An Alternative Characterization of Hidden Regular Variation in Joint Tail Modeling

Grant B. Weller
Daniel S. Cooley

Department of Statistics, Colorado State University, Fort Collins, CO USA

June 12, 2012

Abstract

In modeling the joint upper tail of a multivariate distribution, a fundamental deficiency of classical extreme value theory is the inability to distinguish between asymptotic independence and exact independence. In this work, we examine multivariate threshold modeling based on the framework of regular variation on cones. Tail dependence is described by an angular measure, which in some cases is degenerate on joint tail regions despite strong sub-asymptotic dependence in such regions. The canonical example is a bivariate Gaussian distribution with any correlation less than one. Hidden regular variation (Resnick, 2002), a second-order tail decay on these regions, offers a refinement of the classical theory.

Previous characterizations of random vectors with hidden regular variation are not well-suited for joint tail estimation in finite samples, and estimation approaches thus far have been unable to model both the heavier-tailed regular variation and the hidden regular variation simultaneously. We propose to represent a random vector with hidden regular variation as the sum of independent first- and second-order regular varying pieces. We show our model is asymptotically valid via the concept of multivariate tail equivalence, and illustrate simulation methods with the bivariate Gaussian example. Finally, we outline a framework for estimation from our model via the EM algorithm.

1 Introduction

Classical multivariate extreme value theory provides a theoretical framework for describing the joint upper tail of a random vector. Modeling approaches based on classical theory are based on the limiting distribution of componentwise maxima. An extension to multivariate threshold exceedances is based on the framework of regular variation. This approach describes the limiting joint tail of a random vector as the product of a radial component which decays like a power function and an angular component governed by a limiting angular measure on the unit sphere under a chosen norm. Over the past 15 years, it has been recognized that such an approach can fail in applied modeling of joint tails. The fundamental shortcoming is that the first-order angular measure is degenerate on some joint tail regions, thus masking possible (and potentially strong) dependence structure at sub-asymptotic levels. Ledford and Tawn (1996) provided a first attempt at accounting for this sub-asymptotic dependence, using the example of the bivariate Gaussian distribution with correlation $\rho < 1$.

Following Ledford and Tawn (1996), many papers have offered refinements to the classical theory in attempts to resolve the flaw of the first-order limit. Ledford and Tawn (1997) and Ramos and Ledford (2009) focus specifically on modeling bivariate joint tails in the case that the first-order limit fails to capture dependence. Heffernan and Tawn (2004) offer a conditional approach, while Coles et al. (1999) examine measures of dependence in the asymptotic independence setting. Draisma et al. (2004) and Peng (1999) offer other approaches to joint tail estimation.

From a probabilistic perspective, the concept of hidden regular variation (Resnick, 2002) offers a mathematical structure for describing sub-asymptotic dependence, and is based on a generalization of the methods of Ledford and Tawn (1996, 1997). Hidden regular variation is essentially a second-order regular variation on regions where the first-order limit is degenerate. More treatment is given in Maulik and Resnick (2004), Heffernan and Resnick (2007), and Mitra and Resnick (2010). More recently, De Haan and Zhou (2011) offered a refinement on Ramos and Ledford (2009), offering an alternative polar coordinate transformation for modeling joint tails.

From a modeling standpoint, the joint tail approach of Ledford and Tawn (1997) and Ramos and Ledford (2009) fails to simultaneously account for the first-order limiting tail structure, and it is not immediately

clear how to extend such methods into dimension greater than two. Maulik and Resnick (2004) offer a representation of hidden regular variation as a mixture of a first-order and second-order component. While this provides an asymptotically valid characterization, a mixture representation is clumsy for finite samples and is difficult to justify intuitively.

In this work, we offer a characterization of a random vector with hidden regular variation as the sum of independent first- and second-order pieces. An alternative to Maulik and Resnick (2004), our characterization is more amenable to finite-sample representation and estimation. This representation is asymptotically justified via the concept of multivariate tail equivalence (Maulik and Resnick, 2004). When the hidden measure is finite, we can simulate realizations from our model; when the measure is infinite, we offer a slight adjustment to our simulation methods.

We first review the concepts of multivariate regular variation, hidden regular variation, and tail equivalence. To describe tail dependence in practice, one typically transforms each marginal distribution to a common, heavy-tailed marginal; often, the transformation is to unit Fréchet: $F_Z(z) = \exp\{-z^{-1}\}$ (Ledford and Tawn, 1997; Ramos and Ledford, 2009). The Fréchet marginal case is a special case of multivariate regular variation, which describes the joint tail as decaying like a power function. A decomposition into polar coordinates arises, and tail dependence can be characterized by a limiting angular measure. Hidden regular variation offers a second-order analogue of multivariate regular variation. A polar coordinate decomposition also arises; however, the resulting limiting angular measure is not guaranteed to be finite. Finally, it is through the concept of tail equivalence that we show the asymptotic validity of our model.

1.1 Multivariate Regular Variation

Multivariate regular variation on cones provides a probabilistic framework for describing tail dependence and modeling multivariate threshold exceedances. $\mathfrak{C}$ is a cone of $\mathbb{R}^d$ if for a set $A \in \mathfrak{C}$, $tA \in \mathfrak{C}$ for any $t > 0$. We assume some familiarity with regular variation of functions in the univariate case; the interested reader is referred to Bingham et al. (1989) and de Haan (1970). Let $M_+(\mathfrak{C})$ be the space of Radon measures on $\mathfrak{C}$. Following Resnick (2007), we say that a random vector $\mathbf{Z}$ with values in a subset of $[0, \infty)^d$ and common marginal distributions is regular varying on $\mathfrak{C} = [\mathbf{0}, \boldsymbol{\infty}] \setminus \{\mathbf{0}\}$ with finite limiting measure $\nu \neq 0$ if there exists a function $b(t) \uparrow \infty$ as $t \to \infty$ such that on $\mathfrak{C}$,

$$t\mathbb{P}\left[\frac{\mathbf{Z}}{b(t)} \in \cdot\right] \xrightarrow{v} \nu(\cdot), \tag{1}$$

in $M_+(\mathfrak{C})$ as $t \to \infty$ and $\xrightarrow{v}$ denotes vague convergence of measures (Resnick, 2007).

It follows that there exists $\alpha \geq 0$ such that the limiting measure ν in (1) has the scaling property

$$\nu(cA) = c^{-\alpha}\nu(A), \ c > 0 \tag{2}$$

for any relatively compact set $A \subset \mathfrak{C}$, where $\alpha > 0$ is called the tail index. The joint tail power-function behavior can be seen in (2). The function $b(t)$ is regular varying of order $1/\alpha$, which following Resnick (2002) we denote $b(t) \in RV_{1/\alpha}$.

