# Measuring linear correlation between random vectors

Giovanni Puccetti[1,*]

[1] *Department of Economics, Management and Quantitative Methods, University of Milano, Italy.*
[*] *Corresponding author, email: giovanni.puccetti@unimi.it.*

March 30, 2022

**Abstract**

We introduce a new scalar coefficient to measure linear correlation between random vectors which preserves all the relevant properties of Pearson's correlation in arbitrarily large dimensions. The new measure and its bounds are derived from a mass transportation approach in which the expected inner product of two random vectors is taken as a measure of their covariance and then standardized by the maximal attainable value given their marginal covariance matrices. The new correlation is maximized when the average squared Euclidean distance between the random vectors is minimal and attains value one when, additionally, it is possible to establish an affine relationship between the vectors. In several simulative studies we show the limiting distribution of the empirical estimator of the newly defined index and of the corresponding rank correlation.

A comparative study based on financial data shows that our proposed correlation, though derived from a novel approach, behaves similarly to some of the multivariate dependence notions recently introduced in the literature. Throughout the paper, we also give some auxiliary results of independent interest in matrix analysis and mass transportation theory, including an improvement to the Cauchy-Schwarz inequality for positive definite covariance matrices.

**MSC:** 62H20 (primary), 90B06, 62H30.
**Keywords:** multivariate correlation, cyclical monotonicity, financial time series, dissimilarity measure, multivariate Cauchy-Schwarz inequality.

## 1 Introduction and motivation

To measure linear correlation between non degenerate, square integrable random variables $X$ and $Y$, the most widely known and used measure is probably Pearson's correlation coefficient

$$\rho(X,Y) = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}. \tag{1.1}$$

For two random variables $X$ and $Y$ with distributions functions $F$ and $G$ having finite positive variance, Pearson's correlation satisfies the following remarkable properties (see for instance McNeil et al., 2015, Sect. 7.2.2):

P1. $\rho(X,Y) = \rho(Y,X)$.

P2. $\rho(\alpha_1 X + \beta_1, \alpha_2 Y + \beta_2) = \frac{\alpha_1\alpha_2}{|\alpha_1\alpha_2|}\rho(X,Y)$, for $\alpha_1, \alpha_2 \neq 0, \beta_1, \beta_2 \in \mathbb{R}$.

P3. $\rho(X,Y) \in [\rho_{\min}, \rho_{\max}] \subset [-1,1]$, where

$$\rho_{\min} = \inf\{\rho(X,Y); X \sim F, Y \sim G\}, \quad \rho_{\max} = \sup\{\rho(X,Y); X \sim F, Y \sim G\}.$$

In particular:

(a) $\rho_{\min} < 0 < \rho_{\max}$;

(b) $\rho(X,Y) = \rho_{\max}$ if and only if the pair $(X,Y)$ is comonotonic;

(c) $\rho(X,Y) = \rho_{\min}$ if and only if the pair $(X,Y)$ is countermonotonic;

(d) $\rho_{\max} = 1$ if and only if $Y \stackrel{\mathrm{d}}{=} aX + b, a > 0$;

(e) $\rho_{\min} = -1$ if and only if $Y \stackrel{\mathrm{d}}{=} aX + b, a < 0$;.

P4. If $X$ is independent of $Y$, then $\rho(X,Y) = 0$ (the converse does not hold).

The goal of this paper is to define a measure of linear correlation $\rho(X,Y)$ between two non degenerate, square integrable $\mathbb{R}^d$-valued random *vectors* $X$ and $Y$, so that the above listed properties of Pearson's correlation are maintained for $d > 1$, when the notions of co(unter)monotonicity and equality in type ($Y \stackrel{\mathrm{d}}{=} aX + b$) are replaced by suitable multivariate analogues. For two random vectors $X$ and $Y$ having invertible correlation matrices $\Sigma_X$ and $\Sigma_Y$, and cross-covariance matrix $\Sigma_{XY}$, our novel measure of correlation will be defined as

$$\rho(X,Y) = \frac{\operatorname{tr}(\Sigma_{XY})}{\operatorname{tr}\left((\Sigma_X \Sigma_Y)^{1/2}\right)} .$$

As we will illustrate in the remainder of the paper, the main idea behind the new coefficient is to measure the covariance between the random vectors $X$ and $Y$ in terms of $\operatorname{tr}(\Sigma_{XY})$ and then standardize it by its maximum attainable value given solely the knowledge of their covariance matrices $\Sigma_X$ and $\Sigma_Y$.

The newly introduced correlation $\rho(X,Y)$ measures the similarity between $X$ and $Y$ in terms of a single number in $[-1,1]$. The maximal value of correlation is attained in the case that the average squared Euclidean distance between $X$ and $Y$ is the minimum possible given their marginal distributions. This maximal value is equal to the upper bound 1 if, additionally, it is possible to express $Y$ as an affine function of $X$. Conversely, the minimal value of $\rho(X,Y)$ indicates maximal dissimilarity between $X$ and $Y$. Notice that this interpretation of $\rho$ holds in arbitrary dimensions $d$, including $d = 1$ (hence it holds similarly for Pearson's correlation) and is consistent with the original idea of correlation, as a measure of similarity (distance).

### Existing association measures

To measure the linear correlation between $X$ and $Y$, one could simply use $\Sigma_{XY}$. However, in many statistical applications where the similarity between vectors is of special interest, a scalar measure of association might be useful.

Apart from the early Canonical Correlation measure of Hotelling (1936) and the RV coefficient defined in Robert and Escoufier (1976), the question of measuring multivariate dependence in terms of a single number has raised a limited interest in the literature. Most coefficients surveyed in Josse and Holmes (2016) concern tests of independence between random vectors (or multivariate samples) rather than actually measuring the linear correlation between them. This is also the spirit of Zhu et al. (2017) and Jin and Matteson (2018). Distance correlation as introduced in Székely et al. (2007) belongs to the interval $[0,1]$ with the lower bound 0 representing independence. However, in applications one often needs to distinguish between positive and negative dependence, possibly indicated by positive and negative values. Moreover, a dependence measure should also provide a margin-free measurement of association between vectors, i.e. should be invariant with respect to increasing transformation of the margins.

In general, it seems natural to establish the properties that a dependence measure $\rho(X,Y)$ should satisfy. This is the direction taken by Grothe et al. (2014), where under an axiomatic framework several marginal-invariant measures of association have been introduced as multivariate generalization of Spearman's rho and Kendall's tau rank correlation measures. Recently, Hofert et al. (2019) followed and extended this axiomatic set-up defining so-called collapsed measures of association. In Section 3.3, we will see that our newly defined coefficient, though inspired by a different philosophy, behaves similarly to some of these recently introduced notions. Finally, a new non-parametric measure of association between an arbitrary number of random vectors has been also proposed in Medovikov and Prokhorov (2017) and ordinal pattern dependence as a measure of multivariate dependence has been introduced in Schnurr (2014) and further analyzed in Betken et al. (2021). A concise overview of all approaches used to extend rank correlation coefficients to multivariate spaces is given in Han (2021).

### A mass transportation approach

Our motivation for defining a novel correlation coefficient is driven by an optimization approach, and finds its roots in the theory of mass transportation (Rüschendorf and Rachev, 1990; Rüschendorf, 2013). By writing

$$\mathrm{cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y],$$

one immediately sees that, over the class of all possible pairs $(X, Y)$ of random variables with given marginal distributions, Pearson's correlation is maximized when the mean of the product between $X$ and $Y$ is maximized. This occurs when $X$ and $Y$ are *comonotonic*.

**Definition 1.1.** The set $\Gamma \subset \mathbb{R} \times \mathbb{R}$ is said to be *comonotonic* if it is totally ordered, i.e. for any $(x_1, y_1), (x_2, y_2) \in \Gamma$ one has

$$(x_2 - x_1)(y_2 - y_1) \geq 0. \tag{1.2}$$

Any pair of random variables $(X, Y)$ with a comonotonic support is called *comonotonic*. Similarly, $(X, Y)$ is called *countermonotonic* if the inequality (1.2) holds reversed for its support or, equivalently, if $(X, -Y)$ is comonotonic.

The notions of comonotonicity and countermonotonicity represent two benchmarks in statistical modeling and copula theory, as they represent perfect positive, respectively negative, dependence; see for instance Puccetti and Wang (2015) for an historical survey on the matter or Nelsen (2006), Durante and Sempi (2016) for textbook treatments.

It is well known that $X$ and $Y$ are comonotonic if and only if $(X, Y) \stackrel{\mathrm{d}}{=} (f_1(U), f_2(U))$, for some (common) random variable $U$ and increasing measurable functions $f_1$ and $f_2$; see for instance Dhaene et al. (2002) or Puccetti and Scarsini (2010) and references therein. For comonotonic random variables, the expectation of the product $\mathbb{E}[XY]$ is maximized and Pearson's correlation attains its maximal value $\rho_{\max}$ (property P3b). When the increasing functions $f_1$ and $f_2$ can be chosen to be linear then Pearson's attains value 1, i.e. $\mathrm{cov}(X, Y) = \sqrt{\mathrm{var}(X)\mathrm{var}(Y)}$. In this case, we have that $Y \stackrel{\mathrm{d}}{=} aX + b, a > 0$ and $X$ and $Y$ are called of the same type (property P3d); see for instance Definition A.1 in McNeil et al. (2015).

**Definition 1.2.** The two random variables $X$ and $Y$ are said to be *of the same type* if $Y \stackrel{\mathrm{d}}{=} aX + b$ for some $a > 0$.

Analogously, when $X$ and $Y$ are *countermonotonic* (P3c) (i.e $X$ is an increasing, resp. $Y$ a decreasing function of some common random variable $U$), Pearson's correlation attains its minimum value $\rho_{\min}$, which is $-1$ if and only if $X$ is of the same type of $-Y$ (P3e).

Following this optimization viewpoint, it appears natural to require for a measure of linear correlation between two *vectors* to be maximized (minimized) when the expectation of their *inner product* $\langle \cdot\, , \cdot \rangle$ is maximized (minimized), and this over all possible pairs of vectors having the same multivariate marginal distributions. Using the formula

$$||x - y||^2 = ||x||^2 - 2\langle x, y\rangle + ||y||^2,$$

it follows that the maximization of the expected inner product of two vectors is equivalent to the minimization of the average squared Euclidean distance between their distributions; see Puccetti (2017). As a consequence, the inner product can be interpreted as a measure of similarity between $d$-variate vectors, coherently with the original scope of Pearson's correlation as pursued for instance in the early studies of Hoeffding (1940) and Fréchet (1951). Based on these considerations, in the following we will measure the covariance between two random vectors $X$ and $Y$ by the expectation of their inner product, and we will standardize it properly to obtain a new measure of correlation.

Borrowing the taxonomy introduced in Puccetti and Scarsini (2010), to which we refer for a review of the results illustrated below, pair of vectors maximizing (minimizing) their inner product will be named c-co(unter)monotonic.

**Definition 1.3.** The set $\Gamma \subset \mathbb{R}^d \times \mathbb{R}^d$ is said to be *c-comonotonic* if for any finite number $m > 1$ of points $(x_i, y_i) \in \Gamma, i = 1, \ldots, m$, and any possible permutation $\sigma$ of $\{1, \ldots, m\}$, we have

$$\sum_{i=1}^{m} \langle x_i, y_i \rangle \geq \sum_{i=1}^{m} \langle x_i, y_{\sigma(i)} \rangle. \tag{1.3}$$

Any pair of random vectors $(X, Y)$ with a c-comonotonic support is called *c-comonotonic*. Similarly, the pair of vectors $(X, Y)$ is called *c-countermonotonic* if the inequality (1.3) holds reversed for its support or, equivalently, if $(X, -Y)$ is c-comonotonic.

C-countermonotonicity is a generalization of the notion of comonotonicity for random variables and as such denotes a property of a set $\Gamma$. The term c-comonotonicity stands for *cyclic*-comonotonicity as a c-comonotonic set is in fact a *cyclically monotone set* as defined in Rockafellar (1970).

When $d = 1$, c-co(unter)monotonicity reduces to the standard notion of co(unter)monotonicity between random variables. It is well known that a c-comonotonic pair $(X, Y)$ always exists for any choice of multivariate marginal distributions and that it maximizes the expected inner product $\langle X, Y\rangle$ of its components.

It is quite intuitive that condition (1.3) leads to maximal inner product of $X$ and $Y$. It requires that one cannot increase the average inner product between the points in the support of $X$ and those in the support of $Y$ by selecting a different coupling (transportation plan). As a consequence, our novel correlation coefficient will be maximized by c-comonotonic pairs of vectors and, similarly, minimized by c-countermonotonic ones.

As in the univariate case, there exists a useful characterization of c-comonotonicity. A pair of vector $(X, Y)$ is c-comonotonic if and only if $Y$ almost surely belongs to the subdifferential of a lower semicontinuos convex function $f$ of $X$; see Gangbo and McCann (1996) or Rüschendorf (1996).

