REARRANGED DEPENDENCE MEASURES

BY CHRISTOPHER STROTHMANN[1], HOLGER DETTE[2] AND KARL FRIEDRICH SIBURG[1]

[1]*Department of Mathematics, TU Dortmund University, Vogelpothsweg 87, 44221 Dortmund, Germany, christopher.strothmann@mathematik.tu-dortmund.de; karl.f.siburg@mathematik.tu-dortmund.de*

[2]*Department of Mathematics, Ruhr-University Bochum, Universitätsstraße 150, 44780 Bochum, Germany, holger.dette@rub.de*

Most of the popular dependence measures for two random variables X and Y (such as Pearson's and Spearman's correlation, Kendall's τ and Gini's γ) vanish whenever X and Y are independent. However, neither does a vanishing dependence measure necessarily imply independence, nor does a measure equal to 1 imply that one variable is a measurable function of the other. Yet, both properties are natural desiderata for a convincing dependence measure.

In this paper, we present a general approach to transforming a given dependence measure into a new one which exactly characterizes independence as well as functional dependence. Our approach uses the concept of monotone rearrangements as introduced by Hardy and Littlewood and is applicable to a broad class of measures. In particular, we are able to define a rearranged Spearman's ρ and a rearranged Kendall's τ which do attain the value 1 if, and only if, one variable is a measurable function of the other. We also present simple estimators for the rearranged dependence measures, prove their consistency and illustrate their finite sample properties by means of a simulation study.

1. Introduction. One of the most fundamental problems in statistics is to measure the association between two random variables X and Y based on a sample of independent identically distributed observations $(X_1, Y_1), \dots, (X_n, Y_n)$, and numerous proposals have been made for this purpose. These measures usually vary in the interval $[0, 1]$ or $[-1, 1]$, and vanish if the variables are independent. Moreover, many of these measures, including the frequently used Pearson's and Spearman's correlation, Kendall's τ and Gini's γ, are very powerful to detect linear and monotone dependencies. On the other hand, in general, a vanishing dependence measure (such as Pearson's coefficient) only implies independence of X and Y under quite restrictive additional assumptions (such as a normal distribution), and it is a well known fact that many of these measures cannot detect non-monotone associations.

Several authors have proposed solutions to this problem by introducing alternative dependence measures, but mainly in the context of testing for independence. Among the many contributions, we mention exemplary the early work of Blum, Kiefer and Rosenblatt (1961); Rosenblatt (1975); Schweizer and Wolff (1981); Csörgő (1985) and the more recent papers by Székely, Rizzo and Bakirov (2007); Gretton et al. (2008); Bergsma and Dassios (2014) and Zhang (2019). However, as pointed out by Chatterjee (2021), these measures are designed primarily for testing independence, and not for measuring the strength of the relationship between the variables. In the same paper, a new correlation coefficient is presented, which estimates a (population) measure, say μ, of the dependence between two random variables X and Y with the following properties:

MSC2020 subject classifications: Primary 62H20; secondary 62H05.

Keywords and phrases: measure of dependence, coefficient of correlation, decreasing rearrangement, copula.

(1.1) $0 \le \mu(X,Y) \le 1$

(1.2) $\mu(X,Y) = 0$ if, and only if, X and Y are independent

(1.3) $\mu(X,Y) = 1$ if, and only if, $Y = f(X)$ for some measurable function f.

For continuous distributions, Chatterjee's measure had already been introduced and studied in Dette, Siburg and Stoimenov (2012) who also proposed a kernel based estimator for it. Since its introduction, Chatterjee's correlation coefficient has found considerable attention in the literature (see Cao and Bickel, 2020; Deb, Ghosal and Sen, 2020; Gamboa et al., 2020; Shi, Drton and Han, 2021a,b; Auddy, Deb and Nandy, 2021; Lin and Han, 2021, among others), which underlines the demand for dependence measures possessing the above properties (1.1)–(1.3).

This paper takes a quite different viewpoint on this problem by formulating the following question:

Is it possible to transform a given dependence measure in such a way that the new dependence measure satisfies properties (1.1)–(1.3)?

Our answer to this question is affirmative. More precisely, we will show that there exists a well defined transformation $\mu \mapsto R_\mu$ with the following property. Whenever the dependence measure μ satisfies the axioms (1.1) to (1.3) on the set of *stochastically increasing* continuous distributions, the new dependence measure R_μ will satisfy (1.1) to (1.3) on the set of *all* continuous distributions. By definition, a pair (X,Y) of random variables is stochastically increasing if the function $x \mapsto \mathbb{P}(Y \le y \mid X = x)$ is decreasing for each fixed y (see, e.g. Nelsen, 2006). This property was also discussed earlier in Lehmann (1959) under the term *positive regression dependence*.

The transformed dependence measure R_μ will be called the *rearranged dependence measure*. It turns out that the new transformation is applicable to many of the classical dependence measures and, consequently, enables us to define rearranged dependence measures such as the rearranged Spearman's ρ and the rearranged Kendall's τ, all of which satisfy properties (1.1)–(1.3).

Our approach is based on a classical concept from majorization theory which is called *monotone rearrangement* (see, for instance, Hardy, Littlewood and Pólya, 1988; Ryff, 1965, 1970). In the last decades, monotone rearrangements have found considerable interest in the statistical literature. For example, Dette, Neumeyer and Pilz (2006); Chernozhukov, Fernández-Val and Galichon (2009); Anevski and Fougères (2019); Camirand-Lemyre, Carroll and Delaigle (2022) used this concept to define (smooth) monotone estimates, while Dette and Volgushev (2008); Chernozhukov, Fernández-Val and Galichon (2010) successfully applied rearrangements techniques to define quantile regression estimates without crossing. Recently, Dette and Wu (2019) used monotone rearrangements to detect relevant changes in a (not necessarily monotone) trend of a non-stationary time series.

Our paper is organized as follows. In Section 2, we recall the concept of monotone rearrangements and introduce our transformation of a given dependence measure to a new measure with the desired properties (1.1)–(1.3) in several steps. First, we characterize the dependence measure $\mu(X,Y) = \mu(C)$ in terms of the copula C of the corresponding distribution function of (X,Y). Then we apply a monotone rearrangement to the partial derivative of C with respect to its first argument, which essentially constitutes the conditional distribution[1] $u \mapsto \mathbb{P}(F_Y(Y) \le v \mid F_X(X) = u)$, and integrate it with respect to the conditioning

[1] F_X and F_Y denote the marginal distributions of X and Y, respectively.

coordinate. The resulting rearranged copula is denoted by $C^{\uparrow}$ and, roughly speaking, it can be shown that the *rearranged dependence measure*

$$R_{\mu}(C) := \mu(C^{\uparrow})$$

satisfies the desired properties (1.1)–(1.3). In Section 3, we propose an estimate of the rearranged dependence measure $R_{\mu}(C)$, which is obtained by applying the procedure to the so-called checkerboard copula (see Li et al., 1997, for example). We also prove consistency of the estimate and illustrate the finite sample properties of our approach by means of a small simulation study in Section 4. Finally, all proofs are deferred to appendices which also contain some general results on monotone rearrangements, which will be used for our theoretical arguments.

2. Dependence measures with properties (1.1)–(1.3). In this section, we construct a rearranging transformation which assigns a new measure R_{μ} with the desired properties (1.1)–(1.3) to a given dependence measure μ. We also discuss some further nice properties of the rearranged measure. To be precise, let (X, Y) denote a 2-dimensional random vector with continuous distribution function F and marginal distribution functions F_X and F_Y. The dependence structure of X and Y is completely encoded in the (unique) copula $C = C_{X,Y}$ (see Definition A.1 in the appendix) defined by the equation

$$C(F_X(x), F_Y(y)) = F(x, y)$$

as described, for instance, in Nelsen (2006). The class of all copulas corresponding to continuous 2-dimensional distributions is denoted by $\mathcal{C}$.

2.1. *New dependence measures by monotone rearrangements.* We restrict ourselves to dependence measures which can be represented as a function of the copula[2] and consequently use the notations $\mu(X, Y)$ and $\mu(C)$ interchangeably throughout this paper. The key ingredient is a rearrangement of the conditional distribution functions

$$u \mapsto \mathbb{P}(F_Y(Y) \le v \mid F_X(X) = u) = \partial_1 C(u, v) := \frac{\partial}{\partial u} C(u, v) \tag{2.1}$$

of the vector $(F_X(X), F_Y(Y))$. Note that the partial derivative $\partial_1 C(u, v)$ is only defined almost everywhere. We will suppress this fact in our notation for the remainder of this article.

DEFINITION 2.1. A copula $C \in \mathcal{C}$ is called *stochastically increasing (resp. decreasing)* if $u \mapsto \partial_1 C(u, v)$ is decreasing (resp. increasing) for each v. The class of all stochastically increasing copulas is denoted by $\mathcal{C}^{\uparrow}$. A copula C is called *stochastically monotone* if it is either stochastically increasing or decreasing. Similarly, a random variable Y is stochastically increasing (resp. decreasing/monotone) in X if C_{XY} is stochastically increasing (resp. decreasing/monotone).