The homogeneity property (2) suggests a transformation to polar coordinates. Let $\|\cdot\|$ be any norm on $\mathfrak{C}$, and consider the unit sphere $\aleph = \{\mathbf{z} \in \mathfrak{C} : \|\mathbf{z}\| = 1\}$. Define the bijective transformation $T : \mathbb{R}^d \to [0, \infty) \times \aleph$ via $T(\mathbf{Z}) = (\|\mathbf{Z}\|, \mathbf{Z}\|\mathbf{Z}\|^{-1})$. The measure ν can then be expressed in terms of the new coordinate system $(r, \boldsymbol{\theta})$ via

$$\nu = \nu_\alpha \times H \tag{3}$$

where ν_α is a Pareto measure; i.e. $\nu_\alpha((x, \infty]) = cx^{-\alpha}$, $c > 0$, and H is a non-negative measure on $\aleph$ satisfying the balance condition

$$\int_\aleph \theta_1 H(d\boldsymbol{\theta}) = \int_\aleph \theta_j H(d\boldsymbol{\theta}), \qquad j = 2, ..., d. \tag{4}$$

This condition results from the fact that $\mathbf{Z}$ is assumed to have common marginals; this is what Mitra and Resnick (2010) call the *standard case*. In practice, transformations can be applied to give common marginal tail behavior; see Resnick (2007). the set $\boldsymbol{\Theta} = \{\mathbf{z} \in \mathbb{E} : \|\mathbf{z}\| = 1\}$. H is called the *spectral* measure or *angular* measure. When $\alpha = 1$, a common choice for $\|\cdot\|$ is the L_1 norm, in which case H is a measure on the unit

simplex $\mathcal{N} = \Delta_{d-1} = \{\boldsymbol{z} \in \mathfrak{C} : z_1 + ... + z_d = 1\}$ (Coles and Tawn, 1991; Ballani and Schlather, 2011; Cooley et al., 2010).

The polar coordinate representation (3) of the measure ν can be seen by considering, for any $r > 0$ and Borel set $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

$$
\begin{aligned}
\nu(\{\boldsymbol{z} \in \mathbb{E} : \|\boldsymbol{z}\| > r, \|\boldsymbol{z}\|^{-1}\boldsymbol{z} \in \boldsymbol{\theta}\}) &= r^{-\alpha}\nu(\{r^{-1}\boldsymbol{z} : \|\boldsymbol{z}\| > r, \|\boldsymbol{z}\|^{-1}\boldsymbol{z} \in \boldsymbol{\theta}\}) \\
&= r^{-\alpha}\nu(\{r^{-1}\boldsymbol{z} : \|r^{-1}\boldsymbol{z}\| > 1, \|r^{-1}\boldsymbol{z}\|^{-1}(r^{-1}\boldsymbol{z}) \in \boldsymbol{\theta}\}) \\
&= r^{-\alpha}\nu(\{\boldsymbol{y} \in \mathbb{E} : \|\boldsymbol{y}\| > 1, \|\boldsymbol{y}\|^{-1}\boldsymbol{y} \in \boldsymbol{\theta}\}) \\
&= r^{-\alpha}H(\boldsymbol{\theta}),
\end{aligned}
$$

where $\boldsymbol{y} = r^{-1}\boldsymbol{z}$. Thus with respect to the coordinate system $(\|\mathbf{y}\|, \mathbf{y}\|\mathbf{y}\|^{-1})$, we have that ν is a product measure. Finally, we note that by appropriate choice of normalizing function $b(t)$, $H(\cdot)$ can be made to be a probability measure.

1.2 Hidden Regular Variation and Tail Equivalence

It is possible that the limiting measure ν in (1) places zero mass on pie-shaped regions $\{\mathbf{z} \in \mathfrak{C} : \mathbf{z}\|\mathbf{z}\|^{-1} \subset \mathcal{N}\}$ of the cone $\mathfrak{C}$. In such cases, the normalizing function $b(t)$ obliterates any finer structure of the random variable on such regions, if such a finer structure exists. The corresponding angular measure H thus places zero mass on corresponding regions of the unit sphere $\mathcal{N}$. A classic example is the joint upper tail of a multivariate normal random variable with correlations less than one (Ledford and Tawn, 1996). This prompted Resnick (2002) to formulate the concept of hidden regular variation.

Consider a subcone $\mathfrak{C}_0 \subset \mathfrak{C}$ with $\nu(\mathfrak{C}_0) = 0$. A random vector $\mathbf{Z}$ is said to possess hidden regular variation if, in addition to (1), there exists a non-decreasing function $b_0(t) \uparrow \infty$ with $b(t)/b_0(t) \to \infty$ such that

$$
t\mathbb{P}\left[\frac{\mathbf{Z}}{b_0(t)} \in \cdot\right] \xrightarrow{v} \nu_0(\cdot) \tag{5}
$$

as $t \to \infty$ in $M_+(\mathfrak{C}_0)$. The measure ν_0 decomposes into a product of Pareto measure ν_{α_0} and positive Radon measure H_0 on $\mathcal{N}_0 = \mathcal{N} \cap \mathfrak{C}_0$. The function $b_0(t) \in RV_{1/\alpha_0}$, with $\alpha_0 > \alpha$; thus, $\mathbf{Z}$ has a lighter tail on $\mathfrak{C}_0$ than on $\mathfrak{C}$. As $\mathcal{N}_0$ may not be a relatively compact set of $\mathcal{N}$, H_0 may be either finite or infinite; see (Resnick, 2002; Maulik and Resnick, 2004; De Haan and Zhou, 2011) for details. Finally, ν_0 is homogeneous with tail index α_0, analogous to (2).

The concept of multivariate tail equivalence was introduced in Maulik and Resnick (2004). Consider random vectors $\mathbf{Y}$ and $\mathbf{Z}$ taking values in $[0, \infty)$ with distribution functions F and G, respectively. $\mathbf{Y}$ and $\mathbf{Z}$ are said to be *tail equivalent* on the cone $\mathfrak{C}^* \subseteq \mathfrak{C} = [\mathbf{0}, \infty] \setminus \{\mathbf{0}\}$ if there exists a scaling function $b^*(t) \uparrow \infty$ such that

$$
t\mathbb{P}\left[\frac{\mathbf{Y}}{b^*(t)} \in \cdot\right] \xrightarrow{v} \nu_*(\cdot) \quad \text{and} \quad t\mathbb{P}\left[\frac{\mathbf{Z}}{b^*(t)} \in \cdot\right] \xrightarrow{v} c\nu_*(\cdot) \tag{6}
$$

in $M_+(\mathfrak{C}^*)$ for some constant $c \in (0, \infty)$ and measure ν_* on $\mathfrak{C}^*$. The definition (6) implies that the extremes of samples from $\mathbf{Y}$ and $\mathbf{Z}$ have the same asymptotic properties on $\mathfrak{C}^*$, up to a scaling constant. Following Maulik and Resnick we write $\mathbf{Y} \stackrel{te(\mathfrak{C}^*)}{\sim} \mathbf{Z}$.

The remainder of the paper is structured as follows: in Section 2 we describe the construction of our sum representation and show that our representation is tail equivalent to a random vector with hidden regular variation. Section 3 demonstrates simulation from our representation for a random vector with Gaussian dependence structure. We conclude in Section 4 with a discussion and framework for estimation of our model.

2 Regular Varying Sum Representation

To represent random vectors possessing hidden regular variation, Maulik and Resnick (2004) consider mixtures of random vectors with differing tail indices and angular measures. Maulik and Resnick show such mixtures are tail equivalent to a random vector possessing hidden regular variation. In contrast to Maulik

and Resnick, here we consider sums of such random vectors. We show that such sums can be constructed to be tail equivalent on both $\mathfrak{C}$ and $\mathfrak{C}_0$ to a random vector with hidden regular variation.

Consider a random vector $\mathbf{Z} \in [0,\infty)^d$ which is multivariate regular varying on $\mathfrak{C}$ with limit measure ν; that is for some function $b(t) \uparrow \infty$,

$$t\mathbb{P}\left[\frac{\mathbf{Z}}{b(t)} \in \cdot\right] \xrightarrow{v} \nu(\cdot) \tag{7}$$

as $t \to \infty$ in $M_+(\mathfrak{C})$. Without loss of generality, assume that the resulting angular measure H is a probability measure.

