On the universal calibration of Pareto-type linear combination tests

Parijat Chakraborty, F. Richard Guo, Kerby Shedden, and Stilian Stoev

Department of Statistics, University of Michigan, Ann Arbor

Abstract: It is often of interest to test a global null hypothesis using multiple, possibly dependent, p-values by combining their strengths while controlling the Type I error. Recently, several heavy-tailed combinations tests, such as the harmonic mean test and the Cauchy combination test, have been proposed: they map p-values into heavy-tailed random variables before combining them in some fashion into a single test statistic. The resulting tests, which are calibrated under the assumption of independence of the p-values, have shown to be rather robust to dependence. The complete understanding of the calibration properties of the resulting combination tests of dependent and possibly tail-dependent p-values has remained an important open problem in the area. In this work, we show that the powerful framework of multivariate regular variation (MRV) offers a nearly complete solution to this problem.

We first show that the precise asymptotic calibration properties of a large class of homogeneous combination tests can be expressed in terms of the angular measure – a characteristic of the asymptotic tail-dependence under MRV. Consequently, we show that under MRV, the Pareto-type linear combination tests, which are equivalent to the harmonic mean test, are universally calibrated regardless of the tail-dependence structure of the underlying p-values. In contrast, the popular Cauchy combination test is shown to be universally honest but often conservative; the Tippet combination test, while being honest, is calibrated if and only if the underlying p-values are tail-independent.

One of our major findings is that the Pareto-type linear combination tests are *the only universally calibrated* ones among the large family of possibly non-linear homogeneous heavy-tailed combination tests.

Keywords and phrases: Global hypothesis testing, Cauchy combination test, multivariate regular variation, heavy tails, Pareto.

1. Introduction

It is often of interest to test a global null hypothesis using multiple p-values, each of which is marginally uniformly distributed on the unit interval if the global null holds. Examples are plenty, including set-based analysis in GWAS (Wu et al., 2010), rare-variant analysis in genetics (Liu et al., 2019), meta-analysis (Singh, Xie and Strawderman, 2005), variable and model selection (Meinshausen and Bühlmann, 2010), derandomizing data splitting (Guo and Shah, 2025), to name a few. Depending on the construction of these p-values, they are often (though not always) correlated and their dependence structure is typically unknown. Further, in this paper we focus on the setting where the raw data for constructing these p-values are unavailable and we must treat the p-values themselves as the summary of all the evidence we have against the global null hypothesis. That being said, when available, the raw data can be very valuable for estimating the dependence structure (Guo and Shah, 2025), but it is beyond the scope of this paper.

In the above setting, it is natural to consider a *combination test* that outputs a single p-value by combining the strengths from multiple p-values, an idea that dates back to the early works of Tippett (1937), Fisher (1948), Good (1958), Lancaster (1961) and Simes (1986). Ideally, the combined p-value has more power against the global null than any of the original p-values. While

the early works in this area often assume independence or asymptotic independence of the p-values, more recent development has shifted towards methods that can control the (family-wise) type I error, at least approximately, under a wide variety of (and potentially strong) dependence among the p-values: see, for example, Meng (1994); Wilson (2019); Liu and Xie (2020); Vovk and Wang (2020); DiCiccio, DiCiccio and Romano (2020) and Vovk and Wang (2021).

1.1. Motivation and related work

The most notable recent development is the heavy-tailed combination tests, which combine multiple dependent p-values after transforming them to heavy-tailed random variables such as Pareto or Cauchy (Wilson, 2019; Liu and Xie, 2020). Notably, Liu and Xie (2020) proposed the Cauchy Combination Test (CCT) and prior to them Wilson (2019) proposed the harmonic mean combination test, which dates back to Good (1958). The idea behind both of these tests is to transform the p-values to heavy-tailed Pareto-type scale, and take a linear combination of the resulting variables.

Specifically, let P_i , $i = 1, \dots, d$ be p-values associated with d tests, which have the Uniform(0, 1) distribution, under the composite null hypothesis \mathcal{H}_0 . Consider a cumulative distribution function F and let

$$X_i := F^{-1}(1 - P_i), i = 1, \dots, d.$$

We assume that F has a heavy right tail, i.e.,

$$1 - F(x) \sim L(x)x^{-\beta}$$
, as $x \to \infty$,

for some positive tail exponent β and a slowly varying function L. A function L is said to be slowly varying (at infinity) if $L(tx)/L(t) \to 1$, as $t \to \infty$, for all x > 0 (see e.g. Resnick, 1987). Then, for some positive weights w_i , $i = 1, \dots, d$, one can consider the *linear* combination test statistic:

$$T_{F,w} := \sum_{i=1}^{d} w_i X_i$$
, where $\sum_{i=1}^{d} w_i = 1$.

Observe that now extreme values of the X_i 's correspond to small p-values P_i 's and thus the extremes of the statistic $T_{F,w}$ favor the alternative hypothesis. Therefore, the combination test would reject the composite null hypothesis when $T_{F,w} > \tau$, where the threshold $\tau = \tau_{\alpha}$ is calibrated to a desired level $\alpha \in (0,1)$. The combination test is said to be *calibrated* if

$$\mathbb{P}_0[T_{F,w} > \tau] = \alpha,$$

while it is said to be honest if the above probability under the composite null hypothesis is less than or equal to α . Here \mathbb{P}_0 denotes the probability measure under the global null, i.e., if $\mathcal{H}_{0,i}$, $i=1,\ldots,d$ represent the d many null hypotheses, under \mathbb{P}_0 all of them are true. Picking F to be the Cauchy distribution leads to the CCT (Liu and Xie, 2020), while taking F to be the standard 1-Pareto distribution F(x) = 1 - 1/x, $x \geq 1$ recovers a test equivalent to the harmonic weighted mean of p-values (Wilson, 2019; Good, 1958). The Cauchy Combination Test of Liu and Xie (2020) is particularly elegant since in the case when the X_i 's are independent, the sum-stability property of the Cauchy distribution implies that

$$T_{F,w} \stackrel{d}{=} \left(\sum_{i=1}^{d} w_i\right) \cdot X_1 = X_1,$$

so that $T_{F,w}$ can be calibrated precisely under the null using the standard Cauchy quantile function (see also Example 4.5 below for more details on multivariate stable distributions). A number of simulation and theoretical studies have found that this calibration is robust to dependence in the p-values and in fact asymptotic calibration holds under certain tail-independence conditions on the X_i 's (see e.g. Theorem 1 in Liu and Xie, 2020).

Definition 1.1 (Asymptotic Calibration and Honesty). For a set of thresholds $\tau = \tau_{\alpha}$, we shall say that $T_{F,w}$ is asymptotically calibrated if for the probability under the composite null hypothesis \mathcal{H}_0 , we have:

$$\lim_{\alpha \downarrow 0} \alpha^{-1} \mathbb{P}_0[T_{F,w} > \tau_{\alpha}] = 1.$$

We shall say that it is asymptotically honest if $\limsup_{\alpha \downarrow 0} \alpha^{-1} \mathbb{P}_0[T_{F,w} > \tau_\alpha] \leq 1$.

In many applications, extremely small levels of α are of interest and the above asymptotic notions of calibration and honesty are useful criteria for the control of the Type I error. Henceforth when we talk about calibration and honesty, unless stated otherwise we mean asymptotic calibration and asymptotic honesty, respectively.

The first key problem in this methodological framework is to correctly (asymptotically) calibrate the resulting statistic under the null hypothesis, especially if the underlying p-values are dependent. The seminal works of Wilson (2019) and Liu and Xie (2020) establish the asymptotic calibration either in the independent or asymptotically independent regime, which we define below.

Definition 1.2 (Upper Tail dependence Coefficient and Asymptotic Independence). The (upper) tail-dependence coefficient of the variables X_i and X_j with common distribution function F is defined as:

$$\lambda(X_i, X_j) := \lim_{p \uparrow 1} \mathbb{P}[F(X_i) > p | F(X_j) > p], \tag{1.1}$$

whenever the limit exists. When $\lambda(X_i, X_j) = 0$ we say that X_i, X_j are asymptotically (upper tail) independent. Otherwise, they are asymptotically (upper tail) dependent.

Relation (1.1) corresponds to a bivariate lower-tail dependence coefficient of the copula of the original p-values P_i , $i=1,\cdots,d$ (see, e.g. Joe, 2015). A well-known result dating back to Sibuya (1960) shows that random variables with non-degenerate bivariate normal copula are always tail-independent. Thus the conditions in Theorem 1 of Liu and Xie (2020) amount to assuming bi-variate tail-independence. Since the seminal work of Liu and Xie (2020), there have been studies (Fang et al., 2023; Gui, Jiang and Wang, 2025) that establish the asymptotic calibration of the CCT but they all assume some form of asymptotic tail-independence. On the other hand, tail-dependent p-values, where $\lambda(X_i, X_j) > 0$, arise in many statistical contexts (cf Section 4, below). Thus, it has been of great interest to show whether or not the CCT and other heavy-tailed combination tests that are calibrated under independence remain asymptotically calibrated under potentially arbitrary tail-dependence.

In this work, we show that multivariate regular variation provides an elegant and nearly complete solution to this problem. The recent paper of Gui et al. (2025) used very similar if not identical technical tools from the theory of multivariate regular variation to address the problem of asymptotic calibration of heavy-tailed combination tests. Their elegant and comprehensive analysis investigates also the power of such tests. Our results, derived concurrently and independently, provide a complementary treatment, where the focus is primarily on the asymptotic calibration properties for the broader family of homogeneous combination tests. While we do not study power, we provide a

complete characterization of the *universally calibrated* such tests. The core technical tools underpinning the above results can be traced back to the works of Barbe, Fougères and Genest (2006) and Embrechts, Lambrigger and Wüthrich (2009) in the context of quantifying *extreme value of risk* (see also Yuen, Stoev and Cooley, 2020).

1.2. Main ideas and summary of our results

Multivariate regular variation has been one of the most fundamental notions in extreme value theory (Resnick, 1987; Beirlant et al., 2004; Resnick, 2007; Kulik and Soulier, 2020). It provides a powerful and natural framework for the study of asymptotic dependence of the components of a random vector $X = (X_i)_{i=1}^d$, conditionally on its norm ||X|| being extreme (see Section 2). Specifically, the random vector $X = (X_i)_{i=1}^d$ is said to be multivariate regularly varying (MRV) if there exists a non-zero Borel measure on $\mathbb{R}^d \setminus \{0\}$ and a sequence $a_n \uparrow \infty$, such that, as $n \to \infty$,

$$\mu_n(A) := n\mathbb{P}[X/a_n \in A] \to \mu(A), \tag{1.2}$$

for all μ -continuity sets A, which are bounded away from the origin (cf Sections 2 and A, for more details). By taking a set $A = A_i \cap A_j$, where $A_i := \{X_i > 1\}$, in view (1.1), Relation (1.2) readily implies that

$$\lambda(X_i, X_j) = \lim_{n \to \infty} \frac{\mu_n(A_i \cap A_j)}{\mu_n(A_j)} = \frac{\mu(A_i \cap A_j)}{\mu(A_j)}.$$

where the $\mu(A_i) = \mu(A_j) = \lim_{n\to\infty} n\mathbb{P}[X_1 > a_n]$ play the role of asymptotic scale coefficients relative to the normalization $\{a_n\}$.

This brief argument shows that the tail-dependence coefficients are simply and explicitly expressed in terms of the so-called *exponent measure* μ . More generally, the key Lemma 2.1 below shows that for every continuous non-negative and 1-homogeneous function $h: \mathbb{R}^d \to \mathbb{R}_+$, we have

$$n\mathbb{P}[h(X) > a_n] \to \sigma(h^{\beta}), \quad \text{as } n \to \infty,$$
 (1.3)

where the asymptotic scale $\sigma(h^{\beta})$ can be expressed via μ . Namely, given any fixed norm $\|\cdot\|$ in \mathbb{R}^d the exponent measure μ can be expressed as a product measure of radial and angular components (Theorem 2.2), where the radial component is a an infinite power-law type measure and the angular one is a finite measure σ supported on the unit sphere $S = S_{\|\cdot\|} = \{x \in \mathbb{R}^d : \|x\| = 1\}$. Thus, the asymptotic scale in (1.3) can be written as

$$\sigma(h^{\beta}) = \int_{S} h(\theta)^{\beta} \sigma(d\theta). \tag{1.4}$$

The angular measure σ is sometimes referred to as the *spectral* measure associated with μ in the norm $\|\cdot\|$. The intuition behind $\sigma(d\theta)$ is that it controls what proportion of the extreme values of X cluster along a particular direction $\theta \in S$, conditionally on the event that $\|X\|$ is extreme (cf Theorem 2.2 below for more details).

Relation (1.4) readily solves the asymptotic calibration problem for the general class of *homogeneous* combination tests of the form:

$$T_h(X) := h(X_1, \cdots, X_d),$$

where $h: \mathbb{R}^d \to \mathbb{R}_+$ can be an arbitrary continuous 1-homogeneous function. For example, taking F to be the standard 1-Pareto distribution F(x) = 1 - 1/x, $x \ge 1$, we have $\beta = 1$ and with

$$h(x) := \left(w_1 \cdot x_1 + \dots + w_d \cdot x_d\right)_+,$$

we obtain the Pareto Combination Test (PCT). In this case the measure σ in (1.4) is supported on the non-negative part of the unit sphere $S_+ := S \cap \mathbb{R}^d_+$ and from (1.3) and (1.4) we obtain

$$n\mathbb{P}[h(X) > a_n] \to \int_S h(\theta)\sigma(d\theta) = \int_{S_+} \sum_{i=1}^d w_i \cdot \theta_i \sigma(d\theta) = \sum_{i=1}^d w_i \int_{S_+} \theta_i \sigma(d\theta) = c \sum_{i=1}^d w_i = c,$$

where by the standardization of the marginals, we have that for all $i = 1, \dots, d$:

$$n\mathbb{P}[X_i > a_n] \to c = \int_S (\theta_i)_+ \sigma(d\theta) = \int_{S_+} \theta_i \sigma(d\theta) > 0.$$

The last two relations imply

$$\frac{\mathbb{P}[h(X) > a_n]}{\mathbb{P}[X_1 > a_n]} \to 1, \quad \text{as } n \to \infty,$$

where in fact the sequence $a_n \uparrow \infty$ can be chosen to be arbitrary. This proves that the PCT is *universally calibrated* in the large family of all multivariate regularly varying distributions (cf Corollary 2.1).

Similar ideas lead us to the complete characterization of the asymptotic calibration of the Cauchy combination tests, the Tippet combination test, among others (cf Corollary 2.2 and Example 4.6, below) in the general MRV dependence context.

The main result in this paper is Theorem 3.1. It shows that a homogeneous heavy-tailed combination test is asymptotically calibrated with respect to all MRV dependence structures if and only if it is of linear Pareto-type. In this sense, the linear Pareto-type combination tests are the only universally calibrated ones. The proof of this result relies on a general characterization of integral functionals under constraints, which may be of independent interest (see Section B).

The paper is structured as follows. In Section 2, we outline a framework for the study of the calibration of heavy-tailed combination tests using the powerful notion of multivariate regular variation. In Section 2.1, we begin with a review of the theory of multivariate regular variation and show that it is natural tool to quantify and study the important notion of tail-dependence. We then state a key technical result (Lemma 2.1). Using this result, in Section 2.2, we show how the calibration level of a general homogeneous combination test can be expressed through an integral functional of the angular measure. This readily implies the universal honesty of several popular tests such as the Tippet and Cauchy combination test. In the specific case of the linear Pareto-type combination test we obtain exact asymptotic calibration with respect to all possible angular measures. We refer to this property as to universal calibration. The main result in the paper is presented and proved in Section 3. Namely, it shows that a homogeneous combination test is universally calibrated if and only if it is of linear Pareto-type. In this sense, the linear Pareto-type combination tests (which are equivalent to the weighted harmonic averages) may be seen as universal since they are the only ones that are calibrated under both tail-independence and tail-dependence.

Section 4 provides several analytical and numerical examples illustrating the general theory. The Appendix contains further details on multivariate regular variation as well as a general measure theoretic result used to characterize the universal homogeneous combination tests.

2. Multivariate regular variation and asymptotic calibration of combination tests

In this section, we will review the fundamental notion of multivariate regular variation. This framework, while very well-understood and developed in the literature on extreme value theory (see, e.g. Resnick, 1987; Beirlant et al., 2004; de Haan and Ferreira, 2006; Resnick, 2007; Kulik and Soulier, 2020; Mikosch and Wintenberger, 2024; Resnick, 2024), is perhaps one of the lesser-known notions to the broader statistical community. Here we describe how it provides a very natural and powerful structure for quantifying the asymptotic calibration of combination tests.

2.1. Definition and a key lemma

Definition 2.1. A random vector $X = (X_j)_{j=1}^d$ is said to be multivariate regularly varying if there exists a positive function $b(t) \to \infty$, and a non-zero Borel measure μ on $\mathbb{R}^d \setminus \{0\}$ such that

$$b(t)\mathbb{P}[X \in t \cdot A] \longrightarrow \mu(A) \quad \text{as } t \to \infty$$
 (2.1)

for all Borel sets $A \subset \mathbb{R}^d \setminus \{0\}$ that are bounded away from 0 and $\mu(\partial A) = 0$. Here $\partial A = \overline{A} \setminus A^o$ is the boundary of A. In this case, we write $X \in RV(\mathbb{R}^d, \{b(t)\}, \mu)$.

The measure μ will be referred to as the *exponent* measure of X. It characterizes the asymptotic behavior of the *extremes* of X. For example, it allows us to completely characterize the asymptotic (in)dependence property of the components of the vector X. For simplicity, assume that the vector X is standardized to have asymptotically Pareto marginals as follows:

$$\mathbb{P}[X_i > t] \sim \frac{1}{t}, \text{ as } t \to \infty$$

Then the (upper) tail-dependence coefficient between X_i and X_j is:

$$\begin{split} \Lambda(X_i, X_j) &= \lim_{p \uparrow 1} \mathbb{P}[X_i > F_{X_i}^{\leftarrow}(p) \mid X_j > F_{X_j}^{\leftarrow}(p)] \\ &= \lim_{t \to \infty} t \mathbb{P}[X_i > t, X_j > t] = \lim_{t \to \infty} t \mathbb{P}[X/t \in A_i \cap A_j] = \mu(A_i \cap A_j), \end{split}$$

where $A_i = \{x : x_i > 1\}$. Thus μ is fundamentally related to $\Lambda(X_i, X_j)$, a quantity which characterizes the occurrence of joint (positive) extremes of X_i and X_j . For example, if $\Lambda(X_i, X_j) = 0$, the extremes do not occur simultaneously, and therefore X_i and X_j are said to be asymptotically (upper tail) independent.

Remark 2.1 (asymptotic independence implies MRV). As noted in Gui et al. (2025), it is well-known in the extreme value literature that, for heavy-tailed random vectors, bivariate asymptotic independence implies their multivariate regular variation. In this case, the exponent measure concentrates on the coordinate axes. The result dates back to Berman (1961) see e.g. Relation (8.100) in Beirlant et al. (2004). For an independent treatment see also Theorem A.1 in the Appendix.

Remark 2.2. In view of Remark 2.1, the dependency among p-values assumed in the combination test literature may be cast in the framework of MRV. The seminal paper by Liu and Xie (2020) establishes the asymptotic Type-I error control of the Cauchy Combination Test under the assumption that the p-values arise from a pairwise Gaussian copula. For calibration purposes, this assumption is equivalent to assuming an MRV copula with exponent measure μ concentrated on the axes. This has also been observed in the very recent work Gui et al. (2025).

In the rest of this section, we present a key technical lemma that allows us to establish the asymptotic calibration properties of any homogeneous combination test (Lemma 2.1.) This result relies on the angular (spectral) decomposition of the exponent measure (Theorem 2.2). We shall start, however, with a fundamental result on the general structure of the exponent measure of a regularly varying random vector. Its proof can be found in many comprehensive expositions in the literature (see e.g. Theorem 3.1 in Lindskog, Resnick and Roy, 2014). See also the monographs of Resnick (Resnick, 1987, 2007, 2024), the recent treatment (in Theorem 2.1.3 of Kulik and Soulier, 2020), and the many references therein.

