I N S T I T U T D E S T A T I S T I Q U E

B I O S T A T I S T I Q U E E T

S C I E N C E S A C T U A R I E L L E S

(I S B A)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

2013/26

Measuring Association and Dependence Between Random Vectors

GROTHE, O., SCHNIEDERS, J. and J. SEGERS

Measuring Association and Dependence Between Random Vectors

Oliver Grothe[a,*], Julius Schnieders[a], Johan Segers[b,1]

[a] *University of Cologne, Department of Economic and Social Statistics, Albertus-Magnus-Platz, 50923 Cologne, Germany*
[b] *Université catholique de Louvain, Institut de statistique, biostatistique et sciences actuarielles, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium*

Abstract

Measures of association are suggested between two random vectors. The measures are copula-based and therefore invariant with respect to the univariate marginal distributions. The measures are able to capture positive as well as negative association. In case the random vectors are just random variables, the measures reduce to Kendall's tau or Spearman's rho. Nonparametric estimators, based on ranks, for the measures are derived. Their large-sample asymptotics are derived and their small-sample behaviour is investigated by simulation. The measures are applied to characterize strength and direction of association of northern and southern European bond markets during the recent Euro crisis as well as association of stock markets with bond markets.

Keywords: Association, Copula, Dependence, Kendall's tau, Spearman's rho, U-statistic

[*] Corresponding author

Email addresses: grothe@statistik.uni-koeln.de (Oliver Grothe), schnieders@wiso.uni-koeln.de (Julius Schnieders), johan.segers@uclouvain.be (Johan Segers)

[1] Financial support from the contract "Projet d'Actions de Recherche Concertées" No. 12/17-045 of the "Communauté française de Belgique", granted by the "Académie universitaire Louvain", and from IAP research network grant No. P7/06 of the Belgian governement (Belgian Science Policy) is gratefully acknowledged.

1. Introduction

In many applications, measuring the association between different random vectors is of special interest. However, most of the literature on association of random vectors concerns association *within* a random vector, that is, among its univariate components. Classical measures for bivariate vectors are Spearman's rho, Kendall's tau, Blomqvist's beta, Gini's gamma and several lesser known measures. For most of these, multivariate extensions to higher dimensions exist (Joe, 1990, 1997; Schmid and Schmidt, 2007); see Schmid et al. (2010) for a survey. A measure specifically designed for the multivariate case is Kendall's W coefficient of concordance (Kendall and Babington Smith, 1939), recently studied using copulas by Grothe and Schmid (2011) and Marozzi (2013).

The question of measuring dependence and association *between* random vectors is less investigated. It starts with the canonical correlation measure of Hotelling (1936), a measure of linear dependence between random vectors. Another classical measure is the RV coefficient (Escoufier, 1973; Robert and Escoufier, 1976). Most articles concern tests of independence between random vectors rather than actually measuring the association. For this purpose, matrices of Spearman and Kendall correlation coefficients (El Maache and Lepage, 2003) or other rank-based statistics (Puri et al., 1970) or averages of Spearman's rho and Kendall's tau (Hays, 1960) are used. Other statistics are based on distances of empirical copulas (Quessy, 2010) or densities (Székely et al., 2007; Székely and Rizzo, 2009).

Instead of testing for independence of random vectors, the focus of this paper is the development of margin-free measurement of association *between* random vectors. Such measures may be informative when analyzing comovements or contagion of financial markets (Forbes and Rigobon, 2002). In such applications, the random vectors correspond to, e.g., groups of assets from different markets (like stock markets, commodity and energy markets) or of economic factors as unemployment rate, GDP or inflation rate. Being invariant with respect to increasing transformations of the margins, the measures then quantify the association between groups of variables as a whole instead of the association between individual assets or factors.

The dependence between markets may be of positive or of negative nature and it will be important to be able to make this distinction. Consider for example a vector of returns of assets from the banking sector and a vector of returns from the automobile industry. Association between these vectors will most likely be positive. On the other hand, consider a vector of interest rates and a vector of assets returns. Here, standard economic theory predicts a negative association, since, usually, interest rates are assumed to increase when the stock market is bearish (prices trending down) and decrease when the stock market is bullish (prices trending up). Measures which are interesting for our purpose should therefore be able to distinguish between positive and negative association. As canonical correlation, the RV coefficient and distance correlation are always positive, they are of limited use for our purpose.

We therefore develop two measures of association, generalizations of Spearman's rho and Kendall's tau, which both have the desired properties. In case the univariate marginal distributions of the random vectors are continuous, which is the focus of this paper, the measures depend only on the underlying copulas and are therefore margin-free. The basic idea of the measures is that independence of vectors of length p and q means that the joint $(p+q)$-variate cumulative distribution function (cdf) factorizes into the product of the p-variate and q-variate marginal cdfs. In general, the difference between the $(p+q)$-variate cdf and the product of p-variate and q-variate marginal cdfs contains information on the strength of the dependence and the nature of the association, positive or negative. Integrating out this difference and normalizing it appropriately yields our measures of association.

The definitions of our association measures between random vectors are inspired by the ones for random variables, notably Spearman's rho and Kendall's tau. Association is measured by a normalized integral of the difference of the joint distribution function of the two random vectors and their distribution function under independence, i.e. the product of their marginal distribution functions. In Schmid et al. (2010), multivariate versions of Spearman's rho for association within a random vector are defined via such integrals as well.

The existing and new measures of association somehow average out the pairwise associations of the variables involved. This means that if the measures are positive, positive association is predominant within the variables and vice versa. However, positive values of the measures do not mean that all pairwise associations involved are positive. Neither do negative values of the measures imply negativity of all pairwise associations.

A crucial point when measuring association in this way is how exactly to integrate out the difference between the joint and the marginal cdfs. We propose weightings which lead to generalizations of Spearman's rho and

Kendall's tau. Furthermore, the weightings ensure desirable properties with respect to the concordance ordering, properties which will be discussed in detail. Concordant variables tend to be all large together or all small together (Joe, 1990). For fixed p-variate and q-variate marginal cdfs, more concordant joint $(p+q)$-variate distributions lead to larger values of our measure. Furthermore, if the joint $(p+q)$-variate copula is elliptical, the association measures are increasing functions of the pairwise correlations between components of the first and the second random vectors.

We propose nonparametric inference procedures for our association measures based upon U-statistics. The use of U-statistics in this context is quite natural; for Spearman's rho, it goes back to Hoeffding (1948, p. 318). Because the measures are copula based, the inference procedures depend on the data only through the ranks, lending some robustness to the methodology. We provide explicit estimators, establish their asymptotic normality, and show how to estimate their asymptotic variance in a computationally efficient way. The small-sample performance is assessed through Monte Carlo simulations and the methodology is illustrated via a case study measuring dependence between bond and stock markets as well as between bond markets of northern and southern European countries.

The structure of the paper is as follows: In Section 2, we describe the notation and terminology used within the paper. In Section 3, we define and motivate population versions of our measures. Estimators for the measures and their variances are introduced in Section 4. Section 5 contains the simulation study and an empirical examples with financial data. Software for the estimators, written in R and C, is available on request. Section 6 concludes.

2. Preliminaries

To motivate our approach, we recall Spearman's rho and Kendall's tau for random variables X and Y with continuous distributions. Let their joint and marginal distribution functions be denoted by $F(x) = \Pr[X \leq x]$, $G(y) = \Pr[Y \leq y]$ and $H(x, y) = \Pr[X \leq x, Y \leq y]$, for $x, y \in \mathbb{R}$. Then Spearman's rho can be expressed as

$$\rho_S = 12 \int_{\mathbb{R}} \int_{\mathbb{R}} \{H(x, y) - F(x)\, G(y)\} \,\mathrm{d}x \,\mathrm{d}y$$

while Kendall's tau is given by

$$\tau = 4 \left(\int_{\mathbb{R}^2} H(x,y)\,\mathrm{d}H(x,y) - \int_{\mathbb{R}} \int_{\mathbb{R}} F(x)\,G(y)\,\mathrm{d}F(x)\,\mathrm{d}G(y) \right).$$

For both measures, there exist generalizations to d-dimensional random vectors (Schmid et al., 2010). In all cases, the measures of association are constructed by averaging out the difference between the joint distribution function, $H(x,y)$, and the distribution function in case of independence, $F(x)\,G(y)$.

Now, let p and q be positive integers, let $d = p + q$, and let H be a d-variate cumulative distribution function with p-variate and q-variate margins F and G, i.e. if (X, Y) is a $(p+q)$-variate random vector with distribution function H, then $H(x,y) = \Pr[X \le x, Y \le y]$, $F(x) = \Pr[X \le x]$ and $G(y) = \Pr[Y \le y]$ for $x \in \mathbb{R}^p$ and $y \in \mathbb{R}^q$. Inequalities, orderings and intervals concerning vectors are to be interpreted componentwise.

Our aim is to study the dependence between the random vectors X and Y. Starting point is the observation that if X and Y are independent, then $H(x,y) = F(x)\,G(y)$. As for the measures mentioned above, the difference $H(x,y) - F(x)\,G(y)$ therefore contains information on the strength and the nature of the dependence of X and Y. If this difference tends to be positive, the vectors X and Y are positively associated; otherwise they are negatively associated. Measures of association are obtained by integrating the difference over $(x,y) \in \mathbb{R}^{p+q}$ according to some measure and normalizing the integrated difference in a meaningful way.

Let $H_1, \ldots, H_d$ be the univariate marginal cdfs of H; let $F_1, \ldots, F_p$ be the univariate marginal cdfs of F and $G_1, \ldots, G_q$ those of G. Observe that $H_j = F_j$ for $j \in \{1, \ldots, p\}$ and $H_{p+j} = G_j$ for $j \in \{1, \ldots, q\}$. If all univariate margins H_j are continuous, then H has a unique copula, C, and $H(z) = C(H_1(z_1), \ldots, H_d(z_d))$ for $z \in \mathbb{R}^d$ (Nelsen, 2006). In that case, F and G admit unique copulas too, given by $A(u) = C(u, 1_q)$ and $B(v) = C(1_p, v)$ for $u \in [0,1]^p$ and $v \in [0,1]^q$, where $1_k = (1, \ldots, 1) \in \mathbb{R}^k$. In that case, we have $H(x,y) = C(u,v)$, $F(x) = A(u)$ and $G(y) = B(v)$ with $u_j = F_j(x_j)$ for $j \in \{1, \ldots, p\}$ and $v_j = G_j(y_j)$ for $j \in \{1, \ldots, q\}$.