We will now introduce a procedure transforming an arbitrary copula into a stochastically increasing one. It is based on the monotone rearrangement of a univariate function, which is a classical concept in majorization theory (see, for example, Chong and Rice, 1971; Bennett and Sharpley, 1988). Namely, if λ denotes the Lebesgue measure and $f : [0, 1] \to \mathbb{R}$ is a Borel measurable function, then the *decreasing rearrangement* $f^* : [0, 1] \to \mathbb{R}$ of f is defined by

$$f^*(t) := \inf\{x \mid \lambda(\{t \in [0, 1] \mid f(t) > x\}) \le t\} . \tag{2.2}$$

Obviously, the function f^* is a decreasing function and we have $f^* = f$ whenever f is decreasing and right-continuous.

[2] Any dependence measure $\mu(X, Y)$ induces a dependence measure $\mu(F_X(X), F_Y(Y))$ depending only on the copula. Thus our approach does not imply any restriction.

DEFINITION 2.2. The *stochastically increasing rearrangement*, (SI)-rearrangement in short, of a copula $C \in \mathcal{C}$ is defined as

$$C^{\uparrow}(u,v) := \int_0^u (\partial_1 C)^*(s,v)\,\mathrm{d}s$$

where the rearrangement (2.2) is applied to the first coordinate of $\partial_1 C(u,v)$.

Our next result shows that $C^{\uparrow}$ defines in fact a copula.[3]

THEOREM 2.3. *The (SI)-rearrangement $C^{\uparrow}$ of a copula C is a stochastically increasing copula. Moreover, $C^{\uparrow} = C$ if and only if C is stochastically increasing itself.*

For a given dependence measure μ, we now define a new dependence measure by

$$R_\mu(C) := \mu(C^{\uparrow}). \tag{2.3}$$

We call R_μ the *rearranged dependence measure* obtained from μ. Note that, in general, R_μ differs from μ and hence yields a new measure of dependence. Our main result is the following:

THEOREM 2.4. *Suppose μ is a dependence measure which, when restricted to the set $\mathcal{C}^{\uparrow}$, satisfies the properties (1.1)–(1.3). Then the rearranged dependence measure R_μ satisfies the properties (1.1)–(1.3) on the whole set $\mathcal{C}$.*

REMARK 2.5. Recently, dependence measures with the properties (1.1)–(1.3) have found considerable attention in the literature. For example, Trutschnig (2011) defined the measure

$$\zeta_1(C) = 3\int_0^1\int_0^1 |\partial_1 C(u,v) - v|\,\mathrm{d}u\,\mathrm{d}v,$$

while Dette, Siburg and Stoimenov (2012) and Chatterjee (2021) considered (and proposed estimates for) the measure

$$r(C) = 6\int_0^1\int_0^1 (\partial_1 C(u,v) - v)^2\,\mathrm{d}u\,\mathrm{d}v. \tag{2.4}$$

It will be shown in Appendix B that the stochastically increasing rearrangement captures the entire information about the degree of dependence as defined by these measures in the sense that

$$\zeta_1(C) = \zeta_1(C^{\uparrow}) \text{ as well as } r(C) = r(C^{\uparrow}). \tag{2.5}$$

2.2. *Examples.* In this section, we illustrate the rearrangement approach by a couple of examples. In particular, our method is applicable to construct a rearranged Spearman's ρ or Kendall's τ from the classical measures of concordance. Moreover, we derive some interesting properties of the rearranged dependence measures.

[3]The analogous definition of the stochastically decreasing rearrangement copula $C^{\downarrow}$ is given and discussed in Appendix B.2; see also (Ansari and Rüschendorf, 2021).

EXAMPLE 2.6 (Schweizer-Wolff measures). Let $\Pi(u,v) = uv$ denote the independence copula. Each L^p-norm with $1 \le p < \infty$ defines a so-called Schweizer-Wolff measure

$$\sigma_p(C) := \frac{\|C - \Pi\|_p}{\|C^+ - \Pi\|_p}\,, \tag{2.6}$$

where the copula C^+ is defined by $C^+(u,v) = \min\{u,v\}$ (see Appendix A). The measure σ_1 was considered in Schweizer and Wolff (1981), the general case $p \ge 1$ can be found in Section 5.3.1 of Nelsen (2006). It is easy to see that properties (1.1) and (1.2) hold for σ_p, and it is well known that $\sigma_p(C) = 1$ if and only $Y = f(X)$ for some strictly monotone (and not just measurable) function f (Nelsen, 2006, Sect. 5.3.1). Consequently, σ_p does *not* satisfy property (1.3). On the other hand, it will be shown in Appendix B.4 that the properties (1.1)–(1.3) do hold for the restriction of σ_p to the set $\mathcal{C}^\uparrow$. Therefore, the rearranged Schweizer-Wolff measure

$$R_{\sigma_p}(C) = \frac{\left\|C^\uparrow - \Pi\right\|_p}{\|C^+ - \Pi\|_p}$$

defines a new dependence measure on $\mathcal{C}$ satisfying all the properties (1.1)–(1.3).

EXAMPLE 2.7 (Measures of concordance). Let $\kappa : \mathcal{C} \to [-1,1]$ be a measure of concordance (see Definition A.5). Typical examples include Spearman's ρ, Kendall's τ, Gini's γ, and Blomqvist's β (see Appendix B.5 for a representation of these measures in terms of the copula). We will prove in Appendix B.5 that the measures ρ, τ and γ satisfy (1.1)–(1.3) on the set $\mathcal{C}^\uparrow$ (but not on $\mathcal{C}$). On the other hand, Blomqvist's β does not satisfy (1.3) on $\mathcal{C}^\uparrow$.

Consequently, by Theorem 2.4, the rearranged Spearman's ρ (R_ρ), Kendall's τ (R_τ) and Gini's γ (R_γ) define dependence measures (different from their original measures) satisfying the properties (1.1)–(1.3).

Surprisingly, the Schweizer-Wolff measure σ_1 and Spearman's ρ induce the same rearranged dependence measure.

PROPOSITION 2.8. *We have* $R_{\sigma_1} = R_\rho$.

While a measure of concordance κ measures the strength of the monotone association between two random variables, the corresponding rearranged dependence measure R_κ measures the strength of their (directed) functional relationship. Thus, intuitively, κ should always attain smaller values than R_κ. This heuristic is confirmed by the next theorem, which applies, in particular, to Spearmans ρ and Kendalls τ.

THEOREM 2.9. *Let κ be a measure of concordance satisfying (1.1)–(1.3) on the set $\mathcal{C}^\uparrow$. Then*

$$|\kappa(C)| \le R_\kappa(C)$$

for all $C \in \mathcal{C}$, with equality whenever C is stochastically monotone.

2.3. *Data processing inequality and self-equitability.* Informally, the so-called data processing inequality states that a (random or functional) modification of the input data cannot increase the information contained in the data; see, for example, Cover and Thomas (2006) for an in-depth treatment of the data processing inequality in the context of information theory.

We assume in the following that the dependence measure μ is monotone with respect to the pointwise order on $\mathcal{C}^{\uparrow}$, i.e.

$$C_1 \leq C_2 \implies \mu(C_1) \leq \mu(C_2) \tag{2.7}$$

for all $C_1, C_2 \in \mathcal{C}^{\uparrow}$. Note that this monotonicity condition holds for many dependence measures. For example, (2.7) is satisfied for any concordance measure (see Definition A.5 for a precise definition), the Schweizer-Wolff measures σ_p in (2.6) as well as the measures of complete dependence ζ_1 and r introduced in Remark 2.5.

PROPOSITION 2.10 (Data processing inequality). *Assume that the dependence measure μ satisfies (2.7), and let X, Y, Z be continuous random variables such that Y and Z are conditionally independent given X. Then the data processing inequality*

$$R_\mu(Z, Y) \leq R_\mu(X, Y)$$

holds. In particular, $R_\mu(f(X), Y) \leq R_\mu(X, Y)$ holds for all[4] *measurable functions f.*

Similar to (Geenens and Lafaye de Micheaux, 2020, Proposition 2.1), the data processing inequality also immediately yields an asymmetric version of the so-called self-equitability introduced in Kinney and Atwal (2014).

COROLLARY 2.11. *Assume that μ satisfies (2.7). If f is a measurable function such that X and Y are conditionally independent given $f(X)$, then*

$$R_\mu(f(X), Y) = R_\mu(X, Y)\ .$$

In particular, $R_\mu(g(X), Y) = R_\mu(X, Y)$ holds for all measurable bijections g.

Intuitively, Corollary 2.11 states that, in a regression model $Y = f(X) + \epsilon$, the dependence measure $R_\mu(X, Y)$ depends only on the strength of the noise ϵ and not on the specific form of f. A similar idea is illustrated in Figures 3 and 4 of Junker, Griessenberger and Trutschnig (2021).

