MODELING MULTIPLE RISKS: HIDDEN DOMAIN OF ATTRACTION

ABHIMANYU MITRA AND SIDNEY I. RESNICK

Abstract. A sub-model of multivariate regular variation called hidden regular variation facilitates more accurate estimation of joint tail probabilities in the presence of asymptotic independence. A related concept called *hidden domain of attraction* can sometimes offer similar estimation assistance in circumstances where hidden regular variation is absent. Examples and discussion illustrate strengths and limitations of this concept. We outline estimation techniques where applicable.

1. Introduction

The tail probability of simultaneous exceedence of large thresholds by components of a risk vector serves as a risk measure in applications such as finance [23], environmental protection [29] and hydrology [5, 11]. Multivariate extreme value theory (MEVT) helps approximate such tail probabilities but in common circumstances where asymptotic independence is present the asymptotic technique gives an incorrect tail probability approximation of 0. This paper points out that even when hidden regular variation (HRV) [13, 17, 18, 20, 22, 25, 26] is not applicable, a related concept called *hidden domain of attraction* (HDA) may yield an approximation.

The joint distribution $H(\cdot)$ of a bivariate random vector $\boldsymbol{X} = (X^1, X^2)$ belongs to the maximal domain of attraction of a bivariate extreme value (EV) distribution $G(\cdot)$ if there exist scaling and centering constants $a_n^p > 0$ and $b_n^p \in \mathbb{R}$, $p = 1, 2$, such that for all continuity points $\boldsymbol{x} = (x^1, x^2)$ of G,

$$\lim_{n\to\infty} \left[H(a_n^1 x^1 + b_n^1, a_n^2 x^2 + b_n^2)\right]^n = G(x^1, x^2) \tag{1.1}$$

and both the marginal distributions of $G(\cdot)$, $G^1(\cdot)$ and $G^2(\cdot)$, are non-degenerate extreme value distributions [12, page 208]. Let $M_+(\mathbb{E})$ be the set of Radon measures on $\mathbb{E}$ and denote vague convergence by $\xrightarrow{v}$. The convergence (1.1) is equivalent to vague convergence in $M_+(\mathbb{E})$,

$$nP\Big[\Big((X^1 - b_n^1)/a_n^1, (X^2 - b_n^2)/a_n^2\Big) \in \cdot\Big] \xrightarrow{v} \nu(\cdot) \quad (n \to \infty), \tag{1.2}$$

where depending on the case, $\mathbb{E}$ is one of the domains: $[-\infty, \infty]^2 \setminus \{(-\infty, -\infty)\}$ or $[0, \infty]^2 \setminus \{(0, 0)\}$ or $[-\infty, \infty] \times [0, \infty] \setminus \{(-\infty, 0)\}$ or $[0, \infty] \times [-\infty, \infty] \setminus \{(0, -\infty)\}$. The different domains correspond to the marginal limits in (1.1) being different extreme value distributions. The limit measure $\nu(\cdot)$ in (1.2) is related to the limit distribution $G(\cdot)$ in (1.1) by

$$\nu(\{(z^1, z^2) \in \mathbb{E} : z^1 \le x^1, z^2 \le x^2\}^c) = -\log(G(x^1, x^2)), \quad (\boldsymbol{x} = (x^1, x^2) \in \mathbb{E}), \tag{1.3}$$

which is why $\nu(\cdot)$ is sometimes called the exponent measure. Assuming $[X^1 > u, X^2 > v]$ is a rare event, that is, that u and v are sufficiently large, we use MEVT to approximate the joint tail probability $P(X^1 > u, X^2 > v)$ as

$$P(X^1 > u, X^2 > v) \approx n^{-1}\nu\Big(\big((u - b_n^1)/a_n^1, \infty\big] \times \big((v - b_n^2)/a_n^2, \infty\big]\Big). \tag{1.4}$$

Key words and phrases. Regular variation, maximal domain of attraction, spectral measure, risk sets.

S. Resnick and A. Mitra were partially supported by ARO Contract W911NF-10-1-0289 at Cornell University. S. Resnick also received support from NSA Grant H98230-11-1-0193.

However, in the presence of asymptotic independence [12, page 226], (1.4) approximates this probability as zero. This approximation is often inaccurate and a better approximation is required.

If (1.2) holds with $\mathbb{E} = [0,\infty]^2 \setminus \{(0,0)\}$, $b_n^1 = b_n^2 = 0$ and some $a_n^1, a_n^2 \uparrow \infty$, then H possesses multivariate regular variation (MRV). If X^1 and X^2 are also asymptotically independent, we may improve the approximation of joint tail probabilities if hidden regular variation (HRV) is present; see [21, 22, 25, 26] and the seminal [17, 19]. However, in the standard case HRV requires the distribution of $X^1 \wedge X^2$ to have a regularly varying tail and if this is not the case, HRV is not applicable but perhaps $X^1 \wedge X^2$ has a distribution in some maximal domain of attraction other than the heavy tailed domain. In this case, *hidden domain of attraction* (HDA) may help overcome asymptotic independence.

If the distribution of $\boldsymbol{X}$ does not have MRV but (1.2) still holds, we may standardize and retrieve the regular variation setup by transforming the components of $\boldsymbol{X}$ to $\tilde{\boldsymbol{X}} := \big(U^1(X^1), U^2(X^2)\big)$, where $U^i(\cdot) = 1/(1 - H^i(\cdot))$ and $H^i(\cdot)$ is the distribution of X^i, $i = 1, 2$ ([27, page 265], [12, page 208]). If X^1 and X^2 are asymptotically independent, so are $(\tilde{X}^1, \tilde{X}^2)$. If $\tilde{X}^1 \wedge \tilde{X}^2$ has a regularly varying tail, we seek HRV for $\tilde{\boldsymbol{X}}$; otherwise we check if $\tilde{X}^1 \wedge \tilde{X}^2$ is in a maximal domain of attraction of an extreme value distribution other than the Fréchet and if so, we seek HDA. Formal definitions of HDA are given in Sections 2, 3.1 and a variant is considered in Section 3.2.

It is possible after standardizing that the right endpoints of $\tilde{X}^i$, $i = 1, 2$ are infinite but the right endpoint $\tilde{x}_0$ of $\tilde{X}^1 \wedge \tilde{X}^2$ is finite and when this is true, it appears the HDA technique is not helpful for inference about remote risk regions, since the finite endpoint $\tilde{x}_0$ makes hidden domain of attraction estimation want to localize around $(\tilde{x}_0, \tilde{x}_0)$. For instance, in Example 2.4, X^1, X^2 are each Pareto (so no standardization is needed), asymptotic independence holds, $X^1 \wedge X^2$ is in a domain of attraction with negative extreme value index and $X^1 \wedge X^2 \leq 2$. The HDA procedure estimates

$$P[X^1 > z^1, X^2 > z^2] \approx 1 - (z^1 + z^2)/4,$$

only when $z^1 + z^2 \leq 4$. So while HDA can be defined in such a circumstance, it provides an asymptotic regime that is not demonstrably useful for estimating probabilities of remote risk regions. For estimation, it appears a useful HDA theory requires the right endpoint $\tilde{x}_0 = \infty$ to match the right endpoints of the standardized variables.

Our goal in this paper is to discuss setup, definitions and some estimation procedures. Statistically, of course, we do not know $U^i(\cdot)$, $i = 1, 2$ and this and similar problems must be engaged. We make statistical suggestions in Sections 2.3, 3.1 that are fairly traditional and in Section 3.2 we offer some ideas in a new direction that may stimulate thought. Our methodology is not yet well tested but hopefully is a contribution to discussion of improved risk estimation.

1.1. **Outline.** Section 1.2 reviews frequently used notation. In Section 2, we define hidden domain of attraction for the standard case, when both the components of the risk vector have the same distribution. This assumption, made for didactic purposes, is often unrealistic but could apply either when assessing serial tail dependence by sampling selectively from a stationary time series or if one chooses to standardize so that each marginal distribution is, say, Pareto as in Section 3.1. Section 3 deals with HDA in the non-standard case, where we drop the identical distribution assumption for X^1, X^2. We outline estimation approachs for the circumstance that component random variables and their minima have the same upper bounds. In both Sections 2 and 3, we exhibit examples and, where possible, discuss estimation procedures of limit measures that appear in the asymptotic model. We conclude with remarks, prospects and a summary in Section 4.

1.2. **Notation.** For simplicity, this paper is restricted to two dimensions. For denoting a vector and its components, we use $\boldsymbol{x} = (x^1, x^2)$, $x^p = p$-th component of $\boldsymbol{x}$, $p = 1, 2$. Multivariate intervals or rectangles are denoted $(\boldsymbol{x}, \boldsymbol{y}], [\boldsymbol{x}, \boldsymbol{y}]$, etc where, for instance, $(\boldsymbol{x}, \boldsymbol{y}] = (x^1, y^1] \times (x^2, y^2]$. The vectors of all zeros, all ones and all infinities are denoted by $\mathbf{0} = (0, 0)$, $\mathbf{1} = (1, 1)$ and $\boldsymbol{\infty} = (\infty, \infty)$ respectively. We write $x^{(1)} = x^1 \vee x^2, \qquad x^{(2)} = x^1 \wedge x^2$. So, the superscripts denote components of a vector and the ordered component is denoted by a parenthesis in the superscript.

We express vague convergence [26, page 173] of Radon measures as $\xrightarrow{v}$ and weak convergence of probability measures [4, page 14] as $\Rightarrow$. Denote a point measure with points $\{x_i\}$ in a nice space $\mathbb{F}$ by $\sum_i \epsilon_{x_i}$ where $\epsilon_x(\cdot)$ is the measure with all mass at x:

$$\epsilon_x(B) = \begin{cases} 1, & \text{if } x \in B, \\ 0, & \text{if } x \in B^c, \end{cases} \qquad (x \in \mathbb{F},\, B \subset \mathbb{F}).$$

The set of non-negative Radon measures on a space $\mathbb{F}$ topologized by the vague topology is $M_+(\mathbb{F})$.

For a univariate distribution $F(x)$, set $\bar{F} := 1 - F$; the right endpoint of F is $x_F = \sup\{x : F(x) < 1\}$. The inverse of a non-decreasing function $\psi(x)$ is $\psi^{\leftarrow}(x) := \inf\{y : \psi(y) \geq x\}$. The space of left continuous functions on an interval I with finite right limits is $D_{\text{left}}(I)$ and the right continuous functions with finite left limits is $D(I)$. If the interval I is open on the right, $I^{\dashv}$ is the right closure.

For a multivariate distribution function $H(x^1, x^2)$, $H^p(x)$ is the pth marginal $(p = 1, 2)$.

2. Standard case hidden domain of attraction

Suppose that a distribution $H(\boldsymbol{x})$ of a bivariate random vector $\boldsymbol{X} = (X^1, X^2)$ belongs to the maximal domain of attraction of an extreme value distribution G with G^p non-degenerate, $p = 1, 2$ ([27, page 265], [12, Chapter 6]), $X^1 \stackrel{d}{=} X^2$ and X^1 and X^2 are asymptotically independent so that (1.1) is satisfied with $b_n^1 = b_n^2 = b_n$, $a_n^1 = a_n^2 = a_n > 0$ $(n \in \mathbb{N})$ and $G(x, y) = G^1(x)G^2(y)$ for $x, y \in \mathbb{R}$. Thus, (1.2) becomes

$$nP\left[(\boldsymbol{X} - b_n\mathbf{1})/a_n \in \cdot\right] \xrightarrow{v} \nu(\cdot) \quad \text{in } M_+(\mathbb{E}), \tag{2.1}$$

where $\mathbb{E} = [-\boldsymbol{\infty}, \boldsymbol{\infty}]\backslash\{-\boldsymbol{\infty}\}$ or $\mathbb{E} = [\mathbf{0}, \boldsymbol{\infty}]\backslash\{\mathbf{0}\}$. Since $G(x, y) = G^1(x)G^2(y)$ for $x, y \in \mathbb{R}$, the relation of ν and G given in (1.3) gives for $\boldsymbol{x} \in \mathbb{E}$,

$$\begin{aligned} \nu(\{\mathbf{z} \in \mathbb{E} : z^1 \leq x^1, z^2 \leq x^2\}^c) &= -\log G^1(x^1) + -\log G^2(x^2) \\ &= \nu(\{\mathbf{z} \in \mathbb{E} : z^1 \leq x^1\}^c) + \nu(\{\mathbf{z} \in \mathbb{E} : z^2 \leq x^2\}^c). \end{aligned} \tag{2.2}$$

The standard case contains the additional assumption that $X^1 \stackrel{d}{=} X^2$, which reduces (1.2) to (2.1) and reduces possible choices for $\mathbb{E}$. It would be enough to assume the distributions of X^1 and X^2 are tail equivalent. The cone $\mathbb{E} = [\mathbf{0}, \boldsymbol{\infty}]\backslash\{\mathbf{0}\}$ is chosen only when H has MRV.

From (2.1), the maximal component of $\boldsymbol{X}$ satisfies as $n \to \infty$,

$$nP(X^{(1)} > a_n y + b_n) \to \nu(\{\mathbf{z} \in \mathbb{E} : z^{(1)} > y\}), \qquad (y, y) \in \mathbb{E}, \tag{2.3}$$

so the distribution of $X^{(1)}$ is in a maximal domain of attraction of an extreme value distribution and the distribution of $X^{(1)}$ characterizes a_n and b_n given in (2.1). Using one-dimensional extreme value theory, we can and do choose a_n and b_n in such a way that

$$\psi_\gamma(y) := \left(\nu(\{\mathbf{z} \in \mathbb{E} : z^{(1)} > y\})\right)^{-1} \tag{2.4}$$

takes one of the following forms depending on the extreme value index $\gamma \in \mathbb{R}$:

$$\psi_\gamma(y) = \begin{cases} y^{1/\gamma}, & \text{if } \gamma > 0,\ y > 0, \\ e^y, & \text{if } \gamma = 0,\ y \in \mathbb{R}, \\ (|y|)^{-1/|\gamma|}, & \text{if } \gamma < 0,\ y < 0. \end{cases} \tag{2.5}$$

See [27, page 9] or [12, page 6]. If $\gamma > 0$, set $\mathbb{R}_\gamma = (0, \infty)$; if $\gamma = 0$ set $\mathbb{R}_0 = \mathbb{R}$; and if $\gamma < 0$, $\mathbb{R}_\gamma = (-\infty, 0)$. Let $\mathbb{R}_\gamma^\dashv$ be the right closure of $\mathbb{R}_\gamma$; for instance $\mathbb{R}_0^\dashv = (-\infty, \infty]$. Then $\psi_\gamma : \mathbb{R}_\gamma \mapsto (0, \infty)$ and $\psi_\gamma^\leftarrow : (0, \infty) \mapsto \mathbb{R}_\gamma$ where for $y > 0$,

$$\psi_\gamma^\leftarrow(y) = \begin{cases} y^\gamma, & \text{if } \gamma > 0, \\ \log y, & \text{if } \gamma = 0, \\ -y^{-|\gamma|}, & \text{if } \gamma < 0. \end{cases}$$

At the right endpoint of $\mathbb{R}_\gamma$ we may extend ψ_γ to equal ∞ and then $\psi_\gamma : \mathbb{R}_\gamma^\dashv \to (0, \infty]$ and $\psi_\gamma^\leftarrow : (0, \infty] \mapsto \mathbb{R}_\gamma^\dashv$. We have used the traditional parameterization rather than the Von Mises parameterization used in [12] because it meshes somewhat better with regular variation.

