

## "Nonparametric estimation of extremal dependence"

Kiriliouk, Anna ; Segers, Johan ; Warchol, Michal

### Abstract

There is an increasing interest to understand the dependence structure of a random vector not only in the center of its distribution but also in the tails. Extreme-value theory tackles the problem of modelling the joint tail of a multivariate distribution by modelling the marginal distributions and the dependence structure separately. For estimating dependence at high levels, the stable tail dependence function and the spectral measure are particularly convenient. These objects also lie at the basis of nonparametric techniques for modelling the dependence among extremes in the max-domain of attraction setting. In case of asymptotic independence, this setting is inadequate, and more refined tail dependence coefficients exist, serving, among others, to discriminate between asymptotic dependence and independence. Throughout, the methods are illustrated on financial data.

Document type : *Document de travail (Working Paper)*

## Référence bibliographique

---

Kiriliouk, Anna ; Segers, Johan ; Warchol, Michal. *Nonparametric estimation of extremal dependence*. ISBA Discussion Paper ; 2014/44 (2014) 22 pages

INSTITUT DE STATISTIQUE  
BIOSTATISTIQUE ET  
SCIENCES ACTUARIELLES  
(ISBA)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN



DISCUSSION  
PAPER

2014/44

Nonparametric estimation of extremal dependence

KIRILIOUK, A., SEGERS J. AND M. WARCHOL

# Nonparametric estimation of extremal dependence

Anna Kiriliouk, Johan Segers and Michał Warchol

Université catholique de Louvain

Institut de Statistique, Biostatistique et Sciences Actuarielles

Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium

{anna.kiriliouk,johan.segers,michal.warchol}@uclouvain.be

November 4, 2014

## Abstract

There is an increasing interest to understand the dependence structure of a random vector not only in the center of its distribution but also in the tails. Extreme-value theory tackles the problem of modelling the joint tail of a multivariate distribution by modelling the marginal distributions and the dependence structure separately. For estimating dependence at high levels, the stable tail dependence function and the spectral measure are particularly convenient. These objects also lie at the basis of nonparametric techniques for modelling the dependence among extremes in the max-domain of attraction setting. In case of asymptotic independence, this setting is inadequate, and more refined tail dependence coefficients exist, serving, among others, to discriminate between asymptotic dependence and independence. Throughout, the methods are illustrated on financial data.

## 1 Introduction

Consider a financial portfolio containing three stocks: JP Morgan, Citibank and IBM. We download stock prices from <http://finance.yahoo.com> between January 1, 2000, and December 31, 2013, and convert them to weekly negative log-returns: if  $P_1, \dots, P_n$  is a series of stock prices, the negative log-returns are

$$X_i = -\log(P_i/P_{i-1}), \quad i = 1, \dots, n.$$

These three series of negative log-returns will be denoted by the vectors  $\mathbf{X}_i = (X_{i1}, X_{i2}, X_{i3})$  for  $i = 1, \dots, 729$ . By taking negative log-returns we force (extreme) losses to be in the upper tails of the distribution functions. It is such extreme values, and in particular their simultaneous occurrence, that are the focus of this chapter.

Figure 1 shows the scatterplots of the three possible pairs of negative log-returns on the exponential scale. Specifically, we transformed the negative log-returns to unit-Pareto

margins through

$$\widehat{X}_{ij}^* := \frac{1}{1 - \widehat{F}_j(X_{ij})}, \quad i \in \{1, \dots, n\}, j \in \{1, \dots, d\}, \quad (1)$$

(with  $d = 3$ ) and plotted them on the logarithmic scale; the empirical distribution functions  $\widehat{F}_j$  evaluated at the data are defined as

$$\widehat{F}_j(X_{ij}) := \frac{R_{ij,n} - 1}{n}, \quad i \in \{1, \dots, n\}, j \in \{1, \dots, d\}, \quad (2)$$

where  $R_{ij,n}$  is the rank of  $X_{ij}$  among  $X_{1j}, \dots, X_{nj}$ , i.e.,  $R_{ij,n} = \sum_{l=1}^n \mathbb{1}\{X_{lj} \leq X_{ij}\}$ .

From Figure 1, we observe that joint occurrences of large losses are more frequent for JP Morgan and Citibank (left) than for JP Morgan and IBM (middle) or for Citibank and IBM (right). Given the fact that JP Morgan and Citibank are financial institutions while IBM is an IT company, this is no surprise. In statistical parlance, the pair JP Morgan versus Citibank exhibits the strongest degree of upper *tail dependence*.

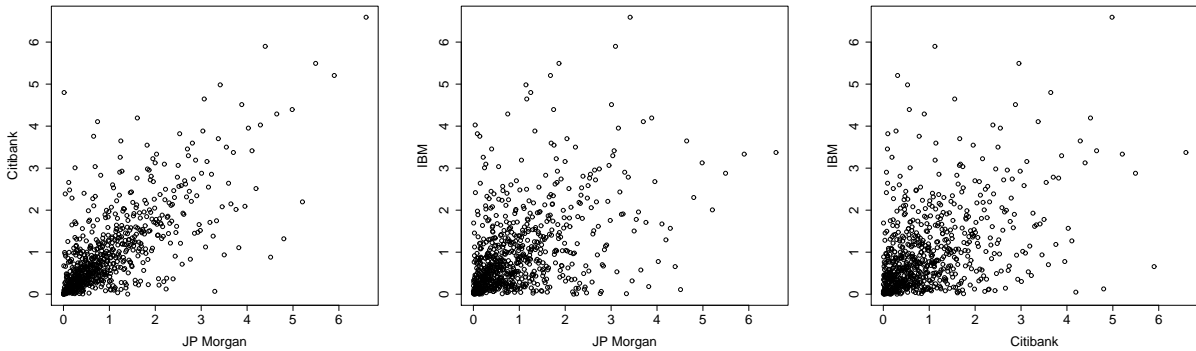


Figure 1: Scatterplots of negative weekly log-returns of stock prices of JP Morgan, Citibank and IBM, plotted on the exponential scale.

The purpose of this chapter is to describe a statistical framework for modelling such tail dependence. Specifically, we wish to model the tail dependence structure of a  $d$ -variate random variable  $\mathbf{X} = (X_1, \dots, X_d)$ , with continuous marginal distribution functions  $F_j(x_j) = \mathbb{P}(X_j \leq x_j)$  and joint distribution function  $F(\mathbf{x}) = \mathbb{P}(X_1 \leq x_1, \dots, X_d \leq x_d)$ .

Preliminary marginal transformations of the variables or changes in the measurement unit should not affect the dependence model. Such invariance is obtained by the probability integral transform, producing random variables  $F_1(X_1), \dots, F_d(X_d)$  that are uniformly distributed on the interval  $(0, 1)$ . Large values of  $X_j$  correspond to  $F_j(X_j)$  being close to unity, whatever the original scale in which  $X_j$  was measured. In order to magnify large values, it is sometimes convenient to consider a further transformation to unit-Pareto margins via  $X_j^* = 1/\{1 - F_j(X_j)\}$ ; note that  $\mathbb{P}[X_j^* > z] = 1/z$  for  $z \geq 1$ .

In practice, the marginal distributions are unknown and need to be estimated. A simple, robust way to do so is via the empirical distribution functions. The data are then reduced to the vectors of ranks as in (1) and (2).

Starting point of the methodology is a mathematical description of the upper tail dependence of the vector  $\mathbf{X}^* = (X_1^*, \dots, X_d^*)$  of standardized variables. There is a close connection with the classical theory on vectors of componentwise maxima, the multivariate version of the Fisher–Tippett–Gnedenko theorem (Fisher and Tippett, 1928; Gnedenko, 1943). In the literature, there is a plethora of objects available for describing the dependence component of the limiting multivariate extreme-value distribution. Out of these, we have singled out the stable tail dependence function  $\ell$  and the spectral measure  $H$  (Section 2). The former is convenient as it yields a direct approximation of the joint tail of the distribution function of  $\mathbf{X}^*$  and thus of  $\mathbf{X}$ ; see (6) below. The latter has the advantage of representing tail dependence as the distribution of the relative magnitudes of the components of  $\mathbf{X}^*$  given that the vector itself is ‘large’; see (11) below.

Nonparametric inference on tail dependence can then be based on the stable tail dependence function and the spectral measure (Section 3). For convenience and in line with most of the literature, the description of the spectral measure estimators is limited to the bivariate case, although the generalization to the general, multivariate case is straightforward.

A weak point of the classical theory is its treatment of asymptotic independence, that is, when the tail dependence coefficient in (25) is equal to zero. Classically, this case is assimilated to the situation where the variables are exactly independent. Such a reduction applying to all bivariate normal distributions with pairwise correlation less than unity, it is clearly inadequate. More refined tail dependence coefficients as well as tests to distinguish asymptotic dependence from asymptotic independence are presented in Section 4.

The present chapter can serve as an introduction to nonparametric multivariate extreme value analysis. Readers looking for in-depth treatments of multivariate extreme-value distributions and their max-domains of attraction can consult, for instance, Chapter 5 in Resnick (1987) or Chapter 6 in de Haan and Ferreira (2006). Some other accessible introductions are the expository papers by de Haan and de Ronde (1998) and Segers (2012), Chapter 8 in Beirlant et al. (2004), and Part II in Falk et al. (2011). The material on asymptotic independence is inspired on Coles et al. (1999).

## 2 Tail dependence

### 2.1 The stable tail dependence function

Let  $\mathbf{X} = (X_1, \dots, X_d)$  be a random vector with continuous marginal distribution functions  $F_j(x_j) = \mathbb{P}(X_j \leq x_j)$  and joint distribution function  $F(\mathbf{x}) = \mathbb{P}(X_1 \leq x_1, \dots, X_d \leq x_d)$ . To study tail dependence, we zoom in on the joint distribution of  $(F_1(X_1), \dots, F_d(X_d))$  in the neighbourhood of its upper endpoint  $(1, \dots, 1)$ . That is, we look at

$$\begin{aligned} 1 - \mathbb{P}[F_1(X_1) \leq 1 - tx_1, \dots, F_d(X_d) \leq 1 - tx_d] \\ = \mathbb{P}[F_1(X_1) > 1 - tx_1 \text{ or } \dots \text{ or } F_d(X_d) > 1 - tx_d], \end{aligned} \quad (3)$$

where  $t > 0$  is small and where the numbers  $x_1, \dots, x_d \in [0, \infty)$  parametrize the relative distances to the upper endpoints of the  $d$  variables. The above probability converges to zero

as  $t \rightarrow 0$  and is in fact proportional to  $t$ :

$$\begin{aligned} \max(tx_1, \dots, tx_d) &\leq \mathbb{P}[F_1(X_1) > 1 - tx_1 \text{ or } \dots \text{ or } F_d(X_d) > 1 - tx_d] \\ &\leq tx_1 + \dots + tx_d. \end{aligned}$$

The *stable tail dependence function*,  $\ell : [0, \infty)^d \rightarrow [0, \infty)$ , is then defined by

$$\ell(\mathbf{x}) := \lim_{t \downarrow 0} t^{-1} \mathbb{P}[F_1(X_1) > 1 - tx_1 \text{ or } \dots \text{ or } F_d(X_d) > 1 - tx_d], \quad (4)$$

for  $\mathbf{x} \in [0, \infty)^d$  (Huang, 1992; Drees and Huang, 1998). The existence of the limit in (4) is an assumption that can be tested (Einmahl et al., 2006).