3. Approximation and estimation. In general, the computation of the rearrangement of a function, and hence the computation of $C^{\uparrow}$, may be a difficult task. In this section, we discuss techniques to approximate $C^{\uparrow}$ and $R_\mu(C)$ and to estimate the rearranged dependence measure R_μ from a sample of independent and identically distributed observations $(X_1, Y_1), \ldots, (X_n, Y_n)$. In principle, one would like to estimate the copula C through a "smooth" statistic, say $\hat{C}_n$, and then apply Definition 2.2 to calculate the rearrangement $\hat{C}_n^{\uparrow}$ and the rearranged dependence measure

$$R_\mu(\hat{C}_n) = \mu(\hat{C}_n^{\uparrow}). \tag{3.1}$$

While various smooth estimators have been proposed (see Fermanian, Radulović and Wegkamp, 2004; Chen and Huang, 2007; Omelka, Gijbels and Veraverbeke, 2009; Genest, Nešlehovà and Rèmillard, 2017, among others), the simultaneous estimation of the rearrangement poses various difficulties. We will now propose a simple solution to this problem.

Our approach is based on an approximation scheme for $C^{\uparrow}$ in the theoretical as well as empirical setting using the concept of checkerboard copulas, thereby circumventing the need to treat partial derivatives explicitly. Checkerboard copulas are an important tool in statistical applications; for a detailed discussion we refer, among others, to Genest, Nešlehovà and

[4]Note that for $R_\mu(f(X), Y)$ to be well-defined, $f(X)$ needs to be a continuous random variable.

Rèmillard (2017) and Junker, Griessenberger and Trutschnig (2021). To be precise let $A = (a_{k\ell})_{k=1,\dots,N_1}^{\ell=1,\dots,N_2} \in \mathbb{R}^{N_1 \times N_2}$ denote a matrix with entries $a_{k\ell}$ satisfying

$$
\begin{aligned}
a_{k\ell} &\geq 0 \quad \text{for all } k = 1,\dots,N_1 \text{ and } \ell = 1,\dots,N_2 \,, \\
\sum_{k=1}^{N_1} a_{k\ell} &= N_1 \quad \text{for all } \ell = 1,\dots,N_2 \,, \\
\sum_{\ell=1}^{N_2} a_{k\ell} &= N_2 \quad \text{for all } k = 1,\dots,N_1 \,.
\end{aligned} \tag{3.2}
$$

Then the function $C^{\#}_{N_1,N_2}(A) : [0,1]^2 \to [0,1]$ defined by

$$
C^{\#}_{N_1,N_2}(A)(u,v) := \sum_{k,\ell=1}^{N_1,N_2} a_{k\ell} \int_0^u \mathbb{1}_{\left[\frac{k-1}{N_1},\frac{k}{N_1}\right)}(s)\,\mathrm{d}s \int_0^v \mathbb{1}_{\left[\frac{\ell-1}{N_2},\frac{\ell}{N_2}\right)}(t)\,\mathrm{d}t \tag{3.3}
$$

is a copula and called the *checkerboard copula of the matrix* A. For a copula C (see Definition A.1) its *induced checkerboard copula* is defined as

$$
C^{\#}_{N_1,N_2}(C) := C^{\#}_{N_1,N_2}(A_{N_1,N_2}) \,, \tag{3.4}
$$

where the elements of the doubly stochastic matrix A_{N_1,N_2} are given by

$$
(A_{N_1,N_2})_{k\ell} := N_1 N_2 \cdot V_C\left(\left[\frac{k-1}{N_1},\frac{k}{N_1}\right] \times \left[\frac{\ell-1}{N_2},\frac{\ell}{N_2}\right]\right) \tag{3.5}
$$

and $V_C(B)$ denotes the measure of the (Borel-)set $B \subset [0,1]^2$ induced by the copula C.

Note that in contrast to most of the literature, we define a (empirical) checkerboard copula also for non-square matrices A satisfying (3.2). For $N = N_1 = N_2$ the representation (3.3) essentially reduces, up to a scaling factor N, to the common definition based on doubly stochastic square matrices (see Genest, Nešlehovà and Rèmillard, 2017; Junker, Griessenberger and Trutschnig, 2021). The consideration of the rectangular case, however, is necessary to address asymmetric dependencies between X and Y resp. Y and X.

We point out that the partial derivatives of the copula $C^{\#}_{N_1,N_2}(A)$ in (3.3) are piecewise constant for fixed $v \in [0,1]$ with

$$
\partial_1 C^{\#}_{N_1,N_2}(A)\left(u, \frac{j}{N_2}\right) = \frac{1}{N_2} \sum_{\ell=1}^{j} a_{k\ell} \quad \text{for } u \in \left[\frac{k-1}{N_1}, \frac{k}{N_1}\right) \,.
$$

Thus, the (SI)-rearrangement satisfies $C^{\#}_{N_1,N_2}(A)^{\uparrow} = C^{\#}_{N_1,N_2}(A)$ if and only if

$$
\sum_{j=1}^{\ell} a_{k_2 j} \leq \sum_{j=1}^{\ell} a_{k_1 j} \tag{3.6}
$$

for all $1 \leq \ell \leq N_2$ and all $1 \leq k_1 \leq k_2 \leq N_1$. In other words, $C^{\#}_{N_1,N_2}(A)^{\uparrow} = C^{\#}_{N_1,N_2}(A)$ if and only if the rows of A are ordered with respect to the majorization ordering of vectors (see Marshall, Olkin and Arnold, 2011). This suggests the following Algorithm 1 for calculating the (SI)-rearrangement (as defined in Definition 2.2) of an arbitrary checkerboard copula.

THEOREM 3.1. *For any matrix* $A \in \mathbb{R}^{N_1 \times N_2}$ *satisfying* (3.2)*, the function* $C^{\#}_{N_1,N_2}(A)^{\uparrow}$ *defined in Algorithm 1 is the (SI)-rearrangement of the checkerboard copula* $C^{\#}_{N_1,N_2}(A)$.

Algorithm 1: Rearranged checkerboard copula

Data: matrix $A \in \mathbb{R}^{N_1 \times N_2}$ with entries satisfying (3.2)
Result: (SI)-rearrangement $C^{\#}_{N_1,N_2}(A)^{\uparrow}$ of the checkerboard copula $C^{\#}_{N_1,N_2}(A)$

(1) Calculate $B_k^{\ell} := \sum_{j=1}^{\ell} a_{kj}$ and set $B_k^0 := 0$.
(2) For every $\ell = 0, \ldots, N_2$, sort B_k^{ℓ} in a decreasing order and denote the result by $\widetilde{B}_k^{\ell}$.
(3) Calculate $a_{k\ell}^{\uparrow}$ iteratively using

$$a_{k\ell}^{\uparrow} := \widetilde{B}_k^{\ell} - \widetilde{B}_k^{\ell-1} \geq 0\,.$$

(4) Define $A^{\uparrow} := (a_{k\ell}^{\uparrow})_{k=1,\ldots,N_1}^{\ell=1,\ldots,N_2}$ and

$$C^{\#}_{N_1,N_2}(A)^{\uparrow} := C^{\#}_{N_1,N_2}(A^{\uparrow})\,.$$

We now turn to the estimation of the population dependence measure $R_\mu(C) = \mu(C^{\uparrow})$ from a sample of independent and identically distributed observations. Because there exists in general no analytic expression for $R_\mu(C)$, this is a challenging task and we proceed in two steps. First, note that the population measure $R_\mu(C)$ can be approximated by $R_\mu(C^{\#}_{N_1,N_2}(C))$ using the induced checkerboard copula $C^{\#}_{N_1,N_2}(C)$ of C defined in (3.4) since

$$C^{\#}_{N_1,N_2}(C)^{\uparrow} \to C^{\uparrow} \tag{3.7}$$

where $C^{\#}_{N_1,N_2}(C)^{\uparrow}$ denotes the rearrangement of $C^{\#}_{N_1,N_2}(C)$. Secondly, we replace the unknown weights in (3.5) by corresponding estimates to obtain an empirical checkerboard copula, which is then rearranged by Algorithm 1.

We begin with the approximation of $C^{\uparrow}$ by the rearranged induced checkerboard copula. Since it is well known that the pointwise convergence is unable to capture complete dependence (see Mikusiński, Sherwood and Taylor, 1992), we consider the finer metrics

$$D_p(C_1, C_2) := \left(\int_0^1 \int_0^1 \left| \partial_1 C_1(u,v) - \partial_1 C_2(u,v) \right|^p \,\mathrm{d}u\,\mathrm{d}v \right)^{\frac{1}{p}} \tag{3.8}$$

for $1 \leq p < \infty$ introduced in Trutschnig (2011).