Theorem 2.1 (The Tail Index Theorem). Let $X = (X_i)_{i=1}^d$ be a random vector in \mathbb{R}^d .

- (i) If $X \in RV(\mathbb{R}^d, \{b(t)\}, \mu)$, then:
 - (a) There exists a $\beta > 0$ referred to as the tail index of X such that $b(t) = \ell(t)t^{\beta}$, for some slowly varying function $\ell: (0, \infty) \to (0, \infty)$.
 - (b) The measure μ is β -homogeneous, i.e., for all t > 0, and all Borel sets A in \mathbb{R}^d that are bounded away from 0:

$$\mu(tA) = t^{-\beta}\mu(A) < \infty. \tag{2.2}$$

(c) If $X \in RV_{\gamma}(\mathbb{R}^d, \{c(t)\}, \nu)$, then

$$\beta = \gamma$$
, $\frac{b(t)}{c(t)} \to a > 0$, and $a\mu(A) = \nu(A)$.

Thus, the tail index β of a RV random vector X is unique.

(ii) Conversely, for every non-zero Borel measure μ on $\mathbb{R}^d \setminus \{0\}$ that satisfies (2.2) for some $\beta > 0$, there exists a random vector $X \in RV(\mathbb{R}^d, \{b(t)\}, \mu)$.

The tail index theorem allows us to write $X \in RV_{\beta}(\mathbb{R}^d, \{b(t)\}, \mu)$, since the tail index $\beta > 0$ is unique. Also, part 3 of Theorem 2.1 shows that the measure μ appearing in (2.1) is, up to rescaling, unique and in particular independent of the choice of the sequence $\{b(t)\}$.

There are several useful equivalent formulations of the notion of regular variation. Next, we recall the very powerful and intuitive polar coordinate perspective.

Theorem 2.2. We have that $X \in RV_{\beta}(\mathbb{R}^d, \{b(t)\}, \mu)$ if and only if for some (any) norm $\|\cdot\|$ in \mathbb{R}^d , the following two conditions hold:

1. For some slowly varying function L at ∞ :

$$\mathbb{P}[\|X\| > t] \sim L(t)t^{-\beta}, \quad t \to \infty.$$

2. As $t \to \infty$, we have

$$\frac{X}{\|X\|} \left| \{ \|X\| > t \} \stackrel{d}{\longrightarrow} \Theta, \right. \tag{2.3}$$

where Θ is a random vector taking values in the unit sphere $S_{\|\cdot\|} := \{x \in \mathbb{R}^d : \|x\| = 1\}$

Moreover, by adopting the polar coordinates $\Psi : \mathbb{R}^d \setminus \{0\} \to S_{\|\cdot\|} \times (0, \infty)$, where $\Psi(x) := (r(x), \theta(x))$, with $r(x) := \|x\|$ and $\theta(x) := x/\|x\|$, we have that

$$\mu \circ \Psi^{-1}(dr, d\theta) = c_{\mu} \beta r^{-\beta - 1} dr \sigma(d\theta),$$

where $c_{\mu} := \mu(B_{\|\cdot\|}(0,1)^c) = \mu(\{r > 1\})$ and σ is the probability distribution of the vector Θ in (2.3).

This result shows that the measure μ , when viewed in polar coordinates, factors into the product of a radial power-law type component and an angular component. Essentially it tells us that radially X behaves like a heavy-tailed random variable and when ||X|| is extreme, the distribution of the directions $X/\|X\|$ is asymptotically governed by σ . As a result σ is called the *angular* probability measure associated with μ . Sometimes, by analogy with the theory if infinite divisible laws, σ is referred to as the spectral measure of μ .

The power and convenience of the correspondence between μ and σ is illustrated by the following result. It yields an explicit expression for the asymptotic scale of any positive homogeneous function of a regularly varying random vector in terms of σ .

Lemma 2.1 (cf Proposition 2.5 in Jansen, Neblung and Stoev (2023)). Let $X \in RV_{\beta}(\mathbb{R}^d, \{b(t)\}, \mu)$ and let σ be the unique angular probability measure associated with the exponent measure μ . For every non-negative continuous 1-homogeneous function $h: \mathbb{R}^d \to \mathbb{R}_+$, we have

$$b(t)\mathbb{P}[h(X) > t] \to c_{\mu}\mathbb{E}[h(\Theta)^{\beta}], \quad as \ t \to \infty,$$

where $c_{\mu} = \mu\{r > 1\}$ and where $\Theta \sim \sigma$ is a random vector on $S_{\|\cdot\|}$ with probability distribution σ .

We end this introductory section with a conceptually important construction of multivariate regularly varying vectors X that can realize all possible asymptotic dependence structures. That is, this example furnishes a constructive proof of the converse claim (ii) in Theorem 2.1.

Example 2.1 (The generalized Breiman's lemma). Assume that Y and $W = (W_i)_{i=1}^d$ are independent dent random variable and vector, respectively. Assume that Y is non-negative and it has a heavy, regularly varying right tail, i.e.,

$$\mathbb{P}[Y > t] \sim L(t)t^{-\beta}$$
, as $t \to \infty$.

for some slowly varying function L.

If $\mathbb{E}[\|W\|^{\beta+\epsilon}] < \infty$, for some $\epsilon > 0$, then $X := (YW_i)_{i=1}^d$ is multivariate regularly varying with exponent β . See for example, Corollary 2.1.14 in Kulik and Soulier (2020). This is a very useful and conceptually fundamental extension of the celebrated result known as Breiman's lemma (see, e.g., Lemma 1.4.3 in Kulik and Soulier, 2020), which was originally formulated for d=1 and $\beta \in (0,1)$ (Proposition 2 in Breiman, 1965). Moreover, Relation (2.3) holds, where for all Borel sets on $S_{\parallel \cdot \parallel}$, we have

$$\mathbb{E}[1_A(\Theta)] = \frac{1}{\mathbb{E}[\|W\|^{\beta}]} \mathbb{E}\left[1_A\left(\frac{W}{\|W\|}\right) \|W\|^{\beta}\right]. \tag{2.4}$$

Notice that for an arbitrary probability measure σ on $S_{\|\cdot\|}$ by taking $W \sim \sigma$ and Y to be standard β-Pareto, i.e., $\mathbb{P}[Y > t] = 1/t^{\beta}$, $t \geq 1$, one obtains that the conditions of Theorem 2.2 trivially hold for X := YW. Indeed, ||X|| = Y, while $W = X/||X|| \stackrel{d}{=} \Theta \sim \sigma$ is independent of Y = ||X||. Thus, the Breiman construction yields instances of multivariate regularly varying random vectors for every possible choice of the asymptotic dependence structures and formally proving part (ii) of Theorem 2.1. See also Section 4, below.

2.2. A general approach to calibrating heavy-tailed combination tests

Let $(p_j)_{j=1}^d$ be a random vector with Uniform (0,1) marginal distributions. We view the p_j 's as p-values under certain null hypotheses. Consider a Pareto-type distribution F such that

$$\overline{F}(x) = 1 - F(x) \sim 1/x$$
, as $x \to \infty$

and transform the *p*-values as follows:

$$X_j := F^{\leftarrow}(1 - p_j), \quad j = 1, \cdots, d.$$

Then, given a vector of weights $w_j \ge 0$ such that $\sum_{j=1}^d w_j = 1$, consider the linear combination test statistic

$$T_w(X) := \sum_{j=1}^d w_j X_j.$$
 (2.5)

So small p-values correspond to large values of T_w . In the special case where F is the standard Cauchy distribution, this leads to the celebrated Cauchy Combination Test. On the other hand, when $w_j = 1/d$ and F is the standard Pareto, this recovers a transformation of the harmonic mean p-value test considered earlier by Wilson (2019) and dating back to Good (1958). In both cases, either under independence or asymptotic independence, it was shown that

$$\lim_{t \to \infty} \frac{\mathbb{P}[T_w(X) > t]}{\mathbb{P}(X_1 > t)} = 1. \tag{2.6}$$

It turns out that the bivariate dependence conditions in Liu and Xie (2020) (cf our Theorem A.1) imply that the vector $X = (X_j)_{j=1}^d$ is multivariate regularly varying with the same exponent measure as a vector of iid Cauchy variables. This is at the heart of the success of the Cauchy combination test. That is, as we shall see below, the test can be asymptotically calibrated in the same way as if the X_j 's were independent Cauchy.

The effectiveness of multivariate regular variation is in the fact that one can compute limits of the type in (2.6) very explicitly in terms of the exponent measure μ or the angular measure σ . This leads to a simple formula for the calibration of any linear combination test under general asymptotic dependence.

Proposition 2.1. Let
$$X = (X_j)_{j=1}^d \in RV_\beta(\mathbb{R}^d, \{b(t)\}, \mu)$$
 such that for all $j = 1, \dots, d$,
$$b(t)\mathbb{P}[X_j > t] \to c > 0, \quad as \ t \to \infty. \tag{2.7}$$

Let also Θ be a random vector on $S_{\|.\|}$ distributed according to the angular measure σ of μ .

Then, $\mathbb{E}[(\Theta_1)_+^{\beta}] = \cdots = \mathbb{E}[(\Theta_d)_+^{\beta}] > 0$, and for all $w_j \in \mathbb{R}$, $j = 1, \dots, d$ we have

$$\lim_{t \to \infty} \frac{\mathbb{P}[T_w(X) > t]}{\mathbb{P}[X_1 > t]} = \frac{1}{\mathbb{E}(\Theta_1)_+^{\beta}} \mathbb{E}\left(\sum_{j=1}^d w_j \Theta_j\right)_+^{\beta}.$$
 (2.8)

Proof. Consider the non-negative 1-homogeneous functions

$$h(x) = \left(\sum_{j=1}^{d} w_j x_j\right)_+, \text{ and } h_j(x) := (x_j)_+, j = 1, \dots, d.$$

Notice that for all t > 0, we have

$$\mathbb{P}[T(w) > t] = \mathbb{P}[h(X) > t], \text{ and } \mathbb{P}[X_j > t] = \mathbb{P}[h_j(X) > t].$$

Lemma 2.1 therefore implies that

$$b(t)\mathbb{P}[T(w) > t] \to c_{\mu}\mathbb{E}[h(\Theta)^{\beta}], \quad \text{and} \quad b(t)\mathbb{P}[X_i > t] \to c_{\mu}\mathbb{E}[h_i(\Theta)^{\beta}],$$
 (2.9)

as $t \to \infty$, where $c_{\mu} = \mu\{r > 1\}$.

Now, assumption (2.7) entails $\mathbb{E}[h_j(\Theta)^{\beta}] = \mathbb{E}[(\Theta_j)_+^{\beta}] = c_{\Theta} > 0$. Thus, taking the ratio of the two limits in (2.9), we obtain (2.8).

Proposition 2.1 leads to a complete characterization of the asymptotic calibration properties of the PCT and the CCT. In the following results, we shall assume that the tail exponent equals one:

$$\beta = 1.$$

Corollary 2.1 (universality of the PCT). Let P_i , $i = 1, \dots, d$ be Uniform(0, 1) distributed random variables and let $X_i := 1/(1 - P_i)$. Assume that $X = (X_i)_{i=1}^d \in RV_{\beta=1}(\mathbb{R}^d, \{b(t)\}, \mu)$. Then, for all $w_i \geq 0$ such that $\sum_{i=1}^d w_i = 1$, we have

$$\lim_{t \to \infty} \frac{\mathbb{P}[T_w(X) > t]}{\mathbb{P}[X_1 > t]} = 1,$$

where T_w is defined in (2.5).

Proof. We shall apply Proposition 2.1 with $\beta = 1$. Observe that the measure μ and hence the angular measure σ are supported on the positive orthant \mathbb{R}^d_+ . Therefore, the Θ_j 's in (2.8) are non-negative, and since $\beta = 1$, by linearity, we obtain

$$\lim_{t\to\infty} \frac{\mathbb{P}[T_w(X)>t]}{\mathbb{P}[X_1>t]} = \frac{1}{\mathbb{E}[\Theta_1]} \sum_{j=1}^d w_j \mathbb{E}[\Theta_j] = \sum_{j=1}^d w_j = 1,$$

since $\mathbb{E}[\Theta_1] = \cdots = \mathbb{E}[\Theta_j]$. This completes the proof.

Corollary 2.2 (universal honesty). Let P_i , $i=1,\dots,d$ be Uniform(0,1) distributed random variables and let $X_i := \tan\left(\pi\left(\frac{1}{2} - P_i\right)\right) \sim standard\ Cauchy$. Assume that $X = (X_i)_{i=1}^d \in RV_{\beta=1}(\mathbb{R}^d, \{b(t)\}, \mu)$. Then, for all $w_j \geq 0$ such that $\sum_{j=1}^d w_j = 1$

$$\lim_{t \to \infty} \frac{\mathbb{P}[T_w(X) > t]}{\mathbb{P}[X_1 > t]} \le 1. \tag{2.10}$$

If the w_j 's are all positive, then in (2.10) we have an equality if and only if the angular measure σ concentrates on $\mathbb{R}^d_+ \cup \mathbb{R}^d_+ = (-\infty, 0]^d \cup [0, \infty)^d$.

Proof. By Proposition 2.1, the following holds:

$$\lim_{t \to \infty} \frac{\mathbb{P}[T_w(X) > t]}{\mathbb{P}[X_1 > t]} = \frac{1}{\mathbb{E}(\Theta_1)_+} \mathbb{E}\left(\sum_{j=1}^d w_j \Theta_j\right)_+ \tag{2.11}$$

By the convexity of $x \mapsto x_+$ and Jensen's inequality,

$$\left(\sum_{j=1}^{d} w_j \Theta_j\right)_{+} \leq \sum_{j=1}^{d} w_j (\Theta_j)_{+}$$

$$\implies \mathbb{E}\left(\sum_{j=1}^{d} w_j \Theta_j\right)_{+} \leq \sum_{j=1}^{d} w_j \mathbb{E}(\Theta_j)_{+} = \left(\sum_{j=1}^{d} w_j\right) \mathbb{E}(\Theta_1)_{+} = \mathbb{E}\left((\Theta_1)_{+}\right)$$

Note that above we have used that

$$1 = \lim_{t \to \infty} b(t) \mathbb{P}(X_j > t) = \lim_{t \to \infty} b(t) \mathbb{P}((X_j)_+ > t) = c_{\mu} \mathbb{E}(\Theta_j)_+ \ \forall j$$

hence, $\mathbb{E}(\Theta_j)_+$ is a constant over j. Thus, by (2.11),

$$\lim_{t \to \infty} \frac{\mathbb{P}[T_w(X) > t]}{\mathbb{P}[X_1 > t]} = \frac{1}{\mathbb{E}(\Theta_1)_+} \mathbb{E}\left(\sum_{j=1}^d w_j \Theta_j\right)_+ \le 1$$
 (2.12)

Hence, the Cauchy combination test is guaranteed to be *conservative* or *honest*. Now, if supp $\sigma \subseteq \mathbb{R}^d_- \cup \mathbb{R}^d_+$,

$$(\Theta_j)_+ = 0 \ \forall \ j \ \text{or} \ (\Theta_j)_+ = \Theta_j \ \forall \ j, \quad \sigma - \text{a.s.}$$

In both the above cases,

$$\left(\sum_{i=1}^d w_i \Theta_i\right)_+ = 0 = \sum_{i=1}^d w_i (\Theta_i)_+ \text{ or } \left(\sum_{i=1}^d w_i \Theta_i\right)_+ = \sum_{i=1}^d w_i \Theta_i = \sum_{i=1}^d w_i (\Theta_i)_+, \quad \sigma - \text{a.s.}$$

Thus,

$$\mathbb{E}\left[\left(\sum_{i=1}^{d} w_i \Theta_i\right)_+\right] = \sum_{i=1}^{d} w_i \mathbb{E}[(\Theta_i)_+] = \mathbb{E}[(\Theta_1)_+]$$

By (2.11),

$$\lim_{t \to \infty} \frac{\mathbb{P}[T_w(X) > t]}{\mathbb{P}[X_1 > t]} = \frac{1}{\mathbb{E}(\Theta_1)_+} \mathbb{E}\left(\sum_{j=1}^d w_j \Theta_j\right)_+ = 1$$

and (asymptotic) calibration holds.

Now, for the converse to hold, one can easily see that Jensen's inequality used in proving honesty, needs to hold with equality almost surely, i.e.,,

$$\lim_{t \to \infty} \frac{\mathbb{P}[T_w(X) > t]}{\mathbb{P}[X_1 > t]} = \frac{1}{\mathbb{E}(\Theta_1)_+} \mathbb{E}\left(\sum_{j=1}^d w_j \Theta_j\right)_+ = \frac{1}{\mathbb{E}\left(\sum_{j=1}^d w_j (\Theta_j)_+\right)} \mathbb{E}\left(\sum_{j=1}^d w_j \Theta_j\right)_+ = 1$$

$$\implies \mathbb{E}\left(\left(\sum_{j=1}^d w_j \Theta_j\right)_+ - \sum_{j=1}^d w_j (\Theta_j)_+\right) = 0$$

$$\implies \left(\sum_{j=1}^d w_j \Theta_j\right)_+ = \sum_{j=1}^d w_j (\Theta_j)_+, \quad \sigma - \text{a.s.}$$

$$(2.13)$$

as the random variable inside the expectation is always non-negative due to Jensen's. This claim can be proved using the general result that if $\exists \{x_1, \ldots, x_n\}$ and $(w_i)_{i=1}^d \ni w_i > 0 \ \forall i$ and $\sum_i w_i = 1$ for which equality in Jensen's holds for a convex function f, then f must be affine over the convex hull of $\{x_i\}$. In our case, $f(x) = x_+$ is affine only in \mathbb{R}_+ and \mathbb{R}_- . Thus, equality in Jensen's implies $\operatorname{Conv}(\Theta_i : i = 1, \ldots, d) \subseteq \mathbb{R}_+ \cup \mathbb{R}_- \implies \Theta_i \in \mathbb{R}_+ \ \forall i$ or $\Theta_i \in \mathbb{R}_- \ \forall i$. However, for completeness, we include an elementary proof below.

Take any $\theta = (\theta_1, \dots, \theta_d)$. Let $\theta_k = \min_i \theta_i$ and $\theta_l = \max_i \theta_i > 0$ (assume). Then,

$$\sum_{j=1}^{d} w_j \theta_j = w^* \theta_k + (1 - w^*) \theta_l$$
where $w^* = \sum_{l=1}^{d} w_j \left(\frac{\theta_j - \theta_l}{\theta_k - \theta_l} \right) \in [0, 1]$

Now, since we assume $w_j > 0 \ \forall j, \ \exists \ \alpha^* \in (0,1] \ni$

$$\alpha^*(\theta_l)_+ = \sum_{j=1}^d w_j(\theta_j)_+$$
 (2.14)

Thus,

$$\left(\sum_{j=1}^{d} w_{j} \theta_{j}\right)_{+} = \sum_{j=1}^{d} w_{j}(\theta_{j})_{+}$$

$$\Rightarrow (w^{*} \theta_{k} + (1 - w^{*}) \theta_{l})_{+} = \alpha^{*}(\theta_{l})_{+} > 0$$

$$\Rightarrow \alpha = w^{*}(\theta_{k}/\theta_{l} - 1) + 1 = \sum_{j=1}^{d} w_{j}(\theta_{j} - \theta_{l})/\theta_{l} + 1 = \sum_{i=1}^{d} w_{j} \theta_{j}/\theta_{l}$$

$$\Rightarrow \sum_{j=1}^{d} w_{j}(\theta_{j})_{+}/\theta_{l} = \sum_{i=1}^{d} w_{j} \theta_{j}/\theta_{l}$$

$$\Rightarrow (\theta_{j})_{-} = 0, \forall j \ni \text{ or } \theta_{j} \geq 0, \ \forall j$$

$$(2.15)$$

As a result, if (2.15) holds, $\exists \theta_i > 0 \implies \theta \in \mathbb{R}^d_+$. Therefore,

$$\left(\sum_{j=1}^{d} w_{j} \theta_{j}\right)_{+} = \sum_{j=1}^{d} w_{j}(\theta_{j})_{+}$$

$$\implies \theta \in \mathbb{R}_{+}^{d} \cup \mathbb{R}_{-}^{d}$$
(2.16)

This means, (2.13) implies

$$\Theta \in \mathbb{R}^d_+ \cup \mathbb{R}^d_-, \quad \sigma - \text{a.s.}$$
 (2.17)

which proves the only if direction and hence completes the proof.