Let (X_1, Y_2), (X_2, Y_2) and (X_3, Y_3) denote independent copies of (X, Y), i.e., independent $(p+q)$-variate random vectors with common distribution function H and margins F and G of X and Y, respectively. The right-continuous survival function of a multivariate cdf K will be denoted by $\bar{K}$, that is, $\bar{K}(x) = \mu_K([x,\infty))$, where μ_K denotes the probability measure defined by $\mu_K((-\infty,x]) = K(x)$. Product measure will be denoted by '$\otimes$'.

Let μ and ν be probability measures on $\mathbb{R}^d$, for instance the laws induced by H or $F \otimes G$ or $H_1 \otimes \cdots \otimes H_d$. In order to study the dependence between X and Y, quantities of the following form will arise naturally:

$$\int_{\mathbb{R}^d} H(x,y)\,\mathrm{d}\mu(x,y) - \int_{\mathbb{R}^d} F(x)\,G(y)\,\mathrm{d}\nu(x,y), \tag{2.1}$$

$$\int_{\mathbb{R}^d} \bar{H}(x,y)\,\mathrm{d}\mu(x,y) - \int_{\mathbb{R}^d} \bar{F}(x)\,\bar{G}(y)\,\mathrm{d}\nu(x,y). \tag{2.2}$$

A natural choice would be to set both μ and ν equal to H. However, even if $p = q = 1$, we would not get a useful measure of association. Up to a multiplicative constant, the result would be equal to a linear combination of Kendall's tau and Spearman's rho proportional to so-called expected quadrant dependence, see Nelsen (2006, Exercise 5.25, page 190). However, this measure does not respect concordance ordering, see Figure 5.5 in the same book. We are grateful to Roger B. Nelsen for having pointed this out.

In the next section, we will consider choices of μ and ν that will lead to multivariate generalizations of Kendall's tau and Spearman's rho.

The following formula will be exceedingly useful: if K and L are two k-variate distribution functions, then

$$\int_{\mathbb{R}^k} K(y)\,\mathrm{d}L(y) = \mu_{(K\otimes L)}(\{(x,y) \in \mathbb{R}^k \times \mathbb{R}^k : x \le y\}) = \int_{\mathbb{R}^k} \bar{L}(x)\,\mathrm{d}K(x), \tag{2.3}$$

where $\mu_{(K\otimes L)}$ denotes the probability measure on $\mathbb{R}^k \times \mathbb{R}^k$ determined by $\mu((-\infty,x] \times (-\infty,y]) = K(x)\,L(y)$ for $x,y \in \mathbb{R}^k$.

3. Measures of association

Measures of association between random vectors X and Y will be defined via specifications of μ and ν in (2.1) and (2.2). Ideally, such measures should reduce to classical univariate measures of association for $p = q = 1$. We therefore consider a measure reducing to Spearman's rho in the next section and a measure reducing to Kendall's tau in Section 3.2. The measures fulfill the following desirable properties, inspired from those of association measures within random vectors (Schmid et al., 2010):

1. The measures take values in the interval $[-1,1]$.

2. The measures are invariant with respect to permutations of the components within X and Y.
3. The measures are invariant with respect to increasing transformations of the components of the random vectors, i.e., they are copula based.
4. If the vectors X and Y are independent, the measures are zero.
5. The measures respect concordance ordering, i.e., for two d-dimensional copulas C_1 and C_2 with the same p- and q-dimensional margins A and B, if $C_1 \prec_{\mathcal{C}} C_2$, then C_1 should have a smaller measure than C_2. Here, $C_1 \prec_{\mathcal{C}} C_2$ is the standard definition of concordance ordering for multivariate copulas (Joe, 1990): $C_1 \prec_{\mathcal{C}} C_2$ if and only if $C_1 \leq C_2$ and $\bar{C}_1 \leq \bar{C}_2$ pointwise. Note that by (2.3), $C_1 \prec_{\mathcal{C}} C_2$ implies $\int C_1 \,\mathrm{d}C_1 \leq \int C_2 \,\mathrm{d}C_1 = \int \bar{C}_1 \,\mathrm{d}C_2 \leq \int \bar{C}_2 \,\mathrm{d}C_2 = \int C_2 \,\mathrm{d}C_2$.

Concerning the last point, compare two scenarios of random vectors X and Y, the two scenarios differing only with respect to the dependence between X and Y. If in one scenario the vector (X, Y) is more concordant than in the other scenario, then the association measures are larger in the more concordant scenario. Furthermore, $C_1 \prec_{\mathcal{C}} C_2$ implies $C_1^I \prec_{\mathcal{C}} C_2^I$ for all subvectors of variables with indices in $I \subset \{1, \ldots, d\}$ (Joe, 1990). Therefore, changing the size of the vectors X and Y in both scenarios, e.g., removing dimensions, does not influence which scenario has the larger dependence measure.

There are also implications regarding pairwise correlations of the involved variables. Let C_1 and C_2 be as above and let them be elliptical copulas with correlation matrices $R_1 = (r_{1,ij})_{ij}$ and $R_2 = (r_{2,ij})_{ij}$, respectively. Then it follows from Das Gupta et al. (1972) that $C_1 \prec_{\mathcal{C}} C_2$ as soon as $r_{2,ij} \geq r_{1,ij}$ for $1 \leq i \leq p$ and $p+1 \leq j \leq d$. Thus, for this rich class of dependence structures, our measures are nondecreasing in all pairwise correlation coefficients of the joint copula between the vectors X and Y.

3.1. Measures inspired by Spearman's rho

Setting μ and ν in (2.1)–(2.2) equal to $F \otimes G$ yields multivariate generalizations of Spearman's rho.

Lemma 3.1.

$$\begin{aligned}\iint \{H(x,y) - F(x)\,G(y)\}\,\mathrm{d}F(x)\,\mathrm{d}G(y) &= \operatorname{cov}(\mathbb{1}_{\{X_1 \le X_2\}},\, \mathbb{1}_{\{Y_1 \le Y_3\}}) \\ &= \operatorname{cov}\big(\bar{F}(X),\, \bar{G}(Y)\big), \\ \iint \{\bar{H}(x,y) - \bar{F}(x)\,\bar{G}(y)\}\,\mathrm{d}F(x)\,\mathrm{d}G(y) &= \operatorname{cov}(\mathbb{1}_{\{X_1 \ge X_2\}},\, \mathbb{1}_{\{Y_1 \ge Y_3\}}) \\ &= \operatorname{cov}\big(F(X),\, G(Y)\big).\end{aligned}$$

Proof. The equalities involving the indicator variables follow from the fact that the joint distribution of $(X_1, Y_1; X_2; Y_3)$ is $H \otimes F \otimes G$. The other equalities follow from equation (2.3).

The proper normalization of the covariances in Lemma 3.1 suggests itself. We define association measures by

$$\rho[X,Y] = \operatorname{cor}\big(F(X),\, G(Y)\big) = \frac{\operatorname{cov}\big(F(X),\, G(Y)\big)}{\sqrt{\operatorname{var}(F(X))\,\operatorname{var}(G(Y))}}, \tag{3.1}$$

$$\bar{\rho}[X,Y] = \operatorname{cor}\big(\bar{F}(X),\, \bar{G}(Y)\big) = \frac{\operatorname{cov}\big(\bar{F}(X),\, \bar{G}(Y)\big)}{\sqrt{\operatorname{var}(\bar{F}(X))\,\operatorname{var}(\bar{G}(Y))}}. \tag{3.2}$$

If $p = q = 1$ and if the univariate margins are continuous, both association measures simplify to the population version of Spearman's rho. For general dimensions p and q, the denominators in (3.1) and (3.2) can be expressed in terms of indicator functions too. By Lemma 3.1, we find for instance

$$\begin{aligned}\operatorname{var}\big(F(X)\big) &= \operatorname{cov}(\mathbb{1}_{\{X_1 \ge X_2\}},\, \mathbb{1}_{\{X_1 \ge X_3\}}), \\ \operatorname{var}\big(\bar{F}(X)\big) &= \operatorname{cov}(\mathbb{1}_{\{X_1 \le X_2\}},\, \mathbb{1}_{\{X_1 \le X_3\}}),\end{aligned}$$

and similarly for $\operatorname{var}\big(G(Y)\big)$ and $\operatorname{var}\big(\bar{G}(Y)\big)$.

Note that the measures in (3.1) and (3.2) are not equal to the correlations of the indicator functions in Lemma 3.1. The latter correlations have the drawback that even if X and Y were perfectly dependent (for instance $p = q$ and $X = Y$), they would not be equal to unity, by independence of Y_3 and $(X_1, Y_1; X_2)$.

If the joint copula of (X, Y) is symmetric in the sense that it is equal to its own survival copula, the measures ρ and $\bar{\rho}$ actually coincide. This is the case for elliptical copulas, for instance the Gaussian copula, and the Frank copula. For other dependence structures, ρ and $\bar{\rho}$ are generally different. Since neither of the two measures seems to be preferable to the other, we suggest to take their average

$$\rho^* = \frac{\rho + \bar{\rho}}{2}.$$

Although this is somehow arbitrary, it is common in the literature on measures *within* random vectors, where also two generalized Spearman measures based on F and $\bar{F}$ arise (Schmid et al., 2010).

3.2. A measure inspired by Kendall's tau

Setting μ equal to H and ν equal to $F \otimes G$ in equations (2.1) and (2.2) yields a multivariate generalization of Kendall's tau.

Lemma 3.2.

$$\begin{aligned}
&\operatorname{cov}\left(\mathbb{1}_{\{X_1 \le X_2\}}, \mathbb{1}_{\{Y_1 \le Y_2\}}\right) \\
&= \int H(x,y)\,\mathrm{d}H(x,y) - \int F(x)\,\mathrm{d}F(x) \int G(y)\,\mathrm{d}G(y) \\
&= \int \bar{H}(x,y)\,\mathrm{d}H(x,y) - \int \bar{F}(x)\,\mathrm{d}F(x) \int \bar{G}(y)\,\mathrm{d}G(y).
\end{aligned}$$

Proof. For the first equality, note that both expressions are equal to

$$\begin{aligned}
&\Pr[X_1 \le X_2, Y_1 \le Y_2] - \Pr[X_1 \le X_2] \Pr[Y_1 \le Y_2] \\
&\qquad = \mathrm{E}[H(X,Y)] - \mathrm{E}[F(X)]\,\mathrm{E}[G(Y)].
\end{aligned}$$

To prove the second equality, apply equation (2.3).