THEOREM 3.2. *For any copula C, the rearranged induced checkerboard copula $C^{\#}_{N_1,N_2}(C)^{\uparrow}$ converges to the rearranged copula $C^{\uparrow}$ with respect to D_p, i.e.*

$$D_p(C^{\#}_{N_1,N_2}(C)^{\uparrow}, C^{\uparrow}) \to 0\,.$$

In particular, $C^{\#}_{N_1,N_2}(C)^{\uparrow}$ converges uniformly towards $C^{\uparrow}$.

In order to carry over the convergence of $C_n^{\uparrow}$ to $C^{\uparrow}$ and establish consistency of the estimator, we require that the underlying dependence measure μ is continuous on $\mathcal{C}^{\uparrow}$ with respect to pointwise convergence, i.e. that

$$C_n \to C \implies \mu(C_n) \to \mu(C) \tag{3.9}$$

holds for all copulas $C_n, C \in \mathcal{C}^{\uparrow}$. We point out that most classical measures are continuous in this sense. In fact, any concordance measure (see Definition A.5), the Schweizer-Wolff measures σ_p in (2.6), as well as the measures of complete dependence ζ_1 and r in Remark 2.5 fulfil our continuity condition[5].

THEOREM 3.3. *If the dependence measure μ satisfies (3.9) then*

$$R_\mu(C^{\#}_{N_1,N_2}(C)^{\uparrow}) \to R_\mu(C)\ .$$

Next, we consider a random sample of independent identically distributed observations $(X_1, Y_1), \dots, (X_n, Y_n)$. Similar to Li, Mikusiński and Taylor (1998) and Junker, Griessenberger and Trutschnig (2021), who considered the case $N_1 = N_2$, we define the empirical checkerboard copula with bandwidth $N_1, N_2 < n$ by

$$\hat{C}^{\#}_{N_1,N_2,n} := C^{\#}_{N_1,N_2}\left(C^{\#}_{n,n}(\hat{A}_n)\right)\ , \tag{3.10}$$

where $\hat{A}_n = (\hat{a}_{ij})$ is the $n \times n$ permutation matrix defined by

$$\hat{a}_{ij} := \begin{cases} 1 & \text{if there exists some } k \text{ with } \operatorname{rank}(X_k) = i \text{ and } \operatorname{rank}(Y_k) = j \\ 0 & \text{else} \end{cases}$$

and $\operatorname{rank}(x_k)$ denotes the rank of x_k among $x_1, x_2, \dots, x_n$. Finally, we define

$$\hat{R}_\mu := R_\mu(\hat{C}^{\#}_{N_1,N_2,n}) \tag{3.11}$$

as an estimator of $R_\mu(C)$, which will be called *rearranged μ-estimate* throughout this paper. The following result shows strong consistency of $\hat{R}_\mu$.

THEOREM 3.4. *Assume that the dependence measure μ fulfils the assumption (3.9), and let $(X_1, Y_1), \dots, (X_n, Y_n)$ denote independent identically distributed random variables with a continuous distribution. If $N_1 := \lfloor n^{s_1} \rfloor$, $N_2 := \lfloor n^{s_2} \rfloor$ with $s_1, s_2 \in (0, 1/2)$, then the estimator defined by (3.11) satisfies*

$$\hat{R}_\mu \to R_\mu(C) \text{ a.s. as } n \to \infty\ .$$

4. Finite sample properties. For a good performance of the estimate $\hat{C}^{\#}_{N_1,N_2,n}$, an appropriate choice of the bandwidths N_1, N_2 will be crucial. These tuning parameters depend sensitively on the form of the underlying unknown copula, and for the finite sample illustrations presented below, we will use the following cross validation principle.

Recall the definition of the empirical checkerboard copula $\hat{C}^{\#}_{N_1,N_2,n}$, and denote its corresponding density by

$$\hat{c}_{N_1,N_2,n}(u,v) := \frac{\partial^2}{\partial u \partial v} \hat{C}^{\#}_{N_1,N_2,n}(u,v)\ . \tag{3.12}$$

We define

$$\mathrm{CV}(N_1, N_2, n) := \int_0^1 \int_0^1 \hat{c}^2_{N_1,N_2,n}(u,v)\, \mathrm{d}u\, \mathrm{d}v - \frac{2}{n} \sum_{i=1}^{n} \hat{c}^{-i}_{N_1,N_2,n-1}(\hat{U}_i, \hat{V}_i)\ ,$$

[5] For ζ_1 and r this follows from (Siburg and Strothmann, 2021, Prop. 3.6).

where $\hat{c}^{-i}_{N_1,N_2,n-1}$ denotes the estimator in (3.12) calculated from the data

$$(X_1,Y_1),\ldots,(X_{i-1},Y_{i-1}),(X_{i+1},Y_{i+1}),\ldots,(X_n,Y_n)$$

and $\hat{U}_i = \frac{1}{n+1}\sum_{j=1}^n I\{X_j \le X_i\}$ and $\hat{V}_i = \frac{1}{n+1}\sum_{j=1}^n I\{Y_j \le Y_i\}$ are the normalized ranks of X_i and Y_i among $X_1,\ldots,X_n$ and $Y_1,\ldots,Y_n$, respectively. The data adaptive choice of the parameters N_1 and N_2 is defined as the minimizer of $\mathrm{CV}(N_1,N_2,n)$ with respect to $N_1,N_2 \in \{\lfloor n^{1/4}\rfloor,\ldots,\lfloor n^{1/2}\rfloor\}$. In cases, where the set of possible bandwidths is very large, we calculate the minimizer in the set $\{\lfloor n^{1/4}\rfloor,\lfloor n^{1/4}\rfloor+2,\ldots,\lfloor n^{1/2}\rfloor\}$ in order to save computational time.

4.1. *Simulation study.* In this section, we present results from a simulation study investigating the performance of the estimator $\hat{R}_\mu$ defined in (3.11). All simulations have been conducted using the statistical software "R" (see R Core Team, 2021) and are based on 1000 replications in each scenario. The package "qad" (see Griessenberger et al., 2021) was used in a slightly adapted form to calculate the matrix $\hat{A}_n$, which is required for the definition of the empirical checkerboard copula in (3.10). As sample sizes we considered $n = 50, 100, 500$ and 1000 and N_1, N_2 were chosen by the cross validation procedure described at the beginning of this section.

4.1.1. *Stochastically increasing distributions.* We begin with a study of the properties of the estimator (3.11) in the rather special case where the underlying copula is stochastically increasing. The corresponding samples have been generated using the package "copula" (see Hofert et al., 2020). As for stochastically monotone copulas we have $R_\mu = \mu$, we can calculate the dependence measure explicitly, and it is also reasonable to compare the new estimator $\hat{R}_\mu$ with commonly used estimators of μ.

The first two scenarios correspond to a 2-dimensional (centred) normal distribution with correlation matrix

$$R = \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix} \tag{3.13}$$

where $r = 0.25$ and $r = 0.75$, respectively. Since for $r > 0$, the corresponding copula, say C_r, is stochastically increasing, the rearranged Spearman's ρ equals

$$R_\rho(C_r) = R(X,Y) = \rho(X,Y) = \frac{6}{\pi}\arcsin\left(\frac{r}{2}\right) \quad (r \ge 0)\,,$$

while the rearranged Kendall's τ equals

$$R_\tau(C_r) = R_\tau(X,Y) = \tau(X,Y) = \frac{2}{\pi}\arcsin(r) \quad (r \ge 0)\,.$$

The third example of a stochastically increasing copulas is a member of both the Archimedean and extreme-value copula families, which are widely applied, both theoretically as well as empirically. More precisely, we consider a Gumbel copula defined by

$$C^G_\theta(u,v) := \exp\left(-\left((-\log u)^\theta + (-\log v)^\theta\right)^{1/\theta}\right)\,, \tag{3.14}$$

where $\theta > 1$ denotes a parameter. Since the Gumbel copula is an extreme-value copula, it is stochastically increasing, where the rearranged Spearman's ρ and Kendall's τ are given by

$$R_\rho(C^G_\theta) = \rho(X,Y) = 12\int_0^1 \frac{1}{\left(1+(t^\theta+(1-t)^\theta)^{1/\theta}\right)^2}\,\mathrm{d}t - 3\,,$$