Remark 2.3. Corollary 2.1 establishes the *universal calibration* of the Pareto Combination Test with respect to all possible tail-dependence scenarios governed by multivariate regular variation. Since the decisions in the PCT are equivalent to those for the (weighted) harmonic means, we arrive at an important universality result for a family of widely used combination tests (Wilson, 2019; Good, 1958).

Remark 2.4. Corollary 2.2 shows that the Cauchy Combination Tests are always asymptotically honest in the multivariate regular variation dependence framework. Moreover, they are calibrated if only if the angular measure of X is supported on the set of vectors with common signs. This characterization result implies that in many popular tail-dependence models, e.g., the multivariate t-distribution, the CCT is *strictly conservative* (see also Section 4).

Remark 2.5. Proposition 2.1 is not new. The question on the limit behavior of sums of dependent heavy-tailed variables has been considered before by many authors especially in the context of financial or insurance risk. For example, the seminal work of Barbe, Fougères and Genest (2006) establishes similar formulae to (2.8). See also Theorem 4.1 in Embrechts, Lambrigger and Wüthrich (2009) and Yuen, Stoev and Cooley (2020) in the context of quantifying extreme Value-at-Risk.

Lemma 2.1 allows us to consider a much more general family of *homogeneous* test statistics. To see this, consider the following Corollary 2.3 which extends Proposition 2.1 but its proof is virtually the same.

Corollary 2.3. Let $h : \mathbb{R}^d \to \mathbb{R}_+$ be a continuous and 1-homogeneous function. Then, under the assumptions of Lemma 2.1, we have

$$\lim_{t \to \infty} \frac{\mathbb{P}[h(X) > t]}{\mathbb{P}[X_1 > t]} = \frac{1}{\mathbb{E}[(\Theta_1)_+^{\beta}]} \mathbb{E}[h(\Theta)^{\beta}].$$

Many classical functions used to combine p-values like min, max or power means are homogeneous. Hence this approach applies to numerous combination tests used in practice. We illustrate this with a small toy example with power means. For a more systematic account of examples and results, see Section 4.

Example 2.2 (Power Means). Say $X \in RV_{\beta}(\mathbb{R}^d_+, \{b(t)\}, \mu)$, the vector of marginally transformed p-values. Consider $h_{\gamma}(X)$, where

$$h_{\gamma}(x) = \left(\sum_{i=1}^{d} w_i x_i^{\gamma}\right)^{1/\gamma}$$

for $\gamma > 0$. Then, by Corollary 2.3,

$$\lim_{t \to \infty} \frac{\mathbb{P}[h_{\gamma}(X) > t]}{\mathbb{P}[X_1 > t]} = \frac{1}{\mathbb{E}[(\Theta_1)_+^{\beta}]} \mathbb{E}[h_{\gamma}(\Theta)^{\beta}] = \frac{1}{\mathbb{E}[(\Theta_1)^{\beta}]} \mathbb{E}\left[\left(\sum_{i=1}^d w_i \Theta_i^{\gamma}\right)^{\beta/\gamma}\right]$$

Now, if $\beta > \gamma$, the function $x \mapsto x^{\beta/\gamma}$ is *strictly* convex. So by Jensen's inequality,

$$\left(\sum_{i=1}^{d} w_i \Theta_i^{\gamma}\right)^{\beta/\gamma} \le \left(\sum_{i=1}^{d} w_i \Theta_i^{\beta}\right),\,$$

where the inequality is strict unless $w_i w_j \Theta_i \Theta_j = 0$, for all $1 \le i \ne j \le d$. Thus,

$$\lim_{t \to \infty} \frac{\mathbb{P}[h_{\gamma}(X) > t]}{\mathbb{P}[X_1 > t]} \le \frac{1}{\mathbb{E}[(\Theta_1)^{\beta}]} \mathbb{E}\left[\left(\sum_{i=1}^d w_i \Theta_i^{\beta}\right)\right] = 1.$$

Notice that, by strict convexity, the last inequality is strict unless $\Theta_i \Theta_j = 0$, $1 \le \forall i \ne j \le d$, a.s., which means that the measure σ concentrating on the axes. That is, the test is always conservative except in the case when the X_i 's are pairwise tail-independent.

So if the index of regular variation is larger than the order of the power mean, the h_{γ} combination test is asymptotically *conservative* in all but the asymptotically independent regimes of MRV, where it is calibrated. Similar calculations show that if the index is smaller, the test is *liberal*.

3. The universality of the Pareto-type linear combination test

In the previous section, we observed an intriguing universality phenomenon for the the linear Pareto-type combination tests. Namely, in Corollary 2.1 it was shown that the linear Pareto-type combination tests are asymptotically *exactly* calibrated for all possible asymptotic dependence structures. In this section, we will characterize this universality phenomenon over the general class of continuous homogeneous combination tests.

Definition 3.1 (universal calibration). Let $X \in RV_1(\mathbb{R}^d_+, \{t\}, \mu)$ be MRV with asymptotically standard 1-Pareto margins. Let also $h : \mathbb{R}^d \to \mathbb{R}_+$ be a 1-homogeneous continuous function. We say that the homogeneous combination test given by the function h is universally calibrated if

$$t\mathbb{P}[h(X) > t] \to 1$$
, as $t \to \infty$,

regardless of the dependence structure, i.e., the exponent measure μ .

Lemma 3.1. The h-combination test is asymptotically calibrated iff

$$d \cdot \mathbb{E}[h(\Theta)] = 1.$$

Proof. Let X be MRV with standard 1-Pareto marginals. Then, for every 1-homogeneous function, we know that

$$t\mathbb{P}[h(X) > t] \to c\mathbb{E}[h(\Theta)], \quad t \to \infty,$$

where $\Theta = (\Theta_i)_{i=1}^d$ is a random vector with probability distribution σ on the unit simplex

$$\Delta = \{(w_i)_{i=1}^d : w_i \ge 0, \sum_i w_i = 1\}.$$

Thus, the h-combination test is universally calibrated iff $c\mathbb{E}\left[h(\Theta)\right]=1, \ \forall \ \sigma \text{ on } \Delta$ Since the marginals are standardized, we have that

$$\mathbb{E}[\Theta_1] = \dots = \mathbb{E}[\Theta_d] = 1/d, \tag{3.1}$$

since $\mathbb{E}[\|\Theta\|_1] = \mathbb{E}[\Theta_1] + \cdots + \mathbb{E}[\Theta_d]$. This implies that

$$t\mathbb{P}[\Theta_i > t] \sim c \cdot (1/d) = 1, \Rightarrow c = d.$$

This proves the claim.

The following is the main theoretical contribution of our paper.

Theorem 3.1. Let $h:[0,\infty)^d \to [0,\infty)$ be a continuous 1-homogeneous function. The h-combination test is universally calibrated if and only if

$$h(x) = \sum_{i=1}^{d} w_i x_i, \quad x = (x_1, \dots, x_d)^{\top} \in \mathbb{R}^d$$

for some $w_i \geq 0$ such that $\sum_i w_i = 1$.

Proof. 'if' part: This direction follows from Corollary 2.1 almost immediately:

$$\lim_{t\to\infty} t \mathbb{P}[\sum_{i=1}^d w_i X_i > t] = \lim_{t\to\infty} \frac{\mathbb{P}[\sum_{i=1}^d w_i X_i > t]}{\mathbb{P}[X_1 > t]} = 1.$$

'only if' part: For this direction, assume that $h:[0,\infty)^d\to\mathbb{R}_+$ is a continuous homogeneous function such that

$$\mathbb{E}_{\sigma}[h(\Theta)] = 1/d,$$

for all probability measures σ on Δ such that relations (3.1) hold. We need to prove that $h(x) = \sum_{i=1}^{d} w_i x_i$ for some $w_i \geq 0$ such that $\sum_i w_i = 1$.

To this end, we shall apply the abstract Theorem B.1 with the space S given by the unit simplex:

$$S := \Delta := \{ (x_i)_{i=1}^d : x_i \ge 0, \sum_{i=1}^d x_i = 1 \}.$$

Let also $g_i: S \to [0,1]$ be the coordinate functions $g_i(x) := x_i, i = 1, \dots, d$ and define $\mathcal{G} = \{g_1, \dots, g_d\}$.

Thus, in the context of Theorem B.1, the probability measures that satisfy the calibration constraints in (3.1) are precisely all φ in the class

$$\mathcal{M}_{1/d}(\mathcal{G}) := \{ \varphi \in \mathcal{M}(\Delta) : (g, \varphi) = 1/d, \ \forall g \in \mathcal{G} \}.$$

Indeed, since $(g_i, \varphi) = 1/d$, we have $1 = \sum_{i=1}^d (g_i, \varphi) = (1, \varphi) = \varphi(\Delta)$, which implies that all $\varphi \in \mathcal{M}_{1/d}$ are probability measures.

Then Theorem B.1 applies as long as condition (B.2) holds, which in our case can be easily shown to hold. Indeed, taking $x_i = e_i$, the i-th unit vector in \mathbb{R}^d , the matrix G in (B.2) is equal to the identity matrix I_d , so trivially the condition is satisfied. Thus, by part (i) of Theorem B.1 or equation (B.3) $h(x) = \sum_{i=1}^d w_i g_i(x) = \sum_{i=1}^d w_i x_i$, where $\sum_{i=1}^d w_i = 1$.

Finally, observe that in our case \mathcal{G} satisfies the anti-dominance condition as proven in Remark B.1. This implies using part (ii) of Theorem B.1, that $w_i \geq 0, \forall i$. Hence,

$$h(x) = \sum_{i=1}^{d} w_i x_i$$
 for some $w_i \ge 0$ and $\sum_{i=1}^{d} w_i = 1$

which proves the claim.

Remark 3.1 (Open problem: Minimal universality dependence classes). Our Theorem 3.1 shows that the Pareto-type linear combination tests are the only homogeneous tests that are universally calibrated with respect to *all possible* tail-dependence structures parameterized by the angular measures. Naturally, one can ask if the Pareto-type linear combination tests remain the only universally

calibrated ones under narrower classes of tail-dependence structures. This is an interesting open problem, which may be possible to address using the abstract characterization result in Theorem B.1 by choosing a suitable measure space therein. Specifically, one may ask what is a minimal class of dependence structures, for which the universally calibrated tests will be of linear Pareto-type. This is an interesting question beyond the scope of the present paper.

Remark 3.2. Theorem 1 of Chen et al. (2023) offers interesting results that are in the spirit of our Theorem 3.1 and the open problem in Remark 3.1. These results are in a somewhat different setting where non-asymptotic calibration is considered.

4. Examples and numerical results

4.1. Multivariate regularly varying models

Multivariate regular variation is typically the rule rather than an exception for random vectors with heavy-tailed marginals. To make this intuition concrete, in this section we review several examples of popular models that exhibit multivariate regular variation. We are not aware of a simple, non-pathological, example of a heavy-tailed random vector that is not multivariate regularly varying.

Example 4.1 (multivariate t-distribution). Let $\nu > 0$ and $G \sim \text{Gamma}(\nu/2, 1/2)$ be a Gamma-distributed random variable with shape $\nu/2$ and rate parameter 1/2. Let also $W \sim \mathcal{N}(0, \Sigma)$ be an independent of G normal random vector in \mathbb{R}^d . Then the random vector

$$X = \frac{W}{\sqrt{G/\nu}}$$

is said to have the multivariate t-distribution with ν degrees of freedom and shape Σ .

Since $Y := (G/\nu)^{-1/2}$ is heavy-tailed with exponent ν , the multivariate t model is a particular instance of the Breiman construction. Specifically, Example 2.1 implies that $X = YW \in RV_{\nu}(\{b(t)\},\mu)$, where Relation (2.4) yields an expression of the corresponding angular distribution. Notice that unless W is concentrated on lower-dimensional subspace, the support of the angular measure σ in this case is the *entire unit sphere*. This, in view of Corollary 2.2, implies that the Cauchy Combination Test is asymptotically conservative, for all but the degenerate multivariate t models. In fact, the upper tail dependence coefficient of the t-copula, which is precisely $\lambda(X_i, X_j)$, can be written as:

$$\lambda(X_i, X_j) = 2T_{\nu+1} \left(-\sqrt{\frac{(\nu+1)(1-\rho_{i,j})}{(1+\rho_{i,j})}} \right), \tag{4.1}$$

where $\rho_{i,j} = \text{Corr}(W_i, W_j)$ and $T_{\nu+1}$ is the distribution function of the standard univariate t-distribution with $(\nu + 1)$ degrees of freedom (see, e.g., page 64 in Joe, 2015).

Example 4.2 (linear heavy-tailed factor models). Let $\beta > 0$ and Z_j , $j = 1, \dots, p$ be iid non-negative¹ random variables with Pareto-type tails. That is,

$$\mathbb{P}[Z_j > t] \sim \frac{1}{t^{\beta}}, \text{ as } t \to \infty.$$

¹The example extends to random variables with two-sided heavy tails, but the formula for the angular measure is slightly more involved.

Let also $A = (a_{i,j})_{d \times p}$ be an arbitrary matrix of constants with non-zero columns vectors $a_j := (a_{i,j})_{i=1}^d$, $j = 1, \dots, p$. Then, considering the column vector $Z := (Z_j)_{j=1}^p$, we have

$$X := AZ \in RV_{\beta}(\{b(t) = t^{\beta}\}, \mu),$$

where the angular measure σ associated with μ (recall Theorem 2.2) is such that

$$\sigma(A) = \frac{1}{\sum_{k=1}^{p} \|a_k\|^{\beta}} \sum_{j=1}^{p} \|a_j\|^{\beta} 1_A \left(\frac{a_j}{\|a_j\|}\right), \tag{4.2}$$

for all Borel sets A on $S_{\|\cdot\|}$ (see also Corollary 2.1.14 in Kulik and Soulier, 2020, for a more general result).

Remark 4.1 (the single large jump heuristic). Example 4.2 illustrates the single large jump heuristic for sums of independent heavy-tailed factors. Namely, the intuition behind it is that the vector $X = a_1 Z_1 + \cdots + a_p Z_p$ is extreme in norm (i.e., conditionally on ||X|| > t) when one and only one of the independent factors $a_j Z_j$ is extreme (in norm), as $t \to \infty$. Hence, the angular distribution of X/||X|| given ||X|| > t converges to a discrete measure with point-masses given by the directions $a_j/||a_j||$ and probability weights proportional to $||a_j||^{\beta}$.

Example 4.3 (max-linear heavy-tailed factor models). Let the Z_j 's and the matrix A be as in Example 4.2. Consider the model

$$X = \bigvee_{j=1}^{p} a_j Z_j =: A \otimes Z,$$

where \bigvee denotes component-wise maxima of the vectors $a_j Z_j$ and the a_j 's are the columns of the matrix A. Thus X is obtained by replacing the '+' operation in the definition of matrix multiplication by a maximum. Interestingly, the single large jump heuristic here entails that $X \in RV_{\beta}(\{t^{\beta}\}, \mu)$, where μ is the same as for the linear model in Example 4.2. Consequently, the corresponding angular measure associated with μ is given by (4.2).

The following two examples illustrate a small part of the rich landscape on the limit theorems for regularly varying random vectors. Specifically, if one considers centered and rescaled component-wise sums (or maxima, respectively), the corresponding limit random vectors will have sum-stable (or max-stable, respectively) distributions. Except in the Gaussian case, these sum-stable (max-stable, respectively) laws are multivariate regularly varying.

Example 4.4 (multivariate max-stable distributions). Fix $\beta > 0$ and let μ be an arbitrary non-zero Borel measure on \mathbb{R}^d , supported on $[0, \infty)^d \setminus \{0\}$ and such that

$$\mu(t \cdot A) = t^{-\beta}\mu(A) < \infty, \tag{4.3}$$

for all t > 0 and Borel $A \subset \mathbb{R}^d$ that are bounded away from 0. Then,

$$F(x) := \exp\{-\mu(\mathbb{R}^d_+ \setminus [0, x])\}, \quad x \in (0, \infty)^d$$
(4.4)

defines a valid cumulative distribution function of a random vector X, which is multivariate regularly varying (see e.g. Chapter 5 in Resnick, 1987). More precisely, we have $X \in RV_{\beta}(\{b(t) = t^{\beta}, \mu)$ and in fact, the random vector X is max-stable. That is, for all integer $n \geq 1$,

$$\bigvee_{i=1}^{n} X(i) \stackrel{d}{=} n^{1/\beta} X,$$

where the X(i)'s are independent copies of X and 'V' denotes the component-wise maximum operation.

The scaling property (4.3) implies that for any fixed norm $\|\cdot\|$ in \mathbb{R}^d , we have

$$F(x) = \mathbb{P}[X \le x] = \exp\Big\{-\int_{S_{\perp}} \Big(\max_{i=1,\dots,d} \frac{\theta_i}{x_i}\Big)^{\beta} H(d\theta)\Big\}, \quad x \in (0,\infty)^d,$$

where $S_+ := S_{\|\cdot\|} \cap [0, \infty)^d$ is the positive part of the unit sphere in the chosen norm $\|\cdot\|$. The angular measure σ associated with the exponent measure μ is a normalized version of H:

$$\sigma(A) = \frac{H(A)}{H(S_+)}, \quad A \subset S_+.$$

Upon centering and transformation of the marginal distributions, the above class of multivariate max-stable laws represent the entire class of *extreme value distributions*. That is, the distributions arising in the limit of centered and rescaled maxima of iid random vectors. For more details, see e.g. Resnick (1987); Beirlant et al. (2004); Resnick (2007).

Remark 4.2. The powerful Poisson random measure perspective (see e.g. Resnick, 1987, 2007) leads to a quick proof of the fact that Relation (4.4) yields a valid distribution function. Indeed, take $\Pi = \{\xi_i, i \in \mathbb{N}\}$ to be a Poisson point process on $\mathbb{R}^d_+ = [0, \infty)^d$ with mean measure μ and define

$$X := \bigvee_{i \in \mathbb{N}} \xi_i.$$

Then, for all $x \in (0, \infty)^d$, we have

$$\mathbb{P}[X \le x] = \mathbb{P}[\Pi([0, x]^c) = 0] = \exp\{-\mu([0, x]^c)\},\tag{4.5}$$

where the last equality follows from the fact that $\Pi(A) \sim \text{Poisson}(\mu(A))$, for every Borel set $A \subset \mathbb{R}^d_+$. This is precisely (4.4).

Notice that this argument does not depend on the scaling property (4.3). The general family of multivariate distributions as in (4.5) are known as max-infinitely divisible distributions and many of them can be multivariate regularly varying (see e.g. Chapter 5 in Resnick, 1987).

Example 4.5 (stable non-Gaussian distributions). Recall that a random vector X in \mathbb{R}^d is said to be sum-stable, if for all positive constants a', a'' there exist positive a and a vector $b \in \mathbb{R}^d$ such that

$$a'X' + a''X'' \stackrel{d}{=} aX + b,$$

where the X' and X'' are independent copies of X (Definition 2.1.1 on page 57 in Samorodnitsky and Taqqu, 1994).

We focus on the simple but rather rich family of *symmetric* stable non-Gaussian distributions. Fix an arbitrary norm $\|\cdot\|$ in \mathbb{R}^d . It is well-known, though not trivial to show, that every symmetric non-Gaussian sum-stable random vector X has a characteristic function of the form:

$$\mathbb{E}[e^{iX^{\top}u}] = \exp\Big\{-\int_{S_{\|\cdot\|}} |\langle u, \theta \rangle|^{\beta} \Gamma(du)\Big\}, \quad \text{where } 0 < \beta < 2$$
(4.6)

(see, e.g., Theorem 2.4.3 in Samorodnitsky and Taqqu, 1994), for some Γ – a finite symmetric measure on the unit sphere $S_{\|\cdot\|}$ in the chosen norm $\|\cdot\|$. (Note that Γ depends on the choice of

the norm.) Conversely, every finite symmetric measure Γ on S yields a characteristic function of an $S\beta S$ random vector X as above.