The appropriate normalization is found by considering the correlation rather than the covariance. We arrive at a multivariate generalization of Kendall's tau:

$$\tau[X,Y] = \operatorname{cor}\left(\mathbb{1}_{\{X_1 \le X_2\}}, \mathbb{1}_{\{Y_1 \le Y_2\}}\right) = \frac{p_{X,Y} - p_X\, p_Y}{\sqrt{p_X(1-p_X)\, p_Y(1-p_Y)}} \tag{3.3}$$

where

$$p_X = \Pr[X_1 \le X_2], \quad p_Y = \Pr[Y_1 \le Y_2], \quad p_{X,Y} = \Pr[X_1 \le X_2, Y_1 \le Y_2]. \tag{3.4}$$

If $p = q = 1$ and if the margins are continuous, we have $p_X = p_Y = 1/2$ and $\tau[X, Y]$ is just Kendall's tau. In the general case, the probabilities p_X, p_Y and $p_{X,Y}$ are connected to a multivariate version of Kendall's tau *within* a random vector (Quessy et al., 2013).

4. Inference

Let (X_i, Y_i), $i \in \{1, \dots, n\}$, be an independent random sample from H. We construct estimators of the measures of association introduced in Section 3. It is convenient to start with inference on the generalization of Kendall's tau.

4.1. Inference on the generalization of Kendall's tau

Inference on $\tau[X, Y]$ in (3.3) boils down to inference on the probabilities p_X, p_Y and $p_{X,Y}$ in (3.4). These can be estimated by U-statistics:

$$\hat{p}_{W,n} = \frac{1}{n(n-1)} \sum_{\substack{i,j \in \{1,\dots,n\} \\ i \neq j}} \mathbb{1}_{\{W_i \leq W_j\}}, \tag{4.1}$$

where 'W' represents 'X', 'Y', or '(X, Y)'. The estimator for $\tau[X, Y]$ then becomes

$$\hat{\tau}_n[X, Y] = f(\hat{p}_{X,n},\, \hat{p}_{Y,n},\, \hat{p}_{(X,Y),n}), \tag{4.2}$$

$$\text{where} \qquad f(x, y, z) = \frac{z - xy}{\sqrt{x(1-x)\, y(1-y)}}. \tag{4.3}$$

By Hoeffding's decomposition theorem (Hoeffding, 1948), we find

$$\sqrt{n}\big(\hat{p}_{W,n} - p_W\big) = \frac{2}{\sqrt{n}} \sum_{i=1}^{n} (h_{1,W}(W_i) - p_W) + o_p(1) \qquad (n \to \infty), \tag{4.4}$$

where the functions $h_{1,X}$, $h_{1,Y}$ and $h_{1,(X,Y)}$ are defined by

$$h_{1,W}(w) = \frac{1}{2}(\Pr[w \leq W] + \Pr[W \leq w]),$$

and with 'w' equal to 'x', 'y' or '(x,y)'. It follows that

$$\sqrt{n}\begin{pmatrix}\hat{p}_{X,n}-p_X\\ \hat{p}_{Y,n}-p_Y\\ \hat{p}_{(X,Y),n}-p_{X,Y}\end{pmatrix} \rightsquigarrow N_3(0,4\Sigma) \qquad (n\to\infty),$$

a three-variate centered Gaussian distribution; $\Sigma \in \mathbb{R}^{3\times 3}$ is the covariance matrix of the random vector $(h_{1,X}(X_1), h_{1,Y}(Y_1), h_{1,(X,Y)}(X_1,Y_1))'$. Asymptotic normality of $\hat{\tau}_n[X,Y]$ then follows from the delta-method:

$$\sqrt{n}\big(\hat{\tau}_n[X,Y]-\tau[X,Y]\big) \rightsquigarrow N(0,\sigma_\tau^2), \qquad (n\to\infty),$$
$$\text{where} \qquad \sigma_\tau^2 = 4(\nabla f)'\Sigma(\nabla f),$$

with $\nabla f \in \mathbb{R}^{3\times 1}$ the gradient of f at the point $(p_X, p_Y, p_{X,Y})'$.

To estimate the asymptotic variance, σ_τ^2, we opt for a simple plug-in approach: estimate ∇f by evaluating the gradient of f at $(\hat{p}_{X,n}, \hat{p}_{Y,n}, \hat{p}_{(X,Y),n})'$ and estimate Σ by the sample covariance matrix of the triples

$$\big(\hat{h}_{1,X,n}(X_i),\ \hat{h}_{1,Y,n}(Y_i),\ \hat{h}_{1,(X,Y),n}(X_i,Y_i)\big), \qquad i\in\{1,\ldots,n\},$$

with

$$\hat{h}_{1,W,n}(W_i) = \frac{1}{n-1}\sum_{\substack{j\in\{1,\ldots,n\}\\ j\neq i}} \frac{1}{2}(\mathbb{1}_{\{W_i\le W_j\}} + \mathbb{1}_{\{W_j\le W_i\}}). \tag{4.5}$$

The estimator of Σ can be motivated by the jackknife methodology (Efron and Stein, 1981), but an easier justification is that up to $O(1/n)$ it is equal to another U-statistic, representing the elements of Σ as estimable parameters of H of degree 3.

4.2. Inference on the generalization of Spearman's rho

By Lemma 3.1, the association measure inspired on Spearman's rho can be expressed in terms of probabilities:

$$\begin{aligned}\rho[X,Y] &= \frac{\operatorname{cov}(\mathbb{1}_{\{X_1\ge X_2\}}, \mathbb{1}_{\{Y_1\ge Y_3\}})}{\sqrt{\operatorname{cov}(\mathbb{1}_{\{X_1\ge X_2\}}, \mathbb{1}_{\{X_1\ge X_3\}})\ \operatorname{cov}(\mathbb{1}_{\{Y_1\ge Y_2\}}, \mathbb{1}_{\{Y_1\ge Y_3\}})}}\\ &= \frac{\Pr[X_1\ge X_2,\ Y_1\ge Y_3]-p_X\,p_Y}{\sqrt{(\Pr[X_1\ge X_2,\ X_1\ge X_3]-p_X^2)\,(\Pr[Y_1\ge Y_2,\ Y_1\ge Y_3]-p_Y^2)}},\end{aligned}$$

with p_X and p_Y as in (3.4). Note that for two independent copies X_1 and X_2 of X, $p(X_1 \geq X_2) = p(X_1 \leq X_2) := p_X$. For $\bar{\rho}[X, Y]$, just replace '$\geq$' by '$\leq$'; for brevity, we will only consider $\rho[X, Y]$ in this subsection.

As in Subsection 4.1, a nonparametric estimator $\hat{\rho}_n[X, Y]$ of $\rho[X, Y]$ can be constructed by replacing the probabilities in the above expression by their sample versions, relying on U-statistics.

Estimation of p_X and p_Y has been treated in Subsection 4.1. The three other probabilities appearing in the above expression for $\rho[X, Y]$ are of the form

$$q_{V,W} = \Pr[V_1 \geq V_2,\, W_1 \geq W_3],$$

with 'V' and 'W' equal to 'X' or 'Y'. These probabilities are estimable parameters of H of degree 3 and they can be estimated by appropriate U-statistics:

$$\hat{q}_{V,W,n} = \frac{1}{n(n-1)(n-2)} \sum_{\substack{\{i,j,k\} \subset \{1,\ldots,n\} \\ \sharp\{i,j,k\}=3}} \mathbb{1}_{\{V_i \geq V_j,\, W_i \geq W_k\}}. \tag{4.6}$$

As before, joint asymptotic normality of the estimators of the five probabilities involved in the expression for $\rho[X, Y]$ follows from Hoeffding's decomposition theorem. The entries of the limiting covariance matrix are estimable parameters of H of degree 5. Employing similar simplifications as above, the U-statistic estimators of these limiting (co)variances can up to $O(1/n)$ be written in a way which requires at most double loops over the sample. Asymptotic normality

$$\sqrt{n}\big(\hat{\rho}_{S,n}[X, Y] - \rho[X, Y]\big) \rightsquigarrow N(0, \sigma_\rho^2) \qquad (n \to \infty)$$

follows from the delta method. The estimation procedure for the asymptotic variance σ_ρ^2 is analogous to the one for σ_τ^2 in Subsection 4.1. However, we give a more technical description of the estimation of ρ in Appendix A.

If the random vectors (X_i, Y_i), $i \in \{1, \ldots, n\}$, are no longer iid but merely form a stationary time series, the asymptotic distributions of the estimators of the association measures will be different from the ones stated above. One possible approach is to represent the association measures and their estimators as smooth functionals of the (empirical) copula. Asymptotic properties of the empirical copula process for weakly dependent, stationary time series together with the functional delta method then yield asymptotic normality of the estimators, albeit with a different asymptotic variance. To estimate their asymptotic variance, block bootstrap procedures could be applied (Bücher and Volgushev, 2013).

5. Numerical examples

5.1. Simulation study

In the simulation study we examine how well the proposed measures and their asymptotic variances may be estimated from finite samples. To this end, we estimate τ and ρ in different setups and compare sample variances of Monte Carlo repetitions with estimates of the asymptotic variances.

For our study, we consider observations of (X, Y) with different dimensions, different sample sizes and different parameters. The dependence structure is alternatively given by the Gaussian copula (Joe, 1997) and by D-vine structures built from bivariate Clayton copulas (Clayton, 1978). The Gaussian copula is defined by

$$C_\Theta(u_1, \ldots, u_d) := \Phi_\Theta\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\right),$$

where Φ_Θ is the distribution function of the multivariate normal distribution with zero mean, unit variances and positive definite correlation matrix $\Theta = (\theta_{ij})_{i,j=1,\ldots,d}$ and Φ^{-1} denotes the quantile function of the univariate standard normal distribution. To reduce the number of parameters in our model, we only consider the case of equi-correlation, i.e., $\theta_{ij} = \theta$ for $i, j = 1, \ldots, d$ and $i \neq j$, where $-1/(d-1) < \theta < 1$.