Further assume that $\mathbf{Z}$ exhibits hidden regular variation on a subcone $\mathfrak{C}_0 \subset \mathfrak{C}$. That is, $\nu(\mathfrak{C}_0) = 0$ and there exists a function $b_0(t) \uparrow \infty$ with $b(t)/b_0(t) \to \infty$ such that

$$t\mathbb{P}\left[\frac{\mathbf{Z}}{b_0(t)} \in \cdot\right] \xrightarrow{v} \nu_0(\cdot) \tag{8}$$

as $t \to \infty$ in $M_+(\mathfrak{C}_0)$.

We construct a sum of regular varying random vectors that is tail equivalent to $\mathbf{Z}$ on both the full cone $\mathfrak{C}$ and the subcone $\mathfrak{C}_0$. Define a random vector $\mathbf{Y} = R\mathbf{W}$ taking values in $[0,\infty)^d$, where $\mathbb{P}(R > r) \sim 1/b^{\leftarrow}(r)$ as $r \to \infty$ and $\mathbf{W} \sim H(\cdot)$, where H and $b(t)$ are as above. Recall that $\nu(\mathfrak{C}_0) = 0$ and thus $H(\mathcal{N} \cap \mathfrak{C}_0) = 0$. We thus define $\mathbf{W}$ to be such that $\mathbb{P}[\mathbf{W} \in A] = 0$ if $H(A) = 0$. Assume that the quantities R and $\mathbf{W}$ are independent. It follows that, on $\mathfrak{C}$

$$t\mathbb{P}\left[\frac{\mathbf{Y}}{b(t)} \in \cdot\right] \xrightarrow{v} \nu(\cdot) \tag{9}$$

as $t \to \infty$ (Maulik and Resnick, 2004).

Now consider a random vector $\mathbf{E} \in (0,\infty]^d$ defined on the same probability space and independent of R and $\mathbf{W}$ which is multivariate regular varying on $\mathfrak{C}_0$ with tail index $\alpha_0 > \alpha$. Assume

$$\mathbb{P}(\|\mathbf{E}\| > r) \sim 1/b_0^{\leftarrow}(r) \text{ as } r \to \infty,$$

and

$$t\mathbb{P}\left[\frac{\mathbf{E}}{b_0(t)} \in \cdot\right] \xrightarrow{v} \nu_0, \tag{10}$$

in $M_+(\mathfrak{C}_0)$, where ν_0 is as above. Further assume that $\mathbf{E}$ is not regular varying of any index $\alpha^* \leq \alpha$ on $\mathfrak{C} \setminus \mathfrak{C}_0$.

To review, we construct $\mathbf{Y}$ to be regular varying with tail index α with support on $\mathfrak{C} \setminus \mathfrak{C}_0$; $\mathbf{Y}$ has no hidden regular variation on $\mathfrak{C}_0$. $\mathbf{E}$ is regular varying on the subcone $\mathfrak{C}_0$ with tail index $\alpha_0 > \alpha$ and limit measure ν_0; that is, $\mathbf{E}$ has the same tail behavior as $\mathbf{Z}$ on $\mathfrak{C}_0$. It can be shown that mixtures of $\mathbf{Y}$ and $\mathbf{E}$ are tail equivalent to $\mathbf{Z}$ on both $\mathfrak{C}$ and $\mathfrak{C}_0$; see (Resnick, 2002; Maulik and Resnick, 2004; Mitra and Resnick, 2010). In many applications, it may be more natural to represent $\mathbf{Z}$ as a sum of the random vectors $\mathbf{Y}$ and $\mathbf{E}$.

Next we show

$$\mathbf{Z} \stackrel{te(\mathfrak{C})}{\sim} \mathbf{Y} + \mathbf{E} \qquad \text{and} \tag{11}$$

$$\mathbf{Z} \stackrel{te(\mathfrak{C}_0)}{\sim} \mathbf{Y} + \mathbf{E}. \tag{12}$$

The result (11) follows from Jessen and Mikosch (2006); we review the proof below. Following a similar argument, we prove (12).

With $\mathbf{Y}$ and $\mathbf{E}$ defined above, we adapt Lemma 3.12 of Jessen and Mikosch (2006) to show tail equivalence on the full cone $\mathfrak{C}$. Consider a relatively compact rectangle $A \in \mathfrak{C}$; that is, A is bounded away from $\mathbf{0}$. This class of sets A generates vague convergence in $\mathfrak{C}$; thus it is sufficient to show

$$\lim_{t \to \infty} t\mathbb{P}\left[\frac{\mathbf{Y} + \mathbf{E}}{b(t)} \in A\right] = \lim_{t \to \infty} t\mathbb{P}\left[\frac{\mathbf{Z}}{b(t)} \in A\right] = \nu(A). \tag{13}$$

Assume without loss of generality that $A = [\mathbf{a}, \mathbf{b}] = \{\mathbf{x} \in \mathfrak{C} : \mathbf{a} \le \mathbf{x} \le \mathbf{b}\}$. For small $\epsilon > 0$, define $\mathbf{a}^{-\epsilon} = (a_1^{-\epsilon}, ..., a_d^{-\epsilon})$, and define $\mathbf{b}^{-\epsilon}$ analogously. Define the rectangles $A^{-\epsilon} = [\mathbf{a}^{-\epsilon}, \mathbf{b}]$ and $A^{\epsilon} = [\mathbf{a}, \mathbf{b}^{-\epsilon}]$. For small ϵ, the rectangles A^{ϵ} and $A^{-\epsilon}$ are relatively compact in $\mathbb{E}$, and $A^{\epsilon} \subset A \subset A^{-\epsilon}$. Finally we note that $\nu(\partial A) = 0$; there is no mass on the edges of A.

For small $\epsilon > 0$ and fixed $t > 0$,

$$\mathbb{P}\left[\frac{\mathbf{Y} + \mathbf{E}}{b(t)} \in A\right] = \mathbb{P}\left[\frac{\mathbf{Y} + \mathbf{E}}{b(t)} \in A, \frac{\|\mathbf{E}\|}{b(t)} > \epsilon\right] + \mathbb{P}\left[\frac{\mathbf{Y} + \mathbf{E}}{b(t)} \in A, \frac{\|\mathbf{E}\|}{b(t)} \le \epsilon\right]$$

$$\le \mathbb{P}\left[\|\mathbf{E}\| > b(t)\epsilon\right] + \mathbb{P}\left[\frac{\mathbf{Y}}{b(t)} \in A^{-\epsilon}\right].$$

Thus

$$\limsup_{t \to \infty} t\mathbb{P}\left[\frac{\mathbf{Y} + \mathbf{E}}{b(t)} \in A\right] \le \limsup_{t \to \infty} t\mathbb{P}\left[\|\mathbf{E}\| > b(t)\epsilon\right] + \limsup_{t \to \infty} t\mathbb{P}\left[\frac{\mathbf{Y}}{b(t)} \in A^{-\epsilon}\right]$$

$$= \limsup_{t \to \infty} t^{1 - \alpha_0/\alpha} \epsilon^{-\alpha_0} + \limsup_{t \to \infty} t\mathbb{P}\left[\frac{\mathbf{Y}}{b(t)} \in A^{-\epsilon}\right]$$

$$= \nu(A^{-\epsilon}) \searrow \nu(A) \text{ as } \epsilon \to 0.$$

For the lower bound, recognize

$$\mathbb{P}\left[\frac{\mathbf{Y} + \mathbf{E}}{b(t)} \in A\right] \ge \mathbb{P}\left[\frac{\mathbf{Y}}{b(t)} \in A^{\epsilon}, \frac{\|\mathbf{E}\|}{b(t)} \le \epsilon\right]$$

$$\ge \mathbb{P}\left[\frac{\mathbf{Y}}{b(t)} \in A^{\epsilon}\right] - \mathbb{P}\left[\|\mathbf{E}\| > b(t)\epsilon\right],$$

and so

$$\liminf_{t \to \infty} t\mathbb{P}\left[\frac{\mathbf{Y} + \mathbf{E}}{b(t)} \in A\right] \ge \liminf_{t \to \infty} t\mathbb{P}\left[\frac{\mathbf{Y}}{b(t)} \in A^{\epsilon}\right] - \liminf_{t \to \infty} t\mathbb{P}\left[\|\mathbf{E}\| > b(t)\epsilon\right]$$

$$= \nu(A^{\epsilon}) \nearrow \nu(A) \text{ as } \epsilon \to 0.$$

Collecting the upper and lower bounds, and using the fact that A is a ν-continuity set, we achieve the desired result

$$t\mathbb{P}\left[\frac{\mathbf{Y} + \mathbf{E}}{b(t)} \in \cdot\right] \xrightarrow{v} \nu(\cdot)$$

in $M_+(\mathfrak{C})$.