$$\tau(C^G_\theta) = \tau(X,Y) = \frac{\theta-1}{\theta}\,.$$

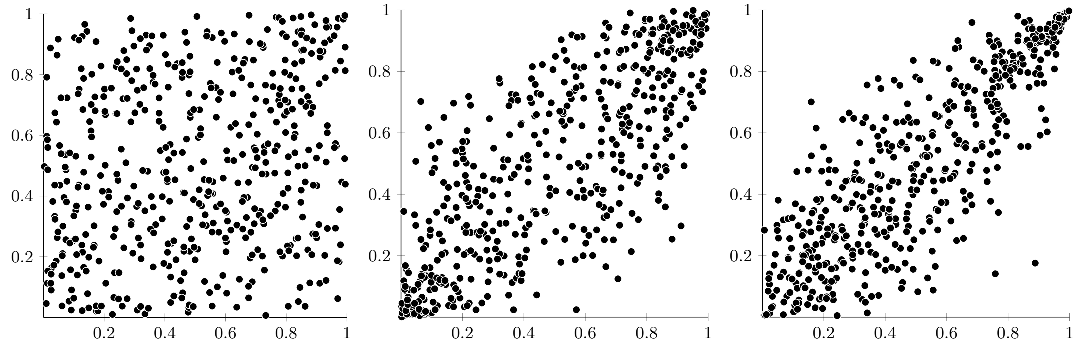

Fig 1: *Scatter plots of data (sample size $n = 500$) from the Gaussian copula with correlation $r = 0.25$ (left panel), $r = 0.75$ (middle panel) and the Gumbel copula with parameter $\theta = 3$ (right panel).*

copula	n	Spearman's ρ R_ρ	$\hat{R}_\rho$	$\lvert\hat{\rho}\rvert$	Kendall's τ R_τ	$\hat{R}_\tau$	$\lvert\hat{\tau}\rvert$
$C_{0.25}$	50	0.239	0.276 (0.132)	0.246 (0.129)	0.161	0.185 (0.090)	0.169 (0.090)
	100		0.263 (0.093)	0.236 (0.096)		0.176 (0.063)	0.160 (0.066)
	500		0.224 (0.043)	0.240 (0.043)		0.150 (0.029)	0.162 (0.029)
	1000		0.226 (0.030)	0.239 (0.029)		0.151 (0.020)	0.160 (0.020)
$C_{0.75}$	50	0.734	0.669 (0.094)	0.721 (0.075)	0.540	0.473 (0.079)	0.538 (0.068)
	100		0.694 (0.063)	0.727 (0.051)		0.496 (0.055)	0.539 (0.047)
	500		0.714 (0.025)	0.732 (0.023)		0.517 (0.023)	0.539 (0.020)
	1000		0.723 (0.017)	0.734 (0.015)		0.527 (0.015)	0.540 (0.014)
C_3^G	50	0.848	0.803 (0.057)	0.839 (0.050)	0.667	0.599 (0.058)	0.668 (0.055)
	100		0.826 (0.040)	0.844 (0.037)		0.628 (0.044)	0.667 (0.041)
	500		0.844 (0.016)	0.848 (0.015)		0.653 (0.019)	0.666 (0.017)
	1000		0.847 (0.011)	0.848 (0.010)		0.659 (0.012)	0.666 (0.012)

TABLE 1

Simulated mean and standard deviation of the rearranged Spearman's ρ estimate $\hat{R}_\rho$, the (absolute) Spearman's rank correlation coefficient $|\hat{\rho}|$ (left part), the rearranged Kendall's τ estimate $\hat{R}_\tau$, (absolute) Kendall's rank correlation coefficient $|\hat{\tau}|$ (right part). The distribution of (X, Y) is given by a centred normal with correlation matrix (3.13) with copula C_r and by a Gumbel copula C_3^G.

In Figure 1, we show scatter plots of data generated from the two Gaussian copulas ($r = 0.25$, $r = 0.75$) and the Gumbel copula ($\theta = 3$), where the sample size is $n = 500$. In Table 1, we present the simulated mean and standard deviation of the rearranged estimate $\hat{R}_\mu$, where μ is either Spearman's ρ (left part) or Kendall's τ (right part). Due to $R_\mu(C) = \mu(C)$ for the three scenarios, the commonly used Spearman's rank correlation coefficient $\hat{\rho}$ and Kendall's rank correlation coefficient $\hat{\tau}$ can also be used to estimate $R_\rho(C)$ and $R_\tau(C)$, respectively. The corresponding results for these estimates are displayed in Table 1 as well (of course, in practice it is not known if the underlying copula is stochastically increasing).

We observe a reasonable behaviour of all rearranged estimates, which improves with increasing sample size. In general, there are only minor differences between the rearranged estimates $\hat{R}_\rho$, $\hat{R}_\tau$ and the non-rearranged estimates $\hat{\rho}$, $\hat{\tau}$, which are mainly caused by a slightly smaller bias of the non-rearranged estimates. For the Gaussian copula with correlation 0.25, the rearranged estimates $\hat{R}_\rho$ and $\hat{R}_\tau$ slightly overestimate their population version R_ρ and R_τ if the sample size is $n = 50$ or 100. For all other scenarios, we observe an underestimation.

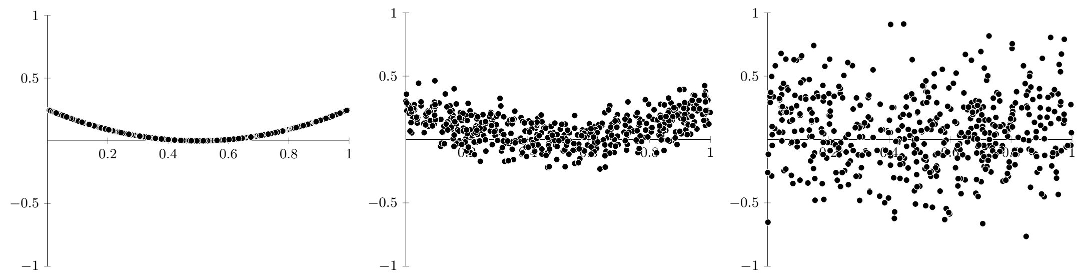

Fig 2: *Scatter plots of a sample of $n = 500$ observations from model (3.15). Left panel: $\sigma = 0$; middle panel: $\sigma = 0.1$; right panel: $\sigma = 0.3$.*

σ	n	Spearman's ρ R_ρ	$\hat{R}_\rho$	$\lvert\hat{\rho}\rvert$	Kendall's τ R_τ	$\hat{R}_\tau$	$\lvert\hat{\tau}\rvert$
0	50	1	0.918 (0.012)	0.155 (0.120)	1	0.732 (0.021)	0.133 (0.102)
	100		0.960 (0.005)	0.105 (0.078)		0.810 (0.013)	0.092 (0.069)
	500		0.992 (0.000)	0.048 (0.036)		0.914 (0.002)	0.041 (0.031)
	1000		0.996 (0.000)	0.032 (0.025)		0.939 (0.001)	0.028 (0.022)
0.1	50	0.580	0.530 (0.116)	0.131 (0.094)	0.404	0.362 (0.085)	0.092 (0.066)
	100		0.550 (0.081)	0.091 (0.068)		0.378 (0.060)	0.063 (0.047)
	500		0.553 (0.035)	0.042 (0.031)		0.381 (0.026)	0.029 (0.021)
	1000		0.559 (0.024)	0.030 (0.022)		0.386 (0.018)	0.020 (0.015)
0.3	50	0.232	0.255 (0.143)	0.113 (0.085)	0.155	0.171 (0.096)	0.078 (0.058)
	100		0.258 (0.098)	0.081 (0.059)		0.173 (0.066)	0.055 (0.040)
	500		0.216 (0.047)	0.037 (0.027)		0.145 (0.032)	0.025 (0.018)
	1000		0.217 (0.033)	0.026 (0.020)		0.146 (0.022)	0.017 (0.013)

TABLE 2

Simulated mean and standard deviation of the rearranged Spearman's ρ estimate $\hat{R}_\rho$, the (absolute) Spearman's rank correlation coefficient $|\hat{\rho}|$ (left part), the rearranged Kendall's τ estimate $\hat{R}_\tau$, (absolute) Kendall's rank correlation coefficient $|\hat{\tau}|$ (right part). The distribution of (X, Y) is given by model (3.15).

4.1.2. *A family of non-stochastically monotone distributions.* In this section, we consider the more common situation where $R_\mu \neq \mu$. To generate data from a family of 2-dimensional distributions with different degrees of dependence, let $X \sim U(0,1)$ denote a uniformly (on the interval $[0,1]$) distributed random variable and $Z \sim \mathcal{N}(0,1)$ a standard normal distributed random variable such that X and Z are independent. We consider the regression model

$$Y := (X - 1/2)^2 + \sigma Z \,, \tag{3.15}$$

where σ is a non-negative constant. A similar model has been studied in Chatterjee (2021) and (3.15) contains perfect functional dependence of X and Y (for $\sigma = 0$) and independence in the limit for $\sigma \to \infty$. The corresponding scatter plots from $n = 500$ independent observations according to model (3.15) with $\sigma = 0$, 0.1 and 0.3 are displayed in Figure 2, while Table 2 shows the simulated mean and standard deviation of the estimates $\hat{R}_\rho$ (for the rearranged Spearman's ρ) and $\hat{R}_\tau$ (for the rearranged Kendall's τ). For $\sigma > 0$ the "true" values of R_ρ and R_τ have been obtained by simulation using a sample of size $n = 1000000$ and bandwidths $N_1 = N_2 = \lfloor n^{0.45} \rfloor$. The empirical results confirm the consistency statement in Theorem 3.4. In the table, we also display the simulated mean of the non-rearranged estimators $|\hat{\rho}|$ and $|\hat{\tau}|$, which do not yield reasonable results.

