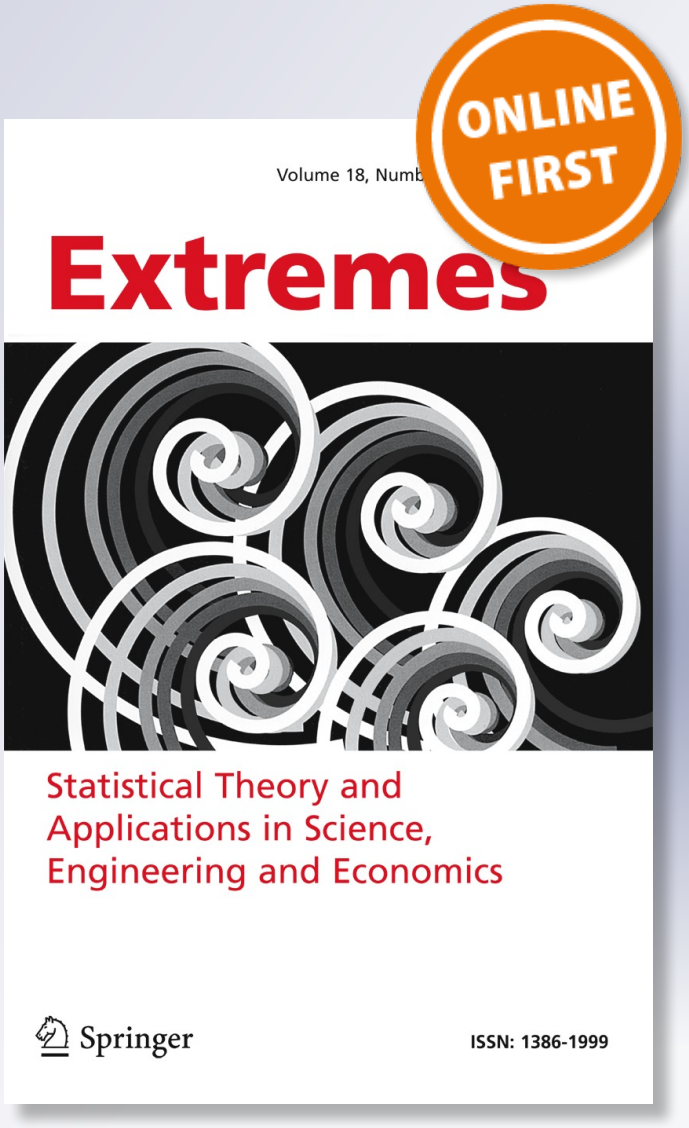
A measure of dependence for stable distributions

Uğur Tuncay Alparslan & John P. Nolan

Extremes
Statistical Theory and Applications in Science, Engineering and Economics

ISSN 1386-1999

Extremes
DOI 10.1007/s10687-015-0233-1



Springer

Extremes
DOI 10.1007/s10687-015-0233-1

A measure of dependence for stable distributions

Uğur Tuncay Alparslan[1] · **John P. Nolan**[1]

Abstract A distance based measure of dependence is proposed for stable distributions that completely characterizes independence for a bivariate stable distribution. Properties of this measure are analyzed, and contrasted with the covariation and codifference. A sample analog of the measure is defined and demonstrated on simulated and real data, including time series and distributions in the domain of attraction of a stable law.

Keywords Stable vectors · Dependence measure · Test statistic

AMS 2000 Subject Classifications MSC 60E07 · MSC 60G52 · MSC 62H20

1 Introduction

A d-dimensional random vector $\mathbf{X} = (X_1, \ldots, X_d)$ is said to be stable if for all $n = 2, 3, 4, \ldots$, there is a constant $a_n > 0$ and a vector $\mathbf{b}_n \in \mathbb{R}^d$ such that $\mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_n \stackrel{d}{=} a_n \mathbf{X} + \mathbf{b}_n$, where $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \ldots$ are i.i.d. copies of $\mathbf{X}$.

The second author was supported by an agreement with Cornell University, Operations Research & Information Engineering under W911NF-12-1-0385 from the Army Research Development and Engineering Command.

✉ John P. Nolan
jpnolan@american.edu

Uğur Tuncay Alparslan
tuncay.alparslan@american.edu

[1] Mathematics and Statistics, American University, Washington, DC 20016, USA

Published online: 20 January 2016

A multivariate stable distribution is usually described by a spectral measure Λ, a finite Borel measure on the unit sphere $\mathbb{S} = \{\mathbf{s} \in \mathbb{R}^d : |\mathbf{s}| = 1\}$, and a shift vector $\boldsymbol{\delta} \in \mathbb{R}^d$. There are multiple parameterizations of stable laws; we will use two, which we call the 0-parameterization and the 1-parameterization. We will say $\mathbf{X} \sim \mathbf{S}(\alpha, \Lambda, \boldsymbol{\delta}; j)$, $j = 0, 1$ if its joint characteristic function is given by

$$\phi(\mathbf{u}) = E \exp(i \langle \mathbf{u}, \mathbf{X} \rangle) = \exp\left(- \int_{\mathbb{S}} \omega \left(\langle \mathbf{u}, \mathbf{s} \rangle | \alpha; j \right) \Lambda(d\mathbf{s}) + i \langle \mathbf{u}, \boldsymbol{\delta} \rangle \right),$$

where

$$\omega(t|\alpha; j) = \begin{cases} |t|^\alpha [1 + i \, \text{sign}\,(t) \tan \frac{\pi\alpha}{2} (|t|^{1-\alpha} - 1)] & \alpha \neq 1, \, j = 0 \\ |t|^\alpha [1 - i \, \text{sign}\,(t) \tan \frac{\pi\alpha}{2}] & \alpha \neq 1, \, j = 1 \\ |t| [1 + i \, \text{sign}\,(t) \frac{2}{\pi} \log |t|] & \alpha = 1, \, j = 0, 1. \end{cases}$$

The 1-parameterization is more commonly used, but because $|\tan(\pi\alpha/2)| \to \infty$ as $\alpha \to 1$, the 1-parameterization is discontinuous in α. Since $\tan(\pi\alpha/2)(|t|^{1-\alpha} - 1) \to \frac{2}{\pi} \log |t|$ as $\alpha \to 1$, the 0-parameterization is a continuous parameterization of multivariate stable laws. If $\mathbf{X} \sim \mathbf{S}(\alpha, \Lambda, \boldsymbol{\delta}_0; 0)$ and $\mathbf{X} \sim \mathbf{S}(\alpha, \Lambda, \boldsymbol{\delta}_1; 1)$, then the shift vectors are related by

$$\boldsymbol{\delta}_1 = \begin{cases} \boldsymbol{\delta}_0 - \tan \frac{\pi\alpha}{2} \int_{\mathbb{S}} \mathbf{s} \, \Lambda(d\mathbf{s}) & \alpha \neq 1 \\ \boldsymbol{\delta}_0 & \alpha = 1. \end{cases}$$

Another way of describing a multivariate stable law is by the use of linear projections. If $\mathbf{X}$ is a stable vector, then every one dimensional projection $\langle \mathbf{u}, \mathbf{X} \rangle = u_1 X_1 + u_2 X_2 + \cdots + u_d X_d$ has a univariate stable distribution, with a constant index of stability α and skewness $\beta(\mathbf{u})$, scale $\gamma(\mathbf{u})$ and shift $\delta(\mathbf{u})$ that depend on the direction $\mathbf{u}$, see Samorodnitsky and Taqqu (1994), Section 2.1. (The converse is true if $\alpha \geq 1$; when $\alpha < 1$ an extra condition is needed for the converse, see the discussion after Lemma 4.1 of Nolan (2010).) We will call the functions $\beta(\cdot)$, $\gamma(\cdot)$ and $\delta(\cdot)$ the projection parameter functions. Since they uniquely determine all one dimensional projections, they determine the joint distribution via the Cramér-Wold device. In this case, we will parameterize $\mathbf{X}$ by these projection parameter functions: $\mathbf{X} \sim \mathbf{S}(\alpha, \beta(\cdot), \gamma(\cdot), \delta(\cdot); j)$, $j = 0, 1$. It is well known that $\gamma(\mathbf{u}) = \left(\int_{\mathbb{S}} |\langle \mathbf{u}, \mathbf{s} \rangle|^\alpha \Lambda(d\mathbf{s}) \right)^{1/\alpha}$ in both parameterizations. One fact we will note here is that these projection parameter functions have scaling properties, in particular $\gamma(r\mathbf{u}) = r\gamma(\mathbf{u})$, so knowing them on the unit sphere determines them everywhere. For the symmetric case, the joint characteristic function is $E \exp(i \langle \mathbf{u}, \mathbf{X} \rangle) = \exp(-\gamma^\alpha(\mathbf{u})) = \exp(-|\mathbf{u}|^\alpha \gamma^\alpha(\mathbf{u}/|\mathbf{u}|))$, so the values of $\gamma(\cdot)$ on $\mathbb{S}$ completely determines the joint distribution.

In Nolan (2010), the projection parameter functions were used to measure the distance between two multivariate stable distributions. Here we will adapt that idea to measure distance between a multivariate stable distribution and the stable law with independent components. Specifically, let $\mathbf{X} = (X_1, X_2)$ be a bivariate α-stable random vector. We assume that the components are normalized: $\gamma_j = \gamma(\mathbf{e}_j) = 1$, $j = 1, 2$, where $\mathbf{e}_j$ is the j-th standard unit vector. Set $\gamma_\perp(\mathbf{u}) = (|u_1|^\alpha + |u_2|^\alpha)^{1/\alpha}$; this is the scale function of any two-dimensional stable distribution having independent components and unit scales. (The distribution can be symmetric or skewed, so

there are multiple stable distributions that have this scale function. Specifying the skewness β_1 and β_2 for each component uniquely determines the joint distribution.) For any $p \in [1, \infty]$, define a measure of independence by

$$\eta_p = \eta_p(X_1, X_2) = \left\| \gamma^\alpha(u_1, u_2) - \gamma_\perp^\alpha(u_1, u_2) \right\|_{L^p(\mathbb{S}, d\mathbf{u})}. \tag{1}$$

Here $d\mathbf{u}$ is surface area on $\mathbb{S}$ (unnormalized, with total mass 2π). Figure 1 illustrates the geometric idea behind η_p.