2.1 Tail Equivalence on $\mathfrak{C}_0$

For (12) to hold, it is sufficient to show the following result:

Theorem. *For $\mathbf{Y}$, $\mathbf{E}$, $b_0(t)$, and ν_0 defined as above,*

$$t\mathbb{P}\left[\frac{\mathbf{Y} + \mathbf{E}}{b_0(t)} \in \cdot\right] \xrightarrow{v} \nu_0(\cdot) \tag{14}$$

as $t \to \infty$ in $M_+(\mathfrak{C}_0)$.

Proof. It suffices to consider any rectangle A_0 which is relatively compact in $\mathfrak{C}_0$, and show that

$$\lim_{t \to \infty} t\mathbb{P}\left[\frac{\mathbf{Y} + \mathbf{E}}{b_0(t)} \in A_0\right] = \nu_0(A_0).$$

Without loss of generality assume $A_0 = [\mathbf{c}, \mathbf{d}] = \{\mathbf{x} \in \mathfrak{C}_0 : \mathbf{c} \le \mathbf{x} \le \mathbf{d}\}$. For small $\epsilon > 0$, define the rectangles $A_0^{-\epsilon} = [\mathbf{c}^{-\epsilon}, \mathbf{d}]$ and $A_0^{\epsilon} = [\mathbf{c}, \mathbf{d}^{-\epsilon}]$, with $\mathbf{c}^{-\epsilon}$ and $\mathbf{d}^{-\epsilon}$ defined analogously to $\mathbf{a}^{-\epsilon}$ and $\mathbf{b}^{-\epsilon}$ above. For small ϵ, A_0^{ϵ} and $A_0^{-\epsilon}$ are relatively compact in $\mathfrak{C}_0$, $A_0^{\epsilon} \subset A_0 \subset A_0^{-\epsilon}$, and $\nu_0(\partial A_0) = 0$.

Recognize that for small $\epsilon > 0$ and fixed $t > 0$,

$$\mathbb{P}\left[\frac{\mathbf{E}}{b_0(t)} \in A_0^\epsilon\right] = \mathbb{P}\left[\frac{\mathbf{E}}{b_0(t)} \in A_0^\epsilon, \frac{\|\mathbf{Y}\|}{b_0(t)} \le \epsilon\right] + \mathbb{P}\left[\frac{\mathbf{E}}{b_0(t)} \in A_0^\epsilon, \frac{\|\mathbf{Y}\|}{b_0(t)} > \epsilon\right]$$

$$\le \mathbb{P}\left[\frac{\mathbf{Y}+\mathbf{E}}{b_0(t)} \in A_0\right] + \mathbb{P}\left[\frac{\|\mathbf{E}\|}{b_0(t)} \ge \|\mathbf{c}\|, \frac{\|\mathbf{Y}\|}{b_0(t)} > \epsilon\right].$$

Thus by definition of $\mathbf{Y}$ and $\mathbf{E}$ and independence,

$$\liminf_{t\to\infty} t\mathbb{P}\left[\frac{\mathbf{Y}+\mathbf{E}}{b_0(t)} \in A_0\right] \ge \liminf_{t\to\infty} t\mathbb{P}\left[\frac{\mathbf{E}}{b_0(t)} \in A_0^\epsilon\right] - \liminf_{t\to\infty} t\mathbb{P}\left[\frac{\|\mathbf{E}\|}{b_0(t)} \ge \|\mathbf{c}\|\right]\mathbb{P}\left[\frac{\|\mathbf{Y}\|}{b_0(t)} > \epsilon\right]$$

$$= \nu_0(A_0^\epsilon) - \liminf_{t\to\infty} t(t^{-1}\|\mathbf{c}\|^{-\alpha_0})(t^{-\alpha/\alpha_0}\epsilon^{-\alpha})$$

$$= \nu_0(A_0^\epsilon) \nearrow \nu_0(A_0) \text{ as } \epsilon \to 0,$$

since A_0 is a ν_0-continuity set.

For the upper bound, we employ the fact that $H(\mathfrak{C}_0) = 0$. For fixed t,

$$\mathbb{P}\left[\frac{\mathbf{Y}+\mathbf{E}}{b_0(t)} \in A_0\right] = \mathbb{P}\left[\frac{\mathbf{Y}+\mathbf{E}}{b_0(t)} \in A_0, \frac{\|\mathbf{Y}\|}{b_0(t)} \le \epsilon\right] + \mathbb{P}\left[\frac{\mathbf{Y}+\mathbf{E}}{b_0(t)} \in A_0, \frac{\|\mathbf{Y}\|}{b_0(t)} > \epsilon\right]$$

$$= I + II$$

Notice that I is bounded above by

$$\mathbb{P}\left[\frac{\mathbf{E}}{b_0(t)} \in A_0^{-\epsilon}\right].$$

Recalling that by construction $\mathbb{P}[\mathbf{Y}/b_0(t) \in A^{-\epsilon}] = 0$,

$$II = \mathbb{P}\left[\frac{\mathbf{Y}+\mathbf{E}}{b_0(t)} \in A_0, \frac{\|\mathbf{Y}\|}{b_0(t)} > \epsilon, \frac{\mathbf{E}}{b_0(t)} \in A_0, \frac{\mathbf{Y}}{b_0(t)} \notin A_0^{-\epsilon}\right]$$

$$+ \mathbb{P}\left[\frac{\mathbf{Y}+\mathbf{E}}{b_0(t)} \in A_0, \frac{\|\mathbf{Y}\|}{b_0(t)} > \epsilon, \frac{\mathbf{E}}{b_0(t)} \notin A_0, \frac{\mathbf{Y}}{b_0(t)} \notin A_0^{-\epsilon}\right]$$

$$\le \mathbb{P}\left[\frac{\|\mathbf{E}\|}{b_0(t)} \ge \|\mathbf{c}\|, \frac{\|\mathbf{Y}\|}{b_0(t)} > \epsilon\right] + \mathbb{P}\left[\frac{\vee_{i=1}^d E_i}{b_0(t)} > \epsilon, \frac{\|\mathbf{Y}\|}{b_0(t)} > \epsilon\right].$$

Then

$$\limsup_{t\to\infty} t\mathbb{P}\left[\frac{\mathbf{Y}+\mathbf{E}}{b_0(t)} \in A_0\right] \le \limsup_{t\to\infty} t\mathbb{P}\left[\frac{\mathbf{E}}{b_0(t)} \in A_0^{-\epsilon}\right] + \limsup_{t\to\infty} t\mathbb{P}\left[\frac{\|\mathbf{E}\|}{b_0(t)} \ge \|\mathbf{c}\|\right]\mathbb{P}\left[\frac{\|\mathbf{Y}\|}{b_0(t)} > \epsilon\right]$$

$$+ \limsup_{t\to\infty} t\mathbb{P}\left[\frac{\vee_{i=1}^d E_i}{b_0(t)} > \epsilon\right]\mathbb{P}\left[\frac{\|\mathbf{Y}\|}{b_0(t)} > \epsilon\right]$$

$$= \nu_0(A_0^{-\epsilon}) + \limsup_{t\to\infty} t(t^{-1}\|\mathbf{c}\|^{-\alpha_0})(t^{-\alpha/\alpha_0}\epsilon^{-\alpha}) + \limsup_{t\to\infty} t(t^{-1}\epsilon^{-\alpha_0})(t^{-\alpha/\alpha_0}\epsilon^{-\alpha})$$

$$= \nu_0(A_0^{-\epsilon}) \searrow \nu_0(A_0) \text{ as } \epsilon \to 0,$$

by independence and ν_0-continuity of A_0.

