induced by the Euclidean topology.

Two sets  $A$  and  $B$  in  $\mathbb{R}^d$  are *bounded away* from each other if  $\bar{A} \cap \bar{B} = \emptyset$ , where  $\bar{A}$  and  $\bar{B}$  are the closures of  $A$  and  $B$ .

1.1.2. *Cones.* We denote by  $\mathbb{E} = [0, \infty]^d \setminus \{\mathbf{0}\}$  and  $\mathbb{D} = [0, \infty)^d \setminus \{\mathbf{0}\}$ , the one point puncturing of the compactified and uncompactified versions of  $\mathbb{R}_+^d$ . The notation  $\mathbb{E}$  and  $\mathbb{D}$  may appear with superscripts denoting subsets of the compactified and uncompactified  $\mathbb{R}_+^d$ , respectively. For example,  $\mathbb{E}^{(l)} = \{\mathbf{x} \in [0, \infty]^d : x_{(l)} > 0\}$  and  $\mathbb{D}^{(l)} = \{\mathbf{x} \in [0, \infty)^d : x_{(l)} > 0\}$ , where  $x_{(l)}$  is the  $l$ th largest component of  $\mathbf{x}$ . Note that  $\mathbb{E}^{(1)} = \mathbb{E}$ .

A set  $\mathbb{C} \subset \mathbb{R}^d$  is a cone if  $\mathbf{x} \in \mathbb{C}$  implies that  $t\mathbf{x} \in \mathbb{C}$  for all  $t > 0$ . Cones in Euclidean space are usually denoted by the blackboard bold symbols  $\mathbb{C}, \mathbb{D}, \mathbb{E}, \mathbb{F}$ , etc. Since one often deals with nonnegative risk vectors, we focus on the case where  $\mathbb{C} \subset [0, \infty)^d$ . We call a subset  $\mathbb{F} \subset \mathbb{C}$  a closed cone in  $\mathbb{C}$  if  $\mathbb{F}$  is a closed subset of  $\mathbb{C}$  as well as a cone, e.g.  $\mathbb{F} = \{\mathbf{0}\}$  or, when  $d = 2$ ,  $\mathbb{F} = \{(t, 0) : t \geq 0\}$ . The complement of the closed cone  $\mathbb{F}$  in  $\mathbb{C}$  is an open cone in  $\mathbb{C}$ ; that is, the complement of  $\mathbb{F}$  is an open subset in  $\mathbb{C}$  as well as a cone.

Fix a closed cone  $\mathbb{C} \subset [0, \infty)^d$  containing  $\mathbf{0}$ , and suppose that  $\mathbb{F} \subset \mathbb{C}$  is a closed cone in  $\mathbb{C}$  containing  $\mathbf{0}$ . Then  $\mathbb{O} := \mathbb{C} \setminus \mathbb{F}$  is an open cone, and  $\mathbb{C}$  and  $\mathbb{F}$  are complete separable metric spaces under the metric  $\text{dist}(\cdot, \cdot)$ . Let  $\mathcal{C}$  denote the Borel  $\sigma$ -algebra of  $\mathbb{C}$ . Clearly,  $\mathbb{O}$  is again a separable metric space (not necessarily complete) equipped with the  $\sigma$ -algebra  $\emptyset = \{B \subset \mathbb{O} : B \in \mathcal{C}\}$ .

Some examples include

- $\mathbb{C} = [0, \infty)^d, \mathbb{F} = \{\mathbf{0}\}$ , and then  $\mathbb{O} = \mathbb{D} = [0, \infty)^d \setminus \{\mathbf{0}\}$ ,
- $\mathbb{C} = [0, \infty)^2 (d = 2), \mathbb{F} = \{(0, x) : x \geq 0\} \cup \{(y, 0) : y \geq 0\}$ , and then  $\mathbb{O} = (0, \infty)^2$ ,
- $\mathbb{C} = [0, \infty)^2 (d = 2), \mathbb{F} = \{(x, 0) : x \geq 0\}$ , and then  $\mathbb{O} = \mathbb{D}_\cap := [0, \infty) \times (0, \infty)$ .

1.1.3. *Regularly varying functions.* A function  $U : [0, \infty) \mapsto [0, \infty)$  is regularly varying with index  $\beta \in \mathbb{R}$  if, for all  $x > 0$ ,

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

We write  $U \in \text{RV}_\beta$ . See [2], [5], and [22].

1.1.4. *Vague convergence of measures.* We express the vague convergence of Radon measures as ‘ $\xrightarrow{v}$ ’ (see [12] and [21, p. 173]) and the weak convergence of probability measures as ‘ $\Rightarrow$ ’ [1, p. 14]. Denote the set of nonnegative Radon measures on a space  $\mathbb{S}$  as  $\mathbb{M}_+(\mathbb{S})$  and the set of all nonnegative continuous functions with compact support from  $\mathbb{S}$  to  $\mathbb{R}_+$  as  $C_K^+(\mathbb{S})$ . Moreover, for any element  $x \in \mathbb{S}$ , the point measure at  $x$  is defined as

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

Vague convergence on  $\mathbb{E}$  has traditionally been used for defining multivariate regular variation. We now explain why continuing this practice is problematic and what should be done.

## 2. $\mathbb{M}^*$ -convergence of measures and regular variation on cones

In this paper we require a definition of multivariate regular variation on cones in Euclidean space which differs from the traditional definition using *vague convergence* of measures. Following [10], we define regular variation based on a notion of convergence of measures we call  $\mathbb{M}^*$ -convergence.

### 2.1. Problems with compactification of $\mathbb{R}^d$

Multivariate regular variation on  $[0, \infty)^d$  is usually defined using vague convergence of Radon measures on  $\mathbb{E} = [0, \infty)^d \setminus \{\mathbf{0}\}$  [21, Section 6.1.4]. The reason for compactifying  $[0, \infty)^d$  and then removing  $\mathbf{0}$  is that this makes sets bounded away from  $\{\mathbf{0}\}$  relatively compact (cf. [21, Section 6.1.3]), and since Radon measures put finite mass on relatively compact sets, this theory is suitable for estimating probabilities of tail regions.

A theory of HRV may require removal of more than just a point. Furthermore, compactifying from  $[0, \infty)^d$  to  $[0, \infty]^d$  introduces problems. For one thing, it is customary to rely heavily on the polar coordinate transform  $\mathbf{x} \mapsto (\|\mathbf{x}\|, \mathbf{x}/\|\mathbf{x}\|)$  which is defined only on  $[0, \infty)^d \setminus \{\mathbf{0}\}$ , and if the state space  $[0, \infty]^d \setminus \{\mathbf{0}\}$  is used, an awkward kluge [21, p. 176] is required to show the equivalence of regular variation in polar and Cartesian coordinates. A workaround is only possible because the limit measure  $\mu$  in (1.1) puts zero mass on lines through infinity  $\{\mathbf{x} : \bigvee_{i=1}^d x_i = \infty\}$ , but absence of mass on lines through  $\infty$  does not necessarily persist for regular variation on other cones [17].

Also, compactification introduces counterintuitive geometric properties. For example, the topology on  $[0, \infty]^d$  can be defined through a homeomorphic map  $[0, \infty]^d \mapsto [0, 1]^d$ , such as

$$z = (z_1, z_2, \dots, z_d) \mapsto \left( \frac{z_1}{1 + z_1}, \dots, \frac{z_d}{1 + z_d} \right).$$

Restrict attention to  $d = 2$ , and consider two parallel lines in  $[0, \infty]^2$  with the same positive and finite slope. These lines both converge to the same point  $(\infty, \infty)$  and, therefore, in the compactified space, these two parallel lines are not bounded away from each other. Interestingly, this is not the case if the lines are horizontal or vertical.

To see the impact that parallel lines not being bounded away from each other can have, recall one of the motivational examples from Section 1 with  $d = 2$ , where the limit measure  $\mu(\cdot)$  in (1.1) is concentrated on the diagonal  $\text{DIAG} := \{(z_1, z_2) \in \mathbb{E} : z_1 = z_2\}$  and we need to approximate the tail probability  $P[|Z_1 - Z_2| > w]$  for a large threshold  $w > 0$ . Of course, if we use multivariate regular variation as in (1.1) to approximate  $P[|Z_1 - Z_2| > w]$ , we approximate  $P[|Z_1 - Z_2| > w]$  as 0. If  $P[Z_1 = Z_2] < 1$ , this approximation is crude. Following the usual definition of HRV, we remove the diagonal  $\text{DIAG}$  and define regular variation on the subcone  $(\text{DIAG})^c := \{(z_1, z_2) \in \mathbb{E} : z_1 \neq z_2\}$ . Since we seek to approximate  $P[|Z_1 - Z_2| > w]$ , we are interested in the set  $A_{>w} := \{(z_1, z_2) \in \mathbb{E} : |z_1 - z_2| > w\}$ . If we define HRV on the subcone  $(\text{DIAG})^c$  as an asymptotic property using vague convergence, we need the set  $A_{>w}$  to be relatively compact in the subcone  $(\text{DIAG})^c$ . However, if the subcone  $(\text{DIAG})^c$  is endowed with the relative topology from the topology on  $[0, \infty]^2$ ,  $A_{>w}$  is not relatively compact since the boundaries of  $A_{>w}$  are the two parallel lines  $\{(z_1, z_2) \in \mathbb{E} : z_1 - z_2 = w\}$  and  $\{(z_1, z_2) \in \mathbb{E} : z_1 - z_2 = -w\}$ , which are both parallel to the diagonal  $\text{DIAG}$ . In the topology of  $[0, \infty]^d$ , the boundaries of the set  $A_{>w}$  are not bounded away from the diagonal  $\text{DIAG}$  and, hence, by Proposition 6.1 of [21, p. 171], the set  $A_{>w}$  is not relatively compact in  $(\text{DIAG})^c$ .

As already observed, horizontal or vertical parallel lines are bounded away from each other in  $[0, \infty]^2$ . If the limit measure  $\mu(\cdot)$  in (1.1) concentrates on the axes, the traditional definition of HRV [20] removes the axes and defines HRV on the cone  $(0, \infty]^2$ . However, risk regions of interest of the form  $(z_1, \infty] \times (z_2, \infty]$  are still relatively compact and we do not encounter a problem as above.

We conclude that a flexible theory of HRV on general cones of  $[0, \infty)^d$  requires considering the possibility that compactification and vague convergence be abandoned. However, without compactification, how do we guarantee that risk sets corresponding to tail events are relatively compact and that their probabilities can be approximated by asymptotic methods? A theory based on  $\mathbb{M}^*$ -convergence of measures sidesteps many difficulties.

### 2.2. $\mathbb{M}^*$ -convergence of measures

We follow the ideas of Hult and Lindskog [10], who removed only a fixed point from a closed set, whereas we remove a closed cone.

As in Section 1.1.2, we fix a closed cone  $\mathbb{C} \subset [0, \infty)^d$  containing  $\mathbf{0}$  and  $\mathbb{F} \subset \mathbb{C}$  is a closed cone in  $\mathbb{C}$  containing  $\mathbf{0}$ . Set  $\mathbb{O} := \mathbb{C} \setminus \mathbb{F}$ , which is an open cone in  $\mathbb{C}$ . Let  $\mathcal{C}$  be the Borel  $\sigma$ -algebra of  $\mathbb{C}$ , and let the  $\sigma$ -algebra in  $\mathbb{O}$  be  $\mathcal{O} = \{B \subset \mathbb{O} : B \in \mathcal{C}\}$ . Denote by  $\mathcal{C}_{\mathbb{F}}$  the set of all bounded, continuous, real-valued functions  $f$  on  $\mathbb{C}$  such that  $f$  vanishes on  $\mathbb{F}^r := \{\mathbf{x} \in \mathbb{C} : \text{dist}(\mathbf{x}, \mathbb{F}) < r\}$  for some  $r > 0$ . The class of Borel measures on  $\mathcal{O}$  that assign finite measure to all  $D \in \mathcal{O}$  which are bounded away from  $\mathbb{F}$  is denoted by  $\mathbb{M}^*(\mathbb{C}, \mathcal{O})$ . Equivalently,  $\mu \in \mathbb{M}^*(\mathbb{C}, \mathcal{O})$  if and only if  $\mu$  is finite on  $\mathbb{C} \setminus \mathbb{F}^r$  for all  $r > 0$ .

**Definition 2.1.** ( $\mathbb{M}^*(\mathbb{C}, \mathbb{O})$ -convergence.) For  $\mu, \mu_n \in \mathbb{M}^*(\mathbb{C}, \mathbb{O})$ ,  $n \geq 1$ ,  $\mu_n$  converges to  $\mu$  in  $\mathbb{M}^*(\mathbb{C}, \mathbb{O})$  if

$$\lim_{n \rightarrow \infty} \mu_n(B) = \mu(B)$$

for all  $B \in \mathcal{O}$  with  $\mu(\partial B) = 0$  and  $B$  bounded away from  $\mathbb{F}$ . We write  $\mu_n \xrightarrow{*} \mu$  in  $\mathbb{M}^*(\mathbb{C}, \mathbb{O})$  as  $n \rightarrow \infty$ .

