AppliedMath

Review

Four Measures of Association and Their Representations in Terms of Copulas

Michel Adès [1], Serge B. Provost [2,*] and Yishan Zang [2]

[1] Département de Mathématiques, Université du Québec à Montréal, Montréal, QC H2X 3Y7, Canada; ades.michel@uqam.ca
[2] Department of Statistical and Actuarial Sciences, The University of Western Ontario, London, ON N6A 5B7, Canada
* Correspondence: provost@stats.uwo.ca

Abstract: Four measures of association, namely, Spearman's ρ, Kendall's τ, Blomqvist's β and Hoeffding's Φ^2, are expressed in terms of copulas. Conveniently, this article also includes explicit expressions for their empirical counterparts. Moreover, copula representations of the four coefficients are provided for the multivariate case, and several specific applications are pointed out. Additionally, a numerical study is presented with a view to illustrating the types of relationships that each of the measures of association can detect.

Keywords: Blomqvist's correlation coefficient; copulas; Hoeffding's dependence index; Kendall's rank correlation coefficient; multivariate measures of dependence; Spearman's rank correlation

1. Introduction

Copula representations and sample estimates of the correlation measures attributed to Spearman, Kendall, Blomqvist and Hoeffding are provided in this paper. All these measures of association depend on the ranks of the observations on each variable. They can reveal the strength of the dependence between two variables that are not necessarily linearly related, as is required in the case of Pearson's correlation. They can as well be applied to ordinal data. While the Spearman, Kendall and Blomqvist measures of association are suitable for observations exhibiting monotonic relationships, Hoeffding's index can also ascertain the extent of the dependence between the variables, regardless of the patterns that they may follow. Thus, these four measures of association prove quite versatile when it comes to assessing the strength of various types of relationships between variables. Moreover, since they are rank-based, they are all robust with respect to outliers. What is more, they can be readily evaluated.

Copulas are principally utilized for modeling dependency features in multivariate distributions. They enable one to represent the joint distribution of two or more random variables in terms of their marginal distributions and a specific correlation structure. Thus, the effect of the dependence between the variables can be separated from the contribution of each marginal. As measures of dependence, copulas have found applications in numerous fields of scientific investigations, including reliability theory, signal processing, geodesy, hydrology, finance and medicine. We now review certain basic definitions and results on the subject.

In the bivariate framework, a copula function is a distribution whose support is the unit square $\mathbb{1}^2 = [0,1]^2$ and whose marginals are uniformly distributed. A more formal definition is now provided.

A function $C : \mathbb{1}^2 \mapsto \mathbb{1}$ is a bivariate copula if it satisfies the two following properties:

1. For every $u,\, v \in \mathbb{1}$,

$$C(u,\,1) = u,\;\; C(1,\,v) = v\,,\text{ and }\; C(u,\,0) = C(0,\,v) = 0.$$

check for updates

Citation: Adès, M.; Provost, S.B.; Zang, Y. Four Measures of Association and Their Representations in Terms of Copulas. *AppliedMath* **2024**, *4*, 363–382. https://doi.org/10.3390/appliedmath4010019

Academic Editors: Dmitriy V. Ivanov and Tommi Sottinen

Received: 16 January 2024
Revised: 9 February 2024
Accepted: 22 February 2024
Published: 2 March 2024

2. For every $u_1, u_2, v_1, v_2 \in \mathbb{1}$ such that $u_1 \le u_2$ and $v_1 \le v_2$,

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \ge 0.$$

This last inequality implies that $C(u,v)$ is increasing in both variables.

We now state a result due to Sklar (Theorem 1) [1].

Theorem 1. *Let $H(x, y)$ be the joint cumulative distribution function of the random variables X and Y whose continuous marginal distribution functions are denoted by $F(x)$ and $G(y)$. Then, there exists a unique bivariate copula $C : \mathbb{1}^2 \mapsto \mathbb{1}$ such that*

$$H(x, y) = C(F(x), G(y)) \tag{1}$$

where $C(\cdot, \cdot)$ is a joint cumulative distribution function having uniform marginals. Conversely, for any continuous cumulative distribution functions $F(x)$ and $G(y)$ and any copula $C(\cdot, \cdot)$, the function $H(\cdot, \cdot)$, as defined in (1), is a joint distribution function with marginal distribution functions $F(\cdot)$ and $G(\cdot)$.

Sklar's theorem provides a technique for constructing copulas. Indeed, the function

$$C(u, v) = H\Big(F^{-1}(u), G^{-1}(v)\Big) \tag{2}$$

is a bivariate copula, where the quasi-inverses $F^{-1}(\cdot)$ and $G^{-1}(\cdot)$ are defined by

$$F^{-1}(u) = \inf\{x|F(x) \ge u\}, \quad u \in (0, 1), \tag{3}$$

and

$$G^{-1}(v) = \inf\{y|G(y) \ge v\}, \quad v \in (0, 1). \tag{4}$$

Copulas are invariant with respect to strictly increasing transformations. More specifically, assuming that X and Y are two continuous random variables whose associated copula is $C(\cdot\,,\cdot)$, and letting $\alpha(\cdot)$ and $\beta(\cdot)$ be two strictly increasing functions and $C_{\alpha,\beta}(\cdot\,,\cdot)$ be the copula obtained from $\alpha(X)$ and $\beta(Y)$, then for all $(u,v) \in \mathbb{1}^2$, one has

$$C_{\alpha,\beta}(u,v) = C(u,v). \tag{5}$$

We shall denote the probability density function corresponding to the copula $C(u_1,u_2)$ by

$$c(u, v) = \frac{\partial^2}{\partial u \partial v} C(u,v). \tag{6}$$

The following relationship between $h(\cdot,\cdot)$, the joint density function of the random variables X and Y as defined in Sklar's theorem, and the associated copula density function $c(\cdot\,,\cdot)$ can then be readily obtained from Equation (1) as

$$h(x,y) = f(x)\,g(y)\,c(F(x),G(y)) \tag{7}$$

where $f(x)$ and $g(y)$ denote the marginal density functions of X and Y, respectively. Accordingly, a copula density function can be expressed as follows:

$$c(u, v) = \frac{h(F^{-1}(u), G^{-1}(v))}{f(F^{-1}(u))\,g(G^{-1}(v))}. \tag{8}$$

Now, given a random sample $(x_1, y_1), \ldots, (x_n, y_n)$ generated from the continuous random vector (X, Y), let

$$(u_i, v_i) = (F(x_i), G(y_i)), \quad i = 1, \ldots, n, \tag{9}$$

where $F(\cdot)$ and $G(\cdot)$ are the usually unknown marginal cumulative distribution functions (cdfs) of X and Y. The empirical marginal cdfs $\hat{F}(\cdot)$ and $\hat{G}(\cdot)$ are then utilized to determine the *pseudo-observations*:

$$(\hat{u}_i, \hat{v}_i) = (\hat{F}(x_i), \hat{G}(y_i)), \quad i = 1, \ldots, n, \tag{10}$$

where the empirical cdfs (ecdfs) are given by $\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{I}(x_i \leq x)$ and $\hat{G}(y) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{I}(y_i \leq y)$, with $\mathcal{I}(\aleph)$ denoting the indicator function which is equal to one if the condition $\aleph$ is verified and zero, otherwise. Equivalently, one has

$$(\hat{u}_i, \hat{v}_i) = (r_i/n, s_i/n), \tag{11}$$

where r_i is the rank of x_i among $\{x_1, \ldots, x_n\}$, and s_i is the rank of y_i among $\{y_1, \ldots, y_n\}$.

The frequencies or probability mass function of an empirical copula can be expressed as

$$\begin{aligned} \hat{c}(u, v) &= \frac{1}{n}\sum_{i=1}^{n} \mathcal{I}(\hat{F}(x_i) = u)\,\mathcal{I}(\hat{G}(y_i) = v) \\ &= \frac{1}{n}\sum_{i=1}^{n} \mathcal{I}(r_i/n = u)\,\mathcal{I}(s_i/n = v), \end{aligned} \tag{12}$$

and the corresponding empirical copula (distribution function) is then given by

$$\begin{aligned} \hat{C}(u, v) &= \frac{1}{n}\sum_{i=1}^{n} \mathcal{I}(\hat{F}(x_i) \leq u)\,\mathcal{I}(\hat{G}(y_i) \leq v) \\ &= \frac{1}{n}\sum_{i=1}^{n} \mathcal{I}(r_i/n \leq u)\,\mathcal{I}(s_i/n \leq v), \end{aligned} \tag{13}$$

which is a consistent estimate of $C(u, v)$. We note that, in practice, the ranks are often divided by $n + 1$ instead of n in order to mitigate certain boundary effects, and that other adjustments that are specified in Section 2 may also be applied. As pointed out by [2], who refers to [3], "Empirical copulas were introduced and first studied by Deheuvels who called them empirical dependence functions".

Additional properties of copulas that are not directly relevant to the results presented in this article are discussed for instance in [4–6].

This article contains certain derivations that do not seem to be available in the literature and also provides missing steps that complete the published proofs. It is structured as follows: Sections 2–5, which, respectively, focus on Spearman's, Kendall's, Blomqvist's and Hoeffding's correlation coefficients, include representations of these measures of dependence in terms of copulas, in addition to providing sample estimates thereof and pointing out related distributional results of interest. The effectiveness of these correlation coefficients in assessing the trends present in five data sets exhibiting distinctive patterns is assessed in a numerical study that is presented in Section 6. Section 7 is dedicated to multivariate extensions of the four measures of association and their copula representations.

To the best of our knowledge, the four major dependence measures discussed here, along with their representations in terms of copulas, have not been previously covered in a single source.

