On the specification of multivariate association measures and their behaviour with increasing dimension

Irène Gijbels[a,*], Vojtěch Kika[a,b], Marek Omelka[b]

[a] Department of Mathematics and Leuven Statistics Research Center (LStat), KU Leuven, Celestijnenlaan 200B, 3001 Leuven, Belgium
[b] Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Charles University, Sokolovská 83, 186 75 Prague, Czech Republic

Abstract

In this paper the interest is to elaborate on the generalization of bivariate association measures, namely Spearman's rho, Kendall's tau, Blomqvist's beta and Gini's gamma, for a general dimension $d \geq 2$. Desirable properties and axioms for such generalizations are discussed, where special attention is given to the impact of the addition of: (i) an independent random variable to a random vector; (ii) a conical combination of all components; (iii) a set of arbitrary random components. Existing generalizations are evaluated with respect to the axiom set. For a d-variate Gini's gamma, a simplified formula is developed, making its analytical computation easier. Further, for Archimedean and meta-elliptical copulas the asymptotic behaviour when the dimension d increases is studied. Nonparametric estimation of the considered generalizations of multivariate association measures is reviewed and a nonparametric estimator of the multivariate Gini's gamma is introduced. The practical use of multivariate association measures is illustrated on a real data example.

Keywords: Archimedean copulas, Association measures, Copulas, Meta-elliptical copulas, Nonparametric estimation.
2010 MSC: Primary 62Hxx, Secondary 62Gxx, 60Exx.

1. Introduction

We have a d-variate random vector $\boldsymbol{X} = (X_1, \ldots, X_d)^\top$ and we want to study the association between its components. More specifically, we are interested in the tendency of the components to simultaneously take large or small values. Originally, association measures were only explored for pairs of random variables. Bivariate association measures were introduced by [3, 9, 12, 30], among others. Later on, attempts to measure dependence within random vectors of general length $d \geq 2$ arose. Some early references include [11] who expanded Spearman's, Blomqvist's and Kendall's bivariate association measures. With increasing popularity of copula theory, the latter became the main tool to study dependence and thus further multivariate association measures were introduced as functionals of a copula. These generalizations are to be found in works of [2, 16, 31], among others. The latter paper, as well as [27] also discussed properties to be expected from a reasonable multivariate association measure. The behaviour of some multivariate association measures in dimension growing to infinity was studied in an Archimedean copula setting by [36].

In this paper, we first elaborate further on the set of desirable properties, axioms, for multivariate association measures and comment on whether these hold for selected multivariate generalizations of bivariate association measures. Special attention is given to a situation when an independent random variable is added to the random vector. One would expect that this would lead to a decrease of an overall association of the starting d-dimensional random vector. We show that not all the generalizations behave as expected despite their analogous way of derivation. For one of the generalizations of Gini's gamma, we provide a simplified formula which avoids d-dimensional integration and makes

[*] Corresponding author. Email address: `irene.gijbels@kuleuven.be`

this measure thus more computationally feasible. Further when duplicating one component of $\boldsymbol{X}$ (or more generally adding a conical combination of all components) one would expect the extended random vector to show an increasing association. We establish for which multivariate association measures such a property holds. Finally, we extend Wysocki's results on asymptotic properties in terms of growing dimension (see [36]) to other measures. In addition we obtain some partial results in this matter for meta-elliptical copulas.

The organization of this paper is as follows. In Section 2, a discussion about a set of axioms to be fulfilled by multivariate association measures is provided. In Section 3 several examples of multivariate association measures are briefly reviewed together with the investigation of the validity of the axioms. Section 4 looks into further properties of the multivariate association measures, including the evolvement for increasing dimension. In Section 5, the limiting behaviour when the dimension tends to infinity is investigated for Archimedean and meta-elliptical copulas. The results for Archimedean copulas complement those of [36]. Section 7 is devoted to estimation of multivariate association measures. An illustrative example is in Section 6, whereas a real data application in Section 8 demonstrates the practical use of the association measures. Section 9 provides an overview of our findings. A study on the multivariate Blomqvists's beta, further illustrative examples and explanations on how to obtain standard errors, as well as additional material on the real data application and the overview are provided in the Supplementary Material.

2. Multivariate copulas and axioms for multivariate association measures

We first introduce necessary tools from copula theory, and next get to a list of axioms for multivariate association measures. In Section 3 we then focus on the discussion of several multivariate measures, in view of the axioms.

2.1. Multivariate copulas

Suppose we have a d-variate random vector $\boldsymbol{X} = (X_1, \ldots, X_d)^\top$ where F is the joint distribution function of $\boldsymbol{X}$ and F_i, $i \in \{1, \ldots, d\}$, are the continuous marginal distribution functions of X_i, $i \in \{1, \ldots, d\}$. Applying Sklar's theorem [29] in higher dimensions, there exists a unique copula function $C_d : [0, 1]^d \to [0, 1]$ such that

$$F(x_1, \ldots, x_d) = C_d(F_1(x_1), \ldots, F_d(x_d)), \quad (x_1, \ldots, x_d)^\top \in \mathbb{R}^d.$$

We denote the set of all d-variate copulas as $\mathrm{Cop}(d)$. Copula C_d is the joint distribution function of the random vector $\boldsymbol{U} = (U_1, \ldots U_d)^\top = (F_1(X_1), \ldots, F_d(X_d))^\top$, that is, with $\boldsymbol{u} = (u_1, \ldots, u_d)^\top \in [0, 1]^d$, $C_d(\boldsymbol{u}) = \Pr(\boldsymbol{U} \leq \boldsymbol{u})$ where inequalities of vectors are understood component-wise. Further, we define the survival function $\overline{K}$ associated to a measurable function $K : [0, 1]^d \to [0, 1]$ as

$$\overline{K}(\boldsymbol{u}) = 1 + \sum_{k=1}^{d} (-1)^k \sum_{1 \leq i_1 < \cdots < i_k \leq d} K_{i_1,\ldots,i_k}(u_{i_1}, \ldots, u_{i_k}) \tag{1}$$

where $K_{i_1,\ldots,i_k}$ denotes the corresponding k-dimensional margin. The way to calculate $\overline{K}$ in (1) will be further referred to as the inclusion-exclusion principle. Since C_d is a distribution function, then $\overline{C}_d(\boldsymbol{u}) = \Pr(\boldsymbol{U} > \boldsymbol{u})$. For the bivariate case ($d = 2$) the survival function of C_2 is $\overline{C}_2(u_1, u_2) = \Pr(U_1 > u_1, U_2 > u_2) = 1 - C_2(u_1, 1) - C_2(1, u_2) + C_2(u_1, u_2)$, for $(u_1, u_2)^\top \in [0, 1]^2$. The survival copula C_d^S is defined as the copula of $\mathbf{1} - \boldsymbol{U}$, that is

$$C_d^S(\boldsymbol{u}) = \Pr(\mathbf{1} - \boldsymbol{U} \leq \boldsymbol{u}) = \Pr(\boldsymbol{U} > \mathbf{1} - \boldsymbol{u}) = \overline{C}_d(\mathbf{1} - \boldsymbol{u}).$$

If C_d and C_d^S coincide, we call C_d radially symmetric. In such case, it follows from the previous equation that

$$\overline{C}_d(\boldsymbol{u}) = C_d(\mathbf{1} - \boldsymbol{u}).$$

We say that copula C_d has a density $c_d : [0, 1]^d \to \mathbb{R}_0^+$ defined as

$$c_d(\boldsymbol{u}) = \frac{\partial^d C_d(\boldsymbol{u})}{\partial u_1 \ldots \partial u_d},$$

if it exists.

The range of values a multivariate copula can take, is restricted by the so-called Fréchet's bounds, namely

$$W_d(\boldsymbol{u}) \leq C_d(\boldsymbol{u}) \leq M_d(\boldsymbol{u}), \quad \forall \boldsymbol{u} \in [0,1]^d,$$

where $W_d(\boldsymbol{u}) = \max(\sum_{i=1}^d u_i - d + 1, 0)$ is the lower Fréchet's bound and copula $M_d(\boldsymbol{u}) = \min(u_i; i \in \{1,\dots,d\})$ is the upper Fréchet's bound. For $d > 2$, however, W_d fails to be a copula, unlike M_d. The copula M_d is called the comonotonicity copula, since it is the copula of a vector $(X_1, g_2(X_1), \dots, g_d(X_1))^\top$ where $g_2, \dots, g_d$ are strictly increasing functions on the support of X_1. For $d = 2$, the lower Fréchet's bound W_2 is a copula and is known as the countermonotonicity copula since it is the copula of a vector $(X_1, X_2)^\top = (X_1, g_1(X_1))^\top$ where g_1 is a strictly decreasing function on the support of X_1. Among the properties of multivariate copulas, is that they are invariant with respect to strictly increasing transformations of the components of $\boldsymbol{X}$. The independence copula $\Pi_d(\boldsymbol{u}) = \prod_{i=1}^d u_i$ corresponds to mutually independent $X_1, \dots, X_d$. Both M_d and Π_d are radially symmetric copulas.

In the multivariate setting, one defines two types of ordering for copulas $A_d, B_d \in \mathrm{Cop}(d)$:

$$\begin{aligned} A_d \leq B_d &\Leftrightarrow \forall \boldsymbol{u} \in [0,1]^d : A_d(\boldsymbol{u}) \leq B_d(\boldsymbol{u}), \\ A_d \leq_C B_d &\Leftrightarrow \forall \boldsymbol{u} \in [0,1]^d : A_d(\boldsymbol{u}) \leq B_d(\boldsymbol{u}) \text{ and } \overline{A}_d(\boldsymbol{u}) \leq \overline{B}_d(\boldsymbol{u}), \end{aligned}$$

being called order and concordance order, respectively. Note that in the bivariate case, the concordance ordering always follows from the ordering itself, since for any copula $C_2 \in \mathrm{Cop}(2)$ it holds that $\overline{C}_2(\boldsymbol{u}) = 1 - u_1 - u_2 + C_2(\boldsymbol{u})$, since $C_2(u_1, 1) = u_1$ and $C_2(1, u_2) = u_2$. Hence for any $A_2, B_2 \in \mathrm{Cop}(2)$ it holds that $A_2 \leq B_2$ implies $\overline{A}_2 \leq \overline{B}_2$.

Taylor [31] introduced a concept of reflections of the d-dimensional unit cube $[0,1]^d$ which happens to be very useful for notation in this context. We say that a mapping $\xi : [0,1]^d \to [0,1]^d$ is a reflection if $\xi(\boldsymbol{u}) = \boldsymbol{v}$ where for $i \in \{1,\dots,d\}$ we have $v_i = u_i$ or $v_i = 1 - u_i$. The set of all d-dimensional reflections is denoted as $\mathcal{R}_d$. An important example of reflection, for $i \in \{1,\dots,d\}$, is an elementary reflection σ_i defined as $\sigma_i(\boldsymbol{u}) = \boldsymbol{v}$ with $v_i = 1 - u_i$ and $v_j = u_j$ for $j \neq i$. Reflection with respect to all components will be denoted as σ, that is $\sigma(\boldsymbol{u}) = \boldsymbol{1} - \boldsymbol{u}$. Reflections can be used to produce new copulas, as shown by [31]. To every copula C_d we can associate the probability measure μ_{C_d} satisfying

$$C_d(u_1, \dots, u_d) = \mu_{C_d}([0, u_1] \times \cdots \times [0, u_d]) = \Pr(U_1 \leq u_1, \dots, U_d \leq u_d).$$

Then for $\xi \in \mathcal{R}_d$ we can define a new copula C_d^ξ via its probability measure given as

$$\mu_{C_d^\xi}(S) = \mu_{C_d}(\xi(S))$$

for S a Borel set of $[0,1]^d$ where reflection applied to a set is understood, for example, as in

$$\sigma_1([0, u_1] \times \cdots \times [0, u_d]) = [1 - u_1, 1] \times [0, u_2] \times \cdots \times [0, u_d].$$

Analogously, we can consider permutations that correspond to changing the order of the components within a random vector.

Taylor [31] further shows that if C_d is the copula of $\boldsymbol{X}$, then C_d^σ is the copula of $-\boldsymbol{X}$, thus also of its (component-wise) strictly increasing transformations $\boldsymbol{1} - \boldsymbol{X}$ and $\boldsymbol{1} - \boldsymbol{U}$. This also means that $C_d^\sigma = C_d^S$, the survival copula associated to C_d.

The copula of any subvector of $\boldsymbol{X}$ of length between 1 and d is called a marginal copula of C_d. Note that by allowing for subvectors of length d, a copula is also a marginal copula of itself. Marginal copulas can be easily expressed from the original copula by setting arguments corresponding to unselected components to 1. Suppose for simplicity that the interest is in the marginal copula $C_{d-1}^{(-i)}$ of $(X_1, \dots, X_{i-1}, X_{i+1}, \dots X_d)^\top$ for $i \in \{1,\dots,d\}$. Then

$$C_{d-1}^{(-i)}(u_1, \dots, u_{i-1}, u_{i+1}, \dots u_d) = C_d(u_1, \dots, u_{i-1}, 1, u_{i+1}, \dots u_d),$$

where C_d is the copula of $(X_1, \dots, X_{i-1}, X_i, X_{i+1}, \dots X_d)^\top$. For lower-dimensional marginal copulas, one could simply iterate this procedure.

Throughout this paper, we often focus on consequences of increasing dimension d. When talking about a sequence of copulas $\{C_d\}_{d=2}^{\infty}$ we understand that these copulas are linked together in the following way. Suppose that $X_1, X_2, \ldots$ are random variables. Then let C_d be the copula of the random vector $(X_1, \ldots, X_d)^\top$. This construction ensures that C_{d_1} is always a marginal copula of C_{d_2} for $d_1 < d_2$ and also that

$$C_{d_1}(u_1, \ldots, u_{d_1}) = C_{d_2}(u_1, \ldots, u_{d_1}, 1, \ldots, 1).$$

One important class of copulas is the class of multivariate Archimedean copulas, thoroughly discussed for example by [14], and for which the definition is recalled in Section S1 of the Supplementary Material. For Archimedean copulas, a sequence $\{C_d\}_{d=2}^{\infty}$ is understood as a sequence of Archimedean copulas sharing a common generator ψ but differing in dimension. That means that if we know that C_d is an Archimedean copula, we can extract its generator and use it to construct C_{d+1} in a unique way. [14] shows that so-called d-monotonicity of ψ is a necessary and sufficient condition for an Archimedean generator to generate some Archimedean copula. See Section S1 for the definition of d-monotonicity.

2.2. *Axioms for multivariate association measures*

For the bivariate case, Rényi [21] introduced a set of axioms for a dependence measure. Later on Scarsini [23] reformulated a set of axioms for what he called a measure of concordance. These axioms can be translated in terms of copulas. Since we do not solely think of associations in terms of concordance or discordance of pairs, we will use the more general term of association measures. Suppose that C_2 is the copula of $(X_1, X_2)^\top$ and denote the copulas of the vectors $(-X_1, X_2)^\top, (X_1, -X_2)^\top$ and $(X_2, X_1)^\top$ by $C_2^{\sigma_1}, C_2^{\sigma_2}$ and C_2^{π}, respectively. Further suppose that $C_{2,m} \in \mathrm{Cop}(2)$ for $m = 1, 2, \ldots$ constitute a sequence of copulas. For a bivariate association measure $\kappa_2 : \mathrm{Cop}(2) \to \mathbb{R}$, one expects the following axioms to hold.

(S_1) (*Normalization*) $\kappa_2(M_2) = 1, \kappa_2(\Pi_2) = 0$.

(S_2) (*Continuity*) If $\lim_{m\to\infty} C_{2,m}(u_1, u_2) = C_2(u_1, u_2), \forall (u_1, u_2)^\top \in [0, 1]^2$, then $\lim_{m\to\infty} \kappa_2(C_{2,m}) = \kappa_2(C_2)$.

(S_3) (*Permutation invariance*) $\kappa_2(C_2^{\pi}) = \kappa_2(C_2)$.

(S_4) (*Ordering*) If $C_{2,1}(u_1, u_2) \le C_{2,2}(u_1, u_2), \forall (u_1, u_2)^\top \in [0, 1]^2$, then $\kappa_2(C_{2,1}) \le \kappa_2(C_{2,2})$.

(S_5) (*Reflection principle*) $\kappa_2(C_2^{\sigma_1}) = \kappa_2(C_2^{\sigma_2}) = -\kappa_2(C_2)$.

Although one can find different sets of axioms for bivariate association measures, Scarsini's axioms appear frequently in the literature and hence we use them as a starting point for discussing possible extensions of association measures into general $d \ge 2$ dimension. We focus on generalizations of bivariate association measures that all can be expressed as functionals of the underlying copula.

We comment on how and why to select a set of axioms for a multivariate association measure κ_d such that these axioms also guarantee fulfilment of Scarsini's axioms (S_1) – (S_5) when restricting to dimension $d = 2$.

Looking back at axioms (S_1) – (S_5), the three required properties (S_1) – (S_3) (normalization, continuity, permutation invariance) are obvious to generalize, only the dimension is changed. For ordering, concordance order becomes necessary. Recall that the condition (S_4) alone in the bivariate setting would lead to having $\forall \boldsymbol{u} \in [0, 1]^d : \Pr_A(\boldsymbol{U} \le \boldsymbol{u}) = C_A(\boldsymbol{u}) \le C_B(\boldsymbol{u}) = \Pr_B(\boldsymbol{U} \le \boldsymbol{u})$ where the subscript of Pr relates to the copula of $\boldsymbol{U}$. Thus one can imagine that for copula B, its probability mass is more concentrated close to $(0, \ldots, 0)^\top$ than for copula A. So for B values more tend to be all simultaneously smaller. However, we equally focus on detecting simultaneously large values and thus we additionally require $\forall \boldsymbol{u} \in [0, 1]^d : \Pr_A(\boldsymbol{U} > \boldsymbol{u}) = \overline{C}_A(\boldsymbol{u}) \le \overline{C}_B(\boldsymbol{u}) = \Pr_B(\boldsymbol{U} > \boldsymbol{u})$. In short, the order of copulas A and B in (S_4) is replaced by concordance order for an ordering axiom in general dimension. Recall that for the bivariate case order and concordance order coincide.

For a generalization of the reflection principle (S_5), [31] proposes two conditions, namely

$$\sum_{\xi \in \mathcal{R}_d} \kappa_d(C_d^{\xi}) = 0 \quad \text{and} \quad \kappa_d(C_d) = \kappa_d(C_d^{\sigma}),$$

recalling that $\mathcal{R}_d$ are all d-dimensional reflections. For arguments for this generalization, we refer to Section 3 in [31]. In the multivariate setting [31] adds another axiom to the generalization of the set $(S_1) - (S_5)$ stating that the sum of association measures for vectors $(X_1, X_2, \dots, X_d)^\top$ and $(-X_1, X_2, \dots, X_d)^\top$ does not depend on X_1 and can be extracted from the association measure of $(X_2, \dots, X_d)^\top$.

We can now summarize these axioms while also adding one axiom not mentioned above. We say that $\kappa_d : \text{Cop}(d) \to \mathbb{R}$ is a multivariate association measure in the sense of axioms $(A_1) - (A_8)$ if it satisfies the following conditions for any d-variate copulas C_d and $C_{d,m}$, $m \in \{1, 2, \dots\}$.

Multivariate axioms

(A_1) (Normalization) $\kappa_d(M_d) = 1, \kappa_d(\Pi_d) = 0$.

(A_2) (Continuity) If $\lim_{m\to\infty} C_{d,m}(\boldsymbol{u}) = C_d(\boldsymbol{u}), \forall \boldsymbol{u} \in [0,1]^d$, then $\lim_{m\to\infty} \kappa_d(C_{d,m}) = \kappa_d(C_d)$.

(A_3) (Permutation invariance) $\kappa_d(C_d^\pi) = \kappa_d(C_d)$ for every permutation π.

(A_4) (Ordering) If $C_{d,1} \preceq_C C_{d,2}$, then $\kappa_d(C_{d,1}) \leq \kappa_d(C_{d,2})$.

(A_5) (Duality) $\kappa_d(C_d^\sigma) = \kappa_d(C_d)$.

(A_6) (Reflection principle) $\sum_{\xi \in \mathcal{R}_d} \kappa_d(C_d^\xi) = 0$.

(A_7) (Transition property) There exists a constant r_{d-1} such that

$$\kappa_d(C_d) + \kappa_d(C_d^{\sigma_1}) = r_{d-1}\kappa_{d-1}(C_{d-1}^{(-1)}).$$

(A_8) (Independent component addition) For X_{d+1} independent of $(X_1, \dots, X_d)^\top$

$$\kappa_d(C_d) > \kappa_{d+1}(C_{d+1}) > 0, \quad \text{or} \quad \kappa_d(C_d) < \kappa_{d+1}(C_{d+1}) < 0, \quad \text{or} \quad \kappa_d(C_d) = \kappa_{d+1}(C_{d+1}) = 0.$$

The reasoning behind axiom (A_8) is that addition of an independent component must be reflected by an association measure moving towards zero, if still possible. A similar axiom is mentioned by [27], however we adjust it to ensure its meaning also for negative values of an association measure.

Note that axioms $(A_1) - (A_8)$ do not explicitly state any limitations for values of such a measure of association. However, an upper bound can be deduced using other axioms since we know that $C_d \leq M_d$ and also $C_d^\sigma \leq M_d = M_d^\sigma$ and thus $C_d \preceq_C M_d$ which means that combining (A_1) and (A_4), we get $\kappa_d(C_d) \leq 1$. However, there is no limitation in terms of a lower bound for $d \geq 3$. [31] suggests that there is no obvious proof for $-1 \leq \kappa_d(C_d)$.

Axioms $(A_1) - (A_8)$ are considered as a minimum set of requirements that a reasonable multivariate association measure should satisfy. One can wonder about other (desirable) properties that one would expect for a multivariate association measure to hold. One of the properties sometimes discussed in the literature is that of irreducibility, which says: For every dimension d and every copula C_d the measure $\kappa_d(C_d)$ cannot be written as a function of lower dimensional measures $\{\kappa_j(C_j); C_j \text{ marginal copula}, j \in \{2, \dots, d-1\}\}$. This issue was discussed in Schmid et al. [27, page 215], who also highlighted that there are exceptions and that such a requirement not really has to apply for all copulas. It might be more logical to say that there does not exist a universal function h such that $\kappa(C_d)$ equals the function h of lower dimensional measures $\{\kappa_j(C_j); C_j \text{ marginal copula}, j \in \{2, \dots, d-1\}\}$, for each $C_d \in \text{Cop}(d)$. In other words the measure $\kappa_d(C_d)$ is really d-dimensional. The property of irreducibility is obviously not satisfied for multivariate association measures constructed via the pairwise approach of Section 3.1.

To simplify the notation, the subscript d of κ_d, denoting the dimension, will sometimes be omitted in the sequel of the text, the dimension being clear from an argument of a functional κ.

In Sections 3 and S2 we investigate a variety of multivariate association measures in the light of axioms $(A_1)-(A_8)$. In Section 4 we further study two specific properties (P_1) and (P_2), formulated as follows. Denote C_d the copula of $\boldsymbol{X} = (X_1, \dots, X_d)^\top$.

(P_1) *Duplication of one component (or more generally adding a conical combination of all components).* When duplicating a component of $\boldsymbol{X}$, say putting $X_{d+1} = X_j$, for some $j \in \{1, \ldots, d\}$, and considering the extended random vector $(\boldsymbol{X}^\top, X_{d+1})^\top$ with copula C_{d+1}, one might expect that $\kappa(C_{d+1}) \geq \kappa(C_d)$, i.e. the association in the extended random vector is not smaller than in $\boldsymbol{X}$. The same can be remarked when adding a conical combination, i.e. $X_{d+1} = \sum_{j=1}^{d} \alpha_j X_j$ with $\alpha_j \geq 0$ for all j.

(P_2) *Effect of adding $d_2 - d$ (with $d_2 > d$) arbitrary components.* For the copula C_{d_2} of $(\boldsymbol{X}^\top, X_{d+1}, \ldots, X_{d_2})^\top$ one would like to provide a non-trivial lower and upper bound for $\kappa(C_{d_2})$ involving $\kappa(C_d)$ and/or the number of added components $d_2 - d$.

Property (P_2) sheds light on how a multivariate association measure evolves when the dimension increases. In Section 5 we go one step further and investigate the limiting behaviour $\lim_{d\to\infty} \kappa(C_d)$.