We can metrize the space  $\mathbb{M}^*(\mathbb{C}, \mathbb{O})$ . One way to metrize the space is to define, for  $\mu, \nu \in \mathbb{M}^*(\mathbb{C}, \mathbb{O})$ ,

$$d_{\mathbb{M}^*}(\mu, \nu) := \int_0^\infty e^{-r} \frac{d_P(\mu^{(r)}, \nu^{(r)})}{1 + d_P(\mu^{(r)}, \nu^{(r)})} dr,$$

where  $\mu^{(r)}$  and  $\nu^{(r)}$  are the restrictions of  $\mu$  and  $\nu$  to  $\mathbb{C} \setminus \mathbb{F}^r$ , and  $d_P$  is the Prohorov metric [19].

Following [10],  $(\mathbb{M}^*(\mathbb{C}, \mathbb{O}), d_{\mathbb{M}^*})$  is a complete separable metric space and the expected analogue of the Portmanteau theorem [1] holds. That is, for  $\mu_n \in \mathbb{M}^*(\mathbb{C}, \mathbb{O})$ ,  $n \geq 0$ , the following statements are equivalent.

- $\mu_n \xrightarrow{*} \mu_0$  in  $\mathbb{M}^*(\mathbb{C}, \mathbb{O})$  as  $n \rightarrow \infty$ .
- For each  $f \in \mathcal{C}_{\mathbb{F}}$ ,  $\int f d\mu_n \rightarrow \int f d\mu_0$  as  $n \rightarrow \infty$ .
- $\limsup_{n \rightarrow \infty} \mu_n(A) \leq \mu_0(A)$  and  $\liminf_{n \rightarrow \infty} \mu_n(G) \geq \mu_0(G)$  for all closed sets  $A \in \mathcal{O}$  and open sets  $G \in \mathcal{O}$  such that  $\overline{G} \cap \mathbb{F} = \emptyset$ .

### 3. Regular and hidden regular variation on cones

We define regular variation on a nested sequence of cones, where each cone is a subset of the previous one. Each cone in the sequence possesses a different regular variation, which remains hidden while studying regular variation on the bigger cones in the sequence.

#### 3.1. Regular variation

We use the concepts of Section 2 to define regular variation.

**Definition 3.1.** Suppose that  $\mathbb{F} \subset \mathbb{C} \subset [0, \infty)^d$ , and that  $\mathbb{F}$  and  $\mathbb{C}$  are closed cones containing  $\mathbf{0}$ . A random vector  $\mathbf{Z} \in \mathbb{C}$  has a distribution with a regularly varying tail on  $\mathbb{O} = \mathbb{C} \setminus \mathbb{F}$  if there exist a function  $b(t) \uparrow \infty$  and a nonzero measure  $\nu(\cdot) \in \mathbb{M}^*(\mathbb{C}, \mathbb{O})$  such that, as  $t \rightarrow \infty$ ,

$$t \mathbb{P} \left[ \frac{\mathbf{Z}}{b(t)} \in \cdot \right] \xrightarrow{*} \nu(\cdot) \quad \text{in } \mathbb{M}^*(\mathbb{C}, \mathbb{O}). \tag{3.1}$$

When there is no danger of confusion, we sometimes use the notation  $\mathbb{M}^*(\mathbb{O})$  to mean  $\mathbb{M}^*(\mathbb{C}, \mathbb{O})$  and say (in an abuse of terminology) that the distribution is regularly varying on  $\mathbb{O}$ .

Definition 3.1 implies that there exists  $\alpha > 0$  such that  $b(\cdot) \in \text{RV}_{1/\alpha}$  and that  $\nu$  has the scaling property

$$\nu(c \cdot) = c^{-\alpha} \nu(\cdot), \quad c > 0. \tag{3.2}$$

This can be derived as in [10, Theorem 3.1]. We define standard multivariate regular variation, HRV, and the conditional extreme value model in terms of Definition 3.1, and attempt to relate each to the way these ideas were first proposed on modifications of compactified spaces.

Some examples include the following.

1. Let  $\mathbb{C} = [0, \infty)^d$  and  $\mathbb{F} = \{\mathbf{0}\}$ . Then  $\mathbb{D} = \mathbb{O} = [0, \infty)^d \setminus \{\mathbf{0}\}$ . Regular variation on  $\mathbb{D}$  is equivalent to regular variation defined in (1.1) on  $\mathbb{E}$ . The definition in (1.1) precludes  $\mu$  having mass on the lines through  $\infty$ . See Appendix A.1.

2. Let  $\mathbb{C} = [0, \infty)^2$  ( $d = 2$ ) and  $\mathbb{F} = \{(x, 0) : x \geq 0\} \cup \{(0, y), y \geq 0\}$ . Then  $\mathbb{O} = (0, \infty)^2$ , the first quadrant with both the  $x$  and  $y$  axes removed. This is the restriction to  $[0, \infty)^2$  of the cone used in the definition of HRV in [20]. For  $d > 2$ , other examples of  $\mathbb{F}$  are given in [17] and in Subsection A.2 we provide a comparison between the regular variation defined in (3.1) on  $\mathbb{D}^{(l)} = [0, \infty)^d \setminus \{\mathbf{x} \in [0, \infty)^d : x_{(l)} > 0\}$  and the regular variation defined in (1.1) on  $\mathbb{E}^{(l)} = [0, \infty)^d \setminus \{\mathbf{x} \in \mathbb{E} : x_{(l)} > 0\}$ , where, recall,  $x_{(l)}$  is the  $l$ th largest component of  $\mathbf{x}$ . The two notions are equivalent provided there is no mass on  $\mathbb{E}^{(l)} \setminus \mathbb{D}^{(l)}$ .
3. Let  $\mathbb{C} = [0, \infty)^2$  ( $d = 2$ ) and  $\mathbb{F} = \{(x, 0) : x \geq 0\}$ . Then  $\mathbb{O} = \{(x, y) \in [0, \infty)^2 : y > 0\} = \mathbb{D}_\cap$ , the first quadrant with the  $x$ -axis removed. This is the restriction to  $[0, \infty)^2$  of the cone used in the definition of the conditional extreme value model [4], [8], [9].
4. Let  $\mathbb{C} = [0, \infty)^2$  ( $d = 2$ ) and  $\mathbb{F} = \{(x, x) : x \geq 0\}$ . Then  $\mathbb{O}$  is the first quadrant with the diagonal removed. This example is suitable for discussing asymptotic full dependence (see [21, p. 195] and [22, p. 294]) and is considered in Example 5.2.

### 3.2. Spectral measures, unit spheres, and semiparametric representations

Regular variation on  $\mathbb{E}$  using the vague convergence definition as in (1.1) allows the polar coordinate transformation  $\mathbf{x} \mapsto (\|\mathbf{x}\|, \mathbf{x}/\|\mathbf{x}\|)$ . Assuming that  $b(\cdot) \in \text{RV}_{1/\alpha}$ , the limit measure has the scaling property, and when this is expressed in polar coordinates, we obtain the following alternative version of (1.1):

$$t \mathbb{P} \left[ \frac{\|\mathbf{Z}\|}{b(t)} \in dx, \frac{\mathbf{Z}}{\|\mathbf{Z}\|} \in d\mathbf{a} \right] \xrightarrow{v} \alpha x^{-\alpha-1} dx \times S^*(d\mathbf{a}).$$

Here  $S^*$  is a finite measure on  $\partial\mathfrak{N} = \{\mathbf{a} \in \mathbb{E} : \|\mathbf{a}\| = 1\}$ , the unit sphere. Fixing  $S^*(\partial\mathfrak{N}) = c$ , we define  $S(\cdot) = S^*(\cdot)/c$ , which becomes a probability measure on  $\partial\mathfrak{N}$  called the spectral or angular measure. So, in polar coordinates, the limit measure  $\mu$  in (1.1) has a semiparametric product structure depending on the parameter  $\alpha$  and the measure  $S$ .

In  $\mathbb{E}$ , the unit sphere  $\partial\mathfrak{N} = \{\mathbf{x} \in \mathbb{E} : \|\mathbf{x}\| = 1\} = \{\mathbf{x} \in \mathbb{E} : \text{dist}(\mathbf{x}, \mathbf{0}) = 1\}$  is compact. However, this may no longer be true when moving to other subcones. For instance, in  $(0, \infty)^2$  the usual unit sphere is not relatively compact. While the polar coordinate transformation still allows this semiparametric representation for other cones, the analogue of  $S$  is no longer necessarily finite and this is a problem for inference. We explain next how to use a change of coordinates different from the polar coordinate transformation to always produce a finite spectral measure analogue. Maulik and Resnick [15] and Mitra and Resnick [17] considered alternatives to the polar coordinate transformation that twist limit measures into a semiparametric form.

Proceed using the context of Definition 3.1. Assume that  $\mathbb{C}$ ,  $\mathbb{F}$ , and  $\mathbb{O}$  are defined as in Definition 3.1. Define  $\mathfrak{N}_\mathbb{O} = \{\mathbf{x} \in \mathbb{O} : \text{dist}(\mathbf{x}, \mathbb{F}) \geq 1\}$ , so  $\mathfrak{N}_\mathbb{O}$  is a subset of  $\mathbb{O}$  bounded away from  $\mathbb{F}$  and  $\bigcup_{\{\theta > 0\}} \theta \mathfrak{N}_\mathbb{O} = \mathbb{O}$ . From Definition 3.1, we have  $0 < \nu(\mathfrak{N}_\mathbb{O}) < \infty$ . The scaling property (3.2) implies that the scaling function  $b(t)$  in (3.1) can be chosen so that  $\nu(\mathfrak{N}_\mathbb{O}) = 1$ . Define the related set  $\partial\mathfrak{N}_\mathbb{O} = \{\mathbf{x} \in \mathbb{O} : \text{dist}(\mathbf{x}, \mathbb{F}) = 1\}$ . Some examples are as follows.

- $\mathbb{C} = [0, \infty)^d$ ,  $\mathbb{F} = \{\mathbf{0}\}$ , and  $\text{dist}(\mathbf{x}, \mathbb{F}) = \|\mathbf{x}\|$ . Then  $\partial\mathfrak{N}_\mathbb{O} = \{\mathbf{x} : \|\mathbf{x}\| = 1\}$ .
- $d = 2$ ,  $\|\mathbf{x}\| = x_1 \vee x_2$ ,  $\mathbb{C} = [0, \infty)^2$ ,  $\mathbb{F} = \{(x, 0) : x \geq 0\} \cup \{(0, y), y \geq 0\}$ , and  $\mathbb{O} = (0, \infty)^2$ . Then  $\partial\mathfrak{N}_\mathbb{O} = \{\mathbf{x} : \text{dist}(\mathbf{x}, \mathbb{F}) = 1\} = \{\mathbf{x} : x_1 \wedge x_2 = 1\}$ .

- $d = 2$ ,  $\|\mathbf{x}\| = x_1 \vee x_2$ ,  $\mathbb{C} = [0, \infty)^2$ , and  $\mathbb{F} = \{(x, 0) : x \geq 0\}$ . Then  $\partial\mathfrak{N}_{\mathbb{O}} = \{\mathbf{x} : \text{dist}(\mathbf{x}, \mathbb{F}) = 1\} = \{(x, y) : x \geq 0, y = 1\}$ .

We transform to an appropriate coordinate system in which the limit measure  $\nu$  in (3.1) is a product of two components: a one-dimensional Pareto measure and a probability measure defined on  $\partial\mathfrak{N}_{\mathbb{O}}$  called the hidden spectral or angular measure. To this end, we first note the following two properties of the distance function  $\text{dist}(\cdot, \mathbb{F})$ .

- Since  $\mathbb{F}$  is a closed subset of  $\mathbb{C}$ ,  $\text{dist}(\mathbf{x}, \mathbb{F}) > 0$  for all  $\mathbf{x} \in \mathbb{O}$  (otherwise,  $\mathbf{x} \in \mathbb{F}$ ).
- Since  $\mathbb{F}$  is a cone,  $\theta \cdot \mathbb{F} = \mathbb{F}$  for  $\theta > 0$ . Hence,  $\text{dist}(\theta\mathbf{x}, \mathbb{F}) = \text{dist}(\theta\mathbf{x}, \theta\mathbb{F}) = \theta \text{dist}(\mathbf{x}, \mathbb{F})$ , that is,  $\text{dist}(\cdot, \mathbb{F})$  is homogeneous of order 1.

A lemma is necessary for the decomposition of the limit measure  $\nu(\cdot)$ . For a set  $A \in [0, \infty)^d$ , we set  $(A)_1 = \{x_1 : \mathbf{x} \in A\}$ .

**Lemma 3.1.** *Suppose that  $h : \mathbb{O} \mapsto (0, \infty) \times \partial\mathfrak{N}_{\mathbb{O}}$  is a continuous bijection such that*

- *for every measurable  $A \subset (0, \infty) \times \partial\mathfrak{N}_{\mathbb{O}}$  with  $\overline{(A)_1} \cap \{0\} = \emptyset$ ,  $h^{-1}(A)$  is bounded away from  $\mathbb{F}$ ,*
- *for every measurable  $B \subset \mathbb{O}$  with  $B$  bounded away from  $\mathbb{F}$ ,  $\overline{(h(B))_1} \cap \{0\} = \emptyset$ .*

*Then the following statements are equivalent for measures  $\mu_t, \mu \in \mathbb{M}^*(\mathbb{C}, \mathbb{O})$ .*

(i) *As  $t \rightarrow \infty$ ,*

$$\mu_t(\cdot) \xrightarrow{*} \mu(\cdot) \text{ in } \mathbb{M}^*(\mathbb{C}, \mathbb{O}).$$

(ii) *For all measurable  $A \subset (0, \infty) \times \partial\mathfrak{N}_{\mathbb{O}}$  such that  $\overline{(A)_1} \cap \{0\} = \emptyset$  and  $\mu \circ h^{-1}(\partial A) = 0$ ,*

$$\mu_t \circ h^{-1}(A) \rightarrow \mu \circ h^{-1}(A).$$

*Proof.* The proof follows the proof of Theorem 2.5 of [10].