Under mild assumptions (e.g. continuity of $X$), such convex $f$ can be made almost everywhere differentiable and a c-comonotonic pair of vectors can be represented as $(X, \nabla f(X))$. The relation $Y \overset{\text{a.s.}}{=} \nabla f(X)$ can be interpreted as the generalization of the notion of increasingness. Suppose that $X$ is continuous. For $d = 1$, the random variables $X$ and $Y$ are comonotonic iff $Y$ is a.s. an increasing function of $X$ (see for instance McNeil et al., 2015, Cor. 7.19). In higher dimensions $d > 1$, the pair $(X, Y)$ is c-comonotonic iff $Y$ is the gradient of a convex differentiable function of $X$; and the gradient of a convex function is exactly an increasing function on the line.

We will see that our novel correlation measure will attain its maximal value $\rho_{\max}$ for c-comonotonic pairs of vectors $(X, Y)$. Similarly to the univariate case, when it is additionally possible to establish an affine relationship between $X$ and $Y$, we will also have $\rho_{\max} = 1$. Theorem 2.1 will justify the following (novel) multivariate definition of equality in type.

**Definition 1.4.** Two $\mathbb{R}^d$-valued random vectors $X$ and $Y$ are said to be of *the same type* if there exists a positive definite $d \times d$ matrix $A$ and $b \in \mathbb{R}^d$ such that $Y \overset{\text{d}}{=} AX + b$.

For $d = 1$, Definition 1.4 returns the univariate notion of random variables of the same type in Definition 1.2. For $d > 1$, notice that if $X_i$ is of the same type of $Y_i$, for all $i = 1, \dots, d$, then the two vectors $X$ and $Y$ are of the same type. However, the above definition is more general as one can immediately see by taking $X = (X_1, X_2)'$ with $X_1 \sim N(0, 1)$ independent of $X_2 \sim U(0, 1)$, and $Y \overset{\text{a.s.}}{=} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} X$. In this case, $X$ and $Y$ are of the same type but not $X_i$ and $Y_i$.

#### Summary and some caveats

A brief summary of the paper follows. In Section 2, we show that the newly proposed correlation coefficient satisfies all the above listed properties P1-P4 stated in a multi-dimensional setting; in Section 3, we provide some simulative studies showing the limiting distribution of its sample estimator and corresponding rank correlation coefficient. We also give a comparative study, based on financial data, with respect to some of the measures of associations existing in the literature. Section 4 concludes the paper with some final remarks and examples. Throughout the paper, we will also give some auxiliary results of independent interest in matrix analysis and mass transportation theory.

Other scalar measures of dependence/association are present in the literature, and in the following they are to be compared with our proposed coefficient. For all such measures, including the one introduced here, one must be aware that a single number cannot generally and univocally determine the dependence of $(X, Y)$, even when $X$ and $Y$ are random variables.

Having said that, there is not a universally acknowledged way of measuring association between random vectors and each measure has its merit and drawbacks. This paper proposes a novel, natural, intuitive, and appealingly simple generalization of Pearson's linear correlation coefficient to measure linear association between random vectors and, in that respect, the proposed coefficient satisfies all relevant properties of Pearson's in arbitrary dimensions and is based on a solid mass transportation approach.

Being based on the minimization of the squared Euclidean distance, the new correlation can only be used for random vectors of the same dimension and is not invariant with respect to permutation of the coordinates within one vector. As such, it is especially suited for three-way (also big) data structures. These limitations are discussed in Section 4 where possible solutions and further extensions/applications are also proposed.

Finally, we note that this paper studies association *between* random vectors, but there is a whole literature concerned about association *within* a random vector. On this latter different topic, we refer the reader to the survey Schmid et al. (2010) and the references therein.

Optimization problems in mass transportation related to this paper have recently gathered a considerable interest with applications in multidimensional risk theory; see for instance Ekeland et al. (2012), Rüschendorf (2012), Carlier et al. (2016), Puccetti (2017), and the more recent Böttcher et al. (2019), and Mordant and Segers (2022). The style of this paper and the applications to follow are taking inspiration from this latter field.

### Notation and preliminaries

Throughout this paper, $d$ and $n$ are positive integers: $d$ always refers to the dimension of random vectors while $n$ indicates (from Section 3 on) the sample size of (simulated) data. For $k \in \mathbb{N}$, we denote by $L_k(\mathbb{R}^d)$ the set of non degenerate $\mathbb{R}^d$-valued random vectors $X = (X_1, \dots, X_d)'$, defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and such that $\mathbb{E}[||X||^k] < \infty$. Let $P_k(\mathbb{R}^d)$ denote the set of the corresponding Borel probability measures on $\mathbb{R}^d$. We write $X \sim F$ to mean that $F$ is the probability distribution of the random vector $X$ and $X \overset{\mathrm{d}}{=} Y$ to mean that $X$ and $Y$ have the same law.

We denote by $\mathcal{M}_d$ the set of real-valued square matrices of order $d$. We use the term positive matrix for a (symmetric) positive semidefinite matrix or strictly positive for a (symmetric) positive definite matrix. For all the matrix analysis results stated in this section, we refer the reader to the standard references Bhatia (2007) and Horn and Johnson (2013).

For a matrix $A \in \mathcal{M}_d$, we denote by $A^{1/2}$ the *principal square root* of $A$, that is any matrix $B$ satisfying the equation $B^2 = A$, *and* whose eigenvalues have positive real part. Such a matrix exists and is unique if $A$ has no nonpositive real eigenvalues. In particular, if $A$ is (strictly) positive then $A^{1/2}$ is unique and (strictly) positive. There exist several methods to compute the principal square root of a matrix; see for instance Higham (1997). In our applications to follow we use the R package `expm` maintained by Martin Mächler or the MATLAB function `sqrtm`. For $2 \times 2$ matrices, an explicit formula is also available.

The geometric mean of two strictly positive matrices $A$ and $B$ is the strictly positive matrix defined as

$$A \# B = A^{1/2} \left( A^{-1/2} B A^{-1/2} \right)^{1/2} A^{1/2}. \tag{1.4}$$

One can prove that $A \# B$ is the unique strictly positive solution of the equation $X A^{-1} X = B$ and that

$$A \# B = B \# A = A(A^{-1}B)^{1/2} = (BA^{-1})^{1/2} A. \tag{1.5}$$

When $A$ and $B$ commute (i.e. $AB = BA$), one has $A \# B = A^{1/2} B^{1/2} = B^{1/2} A^{1/2} = (AB)^{1/2}$.

For $X, Y \in L_2(\mathbb{R}^d)$, let $\Sigma_{XY} \in \mathcal{M}_d$ be the cross-covariance matrix of $X$ and $Y$, defined as

$$(\Sigma_{XY})_{i,j} = \mathrm{cov}(X_i, Y_j), \text{ for } i, j = 1, \dots, d.$$

We also denote by $\Sigma_X = \Sigma_{XX}$ and $\Sigma_Y = \Sigma_{YY}$ the covariance matrices of $X$ and $Y$. $\Sigma_X$ and $\Sigma_Y$ are in general positive but, under the framework of Section 2, they will be required to be invertible, i.e. strictly positive. The following auxiliary result will be used in what follows.

**Lemma 1.1.** *For $A = (a_{ij}) \in \mathcal{M}_d$, and $X, Y \in L_2(\mathbb{R}^d)$, we have*

$$\operatorname{tr}\left(\Sigma_{X(AY)}\right) = \sum_{i=1}^{d} \operatorname{cov}\left(X_i, \sum_{j=1}^{d} a_{ij} Y_j\right) = \sum_{i=1}^{d}\sum_{j=1}^{d} a_{ij} \operatorname{cov}(X_i, Y_j)$$
$$= \sum_{i=1}^{d}\sum_{j=1}^{d} (A \circ \Sigma_{XY})_{ij} = \operatorname{tr}\left(A\Sigma'_{XY}\right).$$

## 2 Linear correlation between random vectors

Let $X$ and $Y$ in $L_2(\mathbb{R}^d)$ have *invertible* correlation matrices $\Sigma_X$ and, respectively, $\Sigma_Y$. Relaxations of this invertibility assumption will be treated in Section 4. We now introduce a new linear correlation coefficient between random vectors and prove that it satisfies properties P1-P4 where the notions of co(unter)monotonicity and equality in type are replaced by their multivariate analogous as given in Definitions 1.3–1.4.

**Definition 2.1.** For two random vectors $X, Y \in L_2(\mathbb{R}^d)$ having invertible correlation matrices $\Sigma_X$ and $\Sigma_Y$, we define the correlation coefficient

$$\rho(X, Y) = \frac{\operatorname{tr}(\Sigma_{XY})}{\operatorname{tr}\left((\Sigma_X \Sigma_Y)^{1/2}\right)}. \tag{2.1}$$

It is straightforward to check that, for $d = 1$, $\rho(X, Y)$ as defined in (2.1) gives Pearson's correlation coefficient (1.1). We also notice that $\rho(X, X) = 1$ and $\rho(X, -X) = -1$. In the following remark, we show that the definition of $\rho(X, Y)$ is well posed in any dimension $d$.

**Remark 2.1.** For the two strictly positive matrices $\Sigma_X, \Sigma_Y$, we have that $\Sigma_X^{-1} \# \Sigma_Y$ is unique and strictly positive and, using (1.5), one can write

$$\Sigma_X \Sigma_X^{-1} \# \Sigma_Y = \Sigma_X \Sigma_X^{-1} (\Sigma_X \Sigma_Y)^{1/2} = (\Sigma_X \Sigma_Y)^{1/2}.$$

Hence the denominator in (2.1) is unique and well-defined. Similarly,

$$\Sigma_X^{-1} \# \Sigma_Y \Sigma_X = (\Sigma_Y \Sigma_X)^{1/2} \Sigma_X^{-1} \Sigma_X = (\Sigma_Y \Sigma_X)^{1/2}.$$

As $\operatorname{tr}\left(\Sigma_X^{-1} \# \Sigma_Y \Sigma_X\right) = \operatorname{tr}\left(\Sigma_X \Sigma_X^{-1} \# \Sigma_Y\right)$, from the previous equations one also obtains

$$\operatorname{tr}\left((\Sigma_X \Sigma_Y)^{1/2}\right) = \operatorname{tr}\left((\Sigma_Y \Sigma_X)^{1/2}\right). \tag{2.2}$$

If $\Sigma_X$ and $\Sigma_Y$ commute (e.g. if they are equi-correlation, hence circulant, matrices), we have that $(\Sigma_X^{-1} \# \Sigma_Y) = \Sigma_X^{-1/2} \Sigma_Y^{1/2}$ and one immediately computes

$$\operatorname{tr}\left(\Sigma_X \Sigma_Y\right)^{1/2} = \operatorname{tr}\left(\Sigma_X^{1/2} \Sigma_Y^{1/2}\right).$$

**Theorem 2.1.** *For a pair of random vectors $X, Y \in L_2(\mathbb{R}^d)$ with invertible correlation matrices $\Sigma_X$ and $\Sigma_Y$ and distribution functions $X \sim F$ and $Y \sim G$, the correlation coefficient $\rho(X, Y)$ defined in (2.1) satisfies the following properties:*

P1. *$\rho(X, Y) = \rho(Y, X)$.*

P2. *$\rho(\alpha_1 X + \beta_1, \alpha_2 Y + \beta_2) = \frac{\alpha_1 \alpha_2}{|\alpha_1 \alpha_2|} \rho(X, Y)$, for $\alpha_1, \alpha_2 \neq 0, \beta_1, \beta_2 \in \mathbb{R}^d$.*

P3. *$\rho(X, Y) \in [\rho_{min}, \rho_{max}] \subset [-1, 1]$, where*

$$\rho_{min} = \inf\{\rho(X, Y); X \sim F, Y \sim G\}, \quad \rho_{max} = \sup\{\rho(X, Y); X \sim F, Y \sim G\}.$$

*In particular:*

*(a)* $\rho_{min} < 0 < \rho_{max}$*;*

*(b)* $\rho(X,Y) = \rho_{max}$ *if and only if the pair* $(X,Y)$ *is c-comonotonic;*

*(c)* $\rho(X,Y) = \rho_{min}$ *if and only if the pair* $(X,Y)$ *is c-countermonotonic;*

*(d)* $\rho_{max} = 1$ *if and only if* $X$ *and* $Y$ *are of the same type.*

*(e)* $\rho_{min} = -1$ *if and only if* $X$ *and* $-Y$ *are of the same type.*

P4. *If* $X$ *is independent of* $Y$*, then* $\rho(X,Y) = 0$ *(the converse does not hold).*

*Proof.* Property P1 follows from (2.2) and $\mathrm{tr}(\Sigma_{XY}) = \mathrm{tr}(\Sigma_{YX})$. For arbitrary $\alpha_1, \alpha_2 \neq 0, \beta_1, \beta_2 \in \mathbb{R}^d$, with $X' = \alpha_1 X + \beta_1$, $Y' = \alpha_2 Y + \beta_2$, one directly checks P2:

$$\rho(\alpha_1 X + \beta_1, \alpha_2 Y + \beta_2) = \frac{\mathrm{tr}(\Sigma_{X'Y'})}{\mathrm{tr}\left((\Sigma_{X'}\Sigma_{Y'})^{1/2}\right)} = \frac{\alpha_1\alpha_2\,\mathrm{tr}(\Sigma_{XY})}{\mathrm{tr}\left((\alpha_1^2\,\alpha_2^2\,\Sigma_X\;\Sigma_Y)^{1/2}\right)}$$

$$= \frac{\alpha_1\alpha_2\,\mathrm{tr}(\Sigma_{XY})}{|\alpha_1\alpha_2|\,\mathrm{tr}\left((\Sigma_X\Sigma_Y)^{1/2}\right)} = \frac{\alpha_1\alpha_2}{|\alpha_1\alpha_2|}\rho(X,Y).$$

In view of this property, in the reminder of the proof we assume without loss of generality that $X$ and $Y$ have null mean vectors.