The following simple result is the motivation for this definition.

Proposition 1 *Let* $\mathbf{X} = (X_1, X_2)$ *be an* $\alpha-$*stable random vector with normalized components,* $\alpha \in (0, 2)$. *Then* $\mathbf{X}$ *has independent components if and only if* $\eta_p = 0$ *for some (every)* $p \in [1, \infty]$.

Proof It is well-known that $\mathbf{X}$ has independent components if and only if the spectral measure is concentrated on the 4 points where the axes intersect the unit circle, e.g. Samorodnitsky and Taqqu (1994), Example 2.3.5. Let $\lambda_{j,+}$ be the spectral mass at $\mathbf{e}_j$ and $\lambda_{j,-}$ be the spectral mass at $-\mathbf{e}_j$, $j = 1, 2$. Then

$$\gamma^\alpha(\mathbf{u}) = |u_1|^\alpha(\lambda_{1,+} + \lambda_{1,-}) + |u_2|^\alpha(\lambda_{2,+} + \lambda_{2,-}).$$

Since $\gamma(\mathbf{e}_j) = \gamma_j = 1$, we must have $\gamma_j^\alpha = (\lambda_{j,+} + \lambda_{j,-}) = 1$, $j = 1, 2$. Thus $\gamma^\alpha(u_1, u_2) = |u_1|^\alpha + |u_2|^\alpha$, and therefore $\eta_p = 0$ for all p. For the converse, the scale function $\gamma(\cdot)$ is continuous, so the only way η_p can be 0 is if $\gamma^\alpha(u_1, u_2) = \gamma_\perp^\alpha(u_1, u_2)$. The following argument shows that the spectral measure is concentrated on $\{\mathbf{e}_1, -\mathbf{e}_1, \mathbf{e}_2, -\mathbf{e}_2\}$. Define $\Lambda_{\text{sym}}(A) = (\Lambda(A) + \Lambda(-A))/2$. This is a symmetric spectral measure with same scale function as Λ. By the uniqueness of the spectral measure, this means the spectral measure Λ_{sym} must be of the form $(\epsilon_{\mathbf{e}_1} + \epsilon_{-\mathbf{e}_1} + \epsilon_{\mathbf{e}_2} + \epsilon_{-\mathbf{e}_2})/2$, the spectral measure of the symmetric stable r. vector with the above

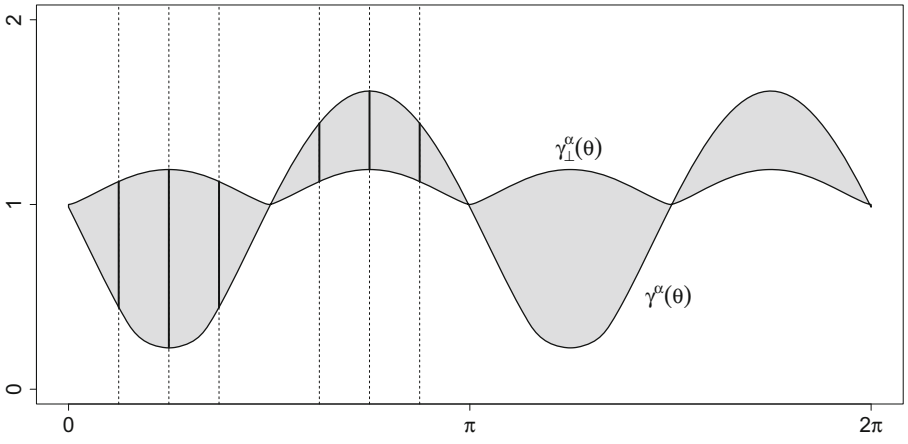


Fig. 1 η_1 is the area between the curves $\gamma^\alpha(\cdot)$ and $\gamma_\perp^\alpha(\cdot)$. The *vertical lines* are discussed when the sample estimator $\widehat{\eta}_p$ is defined below

scale function. By the definition of Λ_{sym}, the original Λ must also be concentrated on $\pm\mathbf{e}_j$, $j = 1, 2$. Thus $\mathbf{X}$ has independent components. □

The idea of a distance based measure of dependence is not new. Hoeffding (1948) and Blum et al. (1961) used distance measures based on empirical cumulative distribution functions, while Csörgő (1985), Székely et al. (2007), and Székely and Rizzo (2009) proposed similar measures based on empirical characteristic functions. In Section 3 we will provide a performance comparison of Brownian distance covariance of Székely and Rizzo (2009) and the measure proposed in this paper. The former has slightly better power than the latter, however the measure being proposed here has advantages in the stable distribution context. First, not only is $\eta_p = 0$ a characterization of independence, but the magnitude of η_p provides a concrete measure of closeness between the density of $\mathbf{X}$ and the independent density. (See Property 10 below for details.) Second, computing the exact (not sample) distance covariance between a given stable distribution and the one with independent components is complicated: it involves integrating over the whole plane the differences between two complex valued characteristic functions. In contrast, the η_p proposed here is simpler: evaluate the L^p distance between two real bounded functions on a bounded interval. And third, the sample version of η_p discussed in Section 3 requires estimating the scale function in several directions. If the hypothesis of independence is rejected, the pattern in those scale estimates gives information about the type of dependence. For example, if the estimated scale functions are roughly the same, this is evidence that the joint distribution is close to isotropic, i.e. the actual spectral measure is closer to a uniform measure than the discrete spectral measure concentrated on the points (1,0), (0,1), (-1,0) and (0,-1).

Here we list several comments about η_p and its properties.

1. The p-norm in Eq. 1 is evaluated as an integral over the unit circle $\mathbb{S}$, not all of $\mathbb{R}^2$. In polar coordinates,

$$\eta_p = \left(2\int_0^{\pi} \left|\gamma^{\alpha}(\cos\theta, \sin\theta) - \gamma_{\perp}^{\alpha}(\cos\theta, \sin\theta)\right|^p d\theta\right)^{1/p}, \tag{2}$$

   where the interval of integration has been reduced by half using the fact that $\gamma(\cdot)$ is π-periodic. Below we will use polar notation for the scale function on the unit circle when it is more convenient: $\gamma(\theta) := \gamma(\cos\theta, \sin\theta)$. In particular, $\gamma_{\perp}^{\alpha}(\theta) = |\cos\theta|^{\alpha} + |\sin\theta|^{\alpha})$.
2. α can be any value in $(0, 2)$ and $\mathbf{X}$ can have symmetric or non-symmetric components, and it can be centered or shifted.
3. η_p is symmetric: $\eta_p(X_1, X_2) = \eta_p(X_2, X_1)$.
4. $\eta_p \geq 0$ by definition, so as with other distance based measures of dependence, there is no notion of positive or negative dependence. Some authors have defined a signed measure of dependence, e.g. the signed covariation of Garel and Kodia (2014), but there is an arbitrariness with the sign. While this assignment may make sense in some cases, e.g. elliptically contoured stable, it doesn't seem to make sense in general. For example, a stable distribution that is a rotation by $\pi/4$ of the bivariate independent component case has probability

concentrated along both lines $y = x$ and $y = -x$; we do not see a meaningful way of assigning a sign to such dependence.

5. The definition makes sense in the Gaussian case: when $\alpha = 2$, the scale function for a bivariate Gaussian distribution with standardized marginals and correlation ρ is $\gamma^2(\mathbf{u}) = 1 + 2\rho u_1 u_2$ and $\gamma_\perp^2(\mathbf{u}) = 1$ on $\mathbb{S}$. Then $\eta_p^p = |2\rho|^p \int_{\mathbb{S}} |u_1 u_2|^p d\mathbf{u}$, so $\eta_p = k_p|\rho|$.
6. In the elliptically contoured/sub-Gaussian stable case with $0 < \alpha < 2$, and shape matrix $(1, \rho; \rho, 1)$, $\gamma^\alpha(\mathbf{u}) = (u_1^2 + 2\rho u_1 u_2 + u_2^2)^{\alpha/2}$, so

$$\eta_p^p = 2\int_0^\pi \left|(1 + 2\rho\cos\theta\sin\theta)^{\alpha/2} - (|\cos\theta|^\alpha + |\sin\theta|^\alpha)\right|^p d\theta.$$

This can be computed numerically, see Fig. 2 for a plot of η_2 in the elliptical case.

7. If the components of $\mathbf{X}$ are not standardized, then define

$$\eta_p(X_1, X_2) = \left\| \gamma^\alpha\left(\frac{u_1}{\gamma_1}, \frac{u_2}{\gamma_2}\right) - \gamma_\perp^\alpha(u_1, u_2) \right\|_{L^p(\mathbb{S}, d\mathbf{u})}.$$

$\gamma(u_1/\gamma_1, u_2/\gamma_2)$ is the scale function of the scaled random vector $\mathbf{Y} = (X_1/\gamma_1, X_2/\gamma_2)$. This normalizes the measure of independence.