The case $\beta = 2$ yields a Gaussian random vector. Interestingly, when $0 < \beta < 2$, the S β S random vector X is multivariate regularly varying with exponent β and angular measure

$$\sigma(A) = \frac{\Gamma(A)}{\Gamma(S_{\|\cdot\|})}, \quad A \subset S_{\|\cdot\|}.$$

Specifically, Theorem 4.4.8 on page 197 in Samorodnitsky and Taqqu (1994) implies that $X \in RV_{\beta}(\{b(t) := t^{\beta}\}, \mu)$, where $\mu(B_{\|\cdot\|}(0, 1)^{c}) = C_{\beta}\Gamma(S_{\|\cdot\|})$ with

$$C_{\beta} = \begin{cases} \frac{1-\beta}{\Gamma(2-\beta)\cos(\pi\beta/2)} &, \ \beta \neq 1\\ 2/\pi &, \ \beta = 1 \end{cases}$$

(cf (1.2.9) on page 17 in Samorodnitsky and Taqqu, 1994).

Remark 4.3 (Aside on notation). Since α is reserved for the level of the Type I error here, we use β to denote the tail exponent. In the literature on non-Gaussian sum-stable distributions (see, e.g. Samorodnitsky and Taqqu, 1994), α stands for the tail-exponent (stability index), while β denotes the skewness parameter.

Remark 4.4. The multivariate Cauchy distribution is a special case of the symmetric 1-stable laws (Samorodnitsky and Taqqu, 1994). See also Example 4.7, below.

4.2. Analytical examples

Example 4.6 (Tippet combination test). Let $P_i \sim \text{Uniform}(0,1)$ and define

$$X_i := -\frac{1}{\log(1 - P_i)}, \quad i = 1, \dots, d.$$

Observe that the X_i 's are standard 1-Fréchet, i.e., $\mathbb{P}[X_i \leq x] = e^{-1/x}, \ x > 0$ and thus they are heavy-tailed:

$$\mathbb{P}[X_i > t] \sim \frac{1}{t}, \text{ as } t \to \infty.$$

Thus, the classical Tippet combination test, which rejects the global null whenever $\min_{i=1,\cdots,d} P_i < 1 - (1-\alpha)^{1/d}, \ \alpha \in (0,1)$ is equivalent to $\max_{i=1,\cdots,d} X_i > -d/\log(1-\alpha)$. Letting $t:=-1/\log(1-\alpha)$, we see that the Tippet test at level α rejects if and only if $h_T(X) > t$, where

$$h_T(X) := \frac{1}{d} \bigvee_{i=1}^d X_i.$$

Since the function h_T is homogeneous, by Lemma 2.1, we readily obtain

$$\lim_{t \to \infty} \frac{\mathbb{P}[h_T(X) > t]}{\mathbb{P}[X_1 > t]} = \frac{1}{\mathbb{E}\Theta_1} \mathbb{E}\left[\frac{1}{d} \bigvee_{i=1}^d \Theta_i\right] \le \frac{1}{\mathbb{E}\Theta_1} \frac{1}{d} \sum_{i=1}^d \mathbb{E}[\Theta_i] = 1,$$

where we used that since the X_i 's are positive, the multivariate regular variation focuses on the positive orthant \mathbb{R}^d_+ .

This argument shows that the Tippet combination test is always asymptotically honest. Notice that $\mathbb{E}[\max_{i=1,\dots,d}\Theta_i] < \mathbb{E}[\sum_{i=1}^d \Theta_i]$, unless $\Theta_i\Theta_j = 0$, $i \neq j$ almost surely. Thus, the Tippet combination test is asymptotically *strictly conservative* in all but the asymptotically independent regime of the *p*-values.

Recall Remark 4.3. The following example provides an alternative and analytically more convenient representation to the class of symmetric β -stable random vectors. Interestingly, when $\beta = 1$, we recover a rich family of models, for which the exact, non-asymptotic, calibration properties of the Cauchy combination test can be thoroughly understood.

For further details on non-Gaussian stable random vectors and processes, we refer the reader to the classical monograph of Samorodnitsky and Taqqu (1994). We will only review some basic notation and facts here.

Example 4.7 (Multivariate S1S laws).

Definition 4.1 (Symmetric β -stable (S β S)). Let $0 < \beta \le 2$. A random variable ξ is said to have a symmetric β -stable (S β S) distribution if

$$\varphi_{\xi}(t) = \mathbb{E}[e^{it\xi}] = e^{-\sigma_{\xi}^{\beta}|t|^{\beta}}, \quad t \in \mathbb{R},$$

for some scale coefficient $\sigma_{\xi} > 0$. We shall denote the scale coefficient σ_{ξ} of ξ as $\|\xi\|_{\beta}$. (Not to be confused with a norm.)

If $0 < \beta < 2$, we have that the S β S random variables are non-Gaussian and heavy-tailed in the sense that

$$\mathbb{P}[\xi > t] \sim c_{\beta} \frac{\|\xi\|_{\beta}^{\beta}}{t^{\beta}}, \quad \text{as } t \to \infty, \tag{4.7}$$

for some constant c_{β}

Definition 4.2 (Multivariate S β S). A random vector $X = (X_i)_{i=1}^d$ is said to be multivariate S β S (or just S β S) if for all $a_j \in \mathbb{R}$, we have that $\sum_{j=1}^d a_j X_j$ is S β S.

This definition is ultimately equivalent to the one discussed in Example 4.5 for the case of symmetric random vectors. The joint characteristic function of S β S random vectors given in (4.6), can be equivalently expressed using the following fact.

A random vector X is S β S if only if there exist $f_j \in L^{\beta}([0,1])$ such that

$$\varphi_X(t_1,\dots,t_d) = \mathbb{E}e^{i\sum_{j=1}^d t_j X_j} = \exp\left\{-\int_{[0,1]} \left|\sum_{j=1}^d t_j f_j(u)\right|^\beta du\right\}$$

for all $t_j \in \mathbb{R}$, $j = 1, \dots, d$. This means in particular that the scale coefficient of the S β S random variable $\xi := \sum_{j=1}^{d} t_j X_j$ equals

$$\left\| \sum_{j=1}^{d} t_j X_j \right\|_{\beta} = \left(\int_{[0,1]} \left| \sum_{j=1}^{n} t_j f_j(u) \right|^{\beta} du \right)^{1/\beta}$$
(4.8)

Conversely, every choice of $f_j \in L^{\beta}([0,1]), \ j=1,\cdots,d$ yields a joint characteristic function of an $S\beta S$ random vector as above.

As discussed in Example 4.5, all non-Gaussian S β S vectors MRV as well. Their angular measure can be expressed as:

$$\sigma(\cdot) = \frac{\int_0^1 \mathbf{1}[f(u)/||f|| \in \cdot] ||f(u)||^{\beta} du}{\int_0^1 ||f(u)||^{\beta} du},$$

where f(u) denotes the vector-valued function $(f_j(u))_{j=1}^d$, $u \in [0,1]$ and $\|\cdot\|$ is the corresponding norm associated with the angular measure.

In the case of $\beta = 1$, the sum-stability of S β S vectors allows one to directly express the calibration properties of the CCTs.

Multivariate 1-stable (generalized multivariate Cauchy distributions). For $\beta = 1$ (S1S), any linear combination is Cauchy. Assume the coordinates have unit scale,

$$||X_j||_1 = \int_0^1 |f_j(u)| du = 1, \qquad j = 1, \dots, d.$$

For weights $w_j \in \mathbb{R}$ with $\sum_{j=1}^d |w_j| = 1$, consider

$$T = \sum_{j=1}^{d} w_j X_j.$$

Then T is Cauchy with scale

$$||T||_1 = \int_0^1 \left| \sum_{j=1}^d w_j f_j(u) \right| du,$$

and, in view of (4.7), the tail ratio satisfies

$$\lim_{t \to \infty} \frac{\mathbb{P}(T > t)}{\mathbb{P}(X_1 > t)} = ||T||_1.$$

By convexity (triangle inequality),

$$||T||_1 \le \sum_{j=1}^d |w_j| \int_0^1 |f_j(u)| du = 1,$$

so rejecting for $T > F_{X_1}^{-1}(1-\alpha)$ yields an asymptotic type-I error $\leq \alpha$.

Exact calibration under spectral positivity. If the spectral functions are *spectrally positive*, i.e.

$$f_i(u)f_j(u) \ge 0$$
 for a.e. $u \in [0,1]$ and all $i, j,$

and the weights are nonnegative $(w_i \geq 0)$, then

$$||T||_1 = \int_0^1 \left| \sum_{j=1}^d w_j f_j(u) \right| du = \sum_{j=1}^d w_j \int_0^1 |f_j(u)| du = \sum_{j=1}^d w_j = 1.$$

Hence T is standard Cauchy, and for every level $\alpha \in (0,1)$,

$$\mathbb{P}(T > F_{X_1}^{-1}(1-\alpha)) = \alpha,$$
21

i.e. the Cauchy combination test is exactly calibrated at all levels. Spectral positivity of the functions implies that the exponent measure is supported on the positive and negative orthants. As a result, Corollary 2.2 already established asymptotic calibration for this copula. However, as we proved, calibration is not just asymptotic, but exact for this case.

Example 4.8 (sample splitting and max-linear combination tests). Data-splitting is a very commonly used method in modern statistics for testing and cross-validation. Below, we use max-stability of Frétchet distributions to demonstrate how, under independence, we can guarantee exact calibration at all levels when combining p-values from overlapping subsets. Moreover, we show that under MRV, we can guarantee asymptotic honesty, with calibration achieved for asymptotic independence.

Setup and screened p-values. Under the global null, let X_1, \ldots, X_n be i.i.d. with continuous cdf F. Fix subsets $I_1, \ldots, I_d \subset \{1, \ldots, n\}$ (screening blocks) and define the Sidak-screened p-values

$$p(I_j) := \operatorname{sidak}_{|I_j|} \left(\min_{i \in I_j} F(X_i) \right) = 1 - \left(1 - \min_{i \in I_j} F(X_i) \right)^{|I_j|}, \quad j = 1, \dots, d.$$

Here, each $p(I_j) \sim \text{Unif}(0,1)$ under H_0 due to independence of $X_i's$. But, we will soon see that even if we do not assume independence, FCT defined below (4.11) is calibrated asymptotically.

Let $\Psi:(0,1)\to(0,\infty)$ be the tail-quantile map

$$\Psi(p) = -\frac{1}{\log(1-p)},$$

so that $Z := \Psi(U)$ is standard 1-Fréchet with cdf $G(x) = e^{-1/x}, x > 0$. Define

$$Z_i := \Psi(F(X_i))$$

$$Y_j := \Psi(p(I_j)), \quad j = 1, \dots, d.$$
(4.9)

Then (Y_i) is a max-linear model driven by the 1-Fréchet factors (Z_i) :

$$Y = A \otimes Z := \bigvee_{j=1}^{n} a_j Z_j, \qquad a_{i,j} = \frac{1}{|I_i|} \mathbf{1}_{\{j \in I_i\}}, \quad i = 1, \dots, d,$$

where \vee denotes component-wise maximum and a_j is the jth column of A.

For nonnegative weights $w \in \mathbb{R}^d_+$ with $\sum_{j=1}^d w_j = 1$, set

$$Y_w := \bigvee_{j=1}^{d} w_j Y_j$$
, and,
 $c_w := \sum_{i=1}^{n} \max_{i=1,\dots,d} (w_i a_{i,j}).$ (4.10)

Using the above definitions, we introduce the Frétchet Combination Test (FCT) which rejects \mathcal{H}_0 at level $\alpha \in (0,1)$ when

$$\frac{Y_w}{c_w} > \Psi(\alpha) = -\frac{1}{\log(1-\alpha)}.$$
(4.11)

If the Z_i 's are independent (equivalently the X_i 's are i.i.d. under H_0), then

$$Y_w \stackrel{d}{=} c_w Z_1,$$

due to the max-stability of the Fréchet distribution. Hence for all $\alpha \in (0,1)$,

$$\mathbb{P}\left(\frac{Y_w}{c_w} > \Psi(\alpha)\right) = \alpha,$$

i.e., the FCT is exactly calibrated.

On the other hand, if the factors $Z = (Z_1, \ldots, Z_n)$ are multivariate regularly varying on $\mathbb{R}^n_+ \setminus \{0\}$ with index 1 and standard 1-Fréchet margins, we have the following result:

Proposition 4.1. Let $Z = (Z_j) \in RV(\mathbb{R}^n_+ \setminus \{0\}, \{b(t) = t\}, \nu)$. Consider the Y_i 's as in (4.9) and assume that for all $j \in \{1, \ldots, n\}$, there exists a k_j such that $j \in I_{k_j}$. Then, for all weights $w_i > 0$, with c_w as in (4.10), we have

$$\lim_{t \to \infty} t \mathbb{P}[Y_w/c_w > t] \le 1.$$

Moreover, in the last relation we have an equality if and only if Z_j , $j = 1, \dots, n$ are asymptotically independent.

For the proof, see Theorem C.1 in the appendix.

4.3. Simulations

4.3.1. Calibration Properties of PCT and CCT

As discussed in Corollaries 2.1 and 2.2, the Pareto combination test (PCT) is always (asymptotically) calibrated regardless of the nature of the exponent measure μ . However, the Cauchy combination test (CCT) is calibrated if and only if μ is supported on the positive or negative orthant. We demonstrate this behavior in the following simulations.

The setup. A statistic used in many methods in classical statistics is the t-statistic/ t-score. Hence, a natural joint structure to consider between these t-distributed random variables is the *multivariate-t*, which is also a MRV distribution.

Below, we simulate $T = (T_1, ..., T_{10}) \sim t_{\nu}(\boldsymbol{\mu} = \mathbf{0}, \Sigma)$ with $n = 10^6$ replications. We consider two types of examples for the shape parameter matrix Σ :

$$\Sigma = \Sigma_{\text{autoreg}} := (\rho_a^{|i-j|})_{d \times d}$$
 and $\Sigma = \Sigma_{\text{exch}} := (\rho_e^{\delta(i,j)})_{d \times d}$,

for some $\rho_a \in (-1,1)$ and $\rho_e \in (-1/(d-1),1)$, where $\delta(i,j)$ is the Kronecker symbol. Observe that for all $\nu > 0$, the components of the multivariate t random vector T are dependent and in fact tail-dependent even if Σ is a diagonal matrix (see Relation (4.1) in Example 4.1). The degree of tail-dependence, however, vanishes as $\nu \to \infty$, provided Σ is non-trivial, which is in line with the asymptotic independence of the non-degenerate multivariate normal vectors.

Assume that we are conducting a coordinate-wise 1 sided t-test: $\mathcal{H}_0: \boldsymbol{\mu} = 0$ and $\mathcal{H}_a: \boldsymbol{\mu} > 0$. First, the T_i are transformed to the corresponding p-values $P_i = 1 - F_t(T_i) \stackrel{\mathcal{H}_0}{\sim} U(0,1)$. These are then converted to Pareto scale or Cauchy scale using $X_i = F^{-1}(1-P_i)$ with the CDF F chosen accordingly. Finally, we consider the test statistic $T_w(X) = \sum_{i=1}^{10} w_i X_i$ with $w_i = 1/10 \ \forall i$. The degree of freedom ν was varied to demonstrate different degrees of tail dependence. Over the 10^6 simulations, the number of rejections were counted at level α , and then we plot the ratio

$$\frac{\widehat{\alpha}}{\alpha} := \frac{\text{Number of Rejections/10}^6}{\alpha}.$$
(4.12)

Figure 1 and 2 provide a first look at the calibration properties of the PCT and CCT for the multivariate t model for several values of the degrees of freedom parameter ν .

Calibration of PCT & CCT under Multivariate- $t_v(0, \Sigma)$ with autoreg Σ

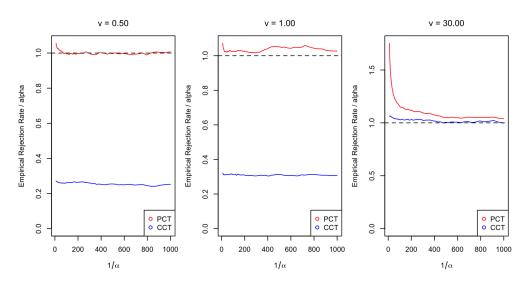


Fig 1: Calibration lineplot comparison of PCT and CCT

Observations. Our main interest is how the above mentioned ratio (4.12) behaves for small values of α . If the ratio is close to 1, the test is asymptotically *calibrated*, if it is significantly less than 1 it is asymptotically *conservative*, otherwise it is asymptotically *liberal*.

In the lineplot of Figure 1, we plot $\hat{\alpha}/\alpha$ as a function of $1/\alpha$. Ideally we would want this curve to go near 1 as $1/\alpha$ increases. We can see that PCT demonstrates this pattern for values of α as large as 0.01, whereas CCT is significantly conservative (the curve is significantly below y = 1 even for small values of α) for small ν .

This supports our Corollaries 2.1 and 2.2. CCT is never going to be truly calibrated- the angular measure σ always has non-zero mass outside $\mathbb{R}^d_+ \cup \mathbb{R}^d_-$. However, the mass decreases with increasing ν as discussed in the setup. This is why for higher values of ν CCT is (approximately) asymptotically calibrated. The curves for PCT do stay near the y=1 line, concurring with Corollary 2.1.

We used autoregressive Σ for our line plots, but the results are similar for any Σ . To illustrate this, the heatmaps we show demonstrate the differences between PCT and CCT again, but now we also use two different Σ - autoregressive and exchangeable. The heatmaps do have more values of ν to also demonstrate the effect of tail dependence as a function of degrees of freedom.

Ideally, we want the heatmaps to approach a lighter shade as we go down any column. But as we can see, CCT stays significantly conservative for low values of ν , similar to what we saw in Fig 1 (a blue shade implies $\hat{\alpha}/\alpha < 1$, hence the test is conservative). PCT on the other hand is (asymptotically) calibrated for all ν as we proved in Corollary 2.1. This trend is preserved

across different Σ . One of the fundamental reasons behind the conservativeness of CCT across Σ is that tail dependence is influenced more by the degrees of freedom ν than Σ . One can intuitively understand this from the way multivariate t is defined - $T = W/\sqrt{G/\nu}$ where $W \sim \mathcal{N}(\mathbf{0}, \Sigma)$ and G is independently distributed to W with $G \sim \operatorname{Gamma}(\nu/2, 1/2)$. While Σ does govern the dependence among the components of W, for upper tail dependence, the fact that we are dividing by a nonnegative random variable G plays a bigger role as G takes smaller values with high probability as ν decreases.

Calibration Heatmaps for Multivariate- $t_v(0, \Sigma)$ with autoreg Σ

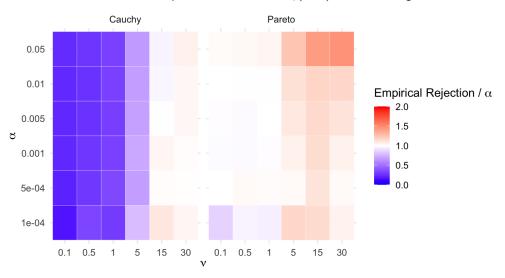


Fig 2: Calibration heatmap comparison of PCT and CCT with autoregressive Σ

Calibration Heatmaps for Multivariate- $t_v(0, \Sigma)$ with exch Σ

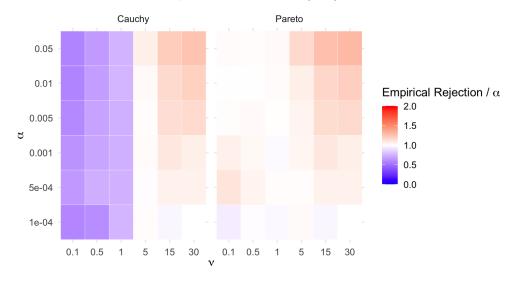


Fig 3: Calibration heatmap comparison of PCT and CCT with exchangeable Σ

4.3.2. Power Analysis

In this subsection, we demonstrate how the power of PCT and CCT relate to each other and can vary drastically over different kinds of alternative hypothesis. While we do not provide a theoretical result for the power of PCT, in these simulations, we can see that PCT performs as good as CCT in all cases. Hence, PCT is the best choice among existing combination tests taking both (asymptotic) calibration and power into account. These findings are supported by the findings in Gui et al. (2025), who provide a theoretical analysis of power under certain assumptions on the alternative.