Given two probability distributions $F, G \in P_2(\mathbb{R}^d)$ with null mean, now define

$$m(F,G) = \inf\left\{\mathrm{tr}(\Sigma_{XY}); X \sim F, Y \sim G\right\}, \tag{2.3a}$$

$$M(F,G) = \sup\left\{\mathrm{tr}(\Sigma_{XY}); X \sim F, Y \sim G\right\}. \tag{2.3b}$$

A standard continuity/compactness argument in mass transportation theory (see e.g. Rüschendorf (1991)) shows that each value $u \in [m(F,G), M(F,G)]$ is attained by some pair of vectors $(X_u, Y_u)$ with $X_u \sim F$ and $Y_u \sim G$. Since $M(F,G)$ is bounded (see for instance (2.5) below), by Theorem 2.2 and Corollary 2.4 in Gangbo and McCann (1996), $\mathrm{tr}(\Sigma_{XY}) = M(F,G)$ holds true if and only if the pair $(X,Y)$ with marginals $F$ and $G$ is c-comonotonic according to Definition 1.3. Analogously, one immediately sees from Lemma 2.1 in Rüschendorf (1996) that $\mathrm{tr}(\Sigma_{XY}) = m(F,G)$ holds if and only if $(X,-Y)$ is c-comonotonic. This shows property P3b-c.

Denoted by $\Sigma_X$ and $\Sigma_Y$ the two correlation matrices of any $X \sim F$ and, respectively, $Y \sim G$, the inequalities

$$-\,\mathrm{tr}\left((\Sigma_X\Sigma_Y)^{1/2}\right) \leq m(F,G) \leq M(F,G) \leq \mathrm{tr}\left((\Sigma_X\Sigma_Y)^{1/2}\right) \tag{2.4}$$

can be derived, applying the formula $||x-y||^2 = ||x||^2 - 2\langle x,y\rangle + ||y||^2$, from the main results of Dowson and Landau (1982) and Olkin and Pukelsheim (1982). When $\Sigma_X$ and $\Sigma_Y$ are invertible, Theorem 4 in Olkin and Pukelsheim (1982) implies that $M(F,G) = \mathrm{tr}\left((\Sigma_X\Sigma_Y)^{1/2}\right)$ if and only if $X$ and $Y$ are of the same type according to Definition 1.4 (property P3d). Notice that $M(F,G) = \mathrm{tr}\left((\Sigma_X\Sigma_Y)^{1/2}\right)$ is attained by the c-comonotonic pair $(X,Y)$, with $Y \overset{\text{a.s.}}{=} TX$ and $T = \Sigma_X^{-1}\#\Sigma_Y$ (see Example 2.1). In this latter case $\rho(X,Y) = 1$. By switching the sign of $T$, one obtains P3e.

From Property P4 (immediate), $\rho(X,Y) = 0$ in case of independence and hence $\rho_{\max} \geq 0$. If $\rho_{\max} = 0$ then the independent pair $(X,Y)$ is optimal, hence c-comonotonic. However, this is not possible as the corresponding law does not belong to the subdifferential of a convex function unless $X$ is degenerate. Hence $\rho_{\max} > 0$ and, similarly, $\rho_{\min} < 0$ (P3a).

We conclude the proof by noting that the converse of P4 does not hold. As an example, take $X \overset{\text{a.s.}}{=} (U,V)'$, $Y \overset{\text{a.s.}}{=} (U,-V)'$, where $U$ and $V$ are two independent copies of $U(0,1)$. We have $\rho(X,Y) = 0$ without $X$ and $Y$ being stochastically independent. □

The main idea behind Definition 2.1 is to measure the covariance between the random vectors $X$ and $Y$ in terms of $\mathrm{tr}(\Sigma_{XY})$ and then standardize it by its maximum attainable value given solely the knowledge of their covariance matrices $\Sigma_X$ and $\Sigma_Y$. An alternative standardization is discussed in Remark 2.3 below.

**Remark 2.2.** For $\mathrm{tr}(\Sigma_{XY}) = \mathrm{tr}\left((\Sigma_X\Sigma_Y)^{1/2}\right)$ to hold, it is necessary that $Y \overset{\mathrm{d}}{=} AX + b$ with A *strictly positive* (property P3d). If, for instance, one takes $Y \overset{\mathrm{a.s.}}{=} AX$ with $A$ assumed to be *only* symmetric, the second equality in

$$\mathrm{tr}\left((\Sigma_X\Sigma_Y)^{1/2}\right) = \mathrm{tr}\left((\Sigma_X A\Sigma_X A)^{1/2}\right) = \mathrm{tr}(\Sigma_X A) = \mathrm{tr}(A\Sigma_X) = \mathrm{tr}(\Sigma_{X(AX)}) = \mathrm{tr}(\Sigma_{XY})$$

might be wrong as the matrix $\Sigma_X A$ might not be the *principal* square root of $(\Sigma_X A\Sigma_X A)$. As an example consider

$$\Sigma_X = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \Sigma_Y = \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}.$$

One can check that the four matrices $A_{1,2} = \pm\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}, A_{3,4} = \pm\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ are all square roots of $\Sigma_Y$. Our framework requires the use of the *principal square root*, that is the one whose eigenvalues have positive real part. Hence

$$\Sigma_Y^{1/2} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix},$$

the latter being a strictly positive matrix, hence delivering an higher trace (covariance).

**Remark 2.3.** Cauchy-Schwarz inequality for random variables gives the trivial bound

$$|\mathrm{tr}(\Sigma_{XY})| \leq \sum_{i=1}^{d} \sqrt{\mathrm{var}(X_i)\mathrm{var}(Y_i)}. \tag{2.5}$$

In view of (2.5), it would have been more intuitive to define a coefficient $\rho'(X,Y)$ measuring correlation between vectors as

$$\rho'(X,Y) := \frac{\mathrm{tr}(\Sigma_{XY})}{\sum_{i=1}^{d} \sqrt{\mathrm{var}(X_i)\mathrm{var}(Y_i)}}.$$

One immediately sees that $\rho'(X,Y)$ still belongs to a closed interval $[\rho_{\min}, \rho_{\max}] \subset [-1,1]$. However, $\rho'(X,Y)$ attains the extremal value 1 ($-1$) if and only if each $X_i$ is of the same type of $Y_i$ ($-Y_i$). This latter case is considered here too restrictive as it implies, for instance, that the pair $(X,Y)$ is $\pi$-co(unter)monotonic (in general, not c-co(unter)monotonic) and $X$ and $Y$ necessarily have the same dependence structure (copula); see Puccetti and Scarsini (2010) for a definition and more details on $\pi$-comonotonicity.

The coefficient $\rho(X,Y)$ defined in (2.1) allows the extremal values $|\rho(X,Y)| = 1$ to be attained for a broader class of marginal distributions. In fact, one can immediately see that in the particular case that $Y_i \overset{\mathrm{d}}{=} a_i X_i + b_i$, for some $a_i > 0, b_i \in \mathbb{R}$, we have that $\Sigma_Y = D\Sigma_X D$, where $D = \mathrm{diag}(a_1,\dots,a_d)$, and $\rho(X,Y) = \rho'(X,Y)$.

In the following corollary, we collect some direct consequences of properties P1-P4. In particular, P3 implies (2.6) below, an improvement of the Cauchy-Schwarz inequality for cross-covariance matrices which, to our knowledge, has not been given attention in multivariate statistics or matrix theory.

**Corollary 2.2** (Improvement of the Cauchy-Schwarz inequality)**.** *The following statements hold true.*

1. *For $X, Y \in L_2(\mathbb{R}^d)$ having invertible covariance matrices $\Sigma_X$ and, respectively, $\Sigma_Y$, we have that*

$$\left|\mathrm{tr}(\Sigma_{XY})\right| \leq \mathrm{tr}\left((\Sigma_X\Sigma_Y)^{1/2}\right) \leq \sum_{i=1}^{d} \sqrt{\mathrm{var}(X_i)\mathrm{var}(Y_i)}. \tag{2.6}$$

2. *For any strictly positive matrices $\Sigma_X, \Sigma_Y$, one has that $(\Sigma_X\Sigma_Y)^{1/2}$ is uniquely defined and*

$$\mathrm{tr}\left((\Sigma_X\Sigma_Y)^{1/2}\right) = \mathrm{tr}\left((\Sigma_Y\Sigma_X)^{1/2}\right).$$

The newly introduced correlation measures the similarity between $X$ and $Y$ in terms of a single number $\rho(X,Y) \in [-1,1]$. The value $\rho(X,Y) = 1$ corresponds to the case when $Y \overset{\text{a.s.}}{=} AX + b$, with $A$ being a positive definite matrix and $b \in \mathbb{R}^d$. The linear relationship $A$ can be extracted by the formula $A = \Sigma_X^{-1} \# \Sigma_Y$, that is $A$ is equal to the geometric mean of $\Sigma_X^{-1}$ and $\Sigma_Y$; see (1.4).

In general, the extremal values 1 and $-1$ for the correlation coefficient (2.1) can only be attained when $X$ and $Y$ are of the same type according to Definition 1.4. We illustrate these properties in the following pedagogical example.

**Example 2.1.** Assume $X$ and $Y$ have bivariate Gaussian distributions $X \sim \mathcal{N}(\mu_X, \Sigma_X)$ and $Y \sim \mathcal{N}(\mu_Y, \Sigma_Y)$, with null mean vectors $\mu_X = \mu_Y = (0,0)'$ and covariance matrices

$$\Sigma_X = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \ \Sigma_Y = \begin{pmatrix} 1 & \frac{2}{5} \\ \frac{2}{5} & 1 \end{pmatrix}.$$

From Corollary 3.2.13 in Rachev and Rüschendorf (1998), one has that $A = \Sigma_X^{-1} \# \Sigma_Y$ is the only strictly positive matrix such that $A \Sigma_X A' = \Sigma_Y$. Define $\hat{Y} \overset{\text{a.s.}}{=} AX \overset{\text{d}}{=} Y$, where

$$A = \Sigma_X^{-1} \# \Sigma_Y = \begin{pmatrix} a & b \\ b & a \end{pmatrix}, \ a = \frac{\sqrt{35} + \sqrt{15}}{10}, \ b = \frac{\sqrt{35} - \sqrt{15}}{10}. \tag{2.7}$$

Being positive definite, $A$ provides a linear optimal transport between the laws of $X$ and $Y$, hence the squared distance $\mathbb{E}||X - \hat{Y}||^2$ is the minimal possible given the two distributions of $X$ and $Y$. Equivalently stated, $X$ and $\hat{Y}$ are c-comonotonic and $\rho(X, \hat{Y})$ attains its maximal value $\rho_{\max}$. Since $Y$ is also of the same type of $X$, we also have that $\rho(X, \hat{Y}) = \rho_{\max} = 1$.

Similarly, due to $-AX \overset{\text{d}}{=} Y$, it also follows that $\rho_{\min} = -1$. In this example, notice that the second inequality in (2.6) is strict, that is

$$\operatorname{tr}\left( (\Sigma_X \Sigma_Y)^{1/2} \right) \simeq 1.9578 < 2 = \sum_{i=1}^{2} \sqrt{\operatorname{var}(X_i)\operatorname{var}(Y_i)},$$

because $X$ and $Y$ are not $\pi$-comonotonic (see Remark 2.3).

Let now define a different linear transformation of $X$, namely $Y' \overset{\text{a.s.}}{=} BX$ with

$$B = \begin{pmatrix} \sqrt{3}/2 & 1/2 \\ c & 4/5 - \sqrt{3}c \end{pmatrix}, \quad c = \frac{\sqrt{3}}{10}(2 + \sqrt{7}). \tag{2.8}$$

After checking that $B \Sigma_X B' = \Sigma_Y$ and hence $\hat{Y} \overset{\text{d}}{=} Y' \overset{\text{d}}{=} Y$, one applies Lemma 1.1 to compute

$$\rho(X, Y') = \rho(X, BX) = \frac{\operatorname{tr}(B)}{\operatorname{tr}\left( (\Sigma_X \Sigma_Y)^{1/2} \right)} = \frac{\sqrt{3}/2 + 4/5 - \sqrt{3}c}{1.9578} \simeq 0.1391. \tag{2.9}$$

Despite the linear relationship between $X$ and $Y \overset{\text{a.s.}}{=} BX$, in this latter case we have that

$$1 = \rho(X, X) \neq \rho(X, BX) \simeq 0.1391.$$

This occurs because $B$ does not possess the optimality property in (4.3).