3. Multivariate association measures and verification of axioms

In this section we discuss two main methods of generalizing bivariate association measures to the case of general dimension $d \geq 2$. One approach uses resulting values of bivariate association measures, referred to as the pairwise approach, whereas the other approach exploits their structure. This also means that the former can be formulated for all bivariate measures at once, the latter requires a specific bivariate measure to start with.

3.1. Pairwise approach

Suppose we have a bivariate association measure κ_2 and create a d-variate association measure as an average of all pairwise measures, i.e.

$$\kappa_d^{\mathrm{PW}}(C_d) = \frac{1}{\binom{d}{2}} \sum_{1 \leq i < j \leq d} \kappa_2(C_2^{i,j}) \tag{2}$$

where $C_2^{i,j}$ is the copula of $(X_i, X_j)^\top$. Let us now check whether such measure fulfils our axioms.

Proposition 1. *Let $\kappa_2 : \mathrm{Cop}(2) \to \mathbb{R}$ be a bivariate measure of association in the sense of axioms $(S_1) - (S_5)$. Define for $d \in \{2, 3, \ldots\}$, the measure $\kappa_d^{PW} : \mathrm{Cop}(d) \to \mathbb{R}$ by (2) and set $r_{d-1} = 2(d-2)/d$ in axiom (A_7). Then $\{(\kappa_d^{PW}, r_d)\}_{d=3}^{\infty}$ fulfils axioms (A_1) to (A_8).*

Proof. Validity of axioms (A_1) to (A_7) was proven by [31], hence we only need to pay attention to axiom (A_8). Suppose that X_{d+1} is independent of $(X_1, \ldots, X_d)^\top$. Then using (S_1)

$$\kappa_{d+1}^{\mathrm{PW}}(C_{d+1}) = \frac{1}{\binom{d+1}{2}} \sum_{\substack{i,j=1 \\ i<j}}^{d+1} \kappa_2(C_2^{i,j}) = \frac{1}{\binom{d+1}{2}} \sum_{\substack{i,j=1 \\ i<j}}^{d} \kappa_2(C_2^{i,j}) + \frac{1}{\binom{d+1}{2}} \sum_{i=1}^{d} \kappa_2(C_2^{i,d+1}) = \frac{\binom{d}{2}}{\binom{d+1}{2}} \kappa_d(C_d) = \frac{d-1}{d+1} \kappa_d(C_d)$$

from which (A_8) follows since $(d-1)/(d+1) < 1$. □

Proposition 1 shows that the pairwise approach leads to a ‘reasonable’ multivariate association measure if the initial bivariate measure is ‘reasonable’ as well. Yet there is a disadvantage resulting from the pairwise nature of this construction. Pairwise-constructed measures will always assign value 0 to vectors having pairwise independent components no matter if there is an association of higher order. Indeed, if all pairs of components of a vector are independent, then all values of $\kappa_2(C_2^{i,j})$ in (2) are zero and hence $\kappa_d^{\mathrm{PW}}(C_d) = 0$.

3.2. Copula approach

We now move towards the generalization for a general dimension $d \geq 2$ based on the intrinsic structure of bivariate association measures, which we first recall for convenience of the reader.

3.2.1. Multivariate Spearman's rho

Bivariate Spearman's rho is defined as [see e.g. 18, p. 167]

$$\rho(X_1, X_2) = 3\left(\Pr\{(X_1 - Y_1)(X_2 - Z_2) > 0\} - \Pr\{(X_1 - Y_1)(X_2 - Z_2) < 0\}\right)$$

where $(X_1, X_2)^\top$, $(Y_1, Y_2)^\top$ and $(Z_1, Z_2)^\top$ are independent and identically distributed random vectors with copula C_2. An alternative expression for $\rho(X_1, X_2)$ is

$$\rho(X_1, X_2) = \frac{\text{cov}(F_1(X_1), F_2(X_2))}{\sqrt{\text{var}(F_1(X_1))}\sqrt{\text{var}(F_2(X_2))}} = \frac{\text{cov}(U_1, U_2)}{\sqrt{\text{var}(U_1)}\sqrt{\text{var}(U_2)}}$$

which can be further expressed using the underlying copula as

$$\rho(C_2) = \frac{\int_{[0,1]^2} u_1 u_2 \, \mathrm{d}C_2(u_1, u_2) - (1/2)^2}{\sqrt{(1/12)}\sqrt{(1/12)}} = \frac{\int_{[0,1]^2} \Pi_2(\boldsymbol{u})\, \mathrm{d}C_2(\boldsymbol{u}) - \int_{[0,1]^2} \Pi_2(\boldsymbol{u})\, \mathrm{d}\Pi_2(\boldsymbol{u})}{\int_{[0,1]^2} \Pi_2(\boldsymbol{u})\, \mathrm{d}M_2(\boldsymbol{u}) - \int_{[0,1]^2} \Pi_2(\boldsymbol{u})\, \mathrm{d}\Pi_2(\boldsymbol{u})}, \tag{3}$$

with $\boldsymbol{u} = (u_1, u_2)^\top$, or equivalently using integration by parts as

$$\rho(C_2) = \frac{\int_{[0,1]^2} C_2(\boldsymbol{u})\, \mathrm{d}\Pi_2(\boldsymbol{u}) - \int_{[0,1]^2} \Pi_2(\boldsymbol{u})\, \mathrm{d}\Pi_2(\boldsymbol{u})}{\int_{[0,1]^2} M_2(\boldsymbol{u})\, \mathrm{d}\Pi_2(\boldsymbol{u}) - \int_{[0,1]^2} \Pi_2(\boldsymbol{u})\, \mathrm{d}\Pi_2(\boldsymbol{u})}, \tag{4}$$

as shown in [25]. Spearman's rho can thus be viewed either as Pearson's correlation coefficient of the uniformly distributed U_1 and U_2 or as the standardized average distance between C_2 and Π_2.

Spearman's rho can be generalized in multiple ways for a general dimension $d \geq 2$. Based on (4), a generalization considered firstly by [35] can be defined as

$$\rho_1(C_d) = \frac{\int_{[0,1]^d} C_d(\boldsymbol{u})\, \mathrm{d}\boldsymbol{u} - \int_{[0,1]^d} \Pi_d(\boldsymbol{u})\, \mathrm{d}\boldsymbol{u}}{\int_{[0,1]^d} M_d(\boldsymbol{u})\, \mathrm{d}\boldsymbol{u} - \int_{[0,1]^d} \Pi_d(\boldsymbol{u})\, \mathrm{d}\boldsymbol{u}} = h_\rho(d)\left\{2^d \int_{[0,1]^d} C_d(\boldsymbol{u})\, \mathrm{d}\boldsymbol{u} - 1\right\} \tag{5}$$

with $h_\rho(d) = (d+1)/\{2^d - (d+1)\}$, where we used that [see e.g. 18, p. 225]

$$\int_{[0,1]^d} M_d(\boldsymbol{u})\, \mathrm{d}\boldsymbol{u} = \frac{1}{d+1}, \qquad \int_{[0,1]^d} \Pi_d(\boldsymbol{u})\, \mathrm{d}\boldsymbol{u} = \frac{1}{2^d}.$$

Similarly, starting from (3), we get to another generalization

$$\rho_2(C_d) = h_\rho(d)\left\{2^d \int_{[0,1]^d} \Pi_d(\boldsymbol{u})\, \mathrm{d}C_d(\boldsymbol{u}) - 1\right\}, \tag{6}$$

introduced (without using the copula theory) by [11]. As a third version of a generalization, [17] considered their average $\rho_3 = (\rho_1 + \rho_2)/2$.

Both ρ_1 and ρ_2 satisfy all the axioms except for the duality axiom (A_5) which is satisfied by ρ_3. This is stated by [27] for axioms $(A_1) - (A_7)$. Axiom (A_8) is satisfied for ρ_1, ρ_2 and ρ_3 as is established in the following proposition.

Proposition 2. *Axiom (A_8) is fulfilled by ρ_1, ρ_2 and ρ_3.*

Proof. Let C_d be a d-variate copula of $(X_1, \ldots, X_d)^\top$ and C_{d+1} be a $(d+1)$-variate copula of $(X_1, \ldots, X_d, X_{d+1})^\top$ where X_{d+1} is independent of $(X_1, \ldots, X_d)^\top$, that is

$$C_{d+1}(u_1, \ldots, u_d, u_{d+1}) = C_d(u_1, \ldots, u_d)u_{d+1}.$$

Spearman's rho ρ_1 can be written as

$$\rho_1(C_{d+1}) = \frac{d+2}{2^{d+1}-(d+2)}\left\{2^{d+1}\int_{[0,1]^{d+1}} C_{d+1}(u_1,\ldots,u_d,u_{d+1})\,\mathrm{d}(u_1\cdot\ldots\cdot u_d\cdot u_{d+1}) - 1\right\}$$
$$= \frac{d+2}{2^{d+1}-(d+2)}\left\{2^{d}\int_{[0,1]^{d}} C_{d}(u_1,\ldots,u_d)\,\mathrm{d}(u_1\cdot\ldots\cdot u_d) - 1\right\}$$

and thus we have

$$\rho_1(C_{d+1}) = \frac{\frac{d+2}{2^{d+1}-(d+2)}}{\frac{d+1}{2^{d}-(d+1)}}\rho_1(C_d) = \frac{d2^d + 2^{d+1} - (d+1)(d+2)}{d2^{d+1} + 2^{d+1} - (d+1)(d+2)}\rho_1(C_d)$$

from which the statement for ρ_1 follows since $\{d2^d + 2^{d+1} - (d+1)(d+2)\}/\{d2^{d+1} + 2^{d+1} - (d+1)(d+2)\} < 1$ for every $d \geq 2$.

The proof for ρ_2 mimics the preceding derivations since

$$\int_{[0,1]^d} \Pi_d(\boldsymbol{u})\,\mathrm{d}C_d(\boldsymbol{u}) = \Pr(\boldsymbol{U} < \boldsymbol{V}) = \Pr(\boldsymbol{V} > \boldsymbol{U}) = \int_{[0,1]^d} \overline{C}_d(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u},$$

where $\boldsymbol{U}$ and $\boldsymbol{V}$ are d-variate random vectors with standard uniform margins and with copulas Π_d and C_d, respectively.

The proof for ρ_3 follows immediately from the statement for ρ_1 and ρ_2. □

Spearman's rho is rather difficult to calculate analytically for many standard copulas, in some cases even in dimension $d = 2$. However, for a copula with a closed form, the bivariate Spearman's rho can generally be calculated via numerical integration techniques. For radially symmetric copulas, ρ_1 and ρ_2 coincide.

3.2.2. Multivariate Kendall's tau

Bivariate Kendall's tau is defined as the difference between the probability of concordance and the probability of discordance [see e.g. 18, p. 158]

$$\tau(X_1, X_2) = \Pr\{(X_1 - Y_1)(X_2 - Y_2) > 0\} - \Pr\{(X_1 - Y_1)(X_2 - Y_2) < 0\} \tag{7}$$

where $(X_1, X_2)^\top$ and $(Y_1, Y_2)^\top$ are independent and identically distributed random vectors with copula C_2. This can be further expressed using the underlying copula as, with $\boldsymbol{u} = (u_1, u_2)^\top \in [0,1]^2$, (see Nelsen [18, Theorem 5.1.3])

$$\tau(C_2) = 4\int_{[0,1]^2} C_2(\boldsymbol{u})\,\mathrm{d}C_2(\boldsymbol{u}) - 1,. \tag{8}$$

For general dimension d Kendall's tau can be generalized based on (8) as

$$\tau(C_d) = \frac{1}{2^{d-1}-1}\left\{2^d\int_{[0,1]^d} C_d(\boldsymbol{u})\,\mathrm{d}C_d(\boldsymbol{u}) - 1\right\} \tag{9}$$

and was firstly proposed by [16].

Another class of generalizations was introduced by [11] which is based on the probabilities of concordance and discordance as in (7). Suppose that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent and identically distributed d-variate random vectors with copula C_d. One can define $D_j = X_j - Y_j$ for $j \in \{1,\ldots,d\}$ and consider

$$\tau_2(C_d, \boldsymbol{w}_d) = \sum_{k=0}^{d} w_{d,k}\Pr((D_1,\ldots,D_d)^\top \in B_{k,d-k}) \tag{10}$$

where $B_{k,d-k}$, for $k \in \{0,\ldots,d\}$, is a subset of $\mathbb{R}^d$ having k positive components and $d-k$ negative components. Let us use a convention that $\binom{n}{m} = 0$ if $m > n$. Conditions on the weights $w_{d,k}$ under which axioms $(A_1) - (A_6)$ are fulfiled, are proposed by [11], specifically

(T_1) $w_{d,k} = w_{d,d-k}$,

(T_2) $w_{d,d} \geq w_{d,d-1} \geq \cdots \geq w_{d,k'}, k' = \lfloor \frac{d+1}{2} \rfloor$,

(T_3) $w_{d,0} = w_{d,d} = 1$,

(T_4) $\sum_{k=0}^{d} \binom{d}{k} w_{d,k} = 0$,

(T_5) $w_{d,k} = \sum_{\ell=2}^{d} \beta_\ell \left\{ \binom{k}{\ell} + \binom{d-k}{\ell} - \left[\binom{d}{\ell} - \binom{k}{\ell} - \binom{d-k}{\ell} \right] / \left[2^{\ell-1} - 1 \right] \right\}, \beta_\ell \geq 0, \ell \in \{2, \ldots, d\}$, such that ($T_3$) is fulfilled.

Note that while condition (T_5) aims to achieve the ordering axiom (A_4), it is not a necessary condition.

Taylor [31] rephrases the generalization in (10) in terms of copulas as

$$\tau_2(C_d, \boldsymbol{w}_d) = \sum_{k=0}^{d} w_{d,k} \sum_{\substack{\xi \in \mathcal{R}_d \\ |S_\xi| = k}} \int_{[0,1]^d} C_d^\xi(\boldsymbol{u}) \, \mathrm{d}C_d^\xi(\boldsymbol{u})$$

and adds another condition on the weights such that τ_2 satisfies the transition property of axiom (A_7), specifically that for every $d \geq 2$ there exists a constant r_d such that

(T_6) $r_d w_{d,k} = w_{d+1,k+1} + w_{d+1,k}, k \in \{0, \ldots, d\}$.

Up to this point, the only undiscussed axiom is thus axiom (A_8) which is under a mild additional condition on the weights shown in Proposition 3.

Proposition 3. *Let the weights for $\tau_2(\cdot, \boldsymbol{w}_d)$ satisfy condition (T_6) with $r_d < 2$. Then axiom (A_8) is fulfilled by $\tau_2(\cdot, \boldsymbol{w}_d)$.*

Proof. Let C_d be a d-variate copula of $(X_1, \ldots, X_d)^\top$ and C_{d+1} be a $(d+1)$-variate copula of $(X_1, \ldots, X_d, X_{d+1})^\top$ where X_{d+1} is independent of $(X_1, \ldots, X_d)^\top$. Each $\zeta \in \mathcal{R}_{d+1}$ can be written in the form $\zeta(u_1, \ldots, u_{d+1}) = (\xi(u_1, \ldots, u_d), \eta(u_{d+1}))$ with $\xi \in \mathcal{R}_d$, $\eta \in \mathcal{R}_1$. Thanks to the independence of the last component

$$C_{d+1}^\zeta(u_1, \ldots, u_{d+1}) = C_d^\xi(u_1, \ldots, u_d) u_{d+1}.$$

Denoting $\boldsymbol{u} = (u_1, \ldots, u_d)^\top$, Kendall's tau τ_2 can be written as

$$\begin{aligned} \tau_2(C_{d+1}, \boldsymbol{w}_{d+1}) &= \sum_{k=0}^{d+1} w_{d+1,k} \sum_{\substack{\zeta \in \mathcal{R}_{d+1} \\ |S_\zeta| = k}} \int_{[0,1]^{d+1}} C_{d+1}^\zeta(\boldsymbol{u}, u_{d+1}) \, \mathrm{d}C_{d+1}^\zeta(\boldsymbol{u}, u_{d+1}) \\ &= \sum_{k=0}^{d} w_{d+1,k} \sum_{\substack{\xi \in \mathcal{R}_d \\ |S_\xi| = k}} \int_{[0,1]^{d+1}} C_d^\xi(\boldsymbol{u}) \, u_{d+1} \, \mathrm{d}C_d^\xi(\boldsymbol{u}) \, \mathrm{d}u_{d+1} + \sum_{k=1}^{d+1} w_{d+1,k} \sum_{\substack{\xi \in \mathcal{R}_d \\ |S_\xi| = k-1}} \int_{[0,1]^{d+1}} C_d^\xi(\boldsymbol{u}) \, u_{d+1} \, \mathrm{d}C_d^\xi(\boldsymbol{u}) \, \mathrm{d}u_{d+1} \end{aligned} \tag{11}$$

and by integrating with respect to the last, independent component together with substitution $\ell = k - 1$ in (11), we get

$$\begin{aligned} \tau_2(C_{d+1}, \boldsymbol{w}_{d+1}) &= \sum_{k=0}^{d} \frac{w_{d+1,k}}{2} \sum_{\substack{\xi \in \mathcal{R}_d \\ |S_\xi| = k}} \int_{[0,1]^d} C_d^\xi(\boldsymbol{u}) \, \mathrm{d}C_d^\xi(\boldsymbol{u}) + \sum_{\ell=0}^{d} \frac{w_{d+1,\ell+1}}{2} \sum_{\substack{\xi \in \mathcal{R}_d \\ |S_\xi| = \ell}} \int_{[0,1]^d} C_d^\xi(\boldsymbol{u}) \, \mathrm{d}C_d^\xi(\boldsymbol{u}) \\ &= \sum_{k=0}^{d} \frac{w_{d+1,k} + w_{d+1,k+1}}{2} \sum_{\substack{\xi \in \mathcal{R}_d \\ |S_\xi| = k}} \int_{[0,1]^d} C_d^\xi(\boldsymbol{u}) \, \mathrm{d}C_d^\xi(\boldsymbol{u}) = \sum_{k=0}^{d} w_{d,k} \frac{r_d}{2} \sum_{\substack{\xi \in \mathcal{R}_d \\ |S_\xi| = k}} \int_{[0,1]^d} C_d^\xi(\boldsymbol{u}) \, \mathrm{d}C_d^\xi(\boldsymbol{u}) = \frac{r_d}{2} \tau_2(C_d, \boldsymbol{w}_d) \end{aligned}$$

which finishes the proof since $r_d/2 < 1$. □

The validity of axioms $(A_1)-(A_7)$ for $\tau(C_d)$ is stated by [27] and the validity of (A_8) is stated in Corollary 1.

Corollary 1. *Axiom* (A_8) *is fulfilled by* τ*, defined in* (9).

Proof. Association measure $\tau(C_d)$ is a special case of $\tau_2(C_d)$ as can be seen as follows. Denoting $\boldsymbol{X}$ and $\boldsymbol{Y}$ identically distributed d-variate random vectors with joint distribution function C_d, we can write

$$\tau(C_d) = \frac{1}{2^{d-1}-1}\left\{2^d \int\limits_{[0,1]^d} C_d(\boldsymbol{u})\,\mathrm{d}C_d(\boldsymbol{u}) - 1\right\} = \frac{1}{2^{d-1}-1}\left\{2^{d-1}[\Pr(\boldsymbol{X}<\boldsymbol{Y}) + \Pr(\boldsymbol{X}>\boldsymbol{Y})] - 1\right\}$$
$$= \frac{1}{2^{d-1}-1}\left\{2^{d-1}[\Pr((D_1,\dots,D_d)\in B_{0,d}) + \Pr((D_1,\dots,D_d)\in B_{d,0})] - 1\right\}$$
$$= \Pr((D_1,\dots,D_d)\in B_{0,d}) + \Pr((D_1,\dots,D_d)\in B_{d,0}) - \frac{1}{2^{d-1}-1}[1 - \Pr((D_1,\dots,D_d)\in B_{0,d}) - \Pr((D_1,\dots,D_d)\in B_{0,d}]$$
$$= \Pr((D_1,\dots,D_d)\in B_{0,d}) + \Pr((D_1,\dots,D_d)\in B_{d,0}) - \frac{1}{2^{d-1}-1}\left[\sum_{k=1}^{d-1}\Pr((D_1,\dots,D_d)\in B_{k,d-k})\right].$$

That is, for $w_{d,k} = 1$ if $k \in \{0, d\}$ and $w_{d,k} = -1/(2^{d-1}-1)$ otherwise, we have $\tau(C_d) = \tau_2(C_d, \boldsymbol{w}_d)$. Taylor [31] stated that for this selection of the weights, condition (T_6) is satisfied with $r_d = (2^d-2)/(2^d-1) < 2$. The assumptions of Proposition 3 are thus satisfied which proves the statement of this corollary. □

Kendall's tau is often a preferred association measure for bivariate random vectors, either because of its interpretation or robustness properties [see e.g. 4]. We can notice that the idea of concordance is incorporated also in the general d-variate definition since $\int_{[0,1]^d} C_d\,\mathrm{d}C_d = \Pr(\boldsymbol{X} \le \boldsymbol{Y})$ where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent random vectors, both having distribution C_d.

3.2.3. Multivariate Gini's gamma

Bivariate Gini's gamma is defined as [see e.g. 18, p. 159 and p. 180]

$$\begin{aligned}\gamma(X_1,X_2) = &\Pr\{(X_1-Y_1)(X_2-Y_2)>0\} - \Pr\{(X_1-Y_1)(X_2-Y_2)<0\}\\ &+ \Pr\{(X_1-Z_1)(X_2-Z_2)>0\} - \Pr\{(X_1-Z_1)(X_2-Z_2)<0\}\end{aligned}$$

where $(X_1,X_2)^\top$, $(Y_1,Y_2)^\top$ and $(Z_1,Z_2)^\top$ are independent random vectors with copulas C_2, M_2 and W_2, respectively, and common margins F_1 (of X_1, Y_1 and Z_1) and F_2 (of X_2, Y_2 and Z_2). This coefficient can be further expressed using the underlying copula as [see 18, p. 181]

$$\gamma(C_2) = 4\int_{[0,1]^2} (M_2(\boldsymbol{u}) + W_2(\boldsymbol{u}))\,\mathrm{d}C_2(\boldsymbol{u}) - 2 \tag{12}$$

and using integration by parts, as

$$\gamma(C_2) = 8\int_{[0,1]^2} C_2(\boldsymbol{u})\,\mathrm{d}\left(\frac{M_2(\boldsymbol{u}) + W_2(\boldsymbol{u})}{2}\right) - 2. \tag{13}$$

Gini's gamma can thus be viewed as a distance between C_2 and the maximal (positive and negative) dependence copulas M_2 and W_2 measured in terms of concordance. For more insight on this interpretation, see Nelsen [18, e.g. Theorem 5.1.13].