The setup. We work with the multivariate-t copula here and test for the location parameter. Formally, $(T_1, \ldots, T_d) \sim t_{\nu}(\boldsymbol{\mu}, \Sigma)$ and we test the hypotheses

$$\mathcal{H}_0: \boldsymbol{\mu} = 0$$
 and $\mathcal{H}_a: \boldsymbol{\mu} > 0$

We use $\Sigma = \Sigma_{\text{autoreg}}$ for our simulations and use two different types of μ to illustrate two cases from the alternative. Figure 4 uses

$$\mu_1 = (\text{effect size}) * v_{\text{bottom}}$$

where $m{v}_{
m bottom}$ is the unit eigenvector of Σ^{-1} for the smallest eigenvalue. Figure 5 uses

$$\boldsymbol{\mu}_2 = (\text{effect size}) * \boldsymbol{v}_{\text{top}}$$

where v_{top} is the unit eigenvector of Σ^{-1} for the largest eigenvalue.

We vary both ν and the dimension d in our simulations to study their effect. Each plot considers different values of effect size (ranging from 0 to 40) to study power as signal strength increases. For each value of the effect size, we simulate $n=10^6$ replicates of $(T_1,\ldots,T_d)\sim t_\nu(\boldsymbol{\mu}_i,\Sigma)$ and convert them to p-values $P_i=1-F_t(X_i)$. Then they are transformed to Pareto or Cauchy using $X_i=F^{-1}(1-P_i)$ with F chosen accordingly. Finally, we consider the statistic $T_w(X)=\sum_{i=1}^d w_i X_i$ with $w_i=1/d$ and, for a test level= α , we reject when $T_w(X)>(1-F)^{-1}(\alpha)$. For these power simulations we choose $\alpha=0.05$.

The power is demonstrated relatively - we illustrate power of PCT and CCT relative to the Likelihood Ratio Test (LRT). LRT is selected as a baseline since it is classically the UMP test. Essentially, we calculate

$$\hat{\beta}_{\mathrm{T}} = \frac{\text{Number of Rejections made by T}}{n}$$

where $T \in \{PCT, CCT, LRT\}$ and plot

$$\frac{\hat{\beta}_{PCT}}{\hat{\beta}_{LRT}}$$
 and $\frac{\hat{\beta}_{CCT}}{\hat{\beta}_{LRT}}$

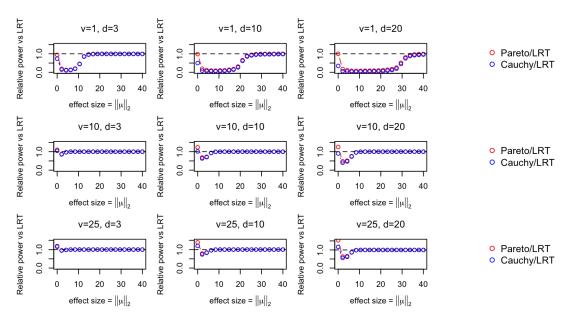


Fig 4: Relative Power of PCT and CCT with $\mu = \mu_1$

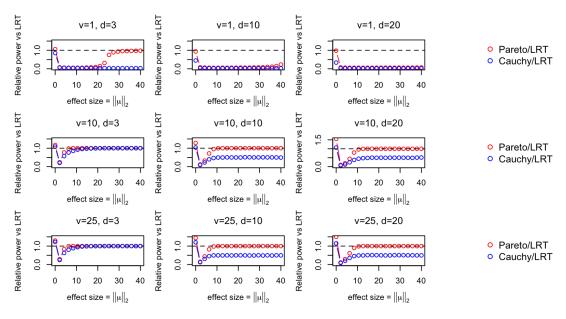


Fig 5: Relative Power of PCT and CCT with $\mu = \mu_2$

Observations. At first glance, we can see that the relative power of PCT is always better or similar to that of CCT. In all plots there is an initial "dip", which suggests that for smaller effect sizes, PCT and CCT struggle with detecting the signal.

However, PCT and CCT both catch up with LRT for higher signal magnitude in Figure 4. In fact, in the first case, the depth of the dip decreases with increasing ν and becomes practically insignificant in low dimensions. This suggests that a lower degree of tail dependence (i.e., high ν) helps both tests to perform better in detecting the signal. With respect to the effect of dimensionality, detection becomes harder in higher dimensions. This is demonstrated by the depth of the dip and the higher effect size needed for PCT and CCT to catch up to LRT.

Figure 5 shows similar trends in terms of effect of dimensionality. However, $\mu = \mu_2$ seems to negatively impact the detection power of the tests quite a bit. PCT performs better than CCT certainly, but in higher dimensions with strong tail dependence, both tests need significantly higher effect size to catch up to LRT. While the depth of the dip in power doesn't show significant change for both the tests, PCT recovers its detection strength more easily than CCT. In cases like $(\nu, d) \in \{(10, 10), (10, 20), (25, 10), (25, 20)\}$, CCT seems to stagnate and never reach LRT's power.

These two figures illustrate that PCT has better power than CCT, but the power depends on the alternative distribution very strongly.

Acknowledgments

We thank Ruodu Wang for an insightful discussion, which brought the result of Berman (1961) to our attention. FRG was supported in part by NSF Grant DMS-2515385. SS is partially supported by the NSF grant CNS/CSE-2319592 "Collaborative Research: IMR: MM-1A: Scalable Statistical Methodology for Performance Monitoring, Anomaly Identification, and Mapping Network Accessibility from Active Measurements".

References

- BARBE, P., FOUGÈRES, A.-L. and GENEST, C. (2006). On the tail behavior of sums of dependent risks. *Astin Bull.* **36** 361–373. MR2312671
- BASRAK, B., MILINČEVIĆ, N. and MOLCHANOV, I. (2025). Foundations of regular variation on topological spaces.
- BASRAK, B. and PLANINIĆ, H. (2019). A note on vague convergence of measures. *Statist. Probab. Lett.* **153** 180–186. MR3979308
- Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004). Statistics of Extremes: Theory and Applications. Wiley, Chichester.
- BERMAN, S. M. (1961). Convergence to Bivariate Limiting Extreme Value Distributions. *Annals of Mathematical Statistics* **32** 733–743.
- BILLINGSLEY, P. (1999). Convergence of probability measures, second ed. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York A Wiley-Interscience Publication. MR1700749
- Breiman, L. (1965). On some limit theorems similar to the arc-sin law. Theory of Probability and its Applications 10 323-331.
- CHEN, Y., LIU, P., TAN, K. S. and WANG, R. (2023). Trade-off between validity and efficiency of merging p-values under arbitrary dependence. *Statistica Sinica* **33** 851–872.
- DE HAAN, L. and FERREIRA, A. (2006). Extreme value theory. Springer Series in Operations Research and Financial Engineering. Springer, New York. An introduction. MR2234156
- DICICCIO, C. J., DICICCIO, T. J. and ROMANO, J. P. (2020). Exact tests via multiple data splitting. Statistics & Probability Letters 166 108865.

- Embrechts, P., Lambrigger, D. D. and Wüthrich, M. V. (2009). Multivariate extremes and the aggregation of dependent risks: examples and counter-examples. *Extremes* 12 107–127. MR2515643
- FANG, Y., CHANG, C., PARK, Y. and TSENG, G. C. (2023). Heavy-tailed distribution for combining dependent p-values with asymptotic robustness. *Statistica Sinica* **33** 1115–1142.
- FISHER, R. A. (1948). Combining independent tests of significance. American Statistician 2 30.
- Good, I. J. (1958). Significance tests in parallel and in series. *Journal of the American Statistical Association* **53** 799–813.
- Gui, L., Jiang, Y. and Wang, J. (2025). Aggregating dependent signals with heavy-tailed combination tests. *Biometrika* asaf038.
- Gui, L., Mao, T., Wang, J. and Wang, R. (2025). Validity and Power of Heavy-Tailed Combination Tests under Asymptotic Dependence.
- Guo, F. R. and Shah, R. D. (2025). Rank-transformed subsampling: inference for multiple data splitting and exchangeable p-values. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 87 256–286.
- Hult, H. and Lindskog, F. (2006). Regular variation for measures on metric spaces. *Publ. Inst. Math. (Beograd) (N.S.)* **80(94)** 121–140. MR2281910 (2008g:28016)
- Janssen, A., Neblung, S. and Stoev, S. (2023). Tail-dependence, exceedance sets, and metric embeddings. *Extremes*.
- JOE, H. (2015). Dependence Modeling with Copulas. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press, Boca Raton, FL.
- Kulik, R. and Soulier, P. (2020). Heavy-tailed time series. Springer Series in Operations Research and Financial Engineering. Springer, New York. MR4174389
- LANCASTER, H. O. (1961). The Combination of Probabilities: An Application of Orthonomal Functions. Australian Journal of Statistics 3 20–33.
- LINDSKOG, F., RESNICK, S. I. and ROY, J. (2014). Regularly varying measures on metric spaces: hidden regular variation and hidden jumps. *Probab. Surv.* 11 270–314. MR3271332
- Liu, Y. and Xie, J. (2020). Cauchy Combination Test: A Powerful Test with Analytic p-Value Calculation under Arbitrary Dependency Structures. *Journal of the American Statistical Association* 115 393–402.
- Liu, Y., Chen, S., Li, B., Zhang, K., Wang, K. and Lin, X. (2019). ACAT: A Fast and Powerful p-Value Combination Method for Rare-Variant Analysis in Sequencing Studies. *American Journal of Human Genetics* **104** 410–421.
- MEERSCHAERT, M. M. (1984). Multivariate Domains of Attraction and Regular Variation, PhD thesis, University of Michigan, Ann Arbor.
- Meinguet, T. and Segers, J. (2010). Regularly varying time series in Banach spaces.
- Meinshausen, N. and Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **72** 417–473.
- Meng, X.-L. (1994). Posterior Predictive p-Values. The Annals of Statistics 22 1142 1160.
- MIKOSCH, T. and WINTENBERGER, O. (2024). Extreme value theory for time series—models with power-law tails. Springer Series in Operations Research and Financial Engineering. Springer, Cham. MR4823721
- RESNICK, S. I. (1987). Extreme Values, Regular Variation and Point Processes. Springer-Verlag, New York.
- RESNICK, S. I. (2007). Heavy-tail phenomena. Springer Series in Operations Research and Financial Engineering. Springer, New York. Probabilistic and statistical modeling. MR2271424

- RESNICK, S. I. (2024). The art of finding hidden risks. Springer, New York. Hidden Regular Variation in the 21st Century.
- Samorodnitsky, G. and Taqqu, M. S. (1994). Stable Non-Gaussian Processes: Stochastic Models with Infinite Variance. Chapman and Hall, New York, London.
- Sibuya, M. (1960). Bivariate extreme statistics. I. Ann. Inst. Statist. Math. Tokyo 11 195–210. MR115241
- SIMES, R. J. (1986). An Improved Bonferroni Procedure for Multiple Tests of Significance. *Biometrika* **73** 751–754.
- SINGH, K., XIE, M. and STRAWDERMAN, W. E. (2005). Combining information from independent sources through confidence distributions.
- TIPPETT, L. H. C. (1937). The methods of statistics-an introduction mainly for experimentalists.
- Vovk, V. and Wang, R. (2020). Combining p-values via averaging. *Biometrika* 107 791–808.
- Vovk, V. and Wang, R. (2021). E-values: Calibration, combination and applications. *The Annals of Statistics* **49** 1736–1754.
- Wilson, D. J. (2019). The harmonic mean p-value for combining dependent tests. *Proceedings of the National Academy of Sciences* **116** 1195–1200.
- Wu, M. C., Kraft, P., Epstein, M. P., Taylor, D. M., Chanock, S. J., Hunter, D. J. and Lin, X. (2010). Powerful SNP-set analysis for case-control genome-wide association studies. *The American Journal of Human Genetics* 86 929–942.
- YUEN, R., STOEV, S. and COOLEY, D. (2020). Distributionally robust inference for extreme Value-at-Risk. *Insurance Math. Econom.* **92** 70–89. MR4079575

Appendix A: The M_0 -topology and multivariate regular variation

This material offers a review of the main ideas on the fundamental concepts of multivariate regular variation for measures. This notion was formally introduced by Mark Marvin Meerschaert (Meerschaert, 1984) though its inception and spirit dates back to the works of William Feller and Jovan Karamata in the middle of the 20th century.

Resnick Resnick (1987) has shaped, developed, and taught many of us about *multivariate regular* variation and its fundamental role in extremes from the perspective of point processes. A series of books and the references therein offer deep probabilistic, and statistical insights to the theory and its applications (de Haan and Ferreira, 2006; Resnick, 2024).

Multivariate regular variation has recently been lifted to the realm of infinite-dimensional metric spaces where topological issues arising from the lack of compactness of bounded sets have been resolved (see e.g. Hult and Lindskog, 2006; Meinguet and Segers, 2010; Basrak and Planinić, 2019). The theory has flourished and matured to enjoy a vast array of applications (see e.g., Kulik and Soulier, 2020; Mikosch and Wintenberger, 2024, and the references therein). Last but not least, the recent work of Basrak, Milinčević and Molchanov (2025) has unified and extended the concept of multivariate regular variation to Suslin spaces.

Multivariate regular variation is by now a polished tool which is equally useful in building monumental projects as well as carving out small intricate frescoes on their facades. It should be considered as indispensable in mainstream statistics as the law of large numbers and the central limit theorem.

A.1. The space M_0

In this section, we follow closely the seminal paper of Hult and Lindskog Hult and Lindskog (2006). Although our focus is on finite-dimensional Euclidean spaces, we adopt the modern language and the M₀-convergence perspective. Thus, *mutatis mutandis*, all results in this section extend to random elements in complete separable metric spaces equipped with a continuous scaling action (Hult and Lindskog, 2006). Extensive expositions can be found in the books Resnick (2007); Kulik and Soulier (2020). The bleeding edge of generality and abstraction is given in Basrak, Milinčević and Molchanov (2025).

Consider the Euclidean space \mathbb{R}^d . Excise its origin $\mathbb{R}_0^d := \mathbb{R}^d \setminus \{0\}$ and equip it with the induced topology. Let $\mathcal{B}_0 := \mathcal{B}(\mathbb{R}_0^d)$ be the Borel σ -field generated by all open sets in \mathbb{R}_0^d .

Let $B_r(x) := \{y \in \mathbb{R}^d : ||x - y|| < r\}$ denote the open ball in \mathbb{R}^d with center x and radius r > 0. For a set $A \subset \mathbb{R}^d$, we write \overline{A} and A° for the closure, interior and let $\partial A := \overline{A} \setminus A^\circ$ be the boundary of A, respectively. We shall say that a set $A \subset \mathbb{R}^d$ is bounded away from the origin (BAFO), if for some $\epsilon > 0$, we have $B_{\epsilon}(0) \cap A = \emptyset$. That is, the BAFO sets are a positive distance away from 0.

Definition A.1 (The \mathbb{M}_0 space and \mathbb{M}_0 -convergence). (i) A measure μ on $(\mathbb{R}_0^d, \mathcal{B}_0)$ is said to be boundedly finite if $\mu(A) < \infty$, for all BAFO Borel sets. Let $\mathbb{M}_0 := \mathbb{M}_0(\mathbb{R}^d)$ denote the collection of all such measures.

(ii) For $\mu, \mu_n \in \mathbb{M}_0$, $n \in \mathbb{N}$, we write $\mu_n \to^{\mathbb{M}_0} \mu$ and say μ_n converges to μ , in the \mathbb{M}_0 -topology, if for all BAFO Borel sets A with $\mu(\partial A) = 0$,

$$\mu_n(A) \longrightarrow \mu(A)$$
, as $n \to \infty$,

where $\partial A := \overline{A} \setminus A^{\circ}$ denotes the boundary of the set A.

Conceptually, it is useful to view the \mathbb{M}_0 -convergence as a type of weak convergence. Let \mathcal{C}_0 denote the class of all bounded and continuous functions $f: \mathbb{R}^d \to \mathbb{R}$ which vanish in a neighborhood of 0. That is, such that f(x) = 0, for all $x \in B_{\epsilon}(0)$ for some $\epsilon > 0$, which means that $\{|f| > 0\}$ is a BAFO set.

Proposition A.1 (Theorem 2.1 in Hult and Lindskog (2006)). We have that $\mu_n \to^{\mathbb{M}_0} \mu$ if and only if $\int_{\mathbb{R}^d} f d\mu_n \to \int_{\mathbb{R}^d} f d\mu$, as $n \to \infty$, for all $f \in \mathcal{C}_0$.

The notion of \mathbb{M}_0 -convergence of sequences of measures can be used to define closed sets in \mathbb{M}_0 and hence a topology on \mathbb{M}_0 . It can be shown that this topology is in fact metrizable. Recall first, that for two *finite* Borel measures μ and ν on \mathbb{R}^d , the Lévy-Prokhorov metric, is:

$$\pi(\mu,\nu) := \inf \Big\{ \epsilon > 0 : \sup_{A \in \mathcal{B}_0} (\mu(A) - \nu(A_{\epsilon})) \lor (\nu(A) - \mu(A_{\epsilon})) \le \epsilon \Big\},\,$$

where $A_{\epsilon} := \bigcup_{x \in A} B_{\epsilon}(x)$ is the ϵ -neighborhood of A and $x \vee y := \max\{x, y\}$.

Following Hult and Lindskog (2006), for every r > 0 and a boundedly finite measure $\mu \in \mathbb{M}_0$, define $\mu^{(r)}$ as the restriction of μ to $B_r(0)^c := \mathbb{R}^d \setminus B_r(0)$. Namely, $\mu^{(r)}$ is the finite measure

$$\mu^{(r)}(A) := \mu(A \setminus B_r(0)), \quad A \in \mathcal{B}_0.$$

Now, for every two boundedly finite measures $\mu, \nu \in \mathbb{M}_0$, define

$$d_{\mathbb{M}_0}(\mu,\nu) := \int_0^\infty e^{-r} \frac{\pi(\mu^{(r)},\nu^{(r)})}{1 + \pi(\mu^{(r)},\nu^{(r)})} dr. \tag{A.1}$$

Proposition A.2 (cf. Theorems 2.3 and 2.4 in Hult and Lindskog (2006)). The functional $d_{\mathbb{M}_0}$ in (A.1) is a metric on \mathbb{M}_0 and $(\mathbb{M}_0, d_{\mathbb{M}_0})$ is a complete separable metric space. Moreover, $\mu_n \to^{\mathbb{M}_0} \mu$ if and only if $d_{\mathbb{M}_0}(\mu_n, \mu) \to 0$, as $n \to \infty$.

For a Portmanteau theorem with equivalent characterizations of the \mathbb{M}_0 -convergence, see Theorem 2.4 in Hult and Lindskog (2006). We conclude this brief review with a characterization of the important notion of relative compactness, which is also reproduced from Hult and Lindskog (2006). Recall that a set of measures $M \subset \mathbb{M}_0$ is said to be relatively compact if its closure is compact. Equivalently, an infinite subset M of a metric space \mathbb{M}_0 is relatively compact if and only if every infinite sequence $\{\mu_n\} \subset M$ has a converging infinite subsequence $\{\mu_{n_k}\}$, whose limit is in \mathbb{M}_0 though not necessarily in M.

Proposition A.3 (Theorem 2.7 in Hult and Lindskog (2006)). A set of measures $M \subset \mathbb{M}_0$ is relatively compact in $(\mathbb{M}_0, d_{\mathbb{M}_0})$ if and only if for some $r_n \downarrow 0$, the following two conditions hold:

1. For all $n \in \mathbb{N}$, we have

$$\sup_{\mu \in M} \mu \Big(\mathbb{R}^d \setminus B_{r_n}(0) \Big) < \infty \tag{A.2}$$

2. For every $\epsilon > 0$, there exist compact sets $C_n \subset \mathbb{R}^d \setminus B_{r_n}(0)$, such that

$$\sup_{\mu \in M} \mu \Big(\mathbb{R}^d \setminus (C_n \cup B_{r_n}(0)) \Big) < \epsilon. \tag{A.3}$$

The necessity of this characterization of relative compactness essentially follows from Proposition A.2 and Prokhorov's characterization of relative compactness for finite measures on complete separable metric spaces Billingsley (1999). The sufficiency is a consequence of Theorem 2.2 in Hult and Lindskog (2006) and yet again Prokhorov's criterion.