2. Spearman's Rank Correlation

Spearman's rank correlation statistic, also referred to as Spearman's ρ, measures the extent to which the relationship between two variables is monotonic—either increasing or decreasing.

First, Spearman's ρ is expressed in terms of a copula denoted by $C(U, V)$. Then, some equivalent representations of Spearman's rank correlation statistic are provided; one of them is obtained by replacing $C(U, V)$ by its empirical counterpart.

Let (X, Y) be a bivariate continuous random vector having $h(x, y)$ as its joint density function, and $F(X)$ and $G(Y)$ denote the respective marginal distribution functions of X and Y.

Theoretically, Spearman's correlation is given by

$$\rho_S = \frac{Cov[F(X), G(Y)]}{\sqrt{Var[F(X)]Var[G(Y)]}} \tag{14}$$

$$= \frac{\int\int_{\mathbb{R}^2} F(x)G(y)h(x,y)\mathrm{d}x\mathrm{d}y - (\int_{\mathbb{R}} F(x)\,\mathrm{d}F(x))(\int_{\mathbb{R}} G(y)\,\mathrm{d}G(y))}{\sqrt{[\int_{\mathbb{R}} F(x)^2\,\mathrm{d}F(x) - (\int_{\mathbb{R}}(F(x)\,\mathrm{d}F(x))^2][\int_{\mathbb{R}} G(y)^2\,\mathrm{d}G(y) - (\int_{\mathbb{R}}(G(y)\,\mathrm{d}G(y))^2]}} \tag{15}$$

$$= \frac{\int_0^1 \int_0^1 u\, v\, c(u,v)\,\mathrm{d}u\mathrm{d}v - (1/2)(1/2)}{\sqrt{(1/12)(1/12)}} \text{ in light of (8),} \tag{16}$$

with the transformation $\{x = F^{-1}(u)$ and $y = G^{-1}(v)\}$ whose Jacobian is the inverse of the Jacobian associated with the following transformation: $\{u = F(x)$ and $v = G(y)\}$, that is, $1/[f(F^{-1}(u))\, g(G^{-1}(v))]$,

$$= 12 \int_0^1 \int_0^1 C(u,v)\,\mathrm{d}u\mathrm{d}v - 3, \tag{17}$$

$$= 12\, E[UV] - 3, \tag{18}$$

where $C(\cdot,\cdot)$ and $c(\cdot,\cdot)$, respectively, denote the copula and copula density function associated with (X, Y), and $\mathbb{R}$ represents the set of real numbers. In [7,8], it is taken as a given that the double integral appearing in (16) can be expressed as that appearing in (17). We now prove that this is indeed the case. First, recall that $\frac{\partial^2 C(u,v)}{\partial u \partial v} = c(u,v)$, the copula density function. On integrating by parts twice, one has

$$\begin{aligned}
\int_0^1 \int_0^1 u\, v\, \mathrm{d}C(u,v) &= \int_0^1 \int_0^1 uv \frac{\partial^2 C(u,v)}{\partial u \partial v} \mathrm{d}v\mathrm{d}u \\
&= \int_0^1 u \left[\int_0^1 v \frac{\partial}{\partial v}\left(\frac{\partial C(u,v)}{\partial u}\right) \mathrm{d}v \right] \mathrm{d}u \\
&= \int_0^1 u \left[v \frac{\partial C(u,v)}{\partial u}\Big|_0^1 - \int_0^1 \frac{\partial C(u,v)}{\partial u} \mathrm{d}v \right] \mathrm{d}u \\
&= \int_0^1 u \left[1 - \int_0^1 \frac{\partial C(u,v)}{\partial u} \mathrm{d}v \right] \mathrm{d}u, \quad \text{as } C(u,1) = u \\
&= \int_0^1 u\mathrm{d}u - \int_0^1 \int_0^1 u \frac{\partial C(u,v)}{\partial u} \mathrm{d}v\mathrm{d}u \\
&= \frac{1}{2} - \int_0^1 \left[uC(u,v)|_0^1 - \int_0^1 C(u,v)\mathrm{d}u \right] \mathrm{d}v \\
&= \frac{1}{2} - \frac{1}{2} + \int_0^1 \int_0^1 C(u,v)\mathrm{d}u\mathrm{d}v, \quad \text{as } C(1,v) = v \\
&= \int_0^1 \int_0^1 C(u,v)\mathrm{d}u\mathrm{d}v.
\end{aligned} \tag{19}$$

Now, let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample generated from the random vector (X, Y), and denote by $\hat{F}(X)$ and $\hat{G}(Y)$ the respective empirical distribution functions of X and Y. Throughout this article, the sample size is assumed to be n. On denoting by R_i and S_j, the rank of X_i among $\{X_1, \ldots, X_n\}$ and the rank of Y_j among $\{Y_1, \ldots, Y_n\}$, respectively, one has $\hat{F}(X_i) = R_i/n \equiv U_i$ and $\hat{G}(Y_j) = S_j/n \equiv V_j$, where U_i and V_j denote the canonical pseudo-observations on each component. Note that the rank averages $\bar{R}$ and $\bar{S}$ are both equal to $(n+1)/2$. Then, Spearman's rank correlation estimator admits the following equivalent representations:

$$\hat{\rho}_S = \frac{\sum_{i=1}^n (R_i - \overline{R})(S_i - \overline{S})}{\sqrt{\sum_{i=1}^n (R_i - \overline{R})^2 \sum_{i=1}^n (S_i - \overline{S})^2}} \tag{20}$$

$$= \frac{(\sum_{i=1}^n R_i S_i) - n\overline{R}\,\overline{S}}{\sqrt{[(\sum_{i=1}^n R_i^2) - n\overline{R}^2][(\sum_{i=1}^n S_i^2) - n\overline{S}^2)]}} \tag{21}$$

$$= \frac{(\sum_{i=1}^n \hat{F}(x_i)\,\hat{G}(y_i)) - n(n+1)^2/4}{\sqrt{[(\sum_{i=1}^n \hat{F}(x_i)^2) - n(n+1)^2/4][(\sum_{i=1}^n \hat{G}(y_i)^2) - n(n+1)^2/4]}} \tag{22}$$

$$= \frac{(\sum_{i=1}^n U_i V_i) - n(n+1)^2/4}{\sqrt{[(\sum_{i=1}^n U_i^2) - n(n+1)^2/4][(\sum_{i=1}^n V_i^2) - n(n+1)^2/4]}} \tag{23}$$

$$= \frac{\sum_{i=1}^n (U_i - \overline{U})(V_i - \overline{V})}{\sqrt{\sum_{i=1}^n (U_i - \overline{U})^2 \sum_{i=1}^n (V_i - \overline{V})^2}}, \tag{24}$$

where $\bar{U} = \sum_{i=1}^n U_i/n$ and $\bar{V} = \sum_{i=1}^n V_i/n$.

Of course, (24) readily follows from (20), and it is seen from either one of these expressions that Spearman's rank correlation is not be affected by any monotonic affine transformation, whether applied to the ranks or the canonical pseudo-observations. As pointed out for instance in [9], the pseudo-observations are frequently taken to be

$$\hat{U}_i = \frac{R_i}{n+1} = \frac{n}{n+1}\hat{F}(x_i) = \frac{1}{n+1}\sum_{k=1}^n \mathcal{I}(x_k \le x_i) \tag{25}$$

and

$$\hat{V}_j = \frac{S_j}{n+1} = \frac{n}{n+1}\hat{G}(y_j) = \frac{1}{n+1}\sum_{k=1}^n \mathcal{I}(y_k \le y_j). \tag{26}$$

Alternatively, one can define the pseudo-observations so that they be uniformly—and less haphazardly—distributed over the unit interval as follows:

$$\tilde{U}_i = \frac{R_i}{n} - \frac{1}{2n} = \hat{F}(x_i) - \frac{1}{2n} \tag{27}$$

and

$$\tilde{V}_j = \frac{S_j}{n} - \frac{1}{2n} = \hat{G}(y_j) - \frac{1}{2n}. \tag{28}$$

In a simulation study, Dias (2022) [10] observed that such pseudo-observations have a lower bias than those obtained by dividing the ranks by $n+1$. What is more, it should be observed that if we extend the pseudo-observations $\tilde{U}_i$, $i = 1, \ldots, n$, and $\tilde{V}_j$, $j = 1, \ldots, n$, by $\frac{1}{2n}$ on each side and assign their respective probability, namely, $\frac{1}{n}$, to each of the n resulting subintervals, the marginal distributions is then uniformly distributed within the interval $[0, 1]$, which happens to be a requirement for a copula density function. However, this is not the case for any other affine transformation of the ranks. The alternative transformations $\frac{\text{rank}-1/3}{n+1/3}$ and $\frac{\text{rank}-1}{n-1}$ were also considered by [10,11], respectively. As established in [10], the pseudo-observation estimators resulting from any of the above-mentioned transformations as well as the canonical pseudo-observations are consistent estimators of the underlying distribution functions.

Kojanovik and Yan (2010) [7] pointed out that $\hat{\rho}_S$, as specified in (21), can also be expressed as

$$\hat{\rho}_S = \frac{12}{n(n+1)(n-1)} \sum_{i=1}^{n} R_i S_i - 3\,\frac{(n+1)}{(n-1)}, \tag{29}$$

where $\hat{\rho}_S$ is a consistent estimator of ρ_S.

Moreover, it can be algebraically shown that, alternatively,

$$\hat{\rho}_S = 1 - 6 \sum_{i=1}^{n} \frac{(R_i - S_i)^2}{n(n^2-1)} \tag{30}$$

when the ranks are distinct integers.