In the multivariate setting Gini's gamma can be generalized in multiple ways into general dimension $d \ge 2$. Starting from (12) and using the inclusion-exclusion principle (see (1)), one obtains

$$\overline{M}_2(u_1,u_2) + \overline{W}_2(u_1,u_2) = 2 - 2u_1 - 2u_2 + M_2(u_1,u_2) + W_2(u_1,u_2)$$

and thus

$$\int_{[0,1]^2} (\overline{M}_2(u_1,u_2) + \overline{W}_2(u_1,u_2))\,\mathrm{d}C_2(u_1,u_2) = \int_{[0,1]^2} (M_2(u_1,u_2) + W_2(u_1,u_2))\,\mathrm{d}C_2(u_1,u_2).$$

Then $\gamma(C_2)$ in (12) can be rewritten as

$$\gamma(C_2) = 2\int_{[0,1]^2} (M_2(u_1,u_2) + W_2(u_1,u_2) + \overline{M}_2(u_1,u_2) + \overline{W}_2(u_1,u_2))\,\mathrm{d}C_2(u_1,u_2) - 2$$

which lead [2] to propose a first generalization in the form

$$\gamma_1(C_d) = \frac{1}{b(d)-a(d)}\left(\int_{[0,1]^d} (M_d(\boldsymbol{u}) + W_d(\boldsymbol{u}) + \overline{M}_d(\boldsymbol{u}) + \overline{W}_d(\boldsymbol{u}))\,\mathrm{d}C_d(\boldsymbol{u}) - a(d)\right), \tag{14}$$

where

$$a(d) = \frac{2}{d+1} + \frac{1}{(d+1)!} + \sum_{j=0}^{d}(-1)^j\binom{d}{j}\frac{1}{(j+1)!}, \qquad b(d) = 2 - \sum_{j=1}^{d-1}\frac{1}{2j}$$

are normalizing constants. If we notice that the probabilistic mass of $(M_2 + W_2)/2$ is uniformly distributed along the diagonals of the unit square $[0,1]^2$, a generalization based on (13), proposed by [31] is

$$\gamma_2(C_d) = \frac{2^d}{2^{d-1}-1}\left(\int_{[0,1]^d}(C_d(\boldsymbol{u}) + C_d^S(\boldsymbol{u}))\,\mathrm{d}\left(\frac{1}{2^d}\sum_{\xi\in\mathcal{R}_d} M_d^{\xi}(\boldsymbol{u})\right) - \frac{1}{2^{d-1}}\right). \tag{15}$$

Calculating γ_2 as originally defined by [31] involves 2^d d-dimensional integrals. However, we can further simplify this formula. By realizing that M_d^{ξ} in the definition of γ_2 is a function that is constant everywhere except on one of the diagonals (depending on ξ) of the unit hypercube. This implies that the integration can be viewed as one-dimensional over the corresponding diagonal. Recall that M_d is the comonotonicity copula corresponding to the situation where all U_i's are equal, i.e., the copula of $(U,\dots,U)^\top$. We know that a reflection ξ reflects some components of the vector $\boldsymbol{U}$. If we denote by S_ξ the set of indices indicating which components were reflected by ξ, we can then write that M_d^{ξ} is the copula of $\boldsymbol{V} = (V_1,\dots,V_d)^\top$ where $V_i = 1-U$ if $i \in S_\xi$ and $V_i = U$ otherwise. Since $\boldsymbol{V}$ depends only on U and ξ, we have for arbitrary but fixed $\xi \in \mathcal{R}_d$

$$\int_{[0,1]^d}(C_d(\boldsymbol{u}) + C_d^S(\boldsymbol{u}))\,\mathrm{d}M_d^{\xi}(\boldsymbol{u}) = \int_0^1 C_d(\xi(u,\dots,u)) + C_d^S(\xi(u,\dots,u))\,\mathrm{d}u.$$

Then we can rewrite γ_2 as

$$\gamma_2(C_d) = \frac{1}{2^{d-1}-1}\left(\sum_{\xi\in\mathcal{R}_d}\int_0^1 (C_d(\xi(u,\dots,u)) + C_d^S(\xi(u,\dots,u)))\,\mathrm{d}u - 2\right), \tag{16}$$

or equivalently, since $C_d^S(\boldsymbol{u}) = \overline{C}_d(\mathbf{1}-\boldsymbol{u})$ and we sum over all reflections

$$\gamma_2(C_d) = \frac{1}{2^{d-1}-1}\left(\sum_{\xi\in\mathcal{R}_d}\int_0^1 (C_d(\xi(u,\dots,u)) + \overline{C}_d(\xi(u,\dots,u)))\,\mathrm{d}u - 2\right), \tag{17}$$

which involves only calculation of one-dimensional integrals.

Let us now focus on the validity of the axioms for these two versions of Gini's gamma. For γ_1 in (14), axioms (A_1)–(A_3) follow easily from its definition. However, two axioms, the duality axiom (A_5) and the axiom (A_8) regarding the independent component addition, are violated as can be seen through the following example.

Example 1. Let C_3 be a trivariate Farlie-Gumbel-Morgenstern copula defined as

$$\begin{aligned} C_3(u_1, u_2, u_3) = u_1 u_2 u_3 \, [1 \, + \, &\alpha(1-u_2)(1-u_3) + \beta(1-u_1)(1-u_3) \\ &+ \gamma(1-u_1)(1-u_2) + \delta(1-u_1)(1-u_2)(1-u_3)] \end{aligned}$$

where the four parameters $\alpha, \beta, \gamma, \delta$ all belong to the interval $[-1, 1]$ and satisfy the inequalities $1 + \epsilon_1\alpha + \epsilon_2\beta + \epsilon_3\gamma > |\delta|$ for $\epsilon_i = \pm 1$ such that $\epsilon_1\epsilon_2\epsilon_3 = 1$. Then a few standard calculations lead to

$$\gamma_1(C_3) = \frac{4}{45}(\alpha + \beta + \gamma).$$

In particular, $\gamma_1(C_3)$ does not depend on δ and if $\alpha = \beta = \gamma = 0$, that is, all the bivariate marginal copulas are independence copulas, then $\gamma_1(C_3) = 0$ no matter what δ is.

We next show that γ_1 does not satisfy axiom (A_8). Define copula C_4 by adding an independent fourth component

$$C_4(u_1, u_2, u_3, u_4) = C_3(u_1, u_2, u_3)u_4.$$

Then we can show that

$$\gamma_1(C_4) = \frac{40\alpha + 40\beta + 40\gamma + 3\delta}{1050}.$$

In particular $\gamma_1(C_4)$ depends on δ and if $\alpha = \beta = \gamma = 0$, that is, all the bivariate marginal copulas are independence copulas, then $\gamma_1(C_4) = 3\delta/1050$. This also means that for $\alpha = \beta = \gamma = 0$ and $\delta \neq 0$, we have $|\gamma_1(C_4)| > \gamma_1(C_3) = 0$. The sign of $\gamma_1(C_4)$ then depends on the sign of δ. In other words, we found an example in which adding an independent component increases association measured by γ_1 which thus fails to satisfy axiom (A_8).

Finally, let C_4^S be a survival copula of C_4. Then we can show that

$$\gamma_1(C_4^S) = \frac{40\alpha + 40\beta + 40\gamma - 3\delta}{1050},$$

that is, unless $\delta = 0$, we have $\gamma_1(C_4) \neq \gamma_1(C_4^S)$ and thus duality axiom (A_5) is also violated.

In conclusion, γ_1 defined in (14) does not fulfil axioms (A_5) and (A_8). Regarding axiom (A_5), this could be easily fixed by considering in the integration $\mathrm{d}(C_d + C_d^S)/2$ instead of $\mathrm{d}C_d$. However, since W_d is not a copula for $d > 2$, γ_1 lacks a clear probabilistic interpretation.

For γ_2 in (15), all axioms are fulfilled with $(A_1) - (A_7)$ discussed by [31] and the validity of (A_8) established in the following proposition.

Proposition 4. *Axiom* (A_8) *is fulfilled by* γ_2*, defined in* (15).

Proof. Let C_d be a d-variate copula of $(X_1, \ldots, X_d)^\top$ and C_{d+1} be a $(d+1)$-variate copula of $(X_1, \ldots, X_d, X_{d+1})^\top$ where X_{d+1} is independent of $(X_1, \ldots, X_d)^\top$. Using equation (16) above, Taylor's generalization of Gini's gamma γ_2 can be written as

$$\begin{aligned} \gamma_2(C_{d+1}) &= \frac{1}{2^d - 1}\left(\sum_{\zeta \in \mathcal{R}_{d+1}} \int_0^1 [C_{d+1}(\zeta(u, \ldots, u)) + C_{d+1}^S(\zeta(u, \ldots, u))] \, \mathrm{d}u - 2 \right) \\ &= \frac{1}{2^d - 1}\left(\sum_{\xi \in \mathcal{R}_d} \int_0^1 [C_{d+1}(\xi(u, \ldots, u), u) + C_{d+1}(\xi(u, \ldots, u), 1-u) \right. \\ &\qquad \left. + C_{d+1}^S(\xi(u, \ldots, u), u) + C_{d+1}^S(\xi(u, \ldots, u), 1-u)] \, \mathrm{d}u - 2 \right). \end{aligned}$$

Using independence of the last component, we get

$$\begin{aligned}
&\gamma_2(C_{d+1}) \\
&= \frac{1}{2^d-1}\left(\sum_{\xi\in\mathcal{R}_d}\int_0^1 [C_d(\xi(u,\dots,u))u + C_d(\xi(u,\dots,u))(1-u) + C_d^S(\xi(u,\dots,u))u + C_d^S(\xi(u,\dots,u))(1-u)]\,\mathrm{d}u - 2\right) \\
&= \frac{1}{2^d-1}\left(\sum_{\xi\in\mathcal{R}_d}\int_0^1 [C_d(\xi(u,\dots,u)) + C_d^S(\xi(u,\dots,u)]\,\mathrm{d}u - 2\right)
\end{aligned}$$

and thus we have

$$\gamma_2(C_{d+1}) = \frac{2^{d-1}-1}{2^d-1}\gamma_2(C_d),$$

which proves the statement since $(2^{d-1}-1)/(2^d-1) < 1$ for every $d \geq 2$. □

Thus, after evaluation by means of axioms, there is certainly a preference for γ_2 over γ_1. Moreover, via (16) we have decreased considerably its computational complexity.

4. Further properties of multivariate association measures

We next investigate properties (P_1) and (P_2) for the discussed multivariate association measures. Herein we can clearly distinguish between Kendall's tau on the one hand and Spearman's rho and Gini's gamma on the other hand. For Kendall's tau we establish results for (P_1) and (P_2), whereas we provide counterexamples for the latter ones.

4.1. Multivariate Kendall's tau

Proposition 5 states that property (P_1) holds for multivariate Kendall's tau.

Proposition 5. *Let C_d be the copula of a random vector $\boldsymbol{X} = (X_1,\dots,X_d)^\top$ and C_{d+1} be the copula of $(\boldsymbol{X}^\top, X_{d+1})^\top$ where $X_{d+1} = X_j$, for some $j \in \{1,\dots,d\}$. Then*

$$\tau(C_d) \leq \tau(C_{d+1}). \tag{18}$$

If $\tau(C_d) < 1$, the concerned inequality in (18) *is a strict inequality, whereas equality in* (18) *happens when $\tau(C_d) = 1$. Statement* (18) *continues to hold if we add a conical combination of the components of $\boldsymbol{X}$ instead of a duplicate.*

Proof. Recall that $\int_{[0,1]^d} C_d(\boldsymbol{u})\,\mathrm{d}C_d(\boldsymbol{u}) = \Pr(\boldsymbol{X} < \boldsymbol{Y})$ where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent and identically distributed random vectors with copula C_d. Then

$$\int_{[0,1]^{d+1}} C_{d+1}(\boldsymbol{u})\,\mathrm{d}C_{d+1}(\boldsymbol{u}) = \Pr(\boldsymbol{X} < \boldsymbol{Y}, X_{d+1} < Y_{d+1}) = \int_{[0,1]^d} C_d(\boldsymbol{u})\,\mathrm{d}C_d(\boldsymbol{u}),$$

and hence

$$\tau(C_{d+1}) = \frac{1}{2^d-1}\left\{2^{d+1}\int\limits_{[0,1]^{d+1}} C_{d+1}(\boldsymbol{u})\,\mathrm{d}C_{d+1}(\boldsymbol{u}) - 1\right\} = \frac{(2^d-2)\tau(C_d)+1}{2^d-1} \geq \tau(C_d),$$

from which statement (18) follows. Note that if and only if $\tau(C_d) = 1$ there is an equality in the last line. From this proof it is evident that statement (18) also holds when adding any conical combination of components of $\boldsymbol{X}$. □

For a general copula C_d, Kendall's tau is bounded as follows

$$\frac{-1}{2^{d-1}-1} \leq \tau(C_d) \leq 1, \tag{19}$$

where the lower bound was introduced by [18], and was shown to be the best possible lower bound by Úbeda-Flores [33, Theorem 5.1]. Note that the lower bound in (19) converges to zero as d tends to infinity. The minimal value of τ is achieved when $\int_{[0,1]^d} C_d(\boldsymbol{u})\,\mathrm{d}C_d(\boldsymbol{u})$ is equal to zero which happens for example for a random vector containing a random variable X and also $-X$.

Proposition 6 states our findings concerning property (P_2) for Kendall's tau.

Proposition 6. *Let C_d be the copula of $\boldsymbol{X} = (X_1, \dots, X_d)^\top$ with corresponding Kendall's tau value $\tau(C_d)$. Then for $d_2 > d$ and any copula C_{d_2} of $(\boldsymbol{X}^\top, X_{d+1}, \dots, X_{d_2})^\top$*

$$\frac{-1}{2^{d_2-1}-1} \le \tau(C_{d_2}) \le \frac{1}{2^{d_2-1}-1}\left\{2^{d_2-1}\left(\tau(C_d) + \frac{1-\tau(C_d)}{2^{d-1}}\right) - 1\right\}. \tag{20}$$

Both the lower and the upper bounds are attainable.

Proof. Recall the definition of $\tau(C_d)$ in (9). First, we can express

$$\int_{[0,1]^d} C_d(\boldsymbol{u})\,\mathrm{d}C_d(\boldsymbol{u}) = \frac{(2^{d-1}-1)\tau(C_d)+1}{2^d}.$$

Further it is easily seen that

$$\int_{[0,1]^d} C_d(\boldsymbol{u})\,\mathrm{d}C_d(\boldsymbol{u}) \ge \int_{[0,1]^{d_2}} C_{d_2}(\boldsymbol{u})\,\mathrm{d}C_{d_2}(\boldsymbol{u})$$

for $d < d_2$ and the equality is achieved, for example, if X_j for $j \in \{d+1, \dots, d_2\}$ is a conical combination of the components of $\boldsymbol{X}$ (see also Proposition 5). That is, the maximal value $\tau(C_{d_2})$ can take is

$$\frac{1}{2^{d_2-1}-1}\left\{2^{d_2}\left(\frac{(2^{d-1}-1)\tau(C_d)+1}{2^d}\right) - 1\right\} = \frac{1}{2^{d_2-1}-1}\left\{2^{d_2-1}\left(\tau(C_d) + \frac{1-\tau(C_d)}{2^{d-1}}\right) - 1\right\}$$

and the minimal value for $\tau(C_{d_2})$ follows from (19) and the subsequent discussion. □

Note that the upper bound in (20) is strictly increasing in d_2, unless $\tau(C_d) = 1$, and converges to $\tau(C_d) + (1 - \tau(C_d))/2^{d-1}$, as d_2 tends to infinity. Also we see that unless $\tau(C_d)$ attains its minimal possible value $-1/(2^{d-1}-1)$, $\tau(C_{d_2})$ can always become positive for d_2 large enough.

4.2. Multivariate Spearman's rho

If a component of a random vector $\boldsymbol{X}$ is duplicated, Spearman's rho can both increase or decrease and thus property (P_1) does not hold. This follows from Example S3 in Section 6 of the Supplementary Material. Additionally, one can observe that in case of Spearman's rho, for example ρ_1, it is not possible to express $\rho_1(C_{d+1})$ using $\rho_1(C_d)$. The key element in ρ_1 is the integral

$$\int_{[0,1]^{d+1}} C_{d+1}(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} = \Pr(U_1 < V_1, \dots U_{d+1} < V_{d+1})$$

with $\boldsymbol{U}$ having a copula C_{d+1} and $\boldsymbol{V}$ having a copula Π_{d+1}. If one considers that C_{d+1} is the copula of $(\boldsymbol{X}^\top, X_d)^\top$, then

$$\int_{[0,1]^{d+1}} C_{d+1}(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} = \Pr(U_1 < V_1, \dots U_d < \min(V_d, V_{d+1})).$$

There is no way to simplify this expression or to express it in terms of $\rho_1(C_d)$. A similar remark holds for $\rho_2(C_d)$ and $\rho_3(C_d)$. In other words knowing the value of $\rho_\ell(C_d)$ (for $\ell \in \{1, 2, 3\}$) where C_d is the copula of $\boldsymbol{X}$ does not determine the value of $\rho_\ell(C_{d+1})$ where C_{d+1} corresponds to a vector $(\boldsymbol{X}^\top, X_{d+1})^\top$ with X_{d+1} being a duplicate of one of the components from $\boldsymbol{X}$. So the finding here is non-conclusive, and in contrast to the conclusive findings for Kendall's tau. The above finding is further illustrated with an example.

Example 2. Let $(X_1, X_2, X_3)^\top$ have a trivariate Farlie-Gumbel-Morgenstern copula C_3 defined in Example 1. Further, let C_4 be the copula of $(X_1, X_2, X_3, X_3)^\top$. Then

$$\int_{[0,1]^3} C_3(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} = \frac{27 + 3\alpha + 3\beta + 3\gamma + \delta}{216}, \qquad \int_{[0,1]^4} C_4(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} = \frac{18 + 3\alpha + 3\beta + 2\gamma + \delta}{216}$$

and thus knowing the values of the dimension, in this case $d = 3$, and Spearman's rho $\rho_1(C_3)$, which is a function of the left integral above, is not sufficient to know how $\rho_1(C_4)$ relates to $\rho_1(C_3)$ (larger or smaller). Different sets of parameters leading to the same value for $\rho_1(C_3)$ can give different values of $\rho_1(C_4)$. Knowing the copula C_3 itself is thus necessary. The same conclusions follow for ρ_2, ρ_3 using the same copulas C_3 and C_4.

It thus remains an open question which random variable X_{d+1} would lead to the largest (or smallest) possible $\rho_\ell(C_{d+1})$ given $\boldsymbol{X}$ and what the largest (or smallest) possible $\rho_\ell(C_{d+1})$ is.

For a general copula C_d, Spearman's rho ρ_ℓ, $\ell \in \{1, 2, 3\}$, is bounded, in particular

$$\frac{2^d - (d+1)!}{d!(2^d - (d+1))} \leq \rho_\ell(C_d) \leq 1$$

where the lower bound was introduced in [16], and is not the best possible lower bound for $d \geq 3$. Note that the lower bound converges to zero as d tends to infinity.

4.3. Multivariate Gini's gamma

We now look into property (P_1) for Gini's gamma γ_2. Here the situation appears to be quite similar as for Spearman's rho. Indeed, recalling (16) we look into the following integrals separately. Note that

$$\sum_{\xi \in \mathcal{R}_{d+1}} \int_0^1 C_{d+1}(\xi(u, \ldots, u))\,\mathrm{d}u = \sum_{\xi \in \mathcal{R}_d} \int_0^1 C_d(\xi(u, \ldots, u))\,\mathrm{d}u + 2 \sum_{\xi \in \mathcal{R}_{d-1}} \int_0^1 C_d(\xi(u, \ldots, u), \min(u, 1-u))\,\mathrm{d}u$$

and similarly

$$\sum_{\xi \in \mathcal{R}_{d+1}} \int_0^1 C^S_{d+1}(\xi(u, \ldots, u))\,\mathrm{d}u = \sum_{\xi \in \mathcal{R}_d} \int_0^1 C^S_d(\xi(u, \ldots, u))\,\mathrm{d}u + 2 \sum_{\xi \in \mathcal{R}_{d-1}} \int_0^1 C^S_d(\xi(u, \ldots, u), \min(u, 1-u))\,\mathrm{d}u.$$

We therefore obtain that

$$\gamma_2(C_{d+1}) = \frac{1}{2^d - 1}\Bigg(\sum_{\xi \in \mathcal{R}_d} \int_0^1 C_d(\xi(u, \ldots, u))\,\mathrm{d}u + 2 \sum_{\xi \in \mathcal{R}_{d-1}} \int_0^1 C_d(\xi(u, \ldots, u), \min(u, 1-u))\,\mathrm{d}u$$
$$+ \sum_{\xi \in \mathcal{R}_d} \int_0^1 C^S_d(\xi(u, \ldots, u))\,\mathrm{d}u + 2 \sum_{\xi \in \mathcal{R}_{d-1}} \int_0^1 C^S_d(\xi(u, \ldots, u), \min(u, 1-u))\,\mathrm{d}u - 2\Bigg)$$

which can be also expressed as

$$\gamma_2(C_{d+1}) = \gamma_2(C_d)\frac{2^{d-1} - 1}{2^d - 1} + 2\frac{\sum_{\xi \in \mathcal{R}_{d-1}} \int_0^1 [C_d(\xi(u, \ldots, u), \min(u, 1-u)) + C^S_d(\xi(u, \ldots, u), \min(u, 1-u))]\,\mathrm{d}u}{2^d - 1}.$$

From this expression one cannot deduce any general conclusion about whether $\gamma_2(C_{d+1})$ is larger or smaller than $\gamma_2(C_d)$. This is again in contrast with the findings for Kendall's tau, but in line with the finding for Spearman's rho. That a property as in (18) cannot hold for Gini's gamma is illustrated by the next counterexample.

Example 3. Consider U and V independent random variables with uniform margins. Let C_4 be the copula of the random vector $(U, U, U, V)^\top$. Direct calculation, starting from, for example (16), then leads to $\gamma_2(C_4) = \frac{3}{7} = 0.43$. Now let C_5 be the copula of the random vector $(U, U, U, V, V)^\top$. Then calculations show that $\gamma_2(C_5) = \frac{6}{15} = 0.4$. In this example we thus have $\gamma_2(C_5) < \gamma_2(C_4)$.

The multivariate association measure γ_2 is bounded, as stated in the following proposition.

Proposition 7. *Let C_d be a d-dimensional copula. Then*

$$\frac{2(1-d)}{d(2^{d-1}-1)} \leq \gamma_2(C_d) \leq 1. \tag{21}$$

Proof. Recall the definition of γ_2 in (16). Using the Fréchet's lower bound W_d, we know that

$$\sum_{\xi \in \mathcal{R}_d} \int_0^1 (C_d(\xi(u,\dots,u)) + C_d^S(\xi(u,\dots,u)))\,\mathrm{d}u \geq 2 \sum_{\xi \in \mathcal{R}_d} \int_0^1 W_d(\xi(u,\dots,u))\,\mathrm{d}u. \tag{22}$$

Straightforward calculations give

$$\begin{aligned} \sum_{\xi \in \mathcal{R}_d} \int_0^1 W_d(\xi(u,\dots,u))\,\mathrm{d}u &= \sum_{j=0}^{d} \binom{d}{j} \int_0^1 \max(j(1-u) + (d-j)u - d + 1, 0)\,\mathrm{d}u \\ &= \int_{\frac{d-1}{d}}^1 (du - d + 1)\,\mathrm{d}u + \int_0^{\frac{1}{d}} (1 - du)\,\mathrm{d}u = \frac{1}{d}, \end{aligned} \tag{23}$$

where we use that W_d is symmetric in its arguments and that function $\max(j(1-u) + (d-j)u - d + 1, 0)$ is different from zero function on $(0,1)$ only if $j = 0$ or $j = d$. Together with the upper bound given from the axioms (A_1) and (A_4), the combination of (22) and (23) gives (21). □

Note that the lower bound in (21) is negative for every $d \geq 2$ and converges to 0 as d tends to infinity. This bound is not necessarily the best possible for $d \geq 3$, since W_d is then even not a copula.

Suppose again that C_d is the copula of $\boldsymbol{X}$ and C_{d+1} is the copula of $(\boldsymbol{X}^\top, X_{d+1})^\top$. Similarly as for ρ_ℓ, $\ell \in \{1,2,3\}$, it is unknown which random variable X_{d+1} would lead to the largest (or smallest) possible $\gamma_2(C_{d+1})$ given $\boldsymbol{X}$ and what the largest (or smallest) possible $\gamma_2(C_{d+1})$ is. Using the same copulas as in Example 2 it can be shown that knowledge of $\gamma_2(C_d)$ is not sufficient to determine $\gamma_2(C_{d+1})$.

4.4. Multivariate association measures based on the pairwise approach

For pairwise type of association measures, property (P_1) also does not hold. To see this consider the following example. Under the assumptions of Proposition 1 and denoting by C_4 the copula of $(X,X,X,-X)^\top$, and by C_5 be the copula of $(X,X,X,-X,-X)^\top$, it is straightforward to show that

$$\kappa_4^{\mathrm{PW}}(C_4) = 0 > \frac{-1}{5} = \kappa_5^{\mathrm{PW}}(C_5).$$

Example S3 in Section 6 of the Supplementary Material contains a situation in which $\kappa_d^{\mathrm{PW}}(C_d) < \kappa_{d+1}^{\mathrm{PW}}(C_{d+1})$ for every $d = 2, \dots$. Hence for a pairwise type of association measure, $\kappa_d^{\mathrm{PW}}(C_{d+1})$ can be larger or smaller than $\kappa_d^{\mathrm{PW}}(C_d)$.

5. Multivariate association measures in increasing dimensions

Of interest is to study how the multivariate association measures evolve when the dimension increases towards infinity. Studying this in full generality is a difficult task since some structure is needed to be able to quantify and interpret the effect of the increasing dimension. So far we are able to study (partially) this effect under two settings: (i) Archimedean copulas; and (ii) meta-elliptical copulas.