A.2. Relative compactness of tail-measures

In this section, we establish a result of independent interest. It shows that the tail-measures of a random vector with regularly varying marginals are relatively compact in the M_0 -topology. As a consequence, this allows us to recover the well-known fact that asymptotic bivariate independence implies multivariate regular variation dating back to Berman (1961) (cf (8.100) in Beirlant et al. (2004)).

Proposition A.4. Let $X = (X_i)_{i=1}^d$ be a random vector. Assume that the marginals of X have regularly varying distributions. Specifically, suppose that for all x > 0 and $i \in [d]$, we have

$$b(t)\mathbb{P}[\pm X_i > tx] \to c_{\pm}x^{-1}, \quad as \ t \to \infty,$$
 (A.4)

where $c_{\pm} \geq 0$ and $c_{+} + c_{-} = 1$, for some monotone non-decreasing function such that $b(t) \rightarrow \infty$. Define the rescaled tail-measures

$$\mu_t(\cdot) := b(t)\mathbb{P}[X/t \in \cdot], \quad t > 1$$

on $(\mathbb{R}_0^d, \mathcal{B}_0)$ and observe that $\mu_t \in \mathbb{M}_0$. Then:

- (i) We have that $b(t) \sim L(t)t$, as $t \to \infty$ for some slowly varying function $L(\cdot)$.
- (ii) The set of rescaled tail-measures $\{\mu_t, t > 1\}$ is relatively compact in the \mathbb{M}_0 -topology. In particular, for every $t_n \to \infty$, there is a measure $\mu \in \mathbb{M}_0(\mathbb{R}^d)$ and a further integer sequence $n_k \to \infty$ such that

$$\mu_{t_{n_k}} \xrightarrow{\mathbb{M}_0} \mu, \quad as \ n_k \to \infty.$$

Proof. If $t_n \not\to \infty$, then one can choose a convergent monotone subsequence. Without loss of generality assume the subsequence is increasing, i.e., $t_{n_k} \uparrow \tau < \infty$. By the monotonicity of b one readily has $\mu_{t_{n_k}} \to^{\mathbb{M}_0} \mu$, as $n_k \to \infty$, for some non-zero μ . Indeed, in this case $b(t_{n_k}) \to b(\tau-)$, and we have $\mu = \mu_{\tau-} := b(\tau-)\mathbb{P}[X/\tau \in \cdot]$. (If t_{n_k} is decreasing, replace $b(\tau-)$ with $b(\tau+)$) The interesting case is when $t_n \to \infty$.

For this case, we use the analogous tightness criteria for boundedly finite measures (Proposition A.3). Note that, for every x > 0, by (A.4), with $A_i := \{u \in \mathbb{R}^d, : |u_i| > 1\}$, we have that

$$\mu_t(x \cdot A_i) = b(t)\mathbb{P}[X/t \in x \cdot A_i] = b(t)\mathbb{P}[|X_i| > xt] \to x^{-1}, \text{ as } t \to \infty.$$

Take any $r_n \downarrow 0$. Then for all $n, \frac{r_n}{d} \bigcap_{i=1}^d A_i^c = \{u \in \mathbb{R}^d : |u_i| \leq r_n/d \ \forall \ i\} \subseteq B_{r_n}(0) \implies \mu_t \Big(\mathbb{R}^d \setminus B_{r_n}(0)\Big) \leq \mu_t \Big(\bigcup_{i=1}^d \frac{r_n}{d} A_i\Big) \ \forall t.$

Using (A.4), $\exists M_n \ni \forall t > M_n$, $\mu_t\left(\frac{r_n}{d}A_i\right) < \frac{d}{r_n} + 1$, $\forall i$. Also, $\forall t \leq M_n$, $\mu_t\left(\frac{r_n}{d}A_i\right) = b(t)\mathbb{P}(|X_i| > \frac{r_n t}{d}) \leq b(M_n)$ as b is non-decreasing. Thus, $\forall r_n \downarrow 0$ and $\forall t > 1$,

$$\mu_t \Big(\mathbb{R}^d \setminus B_{r_n}(0) \Big) \le \mu_t \Big(\bigcup_{i=1}^d \frac{r_n}{d} A_i \Big) \le \sum_{i=1}^d \mu_t \left(\frac{r_n}{d} A_i \right) \le d \left[\left(\frac{d}{r_n} + 1 \right) \vee b(M_n) \right]$$

$$\implies \sup_{t > 1} \mu_t \Big(\mathbb{R}^d \setminus B_{r_n}(0) \Big) < \infty \quad \forall r_n \downarrow 0$$

This proves (A.2) in A.3. For proving (A.3), begin with fixing any $r_n \downarrow 0$ and $\epsilon > 0$. Define $C_{n,\epsilon} = R_n \bigcap_{i=1}^d A_i^c$ where $R_n = R_{n,\epsilon}$ satisfies the following:

- 1. $R_n > \max\left(1, r_n, \frac{2d}{\epsilon}\right)$ 2. If M_{ϵ} is such that $\forall t > M_{\epsilon}$, $\mu_t\left(xA_i\right) \leq \frac{1}{x} + \frac{\epsilon}{2d} \ \forall i \ and \ \forall x > 1$, then R_n be such that $\mathbb{P}(|X_i| > R_n) \leq \frac{\epsilon}{db(M_{\epsilon})} \ \forall i$. Note that here we use Proposition 2.4 in Resnick (2007) which states that (A.4) holds uniformly over $x \in (b, \infty) \ \forall \ b > 0$. Here we take b = 1 when we impose $R_n > 1$.

Observe that, $\mu_t \left(\mathbb{R}^d \setminus (C_{n,\epsilon} \cup B_{r_n}(0)) \right) = \mu_t \left(\bigcup_{i=1}^d R_n A_i \right) \leq \sum_{i=1}^d \mu_t(R_n A_i).$ Then, if $t > M_{\epsilon}$,

$$\mu_t(R_n A_i) \le \frac{1}{R_n} + \frac{\epsilon}{2d} < \frac{\epsilon}{d}$$

(using uniform convergence over $(1, \infty)$ and condition 1 on R)

$$\implies \sum_{i=1}^{d} \mu_t(R_n A_i) \le \epsilon$$

Next, if $1 < t \le M_{\epsilon}$,

$$\mu_t(R_n A_i) = b(t) \mathbb{P}(|X_i| > t R_n) \le b(M_{\epsilon}) \mathbb{P}(|X_i| > R_n) \le \epsilon/d$$
 (using condition 2 on R)

$$\implies \sum_{i=1}^{d} \mu_t(R_n A_i) \le \epsilon$$

Thus, $\forall t > 1, \mu_t \left(\mathbb{R}^d \setminus (C_{n,\epsilon} \cup B_{r_n}(0)) \right) \leq \epsilon$, which finally proves (A.3) in A.3, and hence the relative compactness of $\{\mu_t, t > 1\}$ in M_0 .

Remark A.1. Proposition A.4 is quite useful. As we shall see below, it implies that multivariate regular variation holds whenever the tail-dependence coefficients vanish. This recovers the classical result due to Berman (1961) but it is more widely applicable since it shows the relative compactness of the tail measure for arbitrary random vector with heavy-tailed moarginals.

We start with positive regularly varying random variables and later generalize to all real-valued random variables.

Lemma A.1. Say X, Y are non-negative random variables in $RV_{-1}(b,c)$ for some regularly varying monotone function $b(t) \to \infty$ as $t \to \infty$ and c > 0, i.e., $\forall x > 0$

$$\lim_{t \to \infty} b(t) \mathbb{P}(X > tx) = cx^{-1}, \quad and \quad \lim_{t \to \infty} b(t) \mathbb{P}(Y > tx) = cx^{-1}$$
(A.5)

If they are also asymptotically independent in the upper tail, i.e.,

$$\Lambda(X,Y) := \lim_{p \to 1^{-}} \mathbb{P}\left(X > F_X^{-1}(p)|Y > F_Y^{-1}(p)\right) = 0$$

then,

$$\lim_{t \to \infty} \mathbb{P}(X > t | Y > t) = 0 \tag{A.6}$$

Here F_X, F_Y represent the distribution functions of X and Y respectively while F_X^{-1}, F_Y^{-1} refer to their generalized inverses.

Proof. Let $t \in \mathbb{R}$ and define $p_X(t) = F_X(t)$, $p_Y(t) = F_Y(t)$. Clearly, as $t \to \infty$, $p_X(t) \to 1^-$ and $p_Y(t) \to 1^-$. Now,

$$\mathbb{P}(X > t | Y > t) = \frac{\mathbb{P}(X > t, Y > t)}{\mathbb{P}(Y > t)} = \frac{\mathbb{P}(X > F_X^{-1}(p_X(t)), Y > F_Y^{-1}(p_Y(t)))}{\mathbb{P}(Y > F_Y^{-1}(p_Y(t)))}$$

Note that the above equality does not assume $t = F_X^{-1}(p_X(t)) = F_Y^{-1}(p_Y(t))$. Instead we observe $\mathbb{P}(F_X^{-1}(p_X(t)) < X \le t) = \mathbb{P}(F_Y^{-1}(p_Y(t)) < Y \le t) = 0$, implying that $\{X > t\}$ and $\{X > F_X^{-1}(p_X(t))\}$ are almost surely the same events (same for Y).

Also, the above expressions are all well-defined for every t as the denominator is never exactly zero. This is because we assumed the tail-dependence coefficient Λ to exist which implies X and Y both have supports extending to infinity,i.e.,

$$\sup\{x : \mathbb{P}(X > x) > 0\} = \infty \quad \text{(same for Y)}$$

Next observe that due to (A.5), X and Y are tail equivalent. Indeed,

$$\lim_{t \to \infty} b(t) \mathbb{P}(X > t) = c \text{ and } \lim_{t \to \infty} b(t) \mathbb{P}(Y > t) = c$$

$$\implies \lim_{t \to \infty} \frac{\mathbb{P}(X > t)}{\mathbb{P}(Y > t)} = 1 \text{ or } \lim_{t \to \infty} \frac{1 - p_X(t)}{1 - p_Y(t)} = 1$$
(A.7)

Now, if $p_X(t) \ge p_Y(t)$, then $F_X^{-1}(p_X(t)) \ge F_X^{-1}(p_Y(t))$

$$\implies \mathbb{P}\left(X > F_X^{-1}(p_X(t)), Y > F_Y^{-1}(p_Y(t))\right) \le \mathbb{P}\left(X > F_X^{-1}(p_Y(t)), Y > F_Y^{-1}(p_Y(t))\right)$$

$$\implies \frac{\mathbb{P}\left(X > t, Y > t\right)}{\mathbb{P}\left(Y > t\right)} \le \frac{\mathbb{P}\left(X > F_X^{-1}(p_Y(t)), Y > F_Y^{-1}(p_Y(t))\right)}{\mathbb{P}\left(Y > F_Y^{-1}(p_Y(t))\right)} \tag{A.8}$$

On the other hand, if $p_X(t) < p_Y(t)$, then $F_X^{-1}(p_X(t)) \le F_X^{-1}(p_Y(t))$ so we can't use the above bound. However, we can establish a bound infinitesimally close to the last one:

$$\frac{\mathbb{P}(X > t, Y > t)}{\mathbb{P}(Y > t)} = \frac{\mathbb{P}(X > F_X^{-1}(p_X(t)), Y > F_Y^{-1}(p_Y(t)))}{\mathbb{P}(Y > F_Y^{-1}(p_Y(t)))}
= \frac{\mathbb{P}(X > F_X^{-1}(p_Y(t)), Y > F_Y^{-1}(p_Y(t)))}{\mathbb{P}(Y > F_Y^{-1}(p_Y(t)))} + \frac{\mathbb{P}(F_X^{-1}(p_Y(t)) \ge X > F_X^{-1}(p_X(t)), Y > F_Y^{-1}(p_Y(t)))}{\mathbb{P}(Y > F_Y^{-1}(p_Y(t)))}
\leq \frac{\mathbb{P}(X > F_X^{-1}(p_Y(t)), Y > F_Y^{-1}(p_Y(t)))}{\mathbb{P}(Y > F_Y^{-1}(p_Y(t)))} + \frac{\mathbb{P}(F_X^{-1}(p_Y(t)) \ge X > F_X^{-1}(p_X(t)))}{\mathbb{P}(Y > F_Y^{-1}(p_Y(t)))}
= \frac{\mathbb{P}(X > F_X^{-1}(p_Y(t)), Y > F_Y^{-1}(p_Y(t)))}{\mathbb{P}(Y > F_Y^{-1}(p_Y(t)))} + \frac{p_Y(t) - p_X(t)}{1 - p_Y(t)}
= \frac{\mathbb{P}(X > F_X^{-1}(p_Y(t)), Y > F_Y^{-1}(p_Y(t)))}{\mathbb{P}(Y > F_Y^{-1}(p_Y(t)))} + \frac{1 - p_X(t)}{1 - p_Y(t)} - 1
\leq \frac{\mathbb{P}(X > F_X^{-1}(p_Y(t)), Y > F_Y^{-1}(p_Y(t)))}{\mathbb{P}(Y > F_Y^{-1}(p_Y(t)))} + \frac{1 - p_X(t)}{1 - p_Y(t)} - 1$$
(A.9)

Thus, combining (A.8) and (A.9), we get that for all t,

$$\mathbb{P}(X > t | Y > t) \le \mathbb{P}\left(X > F_X^{-1}(p_Y(t)) \mid Y > F_Y^{-1}(p_Y(t))\right) + \left|\frac{1 - p_X(t)}{1 - p_Y(t)} - 1\right| \tag{A.10}$$

Now the RHS of the above converges to 0 as $t \to \infty$. This is because,

$$\lim_{t \to \infty} \mathbb{P}\left(X > F_X^{-1}(p_Y(t)) \mid Y > F_Y^{-1}(p_Y(t))\right) = \lim_{p \to 1^-} \mathbb{P}\left(X > F_X^{-1}(p) | Y > F_Y^{-1}(p)\right)$$
$$= \Lambda(X, Y) = 0$$

And the second term goes to 0 due to (A.7). Hence,

$$\lim_{t \to \infty} \mathbb{P}\left(X > t | Y > t\right) = 0$$

which proves the claim.

Corollary A.1. Say X, Y are non-negative random variables in $RV_{-1}(b, c_x)$ and $RV_{-1}(b, c_y)$ for some $c_x, c_y > 0$ and some regularly varying monotone function $b(t) \to \infty$, respectively. Also assume that they are asymptotically independent in the upper tail. Then,

$$\lim_{t \to \infty} \mathbb{P}(X/c_x > t | Y/c_y > t) = 0 \tag{A.11}$$

Proof. Clearly, $X \in RV_{-1}(b, c_x), Y \in RV_{-1}(b, c_y) \Longrightarrow X/c_x, Y/c_y \in RV_{-1}(b, 1)$. Moreover, using the fact that $F_{X/c_x}^{-1}(p) = c_x^{-1} F_X^{-1}(p), \ F_{Y/c_y}^{-1}(p) = c_y^{-1} F_Y^{-1}(p),$

$$\Lambda(X,Y) = \Lambda\left(\frac{X}{c_x}, \frac{Y}{c_y}\right) = 0$$

Thus, using Lemma A.1 we are done.

Proposition A.5. Say X, Y are non-negative random variables in $RV_{-1}(b, c)$. If they are also asymptotically independent, i.e., $\Lambda(X, Y) = 0$, then, $(X, Y) \in RV_{-1}(b, \mu_{iid}^+)$ where μ_{iid}^+ is the limit measure concentrated on the positive axes corresponding to the random vector comprised of i.i.d. positive $RV_{-1}(b, c)$ random variables.

Proof. From Lemma A.1 we know that,

$$\begin{split} &\lim_{t \to \infty} \mathbb{P}\left(X > t | Y > t\right) = 0 \\ &\implies \lim_{t \to \infty} \frac{\mathbb{P}\left(X > t, Y > t\right)}{\mathbb{P}\left(Y > t\right)} = 0 \\ &\implies \lim_{t \to \infty} \frac{b(t)\mathbb{P}\left(X > t, Y > t\right)}{b(t)\mathbb{P}\left(Y > t\right)} = 0 \end{split}$$

Now, due to (A.5),

$$\lim_{t \to \infty} b(t) \mathbb{P}(Y > t) = c > 0$$

Combining with the previous equality.

$$\lim_{t \to \infty} b(t) \mathbb{P} (X > t, Y > t) = 0$$

$$\implies \lim_{t \to \infty} b(t) \mathbb{P} ((X, Y) \in t \cdot B_1 \cap B_2) = 0$$

where $B_1 = [1, \infty) \times \mathbb{R}_{\geq 0}$ and $B_2 = \mathbb{R}_{\geq 0} \times [1, \infty)$. Now note that for any $\epsilon > 0$, X/ϵ and $Y/\epsilon \in RV_{-1}(b, c/\epsilon)$. Thus, all the above results hold by replacing (X, Y) by $\left(\frac{X}{\epsilon}, \frac{Y}{\epsilon}\right)$. As a result, $\forall \epsilon > 0$,

$$\lim_{t \to \infty} b(t) \mathbb{P}\left((X, Y) \in t \cdot (\epsilon(B_1 \cap B_2)) \right) = 0 \tag{A.12}$$

Denoting (X,Y) by Z, let $\mu_t(A) := b(t)\mathbb{P}\left(\frac{Z}{t} \in A\right)$ be the rescaled tail measure of Z as defined in Proposition A.4. Thus,

$$\forall \epsilon > 0, \lim_{t \to \infty} \mu_t(\epsilon(B_1 \cap B_2)) = 0. \tag{A.13}$$

Now using Proposition A.4, the above set of rescaled measures is relatively compact, so $\forall t_n \to \infty \exists n_k \to \infty \ni \{\mu_{t_{n_k}}\}$ converges to some measure $\mu' \in \mathbb{M}_0$. To prove the claim it is enough to show that any such μ' is equal to μ_{iid}^+ . This guarantees uniqueness of subsequential limits of μ_t , which in turn implies convergence of μ_t to μ_{iid}^+ .

Then by Proposition A.1, $\forall f \in C_0$, $\int_{\mathbb{R}_0^2} f d\mu_t \longrightarrow \int_{\mathbb{R}_0^2} f d\mu'$ as $t \to \infty$. Consider a closed BAFO rectangle R_1 and an open BAFO rectangle $R_2 \supset R_1$, both not touching the axes. More rigorously, if $A_x := (0, \infty) \times \{0\}$ (the positive X-axis) and $A_y := \{0\} \times (0, \infty)$ (the positive Y-axis), then $R_1 \subset R_2 \subset \mathbb{R}_0^2 \setminus (A_x \cup A_y)$. Now, Urysohn's lemma guarantees us the existence of a continuous function f such that $f \in [0, 1]$, $f \equiv 1$ on R_1 and $\sup (f) = \{x : f(x) > 0\} \subset R_2$. Then,

$$\int_{\mathbb{R}_0^2} f d\mu_t = \int_{R_2} f d\mu_t \le \mu_t(R_2)$$

Let $\{(a,y): y>0\}$ and $\{(x,b): x>0\}$ be the left and bottom edge of R_2 respectively. Then $R_2 \subset (a \wedge b)(B_1 \cap B_2) \implies \mu_t(R_2) \le \mu_t((a \wedge b)(B_1 \cap B_2))$. Thus, by (A.13),

$$\lim_{t \to \infty} \int_{\mathbb{R}_0^2} f d\mu_t \le \lim_{t \to \infty} \mu_t((a \land b)(B_1 \cap B_2)) = 0$$

$$\implies \int_{\mathbb{R}_0^2} f d\mu' = 0$$

$$\implies \int_{B_1} f d\mu' = 0 \implies \mu'(R_1) = 0$$

The last step holds because f is identically 1 on R_1 . Hence, μ' is zero on any closed BAFO rectangle in \mathbb{R}^2_0 which does not touch the axes. Note that $\mathbb{R}^2_0 \setminus (A_x \cup A_y)$ is the countable union of such rectangles, so,

$$\mu'(\mathbb{R}_0^2 \setminus (A_x \cup A_y)) = 0 \tag{A.14}$$

To complete this proof, take a BAFO Borel set $E \ni \mu'(\partial E) = 0$ and let

$$E_x := \{x : (x,0) \in E \cap A_x\}$$
 (intersection of E with X-axis), and $E_y := \{y : (0,y) \in E \cap A_y\}$ (intersection of E with Y-axis) (A.15)

Then,

$$\mu'(E) = \mu'(E_x \times \{0\}) + \mu'(\{0\} \times E_y) + \mu'(E \cap (\mathbb{R}_0^2 \setminus (A_x \cup A_y)))$$

$$= \mu'(E_x \times \mathbb{R}) + \mu'(\mathbb{R} \times E_y) + 0$$

$$= \lim_{k \to \infty} b(t_{n_k}) \mathbb{P}(X/t_{n_k} \in E_x) + \lim_{k \to \infty} b(t_{n_k}) \mathbb{P}(Y/t_{n_k} \in E_y)$$

$$= \mu_c(E_x) + \mu_c(E_y) = \mu_{iid}^+(E)$$

where $d\mu_c := cx^{-2}dx$ is the limit measure of a $RV_{-1}(b,c)$ random variable. Note that the convergence in the third equality holds because E is BAFO Borel implies $E_x \times \mathbb{R}$ is too and $\mu'(\partial(E_x \times \mathbb{R})) = \mu'(\partial E_x \times \mathbb{R}) = \mu'(\partial E_x \times \{0\}) \leq \mu'(\partial E) = 0$.