A D-vine construction of a copula $C(u_1, \ldots, u_d)$ is a special case of pair copula constructions going back to Joe (1996) and further developed in a series of papers (see, e.g., Bedford and Cooke (2001, 2002), Aas et al. (2009), Hobæk Haff et al. (2010), or Czado (2010)). The main idea of pair copula constructions is to decompose a multivariate copula C into a cascade of $d(d-1)/2$ bivariate copulas. Of these bivariate copulas, $d-1$ bivariate copulas directly model bivariate margins of the copula C, whereas the other bivariate copulas indirectly specify the remaining parts of C in terms of conditional distributions. There are many possibilities how to choose which bivariate margins and which conditional distributions should be specified. The possible arrangements may be classified using graph theory. One of the resulting classes is called D-vine structure, where the directly connected bivariate margins of the vector $(u_1, \ldots, u_d)$ are of the type (u_i, u_{i+1}), i.e., specified in a series, and the conditional distributions modeled are all remaining possibilities of the type $(u_i, u_{i+k}|u_{i+1}, u_{i+2}, \dots u_{i+k-1})$ for $k = 2, \dots, d-1$.

For our example we chose d-dimensional D-vine structures built from bivariate Clayton copulas as building blocks. The bivariate Clayton copula is given by

$$C(u_1, u_2) = \left(u_1^{-\theta} + u_2^{-\theta} - 1\right)^{-\frac{1}{\theta}}$$

for $\theta > 0$, where the strength of dependence increases with θ. For $\theta \to 0$, the Clayton copula converges to the independence copula and for $\theta \to \infty$ we get the comonotonicity copula, i.e., the copula of complete dependence. Again, to reduce the numbers of parameters, we use the same θ for all bivariate Clayton copulas. Note however that, in contrast to the Gaussian example, the resulting D-vine structure is not exchangeable, i.e, $C(u, v, w) \neq C(w, v, w)$. For example, the (empirical) Spearman correlation matrix of an 8-dimensional Clayton D-vine construction with parameter 0.5 is

$$\begin{pmatrix} 1 & 0.32 & 0.39 & 0.44 & 0.47 & 0.48 & 0.48 & 0.47 \\ 0.32 & 1 & 0.32 & 0.39 & 0.44 & 0.47 & 0.48 & 0.48 \\ 0.39 & 0.32 & 1 & 0.31 & 0.39 & 0.44 & 0.47 & 0.48 \\ 0.44 & 0.39 & 0.31 & 1 & 0.32 & 0.39 & 0.44 & 0.47 \\ 0.47 & 0.44 & 0.39 & 0.32 & 1 & 0.32 & 0.39 & 0.44 \\ 0.48 & 0.47 & 0.44 & 0.39 & 0.32 & 1 & 0.32 & 0.39 \\ 0.48 & 0.48 & 0.47 & 0.44 & 0.39 & 0.32 & 1 & 0.32 \\ 0.47 & 0.48 & 0.48 & 0.47 & 0.44 & 0.39 & 0.32 & 1 \end{pmatrix},$$

where we report the mean of 100 sample Spearman correlation matrices estimated from samples of size 10 000. Note that only the copulas between neighbored dimensions are Clayton copulas, whereas non-neighbored dimensions are of different type and have no explicit expressions for general values of θ.

Tables 1 to 2 show simulation results based on 10 000 Monte Carlo repetitions. We report results regarding the measures τ and $\rho^* = (\rho + \overline{\rho})/2$. The first four columns in the tables contain the values of the dependence parameters of the respective copulas, the resulting true values of the measures and the sample sizes. The true values of the measures are approximated by the average of 30 results of samples of size 50 000. Comparing the true values to the mean of the estimated associations $m(\hat{\tau})$ or $m(\hat{\rho}^*)$ in column 5 and 6, we observe a finite sample bias, which decreases with increasing sample size.

In particular the estimates of ρ^* are biased for small sample sizes and weak dependence.

In the remaining columns, we report numbers concerning asymptotic standard deviations of the measures, three columns for each measure. First the sample standard deviations of the 10 000 estimates are shown. We compare them to scaled versions of the estimates of the asymptotic standard deviations $\sigma_{\widehat{\rho^*}}$ and $\sigma_{\widehat{\tau}}$. Therefore, means $m(\cdot)$ and sample standard deviations $s(\cdot)$ of the 10 000 estimates of the respective asymptotic standard deviations are shown. We find that the finite sample standard deviations may be well estimated by our estimators of the asymptotic standard deviation for sample sizes larger than or equal to 100.

The values of ρ and $\overline{\rho}$ are equal for copulas C where $C = \overline{C}$. An example is the Gaussian copula. In Figure 1 we show this effect for our examples of a Gaussian copula and the D-vine structure. In the D-vine example, the values of ρ and $\overline{\rho}$ diverge while they are equal in the Gaussian example. The figure further illustrates how the choice of p for fixed $d = p + q = 30$, i.e., the separation of the joint vector in subvectors, effects the measures. We see that the measures hold the overall level of association but vary in p as to be expected.

[Figure 1 about here.]

[Tables 1 and 2 about here.]

5.2. Application

We apply our measures to measure strength and direction of association in financial markets for two examples and compare them to canonical correlation (Hotelling, 1936), the RV coefficient (Escoufier, 1973; Robert and Escoufier, 1976), and distance correlation (Székely et al., 2007; Székely and Rizzo, 2009). In the first example, we consider association between bond and stock markets, which is widely discussed in the financial literature. For example, Gulko (2002) argues that government bonds and stocks are usually positively correlated, but decouple in times of crises when investors search for safe havens like bonds of strong countries. In these times their association is expected to be negative. Another analysis of this effect may be found in Ilmanen (2003). In the second example, the association between south European and north European bonds is analyzed. It is expected that these bond markets were highly positively associated before the recent crisis and that flight-to-quality effects within the crisis might have lead to lesser positive or even negative association since then.

For the bond-stock example, we consider daily returns of the stock market indices of five major countries[2] as well as government bonds indices from The Bank of America Merrill Lynch[3] for the respective countries during the period from January 3, 1996 to November 15, 2012. The returns and squared returns show significant serial correlation as indicated by the Ljung Box test. We therefore apply an ARMA(1,1) GARCH(1,1) filter for further analysis. Figure 2 shows the evolution of the measures, based on a forward-looking moving window with a window size of 150 days. In particular, the first value of each measure is based on the 150 daily returns following January 2, 1996. The last value is estimated from the 150 daily returns from January 31, 2012. The top panel shows that canonical correlation, distance correlation and the RV coefficient exhibit similar patterns of association. Their scaling is quite different, however. Canonical correlation is always highest, RV coefficient always lowest. Distance correlation is somehow between the two, but closer to the canonical correlation in general. The evolution of $\widehat{\tau}_{150}$ and $\widehat{\rho^*}_{150}$ over time is shown in the bottom panel together with pointwise 95% confidence bands calculated from our estimates of the asymptotic variance in the iid case. There are differences between their values which are in general smaller than these of the aforementioned measures. The Kendall measure behaves more smoothed than the Spearman measure. It can be seen that there is a tendency of decreasing association between the markets. The association has been mainly positive from 1996 to 2000 which was a Bullish subperiod of the stock markets. Afterwards, with the beginning of Bearish subperiod from 2000 to 2002, it turned negative and stayed there even during the Bullish period from 2002 to 2006. From 2006 on and with the beginning of the crisis it stayed negative. The measures in the upper panel do not capture this behaviour. There is even a tendency to misinterpretations, since these measures show higher values in the crisis times than in the other times.

[2]All Ordinaries (Australia), CAC 40 (France), DAX (Germany), Nikkei 225 (Japan) and S&P 500 (USA)

[3]The BofA Merrill Lynch Australia Government Index (G0T0), The BofA Merrill Lynch French Government Index (G0F0), The BofA Merrill Lynch Japan Government Index (G0Y0), The BofA Merrill Lynch German Government Index (G0D0) and The BofA Merrill Lynch U.S. Government Index (G0Q0)

[Figure 2 about here.]

In the next empirical example we focus on the association between south and north European bonds markets during the ongoing financial crisis. We consider the daily log-changes of Merril Lynch government bond indices from south European countries (Italy, Portugal, Spain)[4] and north European countries (France, Germany, Netherlands)[5] during the period from January 1, 2007 to November 15, 2012. Again, the data is filtered. We use an AR(1) GARCH(1,1) model to remove existent serial dependence. Figure 3 shows the evolution of the association, based on a forward-looking moving window with a window size of 150 days. Again, the top panel shows canonical correlation, distance correlation and the RV coefficient with similar patterns of association, but quite different scaling. The evolution of $\widehat{\tau}_{150}$ and $\widehat{\rho^*}_{150}$ is shown in the bottom panel. Both panels show a high association of north and south European government bonds at the beginning of the period and decreasing strength of association with the beginning of the year 2008. Economically this is to be expected, since investors regarded the credit-worthiness of the countries in the European Union as homogeneous before 2008, which lead to similar yields on their government bonds. However, starting with the Euro-Crisis, investors started to question the rating of southern European countries, namely Greece, Portugal and Spain, and transferred their money to northern European countries, such as France, Germany and the Netherlands. This lead to significant differences in the yields of southern and northern European government bonds. The lower panel of the figure shows that during periods in 2011 and 2012 the association actually turned negative which again cannot be captured by the measures of the first panel.

[Figure 3 about here.]

[4] The BofA Merrill Lynch Italian Government Index (G0I0), The BofA Merrill Lynch Portugese Government Index (G0U0), and The BofA Merrill Lynch Spanish Government Index (G0E0)

[5] The BofA Merrill Lynch German Government Index (G0D0), The BofA Merrill Lynch French Government Index (G0F0), and The BofA Merrill Lynch Netherlands Government Index (G0N0)

6. Discussion

Many of the results in this note can be extended to a k-tuple of random vectors $X_1, \ldots, X_k$ of dimensions $d_1, \ldots, d_k$ respectively with joint d-variate cdf H, where $d = d_1 + \cdots + d_k$, and marginal cdfs $H_1, \ldots, H_k$.

Another choice for μ and ν in (2.1) and (2.2) would be the product of all univariate margins, that is, $H_1 \otimes \cdots \otimes H_d = (F_1 \otimes \cdots \otimes F_p) \otimes (G_1 \otimes \cdots \otimes G_q)$, yielding another generalization of Spearman's rho, see for instance Grothe et al. (2011). However, as it may be inappropriate to artificially destroy the dependence within each of the vectors X and Y, we have not considered this choice further in this paper.

Further extensions of our measures could involve discrete random variables, allowing for ordinal data. Another point of interest is the assignment of weights to variables, increasing or diminishing their importance within the random vector as a whole.

Testing for independence between the subvectors X and Y based on our measures of association is not a good idea: the tests will not be consistent, i.e., X and Y may be dependent even if (some of) the measures of association are zero. Rather use for instance the test described in Quessy (2010).