5.1. *Archimedean copulas and multivariate association measures in increasing dimensions*

5.1.1. *Spearman's rho in increasing dimension*

Different ways of calculating the multivariate Spearman's rhos ρ_1 and ρ_2 for Archimedean copulas are proposed by [36] who further studied the asymptotic behaviour when the dimension increases to infinity. An infinite dimensional Archimedean copula C_∞ is defined either as

$$C_\infty(\boldsymbol{u}) = \lim_{d\to\infty} \psi\left[\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d)\right], \qquad \forall \boldsymbol{u} \in [0,1]^{\mathbb{N}}$$

or, equivalently, as a measure $\mu_\infty(\prod_{i=1}^{\infty}[0,u_i])$ on an infinite dimensional Hilbert cube $[0,1]^{\mathbb{N}}$. [36] also showed that

$$\lim_{d\to\infty} \rho_1(C_d) = c_1 \in [0,1] \quad \Longleftrightarrow \quad \lim_{d\to\infty} (d+1) \int_{[0,1]^d} C_d(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} = c_1,$$

$$\lim_{d\to\infty} \rho_2(C_d) = c_2 \in [0,1] \quad \Longleftrightarrow \quad \lim_{d\to\infty} (d+1) \int_{[0,1]^d} \Pi_d(\boldsymbol{u})\,\mathrm{d}C_d(\boldsymbol{u}) = c_2.$$

This result can obviously be extended to ρ_3. The existence of the limits is however not discussed and moreover, due to the difficulties with calculating Spearman's rho it remains unknown whether strict positive c_1 and c_2 can be achieved.

5.1.2. *Kendall's tau in increasing dimension*

[8] studied a way how to calculate τ for Archimedean copulas using that an Archimedean copula C_d is the survival copula of a simplex random vector $\boldsymbol{X} = R\boldsymbol{V}$, where R is a positive random variable, independent of the d-variate random vector $\boldsymbol{V}$ which is uniformly distributed on the unit simplex $\Delta^{d-1} \subset \mathbb{R}^d$, as shown by [14]. It then follows that if $\boldsymbol{U}$ is distributed as C_d, then $C_d(\boldsymbol{U})$ has the same distribution as $\psi(R)$ and thus, from (9)

$$\tau(C_d) = \frac{1}{2^{d-1}-1}\left\{-1 + \mathsf{E}\,[\psi(R)]\right\}.$$

Furthermore, C_d has a density if and only if R does. In the latter case,

$$\mathsf{E}\,[\psi(R)] = \frac{(-1)^d}{(d-1)!} \int_0^{\psi^{-1}(0)} r^{d-1}\psi(r)\psi^{(d)}(r)\,\mathrm{d}r.$$

where $\psi^{(d)}$ denotes the d-th derivative of ψ, which exists almost everywhere. This formula allows to calculate Kendall's tau for some specific examples.

Different ways of calculating Kendall's tau for Archimedean copulas are proposed by [36]. He further showed that $\lim_{d\to\infty} \tau(C_d) = 0$. In other words, Kendall's tau cannot capture any association of Archimedean copulas when dimension tends to infinity or, from another point of view, Archimedean copulas are not able to carry any association in very high dimensions.

5.1.3. *Gini's gamma in increasing dimension*

When talking about Gini's gamma, we will only focus on γ_2 (see (15), (16) or (17)) because of its theoretical properties discussed in Section 3. We can express γ_2 for Archimedean copulas using their generator. Let $\{C_d\}$ be a sequence of d-dimensional Archimedean copulas with (the same) generator ψ. Then using formula (17), Definition S1 and the inclusion-exclusion principle we get the following result

Lemma 1. *The multivariate Gini's gamma γ_2 for Archimedean copula C_d with generator ψ can be rewritten as*

$$\gamma_2(C_d) = \frac{1}{2^{d-1}-1}\left(\sum_{j=0}^{d}\binom{d}{j}\int_0^1 \left[J_{1j}(u) + J_{2j}(u)\right]\mathrm{d}u - 2\right)$$

where

$$J_{1j}(u) = \psi(j\psi^{-1}(u) + (d-j)\psi^{-1}(1-u)),$$
$$J_{2j}(u) = 1 + \sum_{k=1}^{d}(-1)^k \sum_{\ell=\max(0,j+k-d)}^{\min(j,k)} \binom{j}{\ell}\binom{d-j}{k-\ell}\psi(\ell\psi^{-1}(u) + (k-\ell)\psi^{-1}(1-u)).$$

Proof. First realize that since Archimedean copulas are exchangeable (invariant with respect to the order of its arguments), we can write

$$\sum_{\xi\in\mathcal{R}_d}\int_0^1 (C_d(\xi(u,\dots,u)) + \overline{C}_d(\xi(u,\dots,u)))\,\mathrm{d}u = \sum_{j=0}^{d}\binom{d}{j}\int_0^1 [C_d(\underbrace{u,\dots,u}_{j},\underbrace{1-u,\dots,1-u}_{d-j}) + \overline{C}_d(\underbrace{u,\dots,u}_{j},\underbrace{1-u,\dots,1-u}_{d-j})]\,\mathrm{d}u.$$

We thus do not need to sum over all the reflections from $\mathcal{R}_d$, but it suffices to group them by the number of arguments j that are not reflected. We next treat the copula and the survival copula parts separately. By the definition of Archimedean copulas

$$C_d(\underbrace{u,\dots,u}_{j},\underbrace{1-u,\dots,1-u}_{d-j}) = \psi(j\psi^{-1}(u) + (d-j)\psi^{-1}(1-u)).$$

For the survival function part, the calculation is a bit more complex. When using the inclusion-exclusion principle, we in fact need to go through all possible subsets of the arguments. However, using again that Archimedean copulas are exchangeable, we only need to know how many times the argument is u and how many times the argument is $1-u$. Thus for fixed $j\in\{0,\dots,d\}$

$$\overline{C}_d(\underbrace{u,\dots,u}_{j},\underbrace{1-u,\dots,1-u}_{d-j}) = 1 + \sum_{k=1}^{d}(-1)^k \sum_{\ell=\max(0,j+k-d)}^{\min(j,k)} \binom{j}{\ell}\binom{d-j}{k-\ell}\psi(\ell\psi^{-1}(u) + (k-\ell)\psi^{-1}(1-u))$$

where the summation over k goes through all the dimensions when using the inclusion-exclusion principle as in (1). Then ℓ denotes the amount of times u is selected to the arguments subset of size k. The limits for ℓ can be seen as following

- ℓ cannot be greater than the subset size, i.e. $\ell \le k$
- u cannot be used more than j times, i.e. $\ell \le j$
- ℓ must be at least 0, i.e. $\ell \ge 0$
- if u is selected ℓ times, the remaining $k-\ell$ arguments will be filled with $1-u$. Thus, we also need to 'have' a sufficient amount of $1-u$ arguments available. i.e. $k-\ell \le d-j$ and thus $\ell \ge j+k-d$.

For ℓ fixed, we only need to select ℓ times u from all j of them and then $k-\ell$ times $1-u$ from all $d-j$ of them which explains the binomial coefficients. No matter what j is, Vandermonde's identity [see e.g. 22, (3) on p. 8] gives us that the total number of summands is

$$1 + \sum_{k=1}^{d}\sum_{\ell=\max(0,j+k-d)}^{\min(j,k)}\binom{j}{\ell}\binom{d-j}{k-\ell} = 2^d$$

which is exactly the number of all subsets of arguments of $\overline{C}_d$. □

The limiting behaviour of γ_2 is established in the following proposition.

Proposition 8. *Let $\{C_d\}$ be a sequence of d-dimensional Archimedean copulas with (the same) generator ψ. Then*

$$\lim_{d\to\infty} \gamma_2(C_d) = 0.$$

Proof. Using (16) we rewrite

$$\gamma_2(C_d) = \frac{2^d}{2^{d-1}-1}\left(\frac{1}{2^d}\sum_{\xi\in\mathcal{R}_d}\int_0^1 (C_d(\xi(u,\ldots,u)) + C_d^S(\xi(u,\ldots,u)))\,\mathrm{d}u - 2^{1-d}\right),$$

and first focus on the term

$$\frac{1}{2^d}\sum_{\xi\in\mathcal{R}_d}\int_0^1 C_d(\xi(u,\ldots,u))\,\mathrm{d}u.$$

We choose an arbitrary element ξ from $\mathcal{R}_d$. Since a copula is a distribution function and thus non-decreasing in its arguments, we can write

$$C_d(\xi(u,\ldots,u)) \le C_d(\max(u,1-u),\ldots,\max(u,1-u)) = \psi(d\cdot\psi^{-1}(\max(u,1-u))), \tag{24}$$

no matter what $\xi\in\mathcal{R}_d$ is. For simplification, denote $u_M = \max(u,1-u)$. Since we also know that $\psi^{-1}(t) = 0$ if and only if $t = 1$, then

$$\lim_{d\to\infty}\psi(d\cdot\psi^{-1}(u_M)) = \begin{cases} 1, & \text{if } u = 0 \text{ or } u = 1,\\ 0 & \text{otherwise.}\end{cases} \tag{25}$$

By the Lebesgue dominated convergence theorem and using (24) and (25), we get

$$\lim_{d\to\infty}\frac{1}{2^d}\sum_{\xi\in\mathcal{R}_d}\int_0^1 C_d(\xi(u,\ldots,u))\,\mathrm{d}u \le \lim_{d\to\infty}\frac{1}{2^d}|\mathcal{R}_d|\int_0^1 \psi(d\cdot\psi^{-1}(u_M))\,\mathrm{d}u = \int_0^1 \lim_{d\to\infty}\psi(d\cdot\psi^{-1}(u_M))\,\mathrm{d}u = 0.$$

Now we focus on calculating

$$\frac{1}{2^d}\sum_{\xi\in\mathcal{R}_d}\int_0^1 C_d^S(\xi(u,\ldots,u))\,\mathrm{d}u.$$

Choose again an arbitrary element ξ from $\mathcal{R}_d$ and write

$$\begin{aligned} C_d^S(\xi(u,\ldots,u)) \le C_d^S(\max(u,1-u),\ldots,\max(u,1-u)) &= \overline{C}_d(1-\max(u,1-u),\ldots,1-\max(u,1-u))\\ &= \overline{C}_d(\min(u,1-u),\ldots,\min(u,1-u))\end{aligned}$$

no matter what $\xi\in\mathcal{R}_d$ is. Now using Lemma S1 and mimicking the proof of Proposition S2 leads to

$$\lim_{d\to\infty}\frac{1}{2^d}\sum_{\xi\in\mathcal{R}_d}\int_0^1 C_d^S(\xi(u,\ldots,u))\,\mathrm{d}u \le \lim_{d\to\infty}\frac{1}{2^d}|\mathcal{R}_d|\int_0^1 \overline{C}_d(u_m,\ldots,u_m)\,\mathrm{d}u = \int_0^1 \lim_{d\to\infty}\overline{C}_d(u_m,\ldots,u_m)\,\mathrm{d}u, = 0,$$

where we denoted $u_m = \min(u,1-u)$. This completes the proof. □

5.2. Meta-elliptical copulas and multivariate association measures in increasing dimensions

Meta-elliptical copulas are copulas with elliptical contours. See for example [1, 6]. For meta-elliptical copulas some partial results on the behaviour of Kendall's tau and Spearman's rho for increasing dimension can be obtained. For Kendall's tau τ for meta-elliptical copulas was studied by Genest et al. [8, Section 2.1]. They show that all meta-elliptical copulas that have the same correlation matrix also share the same Kendall's tau value. In other words, Kendalls'tau only depends on the correlation matrix. Let C_d be the copula of $\boldsymbol{U} = (U_1, \ldots U_d)^\top$, with C_d a meta-elliptical copula with correlation matrix $\boldsymbol{R} = (\varrho_{i,j})$, with $\varrho_{i,j} \in [-1, 1]$. Then it was shown in [8] that

$$\tau(C_d) = \frac{1}{2^{d-1} - 1}\{-1 + 2^d \Pr(\boldsymbol{Z} \geq \boldsymbol{0})\} \tag{26}$$

where $\boldsymbol{Z}$ is d-variate normal distributed random vector with zero mean and correlation matrix $\boldsymbol{R} = (\varrho_{i,j})$.

We next study what happens with (26) when d increases to infinity. This requires that we can evaluate the orthant probability $\Pr(\boldsymbol{Z} \geq \boldsymbol{0})$. It is impossible to say something in general about this probability, i.e. for a meta-elliptical copula with a general correlation structure $\boldsymbol{R}$. However, results can be established for some specific correlation structures. A first structure is when the correlation matrix $\boldsymbol{R}$ takes the form

$$\varrho_{i,j} = \lambda_i \lambda_j,\ i \neq j, \lambda_j \in [-1, 1]. \tag{27}$$

Note that this correlation structure covers, among others, an equicorrelated correlation matrix for which $\varrho_{i,j} = \varrho$ for all $i \neq j$, with $\varrho \in (-1/(d-1), 1)$. A second structure is when the correlation matrix is of a banded type and $(m+2)$-diagonal (with $m \in \mathbb{Z}_{>0}$), and takes the form

$$\varrho_{i,j} = \begin{cases} 1, & \text{if } i = j \\ c_{i,j} & \text{if } |i-j| \in \{1, \cdots, m\} \\ 0, & \text{if } |i-j| > m, \end{cases} \tag{28}$$

where $c_{i,j} = c_{j,i} \in [-1, 1]$ are constants, not all zero, and such that $\boldsymbol{R}$ is a correlation matrix. For $m = 1$ for example, one obtain a so-called tridiagonal matrix, with only non-zero values on the main diagonal and the two adjacent diagonals, i.e. the diagonals just above and below the main diagonal. Proposition 9 establishes that under both particular correlation structures, Kendall's tau tends to zero when d increases to infinity.

Proposition 9. *Let $\{C_d\}$ be a sequence of d-dimensional meta-elliptical copulas with a correlation matrix $\boldsymbol{R} = (\varrho_{i,j})$*

(i) *as in* (27) *where the λ_j satisfy the assumption that there exists $\lambda_0 < 1$ such that $\lambda_j \leq \lambda_0$ for all $j \in \mathbb{Z}_{>0}$;*

or

(ii) *as in* (28).

In both cases, it holds that $\lim_{d\to\infty} \tau(C_d) = 0$.

Proof. Under each of the correlation structures we evaluate the orthant probability $\Pr(\boldsymbol{Z} \geq \boldsymbol{0})$ in (26). Note first of all that for a centered Gaussian random vector $\boldsymbol{Z}$, $\Pr(\boldsymbol{Z} \geq \boldsymbol{0}) = \Pr(\boldsymbol{Z} \leq \boldsymbol{0})$.

(i) Under a correlation structure (27) the probability $\Pr(\boldsymbol{Z} \leq \boldsymbol{0})$ equals

$$\Pr(\boldsymbol{Z} \leq \boldsymbol{0}) = \int_{-\infty}^{\infty} \varphi(t) \prod_{j=1}^{d} \Phi\left(\frac{-\lambda_j t}{\sqrt{1-\lambda_j^2}}\right) \mathrm{d}t, \tag{29}$$

where φ and Φ are respectively the density and cumulative distribution function of the univariate standard normal distribution. This follows from Dunnet and Sobel [5, Expression (5), page 259] or Gupta [10, Expression (34), page 800]. If the parameters λ_j are such that $\lambda_j \leq \lambda_0 < 1$ for all $j \in \mathbb{Z}_{>0}$, then

$$\Phi\Big(\tfrac{-\lambda_j t}{\sqrt{1-\lambda_j^2}}\Big) \leq \Phi\Big(\tfrac{|t|}{\sqrt{1-\lambda_0^2}}\Big), \quad \forall t \in \mathbb{R},\ \forall j \in \mathbb{Z}_{>0},$$

and hence $$0 \le \Pr(\mathbf{Z} \le \mathbf{0}) \le \int_{-\infty}^{\infty} \varphi(t) \left(\Phi\Big(\frac{|t|}{\sqrt{1-\lambda_0^2}} \Big) \right)^d \mathrm{d}t.$$

An application of the Lebesgue dominated convergence theorem gives that the right-hand side converges to zero as $d \to \infty$. From (26) the statement of the proposition holds.

(ii) Assume correlation structure (28). Then the random variables Z_j, $j = k(m+1) - m$, $k \in \left\{1, \ldots, \left\lfloor \frac{d}{m+1} \right\rfloor \right\}$, are independent standard normal variables and thus

$$0 \le \Pr(\mathbf{Z} \le 0) \le \Pr\left(Z_{k(m+1)-m} \le 0; 1 \le k \le \left\lfloor \frac{d}{m+1} \right\rfloor \right) \le \left(\frac{1}{2} \right)^{\left\lfloor \frac{d}{m+1} \right\rfloor},$$

and the right-hand side tends to zero for d tending to infinity. Using (26) finishes the proof. □

For meta-elliptical copulas it does not seem to be possible to derive analytical expressions for Spearman's rho and Gini's gamma for general dimension d. This is even so for the simplest case of an equicorrelation matrix $\boldsymbol{R}$. From Monte-Carlo approximations provided in Schmid and Schmidt [24, p. 5, 7] one might conjecture that Spearman's rho decreases when d increases in the case of an equicorrelation matrix $\boldsymbol{R}$. However, a formal proof for such a result is lacking. Also for Gini's gamma no results for meta-elliptical copulas for increasing dimension d can be given yet.

6. Illustrative example: a four-dimensional Gaussian copula

Let $\boldsymbol{X}$ be a 4-variate random vector with Gaussian copula $C_{4,\Sigma}$ depending on the correlation matrix Σ. We illustrate the impact of the (structure of the) correlation matrix Σ on multivariate association measures, by considering several correlation structures:

- *AR* structure: the correlation values decrease with the distance to the main diagonal;
- *Clusters* structure: the values of the correlation matrix are following two clusters of two variables each;
- *Renegade* structure: one variable is negatively correlated with the other three variables, which are strongly correlated with each other.

The elements of the correlation matrices were selected such that the sum of the six elements above (or equivalently below) the diagonal is the same (equals 2.2) in each structure, with the average of all pairwise Pearson's correlation coefficients equal to 0.37. The specific correlation matrices are:

$$\Sigma_{AR} = \begin{pmatrix} 1.0 & 0.5 & 0.3 & 0.1 \\ 0.5 & 1.0 & 0.5 & 0.3 \\ 0.3 & 0.5 & 1.0 & 0.5 \\ 0.1 & 0.3 & 0.5 & 1.0 \end{pmatrix}, \quad \Sigma_{Cl} = \begin{pmatrix} 1.0 & 0.8 & 0.1 & 0.0 \\ 0.8 & 1.0 & 0.2 & 0.3 \\ 0.1 & 0.2 & 1.0 & 0.8 \\ 0.0 & 0.3 & 0.8 & 1.0 \end{pmatrix}, \quad \Sigma_{Re} = \begin{pmatrix} 1.0 & -0.2 & -0.1 & -0.2 \\ -0.2 & 1.0 & 0.9 & 0.9 \\ -0.1 & 0.9 & 1.0 & 0.9 \\ -0.2 & 0.9 & 0.9 & 1.0 \end{pmatrix}.$$

To calculate the various copula-based multivariate association measures, we use numerical approximation of the involved integrals, using the R package `cubature`. For the pairwise measures, closed form expressions using the elements of the correlation matrix are known, see e.g. [15]. Table 1 summarizes the results of these calculations for the different correlation structures. Recall first, from Sections 5.2 and S2, that Kendall's tau τ and Blomqvist beta β are equal for the Gaussian copula. Furthermore, in case of a copula C that is radially symmetric, which is the case for a Gaussian copula $C = C_{4,\Sigma}$, Spearman's rho association measures $\rho_1(C)$ and $\rho_2(C)$ coincide, and hence $\rho_1(C) = \rho_2(C) = \rho_3(C)$. From Table 1 one can see that for the clusters structure, the copula-based and pairwise association measures convey the same messages. There is a slightly bigger difference between the copula-based and pairwise association measures for the AR structure. The difference is most prevalent for the renegade structure, since the dependence lies mainly in the pairs. Furthermore, the renegade structure shows the lowest copula-based Spearman's rho of the three groups but the largest pairwise Spearman's rho. This indicates the importance of both approaches to obtain a more detailed picture of all dependencies.

In Example S5 in Section S3 we provide some additional material in the context of this example.

Table 1: Multivariate association measures for Gaussian copula with various correlation structure.

	ρ_1=ρ_2=ρ_3	$\tau=\beta$	γ_2	ρ^{PW}	$\tau^{\text{PW}}=\beta^{\text{PW}}$	γ^{PW}
AR	0.34	0.22	0.25	0.35	0.24	0.28
Clusters	0.36	0.28	0.29	0.36	0.26	0.29
Renegade	0.32	0.25	0.26	0.37	0.30	0.32

7. Estimation of multivariate association measures

7.1. Estimation of multivariate association measures

This section studies nonparametric estimation of the multivariate association measures discussed in Section 3. Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be a random sample of d-dimensional random vectors with copula C_d where $\boldsymbol{X}_i = (X_{1,i}, \ldots X_{d,i})^\top$ for $i \in \{1, \ldots n\}$. Throughout this section, the dimension d of a copula C_d is arbitrary but fixed and thus for simplicity of notation we omit the subscript d in C_d. The empirical marginal distribution function for the j-th component is defined as

$$\widehat{F}_{j,n}(x) = \frac{1}{n+1} \sum_{i=1}^{n} \mathbb{1}(X_{j,i} \le x), \tag{30}$$

the pseudo-observations $\widehat{U}_{j,i} = \widehat{F}_{j,n}(X_{j,i})$, and the empirical copula

$$\widehat{C}_n(\boldsymbol{u}) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(\widehat{U}_{1,i} \le u_1, \ldots, \widehat{U}_{d,i} \le u_d). \tag{31}$$

Similarly, we define the empirical survival copula and the empirical survival function as, respectively,

$$\widehat{C}^S_n(\boldsymbol{u}) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(1-\widehat{U}_{1,i} \le u_1, \ldots, 1-\widehat{U}_{d,i} \le u_d), \qquad \widehat{\overline{C}}_n(\boldsymbol{u}) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(\widehat{U}_{1,i} > u_1, \ldots, \widehat{U}_{d,i} > u_d),$$

A fully nonparametric estimator of an association measure $\kappa(C)$ is then mostly constructed as $\widehat{\kappa}_n = \kappa(\widehat{C}_n)$.

Denote by $\ell^\infty([0,1]^d)$ the space of bounded functions from $[0,1]^d$ to $\mathbb{R}$ equipped with the topology of uniform convergence. The following proposition combines results of [25, 28, 32], and also states the results in terms of i.i.d. representations.

Proposition 10 (Empirical copula process convergence)**.** *Let* $\frac{\partial C}{\partial u_j}$ *be continuous in points* $\{\boldsymbol{u} \in [0,1]^d; 0 < u_j < 1\}$ *for* $j \in \{1, \ldots, d\}$*. Then uniformly in* $\boldsymbol{u} \in [0,1]^d$

$$\mathbb{C}_n(\boldsymbol{u}) = \sqrt{n}\left(\widehat{C}_n(\boldsymbol{u}) - C(\boldsymbol{u})\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i(\boldsymbol{u}) + o_P(1), \qquad \overline{\mathbb{C}}_n(\boldsymbol{u}) = \sqrt{n}\left(\widehat{\overline{C}}_n(\boldsymbol{u}) - \overline{C}(\boldsymbol{u})\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \overline{Z}_i(\boldsymbol{u}) + o_P(1),$$

where

$$Z_i(\boldsymbol{u}) = \mathbb{1}(\boldsymbol{U}_i \le \boldsymbol{u}) - C(\boldsymbol{u}) - \sum_{j=1}^{d} \frac{\partial C(\boldsymbol{u})}{\partial u_j} \left[\mathbb{1}(U_{j,i} \le u_j) - u_j\right], \quad \overline{Z}_i(\boldsymbol{u}) = \mathbb{1}(\boldsymbol{U}_i > \boldsymbol{u}) - \overline{C}(\boldsymbol{u}) - \sum_{j=1}^{d} \frac{\partial \overline{C}(\boldsymbol{u})}{\partial u_j} \left[\mathbb{1}(U_{j,i} \le u_j) - u_j\right].$$

From this one can further conclude that the processes $\mathbb{C}_n$ *and* $\overline{\mathbb{C}}_n$ *jointly converge in the weak sense to a centered Gaussian processes* $\mathbb{G}_C(\boldsymbol{u})$ *and* $\mathbb{G}_{\overline{C}}(\boldsymbol{u})$*, respectively, in* $\ell^\infty([0,1]^d)$ *as* $n \to \infty$*. Moreover,*

$$\mathbb{G}_C(\boldsymbol{u}) = \mathbb{B}_C(\boldsymbol{u}) - \sum_{j=1}^{d} \frac{\partial C(\boldsymbol{u})}{\partial u_j} \mathbb{B}_C(\boldsymbol{u}^{(j)}), \qquad \mathbb{G}_{\overline{C}}(\boldsymbol{u}) = \mathbb{B}_{\overline{C}}(\boldsymbol{u}) - \sum_{j=1}^{d} \frac{\partial \overline{C}(\boldsymbol{u})}{\partial u_j} \mathbb{B}_{\overline{C}}(\boldsymbol{u}^{(j)}),$$

where $\boldsymbol{u}^{(j)}$ is a vector with all components except of the j-th replaced by 1. Further, $\mathbb{B}_C$ and $\mathbb{B}_{\overline{C}}$ are Brownian bridges with covariance functions

$$E(\mathbb{B}_C(\boldsymbol{u})\mathbb{B}_C(\boldsymbol{v})) = C(\boldsymbol{u} \wedge \boldsymbol{v}) - C(\boldsymbol{u})C(\boldsymbol{v}), \qquad E(\mathbb{B}_{\overline{C}}(\boldsymbol{u})\mathbb{B}_{\overline{C}}(\boldsymbol{v})) = \overline{C}(\boldsymbol{u} \vee \boldsymbol{v}) - \overline{C}(\boldsymbol{u})\overline{C}(\boldsymbol{v}),$$

where minimum $\wedge$ and maximum $\vee$ are understood component-wise.