Applying Lemma 3.1 with  $h : \mathbf{x} \mapsto (\text{dist}(\mathbf{x}, \mathbb{F}), \mathbf{x}/\text{dist}(\mathbf{x}, \mathbb{F}))$ , we are able to decompose  $\nu$ .

**Proposition 3.1.** *Regular variation on  $\mathbb{O}$  as given in (3.1) is equivalent to*

$$t \mathbb{P} \left[ \left( \frac{\text{dist}(\mathbf{Z}, \mathbb{F})}{b(t)}, \frac{\mathbf{Z}}{\text{dist}(\mathbf{Z}, \mathbb{F})} \right) \in A \right] \rightarrow c\nu_{(\alpha)} \times S_{\mathbb{O}}(A) \tag{3.3}$$

*for all measurable  $A \subset (0, \infty) \times \partial\mathfrak{N}_{\mathbb{O}}$  such that  $\overline{(A)_1} \cap \{0\} = \emptyset$  and  $\nu \circ h^{-1}(\partial A) = 0$ , where  $c > 0$ ,  $S_{\mathbb{O}}(\cdot)$  is a probability measure on  $\partial\mathfrak{N}_{\mathbb{O}}$ , and  $\nu_{(\alpha)}(\cdot)$  is the Pareto measure given by  $\nu_{(\alpha)}((x, \infty)) = x^{-\alpha}$  for  $x > 0$ . Call  $S_{\mathbb{O}}(\cdot)$  the spectral measure on  $\mathbb{O}$ ; it is related to  $\nu(\cdot)$  via*

$$S_{\mathbb{O}}(\Lambda) = \frac{\nu(\{\mathbf{x} \in \mathbb{O} : \text{dist}(\mathbf{x}, \mathbb{F}) \geq 1, \mathbf{x}/\text{dist}(\mathbf{x}, \mathbb{F}) \in \Lambda\})}{\nu(\mathbf{x} \in \mathbb{O} : \text{dist}(\mathbf{x}, \mathbb{F}) \geq 1)}.$$

Since  $(0, \infty) \times \partial\mathfrak{N}_{\mathbb{O}}$  is not a cone, we have not phrased the convergence in (3.3) as  $\mathbb{M}^*$ -convergence, as in (3.1). To do so requires reworking the convergence theory in [10].

**Corollary 3.1.** *The convergence in (3.1) can be equivalently given by the following two conditions.*

- (i) *The distribution of  $\text{dist}(\mathbf{Z}, \mathbb{F})$  is regularly varying on  $(0, \infty)$  following Definition 3.1 with  $\mathbb{C} = [0, \infty)$ .*

(ii) The conditional distribution of  $\mathbf{Z} / \text{dist}(\mathbf{Z}, \mathbb{F})$  given  $\text{dist}(\mathbf{Z}, \mathbb{F}) > t$  converges weakly, i.e.

$$P\left[\frac{\mathbf{Z}}{\text{dist}(\mathbf{Z}, \mathbb{F})} \in \cdot \mid \text{dist}(\mathbf{Z}, \mathbb{F}) > t\right] \Rightarrow S_{\mathbb{O}}(\cdot) \text{ as } t \rightarrow \infty.$$

Some remarks about Proposition 3.1 are in order.

**Remarks 3.1.** (i) *The role of the distance function.* Proposition 3.1 emphasizes that the spectral probability measure  $S_{\mathbb{O}}(\cdot)$  is dependent on the choice of distance function  $\text{dist}(\cdot, \cdot)$ . Corollary 3.1 allows us to use the distance function  $\text{dist}(\cdot, \cdot)$  to detect regular variation on  $\mathbb{O}$ ; see Section 6. However, extending the distance function to a compactified space such as  $[0, \infty]^d$  is difficult and this provides another reason why we deviated from the standard discussion of regular variation using compactified spaces and vague convergence.

(ii) *Connections to prior treatments.*

- (a) Proposition 3.1 of [17] decomposes the limit measure  $\mu^{(l)}(\cdot)$  on  $\mathbb{E}^{(l)}$  (see (A.1) below) by applying the transformation  $T: \mathbf{x} \mapsto (x_{(l)}, \mathbf{x}/x_{(l)})$ , where  $x_{(l)}$  is the  $l$ th largest component of  $\mathbf{x}$ . If we choose  $\mathbb{C} = [0, \infty)^d$ ,  $\mathbb{F} = \{\mathbf{x} \in \mathbb{C}: x_{(l)} = 0\}$ , and  $\mathbb{O} = \{\mathbf{x} \in \mathbb{C}: x_{(l)} > 0\}$ , and choose the  $L_\infty$ -norm when defining  $\text{dist}(\cdot, \cdot)$ , then  $\text{dist}(\mathbf{x}, \mathbb{F}) = x_{(l)}$  and our Proposition 3.1 gives a version of Proposition 3.1 of [17].
- (b) When considering regular variation on the cone  $\mathbb{E}_\Pi := [0, \infty] \times (0, \infty]$ , Heffernan and Resnick [8, Proposition 4] gave a decomposition of their limit measure  $\mu_*(\cdot)$  by applying the transformation  $T: (x, y) \mapsto (y, x/y)$ . If we choose  $\mathbb{C} = [0, \infty)^2$ ,  $\mathbb{F} = [0, \infty) \times \{0\}$ , and  $\mathbb{O} = [0, \infty) \times (0, \infty) =: \mathbb{D}_\Pi$ , and define  $\text{dist}(\cdot, \cdot)$  using the  $L_\infty$ -norm, then  $d((x, y), \mathbb{F}) = y$  and our Proposition 3.1 connects with [8, Proposition 4].
- (c) Proposition 3.1 also relates to the usual polar coordinate characterization of multivariate regular variation on  $\mathbb{E}$  as in [21, p. 173]. Set  $\mathbb{F} = \{\mathbf{0}\}$ , which is a closed cone in  $\mathbb{C} = [0, \infty)^d$ , and  $\text{dist}(\mathbf{x}, \{\mathbf{0}\}) = \|\mathbf{x}\|$ . See also Proposition A.1.

### 3.3. Hidden regular variation

As in Definition 3.1, consider  $\mathbb{F} \subset \mathbb{C} \subset [0, \infty)^d$  with  $\mathbb{F}$  and  $\mathbb{C}$  closed cones containing  $\{\mathbf{0}\}$ . Suppose that  $\mathbb{F}_1$  is another subset of  $\mathbb{C}$  that is a closed cone containing  $\{\mathbf{0}\}$ . Then  $\mathbb{F} \cup \mathbb{F}_1$  is also a closed cone containing  $\{\mathbf{0}\}$ . Set  $\mathbb{O} = \mathbb{C} \setminus \mathbb{F}$  and  $\mathbb{O}_1 = \mathbb{C} \setminus (\mathbb{F} \cup \mathbb{F}_1)$ .

**Definition 3.2.** The distribution of a random vector  $\mathbf{Z} \in \mathbb{C}$  that is regularly varying on  $\mathbb{O}$  with scaling function  $b(t)$  in (3.1) possesses HRV on  $\mathbb{O}_1$  if

- the distribution of  $\mathbf{Z}$  is also regularly varying on  $\mathbb{O}_1$  with scaling function  $b_1(t)$  and limit measure  $\nu_1$ , and
- $b(t)/b_1(t) \rightarrow \infty$  as  $t \rightarrow \infty$ .

Observe that the condition  $b(t)/b_1(t) \rightarrow \infty$  implies that  $\nu$  puts zero mass on  $\mathbb{O}_1$  [20]. From Definition 3.2 and (3.2), it follows that there exists  $\alpha_1 \geq \alpha$  such that  $b_1(\cdot) \in \text{RV}_{1/\alpha_1}$ , and on  $\mathbb{O}_1$ , the limit measure  $\nu_1(\cdot)$  in (3.1) satisfies the scaling property

$$\nu_1(c \cdot) = c^{-\alpha_1} \nu_1(\cdot), \quad c > 0.$$

**Example.** Let  $\mathbf{Z} = (Z_1, Z_2)$  be independent and identically distributed (i.i.d.) unit Pareto random variables. Then the distribution of  $\mathbf{Z}$  is regularly varying on  $\mathbb{D}$  with  $b(t) = t$  and

possesses HRV on  $(0, \infty)^2$  with  $b_1(t) = \sqrt{t}$ . In general, if  $\mathbf{Z} = (Z_1, Z_2)$  has regular variation on  $\mathbb{D} = [0, \infty)^2 \setminus \{\mathbf{0}\}$  with asymptotic independence, the regular variation on  $\mathbb{D}$  estimates joint tail probabilities  $P[Z_1 > x, Z_2 > y]$  for large thresholds  $x, y > 0$  as 0, while HRV offers the possibility of nonzero estimates. Other examples are considered in Section 5.

Some remarks about Definition 3.2 are in order.

**Remarks 3.2.** (i) There is flexibility in choosing  $\mathbb{O}_1$  and this flexibility is useful for defining HRV on a sequence of cones. Cones can be chosen based on the risk regions whose probabilities are required. For example, if  $d = 2$ , we choose the cones  $\mathbb{D}$  and  $\mathbb{D}^{(2)} := \{\mathbf{z} \in [0, \infty)^2 : z_1 \wedge z_2 > 0\}$  if we need the probability that components of the risk vector simultaneously exceed thresholds.

(ii) *Differences with existing notions of HRV.* Previous considerations of HRV relied on vague convergence and compactification and were applied to specific choices of cones. In [7], [15], and [20] HRV on  $(0, \infty)^2$  was considered and in [17] the cone  $\mathbb{E}^{(d)} = \{\mathbf{x} \in [0, \infty)^d : x_{(l)} > 0\}$  was considered. The choice of these specific cones may not provide sufficient flexibility and generality. For example, to estimate  $P[|X - Y| > x]$  when asymptotic full dependence is present, such cones considered previously are of no help. See Example 5.2 below.

3.3.1. *Where to seek HRV?* Suppose that the distribution of  $\mathbf{Z}$  is regularly varying on  $\mathbb{O}$  and that the limit measure  $\nu$  in (3.1) gives zero mass to a subset  $R$  of  $\mathbb{O}$ . Using the asymptotic property of regular variation to estimate  $P[\mathbf{Z} \in tR]$  for large  $t$  means that such an estimate is 0. So we seek another regular variation on a subset of  $R$  which is of lower order.

When seeking HRV our focus is thus on subsets of  $\mathbb{O}$  where the limit measure  $\nu(\cdot)$  gives zero mass. A systematic way to find HRV is facilitated by the following simple proposition.

**Proposition 3.2.** *In Definition 3.1, the support of the limit measure  $\nu$  is a closed cone  $F_\nu \subset \mathbb{C}$  containing  $\mathbf{0}$ .*

*Proof.* Let  $\text{supp}(\nu)$  denote the support of  $\nu$ . By definition,  $\text{supp}(\nu)$  is closed. Let  $\mathbf{x} \in \text{supp}(\nu)$ . We show that, for  $t > 0$ ,  $t\mathbf{x} \in \text{supp}(\nu)$ . For small  $\delta$ , by (3.2),

$$\begin{aligned} \nu((t\mathbf{x} - \delta\mathbf{1}, t\mathbf{x} + \delta\mathbf{1}) \cap \mathbb{C}) &= \nu\left(t\left(\left(\mathbf{x} - \frac{\delta}{t}\mathbf{1}, \mathbf{x} + \frac{\delta}{t}\mathbf{1}\right) \cap \mathbb{C}\right)\right) \\ &= t^{-\alpha}\nu\left(\left(\mathbf{x} - \frac{\delta}{t}\mathbf{1}, \mathbf{x} + \frac{\delta}{t}\mathbf{1}\right) \cap \mathbb{C}\right) \\ &> 0, \end{aligned}$$

since  $\mathbf{x} \in \text{supp}(\nu)$ .

So,  $F_\nu = \text{supp}(\nu)$  is a union of rays emanating from the origin. It is also true that

$$\text{supp}(\nu) = \{t \cdot \text{supp}(S_{\mathbb{O}}), t \geq 0\}.$$

When seeking HRV on a cone smaller than  $\mathbb{O}$ , we conclude that  $\mathbb{O}_\nu := \mathbb{C} \setminus (F \cup F_\nu)$  is the largest possible subcone of  $\mathbb{O}$  where we might find a different regular variation. In practice, guided by the type of risk region whose probability is required, we find a closed cone  $F_1 \subset \mathbb{C}$  containing  $\mathbf{0}$  such that  $F_1 \supset F_\nu$ , set  $\mathbb{O}_1 = \mathbb{C} \setminus (F \cup F_1)$ , and then seek regular variation on  $\mathbb{O}_1$ . Possibly, but not necessarily,  $F_1 = \text{supp}(\nu)$ . Examples are given in [17] and Section 5.