Thus, $\mu' = \mu_{iid}^+$ for every subsequential limit of μ_t , which implies $\mu_t \longrightarrow \mu_{iid}^+$ as $t \to \infty$ which proves the claim.

Corollary A.2. Say X, Y are non-negative random variables in $RV_{-1}(b, c_x)$ and $RV_{-1}(b, c_y)$ respectively. If they are also asymptotically independent, then, $(X, Y) \in RV_{-1}(b, \mu_{indep}^+)$ where μ_{indep}^+ is the limit measure concentrated on the positive axes corresponding to the random vector comprised of independent positive $RV_{-1}(b, c_x)$ and $RV_{-1}(b, c_y)$ random variables.

Proof. Clearly, $X \in RV_{-1}(b, c_x), Y \in RV_{-1}(b, c_y) \Longrightarrow X/c_x, Y/c_y \in RV_{-1}(b, 1)$. Moreover, using the fact that $F_{X/c_x}^{-1}(p) = c_x^{-1} F_X^{-1}(p), \ F_{Y/c_y}^{-1}(p) = c_y^{-1} F_Y^{-1}(p),$

$$\Lambda\left(\frac{X}{c_x}, \frac{Y}{c_y}\right) = \lim_{p \to 1^-} \mathbb{P}\left(\frac{X}{c_x} > F_{X/c_x}^{-1}(p) \middle| \frac{Y}{c_y} > F_{Y/c_y}^{-1}(p)\right) = \Lambda(X, Y) = 0$$

Thus, X/c_x and Y/c_y are asymptotically independent too. By Proposition A.5,

$$\left(\frac{X}{c_x}, \frac{Y}{c_y}\right) \in RV_{-1}(b, \mu_{iid}^+)$$

Now note that, μ_{indep}^+ is given by

$$\mu_{indep}^+(E) = \mu_{c_x}(E_x) + \mu_{c_y}(E_y) \quad \forall \text{ Borel subsets } E \text{ of } \mathbb{R}^2_+ \setminus \{\mathbf{0}\}$$

where E_x , E_y are as in (A.15), $d\mu_{c_x} = c_x u^{-2} du$ and $d\mu_{c_y} = c_y u^{-2} du$. To prove $(X, Y) \in RV_{-1}(b, \mu_{indep}^+)$, using Lemma 6.1 in Resnick (1987), it is enough to show that,

$$\lim_{t\to\infty}b(t)\mathbb{P}\left(\left(\frac{X}{t},\frac{Y}{t}\right)\in[\mathbf{0},\boldsymbol{z}]^c\right)=\mu_{indep}^+([\mathbf{0},\boldsymbol{z}]^c)\quad\forall\;\boldsymbol{z}=(z_1,z_2)\in\mathbb{R}_+^2$$

Indeed,

$$\lim_{t \to \infty} b(t) \mathbb{P}\left(\left(\frac{X}{t}, \frac{Y}{t}\right) \in [\mathbf{0}, \mathbf{z}]^{c}\right)$$

$$= \lim_{t \to \infty} b(t) \mathbb{P}\left(\left(\frac{X/c_{x}}{t}, \frac{Y/c_{y}}{t}\right) \in ([0, z_{1}/c_{x}] \times [0, z_{2}/c_{y}])^{c}\right)$$

$$= \mu_{iid}^{+}(([0, z_{1}/c_{x}] \times [0, z_{2}/c_{y}])^{c})$$

$$= c_{x}z_{1}^{-1} + c_{y}z_{2}^{-1}$$

$$= \mu_{c_{x}}(([\mathbf{0}, \mathbf{z}]^{c})_{x}) + \mu_{c_{y}}(([\mathbf{0}, \mathbf{z}]^{c})_{y}) = \mu_{indep}^{+}([\mathbf{0}, \mathbf{z}]^{c})$$

This proves the claim.

Proposition A.6. Say X, Y are two real random variables with regularly varying upper and lower tails of index -1, i.e. $\exists b(t) \to \infty$ and $c_X^{\pm}, c_Y^{\pm} > 0$ such that $\forall x > 0$,

$$\lim_{t \to \infty} b(t) \mathbb{P}(\pm X > tx) = c_X^{\pm} x^{-1} \quad and \quad \lim_{t \to \infty} b(t) \mathbb{P}(\pm Y > tx) = c_Y^{\pm} x^{-1} \tag{A.16}$$

Suppose they are asymptotically independent in all tails, i.e., following tail dependence coefficients are zero for all combinations of \pm :

$$\Lambda(\pm X, \pm Y) = 0 \tag{A.17}$$

Then, $(X,Y) \in RV_{-1}(b,\mu_{indep})$ where μ_{indep} is the limit measure concentrated on the axes corresponding to the random vector comprised of independent random variables with $RV_{-1}(b,c_X^{\pm})$ and $RV_{-1}(b,c_Y^{\pm})$ tails, respectively.

Proof. Note that (A.17) implies

$$\Lambda(X_{\pm}, Y_{\pm}) = 0 \tag{A.18}$$

where X_+, Y_+ and X_-, Y_- represent the positive and negative parts of X and Y respectively. Indeed, for large p,

$$\{-X > F_{-X}^{-1}(p)\} = \{X < -F_{-X}^{-1}(p)\} = \{X_{-} > F_{-X}^{-1}(p)\}$$

as large p implies $F_{-X}^{-1}(p)$ is positive. Note that due to assumption of regular variation of tails, support of X extends to both $+\infty$ and $-\infty$ so $F_{-X}^{-1}(p)$ is guaranteed to be positive if we take p sufficiently large.

Now, for all x > 0, $F_{-X}(x) = F_{X_{-}}(x)$. Thus, if p is sufficiently large, $F_{X_{-}}^{-1}(p) = F_{-X}^{-1}(p)$. Thus,

$$\{-X>F_{-X}^{-1}(p)\}=\{X_{-}>F_{-X}^{-1}(p)\}=\{X_{-}>F_{X}^{-1}(p)\}$$

Similarly we can conclude that $\{Y > F_Y^{-1}(p)\} = \{Y_+ > F_{Y_+}^{-1}(p)\}$ for large p. Therefore,

$$\begin{split} \Lambda(X_-,Y_+) &= \lim_{p \to 1-} \mathbb{P}(X_- > F_{X_-}^{-1}(p) \big| Y_+ > F_{Y_+}^{-1}(p)) \\ &= \lim_{p \to 1-} \mathbb{P}(-X > F_{-X}^{-1}(p) \big| Y > F_Y^{-1}(p)) = \Lambda(-X,Y) = 0 \end{split}$$

Similarly,

$$\Lambda(X_{-}, Y_{-}) = \Lambda(X_{+}, Y_{+}) = \Lambda(X_{+}, Y_{-}) = 0$$

Observe that (A.16) implies that $X_{\pm} \in RV_{-1}(b, c_X^{\pm})$ and $Y_{\pm} \in RV_{-1}(b, c_Y^{\pm})$. Thus using Corollary A.2, $(X_{\pm}, Y_{\pm}) \in RV_{-1}(b, \mu_{indep}^+)$.

Let $Q_{+,+} = \mathbb{R}^2_+, Q_{+,-} = \mathbb{R}_+ \times \mathbb{R}_-, Q_{-,-} = \mathbb{R}^2_-$ and $Q_{-,+} = \mathbb{R}_- \times \mathbb{R}_+$ denote the four quadrants of \mathbb{R}^2 minus the axes and let $A_x^+, A_y^+, A_x^-, A_y^-$ denote the positive and negative X and Y axis respectively. Next take any BAFO Borel set $E \subset \mathbb{R}^2 \setminus \{0\}$ such that $\mu_{indep}(\partial E) = 0$. Then,

$$\lim_{t \to \infty} b(t) \mathbb{P}((X,Y) \in t \cdot E)$$

$$= \lim_{t \to \infty} b(t) \mathbb{P}((X,Y) \in t \cdot E \cap Q_{+,+}) + \lim_{t \to \infty} b(t) \mathbb{P}((X,Y) \in t \cdot E \cap Q_{+,-})$$

$$+ \lim_{t \to \infty} b(t) \mathbb{P}((X,Y) \in t \cdot E \cap Q_{-,-}) + \lim_{t \to \infty} b(t) \mathbb{P}((X,Y) \in t \cdot E \cap Q_{-,+})$$

$$+ \lim_{t \to \infty} b(t) \mathbb{P}((X,Y) \in t \cdot E \cap A_x^+) + \lim_{t \to \infty} b(t) \mathbb{P}((X,Y) \in t \cdot E \cap A_x^-)$$

$$+ \lim_{t \to \infty} b(t) \mathbb{P}((X,Y) \in t \cdot E \cap A_y^+) + \lim_{t \to \infty} b(t) \mathbb{P}((X,Y) \in t \cdot E \cap A_y^-)$$
(A.19)

if all the limits above exist.

Now observe that $\{(X,Y) \in t \cdot Q_{\pm,\pm}\} = \{(X_{\pm},Y_{\pm}) \in t \cdot Q_{+,+}\}$. As $(X_{\pm},Y_{\pm}) \in RV_{-1}(b,\mu_{indep}^+)$ and μ_{indep}^+ assigns zero mass to any set not intersecting the axes,

$$\lim_{t \to \infty} b(t) \mathbb{P}((X, Y) \in t \cdot Q_{\pm, \pm}) = \lim_{t \to \infty} b(t) \mathbb{P}((X_{\pm}, Y_{\pm}) \in t \cdot Q_{+, +})$$

$$= \mu_{indep}^{+}(Q_{+, +}) = 0$$

$$\implies \lim_{t \to \infty} b(t) \mathbb{P}((X, Y) \in t \cdot E \cap Q_{\pm, \pm}) \le \lim_{t \to \infty} b(t) \mathbb{P}((X_{\pm}, Y_{\pm}) \in t \cdot Q_{+, +}) = 0$$

Thus the first four terms in (A.19) indeed exist and are zero! Let $E_x^+ = \{x \in \mathbb{R}_+ : (x,0) \in E \cap A_x^+\}$. Similarly define E_x^-, E_y^+ and E_y^- . Then,

$$\begin{split} &\lim_{t\to\infty}b(t)\mathbb{P}((X,Y)\in t\cdot E)\\ &=\lim_{t\to\infty}b(t)\mathbb{P}((X,Y)\in t\cdot E\cap A_x^+)+\lim_{t\to\infty}b(t)\mathbb{P}((X,Y)\in t\cdot E\cap A_x^-)\\ &+\lim_{t\to\infty}b(t)\mathbb{P}((X,Y)\in t\cdot E\cap A_y^+)+\lim_{t\to\infty}b(t)\mathbb{P}((X,Y)\in t\cdot E\cap A_y^-)\\ &=\lim_{t\to\infty}b(t)\mathbb{P}((X,Y)\in t\cdot (E_x^+\times\{0\}))+\lim_{t\to\infty}b(t)\mathbb{P}((X,Y)\in t\cdot (E_x^-\times\{0\}))\\ &+\lim_{t\to\infty}b(t)\mathbb{P}((X,Y)\in t\cdot (\{0\}\times E_y^+))+\lim_{t\to\infty}b(t)\mathbb{P}((X,Y)\in t\cdot (\{0\}\times E_y^-))\\ &=\lim_{t\to\infty}b(t)\mathbb{P}((X,Y)\in t\cdot (E_x^+\times\mathbb{R}))+\lim_{t\to\infty}b(t)\mathbb{P}((X,Y)\in t\cdot (E_x^-\times\mathbb{R}))\\ &+\lim_{t\to\infty}b(t)\mathbb{P}((X,Y)\in t\cdot (\mathbb{R}\times E_y^+))+\lim_{t\to\infty}b(t)\mathbb{P}((X,Y)\in t\cdot (\mathbb{R}\times E_y^-))\\ &=\lim_{t\to\infty}b(t)\mathbb{P}(X\in t\cdot E_x^+)+\lim_{t\to\infty}b(t)\mathbb{P}(X\in t\cdot E_x^-)\\ &+\lim_{t\to\infty}b(t)\mathbb{P}(Y\in t\cdot E_y^+)+\lim_{t\to\infty}b(t)\mathbb{P}(Y\in t\cdot E_y^-)\\ &=\mu_{+X}(E_x^+)+\mu_{-X}(E_x^-)+\mu_{+Y}(E_y^+)+\mu_{-Y}(E_y^-)=\mu_{indep}(E) \end{split} \tag{A.20}$$

where $d\mu_{\pm X} = c_X^{\pm} u^{-2} du$ and $d\mu_{\pm Y} = c_Y^{\pm} u^{-2} du$. Note that existence of all the limits involved in the above equalities is justified by the step below it, so no issues regarding existence remain. This proves the claim.

Theorem A.1. Let $X = (X_i)_{i=1}^d$ be a random vector whose marginals have regularly varying distributions with index -1, i.e., \exists a monotone increasing function $b(t) \to \infty$ and $c_{\pm}(i) > 0$ such that

$$\lim_{t \to \infty} b(t) \mathbb{P}\left(\pm X_i > tx\right) = c_{\pm}(i)x^{-1} \quad \forall x > 0 \text{ and } \forall i = 1, \dots, d$$

If $\forall 1 \leq i \neq j \leq d$,

$$\Lambda(\pm X_i, \pm X_j) = 0$$

then, $X \in RV_{-1}(b, \mu_{indep})$, where $\mu_{indep}^{(d)}$ is same as that in Proposition A.6 but in $d \in \mathbb{N}$ dimensions.

Proof. Define $Q_{S_0,S_1,S_{-1}} := \{x \in \mathbb{R}^d : sgn(x_i) = \mathbf{1}[i \in S_1] - \mathbf{1}[i \in S_{-1}] \ \forall i \in [d] \}$ for all $S_0, S_1, S_{-1} \ni S_0 \sqcup S_1 \sqcup S_{-1} = [d], |S_1|, |S_{-1}| \in \{0, 1, \ldots, d\} \text{ and } |S_0| \in \{0, 1, \ldots, d-2\}.$ Here $sgn(z) = \mathbf{1}[z > 0] - \mathbf{1}[z < 0]$. Similar to Proposition A.6, also define $A_i^+, A_i^- \ \forall i \in [d]$ where A_i^+ represents the positive *i*-th axis and A_i^- represents the negative *i*-th axis. Thus, $(Q_{S_0,S_1,S_{-1}})_{(S_0,S_1,S_{-1})}$ take out the axes and partition $\mathbb{R}_0^d \setminus \bigcup_{i=1}^d (A_i^+ \cup A_i^-)$ according to positive, negative and zero coordinates. Now, note that S_0 can take at most d-2 coordinates, so at least two coordinates are always

non-zero. Thus, $\forall S_0, S_1, S_{-1}, \exists k \neq l \in [d] \ni \forall t > 0, \{X \in t \cdot Q_{A_0, S_1, S_{-1}}\} \subset \{(X_k, X_l) \in T_k\}$ $t \cdot (\mathbb{R}^2 \setminus ((A_k^+ \cup A_k^-) \cup (A_l^+ \cup A_l^-)))$. Here we abuse notation a bit: A_i^+, A_i^- were defined to be the i-th axes in d-dimensions, but we use the same notation for the axes in 2-dimensions. Thus,

$$\lim_{t \to \infty} b(t) \mathbb{P} \left(X \in t \cdot \left(\mathbb{R}_0^d \setminus \bigcup_{i=1}^d \left(A_i^+ \cup A_i^- \right) \right) \right) \\
= \sum_{S_0, S_1, S_{-1}} \lim_{t \to \infty} b(t) \mathbb{P} \left(X \in t \cdot Q_{S_0, S_1, S_{-1}} \right) \\
\leq \sum_{S_0, S_1, S_{-1}} \lim_{t \to \infty} b(t) \mathbb{P} \left(\bigcup_{1 \le k \ne l \le d} \left\{ (X_k, X_l) \in t \cdot \left(\mathbb{R}_0^2 \setminus \left((A_k^+ \cup A_k^-) \cup (A_l^+ \cup A_l^-) \right) \right) \right) \right) \\
\leq \sum_{S_0, S_1, S_{-1}} \sum_{1 \le k \ne l \le d} \lim_{t \to \infty} b(t) \mathbb{P} \left((X_k, X_l) \in t \cdot \left(\mathbb{R}_0^2 \setminus \left((A_k^+ \cup A_k^-) \cup (A_l^+ \cup A_l^-) \right) \right) \right) \\
= 0 = \mu_{indep}^{(d)} \left(\mathbb{R}_0^d \setminus \bigcup_{i=1}^d \left(A_i^+ \cup A_i^- \right) \right) \tag{A.21}$$

where (A.21) holds because Proposition A.6 implies $(X_k, X_l) \in RV_{-1}(b, \mu_{indep}^{(2)})$ and,

$$(X_{k}, X_{l}) \in RV_{-1}\left(b, \mu_{indep}^{(2)}\right)$$

$$\implies \lim_{t \to \infty} b(t) \mathbb{P}\left((X_{k}, X_{l}) \in t \cdot \left(\mathbb{R}_{0}^{2} \setminus \left((A_{k}^{+} \cup A_{k}^{-}) \cup (A_{l}^{+} \cup A_{l}^{-})\right)\right)\right)$$

$$= \mu_{indep}^{(2)}\left(\mathbb{R}_{0}^{2} \setminus \left((A_{k}^{+} \cup A_{k}^{-}) \cup (A_{l}^{+} \cup A_{l}^{-})\right)\right) = 0 \quad \forall \ k \neq l$$

Now, take any BAFO Borel set $E \subset \mathbb{R}_0^d$ such that $\mu_{indep}^{(d)}(\partial E) = 0$. Define $E_i^{\pm} = \{x \in \mathbb{R}_{\pm} : x \in E \cap A_i^{\pm}\}$. Then,

$$\begin{split} \lim_{t \to \infty} b(t) \mathbb{P} \left(X \in t \cdot E \right) &= \sum_{i=1}^d \lim_{t \to \infty} b(t) \mathbb{P} \left(X \in t \cdot \left(\{0\}^{i-1} \times E_i^+ \times \{0\}^{d-i} \right) \right) \\ &+ \lim_{t \to \infty} b(t) \mathbb{P} \left(X \in t \cdot \left(\{0\}^{i-1} \times E_i^- \times \{0\}^{d-i} \right) \right) \\ &= \sum_{i=1}^d \lim_{t \to \infty} b(t) \mathbb{P} \left(X \in t \cdot \left(\mathbb{R}^{i-1} \times E_i^+ \times \mathbb{R}^{d-i} \right) \right) \\ &+ \lim_{t \to \infty} b(t) \mathbb{P} \left(X \in t \cdot \left(\mathbb{R}^{i-1} \times E_i^- \times \mathbb{R}^{d-i} \right) \right) \\ &= \sum_{i=1}^d \lim_{t \to \infty} b(t) \mathbb{P} \left(X_i \in t \cdot E_i^+ \right) + \lim_{t \to \infty} b(t) \mathbb{P} \left(X_i \in t \cdot E_i^- \right) \\ &= \sum_{i=1}^d \mu_i^+(E_i^+) + \mu_i^-(E_i^-) = \mu_{indep}^{(d)}(E) \end{split}$$

where $d\mu_i^{\pm} = c_{\pm}(i)x^{-2}dx \ \forall i = 1, \dots, d$. Note that the first two equalities above hold as (A.21) implies there is no mass outside of the axes.