Regarding statistical inference, an alternative to U-statistics would be to express the association measures as copula functionals and to replace the unknown copulas by their empirical versions. The resulting estimators would then be the same, up to $O(1/n)$, as the U-statistic estimators in the paper (see Appendix B for the relation of our estimators to rank statistics). Weak convergence of the empirical process together with the functional delta method would then suffice to find the asymptotic distribution of the estimators also for dependent samples; see for instance Fermanian et al. (2004), Tsukahara (2005), Segers (2012), and Bücher and Volgushev (2013). However, the resulting expressions of the asymptotic variances would be analytically untractable and would motivate computationally intensive resampling procedures, rather than our direct plugin variance estimates; see Quessy (2010) and Quessy et al. (2013). A theoretical objection is that the copula approach would be restricted to copulas whose first-order partial derivatives are continuous on the interior of the unit cube, excluding singular models such as Marshall–Olkin copulas and more generally, extreme-value copulas with discrete spectral measures (Gudendorf and Segers, 2012).

References

Aas, K., Czado, C., Frigessi, A., Bakken, H., 2009. Pair-copula constructions of multiple dependence. Insurance: Mathematics and Economics 44 (2), 182–198.

Bedford, T., Cooke, R. M., 2001. Probability density decomposition for conditionally dependent random variables modeled by vines. Annals of Mathematics and Artificial Intelligence 32 (1-4), 245–268.

Bedford, T., Cooke, R. M., 2002. Vines: A new graphical model for dependent random variables. The Annals of Statistics 30 (4), 1031–1068.

Bücher, A., Volgushev, S., 2013. Empirical and sequential empirical copula processes under serial dependence. Journal of Multivariate Analysis 119 (0), 61–70.

Clayton, D., 1978. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika 65, 141–151.

Czado, C., 2010. Pair-copula constructions of multivariate copulas. In: Jaworski, P., Durante, F., Härdle, W. K., Rychlik, T. (Eds.), Copula Theory and Its Applications. Lecture Notes in Statistics. Springer, Berlin, pp. 93–109.

Efron, B., Stein, C., 1981. The jackknife estimate of variance. The Annals of Statistics 9 (3), 586–596.

El Maache, H., Lepage, Y., 2003. Spearman's rho and Kendall's tau for multivariate data sets. In: Moore, M., Froda, S., Léger, C. (Eds.), Mathematical Statistics and Applications: Festschrift for Constance van Eeden. Institute of Mathematical Statistics, Beachwood, OH, pp. 113–130.

Escoufier, Y., 1973. Le traitement des variables vectorielles. Biometrics 29 (4), 751–760.

Fermanian, J.-D., Radulović, D., Wegkamp, M. H., 2004. Weak convergence of empirical copula processes. Bernoulli 10, 847–860.

Forbes, K. J., Rigobon, R., 2002. No Contagion, Only Interdependence: Measuring Stock Market Comovements. The Journal of Finance 57, 2223–2261.

Grothe, O., Schmid, F., 2011. Kendall's W reconsidered. Communications in Statistics - Simulation and Computation 40, 285–305.

Grothe, O., Schmid, F., Schnieders, J., Segers, J., 2011. Measuring association between random vectors. Tech. rep., University of Cologne and Université catholique de Louvain, arXiv:1107.4381.

Gudendorf, G., Segers, J., 2012. Nonparametric estimation of multivariate extreme-value copulas. Journal of Statistical Planning and Inference 142, 3073–3085.

Gulko, L., 2002. Decoupling. The Journal of Portfolio Management 28 (3), 59–66.

Das Gupta, S., Olkin, I., Savage, L.J., Eaton, M.L., Perlman, M., Sobel, M., 1972. Inequalities on the probability content of convex regions for elliptically contoured distributions. Proc. 6th Berkeley Sympos. Math. Statist. Probab., Univ. Calif. 2, 241–265.

Hays, W. L., 1960. A note on average tau as a measure of concordance. Journal of the American Statistical Association 55, 331–341.

Hobæk Haff, I., Aas, K., Frigessi, A. F., 2010. On the simplified pair-copula construction – Simply useful or too simplistic? Journal of Multivariate Analysis 101 (5), 1296–1310.

Hoeffding, W., 1948. A class of statistics with asymptotically normal distribution. The Annals of Mathematical Statistics 19, 293–325.

Hotelling, H., 1936. Relations between two sets of variates. Biometrika 28, 321–377.

Ilmanen, A., 2003. Stock-bond correlations. The Journal of Fixed Income 13 (2), 55–66.

Joe, H., 1990. Multivariate concordance. Journal of Multivariate Analysis 35, 12–30.

Joe, H., 1996. Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. Distributions with Fixed Marginals and Related Topics, 120–141.

Joe, H., 1997. Multivariate Models and Dependence Concepts. Chapman & Hall, London.

Kendall, M. G. and Babington Smith, B., 1939. The problem of m rankings. The Annals of Mathematical Statistics 10, 275–287.

Marozzi, M., 2013. Testing for concordance between several criteria. Journal of Statistical Computation and Simulation, forthcoming. DOI: 10.1080/00949655.2013.766189

Nelsen, R. B., 2006. An Introduction to Copulas. Vol. 139 of Lecture Notes in Statistics. Springer Verlag.

Puri, M. L., Sen, P. K., Gokhale, D. V., 1970. On a class of rank order tests for independence in multivariate distributions. Sankhyā, Series A 32, 271–298.

Quessy, J.-F., 2010. Applications and asymptotic power of marginal-free tests of stochastic vectorial independence. Journal of Statistical Planning and Inference 140, 3058–3075.

Quessy, J.-F., Saïd, M., Favre, A.-C., 2013. Multivariate Kendall's tau for change-point detection in copulas. The Canadian Journal of Statistics 41, 65–82.

Robert, P., Escoufier, Y., 1976. A unifying tool for linear multivariate statistical methods: the RV-coefficient. Applied Statistics 25 (3), 257–265.

Schmid, F., Schmidt, R., 2007. Multivariate extensions of Spearman's rho and related statistics. Statistics & Probability Letters 77, 407–416.

Schmid, F., Schmidt, R., Blumentritt, T., Gaisser, S., Ruppert, M., 2010. Copula-based measures of multivariate association. In: Jaworski, P., Durante, F., Härdle, W. K., Rychlik, T. (Eds.), Copula Theory and Its Applications. Lecture Notes in Statistics. Springer Berlin Heidelberg, pp. 209–236.

Segers, J., 2012. Asymptotics of empirical copula processes under nonrestrictive smoothness assumptions. Bernoulli 18, 764–782.

Székely, G., Rizzo, M., 2009. Brownian distance covariance. The Annals of Applied Statistics 3 (4), 1236–1265.

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

Tsukahara, H., 2005. Semiparametric estimation in copula models. The Canadian Journal of Statistics 33, 357–375.

Appendix A. Implementation of the generalization of Spearman's rho

Appendix A.1. Implementation of $\hat{q}_{V,W,n}$

A naive implementation of $\hat{q}_{V,W,n}$ in (4.6) involves a triple loop over the sample. This can be simplified to a double loop.

For arrays $(a_{ij})_{i,j=1}^n$ and $(b_{ij})_{i,j=1}^n$, we have

$$\sum_{\substack{\{i,j,k\}\subset\{1,\ldots,n\}\\ \sharp\{i,j,k\}=3}} a_{ij}\, b_{ik} = \sum_{i=1}^n \left\{ \left(\sum_{j\neq i} a_{ij} \right) \left(\sum_{j\neq i} b_{ij} \right) - \sum_{j\neq i} (a_{ij}\, b_{ij}) \right\}.$$

Apply the displayed equality to $\hat{q}_{V,W,n}$ with $a_{ij} = \mathbb{1}_{\{V_i \leq V_j\}}$ and $b_{ik} = \mathbb{1}_{\{W_i \leq W_k\}}$. Only a double loop is needed: outer loop over i, inner loop over j.

Appendix A.2. Estimation and implemetation of the asymptotic variance

The association measure inspired on Spearman's rho can be written as

$$\rho[X,Y] = f(\theta) = \frac{\theta_3 - \theta_1\theta_2}{\sqrt{(\theta_4 - \theta_1^2)(\theta_5 - \theta_2^2)}},$$

for $\theta = (p_X, p_Y, q_{X,Y}, q_{X,X}, q_{Y,Y})'$. The estimator of $\rho[X,Y]$ is then

$$\hat{\rho}_{S,n}[X,Y] = f(\hat{\theta}_n), \qquad \hat{\theta}_n = (\hat{p}_{X,n}, \hat{p}_{Y,n}, \hat{q}_{X,Y,n}, \hat{q}_{X,X,n}, \hat{q}_{Y,Y,n})'$$

where $\hat{p}_{W,n}$ and $\hat{q}_{V,W,n}$ have been defined in (4.1) and (4.6); as before, the symbols 'V' and 'W' denote 'X' or 'Y'. Asymptotic normality of $\hat{\theta}_n$,

$$\sqrt{n}(\hat{\theta}_n - \theta) \rightsquigarrow N_5\big(0, \Sigma\big) \qquad (n \to \infty), \tag{A.1}$$

together with the delta method, yields asymptotic normality of $\hat{\rho}_S[X,Y]$,

$$\begin{aligned}\sqrt{n}(\hat{\rho}_{S,n}[X,Y]-\rho[X,Y]) \quad &\rightsquigarrow \quad N(0,\sigma^2) \qquad (n\to\infty),\\ &\qquad \sigma^2=(\nabla f(\theta))'\,\Sigma\,(\nabla f(\theta))\big),\end{aligned}$$

where $\nabla f(\theta)\in\mathbb{R}^{5\times1}$ is the gradient of f evaluated in θ. The asymptotic variance, σ^2, can be estimated consistently by $\hat{\sigma}_n^2=(\nabla f(\hat{\theta}_n))'\,\hat{\Sigma}_n\,(\nabla f(\hat{\theta}_n))\big)$, provided $\hat{\Sigma}_n$ is a consistent estimator of Σ.