We define a sub-model of MEVT called (standard case) hidden domain of attraction (HDA). We assume $\mathbb{E}$ is either $[-\boldsymbol{\infty}, \boldsymbol{\infty}]\backslash\{-\boldsymbol{\infty}\}$ or $[\mathbf{0}, \boldsymbol{\infty}]\backslash\{\mathbf{0}\}$, and define $\mathbb{E}^0$ as either $(-\boldsymbol{\infty}, \boldsymbol{\infty}]$ or $(\mathbf{0}, \boldsymbol{\infty}]$. It is not always true that $\mathbb{E}^0 \subset \mathbb{E}$ (see Remark 2.2(2)) and we may have $\mathbb{E}^0 = (-\boldsymbol{\infty}, \boldsymbol{\infty}]$ and $\mathbb{E} = [\mathbf{0}, \boldsymbol{\infty}]\backslash\{\mathbf{0}\}$ as in Example 2.4. The choice of $\mathbb{E}^0$ depends on the domain of attraction of the distribution of $X^1 \wedge X^2$ and the choice of $\mathbb{E}$ depends on the domain of $X^1 \vee X^2$.

Definition 2.1. The distribution of $\boldsymbol{X} = (X^1, X^2)$ has standard case hidden domain of attraction on the cone $\mathbb{E}^0$ if (i) $X^1 \stackrel{d}{=} X^2$; (ii) (2.1) and (2.2) hold; (iii) there exist positive scaling and real centering constants $\{c_n\}$ and $\{d_n\}$ and a non-zero measure $\nu^0 \in M_+(\mathbb{E}^0)$ such that in $M_+(\mathbb{E}^0)$,

$$nP\left[(\boldsymbol{X} - d_n\mathbf{1})/c_n \in \cdot\,\right] \stackrel{v}{\to} \nu^0(\cdot) \qquad (n \to \infty). \tag{2.6}$$

This definition requires $X^1 \stackrel{d}{=} X^2$ and that the distribution of $\boldsymbol{X}$ belongs to the maximal domain of attraction of an extreme value product measure G with exponent measure ν. Some other remarks:

Remark 2.2. (1) HRV assumes (2.1) holds on $\mathbb{E} = [\mathbf{0}, \boldsymbol{\infty}]\backslash\{\mathbf{0}\}$ with $b_n = 0$, $a_n \uparrow \infty$ and (2.6) is satisfied on $\mathbb{E}^0 = (\mathbf{0}, \boldsymbol{\infty}]$ with $d_n = 0$, $c_n \uparrow \infty$ with $a_n/c_n \to \infty$ as $n \to \infty$.

(2) For $y\mathbf{1} \in \mathbb{E}^0$, $(y\mathbf{1}, \boldsymbol{\infty}]$ is relatively compact and hence from (2.6) the minimum component of $\boldsymbol{X}$ satisfies,

$$n\bar{H}^{(2)}(c_n y + d_n) := nP[X^{(2)} > c_n y + d_n] \to \nu^0((y, \infty] \times (y, \infty]) = 1/\psi_{\gamma^0}(y) \qquad ((y, y) \in \mathbb{E}^0), \tag{2.7}$$

where for some $\gamma^0 \le \gamma$, $\psi_{\gamma^0}(y)$ is defined in (2.5). The distribution of $X^{(2)}$ belongs to the maximal domain of attraction of an extreme value distribution [12, page 4] with extreme value index $\gamma^0 \in \mathbb{R}$. When HRV exists, $\gamma^0 > 0$, the distribution of $X^{(2)}$ has a regularly varying tail and is hence in the domain of attraction of the Fréchet distribution. HDA allows the additional cases where the distribution of $X^{(2)}$ belongs to the domain of attraction of the Gumbel ($\gamma^0 = 0$) or the Weibull distribution ($\gamma^0 < 0$).

The distribution of $X^{(2)}$ determines the scaling and centering constants $\{c_n\}$ and $\{d_n\}$ and $\mathbb{E}^0$. As illustrated by Example 2.4, even if $\mathbb{E} = [\mathbf{0}, \boldsymbol{\infty}] \setminus \{\mathbf{0}\}$, $\mathbb{E}^0$ could be $(-\boldsymbol{\infty}, \boldsymbol{\infty}]$ and $\mathbb{E}^0$ is not necessarily a subset of $\mathbb{E}$, as was the case for HRV [25].

(3) From (2.3), $X^{(1)}$ belongs to the maximal domain of attraction of an extreme value distribution. Since $X^{(1)} \ge X^{(2)}$, $\gamma \ge \gamma^0$ and the convergence (2.1) puts a restriction on what

convergences are possible in (2.6). For example, if (2.1) is satisfied with $X^{(1)}$ being in the Gumbel domain of attraction, then HRV can never hold on $\mathbb{E}^0$ since the tail of $X^{(2)}$ cannot be heavier than the tail of $X^{(1)}$.

2.1. **Two examples.** Here are two examples of distributions that possess multivariate regular variation, asymptotic independence, hidden domain of attraction but *not* hidden regular variation.

Example 2.3. Suppose, $W_1, W_2 \overset{iid}{\sim} F(\cdot)$ and $Z_1, Z_2 \overset{iid}{\sim} D(\cdot)$, where F and D belong to the univariate maximal domains of attraction of the Fréchet with index $\alpha = 1$ and Gumbel distributions respectively. Let B be a Bernoulli random variable such that $P[B = 1] = 1/2 = 1 - P[B = 0]$. Assume W_1, W_2, Z_1, Z_2, B are independent and define the bivariate random vector $\boldsymbol{X}$ by

$$\boldsymbol{X} = (X^1, X^2) = B(W_1, Z_1) + (1 - B)(Z_2, W_2).$$

We show that the distribution of $\boldsymbol{X}$ has MRV. It suffices [10] to verify that $t_1X^1 \vee t_2X^2$ has a regularly varying tail for any $t_1, t_2 > 0$. This follows easily since F has a regulary varying tail and

$$P[t_1W_1 \vee t_2Z_1 > x] \sim P[t_1W_1 > x], \qquad (x \to \infty).$$

For appropriate $a_n \uparrow \infty$, $nP\left[\boldsymbol{X}/a_n \in \cdot\right] \overset{v}{\to} \nu(\cdot)$, on $\mathbb{E} = [0,\infty]^2 \setminus \{\mathbf{0}\}$ where $\nu\left(\left([0, x^1] \times [0, x^2]\right)^c\right) = \left((x^1)^{-1} + (x^2)^{-1}\right)/2$, and the distribution of $\boldsymbol{X}$ has MRV with asymptotic independence on $\mathbb{E}$.

Furthermore, $X^{(2)}$ belongs to the maximal domain of attraction of a Gumbel distribution and therefore, HRV does not exist. To see this, without loss of generality [1, 27], assume that D is a von-Mises function [27, page 40] in the domain of the Gumbel and for specificity assume the right endpoint of D is infinite. For some constant $c > 0$, the form of the tail is

$$\overline{D}(x) = ce^{-\int_1^x [f_D(y)]^{-1} dy} \quad \text{and} \quad f_D'(x) \overset{x\to\infty}{\to} 0,$$

where $f_D = \bar{D}/D'$ is called the auxiliary function. Likewise, assume without loss of generality [27, page 58] that F satisfies $xF'(x)/\bar{F}(x) \to 1$, where 1 is the index of regular variation of $\bar{F}$. Then $\overline{H^{(2)}}(x) := \overline{F}(x)\overline{D}(x)$ is the tail of a von-Mises function with auxiliary function $f_{H^{(2)}}(t) \sim f_D(t)$, $t \to \infty$. This follows from

$$\frac{1}{f_{H^{(2)}}(t)} := \frac{H^{(2)\prime}(t)}{\overline{H^{(2)}}(t)} = \frac{F'(t)}{\overline{F}(t)} + \frac{D'(t)}{\overline{D}(t)} = \frac{(1 + o(1))}{t} + \frac{1}{f_D(t)}$$

and the fact that $f_D(t)/t \to 0$ as $t \to \infty$. From [27, Corollary 1.7, page 46], $H^{(2)}$ belongs to the maximal domain of attraction of the Gumbel distribution.

Next we make the conventional choices [27, page 40] of scaling and centering constants $\{c_n\}$ and $\{d_n\}$ in (2.7) so that in (2.5) $\psi_{\gamma^0}(\cdot) = e^y$. These choices are $d_n = \left(1/\bar{H}^{(2)}\right)^{\leftarrow}(n)$ and $c_n = f_{H^{(2)}}(d_n)$. Then for $\boldsymbol{x} \in \mathbb{E}^0 = (-\infty, \infty]^2$, as $n \to \infty$,

$$\begin{aligned}
&nP[X^1 > d_n + c_nx^1, X^2 > d_n + c_nx^2] \\
&\quad = \frac{n}{2}\overline{F}\left(d_n + c_nx^1\right)\overline{D}\left(d_n + c_nx^2\right) + \frac{n}{2}\overline{F}\left(d_n + c_nx^2\right)\overline{D}\left(d_n + c_nx^1\right) \\
&\quad = \frac{n}{2}\left(\frac{\overline{F}\left(d_n + c_nx^1\right)}{\overline{F}\left(d_n + c_nx^2\right)}\right)\overline{H^{(2)}}\left(d_n + c_nx^2\right) + \frac{n}{2}\left(\frac{\overline{F}\left(d_n + c_nx^2\right)}{\overline{F}\left(d_n + c_nx^1\right)}\right)\overline{H^{(2)}}\left(d_n + c_nx^1\right) \\
&\quad \to (e^{-x^1} + e^{-x^2})/2.
\end{aligned}$$

The convergence follows from the facts that $\overline{F}$ is regularly varying, $c_n/d_n \to 0$ and (2.7) holds with $\psi_{\gamma^0}(y) = e^y$. Therefore, as in Definition 2.1, the distribution of $\boldsymbol{X}$ has HDA on $\mathbb{E}^0 = (-\infty, \infty]^2$

with limit measure ν^0 such that for $\boldsymbol{x} = (x^1, x^2) \in \mathbb{E}^0$,

$$\nu^0\left((x^1,\infty] \times (x^2,\infty]\right) = (e^{-x^1} + e^{-x^2})/2.$$

This measure ν^0 concentrates on $\{\infty\} \times (-\infty,\infty] \cup (-\infty,\infty] \times \{\infty\}$, the lines through $\boldsymbol{\infty}$, and puts no mass on $(-\boldsymbol{\infty}, \boldsymbol{\infty})$.

We conclude that the distribution of $\boldsymbol{X}$ is regularly varying on $\mathbb{E} = [\mathbf{0}, \boldsymbol{\infty}] \setminus \{\mathbf{0}\}$, has HDA on $\mathbb{E}^0 = (-\boldsymbol{\infty}, \boldsymbol{\infty}]$ and does not have HRV.

Example 2.4. Suppose, $U \sim \text{Uniform}([0,1])$. Define the random vector $\boldsymbol{X}$ as

$$\boldsymbol{X} = (X^1, X^2) = \left(\frac{1}{U}, \frac{1}{1-U}\right).$$

For $\boldsymbol{x} > \mathbf{0}$, $2n > (x^1)^{-1} + (x^2)^{-1}$,

$$\begin{aligned} n(1 - P[1/U \le 2nx^1, 1/(1-U) \le 2nx^2]) &= n(1 - P[U \ge (2nx^1)^{-1}, U \le 1 - (2nx^2)^{-1}]) \\ &= n(1 - (1 - (2nx^2)^{-1} - (2nx^1)^{-1})) \to ((x^1)^{-1} + (x^2)^{-1})/2, \end{aligned}$$

as $n \to \infty$. Therefore, on $\mathbb{E} = [0,\infty]^2 \setminus \{\mathbf{0}\}$, $nP\left[\boldsymbol{X}/2n \in \cdot\right] \xrightarrow{v} \nu(\cdot)$ where the limit measure ν satisfies $\nu\left(\left([0,x^1] \times [0,x^2]\right)^c\right) = ((x^1)^{-1} + (x^2)^{-1})/2$, for $\boldsymbol{x} \in \mathbb{E}$ and thus the distribution of $\boldsymbol{X}$ has MRV with asymptotic independence.

Since $X^1 \wedge X^2$ is bounded above by 2, HRV is absent. However, a little calculation shows $2 - X^1 \wedge X^2$ has a distribution which is regularly varying at 0; that is,

$$P[(2 - X^1 \wedge X^2)^{-1} > x] \sim x^{-1}/2, \quad (x \to \infty),$$

and so there is reason to believe the distribution of (X^1, X^2) has HDA. For $\{(x^1,x^2) \in (-\infty,\infty]^2 : x^1 + x^2 \le 0\}$, and large n,

$$nP[n(X^1 - 2) > x^1,\, n(X^2 - 2) > x^2] = nP\Big[U < \frac{1}{2 + x^1/n},\, U > 1 - \frac{1}{2 + x^2/n}\Big]. \tag{2.8}$$

For $x^1 + x^2 \le 0$ and n large, $1/(2 + x^1/n) \ge 1 - 1/(2 + x^2/n)$ and the above probability is

$$\begin{aligned} n\Big(\frac{1}{2 + x^1/n} - \big(1 - \frac{1}{2 + x^2/n}\big)\Big) &= \frac{n}{2}\Big(\frac{1}{1 + x^1/2n} + \frac{1}{1 + x^2/2n} - 2\Big) \\ &= \frac{n}{2}\Big(1 - \frac{x^1}{2n} + 1 - \frac{x^2}{2n} + O(\frac{1}{n^2}) - 2\Big) \to \frac{1}{4}((-x^1) + (-x^2)), \end{aligned} \tag{2.9}$$

as $n \to \infty$. Similar calculations show that for $\{(x^1,x^2) \in (-\infty,\infty]^2 : x^1 + x^2 > 0\}$, the probability in (2.8) converges to zero. Therefore, the distribution of $\boldsymbol{X}$ has HDA as in Definition 2.1 on $\mathbb{E}^0 = (-\infty,\infty]^2$ and satisfies (2.6) and (2.7). With $\gamma^0 = -1$ and $d_n = 2, c_n = 2/n$, the limit measure ν^0 is

$$\nu^0\left((x^1,\infty] \times (x^2,\infty]\right) = \begin{cases} \frac{1}{2}((-x^1) + (-x^2)), & \text{if } x^1 + x^2 \le 0 \\ 0 & \text{otherwise.} \end{cases}$$

Also, $X^{(2)}$ belongs to the domain of attraction of the reversed Weibull distribution [27, page 59] with extreme value index $\gamma^0 = -1$.