On writing (17) as

$$\rho_S = 12 \int_0^1 \int_0^1 u\, v\, \mathrm{d}C(u,v) - 3, \tag{31}$$

and replacing $C(u,v)$ by $\hat{C}(u,v)$ as defined in (13), the double integral becomes

$$\frac{1}{n} \sum_{i=1}^{n} \int_0^1 u\, \mathrm{d}(\mathcal{I}(r_i/n \le u)) \int_0^1 v\, \mathrm{d}(\mathcal{I}(s_i/n \le v)).$$

For instance, on integrating the first integral by parts, one has

$$u\, \mathcal{I}(r_i/n \le u)\big|_0^1 - \int_0^1 \mathcal{I}(r_i/n \le u)\, \mathrm{d}u = 1 - (1 - r_i/n) = r_i/n.$$

Thus, the resulting estimator of Spearman's rank correlation is given by

$$\hat{\rho}_S = \frac{12}{n^3} \sum_{i=1}^{n} R_i S_i - 3,$$

which is approximately equal to that given in (29).

Now, letting $C_\theta(u,v)$ be a copula whose functional representation is known, and assuming that it is a one-to-one function of the dependence parameter θ, it follows from (17) that

$$\rho_S(\theta) = 12 \iint_{\mathbb{1}^2} C_\theta(u,v)\mathrm{d}u\mathrm{d}v - 3, \tag{32}$$

which provides an indication of the extent to which the variables are monotonically related. Moreover, since $\hat{\rho}_S$, as defined in (21), (29) or (30), tends to $\rho_S(\theta)$, $\hat{\theta} = \rho_S^{-1}(\hat{\rho}_S)$ can serve as an estimate of θ.

It follows from (17) that Spearman's ρ can be expressed as

$$\rho_S = 12 \iint_{\mathbb{1}^2} [C(u,v) - uv]\mathrm{d}u\mathrm{d}v. \tag{33}$$

On replacing $[C(u,v) - uv]$ in (33) by $|C(u,v) - uv|$, one obtains a measure based on the L_1 distance between the copula C and the product copula $\Pi = uv$ ([5]). This is the so-called Schweizer–Wolff's sigma as defined in [12], which is given by

$$\sigma_{X,Y} = \sigma_C = 12 \iint_{\mathbb{1}^2} |C(u,v) - uv|\mathrm{d}u\mathrm{d}v. \tag{34}$$

The expression (34) is a measure of dependence which satisfies the properties of Rényi's axioms [13] for measures of dependence [12], [14] (p. 145).

Note that Pearson's correlation coefficient,

$$\hat{r} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}\,, \tag{35}$$

only measures the strength of a *linear* relationship between X and Y, whereas Spearman's rank correlation ρ_S assesses the strength of *any monotonic* relationship between X and Y. The latter is always well-defined, which is not the case for the former. Both vary between -1 and 1 and $\rho_S = \pm 1$ indicates that Y is either an increasing or a decreasing function of X. Moreover, it should be noted that Pearson's correlation coefficient cannot be expressed in terms of copulas since its estimator is a function of the observations themselves rather than their ranks.

The next three sections include results that were gleaned from the following books among others: [4,5,15,16].

3. Kendall's Rank Correlation Coefficient

Kendall's τ, also referred to as Kendall's rank correlation coefficient, was introduced by [17]. Maurice Kendall also proposed an estimate thereof and published several papers as well as a monograph in connection with certain ordinal measures of correlation. Further historical details are available from [18].

Kendall's τ is a nonparametric measure of association between two variables, which is based on the number of concordant pairs minus the number of discordant pairs. Consider two observations (x_i, y_i) and (x_j, y_j), with $(i, j) \in \{1, \ldots, n\}$ such that $i \neq j$, that are generated from a vector (X, Y) of continuous random variables. Then, for any such assignment of pairs, define each pair as being concordant, discordant or equal, as follows:

- (x_i, y_i) and (x_j, y_j) are *concordant* if
  $\{x_i < x_j$ and $y_i < y_j$ or if $x_i > x_j$ and $y_i > y_j\}$, or equivalently
  $(x_i - x_j)(y_i - y_j) > 0$, i.e., the slope of the line connecting the two points is positive.
- (x_i, y_i) and (x_j, y_j) are *discordant* if
  $\{x_i < x_j$ and $y_i > y_j$ or if $x_i > x_j$ and $y_i < y_j\}$, or equivalently
  $(x_i - x_j)(y_i - y_j) < 0$, i.e., the slope of the line connecting the two points is negative.
- (x_i, y_i) and (x_j, y_j) are equal if $x_i = x_j$ or $y_i = y_j$. Actually, pair equality can be disregarded as the random variables X and Y are assumed to be continuous.

3.1. The Empirical Kendall's τ

Let $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ be a random sample of n pairs arising from the vector (X, Y) of continuous random variables. There are $C_2^n = \binom{n}{2}$ possible ways of selecting distinct pairs (x_i, y_i) and (x_j, y_j) of observations in the sample, with each pair being either concordant or discordant.

Let S_{ij} be defined as follows:

$$S_{ij} = \operatorname{sign}(X_i - X_j)\operatorname{sign}(Y_i - Y_j), \tag{36}$$

where

$$\operatorname{sign}(u) = \begin{cases} -1 & \text{if } u < 0 \\ 0 & \text{if } u = 0 \\ 1 & \text{if } u > 0. \end{cases}$$

Then, the values that S_{ij} can take on are

$$s_{ij} = \begin{cases} -1 & \text{when the pairs are discordant} \\ 0 & \text{when the pairs are neither concordant nor discordant} \\ 1 & \text{when the pairs are concordant.} \end{cases}$$

Kendall's sample $\hat{\tau}$ is defined as follows:

$$\hat{\tau} = \sum_{1 \le i < j \le n} \frac{s_{ij}}{C_2^n} = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} s_{ij}. \tag{37}$$

Alternatively, on letting c denote the number of concordant pairs and d the number of discordant pairs in a given sample of size n, one can express the estimate of Kendall's τ as

$$\hat{\tau} = \frac{c-d}{c+d} = \frac{c-d}{C_2^n} = \frac{2(c-d)}{n(n-1)}. \tag{38}$$

As it is assumed that there can be no equality between pairs, $C_2^n = c + d$, so that

$$\hat{\tau} = \frac{4c}{n(n-1)} - 1 \text{ or, equivalently, } \hat{\tau} = 1 - \frac{4d}{n(n-1)}. \tag{39}$$

In fact, $\hat{\tau}$ is an unbiased estimator of τ. As well, Kendall and Gibbons (1990) [19] (Chapter 5) established that $\mathrm{Var}(\hat{\tau}) = \frac{2(2n+5)}{9n(n-1)}$. A coefficient related to that specified in (39) was discussed in [20–22] in the context of double time series.

3.2. The Population Kendall's τ

Letting (X_1, Y_1) and (X_2, Y_2) be independent and identically distributed random vectors, with the joint distribution function of (X_i, Y_i) being $H(x, y)$, $F(x)$ and $G(y)$ denote the respective distribution functions of X_i and Y_j, $i, j = 1, 2$, and the associated copula be $C(u, v) = H(F^{-1}(u), G^{-1}(v))$, the population Kendall's τ is defined as follows:

$$\begin{aligned}
\tau = \tau_{X,Y} &= \Pr[\text{concordant pairs}] - \Pr[\text{discordant pairs}] \\
&\equiv p_c - p_d \\
&= \Pr[(X_1 - X_2)(Y_1 - Y_2) > 0] - \Pr[(X_1 - X_2)(Y_1 - Y_2) < 0] && (40) \\
&= 2\Pr[(X_1 - X_2)(Y_1 - Y_2) > 0] - 1 && (41) \\
&= 4\Pr[(X_1 < X_2, Y_1 < Y_2)] - 1 && (42) \\
&= 4 \iint_{\mathbb{R}^2} \Pr(X_2 \le x, Y_2 \le y) \mathrm{d}H(x, y) - 1, \text{ with } H(x, y) = C(F(x), G(y)) && (43) \\
&= 4 \iint_{\mathbb{R}^2} H(x, y) c(F(x), G(y)) f(x) g(y) \mathrm{d}x\mathrm{d}y - 1 \\
&= 4 \iint_{\mathbb{I}^2} \frac{H(F^{-1}(u), G^{-1}(v)) c(u, v) f(F^{-1}(u)) g(G^{-1}(v))}{f(F^{-1}(u)) g(G^{-1}(v))} \mathrm{d}u\mathrm{d}v - 1 \\
&= 4 \int_0^1 \int_0^1 C(u, v) \mathrm{d}C(u, v) - 1 && (44) \\
&= 4\,\mathrm{E}[C(U, V)] - 1, && (45)
\end{aligned}$$

where

U and V have a Uniform $(0, 1)$ distribution, their joint cdf being $C(u, v)$;
$u = F_X(x)$ and $v = F_Y(y)$;
$\mathbb{R}^2 \equiv \{(x, y) \mid x \text{ and } y \text{ are real numbers}\}$;
$\mathrm{d}C(u, v) = \frac{\partial^2 C(u,v)}{\partial u \partial v} \mathrm{d}u\mathrm{d}v = c(u, v)\mathrm{d}u\mathrm{d}v$.
Clearly, (41) follows from (40) since

$$\Pr[(X_1 - X_2)(Y_1 - Y_2) < 0] = 1 - \Pr[(X_1 - X_2)(Y_1 - Y_2) > 0].$$

We now state Theorem 5.1.1 from [5]:

Theorem 2. *Let (X_1, Y_1) and (X_2, Y_2) be independent vectors of continuous random variables with joint distributions functions H_1 and H_2, respectively, with common marginals $F(\cdot)$ and $G(\cdot)$. Let*

C_1 and C_2 be the copulas of (X_1, Y_1) and (X_2, Y_2), respectively, so that $H_1(x, y) = C_1(F(x), G(y))$ and $H_2(x, y) = C_2(F(x), G(y))$. Let

$$Q = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0]. \tag{46}$$

Then,

$$Q(C_1, C_2) = 4 \iint_{\mathbb{I}^2} C_2(u_1, u_2) dC_1(u_1, u_2) - 1. \tag{47}$$

If X and Y are continuous random variables whose copula is C, then Equation (44) follows from (40), (46) and (47).