To construct $\hat{\Sigma}_n$, we rely on the main term in the Hoeffding decomposition of the components of $\hat{\theta}_n$. The one for $\hat{p}_{W,n}$ is given in (4.4). For $\hat{q}_{V,W,n}$, we have

$$\sqrt{n}(\hat{q}_{V,W,n}-q_{V,W})=\frac{3}{\sqrt{n}}\sum_{i=1}^{n}\big(g_{1,V,W}(V_i,W_i)-q_{V,W}\big)+o_p(1) \qquad (n\to\infty),$$

for functions $g_{1,V,W}$ to be determined next. First, we need to symmetrize the kernel in the definition of $q_{V,W}$:

$$\begin{aligned}q_{V,W}&=\mathrm{E}\big[g_{V,W}\big((V_1,W_1),(V_2,W_2),(V_3,W_3)\big)\big],\\ \hat{q}_{V,W,n}&=\frac{1}{\binom{n}{3}}\sum_{1\le i<j<k\le n}g_{V,W}\big((V_i,W_i),(V_j,W_j),(V_k,W_k)\big),\end{aligned}$$

where

$$\begin{aligned}&g_{V,W}\big((v_1,w_1),(v_2,w_2),(v_3,w_3)\big)\\ &=\frac{1}{6}\{\mathbb{1}(v_1\ge v_2,w_1\ge w_3)+\mathbb{1}(v_1\ge v_3,w_1\ge w_2)+\\ &\quad\mathbb{1}(v_2\ge v_3,w_2\ge w_1)+\mathbb{1}(v_2\ge v_1,w_2\ge w_3)+\\ &\quad\mathbb{1}(v_3\ge v_1,w_3\ge w_2)+\mathbb{1}(v_3\ge v_2,w_3\ge w_1)\}.\end{aligned}$$

Then $g_{1,V,W}$ is given by

$$\begin{aligned}g_{1,V,W}(v_1,w_1)&=\mathrm{E}\big[g_{1,V,W}\big((v_1,w_1),(V_2,W_2),(V_3,W_3)\big)\big]\\ &=\frac{1}{6}\{\Pr[v_1\ge V_2,w_1\ge W_3]+\Pr[v_1\ge V_3,w_1\ge W_2]+\\ &\quad\Pr[V_2\ge V_3,W_2\ge w_1]+\Pr[V_2\ge v_1,W_2\ge W_3]+\\ &\quad\Pr[V_3\ge v_1,W_3\ge W_2]+\Pr[V_3\ge V_2,W_3\ge w_1]\}.\end{aligned}$$

It follows that Σ in (A.1) is equal to the covariance matrix of the five-dimensional random vector

$$\big(2h_{1,X}(X_1),\, 2h_{1,Y}(Y_1),\, 3g_{1,X,Y}(X_1,Y_1),\, 3g_{1,X,X}(X_1,X_1),\, 3g_{1,Y,Y}(Y_1,Y_1)\big)'.$$

Replacing the functions $h_{1,W}$ and $g_{1,V,W}$ by estimates, we can estimate Σ by the covariance matrix of the quintuples

$$\big(2\hat{h}_{1,X,n}(X_i), 2\hat{h}_{1,Y,n}(Y_i), 3\hat{g}_{1,X,Y,n}(X_i,Y_i), 3\hat{g}_{1,X,X,n}(X_i,X_i), 3\hat{g}_{1,Y,Y,n}(Y_i,Y_i)\big)'$$

for $i \in \{1,\ldots,n\}$. The statistics $\hat{h}_{1,W,n}(W_i)$ have been defined in (4.5). Similarly, for $i \in \{1,\ldots,n\}$, put

$$\hat{g}_{1,V,W,n}(V_i,W_i) = \frac{1}{(n-1)(n-2)} \sum_{j,k:j\neq i,k\neq i,j\neq k} g_{V,W}\big((V_i,W_i),(V_j,W_j),(V_k,W_k)\big).$$

By symmetry, the six terms in the definition of $g_{V,W}$ can be reduced to three terms:

$$\hat{g}_{1,V,W,n}(V_i,W_i) = \frac{1}{(n-1)(n-2)} \times \sum_{j,k:j\neq i,k\neq i,j\neq k} \frac{1}{3}\{\mathbb{1}_{\{V_i\geq V_j,W_i\geq W_k\}} + \mathbb{1}_{\{V_j\geq V_k,W_j\geq W_i\}} + \mathbb{1}_{\{V_k\geq V_i,W_k\geq W_j\}}\} \quad \text{(A.2)}$$

A naive implementation of $\hat{g}_{1,V,W,n}(V_i,W_i)$ would require a double loop (over j and k), so that the calculation of $\hat{\Sigma}_n$ would require even a triple loop. However, simplification is possible, leading to just a double loop required for $\hat{\Sigma}_n$. We treat each of the three sums in (A.2) in turn.

- First sum: Apply the identity

$$\sum_{j,k:j\neq i,k\neq i,j\neq k} a_{ij}\, b_{ik} = \Big(\sum_{j\neq i} a_{ij}\Big)\Big(\sum_{j\neq i} b_{ij}\Big) - \sum_{j\neq i} a_{ij}\, b_{ij}$$

  to $a_{ij} = \mathbb{1}_{\{V_i\geq V_j\}}$ and $b_{ik} = \mathbb{1}_{\{W_i\geq W_k\}}$.

- Second sum: Write

$$\sum_{j,k:j\neq i,k\neq i,j\neq k} a_{jk}\, b_{ji} = \sum_{j:j\neq i}\left(\sum_{k:k\neq j} a_{jk} - a_{ji}\right) b_{ji},$$

  applied to $a_{jk} = \mathbb{1}_{\{V_j \geq V_k\}}$ and $b_{ji} = \mathbb{1}_{\{W_j \geq W_i\}}$. The sums $s_j = \sum_{k\neq j} a_{jk}$ do not depend on i; they can be calculated in advance, outside the loop over i.

- Third sum: Write

$$\sum_{j,k:j\neq i,k\neq i,j\neq k} a_{ki}\, b_{kj} = \sum_{k:k\neq i} a_{ki}\left(\sum_{j:j\neq k} b_{kj} - b_{ki}\right),$$

  applied to $a_{ki} = \mathbb{1}_{\{V_k \geq V_i\}}$ and $b_{kj} = \mathbb{1}_{\{W_k \geq W_j\}}$. The sums $t_k = \sum_{j\neq k} b_{jk}$ do not depend on i; they can be calculated in advance, outside the loop over i.

To estimate the asymptotic variance of $\rho^*[X,Y]$ consider $\widehat{\rho^*}[X,Y] = g(\widehat{\theta}_n)$ with $\theta = (p_X, p_Y, q_{X,Y}, q_{X,X}, q_{Y,Y}, q_{-X,-Y}, q_{-X,-X}, q_{-Y,-Y})$ and proceed as above.

Appendix B. Relationship of the proposed estimators with ranks

Our measures rely on indicators of the form $\mathbb{1}(X_i \leq X_j)$, which only depend on ranks. In this section we express our estimators in terms of multivariate ranks. Let $Z_1 = (X_1, Y_1), \ldots, Z_n = (X_n, Y_n)$ be a random sample in $\mathbb{R}^p \times \mathbb{R}^q$. Consider the following multivariate generalizations of the ranks:

$$R_j^X = \sum_{i=1}^n \mathbb{1}(X_i \leq X_j) - 1 \text{ and } S_i^X = \sum_{j=1}^n \mathbb{1}(X_i \leq X_j) - 1.$$

Subtracting 1 will be convenient. Similarly define R_j^Y, S_j^Y, R_j^Z and S_j^Z. Put $\delta_{rs} = 1$ if $r = s$ and $\delta_{rs} = 0$ otherwise. Summing the ranks yields

$$R_\bullet^X = \sum_{j=1}^n R_j^X \text{ and } S_\bullet^X = \sum_{i=1}^n S_i^X.$$

Lemma Appendix B.1 (Kendall's tau). *We have*

$$\hat{\tau}_n[X,Y] = \frac{n(n-1)R_\bullet^{(X,Y)} - R_\bullet^X\, R_\bullet^Y}{\sqrt{R_\bullet^X\,\{n(n-1) - R_\bullet^X\}\, R_\bullet^Y\,\{n(n-1) - R_\bullet^Y\}}},$$

and similarly with 'R' replaced by 'S'.

Proof. Since $\sum_{i=1}^n \sum_{j=1}^n \delta_{ij} = n$, we have

$$n(n-1)\,\hat{p}_X = \sum_{i=1}^n \sum_{j=1}^n \mathbb{1}(X_i \le X_j)(1-\delta_{ij}) = R_\bullet^X = S_\bullet^X.$$

Similarly for $n(n-1)\,\hat{p}_Y$ and $n(n-1)\,\hat{p}_{(X,Y)}$. It follows that

$$\begin{aligned}\hat{\tau}_n[X,Y] &= \frac{\hat{p}_Z - \hat{p}_X\,\hat{p}_Y}{\sqrt{\hat{p}_X\,(1-\hat{p}_X)\,\hat{p}_Y\,(1-\hat{p}_Y)}} \\ &= \frac{n(n-1)S_\bullet^{(X,Y)} - S_\bullet^X\, S_\bullet^Y}{\sqrt{S_\bullet^X\,\{n(n-1) - S_\bullet^X\}\, S_\bullet^Y\,\{n(n-1) - S_\bullet^Y\}}}\end{aligned}$$

or the same expression with 'S' replaced by 'R'. This completes the proof.

Observe that $\hat{p}_X$ is, up to $O(1/n)$, equal to $\int \hat{A}_n \,\mathrm{d}\hat{A}_n$, with $\hat{A}_n$ the empirical copula of the sample $X_1, \ldots, X_n$, assuming there are no ties. Similarly for $\hat{p}_Y$ and $\hat{p}_{X,Y}$. As a consequence, our estimator for tau is, up to $O(1/n)$, equal to the expression for tau in terms of copula integrals upon replacing the unknown copulas by their empirical counterparts.

For integer $0 \le k \le n$, put $(n)_k = n(n-1)\cdots(n-k+1)$.

Lemma Appendix B.2 (Spearman's rho). *We have*

$$\hat{\rho}_n[X,Y] = \frac{\frac{1}{(n)_3}\sum_{i=1}^n R_i^X\, R_i^Y - \frac{1}{(n)_2}R_\bullet^X \frac{1}{(n)_2}R_\bullet^Y - \frac{1}{(n)_3}R_\bullet^{(X,Y)}}{\sqrt{\{\frac{1}{(n)_3}\sum_{i=1}^n R_i^X(R_i^X - 1) - (\frac{1}{(n)_2}R_\bullet^X)^2\}\,\{\mathit{idem\ for}\ Y\}}}.$$

We find that $\hat{\rho}_n[X,Y]$ is up to $O(1/n)$ equal to the sample correlation between the pairs (R_i^X, R_i^Y) for $i = 1, \ldots, n$. Similarly, $\hat{\hat{\rho}}_n[X,Y]$ is up to $O(1/n)$ equal to the sample correlation between the pairs (S_i^X, S_i^Y).