The probability in (3) represents the situation where *at least one* of the variables is large: for instance, the sea level exceeds a critical height at one or more locations. Alternatively, we might be interested in the situation where *all* variables are large simultaneously. Think of the prices of all stocks in a financial portfolio going down together. The *tail copula*,  $R : [0, \infty)^d \rightarrow [0, \infty)$ , is defined by

$$R(\mathbf{x}) := \lim_{t \downarrow 0} t^{-1} \mathbb{P}[F_1(X_1) > 1 - tx_1, \dots, F_d(X_d) > 1 - tx_d], \quad (5)$$

for  $\mathbf{x} \in [0, \infty)^d \setminus \{(\infty, \dots, \infty)\}$ . Again, existence of the limit is an assumption. In the bivariate case ( $d = 2$ ), the functions  $\ell$  and  $R$  are directly related by  $R(x, y) = x + y - \ell(x, y)$ . The difference between  $\ell$  and  $R$  is visualized in Figure 2 for the log-returns of the stock prices of JP Morgan versus Citibank. From now on, we will focus on the function  $\ell$  because of its direct link to the joint distribution function; see (6).

It is convenient to transform the random variables  $X_1, \dots, X_d$  to have the unit-Pareto distribution via the transformations  $X_j^* = 1/\{1 - F_j(X_j)\}$  for  $j = 1, \dots, d$ . The existence of  $\ell$  in (4) is equivalent to the statement that the joint distribution function,  $H_0$ , of the random vector  $\mathbf{X}^* = (X_1^*, \dots, X_d^*)$  is in the max-domain of attraction of a  $d$ -variate extreme-value distribution, say  $G_0$ , with unit-Fréchet margins, i.e.,  $G_{0j}(z_j) = \exp(-1/z_j)$  for  $z_j > 0$ . The link between  $\ell$  and  $G_0$  is given by

$$\begin{aligned} \ell(\mathbf{x}) &= \lim_{t \downarrow 0} t^{-1} [1 - \mathbb{P}\{X_1^* \leq 1/(x_1 t), \dots, X_d^* \leq 1/(x_d t)\}] \\ &= \lim_{n \rightarrow \infty} -\log H_0^n(n/x_1, \dots, n/x_d) \\ &= -\log G_0(1/x_1, \dots, 1/x_d), \end{aligned}$$

so that  $G_0(\mathbf{z}) = \exp\{-\ell(1/z_1, \dots, 1/z_d)\}$ . Since, for “large”  $z_1, \dots, z_d$  we have

$$H_0(\mathbf{z}) \approx \exp\left\{-\frac{1}{n} \ell\left(\frac{n}{x_1}, \dots, \frac{n}{x_d}\right)\right\} = \exp\{-\ell(1/z_1, \dots, 1/z_d)\}, \quad (6)$$

we see that estimating the stable tail dependence function  $\ell$  is key to estimation of the tail of  $H_0$ , and thus, after marginal transformation, of  $F$ . If, in addition, the marginal distributions  $F_1, \dots, F_d$  are in the max-domains of attraction of the univariate extreme-value distributions  $G_1, \dots, G_d$ , then  $F$  is in the max-domain of attraction of the  $d$ -variate extreme-value distribution  $G(\mathbf{x}) = \exp\{-\ell(-\log G_1(x_1), \dots, -\log G_d(x_d))\}$ .

Every stable tail dependence function  $\ell$  has the following properties:

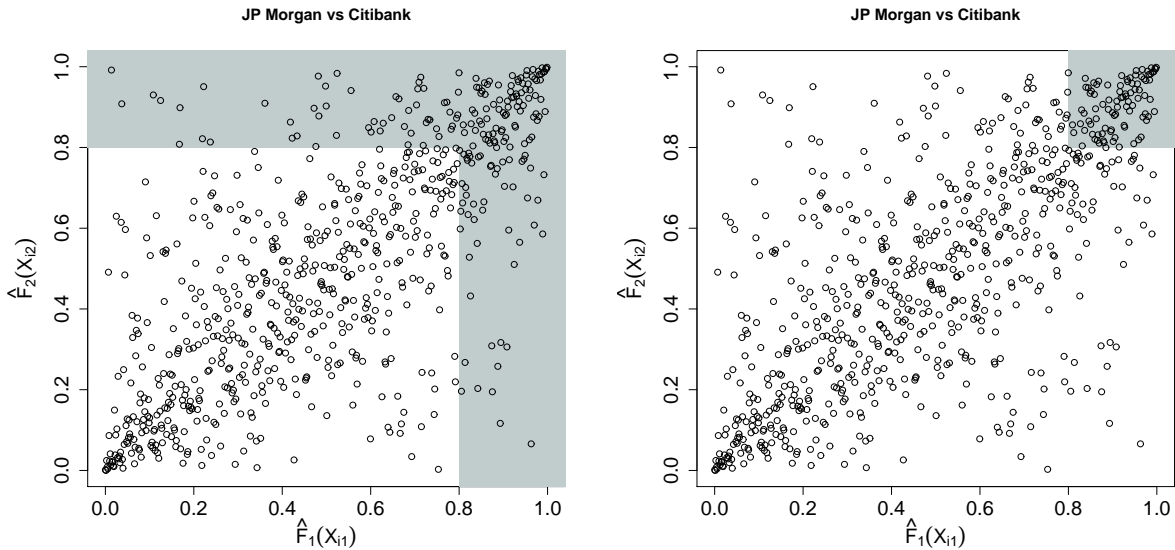


Figure 2: Scatterplots of negative weekly logreturns of stock prices of JP Morgan versus Citibank, transformed using ranks via (2). Left: the region where at least one variable is large, inspiring the definition of the stable tail dependence function  $\ell$  in (4). Right: the region where both variables are large, inspiring the definition of the tail copula  $R$  in (5).

1.  $\max(x_1, \dots, x_d) \leq \ell(\mathbf{x}) \leq x_1 + \dots + x_d$ , and in particular  $\ell(0, \dots, 0, x_j, 0, \dots, 0) = x_j$  for  $j = 1, \dots, d$ ;
2. convexity, that is,  $\ell(t\mathbf{x} + (1-t)\mathbf{y}) \leq t\ell(\mathbf{x}) + (1-t)\ell(\mathbf{y})$ , for  $t \in [0, 1]$ ;
3. order-one homogeneity:  $\ell(ax_1, \dots, ax_d) = a\ell(x_1, \dots, x_d)$ , for  $a > 0$ ;

(Beirlant et al., 2004, page 257). When  $d = 2$ , these three properties characterize the class of stable tail dependence functions. When  $d \geq 3$ , a function satisfying these properties is not necessarily a stable tail dependence function.

The two boundary cases are complete dependence,  $\ell(\mathbf{x}) = \max(x_1, \dots, x_d)$ , and independence,  $\ell(\mathbf{x}) = x_1 + \dots + x_d$ . Another example is the *logistic model*,

$$\ell(\mathbf{x}; \theta) = (x_1^\theta + \dots + x_d^\theta)^{1/\theta}, \quad \theta \in [1, \infty). \quad (7)$$

As  $\theta \rightarrow 1$ , extremes become independent and as  $\theta \rightarrow \infty$ , extremes become completely dependent. The copula of the extreme-value distribution  $G_0$  with this stable tail dependence function is the Gumbel–Hougaard copula.

The stable tail dependence function is related to two other functions:

- The *exponent measure function*  $V(\mathbf{z}) := -\log G_0(\mathbf{z}) = \ell(1/z_1, \dots, 1/z_d)$ , for  $\mathbf{z} \in (0, \infty]^d$  (Coles and Tawn, 1991).
- The *Pickands dependence function*  $A$ , which is just the restriction of  $\ell$  to the unit simplex  $\{\mathbf{x} \in [0, \infty)^d : x_1 + \dots + x_d = 1\}$  (Pickands, 1981). In dimension  $d = 2$ , it is common to set  $A(t) = \ell(1-t, t)$  for  $t \in [0, 1]$ .

## 2.2 The angular or spectral measure

Another insightful approach for modelling extremal dependence of a  $d$ -dimensional random vector is to look at the contribution of each of the  $d$  components to their sum conditionally on the sum being large.

Consider first the bivariate case. In Figure 3 we show pseudo-random samples from the Gumbel copula, which is related to the logistic model (7), at two different values of the dependence parameter  $\theta$ : strong dependence ( $\theta = 2.5$ ) at the top and weak dependence ( $\theta = 1.25$ ) at the bottom. For convenience, the axes in the scatter plots on the left are on the logarithmic scale. The points for which the sum of the coordinates belong to the top-10% are shown in black. In the middle panels we plot the ratio of the first component to the sum (horizontal axis) against the sum itself (vertical axis). On the right, finally, histograms are given of these ratios for those points for which the sum exceeded the threshold in the plots on the left and in the middle. Strong extremal dependence leads to the ratios being close to 0.5 (top), whereas low extremal dependence leads to the ratios being close to either 0 or 1 (bottom).