To summarize: The distribution of $\boldsymbol{X}$ has MRV with asymptotic independence, does not have HRV but does have HDA and furthermore, $\mathbb{E}^0$ is not a subset of $\mathbb{E}$.

Note the HDA asymptotics in Example 2.4 given in (2.9) localize around the point $(2,2)$. Using HDA asymptotics to approximate $P[\boldsymbol{X} > \boldsymbol{z}]$ for large z^i, $i = 1,2$ requires setting $z^i = 2 + x^i/n$, $i = 1,2$ while maintaining z^i, $i = 1,2$ large. This is incompatible with the requirement $0 \geq x^1 + x^2 = n(z^1 + z^2 - 4)$ required for a non-zero risk estimate. When $X^1 \wedge X^2$ has a smaller right endpoint than X^i, $i = 1,2$, the HDA asymptotics do not seem to help with computing probabilities of remote risk regions. In Example 2.3, this was not an issue.

2.2. **Semi-parametric structure of ν^0.** Recall the distribution of the minimum component $X^{(2)}$ is $H^{(2)}$ and the limit measure in (2.6) is ν^0 and $\nu^0((y\mathbf{1}, \boldsymbol{\infty}]) = 1/\psi_{\gamma^0}(y)$. In this section we need to assume the right endpoint of $H^{(2)}$ is the same as the right endpoints of the common distribution of X^i, $i = 1,2$, thus avoiding the context of Example 2.4. With this matching of endpoints, the limit measure ν^0 in (2.6) has semi-parametric structure which both characterizes the limits in (2.6) and aids estimation (as in [21, 22] for HRV).

Assume that c_n and d_n are chosen so that for $\gamma^0 \in \mathbb{R}$, ψ_{γ^0} has the form given in (2.5). Define

$$U^{(2)}(x) = 1/(1 - H^{(2)}(x)), \tag{2.10}$$

and for $y \in \mathbb{R}_{\gamma^0}$, (2.7) becomes

$$U^{(2)}(c_n y + d_n)/n \to \psi_{\gamma^0}(y), \tag{2.11}$$

and $c_n y + d_n \to x_{H^{(2)}}$. The next Proposition identifies the semi-parametric structure of ν^0.

Proposition 2.5. *Suppose $X^1 \stackrel{d}{=} X^2$, the right endpoint of $H^{(2)}$ is the same as the right endpoints of the common distribution of X^i, $i = 1,2$ and $U^{(2)}$ is strictly increasing in a left neighborhood of $x_{H^{(2)}}$. Then the distribution of $\boldsymbol{X}$ satisfies* (2.6) *iff*

(1) $X^{(2)}$ has a distribution $H^{(2)}$ in a univariate maximal domain of attraction so that for some $c_n > 0$, $d_n \in \mathbb{R}$, (2.7) *or* (2.11) *holds, and*

(2) regular variation on $(\mathbf{0}, \boldsymbol{\infty}]$ holds for the distribution of $(U^{(2)}(X^1), U^{(2)}(X^2))$:

$$nP\left[n^{-1}\left(U^{(2)}(X^1), U^{(2)}(X^2)\right) \in \cdot\right] \stackrel{v}{\to} \tilde{\nu}^0(\cdot) \tag{2.12}$$

where $U^{(2)}(\cdot)$ is defined in (2.10) *and $\tilde{\nu}^0(\cdot)$ is a Radon measure on $(\mathbf{0}, \boldsymbol{\infty}]$ that is related to the limit measure $\nu^0(\cdot)$ in* (2.6) *by*

$$\tilde{\nu}^0((\boldsymbol{x}, \boldsymbol{\infty}]) = \nu^0\Big(\big(\psi^{\leftarrow}_{\gamma^0}(x^1), \infty\big] \times \big(\psi^{\leftarrow}_{\gamma^0}(x^2), \infty\big]\Big), \qquad \boldsymbol{x} \in (\mathbf{0}, \boldsymbol{\infty}]. \tag{2.13}$$

The measure $\tilde{\nu}^0(\cdot)$ satisfies the scaling property on $(\mathbf{0}, \boldsymbol{\infty}]$:

$$\tilde{\nu}^0(c\cdot) = c^{-1}\tilde{\nu}^0(\cdot), \qquad c > 0. \tag{2.14}$$

Corollary 2.7 below shows that the limit measure $\tilde{\nu}^0$ is determined by a probability measure S^0 on a certain space $\delta\aleph^{(2)}$ to be explained and the family of limits in (2.12) is indexed by probability measures on $\delta\aleph^{(2)}$. The probability measure S^0 determines $\tilde{\nu}^0$ and γ^0 gives $\psi_{\gamma^0}(\cdot)$ from (2.5) and then (2.13) yields ν^0. This shows the semi-parametric structure.

Proof of Proposition 2.5. Assume (1) and (2); that is, assume (2.7) or (2.11) as well as (2.12). For $x^p \in \mathbb{R}_{\gamma^0}$, $p = 1,2$, we have $c_n x^p + d_n \uparrow x_{H^{(2)}}$ and $x_{H^{(2)}}$ is the same as the right endpoint of the distribution of X^p, $p = 1,2$ so

$$nP\Big[\frac{X^p - d_n}{c_n} > x^p; p = 1,2\Big] = nP\Big[\frac{U^{(2)}(X^p)}{n} > \frac{U^{(2)}(c_n x^p + d_n)}{n}; \, p = 1,2\Big]$$

$$\to \tilde{\nu}^0\left(\left(\psi_{\gamma^0}(x^1),\infty\right]\times\left(\psi_{\gamma^0}(x^2),\infty\right]\right) =: \nu^0((\boldsymbol{x},\boldsymbol{\infty}]),$$

where the convergence follows from (2.11) and (2.12). Hence, (2.6) holds.

Conversely, if (2.6) holds, then (2.11) is also valid. Invert (2.11) and $U^{(2)}:(-\infty,x_{H^{(2)}})\mapsto(0,\infty)$ satisfies

$$(U^{(2)\leftarrow}(yn)-d_n)/c_n\to\psi^{\leftarrow}_{\gamma^0}(y),\qquad y>0. \tag{2.15}$$

So for $\boldsymbol{x}>\mathbf{0}$,

$$\begin{aligned}
nP[U^{(2)}(X^p)\ge nx^p,\,p=1,2] &= nP[X^p\ge U^{(2)\leftarrow}(nx^p),\,p=1,2]\\
&= nP\Big[\frac{X^p-d_n}{c_n}\ge (U^{(2)\leftarrow}(nx^p)-d_n)/c_n;\,p=1,2\Big]\\
&\to \nu^0([\psi^{\leftarrow}_{\gamma^0}(x^1),\infty]\times[\psi^{\leftarrow}_{\gamma^0}(x^2),\infty]) =: \tilde{\nu}^0([\boldsymbol{x},\boldsymbol{\infty}]).
\end{aligned}$$

□

Remark 2.6. As patiently and clearly pointed out by the editors and referees, Proposition 2.5 can fail without the assumption that $x_{H^{(2)}}$ matches the right endpoint of the common distribution of X^p, $p=1,2$. Absent this match, if (2.6) holds, (2.12) can fail and conversely (1) and (2) of the Proposition can fail to imply (2.6) (counterexample graciously provided by the associate editor).

If $x_{H^{(2)}}$ is smaller than the right endpoints of X^p, it is possible that (2.6) holds as well as regular variation (2.12), provided

$$P[X^p\ge x_{H^{(2)}};\,p=1,2]=0. \tag{2.16}$$

This is the situation in Example 2.4 where $x_{H^{(2)}}=2$, and $P[X^1>2,X^2>2]=0$. Extend $U^{(2)}$ beyond $x_{H^{(2)}}$ by setting $U^{(2)}(x)=\infty$ if $x\ge x_{H^{(2)}}$. For Example 2.4,

$$U^{(2)}(x)=\begin{cases}\frac{1}{2x^{-1}-1}, & \text{if } x\le 2,\\ \infty, & \text{if } x>2,\end{cases}$$

and for $\boldsymbol{x}>\mathbf{0}$,

$$\begin{aligned}
&nP[U^{(2)}(X^p)\ge nx^p;\,p=1,2]\\
&= nP[U^{(2)}(X^p)\ge nx^p,\,p=1,2;\,U\le 1/2] + nP[U^{(2)}(X^p)\ge nx^p,\,p=1,2;\,U>1/2]. \qquad (2.17)
\end{aligned}$$

For the second term in (2.17), on $[U>1/2]$, $U^{(2)}(X^1)\ge nx^1$ means $1/2<U\le\frac{1}{2}(1+1/(nx^1))$ and $1/(1-U)=X^2\ge 2$ so $U^{(2)}(X^2)=\infty\ge nx^2$. Therefore, the second term in (2.17) becomes $nP[1/2<U\le\frac{1}{2}(1+1/(nx^1))\to 1/(2x^1)$ as $n\to\infty$. Analogous analysis of the first term of (2.17) yields the final result

$$nP[U^{(2)}(X^p)\ge nx^p;\,p=1,2]\to 1/(2x^1)+1/(2x^2),\quad (n\to\infty,\,\boldsymbol{x}>\mathbf{0}).$$

The limit measure concentrates on lines through $\boldsymbol{\infty}$, namely on $(0,\infty]\times\{\infty\}\cup\{\infty\}\times(0,\infty]$.

In fact, the limit measure *must* concentrate on the lines through ∞ if (i) $x_{H^{(2)}}$ is smaller than the right endpoints of X^p; (ii) (2.16) holds; and (iii) standardized regular variation (2.12) holds. For example, it is easy to show for $0<a<b<\infty$ and $y>0$, that

$$\lim_{n\to\infty} nP[a<U^{(2)}(X^1)/n\le b,\,U^{(2)}(X^2)/n>y] = \tilde{\nu}^0\big((a,b]\times(y,\infty]\big)$$

is independent of y. Existence of the limit is assumed in (iii). Because of (2.16), we may write the left side as

$$nP[a<U^{(2)}(X^1)/n\le b, U^{(2)}(X^2)/n>y, X^1>x_{H^{(2)}}]$$

$$+ nP[a < U^{(2)}(X^1)/n \le b,\, U^{(2)}(X^2)/n > y, X^2 > x_{H^{(2)}}] = I + II.$$

For II, $X^2 > x_{H^{(2)}}$ implies $U^{(2)}(X^2) = \infty$ which implies $U^{(2)}(X^2)/n > y$ so that

$$II = nP[a < U^{(2)}(X^1)/n \le b, X^2 > x_{H^{(2)}}],$$

independent of y. The term I is quickly seen to be 0. We conclude $\tilde{\nu}^0$ concentrates on the lines through $\boldsymbol{\infty}$.

If $x_{H^{(2)}}$ is smaller than the right endpoints of X^p but in contrast to (2.16) we have

$$P[X^p \ge x_{H^{(2)}};\, p = 1, 2] > 0, \tag{2.18}$$

then both (2.12) and (2.6) fail. For $\boldsymbol{x} > \mathbf{0}$, and $p = 1, 2$,

$$[U^{(2)}(X^p) \ge nx^p] = [U^{(2)}(X^p) \ge nx^p, X^p < x_{H^{(2)}}] \cup [X^p \ge x_{H^{(2)}}] \supset [X^p \ge x_{H^{(2)}}]$$

so that (2.18) implies

$$nP[U^{(2)}(X^p) \ge nx^p;\, p = 1, 2] \ge nP[X^p \ge x_{H^{(2)}};\, p = 1, 2] \to \infty,$$

and regular variation on $(\mathbf{0}, \boldsymbol{\infty}]$ fails. Furthermore, (2.18) implies $0 < P[X^{(2)} \ge x_{H^{(2)}}] = P[X^{(2)} = x_{H^{(2)}}]$ which rules out (2.7) and hence (1) of Proposition 2.5 and (2.6). □

The scaling property (2.14) allows us to express (2.12) in an alternate coordinate system that transforms the limit measure into a product. From (2.12), $(U^{(2)}(X^1), U^{(2)}(X^2))$ has regular variation on $(\mathbf{0}, \boldsymbol{\infty}]$ and using (2.5) and (2.13), we get

$$\tilde{\nu}^0\left((\mathbf{1}, \boldsymbol{\infty}]\right) = \nu^0\left(\left(\left(\psi_{\gamma^0}\right)^{\leftarrow}(1), \infty\right]^2\right) = \left[\psi_{\gamma^0}\left(\left(\psi_{\gamma^0}\right)^{\leftarrow}(1)\right)\right]^{-1} = 1.$$

The scaling property (2.14) extends this to $\tilde{\nu}^0([\mathbf{1}, \boldsymbol{\infty}]) = 1$ and Proposition 3.1 of [22] and Proposition 2.5 yield the equivalent convergence in alternate coordinates given in (2.19) below.