Estimation of Spearman's rho

The following estimators for Spearman's rho were proposed by [25] as

$$\widehat{\rho}^*_{1n} = h_\rho(d)\left\{2^d \int\limits_{[0,1]^d} \widehat{C}_n(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} - 1\right\} = h_\rho(d)\left\{\frac{2^d}{n}\sum_{i=1}^n\prod_{j=1}^d(1-\widehat{U}_{j,i}) - 1\right\},$$
$$\widehat{\rho}^*_{2n} = h_\rho(d)\left\{2^d \int\limits_{[0,1]^d} \Pi(\boldsymbol{u})\,\mathrm{d}\widehat{C}_n(\boldsymbol{u}) - 1\right\} = h_\rho(d)\left\{\frac{2^d}{n}\sum_{i=1}^n\prod_{j=1}^d\widehat{U}_{j,i} - 1\right\}$$
$$\widehat{\rho}^*_{3n} = \frac{\widehat{\rho}^*_{1n} + \widehat{\rho}^*_{2n}}{2}.$$

Note that [25] originally used division by n instead of $n+1$ in (30). However, as pointed out by [19], estimators $\widehat{\rho}^*_{\ell n}$ for $\ell \in \{1,2,3\}$ can take values out of the parameter space, for example, they can be greater than 1. [19] further suggest to define $\widehat{\rho}_{\ell n}$ as the ratio $\widehat{\rho}_{\ell n} = \widehat{\rho}^*_{\ell n}/m_{d,n}$ where $m_{d,n}$ is the maximal possible value $\widehat{\rho}^*_{\ell n}$ can take. That is

$$m_{d,n} = h_\rho(d)\left(\frac{2^d}{n(n+1)^d}\sum_{j=1}^n j^d - 1\right)$$

which is achieved by evaluating $\widehat{\rho}^*_{\ell n}$ for a sample from the comonotonicity copula M, that is if $\widehat{U}_{1,i} = \cdots = \widehat{U}_{d,i}$ for every $i \in \{1,\ldots,n\}$ almost surely. Moreover, since $m_{d,n}$ converges to 1 as $n \to \infty$, the asymptotic theory is valid also for $\widehat{\rho}_{\ell n}$.

[25] state that under the assumptions and notation of Proposition 10, as $n \to \infty$,

$$\sqrt{n}(\widehat{\rho}^*_{\ell n} - \rho_\ell(C)) \xrightarrow{D} \mathcal{N}(0,\sigma^2_\ell), \qquad \text{for } \ell \in \{1,2\},$$

where the asymptotic variances are given as

$$\sigma^2_1 = 2^{2d}h^2_\rho(d)\int\limits_{[0,1]^d}\int\limits_{[0,1]^d} \mathsf{E}\{\mathbb{G}_C(\boldsymbol{u})\mathbb{G}_C(\boldsymbol{v})\}\,\mathrm{d}\boldsymbol{u}\,\mathrm{d}\boldsymbol{v}, \qquad \sigma^2_2 = 2^{2d}h^2_\rho(d)\int\limits_{[0,1]^d}\int\limits_{[0,1]^d} \mathsf{E}\{\mathbb{G}_{\overline{C}}(\boldsymbol{u})\mathbb{G}_{\overline{C}}(\boldsymbol{v})\}\,\mathrm{d}\boldsymbol{u}\,\mathrm{d}\boldsymbol{v}. \tag{32}$$

[25] further remark that asymptotic normality of $\widehat{\rho}_{3n}$ can be established in a similar way using joint convergence of processes $C_n(\boldsymbol{u})$ and $\overline{C}_n(\boldsymbol{u})$.

Estimation of Kendall's tau

The estimator of τ is proposed by [8] as

$$\widehat{\tau}_n = \frac{1}{2^{d-1}-1}\left\{\frac{2^d}{n(n-1)}\left[\sum_{i\neq j}\mathbb{1}\left(\boldsymbol{X}_i \leq \boldsymbol{X}_j\right)\right] - 1\right\}, \tag{33}$$

since $\int_{[0,1]^d} C(\boldsymbol{u})\,\mathrm{d}C(\boldsymbol{u}) = \Pr(\boldsymbol{X} \leq \boldsymbol{Y})$, where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are iid d-variate random vectors with copula C. [8] establish that

$$\sqrt{n}(\widehat{\tau}_n - \tau(C)) \xrightarrow{D} \mathcal{N}(0,\sigma^2_\tau),$$

as $n \to \infty$, where

$$\sigma_\tau^2 = \left(\frac{2^d}{2^{d-1}-1}\right)^2 \operatorname{var}(C(\boldsymbol{U}) + \overline{C}(\boldsymbol{U})). \tag{34}$$

Note that for a sample from the comonotonicity copula, $\widehat{\tau}_n = 1$ since in that case $\boldsymbol{X}_i < \boldsymbol{X}_j$ or $\boldsymbol{X}_i > \boldsymbol{X}_j$ for every pair (i, j) such that $i \neq j$ and thus exactly one half of the $n(n-1)$ indicators in (33) is equal to 1.

Estimation of Gini's gamma

[31] proposed the population version γ_2 in (15), but did not discuss its estimation. A natural nonparametric estimator of $\gamma_2(C)$ is $\widehat{\gamma}^*_{2n} = \gamma_2(\widehat{C}_n)$. Recall that S_ξ is the subset of indices $\{1, \ldots, d\}$ indicating which components were reflected ξ. As the pseudo-observations $\widehat{U}_{j,i}$ take values between zero and one, then with a slight abuse of notations by $\max_{j \in S_\xi}(\widehat{U}_{j,i})$ (respectively $\min_{j \in S_\xi}(\widehat{U}_{j,i})$) we will understand 0 (respectively 1) if S_ξ is an empty set.

Lemma 2. *The nonparametric estimator $\widehat{\gamma}^*_{2n}$ of $\gamma_2(C)$ can be expressed as*

$$\widehat{\gamma}^*_{2n} = \frac{1}{2^{d-1}-1}\left(\sum_{\xi \in \mathcal{R}_d} (I_1(\widehat{C}_n, \xi) + I_2(\widehat{C}_n, \xi)) - 2\right)$$

where

$$I_1(\widehat{C}_n, \xi) = \frac{1}{n}\sum_{i=1}^n (1 - \max_{j \in S_\xi}(\widehat{U}_{j,i}) - \max_{k \notin S_\xi}(\widehat{U}_{k,i}))_+, \qquad I_2(\widehat{C}_n, \xi) = \frac{1}{n}\sum_{i=1}^n (\min_{j \in S_\xi}(\widehat{U}_{j,i}) + \min_{k \notin S_\xi}(\widehat{U}_{k,i}) - 1)_+.$$

Proof. Using (16), we can write

$$\widehat{\gamma}^*_{2n} = \gamma_2(\widehat{C}_n) = \frac{1}{2^{d-1}-1}\left(\sum_{\xi \in \mathcal{R}_d} \int_0^1 (\widehat{C}_n(\xi(u, \ldots, u)) + \widehat{C}^S_n(\xi(u, \ldots, u)))\, \mathrm{d}u - 2\right).$$

First, we fix $\xi \in \mathcal{R}_d$ and treat the summands in the above integral separately. Then by plugging in the empirical copula (31) and using the convention that a product over an empty index set is equal to 1, we get

$$\begin{aligned}
\int_0^1 \widehat{C}_n(\xi(u, \ldots, u))\, \mathrm{d}u &= \frac{1}{n}\sum_{i=1}^n \int_0^1 \prod_{j \in S_\xi} \mathbb{1}(\widehat{U}_{j,i} \le 1 - u) \prod_{k \notin S_\xi} \mathbb{1}(\widehat{U}_{k,i} \le u)\, \mathrm{d}u \\
&= \frac{1}{n}\sum_{i=1}^n \int_0^1 \mathbb{1}(\max_{k \notin S_\xi}(\widehat{U}_{k,i}) \le u \le \min_{j \in S_\xi}(1 - \widehat{U}_{j,i}))\, \mathrm{d}u \\
&= \frac{1}{n}\sum_{i=1}^n \int_0^1 \mathbb{1}(\max_{k \notin S_\xi}(\widehat{U}_{k,i}) \le u \le 1 - \max_{j \in S_\xi}(\widehat{U}_{j,i}))\, \mathrm{d}u = I_1(\widehat{C}_n, \xi).
\end{aligned}$$

Similarly

$$\begin{aligned}
\int_0^1 \widehat{C}^S_n(\xi(u, \ldots, u))\, \mathrm{d}u &= \frac{1}{n}\sum_{i=1}^n \int_0^1 \prod_{j \in S_\xi} \mathbb{1}(1 - \widehat{U}_{j,i} \le 1 - u) \prod_{k \notin S_\xi} \mathbb{1}(1 - \widehat{U}_{k,i} \le u)\, \mathrm{d}u \\
&= \frac{1}{n}\sum_{i=1}^n \int_0^1 \mathbb{1}(1 - \min_{k \notin S_\xi}(\widehat{U}_{k,i}) \le u \le \min_{j \in S_\xi}(\widehat{U}_{j,i}))\, \mathrm{d}u = I_2(\widehat{C}_n, \xi),
\end{aligned}$$

which completes the proof. □

In this case, similarly as for estimators of Spearman's rho, room for improvement is to be seen through the following example.

Example 4 (γ_2 estimator for maximal dependence)**.** Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be a random sample of d-dimensional random vectors from the d-variate comonotonicity copula M. Then $\widehat{U}_{1,i} = \cdots = \widehat{U}_{d,i}$ for every $i \in \{1, \ldots, n\}$ almost surely. Then

$$I_1(\widehat{M}_n, \xi) = \begin{cases} \frac{1}{n}\sum_{i=1}^n (1-\widehat{U}_{1,i})_+ = \frac{1}{n}\sum_{i=1}^n (1-\widehat{U}_{1,i}) & \text{if } S_\xi = \emptyset \text{ or } S_\xi = \{1,\ldots,d\}, \\ \frac{1}{n}\sum_{i=1}^n (1-2\widehat{U}_{1,i})_+ & \text{otherwise} \end{cases}$$

and similarly

$$I_2(\widehat{M}_n, \xi) = \begin{cases} \frac{1}{n}\sum_{i=1}^n (\widehat{U}_{1,i})_+ = \frac{1}{n}\sum_{i=1}^n \widehat{U}_{1,i} & \text{if } S_\xi = \emptyset \text{ or } S_\xi = \{1,\ldots,d\}, \\ \frac{1}{n}\sum_{i=1}^n (2\widehat{U}_{1,i}-1)_+ & \text{otherwise.} \end{cases}$$

By adding these terms together, we get

$$I_1(\widehat{M}_n, \xi) + I_2(\widehat{M}_n, \xi) = \begin{cases} 1 & \text{if } S_\xi = \emptyset \text{ or } S_\xi = \{1,\ldots,d\}, \\ \frac{1}{n}\sum_{i=1}^n |1-2\widehat{U}_{1,i}| & \text{otherwise.} \end{cases}$$

Recall that in total we have 2^d reflections in $\mathcal{R}_d$ and thus

$$\widehat{\gamma}^*_{2n} = \frac{1}{2^{d-1}-1}\left(2\cdot 1 + (2^d-2)\frac{1}{n}\sum_{i=1}^n |1-2\widehat{U}_{1,i}| - 2\right) = \frac{2}{n}\sum_{i=1}^n |1-2\widehat{U}_{1,1}|.$$

Without loss of generality, assume that $\widehat{U}_{1,i} = \cdots = \widehat{U}_{d,i} = \frac{i}{n+1}$ for every $i \in \{1, \ldots, n\}$. Then

$$\widehat{\gamma}^*_{2n} = \frac{2}{n}\sum_{i=1}^n \left|1 - \frac{2i}{n+1}\right| = \begin{cases} \frac{n}{n+1} & \text{if } n \text{ even,} \\ \frac{n-1}{n} & \text{if } n \text{ odd.} \end{cases}$$

The previous example shows that the range of values for $\widehat{\gamma}^*_{2n}$ is not wide enough and one could consider an alternative and asymptotically equivalent estimator

$$\widehat{\gamma}_{2n} = \begin{cases} \frac{n+1}{n}\widehat{\gamma}^*_{2n} & \text{if } n \text{ even,} \\ \frac{n}{n-1}\widehat{\gamma}^*_{2n} & \text{if } n \text{ odd.} \end{cases}$$

Asymptotic normality of $\widehat{\gamma}_{2n}$ is established in Proposition 11.

Proposition 11. *Suppose that the assumptions of Proposition 10 are satisfied. Then*

$$\sqrt{n}\left(\widehat{\gamma}_{2n} - \gamma_2(C)\right) \xrightarrow{D} \mathcal{N}\left(0, \sigma^2_{\gamma_2}\right),$$

as $n \to \infty$, *where*

$$\sigma^2_{\gamma_2} = \frac{1}{(2^{d-1}-1)^2}\sum_{\xi\in\mathcal{R}_d}\sum_{\xi'\in\mathcal{R}_d}\int_0^1\int_0^1 \mathsf{E}\Big\{\Big[\mathbb{G}_C(\xi(u,\ldots,u)) + \mathbb{G}_{\overline{C}}(\xi(1-u,\ldots,1-u))\Big]$$
$$\Big[\mathbb{G}_C(\xi'(v,\ldots,v)) + \mathbb{G}_{\overline{C}}(\xi'(1-v,\ldots,1-v))\Big]\Big\}\,\mathrm{d}u\,\mathrm{d}v.$$

Proof. This proof is analogous to the proof of Theorem 3 for $\widehat{\rho}_{1n}$ in [25]. Rewrite

$$\sqrt{n}\left(\widehat{\gamma}^*_{2n} - \gamma_2\right) = \frac{1}{2^{d-1}-1}\sum_{\xi\in\mathcal{R}_d}\int_0^1 \mathbb{C}_n(\xi(u,\ldots,u)) + \overline{\mathbb{C}}_n(\xi(1-u,\ldots,1-u))\,\mathrm{d}u.$$

Note that the mapping on the set of bounded functions $\ell^\infty([0,1]^d)$ given by

$$\alpha \mapsto \frac{1}{2^{d-1}-1} \sum_{\xi \in \mathcal{R}_d} \int_0^1 \alpha(\xi(u,\ldots,u)) + \overline{\alpha}(\xi(1-u,\ldots,1-u))\, \mathrm{d}u$$

is continuous. Thus by the continuous mapping theorem [see e.g. Theorem 1.3.6 in 34] and the joint convergence of the processes C_n and $\overline{C}_n$ given in Proposition 8 the quantity $\sqrt{n}\left(\widehat{\gamma}^*_{2n} - \gamma_2(C)\right)$ converges in distribution to the random variable

$$Z = \frac{1}{2^{d-1}-1} \sum_{\xi \in \mathcal{R}_d} \int_0^1 \mathbb{G}_C(\xi(u,\ldots,u)) + \mathbb{G}_{\overline{C}}(\xi(1-u,\ldots,1-u))\, \mathrm{d}u.$$

Now the normality of Z follows from the fact that $\mathbb{G}_C$ and $\mathbb{G}_{\overline{C}}$ are Gaussian processes. Finally with the help of Fubini's theorem one can calculate the mean and variance of Z. The result for $\widehat{\gamma}_{2n}$ follows immediately. □

7.2. Standard errors of the estimators

In real data settings we not only want the estimate the association measures but also are interested in the accuracy of this estimation. One way to measure accuracy is by providing standard errors for the estimations.

A first approach to get approximate standard errors is by using standard nonparametric bootstrap procedures. See for example [25] or [26] for such an approach. An alternative to resampling methods is to rely on the asymptotic normality results for the estimators in Section 7.1, and to work towards an estimator for the square root of the asymptotic variance. In doing so one could rely on Proposition 10, which allows to obtain asymptotic representations of estimators of multivariate association measures. One then focuses on finding the standard errors of the main terms in these approximations. Finally one replaces the unknown $\boldsymbol{U}_i$ and $\frac{\partial C(\boldsymbol{u})}{\partial u_j}$ with appropriate estimates. We discuss here for the estimator of Kendall's tau the approach of using the asymptotic distributional result. For similar results for estimators of other multivariate association measures, we refer the reader to Section S4 in the Supplementary Material.

Recall the asymptotic variance of the estimator of Kendall's tau in (34). The variance $\mathrm{var}(C(\boldsymbol{U}) + \overline{C}(\boldsymbol{U}))$ can be estimated, using ideas from U-statistics, by the sample variance of $W_1,\ldots,W_n$, i.e.

$$\widehat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^n (W_i - \overline{W}_n)^2, \quad \text{where} \quad \overline{W}_n = \frac{1}{n}\sum_{i=1}^n W_i, \tag{35}$$

with

$$W_i = \frac{1}{n-1}\sum_{j=1, j\neq i}^n \mathbb{1}(\boldsymbol{X}_j \le \boldsymbol{X}_i) + \frac{1}{n-1}\sum_{j=1, j\neq i}^n \mathbb{1}(\boldsymbol{X}_j \ge \boldsymbol{X}_i).$$

Consequently, the standard error of $\widehat{\tau}_n$ is estimated by $\frac{2^d}{2^{d-1}-1}\,\frac{\widehat{\sigma}_n}{\sqrt{n}}$.

8. Real data application

We now shortly demonstrate the use of the studied association measures in a real data example. The data discussed in this section are publicly available data on so called Environmental Quality Index (EQI) and can be downloaded from the website of United States Environmental Protection Agency `https://edg.epa.gov/data/Public/ORD/NHEERL/EQI`. EQI is being produced based on variables from five domains – air, water, land, built and sociodemographic. The dataset consists of 3141 observations, representing different locations across the United States, and 219 variables. For further information about the variables and a discussion about the data sources, we refer to [13].

For purpose of this real data application, we chose four groups of variables, three using variables from the same domain (air, water and land) and one combining variables from these three domains. Levels of chemical substances measured in the domains were considered. Since we assume that all the variables are continuous, we restricted to variables with a high percentage of unique values in the dataset. In the air domain, all variables considered have no ties in their observations. In the water and land domains, all variables included in our analysis have more than 90% unique values among the observations. The considered variables are listed in Table 2. To gain some insight in the data we provide in Figure S3 a heatmap of the empirical pairwise Spearman's rho values. A first observation is that all nine variables in the air domain are positively associated. This is also mostly the case, for three of the four variables, in the land domain. In this domain herbicides are less associated with the metal elements. The nine variables in the water domain are either positively or negatively related with the other variables in that domain. Note also that none of the 22 variables have a negative (nor positive) association with *all* other 21 variables. For some of the variables, one can notice that there is an association between the variable and all variables from another domain. Examples here are the variables Nitrate, Sodium and Mercury of the water domain that all seem to be negatively associated with all variables of the air domain.

Table 2: List of selected variables. Units do not affect the analysis and are thus omitted.

Domain	Dimension	Variables
Air	9	Acrolein, Acrylonitrile, Carbon disulfide, Chlorobenzene, Glycol ethers, Methanol, Methyl isobutyl ketone, Polycyclic organic matter/polycyclic aromatic hydrocarbons, Selenium compounds
Water	9	Ammonium, Calcium, Chloride, Magnesium, Nitrate, Potassium, Sodium, Sulfate (all in precipitation), Mercury (deposited)
Land	4	Lead, Zinc, Copper, Herbicides
Combined	22	All the above variables

The following estimators of multivariate association measure are considered for the EQI dataset: $\widehat{\rho}_n^{\mathrm{PW}}$, $\widehat{\tau}_n^{\mathrm{PW}}$, $\widehat{\beta}_n^{\mathrm{PW}}$, $\widehat{\gamma}_n^{\mathrm{PW}}$, $\widehat{\rho}_{3n}$, $\widehat{\tau}_n$, $\widehat{\beta}_n$ and $\widehat{\gamma}_{2n}$. For the estimator $\widehat{\beta}_n$ of Blomqvist's beta we refer the reader to Section S2 of the Supplementary Material. All the calculations for the estimators were done in the statistical software R [20]. For each of the estimated association measures we also calculate the estimated standard errors, using a standard nonparametric bootstrap procedure. We also calculated the standard errors using the estimators discussed in Sections 7.2 and S4, which were very similar (and hence are not included in the table).

Table 3: Estimated multivariate association measures for the EQI dataset and their corresponding bootstrap standard errors between brackets.

Domain	d	$\widehat{\rho}_n^{\mathrm{PW}}$	$\widehat{\tau}_n^{\mathrm{PW}}$	$\widehat{\beta}_n^{\mathrm{PW}}$	$\widehat{\gamma}_n^{\mathrm{PW}}$	$\widehat{\rho}_{3n}$	$\widehat{\tau}_n$	$\widehat{\beta}_n$	$\widehat{\gamma}_{2n}$
Air	9	0.60 (0.007)	0.44 (0.006)	0.44 (0.008)	0.49 (0.006)	0.44 (0.008)	0.29 (0.007)	0.29 (0.008)	0.29 (0.006)
Water	9	0.05 (0.005)	0.04 (0.004)	−0.01 (0.003)	0.03 (0.004)	0.04 (0.004)	0.03 (0.002)	0.01 (0.002)	0.02 (0.002)
Land	4	0.33 (0.011)	0.24 (0.008)	0.23 (0.010)	0.27 (0.009)	0.30 (0.011)	0.21 (0.008)	0.20 (0.011)	0.23 (0.009)
Combined	22	0.11 (0.003)	0.09 (0.002)	0.08 (0.003)	0.09 (0.002)	0.00 (0.000)	0.00 (0.000)	−0.00 (0.000)	0.00 (0.000)
AirPlus	12	0.50 (0.007)	0.36 (0.005)	0.36 (0.007)	0.40 (0.006)	0.23 (0.008)	0.15 (0.005)	0.15 (0.007)	0.14 (0.006)
WaterClus5	5	0.45 (0.008)	0.33 (0.006)	0.34 (0.010)	0.37 (0.007)	0.41 (0.007)	0.30 (0.006)	0.30 (0.009)	0.31 (0.007)
WaterClus6	6	0.30 (0.007)	0.22 (0.005)	0.21 (0.009)	0.24 (0.006)	0.22 (0.006)	0.15 (0.005)	0.12 (0.008)	0.15 (0.006)

Table 3 presents the estimated multivariate association measures, with within brackets the estimated standard errors. We first look into the domains listed in Table 2. See the first block of rows in Table 3. For variables in the air domain, one can see moderately strong association among the variables, covered by both types of the association measure (pairwise-based and copula-based). Pairwise-based association measures, however, show that the strongest association lies in the pairwise structure of these variables and higher-order association is lower. The variables in the water domain, that is the concentration of the considered chemical substances in water, on the other hand, possess, very low levels of association, measured by all 8 multivariate association measures. In addition, the value of $\widehat{\beta}_n^{\mathrm{PW}}$ for this domain is even negative, unlike all the remaining association measures. The variables in the land domain, similarly to the air domain, are moderately associated, with the estimated association measures ranging from 0.20 to 0.33. Nevertheless, in this case pairwise-based and copula-based association measures are very alike, suggesting that the higher-order association is of a similar level as the pairwise associations. Finally, the estimated association measures for the set of variables from various domains are low, showing that these pollutants from 3 domains are almost not associated. Also note that for the considered domains, Spearman's rho is the largest among the association measures within each of the two approaches.

So far we focused on estimated bivariate association measures (in Figure S3) and estimated multivariate association measures for a whole set of variables in one domain, or in the combined set of variables of all domains. To gain further insights of the associations in the data, one can investigate various subsets of the variables. The selection of such subsets might come either from background knowledge on the topic of the study, the pairwise correlation structure or might emerge from an automatized procedure looking for such subsets. The latter is studied in [7] in the context of cluster analysis.