This example shows that, in contrast to Pearson's correlation in dimension $d = 1$, the measure is not (and, of course, cannot be) invariant with respect to general multivariate linear transformations of the vectors, i.e., in general $\rho(X,Y)$ is different from $\rho(X, BY)$. This is a direct consequence of our transportation approach. As already remarked in the Introduction, under the squared Euclidean distance optimal transportation plans are subdifferentials (gradients) of convex functions. If an optimal plan is also linear, it must necessarily come in the form of a symmetric positive definite matrix, while the above defined matrix $B$ is not even symmetric.

## 3 Sample and rank correlation

In this section we present some simulation studies to show how the newly proposed correlation coefficient can be estimated from finite samples. We also introduce the corresponding rank correlation coefficient and show similar limiting properties. Finally, we provide a comparative example with respect to the multivariate dependence notions existing in the literature.

### 3.1 Sample Correlation

We represent $n$ independent realizations of the random vectors $X, Y \in L_2(\mathbb{R}^d)$ by the $n \times d$ matrices $\hat{X}, \hat{Y}$, which (to simplify notation) we assume to be column-centered. We define $\hat{\rho}(\hat{X}, \hat{Y})$, the empirical estimator of $\rho(X, Y)$, as

$$\hat{\rho}(\hat{X}, \hat{Y}) = \frac{\operatorname{tr}(\hat{\Sigma}_{XY})}{\operatorname{tr}\left((\hat{\Sigma}_X \hat{\Sigma}_Y)^{1/2}\right)}, \tag{3.1}$$

where $\hat{\Sigma}_X = \frac{1}{n-1}\hat{X}'\hat{X}$, $\hat{\Sigma}_Y = \frac{1}{n-1}\hat{Y}'\hat{Y}$ and $\hat{\Sigma}_{XY} = \frac{1}{n-1}\hat{X}'\hat{Y}$ are the empirical non-biased estimators of respectively $\Sigma_X, \Sigma_Y$ and $\Sigma_{XY} = \Sigma'_{YX}$. Similarly, define

$$\hat{Z} = (\hat{X}, \hat{Y}), \hat{\Sigma}_Z = \frac{1}{n-1}\hat{Z}'\hat{Z} = \begin{pmatrix} \hat{\Sigma}_X & \hat{\Sigma}_{XY} \\ \hat{\Sigma}_{YX} & \hat{\Sigma}_Y \end{pmatrix}, \text{ and } \Sigma_Z = \begin{pmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{pmatrix}.$$

Using the notation introduced above, the empirical estimator $\hat{\rho}(\hat{X}, \hat{Y})$ can be written as a function $f$ of the sample correlation $\hat{\Sigma}_Z$. In fact, we have $\hat{\rho}(\hat{X}, \hat{Y}) = f(\hat{\Sigma}_Z)$ and $\rho(X, Y) = f(\Sigma_Z)$, where $f : \mathcal{M}_{2d} \to \mathbb{R}$ is defined by $f : \mathcal{M} \to \mathbb{R}$ as

$$f\begin{pmatrix} A & B \\ B' & C \end{pmatrix} = \frac{\operatorname{tr}(B)}{\operatorname{tr}\left((AC)^{1/2}\right)}.$$

At this point, from Theorems 3.1.3–5 in Kollo and von Rosen (2005) or Lemma 3.1 in Robert et al. (1985), one obtains that $\hat{\rho}(\hat{X}, \hat{Y})$ has a Gaussian, non biased asymptotic distribution.

**Theorem 3.1.** *Let $X, Y \in L_4(\mathbb{R}^d)$ with invertible covariance matrices $\Sigma_X, \Sigma_Y$. If $|\rho(X, Y)| < 1$, we have, for $n \to \infty$, that*

$$\sqrt{n}\left(\hat{\rho}(\hat{X}, \hat{Y}) - \rho(X, Y)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2),$$

*for $\sigma^2 > 0$.*

Following the theory in Kollo and von Rosen (2005, Sect. 3.1.2), the asymptotic variance of $\hat{\rho}(\hat{X}, \hat{Y})$ is given by

$$\sigma^2 = \left(\frac{df(H)}{dH}\Big|_{H=\Sigma_Z}\right)' \Pi \left(\frac{df(H)}{dH}\Big|_{H=\Sigma_Z}\right),$$

where $\Pi$ is defined in (3.1.5) of Kollo and von Rosen (2005); see also Lemma 3.1 in Robert et al. (1985). Due to the presence of the square root functional, the explicit computation of $\sigma^2$ appears to be very involved if at all possible. To approximate the asymptotic variance, one can compute its bootstrap estimator, which performs well on the average and seems to be accurate enough for large sample sizes; see the results in Table 1.

We now provide a simulation example showing the asymptotic normality of $\hat{\rho}(\hat{X}, \hat{Y})$ as well as estimates of its asymptotic variance. We simulate from a pair $(X, Y)$ of bivariate ($d = 2$) Gaussian vectors having the marginal distributions $X \sim \mathcal{N}(\mu_1, \Sigma_X)$ and $Y \sim \mathcal{N}(\mu_2, \Sigma_Y)$ as described in Example 2.1, and we sample the value $\rho(X, Y) - \hat{\rho}(\hat{X}, \hat{Y})$ from different scenarios:

- *Scenario I.* $X$ and $Y$ are independent. In this case, we have $\rho(X, Y) = 0$.

- *Scenario B.* $Y \overset{\text{a.s.}}{=} BX$ with $B$ as in (2.8) yielding $\rho(X, Y) \simeq 0.1391$; see (2.9).

- *Scenario* $A_\xi$. $Y \overset{\text{a.s.}}{=} A_\xi X$ with $A_\xi = A + \begin{pmatrix} 0 & \xi \\ -\xi & 0 \end{pmatrix}$ being a $\xi$-perturbation of the matrix $A$ given in (2.7). Recalling that $\Sigma_X = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, we apply Lemma 1.1 to obtain, for $\xi = 0.1$, that

$$\rho(X,Y) = \rho(X, A_\xi X) = \frac{\operatorname{tr}(\Sigma_{X(A_\xi X)})}{\operatorname{tr}\left((\Sigma_X A_\xi \Sigma_X A'_\xi)^{1/2}\right)} = \frac{\operatorname{tr}(A_\xi \Sigma_X)}{\operatorname{tr}\left((\Sigma_X A_\xi \Sigma_X A'_\xi)^{1/2}\right)} = \frac{\operatorname{tr}(A_\xi)}{\operatorname{tr}\left((A_\xi A'_\xi)^{1/2}\right)} \simeq 0.9948.$$

Our results for different values $n$ of the sample size are collected in Table 1 and Figure 1 and fully confirm the asymptotic distribution given in Theorem 3.1. In the tables to follow, we use $\rho$ and $\hat{\rho}$ to indicate $\rho(X,Y)$ and $\hat{\rho}(\hat{X},\hat{Y})$.

| *Scenario I* | $(\rho - \hat{\rho})$ | CI | $\sqrt{n}\,\mathrm{sd}(\rho-\hat{\rho})$ | BS | BS CI |
|---|---|---|---|---|---|
| $n = 100$ | $2.2988e-04$ | $[-0.1424, 0.1438]$ | 0.7284 | 0.7010 | $[0.3855, 1.0754]$ |
| $n = 1000$ | $7.7261e-05$ | $[-0.0446, 0.0445]$ | 0.7202 | 0.7205 | $[0.6175, 0.8327]$ |
| $n = 10000$ | $7.9865e-06$ | $[-0.0142, 0.0141]$ | 0.7221 | 0.7218 | $[0.6883, 0.7557]$ |
| *Scenario B* | $(\rho - \hat{\rho})$ | CI | $\sqrt{n}\,\mathrm{sd}(\rho-\hat{\rho})$ | BS | BS CI |
| $n = 100$ | $-4.8786e-04$ | $[-0.1991, 0.1888]$ | 0.9905 | 0.9379 | $[0.5123, 1.4534]$ |
| $n = 1000$ | $-3.4895e-05$ | $[-0.0606, 0.0603]$ | 0.9765 | 0.9773 | $[0.8315, 1.1339]$ |
| $n = 10000$ | $-1.5957e-05$ | $[-0.0192, 0.0191]$ | 0.9772 | 0.9801 | $[0.9335, 1.0271]$ |
| *Scenario* $A_\xi$ | $(\rho - \hat{\rho})$ | CI | $\sqrt{n}\,\mathrm{sd}(\rho-\hat{\rho})$ | BS | BS CI |
| $n = 100$ | $9.4993e-05$ | $[-2.1148e-04, 6.3764e-04]$ | 0.0023 | 0.0023 | $[0.0009, 0.0044]$ |
| $n = 1000$ | $9.5102e-06$ | $[-1.1062e-04, 1.5519e-04]$ | 0.0022 | 0.0022 | $[0.0015, 0.0028]$ |
| $n = 10000$ | $8.9191e-07$ | $[-3.9772e-05, 4.4011e-05]$ | 0.0021 | 0.0021 | $[0.0019, 0.0024]$ |

Table 1: Values of $(\rho - \hat{\rho})$ estimated over $10^5$ repetitions of a simulated sample of size $n$ from a bivariate ($d = 2$) Gaussian pair $(X,Y)$ as described in Section 3.1. Estimates are computed with their 95% empirical confidence interval (CI) and their standard deviation $\mathrm{sd}(\rho-\hat{\rho})$ multiplied by $\sqrt{n}$. Estimates of standard deviation are compared with the corresponding bootstrap estimates (BS) obtained by resampling $n/10$ times from a single simulated sample of size $n$. We show average estimate over $10^4$ repetitions with their 95% empirical confidence interval (BS CI).

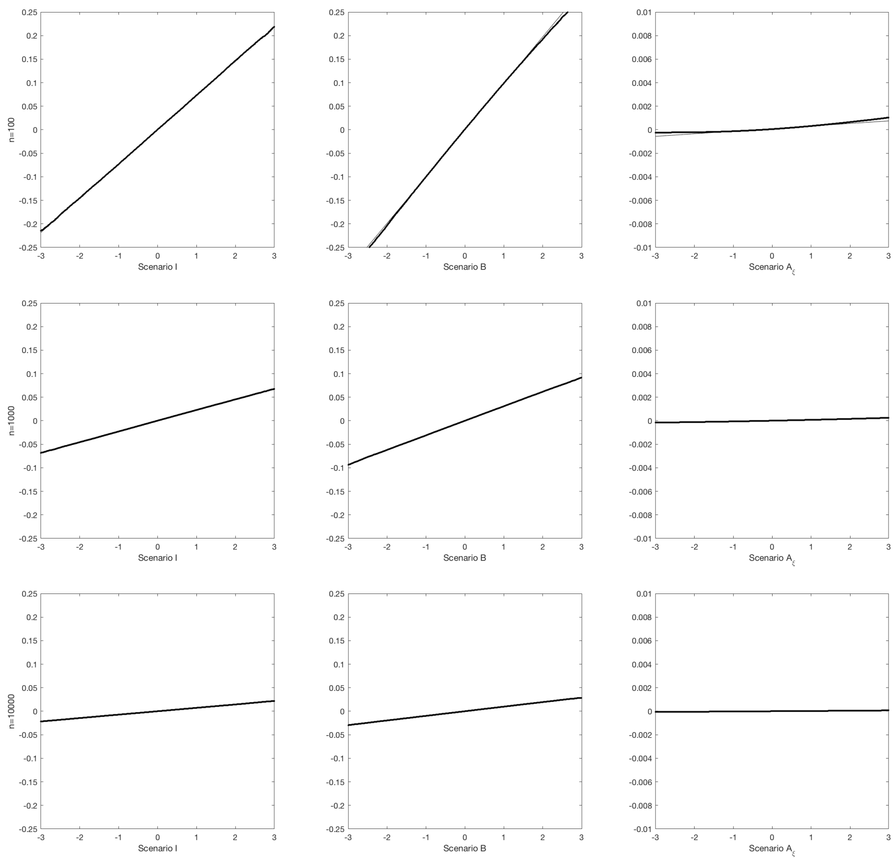


Figure 1: QQ plot versus Gaussian theoretical quantiles for $10^5$ estimates of $(\rho - \hat{\rho})$ coming from $n = 10^2$ (first row), $n = 10^3$ (second row), $n = 10^4$ (third row) simulations in Scenario $I$ (first column); $B$ (second); $A_\xi$ (third) as defined in Section 3.1. Notice the different scale of scenario $A_\xi$.