Finally, putting together the upper and lower bounds yields the desired result (14). $\square$

Remark 1. *Heuristically, the scaled random vector $(\mathbf{Y}+\mathbf{E})/b_0(t)$ can only land in $\mathfrak{C}_0$ when $\|\mathbf{Y}\|$ is small and $\|\mathbf{E}\|$ is large. Suitably normalized large values of $\mathbf{Y}$ will converge to points outside of $\mathfrak{C}_0$, and by independence, the probability of $\mathbf{Y}$ and $\mathbf{E}$ being simultaneously large is asymptotically negligible.*

Remark 2. *The proof relies on $\mathbf{Y}$ being constructed in such a way that $\mathbb{P}[\mathbf{Y} \in \mathfrak{C}_0] = 0$. Such a condition gives convergence to the measure ν_0 on $\mathfrak{C}_0$. The result may not hold in general if $\mathbf{Y}$ has angular measure H only in the limit and exhibits hidden regular variation on $\mathfrak{C}_0$. We do not impose such additional conditions on the construction of $\mathbf{E}$.*

Remark 3. *Because the measure H can be made to be a probability measure, simulation of realizations of the random vector $\mathbf{Y}$ is often quite tractable, especially in low dimensions. The angular measure H_0 of $\mathbf{E}$ may be infinite on $\mathfrak{C}_0$, making simulation more difficult. In some cases, H_0 can be made to be a probability measure under an alternative transformation (Mitra and Resnick, 2010; De Haan and Zhou, 2011). This still may pose difficulty in simulation; see Section 3 for an example.*

3   Example: Bivariate Gaussian Dependence

We now demonstrate an example of $\mathbf{Z}$ for which we can simulate a tail equivalent representation $\mathbf{Y} + \mathbf{E}$. We explore the case of asymptotic independence plus hidden regular variation in dimension $d = 2$, with the bivariate normal distribution as the classical example (Ledford and Tawn, 1996). Through simulation, we see that the sum representation results in a more realistic representation of the random vector $\mathbf{Z}$. Here the resulting hidden measure is not finite, and difficulty arises near the boundaries of the subcone. We review previous attempts to address this difficulty, and propose a novel method which accommodates our sum representation.

Up to this point, we have considered polar coordinate transformations defined by any norm $\|\cdot\|$ on $\mathfrak{C}$. Unless noted otherwise, in this section we consider the L_1 norm transformation defined by $r = z_1 + z_2$ and $w = z_1/r$. This transformation is common in most multivariate extreme value analyses with Fréchet marginal variables.

Consider the bivariate random vector $\mathbf{Z} = (Z_1, Z_2)$, where $Z_i = -1/\log \Phi(X_i)$, $i = 1, 2$, and (X_1, X_2) follow a bivariate Gaussian distribution with correlation $\rho < 1$. Here $\Phi(\cdot)$ is the standard Gaussian distribution function. Sibuya (1960) showed that asymptotic independence holds; i.e., we can find $b(t) \in RV_1$ such that

$$t\mathbb{P}\left[\frac{\mathbf{Z}}{b(t)} \in \cdot\right] \xrightarrow{v} \nu = \nu_1 \times H, \tag{15}$$

in $M_+(\mathfrak{C})$, where H consists of point masses at the axes $\mathcal{N} \cap \{\mathbf{x} \in \mathfrak{C} : x_1 \wedge x_2 = 0\}$. If $b(t) = 2t$, H is a probability measure with point masses of $1/2$ at $w = 0$ and $w = 1$.

An exploration of the second-order regular variation of $\mathbf{Z}$ was provided by Ledford and Tawn (1996, 1997). Ledford and Tawn formulate this in terms of the joint survivor function $\bar{F}(z_1, z_2) := \mathbb{P}[Z_1 > z_1, Z_2 > z_2]$. Ledford and Tawn (1997) show

$$\bar{F}(z_1, z_2) \sim (z_1 z_2)^{-1/(1+\rho)} \mathcal{L}(z_1, z_2; \rho)(1 + O[1/\log\{\min(z_1, z_2)\}]), \tag{16}$$

where $\mathcal{L}(z_1, z_2)$ is a slowly varying function given by (Ledford and Tawn, 1996) satisfies

$$\lim_{t \to \infty} \frac{\mathcal{L}(tz_1, tz_2)}{\mathcal{L}(t, t)} = 1.$$

Here, the function $g(z_1, z_2) \equiv 1$ (Ledford and Tawn, 1997). Ledford and Tawn (1996) derive

$$\mathcal{L}(t, t; \rho) = (1 + \rho)^{3/2}(1 - \rho)^{-1/2}(4\pi \log t)^{-\rho/(1+\rho)}. \tag{17}$$

The random vector $\mathbf{Z}$ also exhibits hidden regular variation. Consider a set $[\mathbf{z}, \boldsymbol{\infty}]$ for $\mathbf{z} = (z_1, z_2)$ with $z_1, z_2 > 0$. One can show

$$t\mathbb{P}\left[\frac{\mathbf{Z}}{b_0(t)} \in [\mathbf{z}, \boldsymbol{\infty}]\right] \longrightarrow (z_1 z_2)^{-1/2\eta} =: \nu_0([\mathbf{z}, \boldsymbol{\infty}]) \tag{18}$$

as $t \to \infty$, where the function $b_0(t) := 2U^{\leftarrow}(t)$, with

$$U(t) = \frac{(2t)^{1/\eta}}{\mathcal{L}(2t, 2t)}, \tag{19}$$

for $\mathcal{L}$ given by (17). Ledford and Tawn refer to $\eta = (1 + \rho)/2 \in (0, 1]$ as the coefficient of tail dependence.

7

It is easily shown for sets of the form $A(r, B) = \{\mathbf{z} \in \mathfrak{C}_0 : \|\mathbf{z}\| > r, \mathbf{z}\|\mathbf{z}\|^{-1} \in B\}$ that

$$\nu_0(dr \times dw) = r^{-1-1/\eta} H_0(dw),$$

where

$$H_0(dw) = \frac{1}{4\eta}\{w(1-w)\}^{-1-1/2\eta}; \tag{20}$$

see, for example, Beirlant et al., 2004, chapter 9. Note that $H_0(\mathcal{N}_0) = \int_{(0,1)} H_0(dw) = +\infty$, thus the hidden measure ν_0 is infinite on $\mathfrak{C}_0$.