8. A stable random vector $\mathbf{X} = (X_1, \ldots, X_d)$ has mutually independent components if and only if all pairs are independent, e.g. Corollary 3.5.4 in Samorodnitsky and Taqqu (1994). Hence the components of $\mathbf{X}$ are mutually independent if and only if $\eta_p(X_i, X_j) = 0$ for all $i > j$.
9. Since $L^p(\mathbb{S}, d\mathbf{s})$ is a finite measure space, Holder's Inequality shows for $1 \le p \le \infty$, $\eta_1 \le (2\pi)^{1+1/p}\eta_p$. Also, Section 2 below shows $|\gamma^\alpha(\mathbf{u}) - \gamma_\perp^\alpha(\mathbf{u})| \le 2$, so for any $q > 1$, $\eta_q = \left(\int |\gamma^\alpha(\mathbf{u}) - \gamma_\perp^\alpha(\mathbf{u})|^q\right)^{1/q} \le \left(2^{q-1}\int |\gamma^\alpha(\mathbf{u}) - \gamma_\perp^\alpha(\mathbf{u})|\right)^{1/q} = 2^{1-1/q}\eta_1^{1/q}$. Hence, η_p is small for some p

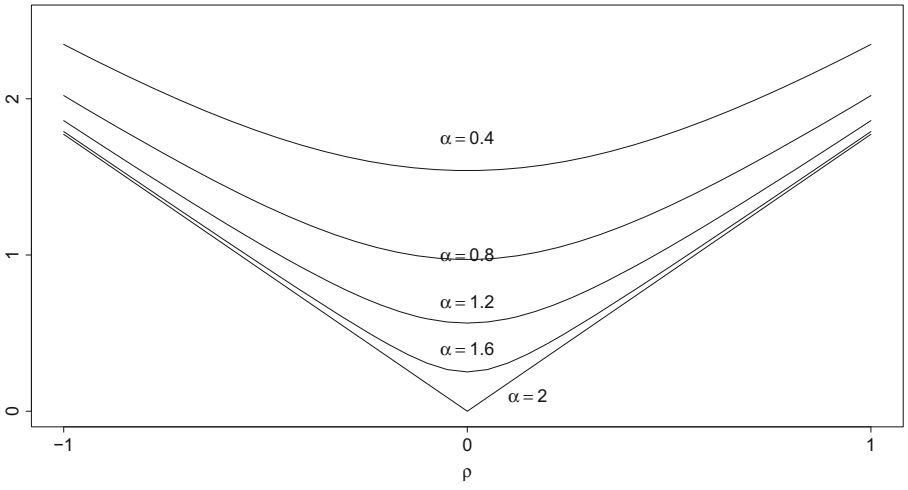


Fig. 2 Plot of η_2 for α-stable elliptically contoured stable distribution with shape parameter ρ

if and only if it is small for all q. We will focus on the case $p = 2$ below, particularly in the sample analog $\widehat{\eta}_2$.

10. η_p measures how far the scale function of $\mathbf{X}$ is from the scale function of a stable random vector with independent components. When $\mathbf{X}$ is symmetric, not only is $\eta_p = 0$ a characterization of independence, but more generally the size of η_p is a measure of closeness between $\mathbf{X}$ and the independent case. Let $f(\cdot)$ be the density of $\mathbf{X}$ and let $f_\perp(\cdot)$ be the density of the independent components case (both having normalized components), then Nolan (2010) shows

$$\sup_{\mathbf{x}\in\mathbb{R}^2} |f(\mathbf{x}) - f_\perp(\mathbf{x})| \leq k_\alpha \|\gamma(\mathbf{u}) - \gamma_\perp(\mathbf{u})\|_1.$$

When $\alpha \in (0, 1]$, the right hand side can be bounded by η_p. To see this, Section 2 shows $\gamma(\mathbf{u})$ and $\gamma_\perp(\mathbf{u})$ are in a bounded interval $[0, R_\alpha]$. For α in the range (0,1] and $s, t \in [0, R_\alpha]$, $|s - t| \leq (R_\alpha^{1-\alpha}/\alpha)|s^\alpha - t^\alpha|$, so

$$\begin{aligned}\|\gamma(\mathbf{u}) - \gamma_\perp(\mathbf{u})\|_1 &= \int |\gamma(\mathbf{u}) - \gamma_\perp(\mathbf{u})| d\mathbf{s} \\ &\leq c_\alpha \int |\gamma^\alpha(\mathbf{u}) - \gamma_\perp^\alpha(\mathbf{u})| d\mathbf{s} = c_\alpha \eta_1 \leq c'_\alpha \eta_p.\end{aligned}$$

When $1 < \alpha \leq 2$, a similar result can be shown by modifying the arguments in Nolan (2010). In the non-symmetric case, showing that the respective densities $f(\mathbf{x})$ and $f_\perp(\mathbf{x})$ are close requires an additional condition, i.e. that $\|\beta(\mathbf{u}) - \beta_\perp(\mathbf{u})\|_1$ is also small.

11. The same idea can be used to compare the distribution given by $\gamma^\alpha(\cdot)$ to a different model, other than the independent one, by replacing $\gamma_\perp^\alpha(\cdot)$ with the appropriate scale function. For example, using $\gamma_{iso}^\alpha(\cdot) = 1$ in place of $\gamma_\perp^\alpha(\cdot)$ in Eq. 1 would measure distance from the isotropic distribution. Modification of the sample measure $\widehat{\eta}_2$ in Section 3 is straightforward.

The next section examines the scale function, the proposed measure η, and compares with covariation and co-difference. In the third section, a sample analogue is defined and demonstrated with simulated and real data with bivariate and multivariate data, time series and a modification is given for vectors in the domain of attraction of a stable law.

2 Properties of $\gamma^\alpha(\cdot)$, η_p, covariation, and co-difference

In multivariate extreme value theory, Pickands (1981) defined a function $A(\cdot)$ that characterizes the joint distributon in a bivariate extreme value distribution. The Pickands function is convex and satisfies the bound: $\max(t, 1 - t) \leq A(t) \leq 1$, $0 \leq t \leq 1$. Below we seek corresponding bounds for the stable scale function $\gamma^\alpha(\cdot)$, and then explore previous measures of dependence for stable distributions.

Let $\mathbf{X}$ be a bivariate α-stable random vector with spectral measure Λ and scale function $\gamma(\cdot)$. Throughout this section we assume that $\mathbf{X}$ has normalized components:

$\gamma_1 = \gamma_2 = 1$. As above, let $\gamma(\theta) = \gamma(\cos\theta, \sin\theta)$, $\theta \in [0, 2\pi]$. In polar coordinates,

$$\gamma^\alpha(\theta) = \int_0^{2\pi} |\langle(\cos\theta, \sin\theta), (\cos\phi, \sin\phi)\rangle|^\alpha \Lambda(\phi) = \int_0^{2\pi} |\cos(\theta - \phi)|^\alpha \Lambda(\phi).$$

First, we seek envelope functions:

$$\gamma^\alpha_{\min}(\theta) := \inf_\gamma \gamma^\alpha(\theta) \quad \text{and} \quad \gamma^\alpha_{\max}(\theta) := \sup_\gamma \gamma^\alpha(\theta),$$

where the inf and sup are taken over all valid scale functions for a bivariate $\alpha-$stable random vector with normalized components. Proposition 2 gives an explicit formula for $\gamma^\alpha_{\max}(\theta)$ and Proposition 3 gives a candidate for $\gamma^\alpha_{\min}(\theta)$; both depend only on α.

Proposition 2 *For normalized components, we have* $0 \le \gamma^\alpha(\theta) \le \gamma^\alpha_{\max}(\theta)$, *where*

$$\gamma^\alpha_{\max}(\theta) = \begin{cases} (|\cos\theta| + |\sin\theta|)^\alpha & 1 \le \alpha < 2 \\ |\cos\theta|^\alpha + |\sin\theta|^\alpha & 0 < \alpha < 1. \end{cases}$$

Proof Since the distribution has normalized components, $\gamma(1, 0) = \gamma(0, 1) = 1$. The proof is easier using rectangular coordinates: write $\mathbf{u} \in \mathbb{S}$ as $(u_1, u_2) = (\cos\theta, \sin\theta)$. When $0 < \alpha \le 1$,

$$\begin{aligned} \gamma^\alpha(u_1, u_2) &= \int_{\mathbb{S}} |u_1 s_1 + u_2 s_2|^\alpha \Lambda(d\mathbf{s}) \\ &\le \int_{\mathbb{S}} |u_1 s_1|^\alpha \Lambda(d\mathbf{s}) + \int_{\mathbb{S}} |u_2 s_2|^\alpha \Lambda(d\mathbf{s}) \\ &= |u_1|^\alpha \gamma^\alpha(1, 0) + |u_2|^\alpha \gamma^\alpha(0, 1) = |u_1|^\alpha + |u_2|^\alpha = \gamma^\alpha_\perp(u_1, u_2) \end{aligned}$$

When $\alpha \ge 1$, using the triangle inequality,

$$\begin{aligned} \gamma(u_1, u_2) &= \left(\int_{\mathbb{S}} |u_1 s_1 + u_2 s_2|^\alpha \Lambda(d\mathbf{s}) \right)^{1/\alpha} \\ &\le \left(\int_{\mathbb{S}} |u_1 s_1|^\alpha \Lambda(d\mathbf{s}) \right)^{1/\alpha} + \left(\int_{\mathbb{S}} |u_2 s_2|^\alpha \Lambda(d\mathbf{s}) \right)^{1/\alpha} \\ &= |u_1|\gamma(1, 0) + |u_2|\gamma(0, 1) = |u_1| + |u_2| \end{aligned}$$

The expressions for $\gamma_{\max}(\cdot)$ are sharp. To see this, set

$$\begin{aligned} \gamma_+(\theta) &= 2^{1/2} |\cos(\theta - \pi/4)| \\ \gamma_-(\theta) &= 2^{1/2} |\cos(\theta - 3\pi/4)| \end{aligned}$$

Note that $\gamma_+(\cdot)$ corresponds to a bivariate stable distribution with exact positive linear dependence ($X_2 = X_1$, spectral measure with mass of weight $2^{\alpha/2}$ on the diagonal) and $\gamma_-(\cdot)$ corresponds to exact negative linear dependence ($X_2 = -X_1$, mass of weight $2^{\alpha/2}$ on the anti-diagonal). When $\alpha \ge 1$, $|\cos\theta| + |\sin\theta| = \max(\gamma_-(\theta), \gamma_+(\theta))$, so the upper bound is achieved by $\gamma_-(\theta)$ in the first and third quadrants and $\gamma_+(\theta)$ in the second and fourth quadrants. When $\alpha \le 1$, the upper bound is achieved by the independent component case. □

We conjecture that a sharp lower bound $\gamma^{\alpha}_{\min}(\theta)$ is given by

$$\gamma^{\alpha}_{*}(\theta) = \begin{cases} \max(g(\theta), g(\pi/2-\theta)) & \alpha < 1 \\ \min(\gamma^{\alpha}_{-}(\theta), \gamma^{\alpha}_{+}(\theta)) & \alpha \geq 1, \end{cases}$$

where $g(\theta) = |\cos\theta|^{\alpha}(1 - |\tan\theta|^{\alpha})$. The following result shows that $\gamma^{\alpha}_{\min}(\cdot)$ is less than or equal to $\gamma^{\alpha}_{*}(\cdot)$.