We learn that the distribution of the ratios of the components to their sum, given that the sum is large, carries information about dependence between large values. This distribution is called the *angular* or *spectral measure* (de Haan and Resnick, 1977) and is denoted here by  $H$ . The reference lines in Figure 3 represent the spectral density associated with the Gumbel copula with parameter  $\theta > 1$ , that is,

$$\frac{dH_\theta}{dw}(w) = \frac{\theta - 1}{2} \{w(1 - w)\}^{-1-\theta} \{w^{-\theta} + (1 - w)^{-\theta}\}^{1/\theta-2}, \quad 0 < w < 1.$$

In Figure 4 we show the same plots as in Figure 3 but now for the pairs of negative log-returns of the stock prices of JP Morgan versus Citibank (top) and JP Morgan versus IBM (bottom). For each of the two pairs we do the following. Let  $(X_i, Y_i)$ ,  $i = 1, \dots, n$ , denote a bivariate sample. First, we transform the data to the unit-Pareto scale using (1) and (2), i.e.,

$$\widehat{X}_i^* = n/(n + 1 - R_{i,X}) \quad \text{and} \quad \widehat{Y}_i^* = n/(n + 1 - R_{i,Y}), \quad i = 1, \dots, n. \quad (8)$$

Here  $R_{i,X}$  is the rank of  $X_i$  among  $X_1, \dots, X_n$  and  $R_{i,Y}$  is the rank of  $Y_i$  among  $Y_1, \dots, Y_n$ . Next, we construct an approximate sample from the spectral measure  $H$  by setting

$$\widehat{S}_i = \widehat{X}_i^* + \widehat{Y}_i^* \quad \text{and} \quad \widehat{W}_i = \widehat{X}_i^*/(\widehat{X}_i^* + \widehat{Y}_i^*), \quad i = 1, \dots, n. \quad (9)$$

The data in the pseudo-polar representation (9) are shown in the middle plots. The histograms of the ratios  $\widehat{W}_i$  for those points for which the sum  $\widehat{S}_i$  is above the 95% sample quantile are displayed on the right-hand side. The solid black lines show the smoothed spectral density estimator in (21) below.

For JP Morgan versus Citibank, the ratios are spread equally across the unit interval, suggesting extremal dependence. In contrast, for JP Morgan versus IBM, the spectral density has peaks at the boundaries 0 and 1, suggesting weak extremal dependence.

In the general,  $d$ -dimensional setting we transform the random vector  $\mathbf{X}^* = (X_1^*, \dots, X_d^*) = (1/\{1 - F_j(X_j)\}) : j = 1, \dots, d$  as follows:

$$S = X_1^* + \dots + X_d^* \quad \text{and} \quad \mathbf{W} = (X_1^*/S, \dots, X_d^*/S). \quad (10)$$



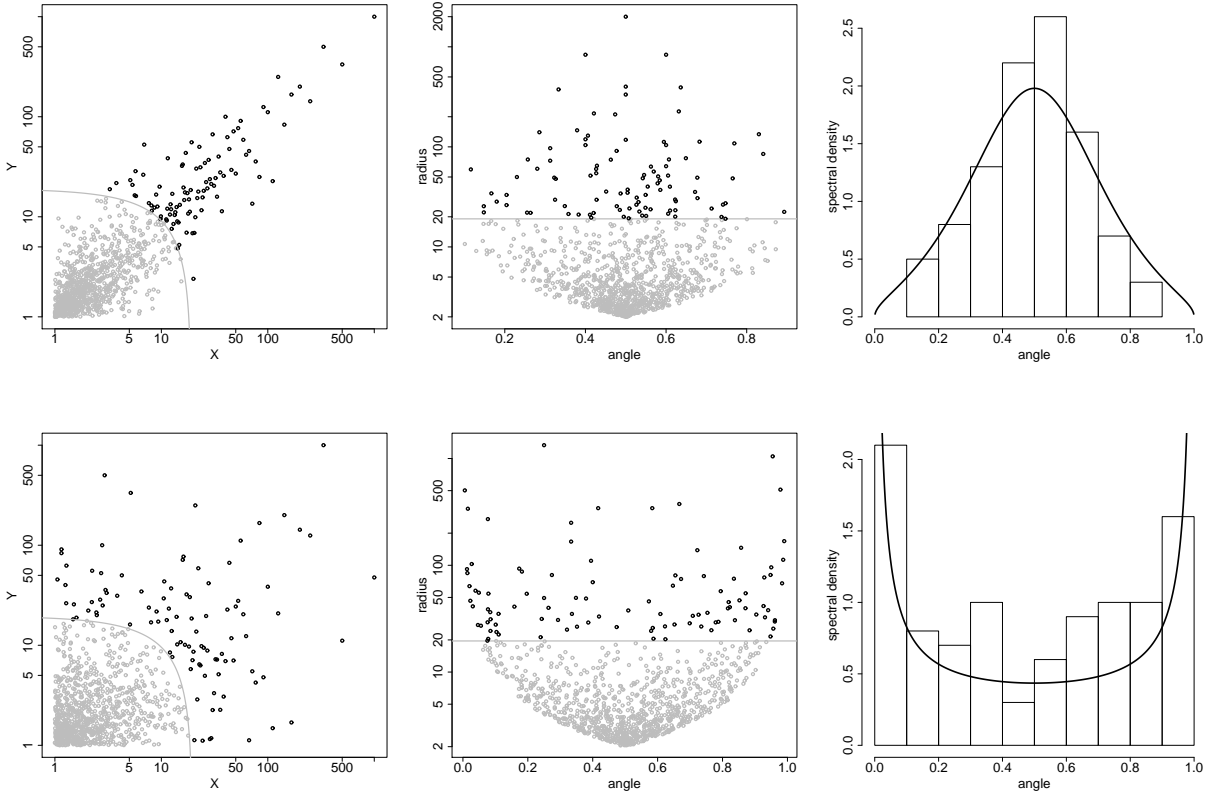


Figure 3: Pseudo-random samples from the Gumbel copula;  $\theta = 2.5$  (top) and  $\theta = 1.25$  (bottom). Left: the data transformed to unit-Pareto margins (log scale); middle: further transformation into pseudo-polar coordinates; right: histogram of the angles of the extreme pairs and the true spectral density (solid line).

The spectral measure is defined as the asymptotic distribution of the vector of ratios  $\mathbf{W}$  given that the sum  $S$  is large:

$$\mathbb{P}[\mathbf{W} \in \cdot \mid S > u] \xrightarrow{d} H(\cdot), \quad \text{as } u \rightarrow \infty. \quad (11)$$

The support of  $H$  is included in the unit simplex  $\Delta_{d-1} = \{\mathbf{w} \in [0, \infty)^d : w_1 + \dots + w_d = 1\}$ . In case of asymptotic independence, the spectral measure is concentrated on the  $d$  vertices of  $\Delta_{d-1}$ , whereas in case of complete asymptotic dependence it reduces to a unit point mass at the center of  $\Delta_{d-1}$ .

The spectral measure is related to the stable tail dependence function via

$$\ell(\mathbf{x}) = d \int_{\Delta_{d-1}} \max_{j=1, \dots, d} \{w_j x_j\} H(d\mathbf{w}). \quad (12)$$

The property that  $\ell(0, \dots, 0, 1, 0, \dots, 0) = 1$  implies the moment constraints

$$\int_{\Delta_{d-1}} w_j H(d\mathbf{w}) = \frac{1}{d}, \quad j = 1, \dots, d. \quad (13)$$

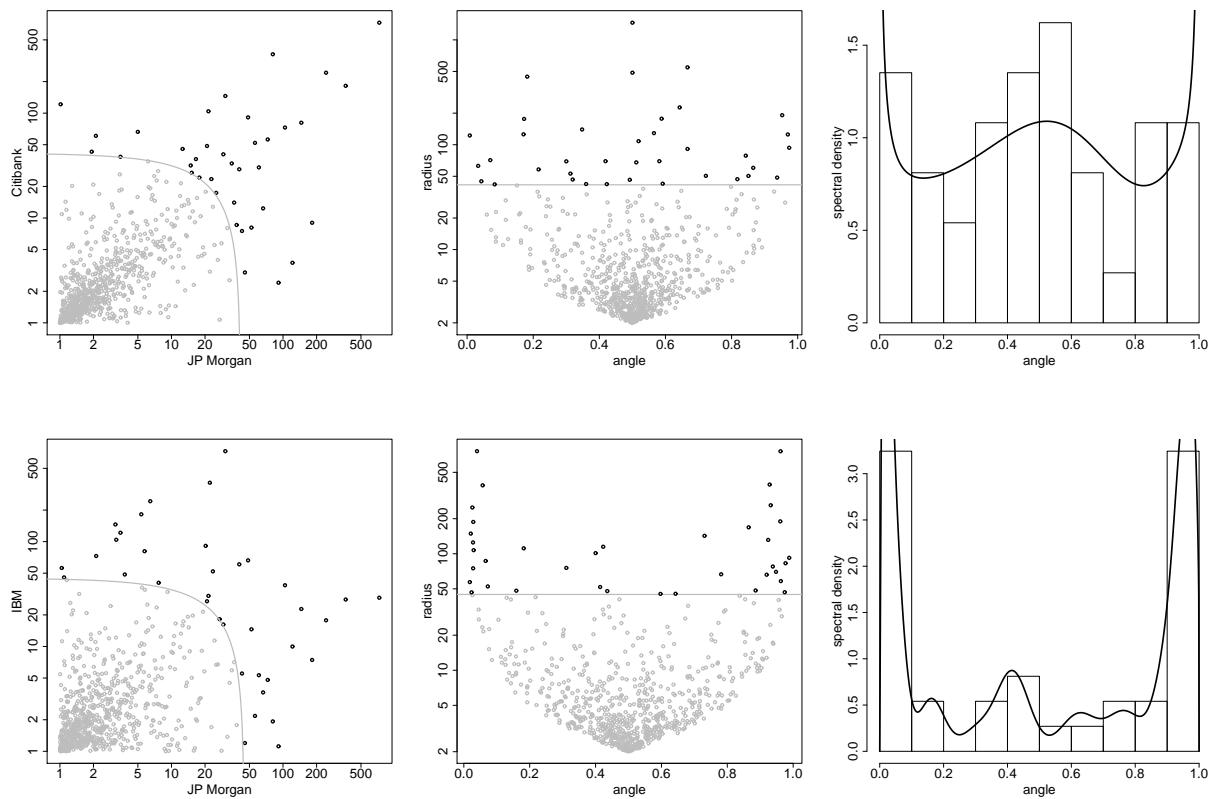


Figure 4: Log-returns of stock prices of JP Morgan versus Citibank (top) and JP Morgan versus IBM (bottom). Left: the data transformed to unit-Pareto margins (log scale); middle: further transformation into pseudo-polar coordinates; right: histogram of the angles of the extreme pairs and kernel density estimator (solid line).