3.3.2. *Regular variation on a sequence of cones.* Having found regular variation on  $\mathbb{O}$  with HRV on  $\mathbb{O}_1$ , we ask: should we stop here? There might be a subcone  $\mathbb{O}_2$  of  $\mathbb{O}_1$ , where  $\nu_1(\cdot)$  gives zero mass and, hence, there might exist a different regular variation on  $\mathbb{O}_2$ .

To proceed further, remove the support of  $\nu_1$  from  $\mathbb{O}_1$  and let  $\mathbb{O}_{\nu_1} := \mathbb{O}_1 \setminus \text{supp}(\nu_1)$  be the largest possible subcone of  $\mathbb{O}_1$  where we might find a different regular variation. So we choose  $\mathbb{F}_2 \supset \text{supp}(\nu_1)$ , set  $\mathbb{O}_2 = \mathbb{C} \setminus (\mathbb{F} \cup \mathbb{F}_1 \cup \mathbb{F}_2)$ , and seek regular variation on  $\mathbb{O}_2$  with scaling function  $b_2(t)$  such that  $b_1(t)/b_2(t) \rightarrow \infty$  as  $t \rightarrow \infty$ . This last condition guarantees that the regular variation on  $\mathbb{O}_2$  is of lower order than the regular variation on either  $\mathbb{O}$  or  $\mathbb{O}_1$  and hidden from both higher-order regular variations. This process of discovery may be continued as long as on each new cone regular variation is found. Example 5.3 below shows this discovery process may lead to an infinite sequence of cones.

From our definition of HRV, at each stage of the discovery process we have some flexibility in choosing the next cone where we seek HRV. Example 5.3 below shows that it may be impractical to analyze HRV on every possible cone as the discovery process may lead to an infinite sequence of cones. A more practical approach is to decide on a particular finite sequence of cones based on the risk regions of interest; see Remarks 3.2(i). For example, if we are interested in estimating joint tail probabilities, we might consider only the sequence of cones

$$\begin{aligned} \mathbb{D} &= [0, \infty)^d \setminus \{\mathbf{0}\} \supset \mathbb{D}^{(2)} \\ &= \{\mathbf{x} \in [0, \infty)^d : x_{(2)} > 0\} \supset \dots \supset \mathbb{D}^{(d)} \\ &= \{\mathbf{x} \in [0, \infty)^d : x_{(d)} > 0\}; \end{aligned}$$

cf. [17].

#### 4. Remarks on other models of multivariate regular variation

Despite the fact that most common examples in heavy tail analysis start by analyzing convergence on the cone  $[0, \infty)^d \setminus \{\mathbf{0}\}$ , this need not always be the case. For example, the standard case of the conditional extreme value (CEV) model [4], [8] is regular variation (3.1) with  $b(t) = t$  on the cone  $\mathbb{D}_\Gamma := \mathbb{O} = [0, \infty) \times (0, \infty)$  with  $\mathbb{C} = [0, \infty)^2$  and  $\mathbb{F} = [0, \infty) \times \{0\}$ .

##### 4.1. The CEV model

The CEV model, introduced in [9], is an alternative to classical multivariate extreme value theory (MEVT). In contrast with classical MEVT which implies that all marginals are in a maximal domain of attraction, in the CEV model only a particular subset of the random vector is assumed to be in a maximal domain of attraction. For convenience, we restrict attention to  $d = 2$ .

The CEV model as formulated in [4] and [8] allows variables to be centered as well as scaled. To make comparisons with models of regular variation on the first quadrant easy, we recall the vague convergence definition using only scaling functions.

**Definition 4.1.** Suppose that  $\mathbf{Z} := (Z_1, Z_2) \in \mathbb{R}_+^2$  is a random vector, and that there exist functions  $a_1(t)$  and  $a_2(t) > 0$  and a nonnull Radon measure  $\mu$  on Borel subsets of  $\mathbb{E}_\Gamma := [0, \infty) \times (0, \infty]$  such that, in  $\mathbb{M}_+(\mathbb{E}_\Gamma)$ ,

$$t \mathbb{P} \left[ \left( \frac{Z_1}{a_1(t)}, \frac{Z_2}{a_2(t)} \right) \in \cdot \right] \xrightarrow{v} \mu(\cdot). \tag{4.1}$$

Additionally, assume that  $\mu$  satisfies the following nondegeneracy conditions:

- (i)  $\mu([0, x] \times (y, \infty))$  is a nondegenerate distribution in  $x$ ,
- (ii)  $\mu([0, x] \times (y, \infty)) < \infty$ .

Also, assume that

- (iii)  $H(x) := \mu([0, x] \times (1, \infty))$  is a probability distribution.

Then  $Z$  satisfies a CEV model and we write  $Z \in \text{CEV}(a_1, a_2)$ .

The general CEV model, provided the limit measure is not a product, can be standardized to have standard regular variation on the cone  $\mathbb{E}_\Gamma$  [4, p. 236]. Following the theme of examples 1 and 2 at the beginning of Section 3, if no mass exists on the lines through  $\infty$ , and  $a_1 = a_2$ , then the vague convergence definition of the CEV model on  $\mathbb{E}_\Gamma$  is the same as the  $\mathbb{M}^*$  definition on regular variation  $\mathbb{D}_\Gamma$ .

The issue of mass on the lines through  $\infty$  is significant since if mass on these lines is allowed, there is a statistical identifiability problem in the sense that in  $\mathbb{E}_\Gamma$  it is possible to have two different limits in (4.1) under two different normalizations: one limit puts mass on the lines through  $\infty$  and the other does not. Restricting to  $\mathbb{D}_\Gamma$  resolves the identifiability problem because in this space, limits are unique. See Example 5.4 below.

**4.2. Nonstandard regular variation**

Standard multivariate regular variation on  $\mathbb{E}$  requires the same normalizing function to scale all components and is a convenient starting place for theory, but unrealistic for many applications as it makes all one-dimensional marginal distributions tail equivalent. Nonstandard regular variation [21, Section 6.5.6] allows different normalizing functions for vector components and, hence, permits each marginal distribution tail to have a different tail index. When the components of the risk vector have different tail indices, nonstandard regular variation is sensitive to the different tail strengths. On  $\mathbb{E}$  or  $\mathbb{D}$ , nonstandard regular variation takes the form

$$t \mathbb{P} \left[ \left( \frac{Z_i}{a_i(t)}, i = 1, \dots, d \right) \in \cdot \right] \rightarrow \nu(\cdot)$$

for scaling functions  $a_i(t) \uparrow \infty, i = 1, \dots, d$ , where convergence is vague for  $\mathbb{E}$  and in  $\mathbb{M}^*$  for  $\mathbb{D}$ . If convergence is in  $\mathbb{E}$  and there is no mass on the lines through  $\infty$ , the difference between convergence in  $\mathbb{E}$  and  $\mathbb{D}$  vanishes.

In cases where  $a_i(t)/a_{i+1}(t) \rightarrow 0$ , it is sometimes possible to compare the information from nonstandard regular variation with what can be obtained from HRV. Sometimes HRV provides more detailed information. Consider the following example.

**Example 4.1.** Suppose that  $X_1, X_2$ , and  $X_3$  are independent random variables, where  $X_1$  is Pareto(1),  $X_2$  is Pareto(3), and  $X_3$  is Pareto(4). Furthermore, assume that  $B$  is a Bernoulli( $\frac{1}{2}$ ) random variable independent of  $X_1, X_2$ , and  $X_3$ . Define

$$Z = (Z_1, Z_2) := B(X_1, X_3) + (1 - B)(X_2, X_2).$$

Nonstandard regular variation on  $\mathbb{E}$  or  $\mathbb{D}$  is given by

$$t \mathbb{P} \left[ \left( \frac{Z_1}{t}, \frac{Z_2}{t^{1/3}} \right) \in dx dy \right] \rightarrow \frac{1}{2} x^{-2} dx \varepsilon_0(dy) + \frac{1}{2} \varepsilon_0(dx) 3y^{-4} dy,$$

where  $\varepsilon_0$  indicates the point measure at 0 and the limit measure concentrates on the two axes.

The standardized version on  $\mathbb{E}$  or  $\mathbb{D}$  is

$$t \mathbb{P} \left[ \left( \frac{Z_1}{t}, \frac{Z_2^3}{t} \right) \in dx dy \right] \rightarrow \frac{1}{2} x^{-2} dx \varepsilon_0(dy) + \frac{1}{2} \varepsilon_0(dx) y^{-2} dy.$$

Now HRV can also be sought and we obtain

$$t \mathbb{P} \left[ \left( \frac{Z_1}{t^{3/7}}, \frac{(Z_2)^3}{t^{3/7}} \right) \in dx dy \right] \rightarrow \frac{1}{2} x^{-2} dx 4y^{-5} dy$$

on the space  $\mathbb{E}^{(2)}$ . Note that this form of regular variation completely ignores the existence of the completely dependent component of  $\mathbf{Z}$  given by  $(X_2, X_2)$ . Alternatively, if we pursue regular variation and HRV on a sequence of cones as defined in this paper then we observe the following convergences as  $t \rightarrow \infty$ .

- On  $\mathbb{D}$ , we have

$$t \mathbb{P} \left[ \frac{\mathbf{Z}}{t} \in dx dy \right] \rightarrow \frac{1}{2} x^{-2} dx \varepsilon_0(dy).$$

- On  $\mathbb{D} \setminus \{x\text{-axis}\}$ , we have

$$t \mathbb{P} \left[ \frac{\mathbf{Z}}{t^{1/3}} \in dx dy \right] \rightarrow \frac{3}{2} x^{-4} dx \varepsilon_x(dy).$$

- On  $\mathbb{D} \setminus [\{x\text{-axis}\} \cup \{\text{diagonal}\}]$ , we have

$$t \mathbb{P} \left[ \frac{\mathbf{Z}}{t^{1/4}} \in dx dy \right] \rightarrow \frac{1}{2} \varepsilon_0(dx) 4y^{-5} dy.$$

- Finally, on  $\mathbb{D} \setminus [\{x\text{-axis}\} \cup \{\text{diagonal}\} \cup \{y\text{-axis}\}]$ , we have

$$t \mathbb{P} \left[ \frac{\mathbf{Z}}{t^{1/5}} \in dx dy \right] \rightarrow \frac{1}{2} x^{-2} dx 4y^{-5} dy.$$

This analysis provides more detailed information compared with first standardizing and then seeking HRV on  $(0, \infty)^2$ . However, this is not always the case and sometimes nonstandard regular variation is better suited to explaining the model structure.

**Example 4.2.** Suppose that  $X$  is a standard Pareto random variable and that  $\mathbf{Z} = (X, X^2)$  [21, Example 6.3]. Using the obvious nonstandard scaling for the coordinates leads to

$$t \mathbb{P} \left[ \left( \frac{X}{t}, \frac{X^2}{t^2} \right) \in \cdot \right] \xrightarrow{v} v_{(1)} \circ T^{-1} \quad \text{as } t \rightarrow \infty, \tag{4.2}$$

where  $T: (0, \infty] \rightarrow (0, \infty] \times (0, \infty]$  is defined by  $T(x) = (x, x^2)$  and  $v_{(1)}(dx) = x^{-2} dx$ ,  $x > 0$ . The limit measure concentrates on  $\{(x, x^2): x > 0\}$ . Using the same normalization on both coordinates is not so revealing. With the heavier normalization, we have, on  $\mathbb{M}^*([0, \infty)^2, [0, \infty)^2 \setminus \{\mathbf{0}\})$ ,

$$t \mathbb{P} \left[ \frac{\mathbf{Z}}{t^2} \in \cdot \right] \xrightarrow{*} \varepsilon_0 \times v_{(1)} \quad \text{as } t \rightarrow \infty, \tag{4.3}$$

and with the lighter normalization, for  $x > 0$  and  $y > 0$ ,

$$t \mathbb{P} \left[ \frac{X}{t} > x, \frac{X^2}{t} > y \right] \rightarrow x^{-1} \quad \text{for all } y > 0,$$

so that

$$t \mathbb{P} \left[ \frac{\mathbf{Z}}{t} \in \cdot \right] \xrightarrow{v} \nu_{(1)} \times \varepsilon_\infty \quad \text{on } \mathbb{E}. \tag{4.4}$$

Neither (4.3) nor (4.4) approach the explanatory power of (4.2). Moreover, since our modeling excludes points at  $\infty$ , the convergence in (4.4) is not equivalent to any  $\mathbb{M}^*$ -convergence.

### 5. Examples

The definition of HRV given in this paper changes the definition of convergence, but, more importantly, allows general cones compared with the original definition of HRV [17], [20]; see Remarks 3.2 and (A.2) below. Here we provide examples to illustrate use and highlight subtleties, and give risk sets, whose probabilities can be approximated using our general concept of HRV but not the existing notion.

**Example 5.1.** (*Diversification of risk.*) Suppose that we invest in two financial instruments  $I_1$  and  $I_2$ , and, for a given time horizon, let  $Z_1$  and  $Z_2$  denote future losses associated with each unit of  $I_1$  and  $I_2$ , respectively. Let  $\mathbf{Z} = (Z_1, Z_2)$  be regularly varying on  $\mathbb{O} = \mathbb{D} = [0, \infty)^2 \setminus \{\mathbf{0}\}$  with limit measure  $\nu(\cdot)$ . We earn risk premia of  $\$l_1$  and  $\$l_2$  for investing in each unit of  $I_1$  and  $I_2$ .