3.3. Marginal Probability of S_{ij}

The marginal probability of S_{ij} is

$$p_{S_{ij}}(s_{ij}) = \begin{cases} p_c, & s_{ij} = 1 \\ p_d, & s_{ij} = -1 \\ 1 - p_c - p_d, & s_{ij} = 0\,. \end{cases}$$

Gibbons and Chakraborti (2003) [15] proved that

$$E(S_{ij}) = 1\, p_c + (-1)\, p_d = \tau.$$

3.4. Certain Properties of τ

- The correlation coefficient τ is invariant with respect to strictly increasing transformations.
- If X_1 and Y_1 are independent, then the value of τ is zero:

$$\begin{aligned} \tau(X_1, Y_1) &= 2 \Pr[(X_1 - X_2)(Y_1 - Y_2) > 0] - 1 \\ &= 2\{\Pr[X_1 - X_2 > 0, Y_1 - Y_2 > 0] + \Pr[X_1 - X_2 < 0, Y_1 - Y_2 < 0]\} - 1 \\ &= 2\left(\frac{1}{4} + \frac{1}{4}\right) - 1 = 0. \end{aligned}$$

- Kendall's τ takes on values in the interval $[-1, 1]$.
- As stated in [4], when the number of discordant pairs is 0, the value of τ is maximum and equals 1, which means a perfect relationship; the variables are then comonotonic, i.e., one variable is an increasing transform of the other; if the variables are counter-monotonic, i.e., one variable is a decreasing transform of the other, the correlation coefficient τ equals -1. Note that these two properties do not hold for Pearson's correlation coefficient. Moreover, it proves more appropriate to make use of Kendall's τ when the joint distribution is not Gaussian.

4. Blomqvist's Correlation Coefficient

Blomqvist (1950) [23] proposed a measure of dependence that was similar in its structure to Kendall's correlation coefficient, except that in this instance, medians were utilized. Blomqvist's correlation coefficient can be defined as follows:

$$\begin{aligned} \beta = \beta_{X,Y} = P[(X - F_X^{-1}(1/2))(Y - G_Y^{-1}(1/2)) > 0] \\ - P[(X - F_X^{-1}(1/2))(Y - G_Y^{-1}(1/2)) < 0], \end{aligned} \tag{48}$$

where $F_X^{-1}(1/2) \equiv \tilde{x}$ and $G_Y^{-1}(1/2) \equiv \tilde{y}$ are the respective medians of X and Y, which explains why this coefficient is also known as the median correlation coefficient.

Now, letting X and Y be continuous random variables whose joint cdf is $H(\cdot,\cdot)$, $F(\cdot)$ and $G(\cdot)$ denote the respective marginal cdfs, and $C(\cdot,\cdot)$ be the associated copula, then,

$$F(\tilde{x}) = F(F_X^{-1}(1/2)) = 1/2,\ G(\tilde{y}) = G(G_Y^{-1}(1/2)) = 1/2,$$

and

$$\begin{aligned}
\beta = \beta_{X,Y} &= 2\,\mathrm{Pr}[(X - F_X^{-1}(1/2))(Y - G_Y^{-1}(1/2)) > 0] - 1 && (49)\\
&= 2\,\{\mathrm{Pr}[X < F_X^{-1}(1/2), Y < G_Y^{-1}(1/2)] + \mathrm{Pr}[X > F_X^{-1}(1/2), Y > G_Y^{-1}(1/2)]\} - 1 && (50)\\
&= 4\,H(F_X^{-1}(1/2), G_Y^{-1}(1/2)) - 1 && (51)\\
&= 4\,C(1/2, 1/2) - 1. && (52)
\end{aligned}$$

In the development of these equations, the following relationships were utilized in addition to $H(x,y) = C(F(x), G(y))$:

$$\begin{aligned}
&P[(X - F_X^{-1}(1/2))(Y - G_Y^{-1}(1/2)) > 0]\\
&= P[X - F_X^{-1}(1/2) > 0, Y - G_Y^{-1}(1/2) > 0] \qquad (53)\\
&\quad + P[X - F_X^{-1}(1/2) < 0, Y - G_Y^{-1}(1/2) < 0];
\end{aligned}$$

$$P[X > F_X^{-1}(1/2), Y > G_Y^{-1}(1/2)] = P[X < F_X^{-1}(1/2), Y < G_Y^{-1}(1/2)]. \qquad (54)$$

4.1. Estimation of β

Let $\tilde{x}_n$ and $\tilde{y}_n$ be the respective medians of the samples $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$. The computation of Blomqvist's correlation coefficient is based on a 2×2 contingency table that is constructed from these two samples.

According to Blomqvist's suggestion, the $x\,y$-plane is divided into four regions by drawing the lines $x = \tilde{x}_n$ and $y = \tilde{y}_n$. Let n_1 and n_2 be the number of points belonging to the first or third quadrant and to the second or fourth quadrant, respectively.

Blomqvist's sample β_n or the median correlation coefficient is defined by

$$\beta_n = \frac{n_1 - n_2}{n_1 + n_2} = 2\,\frac{n_1}{n_1 + n_2} - 1. \qquad (55)$$

If the sample size n is even, then clearly, no sample points fall on the lines $x = \tilde{x}_n$ and $y = \tilde{y}_n$. Moreover, n_1 and n_2 are then both even. However, if n is odd, then one or two sample points must fall on the lines $x = \tilde{x}_n$ and $y = \tilde{y}_n$. In the first case (a single point lying on a median), Blomqvist proposed that this point shall not be counted. For the second case, one point has to fall on each line; then, one of the points is assigned to the quadrant touched by both points, while the other is not counted.

Genest et al. (2013) [24] provided an accurate interpretation of β_n as "the difference between the proportion of sample points having both components either smaller or greater than their respective medians, and the proportion of the other sample points". Finally, as pointed out by [23], the definition of β_n as given in (55) was not new [25]; however, its statistical properties had not been previously fully investigated.

4.2. Some Properties of Blomqvist's Correlation Coefficient

- The coefficient β is invariant under strictly increasing transformations of X and Y.
- The correlation coefficient β takes on values in the interval $[-1, 1]$.
- If X and Y are independent, then $C(1/2, 1/2) = F(1/2)G(1/2) = 1/4$, and $\beta = 0$.

5. Hoeffding's Dependence Index

To measure the strength of relationships that are not necessarily monotonic, one may make use of Hoeffding's dependence coefficient. Letting $H(X, Y)$ denote the joint distribution function of X and Y, and $F(X)$ and $G(Y)$ stand for the marginal distribution functions of X and Y, Hoeffding's nonparametric rank statistic for testing bivariate independence is based on

$$D(x, y) = H(x, y) - F(x)G(y), \tag{56}$$

which is equal to zero if and only if X and Y are independently distributed.

The nonparametric estimator of the quantity $\mathcal{D}^2 = 30 \int D^2(x, y)\mathrm{d}H(x, y)$ results in the statistic

$$\hat{\mathcal{D}}^2 = 30\,\frac{Q - 2(n-2)R + (n-2)(n-3)S}{n(n-1)(n-2)(n-3)(n-4)}, \tag{57}$$

where

$$Q = \sum_{i=1}^{n} (R_i - 1)(R_i - 2)(S_i - 1)(S_i - 2), \tag{58}$$

$$R = \sum_{i=1}^{n} (R_i - 2)(S_i - 2)C_i, \tag{59}$$

and

$$S = \sum_{i=1}^{n} (C_i - 1)C_i, \tag{60}$$

with R_i and S_j representing the rank of X_i among $\{X_1, \ldots, X_n\}$ and the rank of Y_j among $\{Y_1, \ldots, Y_n\}$, respectively, and C_i denoting the number of bivariate observations (X_j, Y_j) for which $X_j \leq X_i$ and $Y_j \leq Y_i$.

We now state Hoeffding's Lemma [26]: *Let X and Y be random variables with joint distribution function $H(x, y)$ and marginal distributions $F(x)$ and $G(y)$. If $E(XY)$ and $E(X)E(Y)$ are finite, then*

$$\mathrm{Cov}(X, Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [H(x, y) - F(x)G(y)]\mathrm{d}x\mathrm{d}y. \tag{61}$$

This result became known when it was cited by [27]. Refs. [28,29] discussed multivariate versions of this lemma.

The correlation coefficient is thus given by

$$\mathrm{Cor}(X, Y) = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [H(x, y) - F(x)G(y)]\mathrm{d}x\mathrm{d}y}{\sqrt{Var(X)}\sqrt{Var(Y)}} \tag{62}$$

or

$$\mathrm{Cor}(X, Y) = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [C(F(x), G(y)) - F(x)G(y)]\mathrm{d}x\mathrm{d}y}{\sqrt{Var(X)}\sqrt{Var(Y)}}, \tag{63}$$

with (63) resulting from Sklar's theorem.

Invoking Hoeffding's lemma, Hofert et al. (2019) [16] (p. 47) pointed out two fallacies about the uniqueness and independence of random variables. Hoeffding appealed to his lemma to identify the bivariate distributions with given marginal distribution functions $F(x)$ and $G(y)$, which minimize or maximize the correlation between X and Y.