The fact that the hidden angular measure is infinite poses difficulty in finite-sample simulation of the joint tail of $\mathbf{Z}$. Because the hidden measure diverges near the endpoints of $\mathcal{N}_0$, one always encounters difficulty near the axes of $\mathfrak{C}$. Several authors have explored ways to remedy this problem. Mitra and Resnick (2010) propose an alternative transformation $\tilde{T}(z_1, z_2) = (z_{(2)}, \mathbf{z}/z_{(2)})$, where $z_{(2)} = \min(z_1, z_2)$. The limit measure ν_0 can then be decomposed into the product of a Pareto measure and probability measure on $\tilde{\mathcal{N}} = \{\mathbf{z} \in \mathfrak{C}_0 : z_{(2)} = 1\}$. However, this approach avoids behavior near the axes, and it is not clear how one would simulate random vectors which are tail equivalent to $\mathbf{Z}$ under this alternative representation.

More recently, De Haan and Zhou (2011) offered an alternative for characterizing the random vector $\mathbf{Z}$. De Haan and Zhou cleverly define a transformation on $\mathfrak{C}_0$ via $T^*(z_1, z_2) = (s, v)$, where $s = (z_1^{-1} + z_2^{-1})^{-1}$ and $v = s/z_1$. They then show that the limiting measure of the joint tail of the normalized vector $\mathbf{Z}^{1/\eta}$ can be decomposed into the product of a Pareto measure and a finite measure H^* on $\mathcal{N}^* = \{\mathbf{z} \in \mathfrak{C}_0 : (z_1^{-1} + z_2^{-1})^{-1} = 1\}$. Specifically for the Gaussian dependence example, H^* is proportional to a Beta distribution with parameters $(1/2, 1/2)$. While De Haan and Zhou (2011) offer a method for constructing a random vector which is tail equivalent to $\mathbf{Z}$, their simulation method still encounters problems near the axes of $\mathfrak{C}$, which we illustrate in Section 3.1.

3.1 Simulation from Sum Representation

Because the hidden measure with density $H_0(dw)$ given by (20) is infinite on $(0, 1)$, one cannot simulate from it directly. As an alternative, we propose an approximation to $H_0(dw)$ by restricting the subcone $\mathfrak{C}_0$ to $\mathfrak{C}_0^\epsilon = \{\mathbf{z} \in \mathfrak{C}_0 : z_1\|\mathbf{z}\|^{-1} \in \mathcal{N}_0^\epsilon\}$, where $\mathcal{N}_0^\epsilon = [\epsilon, 1-\epsilon]$ for some $\epsilon \in (0, 1/2)$. The density (20) can then be made to be a probability density on $\mathcal{N}_0^\epsilon$ via $H_0^\epsilon(dw) = H_0(dw)/H_0(\mathcal{N}_0^\epsilon)$ for $w \in \mathcal{N}_0^\epsilon$. One can then simulate realizations from H_0^ϵ via an accept-reject algorithm or other sampling method.

We proceed to simulate realizations of $\hat{\mathbf{Z}} = \mathbf{Y} + \hat{\mathbf{E}}$ which is tail equivalent to $\mathbf{Z}$ on $\mathfrak{C}$ and $\mathfrak{C}_0^\epsilon$. Define $\mathbf{Y}$ as follows: let R follow a Pareto distribution with $\mathbb{P}[R > r] = 2/r$ for $r \geq 2$. Draw a Bernoulli(0.5) random variable B independently of R, and let $Y_1 = RB$, $Y_2 = R(1 - B)$. For a fixed sample size n, draw R_0 independently of $\mathbf{Y} = (Y_1, Y_2)^T$ with R_0 such that

$$\mathbb{P}[R_0 > x] = \begin{cases} d_{\epsilon,n} x^{-1/\eta} & \text{if } x > (d_{\epsilon,n})^\eta \\ 1 & \text{otherwise,} \end{cases} \quad \text{where } d_{\epsilon,n} = (2U^\leftarrow(n))^{1/\eta}\left\{\frac{H_0(\mathcal{N}_0^\epsilon)}{n}\right\}.$$

Draw n independent realizations of W_0 from the density $H_0^\epsilon(dw)$ independently of $\mathbf{Y}$ and R_0. Define $\hat{\mathbf{E}} = (\hat{E}_1, \hat{E}_2)$ via

$$\hat{E}_1 = R_0 W_0 \quad \text{and} \quad \hat{E}_2 = R_0(1 - W_0).$$

Then for any set $A(r, B) = \{\mathbf{z} \in \mathfrak{C}_0^\epsilon : \|\mathbf{z}\| > r, z_1\|\mathbf{z}\|^{-1} \in B\}$ with B a Borel set of $\mathcal{N}_0^\epsilon$,

$$
\begin{aligned}
n\mathbb{P}\left[\frac{\hat{\mathbf{E}}}{b_0(n)} \in A_0(r, B)\right] &= n\mathbb{P}\left[\frac{R_0}{2U^\leftarrow(n)} > r, W_0 \in B\right] \\
&= n\mathbb{P}\left[R_0 > 2rU^\leftarrow(n)\right]\mathbb{P}[W_0 \in B] \\
&= n\left[d_{\epsilon,n}(2rU^\leftarrow(n))^{-1/\eta}\right]\frac{H_0(B)}{H_0(\mathcal{N}_0^\epsilon)} \\
&= r^{-1/\eta}H_0(B)
\end{aligned}
$$

8

for $r > \{H_0(\mathcal{N}_0^\epsilon)/n\}^\eta$, which is precisely the decomposition of ν_0 in (18).

When examining the limiting measure of a set in the full subcone $\mathfrak{C}_0$ which is not completely contained in $\mathfrak{C}_0^\epsilon$, a bias is induced by the choice of ϵ. To see this, extend the restricted hidden measure via $H_0^\epsilon\{(0,\epsilon)\} = H_0^\epsilon\{(1-\epsilon,1)\} \equiv 0$, and consider a set $A = [\mathbf{z}, \infty]$ for $z_1, z_2 > 0$. Note that one can choose n and ϵ such that $\mathbf{z} \in \mathfrak{C}_0^\epsilon$ and $z_1 + z_2 > \{H_0(\mathcal{N}_0^\epsilon)/n\}^\eta$, and in this case we have

$$
\begin{aligned}
n\mathbb{P}\left[\frac{\hat{\mathbf{E}}}{b^*(n)} \in A\right] &= n \int_0^1 \int_{\frac{b^*(n)z_1}{w} \vee \frac{b^*(n)z_2}{1-w}}^{\infty} \eta^{-1} d_{\epsilon,n} r^{-(1+1/\eta)} dr H_0^\epsilon(dw) \\
&= \int_0^1 \left\{\frac{w}{z_1} \bigwedge \frac{1-w}{z_2}\right\}^{1/\eta} H_0(\mathcal{N}_0^\epsilon) H_0^\epsilon(dw) \\
&= \int_\epsilon^{1-\epsilon} \left\{\frac{w}{z_1} \bigwedge \frac{1-w}{z_2}\right\}^{1/\eta} H_0(dw) \\
&= (z_1 z_2)^{-1/2\eta} - \frac{1}{2}\left(\frac{\epsilon}{1-\epsilon}\right)^{1/2\eta} [z_1^{-1/\eta} + z_2^{-1/\eta}] \\
&= \nu_0([\mathbf{z}, \infty]) - B(\epsilon, \mathbf{z}),
\end{aligned}
\tag{21}
$$

where the bias term $B(\epsilon, \mathbf{z})$ can be made arbitrarily small via choice of ϵ.