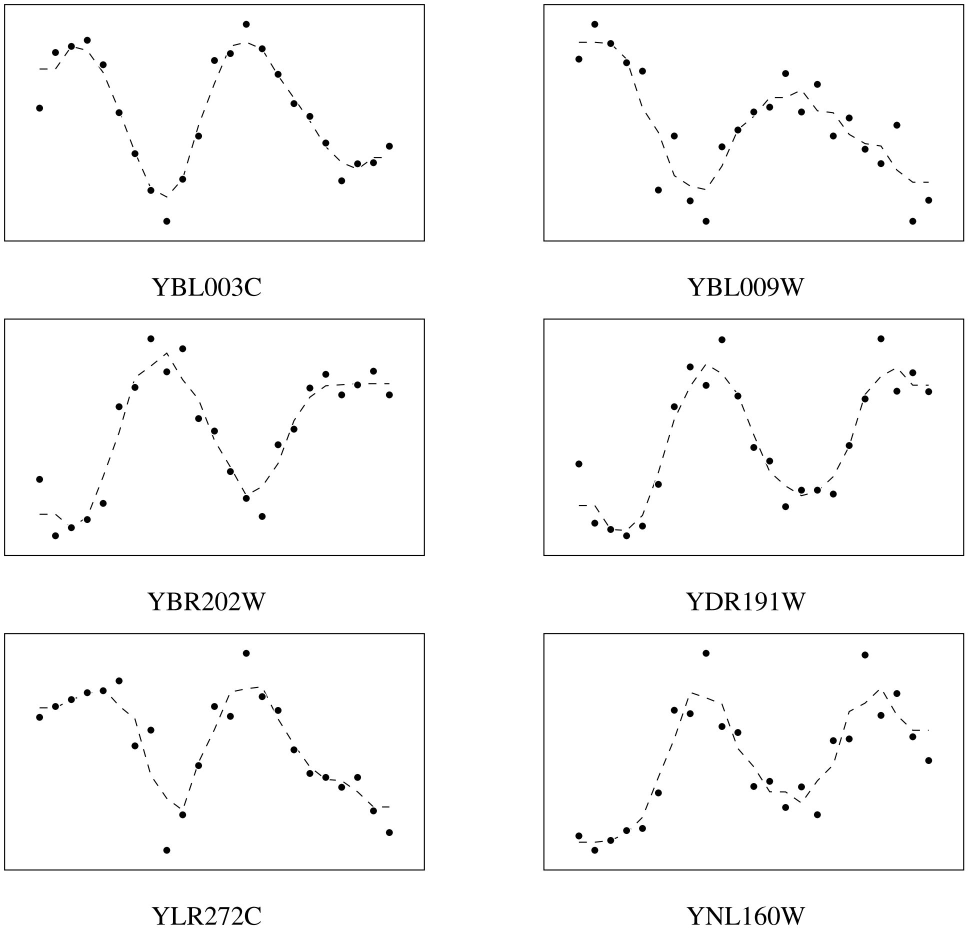

Fig 3: *Transcript levels of the top genes, which were selected by the FDR procedure based on the rearranged Spearman's rank correlation coefficient, but not by the FDR procedure based on Spearman's rank correlation coefficient. (*$\alpha = 0.05$*). The dashed lines represent the* 3*-nearest neighbour regression estimates.*

4.2. *Data example.* In this section we briefly revisit a data example which was investigated by Chatterjee (2021) to study the performance of his correlation coefficient in the analysis of yeast gene expression data. The data consists of the expressions of 6223 yeast genes and was originally analyzed by Spellman et al. (1998) who tried to identify genes whose transcript levels oscillate during the cell cycle. For each gene, the gene expression was observed at 23 time points. Because the number of genes is large, visual inspection is not possible and Reshef et al. (2011) proposed to use the MIC and MINE correlation coefficient to analyze the data. Chatterjee (2021) compared the performance of his correlation coefficient with these measures and demonstrated some advantages of his approach. We will now provide a brief illustration analyzing this type of data with a rearranged dependence measure to demonstrate the ability of our approach to also detect non-monotone dependencies. We begin with an analysis of the rearranged Spearman's rank coefficient $\hat{R}_\rho$. After that, we provide a very brief comparison of $\hat{R}_\rho$ with Chatterjee's correlation coefficient.

To be precise, we consider the curated data set (available through the R-package "minerva") of 4381 genes. For each gene, we perform a permutation test based on Spearman's

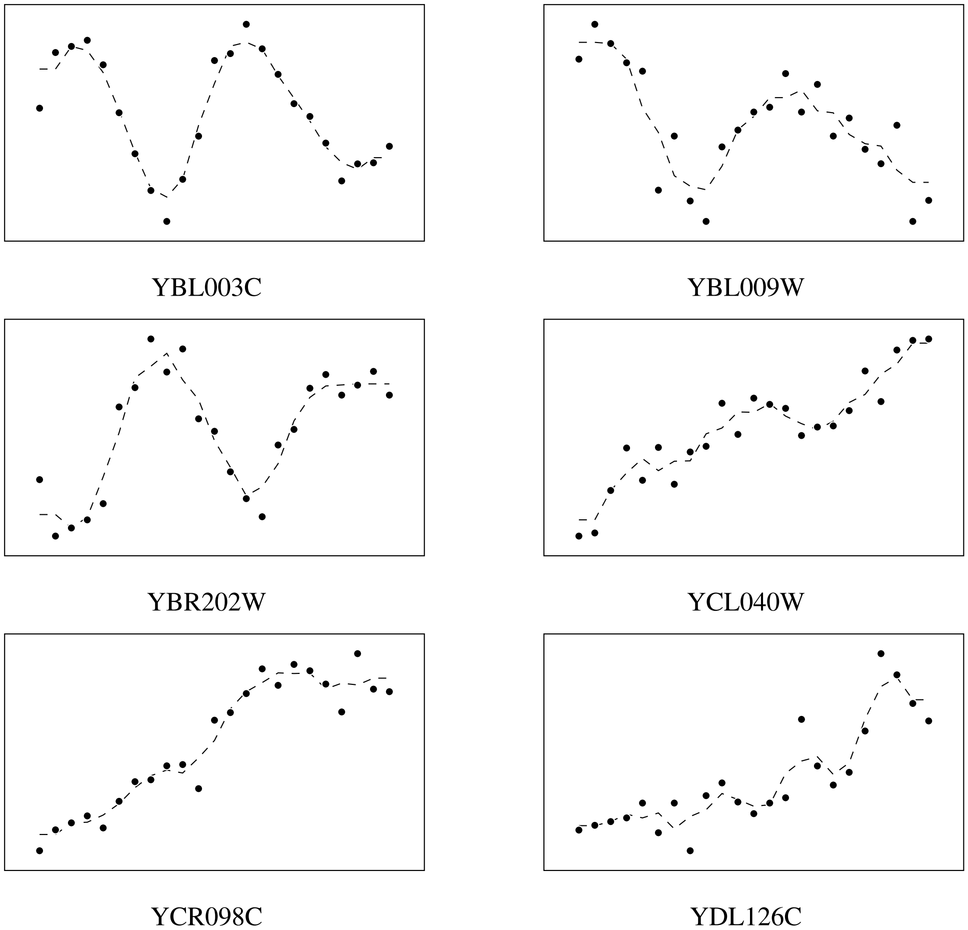

Fig 4: *Transcript levels of genes, which were selected by the rearranged Spearman's rank correlation coefficient. The figure shows the* 6 *top genes with the smallest* p*-values. The dashed lines represent the* 3*-nearest neighbour regression estimates.*

rank correlation for the hypotheses

$$H_0 : \rho = 0 \text{ versus } H_1 : \rho > 0$$

and a permutation test based on the statistic $\hat{R}_\rho$ for the hypotheses

$$H_0 : R_\rho = 0 \text{ versus } H_1 : R_\rho > 0 \, , \tag{3.16}$$

where we use 10000 permutations. The corresponding p-values are used to identify the significant genes using the Benjamini–Hochberg FDR procedure with a false discovery rate of 0.05 (see, Benjamini and Hochberg, 1995). To concentrate on non-monotone dependencies, we exclude from those genes selected by the FDR procedure based on the rearranged Spearman's rank correlation all genes which are also detected by Spearman's rank correlation. This results in 84 remaining genes. In Figure 3 we display the transcript levels of the top 6 genes with the smallest p-values from the remaining data. We observe that the FDR procedure based on the rearranged Spearman's rank correlation identifies additional dependencies, which are oscillating and are not found if the analysis is based on Spearman's rank correlation. A similar observation was made by Chatterjee (2021) for his rank correlation coefficient, who used 4 alternative tests to exclude genes with a monotone behaviour (a gene was excluded, whenever one of these tests identified it as significant). Because both procedures are based on