Taking these constraints into account helps to improve the efficiency of nonparametric estimators of  $H$  (Section 3.1).

Any probability measure on the unit simplex satisfying (13) can be shown to be a valid spectral measure. The family of stable tail dependence functions is then characterized by (12) via a certain integral transform of such measures.

## 3 Estimation

### 3.1 Estimating the spectral measure

Let  $(X_i, Y_i)$ ,  $i = 1, \dots, n$ , be independent copies of a random vector  $(X, Y)$  with distribution function  $F$  and continuous marginal distribution functions  $F_X$  and  $F_Y$ . Assume that  $F$  has a stable tail dependence function  $\ell$  as in (24). The function  $\ell$  can be represented by the spectral measure  $H$  via (12), which in the bivariate case specializes to

$$\ell(x, y) = 2 \int_0^1 \max\{wx, (1-w)y\} H(dw), \quad (x, y) \in [0, \infty)^2, \quad (14)$$

the unit simplex in  $\mathbb{R}^2$  being identified with the unit interval,  $[0, 1]$ . The moment constraints in (13) simplify to

$$\int_0^1 w H(dw) = \int_0^1 (1-w) H(dw) = \frac{1}{2}. \quad (15)$$

The aim here is to estimate the spectral measure  $H$ . The starting point is the limit relation (11). Standardize the margins to the unit-Pareto distribution via  $X^* = 1/\{1 - F_X(X)\}$  and  $Y^* = 1/\{1 - F_Y(Y)\}$  and write

$$S = X^* + Y^*, \quad W = X^*/(X^* + Y^*).$$

Equation (11) then becomes

$$\mathbb{P}[W \in \cdot \mid S > u] \xrightarrow{d} H(\cdot), \quad \text{as } u \rightarrow \infty. \quad (16)$$

The spectral measure  $H$  can be interpreted as the limit distribution of the angular component,  $W$ , given that the radial component,  $S$ , is large.

Mimicking the transformation from  $(X, Y)$  via  $(X^*, Y^*)$  to  $(S, W)$  in a given data-set, it is convenient to standardize the data to the unit-Pareto distribution via (8) and then switch to pseudo-polar coordinates using (9). Different choices for the angular components are possible; for example, Einmahl and Segers (2009) use  $\arctan(X^*/Y^*)$ . For the radial component one can take in general the  $L_p$  norm,  $p \in [1, +\infty]$ ; Einmahl et al. (2001) take the  $L_\infty$  norm and de Carvalho et al. (2013) take the  $L_1$  norm. Note however that different normalizations lead to different spectral measures.

For estimation purposes, transform pairs  $(X_i, Y_i)$  into  $(\widehat{S}_i, \widehat{W}_i)$  as in (8) and (9). Set the threshold  $u$  in (16) to  $n/k$ , where  $k = k_n \in (0, n]$  is an intermediate sequence such that  $k \rightarrow \infty$  and  $k/n \rightarrow 0$  as  $n \rightarrow \infty$ . Let  $I_n = \{i = 1, \dots, n : \widehat{S}_i \geq n/k\}$  denote the set of indices that correspond to the observations with pseudo-radial component  $\widehat{S}_i$  above a high threshold, and let  $N_n = |I_n|$  denote its cardinality. Here we specify  $u$  as the 95% quantile of  $\widehat{S}_i$ .

In the literature, three nonparametric estimators of the spectral measure have been proposed. All three are of the form

$$\widehat{H}_l(x) = \sum_{i \in I_n} \widehat{p}_{l,i} \mathbb{1}\{\widehat{W}_i \leq x\}, \quad x \in [0, 1], \quad l = 1, 2, 3. \quad (17)$$

The estimators distinguish themselves in the way the weights  $\widehat{p}_{l,i}$  are defined.

For the *empirical spectral measure* (Einmahl et al., 2001), the weights are set to be constant, i.e.,  $\widehat{p}_{1,i} = 1/N_n$ . The estimator  $\widehat{H}_1$  (17) becomes an empirical version of (16). However, this estimator does not necessarily satisfy the moment constraints in (15). This is a motivation for the two other estimators where the moment constraint is enforced by requiring that  $\sum_{i \in I_n} \widehat{W}_i \widehat{p}_{l,i} = 1/2$ .

The *maximum empirical likelihood estimator* (Einmahl and Segers, 2009),  $\widehat{H}_2(x)$ , has probability masses  $\widehat{p}_{2,i}$  solving the optimization problem

$$\begin{aligned} \max_{\mathbf{p} \in \mathbb{R}_+^k} \quad & \sum_{i \in I_n} \log p_i \\ \text{s.t.} \quad & \sum_{i \in I_n} p_i = 1 \\ & \sum_{i \in I_n} \widehat{W}_i p_i = 1/2. \end{aligned} \quad (18)$$

By construction, the weights  $\widehat{p}_{2,i}$  satisfy the moment constraints. The optimization problem in (18) can be solved by the method of Lagrange multipliers. In the nontrivial case where  $1/2$  is in the convex hull of  $\{\widehat{W}_i : i \in I_n\}$ , the solution is given by

$$\widehat{p}_{2,i} = \frac{1}{N_n} \frac{1}{1 + \lambda(\widehat{W}_i - 1/2)}, \quad i \in I_n,$$

where  $\lambda \in \mathbb{R}$  is the Lagrange multiplier associated to the second equality constraint in (18), defined implicitly as the solution to the equation

$$\frac{1}{N_n} \sum_{i \in I_n} \frac{\widehat{W}_i - 1/2}{1 + \lambda(\widehat{W}_i - 1/2)} = 0.$$

Note that in (18), it is implicitly assumed that  $\widehat{p}_{2,i} > 0$ .

Another estimator that satisfies the moment constraints is the *maximum Euclidean likelihood estimator* (de Carvalho et al., 2013). The probability masses  $\widehat{p}_{3,i}$  solve the following optimization problem:

$$\begin{aligned} \max_{\mathbf{p} \in \mathbb{R}^{I_n}} \quad & -\frac{1}{2} \sum_{i \in I_n} (N_n p_i - 1)^2 \\ \text{s.t.} \quad & \sum_{i \in I_n} p_i = 1 \\ & \sum_{i \in I_n} \widehat{W}_i p_i = 1/2. \end{aligned} \tag{19}$$

This quadratic optimization problem with linear constraints can be solved explicitly with the method of Lagrange multipliers, yielding

$$\widehat{p}_{3,i} = \frac{1}{N_n} \{1 - (\overline{W} - 1/2) S_W^{-2} (\widehat{W}_i - \overline{W})\}, \quad i \in I_n, \tag{20}$$

where  $\overline{W}$  and  $S_W^2$  denote the sample mean and sample variance of  $\widehat{W}_i$ ,  $i \in I_n$ , respectively, that is,

$$\overline{W} = \frac{1}{N_n} \sum_{i \in I_n} \widehat{W}_i, \quad S_W^2 = \frac{1}{N_n} \sum_{i \in I_n} (\widehat{W}_i - \overline{W})^2.$$

The weights  $\widehat{p}_{3,i}$  could be negative, but this does usually not occur as the weights are all nonnegative with probability tending to one.

It is shown in Einmahl and Segers (2009) and de Carvalho et al. (2013) that  $\widehat{H}_2(x)$  and  $\widehat{H}_3(x)$  are more efficient than  $\widehat{H}_1(x)$ . Moreover, asymptotically there is no difference between the maximum empirical likelihood or maximum Euclidean likelihood estimators. The maximum Euclidean likelihood estimator is especially convenient as the weights  $\widehat{p}_{3,i}$  are given explicitly.

For the stock market data from JP Morgan, Citibank and IBM, Figure 5 shows the empirical spectral measure  $\widehat{H}_1(x)$  and the maximum Euclidean estimator  $\widehat{H}_3(x)$  for the two pairs involving JP Morgan. In each case, the threshold  $u$  is set to be the 95% quantile of the sample of radii  $\widehat{S}_i$ . Enforcing the moment constraints makes a small but noticeable difference. Tail dependence is strongest for the pair JP Morgan versus Citibank, with spectral mass distributed approximately uniformly over  $[0, 1]$ . For the pair JP Morgan versus IBM,

tail dependence is weaker, the spectral measure being concentrated mainly near the two endpoints, 0 and 1.

The nonparametric estimators of the spectral probability measure can be smoothed in such a way that the smoothed versions still obey the moment constraints. This can be done using kernel smoothing techniques, although some care is needed since the spectral measure is defined on a compact interval. In de Carvalho et al. (2013) the estimator is constructed by combining Beta distributions with the weights,  $\widehat{p}_{3,i}$ , of the maximum Euclidean likelihood estimate. To ensure that the estimated measure obeys the marginal moment constraints, it is imposed that the mean of each smoother equals the observed pseudo-angle. The Euclidean spectral density estimator is defined as

$$\widetilde{h}_3(x) = \sum_{i \in I_n} \widehat{p}_{3,i} \beta(x; \widehat{W}_i \nu, (1 - \widehat{W}_i) \nu), \quad x \in (0, 1), \quad (21)$$

where  $\nu > 0$  is the smoothing parameter, to be chosen via cross-validation, and

$$\beta(x; p, q) = \left\{ \int_0^1 u^{p-1} (1-u)^{q-1} du \right\}^{-1} x^{p-1} (1-x)^{q-1} \mathbb{1}_{[0,1]}(x),$$

is the Beta density with parameters  $p, q > 0$ .

For the financial data, the realized spectral density estimator  $\widetilde{h}_3(x)$  is plotted on the right-hand panel of Figure 4. The picture is in agreement with the one in Figure 5. The distribution of the spectral mass points to asymptotic dependence for JP Morgan versus Citibank and to weaker asymptotic dependence for JP Morgan versus IBM.

Integrating out the estimated spectral density yields a smoothed version of the empirical spectral measure or its variants: for the maximum Euclidean likelihood estimator, we get

$$\widetilde{H}_3(x) = \int_0^x \widetilde{h}_3(v) dv = \sum_{i \in I_n} \widehat{p}_{3,i} \mathcal{B}(x; \widehat{W}_i \nu, (1 - \widehat{W}_i) \nu), \quad x \in [0, 1], \quad (22)$$

with  $\mathcal{B}(x; p, q) = \int_0^x \beta(y; p, q) dy$  the regularized incomplete beta function. As

$$\int_0^1 x \widetilde{h}_3(x) dx = \sum_{i \in I_n} \widehat{p}_{3,i} \frac{\nu \widehat{W}_i}{\nu \widehat{W}_i + \nu(1 - \widehat{W}_i)} = \sum_{i \in I_n} \widehat{p}_{3,i} \widehat{W}_i = 1/2, \quad (23)$$

the moment constraint (15) is satisfied.