Suppose that we invest in  $a_1$  units of  $I_1$  and  $a_2$  units of  $I_2$ . A possible risk measure for this portfolio is  $\mathbb{P}[a_1 Z_1 > x, a_2 Z_2 > y]$  for two large thresholds  $x$  and  $y$ . This risk measure quantifies the tail dependence of  $Z_1$  and  $Z_2$ . The greater the tail dependence between  $Z_1$  and  $Z_2$ , the greater should be our reserve capital requirement so that we guard against the extreme situation that both investments go awry. For this risk measure, the best circumstance is if risk contagion is absent; that is,  $Z_1$  and  $Z_2$  are asymptotically independent so that the measure  $\nu(\cdot)$  is concentrated on the axes [21, p. 192], since then  $\mathbb{P}[a_1 Z_1 > x, a_2 Z_2 > y]$  is estimated to be 0 if HRV is absent and even if HRV exists on the cone  $(0, \infty)^2$  according to Definition 3.2, the estimate of  $\mathbb{P}[a_1 Z_1 > x, a_2 Z_2 > y]$  for large thresholds  $x$  and  $y$  should be small compared to the case where  $Z_1$  and  $Z_2$  are not asymptotically independent.

In place of asymptotic independence, suppose instead that  $\mathbf{Z}$  is regularly varying on  $\mathbb{D}$  as in Definition 3.1 with limit measure  $\nu(\cdot)$ , whose support is

$$\{(u, v) \in \mathbb{D} : 2u \leq v\} \cup \{(u, v) \in \mathbb{D} : u \geq 2v\}$$

so that  $\nu(\cdot)$  puts zero mass on

$$\text{CONE} := \{(u, v) \in \mathbb{D} : 2u > v, 2v > u\} = \bigcup_{x>0, y>0} \{(u, v) : 2u - v > x, 2v - u > y\}.$$

Though this  $\mathbf{Z}$  does not possess asymptotic independence, we can still build a portfolio of two financial instruments that have asymptotically independent risks.

Define two new financial instruments  $W_1 = 2I_1 - I_2$  (buy two units of  $I_1$  and sell a unit of  $I_2$ , assuming such transactions are allowed) and  $W_2 = 2I_2 - I_1$ . The loss associated with  $W_1$  is  $L_{W_1} := 2Z_1 - Z_2$  and the loss associated with  $W_2$  is  $L_{W_2} := 2Z_2 - Z_1$ . We earn the same

risk premium  $a_1l_1 + a_2l_2$  in the following two cases.

Case 1: invest in  $a_1$  units of  $I_1$  and  $a_2$  units of  $I_2$ .

Case 2: invest in  $c_1 = (2a_1 + a_2)/3$  units of  $W_1$  and  $c_2 = (a_1 + 2a_2)/3$  units of  $W_2$ .

A measure of risk of the portfolio in case 1 is  $P[a_1Z_1 > x, a_2Z_2 > y]$ . Since asymptotic independence is absent,

$$t P \left[ \frac{a_1 Z_1}{b(t)} > x, \frac{a_2 Z_2}{b(t)} > y \right] \rightarrow \nu((x, y), \infty) > 0 \quad \text{as } t \rightarrow \infty,$$

and we expect the risk probability to be not too small. However, the risk measure for case 2 should be rather small. Let  $T(u, v) = (2u - v, 2v - u)$ . Then the risk measure for case 2 is

$$P[c_1(2Z_1 - Z_2) > x, c_2(2Z_2 - Z_1) > y] = P \left[ T(Z_1, Z_2) > \left( \frac{x}{c_1}, \frac{y}{c_2} \right) \right].$$

Since, as  $t \rightarrow \infty$ ,

$$t P \left[ T(Z_1, Z_2) > \left( \frac{b(t)x}{c_1}, \frac{b(t)y}{c_2} \right) \right] \rightarrow \nu \left( T^{-1} \left( \left( \left( \frac{x}{c_1}, \frac{y}{c_2} \right), \infty \right) \right) \right)$$

and  $T^{-1}((x/c_1, y/c_2), \infty) \subset \text{CONE}$ , the risk measure for case 2 is small and the losses are asymptotically independent. Hence, the investment portfolio in case 2 achieves more diversification of risk than the portfolio in case 1, and both earn the same amount of risk premium.

How do we construct a risk vector  $Z$  whose distribution is regularly varying and whose limit measure  $\nu$  concentrates on  $\mathbb{D} \setminus \text{CONE}$ ? Suppose that  $R_1, R_2, U_1, U_2$ , and  $B$  are independent, and that  $R_1$  is Pareto(1),  $R_2$  is Pareto(2),  $U_1 \sim U((0, \frac{1}{3}) \cup (\frac{2}{3}, 1))$ ,  $U_2 \sim U(\frac{1}{3}, \frac{2}{3})$ , and  $B$  is a Bernoulli random variable taking the values 0 and 1 with equal probability. Define

$$Z = BR_1(U_1, 1 - U_1) + (1 - B)R_2(U_2, 1 - U_2).$$

Because of Proposition 3.1 applied with the  $L_1$  metric,  $Z$  is regularly varying on  $\mathbb{D}$  with index 1 and angular measure concentrating on  $[0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$ . Hence, the limit measure  $\nu$  concentrates on  $\mathbb{D} \setminus \text{CONE}$ . At scale  $t^{1/2}$ ,  $Z$  is also regularly varying on  $\text{CONE}$  with uniform angular measure.

**Example 5.2.** (*Asymptotic full dependence.*) For convenience, in this example we restrict attention to  $d = 2$ . HRV [20] was designed to deal with asymptotic independence [21, p. 322] where  $\nu(\cdot)$  concentrates on the axes. For asymptotic full dependence, the limit measure  $\nu(\cdot)$  concentrates on the diagonal  $\text{DIAG} := \{(z, z) : z \geq 0\}$  [21, p. 195] and previous treatments did not deal with this or related degeneracies where the limit measure  $\nu(\cdot)$  concentrates on a finite number of rays other than the axes.

Consider the following example. Suppose that  $X_1, X_2$ , and  $X_3$  are i.i.d. with common distribution Pareto(2). Let  $B$  be a Bernoulli random variable independent of  $\{X_i : i = 1, 2, 3\}$ , and let  $P[B = 0] = P[B = 1] = \frac{1}{2}$ . Construct the random vector  $Z$  as

$$Z = (Z_1, Z_2) = B((X_1)^2, (X_1)^2) + (1 - B)(X_2, X_3).$$

So,  $\mathbf{Z}$  is regularly varying on  $\mathbb{O} = \mathbb{D}$  with scaling function  $b(t) = t$  and limit measure

$$\nu([\mathbf{0}, (u, v)]^c) = \frac{1}{2}(u \wedge v)^{-1}, \quad (u, v) \in \mathbb{D}.$$

The measure  $\nu(\cdot)$  concentrates on DIAG and satisfies  $\nu(\{(u, v) \in \mathbb{D} : |u - v| > x\}) = 0$  for  $x > 0$ , and, as a result, we estimate risk probabilities like  $P[|Z_1 - Z_2| > x]$  for large thresholds  $x$  as 0. We gain precision from HRV.

Define the cone

$$\mathbb{O}_1 = [0, \infty)^2 \setminus (\{\mathbf{0}\} \cup \text{DIAG}) = \{(u, v) \in \mathbb{D} : |u - v| > 0\}.$$

The distribution of  $\mathbf{Z}$  has HRV on  $\mathbb{O}_1$  with scaling function  $b_1(t) = t^{1/2}$  and limit measure

$$\nu_1(du dv) = u^{-3} du \varepsilon_0(dv) + \varepsilon_0(du) v^{-3} dv, \quad (u, v) \in \mathbb{O}_1,$$

the measure that concentrates mass on the axes and which is restricted to  $\mathbb{O}_1$ . This limit results from the second summand in  $\mathbf{Z}$ ,  $(1 - B)(X_2, X_3)$ ; the first summand  $B(X_1^2, X_1^2)$  contributes nothing to the limit due to the restriction to  $\mathbb{O}_1$ . So, for instance, for  $(u, v) \in \mathbb{O}_1$  and  $x > 0$ ,

$$\nu_1([\mathbf{0}, (u, v)]^c \cap \{(u, v) \in \mathbb{O}_1 : |u - v| > x\}) = \frac{1}{2}((u \vee x)^{-2} + (v \vee x)^{-2})$$

and, for some large  $t > 0$ , letting  $u \downarrow 0$  and  $v \downarrow 0$ , we see that

$$P[|Z_1 - Z_2| > x] \approx t^{-1} \left( \frac{x}{b_1(t)} \right)^{-2}.$$

Statistical estimates of the risk region probability replace  $b_1(t)$  by a statistic; see Section 6.

The fact that  $\nu_1(\cdot)$  concentrates on the axes suggests seeking a further HRV property on a cone smaller than  $\mathbb{O}_1$ . If we need risk probabilities of the form  $P[Z_1 - Z_2 > x, Z_2 > y]$  for large thresholds  $x$  and  $y$ , we seek the HRV property, say,  $\mathbb{O}_2 = \{(u, v) \in \mathbb{O}_1 : u, v > 0\}$  or a subcone of  $\mathbb{O}_2$ .

As an example of why risk probabilities like  $P[Z_1 - Z_2 > x]$  arise, imagine investing in financial instruments  $I_1$  and  $I_2$  that have risks  $Z_1$  and  $Z_2$  per unit of investment and suppose that these risks have asymptotic full dependence.

For any  $a_1, a_2, c > 0$ , asymptotic full dependence of  $Z_1$  and  $Z_2$  implies that  $P[a_1 Z_1 + a_2 Z_2 > x]$  should be bigger than  $P[c(Z_1 - Z_2) > x]$ , provided  $x$  is large. So, if  $I_1 > I_2$ , it is less risky to invest in the financial instrument  $I_1 - I_2$  rather than investing in both  $I_1$  and  $I_2$ . Obviously, investing in the financial instrument  $I_1 - I_2$  requires us to measure risks associated with this portfolio, which leads to the need to evaluate  $P[Z_1 - Z_2 > x]$  for large thresholds  $x$ .

In summary, this example shows how a more flexible definition of HRV possibly allows computation of risk probabilities in the presence of asymptotic full dependence.

HRV was originally defined for  $d = 2$  and for two cones [20], and then extended to a finite sequence of cones [17]. In this paper our definition of HRV allows the possibility that we progressively find HRV on an infinite sequence of cones. We now present an example for  $d = 2$  where this is indeed the case. An infinite sequence of cones may create problems for risk estimation which we discuss afterwards.

**Example 5.3.** (*Infinite sequence of cones.*) Suppose that  $\{X_i, i \geq 1\}$  are i.i.d. random variables with common Pareto(1) distribution. Let  $\{Y_1, Y_2\}$  be i.i.d. with common Pareto(2) distribution,

and suppose that  $\{B_i, i \geq 1\}$  is an infinite sequence of random variables with  $P[B_i = 1] = 1 - P[B_i = 0] = 2^{-i}$  and  $\sum_{i=1}^\infty B_i = 1$ . (For instance, let  $T$  be the index of the first success in an i.i.d. sequence of Bernoulli trials and then set  $B_i = \mathbf{1}_{\{T=i\}}, i \geq 1$ .) Assume that  $\{X_i, i \geq 1\}, \{Y_1, Y_2\}$ , and  $\{B_i : i \geq 1\}$  are mutually independent. Define the random vector  $\mathbf{Z}$  as

$$\mathbf{Z} = (Z_1, Z_2) = B_1(Y_1, Y_2) + \sum_{i=1}^\infty B_{i+1}((X_i)^{1/(2-2^{-(i-1)})}, 2^{i-1}(X_i)^{1/(2-2^{-(i-1)})}).$$

So,  $\mathbf{Z}$  has regular variation on the cone  $\mathbb{O} = \mathbb{O}_0 = \mathbb{D} = [0, \infty)^d \setminus \{\mathbf{0}\}$  with index of regular variation  $\alpha = 1$ , scaling function  $b(t) = t$ , and limit measure  $\nu(\cdot)$  concentrating on the diagonal  $\text{DIAG} := \{(x, x) : x \in [0, \infty)\}$ . This comes from the  $i = 1$  term in the series. We remove the diagonal and find HRV on  $\mathbb{O}_1 = \mathbb{D} \setminus \text{DIAG}$  with  $b_1(t) = t^{2/3}, \alpha_1 = \frac{3}{2}$ , and limit measure  $\nu_1$  concentrating on the ray  $\{(x, 2x) : x \geq 0\}$ . Progressively seeking HRV on successive cones, we find at the  $(i + 1)$ th step of our analysis that  $\mathbf{Z}$  has regular variation on the cone

$$\mathbb{O}_i = \mathbb{D} \setminus \left[ \bigcup_{j=1}^i \{(x, 2^{j-1}x) : x \in [0, \infty)\} \right] \tag{5.1}$$

with index of regular variation  $\alpha_i = 2 - 2^{-i}$  and limit measure  $\nu_i(\cdot)$  concentrating on

$$\{(x, 2^i x) : x \in [0, \infty)\}.$$

Selection of cones must be guided by the type of risk probability needed. Consider trying to estimate  $P[Z_1 - Z_2 > x]$  for large thresholds  $x$  using the cones  $\mathbb{O}_i, i \geq 0$ , given in (5.1). At the  $(i + 1)$ th stage, using cone  $\mathbb{O}_i$ , the limit measure  $\nu_i(\cdot)$  puts zero mass on the cone  $\{(u, v) : u > v\}$ . So, even after a million HRV steps, we will estimate  $P[Z_1 - Z_2 > x]$  for large thresholds  $x$  as 0, which is clearly wrong due to the definition of  $\mathbf{Z}$  since

$$P[Z_1 - Z_2 > x] \geq P[B_1 = 1]P[Y_1 - Y_2 > x] > 0.$$

An alternative procedure seeks regular variation on the cone  $\{(u, v) : u > v\}$ , and this leads to somewhat more reasonable estimates of  $P[Z_1 - Z_2 > x]$  for large thresholds  $x$  since in this case the regular variation with the Pareto(2) variables is captured.