Proof. We have

$$
\begin{aligned}
&n(n-1)(n-2)\hat{\hat{q}}_{X,Y} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} \underbrace{\mathbb{1}(X_i \le X_j)}_{=A_{ij}} \underbrace{\mathbb{1}(Y_i \le Y_k)}_{=B_{ik}} (1-\delta_{ij})(1-\delta_{ik})(1-\delta_{jk}) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}(1-\delta_{ij}) \sum_{k=1}^{n} B_{ik}(1-\delta_{ik}) - \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}\, B_{ij}\,(1-\delta_{ij}) \\
&= \sum_{i=1}^{n}\Bigg(\sum_{j=1}^{n} A_{ij} - \underbrace{A_{ii}}_{=1}\Bigg)\Bigg(\sum_{k=1}^{n} B_{ik} - \underbrace{B_{ii}}_{=1}\Bigg) - \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}\, B_{ij} + \sum_{i=1}^{n} \underbrace{A_{ii}\, B_{ii}}_{=1} \\
&= \sum_{i=1}^{n} S_i^X\, S_i^Y - \sum_{i=1}^{n} S_i^{(X,Y)}.
\end{aligned}
$$

Note that the second term is of smaller order than the first one, $O(n^2)$ compared to $O(n^3)$. A particular case of the preceding formula is

$$
n(n-1)(n-2)\hat{\hat{q}}_{X,X,n} = \sum_{i=1}^{n}(S_i^X)^2 - \sum_{i=1}^{n} S_i^X = \sum_{i=1}^{n} S_i^X(S_i^X - 1),
$$

and similarly for $\hat{\hat{q}}_{Y,Y,n}$. To find the expressions for $\hat{q}$, it suffices to replace '$\le$' by '$\ge$', that is, 'S' by 'R', e.g.

$$
n(n-1)(n-2)\hat{q}_{X,Y,n} = \sum_{i=1}^{n} R_i^X\, R_i^Y - \sum_{i=1}^{n} R_i^{(X,Y)}.
$$

It follows that

$$
\begin{aligned}
\hat{\rho}_n[X,Y] &= \frac{\hat{q}_{X,Y} - \hat{p}_Y\,\hat{p}_Y}{\sqrt{(\hat{q}_{X,X} - \hat{p}_X^2)\,(\hat{q}_{Y,Y} - \hat{p}_Y^2)}} \\
&= \frac{\frac{1}{(n)_3}\sum_{i=1}^{n} R_i^X\, R_i^Y - \frac{1}{(n)_2} R_\bullet^X \frac{1}{(n)_2} R_\bullet^Y - \frac{1}{(n)_3} R_\bullet^{(X,Y)}}{\sqrt{\{\frac{1}{(n)_3}\sum_{i=1}^{n} R_i^X(R_i^X - 1) - (\frac{1}{(n)_2} R_\bullet^X)^2\}(\text{idem for } Y)}}.
\end{aligned}
$$

To see the statement about the correlation, multiply the numerator and the denominator with $(n-1)^2$. This completes the proof.

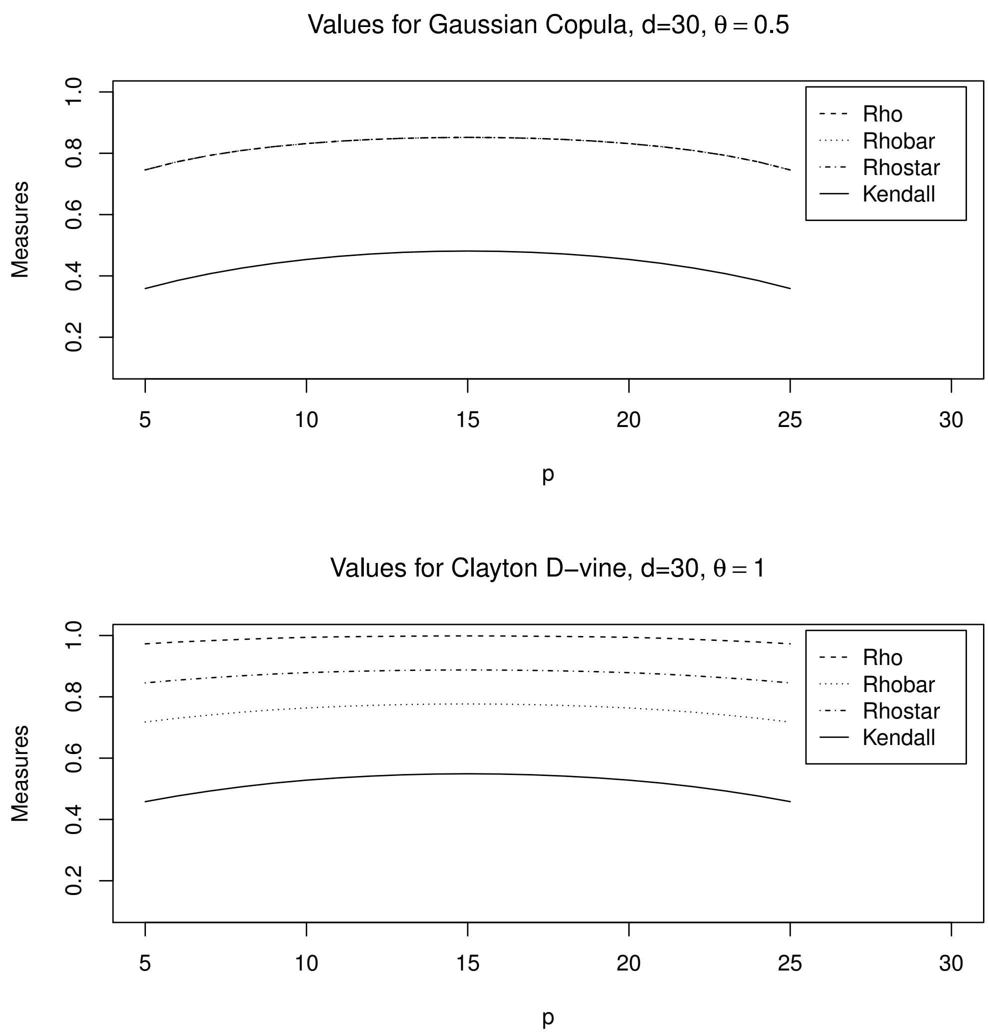

Figure 1: The association measures for $d = p + q = 30$ for varying values of p for the Gaussian copula and the D-vine structure. Values are means of 10 000 repetitions of sample size 1 000. In the Gaussian example, the two Spearman measures ρ and $\overline{\rho}$ coincide with the upper line, their equality following from the symmetry of the copula.

θ	τ	ρ^*	n	$m(\widehat{\tau}_n)$	$m(\widehat{\rho^*}_n)$	$s(\widehat{\tau}_n)$	$\frac{m(\hat{\sigma}_{\hat{\tau}})}{\sqrt{n}}$	$\frac{s(\hat{\sigma}_{\hat{\tau}})}{\sqrt{n}}$	$s(\widehat{\rho^*}_n)$	$\frac{m(\hat{\sigma}_{\widehat{\rho^*}})}{\sqrt{n}}$	$\frac{s(\hat{\sigma}_{\widehat{\rho^*}})}{\sqrt{n}}$
Two 4-dimensional vectors											
-0.1	-0.038	-0.234	50	-0.035	-0.146	0.009	0.010	0.004	0.067	0.075	0.020
			100	-0.037	-0.191	0.007	0.007	0.002	0.041	0.041	0.008
			1000	-0.039	-0.230	0.002	0.002	0.000	0.012	0.012	0.001
0.2	0.176	0.388	50	0.143	0.426	0.068	0.071	0.013	0.115	0.106	0.016
			100	0.160	0.408	0.047	0.047	0.006	0.083	0.078	0.009
			1000	0.174	0.390	0.014	0.014	0.001	0.026	0.026	0.001
0.5	0.445	0.723	50	0.430	0.738	0.064	0.066	0.008	0.066	0.065	0.013
			100	0.438	0.730	0.043	0.044	0.004	0.047	0.046	0.007
			1000	0.444	0.724	0.013	0.013	0.000	0.015	0.015	0.001
0.8	0.695	0.913	50	0.691	0.919	0.041	0.042	0.006	0.025	0.025	0.007
			100	0.693	0.915	0.027	0.028	0.003	0.017	0.017	0.004
			1000	0.694	0.913	0.008	0.008	0.000	0.005	0.005	0.000
3-dimensional and 4-dimensional vector											
-0.1	-0.053	-0.234	50	-0.056	-0.146	0.015	0.017	0.006	0.081	0.085	0.020
			100	-0.058	-0.191	0.010	0.012	0.003	0.050	0.051	0.010
			1000	-0.055	-0.230	0.003	0.003	0.000	0.015	0.015	0.001
0.2	0.173	0.365	50	0.147	0.404	0.069	0.070	0.010	0.115	0.106	0.015
			100	0.161	0.386	0.046	0.046	0.005	0.080	0.077	0.008
			1000	0.173	0.368	0.014	0.014	0.000	0.026	0.025	0.001
0.5	0.433	0.700	50	0.420	0.716	0.065	0.066	0.008	0.071	0.068	0.014
			100	0.427	0.708	0.043	0.044	0.004	0.049	0.048	0.007
			1000	0.433	0.701	0.013	0.013	0.000	0.015	0.015	0.001
0.8	0.683	0.903	50	0.681	0.910	0.042	0.043	0.007	0.027	0.027	0.008
			100	0.682	0.906	0.028	0.029	0.003	0.019	0.019	0.004
			1000	0.683	0.903	0.009	0.009	0.000	0.006	0.006	0.000

Table 1: Simulations from a Gaussian copula with parameter θ and sample size n. The true values of the measures τ and ρ^* are approximated by the means of 30 repetitions of sample size 50 000. The remaining numbers are means or standard deviations of 10 000 repetitions. The columns 7, 8 and 9 refer to the (asymptotic) variance of the τ estimator: $s(\widehat{\tau}_n)$ is the sample standard deviation of the estimates of τ, while $\frac{m(\hat{\sigma}_{\hat{\tau}})}{\sqrt{n}}$ and $\frac{s(\hat{\sigma}_{\hat{\tau}})}{\sqrt{n}}$ are mean and standard deviation of the estimates of the asymptotic variance scaled to the final sample size. The remaining columns refer to the analogous results regarding the ρ^* estimator.