### 3.2 Estimating the stable tail dependence function

Consider again the general, multivariate case. Given a random sample  $\mathbf{X}_i = (X_{i1}, \dots, X_{id})$ ,  $i = 1, \dots, n$ , the aim is to estimate the stable tail dependence function  $\ell$  in (4) of the common distribution  $F$  with continuous margins  $F_1, \dots, F_d$ . A straightforward nonparametric estimator can be defined as follows. Let  $k = k_n \in \{1, \dots, n\}$  be such that  $k \rightarrow \infty$  and  $k/n \rightarrow 0$  as  $n \rightarrow \infty$ . By replacing  $\mathbb{P}$  by the empirical distribution function,  $t$  by  $k/n$ , and

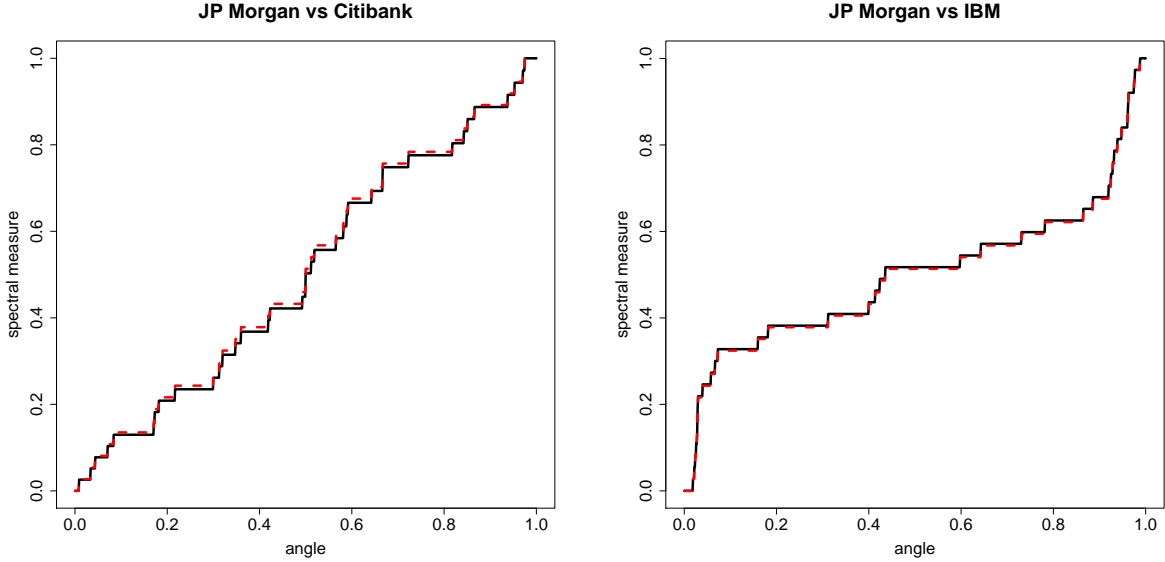


Figure 5: Estimated empirical spectral measures for JP Morgan versus Citibank (left) and JP Morgan versus IBM (right). The solid line and the dashed line correspond to  $\widehat{H}_3(x)$  and  $\widehat{H}_1(x)$  respectively.

$F_1, \dots, F_d$  by  $\widehat{F}_1, \dots, \widehat{F}_d$  as defined in (2), we obtain the *empirical tail dependence function* (Huang, 1992; Drees and Huang, 1998)

$$\begin{aligned} \widehat{\ell}(\mathbf{x}) &:= \frac{n}{k} \frac{1}{n} \sum_{i=1}^n \mathbb{1} \left\{ \widehat{F}_1(X_{i1}) > 1 - \frac{kx_1}{n} \text{ or } \dots \text{ or } \widehat{F}_d(X_{id}) > 1 - \frac{kx_d}{n} \right\} \\ &= \frac{1}{k} \sum_{i=1}^n \mathbb{1} \{ R_{i1,n} > n + 1 - kx_1 \text{ or } \dots \text{ or } R_{id,n} > n + 1 - kx_d \}. \end{aligned}$$

Under minimal assumptions, the estimator is consistent and asymptotically normal with a convergence rate of  $\sqrt{k}$  (Einmahl et al., 2012; Bücher et al., 2014). Alternatively, the marginal distributions might be estimated by  $\widehat{F}_{j,2}(X_{ij}) = R_{ij,n}/n$  or  $\widehat{F}_{j,3}(X_{ij}) = (R_{ij,n} - 1/2)/n$ , resulting in estimators that are asymptotically equivalent to  $\widehat{\ell}$ .

Another estimator of  $\ell$  can be defined by estimating the spectral measure,  $H$ , and applying the transformation in (12), an idea going back to Capéraà and Fougères (2000). In the bivariate case, we can replace  $H$  in (14) by one of the three nonparametric estimators  $\widehat{H}_l$  studied in Section 3.1. Recall the notations in (8) and (9). If we modify the definition of the index set  $I_n$  to  $I_n = \{i = 1, \dots, n : \widehat{S}_i > \widehat{S}_{(k+1)}\}$ , where  $\widehat{S}_{(k+1)}$  denotes the  $(k+1)$ -th largest observation of the  $\widehat{S}_i$ 's, then there are exactly  $k$  elements in the set  $I_n$ . Starting from the empirical spectral measure  $\widehat{H}_1$ , we obtain the estimator

$$\widehat{\ell}_{CF}(t) = \frac{1}{k} \sum_{i \in I_n} \max \left\{ \widehat{W}_i x, (1 - \widehat{W}_i) y \right\}.$$

For the maximum empirical or Euclidean likelihood estimators  $\widehat{H}_2$  and  $\widehat{H}_3$ , one needs to

replace the factor  $1/k$  by the weights  $\widehat{p}_{2,i}$  and  $\widehat{p}_{3,i}$ , respectively.

A way to visualize the function  $\ell$  or an estimator thereof is via the level sets  $\mathcal{D}_c := \{(x, y) : \ell(x, y) = c\}$  for a range of value of  $c > 0$ . Note that the level sets are equal to the lines  $x + y = c$  in case of asymptotic independence and to the elbow curves  $\max(x, y) = c$  in case of complete asymptotic dependence. Likewise, a plot of the level sets of an estimator of  $\ell$  can be used as a graphical diagnostic of asymptotic (in)dependence; see de Haan and de Ronde (1998) or de Haan and Ferreira (2006, Section 7.2).

We plot the lines  $\mathcal{D}_c$  for  $c \in \{0.2, 0.4, 0.6, 0.8, 1\}$  and  $k = 50$  of  $\widehat{\ell}$  and  $\widehat{\ell}_{CF}$  for the weekly log-returns of JP Morgan versus Citibank and JP Morgan versus IBM in Figure 6. The level sets for JP Morgan versus IBM resemble the straight lines  $x + y = c$  much more closely than the level sets for JP Morgan versus Citibank do. The estimator based on the spectral measure,  $\widehat{\ell}_{CF}$ , acts as a smooth version of the empirical tail dependence function  $\widehat{\ell}$ .

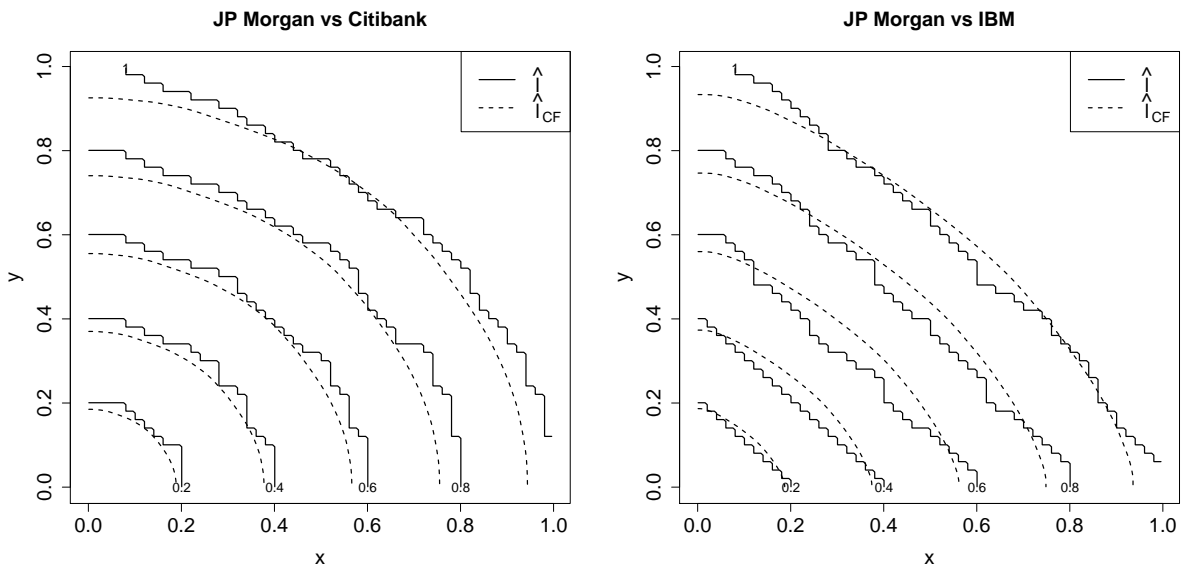


Figure 6: Level sets  $\mathcal{D}_c$  of  $\widehat{\ell}(x, y)$  and  $\widehat{\ell}_{CF}$  for  $c \in \{0.2, 0.4, 0.6, 0.8, 1\}$ .

Nonparametric estimators for  $\ell$  can also serve as stepping stone for semiparametric inference. Assume that  $\ell = \ell_{\theta_0} \in \{\ell_{\theta} : \theta \in \Theta\}$ , a finite-dimensional parametric family. Then one could estimate  $\theta_0$  by minimizing a distance or discrepancy measure between  $\widehat{\ell}$  and members of the parametric family (Einmahl et al., 2012, 2014).