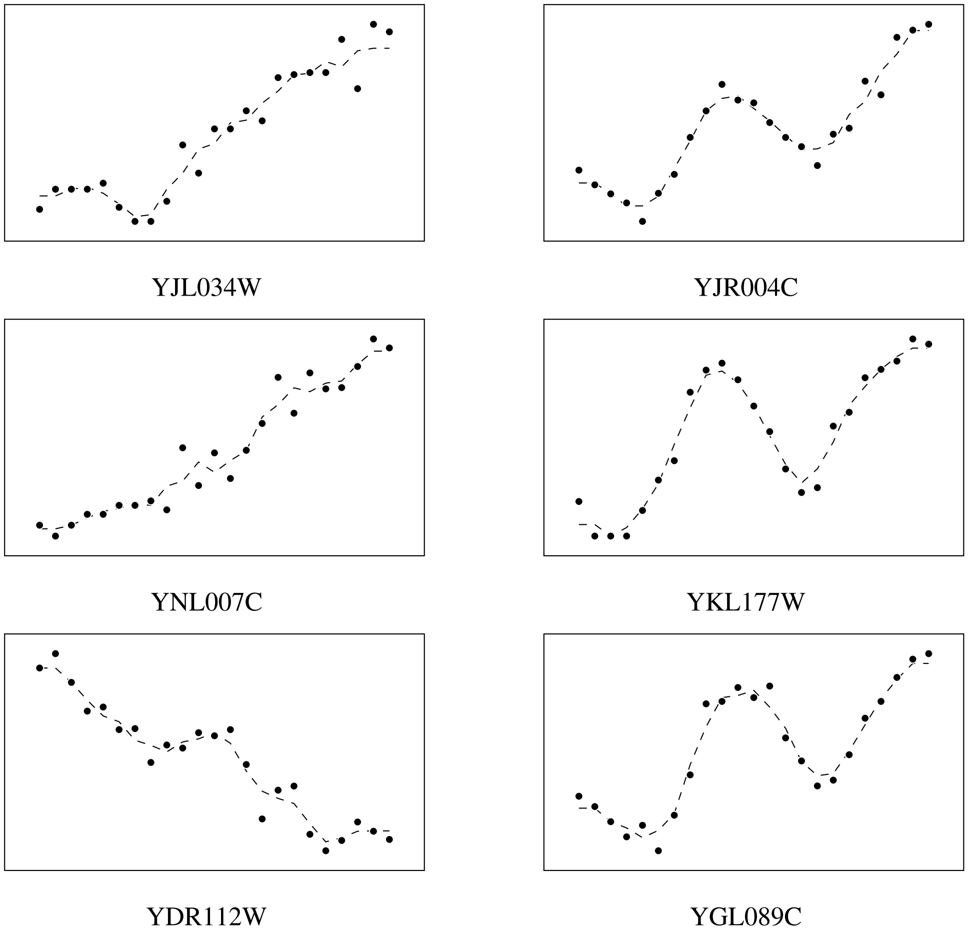

Fig 5: *Transcript levels of genes, which were selected by the Chatterjee's correlation coefficient. The figure shows the* 6 *genes with the smallest* p*-values. The dashed lines represent the* 3*-nearest neighbour regression estimates.*

different dependence measures the finally identified 6 top genes do not necessarily coincide (only the gene YBL003C was selected by our and Chatterjee's procedure). However, all 6 top genes found by Chatterjee (2021) are also selected by the FDR procedure based on rearranged Spearman's rank correlation and vice versa. Moreover, the qualitative conclusion from both methods is same. Both methods are able to identify non-monotone (in the concrete example oscillating) associations.

We conclude with a brief comparison of the FDR procedures based on the rearranged Spearman's and Chatterjee's rank correlation coefficient, if they are used without sorting out monotone dependencies by preliminary analysis. In Figures 4 and 5, we display the transcript levels of the 6 genes with the smallest p-values after running the FDR procedure based on the two dependency measures. We observe again that both methods are able to identify non-monotone associations. Interestingly the top three genes identified by the rearranged Spearman's rank correlation with the smallest three p-values exhibit an oscillating transcript level while it looks more monotone for the next three genes. For the FDR procedure based on Chatterjee's rank correlation the picture is not so clear.

Acknowledgements. C. Strothmann gratefully acknowledges financial support from the German Academic Scholarship Foundation. The work of H. Dette was supported by the DFG Research unit 5381 *Mathematical Statistics in the Information Age*.

REFERENCES

ANEVSKI, D. and FOUGÈRES, A.-L. (2019). Limit properties of the rearrangement for density and regression function estimation. *Bernoulli* **25** 549 – 583. MR3892329

ANSARI, J. and RÜSCHENDORF, L. (2021). Sklar's theorem, copula products, and ordering results in factor models. *Depend. Model.* **9** 267–306. MR4327840

AUDDY, A., DEB, N. and NANDY, S. (2021). Exact Detection Thresholds for Chatterjee's Correlation. *https://arxiv.org/abs/2104.15140*.

BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. Roy. Statist. Soc. Ser. B* **57** 289-300. MR1325392

BENNETT, C. and SHARPLEY, R. C. (1988). *Interpolation of Operators*. Academic Press, Boston. MR0928802

BERGSMA, W. and DASSIOS, A. (2014). A consistent test of independence based on a sign covariance related to Kendall's tau. *Bernoulli* **20** 1006 – 1028. MR3178526

BLUM, J. R., KIEFER, J. and ROSENBLATT, M. (1961). Distribution Free Tests of Independence Based on the Sample Distribution Function. *Ann. Math. Statist.* **32** 485 – 498. MR0125690

CAMIRAND-LEMYRE, F., CARROLL, R. J. and DELAIGLE, A. (2022). Semiparametric Estimation of the Distribution of Episodically Consumed Foods Measured with Error. *To appear in: J. Amer. Statist. Assoc.*

CAO, S. and BICKEL, P. J. (2020). Correlations with tailored extremal properties. *http://arxiv.org/abs/2008.10177*.

CHATTERJEE, S. (2021). A New Coefficient of Correlation. *J. Amer. Statist. Assoc.* **116** 2009-2022. MR4353729

CHEN, S. X. and HUANG, T. M. (2007). Nonparametric Estimation of Copula Functions for Dependence Modelling. *Canad. J. Statist.* **35** 265–282. MR2393609

CHERNOZHUKOV, V., FERNÁNDEZ-VAL, I. and GALICHON, A. (2009). Improving point and interval estimators of monotone functions by rearrangement. *Biometrika* **96** 559-575. MR2538757

CHERNOZHUKOV, V., FERNÁNDEZ-VAL, I. and GALICHON, A. (2010). Quantile and Probability Curves Without Crossing. *Econometrica* **78** 1093-1125. MR2667913

CHONG, K. M. (1974). Some extensions of a theorem of Hardy, Littlewood and Pólya and their applications. *Canadian J. Math.* **26** 1321–1340. MR0352377

CHONG, K. M. and RICE, N. M. (1971). *Equimeasurable Rearrangements of Functions*. Queen's University, Kingston. MR0372140

COVER, T. M. and THOMAS, J. A. (2006). *Elements of Information Theory*, 2nd ed. Wiley-Interscience, Hoboken. MR2239987

CSÖRGŐ, S. (1985). Testing for independence by the empirical characteristic function. *J. Multivariate Anal.* **16** 290-299. MR0793494

DARSOW, W. F., NGUYEN, B. and OLSEN, E. T. (1992). Copulas and Markov processes. *Illinois J. Math.* **36** 600–642. MR1215798

DAY, P. W. (1972). Rearrangement inequalities. *Canadian J. Math.* **24** 930–943. MR0310156

DAY, P. W. (1973). Decreasing rearrangements and doubly stochastic operators. *Trans. Amer. Math. Soc.* **178** 383–383. MR0318962

DEB, N., GHOSAL, P. and SEN, B. (2020). Measuring association on topological spaces using kernels and geometric graphs. *http://arxiv.org/abs/2010.01768*.

DETTE, H., NEUMEYER, N. and PILZ, K. F. (2006). A simple nonparametric estimator of a strictly monotone regression function. *Bernoulli* **12** 469-490. MR2232727

DETTE, H., SIBURG, K. F. and STOIMENOV, P. A. (2012). A copula-based non-parametric measure of regression dependence. *Scand. J. Stat.* **40** 21–41. MR3024030

DETTE, H. and VOLGUSHEV, S. (2008). Non-crossing non-parametric estimates of quantile curves. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **70** 609-627. MR2420417

DETTE, H. and WU, W. (2019). Detecting relevant changes in the mean of nonstationary processes, a mass excess approach. *Ann. Statist.* **47** 3578 – 3608. MR4025752

DURANTE, F. and PAPINI, P. L. (2009). Componentwise concave copulas and their asymmetry. *Kybernetika (Prague)* **45** 1003–1011. MR2650079

DURANTE, F. and SEMPI, C. (2016). *Principles of Copula Theory*. CRC Press, Boca Raton. MR3443023

FERMANIAN, J. D., RADULOVIĆ, D. and WEGKAMP, M. (2004). Weak convergence of empirical copula processes. *Bernoulli* **10** 847 – 860. MR2093613

FUCHS, S., MCCORD, Y. and SCHMIDT, K. D. (2018). Characterizations of copulas attaining the bounds of multivariate Kendall's Tau. *J. Optim. Theory Appl.* **178** 424–438. MR3825632

GAMBOA, F., GREMAUD, P., KLEIN, T. and LAGNOUX, A. (2020). Global sensitivity analysis: A new generation of might estimators based on rank statistics. *http://arxiv.org/abs/2003.01772*.