This example is not likely to arise in practice, but the moral of the story is that the choice of sequence of cones when defining HRV must be guided by the kind of risk sets considered. For example, if we are interested in only joint tail probabilities, a possible choice of sequence of cones is  $\mathbb{D} = \mathbb{D}^{(1)} \supset \mathbb{D}^{(2)} \supset \dots \supset \mathbb{D}^{(d)}$ , where

$$\mathbb{D}^{(l)} = \{\mathbf{x} \in [0, \infty)^d : x_{(l)} > 0\}$$

and  $x_{(l)}$  is the  $l$ th largest component of  $\mathbf{x}$  [17].

**Example 5.4.** (*The CEV model and mass on the lines through  $\infty$* .) If we consider the CEV model on  $\mathbb{E}_\square$ , there can exist two different limits in (4.1) under two different normalizations. This problem disappears if we restrict convergence to  $\mathbb{D}_\square$ .

Suppose that  $Y$  is Pareto(1) and  $B$  is a Bernoulli random variable with  $P[B = 1] = P[B = 0] = \frac{1}{2}$ . Define

$$\mathbf{Z} = (Z_1, Z_2) = B(Y, Y) + (1 - B)(\sqrt{Y}, Y).$$

Then the following two convergences hold in  $\mathbb{E}_\Gamma$ : for  $x \geq 0$  and  $y > 0$ ,

$$\begin{aligned} \nu_1([0, x] \times (y, \infty]) &:= \lim_{t \rightarrow \infty} t \mathbb{P} \left[ \left( \frac{Z_1}{t}, \frac{Z_2}{t} \right) \in [0, x] \times (y, \infty] \right] = \frac{1}{2} \left( \frac{1}{y} - \frac{1}{x} \right)_+ + \frac{1}{2y}, \\ \nu_2([0, x] \times (y, \infty]) &:= \lim_{t \rightarrow \infty} t \mathbb{P} \left[ \left( \frac{Z_1}{\sqrt{t}}, \frac{Z_2}{t} \right) \in [0, x] \times (y, \infty] \right] = \frac{1}{2} \left( \frac{1}{y} - \frac{1}{x^2} \right)_+. \end{aligned} \tag{5.2}$$

So  $\mathbf{Z}$  follows a CEV model on  $\mathbb{E}_\Gamma$  with two different scalings. Note that  $\nu_1$  does not put any mass on lines through  $\infty$ , but  $\nu_2$  does:  $\nu_2(\{\infty\} \times (y, \infty]) = 1/2y$ . If we restrict convergence to  $\mathbb{D}_\Gamma$ , limits are unique and  $\mathbf{Z}$  is regularly varying on  $\mathbb{D}_\Gamma$  with limit measure  $\nu_1$  given by (5.4) restricted to  $\mathbb{D}_\Gamma$ .

In some cases, our approach may not identify the hidden structure. As pointed out by an astute referee, this can happen if there are uncountably many distinct indices of regular variation along distinct rays. We now provide an example in two dimensions.

**Example 5.5.** (*Different tail rates in different directions.*) Suppose that  $(R, U)$  is a bivariate random vector with probability density function given by

$$f_{R,U}(r, u) = 4r^{-(u+2)}((u + 1) \log r - 1), \quad r > 2, 0 < u < 1.$$

Now define  $\mathbf{Z} = (RU, R(1-U))$ . We can check that  $\mathbf{Z}$  is regularly varying on  $\mathbb{D} = [0, \infty)^2 \setminus \{\mathbf{0}\}$  by checking this for the polar coordinate version  $(R, U)$  as follows. For any  $r > 2$  and  $0 \leq a < b \leq 1$ ,

$$\mathbb{P}[R > r, U \in [a, b]] = 4r^{-(a+1)} - 4r^{-(b+1)}.$$

Thus, if we define  $b_a(t) = t^{1/(a+1)}$  for  $0 \leq a \leq 1$ , we have, for  $r > 2$ ,  $\|(x, y)\| = x + y$  and  $T_{a,b} = \{(u, 1 - u) : a \leq u \leq b\}$ ,

$$\lim_{t \rightarrow \infty} t \mathbb{P} \left( \frac{\|\mathbf{Z}\|}{b_a(t)} > r, \frac{\mathbf{Z}}{\|\mathbf{Z}\|} \in T_{a,b} \right) = 4\nu_{a+1}(r, \infty) \times \varepsilon_{(a,1-a)}(T_{a,b}),$$

where  $\nu_\alpha(r, \infty) = r^{-\alpha}$  for  $r > 0$  and  $\alpha > 0$ . Hence, if we choose the cone

$$\mathbb{D}_{a,b} = \left\{ (x, y) \in \mathbb{D} : x + y > 0, \frac{x}{x + y} \in [a, b] \right\}$$

as our first cone for seeking regular variation then we get  $\mathbf{Z}$  to be regularly varying on  $\mathbb{D}_{a,b}$  with normalizing function  $b_a(t) = t^{1/(a+1)}$  and limit measure  $\nu_{a+1} \times \varepsilon_{(a,1-a)}$  concentrated on the line  $L_a = \{(x, y) \in \mathbb{D} : x/(x + y) = a\}$ . Specifically,  $\mathbf{Z}$  has regular variation on  $\mathbb{D} = \mathbb{D}_{0,1}$  with  $b(t) = t$  and limit measure concentrated on the  $y$ -axis  $L_0$ .

After finding regular variation on  $\mathbb{D}_{a,b}$  with limit measure concentrated on  $L_a$ , if in the next step we take the cone  $\mathbb{D}_{a,b} \setminus L_a$ , an HRV cannot be found. This gives an example where we have a regularly varying tail in any direction and examination of further hidden regularly varying structure is not possible.

### 6. Estimating the spectral measure and its support

We have defined regular variation on a big cone  $\mathbb{O}_0 \subset \mathbb{R}^d$  along with HRV in a nested sequence of subcones  $\mathbb{O}_0 \supset \mathbb{O}_1 \supset \mathbb{O}_2 \supset \dots$ . We now propose strategies for deciding whether HRV is consistent with a given data set and, if so, how to estimate probabilities of sets pertaining

to joint occurrence of extreme or high values. We proceed as follows.

- Specify a fixed finite sequence of cones pertinent to the problem and seek HRV sequentially on these cones. We discuss this in Section 6.1 which follows ideas proposed in [17, Section 3].
- If the sequence of cones is not clear, proceed by estimating the support of the limit measure at each step, removing it, and seeking HRV on the complement of the support. Then estimate the hidden limit measure using semiparametric techniques similar to those used in Section 6.1.

**6.1. Specified sequence of cones**

Suppose that  $\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n$  are i.i.d. random vectors in  $\mathbb{C} \subset [0, \infty)^d$  whose common distribution has a regularly varying tail on  $\mathbb{O}$  according to Definition 3.1 with normalizing function  $b(\cdot)$  and limit measure  $\nu(\cdot)$ . Also, assume that we have a specified sequence of cones  $\mathbb{O} = \mathbb{O}_0 \supset \mathbb{O}_1 \supset \mathbb{O}_2 \supset \dots$  where we seek regular variation. Such a sequence of cones is known and fixed.

We provide an estimate for the limit measure of regular variation on  $\mathbb{O}$  and the same method can be applied to find limit measures for HRV on the subcones.

Now, according to Corollary 3.1, regular variation on  $\mathbb{O}$  as above is equivalent to assuming that  $P[\text{dist}(\mathbf{Z}, \mathbb{F}) > x]$  is regularly varying at  $\infty$  with some exponent  $\alpha > 0$  and normalizing function  $b(t)$ , and

$$P\left[\frac{\mathbf{Z}}{\text{dist}(\mathbf{Z}, \mathbb{F})} \in \cdot \mid \text{dist}(\mathbf{Z}, \mathbb{F}) > t\right] \Rightarrow S_{\mathbb{O}}(\cdot) \quad \text{as } t \rightarrow \infty \tag{6.1}$$

in  $P(\partial\mathfrak{S}_{\mathbb{O}})$ , the class of all probability measures on  $\partial\mathfrak{S}_{\mathbb{O}} = \{\mathbf{x} \in \mathbb{O} : \text{dist}(\mathbf{x}, \mathbb{F}) = 1\}$ . We take  $b(t) := F_d^{\leftarrow}(1 - 1/t)$ , where  $\text{dist}(\mathbf{Z}, \mathbb{F})$  has distribution function  $F_d$ . Thus, we estimate  $\nu$  by estimating  $\alpha$  and  $S_{\mathbb{O}}$  separately. Considering  $d_i^{\mathbb{F}} := \text{dist}(\mathbf{Z}_i, \mathbb{F})$ ,  $i = 1, \dots, n$ , as i.i.d. from a regularly varying distribution on  $(0, \infty)$ , the exponent  $\alpha$  can be estimated using the Hill, Pickands, or QQ estimator [5], [21].

We now present an outline of how to obtain an empirical estimator of  $S_{\mathbb{O}}$  following [21].

**Proposition 6.1.** *Assume that the common distribution of the i.i.d. random vectors  $\mathbf{Z}_1, \dots, \mathbf{Z}_n$  satisfies Definition 3.1 and (3.1). As  $n \rightarrow \infty$ ,  $k \rightarrow \infty$ , and  $n/k \rightarrow \infty$ , we have, in  $P(\partial\mathfrak{S}_{\mathbb{O}})$ ,*

$$S_n(\cdot) := \frac{\sum_{i=1}^n \varepsilon_{(d_i^{\mathbb{F}}/b(n/k), \mathbf{Z}_i/d_i^{\mathbb{F}})}((1, \infty) \times \cdot)}{\sum_{i=1}^n \varepsilon_{d_i^{\mathbb{F}}/b(n/k)}(1, \infty)} \Rightarrow S_{\mathbb{O}}(\cdot). \tag{6.2}$$

*Proof.* We have  $\{d_i^{\mathbb{F}}, 1 \leq i \leq n\}$  i.i.d. regularly varying random variables from a distribution  $F_d$  on  $(0, \infty)$  with regularly varying tail and norming function  $b(t) = F_d^{\leftarrow}(1 - 1/t)$ , by Theorem A.1. Thus, for  $x > 0$ ,  $(n/k) P[d_i^{\mathbb{F}}/b(n/k) > x] \rightarrow cx^{-\alpha}$ , and from [21, p. 139], this is equivalent to  $(1/k) \sum_{i=1}^n \varepsilon_{d_i^{\mathbb{F}}/b(n/k)}(1, \infty) \Rightarrow c$ , and to prove (6.2), it suffices to show that, in  $M_+(\partial\mathfrak{S}_{\mathbb{O}})$ ,

$$\frac{1}{k} \sum_{i=1}^n \varepsilon_{\mathbf{Z}_i/d_i^{\mathbb{F}}}(\cdot) \mathbf{1}_{[d_i^{\mathbb{F}}/b(n/k) > 1]} \Rightarrow S_{\mathbb{O}}(\cdot). \tag{6.3}$$

The counting function in (6.3) only counts  $\mathbf{Z}_i/d_i^{\mathbb{F}}$  such that  $d_i^{\mathbb{F}}/b(n/k) > 1$ . The distribution of such random elements is  $P[\mathbf{Z}_i/d_i^{\mathbb{F}} \in \cdot \mid d_i^{\mathbb{F}}/b(n/k) > 1]$  [22, p. 212], and (6.1) holds. Using [21, Theorem 5.3ii, p. 139] and the style of argument in [22, p. 213], we obtain (6.3).