Figure 1 shows $n = 2500$ simulated realizations of $\mathbf{Z}$ and $\hat{\mathbf{Z}}$ for $\epsilon = 10^{-3}, 10^{-2}$ and correlations of $\rho = 0.8, 0.5, 0.2$, as well as $\mathbf{Z}^*$ of De Haan and Zhou (2011) and $\mathbf{Y}$, the limiting first-order piece. Note that $\hat{\mathbf{Z}}$ appears to capture the tail dependence structure of $\mathbf{Z}$ better than $\mathbf{Z}^*$ of De Haan and Zhou (2011). The primary difference between $\hat{\mathbf{Z}}$ simulations with $\epsilon = 0.01$ and $\epsilon = 0.1$ is the number of points near the axes of the cone $\mathfrak{C}$. As ϵ decreases we see more large points with angular components near 0 and 1 in finite samples. This is due to the increase in scale parameter of R_0 induced by smaller ϵ; see Section 3.2.

In Table 1, we provide a comparison of $\mathbf{Z}$ and $\hat{\mathbf{Z}}$ by examining the empirical average number of points in specific sets over 250 simulations of $n = 2500$ points from both $\mathbf{Z}$ and $\hat{\mathbf{Z}}$ for $\rho = 0.5$ and $\epsilon = 10^{-3}, 10^{-2}$. Note that for $\epsilon = 0.01$, the convergence to the limiting measure is slow for regions that are near the axes of the cone $\mathfrak{C}$. However, for sets of the form $A = [\mathbf{z}, \infty]$, choosing ϵ small results in near-unbiased estimation of $\nu_0(A)$ by $\hat{\mathbf{Z}}$. For sets that are near the axes, choosing ϵ slightly larger results in faster convergence to the limiting measure (in terms of z_1, z_2), particularly when considering marginal distributions. The trade-off is that the bias is greater for sets on $\mathfrak{C}_0$; see Section 3.2. For comparison, we also show results from $\mathbf{Z}^*$ of De Haan and Zhou (2011) and $\mathbf{Y}$, the first-order piece of $\hat{\mathbf{Z}}$. We note that $\mathbf{Z}^*$ results in more points than expected in $\mathbf{Z}$ in most regions of the cone $\mathfrak{C}$. This is likely due to the slowly varying function (17), which is not accounted for by De Haan and Zhou. Of course, the first-order approximation $\mathbf{Y}$ fails to capture any of the distribution of $\mathbf{Z}$ on $\mathfrak{C}_0$.
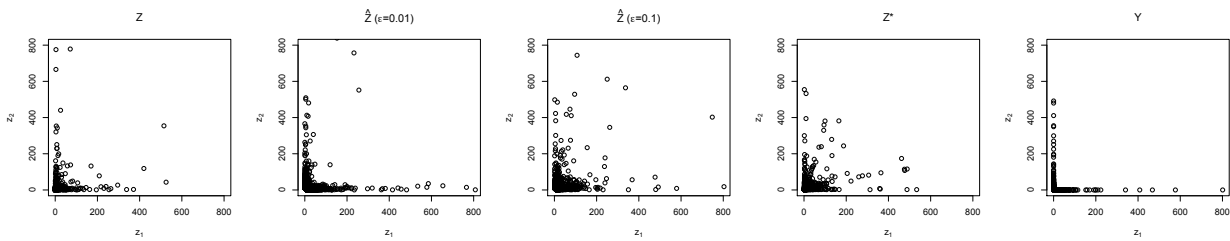

Figure 1: Simulation of $n = 2500$ points from (left to right) $\mathbf{Z}$ with Gaussian dependence structure ($\rho = 0.5$), $\hat{\mathbf{Z}} = \mathbf{Y} + \hat{\mathbf{E}}$ for $\epsilon = 0.01, 0.1$, $\mathbf{Z}^*$ of De Haan and Zhou (2011), and $\mathbf{Y}$, the first-order limiting measure.

3.2 Choice of ϵ

As Figure 1 and Table 1 indicate, one drawback of our construction $\hat{\mathbf{Z}}$ is that it results in significantly more points near the axes than we see in $\mathbf{Z}$. This is not surprising when one considers the limiting marginal

Table 1: Summary statistics from 250 simulations of $n = 2500$ points from $\mathbf{Z}$, a bivariate random vector with Fréchet marginals and Gaussian dependence with $\rho = 0.5$ and $\hat{\mathbf{Z}}$ as constructed in Section 3.1 with $\epsilon = 10^{-3}, 10^{-2}$. For comparison, we also show a summary of $\mathbf{Z}^*$ of De Haan and Zhou (2011) and of $\mathbf{Y}$, the first-order approximation to $\mathbf{Z}$. Numbers reported are empirical means and simulation-based 95% intervals.

	Number of points in the set $[\mathbf{z}, \infty]$ with $z_1 = z_2 = z$			Number of points with $Z_2 > z$		
z	100	250	500	500	1000	2000
$\mathbf{Z}$	3.20 $(1, 6)$	0.90 $(0, 3)$	0.36 $(0, 2)$	4.80 $(2, 8)$	2.39 $(0, 6)$	1.29 $(0, 3)$
$\hat{\mathbf{Z}}$ ($\epsilon = 0.01$)	3.97 $(1, 7)$	1.08 $(0, 3)$	0.36 $(0, 2)$	9.39 $(5, 14)$	4.19 $(2, 8)$	1.89 $(0, 4)$
$\hat{\mathbf{Z}}$ ($\epsilon = 0.1$)	2.89 $(1, 6)$	0.84 $(0, 3)$	0.3 $(0, 2)$	6.03 $(2, 10)$	2.86 $(0, 6)$	1.42 $(0, 4)$
$\mathbf{Z}^*$	11.29 $(7, 16)$	3.21 $(1, 6)$	1.14 $(0, 3)$	8.48 $(5, 14)$	3.97 $(1, 7)$	1.93 $(0, 4)$
$\mathbf{Y}$	0	0	0	5.00 $(2, 9)$	2.37 $(0, 5)$	1.16 $(0, 3)$

measure of $\hat{\mathbf{E}}$:

$$
\begin{aligned}
n\mathbb{P}\left[\frac{\hat{E}_1}{b^*(n)} > z_1\right] &= \int_0^1 \left(\frac{w}{z_1}\right)^{1/\eta} H_0(\mathcal{N}_0^\epsilon) H_0^\epsilon(dw) \\
&= \int_\epsilon^{1-\epsilon} \left(\frac{w}{z_1}\right)^{1/\eta} H_0(dw) \\
&= z_1^{-1/\eta} \int_\epsilon^{1-\epsilon} w^{1/\eta} H_0(dw) \\
&= \frac{1}{2} z_1^{-1/\eta} \left\{ \left(\frac{\epsilon}{1-\epsilon}\right)^{-1/2\eta} - \left(\frac{\epsilon}{1-\epsilon}\right)^{1/2\eta} \right\}.
\end{aligned}
\tag{22}
$$

For very large z_1, this is negligible compared to the heavier-tailed Y_1 piece of $\hat{Z}_1$, which has limit measure z_1^{-1}. However, for small ϵ the scaling factor in (22) is quite large, and plays a significant role in finite samples. This difficulty can be alleviated by choosing a slightly larger ϵ, which will reduce the magnitude of the scaling factor in (22).

The drawback of choosing a larger value for ϵ is that is increases the bias term in (21). That is, for sets in $\mathfrak{C}_0$ for which smaller ϵ results in greater coverage by $\mathfrak{C}_0^\epsilon$, a larger ϵ increases the rate of convergence to the limiting measure, but also decreases the accuracy of the approximation to the limiting measure of such a set. Thus the choice of ϵ involves a trade-off between the marginal behavior of $\hat{\mathbf{Z}}$ and the size of the restricted subcone $\mathfrak{C}_0^\epsilon$.