Hoeffding's Φ^2

Hoeffding (1940) [26] defined the stochastic dependence index of the random variables X and Y as

$$\Phi^2_{X,Y} = 90 \int_0^1 \int_0^1 (C(u, v) - uv)^2 \mathrm{d}u\mathrm{d}v, \tag{64}$$

where

$$\Phi^2_{X,Y} = \begin{cases} 0 & \text{in the case of independence since then } C(u,v) = uv, \\ 1 & \text{in the case of monotone dependence,} \\ \Phi^2 \in (0,1) & \text{otherwise.} \end{cases}$$

Hoeffding (1940) [26] showed that $\Phi^2_{X,Y}$ takes the value one in the cases of monotonically increasing and monotonically decreasing continuous functional dependence; it is otherwise less than one and greater than zero.

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be a simple random sample generated from the two-dimensional random vector $\mathbf{X}$ whose distribution function and copula are denoted by $H(\cdot)$ and $C(\cdot)$, respectively, and assumed to be unknown. The copula C is then estimated by the empirical copula $\hat{C}_n$, which is defined as

$$\hat{C}_n(\mathbf{u}) = \frac{1}{n}\sum_{j=1}^{n}\prod_{i=1}^{2} I(\hat{U}_{ij} \le u_i) \quad \text{for } \mathbf{u} = (u_1, u_2) \in \mathbb{I}^2, \tag{65}$$

with the pseudo-observations $\hat{U}_{ij} = \hat{F}_i(X_{ij})$ for $i = 1, 2$, and $j = 1, \ldots, n$ and $\hat{F}_i(x) = \frac{1}{n}\sum_{j=1}^{n} I(X_{ij} \le x)$, $x \in \mathbb{R}$. Since $\hat{U}_{ij} = \frac{1}{n}(\text{rank of } X_{ij} \text{ in } X_{i1}, \ldots, X_{in})$, statistical inference is based on the ranks of the observations.

A nonparametric estimator of Φ^2 is then obtained by replacing the copula $C(\cdot)$ in (64) by the empirical copula $\hat{C}_n(\cdot)$, i.e.,

$$\hat{\Phi}^2_n := \Phi^2(\hat{C}_n) = 90\iint_{\mathbb{I}^2}\{\hat{C}_n(\mathbf{u}) - \Pi(\mathbf{u})\}^2 \mathrm{d}\mathbf{u}, \tag{66}$$

where $\Pi(\mathbf{u}) = u_1 u_2$ denotes the independence copula.

As explained in [30], this estimator can be evaluated as follows:

$$\hat{\Phi}^2_n = 90\left\{\frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\prod_{i=1}^{2}(1 - \max\{\hat{U}_{ij}, \hat{U}_{ik}\}) - \frac{1}{2n}\sum_{j=1}^{n}\prod_{i=1}^{2}(1 - \hat{U}^2_{ij}) + \left(\frac{1}{3}\right)^2\right\}. \tag{67}$$

The asymptotic distribution of $\hat{\Phi}^2_n$ can be deduced from the asymptotic behavior of the empirical copula process which, for instance, has been discussed by [31–33].

The quantity $\Phi^2_{X,Y}$ was introduced by [34] without the normalizing factor 90, as a distribution-free statistic for testing the independence of X and Y.

Referring to [12], Nelsen (2006) [5] (p. 210) states that "... any L_p distance should yield a symmetric nonparametric measure of dependence". For any p, $1 < p < \infty$, the L_p distance between the copula $C(\cdot)$ and the product copula $\Pi(\cdot)$ is given by the following expression:

$$\left(k_p \iint_{\mathbb{I}^2} |C(u,v) - uv|^p \mathrm{d}u\mathrm{d}v\right)^{\frac{1}{p}}, \tag{68}$$

where k_p is the normalizing factor. On letting $p = 2$, one obtains $\Phi_{X,Y}$.

6. Illustrative Examples

In order to compare the measures of association discussed in the previous sections, five two-dimensional data sets exhibiting different patterns that will be identified by the letters A, B, C, D and E, are considered. The first one was linearly decreasing, in which case Pearson's correlation ought to be the most appropriate coefficient. The strictly monotonic pattern of the second set ought to be readily detected by Spearman's, Kendall's and Blomqvist's coefficients, whose performance is assessed when applied to the fourth pattern, which happens to be piecewise monotonic. In the case of patterns C and E, whose points exhibit distinctive patterns, Hoeffding's measure of dependence is expected to be more suitable than any of the other measures of association.

First, 500 random values of x, denoted by $\mathcal{S}$, were generated within the interval $(-3, 3)$. Now, let

$f_A(x) = -x/5 + 1 + \epsilon$,
$f_B(x) = -x^5 + \epsilon$,
$f_C(x) = \sin(x) + \epsilon$,
$f_D(x) = -\sqrt{|x^{3/2}|} + \epsilon$ and

$f_E = \tan(x)^3 + \epsilon,$

where ϵ represents a slight perturbation consisting of a multiple of random values generated from a uniform distribution on the interval $[-1, 1]$. The five resulting data sets, $A = \{(x, f_A(x)) | x \in \mathcal{S}\}$, $B = \{(x, f_B(x)) | x \in \mathcal{S}\}$, $C = \{(\cos(x), f_C(x)) | x \in \mathcal{S}\}$, $D = \{(x, f_D(x)) | x \in \mathcal{S}\}$ and $E = \{(x, f_E(x)) | x \in \mathcal{S}\}$ are plotted in Figures 1–5.

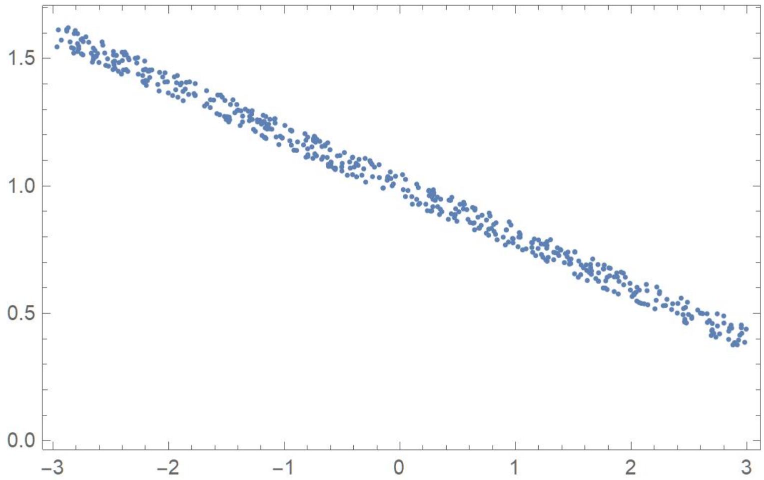

Figure 1. Plot of data set A.

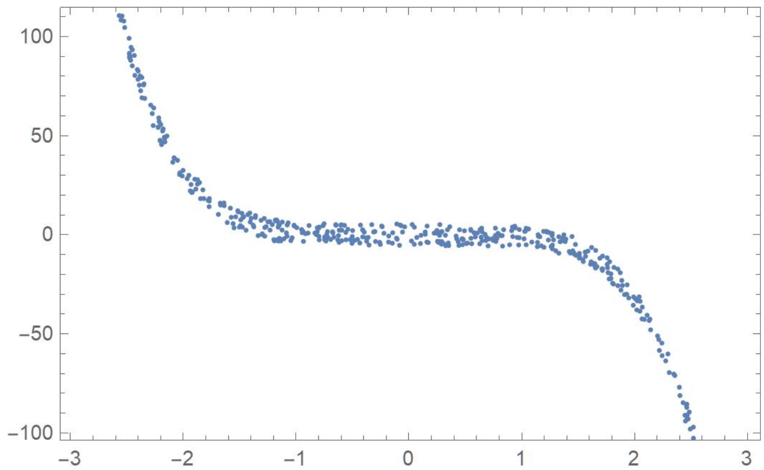

Figure 2. Plot of data set B.

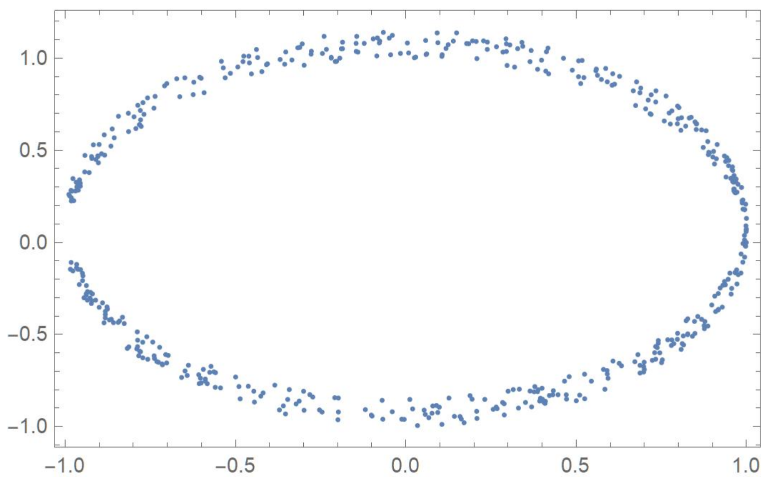

Figure 3. Plot of data set C.