### 3.2 Rank correlation

Similarly to Pearson's correlation, and when the random vectors $X$ and $Y$ are not of the same type, the maximal attainable correlation $\rho_{\max}$ might be less than 1, and the minimal attainable correlation $\rho_{\min}$ might be bigger than $-1$. This for instance occurs when one vector is measured on a different scale, as in the following example.

**Example 3.1.** Assume $X \sim \mathcal{N}(\mu_X, \Sigma_X)$ and let the components of $Y = (Y_1, Y_2)'$ be independent and have a LogNormal distribution, that is $\ln Y_i \sim \mathcal{N}(0, \sigma^2)$, $i = 1, 2$, with $\sigma > 0$. Note that the function $f_+ : \mathbb{R}^2 \to \mathbb{R}$, $f_+(x_1, x_2) = \frac{1}{\sigma}(e^{\sigma x_1} + e^{\sigma x_2})$ is differentiable and convex, and has gradient $\nabla f_+(x_1, x_2) = (e^{\sigma x_1}, e^{\sigma x_2})'$. Since $\nabla f_+(X) \sim Y$, the maximal average inner product between $X$ and $Y$ is attained when $Y \overset{\text{a.s.}}{=} \nabla f_+(X)$ and is given by

$$\sup\{\operatorname{tr}(\Sigma_{\tilde{X}\tilde{Y}})\,;\tilde{X} \sim X, \tilde{Y} \sim Y\} = \mathbb{E}[X_1(e^{\sigma X_1} - e^{\sigma^2/2}) + X_2(e^{\sigma X_2} - e^{\sigma^2/2})] = 2\mathbb{E}[X_1(e^{\sigma X_1})].$$

Since

$$\Sigma_X = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \Sigma_Y = \begin{pmatrix} e^{\sigma^2}(e^{\sigma^2} - 1) & 0 \\ 0 & e^{\sigma^2}(e^{\sigma^2} - 1) \end{pmatrix},$$

one directly computes

$$\operatorname{tr}\left((\Sigma_X \Sigma_Y)^{1/2}\right) = \operatorname{tr}\left(\Sigma_Y^{1/2}\right) = 2\sqrt{e^{\sigma^2}(e^{\sigma^2} - 1)}.$$

Thus

$$\rho_{\max} = \frac{\mathbb{E}[X_1(e^{\sigma X_1})]}{\sqrt{e^{\sigma^2}(e^{\sigma^2} - 1)}}.$$

Taking $f_-(x_1, x_2) = -\frac{1}{\sigma}(e^{-\sigma x_1} - e^{-\sigma x_2})$, one analogously computes $\rho_{\min} = -\rho_{\max}$. This latter equality holds in general when $X$ is symmetric.

In Figure 2, we illustrate the range $[\rho_{\min}, \rho_{\max}]$ of attainable correlations versus the value of $\sigma > 0$. One immediately sees that $\rho_{\max} < 1$ and $\rho_{\min} > -1$ because $Y$ is not of the same type of $X$. Moreover, $\rho_{\max}$ can be made arbitrary small (positive) for enough large values of $\sigma$, yet the pair of vectors attaining $\rho_{\max}$ is always c-comonotonic. In the limit when $\sigma \to 0$, $\rho_{\max}$ converges to 1 (thus $\rho_{\min}$ to $-1$) because in this limiting case $Y$ converge to a degenerate distribution (which can be made approximately of the same type of $X$).

In Figure 2, we also notice that the sample correlation estimator is not in $[\rho_{\min}, \rho_{\max}]$. This is not a design error of the measure but just estimation uncertainty since $\rho_{\min}$ and $\rho_{\max}$ are unknown given only the data (the estimator is nevertheless always bounded by $[-1, 1]$). It is known (see for instance Lai et al., 1999) that for the LogNormal distribution the sample correlation is affected by a large bias for smaller sample sizes, a bias that can be reduced only after a large number of observations – typically in the order of millions. This behavior is evident in Figure 2, where we show estimates for the sample correlation coefficient $\hat{\rho}(\hat{X}, \hat{Y})$ for simulations coming from the pair $(X, Y)$ in Example 3.1. An upward bias is evident for $n = 10^3$ simulations, while the figure becomes consistent with Theorem 3.1 only at $n = 10^6$.

Example 3.1 shows that, for general random vectors, the range of attainable correlations might be strictly included in the interval $[-1, 1]$ and a small value of correlation does not necessarily imply a weak dependence. This warning can be ignored when one works within families of probability distributions closed under affine transformations, e.g. elliptical distributions. In such cases, the maximal (minimal) attainable correlation is always $(-)1$; see Example 2.1.

In order to avoid such cases where the correlation coefficient is not able to detect c-comonotone vectors observed on different scales, or fails to provide accurate estimates with a reasonable number of observations, it might be senseful to measure *rank correlation*. Recall our notation that the $n \times d$ matrices $\hat{X}, \hat{Y}$ contain $n$ independent realizations of the random vectors $X, Y \in L_2(\mathbb{R}^d)$. We define the *rank correlation* coefficient

$$\hat{\rho}_r(\hat{X}, \hat{Y}) = \hat{\rho}(\hat{X}^r, \hat{Y}^r), \tag{3.2}$$

where $\hat{X}^r, \hat{Y}^r$ are the columnwise rank values of the two set of observations $\hat{X}, \hat{Y}$.

The rank correlation $\hat{\rho}_r(\hat{X}, \hat{Y})$ preserves all the asymptotic properties of the sample correlation coefficient $\hat{\rho}(\hat{X}, \hat{Y})$, but has the additional advantage of being invariant with respect to increasing transformations of the marginal components of the vectors $X$, $Y$. Hence, it is able to detect c-comonotonicity between vectors up to increasing variations of scale. In Example 3.1 the rank correlation coefficient (3.2) detects one correlation, i.e. c-comonotonicity of the pair $(X, Y)$, at any level of $\sigma$ and number of observations $n$; see Figure 2.

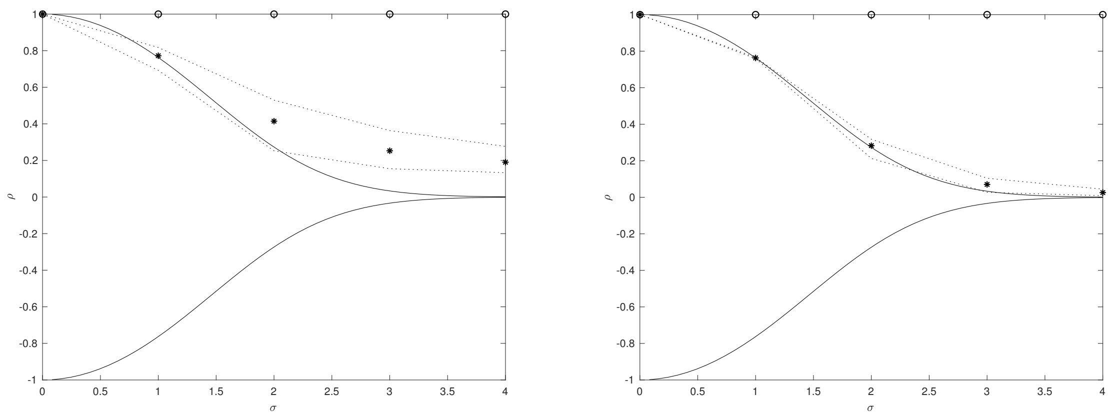


Figure 2: The solid lines represent maximum and minimum attainable correlations (versus the choice of the parameter $\sigma$) between the vectors $X$ and $Y$ as described in Example 3.1. The black stars are the values of the sample correlation $\hat{\rho}(\hat{X}, \hat{Y})$, while the white dots are the values of the rank correlation $\hat{\rho}_r(\hat{X}, \hat{Y})$ under maximal dependence. These estimates are averages computed with their 95% empirical confidence interval (dashed) over $10^4$ repetitions of a simulated sample of size $n = 10^3$ (left) and $n = 10^6$ (right).

We conclude this section by a second simulation study, where we show that, besides being invariant with respect to increasing transformations of the marginals, the rank correlation $\hat{\rho}_r(\hat{X}, \hat{Y})$ preserves all the asymptotic properties of the sample correlation coefficient $\hat{\rho}(\hat{X}, \hat{Y})$. We simulate from a pair $(X, Y)$ following the probabilistic model

$$Y = \sigma X + \sqrt{1 - \sigma^2} Z, \tag{3.3}$$

for a given value of $\sigma \in [-1, 1]$. $Z$ has a centered 5-dimensional Student's t distribution $Z \sim t_{4.5}(0, I_d)$, and is independent of $X \sim t_{4.5}(0, \Sigma_X)$, with

$$\Sigma_X = \begin{pmatrix} 1 & 0.32 & 0.39 & 0.44 & 0.47 \\ 0.32 & 1 & 0.32 & 0.39 & 0.44 \\ 0.39 & 0.32 & 1 & 0.31 & 0.39 \\ 0.44 & 0.39 & 0.31 & 1 & 0.32 \\ 0.47 & 0.44 & 0.39 & 0.32 & 1 \end{pmatrix}.$$

One easily computes

$$\rho(X, Y) = \frac{\sigma \operatorname{tr}(\Sigma_X)}{\operatorname{tr}\left((\sigma^2 \Sigma_X^2 + (1 - \sigma^2)\Sigma_X)^{1/2}\right)}.$$

Our results (collected in Table 2 and Figure 3) are coherent with the unbiasedness and asymptotic normality of $\hat{\rho}(\hat{X}, \hat{Y})$ and $\hat{\rho}_r(\hat{X}, \hat{Y})$.

### 3.3 Comparison with other association measures used in finance

We now apply our measure of correlation in a real financial example and we compare it to the multivariate measures defined in Grothe et al. (2014) and in Hofert et al. (2019).

|  | $\sigma = -0.5$ | CI | $\sigma = 0$ | CI | $\sigma = 0.9$ | CI |
|---|---|---|---|---|---|---|
| $\hat{\rho}$ | $-0.5163$ | $[-0.5557, -0.4782]$ | -0.0001 | $[-0.0303, 0.0299]$ | 0.9018 | $[0.8862, 0.9167]$ |
| $\hat{\rho}_r$ | $-0.5033$ | $[-0.5291, -0.4770]$ | -0.0001 | $[-0.0297, 0.0296]$ | 0.8734 | $[0.8618, 0.8844]$ |

Table 2: Values of $\hat{\rho}(\hat{X}, \hat{Y})$ and $\hat{\rho}_r(\hat{X}, \hat{Y})$ estimated over $10^4$ repetitions of a simulated sample of size $n = 10^3$ from the pair (X,Y) as defined in (3.3). Estimates are computed with their 95% empirical confidence interval (CI).

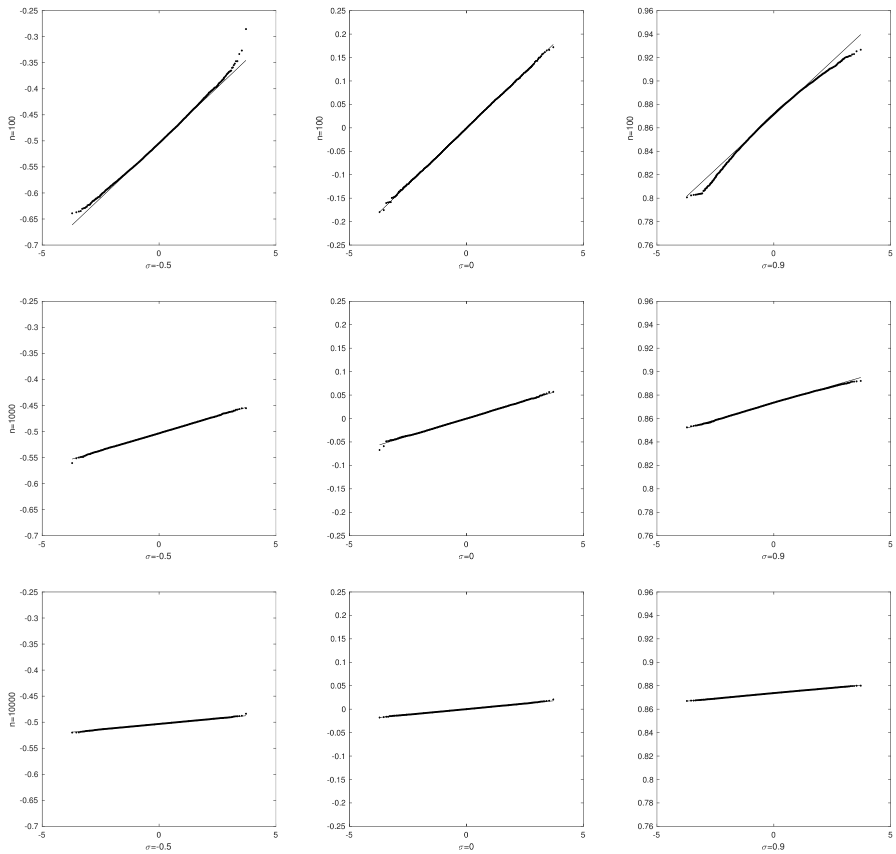


Figure 3: QQ plot versus Gaussian theoretical quantiles for $10^5$ estimates of $\hat{\rho}_r(\hat{X}, \hat{Y})$ coming from different sample sizes (in row) and for different values of $\sigma$ (in column) for the simulation study in Section 3.2. Notice the different scale used for $\sigma = 0.9$.