## 4 Asymptotic independence

In this section we will focus on the bivariate case. Let  $(X_i, Y_i)$ ,  $i = 1, \dots, n$ , be independent copies of a random vector  $(X, Y)$ , with distribution function  $F$  having continuous margins  $F_X$  and  $F_Y$ . Recall from (4) that the stable tail dependence function (provided it exists) is given by

$$\ell(x, y) = \lim_{t \downarrow 0} t^{-1} \mathbb{P}[1 - F_X(X) \leq xt \text{ or } 1 - F_Y(Y) \leq yt], \quad (x, y) \in [0, \infty)^2. \quad (24)$$

After marginal standardization, the pair of component-wise sample maxima converges weakly to the bivariate max-stable distribution  $G_0$  given by  $G_0(z_1, z_2) = \exp\{-\ell(1/z_1, 1/z_2)\}$  for  $(z_1, z_2) \in (0, \infty)^2$ ; see (6). *Asymptotic independence* of the sample maxima,  $G_0(z_1, z_2) = \exp(-1/z_1) \exp(-1/z_2)$  for all  $(z_1, z_2)$ , is equivalent to  $\ell(x, y) = x + y$  for all  $(x, y)$ . The opposite case,  $\ell(x, y) < x + y$  for some  $(x, y)$ , is referred to as *asymptotic dependence*.

Sibuya (1960) already observed that bivariate normally distributed vectors are asymptotically independent as soon as the correlation is less than unity. We find that in case of asymptotic independence, we cannot rely on the function  $\ell$  to quantify the amount of dependence left above high but finite thresholds.

Tail dependence coefficients serve to quantify the amount of tail dependence and to distinguish between asymptotic dependence versus independence (Subsection 4.1). To decide which of the two situations applies for a data set, a number of testing procedures are available (Subsection 4.2).

## 4.1 Tail dependence coefficients

A first measure for the strength of asymptotic dependence is the *tail dependence coefficient*

$$\chi := \lim_{u \uparrow 1} \mathbb{P}[F_X(X) > u \mid F_Y(Y) > u] = R(1, 1) = 2 - \ell(1, 1) \quad (25)$$

(Coles et al., 1999). By the properties of  $\ell$  (Section 2.1), we have  $\ell(x, y) = x + y$  for all  $(x, y)$  if and only if  $\ell(1, 1) = 2$ . That is, asymptotic independence is equivalent to  $\chi = 0$ , whereas asymptotic dependence is equivalent to  $\chi \in (0, 1]$ .

To estimate  $\chi$ , we will first write it as the limit of a function  $\chi(u)$ , so that  $\lim_{u \uparrow 1} \chi(u) = \chi$ ; the function  $\chi(u)$  is defined as

$$\chi(u) := 2 - \frac{1 - \mathbb{P}[F_X(X) < u, F_Y(Y) < u]}{1 - u}, \quad 0 \leq u < 1.$$

To estimate  $\chi(u)$  from a sample  $(X_1, Y_1), \dots, (X_n, Y_n)$ , simply replace  $F_X$  and  $F_Y$  by their empirical counterparts in (2):

$$\widehat{\chi}(u) := 2 - \frac{1}{1 - u} \left( 1 - \frac{1}{n} \sum_{i=1}^n \mathbb{1} \left( \widehat{F}_X(X_i) < u, \widehat{F}_Y(Y_i) < u \right) \right).$$

For the stock price log-returns of JP Morgan, Citibank, and IBM, the estimated tail dependence coefficients  $\widehat{\chi}(u)$  of the three possible pairs are shown in Figure 7 as a function of  $u \geq 0.8$ . The plot confirms the earlier finding that tail dependence is stronger between the two banking stocks, JP Morgan and Citibank, than between either of the two banks and IBM.

If the variables are asymptotically independent,  $\chi = 0$ , we need a second measure quantifying the extremal dependence. The *coefficient of tail dependence*,  $\eta$ , can be motivated as follows (Ledford and Tawn, 1996). Since  $\mathbb{P}[1/\{1 - F_X(X)\} > t] = 1/t$  for  $t \geq 1$ , we have

$$\mathbb{P} \left[ \frac{1}{1 - F_X(X)} > t, \frac{1}{1 - F_Y(Y)} > t \right] \propto \begin{cases} t^{-1} & \text{for asymptotic dependence,} \\ t^{-2} & \text{for exact independence.} \end{cases}$$



A model that links these two situations is given by the assumption that the joint survival function above is regularly varying at  $\infty$  with index  $-1/\eta$ , i.e.,

$$\mathbb{P} \left[ \frac{1}{1 - F_X(X)} > t, \frac{1}{1 - F_Y(Y)} > t \right] = t^{-1/\eta} \mathcal{L}(t), \quad (26)$$

where  $\mathcal{L}$  is a slowly varying function, that is,  $\mathcal{L}(tr)/\mathcal{L}(t) \rightarrow 1$  as  $t \rightarrow \infty$  for fixed  $r > 0$ ; see for example Resnick (1987).

For large  $t$ , if we treat  $\mathcal{L}$  as constant, there are three cases of non-negative extremal association to distinguish:

1.  $\eta = 1$  and  $\mathcal{L}(t) \rightarrow c > 0$ : asymptotic dependence;
2.  $1/2 < \eta < 1$ : positive association within asymptotic independence;
3.  $\eta = 1/2$ : (near) perfect independence.

Values of  $\eta < 1/2$  occur when there is negative association between the two random variables at extreme levels.

In order to estimate  $\eta$ , we first define the structure variable

$$T := \min \left( \frac{1}{1 - F_X(X)}, \frac{1}{1 - F_Y(Y)} \right). \quad (27)$$

For a high threshold  $u$  and for  $t \geq 0$ , we have, by slow variation of  $\mathcal{L}$ ,

$$\mathbb{P}[T > u + t \mid T > u] = \frac{\mathcal{L}(u + t)}{\mathcal{L}(u)} (1 + t/u)^{-1/\eta} \approx (1 + t/u)^{-1/\eta}.$$

On the right-hand side we recognize a generalized Pareto distribution with shape parameter  $\eta$ . Given the sample  $(X_1, Y_1), \dots, (X_n, Y_n)$ , we can estimate  $\eta$  with techniques from univariate extreme value theory applied to the variables

$$\widehat{T}_i := \min \left( \frac{1}{1 - \widehat{F}_X(X_i)}, \frac{1}{1 - \widehat{F}_Y(Y_i)} \right), \quad i = 1, \dots, n.$$

One possibility consists of fitting the generalized Pareto distribution to the sample of excesses  $\widehat{T}_i - u$  for those  $i$  for which  $\widehat{T}_i > u$  (Ledford and Tawn, 1996, page 180). Alternatively, one can use the Hill estimator (Hill, 1975)

$$\widehat{\eta}_H := \frac{1}{k} \sum_{i=1}^k \log \left( \frac{\widehat{T}_{(i)}}{\widehat{T}_{(k+1)}} \right),$$

where  $\widehat{T}_{(1)} \geq \dots \geq \widehat{T}_{(n)}$  denote the order statistics of  $\widehat{T}_1, \dots, \widehat{T}_n$ . Figure 7 shows the estimators  $\widehat{\eta}_H$  of the coefficient of tail dependence between JP Morgan and Citibank, JP Morgan and IBM, and Citibank and IBM for a decreasing number of order statistics. The estimators suggest asymptotic dependence for the log-returns of JP Morgan versus Citibank.

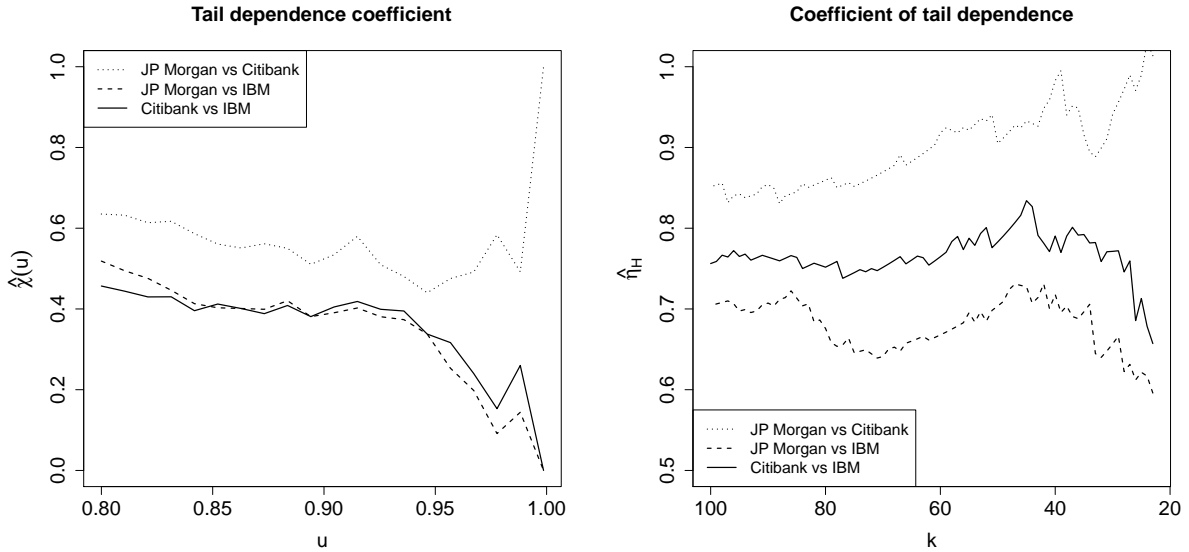


Figure 7: Estimators  $\hat{\chi}(u)$  (left) and  $\hat{\eta}_H$  (right) for the three pairs of log-returns.

Another frequently used measure in case of asymptotic independence can be obtained from (26). Taking logarithms on both sides and using slow variation of  $\mathcal{L}$ , we obtain

$$\frac{1}{\eta} = \lim_{u \uparrow 1} \frac{\log \mathbb{P}[F_X(X) > u, F_Y(Y) > u]}{\log \mathbb{P}[F_X(X) > u]}.$$

Coles et al. (1999) proposed the coefficient

$$\bar{\chi} := \lim_{u \uparrow 1} \bar{\chi}(u) = \lim_{u \uparrow 1} \frac{2 \log \mathbb{P}[F_X(X) > u]}{\log \mathbb{P}[F_X(X) > u, F_Y(y) > u]} - 1 = 2\eta - 1.$$

As  $0 < \eta \leq 1$ , we have  $-1 < \bar{\chi} \leq 1$ . For a bivariate normal dependence structure,  $\bar{\chi}$  is equal to the correlation coefficient  $\rho$ . Together, the two coefficients  $\chi$  and  $\bar{\chi}$  contain complementary information on the amount of tail dependence:

- If  $\chi = 0$  and  $-1 < \bar{\chi} \leq 1$ , the variables  $X$  and  $Y$  are asymptotically independent and  $\bar{\chi}$  quantifies the degree of dependence. The cases  $\bar{\chi} > 0$ ,  $\bar{\chi} = 0$ , and  $\bar{\chi} < 0$  correspond to positive association, near independence, and negative association, respectively.
- If  $\bar{\chi} = 1$  and  $0 < \chi \leq 1$ , the variables are asymptotically dependent, and  $\chi$  quantifies the degree of dependence.