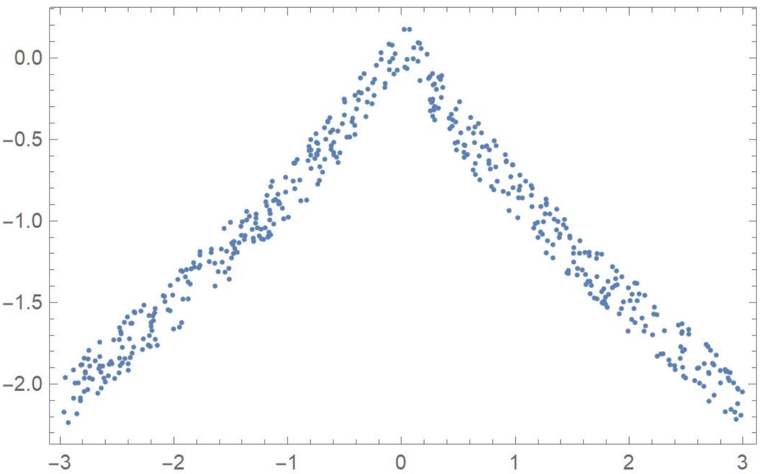

Figure 4. Plot of data set D.

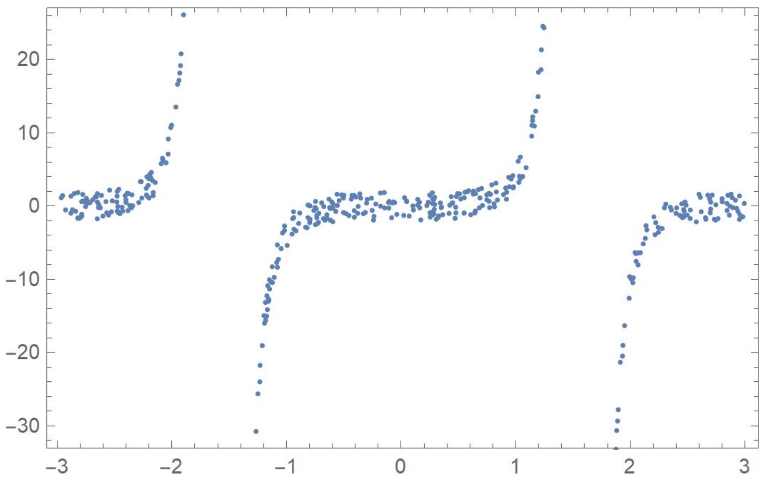

Figure 5. Plot of data set E.

We then evaluated Spearman's, Kendall's, Blomqvist's and Hoeffding's statistics, as well as Pearson's sample correlation coefficient for each data set. Their numerical values and associated p-values are reported in Table 1.

Table 1. Five statistics and their associated p-values.

Statistics and p-Values	A	B	C	D	E
Spearman	$\{-0.9963, 0\}$	$\{-0.9218, 0\}$	$\{0.0022, 0.9602\}$	$\{0.1028, 0.0215\}$	$\{-0.0745, 0.0961\}$
Kendall	$\{-0.9456, 0\}$	$\{-0.8072, 0\}$	$\{0.0071, 0.8136\}$	$\{0.0919, 0.0021\}$	$\{-0.0350, 0.2419\}$
Blomqvist	$\{-0.9600, 0\}$	$\{-0.6160, 0\}$	$\{0.0320, 0.4209\}$	$\{0.0720, 0.0891\}$	$\{0.0640, 0.1283\}$
Hoeffding	$\{0.8679, 0\}$	$\{0.5302, 0\}$	$\{0.0472, 0\}$	$\{0.1902, 0\}$	$\{0.0104, 0\}$
Pearson	$\{-0.9964, 0\}$	$\{-0.8207, 0\}$	$\{0.0202, 0.6529\}$	$\{0.0555, 0.2152\}$	$\{-0.0092, 0.8377\}$

Hoeffding's statistic strongly rejects the null hypothesis of independence since the p-values are all virtually equal to zero. This correctly indicates that, in all five cases, the variables are functionally related.

As anticipated, Pearson's correlation coefficient is larger in absolute value in the case of a linear relationship (data set A) with a value of -0.9964, than in the case of a monotonic relationship (data set B) with a value of -0.8207.

Spearman's, Kendall's and Blomqvist's statistics readily detect the monotonic relationships that data sets A and B exhibit. Interestingly, in the case of data set D, which happens to be monotonically increasing and then decreasing, at the 5% significance level, both Spearman's and Kendall's statistics manage to reject the independence assumption.

7. Multivariate Measures of Association

7.1. Blomqvist's β

Consider the random vector $(X_1, X_2, \ldots, X_n)$ whose joint distribution function is $F(\mathbf{x}) = P(X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n)$ and marginal continuous distribution functions are $F_i(x_i) = P(X_i \le x_i)$ for $x_i \in \mathbb{R}$, $i \in \{1, 2, \ldots, n\}$. We now state Sklar's Theorem for the multivariate case:

Let F be an n-dimensional continuous distribution function with continuous marginal distribution functions $(F_1, F_2, \ldots, F_n)$. Then, there exists a unique n-copula $C : \mathbb{1}^n \to \mathbb{1}$ such that

$$F(x_1, x_2, \ldots, x_n) = C(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)). \tag{69}$$

Conversely, if C is an n-copula and $F_1, F_2, \ldots, F_n$ are continuous distribution functions, then the function F is an n-dimensional distribution function with marginal distribution functions $(F_1, F_2, \ldots, F_n)$ [5] (Theorem 2.10.9, p. 46).

Clearly, the copula $C(\cdot)$ in Equation (69) is the joint distribution function of the random variables $U_i = F_i(x_i)$, $i \in \{1, 2, \ldots, n\}$. Observe that $C(\mathbf{u}) = P(\mathbf{U} \le \mathbf{u}) = F(F_1^{-1}(u_1), F_2^{-1}(u_2), \ldots, F_n^{-1}(u_n))$ for all $\mathbf{u} = (u_1, \ldots, u_n) \in \mathbb{1}^n$.

Letting $W^n(\mathbf{u}) = \max(u_1 + u_2 + \cdots + u_n - n + 1, 0)$ and $M^n(\mathbf{u}) = \min(u_1, u_2, \ldots, u_n)$, the *Fréchet–Hoeffding inequality*,

$$W^n(\mathbf{u}) \le C(\mathbf{u}) \le M^n(\mathbf{u}), \tag{70}$$

provides lower and upper bounds for any copula. This inequality is attributed to [26,35]. We note that a related result appeared in [36].

The Fréchet–Hoeffding upper bound is a copula when the random variables are perfectly positively dependent, i.e., they are comonotonic. However, the lower bound is a copula only in the bivariate case [8].

Blomqvist's β, as given in Equation (52), can also be expressed as

$$\beta = \frac{C(1/2, 1/2) - \Pi(1/2, 1/2) + \bar{C}(1/2, 1/2) - \bar{\Pi}(1/2, 1/2)}{M(1/2, 1/2) - \Pi(1/2, 1/2) + \bar{M}(1/2, 1/2) - \bar{\Pi}(1/2, 1/2)}. \tag{71}$$

where $\Pi^n(\mathbf{u}) = u_1 u_2 \cdots u_n$, and the survival function $\bar{C}(\mathbf{u}) = P(\mathbf{U} > \mathbf{u})$. When C is a copula involving n random variables, Equation (71) can be generalized as follows:

$$\begin{aligned} \beta &= \frac{C(\mathbf{1/2}) - \Pi(\mathbf{1/2}) + \bar{C}(\mathbf{1/2}) - \bar{\Pi}(\mathbf{1/2})}{M(\mathbf{1/2}) - \Pi(\mathbf{1/2}) + \bar{M}(\mathbf{1/2}) - \bar{\Pi}(\mathbf{1/2})} \\ &= k_n \left(C(\mathbf{1/2}) + \bar{C}(\mathbf{1/2}) - 2^{1-n} \right), \end{aligned} \tag{72}$$

where $\mathbf{1/2} = (1/2, 1/2, \ldots, 1/2)$, $k_n = \frac{2^{n-1}}{2^{n-1}-1}$, $\Pi(\mathbf{1/2}) = 2^{-n}$, and $M(\mathbf{1/2}) = 1/2$. When $n = 2$, one has $C(1/2, 1/2) = \bar{C}(1/2, 1/2)$ for any copula; however, this is not the case for $n \ge 3$. The coefficient β can be interpreted as the normalized average distance between the copula C and the independence copula Π.

Ref. [37] utilized the multivariate Blomqvist measure of dependence to analyze main GDP (gross domestic product) aggregates per capita in the European Union, Germany and Portugal for the period 2008–2019.

7.2. Spearman's ρ

In the bivariate case, Spearman's rank correlation can be expressed as

$$\rho_S = \frac{E(UV) - E(U)E(V)}{\sqrt{\mathrm{Var}(U)\mathrm{Var}(V)}}, \tag{73}$$

where U and V are uniformly distributed, so that $E(U) = E(V) = 1/2$ and $\mathrm{Var}(U) = \mathrm{Var}(V) = 1/12$. As previously established,

$$\begin{aligned}\rho_S &= \frac{\int_0^1 \int_0^1 u\, v\, \mathrm{d}C(u,v) - (1/2)^2}{1/12} \\ &= 12 \int_0^1 \int_0^1 C(u,v)\, \mathrm{d}u \mathrm{d}v - 3,\end{aligned} \tag{74}$$

where C is the joint distribution of (U,V).