Proposition 3 *For normalized components,* $0 \leq \gamma^{\alpha}_{\min}(\theta) \leq \gamma^{\alpha}_{*}(\theta)$.

Proof Consider the $\alpha < 1$ case. First assume θ_0 is in the interval $(0, \pi/4)$. For $\lambda_1 > 0$, define $\theta_1 = \theta_0 + \pi/2$ and $\theta_2 = \arctan\{[(1 - \lambda_1|\cos\theta_0|^{\alpha})/(1-\lambda_1|\sin\theta_0|^{\alpha})]^{-1/\alpha}\}$, and $\lambda_2 = (1-\lambda_1|\cos\theta_0|^{\alpha})/|\sin\theta_2|^{\alpha}$. When $0 < \lambda_1 < \lambda^* := 1/|\cos\theta_0|^{\alpha}$, $\lambda_2 > 0$, and calculation shows the bivariate stable distribution $\mathbf{X}$ having the two point spectral measure with mass λ_j at θ_j has standardized components. For large λ_1, most of the mass is at θ_1, which is perpendicular to θ_0; the second point mass is placed to get normalized components. The corresponding scale function is $\gamma^{\alpha}(\phi) = |\cos(\phi-\theta_1)|^{\alpha}\lambda_1 + |\cos(\phi-\theta_2)|^{\alpha}\lambda_2$. So $\gamma^{\alpha}(\theta_0) = 0\cdot\lambda_1 + |\cos(\theta_0-\theta_2)|^{\alpha}\lambda_2$. As $\lambda_1 \uparrow \lambda^*$, more calculations show that $\gamma^{\alpha}(\theta_0) \downarrow g^{\alpha}(\theta_0)$. For θ_0 in the interval $(\pi/4, \pi/2)$, the argument can be reflected to get the $g(\pi/2-\theta_0)$ bound. For the rest of the interval $(\pi/2, 2\pi)$, the argument can be shifted from the first quadrant.

When $\alpha \geq 1$, $\gamma^{\alpha}_{*}(\cdot)$ is achieved by $\gamma^{\alpha}_{-}(\cdot)$ in the first and third quadrants, and by $\gamma^{\alpha}_{+}(\cdot)$ in the second and fourth quadrants. □

For any scale function with normalized components, $\gamma(\theta) \leq \gamma_{\max}(\pi/4) = 2^{1/2}$ when $1 \leq \alpha \leq 2$, and $\gamma(\theta) \leq \gamma_{\max}(\pi/4) = 2^{1/\alpha-1/2}$ when $0 < \alpha \leq 1$. This latter term is unbounded as $\alpha \downarrow 0$. However $\gamma^{\alpha}(\theta) \leq 2$ for all α and all θ; this is why we used the α-th power in the definition of η_p. Figure 3 shows the upper envelope function $\gamma^{\alpha}_{\max}(\cdot)$ and the proposed lower envelope $\gamma^{\alpha}_{*}(\cdot)$.

Unlike the Pickands function, convexity is not necessary for $\gamma^{\alpha}(\cdot)$ to be a valid scale function. $\gamma^{\alpha}(\cdot)$ must be of positive type, or equivalently, $\exp(-\gamma^{\alpha}(\cdot))$ must be non-negative definite. We are not aware of any intrinsic characterization of a function $\gamma^{\alpha}(\cdot)$ that guarantees this.

Since $\gamma^{\alpha}(\cdot) \leq 2$ for any scale function with normalized components, η_p is always bounded by 4π, but this bound is not sharp. Numerical experiments lead us to conjecture that for any $0 < \alpha \leq 2$, any $p \geq 1$, and any normalized bivariate α-stable distribution (X_1, X_2), $0 \leq \eta_p(X_1, X_2) \leq \eta^*_p$, where

$$\eta^*_p = \left\|\gamma^{\alpha}_{+}(\theta) - \gamma^{\alpha}_{\perp}(\theta)\right\|_p = \left\|\gamma^{\alpha}_{-}(\theta) - \gamma^{\alpha}_{\perp}(\theta)\right\|_p.$$

In words, the farthest distributions away from independence are the exact linear dependent cases. These values of η^*_p can be compute numerically.

We briefly compare η_p to covariation and co-difference. For $\alpha > 1$, the first equality below is the definition of covariation, and Example 2.7.3 in Samorodnitsky and Taqqu (1994) shows the second equality:

$$[X_1, X_2]_{\alpha} = \int_{\mathbb{S}} s_1 s_2^{<\alpha-1>} \Lambda(d\mathbf{s}) = \frac{1}{\alpha} \left.\frac{\partial \gamma^{\alpha}(u_1, u_2)}{\partial u_1}\right|_{(u_1=0, u_2=1)}. \tag{3}$$

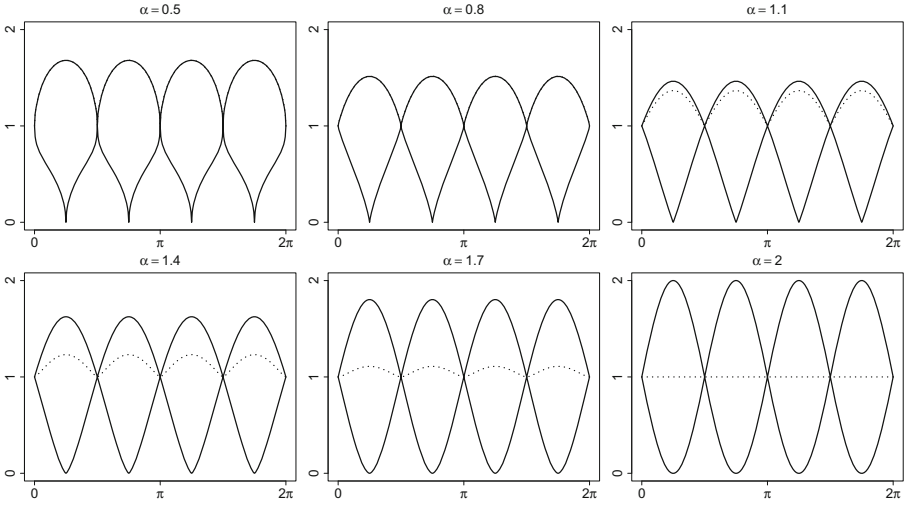


Fig. 3 Envelope functions $\gamma^{\alpha}_{\max}(\theta)$ (*top curves*) and conjectured $\gamma^{\alpha}_{\min}(\theta)$ (*bottom curves*) for different α. For comparison to the independent case, $\gamma^{\alpha}_{\perp}(\theta)$ is also shown as a dotted curve; when $\alpha \leq 1$, the independent scale function is identical to the upper bound

Thus the covariation depends only on the behavior of $\gamma^{\alpha}(\cdot, \cdot)$ near the point $(0, 1)$. If X_1 and X_2 are independent, then $[X_1, X_2]_{\alpha} = 0$. The discussion below shows that converse is false: there are many dependent distributions where $[X_1, X_2]_{\alpha} = 0$.

The co-difference is defined for symmetric α-stable vectors, and can be written as

$$\tau = \tau(X_1, X_2) = \gamma^{\alpha}(1, 0) + \gamma^{\alpha}(0, 1) - \gamma^{\alpha}(1, -1),$$

and is defined for any $\alpha \in (0, 2)$. It uses the values of the scale function at three particular points. If X_1 and X_2 are independent, then $\tau = 0$; when $\alpha \leq 1$, the converse is also true, see Section 2.10 of Samorodnitsky and Taqqu (1994). When $\alpha > 1$, the converse is false, however if both $\tau(X_1, X_2) = 0$ and $\tau(X_1, -X_2) = 0$, then X_1 and X_2 are independent, see Section 2 of Rosiński and Zak (1997). A direct proof of this is straightforward: the condition $\tau(X_1, X_2) = 0$ is equivalent to

$$\int_{\mathbb{S}} |s_1|^{\alpha} \Lambda(d\mathbf{s}) + \int_{\mathbb{S}} |s_1|^{\alpha} \Lambda(d\mathbf{s}) = \int_{\mathbb{S}} |s_1 - s_2|^{\alpha} \Lambda(d\mathbf{s}),$$

and the condition $\tau(X_1, -X_2) = 0$ is equivalent to

$$\int_{\mathbb{S}} |s_1|^{\alpha} \Lambda(d\mathbf{s}) + \int_{\mathbb{S}} |s_1|^{\alpha} \Lambda(d\mathbf{s}) = \int_{\mathbb{S}} |s_1 + s_2|^{\alpha} \Lambda(d\mathbf{s}).$$

By Lemma 2.7.14 (2) in Samorodnitsky and Taqqu (1994), these two conditions together are equivalent to $s_1 s_2 = 0$ Λ-a.e., i.e. X_1 and X_2 are independent. As a consequence, one can define $\tau^2(X_1, X_2) + \tau^2(X_1, -X_2)$ as a measure that characterizes dependence. However the sample analog of this has less power then the sample analog of η_2 proposed below.

Chapter 2 of Samorodnitsky and Taqqu (1994) gives properties of covariation and co-difference. Here we elaborate some on these properties, with attention to when they can be zero. Let Q_j be the open j-th quadrant, $j = 1, 2, 3, 4$.