Although the graphical procedures in this section are easy to use, it is difficult to draw a conclusion from the plots, and it might be better to consider formal tests of asymptotic independence (Section 4.2). For an overview of more recent and advanced techniques in asymptotic independence, see for example de Carvalho and Ramos (2012) or Bacro and Toulemonde (2013).

## 4.2 Testing for asymptotic (in)dependence

Above we have called the variables  $X$  and  $Y$  asymptotically independent if  $\ell(x, y) = x + y$ . A sufficient condition for asymptotic independence is that the coefficient of tail dependence,  $\eta$ , in (26) is less than unity.

These observations yield two possible approaches to test for asymptotic (in)dependence, depending on the choice of the null hypothesis. Suppose the function  $\mathcal{L}$  in (26) converges to a positive constant.

- Testing  $H_0 : \eta = 1$  versus  $H_1 : \eta < 1$  amounts to choosing asymptotic dependence as the null hypothesis.
- Testing  $H_0 : \forall(x, y), \ell(x, y) = x + y$  versus  $H_1 : \exists(x, y), \ell(x, y) < x + y$  means that asymptotic independence is taken as the null hypothesis.

**Asymptotic dependence as the null hypothesis.** In order to test for  $H_0 : \eta = 1$  versus  $H_1 : \eta < 1$ , the idea is to estimate  $\eta$  and to reject  $H_0$  if the estimated value of  $\eta$  is below some critical value. The estimator chosen in Draisma et al. (2004) is the maximum likelihood estimator,  $\hat{\eta}_{\text{MLE}}$ , obtained by fitting a generalized Pareto distribution to the  $k$  highest order statistics of the sample  $\hat{T}_1, \dots, \hat{T}_n$  in (27). The effective sample size,  $k$ , tends to infinity at a slower rate than  $n$ .

In order to derive the asymptotic distribution of  $\hat{\eta}_{\text{MLE}}$ , condition (26) needs to be refined. Consider the function

$$q(x, y) := \mathbb{P}[1 - F_X(X) < x, 1 - F_Y(Y) < y].$$

Note that the tail dependence coefficient  $\chi$  in (25) is given by  $\chi = \lim_{t \downarrow 0} q(t, t)/t$ . Equation (26) implies that the function  $t \mapsto q(t, t)$  is regularly varying at zero with index  $1/\eta$ , i.e.,  $\lim_{t \downarrow 0} q(tx, tx)/q(t, t) = x^{1/\eta}$  for  $x > 0$ . In Draisma et al. (2004), this relation is refined to *bivariate regular variation*: it is assumed that

$$\lim_{t \downarrow 0} \frac{q(tx, ty)}{q(t, t)} = c(x, y), \quad (x, y) \in (0, \infty)^2. \quad (28)$$

The limit function  $c$  is homogeneous of order  $1/\eta$ , i.e.,  $c(ax, ay) = a^{1/\eta}c(x, y)$  for  $a > 0$ . Clearly,  $c(1, 1) = 1$ . To control the bias of  $\hat{\eta}_{\text{MLE}}$ , a second-order refinement of (28) is needed: assume the existence of the limit

$$c_1(x, y) := \lim_{t \downarrow 0} \frac{\frac{q(tx, ty)}{q(t, t)} - c(x, y)}{q_1(t)},$$

for all  $x, y \geq 0$  with  $x + y > 0$ . Here,  $0 < q_1(t) \rightarrow t$  as  $t \rightarrow 0$ , while the limit function  $c_1$  is neither constant nor a multiple of  $c$ . The convergence is assumed to be uniform on  $\{(x, y) \in [0, \infty)^2 : x^2 + y^2 = 1\}$ .

Suppose that the function  $c$  has partial derivatives  $c_x := \partial c(x, y)/\partial x$  and  $c_y := \partial c(x, y)/\partial y$ . Let  $k$  be a sequence such that  $\sqrt{k} q_1(q^{-1}(k/n, k/n)) \rightarrow 0$  as  $n \rightarrow \infty$ , where  $q^{-1}(u, u) = t$  if and only if  $q(t, t) = u$ . Then  $\hat{\eta}_{\text{MLE}}$  is asymptotically normal,

$$\sqrt{k}(\hat{\eta}_{\text{MLE}} - \eta) \xrightarrow{d} \mathcal{N}(0, \sigma^2(\eta)),$$

the asymptotic variance being

$$\sigma^2(\eta) = (1 + \eta)^2 (1 - \chi) (1 - 2\chi c_x(1, 1)c_y(1, 1)).$$

Under the null hypothesis  $H_0 : \eta = 1$ , consistent estimators of  $\chi$  and  $c_x(1, 1)$  are given by

$$\hat{\chi} := \frac{k\hat{T}_{(k+1)}}{n}, \quad \hat{c}_x(1, 1) := \frac{\hat{p}^{5/4}}{n} \left[ \hat{T}_{(k+1)}^{(n, \hat{p}^{-1/4})} - \hat{T}_{(k+1)} \right],$$

where  $\hat{p} := k/\hat{\chi}$  and  $\hat{T}_{(k+1)}^{(n, u)}$  is the  $(k+1)$ th largest observation of

$$\hat{T}_i^{(n, u)} := \min \left( \frac{(1+u)}{1 - \hat{F}_X(X_i)}, \frac{1}{1 - \hat{F}_Y(Y_i)} \right), \quad i = 1, \dots, n.$$

The estimator  $\hat{c}_y(1, 1)$  is defined analogously to  $\hat{c}_x(1, 1)$ . The null hypothesis  $H_0 : \eta = 1$  is rejected at significance level  $\alpha$  if

$$\hat{\eta}_{\text{MLE}} \leq 1 - \frac{\hat{\sigma}}{\sqrt{k}} \Phi^{-1}(1 - \alpha), \quad (29)$$

where  $\hat{\sigma} = \hat{\sigma}(1)$  or  $\hat{\sigma} = \hat{\sigma}(\hat{\eta}_{\text{MLE}})$ .

Figure 8 shows the estimates of  $\hat{\eta}_{\text{MLE}}$  for a varying number of order statistics  $k$ , together with the lines defining the critical regions for the two tests in (29). We used the function `gpd.fit` from the `ismev` package (Heffernan and Stephenson, 2012). We see clearly that for JP Morgan versus Citibank, for every threshold, asymptotic dependence cannot be rejected; for JP Morgan versus IBM, asymptotic dependence is rejected at every threshold; and for Citibank versus IBM, results vary depending on the value of  $k$ .

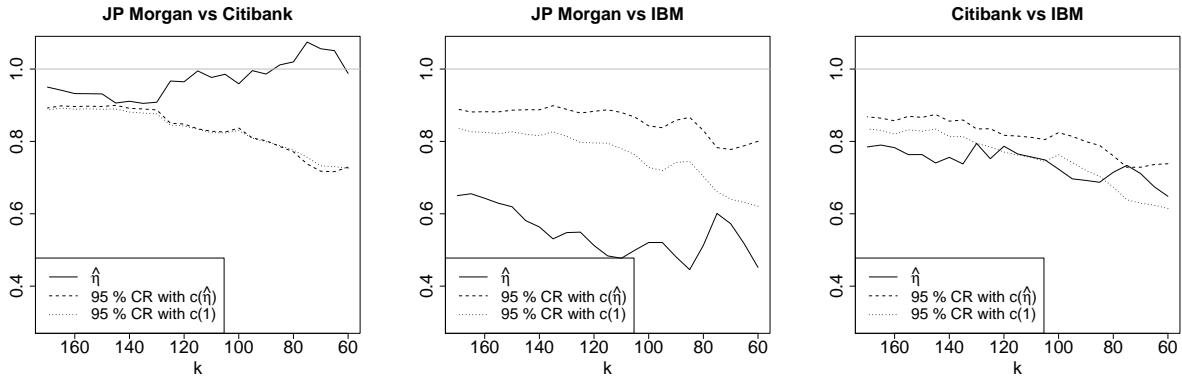


Figure 8: Estimates  $\hat{\eta}$  (solid lines) obtained by maximum likelihood based on excesses over the threshold  $\hat{T}_{(k+1)}$ , together with pointwise 95% critical regions (CR).

**Asymptotic independence as the null hypothesis.** Another approach for deciding between asymptotic dependence and asymptotic independence is to assume a null hypothesis of asymptotic independence, i.e.,  $H_0 : \forall(x, y), \ell(x, y) = x + y$ . Recall the nonparametric

estimator  $\widehat{\ell}$  in Section 3.2. A natural approach would be to reject  $H_0$  as soon as the difference between  $\widehat{\ell}$  and the function  $(x, y) \mapsto x + y$  is too large. However, from Einmahl et al. (2006) it follows that, under  $H_0$ ,

$$\sup_{x, y \in [0, 1]} \sqrt{k} \left| \widehat{\ell}(x, y) - (x + y) \right| \xrightarrow{\mathbb{P}} 0.$$

The limit being degenerate, it is not possible to compute critical values.