This proves the claim.

Appendix B: A characterization result for integral functionals under linear constraints

Let (S, \mathcal{S}) be a measurable space and denote by $\mathcal{M}(S)$ the set of all finite positive measures on it. Let also $\mathbb{B}_+(S) := \mathbb{B}_+(S, \mathcal{S})$ denote the class of all non-negative and bounded measurable functions $f: S \to \mathbb{R}_+$. For $\varphi \in \mathcal{M}(S)$ and $f \in \mathbb{B}_+(S)$, we shall write

$$(f,\varphi) := \int_{S} f(x)\varphi(dx).$$

Definition B.1. We say that a finite set of non-negative functions $\mathcal{G} := \{g_i, i = 1, \dots, d\} \subset \mathbb{B}_+(S)$ satisfies the anti-dominance condition if for all \mathcal{I} , $\emptyset \neq \mathcal{I} \subsetneq \{1, \dots, d\}$, we have

$$\sum_{i \in \mathcal{I}} \lambda_i g_i(\cdot) \not \leq \sum_{j \in \mathcal{I}^c} \lambda_j g_j(\cdot),$$

for all $\lambda_i \geq 0$ such that $\sum_{i \in \mathcal{I}} \lambda_i > 0$.

The above anti-dominance condition means that among the *non-negative* linear combinations of the functions in \mathcal{G} only the trivial one can be dominated pointwise by another linear combination of the functions in \mathcal{G} . This anti-dominance property readily implies the following.

Lemma B.1. Suppose that a finite set of bounded non-negative measurable functions $\mathcal{G} = \{g_1, \dots, g_d\} \subset \mathbb{B}_+(S)$ satisfies the anti-dominance condition. If for some weights $w_i \in \mathbb{R}$, we have that

$$h(\cdot) = \sum_{i=1}^{d} w_i g_i(\cdot) \in \mathbb{B}_+(S), \tag{B.1}$$

Then, $w_i \geq 0$, for all $i = 1, \dots, d$.

Proof. Suppose that (B.1) holds where $w_{i_0} < 0$ for some $i_0 \in \{1, \dots, d\}$. Then, let $\mathcal{I} := \{i : w_i < 0\}$ and observe that since h and the g_i 's are all non-negative, then $\mathcal{I}^c = \{j : w_j \geq 0\}$ is non-empty. Thus $\emptyset \neq \mathcal{I} \subsetneq \{1, \dots, d\}$. On the other hand, Relation (B.1) can be equivalently written as

$$h(x) = \sum_{i \in \mathcal{I}^c} w_j g(x) - \sum_{i \in \mathcal{I}} |w_i| g_i(x), \quad x \in S.$$

This, since h is a non-negative function, entails that

$$\sum_{i \in \mathcal{I}} |w_i| g_i(x) \le \sum_{j \in \mathcal{I}^c} w_j g(x), \quad \forall x \in S,$$

where $|w_{i_0}| > 0$ for some $i_0 \in \mathcal{I}$. This contradicts the anti-dominance condition.

Remark B.1. While the anti-dominance condition may appear to be stringent, in some cases it is very easy to verify. Indeed, suppose that

$$S = \{(u_i)_{i=1}^d : u_i \ge 0, \sum_{i=1}^d u_i = 1\}$$

is the non-negative unit simplex. Let also $g_i(u) = u_i$, $u \in S$ be the coordinate functions. Then, clearly for no choice of $\lambda_i \geq 0$, and a non-empty set $\mathcal{I} \subsetneq \{1, \dots, d\}$ such that $\sum_{i \in \mathcal{I}} \lambda_i > 0$, can we have

$$\sum_{i \in \mathcal{I}} \lambda_i u_i \le \sum_{j \in \mathcal{I}^c} \lambda_j u_j, \quad \forall u = (u_i)_{i=1}^d \in S.$$

Indeed, this inequality is violated by taking $u_{j_0} \downarrow 0$, for some $j_0 \in \mathcal{I}^c$ with $\lambda_{j_0} > 0$.

Theorem B.1. Let $\mathcal{G} = \{g_1, \dots, g_d\}$ be a class of non-negative bounded measurable functions on (S, \mathcal{S}) . For a constant c > 0 define the set of positive finite measures:

$$\mathcal{M}_c(\mathcal{G}) := \{ \varphi \in \mathcal{M}(S) : (g, \varphi) = c, \forall g \in \mathcal{G} \}.$$

Suppose that for some $\{x_1, \dots, x_d\} \subset S$, we have that the matrix $G := (g_i(x_j))_{d \times d}$ is non-singular and that the vector $\mathbf{1} = (1 \dots 1)^{\top} \in \mathbb{R}^d$ belongs to the interior of the cone

$$G(\mathbb{R}^d_+) := \{ y : y = Gz, \ z \in \mathbb{R}^d_+ \},$$
 (B.2)

where \mathbb{R}^d_+ denotes $[0,\infty)^d$.

(i) If for some $h \in \mathbb{B}_+(S, \rho)$ we have that $(h, \varphi) = c$ for all $\varphi \in \mathcal{M}_c(\mathcal{G})$, then

$$h(x) = \sum_{i=1}^{d} \lambda_i g_i(x), \ (\forall x \in S) \quad with \ \lambda_i \in \mathbb{R}, \ such \ that \qquad \sum_{i=1}^{d} \lambda_i = 1.$$
 (B.3)

(ii) If, moreover, \mathcal{G} satisfies the anti-dominance condition, then (B.3) holds with $\lambda_i \geq 0$, $i = 1, \dots, d$.

Proof. For simplicity, and without loss of generality we will assume that c = 1. Assume that there $h \in \mathbb{B}_+(S)$ is such that (h, μ) for all $\mu \in \mathcal{M}_c(\mathcal{G})$. We will prove part (i) in two steps.

Step 1. Consider any set $\{y_i, i = 1, \dots, m\}$ containing the fixed set of points $\{x_1, \dots, x_d\}$ and define the matrix

$$D = (g_i(y_j))_{d \times m}.$$

Notice that G is a sub-matrix of D, obtained by selecting the d columns of D that correspond to the set $\{x_1, \dots, x_d\}$.

By assumption, we have that **1** is an interior point of $G(\mathbb{R}^d_+)$ and hence, **1** is also an interior point of $D(\mathbb{R}^m_+) \supset G(\mathbb{R}^d_+)$.

We will show that

$$D\mu = 1$$
, for some $\mu \in (0, \infty)^m$ (B.4)

that is, the vector μ has all positive entries.

Let $\mu_0 = (\mu_0(1), \dots, \mu_0(m)) \in (0, \infty)^m$ be an arbitrary vector of strictly positive entries. Since $\mathbf{1} \in D(\mathbb{R}_+^m)^\circ$, there exists a sufficiently small $\delta > 0$, and a $\mu_\delta \in \mathbb{R}_+^m$, such that $D\mu_\delta = \mathbf{1} - \delta D\mu_0$. Indeed, this follows from the facts that for all $\epsilon > 0$, there exists a $\delta > 0$ such that $\mathbf{1} - \delta D\mu_0 \in B_1(\epsilon)$ where $B_1(\epsilon) \subset D(\mathbb{R}_+^m)$.

Now, define

$$\mu := \mu_{\delta} + \delta \mu_0.$$

Observe that by construction $\mu \in (0, \infty)^m$ has all positive entries and

$$D\mu = \mathbf{1} - \delta D\mu_0 + \delta D(\mu_0) = \mathbf{1}.$$

This completes the proof of (B.4). We shall use this fact in the following step of the proof.

Step 2. Note that every $\nu \in \mathbb{R}^m_+$ corresponds to a measure

$$\varphi_{\nu}(du) := \sum_{\substack{i=1\\43}}^{m} \nu_{i} \epsilon_{\{y_{i}\}}(du),$$

where $\epsilon_{\{y\}}(A) = 1_A(y)$, $A \in \mathcal{S}$ is the unit mass measure at the singleton $\{y\}$. With this correspondence, we have that

$$(h, \varphi_{\nu}) = \mathbf{h}^{\top} \nu,$$

where $\mathbf{h} := (h(y_j))_{j=1}^m$. Thus, the assumptions of the theorem entail

$$\mathbf{h}^{\top} \nu = 1$$
, for all $\nu \in \mathbb{R}^m_+$ such that $D\nu = \mathbf{1}$

We will show that $\mathbf{h} \in V_{\mathcal{G}} := \operatorname{span}(\mathbf{g}_i, i = 1, \dots, d)$, where $\mathbf{g}_i := (g_i(y_j))_{i=1}^m$. Suppose that

$$\mathbf{h}_0 := \operatorname{Proj}_{V_{\mathcal{C}}}(\mathbf{h}).$$

Define the vector

$$\nu_{\epsilon} := \mu + \epsilon (\mathbf{h} - \mathbf{h}_0),$$

and notice that since by construction μ has positive entries, there is an $\epsilon > 0$, such that $\nu_{\epsilon} \in \mathbb{R}^{m}_{+}$. Then, since $\mathbf{h} - \mathbf{h}_{0} \perp \mathbf{g}_{\mathbf{i}}$, we obtain $D\nu_{\epsilon} = D\mu = \mathbf{1}$. This, by assumption implies

$$\mathbf{h}^{\top} \nu_{\epsilon} = 1.$$

Since by assumption we also have $\mathbf{h}^{\top}\mu = 1$, it follows that

$$0 = \mathbf{h}^{\top}(\nu_{\epsilon} - \mu) = \epsilon \mathbf{h}^{\top}(\mathbf{h} - \mathbf{h}_{0}).$$

This, however, since $\epsilon > 0$, implies that $\mathbf{h} - \mathbf{h}_0 = 0$. Indeed, since $\mathbf{h}_0 \in V_{\mathcal{G}} \perp \mathbf{h} - \mathbf{h}_0$, it follows that

$$0 = \mathbf{h}^{\top}(\mathbf{h} - \mathbf{h}_0) = (\mathbf{h} - \mathbf{h}_0)^{\top}(\mathbf{h} - \mathbf{h}_0) = \|\mathbf{h} - \mathbf{h}_0\|^2.$$

We have thus shown that $\mathbf{h} = \mathbf{h}_0 = \operatorname{Proj}_{V_{\mathcal{G}}}(\mathbf{h})$. This means that there exist coefficients $\lambda_i \in \mathbb{R}, i = 1, \dots, d$, possibly dependent on the set $\{y_j\}$, such that

$$h(y_j) = \sum_{i=1}^d \lambda_i g_i(y_j), \quad \text{for all } j = 1, \dots, m.$$
(B.5)

It remains to show that the coefficients λ_i do not depend on the choice of the $\{y_j\}$'s.

Notice, however, that we started with a fixed set $\{x_i, i = 1, \dots, d\} \subset \{y_j, j = 1, \dots, m\}$, such that the matrix $G = (g_i(x_j))_{d \times d}$ is invertible. By focusing on a subset of the equations in (B.5), we obtain $\lambda G = \widetilde{\mathbf{h}}^\top$, where $\widetilde{\mathbf{h}} = (h(x_i), i = 1, \dots, d)$. Hence $\lambda = \widetilde{\mathbf{h}}^\top G^{-1}$, which demonstrates the uniqueness of the vector $\lambda = (\lambda_i, i = 1, \dots, d)$. This completes the proof of part (i).

Part (ii) follows from Lemma B.1 due to the anti-dominance condition.

Appendix C: Calibration of the Fréchet Combination Test under MRV

Theorem C.1. Let $Z = (Z_j) \in RV(\mathbb{R}^n_+ \setminus \{0\}, \{b(t) = t\}, \nu)$. Consider the Y_i 's as in (4.9) and assume that $\forall j \in 1, \ldots, n, \exists k_j \ni j \in I_{k_j}$. Then, for all weights $w_i > 0$, with c_w as in (4.10), we have

$$\lim_{t \to \infty} t \mathbb{P}[Y_w/c_w > t] \le 1.$$

Moreover, in the last relation we have an equality if and only if Z_j , $j = 1, \dots, n$ are asymptotically independent.

This result entails that the max-linear Fréchet combination test is always honest under the mild assumption of multivariate regular variation of the vector $Z = (Z_j)_{j=1}^n$. Moreover, the test is asymptotically calibrated if and only if the Z_j 's are asymptotically independent.

Proof of Theorem C.1. By Lemma 2.1, we have

$$\lim_{t \to \infty} t \mathbb{P}[Y_w/c_w > t] = \frac{1}{c_w} \int_{\Delta} \max_{j=1,\dots,n} a_w(j) u_j \sigma(du),$$

where

$$a_w(j) := \max_{i=1,\dots,d} w_i \frac{1}{|I_i|} 1_{I_i}(j),$$

and where $\sigma(du)$ is the angular measure on Δ associated with ν , such that

$$\int_{\Lambda} u_i \sigma(du) = 1, \quad i = 1, \dots, n.$$
(C.1)

Now,

$$\max_{j} a_{w}(j)u_{j} \leq \sum_{j=1}^{n} a_{w}(j)u_{j} \quad \forall u \in \Delta$$

$$\implies \int_{\Delta} \max_{j=1,\dots,n} a_{w}(j)u_{j}\sigma(du) \leq \sum_{j=1}^{n} a_{w}(j) \int_{\Delta} u_{j}\sigma(du) = \sum_{j=1}^{n} a_{w}(j) = c_{w}$$

$$\implies \lim_{t \to \infty} t\mathbb{P}[Y_{w}/c_{w} > t] = \frac{1}{c_{w}} \int_{\Delta} \max_{j=1,\dots,n} a_{w}(j)u_{j}\sigma(du) \leq 1 \tag{C.2}$$

This proves the desired inequality. Now to argue the condition for equality above, observe that equality holds iff

$$\max_{j} a_{w}(j)u_{j} = \sum_{j=1}^{n} a_{w}(j)u_{j} \quad \sigma - a.e, \text{ or}$$

$$\sigma \left(\left\{ u : \max_{j} a_{w}(j)u_{j} = \sum_{j=1}^{n} a_{w}(j)u_{j} \right\} \right) = \sigma(\Delta)$$
(C.3)

Take any $u \in \Delta$ such that (C.3) holds. Say $j_u = \operatorname{argmax}_j a_w(j)u_j$. Then, (C.3) implies

$$\sum_{j \neq j_u} a_w(j) u_j = 0 \implies \forall j \neq j_u, \ a_w(j) u_j = 0$$
 (C.4)

Now, recall that we assumed $\forall j \in 1, ..., n, \exists k_j \ni j \in I_{k_j}$, i.e., every index belongs to some subset I_j . Also, $w_i > 0 \ \forall i$. Thus,

$$a_w(j) = \bigvee_{i=1}^d \frac{w_i}{|I_i|} \mathbf{1}_{I_i}(j) > 0 \quad \forall \ j \in [n]$$

Hence, (C.4) implies

$$\forall i \neq j_u, u_i = 0$$

As $u \in \Delta$, $\sum_{j=1}^{n} u_j = 1 \implies u_{j_u} = 1$, or $u = e_{j_u}$, the j_u -th positive unit direction in \mathbb{R}^n . Thus,

$$\forall u \ni \max_{j} a_{w}(j)u_{j} = \sum_{j=1}^{n} a_{w}(j)u_{j}, u \in \{e_{i} : i = 1, \dots, n\}, \text{ or}$$
$$\{u : \max_{j} a_{w}(j)u_{j} = \sum_{j=1}^{n} a_{w}(j)u_{j}\} \subseteq \{e_{i} : i = 1, \dots, n\}$$

But, for equality to hold in (C.2), (C.3) holds $\sigma - a.e.$. Hence,

$$\sigma(\Delta) = \sigma\left(\left\{u : \max_{j} a_{w}(j)u_{j} = \sum_{j=1}^{n} a_{w}(j)u_{j}\right\}\right) \leq \sigma\left(\left\{e_{i} : i = 1, \dots, n\right\}\right) \leq \sigma(\Delta)$$

$$\implies \sigma\left(\left\{e_{i} : i = 1, \dots, n\right\}\right) = \sigma(\Delta) \text{ or } \sigma\left(\left\{e_{i} : i = 1, \dots, n\right\}^{c}\right) = 0$$

Thus, equality, or asymptotic calibration, holds iff the support of the angular measure σ is contained within the extremal points of Δ which is equivalent to saying that the limit measure ν has non-zero mass only on the (positive) axes of \mathbb{R}^n .

Now, for any $1 \le i < j \le n$, take $p \in [0,1]$ sufficiently large such that $F_{Z_i}^{-1}(p) = F_{Z_j}^{-1}(p) > 0$. Note that equality between the quantiles holds because both Z_i and Z_j are 1-Frétchet. Then,

$$\mathbb{P}\left(Z_{i} > F_{Z_{i}}^{-1}(p), Z_{j} > F_{Z_{j}}^{-1}(p)\right)$$

$$\leq \mathbb{P}\left(Z \in \mathbb{R}_{+}^{i-1} \times \left(F_{Z_{i}}^{-1}(p), \infty\right) \times \mathbb{R}_{+}^{j-i-1} \times \left(F_{Z_{j}}^{-1}(p), \infty\right) \times \mathbb{R}_{+}^{n-j}\right)$$
Let $t_{p} = F_{Z_{i}}^{-1}(p) = F_{Z_{j}}^{-1}(p) \implies \lim_{p \to 1^{-}} t_{p} = \infty$. Thus,
$$b(t_{p})\mathbb{P}\left(Z_{i} > F_{Z_{i}}^{-1}(p), Z_{j} > F_{Z_{j}}^{-1}(p)\right)$$

$$\leq b(t_{p})\mathbb{P}\left(\frac{Z}{t_{p}} \in \mathbb{R}_{+}^{i-1} \times (1, \infty) \times \mathbb{R}_{+}^{j-i-1} \times (1, \infty) \times \mathbb{R}_{+}^{n-j}\right)$$

$$\implies \lim_{p \to 1^{-}} b(t_{p})\mathbb{P}\left(Z_{i} > F_{Z_{i}}^{-1}(p), Z_{j} > F_{Z_{j}}^{-1}(p)\right)$$

$$\leq \lim_{p \to 1^{-}} b(t_{p})\mathbb{P}\left(\frac{Z}{t_{p}} \in \mathbb{R}_{+}^{i-1} \times (1, \infty) \times \mathbb{R}_{+}^{j-i-1} \times (1, \infty) \times \mathbb{R}_{+}^{n-j}\right)$$

$$= \nu\left(\mathbb{R}_{+}^{i-1} \times (1, \infty) \times \mathbb{R}_{+}^{j-i-1} \times (1, \infty) \times \mathbb{R}_{+}^{n-j}\right) = 0$$

Now since $Z_i's$ are standard 1-Frétchet,

$$\lim_{t \to \infty} b(t) \mathbb{P}\left(Z_j > t\right) = 1 \implies \lim_{p \to 1^-} b(t_p) \mathbb{P}\left(Z_j > F_{Z_j}^{-1}(p)\right) = 1 \text{ or}$$
$$b(t_p) \sim \left(\mathbb{P}\left(Z_j > F_{Z_j}^{-1}(p)\right)\right)^{-1} \text{ as } p \to 1-$$

Thus,

$$\lim_{p \to 1^{-}} b(t_p) \mathbb{P}\left(Z_i > F_{Z_i}^{-1}(p), Z_j > F_{Z_j}^{-1}(p)\right) = 0$$

$$\implies \Lambda(Z_i, Z_j) = \lim_{p \to 1^{-}} \frac{\mathbb{P}\left(Z_i > F_{Z_i}^{-1}(p), Z_j > F_{Z_j}^{-1}(p)\right)}{\mathbb{P}\left(Z_j > F_{Z_j}^{-1}(p)\right)} = 0$$

i.e., $Z_i^\prime s$ are asymptotically independent.

This proves that the support of ν concentrated on the axes implies Z is asymptotically independent. The other direction is proved by Proposition A.5. Thus, equality holds in (C.2) iff Z is asymptotically independent.