For the covariation we restrict to $\alpha > 1$. The integrand in Eq. 3 is zero precisely at the "poles" $(s_1, s_2) = (1, 0), (0, 1), (-1, 0), (0, -1)$, which is precisely the support of any independent stable vector. So independence implies 0 covariation. But there are many other ways this can happen. The integrand function above is strictly positive on $Q_1 \cup Q_3$, and strictly negative on $Q_2 \cup Q_4$. Hence, if Λ is supported on $Q_1 \cup Q_3$, $[X_1, X_2]_\alpha > 0$; if Λ is supported on $Q_2 \cup Q_4$, $[X_1, X_2]_\alpha < 0$. Furthermore, if support Λ satisfies either condition, then normalizing the components of the corresponding distribution also has support in the corresponding region. So, there are many distributions with positive covariation and many with negative covariation, even if we restrict to normalized components. Furthermore, the integral definition shows that covariation is linear in the spectral measure: for $c_0, c_1 \geq 0$, the covariation corresponding to $c_0\Lambda_0 + c_1\Lambda_1$ is the sum of c_0 times the covariation corresponding to Λ_0 plus c_1 times the covariation corresponding to Λ_1. Thus if Λ_0 has positive covariation and Λ_1 has negative covariation, setting $c_j = |\text{covariation of } \Lambda_{1-j}| \,/(|\text{covariation of } \Lambda_0| + |\text{covariation of } \Lambda_1|)$, $c_0\Lambda_0 + c_1\Lambda_1$ has covariation 0. Since this is a convex combination, if Λ_0 and Λ_1 have normalized components, so does the sum. Hence, there are many normalized dependent distributions with zero covariation, much like in the case of ordinary correlation, where there are important classes of models in which $\rho = 0$ does not imply independence.

We conclude the discussion of covariation by showing that the range of the covariation is [-1,1] when (X_1, X_2) are jointly stable with unit scales. This follows by combining (a) $|[X_1, X_2]_\alpha| \leq 1$ by Property 2.8.4 of Samorodnitsky and Taqqu (1994); (b) the bounds are achieved: straightforward calculation shows that exact positive dependence give covariation +1 and exact negative correlation gives covariation -1; and (c) linearity of covariation in the spectral measure.

For the co-difference, any $\alpha \in (0, 2]$ is allowed. Scaling shows $\tau = \gamma^\alpha(1, 0) + \gamma^\alpha(0, 1) - 2^{\alpha/2}\gamma^\alpha(1/\sqrt{2}, -1/\sqrt{2})$, so it suffices to consider $\gamma(\cdot)$ on the unit circle. Expressing the spectral measure in polar coordinates also gives a polar expression for τ:

$$\begin{aligned}\tau &= \gamma^\alpha(0) + \gamma^\alpha(\pi/2) - 2^{\alpha/2}\gamma^\alpha(3\pi/4) \\ &= \int_0^{2\pi} \left[|\cos\phi|^\alpha + |\sin\phi|^\alpha - 2^{\alpha/2}|\cos(3\pi/4 - \phi)|^\alpha \right] \Lambda(d\phi). \end{aligned} \tag{4}$$

The integrand above is 0 at the "poles", so independence implies 0 co-difference. Some special cases are straightforward to compute. In the elliptical case, $\gamma_{\text{ellip}}(\mathbf{u}) = 1 + 2\rho u_1 u_2$, so $\tau_{\text{ellip}} = 1 + 1 - (1 + 2\rho(-1)) = 2 - (2(1 - \rho))^{\alpha/2}$. The range of τ_{ellip} is $[2 - 2^\alpha, 2]$; the lower bound is negative if and only if $\alpha > 1$. In the isotropic case, $\rho = 0$ and $\tau = 2 - 2^{\alpha/2}$. For $\gamma_+(\cdot)$, $\tau_+ = 1 + 1 - 2^{\alpha/2}|\cos\pi/2|^\alpha = 2$. This achieves max. For $\gamma_-(\cdot)$, $\tau_- = 1 + 1 - 2^{\alpha/2}2^{\alpha/2}|\cos 0|^\alpha = 2 - 2^\alpha$. This is positive for $\alpha < 1$, zero for $\alpha = 1$, and negative for $\alpha > 1$.

When $\alpha > 1$, the integrand function in Eq. 4 is strictly positive in $Q_1 \cup Q_3$ and strictly negative in $Q_2 \cup Q_4$. So, as with the covariation above, any spectral

measure concentrated in $Q_1 \cup Q_3$ has strictly positive co-difference, and any spectral measure concentrated in $Q_2 \cup Q_4$ has strictly negative co-difference. As above, if Λ satisfies either of the conditions on the support, then normalizing the coordinates gives a spectral measure with support that satisfies the same condition.

We can also follow the argument above for covariation of sums to co-difference. Here we use notation τ_Λ for the co-difference for (X_1, X_2) having spectral measure Λ. It is simple to see that for spectral measures Λ_0 and Λ_1 and non-negative c_0, c_1, $\tau_{c_0\Lambda_0+c_1\Lambda_1} = c_0\tau_{\Lambda_0} + c_1\tau_{\Lambda_1}$. In particular, if $\tau_{\Lambda_0} < 0$, $\tau_{\Lambda_1} > 0$ and $c_j = |\tau_{\Lambda_{1-j}}|/(|\tau_{\Lambda_0}| + |\tau_{\Lambda_1}|)$, then $\tau_{c_0\Lambda_0+c_1\Lambda_1} = 0$. This shows that for a fixed α, the range of τ is an interval. Combined with the previous example, this gives many examples with $\alpha > 1$ where $\tau = 0$, but X_1 and X_2 are dependent. If Λ_0 and Λ_1 have unit scales, then so does the convex combination $c_0\Lambda_0 + c_1\Lambda_1$.

When $\alpha \leq 1$, the situation is different. In this case, the integrand in Eq. 4 is strictly positive off the "poles", so $\tau \geq 0$. The only way τ can be zero is therefore when the components are independent.

Combining the results above, we have shown that for a fixed α, $\min(0, 2 - 2^\alpha) \leq \tau \leq 2$. For $\alpha \geq 1$, all values in this region can be achieved by an elliptical stable law with some ρ. For $\alpha < 1$, all values can be achieved with spectral measures the sum of an independent and exact positive independence. Most values of τ can be achieved by many different distributions.

3 Measuring dependence in a bivariate sample

3.1 Estimation of η and the significance level

In this section we consider the statistical problem of determining the independence of the components of a bivariate stable law using a sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ of α-stable vectors. We start by normalizing the data: estimate the parameters $(\widehat{\alpha}_j, \widehat{\beta}_j, \widehat{\gamma}_j, \widehat{\delta}_j)$ of each component, $j = 1, 2$ using maximum likelihood as in Nolan (2001). Then use a pooled estimate of α: set $\widehat{\alpha} = (\widehat{\alpha}_1 + \widehat{\alpha}_2)/2$. To avoid numerical problems around $\alpha = 1$, we will only consider the 0-parameterization in this section. Then normalize the data $\mathbf{Y}_i = ((X_{i,1} - \delta_1)/\gamma_1, (X_{i,2} - \delta_2)/\gamma_2)$, $i = 1, \ldots, n$. Since MLE is used to estimate $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$, standard convergence results for maximum likelihood estimators apply to $\widehat{\alpha}$ and to $\widehat{\gamma}_j$ below.

We note that $\gamma(-\mathbf{u}) = \gamma(\mathbf{u})$, so as in Eq. 2 it suffices to restrict the θ_j's to be in the interval $[0, \pi]$. For a set of angles $0 \leq \theta_1 < \theta_2 < \cdots < \theta_m \leq \pi$, define $\widehat{\gamma}_j = \widehat{\gamma}(\cos\theta_j, \sin\theta_j) =$ scale of the projected data set $\langle \mathbf{Y}_i, (\cos\theta_j, \sin\theta_j)\rangle$, $i = 1, \ldots, n$ and $\gamma_{\perp,j}^{\widehat{\alpha}} = |\cos\theta_j|^{\widehat{\alpha}} + |\sin\theta_j|^{\widehat{\alpha}}$, $j = 1, \ldots, m$. We will discuss the choice of grid below.

A straightforward (unnormalized) sample approximation to $\eta_2(Y_1, Y_2)$ for the normalized data set is

$$\widehat{\eta}_2^2 = \widehat{\eta}_2^2(m, n) = \sum_{j=1}^{m} \left(\widehat{\gamma}_j^{\widehat{\alpha}} - \gamma_{\perp,j}^{\widehat{\alpha}}\right)^2. \tag{5}$$

We propose using this as a test statistic to determine the independence of a bivariate data set. The distribution of $\widehat{\eta}_2$ can be approximated using simulation.

For a formal test of independence, the null hypothesis is H_0: $\mathbf{X}$ has independent components. We compute the sample measure of independence $\widehat{\eta}_2$ and reject H_0 if $\widehat{\eta}_2 > c_\epsilon = c_\epsilon(\alpha, \beta_1, \beta_2, n, \theta_1, \ldots, \theta_m)$, where the ϵ is the desired significance level. An R program to compute $\widehat{\eta}_2$ and estimate c_ϵ has been written. The simulation is straightforward: for a single sample, independent terms can be simulated using the univariate simulation method of Chambers et al. (1976), and then the univariate parameters $(\widehat{\alpha}_j, \widehat{\beta}_j, \widehat{\gamma}_j, \widehat{\delta}_j)$ can be estimated using maximum likelihood and the sample can be normalized. Then $\widehat{\eta}_2$ is computed using Eq. 5. Repeating this M times gives a sample from the distribution of $\widehat{\eta}_2$ under the null hypothesis; the $(1-\epsilon)$ quantile of this is used for the critical value c_ϵ. The appendix gives an approximation to the critical values that is accurate for most practical values of the parameters, e.g. $0.5 \leq \alpha \leq 2, \epsilon \leq 0.3, n \geq 100, 1 \leq k \leq 10$.