The estimator  $S_n$  of  $S_{\mathbb{O}}$  in Proposition 6.1 relies on  $b(t)$ , which is typically unknown, but  $b(n/k)$  can be estimated. Order  $d_1^{\mathbb{F}}, \dots, d_n^{\mathbb{F}}$  as  $d_{(1)}^{\mathbb{F}} \geq \dots \geq d_{(n)}^{\mathbb{F}}$ , and let  $d_{(k+1)}^{\mathbb{F}}/b(n/k) \xrightarrow{p} 1$  so  $d_{(k+1)}^{\mathbb{F}}$  is a consistent estimator of  $b(n/k)$  as  $n \rightarrow \infty, k \rightarrow \infty$ , and  $n/k \rightarrow \infty$  [21, p. 81]. Hence, we replace  $b(n/k)$  by  $d_{(k+1)}^{\mathbb{F}}$  and propose the estimator  $\hat{S}_n$  for  $S_{\mathbb{O}}$  in  $S_n$  as follows:

$$\hat{S}_n(\cdot) := \frac{\sum_{i=1}^n \varepsilon_{\{d_i^{\mathbb{F}}/d_{(k+1)}^{\mathbb{F}}, \mathbf{Z}_i/d_i^{\mathbb{F}}\}}((1, \infty) \times \cdot)}{\sum_{i=1}^n \varepsilon_{\{d_i^{\mathbb{F}}/d_{(k+1)}^{\mathbb{F}}\}}(1, \infty)} = \frac{1}{k} \sum_{i=1}^n \mathbf{1}_{[d_i^{\mathbb{F}}/d_{(k+1)}^{\mathbb{F}} > 1]} \varepsilon_{\mathbf{Z}_i/d_i^{\mathbb{F}}}(\cdot). \tag{6.4}$$

**Proposition 6.2.** Define  $\hat{S}_n$  as in (6.4). As  $n \rightarrow \infty, k \rightarrow \infty$ , and  $n/k \rightarrow \infty$ ,  $\hat{S}_n \Rightarrow S_{\mathbb{O}}$  in  $P(\partial\mathfrak{N}_{\mathbb{O}})$ .

*Proof.* Use the continuous mapping theorem and Proposition 6.1 as in, for instance, [6].

Thus, when  $\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n$  are i.i.d. random vectors in  $\mathbb{C} \subset [0, \infty)^d$  which have a regularly varying distribution on  $\mathbb{O}$ , we can estimate both  $\alpha$  and the spectral measure  $S_{\mathbb{O}}$ . If we have a specified finite sequence of cones  $\mathbb{O} := \mathbb{O}_0 \supset \mathbb{O}_1 \supset \mathbb{O}_2 \supset \dots \supset \mathbb{O}_m$ , then we sequentially estimate the limit measure by separately estimating the spectral measure and the index.

**6.2. Support estimation**

Suppose that  $\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n$  are i.i.d. random vectors in  $\mathbb{C} \subset [0, \infty)^d$  whose common distribution is regularly varying on  $\mathbb{O}$  according to Definition 3.1 with normalizing function  $b(\cdot)$  and limit measure  $\nu(\cdot)$ . Without a sequence of cones where HRV can be sought, the task of looking for appropriate cones where HRV may exist is challenging. One clear strategy is to identify the support of  $\nu$ , which we call  $\text{supp}(\nu)$ , and then seek HRV on the complement of the support. Since

$$\text{supp}(\nu) = \{t \cdot \text{supp}(S_{\mathbb{O}}), t \geq 0\},$$

it suffices to determine the support of  $S_{\mathbb{O}}$ .

We propose estimating the support of the spectral measure  $S_{\mathbb{O}}$  with a point cloud, that is, a discrete random closed set.

**Proposition 6.3.** Suppose that  $\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n$  are i.i.d. random vectors in  $\mathbb{C} \subset [0, \infty)^d$  whose common distribution is regularly varying on  $\mathbb{O}$  with normalizing function  $b(\cdot)$  and limit measure  $\nu(\cdot)$ . As  $n \rightarrow \infty, k \rightarrow \infty$ , and  $n/k \rightarrow \infty$ ,

$$\text{supp}_{k,n} = \left\{ \frac{\mathbf{Z}_i}{d_i^{\mathbb{F}}} : d_i^{\mathbb{F}} > d_{(k+1)}^{\mathbb{F}}, i = 1, \dots, n \right\} \Rightarrow \text{supp}(S_{\mathbb{O}}). \tag{6.5}$$

Convergence in (6.5) occurs in the space of closed sets under the Fell topology or the space of compact sets in the Hausdorff topology [18].

*Proof of Proposition 6.3.* To show (6.5), from [18, Proposition 6.10, page 87], it suffices to show that, for any  $h \in C_K^+( \partial\mathfrak{N}_{\mathbb{O}} )$ ,

$$\mathbb{E} \left( \sup_i \left\{ h \left( \frac{\mathbf{Z}_i}{d_i^{\mathbb{F}}} \right) : d_i^{\mathbb{F}} > d_{(k+1)}^{\mathbb{F}}, 1 \leq i \leq n \right\} \right) \rightarrow \sup_x \{h(x) : x \in \text{supp}(S_{\mathbb{O}})\}. \tag{6.6}$$

From Proposition 6.2,

$$\hat{S}_n(\cdot) := \frac{1}{k} \sum_{i=1}^n \mathbf{1}_{[d_i^{\mathbb{F}}/d_{(k+1)}^{\mathbb{F}} > 1]} \varepsilon_{\mathbf{Z}_i/d_i^{\mathbb{F}}}(\cdot) \Rightarrow S_{\mathbb{O}},$$

and, from the continuous mapping theorem, for any  $h \in C_K^+(\partial\mathfrak{N}_\mathbb{O})$ , in  $\mathbb{P}(\mathbb{R})$ , the class of probability measures on  $\mathbb{R}$ , we obtain

$$\hat{S}_n \circ h^{-1} = \frac{1}{k} \sum_{i=1}^n \mathbf{1}_{[d_i^{\mathbb{F}}/d_{(k+1)}^{\mathbb{F}} > 1]} \varepsilon_{h(\mathbf{Z}_i/d_i^{\mathbb{F}})}(\cdot) \Rightarrow S_\mathbb{O} \circ h^{-1}. \tag{6.7}$$

If the  $F_n$ ,  $n \geq 0$ , are probability measures on  $\mathbb{R}$  with bounded support and  $F_n \Rightarrow F_0$ , then

$$x_{F_n} := \sup_x \{x : F_n(x) < 1\} \rightarrow \sup_x \{x : F(x) < 1\} =: x_F.$$

Applying this remark to (6.7) and using the continuous mapping theorem yields, as  $n \rightarrow \infty$ ,  $k \rightarrow \infty$ , and  $n/k \rightarrow \infty$ ,

$$\sup_i \left\{ h\left(\frac{\mathbf{Z}_i}{d_i^{\mathbb{F}}}\right) : d_i^{\mathbb{F}} > d_{(k+1)}^{\mathbb{F}}, 1 \leq i \leq n \right\} \Rightarrow \sup_x \{h(x) : x \in \text{supp}(S_\mathbb{O})\}.$$

Since  $h \in C_K^+(\partial\mathfrak{N}_\mathbb{O})$  is always bounded above, use dominated convergence applied to convergence in distribution to get (6.6).

Proposition 6.3 provides an estimate of  $\text{supp}(S_\mathbb{O})$  and, hence, of  $\text{supp}(\nu)$ . In principle, we can remove the estimated support from  $\mathbb{O}$  and look for HRV in the complement. How well this works in practice remains to be seen. For one thing, the estimated support set of  $S_\mathbb{O}$  is always discrete, meaning that the estimated support of  $\nu$  is a finite set of rays. With a large data set, we might be able to get a fair idea about the support of the distribution and where to look for further HRV. If there was reason to believe or hope that the support of  $S_\mathbb{O}$  is convex, our estimation procedure could be modified by taking the convex hull of the points in (6.5).

### 7. Conclusion

Our treatment of regular variation on cones which is determined by the support of the limit measures unifies under one theoretical umbrella several related concepts: asymptotic independence, asymptotic full dependence, and the conditional extreme value model. Our approach highlights the structural similarities of these concepts while making plain in what ways the cases differ. Furthermore, the notion of  $\mathbb{M}^*$ -convergence introduced in Section 2.2 provides a tool to deal with the generalized notion of regular variation given here. Generalizing this notion of convergence and analyzing its properties admits potential for further research.

It is always an ambitious undertaking to statistically identify lower-order behavior and this project has not attempted data analysis or tested the feasibility of the statistical methods discussed in Section 6. It is clear that further work is required, particularly for the case where the support of the limit measures must be identified from data. One can imagine that, for high-dimensional data whose dimension is of the order of hundreds, sophistication is required to pursue successive cones where regular variation exists.

### Appendix A

#### A.1. Regular variation on $\mathbb{E} := [0, \infty]^d \setminus \{0\}$ versus $\mathbb{D} = [0, \infty)^d \setminus \{0\}$

We verify that the traditional notion of multivariate regular variation given in (1.1) on  $\mathbb{E}$  is equivalent to Definition 3.1 if we choose  $\mathbb{C} = [0, \infty)^d$  and  $\mathbb{F} = \{0\}$ . This yields  $\mathbb{O} = \mathbb{D}$ .

**Theorem A.1.** *Regular variation on  $\mathbb{D}$  according to Definition 3.1 is equivalent to the traditional notion of multivariate regular variation given in (1.1), and the limit measures  $\nu(\cdot)$  of (3.1) and  $\mu(\cdot)$  of (1.1) are equal on  $\mathbb{D}$ . Moreover,  $\mu(\cdot)$  puts zero measure on  $\mathbb{E} \setminus \mathbb{D}$ .*

*Proof.* First we show that the standard notion of multivariate regular variation on  $\mathbb{E}$  given in (1.1) implies (3.1) in  $\mathbb{M}^*(\mathbb{D})$ . Let  $\nu(\cdot)$  be a measure on  $\mathbb{D}$  such that  $\nu(\cdot) = \mu(\cdot)$ . From [21, p. 176], we find that  $\mu(\mathbb{E} \setminus \mathbb{D}) = 0$ . So, since  $\mu(\cdot) \neq 0$  and nondegenerate,  $\nu(\cdot) \neq 0$  and nondegenerate.

For  $B \subset \mathbb{D}$ , note that  $\partial B = \bar{B} \setminus B^o$  is defined with respect to the relative topology on  $\mathbb{D}$  and, hence,  $\partial B \subset \mathbb{D}$ . Thus,  $\nu(\partial B) = 0$  implies that  $\mu(\partial B) = 0$ . Also, since  $[0, \infty]^d$  is a compact space, any set  $B \subset \mathbb{D}$  bounded away from  $\{\mathbf{0}\}$  is a relatively compact set in  $\mathbb{E}$  [21, Proposition 6.1, p. 171]. Therefore, by definition,  $\nu(\cdot) \in \mathbb{M}^*([0, \infty)^d, \mathbb{D})$  and, by (1.1), for any  $B \subset \mathbb{D}$  bounded away from  $\{\mathbf{0}\}$  and  $\nu(\partial B) = 0$ ,

$$t \mathbb{P} \left[ \frac{\mathbf{Z}}{b(t)} \in B \right] \rightarrow \mu(B) = \nu(B).$$

So, (3.1) holds with  $\mathbb{C} = [0, \infty)^d$ ,  $\mathbb{O} = \mathbb{D}$ ,  $b$  is the same as in (1.1), and  $\nu$  is the restriction of  $\mu$  to  $\mathbb{D}$ .

Conversely, we show that Definition 3.1 and (3.1) with  $\mathbb{O} = \mathbb{D}$  implies the traditional notion of multivariate regular variation on  $\mathbb{E}$  in (1.1). Define a measure  $\mu(\cdot)$  on  $\mathbb{E}$  as  $\mu(\cdot) = \nu(\cdot \cap \mathbb{D})$ . A relatively compact set  $B$  of  $\mathbb{E}$  must be bounded away from  $\{\mathbf{0}\}$  [21, Proposition 6.1, p. 171]. So, from the definition of  $\mu(\cdot)$ , it is Radon. Note that  $\partial B = \bar{B} \setminus B^o$  is defined with respect to the topology on  $\mathbb{E}$ , but  $\partial(B \cap \mathbb{D})$  is defined with respect to the relative topology on  $\mathbb{D}$ . Also, from the definition of  $\mu(\cdot)$ ,  $\mu(\partial B) = 0$  implies that  $\nu(\partial(B \cap \mathbb{D})) = \nu(\partial B \cap \mathbb{D}) = \mu(\partial B) = 0$ . Therefore, from (3.1), for any relatively compact set  $B$  of  $\mathbb{E}$  such that  $\mu(\partial B) = 0$ , as  $t \rightarrow \infty$ ,

$$t \mathbb{P} \left[ \frac{\mathbf{Z}}{b(t)} \in B \right] = t \mathbb{P} \left[ \frac{\mathbf{Z}}{b(t)} \in B \cap \mathbb{D} \right] \rightarrow \nu(B \cap \mathbb{D}) = \mu(B).$$

The first equality above holds since  $\mathbf{Z} \in [0, \infty)^d$ . Hence, vague convergence in (1.1) holds with the same  $b$  as in (3.1) and with  $\mu$  as the extension of  $\nu$  from  $\mathbb{D}$  to  $\mathbb{E}$ .

Regular variation on  $\mathbb{D}$  can also be expressed in terms of the polar coordinate transformation. As at the beginning of this section, set  $\mathbb{C} = [0, \infty)^d$ ,  $\mathbb{F} = \{\mathbf{0}\}$ , and  $\mathbb{O} = \mathbb{D}$ .