Employing the same notation as in the previous section, we now present different versions of Spearman's ρ for the multivariate case:

(i) Kendall (1970) [38]:

$$\rho_{S4} = h_2 (4 \sum_{i<j} (C_2^n)^{-1} \int_{[0,1]^n} C_{ij}(u,v)\, \mathrm{d}u \mathrm{d}v - 1); \tag{75}$$

(ii) Ruymgaart and van Zuijlen (1978) [39]:

$$\rho_{S1} = h_n (2^n \int_{[0,1]^n} C(\mathbf{u})\, \mathrm{d}\mathbf{u} - 1); \tag{76}$$

(iii) Joe (1990) [40]:

$$\rho_{S2} = h_n (2^n \int_{[0,1]^n} \mathbf{\Pi}(\mathbf{u})\, \mathrm{d}C(\mathbf{u}) - 1); \tag{77}$$

(iv) Nelsen (2002) [2]:

$$\rho_{S3} = h_n \{2^{n-1} (\int_{[0,1]^n} C(\mathbf{u})\, \mathrm{d}\mathbf{\Pi}(\mathbf{u}) + \int_{[0,1]^n} \mathbf{\Pi}(\mathbf{u})\, \mathrm{d}C(\mathbf{u}) - 1)\}; \tag{78}$$

where $h_n = \frac{1+n}{2^n - (1+n)}$, $C_2^n = \frac{n!}{2!(n-2)!}$, and $\mathbf{u} = (u_1, u_2, \ldots, u_n)$.

We observe that ρ_{S3} appears in [41] (p. 227) as a measure of average upper and lower orthant dependence, and that ρ_{S4} constitutes the population version of the weighted average pairwise Spearman's rho given in Chapter 6 of [38], where $C_{ij}(u,v)$ is the bivariate marginal copula [6] (p. 22).

As obtained by [41] (p. 228), a lower bound for ρ_{Si}, $i \in \{1,2,3\}$, is given by

$$\frac{2^n - (n+1)!}{n!\{2^n - (n+1)\}} \text{ for } n \geq 2.$$

For $n = 3$, this lower bound is at least equal to $-4/3$, and for $n = 2$, we have $\rho_{S1} = \rho_{S2} = \rho_{S4}$. As noted by [42] (p. 787), the aforementioned lower bound may fail to be the best possible.

Spearman's rank correlation can also be expressed as follows for the bivariate case:

$$\rho_S = \frac{\int_{[0,1]^2} C(u,v)\, \mathrm{d}u \mathrm{d}v - \int_{[0,1]^2} \Pi(u,v)\, \mathrm{d}u\, \mathrm{d}v}{\int_{[0,1]^2} M(u,v)\, \mathrm{d}u\, \mathrm{d}v - \int_{[0,1]^2} \Pi(u,v)\, \mathrm{d}u \mathrm{d}v} \tag{79}$$

$$= \frac{\int_{[0,1]^2} uv\, \mathrm{d}C(u,v) - \int_{[0,1]^2} uv\, \mathrm{d}\Pi(u,v)}{\int_{[0,1]^2} uv\, \mathrm{d}M(u,v) - \int_{[0,1]^2} uv\, \mathrm{d}\Pi(u,v)}, \tag{80}$$

where $\int_{[0,1]^2} M(u,v)\, \mathrm{d}u \mathrm{d}v = 1/3$ and $\int_{[0,1]^2} \Pi(u,v)\, \mathrm{d}u \mathrm{d}v = 1/4$. It is readily seen that representation (79) coincides with that given in (74). The coefficient ρ_S can be interpreted as the normalized average distance between the copula C and the independence copula Π.

Equation (79) suggests the following natural generalization for the multivariate case:

$$\rho_S = \frac{\int_{[0,1]^n} C(\mathbf{u})\,\mathrm{d}\mathbf{u} - \int_{[0,1]^n} \Pi(\mathbf{u})\,\mathrm{d}\mathbf{u}}{\int_{[0,1]^n} M(\mathbf{u})\,\mathrm{d}\mathbf{u} - \int_{[0,1]^n} \Pi(\mathbf{u})\,\mathrm{d}\mathbf{u}}, \tag{81}$$

which, incidentally, agrees with the representation specified in Equation (76).

For instance, Liebscher (2021) [43] made use of the multivariate Spearman measure of correlation to determine the dependence of a response variable on a set of regressor variables in a nonparametric regression model.

7.3. Kendall's τ

Joe (1990) [40] provides the following representation of Kendall's τ for the multivariate case:

$$\tau_{n,c} = (2^{n-1} - 1)^{-1}\{2^n \int_{[0,1]^n} C(\mathbf{u})\mathrm{d}C(\mathbf{u}) - 1\}, \tag{82}$$

which also appears in [5] (p. 231) as measure of average multivariate total positivity. In fact, Equation (82) generalizes Equation (44), i.e.,

$$\tau_{2,c} = 4\int_0^1 \int_0^1 C(u_1, u_2)\mathrm{d}C(u_1, u_2) - 1. \tag{83}$$

Nelsen (1996) [41] also notes that a lower bound for $\tau_{n,c}$ is given by

$$-(2^{n-1} - 1)^{-1} \text{ since } \int_{[0,1]^n} C(\mathbf{u})\,\mathrm{d}C(\mathbf{u}) \geq 0. \tag{84}$$

As shown by [42] (Theorem 5.1), and reported by [44] (p. 218), this lower bound is attained if at least one of the bivariate margins of the copula C equals W (the Fréchet–Hoeffding lower bound).

Kendall and Babington Smith (1940) [45] introduced an extension of Kendall's τ as a coefficient of agreement among $n \geq 2$ rankings. Another generalization is proposed in [46].

A test based on the nonparametric estimator of the multivariate extension of Kendall's τ was utilized in [47] to establish links between innovation and higher education in certain regions.

7.4. Hoeffding's Φ^2

Using the same notation as in the previous sections, one can express Hoeffding's dependence index as follows for the bivariate case:

$$\begin{aligned}\Phi^2_{X,Y} &= 90\int_0^1 \int_0^1 \{C(u,v) - uv\}^2 \mathrm{d}u\mathrm{d}v \\ &= 90\int_{[0,1]^2} \{C(u,v) - \Pi(u,v)\}^2 \mathrm{d}u\mathrm{d}v,\end{aligned} \tag{85}$$

where

$$\Phi^2_{X,Y} = \begin{cases} 0 & \text{in the case of stochastic independence,} \\ 1 & \text{in the case of monotone dependence,} \\ (0,1), & \text{otherwise.} \end{cases} \tag{86}$$

Observe that $\Phi^2_{X,Y} = 0$ if and only if $C = \Pi$.

For the multivariate case, Φ^2 is defined as

$$\Phi^2 = h_n \int_{[0,1]^n} \{C(\mathbf{u}) - \Pi(\mathbf{u})\}^2 \mathrm{d}\mathbf{u}, \tag{87}$$

where $h_n = [\int_{[0,1]^n} \{M(\mathbf{u})\,\mathrm{d}\mathbf{u} - \Pi(\mathbf{u})\,\mathrm{d}\mathbf{u}\}^2]^{-1}$ is the normalizing constant.

Gaißer et al. (2010) [30] determined that the inverses of the normalizing constants for the upper and lower bounds are, respectively, given by

$$\begin{aligned}(h_n)^{-1} &= \int_{[0,1]^n} \{M(\mathbf{u})\,\mathrm{d}\mathbf{u} - \Pi(\mathbf{u})\}^2 \mathrm{d}\mathbf{u} \\ &= \frac{2}{(n+1)(n+2)} - \frac{2^{-n} n!}{\prod_{i=0}^{n}(i+1/2)} + (1/3)^n \end{aligned} \tag{88}$$

and

$$\begin{aligned}(g_n)^{-1} &= \int_{[0,1]^n} \{W(\mathbf{u})\,\mathrm{d}\mathbf{u} - \Pi(\mathbf{u})\}^2 \mathrm{d}\mathbf{u} \\ &= \frac{2}{(n+2)!} - 2\sum_{i=0}^{n} C_i^n (-1)^i \frac{1}{(n+1+i)!} + (1/3)^n, \end{aligned} \tag{89}$$

where $C_i^n = \frac{n!}{i!(n-i)!}$.

For instance, Medovikov and Prokhorov (2017) [48] made use of Hoeffding's multivariate index to determine the dependence structure of financial assets and evaluate the risk of contagion.

7.5. Note

As pointed out by [49], Spearman's ρ and Blomqvist's β can be expressed as follows:

$$k_n(C) = \alpha_n \left\{ \int_{[0,1]^n} (C + \sigma^*) \mathrm{d}\mu_n - \frac{1}{2^{n-1}} \right\}, \tag{90}$$

where μ_n is a probability measure on $[0,1]^n$, whereas Kendall's τ has the following representation:

$$\tau_n(C) = \alpha_n \left\{ \int_{[0,1]^n} C\,\mathrm{d}C - \frac{1}{2^n} \right\}, \tag{91}$$

with

$$\alpha_n = \frac{(1+n)2^{n-1}}{2^n - (1+n)}, \text{ for Spearman's rho,}$$

$$\alpha_n = \frac{2^{n-1}}{2^{n-1} - 1}, \text{ for Blomqvist's beta,}$$

and

$$\alpha_n = \frac{2^n}{2^{n-1} - 1}, \text{ for Kendall's tau.}$$

8. Conclusions

Bivariate and multivariate measures of dependence originally due to Spearman, Kendall, Blomqvist and Hoeffding, as well as related results of interest such as their sample estimators and representations in terms of copulas, were discussed in this paper. Various recent applications were also pointed out. Additionally, a numerical study corroborated the effectiveness of these coefficients of association in assessing dependence with respect to five sets of generated data exhibiting various patterns.

A potential avenue for future research would consist in studying matrix-variate rank correlation measures as was achieved very recently by [50] for the case of Kendall's τ.