In this application we looked at the following clusters of variables. A first cluster consists of five variables within the water domain, namely Nitrate, Ammonium, Magnesium, Potassium and Calcium, and is referred to as *WaterClus5*. A second cluster, the *WaterClus6* group, is this group extended with the variable Sulfate. A third cluster, *AirPlus*, consists of all variables of the air domain together with the variables, Sulfate, Mercury and Chloride from the water domain. See Section S5 in the Supplementary Material for some motivation for looking into these clusters. Estimated multivariate association measures for these three subgroups are presented in the last three rows of Table 3. Note that the addition of the variable Sulfate to *WaterClus5* results in a considerable drop in the estimated association measure. From this one could decide for further analysis to focus on the cluster *WaterClus5* instead of on the cluster *WaterClus6*. Furthermore, it is noted that the estimated multivariate association measures for the *WaterClus5* cluster are very close to these for the set of all nine air variables. The overall association in the cluster *AirPlus* is very comparable to that of the cluster *WaterClus6*.

Another aspect of estimating multivariate association measures is the computational cost. In general, estimation of the copula-based β is the fastest one, followed by estimation of ρ_3. Both these estimators are also not largely affected by the dimension of the data. The estimators of the pairwise-based association measures are also fast to compute. On the other hand, estimation time of γ_2 dramatically increases with dimension, corresponding to the need of calculating all the 2^d reflections of the data. Also calculation of $\widehat{\tau}_n$ is rather computationally intensive for large sample sizes since it is based on pairwise comparison of the observations, see (33).

9. Multivariate association measures: overview and recommendation

From all investigations, it is clear that a complete comparison of the different association measures is complex since there are various aspects to be considered and taken into account. These include: (i) the interpretation that can be given to the measures; (ii) the level of complexity of the dependence structure that the measure can potentially capture; (iii) do the measures satisfy axioms (A_1) – (A_8)?; (iv) do the measures increase (or stay the same) when a duplicated component or, more generally, a conical combination of all components is added?; (v) do we know the behaviour when a number of arbitrary components is added?; (vi) how much do we know about the behaviour when the number of components tends to infinity?; (vii) how easy to calculate is the measure for a given copula C_d?; (viii) what about the computational cost to calculate the empirical versions of the association measures?

On some of these aspects certain association measures score well, whereas they score to a lesser extend on other aspects. In Table S2 we recall in the first row where to find the definition of the association measure, and indicate in the subsequent rows whether an association measure performs well (+), neutral (+/−) or not so good (−) with respect to the mentioned aspects. Some caution is needed when consulting such a table: (1) it only summarizes our findings; (2) indications are rough and might be subjective.

In terms of analytical calculation, it is easiest to calculate Blomqvist's beta since it can be calculated for any copula. For all other measures, the ability to calculate a measure analytically depends on the fact whether the corresponding integral containing the copula can be calculated. For ρ_ℓ, $\ell \in \{1, 2, 3\}$, a d-dimensional integral is to be calculated. This might be difficult already in dimension $d = 2$ and not all well-known copulas have an analytical expression of Spearman's rho in such case. For Kendall's tau, a d-dimensional integral also needs to be calculated, yet can be simplified for some copulas For Gini's gamma γ_2 measure, one needs to calculate 2^d, fortunately one-dimensional, integrals.

In terms of interpretability, property (P_1) and knowledge about behaviour in increasing dimension Kendall's tau and Blomqvist's beta perform better, whereas for these aspects Spearman's rho and Gini's gamma score less well. It is to be noted though that Blomqvist's beta might be a too simple association measure. In terms of computational cost, and keeping in mind axioms (A_1) – (A_8), Spearman's rho association measure ρ_3 and Blomqvist's beta are preferable above Kendall's tau and Gini's gamma γ_2.

In multivariate data analysis, it is recommendable to start with a summary of bivariate association measures followed by calculation of multivariate association measures, possibly in various dimensions. This practical approach allows to gain insights in the overall dependencies in the data.

Acknowledgments

This research is supported by Research Fund KU Leuven (projects GOA/12/014 and C16/20/002). The research of the third author was supported by the grant GACR 19−00015S. The authors are grateful to the Editor-in-Chief, an Associate Editor, and two reviewers for their valuable comments which led to an improvement of the paper.

References

[1] B. Abdous, C. Genest, B. Rémillard, Dependence properties of meta-elliptical distributions, in: P. Duchesne, B. Rémillard (Eds.), Statistical Modeling and Analysis for Complex Data Problems, Springer, Boston, MA, 2005, pp. 1–15.

[2] J. Behboodian, A. Dolati, M. Úbeda-Flores, A multivariate version of Gini's rank association coefficient, Statistical Papers 48 (2007) 295–304.

[3] N. Blomqvist, On a measure of dependence between two random variables, The Annals of Mathematical Statistics 21 (1950) 593–600.

[4] C. Croux, C. Dehon, Influence functions of the Spearman and Kendall correlation measures, Statistical Methods & Applications 19 (2010) 497–515.

[5] C. W. Dunnet, M. Sobel, Approximations to the probability integral and certain percentage points of a multivariate analogue of Student's t-distribution, Biometrika 42 (1955) 258–267.

[6] H.-B. Fang, K.-T. Fang, The meta-elliptical distributions with given marginals, Journal of Multivariate Analysis 82 (2002) 1–16.

[7] S. Fuchs, F. D. Lascio, F. Durante, Dissimilarity functions for rank-based hierarchical clustering of continuous variables, 2020. ArXiv:2007.04799v1.

[8] C. Genest, J. Nešlehová, N. B. Ghorbal, Estimators based on Kendall's tau in multivariate copula models, Australian & New Zealand Journal of Statistics 53 (2011) 157–177.

[9] C. Gini, L'Ammontare e la Composizione della Ricchezza delle Nazioni, F. Bocca, Torino, 1914.

[10] S. Gupta, Probability integrals of multivariate normal and multivariate t, The Annals of Mathematical Statistics 34 (1963) 792–828.

[11] H. Joe, Multivariate Concordance, Journal of Multivariate Analysis 35 (1990) 12–30.

[12] M. G. Kendall, A new measure of rank correlation, Biometrika 30 (1938) 81–93.

[13] D. Lobdell, J. Jagai, K. Rappazzo, L. C. Messer, Data Sources for an Environmental Quality Index: Availability, Quality, and Utility, American Journal of Public Health 101 (Suppl 1) (2011) 277–85.

[14] A. J. McNeil, J. Nešlehová, Multivariate Archimedean copulas, d-monotone functions and ℓ_1-norm symmetric distributions, The Annals of Statistics 37 (2009) 3059–3097.

[15] C. Meyer, The bivariate normal copula, Communications in Statistics - Theory and Methods 42 (2013) 2402–2422.

[16] R. B. Nelsen, Nonparametric measures of multivariate association, in: L. Rüschendorf, B. Schweizer, M. D. Taylor (Eds.), Distributions with Fixed Marginals and Related Topics, volume 28 of *Lecture Notes–Monograph Series*, Institute of Mathematical Statistics, 1996, 2nd edition, pp. 223–232.

[17] R. B. Nelsen, Concordance and copulas: A survey, in: C. M. Cuadras, J. Fortiana, J. A. Rodriguez-Lallena (Eds.), Distributions With Given Marginals and Statistical Modelling, Springer, Dordrecht, 2002, pp. 169–177.

[18] R. B. Nelsen, An Introduction to Copulas, Springer, New York, 2006.

[19] A. Pérez, M. Prieto-Alaiz, A note on nonparametric estimation of copula-based multivariate extensions of Spearmans rho, Statistics & Probability Letters 112 (2016) 41–50.

[20] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2018.

[21] A. Rényi, On measures of dependence, Acta Mathematica Hungarica 10 (1959) 441–451.

[22] J. Riordan, Combinatorial Identities, Robert E. Krieger Publishing Company, Huntigton, New York, 1979.

[23] M. Scarsini, On measures of concordance, Stochastica 8 (1984) 201–218.

[24] F. Schmid, R. Schmidt, Bootstrapping Spearman's multivariate rho, in: A. Rizzi, M. Vichi (Eds.), Proceedings of COMPSTAT, 2006, pp. 759–766.

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

[26] F. Schmid, R. Schmidt, Nonparametric inference on multivariate versions of Blomqvist's beta and related measures of tail dependence, Metrika 66 (2007) 323–354.

[27] F. Schmid, R. Schmidt, T. Blumentritt, S. Gaißer, M. Ruppert, Copula-Based Measures of Multivariate Association, in: P. Jaworski, F. Durante, W. Härdle, T. Rychlik (Eds.), Copula Theory and Its Applications: Proceedings of the Workshop Held in Warsaw, 25-26 September 2009, 2010, pp. 209–236.

[28] J. Segers, Asymptotics of empirical copula processes under non-restrictive smoothness assumptions, Bernoulli 18 (2012) 764–782.

[29] A. Sklar, Fonctions de répartition à n dimensions et leurs marges, Publications de l'Institut de Statistique de l'Université de Paris 8 (1959) 229–231.

[30] C. Spearman, The proof and measurement of association between two things, The American Journal of Psychology 15 (1904) 72–101.

[31] M. D. Taylor, Multivariate measures of concordance, Annals of the Institute of Statistical Mathematics 59 (2007) 789–806.

[32] H. Tsukahara, Semiparametric estimation in copula models, Canadian Journal of Statistics 33 (2005) 357–375.

[33] M. Úbeda-Flores, Multivariate versions of Blomqvist's beta and Spearman's footrule, Annals of the Institute of Statistical Mathematics 57 (2005) 781–788.

[34] A. W. van der Vaart, J. A. Wellner, Weak Convergence and Empirical Processes, Springer, New York, 1996.
[35] E. F. Wolff, N-dimensional measures of dependence, Stochastica 4 (1980) 175–188.
[36] W. Wysocki, Kendall's tau and Spearman's rho for n-dimensional Archimedean copulas and their asymptotic properties, Journal of Nonparametric Statistics 27 (2015) 442–459.

Supplementary Material to the paper
"On the specification of multivariate association measures and their behaviour with increasing dimension"
by

Irène Gijbels[a,*], Vojtěch Kika[a,b], Marek Omelka[b]

[a] Department of Mathematics and Leuven Statistics Research Center (LStat), KU Leuven, Celestijnenlaan 200B, 3001 Leuven, Belgium
[b] Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Charles University, Sokolovská 83, 186 75 Prague, Czech Republic

This Supplementary Material contains the following items.

- Section S1 recalls the definition of an Archimedean copula, and a brief statement on Pearson's correlation coefficient.
- Section S2 studies multivariate Blomqvist's beta, and thus complements the study of the multivariate association measures presented in the paper.
- In Section S3 multivariate association measures are calculated (mainly analytically) in a set of examples.
- Section S4 discusses further how to obtain standard errors for the estimators of the multivariate association measures presented in Section 7. This section complements Section 7.2 of the paper.
- In Section S5 we provide some additional material regarding the real data application in Section 8.
- Section S6 provides a table with overview of the overall findings regarding the various association measures.

S1. Preliminaries

S1.1. Class of Archimedian copulas

Definition S1 (Archimedean copula)**.**

A nonincreasing and continuous function $\psi : [0,\infty) \to [0,1]$ which satisfies the conditions $\psi(0) = 1$, $\lim_{x\to\infty} \psi(x) = 0$ and is strictly decreasing on $[0, \inf\{x : \psi(x) = 0\})$ is called an Archimedean generator. A d-dimensional copula C_d is called Archimedean if it for any $\boldsymbol{u} \in [0,1]^d$ permits the representation

$$C_d(\boldsymbol{u}) = \psi\left[\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d)\right]$$

for some Archimedean generator ψ and its inverse $\psi^{-1} : (0,1] \to [0,\infty)$ where, by convention, $\psi(\infty) = 0$ and $\psi^{-1}(0) = \inf\{u : \psi(u) = 0\}$.

In paper [4] it was shown that the d-monotonicity property of ψ characterizes it as a generator for an Archimedean copula. The meaning of the d-monotonicity property is formally stated in Definition S2.

*Corresponding author. Email address: `irene.gijbels@kuleuven.be`

Definition S2 (d-monotonicity)**.**

A real function f is d-monotone on the interval $[0, \infty)$, where $d \geq 2$, if it is continuous on $[0, \infty)$ and differentiable on $(0, \infty)$ up to the order $d-2$ and the derivatives satisfy

$$(-1)^k f^{(k)}(x) \geq 0, \quad \text{for } k = 0, 1, \ldots d-2$$

for any $x \in (0, \infty)$ and further if $(-1)^{d-2} f^{(d-2)}$ is non-increasing and convex in $(0, \infty)$. If f has derivatives of all orders in $(0, \infty)$ and if $(-1)^k f^{(k)}(x) \geq 0$ for any $x \in (0, \infty)$ and any $k = 0, 1, \ldots$, then f is called completely monotone.

When studying the behaviour for increasing dimension, we consider a sequence of Archimedean copulas C_d with the same generator ψ. Since ψ needs to generate an Archimedean copula in any dimension d, the function ψ then needs to be completely monotone.

S1.2. Pearson correlation coefficient

A well-known bivariate dependence measure is Pearson's correlation coefficient ρ_P defined as

$$\rho_P(X_1, X_2) = \frac{\text{cov}(X_1, X_2)}{\sqrt{\text{var}(X_1)}\sqrt{\text{var}(X_2)}}$$

provided that $0 < \text{var}(X_1), \text{var}(X_2) < \infty$. Herein we omitted writing the subscript 2 for the dimension. It is well-known that this coefficient measures only linear dependence between random variables, is very sensitive to outliers and is not invariant with respect to all strictly increasing transformations of X_1 and X_2 and thus cannot be expressed as a functional of a copula. We therefore do not consider it in this paper. See [1] for more details on drawbacks of ρ_P.

S2. Multivariate Blomqvist's beta

S2.1. Verification of axioms

In the bivariate case, let C_2 be the copula of $(X_1, X_2)^\top$ with X_1, X_2 having medians $\text{med}(X_1)$ and $\text{med}(X_2)$, respectively. Blomqvist's beta is then defined as [see e.g. 5, p. 182]

$$\beta(X_1, X_2) = \Pr\{(X_1 - \text{med}(X_1))(X_2 - \text{med}(X_2)) > 0\} - \Pr\{(X_1 - \text{med}(X_1))(X_2 - \text{med}(X_2)) < 0\}.$$

Using copula notation, we have

$$\beta(C_2) = 4C_2(1/2, 1/2) - 1 = \frac{C_2(1/2, 1/2) - \Pi_2(1/2, 1/2) + \overline{C}_2(1/2, 1/2) - \overline{\Pi}_2(1/2, 1/2)}{M_2(1/2, 1/2) - \Pi_2(1/2, 1/2) + \overline{M}_2(1/2, 1/2) - \overline{\Pi}_2(1/2, 1/2)}, \tag{S.1}$$

as shown by [7]. Blomqvist's beta can be viewed as the normalized difference between C_2 and Π_2 at the point $(1/2, 1/2)^\top$.

For general dimension d, Blomqvist's beta can be generalized based on (S.1) as

$$\beta(C_d) = \frac{C_d(\mathbf{1/2}) - \Pi_d(\mathbf{1/2}) + \overline{C}_d(\mathbf{1/2}) - \overline{\Pi}_d(\mathbf{1/2})}{M_d(\mathbf{1/2}) - \Pi_d(\mathbf{1/2}) + \overline{M}_d(\mathbf{1/2}) - \overline{\Pi}_d(\mathbf{1/2})} = \frac{2^{d-1}}{2^{d-1} - 1}\left\{C_d(\mathbf{1/2}) + \overline{C}_d(\mathbf{1/2}) - 2^{1-d}\right\}, \tag{S.2}$$

where $\mathbf{1/2} = (1/2, \ldots, 1/2)^\top$. This generalization was considered by [9].

Similarly as for Kendall's tau, [3] introduced a broad class of generalizations for Blomqvist's beta. The validity of axioms $(A_1) - (A_7)$ for β is stated by [8] and the validity of (A_8) for β is established below.

Proposition S1. *Axiom* (A_8) *is fulfilled by* β*, defined in* (S.2)*.*

The proof of Proposition S1 is according to the following lines. Let C_d be a d-variate copula of $(X_1, \ldots, X_d)^\top$ and C_{d+1} be a $(d+1)$-variate copula of $(X_1, \ldots, X_d, X_{d+1})^\top$ where X_{d+1} is independent of $(X_1, \ldots, X_d)^\top$. Blomqvist's beta β can be written as

$$\beta(C_{d+1}) = \frac{2^d}{2^d - 1} \left\{ C_{d+1}(\mathbf{1/2}) + \overline{C}_{d+1}(\mathbf{1/2}) - 2^{-d} \right\} = \frac{2^{d-1}}{2^d - 1} \left\{ C_d(\mathbf{1/2}) + \overline{C}_d(\mathbf{1/2}) - 2^{1-d} \right\}$$

and thus we have

$$\beta(C_{d+1}) = \frac{\frac{2^{d-1}}{2^d-1}}{\frac{2^{d-1}}{2^{d-1}-1}} \beta(C_d) = \frac{2^{d-1}-1}{2^d-1} \beta(C_d)$$

Since for every $d \geq 2$ we have $(2^{d-1} - 1)/(2^d - 1) < 1$, it is clear that Axiom (A_8) holds for Blomqvist's beta.

Blomqvist's beta differs from other association measures mentioned in the paper by its simplicity which is an advantage in terms of computational complexity but a severe disadvantage in terms of information captured.

S2.2. Further properties

Statement (18) also holds for Blomqvist's beta. This can be seen as follows. First note that $C_d(\mathbf{1/2}) + \overline{C}_d(\mathbf{1/2}) = \Pr(X_1 < \text{med}(X_1), \ldots, X_d < \text{med}(X_d)) + \Pr(X_1 > \text{med}(X_1), \ldots, X_d > \text{med}(X_d))$. Hence one obtains

$$\begin{aligned} C_{d+1}(\mathbf{1/2}) + \overline{C}_{d+1}(\mathbf{1/2}) &= \Pr(X_1 < \text{med}(X_1), \ldots, X_{d+1} < \text{med}(X_{d+1})) + \Pr(X_1 > \text{med}(X_1), \ldots, X_{d+1} > \text{med}(X_{d+1})) \\ &= C_d(\mathbf{1/2}) + \overline{C}_d(\mathbf{1/2}) \end{aligned}$$

and recalling (S.2), the statement follows. Furthermore, remarks about the strict inequality and the equality in (18), similar to these formulated in Proposition 5 for Kendall's tau, can be made for Blomqvist's beta.

Regarding property (P_2) we mention that the statement in Proposition 6 also holds for Blomqvist's beta. The proof of this is along the same lines as the proof of the statement for Kendall's tau. Also for Blomqvist's beta the lower and the upper bounds are attainable. Similar remarks regarding the upper bound in (20), as made for Kendall's tau, are valid for $\beta(C_d)$.

S2.3. Multivariate Blomqvist's beta in increasing dimension

S2.3.1. Archimedean copulas and multivariate Blomqvist's beta in increasing dimensions

Blomqvist's beta for Archimedean copulas can be easily expressed using their generator. Let $\{C_d\}$ be a sequence of d-dimensional Archimedean copulas with (the same) generator ψ. Then using Definition S1 and the inclusion-exclusion principle (see (1)) we get

$$\begin{aligned} \beta(C_d) &= \frac{2^{d-1}}{2^{d-1} - 1} \left\{ C_d(\mathbf{1/2}) + \overline{C}_d(\mathbf{1/2}) - 2^{1-d} \right\} \\ &= \frac{2^{d-1}}{2^{d-1} - 1} \left\{ \psi\left[d \cdot \psi^{-1}\left(\frac{1}{2}\right)\right] + \sum_{j=0}^{d} (-1)^j \binom{d}{j} \psi\left[j \cdot \psi^{-1}\left(\frac{1}{2}\right)\right] - 2^{1-d} \right\}. \end{aligned}$$

Blomqvist's beta β for Archimedean copulas tends to zero when the dimension d increases, as is established in Proposition S2. The proof of this proposition relies on the following lemma.

Lemma S1. *Let $\{Y_k\}_{k=1}^\infty$ be a sequence of iid random variables from a standard exponential distribution with rate parameter 1 and W be a non-negative random variable such that $\{Y_k\}_{k=1}^\infty$ and W are independent. Also suppose that $\{x_k\}_{k=1}^\infty$ is a sequence of non-negative numbers such that $x = \sup_{k \in \mathbb{N}} x_k < \infty$. Then*

$$\lim_{d \to \infty} \Pr(Y_1/W \leq x_1, \ldots, Y_d/W \leq x_d) = 0.$$

The proof of the lemma is as follows. Using that the x_j's are bounded by x, we get

$$\Pr(Y_1/W \leq x_1, \ldots, Y_d/W \leq x_d) = \Pr(Y_1 \leq Wx_1, \ldots, Y_d \leq Wx_d) = \mathsf{E}\left(\prod_{j=1}^{d}\left(1-e^{-x_jW}\right)\right) \leq \mathsf{E}\left(\left(1-e^{-xW}\right)^d\right).$$

Since $1 - e^{-xW}$ can only take values in the interval $[0, 1)$, by the Lebesgue dominated convergence theorem

$$\lim_{d\to\infty} \Pr(Y_1/W \leq x_1, \ldots, Y_d/W \leq x_d) \leq \mathsf{E}\left(\lim_{d\to\infty}\left(1-e^{-xW}\right)^d\right) = 0.$$

Proposition S2 states that for a sequence of Archimedian copulas, Blomqvist's beta tends to zero when the dimension increases to infinity.

Proposition S2. *Let $\{C_d\}$ be a sequence of d-dimensional Archimedean copulas with (the same) generator ψ. Then*

$$\lim_{d\to\infty} \beta(C_d) = 0.$$

That this results holds can be seen as follows. Recall the definition of β in (S.2). Since obviously $\lim_{d\to\infty} 2^{d-1}/(2^{d-1}-1) = 1$ and $\lim_{d\to\infty} 2^{1-d} = 0$, we only need to show

$$\lim_{d\to\infty} C_d(\mathbf{1/2}) = 0, \tag{S.3}$$

$$\lim_{d\to\infty} \overline{C}_d(\mathbf{1/2}) = 0. \tag{S.4}$$

Let ψ be the generator of C_d and also recall Definition S1. Then we can express

$$C_d(\mathbf{1/2}) = \psi\left[d \cdot \psi^{-1}\left(\frac{1}{2}\right)\right],$$

where we know that $\psi^{-1}(1/2) > 0$ since ψ^{-1} is strictly decreasing and $\psi^{-1}(1) = 0$. Also $\lim_{x\to\infty} \psi(x) = 0$. Combining these two, we get (S.3)

$$\lim_{d\to\infty} C_d(\mathbf{1/2}) = \lim_{d\to\infty} \psi\left[d \cdot \psi^{-1}\left(\frac{1}{2}\right)\right] = 0.$$

For calculating the limit of the survival function, another concept is needed. Following the relation stated by [4], a survival function of an Archimedean copula is the joint distribution function of a random vector $(Y_1/W, \ldots, Y_d/W)^\top$ where $Y_1, \ldots, Y_d$ are iid random variables from Exp(1) independent of W which is a non-negative random variable. Thus we can apply Lemma S1 (see the Appendix) with $x_j = \text{med}(Y_1/W)$ for every $j \in \{1, 2, \ldots\}$ and obtain that

$$\lim_{d\to\infty} \overline{C}_d(\mathbf{1/2}) = \lim_{d\to\infty} \Pr(Y_1/W \leq \text{med}(Y_1/W), \ldots, Y_d/W \leq \text{med}(Y_1/W)) = 0,$$

which leads to (S.4).

S2.3.2. Meta-elliptical copulas and multivariate Blomqvist's beta in increasing dimensions

From Schmid and Schmidt [7, Proposition 8 + proof] we know that Blomqvist's beta for the meta-elliptical copula C_d also equals the expression in (26). Consequently for a meta-elliptical copula C_d it holds that $\tau(C_d) = \beta(C_d)$. Hence the results for Kendall's tau discussed in Section 5.2 also hold for Blomqvist's beta.