To specify the coordinates, we need the following: Let ν_1 be a Pareto measure on $(0, \infty]$ satisfying $\nu_1((y, \infty]) = y^{-1}$ for $y > 0$. Since $U^{(2)}(\cdot)$ is non-decreasing, $U^{(2)}(X^{(2)}) = U^{(2)}(X^1) \wedge U^{(2)}(X^2)$. Set

$$\delta\aleph^{(2)} = \{\boldsymbol{x} \in (\mathbf{0}, \boldsymbol{\infty}] : x^{(2)} = 1\}.$$

Corollary 2.7. *If the hypotheses of Proposition 2.5 hold, the convergence in* (2.12) *is equivalent to*

$$nP\left[\left(\frac{U^{(2)}\left(X^{(2)}\right)}{n}, \left(\frac{U^{(2)}(X^1)}{U^{(2)}(X^{(2)})}, \frac{U^{(2)}(X^2)}{U^{(2)}(X^{(2)})}\right)\right) \in \cdot\right] \xrightarrow{v} \nu_1 \times S^0(\cdot), \tag{2.19}$$

on $(0, \infty] \times \delta\aleph^{(2)}$, *where* S^0 *is a probability measure on* $\delta\aleph^{(2)}$. *The relation between* $\tilde{\nu}^0$ *in* (2.12) *and* S^0 *is*

$$\tilde{\nu}^0\left(\left\{\boldsymbol{x} \in (0, \infty]^2 : x^{(2)} \ge r, \boldsymbol{x}/x^{(2)} \in \Lambda\right\}\right) = r^{-1}S^0(\Lambda), \tag{2.20}$$

for $r > 0$ *and Borel sets* $\Lambda \subset \delta\aleph^{(2)}$.

The probability measure S^0, called the *standardized hidden spectral measure*, is

$$P\left[\frac{\left(U^{(2)}(X^1), U^{(2)}(X^2)\right)}{U^{(2)}(X^{(2)})} \in \cdot \,\middle|\, X^{(2)} > t\right] \Rightarrow S^0(\cdot), \tag{2.21}$$

where the convergence holds as $t \to x_{H^{(2)}} = \sup\{y \in \mathbb{R} : H^{(2)}(y) < 1\}$. For HRV, a similar spectral measure was defined in [22].

In MEVT, a change to Pareto scale and then a change to polar coordinates $\boldsymbol{x} \mapsto (\|\boldsymbol{x}\|, \boldsymbol{x}/\|\boldsymbol{x}\|)$ produces a product measure and spectral measure ([12, page 214], [27, Chapter 5], [26, Chapter 6]). However on smaller cones, the polar coordinate transformation is problematic, since $\{\boldsymbol{x} > \mathbf{0} : \|\boldsymbol{x}\| \geq 1\}$ is not relatively compact in $(\mathbf{0}, \boldsymbol{\infty}]$ and hence a Radon measure on $\{\boldsymbol{x} > \mathbf{0} : \|\boldsymbol{x}\| \geq 1\}$ is not necessarily finite. This is remedied by using $\delta\aleph^{(2)}$.

2.3. **Estimation.** Recall the standard case assumes marginal distributions are the same. To estimate joint tail probabilities, we first estimate the limit measure ν^0 given in (2.6). Let, $\{\boldsymbol{X}, \boldsymbol{X}_i, i = 1, 2, \cdots, n\}$ be iid with a common distribution satisfying (2.6). From (2.6),

$$\frac{1}{k}\sum_{i=1}^{n} \epsilon_{\left(\frac{\boldsymbol{X}_i - d(n/k)\mathbf{1}}{c(n/k)}\right)}(\cdot) \Rightarrow \nu^0(\cdot) \qquad (k \to \infty, n/k \to \infty), \tag{2.22}$$

in $M_+(\mathbb{E}^0)$ [26, page 139]. From (2.7), the distribution of $X^{(2)}$ determines c_n and d_n. The iid data $\{X_i^{(2)} : i = 1, 2, \cdots, n\}$ allow estimates ([12], [26, page 93]) of $c(n/k)$ and $d(n/k)$, denoted by $\hat{c}(n/k)$ and $\hat{d}(n/k)$, satisfying

$$\frac{c(n/k)}{\hat{c}(n/k)} \xrightarrow{P} 1, \qquad \frac{d(n/k) - \hat{d}(n/k)}{c(n/k)} \xrightarrow{P} 0; \tag{2.23}$$

Therefore, we get the joint convergence

$$\left(\frac{1}{k}\sum_{i=1}^{n} \epsilon_{\left(\frac{\boldsymbol{X}_i - d(n/k)\mathbf{1}}{c(n/k)}\right)}, \frac{d(n/k) - \hat{d}(n/k)}{c(n/k)}, \frac{c(n/k)}{\hat{c}(n/k)}\right) \Rightarrow \big(\nu^0(\cdot), 0, 1\big) \tag{2.24}$$

in $M_+(\mathbb{E}^0) \times \mathbb{R}^2$. Apply the almost surely continuous map $(\nu(\cdot), b, a) \mapsto \nu(a[(\cdot) + b\mathbf{1}])$ in (2.24) and we get the following proposition:

Proposition 2.8. *Let, $\{\boldsymbol{X}, \boldsymbol{X}_i, i \geq 1\}$ be iid with common distribution satisfying* (2.6)*. Then,*

$$\widehat{\nu_n^0}(\cdot) := \frac{1}{k}\sum_{i=1}^{n} \epsilon_{\left(\frac{\boldsymbol{X}_i - \hat{d}(n/k)\mathbf{1}}{\hat{c}(n/k)}\right)}(\cdot) \Rightarrow \nu^0(\cdot) \quad \textit{in } M_+(\mathbb{E}^0), \qquad (k \to \infty, n/k \to \infty). \tag{2.25}$$

Estimation of $\nu^0(\cdot)$ in Proposition 2.8 does not exploit any semi-parametric structure discussed in Section 2.2 and has the disadvantage that there is no guarantee the estimator $\widehat{\nu_n^0}(\cdot)$ is a member of the class of possible limit measures. An alternative method, assuming the hypotheses of Proposition 2.5, estimates the extreme value index γ^0 of the distribution of $X^{(2)}$ as well as the standardized hidden spectral measure S^0. Since $\{X_j^{(2)} : j = 1, 2, \cdots, n\}$ is iid data, estimating γ^0 of $X^{(2)}$ is a known procedure [12, page 65] so we turn to estimating $S^0(\cdot)$. A modification of the ranks method [12, 13, 16, 26] to obtain an estimator of $\tilde{\nu}^0(\cdot)$ is as follows. For $i = 1, 2, \cdots, n$ and $p = 1, 2$, define

$$R_i^p = \big|\big\{j \leq n : X_j^{(2)} \geq X_i^p\big\}\big|. \tag{2.26}$$

where $|\cdot|$ denotes size of a set. Then for $i = 1, 2, \cdots, n$, $p = 1, 2$, $0 \leq R_i^p \leq n$ and since $R_i^1 \vee R_i^2 = |\{j : X_j^{(2)} \geq X_i^{(2)}\}|$, $1 \leq R_i^1 \vee R_i^2 \leq n$. Proposition 2.9 gives an estimator of $\tilde{\nu}^0$ which we modify to get an estimator of $S^0(\cdot)$.

Proposition 2.9. *Under the hypotheses of Proposition 2.5, we have in $M_+(\mathbf{0}, \boldsymbol{\infty}]$,*

$$\widehat{\tilde{\nu}_n^0}(\cdot) := \frac{1}{k}\sum_{i=1}^{n} \epsilon_{\left(k/R_i^1,\, k/R_i^2\right)}(\cdot) \Rightarrow \tilde{\nu}^0(\cdot) \qquad (k \to \infty, n/k \to \infty), \tag{2.27}$$

where (2.12) *defines* $\tilde{\nu}^0(\cdot)$ *and* (2.26) *defines* R_i^1 *and* R_i^2.

Proof. From (2.6) and the definition of $\psi_{\gamma^0}(\cdot)$ given in (2.5), we have in $D(\mathbb{R}^{\dashv}_{\gamma^0})$,

$$\frac{1}{k}\sum_{i=1}^{n}\epsilon_{\left((X_i^{(2)}-d(n/k))/c(n/k)\right)}((x,\infty]) \Rightarrow 1/\psi_{\gamma^0}(x).$$

Hence [26, page 58], inverse functions also converge in distribution in $D_{\text{left}}((0,\infty])$, the space of left continuous functions with finite right limits,

(2.28)
$$\inf\{x : \frac{1}{k}\sum_{i=1}^{n}\epsilon_{\left((X_i^{(2)}-d(n/k))/c(n/k)\right)}((x,\infty]) \le 1/s\} \Rightarrow \inf\{x : 1/\psi_{\gamma^0}(x) \le 1/s\} = \left(\psi_{\gamma^0}\right)^{\leftarrow}(s).$$

Write the order statistics of $\{X_1^{(2)},\ldots,X_n^{(2)}\}$ as $X_{(1)}^{(2)} \ge \cdots \ge X_{(n)}^{(2)}$ and observe the left side of (2.28) is

$$\inf\{x : \sum_{i=1}^{n}\epsilon_{\left((X_i^{(2)}-d(n/k))/c(n/k)\right)}((x,\infty]) \le k/s\} = \left(\frac{X_{(\lceil k/s\rceil)}^{(2)} - d(n/k)}{c(n/k)}\right). \tag{2.29}$$

From (2.22), (2.28) and (2.29) we get that as $k\to\infty$ and $n/k\to\infty$,

$$\begin{aligned}\Big(\frac{1}{k}\sum_{i=1}^{n}\epsilon_{\left(\frac{\boldsymbol{X}_i - d(n/k)\mathbf{1}}{c(n/k)}\right)}(\cdot), &\frac{X_{(\lceil k/s\rceil)}^{(2)} - d(n/k)}{c(n/k)}, \frac{X_{(\lceil k/t\rceil)}^{(2)} - d(n/k)}{c(n/k)}\Big)\\ &\Rightarrow \left(\nu^0(\cdot), \psi_{\gamma^0}^{\leftarrow}(s), \psi_{\gamma^0}^{\leftarrow}(t)\right)\end{aligned} \tag{2.30}$$

in $M_+(\mathbb{E}^0)\times D_{\text{left}}(0,\infty]\times D_{\text{left}}(0,\infty]$. Using the scaling technique as in [26, page 311] we get from (2.30) that as $k\to\infty$ and $n/k\to\infty$,

$$\frac{1}{k}\sum_{i=1}^{n}1_{\left\{X_i^1 > X_{(\lceil k/s\rceil)}^{(2)}, X_i^2 > X_{(\lceil k/t\rceil)}^{(2)}\right\}} \Rightarrow \nu^0\left((\psi_{\gamma^0}^{\leftarrow}(s),\infty]\times(\psi_{\gamma^0}^{\leftarrow}(t),\infty]\right) = \tilde{\nu}^0\left((s,\infty]\times(t,\infty]\right), \tag{2.31}$$

in $D_{\text{left}}(0,\infty])\times D_{\text{left}}((0,\infty])$. The left side of (2.31) is

$$\frac{1}{k}\sum_{i=1}^{n}1_{\left\{R_i^1 < k/s,\ R_i^2 < k/t\right\}} = \frac{1}{k}\sum_{i=1}^{n}1_{\left\{s<k/R_i^1,\ t<k/R_i^2\right\}},$$

and this suffices to prove (2.27). □

Proposition 2.9 yields an estimator of the limit measure $\nu_1\times S^0(\cdot)$ and then an estimator of $S^0(\cdot)$.

Corollary 2.10. *Under the hypotheses of Proposition 2.5, the convergence in* (2.27) *is equivalent to*

$$\widehat{\nu_1\times S^0}_n(\cdot) := \frac{1}{k}\sum_{i=1}^{n}\epsilon_{\left(\frac{k}{R_i^1\vee R_i^2}, \left(\frac{R_i^1\vee R_i^2}{R_i^1}, \frac{R_i^1\vee R_i^2}{R_i^2}\right)\right)}(\cdot) \Rightarrow \nu_1\times S^0(\cdot) \tag{2.32}$$

in $M_+((0,\infty]\times\delta\aleph^{(2)})$*, where* $\nu_1\times S^0(\cdot)$ *is given in Corollary 2.7. A consistent estimator of* $S^0(\cdot)$ *in* $M_+(\delta\aleph^{(2)})$ *is*

$$\hat{S}_n^0(\cdot) := \frac{\sum_{i=1}^n \epsilon_{\left(\frac{k}{R_i^1\vee R_i^2},\left(\frac{R_i^1\vee R_i^2}{R_i^1},\frac{R_i^1\vee R_i^2}{R_i^2}\right)\right)}([1,\infty]\times\cdot)}{\sum_{i=1}^n \epsilon_{\frac{k}{R_i^1\vee R_i^2}}([1,\infty])} \Rightarrow S^0(\cdot). \tag{2.33}$$

Proof. The proof uses Corollary 2.7 and follows exactly similar steps as that of Proposition 3.7 of [21]. It is based on the map $\boldsymbol{x}\mapsto(x^{(2)},\boldsymbol{x}/x^{(2)})$. □

It is possible that $R_i^j=0$ for some $j=1,2$ and some $i=1,2,\cdots,n$ and thus division by zero may be indicated in (2.33). Though theoretically justified, this is not desirable when writing code for an estimator. The continuous bijection $T:\delta\aleph^{(2)}\mapsto[0,1]$ given by $T:\boldsymbol{x}\mapsto x^2/(x^1+x^2)$ provides a remedy. With the conventions that $\infty/\infty=1$ and $1/\infty=0$, this transformation, (2.33) yields

$$\frac{\sum_{i=1}^n \epsilon_{\left(\frac{k}{R_i^1\vee R_i^2},\left(\frac{R_i^1}{R_i^1+R_i^2}\right)\right)}([1,\infty]\times\cdot)}{\sum_{i=1}^n \epsilon_{\frac{k}{R_i^1\vee R_i^2}}([1,\infty])} \Rightarrow S^0\circ T^{-1}(\cdot) \tag{2.34}$$

in $M_+([0,1])$. Since T is a continuous bijection, we retrieve S^0 from $S^0\circ T^{-1}$.

3. Non-standard hidden domain of attraction

To provide broader scope for applications, the non-standard case no longer assumes that $X^1\stackrel{d}{=}X^2$. With the goal being to approximate marginal and joint tail probabilities we consider how to frame definitions so as to construct an estimation theory leading to risk estimates. We present two approaches. The first is a non-parametric rank based method suitable for when the distribution of the minimum component of the standardized variables has an infinite endpoint and is in the domain of attraction of the Gumbel. The second approach departs from traditional MEVT by requiring that both components in (2.1) have the same centering and scaling but permitting the limit measure to have one zero marginal. This second approach may stimulate alternative methodologies.