In Hüsler and Li (2009) another estimator of  $\ell$  is proposed, based on a division of the sample into two sub-samples. Note first that we can write  $\widehat{\ell}$  as

$$\widehat{\ell}(x, y) = \frac{1}{k} \sum_{i=1}^n \mathbb{1} \left\{ X_i > \widehat{F}_X^{-1} \left( 1 - \frac{kx}{n} \right) \text{ or } Y_i > \widehat{F}_Y^{-1} \left( 1 - \frac{ky}{n} \right) \right\}.$$

For convenience, assume that  $n$  is even. The first sub-sample  $X_1, \dots, X_{n/2}$  is compared with the  $[kx]$ -th largest order statistics of the second sub-sample  $X_{n/2+1}, \dots, X_n$ ; in other words, we use the estimator

$$\begin{aligned} \widetilde{\ell}_n(x, y) &= \frac{1}{k} \sum_{i=1}^{n/2} \mathbb{1} \left\{ X_i > \widehat{F}_X^{-1} \left( 1 - \frac{2kx}{n} \right) \text{ or } Y_i > \widehat{F}_Y^{-1} \left( 1 - \frac{2ky}{n} \right) \right\} \\ &= \frac{1}{k} \sum_{i=1}^{n/2} \mathbb{1} \left\{ \widetilde{R}_{i,X} > n/2 + 1 - kx \text{ or } \widetilde{R}_{i,Y} > n/2 + 1 - ky \right\}, \end{aligned}$$

where  $\widetilde{R}_{i,X}$  for  $i = 1, \dots, n/2$  denotes the rank of  $X_i$  among  $X_{n/2+1}, \dots, X_n$ . Under  $H_0$ , the estimator  $\widetilde{\ell}_n(x, y)$  converges in probability to  $x + y$ , for fixed  $x, y > 0$ . Define

$$D_n(x, y) := \sqrt{k} \{ \widetilde{L}_n(x, y) - (x + y) \}, \quad x, y \in [0, 1].$$

Then under  $H_0$ , assuming certain regularity conditions on  $\ell$  and  $k$  (Hüsler and Li, 2009, page 992), if  $k = k_n \rightarrow \infty$  as  $n \rightarrow \infty$ ,

$$\{D_n(x, y)\}_{x, y \in [0, 1]} \xrightarrow{d} \{W_1(2x) + W_2(2y)\}_{x, y \in [0, 1]}, \quad \text{as } n \rightarrow \infty,$$

where  $W_1$  and  $W_2$  are two independent Brownian motions. By the continuous mapping theorem,

$$\begin{aligned} T_{I,n} &:= \int_{[0,1]^2} D_n(x, y)^2 dx dy \xrightarrow{d} \int_{[0,1]^2} (W_1(2x) + W_2(2y))^2 dx dy =: T_I, \\ T_{S,n} &:= \sup_{x, y \in [0, 1]} |D_n(x, y)| \xrightarrow{d} \sup_{x, y \in [0, 1]} |W_1(2x) + W_2(2y)| =: T_S. \end{aligned}$$

We reject the null hypothesis of asymptotic independence at significance level  $\alpha$  if  $T_{I,n} > Q_{T_I}(1 - \alpha)$  or if  $T_{S,n} > Q_{T_S}(1 - \alpha)$ , where  $Q_{T_I}$  and  $Q_{T_S}$  represent the quantile functions of  $T_I$  and  $T_S$  respectively. Hüsler and Li (2009) compute  $Q_{T_I}(0.95) = 6.237$  and  $Q_{T_S}(0.95) = 4.956$ .

Figure 9 shows the values of the test statistics  $T_{I,n}$  and  $T_{S,n}$  for the three pairs of stock returns, together with the pointwise critical regions for a range of  $k$ -values. The results are in agreement with Figure 8: asymptotic independence is rejected for JP Morgan versus Citibank, asymptotic independence cannot be rejected for JP Morgan versus IBM, and conclusions for Citibank versus IBM depend on the value of  $k$ .

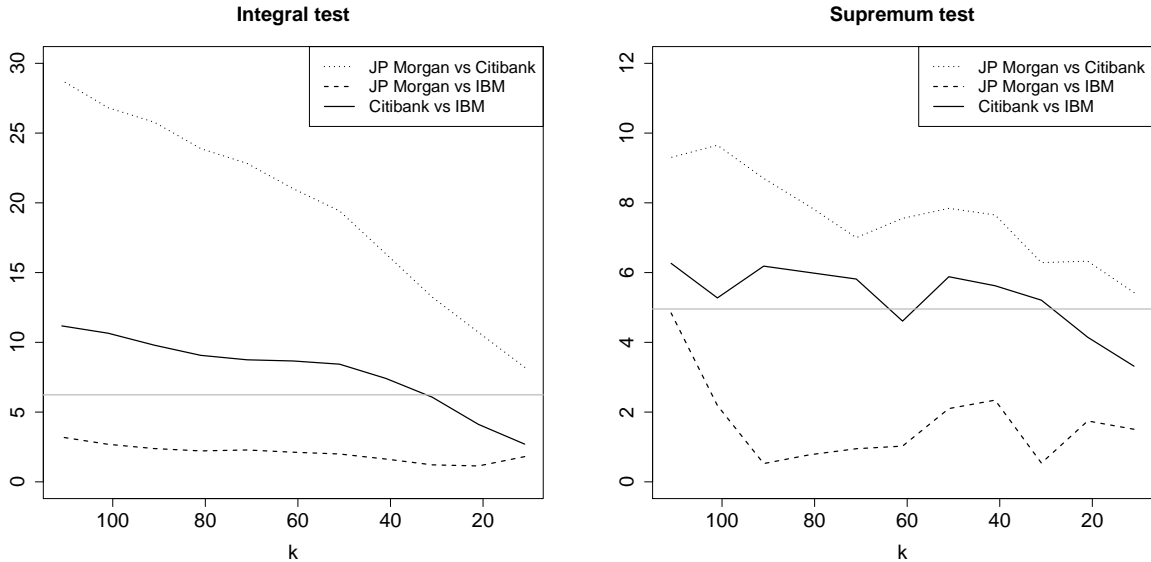


Figure 9: Test statistics  $T_{I,n}$  (left) and  $T_{S,n}$  (right) for a varying number of order statistics  $k$ . The horizontal lines define the critical regions  $Q_{T_I}(0.95) = 6.237$  and  $Q_{T_S}(0.95) = 4.956$ .

## Acknowledgments

The authors gratefully acknowledge funding by contract “Projet d’Actions de Recherche Concertées” No. 12/17-045 of the “Communauté française de Belgique” and by IAP research network Grant P7/06 of the Belgian government (Belgian Science Policy). Anna Kiriliouk and Michał Warchoł gratefully acknowledge funding from the Belgian Fund for Scientific Research (F.R.S. - FNRS).

## REFERENCES

- Bacro, J.-N. and G. Toulemonde (2013). Measuring and modelling multivariate and spatial dependence of extremes. *Journal de la Société Française de Statistique* 154(2), 139–155.
- Beirlant, J., Y. Goegebeur, J. Segers, and J. Teugels (2004). *Statistics of Extremes: Theory and Applications*. Wiley.
- Bücher, A., J. Segers, and S. Volgushev (2014). When uniform weak convergence fails: empirical processes for dependence functions and residuals using epi- and hypographs. *The Annals of Statistics* (to appear), <http://arxiv.org/abs/1305.6408>.
- Capéraà, P. and A.-L. Fougères (2000). Estimation of a bivariate extreme value distribution. *Extremes* 3(4), 311–329.
- Coles, S., J. Heffernan, and J. Tawn (1999). Dependence measures for extreme value analyses. *Extremes* 2(4), 339–365.

- Coles, S. and J. Tawn (1991). Modelling extreme multivariate events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 53, 377–392.
- de Carvalho, M., B. Oumow, J. Segers, and M. Warchoř (2013). A euclidean likelihood estimator for bivariate tail dependence. *Communications in Statistics - Theory and Methods* 42(7), 1176–1192.
- de Carvalho, M. and A. Ramos (2012). Bivariate extreme statistics, ii. *REVSTAT - Statistical Journal* 11(1), 83–107.
- de Haan, L. and J. de Ronde (1998). Sea and wind: Multivariate extremes at work. *Extremes* 1, 7–45.
- de Haan, L. and A. Ferreira (2006). *Extreme Value Theory: an Introduction*. Springer-Verlag Inc.
- de Haan, L. and S. I. Resnick (1977). Limit theory for multivariate sample extremes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 40(4), 317–337.
- Draisma, G., H. Drees, A. Ferreira, and L. De Haan (2004). Bivariate tail estimation: dependence in asymptotic independence. *Bernoulli* 10(2), 251–280.
- Drees, H. and X. Huang (1998). Best attainable rates of convergence for estimators of the stable tail dependence function. *Journal of Multivariate Analysis* 64(1), 25–47.
- Einmahl, J., L. de Haan, and D. Li (2006). Weighted approximations of tail copula processes with application to testing the bivariate extreme value condition. *The Annals of Statistics* 34, 1987–2014.
- Einmahl, J., A. Kiriliouk, A. Krajina, and J. Segers (2014). An M-estimator of spatial tail dependence. Available on <http://arxiv.org/abs/1403.1975>.
- Einmahl, J., A. Krajina, and J. Segers (2012). An M-estimator for tail dependence in arbitrary dimensions. *The Annals of Statistics* 40, 1764–1793.
- Einmahl, J. H., L. de Haan, and V. I. Piterbarg (2001). Nonparametric estimation of the spectral measure of an extreme value distribution. *The Annals of Statistics* 29(5), 1401–1423.
- Einmahl, J. H. J. and J. Segers (2009). Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. *The Annals of Statistics* 37(5B), 2953–2989.
- Falk, M., J. Hüsler, and R.-D. Reiss (2011). *Laws of small numbers: extremes and rare events* (extended ed.). Birkhäuser/Springer Basel AG, Basel.
- Fisher, R. A. and L. H. C. Tippett (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society*, Volume 24, pp. 180–190. Cambridge Univ Press.

- Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of mathematics*, 423–453.
- Heffernan, J. E. and A. G. Stephenson (2012). *ismev: An Introduction to Statistical Modeling of Extreme Values*. R package version 1.38.
- Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics* 3, 1163–1174.
- Huang, X. (1992). *Statistics of bivariate extreme values*. Ph. D. thesis, Tinbergen Institute Research Series.
- Hüsler, J. and D. Li (2009). Testing asymptotic independence in bivariate extremes. *Journal of Statistical Planning and Inference* 139, 990–998.
- Ledford, A. W. and J. A. Tawn (1996). Statistics for near independence in multivariate extreme values. *Biometrika* 83, 169–187.
- Pickands, J. (1981). Multivariate extreme value distributions. In *Bulletin of the International Statistical Institute*, pp. 859–878.
- Resnick, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer, New York.
- Segers, J. (2012). Max-stable models for multivariate extremes. *REVSTAT Statistical Journal* 10, 61–82.
- Sibuya, M. (1960). Bivariate extreme statistics, I. *Annals of the Institute of Statistical Mathematics* 11(2), 195–210.