In Grothe et al. (2014), the authors develop multivariate generalizations of Spearman's rho and Kendall's tau, measuring association in terms of a normalized integral of the difference between the joint distribution of $(X, Y)$ and the joint distribution under independence. These measures are referred to as *multivariate Spearman* and *multivariate Kendall* in the following.

In Hofert et al. (2019), dependence between $X$ and $Y$ is measured in terms of

$$\rho(S(X), S(Y)), \tag{3.4}$$

that is the correlation between the so-called collapsed variables $S(X)$ and $S(Y)$, where the function $S$ can take different forms. These measures will be referred to as *collapsed measures of association*. The above mentioned papers operate within an axiomatic framework, where the desirable properties to be satisfied by a general multivariate measure are listed a priori. We further discuss this point in Section 4.

To make our comparison effective, we use the same dataset as in Grothe et al. (2014, Sect. 5.2) which was kindly provided by Oliver Grothe. Thus, 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; see also Ilmanen (2003).

We consider daily returns of the stock market indices $(X_1, \dots, X_5)'$ of five major countries as well as government bonds indices $(Y_1, \dots, Y_5)'$ for the respective countries. Full details of the study are to be found in Grothe et al. (2014). Figure 4, top, shows the evolution of multivariate Kendall and Spearman based on a 150-day forward-looking window with start date from Jan 3, 1996 to Jan 31, 2012, as compared to our correlation (2.1) and rank correlation (3.2) coefficients. Figure 4, middle, shows the same analysis for two different collapsed measures of association, defined by the arithmetic mean of the vector components and the maximum collapsing functions. Finally, Figure 4, bottom, shows Canonical Correlation (Hotelling, 1936), the RV coefficient (Robert and Escoufier, 1976), distance correlation (Székely et al., 2007) and the (Euclidean) distance collapsing function.

It is interesting to see how our newly defined correlation indexes, though derived from a totally different approach, behave very similarly to the multivariate Spearman, mean and maximum collapsing function. The comments in Grothe et al. (2014) hence remain valid, indicating that these association measures are more suitable to describe association with respect to classical ones. This is not surprising as these latter measures are standardized, clearly quantify positive and negative association and satisfy a list of properties as indicated by the respective authors. Also correlation and rank correlation are very similar, suggesting that the the use of simple correlation might be deemed appropriate.

As already noted in Hofert et al. (2019), the distance collapsing function is similar to the sample version of distance correlation and hence behaves similarly to Canonical Correlation and the RV coefficient.

Typically, association measures restricted to $[0, 1]$ are not able to distinguish positive from negative dependence. Consider for instance the two vectors $X = (X_1, X_2)'$ and $Y = (-X_1, -X_2)'$. It is immediate to compute $\rho(X, Y) = -1$ as the two vectors are c-countermonotonic. For Canonical Correlation

$$\chi(X, Y) = \sup_{u,v \in \mathbb{R}^d} \rho(u'X, v'Y),$$

one obtains $\chi(X, Y) = 1$ attained by $u = (1, 1)', v = (-1, -1)'$.

We also notice that in the case the function $S(X) = \sum_{i=1}^{d} X_i$ is simply the sum operator, the collapsed measure of association defined in (3.4) is computed as

$$\rho\left(\sum_{i=1}^{d} X_i, \sum_{j=1}^{d} Y_j\right) = \sum_{i=1}^{d}\sum_{j=1}^{d} \rho(X_i, Y_j),$$

that is as the sum of possible correlation between pairs of coordinates of $X$ and $Y$. Though in principle it is appealing to have a measure which is invariant with respect to permutations of the coordinates and that can be applied to vectors of different dimensions, there are cases in which our optimization approach appears to be more adequate than those existing in the literature. We provide a pedagogical example to see how and why the newly-proposed measure materially differs from the other available measures of vector dependence.

**Example 3.2.** Let $X$ have a Gaussian distribution $X \sim \mathcal{N}((0, 0)', \Sigma_X)$ and $Y \overset{\text{a.s.}}{=} AX$, where

$$\Sigma_X = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \; A = \begin{pmatrix} a & 1 \\ 1 & 1 \end{pmatrix}, \text{ with } a > 1.$$

Since $A$ is a strictly positive matrix, we have $\rho(X, Y) = \rho_{\max} = 1$, for any value of $a > 1$. Having $\rho = \rho_{\max}$ means that the interdependence between $X$ and $Y$ yields the minimal possible (squared Euclidean) distance between their distributions. This is coherent with the original meaning of correlation as a measure of similarity: given their marginal laws, $X$ and $Y$ are the closest possible random vectors.

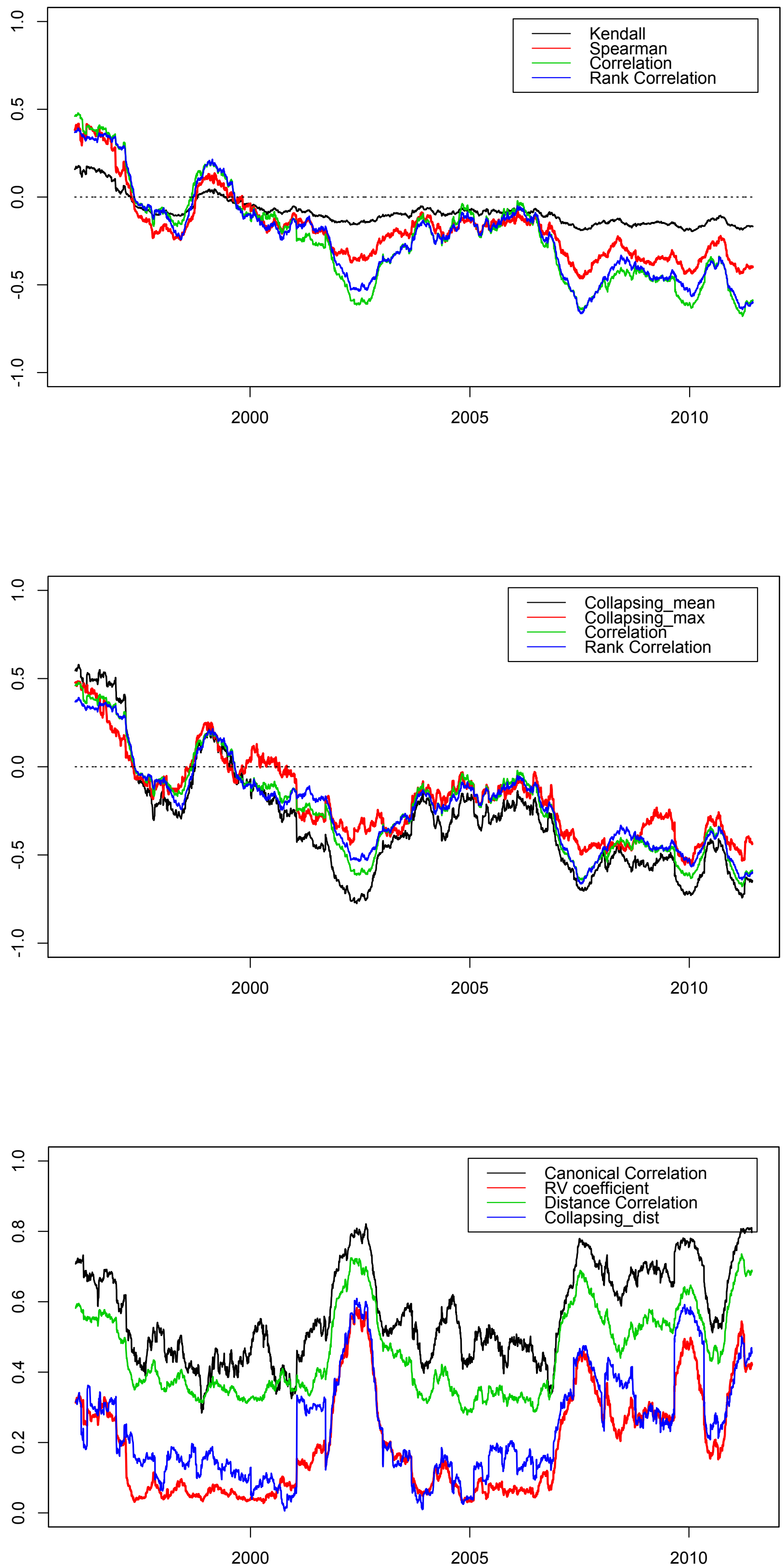


Figure 4: Time evolution of the association between bond and stock markets, measured by different multivariate measures of association as described in Section 3.3.

Other measures of association fail to detect this minimal distance case, not because they are bad measures but just because they are based on different approaches. Using the collapsed measure of association in Hofert et al. (2019) with the mean collapsing function, elementary calculations yield

$$\rho_{mean} = \rho\left(\frac{X_1+X_2}{2}, \frac{Y_1+Y_2}{2}\right) = \rho\left(\frac{X_1+X_2}{2}, \frac{(a+1)X_1+2X_2}{2}\right) = \frac{a+3}{\sqrt{2}\sqrt{a(a+2)+5}}.$$

In Figure 5, left, we plot the function $\rho_{mean}$ against the value of $a$. For $a > 1$, the function is always strictly less than 1 (and goes to the limit $\sqrt{2}/2$ for $a \to \infty$). In Figure 5, right, we estimated the multivariate Kendall as introduced in Grothe et al. (2014) from a set of 1000 independent simulations from $(X, Y)$, for varying values of $a$. Even at $a = 1$, the value of Kendall is $\sqrt{3}/3 < 1$, while our sample coefficient is always 1 for any value of $a > 1$.

We stress again that this is only a pedagogical example to highlight the differences between different measures of association and their different domains of application – does not assess the goodness of one with respect to the others. Despite the mathematical example, in the real financial case study illustrated above, all the mentioned measures behave similarly.

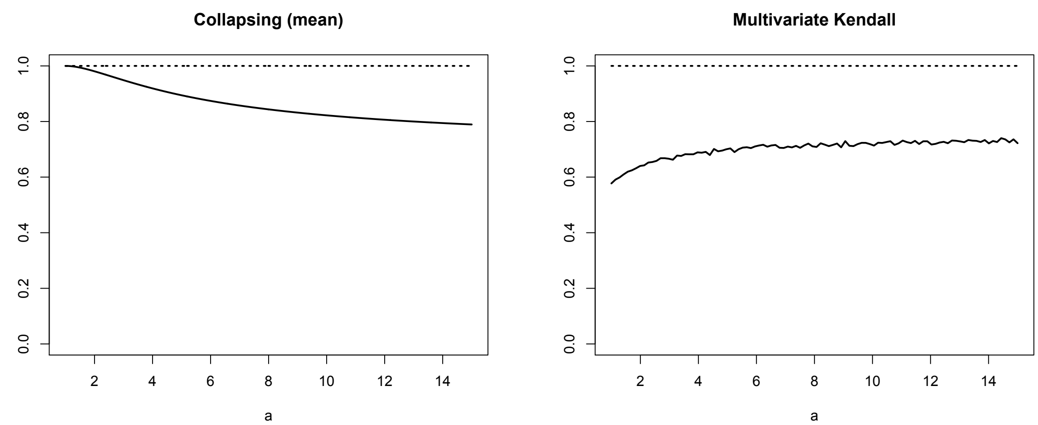


Figure 5: Values of the collapsed measure of association (with mean collapsing function) and multivariate Kendall against the value of the parameter $a$ in Example 3.2. The dashed line represent the value of (estimated) correlation, which is always 1.

# 4 Final discussion and further extensions

Fully inspired by the structure of Pearson's correlation, in this paper we introduce the coefficient

$$\rho(X, Y) = \frac{\operatorname{tr}(\Sigma_{XY})}{\operatorname{tr}\left((\Sigma_X \Sigma_Y)^{1/2}\right)}$$

to measure linear correlation between two non degenerate, square integrable $\mathbb{R}^d$-valued random vectors $X$ and $Y$ having invertible covariance matrices $\Sigma_X$ and $\Sigma_Y$, and cross-covariance matrix $\Sigma_{XY}$. The new measure preserves all the relevant properties of Pearson's in arbitrary dimension.

The rationale behind our approach is to measure the covariance between two random vectors $X$ and $Y$ in terms of $\operatorname{tr}(\Sigma_{XY})$ and then standardize it by its maximum attainable value given solely the knowledge of the covariance matrices of $X$ and $Y$. The novel coefficient is thus maximized when the average inner product between $X$ and $Y$ is maximized (over all the pairs having the same marginal distributions) or, equivalently, when their average squared Euclidean distance is minimized. This optimization approach relies on the notion of c-co(unter)monotonicity and finds its roots in the theory of mass transportation. The new measure is also consistent with the original idea of correlation as a measure of similarity intended as minimal distance.