3.1. **Standardization and ranks.**

3.1.1. *Assumptions.* Assume $\boldsymbol{X},\boldsymbol{X}_1,\boldsymbol{X}_2,\ldots,$ are iid with common distribution satisfying (1.2) with asymptotic independence, where the convergence in (1.2) is in $M_+(\mathbb{E})$ and $\mathbb{E}$ depends on the support of the limiting extreme value distribution in (1.1). Marginal convergence to non-degenerate limits also holds

$$nP[\frac{X^p-b_n^p}{a_n^p}>x^p]\to\psi_{\gamma^p}(x),\quad p=1,2;\ n\to\infty, \tag{3.1}$$

where for some $\gamma^p\in\mathbb{R}$, $\psi_{\gamma^p}(x)$ is defined in (2.5) and $G_{\gamma^p}(x)=\exp\{-1/\psi_{\gamma^p}(x)\}$ is the limiting extreme value distribution corresponding to maxima of the pth marginal.

Standardization: Assuming (1.2), define

$$U^p(x)=\frac{1}{P[X^p>x]},\quad p=1,2, \tag{3.2}$$

and the standardized variables

$$\tilde{X}^p=U^p(X^p),\quad p=1,2;\qquad \tilde{\boldsymbol{X}}=\left(\tilde{X}^1,\tilde{X}^2\right). \tag{3.3}$$

Then for the standardized variables (1.2) becomes regular variation

$$nP[\tilde{\boldsymbol{X}}/n \in \cdot\,] \xrightarrow{v} \tilde{\nu}(\cdot), \qquad (n \to \infty) \tag{RegVar}$$

on $[\mathbf{0}, \boldsymbol{\infty}] \setminus \{\mathbf{0}\}$, where we assume asymptotic independence so

$$\tilde{\nu}([\mathbf{0}, \boldsymbol{x}]^c) = (x^1)^{-1} + (x^2)^{-1}, \quad \boldsymbol{x} > \mathbf{0}. \tag{3.4}$$

Since the distributions of $\tilde{X}^p$, $p = 1, 2$ are asymptotically Pareto, their right endpoints are infinite.

3.1.2. *The hidden domain.* Now assume further that there exist $d(t) \in \mathbb{R}$, $c(t) > 0$ such that on either $\mathbb{E}^0 = (-\boldsymbol{\infty}, \boldsymbol{\infty}]$ or $\mathbb{E}^0 = (\mathbf{0}, \boldsymbol{\infty}]$ we have

$$nP\Big[\frac{\tilde{\boldsymbol{X}} - d(n)\mathbf{1}}{c(n)} \in \cdot\,\Big] \xrightarrow{v} \tilde{\nu}^0(\cdot). \tag{3.5}$$

If the right endpoint of $\tilde{X}^1 \wedge \tilde{X}^2$ is infinite, we can standardize this convergence so it becomes regular variation by defining

$$\tilde{U}^{(2)}(x) = \frac{1}{P[\tilde{X}^1 \wedge \tilde{X}^2 > x]}, \tag{3.6}$$

and shifting to the doubly standardized variables

$$\tilde{\tilde{X}}^p = \tilde{U}^{(2)}(\tilde{X}^p),\ p = 1, 2; \quad \tilde{\tilde{\boldsymbol{X}}} = \big(\tilde{\tilde{X}}^1, \tilde{\tilde{X}}^2\big). \tag{3.7}$$

Then regular variation holds on $(\mathbf{0}, \boldsymbol{\infty}]$,

$$nP[\tilde{\tilde{\boldsymbol{X}}}/n \in \cdot\,] \xrightarrow{v} \tilde{\tilde{\nu}}^0(\cdot), \tag{3.8}$$

and $\tilde{\tilde{\nu}}^0(\cdot)$ satisfies the scaling property as in (2.14). From (3.5) there is $\tilde{\gamma}^0 \in \mathbb{R}$, such that

$$\tilde{U}^{(2)}\big(c(n)x + d(n)\big)/n \to \psi_{\tilde{\gamma}^0}(x), \tag{3.9}$$

where $\psi_{\tilde{\gamma}^0}(x)$ is defined in (2.5), and this means $\tilde{X}^1 \wedge \tilde{X}^2$ has a distribution in a univariate domain of attraction of an extreme value distribution $G_{\tilde{\gamma}^0}(x)$. Assuming the right endpoint of $\tilde{X}^1 \wedge \tilde{X}^2$ is infinite implies $\tilde{\gamma}^0 \geq 0$. We are using the mnemonic that a tilde represents a level of standardization.

3.1.3. *Why ranks are a natural tool.* The functions $U^p, p = 1, 2$ and $\tilde{U}^{(2)}$ in (3.2), (3.3) and (3.6), (3.7) are statistically unknown. It is natural in (3.2) to replace $U^p(x) = 1/P[X^p > x]$ by an empirical version based on a random sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ of size n,

$$\hat{U}^p(x) = \frac{1}{\frac{1}{n}|\{j \leq n : X_j^p \geq x\}|},$$

so that

$$\tilde{X}_i^p = U^p(X_i^p) \approx \hat{U}^p(X_i^p) =: \hat{\tilde{X}}_i^p = n/r_i^p,$$

and

$$r_i^p := |\{j \leq n : X_j^p \geq X_i^p\}| \tag{3.10}$$

is the anti-rank of X_i^p among $X_1^p, \ldots, X_n^p$. So the observable n/r_i^p is a surrogate for the unknown standardized variable $\tilde{X}_i^p$.

Similarly,

$$\tilde{U}^{(2)}(x) := \frac{1}{P[\tilde{X}_1^1 \wedge \tilde{X}_1^2 > x]} \approx \hat{\tilde{U}}^{(2)}(x) := \frac{1}{\frac{1}{n}|\{j : \tilde{X}_j^1 \wedge \tilde{X}_j^2 \geq x\}|}$$

and replacing $\tilde{X}_j^p$ by $\hat{\tilde{X}}_j^p = n/r_j^p$, this is

$$\approx \frac{1}{\frac{1}{n}|\{j : n/r_j^1 \wedge n/r_j^2 \geq x\}|}.$$

Thus,

$$\begin{aligned}\hat{\tilde{X}}_i^p =& \tilde{U}^{(2)}(\tilde{X}_i^p) \approx \tilde{U}^{(2)}(n/r_i^p)\\ \approx& \hat{\tilde{U}}^{(2)}(n/r_i^p) = \frac{1}{\frac{1}{n}|\{j : n/r_j^1 \wedge n/r_j^2 \geq n/r_i^p\}|} =: n/R_i^p,\end{aligned}$$

where

$$R_i^p = \{j \leq n : \frac{1}{r_j^1} \wedge \frac{1}{r_j^2} \geq \frac{1}{r_i^p}\}.$$

So we expect estimation to use the statistics r_i^p, R_i^p, $i = 1, \ldots, n$; $p = 1, 2$.

3.1.4. *Required limit theorems.* We now give two results which underpin our non-parametric estimation method based on ranks. We assume $\tilde{\gamma}^0 = 0$ in (3.9) as this is consistent with the needed assumption that the right endpoint of $\tilde{X}^1 \wedge \tilde{X}^2$ is infinite. (Assuming $\tilde{\gamma}^0 > 0$ would reproduce hidden regular variation.) We rely on asymptotic normality for uniform intermediate order statistics ([12, 28] and [24, page 256]).

Proposition 3.1. *Suppose* (1.2)*, asymptotic independence and* (3.5) *hold and consider the case that* $\tilde{\gamma}^0 = 0$ *and* $c(n)x + d(n) \to \infty$ *as* $n \to \infty$*. This means that the distribution of* $\tilde{X}^1 \wedge \tilde{X}^2$ *is in the domain of attraction of the Gumbel distribution* G_0*, that* $d(\cdot)$ *is* Π*-varying and* $c(\cdot)$ *is slowly varying at* ∞ *[9, 12, 27]. Suppose further that for* $p = 1, 2$ *the functions* U^p *are strictly increasing on the support of the distribution of* X^p *so that*

$$r_i^p := |\{j \leq n : X_j^p \geq X_i^p\}| = |\{j \leq n : \tilde{X}_j^p \geq \tilde{X}_i^p\}|.$$

Then for any sequence $k = k(n)$ *satisfying* $k \to \infty$*,* $k/n \to 0$*, we have in* $M_+(-\boldsymbol{\infty}, \boldsymbol{\infty}]$*,*

$$\hat{\tilde{\nu}}^0(\cdot) := \frac{1}{k}\sum_{k=1}^n \epsilon_{\left(\frac{(n/r_i^1, n/r_i^2) - d(\frac{n}{k})\mathbf{1}}{c(\frac{n}{k})}\right)}(\cdot) \Rightarrow \tilde{\nu}^0(\cdot). \tag{3.11}$$

Comparing (3.11) and (3.5), we see that n/r_i^p, $p = 1, 2$ acts as the surrogate for the unobservable $\tilde{X}_i^p$, $p = 1, 2$. We write the order statistics of $\tilde{X}_i^p, 1 \leq i \leq n$ as $\tilde{X}_{(1)}^p \geq \cdots \geq \tilde{X}_{(n)}^p$.

Proof. For $y \in \mathbb{R}$, let $\lceil y \rceil$ be the smallest integer greater than or equal to y. For fixed $\boldsymbol{x} \in \mathbb{E}^0$, observe that

$$\frac{n/r_i^p - d(\frac{n}{k})}{c(\frac{n}{k})} > x^p \quad \text{iff } r_i^p < \frac{n}{c(\frac{n}{k})x^p + d(\frac{n}{k})} =: k_1^* = k_1^*(n, x) \text{ iff } r_i^p < \lceil k_1^* \rceil =: k^* = k^*(n, x),$$

which is equivalent to

$$\tilde{X}_i^p > \tilde{X}_{(k^*)}^p. \tag{3.12}$$

The function $d(\cdot)$ is Π-varying and hence slowly varying and also $d(t)/c(t) \to \infty$, $t \to \infty$ ([12], [27, Chapter 1]), and hence for fixed $\boldsymbol{x}$, as $n \to \infty$,

$$k^* \sim k_1^* = \frac{k \cdot \frac{n}{k}}{c(\frac{n}{k})x^p + d(\frac{n}{k})} \sim k\frac{\frac{n}{k}}{d(\frac{n}{k})} \to \infty,$$

and since $c(t)x + d(t) \to \infty$,

$$k^*/n \sim k_1^*/n = \frac{1}{c(\frac{n}{k})x^p + d(\frac{n}{k})} \to 0.$$

So the sequence k^* satisfies $k^* \to \infty$ and $k^*/n \to 0$ as required for an intermediate sequence. Note (3.12) is the same as

$$\frac{\tilde{X}_i^p - d(\frac{n}{k})}{c(\frac{n}{k})} > \frac{\tilde{X}_{(k^*)}^p - d(\frac{n}{k})}{c(\frac{n}{k})}$$

and

$$\begin{aligned}\frac{\tilde{X}_{(k^*)}^p - d(\frac{n}{k})}{c(\frac{n}{k})} =& \frac{\frac{\tilde{X}_{(k^*)}^p}{n/k_1^*} \cdot (c(\frac{n}{k})x^p + d(\frac{n}{k})) - d(\frac{n}{k})}{c(\frac{n}{k})} \\ =& \frac{\tilde{X}_{(k^*)}^p}{n/k_1^*}x^p + \frac{d(\frac{n}{k})}{c(\frac{n}{k})}\Big(\frac{\tilde{X}_{(k^*)}^p}{n/k_1^*} - 1\Big) = (1 + o_p(1))x^p + \frac{d(\frac{n}{k})}{c(\frac{n}{k})\sqrt{k^*}}\sqrt{k^*}\Big(\frac{\tilde{X}_{(k^*)}^p}{n/k_1^*} - 1\Big) \\ =& (1 + o_p(1))x^p + o(1)O_p(1) \to x^p.\end{aligned}$$

To verify these claims, note the $o_p(1)$ term results from $\tilde{X}_{(k^*)}^p/(n/k^*) \xrightarrow{P} 1$ for example from [12, page 41] or [26, page 297]. Since $c(\frac{n}{k})x^p + d(\frac{n}{k}) \sim d(\frac{n}{k})$, the term

$$\frac{d(\frac{n}{k})}{c(\frac{n}{k})\sqrt{k^*}} \sim \frac{d(\frac{n}{k})\sqrt{d(\frac{n}{k})}}{c(\frac{n}{k})\sqrt{n}} = \frac{d(\frac{n}{k})^{3/2}}{\sqrt{k}c(\frac{n}{k})\sqrt{n/k}}$$

and $d^{3/2}(t)/c(t)$ is slowly varying, $d^{3/2}(t)/\sqrt{t}c(t) \to 0$ and therefore $d(\frac{n}{k})/(c(\frac{n}{k})\sqrt{k^*}) = o(1) \to 0$. Finally, $\sqrt{k^*}\Big(\frac{\tilde{X}_{(k^*)}^p}{n/k^*} - 1\Big)$ is asymptotically normal, and hence $O_p(1)$ by the result of Smirnov ([28], [12, page 41], [26, Section 9.1.2]) but we have to show we can ignore the difference between k_1^* and k^*. This is readily accomplished since

$$\begin{aligned}\sqrt{k^*}\Big(\frac{\tilde{X}_{(k^*)}^p}{n/k_1^*} - 1\Big) =& \frac{\sqrt{k^*}}{n/k_1^*}n/k^*\Big(\frac{\tilde{X}_{(k^*)}^p - n/k^* + (n/k^* - n/k_1^*)}{n/k^*}\Big) \\ =& (1 + o(1))\sqrt{k^*}\Big(\frac{\tilde{X}_{(k^*)}^p}{n/k^*} - 1\Big) + (1 + o(1))\sqrt{k^*}\frac{(k_1^* - k^*)}{k_1^*}\end{aligned}$$

and the second term goes to zero since $|k_1^* - k^*| \le 1$.