While the infinite hidden angular measure of a Fréchet-marginal random vector with Gaussian dependence poses difficulty in simulation, our sum representation of $\mathbf{Z}$ in terms of independent $\mathbf{Y}$ and $\mathbf{E}$ provides several advantages over previous approaches. We are able to capture not only the first-order limit on the whole cone $\mathfrak{C}$, but also the hidden regular varying piece on the subcone $\mathfrak{C}_0^\epsilon$. We can choose ϵ such that the restricted subcone $\mathfrak{C}_0^\epsilon$ becomes arbitrarily close to $\mathfrak{C}_0$. Choosing ϵ involves a trade off between bias in the limiting measure of sets not fully contained in $\mathfrak{C}_0^\epsilon$, and the level at which the limiting measure is a good approximation for finite samples.

We also point out here that our representation $\mathbf{Y} + \mathbf{E}$ offers an advantage in finite samples over a mixture of $\mathbf{Y}$ and $\mathbf{E}$, as proposed by Maulik and Resnick (2004). In applications in which marginal distributions are transformed to be unit Fréchet, no observation has any component exactly equal to zero. This is also the case for our sum representation, but is not a feature of the mixture characterization. Finally, it is quite natural to think of tail observations from a random vector $\mathbf{Z}$ as a sum of first- and second-order pieces. Viewing tail observations as arising from a mixture distribution is not as intuitive.

4 Summary and Discussion

This work presents a new representation of a multivariate regular varying random vector with hidden regular variation, in terms of a sum of independent regular varying pieces. We have shown our representation to be asymptotically justified via the concept of multivariate tail equivalence. An illustration of simulation from our model was provided using the bivariate Gaussian as an example. The infinite hidden measure of

this example introduced difficulty in simulation; however, we can still simulate the lighter-tailed piece on a restricted subcone. Our sum representation shares features with real data in applications and provides an intuitive model for the joint tail of a random vector.

The sum representation provides a framework for maximum-likelihood estimation and likelihood-based model selection procedures. Finite samples of random vectors from random vectors exhibiting hidden regular variation can be viewed as arising from the sum of components $\mathbf{Y}$ and $\mathbf{E}$, whereas finite samples cannot be reconciled with the mixture representation of Maulik and Resnick (2004). Our current work is to develop appropriate statistical procedures.

Specifically, as only realizations of $\mathbf{Z}$ are observed, we treat $\mathbf{Y}$ and $\mathbf{E}$ as unobserved latent variables and employ the EM algorithm (Dempster et al., 1977). One can write down a complete log-likelihood for a parameter vector θ governing the tails of $\mathbf{Y}$ and $\mathbf{E}$:

$$\ell(\boldsymbol{\theta}|\mathbf{z}, \mathbf{y}, \mathbf{e}) = \ell_{\mathbf{Y}}(\theta|\mathbf{y}) + \ell_{\mathbf{E}}(\theta|\mathbf{e}) \tag{23}$$

based on limiting point process results for $\mathbf{Y}$ and $\mathbf{E}$ (Resnick, 2007). Conditional on $\mathbf{z}$ and a fixed value of the parameter vector $\boldsymbol{\theta}^{(k)}$, one can view the conditional density of unobserved $\mathbf{Y}$ and $\mathbf{E}$ as

$$p(\mathbf{y}, \mathbf{e}|\mathbf{z}, \boldsymbol{\theta}^{(k)}) = \frac{p_{\mathbf{Y}}(\mathbf{y}|\boldsymbol{\theta}^{(k)})p_{\mathbf{E}}(\mathbf{z} - \mathbf{y}|\boldsymbol{\theta}^{(k)})}{p_{\mathbf{Z}}(\mathbf{z}|\boldsymbol{\theta}^{(k)})}$$

$$\propto p_{\mathbf{Y}}(\mathbf{y}|\boldsymbol{\theta}^{(k)})p_{\mathbf{E}}(\mathbf{z} - \mathbf{y}|\boldsymbol{\theta}^{(k)}). \tag{24}$$

One can speculate that in many cases, one can simulate realizations of $\mathbf{Y}$ and $\mathbf{E}$ from (24) without much difficulty. A Monte Carlo Expectation-Maximization algorithm could then be used to iteratively compute and maximize

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) = \int \ell(\boldsymbol{\theta}|\mathbf{z}, \mathbf{y}, \mathbf{e})p(\mathbf{y}, \mathbf{e}|\mathbf{z}, \boldsymbol{\theta}^{(k)})d\mathbf{y}d\mathbf{e}. \tag{25}$$

Estimation of this model via the EM algorithm is a direction for future research.

References

Ballani, F. and Schlather, M. (2011). A construction principle for multivariate extreme value distributions. *Biometrika*, 98(3):633–645.

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., Waal, D. D., and Ferro, C. (2004). *Statistics of Extremes: Theory and Applications.* Wiley, New York.

Bingham, N., Goldie, C., and Teugels, J. (1989). *Regular variation*, volume 27. Cambridge Univ Pr.

Coles, S., Heffernan, J., and Tawn, J. (1999). Dependence measures for extreme value analysis. *Extremes*, 2:339–365.

Coles, S. and Tawn, J. (1991). Modeling multivariate extreme events. *Journal of the Royal Statistical Society, Series B*, 53:377–92.

Cooley, D., Davis, R., and Naveau, P. (2010). The pairwise beta distribution: A flexible parametric multivariate model for extremes. *Journal of Multivariate Analysis*, 101(9):2103–2117.

de Haan, L. (1970). *On regular variation and its application to the weak convergence of sample extremes.* Mathematisch Centrum.

De Haan, L. and Zhou, C. (2011). Extreme residual dependence for random vectors and processes. *Advances in Applied Probability*, 43(1):217–242.

Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–38.

Draisma, G., Drees, H., Ferreira, A., and De Haan, L. (2004). Bivariate tail estimation: dependence in asymptotic independence. *Bernoulli*, 10(2):251–280.

Heffernan, J. and Resnick, S. (2007). Limit laws for random vectors with an extreme component. *The Annals of Applied Probability*, 17(2):537–571.

Heffernan, J. E. and Tawn, J. A. (2004). A conditional approach for multivariate extreme values. *Journal of the Royal Statistical Society, Series B*, 66:497–546.

Jessen, A. and Mikosch, T. (2006). *Regularly varying functions.* University of Copenhagen, laboratory of Actuarial Mathematics.

Ledford, A. and Tawn, J. (1997). Modelling depedence within joint tail regions. *Journal of the Royal Statistical Society, Series B*, B:475–499.

Ledford, A. W. and Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika*, 83:169–187.

Maulik, K. and Resnick, S. (2004). Characterizations and examples of hidden regular variation. *Extremes*, 7(1):31–67.

Mitra, A. and Resnick, S. (2010). Hidden regular variation: Detection and estimation. *Arxiv preprint arXiv:1001.5058.*

Peng, L. (1999). Estimation of the coefficient of tail dependence in bivariate extremes. *Statistics and Probability Letters*, 43:399–409.

Ramos, A. and Ledford, A. (2009). A new class of models for bivariate joint tails. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(1):219–241.

Resnick, S. (2002). Hidden regular variation, second order regular variation and asymptotic independence. *Extremes*, 5(4):303–336.

Resnick, S. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling.* Springer Series in Operations Research and Financial Engineering. Springer, New York.

Sibuya, M. (1960). Bivariate extremal distribution. *Ann. Inst. Statist. Math.*, 11:195–210.