We now discuss the choice of grid points $\theta_1, \ldots, \theta_m$. A uniform grid is not optimal: because of the normalizing, $\widehat{\gamma}(\mathbf{u}) = 1$ at $\mathbf{u} \in \{(1, 0), (0, 1), (-1, 0), (0, -1)\}$ so we get no information including those points in the grid. Furthermore, when angles θ_j and θ_k are close, the corresponding directional scales are correlated. Figure 4 shows a plot of the covariance surface in one case. The discussion in Section 2 shows that the maximum distance between $\gamma(\theta)$ and $\gamma_\perp(\theta)$ can occur at the angles $\pi/4$ and $3\pi/4$. A reasonable choice is to pick an integer k and then define $\theta_j = j\pi/(2k+2)$, $j = 1, \ldots, k$ and $\theta_{j+k} = \theta_j + \pi/2$. This gives a grid of length $m = 2k$ that is uniformly spread on the interior of the first and second quadrant. The vertical lines in Fig. 1 shows the $k = 3$ case. The choice of k is complicated for several reasons. There is little information gained by getting close to the points (1,0), (0,1) and (-1,0) where the scale is fixed at 1. As k increases, the estimators of $\widehat{\gamma}_j$ and $\widehat{\gamma}_i$ are

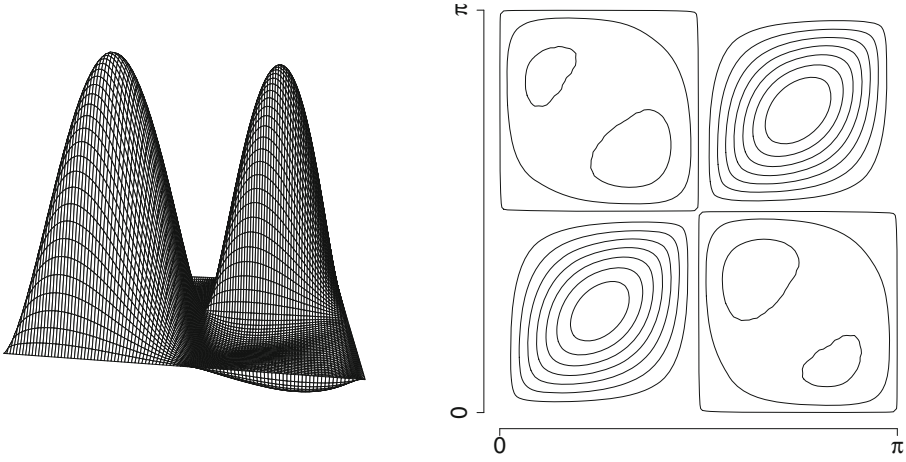


Fig. 4 Plot of the empirical covariance of $\gamma^\alpha(\cdot)$ for $\alpha = 1.5$, $\beta_1 = \beta_2 = 0$. The plot was generated by running $M = 20000$ simulations of sample size $n = 1000$ with independent components

highly correlated. Unless n is very large, the variability in the estimators $\widehat{\gamma}_j$ will outweigh the information gained by adding more grid points. Simulations discussed in the Appendix suggest $k = 5$ is a reasonable grid size for most practical applications.

Table 1 shows estimates of the power for detecting dependence for several types of dependence. The isotropic column is when the data is radially symmetric, the third, fourth and fifth columns are counterclockwise rotations of the independent spectral measure by the stated angle, the last column shows exact linear positive dependence (all the mass of the spectral measure on the diagonal line $y = x$). (The simulation method for the isotropic case uses the representation $\mathbf{X} = A^{1/2}\mathbf{Z}$, where A is a positive $(\alpha/2)$-stable term and $\mathbf{Z}$ is N$(0, I)$, see section 2.5 of Samorodnitsky and Taqqu (1994); the other cases can be simulated using Example 2.3.6 in Samorodnitsky and Taqqu (1994).)

The isotropic case is a modest departure from independence, and the table shows that to reliably detect this kind of dependence, sample sizes in the several hundreds are required. The third column shows that when the spectral measure is concentrated on the diagonal and anti-diagonal lines, $\widehat{\eta}_2$ reliably detects dependence with sample sizes on the order of 100 or more. As the rotation lessens, the corresponding distribution is closer and closer to independence, and larger samples are needed to reliably detect dependence. The last column shows that quite small samples of size $n = 25$ are sufficient to detect dependence when the data is concentrated on the diagonal line.

Further simulations were performed to compare the power of $\widehat{\eta}_2$ to the power of the distance covariance of Székely et al. (2007) and the classical correlation based test in the R package `energy`. Specifically, we examined what sample size is needed to get an empirical power of 0.95 with a significance level of $\epsilon = 0.05$ for the cases considered in Table 1. Since the last column of that table is always 1, even for samples of size as small as $n = 25$, we exclude that from our analysis. Table 2 shows the results. The first row is extracted from Table 1. For the other methods we ran simulations for various values of n and interpolation was used to get the entries in the table. For distance covariance, the same power can be achieved with a smaller sample size, with the same pattern of increasing sample size when the distribution is closer to independence. If Pearson, Spearman or Kendall correlation is used in the correlation test, sample sizes of over 50000 are needed in all cases (our simulations

Table 1 Power of $\widehat{\eta}_2$ for various types of dependence when $\alpha = 1.5$, $k = 5$, and $\beta_1 = \beta_2 = 0$

Sample size n	Isotropic	Independent $\circlearrowleft \pi/4$	Independent $\circlearrowleft \pi/8$	Independent $\circlearrowleft \pi/16$	Exact linear dependence
25	0.191	0.322	0.243	0.213	1
50	0.223	0.624	0.381	0.183	1
100	0.344	0.918	0.644	0.214	1
200	0.636	0.998	0.937	0.440	1
300	0.874	1	0.997	0.627	1
400	0.960	1	1	0.791	1
500	0.989	1	1	0.893	1

Table 2 Approximate sample size needed to achieve power 0.95 for $\widehat{\eta}_2$, distance covariance, and correlation test using different correlation measures. In all cases, $\epsilon = 0.05$, $\alpha = 1.5$, $k = 5$, and $\beta_1 = \beta_2 = 0$

Test	Isotropic	Independent $\circlearrowleft \pi/4$	Independent $\circlearrowleft \pi/8$	Independent $\circlearrowleft \pi/16$
$\widehat{\eta}_2$	400	150	225	600
Distance covariance	250	100	150	300
Correlation test	> 50000	> 50000	> 50000	> 50000

went up to $n = 50000$ and none of these test reached the desired power). Clearly, correlation tests have very low power for stable laws. If one is solely interested in deciding whether components are independent or not, then the distance covariance is more efficient and it has the advantage of not assuming joint stability. However, there is still value to the proposed method. First, one should be wary of working with small data sets when heavy tails are claimed, and with sample sizes in the several hundreds, both the proposed method and distance covariance have high power. Second, there are advantages discussed in the first section above: when independence is rejected, the values of $\widehat{\gamma}_j$ give information about the actual distribution. The distance covariance test does not provide any such information.

We now mention an alternative statistic. Since we are using maximum likelihood estimation of the parameters, they are asymptotically normal. Using the delta method, the $\widehat{\gamma}_j^{\widehat{\alpha}}$ are also asymptotically normal. We have been unable to derive an analytic expression for $\Sigma = \Sigma(\alpha, \beta_1, \beta_2, \theta_1, \ldots, \theta_m) = [\sigma_{i,j}]_{i,j=1}^m$, the covariance matrix of $\widehat{\gamma}_j^{\widehat{\alpha}}$, $j = 1, \ldots, m$. If an analytic form of this covariance was known, then a χ^2 type of statistic could be used:

$$\widehat{\eta}_{\Sigma,2}^2 = (\widehat{\boldsymbol{\gamma}}^{\widehat{\alpha}} - \boldsymbol{\gamma}_{\perp}^{\widehat{\alpha}})^T \, \Sigma^{-1} \, (\widehat{\boldsymbol{\gamma}}^{\widehat{\alpha}} - \boldsymbol{\gamma}_{\perp}^{\widehat{\alpha}}),$$

where $\widehat{\boldsymbol{\gamma}}^{\widehat{\alpha}} = (\widehat{\gamma}_1^{\widehat{\alpha}}, \ldots, \widehat{\gamma}_m^{\widehat{\alpha}})^T$ and $\boldsymbol{\gamma}_{\perp}^{\widehat{\alpha}} = (\gamma_{\perp,1}^{\widehat{\alpha}}, \ldots, \gamma_{\perp,m}^{\widehat{\alpha}})^T$. Under the null hypothesis, this is the square of the Mahalanobis distance between the data and a stable distribution with independent components, and has a $\chi^2(m)$ sampling distribution. This approach would have the advantage of using familiar tabulated critical values.

3.2 Application to financial data

Next we analyze some financial data. Closing price data on two pharmaceutical stocks, Pfizer (symbol PFE) and Merck (MRK), was gathered for the five year time period January 1, 2010 to December 31, 2014 resulting in 1257 prices. For each stock, log returns were computed and stable parameters were estimated using maximum likelihood for each company. For Pfizer, $\widehat{\alpha} = 1.748$, $\widehat{\beta} = 0.0000$, $\widehat{\gamma} = 0.0070$ and $\widehat{\delta} = 0.00585$; for Merck, $\widehat{\alpha} = 1.735$, $\widehat{\beta} = -0.0852$, $\widehat{\gamma} = 0.0070$ and $\widehat{\delta} = 0.000684$. The indices are close, so we used $\alpha = (1.748 + 1.735)/2 = 1.7415$ and computed the test statistic $\widehat{\eta}_2 = 1.3381$. Simulations with 1256 independent stable terms shows that this value is highly significant: the critical value for $p = 0.01$ is 0.241, so we

reject the null hypothesis that the terms are independent. This is not surprising as PFE and MRK are in the same sector. We repeated this procedure with PFE and Walmart (WMT), which has a similar α and also reject independence at the p=0.01 level. In fact, we computed $\widehat{\eta}_2$ for every pair of stocks in the Dow Jones 30 index, and all reject independence at this level. These results should be interpreted cautiously, as estimates of α varied between 1.62 and 1.87 for different stocks and there were $C(30, 2) = 435$ comparisons made.

3.3 Higher dimensions

For $\mathbf{X} = (X_1, \ldots, X_d)$, $d > 2$, we can apply this method to each pair of coordinates and assess the d-dimensional data set. Specifically, for each pair of indices we compute $\eta_2(i, j) = \eta_2(X_i, X_j)$. Recall from above that the d dimensional data set is independent if and only if each pair of components is independent. Perhaps more important is the view of the joint dependence structure in a multivariate data set given by the η_2 matrix.

One application of this idea is in dimension reduction. If the η matrix shows obvious structure, e.g. blocks of dependence, then one can split the d dimensional modeling problem into two or more lower dimensional problems. This is illustrated in Fig. 5, which shows a greyscale plot of the pairwise values of η_2 for a simulated 10 dimensional data set that is $\alpha = 1.3$ stable. The first four coordinates of the simulated vector are independent, the next three are from an elliptical stable distribution with shape matrix $R = (1, 0.5, 0.25; 0.5, 1, 0.5; 0.25, 0.5, 1)$, and the last three are from a discrete spectral measure with mass spread around the first octant. The simulation method for the dependent elliptically contoured terms uses the representation $\mathbf{X} = A^{1/2}\mathbf{Z}$, where A is a positive $(\alpha/2)$-stable term and $\mathbf{Z}$ is N$(0, R)$, see section 2.5 of Samorodnitsky and Taqqu (1994). The dependent block with discrete spectral measure is simulated using Example 2.3.6 in Samorodnitsky and Taqqu (1994).