To summarize, we have shown for $\boldsymbol{x} \in \mathbb{E}^0$,

$$\frac{\tilde{X}_{(k^*(n,x^p))}^p - d(\frac{n}{k})}{c(\frac{n}{k})} \xrightarrow{P} x^p, \quad p = 1, 2;\ n \to \infty. \tag{3.13}$$

Apply ([26, Theorem 5.3(ii), page 139]) to (3.5) and we get when $\tilde{\gamma}^0 = 0$,

$$q_n(\cdot) := \frac{1}{k}\sum_{i=1}^n \epsilon_{\frac{\tilde{\boldsymbol{X}}_i - d(\frac{n}{k})\mathbf{1}}{c(\frac{n}{k})}}(\cdot) \Rightarrow \tilde{\nu}^0(\cdot), \qquad \text{in } M_+(-\boldsymbol{\infty}, \boldsymbol{\infty}]. \tag{3.14}$$

Using (3.12) and (3.13), we can write,

$$\hat{\tilde{\nu}}^0(\boldsymbol{x}, \boldsymbol{\infty}] = \frac{1}{k}\sum_{k=1}^n \epsilon_{\Big(\frac{(n/r_i^1, n/r_i^2) - d(\frac{n}{k})\mathbf{1}}{c(\frac{n}{k})}\Big)}(\boldsymbol{x}, \boldsymbol{\infty}] = \frac{1}{k}\sum_{i=1}^n 1_{[\tilde{X}_i^p > \tilde{X}_{(k^*(n,x^p))}^p;\, p=1,2]}$$

$$=\frac{1}{k}\sum_{i=1}^{n}\epsilon_{\frac{\tilde{\boldsymbol{X}}_i-d(\frac{n}{k})\mathbf{1}}{c(\frac{n}{k})}}\left(\Big(\frac{\tilde{X}^p_{(k^*(n,x^p))}-d(\frac{n}{k})}{c(\frac{n}{k})};p=1,2\Big),\boldsymbol{\infty}\right]$$
$$=q_n\left(\Big(\frac{\tilde{X}^p_{(k^*(n,x^p))}-d(\frac{n}{k})}{c(\frac{n}{k})};p=1,2\Big),\boldsymbol{\infty}\right]=q_n\Big(\boldsymbol{x}+\big(o_p(1),o_p(1)\big),\boldsymbol{\infty}\Big].$$

This representation, plus a scaling argument and (3.14) yield the claimed result. □

Corollary 3.2. *Assume the hypotheses of Proposition 3.1 and define*

$$m_j=\frac{1}{r^1_j}\wedge\frac{1}{r^2_j},\quad \textit{and}\quad R^p_i:=|\{j\le n:m_j\ge\frac{1}{r^p_i}\}|.$$

Then for any sequence $k=k(n)$ satisfying $k\to\infty$, $k/n\to 0$ we have

$$\hat{\tilde{\nu}}^0(\cdot):=\frac{1}{k}\sum_{i=1}^{n}\epsilon_{\left(k/R^1_i,k/R^2_i\right)}(\cdot)\Rightarrow\tilde{\nu}^0(\cdot), \tag{3.15}$$

in $M_+(\mathbf{0},\boldsymbol{\infty}]$, where $\tilde{\tilde{\nu}}^0(\cdot)$ satisfies the scaling property in (2.14) *and for $\boldsymbol{x}>\mathbf{0}$,*

$$\tilde{\tilde{\nu}}^0(\boldsymbol{x},\boldsymbol{\infty}]=\tilde{\nu}^0\big((-\log x^1,\infty]\times(-\log x^2,\infty]\big). \tag{3.16}$$

Proof. Applying the definition of R^p_i, we observe for $x^p>0$, $p=1,2$,

$$k/R^p_i>x^p\quad\text{iff}\quad R^p_i<k/x^p\le\big\lceil k/x^p\big\rceil\quad\text{iff}\quad 1/r^p_i>m_{(\lceil k/x^p\rceil)},$$

where the order statistics of $m_j, 1\le j\le n$ are $m_{(1)}\ge\cdots\ge m_{(n)}$. Thus, write

$$\begin{aligned}\hat{\tilde{\nu}}^0(\boldsymbol{x},\infty]=&\frac{1}{k}\sum_{i=1}^{n}\epsilon_{\left(k/R^1_i,k/R^2_i\right)}(\boldsymbol{x},\boldsymbol{\infty}]=\frac{1}{k}\sum_{i=1}^{n}1_{[1/r^p_i>m_{(\lceil k/x^p\rceil)};p=1,2]}\\=&\frac{1}{k}\sum_{i=1}^{n}1_{\left[\frac{n/r^p_i-d(\frac{n}{k})}{c(\frac{n}{k})}>\frac{nm_{(\lceil k/x^p\rceil)}-d(\frac{n}{k})}{c(\frac{n}{k})};p=1,2\right]}\\=&\hat{\tilde{\nu}}^0\left(\Big(\frac{nm_{(\lceil k/x^1\rceil)}-d(\frac{n}{k})}{c(\frac{n}{k})},\frac{nm_{(\lceil k/x^2\rceil)}-d(\frac{n}{k})}{c(\frac{n}{k})}\Big),\boldsymbol{\infty}\right]\end{aligned} \tag{3.17}$$

We know from Proposition 3.1 that $\hat{\tilde{\nu}}^0\Rightarrow\tilde{\nu}^0$ in $M_+(-\boldsymbol{\infty},\boldsymbol{\infty}]$ and applying the minimum functional $\boldsymbol{x}\mapsto x^1\wedge x^2$ to (3.11) yields

$$\frac{1}{k}\sum_{i=1}^{n}\epsilon_{\frac{nm_i-d(\frac{n}{k})}{c(\frac{n}{k})}}(x,\infty]\Rightarrow\tilde{\nu}^0\{\boldsymbol{x}\in(-\boldsymbol{\infty},\boldsymbol{\infty}):x^1\wedge x^2>x\},\quad x\in\mathbb{R}, \tag{3.18}$$

and we assume $c(\cdot)$ and $d(\cdot)$ are chosen in the standard way in (3.9) so that $\tilde{\nu}^0\{\boldsymbol{x}\in(-\boldsymbol{\infty},\boldsymbol{\infty}):x^1\wedge x^2>x\}=e^{-x}$, $x\in\mathbb{R}$. The convergence in (3.18) is of a family of monotone functions with a continuous limit and taking inverses we get

$$\frac{nm_{(\lceil kx\rceil)}-d(\frac{n}{k})}{c(\frac{n}{k})}\Rightarrow-\log x,\quad x>0. \tag{3.19}$$

Combine (3.17), (3.11), (3.19) to get the result. □

3.1.5. *Estimating the probability of a risk region.* Suppose a risk vector $\boldsymbol{X}$ satisfies (1.2) with asymptotic independence and (3.5). How do the asymptotic results presented in the previous section aid in estimating $P[\boldsymbol{X} > \boldsymbol{x}]$ for extreme $\boldsymbol{x}$? Continue to suppose the assumptions of Proposition 3.1 so in particular $\tilde{\gamma}^0 = 0$ and $c(n)x + d(n) \to \infty$, for $x \in \mathbb{R}$.

Transform the tail probability:

$$\begin{aligned} P[\boldsymbol{X} >\boldsymbol{x}] &= P[U^p(X^p) > U^p(x^p);\, p = 1, 2] = P[\tilde{\boldsymbol{X}} > \big(U^p(x^p);\, p = 1, 2\big)] \\ &= P\Big[\frac{\tilde{\boldsymbol{X}} - d(\frac{n}{k})\mathbf{1}}{c(\frac{n}{k})} > \big(\frac{U^p(x^p) - d(\frac{n}{k})}{c(\frac{n}{k})};\, p = 1, 2\big)\Big] \\ &\approx \frac{k}{n}\tilde{\nu}^0\Big(\big(\frac{U^p(x^p) - d(\frac{n}{k})}{c(\frac{n}{k})};\, p = 1, 2\big), \boldsymbol{\infty}\Big] = \frac{k}{n}\tilde{\tilde{\nu}}^0\Big(\Big(\exp\{-\big(\frac{U^p(x^p) - d(\frac{n}{k})}{c(\frac{n}{k})}\big)\};\, p = 1, 2\Big), \boldsymbol{\infty}\Big], \end{aligned}$$

using (3.16). The limit measure $\tilde{\tilde{\nu}}^0$ is unknown so we replace it by $\hat{\tilde{\tilde{\nu}}}^0$ from (3.15).

Although $\hat{\tilde{\tilde{\nu}}}^0$ is observable, the overall expression still involves the unknowns $c(\cdot), d(\cdot), U^p;\, p = 1, 2$. From (3.19), we can set

$$\hat{d}(n/k) = nm_{(k)}, \quad \hat{c}(n/k) = n(m_{(\lceil ke \rceil)} - m_{(k)}).$$

The unknowns U^p must be estimated from marginal convergence in (1.2) using, say peaks over threshold methodology. This is a one dimensional issue for which there is much experience. Alternatively, as in Section 2.3, the scaling property of $\tilde{\tilde{\nu}}^0$ can be used to define a spectral distribution and Corollary 3.2 modified to define a spectral distribution estimator.

The practicality of these methods must be explored and software implementation is needed. For cases where the assumptions $\tilde{\gamma}^0 = 0$, $c(n)x + d(n) \to \infty$ fail, it is not clear how to proceed with joint tail estimation. There are literature and results ([2, 3]) but relevance to joint tail risk is not presently apparent to us.

3.2. **An untraditional approach.** In this section, we deviate from the traditional MEVT treatment by requiring that both components in (2.1) have the same centering and scaling but permitting the limit measure to have one zero marginal. By a zero marginal, we mean that either the limit measure $\nu(\cdot)$ in (2.1) has the property

$$\nu\left(\left\{\boldsymbol{z} \in \mathbb{E} : z^2 > y\right\}\right) = 0 \qquad ((y, y) \in \mathbb{E}) \tag{3.20}$$

or the same holds with z^1 in place of z^2. This could happen if the tail of X^2 is lighter than that of X^1 or vice versa. If $\mathbb{E} = [\mathbf{0}, \boldsymbol{\infty}] \setminus \{\mathbf{0}\}$, (3.20) means ν must concentrate either on one of the two axes emanating from $\mathbf{0}$. We explore the implications of these assumptions.

In the non-standard case, if we assume (2.1), the limit measure ν may satisfy:

(i) ν has a zero second marginal: for $(y, y) \in \mathbb{E}$, $\nu\left(\left\{\boldsymbol{z} \in \mathbb{E} : z^2 > y\right\}\right) = 0$;
(ii) ν has a zero first marginal: for $(y, y) \in \mathbb{E}$, $\nu\left(\left\{\boldsymbol{z} \in \mathbb{E} : z^1 > y\right\}\right) = 0$,
(iii) the cases (i) and (ii) do not hold, but $\nu\left(\left\{\boldsymbol{z} \in \mathbb{E} : z^1 > x, z^2 > y\right\}\right) = 0$ for $(x, y) \in \mathbb{E}$,
(iv) for $(x, y) \in \mathbb{E}$, $\nu\left(\left\{\boldsymbol{z} \in \mathbb{E} : z^1 > x, z^2 > y\right\}\right) > 0$.

Case (iv) means (2.1) yields non-zero estimates of the marginal and joint tail probabilities, so we have no need for HDA. The definition and analysis of HDA in case (iii) is the same as the standard case discussed in Section 2 since the two components are normalized the same. The definition and analysis of HDA are very similar for cases (i) and (ii) so focus only on case (i).

3.2.1. *Definition.* Our definition of HDA borrows a basic idea of the conditional extreme value (CEV) model [6–8, 14, 15] without imposing the CEV non-degeneracy assumption. A relevant state space is $\mathbb{E}^{\sqcap}$ where either $\mathbb{E}^{\sqcap} = [-\infty, \infty] \times (-\infty, \infty]$ or $\mathbb{E}^{\sqcap} = [0, \infty] \times (0, \infty]$.

Definition 3.3. The distribution of $\boldsymbol{X} = (X^1, X^2)$ has hidden domain of attraction on the cone $\mathbb{E}^{\sqcap}$ if (2.1) holds with the limit measure ν, the second marginal of ν is a zero measure and in addition, there exist constants $e_n > 0$ and $f_n \in \mathbb{R}$ and a non-zero measure $\nu^{\sqcap} \in M_+(\mathbb{E}^{\sqcap})$ such that as $n \to \infty$,

$$nP\left[\left(\frac{\mathbf{X} - f_n\mathbf{1}}{e_n}\right) \in \cdot\right] \xrightarrow{v} \nu^{\sqcap}(\cdot) \qquad \text{in } M_+(\mathbb{E}^{\sqcap}). \tag{3.21}$$

Note (3.21) does not preclude a similar convergence with different norming constants from holding on a smaller cone, say $\mathbb{E}^0$. See Definition 3.5 and Example 3.8. From (3.21) it follows that for $(y, y) \in \mathbb{E}^{\sqcap}$, as $n \to \infty$,

$$nP[X^2 > e_n y + f_n] \to \nu^{\sqcap}\left(\{(u, v) \in E^{\sqcap} : v > y\}\right), \tag{3.22}$$

and the distribution of X^2, the second component of $\boldsymbol{X}$, belongs to the maximal domain of attraction of an extreme value distribution [12, page 4]. Using one-dimensional extreme value theory, there exists $\gamma^{\sqcap} \in \mathbb{R}$ such that $\psi_{\gamma^{\sqcap}}(y) = 1/\nu^{\sqcap}\big(\{(u, v) \in E^{\sqcap} : v > y\}\big)$ takes the form in (2.5) and the parameter $\gamma^{\sqcap}$ is the extreme value index of X^2. We always choose $\{e_n\}$ and $\{f_n\}$ so that $\psi^{\sqcap}$ takes the form in (2.5).

Remark 3.4. Some comments about Definition 3.3.

(1) Definition 3.3 often permits non-zero approximation of joint tail probabilities when asymptotic independence holds in (1.2) with each component normalized differently.
(2) Since (2.1) holds, so does (2.3) and therefore the maximum component $X^{(1)}$ belongs to the maximal domain of attraction of some extreme value distribution. Since $X^{(1)} \geq X^2$, the convergence relation (2.1) constrains the possible convergences in (3.21). For example, if (2.3) has $X^{(1)}$ in the Gumbel domain of attraction, then the distribution of X^2 cannot have a regularly varying tail.
(3) The distribution of X^2 determines the cone $\mathbb{E}^{\sqcap}$ and (3.22) yields the scaling and centering constants $\{e_n\}$ and $\{f_n\}$. If the distribution of X^2 is in the Fréchet domain of attraction, $\mathbb{E}^{\sqcap} = [0, \infty] \times (0, \infty]$ and otherwise we take $\mathbb{E}^{\sqcap} = [-\infty, \infty] \times (-\infty, \infty]$.

There are two possibilities for the limit measure $\nu^{\sqcap}$ in (3.21):

(i) the limit measure $\nu^{\sqcap}$ puts zero mass on all sets $(x, \infty] \times (y, \infty]$ for $(x, y) \in \mathbb{E}^{\sqcap}$; or
(ii) the limit measure $\nu^{\sqcap}$ puts non-zero mass on one of the sets $(x, \infty] \times (y, \infty]$ for $(x, y) \in \mathbb{E}^{\sqcap}$.