S2.4. Estimation of Blomqvist's beta

The estimator of β is proposed by [7] as

$$\widehat{\beta}_n = \beta(\widehat{C}_n) = \frac{2^{d-1}}{2^{d-1}-1}\left\{\widehat{C}_n(\mathbf{1/2}) + \widehat{\overline{C}}_n(\mathbf{1/2}) - 2^{1-d}\right\}$$

If the partial derivatives $\frac{\partial C}{\partial u_i}$ and $\frac{\partial \overline{C}}{\partial u_i}$ are continuous at $\mathbf{1/2}$, it follows that

$$\sqrt{n}(\widehat{\beta}_n - \beta(C)) \xrightarrow{D} \mathcal{N}(0, \sigma_\beta^2),$$

as $n \to \infty$ where $\sigma_\beta^2 = (2^{d-1}/(2^{d-1}-1))^2\,\mathsf{E}\{\mathbb{G}_C(\mathbf{1/2}) + \mathbb{G}_{\bar{C}}(\mathbf{1/2})\}^2$, as shown by [7]. Note again that [7] originally used division by n instead of $n+1$ in (30). This, however, does not affect validity of the asymptotic result.
Note that for a sample from the comonotonicity copula, $\widehat{\beta}_n = 1$ since $\widehat{U}_{1,i} = \cdots = \widehat{U}_{d,i}$ for every $i \in \{1,\ldots,n\}$ and thus $\widehat{C}_n(\mathbf{1/2}) + \widehat{\overline{C}}_n(\mathbf{1/2}) = 1$.

S3. Illustrative examples

In this section, we provide examples to illustrate the calculations needed to obtain population versions of the covered multivariate association measures.

Example S1 (Farlie-Gumbel-Morgenstein copula)**.** Let C_d be a d-dimensional Farlie-Gumbel-Morgenstern copula defined as

$$C_d(\boldsymbol{u}) = u_1u_2\ldots u_d\left[1 + \sum_{j=2}^{d}\sum_{1\le k_1<\cdots<k_j\le d}\alpha_{k_1,\ldots,k_j}\left(1-u_{k_1}\right)\ldots\left(1-u_{k_j}\right)\right] \tag{S.5}$$

where parameters satisfy the following 2^d conditions

$$1 + \sum_{j=2}^{d}\sum_{1\le k_1<\cdots<k_j\le d}\alpha_{k_1,\ldots,k_j}\epsilon_{k_1}\cdots\epsilon_{k_j} \ge 0, \quad \forall\epsilon_1,\ldots,\epsilon_d \in \{-1,1\}.$$

In case $d = 3$, these conditions can be alternatively rewritten as in Example 1. To derive the formulas for association measures assume for a moment that for every j in (S.5), there is only one non-zero parameter α_j and thus the copula $C_d(\boldsymbol{u})$ reduces to the copula of the form

$$C_d^*(\boldsymbol{u}) = u_1u_2\ldots u_d\left[1 + \sum_{j=2}^{d}\alpha_j\prod_{k=1}^{j}(1-u_k)\right].$$

This copula has a density

$$c_d^*(\boldsymbol{u}) = 1 + \sum_{j=2}^{d}\alpha_j\prod_{k=1}^{j}(1-2u_k)$$

and we can calculate

$$\int_{[0,1]^d} C_d^*(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} = 2^{-d} + 2^{-d}\sum_{j=2}^{d}\alpha_j 3^{-j},$$

$$\int_{[0,1]^d} \Pi_d(\boldsymbol{u})c_d^*(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} = 2^{-d} + 2^{-d}\sum_{j=2}^{d}\alpha_j(-1)^j 3^{-j}.$$

From these, the expressions for $\rho_1(C_d^*)$ and $\rho_2(C_d^*)$ follow. Returning to the general case, note first that all $\alpha_{k_1,\cdots,k_j}$ play similar roles to α_j, and hence

$$\rho_1(C_d) = \frac{d+1}{2^d-(d+1)}\left(\sum_{j=2}^{d} 3^{-j} \sum_{1\le k_1<\cdots<k_j\le d} \alpha_{k_1,\dots,k_j}\right),$$

$$\rho_2(C_d) = \frac{d+1}{2^d-(d+1)}\left(\sum_{j=2}^{d} (-1)^j 3^{-j} \sum_{1\le k_1<\cdots<k_j\le d} \alpha_{k_1,\dots,k_j}\right)$$

and thus

$$\rho_3(C_d) = \frac{d+1}{2^d-(d+1)}\left(\sum_{\substack{j=2\\ j \text{ even}}}^{d} 3^{-j} \sum_{1\le k_1<\cdots<k_j\le d} \alpha_{k_1,\dots,k_j}\right).$$

Kendall's tau is based on the integral

$$\begin{aligned}\int_{[0,1]^d} C_d^*(\boldsymbol{u})c_d^*(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} &= \int_{[0,1]^d} \Pi_d(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} + \int_{[0,1]^d} \sum_{j=2}^{d} \alpha_j \Pi_d(\boldsymbol{u}) \prod_{\ell=1}^{j} (1-u_\ell)\,\mathrm{d}\boldsymbol{u} \\ &+ \int_{[0,1]^d} \sum_{k=2}^{d} \alpha_k \Pi_d(\boldsymbol{u}) \prod_{m=1}^{k} (1-2u_m)\,\mathrm{d}\boldsymbol{u} \\ &+ \int_{[0,1]^d} \Pi_d(\boldsymbol{u}) \left[\sum_{j=2}^{d} \alpha_j \prod_{\ell=1}^{j}(1-u_\ell)\right]\left[\sum_{k=2}^{d} \alpha_k \prod_{m=1}^{k}(1-2u_m)\right]\mathrm{d}\boldsymbol{u}.\end{aligned}$$

The last integral always contains at least one factor of the form $\int_0^1 u(1-u)(1-2u)\,\mathrm{d}u = 0$ and thus one obtains

$$2^d \int_{[0,1]^d} C_d^*(\boldsymbol{u})c_d^*(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} = 1 + 2 \sum_{\substack{j=2\\ j \text{ even}}}^{d} \alpha_j 3^{-j}$$

which, returning to the general case, leads to

$$\tau(C_d) = \frac{2}{2^{d-1}-1}\left[\sum_{\substack{j=2\\ j \text{ even}}}^{d} 3^{-j} \sum_{1\le k_1<\cdots<k_j\le d} \alpha_{k_1,\dots,k_j}\right].$$

In what follows we show that

$$\gamma_2(C_d) = \frac{2}{2^{d-1}-1} \sum_{\substack{j=2\\ j \text{ even}}}^{d} 2^j \frac{(j!)^2}{(2j+1)!} \sum_{1\le k_1<\cdots<k_j\le d} \alpha_{k_1,\dots,k_j},$$

$$\beta(C_d) = \frac{1}{2^{d-1}-1} \sum_{\substack{j=2\\ j \text{ even}}}^{d} 2^{-j} \sum_{1\le k_1<\cdots<k_j\le d} \alpha_{k_1,\dots,k_j}.$$

The detailed derivation of these expressions goes as follows. Consider further simplification of C_d in a form

$$C_{d,j}^*(\boldsymbol{u}) = u_1 u_2 \dots u_d \left[1 + \alpha_j \prod_{k=1}^{j}(1-u_k)\right] = \Pi_d(\boldsymbol{u}) + \alpha_j \Pi_d(\boldsymbol{u}) \prod_{k=1}^{j}(1-u_k).$$

Such a simplification could also be used while calculating ρ_1, ρ_2 and ρ_3 since these do not contain expressions $\alpha_j\alpha_k$ for $j \neq k$ which are however involved in the calculation of τ. Denote

$$\widetilde{C}_{d,j}(\boldsymbol{u}) = \Pi_d(\boldsymbol{u}) \prod_{k=1}^{j} (1-u_k) = \prod_{k=1}^{j} p(u_k) \prod_{\ell=j+1}^{d} u_\ell$$

with $p(u) = u(1-u) = p(1-u)$. Now calculate

$$\begin{aligned}
K(d,j) &= \sum_{\xi\in\mathcal{R}_d} \int_0^1 \widetilde{C}_{d,j}(\xi(u,\dots,u))\,\mathrm{d}u = 2^j \sum_{\xi\in\mathcal{R}_{d-j}} \int_0^1 \widetilde{C}_{d,j}(\underbrace{u,\dots,u}_{j}, \xi(\underbrace{u,\dots,u}_{d-j}))\,\mathrm{d}u \\
&= 2^j \sum_{m=0}^{d-j} \binom{d-j}{m} \int_0^1 [p(u)]^j u^m (1-u)^{d-j-m}\,\mathrm{d}u \\
&= 2^j \sum_{m=0}^{d-j} \frac{(d-j)!}{m!(d-j-m)!} \frac{(d-m)!(j+m)!}{(d+j+1)!} \\
&= 2^j \frac{(j!)^2}{(2j+1)!} \frac{1}{\binom{d+j+1}{d-j}} \sum_{m=0}^{d-j} \binom{j+m}{m}\binom{d-m}{d-j-m} \\
&= 2^j \frac{(j!)^2}{(2j+1)!}
\end{aligned}$$

where the index m denotes the number of components not reflected by ξ and where Vandermonde's identity [see e.g. 6, (3c) on p. 9] is used. Denote by $\overline{\widetilde{C}}_{d,j}$ the contribution of $\widetilde{C}_{d,j}$ to the survival function $\overline{C}^*_{d,j}(\boldsymbol{u})$ of $C^*_{d,j}(\boldsymbol{u})$. Then

$$\overline{\widetilde{C}}_{d,j}(\boldsymbol{u}) = \sum_{k=j}^{d} (-1)^k \sum_{j<i_{j+1}<\dots<i_k\le d} \widetilde{C}_{k,j}(u_1,\dots,u_j,u_{i_{j+1}},\dots,u_{i_k}) = \sum_{k=j}^{d} (-1)^k \sum_{j<i_{j+1}<\dots<i_k\le d} p(u_1)\cdot\ldots\cdot p(u_j) u_{i_{j+1}}\cdot\ldots\cdot u_{i_k}$$

and thus

$$\begin{aligned}
\overline{K}(d,j) &= \sum_{\xi\in\mathcal{R}_d} \int_0^1 \overline{\widetilde{C}}_{d,j}(\xi(u,\dots,u))\,\mathrm{d}u = 2^j \sum_{\xi\in\mathcal{R}_{d-j}} \int_0^1 \overline{\widetilde{C}}_{d,j}(\underbrace{u,\dots,u}_{j}, \xi(\underbrace{u,\dots,u}_{d-j}))\,\mathrm{d}u \\
&= 2^j \sum_{k=j}^{d} (-1)^k \sum_{j<i_{j+1}<\dots<i_k\le d} \sum_{\xi\in\mathcal{R}_{d-j}} 2^{d-k} \int_0^1 [p(u)]^j \xi_{i_{j+1}}(u)\dots\xi_{i_k}(u)\,\mathrm{d}u \\
&= 2^j \sum_{k=j}^{d} (-1)^k \sum_{j<i_{j+1}<\dots<i_k\le d} K(k,j) 2^{d-k} = 2^j \frac{(j!)^2}{(2j+1)!} \sum_{k=j}^{d} (-1)^k \binom{d-j}{k-j} 2^{d-k} \\
&= 2^j \frac{(j!)^2}{(2j+1)!} (-1)^j.
\end{aligned}$$

In the last equality we used the binomial theorem

$$\sum_{k=j}^{d} (-1)^k \binom{d-j}{k-j} 2^{d-k} \overset{\ell=k-j}{=} (-1)^j \sum_{\ell=0}^{d-j} (-1)^\ell \binom{d-j}{\ell} 2^{d-j-\ell} = (-1)^j.$$

Plugging $K(d,j)$ and $\overline{K}(d,j)$ into (17) and moving back from $\widetilde{C}_{d,j}$ to C_d, we obtain

$$\gamma_2(C_d) = \frac{2}{2^{d-1}-1} \sum_{\substack{j=2 \\ j \text{ even}}}^{d} 2^j \frac{(j!)^2}{(2j+1)!} \sum_{1\le k_1<\dots<k_j\le d} \alpha_{k_1,\dots,k_j}.$$

Finally, $\widetilde{C}_{d,j}(\mathbf{1/2}) = 2^{-d-j}$ and

$$\overline{\widetilde{C}}_{d,j}(\mathbf{1/2}) = \sum_{k=j}^{d}(-1)^k \sum_{j<i_{j+1}<\cdots<i_k\leq d} 2^{-j-k} = 2^{-d-j}\sum_{k=j}^{d}(-1)^k\binom{d-j}{k-j}2^{d-k} = 2^{-d-j}(-1)^j$$

which, while moving back from $\widetilde{C}_{d,j}$ to C_d, leads to

$$\beta(C_d) = \frac{1}{2^{d-1}-1}\sum_{\substack{j=2\\ j\text{ even}}}^{d} 2^{-j}\sum_{1\leq k_1<\cdots<k_j\leq d}\alpha_{k_1,\ldots,k_j}.$$

Example S2 (Block independence copula)**.** Let C_d be a copula characterized by a density $c_d(\boldsymbol{u}) = 2^{d-1}$ in the two cubes $[0,1/2]^d$ and $[1/2,1]^d$ and 0 elsewhere. Recall that the density of the independence copula Π_d is constant on $[0,1]^d$. Copula C_d can thus be seen as a 'block independence' copula. By integrating over the density, the copula itself has a form

$$C_d(\boldsymbol{u}) = \begin{cases} \frac{1}{2} + 2^{d-1}\prod_{j=1}^{d}(u_j - \frac{1}{2}), & \text{if } u_1 > \frac{1}{2},\ldots,u_d > \frac{1}{2},\\ \frac{1}{2}\prod\limits_{\substack{j=1\\ u_j<1/2}}^{d} 2u_j, & \text{otherwise.}\end{cases}$$

Calculating Spearman's rho ρ_2 is based on the integral

$$\int_{[0,1]^d}\Pi_d(\boldsymbol{u})c_d(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} = 2^{d-1}\left[\int_{[0,\frac{1}{2}]^d}\Pi_d(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} + \int_{[\frac{1}{2},1]^d}\Pi_d(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}\right] = 2^{-2d-1}(1+3^d)$$

and thus

$$\rho_1(C_d) = \rho_2(C_d) = \rho_3(C_d) = \frac{d+1}{2^d-(d+1)}\left[2^{-2d-1}(1+3^d)-1\right]$$

where the result for ρ_1, and consequently for ρ_3 comes from obvious radial symmetry of C_d. Similarly, Kendall's tau τ is based on the integral

$$\int_{[0,1]^d}C_d(\boldsymbol{u})c_d(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} = 2^{2d-2}\int_{[0,\frac{1}{2}]^d}\Pi_d(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} + \int_{[\frac{1}{2},1]^d}\left[2^{d-2} + 2^{2d-2}\prod_{j=1}^{d}\left(u_j - \frac{1}{2}\right)\right]\mathrm{d}\boldsymbol{u} = 2^{-d-1} + 2^{-2}$$

and thus

$$\tau(C_d) = \frac{1}{2}.$$

Gini's gamma γ_2 is based on $\sum_{\xi\in\mathcal{R}_d}\int_0^1(C_d(\xi(u,\ldots,u)) + C_d^S(\xi(u,\ldots,u)))\,\mathrm{d}u$ which can be, using radial symmetry and exchangeability of C_d rewritten as

$$2\sum_{j=0}^{d}\binom{d}{j}\int_0^1 C_d(\underbrace{u,\ldots,u}_{j},\underbrace{1-u,\ldots,1-u}_{d-j})\,\mathrm{d}u = \frac{1}{2}\sum_{j=1}^{d-1}\binom{d}{j}\frac{d+2}{(j+1)(d-j+1)} + 1 + \frac{2}{d+1}.$$

Further using that

$$\sum_{j=1}^{d-1}\binom{d}{j}\frac{d+2}{(j+1)(d-j+1)} = \frac{1}{d+1}\sum_{j=1}^{d-1}\binom{d+2}{j+1} = \frac{1}{d+1}\left[2^{d+2}-2-2(d+2)\right],$$

one gets to the final expression

$$\gamma_2(C_d) = \frac{2(2^d - d - 1)}{(2^{d-1} - 1)(d + 1)}$$

and finally Blomqvist's beta is obvious to evaluate as

$$\beta(C_d) = 1.$$

Further notice that whereas ρ_1, ρ_2, ρ_3 and γ_2 converge to 0 when $d \to \infty$, τ and β remain constant for every $d \geq 2$.

Example S3. Let C_d be the copula of $(X, \ldots, X, -X)^\top$, that is

$$C_d(\boldsymbol{u}) = \max\left(\min_{1,\ldots,d-1}(u_j) + u_d - 1, 0\right).$$

Copula C_d is radially symmetric which can be seen through the following

$$\begin{aligned} C^S(\boldsymbol{u}) &= \Pr(1 - U \leq u_1, \ldots, 1 - U \leq u_{d-1}, U \leq u_d) \\ &= \Pr(\max_{1,\ldots,d-1}(1 - u_j) \leq U \leq u_d) = \max(0, u_d - \max_{1,\ldots,d-1}(1 - u_j)) = C_d(\boldsymbol{u}). \end{aligned}$$

Because of the radial symmetry, to calculate γ_2 one only needs the expression $\sum_{\xi \in \mathcal{R}_d} \int_0^1 C_d(\xi(u, \ldots, u))\,\mathrm{d}u$. Since C_d is the copula of $(U, \ldots, U, 1 - U)^\top$, it is straightforward to see that

$$C_d(\xi(u, \ldots, u)) = \begin{cases} \max(2u - 1, 0), & \text{if } |S_\xi| = 0, \\ \max(1 - 2u, 0), & \text{if } |S_\xi| = d, \\ 0, & \text{otherwise.} \end{cases}$$

Moreover, $\int_{[0,1]^d} C_d(\boldsymbol{u})\,\mathrm{d}C_d(\boldsymbol{u}) = \Pr(X < Y, \ldots, X < Y, -X < -Y) = 0$, with Y being an independent copy of X and also $C_d(\mathbf{1/2}) = \overline{C}_d(\mathbf{1/2}) = 0$ and thus

$$\gamma_2(C_d) = \beta(C_d) = \tau(C_d) = -\frac{1}{2^{d-1} - 1}. \tag{S.6}$$

Further notice that $u_1, \ldots, u_{d-1}$ are playing a symmetric role in C_d and we can divide $[0, 1]^d$ into $d - 1$ parts depending on which argument from $u_1, \ldots, u_{d-1}$ is minimal, that is into the parts $S_j = \{\boldsymbol{u} \in [0, 1]^d; u_j = \min(u_1, \ldots, u_{d-1})\}$, for $j \in \{1, \ldots, d - 1\}$. Then for $j = 1$

$$\begin{aligned} \int_{S_1} C(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} = \int_{S_1} \max(u_1 + u_d - 1, 0)\,\mathrm{d}\boldsymbol{u} &= \int_0^1 \int_{1-u_1}^1 \int_{u_1}^1 \cdots \int_{u_1}^1 (u_1 + u_d - 1)\,\mathrm{d}u_2 \ldots \mathrm{d}u_{d-1}\,\mathrm{d}u_d\,\mathrm{d}u_1 \\ &= \frac{1}{(d - 1)d(d + 1)}. \end{aligned}$$

From exchangeability within the first $d - 1$ components of the random vector it follows that $\int_{S_1} C(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} = \cdots = \int_{S_{d-1}} C(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}$ and thus $\int_{[0,1]^d} C(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} = 1/[d(d + 1)]$ which further implies, together with the radial symmetry, that

$$\rho_1(C_d) = \rho_2(C_d) = \rho_3(C_d) = \frac{2^d - d^2 - d}{d2^d - d^2 - d}. \tag{S.7}$$

Note that while both functions in (S.6) and (S.7) tend to 0 as $d \to \infty$, the function in (S.6) is negative for any $d \geq 2$ whereas the function in (S.7) is negative for $d \in \{2, 3, 4\}$ and positive for $d \in \{5, 6, \ldots\}$. In case of γ_2, β and τ, the negative sign of these association measures for d large could be considered counter-intuitive since except for one component, one is dealing with the comonotonicity copula. On the other hand, ρ_ℓ for $\ell \in \{1, 2, 3\}$ is decreasing as a function of d from $d = 9$ which could be considered counter-intuitive as well since one keeps adding the same random variable X. Any of the pairwise association measures, based on a bivariate association measure satisfying at least (S_1) and (S_5), will be equal to $(d-4)/d$ in this example and is thus positive for $d \geq 5$ and converges to 1 as $d \to \infty$. This example thus clearly illustrates the important different behaviour of multivariate association measures obtained via the two approaches: all pairwise association measures tend to 1 as d tends to infinity, whereas all multivariate generalizations based on the copula approach tend to 0 for d tending to infinity.

Example S4 (Clayton copula)**.** Let C_d be a d-variate Clayton family copula defined as

$$C_d(\boldsymbol{u}) = \left(\sum_{j=1}^{d} u_j^{-\theta} - d + 1 \right)^{-1/\theta}$$

for $\theta > 0$.

Kendall's tau can be calculated as

$$\tau(C_d) = \frac{1}{2^{d-1} - 1} \left\{ -1 + 2^d \prod_{j=0}^{d-1} \frac{1 + j\theta}{2 + j\theta} \right\},$$

as can be seen in Example 1 in [2]. Realising that all factors in the above product are positive and using $1 - x \leq \exp(-x)$ for $x \geq 0$, we can focus on the asymptotic behaviour for increasing dimension. For any $\theta > 0$

$$\lim_{d\to\infty} \prod_{j=0}^{d-1} \frac{1 + j\theta}{2 + j\theta} = \lim_{d\to\infty} \prod_{j=0}^{d-1} \left(1 - \frac{1}{2 + j\theta} \right) \leq \lim_{d\to\infty} \prod_{j=0}^{d-1} \exp\left(\frac{-1}{2 + j\theta} \right)$$
$$= \lim_{d\to\infty} \exp\left(-\sum_{j=0}^{d-1} \frac{1}{2 + j\theta} \right) = \exp\left(-\sum_{j=0}^{\infty} \frac{1}{2 + j\theta} \right) = 0.$$

This means that with increasing dimension, τ of Clayton copula tends to 0 which is in agreement with results of Wysocki [10, Theorem 7]. For Blomqvist's beta we have

$$\beta(C_d) = \frac{2^{d-1}}{2^{d-1} - 1} \left\{ (d2^\theta - d + 1)^{-1/\theta} + \sum_{j=0}^{d} (-1)^j \binom{d}{j} (j2^\theta - j + 1)^{-1/\theta} - 2^{1-d} \right\}.$$

In Fig. S1, Kendall's tau and Blomqvist's beta are plotted as a function of dimension. Moreover, an approximation of Spearman's rho ρ_1 via Monte Carlo integration is added. It appears as if $\lim_{d\to\infty} \rho_1(C_d) = c_1 \in (0, 1)$, which gives a hint that strictly positive constants c_1 and c_2 in Section 5.1.1 can indeed be achieved. For Kendall's τ, $\lim_{d\to\infty} \tau(C_d) = 0$ follows the theoretical result. The convergence of $\beta(C_d)$ to zero seems to be slower, as evidenced by Fig. S1.

Example S5 (Four-dimensional Gaussian copula (continued))**.** Recall the setting of a four-dimensional Gaussian copula presented in Section 6. We further consider a correlation structure

$$\Sigma_{Cl} = \begin{pmatrix} 1.00 & \varrho & 0.05 & 0.00 \\ \varrho & 1.00 & 0.05 & -0.05 \\ 0.05 & 0.05 & 1.00 & \varrho \\ 0.00 & -0.05 & \varrho & 1.00 \end{pmatrix},$$

with varying parameter ϱ describing the correlation within the two clusters. Results from a numerical approximation of the association measure, as a function of ρ for all positive values of ρ, are depicted in Figure S2. Note that all association measures increase with increasing value of ρ. For Spearman's who the increase is almost linear while for Kendall, tau, Blomqvist beta and Gini's gamma the increase has a slightly quadratic appearance.