The blocks are independent of each other. The upper left plot shows the estimated $\widehat{\eta}_2$ matrix for a modest sample size of $n = 150$, where the dependence structure is visible, but not yet sharp. In upper right plot, the sample size is increased to $n = 4000$, and now the dependence structure is sharp. In the bottom left plot, the coordinates are randomly permuted to hide the dependence structure. Finally, the bottom right plot shows the reordered (from the permuted data) $\widehat{\eta}_2$ matrix, with the ordering coming from a clustering algorithm. (Specifically, we used the R function `hclust(dist(·))` applied to the $\widehat{\eta}_2$ matrix.) This approach may be useful in applications to discover structure in a heavy tailed multivariate data set.

3.4 Application to times series

Given a univariate time series $X_1, X_2, \ldots, X_n$ with stable error terms, the above definition of independence can be used to define an analogy of the autocorrelation function (ACF). The approach is similar: compute the dependence measure $\eta_{2,h}$ for lagged pairs (X_{t+h}, X_t), $t = 1, \ldots, n - h$. Plot $\eta_{2,h}$ as a function of h and show a simulation derived threshold value as in the standard ACF plot.

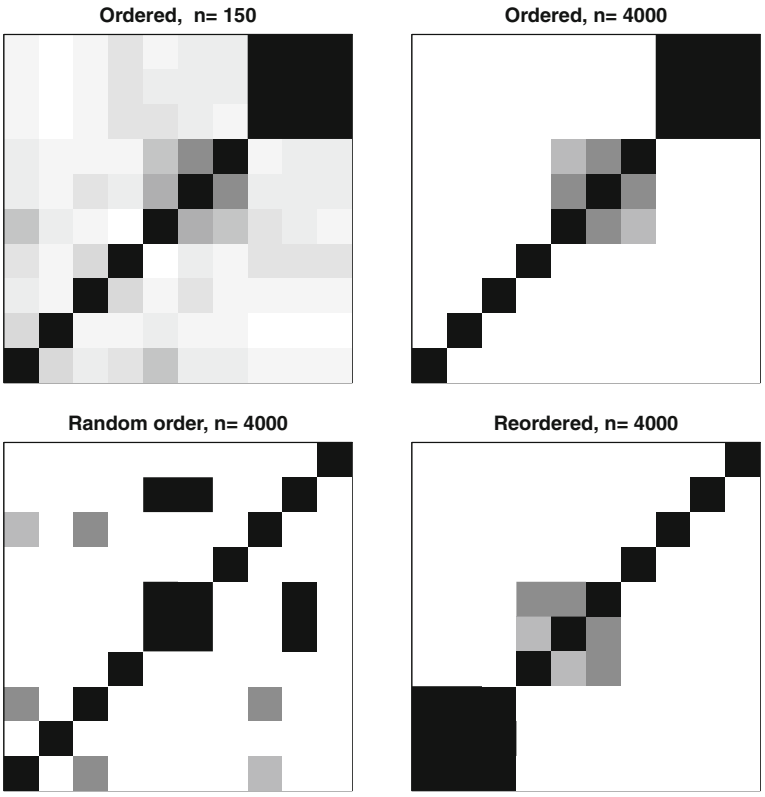


Fig. 5 Plot of $\widehat{\eta}_2$ for a simulated 10 dimensional data set with $\alpha = 1.3$, and dependence structure as described in the text

Figure 6 shows an example of this with simulated data. In this case, an AR(1) process is simulated: $X_t = (1/2)X_{t-1} + Z_t$, where Z_t are normalized i.i.d. symmetric α-stable. The geometric decay of $\widehat{\eta}_2$ is indicative of an AR model for the serial dependence.

This method is applied to the financial returns of Merck stock that was examined above. Figure 7 shows that there is no evidence for dependence among the lagged values of the returns.

While generally the η plot and the ACF plot look similar, the latter is sensitive to extreme values. This is illustrated in Fig. 8, where a simulated time series with independent terms is analyzed on the left side of the plot using both the η plot and a standard ACF plot. Then one point is changed in the time series in the following way: we looked for the maximum value in the time series (which was 204.62 in this simulation), and replaced a value 15 time periods away with 80 % of this maximum. The η plot and the standard ACF were then graphed for this altered time series on the right set of plots. For the modified data set, the ACF plot has a strong spike at lag 15, whereas the η plot does not. Thus one changed value significantly changes the ACF, but not the η plot.

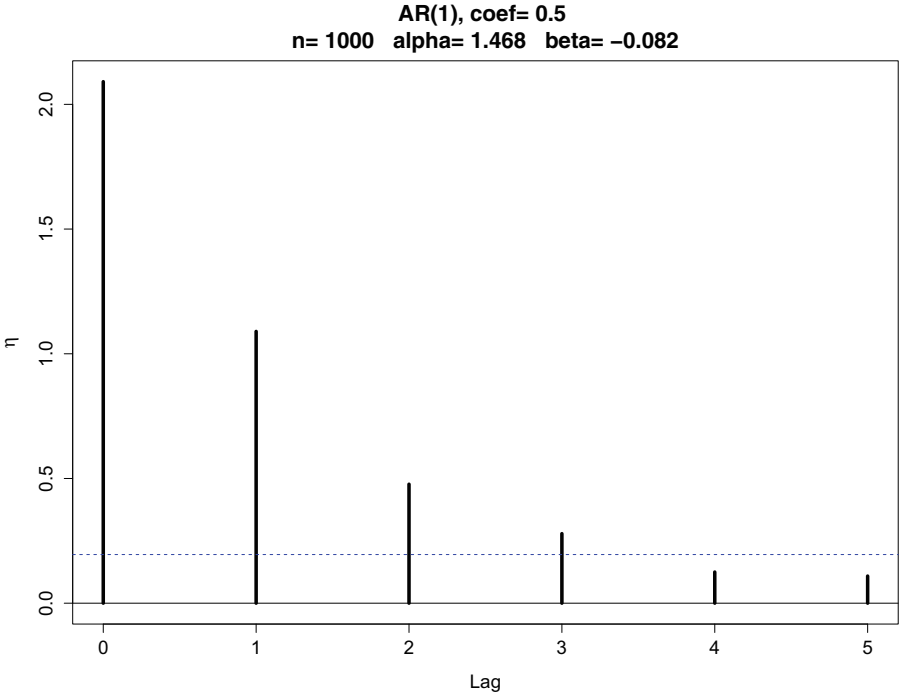


Fig. 6 $\eta_2(X_i, X_{i+j})$ for a simulated AR(1) time series. The *dashed horizontal line* is the critical value for rejecting independence when $n = 1000$, $\alpha = 1.5$, $\beta_1 = \beta_2 = 0$, and $\epsilon = 0.05$

3.5 Vectors in the domain of attraction of a stable law

In the preceding sections we assumed the bivariate vector was stable. We now show how the method can be adapted to vectors in the domain of attraction of a stable law. We will use the notation $\mathbf{X} \in \text{DOA}(\mathbf{S}(\alpha, \beta(\cdot), \gamma(\cdot), \delta(\cdot); 0))$ when $\mathbf{X}$ is in the domain of attraction of a $\mathbf{S}(\alpha, \beta(\cdot), \gamma(\cdot), \delta(\cdot); 0)$ law. Our approach is straightforward: use

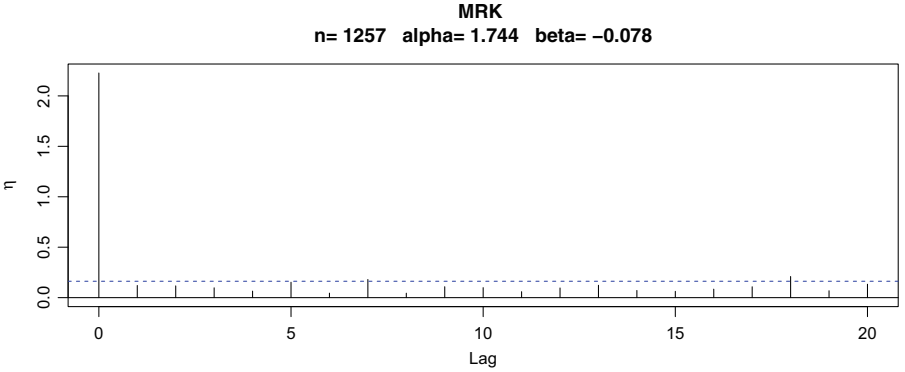


Fig. 7 $\eta_2(X_i, X_{i+h})$ for Merck returns for $\epsilon = 0.05$

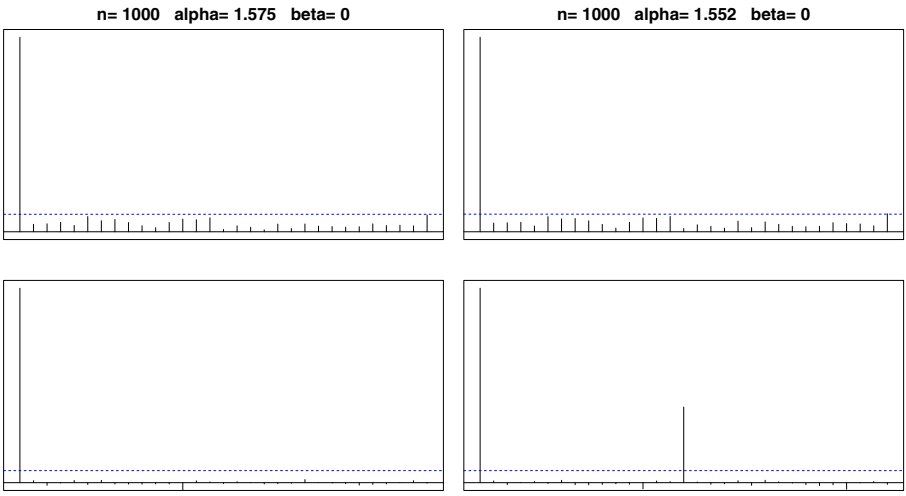


Fig. 8 $\eta_2(X_i, X_{i+j})$ for a simulated stable time series with $\epsilon = 0.05$, $\alpha = 1.5$, and independent terms. The *top left* shows a plot of $\widehat{\eta}_2$ for the original time series, the *bottom left* shows the standard ACF plot for the same data. On the right side, one point in the time series was changed

any univariate tail estimator for $\widehat{\alpha}$ and the scale function and then compute Eq. 5 as above.