Case (ii) typically leads to a solution to the problem of estimating joint tail probabilities and there is no reason to seek further instances of HDA. In case (i), $\nu^{\sqcap}(\cdot)$ fails to provide non-zero estimates of joint tail probabilities but HDA could still exist on a smaller cone such as $\mathbb{E}^0$ and if this is true we have a potential approximation. We formalize this idea in Definition 3.5, where the state space $\mathbb{E}^0$ is either $(-\infty, \infty]^2$ or $(0, \infty]^2$.

Definition 3.5. The distribution of $\boldsymbol{X} = (X^1, X^2)$ has hidden domain of attraction on the cones $\mathbb{E}^{\sqcap}$ and $\mathbb{E}^0$ if Definition 3.3 holds, the limit measure $\nu^{\sqcap}$ in (3.21) puts zero mass on sets of the form $(x, \infty] \times (y, \infty]$ for $(x, y) \in \mathbb{E}^{\sqcap}$, and in addition, there exist centering and scaling constants $\{c_n\}$ and $\{d_n\}$ and a non-zero measure $\nu^0 \in M_+(\mathbb{E}^0)$ such that (2.6) holds.

As noted for (2.7), the scaling and centering constants $\{c_n\}$ and $\{d_n\}$ in (2.6) are characterized by the distribution of $X^{(2)}$, the minimum component of $\boldsymbol{X}$, and there is γ^0 and ψ_{γ^0} given in (2.5) related to ν^0. Also, whether $\mathbb{E}^0$ in (2.6) is $(-\boldsymbol{\infty}, \boldsymbol{\infty}]$ or $(\mathbf{0}, \boldsymbol{\infty}]$ is determined by the distribution of $X^{(2)}$.

3.2.2. *Semi-parametric structure of $\nu^{\sqcap}$.* When endpoints match appropriately, both limit measures $\nu^{\sqcap}$ of (3.21) and ν^0 of (2.6) have semi-parametric structure. Assuming the right endpoints of X^1, X^2 and $X^1 \wedge X^2$ are the same, the semi-parametric structure of ν^0 was discussed in Section 2.2 and we discuss only the semi-parametric structure of $\nu^{\sqcap}$.

The distributions of $\boldsymbol{X}$ and X^2 are H and H^2 and the right endpoint of H^2 is x_{H^2}. Define

$$U^2(x) = 1/(1 - H^2(x)). \tag{3.23}$$

The following proposition relates (3.21) to a regular variation condition on $[0, \infty] \times (0, \infty]$. Its proof is similar to that of Proposition 2.5 and is omitted.

Proposition 3.6. *Assume the distributions of X^1, X^2 have the same right endpoint and are strictly increasing in a left neighborhood of the common right endpoint. Then* (3.21) *holds iff*

(1) X^2 has a distribution in a maximal domain of attraction so that for some $e_n > 0$, $f_n \in \mathbb{R}$ (3.22) *holds, and*

(2) regular variation on the cone $[0, \infty] \times (0, \infty]$ holds for the distribution of $(U^2(X^1), U^2(X^2))$:

$$nP\left[n^{-1}\left(U^2(X^1), U^2(X^2)\right) \in \cdot\right] \xrightarrow{v} \tilde{\nu}^{\sqcap}(\cdot) \qquad \text{in } M_+([0, \infty] \times (0, \infty]), \tag{3.24}$$

where (3.23) *defines $U^2(\cdot)$ and $\tilde{\nu}^{\sqcap}(\cdot)$ is a Radon measure on $[0, \infty] \times (0, \infty]$.*

The limit measure $\tilde{\nu}^{\sqcap}(\cdot)$ is related to $\nu^{\sqcap}(\cdot)$ in (3.21) *by the following: for $(x^1, x^2) \in [0, \infty] \times (0, \infty]$,*

$$\begin{aligned} \tilde{\nu}^{\sqcap}\big((x^1, \infty] \times (x^2, \infty]\big) &= \nu^{\sqcap}\Big(\big(\psi^{\leftarrow}_{\gamma^{\sqcap}}(x^1), \infty\big] \times \big(\psi^{\leftarrow}_{\gamma^{\sqcap}}(x^2), \infty\big]\Big), \\ \tilde{\nu}^{\sqcap}\big([0, x^1] \times (x^2, \infty]\big) &= \nu^{\sqcap}\Big(\big\{\boldsymbol{z} \in \mathbb{E}^{\sqcap} : z^1 \le \psi^{\leftarrow}_{\gamma^{\sqcap}}(x^1), z^2 > \psi^{\leftarrow}_{\gamma^{\sqcap}}(x^2)\big\}\Big). \end{aligned} \tag{3.25}$$

The scaling property $\tilde{\nu}^{\sqcap}(c\cdot) = c^{-1}\tilde{\nu}^{\sqcap}(\cdot)$, $c > 0$, holds for $\tilde{\nu}^{\sqcap}(\cdot)$ on $[0, \infty] \times (0, \infty]$,

On the semi-parametric structure of $\nu^{\sqcap}$: We will see that a probability measure $S^{\sqcap}$ on $[0, \infty]$ determines the limit measure $\tilde{\nu}^{\sqcap}$. The parameter $\gamma^{\sqcap}$ and the probability measure $S^{\sqcap}$ on $[0, \infty]$ determine $\nu^{\sqcap}$, since given $\gamma^{\sqcap}$ and measure $S^{\sqcap}$, we get the function $\psi_{\gamma^{\sqcap}}(\cdot)$ in (2.5) and $\tilde{\nu}^{\sqcap}$ which through (3.25) determines $\nu^{\sqcap}$.

The method that shows $\tilde{\nu}^0([1, \infty]^2) = 1$ also shows $\tilde{\nu}^{\sqcap}([0, \infty] \times [1, \infty]) = 1$. Proposition 4 of [14] and Proposition 3.6 give a convergence relation in new coordinates.

Corollary 3.7. *Under the hypotheses of Proposition 3.6, convergence in* (3.21) *is equivalent to*

$$nP\left[\left(\frac{U^2\left(X^2\right)}{n}, \frac{U^2(X^1)}{U^2(X^2)}\right) \in \cdot\right] \xrightarrow{v} \nu_1 \times S^{\sqcap}(\cdot) \qquad (\text{in } M_+((0, \infty] \times [0, \infty])), \tag{3.26}$$

where ν_1 is a Pareto measure on $(0, \infty]$ satisfying $\nu_1((x, \infty]) = x^{-1}$ for $x > 0$, and $S^{\sqcap}$ is a probability measure on $[0, \infty]$, called the standardized hidden spectral measure. The relation between $\tilde{\nu}^{\sqcap}$ given in (3.24) *and $S^{\sqcap}$ is*

$$\tilde{\nu}^{\sqcap}\left(\left\{\boldsymbol{x} \in [0, \infty] \times (0, \infty] : x^2 \ge r, x^1/x^2 \in \Lambda\right\}\right) = r^{-1}S^{\sqcap}(\Lambda), \tag{3.27}$$

which holds for all $r > 0$ and all Borel sets $\Lambda \subset [0, \infty]$.

3.2.3. *Examples.* We give examples of distributions of $\boldsymbol{X} = (X^1, X^2)$, $X^1 \stackrel{d}{\neq} X^2$, and which have the kind of HDA considered in this section. In Example 3.8, the limit measure $\nu^{\sqcap}$ of (3.21) puts zero mass on all sets of the form $(x^1, \infty] \times (x^2, \infty]$ for $\boldsymbol{x} \in \mathbb{E}^{\sqcap}$ and HDA also holds on $\mathbb{E}^0$. In Example 3.9, $\nu^{\sqcap}$ of (3.21) puts non-zero mass on sets of the form $(x^1, \infty] \times (x^2, \infty]$ for $\boldsymbol{x} \in \mathbb{E}^{\sqcap}$.

Example 3.8. Suppose $X^1 \sim \exp(1)$, $X^2 \sim \exp(2)$ and (X^1, X^2) are independent. Hence (X^1, X^2) are also asymptotically independent and to estimate joint tail probabilities, we proceed as follows. For $\boldsymbol{x} \in \mathbb{R}^2$ and n sufficiently large,

$$\begin{aligned} n\big(1 - P\big[X^1 - \log n \le x^1, X^2 - \log n \le x^2\big]\big) &= n\big[1 - \big(1 - e^{-(\log n + x^1)}\big)\big(1 - e^{-2(\log n + x^2)}\big)\big] \\ &\sim n\Big(e^{-(\log n + x^1)} + e^{-2(\log n + x^2)}\Big) \to e^{-x^1} \quad \text{as } n \to \infty, \end{aligned}$$

which implies (2.1) holds on $\mathbb{E} = [-\infty, \infty]^2 \setminus \{(-\infty, -\infty)\}$ with $\nu\left(\left([-\infty, x^1] \times [-\infty, x^2]\right)^c\right) = e^{-x^1}$ and ν puts mass only on $(-\infty, \infty] \times \{-\infty\}$, the horizontal line through $-\boldsymbol{\infty}$, and ν has zero second marginal. So we seek HDA on $\mathbb{E}^{\sqcap}$. For $\frac{1}{2}\log n + x^p > 0$, $p = 1, 2$ we have

$$nP\left[X^1 - \frac{\log n}{2} \le x^1, X^2 - \frac{\log n}{2} > x^2\right] = n\left(1 - e^{-(\frac{\log n}{2} + x^1)}\right) e^{-2(\frac{\log n}{2} + x^2)} \to e^{-2x^2},$$

as $n \to \infty$, and

$$nP\left[X^1 - \frac{\log n}{2} > x^1, X^2 - \frac{\log n}{2} > x^2\right] = ne^{-(\frac{\log n}{2} + x^1)} e^{-2(\frac{\log n}{2} + x^2)} \to 0.$$

Thus, HDA exists on $\mathbb{E}^{\sqcap} = [-\infty, \infty] \times (-\infty, \infty]$ with limit measure $\nu^{\sqcap}$, where $\nu^{\sqcap}([-\infty, x^1] \times (x^2, \infty]) = e^{-2x^2}$ and $\nu^{\sqcap}((x^1, \infty] \times (x^2, \infty]) = 0$ for $\boldsymbol{x} \in \mathbb{E}^{\sqcap}$. So $\nu^{\sqcap}$ concentrates on $\{-\infty\} \times (-\infty, \infty]$, the vertical line through $-\boldsymbol{\infty}$. After peeling away both lines through $-\boldsymbol{\infty}$, we look for HDA on $\mathbb{E}^0$. A hint for how to proceed is provided by $X^1 \wedge X^2 \sim \exp(3)$. Note that as $n \to \infty$,

$$nP\left[X^1 - \frac{\log n}{3} > x^1, X^2 - \frac{\log n}{3} > x^2\right] = ne^{-(\frac{\log n}{3} + x^1)} e^{-2(\frac{\log n}{3} + x^2)} \to e^{-(x^1 + 2x^2)}.$$

Thus HDA exists on $\mathbb{E}^0 = (-\infty, \infty]^2$ with limit measure ν^0, where $\nu^0((x^1, \infty] \times (x^2, \infty]) = e^{-(x^1 + 2x^2)}$ for $\boldsymbol{x} \in \mathbb{E}^0$.

To summarize Example 3.8, Definition 3.5 holds and HDA holds on both the cones $\mathbb{E}^{\sqcap}$ and $\mathbb{E}^0$, but the HDA on $\mathbb{E}^{\sqcap}$ is not informative for calculating risk probabilities where both components of the risk vector are large.

Example 3.9. Suppose E_1, E_2, E_3 are iid $\exp(1)$ random variables independent of $B \sim \text{Bernoulli}(1/2)$ and define $\boldsymbol{X}$ as

$$\boldsymbol{X} = B(E_1, E_3/3) + (1 - B)(E_2/2, E_2/2).$$

Then $\boldsymbol{X}$ possesses asymptotic independence since for $\boldsymbol{x} \in \mathbb{R}^2$

$$n\big(1 - P[X^1 - \log n \le x^1, X^2 - (1/2)\log n \le x^2]\big) \to \big(\exp\{-x^1\} + \exp\{-2x^2\}\big)/2.$$

How can we estimate joint tail probabilities? Note the right endpoints of X^1 and X^2 are the same. For $x^1 \wedge x^2 + \log n > 0$,

$$\begin{aligned} &n\left(1 - P\left[X^1 - \log n \le x^1, X^2 - \log n \le x^2\right]\right) \\ &\quad = n\left[1 - \left[\left(1 - e^{-(\log n + x^1)}\right)\left(1 - e^{-3(\log n + x^2)}\right)/2 + \left(1 - e^{-2(\log n + x^1 \wedge x^2)}\right)/2\right]\right] \to e^{-x^1}/2, \end{aligned}$$

which implies (2.1) holds on $\mathbb{E} = [-\infty,\infty]^2 \setminus \{(-\infty,-\infty)\}$ with $\nu\left(\left([-\infty,x^1]\times[-\infty,x^2]\right)^c\right) = e^{-x^1}/2$ and ν concentrates on $(-\infty,\infty]\times\{-\infty\}$. Thus ν has zero second marginal and we seek HDA on $\mathbb{E}^{\sqcap}$. As $n\to\infty$,

$$\begin{aligned} &nP\left[X^1 - \frac{\log n}{2} \le x^1, X^2 - \frac{\log n}{2} > x^2\right] \\ &= n\left[\frac{1}{2}\left(1 - e^{-(\frac{\log n}{2}+x^1)}\right)e^{-3(\frac{\log n}{2}+x^2)} + \frac{1}{2}\left(e^{-2(\frac{\log n}{2}+x^2)} - e^{-2(\frac{\log n}{2}+x^1)}\right)1_{\{x^2<x^1\}}\right] \\ &\to \frac{1}{2}\left(e^{-2x^2} - e^{-2x^1}\right)1_{\{x^2<x^1\}} \end{aligned}$$

and therefore, $nP[X^1 - \frac{1}{2}\log n > x^2] \to \frac{1}{2}\exp\{-2x^2\}$ and consequently,

$$nP\left[X^1 - \frac{\log n}{2} > x^1, X^2 - \frac{\log n}{2} > x^2\right] \to \frac{1}{2}e^{-2(x^1\vee x^2)}.$$

Thus, HDA exists on $\mathbb{E}^{\sqcap} = [-\infty,\infty]\times(-\infty,\infty]$ with limit measure $\nu^{\sqcap}$, where $\nu^{\sqcap}([-\infty,x^1]\times(x^2,\infty]) = \frac{1}{2}\left(e^{-2x^2} - e^{-2x^1}\right)1_{\{x^2<x^1\}}$ and $\nu^{\sqcap}((x^1,\infty]\times(x^2,\infty]) = \frac{1}{2}e^{-2(x^1\vee x^2)}$ for $\boldsymbol{x}\in\mathbb{E}^{\sqcap}$. In fact, $\nu^{\sqcap}$ concentrates on the line $\{(x,x) : x\in(-\infty,\infty)\}$.