We propose an empirical estimator, establish its asymptotic distribution and illustrate the concept in simulations as well as in a financial application. Since (same as Pearson's correlation) the original estimator may be biased depending on the actual distribution of the multivariate margins, we also propose the use of ranks and derive the respective results for the rank correlation measure.

While we learn in introductory courses that the normalization of Pearson's correlation coefficient is due to the Cauchy-Schwarz inequality, in this paper we propose a sharper bound in the multivariate case, which is also interesting per se (for positive definite covariance matrices).

There is not a universally acknowledged way of measuring association between random vectors but rather many measures motivated by different approaches and scopes of application. Each measure has its merit and disadvantages, and the one proposed here makes no exception. In the following concluding remarks, we discuss further extensions of the mathematical framework given in this paper and its relationship with the current literature.

**Non-invertible covariance matrices**

We give some insight into the case in which one or both covariance matrices $\Sigma_X$ and $\Sigma_Y$ are not invertible. As long as $\Sigma_X$ is invertible, if follows from the results in Dowson and Landau (1982) and Olkin and Pukelsheim (1982) that there still exist a c-comonotonic pair $(\tilde{X}, \tilde{Y})$ with the same marginal covariances and such that $\mathrm{tr}(\Sigma_{\tilde{X},\tilde{Y}}) = \mathrm{tr}\left((\Sigma_X\Sigma_Y)^{1/2}\right)$. The existence of such pair does not rely on the invertibility of $\Sigma_Y$ while its uniqueness holds under an extra continuity assumption for $X$ (this follows easily using for instance Corollary 14 in McCann (1995)). Thus, when $\Sigma_X$ is invertible but $\mathrm{rank}(\Sigma_Y) < \mathrm{rank}(\Sigma_X)$, it makes sense to postulate $\rho(Y, X) = \rho(X, Y)$.

As an example, consider elliptical distributions $F_X$, $F_Y$ of Gaussian type with null mean vectors and covariance matrices

$$\Sigma_X = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \Sigma_Y = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Without loss of generality we can represent $X \overset{\text{a.s.}}{=} (X_1, X_1)'$, where $X_1 \sim \Phi$ has a standard univariate normal distribution, while $Y \overset{\text{a.s.}}{=} (Y_1, Y_2)'$ with $Y_1, Y_2 \sim \Phi$ and $Y_1$ independent of $Y_2$. $\Sigma_X$ is not invertible. By Theorem 2.1 in Puccetti and Wang (2015), the maximum

$$M(F_X, F_Y) = \sup\{\mathrm{tr}(\Sigma_{XY}); X \sim F_X, Y \sim F_Y\} = \sup\{\mathbb{E}[X_1(Y_1 + Y_2)] : X_1 \sim \Phi, Y \sim F_Y\}$$

is attained when the random variables $X_1$ and $(Y_1 + Y_2)$ are comonotonic, that is when

$$Y_1 + Y_2 \overset{\text{a.s.}}{=} \sqrt{2}\, X_1. \tag{4.1}$$

Take $Z \sim \Phi$ independent of $X$. Clearly, the vector $\hat{Y}$ defined as

$$\hat{Y} \overset{\text{a.s.}}{=} \begin{pmatrix} \frac{\sqrt{2}}{2}X_1 + \frac{\sqrt{2}}{2}Z \\ \frac{\sqrt{2}}{2}X_1 - \frac{\sqrt{2}}{2}Z \end{pmatrix}$$

has distribution $F_Y$ and satisfies (4.1). It follows that

$$\mathrm{tr}(\Sigma_{X\hat{Y}}) = \mathrm{tr}\left((\Sigma_X\Sigma_Y)^{1/2}\right) = \mathrm{tr}\left(\Sigma_X^{1/2}\right) = \mathrm{tr}\begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} = \sqrt{2},$$

which gives $\rho(X, \hat{Y}) = 1$ coherently with the fact that $(X, \hat{Y})$ is c-comonotonic.

One finds the same result, and more easily, by swapping $\Sigma_X$ and $\Sigma_Y$, thus assuming

$$\Sigma_X = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \ \Sigma_Y = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Similarly to Example 2.1, the vector $(X, \hat{Y})$ with $\hat{Y} \overset{\text{a.s.}}{=} \Sigma_X^{-1} \# \Sigma_Y X = \Sigma_Y^{1/2} X$ is the only c-comonotonic pair yielding

$$\hat{Y}_1 \overset{\text{a.s.}}{=} \hat{Y}_2 \overset{\text{a.s.}}{=} \frac{\sqrt{2}}{2}(X_1 + X_2),$$

and then, again,

$$\text{tr}(\Sigma_{X\hat{Y}}) = \text{tr}\left((\Sigma_X \Sigma_Y)^{1/2}\right) = \sqrt{2}.$$

If $\Sigma_X$ is not invertible and $\text{rank}(\Sigma_X) \geq \text{rank}(\Sigma_Y)$, unfortunately the rationale behind definition (2.1) does not make sense since there might exist infinitely many c-comonotonic pairs. As an example, consider $F_X$ and $F_Y$ of Gaussian type with null mean vectors and

$$\Sigma_X = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \; \Sigma_Y = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix},$$

the support of $F_X$ is the set $\Gamma_x = \{(x_1, x_2)' \in \mathbb{R}^2; x_1 = x_2\}$ while the support of $F_Y$ is the set $\Gamma_y = \{(x_1, x_2)' \in \mathbb{R}^2; x_1 = -x_2\}$. It is immediate to see that (1.3) always holds with equality for any set of points in $\Gamma_x \times \Gamma_y$ and therefore any random pair with marginals $F_X$ and $F_Y$ is c-comonotonic *and* c-countermonotonic at the same time.

**Permutation of vector components and vectors with different dimensions**

As already remarked, we introduced the coefficient (2.1) with the aim of preserving the properties of Pearson's correlation in a multivariate setting. It is immediate to show that our correlation satisfies axioms 1., 4., 5. in Grothe et al. (2014), and it sufficient to compute the corresponding rank coefficient (3.2) to have also invariance with respect to increasing transformations (axiom 3.).

Thus, the only exception with respect to the framework of Grothe et al. (2014) is that our proposed correlation can be computed only for vectors of the same dimension and it is not invariant on permutations of the coordinates within one vector. This is a natural implication of our approach based on the minimization of the squared Euclidean distance, which implicitly requires that the $i$-th coordinates of $X$ and $Y$ have a comparable meaning/interpretation as in the case study outlined in Section 3.3.

For instance, we believe that our new correlation coefficient finds a natural application as a measure of similarity for the clustering of three-way data structures. So called *three-way* data structures occur in different application domains from the simultaneous observation of the same attributes on a set of statistical units in different situations or locations. These include data coming from multivariate longitudinal data or spatio-temporal data, multivariate time series or data collecting multivariate repeated measurements. Different solutions for clustering three-way data have been proposed in the statistical literature (see Viroli, 2011, for a review), often implying some dimension reduction techniques (e.g. principal component analysis) to convert the three-way dataset to a two-way dataset, and hence apply conventional clustering techniques. When one has a three-way structured dataset and want to take into account the covariances at the level of attributes contained in the data, it could be useful to define a dissimilarity measure between two statistical units based on (2.1) and hence apply a suitable clustering technique.

Generally, multivariate observations are arranged in a three-way dataset where for each of the $r$ statistical units a set of $n$ simultaneous observations of the same $p$ variables are available. A dissimilarity measure between two sets of observations $\hat{X}$ and $\hat{Y}$ coming from two statistical units $X$ and $Y$ can then be defined in a classical way by

$$d(X, Y) = \frac{1 - \hat{\rho}(\hat{X}, \hat{Y})}{2}, \tag{4.2}$$

and at this point one can use for instance the preferred hierarchical clustering in the agglomerative approach, as described in Everitt et al. (2011).

In principle, swapping the coordinates of one vector might highly affect correlation. For instance, consider the vector $X = (X_1, X_2)'$, with $X_1$ independent of $X_2$, and swap its components to obtain $Y = (X_2, X_1)'$. One has

$\rho(X,X)=1$, while $\rho(X,Y)=0$. Swapping the components might even switch the sign of a 1-correlation, as the following example shows.

**Example 4.1.** Consider the two vectors

$$X = \begin{pmatrix} X_1 \\ -X_1 + \sqrt{\xi} Z \end{pmatrix}, \quad Y = \begin{pmatrix} X_1 \\ -X_1 + \sqrt{\xi} Z \end{pmatrix},$$

where $X_1$ and $Z$ are two i.i.d. standard normal and $0 < \xi < 1$. Since $Y \overset{\text{a.s.}}{=} X$, we have $\rho(X,Y)=1$.

By swapping the components of $X$ and considering instead

$$X' = \begin{pmatrix} -X_1 + \sqrt{\xi} Z \\ X_1 \end{pmatrix}, \quad Y = \begin{pmatrix} X_1 \\ -X_1 + \sqrt{\xi} Z \end{pmatrix},$$

one computes

$$\Sigma_{X'} = \begin{pmatrix} 1+\xi & -1 \\ -1 & 1 \end{pmatrix}, \quad \Sigma_Y = \begin{pmatrix} 1 & -1 \\ -1 & 1+\xi \end{pmatrix}, \quad \text{and} \quad \Sigma_{X'Y} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Using the formula for the square root of a $2 \times 2$ matrix one obtains

$$\begin{aligned} \rho(X',Y) = \frac{\operatorname{tr}(\Sigma_{X'Y})}{\operatorname{tr}\left((\Sigma_{X'}\Sigma_Y)^{1/2}\right)} &= \frac{-2}{\operatorname{tr}\left(\begin{pmatrix} 2+\xi & -2-2\xi \\ -2 & 2+\xi \end{pmatrix}^{1/2}\right)} \\ &= \frac{-2}{\operatorname{tr}\left(\begin{pmatrix} (1+\xi)^{1/2} & -(1+\xi)^{1/2} \\ -(1+\xi)^{-1/2} & (1+\xi)^{1/2} \end{pmatrix}\right)} = \frac{-2}{2(1+\xi)^{1/2}} = -(1+\xi)^{-1/2}, \end{aligned}$$

which can be made arbitrarily close to $-1$. Similarly to Remark 2.2, we point out that there exists four matrices $B$ with the property that $B^2 = \Sigma_{X'}\Sigma_Y = \begin{pmatrix} 2+\xi & -2-2\xi \\ -2 & 2+\xi \end{pmatrix}$, but the one used above is the only having positive real eigenvalues, hence it is the principal square root.

Extreme behaviors like the one just illustrated, however, might be unrealistic and restricted to theoretical (counter)examples. In Figure 6 we plot the time evolution of correlation between bond and stock market, analogously to what done in Figure 4, but permuting the coordinates of the stock market indices in three different ways. All the resulting correlations are visually overlapping.

A possible alternative to make the measure invariant versus permutations could be to transform $X$ and $Y$ by a PCA with descending principal components before applying the measure. In Figure 7 we plot the time evolution of correlation between bond and stock market using this method with all components and with only the principal component. The overall process (PCA + measure) is invariant with respect to permutations of variables within the original data, but poses a sign issue: in the figure the sign of the obtained correlation has been adjusted to the sign of the original one.

This technique could also be used to make correlation computable also for vectors of different dimensions, by choosing for each of them only $k$ principal components. However, this would affect the original interpretation and properties of the measure as illustrated in this paper.

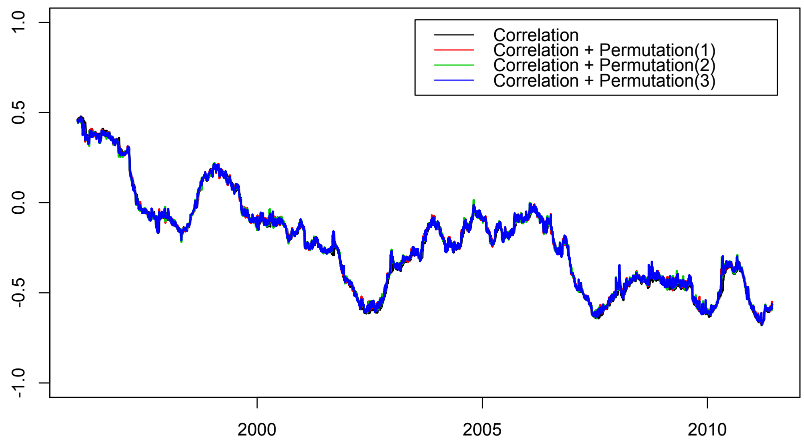


Figure 6: Time evolution of the newly defined correlation between bond and stock markets, when: (*top figure*) the coordinates $X_i$ of the stock market vector are kept the same (black line) or permuted (red line: $1 \to 2$,$2 \to 3$,$3 \to 4$,$4 \to 5$,$5 \to 1$; green line: $1 \to 3$,$2 \to 4$,$3 \to 5$,$4 \to 1$,$5 \to 2$; blue line: $1 \to 3$,$2 \to 5$,$3 \to 4$,$4 \to 2$,$5 \to 1$;); (*bottom figure*).