**Proposition A.1.** *Regular variation on  $\mathbb{O}$  as given in Definition 3.1 is equivalent to the condition*

$$t \mathbb{P} \left[ \left( \frac{\|\mathbf{Z}\|}{b(t)}, \frac{\mathbf{Z}}{\|\mathbf{Z}\|} \right) \in A \right] \rightarrow \nu_{(\alpha)} \times S_{\mathbb{O}}(A)$$

for all measurable  $A \subset (0, \infty) \times \partial \mathfrak{S}_{\mathbb{O}}$  such that  $\overline{(A)_1} \cap \{0\} = \emptyset$  and  $\nu \circ h^{-1}(\partial A) = 0$ , where  $(A)_1$  is the projection of  $A$  on its first coordinate,  $h(\cdot)$  is a function defined by  $h: \mathbf{x} \mapsto (\|\mathbf{x}\|, \mathbf{x}/\|\mathbf{x}\|)$ ,  $\partial \mathfrak{S}_{\mathbb{O}} = \{\mathbf{x} \in \mathbb{O}: \|\mathbf{x}\| = 1\}$ ,  $S_{\mathbb{O}}(\cdot)$  is a probability measure on  $\partial \mathfrak{S}_{\mathbb{O}}$ , and  $\nu_{(\alpha)}(\cdot)$  is a Pareto measure given by  $\nu_{(\alpha)}((x, \infty)) = x^{-\alpha}$  for  $x > 0$ . The probability measure  $S_{\mathbb{O}}(\cdot)$  is called the spectral measure and is related to  $\nu(\cdot)$  by the relation

$$S_{\mathbb{O}}(\Lambda) = \nu \left( \left\{ \mathbf{x} \in \mathbb{O}: \|\mathbf{x}\| \geq 1, \frac{\mathbf{x}}{\|\mathbf{x}\|} \in \Lambda \right\} \right).$$

*Proof.* This is a special case of Proposition 3.1.

**A.2. Regular variation on  $\mathbb{E}^{(l)} = [0, \infty]^d \setminus \{x \in \mathbb{E} : x_{(l)} > 0\}$  versus  $\mathbb{D}^{(l)} = [0, \infty)^d \setminus \{x \in [0, \infty)^d : x_{(l)} > 0\}$**

Recall that  $x_{(l)}$  is the  $l$ th largest component of  $x$ ,  $l = 1, 2, \dots, d$ . HRV using  $\mathbb{E}^{(l)}$  is considered in [16]. Unlike the situation in subsection A.1, here limit measures can put mass on  $\mathbb{E}^{(l)} \setminus \mathbb{D}^{(l)}$  as found in [16]. We compare regular variation in  $\mathbb{E}^{(l)}$  using the traditional vague convergence definition in which the vague convergence in (1.1) is assumed to hold in  $\mathbb{M}_+(\mathbb{E}^{(l)})$  with regular variation given in (3.1) in  $\mathbb{M}^*(\mathbb{C}, \mathbb{O})$ , where  $\mathbb{C} = [0, \infty)^d$ ,  $\mathbb{F} = \{x \in [0, \infty)^d : x_{(l)} = 0\}$ , and  $\mathbb{O} = \mathbb{D}^{(l)} = \mathbb{C} \setminus \mathbb{F}$ .

**Theorem A.2.** *Regular variation on  $\mathbb{M}^*(\mathbb{D}^{(l)})$  is equivalent to the traditional vague convergence notion of regular variation in  $\mathbb{M}_+(\mathbb{E}^{(l)})$  if the limit measure  $\mu(\cdot)$  given in the  $\mathbb{M}_+(\mathbb{E}^{(l)})$  analogue of (1.1) does not give any mass to the set  $\mathbb{E}^{(l)} \setminus \mathbb{D}^{(l)}$ . In this case, the limit measures  $\nu(\cdot)$  of (3.1) and  $\mu(\cdot)$  of (1.1) are equal on  $\mathbb{D}^{(l)}$ .*

*Proof.* Suppose that, for a random vector  $Z$ , there exist a function  $b^{(l)}(t) \uparrow \infty$  and a nonnegative, nondegenerate Radon measure  $\mu^{(l)}(\cdot) \neq 0$  on  $\mathbb{E}^{(l)}$  such that, in  $\mathbb{M}_+(\mathbb{E}^{(l)})$ ,

$$t P \left[ \frac{Z}{b^{(l)}(t)} \in \cdot \right] \xrightarrow{v} \mu^{(l)}(\cdot), \tag{A.1}$$

and the limit measure  $\mu^{(l)}(\cdot)$  does not give any mass to  $\mathbb{E}^{(l)} \setminus \mathbb{D}^{(l)}$ . Define a measure  $\chi(\cdot)$  on  $\mathbb{D}^{(l)}$  as  $\chi(\cdot) = \mu^{(l)}(\cdot)$ . Since  $\mu^{(l)}(\cdot) \neq 0$  is nonnegative, nondegenerate, and  $\mu^{(l)}(\mathbb{E}^{(l)} \setminus \mathbb{D}^{(l)}) = 0$ , the measure  $\chi(\cdot) \neq 0$  is nonnegative and nondegenerate. The subsets of  $\mathbb{D}^{(l)}$  bounded away from  $(\mathbb{D}^{(l)})^c = [0, \infty)^d \setminus \mathbb{D}^{(l)}$  are relatively compact in  $\mathbb{E}^{(l)}$ . Therefore, using the fact that  $\mu^{(l)}(\cdot)$  is Radon and the definition of  $\chi(\cdot)$ , it follows that  $\chi(\cdot)$  gives finite measure to sets bounded away from  $(\mathbb{D}^{(l)})^c$ . From the definition of  $\mathbb{M}^*$ -convergence, it follows that  $Z$  satisfies (3.1) with the scaling function  $b(\cdot) = b^{(l)}(\cdot)$  and the limit measure  $\nu(\cdot) = \chi(\cdot)$ .

Conversely, suppose that a random vector  $Z$  satisfies (3.1) in  $\mathbb{M}^*(\mathbb{C}, \mathbb{O})$  with  $\mathbb{C} = [0, \infty)^d$  and  $\mathbb{O} = \mathbb{D}^{(l)}$ . Define a measure  $\mu(\cdot)$  on  $\mathbb{E}^{(l)}$  as  $\mu(\cdot) = \nu(\cdot \cap \mathbb{D}^{(l)})$ . Since  $\nu(\cdot) \neq 0$  and is nonnegative and nondegenerate, so is  $\mu(\cdot)$ . A subset of  $\mathbb{E}^{(l)}$  is relatively compact in  $\mathbb{E}^{(l)}$  if and only if it is bounded away from  $\{x \in [0, \infty)^d : x_{(l)} = 0\}$ . Since  $\nu(\cdot)$  gives finite mass to sets bounded away from  $(\mathbb{D}^{(l)})^c$ , from the definition of  $\mu(\cdot)$ , it follows that  $\mu(\cdot)$  is a Radon measure. From the description of the compact sets in  $\mathbb{E}^{(l)}$ , it follows that  $Z$  also satisfies (A.1) with the scaling function  $b^{(l)}(\cdot) = b(\cdot)$  and the limit measure  $\mu^{(l)}(\cdot) = \mu(\cdot)$  [21, Theorem 3.2, p. 52].

The set  $\mathbb{E}^{(l)} \setminus \mathbb{D}^{(l)} = \{x \in \mathbb{E}^{(l)} : \|x\| = \infty\}$  is the union of the lines through  $\infty$ . We emphasize that there exist examples of random vectors  $Z$  which satisfy (A.1) and the limit measure  $\mu^{(l)}(\cdot)$  gives positive measure on the set  $\mathbb{E}^{(l)} \setminus \mathbb{D}^{(l)}$  [17].

**A.3. Regular variation on  $\mathbb{E}_\Gamma = [0, \infty] \times (0, \infty)$  versus  $\mathbb{D}_\Gamma = [0, \infty) \times (0, \infty)$**

Recall the CEV model from Section 4.1.

**Proposition A.2.** *The following statements are equivalent.*

- (i)  $Z \in \text{CEV}(b_1, b_2)$  with limit measure  $\mu(\cdot)$  and  $b_1 \sim b_2$  with

$$\mu([0, \infty] \times \{\infty\} \cup \{\infty\} \times (0, \infty]) = 0. \tag{A.2}$$

- (ii)  $Z$  is regularly varying on  $\mathbb{D}_\Gamma$  according to (3.1) with normalizing function  $b_1$  and limit measure  $\nu$  which does not concentrate on  $\{0\} \times (0, \infty)$ .

Also, if either of (i) or (ii) holds, then  $\mu(\cdot) = \nu(\cdot)$  on  $\mathbb{D}_\Gamma$ .

*Proof.* (i) implies (ii). Since  $b_1 \sim b_2$ , (4.1) implies that  $\mathbf{Z} \in \text{CEV}(b_1, b_1)$ . Now, (4.1) implies that, for all relatively compact Borel sets  $B$  in  $\mathbb{D}_\Gamma \subset \mathbb{E}_\Gamma$  with  $\mu(\partial B) = 0$ ,

$$t \mathbb{P} \left[ \frac{\mathbf{Z}}{b_1(t)} \in B \right] \rightarrow \mu(B) \quad \text{as } t \rightarrow \infty.$$

Clearly,  $B$  is bounded away from  $\mathbb{F}$ . Also,  $\mu$  is nonnull and satisfies (A.2). Thus,  $\nu(\cdot) = \mu(\cdot)|_{\mathbb{D}_\Gamma}$  is nonnegative and nondegenerate on  $\mathbb{D}_\Gamma$ . Hence,  $\mathbf{Z}$  is regularly varying on  $\mathbb{D}_\Gamma$  with limit measure  $\nu$ . The nondegeneracy condition given in Definition 4.1(i) for the CEV model implies that  $\mu$  cannot concentrate on  $\{0\} \times (0, \infty)$ . Conversely, if (ii) implies (i), extend  $\nu$  to a measure  $\mu$  on  $\mathbb{E}_\Gamma$  which satisfies (A.2).

**Remark A.1.** We can drop the condition that  $\mu$  does not concentrate on  $\{0\} \times (0, \infty)$  in statement (ii) of Proposition A.2, if we drop condition (i) from Definition 4.1 of the CEV model.

### Acknowledgements

B. Das was partially supported by RiskLab, Switzerland, and the IRTG *Stochastic Models for Complex Processes*, Berlin–Zurich. S. Resnick and A. Mitra were partially supported by the ARO, contract number W911NF-10-1-0289, at Cornell University. S. Resnick was also partially supported by the NSA, grant number H98230-11-1-0193, at Cornell. The authors would like to thank the anonymous referee for insightful comments which led to an improvement of the paper.

### References

- [1] BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd edn. John Wiley, New York.
- [2] BINGHAM, N. H., GOLDIE, C. M. AND TEUGELS, J. L. (1989). *Regular Variation* (Encyclopedia Math. Appl. 27). Cambridge University Press.
- [3] CAI, J.-J., EINMAHL, J. H. J. AND DE HAAN, L. (2011). Estimation of extreme risk regions under multivariate regular variation. *Ann. Statist.* **39**, 1803–1826.
- [4] DAS, B. AND RESNICK, S. (2011). Conditioning on an extreme component: model consistency with regular variation on cones. *Bernoulli* **17**, 226–252.
- [5] DE HAAN, L. AND FERREIRA, A. (2006). *Extreme Value Theory*. Springer, New York.
- [6] DE HAAN, L. AND RESNICK, S. (1993). Estimating the limit distribution of multivariate extremes. *Commun. Statist. Stoch. Models* **9**, 275–309.
- [7] HEFFERNAN, J. AND RESNICK, S. (2005). Hidden regular variation and the rank transform. *Adv. Appl. Prob.* **37**, 393–414.
- [8] HEFFERNAN, J. E. AND RESNICK, S. I. (2007). Limit laws for random vectors with an extreme component. *Ann. Appl. Prob.* **17**, 537–571.
- [9] HEFFERNAN, J. E. AND TAWN, J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *J. R. Statist. Soc. B* **66**, 497–546.
- [10] HULT, H. AND LINDSKOG, F. (2006). Regular variation for measures on metric spaces. *Publ. Inst. Math. (Beograd) (N.S.)* **80(94)**, 121–140.
- [11] JOE, H. AND LI, H. (2011). Tail risk of multivariate regular variation. *Methodology Comput. Appl. Prob.* **13**, 671–693.
- [12] KALLENBERG, O. (1983). *Random Measures*, 3rd edn. Akademie, Berlin.
- [13] LEDFORD, A. W. AND TAWN, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika* **83**, 169–187.
- [14] LEDFORD, A. W. AND TAWN, J. A. (1998). Concomitant tail behaviour for extremes. *Adv. Appl. Prob.* **30**, 197–215.
- [15] MAULIK, K. AND RESNICK, S. (2005). Characterizations and examples of hidden regular variation. *Extremes* **7**, 31–67.
- [16] MITRA, A. AND RESNICK, S. I. (2010). Hidden regular variation: detection and estimation. Preprint. Available at <http://arxiv.org/abs/1001.5058v2>.

- [17] MITRA, A. AND RESNICK, S. I. (2011). Hidden regular variation and detection of hidden risks. *Stoch. Models* **27**, 591–614.
- [18] MOLCHANOV, I. (2005). *Theory of Random Sets*. Springer, London.
- [19] PROHOROV, Y. V. (1956). Convergence of random processes and limit theorems in probability theory. *Teor. Veroyat. Primen.* **1**, 177–238.
- [20] RESNICK, S. (2002). Hidden regular variation, second order regular variation and asymptotic independence. *Extremes* **5**, 303–336.
- [21] RESNICK, S. (2007). *Heavy Tail Phenomena: Probabilistic and Statistical Modeling*. Springer, New York.
- [22] RESNICK, S. (2008). *Extreme Values, Regular Variation and Point Processes*. Springer, New York.