Since $\nu^{\sqcap}((x^1,\infty]\times(x^2,\infty]) > 0$ for $\boldsymbol{x}\in\mathbb{E}^{\sqcap}$, we do not seek HDA on $\mathbb{E}^0$.

3.2.4. *Estimation methods.* To estimate joint tail probabilities, we require an estimate of the limit measure $\nu^{\sqcap}$ given in Definition 3.3 and possibly ν^0 given in Definition 3.5. Assuming the distributions of $X^1, X^2, X^1\wedge X^2$ have the same right endpoint, estimation of ν^0 follows the same steps as in Section 2.3 so we focus on estimating $\nu^{\sqcap}$. Let $\{\boldsymbol{X}, \boldsymbol{X}_i, i = 1,2,\cdots,n\}$ be iid where the distribution of $\boldsymbol{X}$ satisfies (3.21). From (3.21) we get [26, page 139] in $M_+(\mathbb{E}^{\sqcap})$

$$\frac{1}{k}\sum_{i=1}^{n}\epsilon_{\left(\frac{X_i^1 - f(n/k)}{e(n/k)}, \frac{X_i^2 - f(n/k)}{e(n/k)}\right)}(\cdot) \Rightarrow \nu^{\sqcap}(\cdot) \qquad (k\to\infty,\ n/k\to\infty). \tag{3.28}$$

We know from (3.22) that the distribution of X^2 characterizes $\{e_n\}$ and $\{f_n\}$ and from the iid data $\{X_i^2 : i = 1,2,\cdots,n\}$, we can construct estimators of $e(n/k)$ and $f(n/k)$ denoted by $\hat{e}(n/k)$ and $\hat{f}(n/k)$ [26, page 93] such that

$$\frac{e(n/k)}{\hat{e}(n/k)} \xrightarrow{P} 1, \qquad \frac{f(n/k) - \hat{f}(n/k)}{e(n/k)} \xrightarrow{P} 0. \tag{3.29}$$

Since the limits in (3.29) are constants, we get joint convergence in $M_+(\mathbb{E}^{\sqcap})\times\mathbb{R}^2$,

$$\left(\frac{1}{k}\sum_{i=1}^{n}\epsilon_{\left(\frac{X_i^1 - f(n/k)}{e(n/k)}, \frac{X_i^2 - f(n/k)}{e(n/k)}\right)}, \frac{f(n/k) - \hat{f}(n/k)}{e(n/k)}, \frac{e(n/k)}{\hat{e}(n/k)}\right) \Rightarrow \left(\nu^{\sqcap}(\cdot), 0, 1\right) \tag{3.30}$$

Apply the continuous mapping theorem to (3.30) using the map $(\nu(\cdot), b, a) \mapsto \nu(a[(\cdot) + b])$ to get in $M_+(\mathbb{E}^{\sqcap})$

$$\widehat{\nu_n^{\sqcap}}(\cdot) := \frac{1}{k}\sum_{i=1}^{n}\epsilon_{\left(\frac{X_i^1 - \hat{f}(n/k)}{\hat{e}(n/k)}, \frac{X_i^2 - \hat{f}(n/k)}{\hat{e}(n/k)}\right)}(\cdot) \Rightarrow \nu^{\sqcap}(\cdot) \qquad (k\to\infty,\ n/k\to\infty). \tag{3.31}$$

As in Section 2.3, where applicable, we exploit the semi-parametric structure of $\nu^{\sqcap}$. We assume the hypotheses of Proposition 3.6 and estimate $\gamma^{\sqcap}$ and the standardized hidden spectral measure $S^{\sqcap}$. The parameter $\gamma^{\sqcap}$ is the extreme value index of the distribution of X^2 so estimating this

from iid data $\{X_i^2 : i = 1, 2, \cdots, n\}$ is standard [12, page 65]. We obtain an estimator of $S^{\sqcap}(\cdot)$ by modifying (2.26) to account for the difference between $\mathbb{E}^0$ and $\mathbb{E}^{\sqcap}$. To help estimate $S^{\sqcap}(\cdot)$ we first construct a consistent estimator of $\tilde{\nu}^{\sqcap}(\cdot)$ defined in (3.24). Define

$$r_i^{1,2} := \Big| \big\{ j : X_j^2 \geq X_i^1 \big\} \Big| \quad \text{and} \quad r_i^{2,2} := \Big| \big\{ j : X_j^2 \geq X_i^2 \big\} \Big|, \qquad (i = 1, 2, \cdots, n,) \tag{3.32}$$

where $|\cdot|$ denotes size of a set. Note the definition in (3.32) is a little different than the antirank formulation in (3.10). An estimator of $\tilde{\nu}^{\sqcap}$ is obtained from the convergence in $M_+([0,\infty] \times (0,\infty])$:

$$\widehat{\tilde{\nu}_n^{\sqcap}}(\cdot) := \frac{1}{k} \sum_{i=1}^{n} \epsilon_{\left(k/r_i^{1,2},\, k/r_i^{2,2}\right)}(\cdot) \Rightarrow \tilde{\nu}^{\sqcap}(\cdot) \qquad (k \to \infty, n/k \to \infty). \tag{3.33}$$

The verification of (3.33) follows the steps used in the proof of Proposition 2.9. Changing coordinate systems in (3.33) leads to an estimator of $\nu_1 \times S^{\sqcap}(\cdot)$ from the convergence in $M_+((0,\infty] \times [0,\infty])$

$$\widehat{\nu_1 \times S^{\sqcap}}_n(\cdot) := \frac{1}{k} \sum_{i=1}^{n} \epsilon_{\left(k/r_i^{2,2}, r_i^{2,2}/r_i^{1,2}\right)}(\cdot) \Rightarrow \nu_1 \times S^{\sqcap}(\cdot) \qquad (k \to \infty, n/k \to \infty) \tag{3.34}$$

and this produces an estimator of $S^{\sqcap}$ since in $M_+([0,\infty])$,

$$\widehat{S_n^{\sqcap}}(\cdot) := \frac{\sum_{i=1}^{n} \epsilon_{\left(k/r_i^{2,2}, r_i^{2,2}/r_i^{1,2}\right)} \left([1,\infty] \times \cdot\right)}{\sum_{i=1}^{n} \epsilon_{k/r_i^{2,2}}([1,\infty])} \Rightarrow S^{\sqcap}(\cdot) \qquad (k \to \infty, n/k \to \infty). \tag{3.35}$$

This estimator may be modified as in (2.34) using the continuous bijection $TR : [0,\infty] \mapsto [0,1]$ defined by $TR : x \mapsto x/(1+x)$ to get in $M_+([0,1])$,

$$\widehat{S_n^{\sqcap}}(\cdot) \circ TR^{-1} \Rightarrow S^{\sqcap} \circ TR^{-1}(\cdot). \tag{3.36}$$

This summarizes how to obtain consistent estimators for extreme value index $\gamma^{\sqcap}$ and the standardized hidden spectral measure $S^{\sqcap}$.

4. Concluding remarks

Sections 2 and 3.1 assume the classical MEVT assumption that (X^1, X^2) has a joint distribution in the domain of attraction of a multivariate extreme value distribution with non-degenerate univariate extreme value marginals. Section 3.2 departs from the usual approach and assumes something that implies both $X^{(1)}$ and X^2 are in a domain of attraction. The second approach is suggested by the conditional extreme value (CEV) model (without the needed non-degeneracy conditions used in the CEV model) and though this approach sometimes seems ad hoc, it potentially can be useful. For either approach, detection and estimation techniques can be devised.

We restricted discussion to two dimensions and as observed for HRV [21], extensions to higher dimensions are not always straightforward and involve subtleties. In particular, in higher dimensions there are many more ways domains of attraction could be hidden and many more subspaces to explore for behavior that helps to estimate risk probabilities. The existence of mass on lines through $\boldsymbol{\infty}$ in limiting measures is a troubling issue and raises the possibility both that the convergence to types theorem may fail and that approximations might not be very accurate.

We showed some examples where HRV is absent but HDA is present and this broadens the ideas behind searching for hidden risks. However, currently it appears that for HDA to be useful for multivariate tail probability estimation, endpoints of the distributions of marginal distributions and minimum components must appropriately match. Without this match, the relevance of this theory to risk estimation where multiple components are large is not obvious.

We believe the methods outlined here can help fill gaps in theory needed to provide improved estimates of probability of simultaneous exceedance by components of a risk vector. We have not yet implemented nor extensively tested the methods nor demonstrated utility by analyzing data. Further, our estimators are only provably consistent. More formal statistical theory is needed to turn exploratory methods into confirmatory ones.

Acknowledgement: We are grateful to two resourceful and conscientious referees and a wise associate editor who contributed many helpful criticisms and suggestions that resulted in a substantially improved and corrected version of this work.

References

[1] A. A. Balkema and L. de Haan. On R. von Mises' condition for the domain of attraction of $\exp(-e^{-x})$[1]. *Ann. Math. Statist.*, 43:1352–1354, 1972. ISSN 0003-4851.

[2] A. A. Balkema and L. de Haan. Limit distributions for order statistics. I. *Teor. Verojatnost. i Primenen.*, 23(1):80–96, 1978. ISSN 0040-361x.

[3] A. A. Balkema and L. de Haan. Limit distributions for order statistics. II. *Teor. Verojatnost. i Primenen.*, 23(2):358–375, 1978. ISSN 0040-361x.

[4] P. Billingsley. *Convergence of Probability Measures*. John Wiley & Sons Inc., New York, second edition, 1999. ISBN 0-471-19745-9. A Wiley-Interscience Publication.

[5] J.T. Bruun and J.A. Tawn. Comparison of approaches for estimating the probability of coastal flooding. *J. R. Stat. Soc., Ser. C, Appl. Stat.*, 47(3):405–423, 1998.

[6] B. Das and S.I. Resnick. Conditioning on an extreme component: Model consistency with regular variation on cones. *Bernoulli*, 17(1):226–252, 2011. ISSN 1350-7265. doi: 10.3150/10-BEJ271.

[7] B. Das and S.I. Resnick. Detecting a conditional extreme value model. *Extremes*, 14(1):29–61, 2011.

[8] B. Das, A. Mitra, and S. Resnick. Living on the multi-dimensional edge: Seeking hidden risks using regular variation. *Advances in Applied Probability*, 45(1), 2013. ArXiv e-prints 1108.5560.

[9] L. de Haan. *On Regular Variation and Its Application to the Weak Convergence of Sample Extremes*. Mathematisch Centrum Amsterdam, 1970.

[10] L. de Haan. A characterization of multidimensional extreme-value distributions. *Sankhyā Ser. A*, 40(1):85–88, 1978. ISSN 0581-572X.

[11] L. de Haan and J. de Ronde. Sea and wind: multivariate extremes at work. *Extremes*, 1(1): 7–46, 1998.

[12] L. de Haan and A. Ferreira. *Extreme Value Theory: An Introduction*. Springer-Verlag, New York, 2006.

[13] J.E. Heffernan and S.I. Resnick. Hidden regular variation and the rank transform. *Adv. Appl. Prob.*, 37(2):393–414, 2005.

[14] J.E. Heffernan and S.I. Resnick. Limit laws for random vectors with an extreme component. *Ann. Appl. Probab.*, 17(2):537–571, 2007. ISSN 1050-5164. doi: 10.1214/105051606000000835.

[15] J.E. Heffernan and J.A. Tawn. A conditional approach for multivariate extreme values (with discussion). *JRSS B*, 66(3):497–546, 2004.

[16] Xin Huang. *Statistics of Bivariate Extreme Values*. Ph.D. thesis, Tinbergen Institute Research Series 22, Erasmus University Rotterdam, Postbus 1735, 3000DR, Rotterdam, The Netherlands, 1992.

[17] A.W. Ledford and J.A. Tawn. Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187, 1996. ISSN 0006-3444.

[18] A.W. Ledford and J.A. Tawn. Modelling dependence within joint tail regions. *J. Roy. Statist. Soc. Ser. B*, 59(2):475–499, 1997. ISSN 0035-9246.

[19] A.W. Ledford and J.A. Tawn. Concomitant tail behaviour for extremes. *Adv. in Appl. Probab.*, 30(1):197–215, 1998. ISSN 0001-8678.

[20] K. Maulik and S.I. Resnick. Characterizations and examples of hidden regular variation. *Extremes*, 7(1):31–67, 2005.

[21] A. Mitra and S.I. Resnick. Hidden Regular Variation: Detection and Estimation. *Arxiv preprint arXiv:1001.5058*, 2010.

[22] A. Mitra and S.I. Resnick. Hidden regular variation and detection of hidden risks. *Stochastic Models*, 27(4):591–614, 2011.

[23] S.-H Poon, M. Rockinger, and J. Tawn. Modelling extreme-value dependence in international stock markets. *Statist. Sinica*, 13(4):929–953, 2003. ISSN 1017-0405. Statistical applications in financial econometrics.

[24] S.I. Resnick. *A Probability Path*. Birkhäuser, Boston, 1999.

[25] S.I. Resnick. Hidden regular variation, second order regular variation and asymptotic independence. *Extremes*, 5(4):303–336 (2003), 2002. ISSN 1386-1999.

[26] S.I. Resnick. *Heavy Tail Phenomena: Probabilistic and Statistical Modeling*. Springer Series in Operations Research and Financial Engineering. Springer-Verlag, New York, 2007. ISBN: 0-387-24272-4.

[27] S.I. Resnick. *Extreme Values, Regular Variation and Point Processes*. Springer, New York, 2008. ISBN 978-0-387-75952-4. Reprint of the 1987 original.

[28] N. V. Smirnov. Limit distributions for the terms of a variational series. *Trudy Mat. Inst. Steklov.*, 25:60, 1949. ISSN 0371-9685.

[29] R.L. Smith. Statistics of extremes, with applications in environment, insurance and finance. In B. Finkenstadt and H. Rootzén, editors, *SemStat: Seminaire Europeen de Statistique, Exteme Values in Finance, Telecommunications, and the Environment*, pages 1–78. Chapman-Hall, London, 2003.

Abhimanyu Mitra, School of ORIE, Cornell University, Ithaca, NY 14853
E-mail address: am492@cornell.edu

Sidney I. Resnick, School of ORIE, Cornell University, Ithaca, NY 14853
E-mail address: sir1@cornell.edu