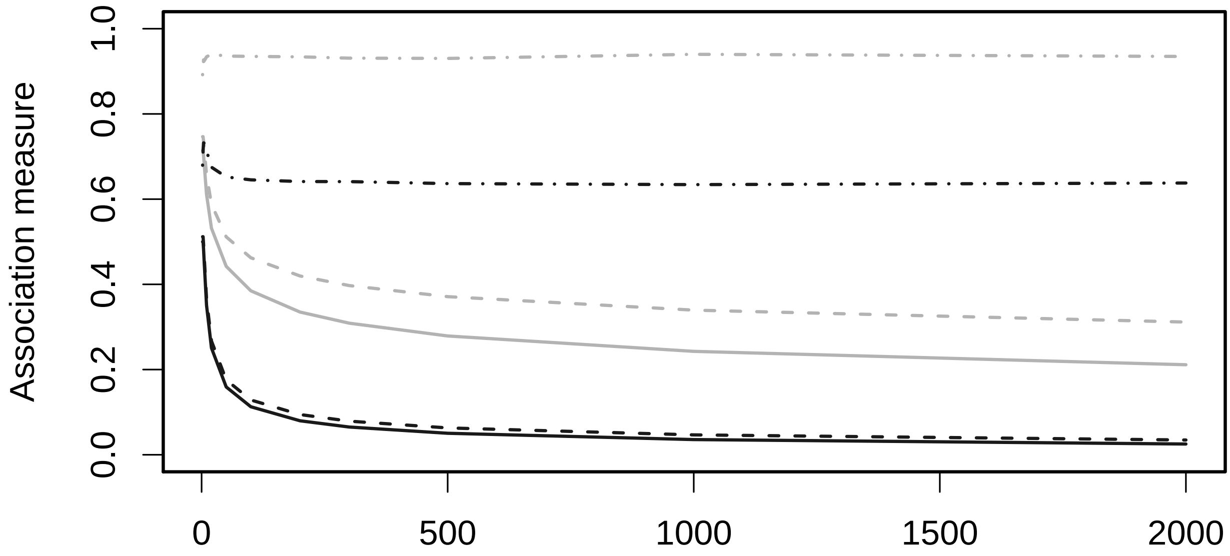

Fig. S1: Kendall's tau (full line), Blomqvist's beta (dashed line) and an approximation of Spearman's rho ρ_1 (dot-dashed line) for Clayton copula with parameters 2 (black line) and 5 (grey line) as a function of the dimension of the copula.

S4. Standard errors of the estimators: additional material

We present here methods for calculating standard errors of estimators of other multivariate association measures, in line with the material provided in Section 7.2.

S4.1. Standard error of estimator of Blomqvist's beta

The asymptotic variance for the estimator of Blomqvist's beta is

$$\sigma_\beta^2 = \left(\frac{2^{d-1}}{2^{d-1}-1}\right)^2 \mathsf{E}\,\{\mathbb{G}_C(\mathbf{1/2}) + \mathbb{G}_{\bar{C}}(\mathbf{1/2})\}^2.$$

Herein the expectation $\mathsf{E}\,\{\mathbb{G}_C(\mathbf{1/2}) + \mathbb{G}_{\bar{C}}(\mathbf{1/2})\}^2$ can be estimated by (35). But now W_i is given by

$$W_i = Z_i + \widetilde{Z}_i,$$

where

$$Z_i = \mathbb{1}(\widehat{\boldsymbol{U}}_i \leq \mathbf{1/2}) - \sum_{j=1}^{d} \widehat{C}_{jn}\,(\mathbf{1/2})\,\mathbb{1}(\widehat{U}_{j,i} \leq 1/2),$$

$$\widetilde{Z}_i = \mathbb{1}(\widehat{\boldsymbol{U}}_i > \mathbf{1/2}) - \sum_{j=1}^{d} \widehat{\overline{C}}_{jn}(\mathbf{1/2})\,\mathbb{1}(\widehat{U}_{j,i} \leq 1/2).$$

and $\widehat{C}_{jn}\,(\boldsymbol{u})$, $\widehat{\overline{C}}_{jn}(\boldsymbol{u})$ are the estimates of $\frac{\partial C(\boldsymbol{u})}{\partial u_j}$, $\frac{\partial \bar{C}(\boldsymbol{u})}{\partial u_j}$ respectively given by

$$\widehat{C}_{jn}\,(\boldsymbol{u}) = \frac{\widehat{C}_n\left(\boldsymbol{u} + \frac{1}{\sqrt{n}}\,\boldsymbol{e}_j\right) - \widehat{C}_n\left(\boldsymbol{u} - \frac{1}{\sqrt{n}}\,\boldsymbol{e}_j\right)}{\frac{2}{\sqrt{n}}}, \qquad \widehat{\overline{C}}_{jn}\,(\boldsymbol{u}) = \frac{\widehat{\overline{C}}_n\left(\boldsymbol{u} + \frac{1}{\sqrt{n}}\,\boldsymbol{e}_j\right) - \widehat{\overline{C}}_n\left(\boldsymbol{u} - \frac{1}{\sqrt{n}}\,\boldsymbol{e}_j\right)}{\frac{2}{\sqrt{n}}}$$

with $\boldsymbol{e}_j$ being the j-th canonical vector, i.e. vector of zeroes with the j-th component equal to one.

Finally the standard error of $\widehat{\beta}_n$ is estimated by $\frac{2^d}{2^{d-1}-1}\,\frac{\widehat{\sigma}_n}{\sqrt{n}}$.

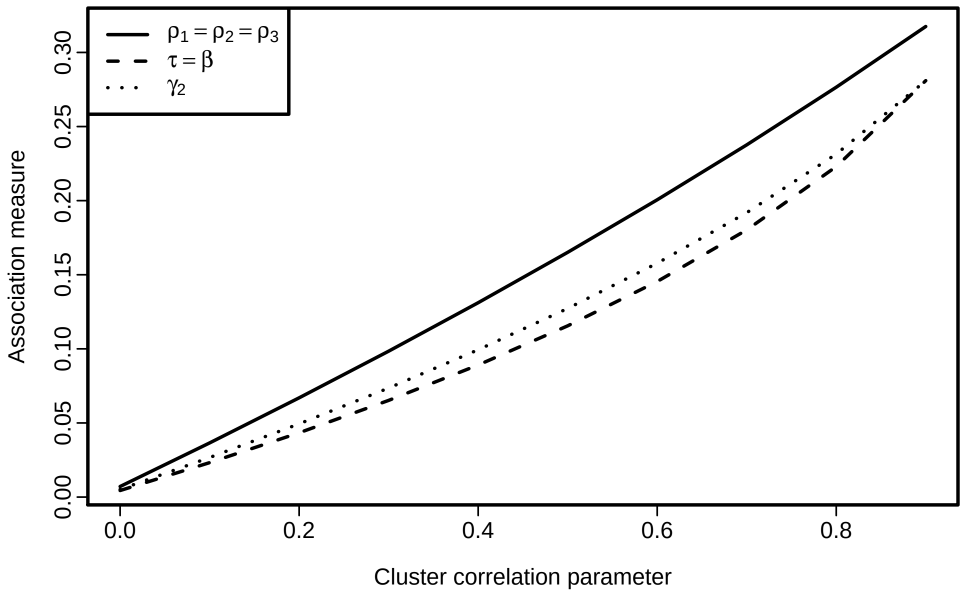

Fig. S2: Association measures varying with correlation within clusters.

S4.2. Standard error of estimators of Spearman's rho measures

The expressions for the asymptotic variances of the estimators $\widehat{\rho}_{1n}$ and $\widehat{\rho}_{2n}$ are in (32). The asymptotic variance for the average estimator $\widehat{\rho}_{3n}$ is then

$$\sigma_3^2 = 2^{2d} h_\rho^2(d) \int_{[0,1]^d} \int_{[0,1]^d} \frac{1}{4} \,\mathsf{E}\left[\{\mathbb{G}_C(\boldsymbol{u}) + \mathbb{G}_{\overline{C}}(\boldsymbol{u})\} \{\mathbb{G}_C(\boldsymbol{v}) + \mathbb{G}_{\overline{C}}(\boldsymbol{v})\} \right] \mathrm{d}\boldsymbol{u}\, \mathrm{d}\boldsymbol{v}.$$

Now the integral in the expression for σ_1^2, the asymptotic variance of $\widehat{\rho}_{1n}$ can be estimated by (35) with W_i replaced by $W_i^{(1)}$ given by

$$W_i^{(1)} = \prod_{j=1}^{d} (1 - \widehat{U}_{j,i}) - \sum_{j=1}^{d} A_j^{(1)}(\widehat{U}_{j,i}),$$

where

$$A_j^{(1)}(u) = \frac{1}{n-1} \sum_{i'=1}^{n} \mathbb{1}(\widehat{U}_{j,i'} > u) \prod_{k=1, k\neq j}^{d} (1 - \widehat{U}_{k,i'}).$$

Similarly the integral in the expression for σ_2^2 can be estimated by (35) with W_i replaced by $W_i^{(2)}$ given by

$$W_i^{(2)} = \prod_{j=1}^{d} \widehat{U}_{j,i} - \sum_{j=1}^{d} A_j^{(2)}(\widehat{U}_{j,i}),$$

where

$$A_j^{(2)}(u) = \frac{-1}{n-1} \sum_{i'=1}^{n} \mathbb{1}(\widehat{U}_{j,i'} > u) \prod_{k=1, k\neq j}^{d} \widehat{U}_{k,i'}.$$

Finally the integral in the expression for σ_3^2 can be estimated by (35) with $W_i = \frac{W_i^{(1)} + W_i^{(2)}}{2}$.

S4.3. Standard error of estimators of Gini's gamma measures

The estimator of the standard error of $\widehat{\gamma}_{2n}$ is given by $\frac{1}{2^{d-1}-1}\,\frac{\widehat{\sigma}_n}{\sqrt{n}}$, where $\widehat{\sigma}_n^2$ is given by (35) with $W_i = \widetilde{W}_i^{(1)} + \widetilde{W}_i^{(2)}$ and $\widetilde{W}_i^{(1)}$, $\widetilde{W}_i^{(2)}$ are defined in the following way

$$\widetilde{W}_i^{(1)} = \sum_{\xi\in\mathcal{R}_d} \Big(1 - \max_{k\in S_\xi}(\widehat{U}_{k,i}) - \max_{k\notin S_\xi}(\widehat{U}_{k,i})\Big)_+ - \sum_{j=1}^{d} A_j^{(1)}(\widehat{U}_{j,i}),$$

$$\widetilde{W}_i^{(2)} = \sum_{\xi\in\mathcal{R}_d} \Big(\min_{k\in S_\xi}(\widehat{U}_{k,i}) + \min_{k\notin S_\xi}(\widehat{U}_{k,i}) - 1\Big)_+ - \sum_{j=1}^{d} A_j^{(2)}(\widehat{U}_{j,i}).$$

The definition of $A_j^{(1)}$ and $A_j^{(2)}$ is rather tricky. Denote

$$\widehat{\boldsymbol{U}}_i^{j+} = (\widehat{U}_{1,i}^{j+}, \ldots, \widehat{U}_{d,i}^{j+})^\top = \widehat{\boldsymbol{U}}_i + \tfrac{1}{\sqrt{n}}\,\boldsymbol{e}_j \quad \text{and} \quad \widehat{\boldsymbol{U}}_i^{j-} = (\widehat{U}_{1,i}^{j-}, \ldots, \widehat{U}_{d,i}^{j-})^\top = \widehat{\boldsymbol{U}}_i - \tfrac{1}{\sqrt{n}}\,\boldsymbol{e}_j.$$

Then, we define

$$\begin{aligned}
A_j^{(1)}(u) = {}& \frac{1}{2\sqrt{n}} \sum_{i'=1}^{n} \sum_{\xi\in\mathcal{R}_d, j\in S_\xi} \Big(1 - \max\Big\{\max_{k\in S_\xi}(\widehat{U}_{k,i'}^{j-}), u\Big\} - \max_{k\notin S_\xi}(\widehat{U}_{k,i'}^{j-})\Big)_+ - \Big(1 - \max\Big\{\max_{k\in S_\xi}(\widehat{U}_{k,i'}^{j+}), u\Big\} - \max_{k\notin S_\xi}(\widehat{U}_{k,i'}^{j+})\Big)_+ \\
&+ \frac{1}{2\sqrt{n}} \sum_{i'=1}^{n} \sum_{\xi\in\mathcal{R}_d, j\notin S_\xi} \Big(1 - \max_{k\in S_\xi}(\widehat{U}_{k,i'}^{j-}) - \max\Big\{\max_{k\notin S_\xi}(\widehat{U}_{k,i'}^{j-}), u\Big\}\Big)_+ - \Big(1 - \max_{k\in S_\xi}(\widehat{U}_{k,i'}^{j+}) - \max\Big\{\max_{k\notin S_\xi}(\widehat{U}_{k,i'}^{j+}), u\Big\}\Big)_+ \\
A_j^{(2)}(u) = {}& \frac{1}{2\sqrt{n}} \sum_{i'=1}^{n} \sum_{\xi\in\mathcal{R}_d, j\in S_\xi} \Big(\min\Big\{\min_{k\in S_\xi}(\widehat{U}_{k,i'}^{j+}), u\Big\} + \min_{k\notin S_\xi}(\widehat{U}_{k,i'}^{j+}\} - 1\Big)_+ - \Big(\min\Big\{\min_{k\in S_\xi}(\widehat{U}_{k,i'}^{j-}), u\Big\} + \min_{k\notin S_\xi}(\widehat{U}_{k,i'}^{j-}\} - 1\Big)_+ \\
&+ \frac{1}{2\sqrt{n}} \sum_{i'=1}^{n} \sum_{\xi\in\mathcal{R}_d, j\notin S_\xi} \Big(\min_{k\in S_\xi}(\widehat{U}_{k,i'}^{j+}\} + \min\Big\{\min_{k\notin S_\xi}(\widehat{U}_{k,i'}^{j+}), u\Big\} - 1\Big)_+ - \Big(\min_{k\in S_\xi}(\widehat{U}_{k,i'}^{j-}\} + \min\Big\{\min_{k\notin S_\xi}(\widehat{U}_{k,i'}^{j-}), u\Big\} - 1\Big)_+ .
\end{aligned}$$

As can be seen this approach to obtain standard errors becomes rather involved in the case of this association measure. For Gini's gamma it is more convenient, from computational point of view, to approximate the standard error of $\widehat{\gamma}_{2n}$ using standard nonparametric bootstrap.

The expressions for the estimated standard errors in case of Kendall's tau, Blomqvist beta, and Spearman's rho (see Sections 7.2, S4.1 and S4.2) on the other hand are advantageous from computational point of view.

S4.4. Standard errors of pairwise-based association measures

We next discuss how to obtain estimates of standard errors for estimators of multivariate association measures based on the pairwise approach discussed in Section 3.1. Let the indices $j,k \in \{1,\ldots,d\}$ be fixed for a moment and $\widehat{\kappa}_n(C_2^{j,k})$ be the estimator of $\kappa(C_2^{j,k})$. As described above one can construct the corresponding variables $W_i^{(j,k)}$ that would be used to estimate the standard deviation of $\widehat{\kappa}_n(C_2^{j,k})$. Then the standard error of $\widehat{\kappa}_n^{PW} = \frac{1}{\binom{d}{2}} \sum_{1\le i<j\le d} \widehat{\kappa}_n(C_2^{j,k})$ is estimated by $a_d\,\frac{\widehat{\sigma}_n}{\sqrt{n}}$, where $\widehat{\sigma}_n^2$ is given by (35) with $W_i = \frac{1}{\binom{d}{2}} \sum_{1\le i<j\le d} W_i^{(j,k)}$ and a_d is an appropriate constant depending only on the dimension d (e.g. $a_d = 2^d/\{2^{d-1}-1\}$ for Kendall's tau and Blomqvist beta).

S5. Real data application: additional material

A heatmap of the empirical pairwise Spearman's rho values is provided in Figure S3.

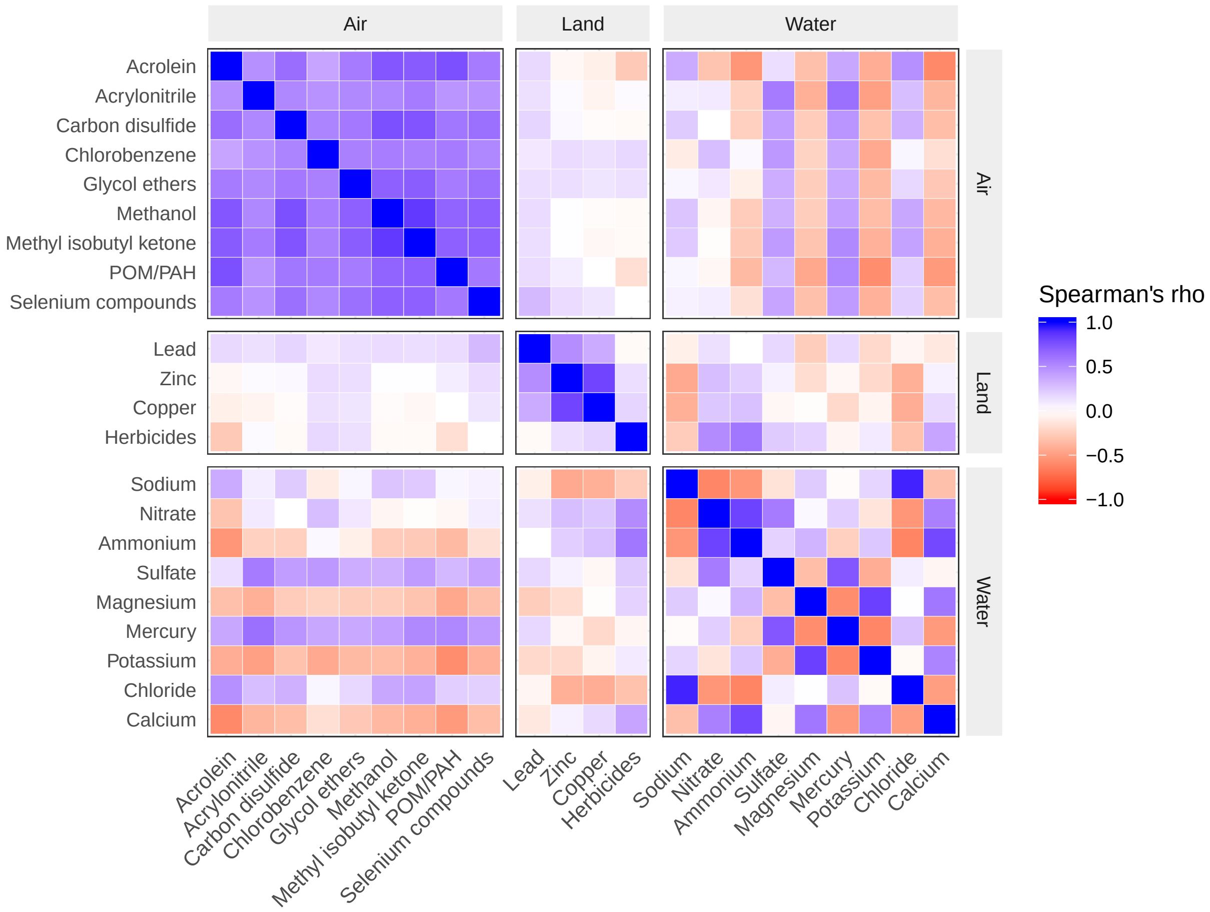

Fig. S3: Spearman's rho heatmap for all the variables used in the data example.

Table S1: Estimated multivariate association measures for triplets within WaterClus6 variables, sorted by value of $\widehat{\rho}_{3n}$.

Variable 1	Variable 2	Variable 3	$\widehat{\rho}_{3n}$	$\widehat{\tau}_n$	$\widehat{\beta}_n$	$\widehat{\gamma}_{2n}$
Nitrate	Ammonium	Calcium	0.71	0.52	0.60	0.60
Magnesium	Potassium	Calcium	0.64	0.46	0.46	0.52
Ammonium	Magnesium	Calcium	0.56	0.41	0.47	0.47
Nitrate	Ammonium	Sulfate	0.52	0.39	0.37	0.41
Ammonium	Potassium	Calcium	0.52	0.37	0.38	0.42
Ammonium	Magnesium	Potassium	0.46	0.34	0.31	0.37
Nitrate	Magnesium	Calcium	0.39	0.27	0.32	0.32
Nitrate	Ammonium	Magnesium	0.38	0.28	0.32	0.32
Nitrate	Sulfate	Calcium	0.36	0.26	0.22	0.27
Nitrate	Potassium	Calcium	0.31	0.21	0.21	0.25
Ammonium	Sulfate	Calcium	0.31	0.24	0.23	0.25
Nitrate	Ammonium	Potassium	0.30	0.22	0.21	0.25
Nitrate	Magnesium	Potassium	0.23	0.18	0.17	0.20
Nitrate	Sulfate	Magnesium	0.09	0.07	0.05	0.06
Sulfate	Magnesium	Calcium	0.07	0.06	0.03	0.05
Ammonium	Sulfate	Magnesium	0.05	0.05	0.03	0.04
Sulfate	Potassium	Calcium	0.02	0.02	-0.05	-0.01
Sulfate	Magnesium	Potassium	0.02	0.04	-0.01	0.02
Ammonium	Sulfate	Potassium	0.00	0.01	-0.05	-0.02
Nitrate	Sulfate	Potassium	0.00	0.02	-0.05	-0.01

We investigate further the association within the water domain. We in particular look into the six variables Nitrate, Ammonium, Magnesium, Potassium, Calcium, and Sulfate. In Table S1 we present the association measures when considering groups of 3 variables from this set, ordered by descending values according to the estimate $\widehat{\rho}_{3n}$. There are $\binom{6}{3} = 20$ such subsets (triplets). It appears that the highest associations in the triplets occur when the variables Nitrate, Ammonium, Magnesium, Potassium and Calcium are involved, whereas the lower trivariate associations all involve sulfate. This led to consider the two subgroups: the group *WaterClus5* of the five variables, Nitrate, Ammonium, Magnesium, Potassium and Calcium, within the water domain; and this group extended with the variable Sulfate leading to the *WaterClus6* group. Similar investigations (not detailed here) led us to consider as a cluster all variables of the air domain together with the variables, Sulfate, Mercury and Chloride from the water domain. We call this cluster of 12 variables, the *AirPlus* cluster.

S6. Multivariate association measures: table with overview

Table S2 contains an overview of how the different multivariate association measures score on the various aspects mentioned in Section 9. Some caution is needed when reading this table.

Table S2: Summary on multivariate association measures and properties.

Aspect	Pairwise association measure	Spearman's rho			Kendall's tau $\tau(C_d)$	Blomqvist's beta $\beta(C_d)$	Gini's gamma	
		$\rho_1(C_d)$	$\rho_2(C_d)$	$\rho_3(C_d)$ $= (\rho_1 + \rho_2)/2$			$\gamma_1(C_d)$	$\gamma_2(C_d)$
defined in	(2)	(5)	(6)		(9)	(S.2)	(14)	(15)/(16)
clarity of interpretation	+	−	−	−	+	+	−	−
complexity	−	+	+	+	+	−	+	+
axioms $(A_1) - (A_8)$	provided κ_2 does	not (A_5)	not (A_5)	all	all	all	not (A_5) & (A_8)	all
Property (P_1)	no	no	no	no	yes	yes	no	no
Property (P_2)	no	no	no	no	yes	yes	no	no
behaviour $d \to \infty$	not studied	+/−	+/−	+/−	+	+	+/−	+/−
analytical computation	+/−	−	−	−	+/−	+	−	+/−
computational cost	+	+	+	+	+/−	+	not studied	−

References

[1] P. Embrechts, A. McNeil, D. Straumann, Correlation and dependence in risk management: Properties and pitfalls, in: RISK Management: Value at Risk and Beyond, Cambridge University Press, 2002, pp. 176–223.

[2] C. Genest, J. Nešlehová, N. B. Ghorbal, Estimators based on Kendall's tau in multivariate copula models, Australian & New Zealand Journal of Statistics 53 (2011) 157–177.

[3] H. Joe, Multivariate Concordance, Journal of Multivariate Analysis 35 (1990) 12–30.

[4] A. J. McNeil, J. Nešlehová, Multivariate Archimedean copulas, d-monotone functions and ℓ_1-norm symmetric distributions, The Annals of Statistics 37 (2009) 3059–3097.

[5] R. B. Nelsen, An Introduction to Copulas, Springer, New York, 2006.

[6] J. Riordan, Combinatorial Identities, Robert E. Krieger Publishing Company, Huntigton, New York, 1979.

[7] F. Schmid, R. Schmidt, Nonparametric inference on multivariate versions of Blomqvist's beta and related measures of tail dependence, Metrika 66 (2007) 323–354.

[8] F. Schmid, R. Schmidt, T. Blumentritt, S. Gaißer, M. Ruppert, Copula-Based Measures of Multivariate Association, in: P. Jaworski, F. Durante, W. Härdle, T. Rychlik (Eds.), Copula Theory and Its Applications: Proceedings of the Workshop Held in Warsaw, 25-26 September 2009, 2010, pp. 209–236.

[9] M. Úbeda-Flores, Multivariate versions of Blomqvist's beta and Spearman's footrule, Annals of the Institute of Statistical Mathematics 57 (2005) 781–788.

[10] W. Wysocki, Kendall's tau and Spearman's rho for n-dimensional Archimedean copulas and their asymptotic properties, Journal of Nonparametric Statistics 27 (2015) 442–459.