Author Contributions: Conceptualization, M.A. and S.B.P.; methodology, M.A. and S.B.P.; software, S.B.P. and Y.Z.; validation, M.A. and S.B.P.; formal analysis, M.A., S.B.P. and Y.Z.; investigation, M.A., S.B.P. and Y.Z.; resources, M.A. and S.B.P.; writing—original draft preparation, M.A. and S.B.P.; writing—review and editing, M.A. and S.B.P.; visualization, S.B.P. and Y.Z.; supervision, M.A. and

S.B.P.; project administration, M.A. and S.B.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Sciences and Engineering Research Council of Canada, grant number R0610A.

Data Availability Statement: No new data were created or analyzed in this study.

Acknowledgments: We would like to express our sincere thanks to two reviewers for their insightful comments and suggestions. The financial support of the Natural Sciences and Engineering Research Council of Canada is gratefully acknowledged by the second author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Sklar, A. Fonctions de répartition à *n* dimensions et leurs marges. *Publ. l'Institut Stat. l'Université Paris* **1959**, *8*, 229–231.
2. Nelsen, R.B. Concordance and copulas: A survey. In *Distributions with Given Marginals and Statistical Modelling*; Cuadras, C.M., Fortiana, J., Rodriguez-Lallena, J.A., Eds.; Kluwer Academic Publishers: London, UK, 2002; pp. 169–177.
3. Deheuvels, P. La fonction de dépendance empirique et ses propriétés, un test non-paramétrique d'indépendance. *Bullettin l'Académie R. Belg.* **1979**, *65*, 274–292. [CrossRef]
4. Joe, H. *Multivariate Models and Dependence Concepts*; Chapman & Hall: Boca Raton, FL, USA, 1997.
5. Nelsen, R.B. *An Introduction to Copulas*, 2nd ed.; Springer: New York, NY, USA, 2006.
6. Cherubini, U.; Fabio, G.; Mulinacci, S.; Romagno, S. *Dynamic Copula Methods in Finance*; John Wiley & Sons: New York, NY, USA, 2012.
7. Kojadinovic, I.; Yan, J. Comparison of three semiparametric methods for estimating dependence parameters in copula models. *Math. Econ.* **2010**, *47*, 52–63. [CrossRef]
8. Schmid, F.; Schmidt, R. Nonparametric inference on multivariate versions of Blomqvist's beta and related measures of tail dependence. *Metrika* **2007**, *66*, 323–354. [CrossRef]
9. Genest, C.; Ghoudi, K.; Rivest, L.-P. A semiparametric estimation procedure of dependence parameters in multivariate families of distribution. *Biometrika* **1995**, *82*, 543–552. [CrossRef]
10. Dias, A. Maximum pseudo-likelihood estimation in copula models for small weakly dependent samples. *arXiv* **2022**, arXiv:2208.01322.
11. Kerman, J. A closed-form approximation for the median of the beta distribution. *arXiv* **2011**, arXiv:1111.0433.
12. Schweizer, B.; Wolff, E.F. On nonparametric measures of dependence for random variables. *Ann. Stat.* **1981**, 9, 879–885. [CrossRef]
13. Rényi, A. On measures of dependence. *Acta Math. Hung.* **1959**, *10*, 441–451. [CrossRef]
14. Lai, C.D.; Balakrishnan, N. *Continuous Bivariate Distributions*, 2nd ed.; Springer: New York, NY, USA, 2009.
15. Gibbons, J.D.; Chakraborti, S. *Nonparametric Statistical Inference*, 4th ed.; Revised and Expanded; Statistics Textbooks and Monographs; Marcel Dekker, Inc.: New York, NY, USA, 2003; Volume 168.
16. Hofert, M.; Kojadinovic, I.; Mächler, M.; Yan, J. *Elements of Copula Modeling with R*; Springer: New York, NY, USA, 2019.
17. Kendall, M.G. A new measure of rank correlation. *Biometrika* **1938**, *30*, 81–93. [CrossRef]
18. Kruskal, W.H. Ordinal measures of association. *J. Am. Stat. Assoc.* **1958**, 53, 814–861. [CrossRef]
19. Kendall, M.G.; Gibbons, J.D. *Rank Correlation Methods*, 5th ed.; Griffin: London, UK, 1990.
20. Lipps, G.F. Die Betimmung der Abhängigkeit zwischen den Merkmalen eines Gegestándes. *Berichte Königlich Säsischen Ges. Wiss.* **1905**, *57*, 1–32.
21. Lipps, G.F. *Die Psychischen Massmethoden*; (No. 10); F. Vieweg und Sohn: Braunschweig, Germany, 1906.
22. Fechner, G.T. *Kollektivmasslehre*; Engelmann: Kogarah, Australia, 1897.
23. Blomqvist, N. On a measure of dependence between two random variables. *Ann. Math. Stat.* **1950**, *21*, 593–600. [CrossRef]
24. Genest, C.; Carabarín-Aguirre, A.; Harvey, F. Copula parameter estimation using Blomqvist's beta. *J. Société Française Stat.* **2013**, *154*, 5–24.
25. Mosteller, F. On some useful "inefficient" statistics. *Ann. Math. Stat.* **1946**, *17*, 377–408. [CrossRef]
26. Hoeffding, W. Masstabinvariante Korrelationstheorie. *Schriften Math. Instituts Instituts Angew. Math. Universität Berlin* **2012**, *5*, 181–233.
27. Lehmann, E.L. Some concepts of dependence. *T Ann. Math. Stat.* **1966**, *37*, 1137–1153. [CrossRef]
28. Jogdeo, K. Characterizations of independence in certain families of bivariate and multivariate distributions. *Ann. Math. Stat.* **1968**, *39*, 433–441. [CrossRef]
29. Block, H.W.; Fang, Z. A multivariate extension of Hoeffding's lemma. *Ann. Probab.* **1988**, *16*, 1803–1820. [CrossRef]
30. Gaißer, S.; Ruppert, M.; Schmid, F. A multivariate version of Hoeffding's phi-square. *J. Multivar. Anal.* **2010**, *101*, 2571–2586. [CrossRef]
31. Gänssler, P.; Stute, W. *Seminar on Empirical Processes;* DMV Seminar 9; Birkhäuser: Basel, Switzerland, 1987.
32. Van der Vaart, A.W.; Wellner, J.A. *Weak Convergence and Empirical Processes*; Springer: New York, NY, USA, 1996.
33. Tsukahara, H. Semiparametric estimation in copula models. *Can. J. Stat.* **2005**, *33*, 357–375. [CrossRef]

34. Blum, J.R.; Kiefer, J.; Rosenblatt, M. Distribution free tests of independence based on the sample distribution function. *Ann. Math. Stat.* **1961**, *32*, 485–498. [CrossRef]
35. Fréchet, M. Sur les tableaux de corrélations dont les marges sont données. *Ann. L'Université Lyon* **1951**, *9*, 53–77.
36. Fréchet, M. Généralisations du théorème des probabilités totales. *Fundam. Math.* **1935**, *25*, 379–387. [CrossRef]
37. Ferreira, H.; Ferreira, M. Multivariate medial correlation with applications. *Depend. Model.* **2020**, *8*, 361–372. [CrossRef]
38. Kendall, M.G. *Rank Correlation Methods*; Griffin: London, UK, 1970.
39. Ruymgaart, F.H.; van Zuijlen, M.C.A. Asymptotic Normality of Multivariate Linear Rank Statistics in the Non-I.I.D. Case. *Ann. Stat.* **1978**, *6*, 588–602. [CrossRef]
40. Joe, H. Multivariate Concordance. *J. Multivar. Anal.* **1990**, *35*, 12–30. [CrossRef]
41. Ne1sen, R.B. Nonparametric measures of multivariate association. In *Distributions with Fixed Marginals and Related Topics*; Rüschendorf, L., Schweizer, B., Taylor, M.D., Eds.; Institute of Mathematical Statistics: Hayward, CA, USA, 1996; pp. 223–232.
42. Úbeda-Flores, M. Multivariate versions of Blomqvist's Beta and Spearman's footrule. *Ann. Inst. Stat. Math.* **2005**, *57*, 781–788. [CrossRef]
43. Liebscher, E. On a multivariate version of Spearman's correlation coefficient for regression: Properties and Applications. *Asian J. Stat. Sci.* **2021**, *1*, 123–150.
44. Schmid, F.; Schmidt, R.; Blumentritt, T.; Gaißer, S.; Ruppert, M. Copula-Based Measures of Multivariate Association in Copula Theory and Its Applications. In *Lecture Notes in Statistics*; Jaworski, P., Durante, F., Härdle, W., Rychlik, T. Eds.; Springer: Berlin, Germany, 2010; Volume 198.
45. Kendall, M.G.; Babington Smith, B. On the method of paired comparisons. *Biometrika* **1940**, *31*, 324–345. [CrossRef]
46. Genest, C.; Nešlehová, J.; Ghorbal, N.B. Estimators based on Kendall's tau in multivariate copula models. *Aust. N. Z. J. Stat.* **2011**, *53*, 157–177. [CrossRef]
47. Ascorbebeitia, J.; Ferreira, E.; Orbe, S. Testing conditional multivariate rank correlations: The effect of institutional quality on factors influencing competitiveness. *TEST* **2022**, *31*, 931–949. [CrossRef]
48. Medovikov, I.; Prokhorov, A. A new measure of vector dependence, with applications to financial risk and contagion. *J. Financ. Econom.* **2017**, *15*, 474–503. [CrossRef]
49. Taylor, M.D. Multivariate measures of concordance for copulas and their marginals. *Depend. Model.* **2016**, *4*, 224–236. [CrossRef]
50. McNeil, A.J.; Nešlehová, J.G.; Smith, A.D. On Attainability of Kendall's Tau Matrices and Concordance Signatures. *J. Multivar. Anal.* **2022**, *191*, 105033. [CrossRef]