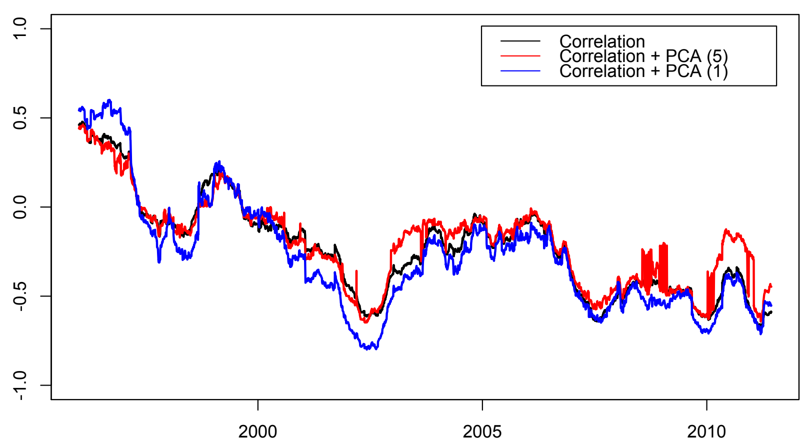


Figure 7: Time evolution of the newly defined correlation between bond and stock markets, when a PCA is first applied to the components of the two vectors, considering all components (red line) and only the principal component (blue line).

### The range of attainable correlations

Example 3.1 shows that, for general random vectors, the range of attainable correlations might be strictly included in the interval $[-1, 1]$ and, consequently, a small value of correlation does not necessarily imply a weak dependence. For the value of $\rho(X, Y)$ to be fully meaningful, one should know or at least should be able to estimate the range $[\rho_{\min}, \rho_{\max}]$, with

$$\rho_{\min} = \inf\{\rho(X, Y); X \sim F, Y \sim G\}, \quad \rho_{\max} = \sup\{\rho(X, Y); X \sim F, Y \sim G\}.$$

For distributions with null mean, computing $\rho_{\max}$ is equivalent to computing (2.3b), that is solving the mass transportation problem

$$M(F, G) = \sup \{\mathrm{tr}(\Sigma_{XY}); X \sim F, Y \sim G\}.$$

For distributions on the line, i.e. for $d = 1$, the value of $M(F, G)$ and therefore $\rho_{\max}$ can be easily computed since the maximum correlation for two random variables is attained if and only if they are comonotonic.

On a multivariate space ($d > 1$), an analytical computation of $M(F, G)$ and $\rho_{\max}$ is restricted to a few cases, including the case of Elliptical distributions described below. As illustrated in Puccetti (2017), the computation of $M(F, G)$ is equivalent to the computation of the so-called $L^2$-*Wasserstein distance* between distributions. A numerical estimate of $\rho_{\max}$ can therefore be obtained by using the broad variety of methods described for instance in Puccetti (2017), Peyré and Cuturi (2019) and references therein.

Using rank correlation as in Section 3.2 allows one to avoid the cases where a change of scale reduces the range of attainable correlations. In fact, it is easy to show that the two vectors are $\pi$-comonotonic, one has $\rho_{\max} = 1$;. A pair of random vectors $(X, Y)$ is said to be $\pi$-comonotonic if the marginal components $X_i$ and $Y_i$ are both increasing function of a common random factor $Z_i$, for each $1 \leq i \leq d$. When $d = 1$, this definition is exactly that of comonotonicity between random variables. When $d > 1$, a c-comonotonic random vector is also $\pi$-comonotonic, but not viceversa; see Puccetti and Scarsini (2010) for more details on $\pi$-comonotonicity.

However, even in the case of rank correlation, there still exist cases in which the maximal attainable (rank) correlation is smaller than one, and a numerical estimate of $\rho_{\max}$ is called for. Analogous considerations hold for $\rho_{\min}$ (by replacing comonotonicity by countermonotonicity notions).

These warning can be ignored when one works within families of probability distributions closed under affine transformations, e.g. elliptical distributions. In such cases, the maximal (minimal) attainable correlation is always $(-)1$. For example, consider $X \sim \mathcal{E}(\mu_X, \Sigma_X, \psi)$ having an elliptical distribution with generator $\psi$ and $Y \sim \mathcal{E}(\mu_Y, \Sigma_Y, \psi)$. Then, it is sufficient to put $\hat{Y} \overset{\text{a.s.}}{=} A(X - \mu_X) + \mu_Y$, with $A = \Sigma_X^{-1} \# \Sigma_Y$ to obtain $\hat{Y} \sim Y$ and $\rho(X, \hat{Y}) = 1$. The linear function $A$ is the only one such that

$$[A(X - \mu_X) + \mu_Y] \sim Y,$$

and

$$\mathbb{E} ||\, [A(X - \mu_X) + \mu_Y] - Y \,||^2 \leq \mathbb{E} ||\, T(X) - Y \,||^2, \tag{4.3}$$

for all possible measurable transformations $T(X) \sim Y$. Switching the sign of $A$, one similarly obtains $\rho(X, \hat{Y}) = -1$.

As our bottom line, we would like to state the words of McNeil et al. (2015): *Correlation plays a central role in financial theory, but it is important to realize that the concept is only really a natural one in the context of multivariate normal or, more generally, elliptical models.*

### Open problems

It would be of interest to compute (at least numerically) the asymptotic variance $\sigma^2$ in Theorem 3.1, thus avoiding bootstrap procedures to make inference. Unfortunately, the presence of the square root functional in (2.1) seems to make the application of the delta method (as illustrated in Kollo and von Rosen (2005) and performed for instance in Robert et al. (1985); Grothe et al. (2014); Hofert et al. (2019)) particularly challenging. Also, finding the exact distribution of (3.1) for finite sample would be valuable. We hope that these open problems will stimulate further research on the topic.

**Acknowledgments**

The author would like to thank two anonymous reviewers for their suggestions on an earlier version of this paper, and Oliver Grothe who kindly provided the code and data to perform the comparative study included in Section 3.3.

# References

Betken, A., H. Dehling, I. Nüß gen, and A. Schnurr (2021). Ordinal pattern dependence as a multivariate dependence measure. *J. Multivariate Anal. 186*, Paper No. 104798, 19.

Bhatia, R. (2007). *Positive Definite Matrices.* Princeton University Press.

Böttcher, B., M. Keller-Ressel, and R. L. Schilling (2019). Distance multivariance: new dependence measures for random vectors. *Ann. Statist. 47*(5), 2757–2789.

Carlier, G., V. Chernozhukov, and A. Galichon (2016). Vector quantile regression: an optimal transport approach. *Ann. Statist. 44*(3), 1165–1192.

Dhaene, J., M. Denuit, M. J. Goovaerts, R. Kaas, and D. Vyncke (2002). The concept of comonotonicity in actuarial science and finance: theory. *Insurance Math. Econom. 31*(1), 3–33.

Dowson, D. C. and B. V. Landau (1982). The Fréchet distance between multivariate normal distributions. *J. Multivariate Anal. 12*(3), 450–455.

Durante, F. and C. Sempi (2016). *Principles of Copula Theory.* CRC Press, Boca Raton, FL.

Ekeland, I., A. Galichon, and M. Henry (2012). Comonotonic measures of multivariate risks. *Math. Finance 22*(1), 109–132.

Everitt, B. S., S. Landau, M. Leese, and D. Stahl (2011). *Cluster Analysis* (Fifth ed.). John Wiley & Sons, Chichester.

Fréchet, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon. Sect. A. (3) 14*, 53–77.

Gangbo, W. and R. J. McCann (1996). The geometry of optimal transportation. *Acta Math. 177*(2), 113–161.

Grothe, O., J. Schnieders, and J. Segers (2014). Measuring association and dependence between random vectors. *J. Multivariate Anal. 123*, 96–110.

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

Han, F. (2021). On extensions of rank correlation coefficients to multivariate spaces. *Bernoulli News 28*, 7–11.

Higham, N. J. (1997). Stable iterations for the matrix square root. *Numer. Algorithms 15*(2), 227–242.

Hoeffding, W. (1940). Massstabinvariante korrelationtheorie. *Schriften Math. Inst. Univ. Berlin 5*(5), 181–233.

Hoeffding, W. (1994). Scale–invariant correlation theory. In N. Fisher and P. Sen (Eds.), *The Collected Works of Wassily Hoeffding*, pp. 57–107. Springer New York. This is the english translation of Hoeffding (1940).

Hofert, M., W. Oldford, A. Prasad, and M. Zhu (2019). A framework for measuring association of random vectors via collapsed random variables. *J. Multivariate Anal. 172*, 5–27.

Horn, R. A. and C. R. Johnson (2013). *Matrix Analysis* (Second ed.). Cambridge University Press.

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika 28*(3-4), 321.

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

Jin, Z. and D. S. Matteson (2018). Generalizing distance covariance to measure and test multivariate mutual dependence via complete and incomplete V-statistics. *J. Multivariate Anal. 168*, 304–322.

Josse, J. and S. Holmes (2016). Measuring multivariate association and beyond. *Stat. Surv. 10*, 132–167.

Kollo, T. and D. von Rosen (2005). *Advanced Multivariate Statistics with Matrices.* Springer, Dordrecht.

Lai, C., J. Rayner, and T. P. Hutchinson (1999). Robustness of the sample correlation – the bivariate lognormal case. *J. Appl. Math. Decis. Sci. 3*(1), 7–19.

McCann, R. J. (1995). Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J. 80*(2), 309–323.

McNeil, A. J., R. Frey, and P. Embrechts (2015). *Quantitative Risk Management. Concepts, Techniques and Tools. Revised edition.* Princeton University Press.

Medovikov, I. and A. Prokhorov (2017). A new measure of vector dependence, with applications to financial risk and contagion. *J. Financ. Econom. 15*(3), 474–503.

Mordant, G. and J. Segers (2022). Measuring dependence between random vectors via optimal transport. *J. Multivariate Anal. 189*, Paper No. 104912, 22.

Nelsen, R. B. (2006). *An Introduction to Copulas* (Second ed.). New York: Springer.

Olkin, I. and F. Pukelsheim (1982). The distance between two random vectors with given dispersion matrices. *Linear Algebra Appl. 48*, 257–263.

Peyré, G. and M. Cuturi (2019). Computational optimal transport. *Foundations and Tends in Machine Learning 11*(5–6), 355–607.

Puccetti, G. (2017). An algorithm to approximate the optimal expected inner product of two vectors with given marginals. *J. Math. Anal. Appl. 451*, 132–145.

Puccetti, G. and M. Scarsini (2010). Multivariate comonotonicity. *J. Multivariate Anal. 101*(1), 291–304.

Puccetti, G. and R. Wang (2015). Extremal dependence concepts. *Stat. Science 30*(4), 485–517.

Rachev, S. T. and L. Rüschendorf (1998). *Mass Transportation Problems. Vol. I-II.* New York: Springer-Verlag.

Robert, P., R. Cléroux, and N. Ranger (1985). Some results on vector correlation. *Comput. Statist. Data Anal. 3*(1), 25–32.

Robert, P. and Y. Escoufier (1976). A unifying tool for linear multivariate statistical methods: the *RV*-coefficient. *J. Roy. Statist. Soc. Ser. C Appl. Statist. 25*(3), 257–265.

Rockafellar, R. T. (1970). On the maximal monotonicity of subdifferential mappings. *Pacific J. Math. 33*(1), 209–216.

Rüschendorf, L. (1991). Fréchet-bounds and their applications. In *Advances in Probability Distributions with Given Marginals*, pp. 151–187. Dordrecht: Kluwer Acad. Publ.

Rüschendorf, L. (1996). On $c$-optimal random variables. *Statist. Probab. Lett. 27*(3), 267–270.

Rüschendorf, L. (2012). Worst case portfolio vectors and diversification effects. *Finance Stoch. 16*(1), 155–175.

Rüschendorf, L. (2013). *Mathematical Risk Analysis. Dependence, Risk Bounds, Optimal Allocations and Portfolios*. Heidelberg: Springer.

Rüschendorf, L. and S. T. Rachev (1990). A characterization of random variables with minimum $L^2$-distance. *J. Multivariate Anal. 32*(1), 48–54.

Schmid, F., R. Schmidt, T. Blumentritt, S. Gaiß er, and M. Ruppert (2010). Copula-based measures of multivariate association. In *Copula theory and its applications*, pp. 209–236. Springer, Heidelberg.

Schnurr, A. (2014). An ordinal pattern approach to detect and to model leverage effects and dependence structures between financial time series. *Statist. Papers 55*(4), 919–931.

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

Viroli, C. (2011). Model based clustering for three-way data structures. *Bayesian Anal. 6*(4), 573–602.

Zhu, L., K. Xu, R. Li, and W. Zhong (2017). Projection correlation between two random vectors. *Biometrika 104*(4), 829–843.