θ	τ	ρ^*	n	$m(\widehat{\tau}_n)$	$m(\widehat{\rho^*}_n)$	$s(\widehat{\tau}_n)$	$\frac{m(\hat{\sigma}_{\widehat{\tau}})}{\sqrt{n}}$	$\frac{s(\hat{\sigma}_{\widehat{\tau}})}{\sqrt{n}}$	$s(\widehat{\rho^*}_n)$	$\frac{m(\hat{\sigma}_{\widehat{\rho^*}})}{\sqrt{n}}$	$\frac{s(\hat{\sigma}_{\widehat{\rho^*}})}{\sqrt{n}}$
Two 4-dimensional vectors											
0.5	0.429	0.761	50	0.410	0.781	0.075	0.076	0.012	0.069	0.063	0.015
			100	0.420	0.770	0.051	0.050	0.006	0.049	0.046	0.008
			1000	0.429	0.762	0.015	0.015	0.001	0.015	0.015	0.001
1.0	0.600	0.892	50	0.592	0.903	0.059	0.061	0.009	0.038	0.036	0.012
			100	0.596	0.897	0.041	0.040	0.004	0.027	0.026	0.006
			1000	0.600	0.892	0.012	0.012	0.000	0.008	0.008	0.001
2.0	0.716	0.935	50	0.712	0.942	0.047	0.047	0.008	0.024	0.023	0.008
			100	0.714	0.939	0.031	0.032	0.004	0.016	0.016	0.004
			1000	0.716	0.936	0.009	0.010	0.000	0.005	0.005	0.000
5.0	0.843	0.973	50	0.842	0.976	0.030	0.031	0.006	0.011	0.010	0.004
			100	0.843	0.975	0.020	0.021	0.003	0.007	0.007	0.002
			1000	0.843	0.973	0.006	0.006	0.000	0.002	0.002	0.000
3-dimensional and 4-dimensional vector											
0.5	0.407	0.714	50	0.390	0.734	0.074	0.073	0.010	0.074	0.068	0.014
			100	0.399	0.724	0.049	0.049	0.005	0.052	0.050	0.008
			1000	0.407	0.715	0.015	0.015	0.000	0.016	0.016	0.001
1.0	0.595	0.880	50	0.587	0.892	0.058	0.059	0.009	0.039	0.037	0.011
			100	0.591	0.886	0.039	0.039	0.004	0.027	0.026	0.006
			1000	0.594	0.881	0.012	0.012	0.000	0.008	0.008	0.001
2.0	0.723	0.938	50	0.720	0.944	0.044	0.046	0.007	0.023	0.022	0.008
			100	0.721	0.941	0.030	0.031	0.004	0.016	0.016	0.004
			1000	0.723	0.938	0.009	0.009	0.000	0.005	0.005	0.000
5.0	0.853	0.976	50	0.852	0.979	0.029	0.030	0.006	0.010	0.010	0.004
			100	0.852	0.977	0.019	0.020	0.003	0.007	0.007	0.002
			1000	0.853	0.976	0.006	0.006	0.000	0.002	0.002	0.000

Table 2: Simulations from a D-vine copula built from bivariate Clayton copulas with parameter θ and sample size n. The true values of the measures τ and ρ^* are approximated by the means of 30 repetitions of sample size 50 000. The remaining numbers are means or standard deviations of 10 000 repetitions. The columns 7, 8 and 9 refer to the (asymptotic) variance of the τ estimator: $s(\widehat{\tau}_n)$ is the sample standard deviation of the estimates of τ, while $\frac{m(\hat{\sigma}_{\widehat{\tau}})}{\sqrt{n}}$ and $\frac{s(\hat{\sigma}_{\widehat{\tau}})}{\sqrt{n}}$ are mean and standard deviation of the estimates of the asymptotic variance scaled to the final sample size. The remaining columns refer to the analogous results regarding the ρ^* estimator.

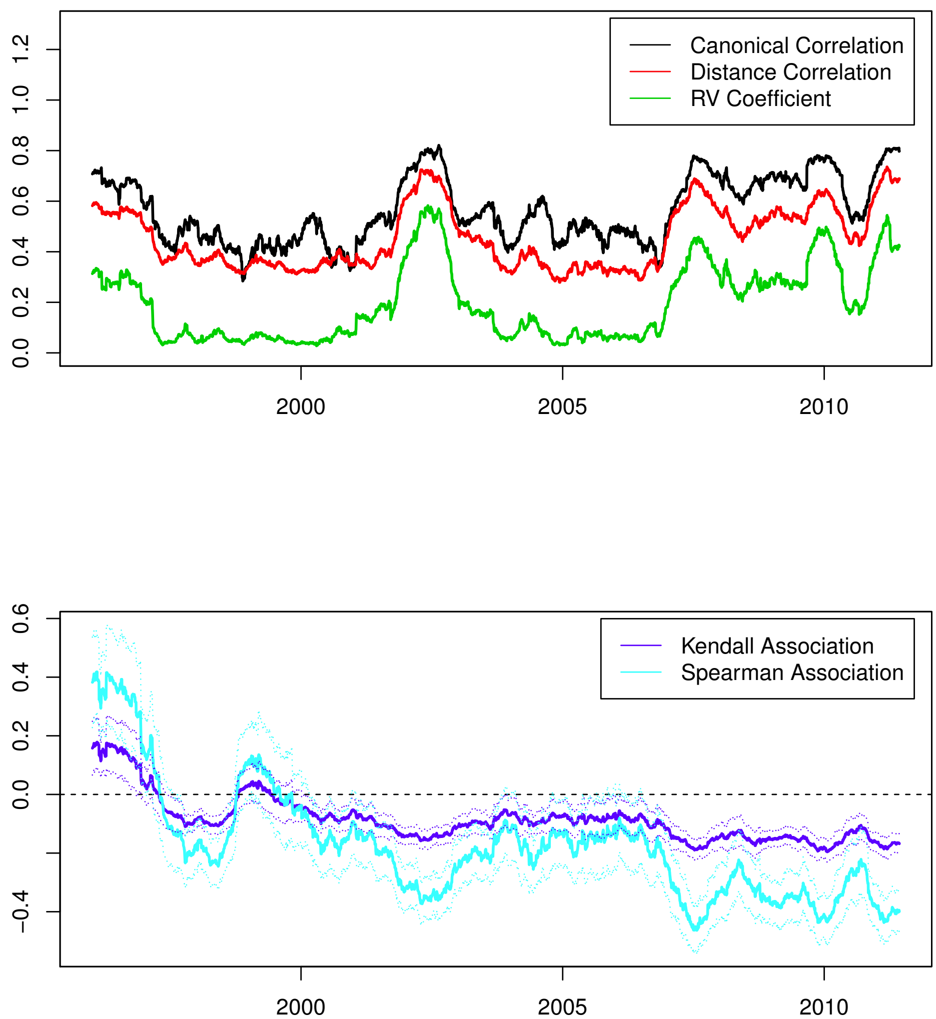

Figure 2: Evolution of the association between the bond and stock markets, measured by their canonical correlation, their distance correlation and their RV coefficient (top panel) and by the introduced measures of association τ and ρ (bottom panel). The analysis is based on a moving window approach with a window size of 150 days. The 95% pointwise confidence bands shown in the bottom panel are calculated from the iid estimates of the asymptotic variance.

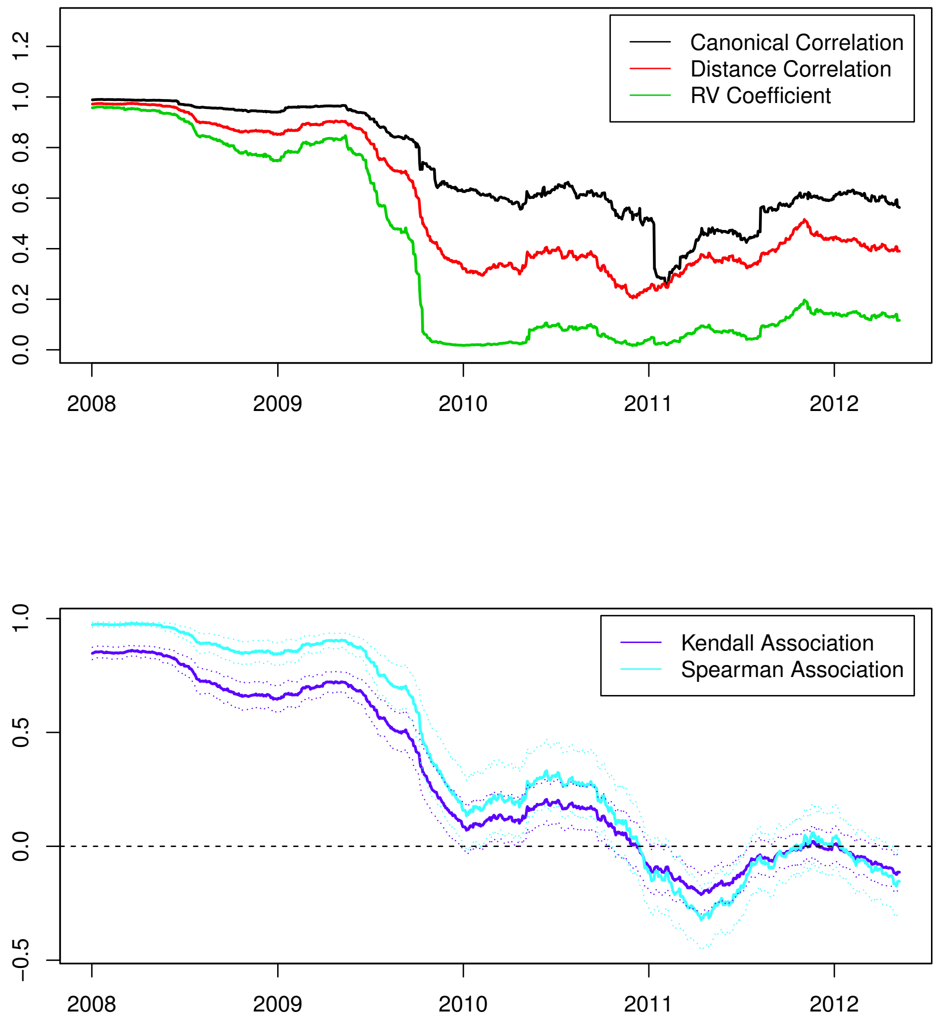

Figure 3: Evolution of the association between the southern and northern European bond markets. Association is measured by canonical correlation, distance correlation and the RV coefficient (top panel) as well as by our introduced measures of association (bottom panel). The analysis is based on a moving window approach with a forward looking window of size 250 days.