There are a range of methods to choose here; we use the simplest to illustrate the approach. Sort the data and look at $\log x$ vs. the log of $1 - \widehat{F}(x)$, the complement of the empirical distribution function, beyond some threshold. A simple linear regression will give an estimate of the tail index and the scale. A similar approach can be done for the lower tail. For simplicity, we will assume that the data is two sided, and we average the tail indices to get $\widehat{\alpha}$ and scale $\widehat{\gamma}$.

Equipped with a one dimensional estimate of the tail index and scale, we proceed as we did in the exact stable case: (i) analyze the marginals and average the resulting tail indices to get an estimate of α; (ii) pick a grid of directions $\theta_1, \ldots, \theta_m$; (iii) project the data along each direction, getting estimates of the scale $\widehat{\gamma}_1, \ldots, \widehat{\gamma}_m$; (iv) compute $\widehat{\eta}_2$ using Eq. 5.

To compute a critical value, we use non-parametric bootstrapping: generate M data sets, each time generating independent vectors, with each component sampled with replacement from the sample; compute $\widehat{\eta}_2$ for this bootstrap sample; tabulate the values of $\widehat{\eta}_2$ and find the appropriate quantile.

To test this method, we reproduced Table 1 for the domain of attraction case. We generated several data sets based on independent Pareto terms: we simulated $\mathbf{X} = (X_1 - X_1', X_2 - X_2')$, where each term is independent Pareto($\alpha = 1.5$). This is a symmetric r. vector with independent components that is in the domain of attraction of an independent $\alpha = 1.5$ stable law. For each n, this was simulated $M = 1000$ times and $\widehat{\eta}_2$ was calculated using the tail estimator modification. In this and the following examples, $k = 5$ grid points were used in each quadrant for a total of $m = 2k = 10$ grid points and 10 % of the tails were used to estimate the tail index and scale of both the lower and the upper tails. From these simulations, the 0.95-percentile was

Table 3 Power of the statistic $\widehat{\eta}_2$ for $\mathbf{X}$ in the domain of attraction of an $\alpha = 1.5$ stable law, see the text for data description

Sample size n	Isotropic	Independent $\circlearrowleft \pi/4$	Independent $\circlearrowleft \pi/8$	Independent $\circlearrowleft \pi/16$	Exact linear dependence
100	0.253	0.057	0.049	0.058	0.161
200	0.708	0.025	0.040	0.049	0.342
300	0.844	0.010	0.013	0.023	0.481
400	0.940	0.011	0.020	0.022	0.995
500	0.956	0.011	0.007	0.018	1
600	0.986	0.024	0.013	0.028	1
700	0.988	0.023	0.003	0.009	1
800	0.995	0.258	0.012	0.019	1
900	0.998	0.284	0.013	0.011	1
1000	0.993	0.498	0.006	0.009	1
2000	1	0.996	0.376	0.008	1
3000	1	1	0.876	0.003	1
4000	1	1	0.989	0.003	1
5000	1	1	1	0.004	1

calculated and used as critical values for each of the cases considered in Table 3. For each column and row in that table we simulated $M = 1000$ data sets of the specified type and computed the fraction of times that independence is rejected at the 95 % level. For the column labeled isotropic, we generated $R \sim \text{Pareto}(\alpha = 1.5)$ and $\theta \sim \text{Uniform}(0, 2\pi)$ and set $\mathbf{X} = (R\cos\theta, R\sin\theta)$. For the next three columns, we generated independent terms as in the critical value calculations and then rotated those vectors by the specified angles. For the exact linear dependence, we simulated $\mathbf{X} = \epsilon(R, R)$, where $\epsilon \pm 1$ with probability 1/2 and $R \sim \text{Pareto}(\alpha = 1.5)$.

This table starts at higher values of n than the earlier table because we could not get consistent results with smaller sample sizes. Recall that the tail estimator is based on the bottom and top 10 % of a sample, and without sample sizes in the hundreds, little can be done. In fact, the high sample variability when there are small samples leads to some of the columns of Table 3 being non-monotonic. The power values are much lower than the ones in the exact stable case because most of the data is not being used by the tail estimator. We note that these results are dependent on multiple factors: the tail estimation method, the fraction of the tail used to estimate the index and scale, the marginal distributions of the data, etc. Finally, we note that the this method will be unreliable if the tail threshold is not chosen well or the available data doesn't show the limiting behavior clearly.

Acknowledgments The authors would like to thank Alyssa Cuyjet for performing simulations described in the Appendix. The authors would like to thank the reviewers for suggestions that improved the paper.

Conflict of interests The authors declare that they have no conflict of interest.

Appendix

To prevent having to do time consuming simulations to establish critical values, large scale simulations of critical values were run for a range of parameter values. The ranges were $\alpha \in \{0.5, 0.6, \ldots, 1.9, 2\}$, $(\beta_1, \beta_2) \in \{0, 1/2, 1\}^2$, $n \in \{100, 200, \ldots, 900, 1000, 2000, 3000, 4000\}$ and the grid had $k \in \{1, 2, \ldots, 10\}$ points uniformly spread in the first and second quadrant as described above. A total of 1479 sets of critical values were generated for quantile levels $\epsilon \in \{0.30, 0.20, 0.10, 0.05, 0.025, 0.01, 0.001\}$. each from simulation of size $M = 1000$. Combining heuristics and ad hoc fitting lead to the following approximation to the critical value:

$$c_\epsilon(\alpha, \beta_1, \beta_2, n, k) = D_0(\alpha, \beta_1 + \beta_2)\sigma_{\widehat{\gamma}}(\alpha)\sqrt{\frac{(k+1)\chi^2_{1-\epsilon,2}}{n}} + D_1(\alpha, \beta_1 + \beta_2),$$

where $\sigma_{\widehat{\gamma}}(\alpha)$ is the asymptotic standard deviation of the maximum likelihood estimation of the scale parameter γ for a standardized univariate symmetric stable law with index α (available from Nolan (2001)), $\chi^2_{1-\epsilon,2}$ is the standard χ^2 critical value for 2 degrees of freedom, and

$$\begin{aligned}
D_0(\alpha, b) &= d_0(\alpha)d_1(b)d_2(\alpha) \\
D_1(\alpha, b) &= d_0(\alpha)[d_1(b)d_3(\alpha) + d_4(b)] \\
d_0(\alpha) &= 0.1666\alpha + 0.95 \\
d_1(b) &= 1 - 0.0719b + 0.0083b^2 \\
d_2(\alpha) &= 0.01342 + 0.69421\alpha - 0.09149\alpha^2 \\
d_3(\alpha) &= -0.0387 + 0.124\alpha - 0.0558\alpha^2 \\
d_4(b) &= 0.007163b - 0.00187b^2.
\end{aligned}$$

To choose grid size k, further simulations were done for several different types of bivariate dependent stable components. For dependent distributions that are far away from independence or had maximum difference between $\gamma(\cdot)$ and $\gamma_\perp(\cdot)$ near $\pi/4$ or $3\pi/4$, high empirical power is reached with $k = 1$. However for some types of dependency, empirical power increases as k increases through $k = 1, 2, 3, 4, 5$, and plateaus after that. Two particular cases that show this are independence rotated by $\pi/16$ and the example in Proposition 3 with $\theta_0 = \pi/16$, both with sample size $n = 100$.

References

Blum, J.R., Kiefer, J., Rosenblatt, M.: Distribution free tests of independence based on the sample distribution function. Ann. Math. Statist **32**, 485–498 (1961)

Chambers, J., Mallows, C., Stuck, B.: A method for simulating stable random variables. J. Am. Stat. Assoc **71**(354), 340–344 (1976). Correction in JASA 82 (1987), 704

Csörgő, S.: Testing for independence by the empirical characteristic function. J. Multivariate Anal **16**(3), 290–299 (1985)

Garel, B., Kodia, B.: Estimation and comparison of signed symmetric covariation coefficient and generalized association parameter for alpha-stable dependence modeling. Communication in Statistics **54**, 252–276 (2014)

Hoeffding, W.: A non-parametric test of independence. Ann. Math. Statistics **19**, 546–557 (1948)

Nolan, J.P.: Maximum likelihood estimation of stable parameters. In: Barndorff-Nielsen, O.E., Mikosch, T., Resnick, S.I. (eds.) Lévy processes: theory and applications, pp. 379–400. Birkhäuser, Boston (2001)

Nolan, J.P.: Metrics for multivariate stable distributions. In: Stability in probability, vol. 90 of Banach Center Publ., pp.83-102. Polish Acad. Sci. Inst. Math.Warsaw (2010)

Pickands, J.: Multivariate extreme value distributions. In: Proceedings of the 43rd session of the International Statistical Institute, vol.2 (Buenos Aires, 1981), vo. 49, pp. 859–878, 894–902. With a discussion (1981)

Rosiński, J., Zak, T.: The equivalence of ergodicity of weak mixing for infinitely divisible processes. J. Theoret. Probab. **10**(1), 73–86 (1997)

Samorodnitsky, G., Taqqu, M.: Stable Non-Gaussian Random Processes. Chapman and Hall, New York (1994)

Székely, G.J., Rizzo, M.L.: Brownian distance covariance. Ann. Appl. Stat **3**(4), 1236–1265 (2009)

Székely, G.J., Rizzo, M.L., Bakirov, N.K.: Measuring and testing dependence by correlation of distances. Ann. Statist **35**(6), 2769–2794 (2007)