GEENENS, G. and LAFAYE DE MICHEAUX, P. (2020). The Hellinger Correlation. *J. Amer. Statist. Assoc.* 1–15.

GENEST, C., NEŠLEHOVÀ, J. G. and RÈMILLARD, B. (2017). Asymptotic behavior of the empirical multilinear copula process under broad conditions. *J. Multivariate Anal.* **159** 82-110. MR3668549

GRETTON, A., FUKUMIZU, K., TEO, C., SONG, L., SCHÖLKOPF, B. and SMOLA, A. (2008). A Kernel Statistical Test of Independence. In *Advances in Neural Information Processing Systems* (J. PLATT, D. KOLLER, Y. SINGER and S. ROWEIS, eds.) **20**. Curran Associates, Inc.

GRIESSENBERGER, F., JUNKER, R. R., PETZEL, V. and TRUTSCHNIG, W. (2021). qad: Quantification of asymmetric dependence. R package version 1.0.0 available at https://CRAN.R-project.org/package=qad.

HARDY, G. H., LITTLEWOOD, J. E. and PÓLYA, G. (1988). *Inequalities*. *Cambridge Mathematical Library*. Cambridge University Press, Cambridge. Reprint of the 1952 edition. MR944909

HOFERT, M., KOJADINOVIC, I., MÄCHLER, M. and YAN, J. (2020). copula: Multivariate dependence with copulas. R package version 1.0-1 available at https://CRAN.R-project.org/package=copula.

JUNKER, R. R., GRIESSENBERGER, F. and TRUTSCHNIG, W. (2021). Estimating scale-invariant directed dependence of bivariate distributions. *Comput. Statist. Data Anal.* **153** 107058. MR4141460

KINNEY, J. B. and ATWAL, G. S. (2014). Equitability, mutual information, and the maximal information coefficient. *Proc. Natl. Acad. Sci. USA* **111** 3354–3359. MR3200177

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York. MR0107933

LI, X., MIKUSIŃSKI, P. and TAYLOR, M. D. (1998). Strong Approximation of Copulas. *J. Math. Anal. Appl.* **225** 608-623. MR1644300

LI, X., MIKUSIŃSKI, P., SHERWOOD, H. and TAYLOR, M. D. (1997). On approximation of copulas. In *Distributions with given Marginals and Moment Problems* (V. Beneš and J. Štěpán, eds.) 107–116. Springer, Dordrecht. MR1614663

LIN, Z. and HAN, F. (2021). On boosting the power of Chatterjee's rank correlation. *http://arxiv.org/abs/2108.06828*.

MARSHALL, A. W., OLKIN, I. and ARNOLD, B. C. (2011). *Inequalities: Theory of Majorization and Its Applications*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2759813

MIKUSIŃSKI, P., SHERWOOD, H. and TAYLOR, M. (1992). Shuffles of min. *Stochastica* **13** 61–74. MR1197328

NELSEN, R. B. (2006). *An Introduction to Copulas*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2197664

OLSEN, E. T., DARSOW, W. F. and NGUYEN, B. (1996). Copulas and Markov operators. In *Distributions with fixed marginals and related topics* 244–259. Institute of Mathematical Statistics, Hayward. MR1485536

OMELKA, M., GIJBELS, I. and VERAVERBEKE, N. (2009). Improved kernel estimation of copulas: Weak convergence and goodness-of-fit testing. *Ann. Statist.* **37** 3023 – 3058. MR2541454

RESHEF, D. N., RESHEF, Y. A., FINUCANE, H. K., GROSSMAN, S. R., MCVEAN, G., TURNBAUGH, P. J., LANDER, E. S., MITZENMACHER, M. and SABETI, P. C. (2011). Detecting Novel Associations in Large Data Sets. *Science* **334** 1518-1524.

ROSENBLATT, M. (1975). A Quadratic Measure of Deviation of Two-Dimensional Density Estimates and A Test of Independence. *Ann. Statist.* **3** 1 – 14. MR0428579

RYFF, J. V. (1965). Orbits of L^1-functions under doubly stochastic transformations. *Trans. Amer. Math. Soc.* **117** 92–100. MR0209866

RYFF, J. V. (1970). Measure preserving transformations and rearrangements. *J. Math. Anal. Appl.* **31** 449–458. MR0419734

SCHWEIZER, B. and WOLFF, E. F. (1981). On nonparametric measures of dependence for random variables. *Ann. Statist.* **9** 879–885. MR0619291

SHI, H., DRTON, M. and HAN, F. (2021a). On the power of Chatterjee's rank correlation. *To appear in: Biometrika*.

SHI, H., DRTON, M. and HAN, F. (2021b). On Azadkia-Chatterjee's conditional dependence coefficient. *http://arxiv.org/abs/2108.06827*.

SIBURG, K. F. and STROTHMANN, C. (2021). Stochastic monotonicity and the Markov product for copulas. *J. Math. Anal. Appl.* **503** 125348. MR4263102

SPELLMAN, P. T., GAVIN, S., ZHANG, M. Q., IYER, V. R., ANDERS, K., EISEN, M. B., BROWN, P. O., BOTSTEIN, D. and FUTCHER, B. (1998). Comprehensive Identification of Cell Cycle-regulated Genes of the Yeast Saccharomyces cerevisiae by Microarray Hybridization. *Mol. Biol. Cell* **9** 3273-3297.

SZÉKELY, G. J., RIZZO, M. L. and BAKIROV, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35** 2769 – 2794. MR2382665

R CORE TEAM (2021). R: A Language and environment for statistical computing R Foundation for Statistical Computing, Vienna.

TRUTSCHNIG, W. (2011). On a strong metric on the space of copulas and its induced dependence measure. *J. Math. Anal. Appl.* **384** 690–705. MR2825218

ZHANG, K. (2019). BET on Independence. *J. Amer. Statist. Assoc.* **114** 1620-1637. MR4047288

APPENDIX A: PRELIMINARIES

In this section, we present some basic facts about copulas and monotone rearrangements, which will be frequently used throughout the proofs of our results in Appendix B and C. We start with the definition of a copula, which is a bivariate distribution function on the unit square with uniform univariate margins.

DEFINITION A.1. A function $C:[0,1]^2\to[0,1]$ is called a copula if

1. C is grounded, i.e. $C(0,v)=C(u,0)=0$ for all $u,v\in[0,1]$
2. C has uniform margins, i.e. $C(1,u)=C(u,1)=u$ for all $u\in[0,1]$
3. C is 2-increasing, i.e. the C-volume of every rectangle $R=[u_1,u_2)\times[v_1,v_2)$ is nonnegative:

$$V_C(R):=C(u_2,v_2)-C(u_1,v_2)-C(u_2,v_1)+C(u_1,v_1)\geq 0\,.$$

The set of all copulas is denoted by $\mathcal{C}$. We refer to the lower Fréchet-Hoeffding bound by $C^-(u,v):=\max\{u+v-1,0\}$, to the independence (or product) copula by $\Pi(u,v):=uv$, and to the upper Fréchet-Hoeffding bound by $C^+(u,v):=\min\{u,v\}$. Any copula C satisfies $C^-\leq C\leq C^+$.

DEFINITION A.2. The Markov product of two copulas C and D is defined as the copula

$$(C*D)(u,v):=\int_0^1 \partial_2 C(u,t)\partial_1 D(t,v)\,\mathrm{d}t\,.$$

A comprehensive review of the Markov product can be found in Durante and Sempi (2016).

DEFINITION A.3. A linear operator $T:L^1([0,1])\to L^1([0,1])$ is called a Markov operator if

1. T is positive, i.e. $Tf\geq 0$ whenever $f\geq 0$
2. $T\mathbb{1}_{[0,1]}=\mathbb{1}_{[0,1]}$
3. T preserves the integral, i.e. $\int_0^1 Tf(t)\,\mathrm{d}t=\int_0^1 f(t)\,\mathrm{d}t$ for all $f\in L^1([0,1])$.

The following result shows that copulas and Markov operators are closely linked and that the composition of Markov operators corresponds to the Markov product of copulas. A proof can be found in Olsen, Darsow and Nguyen (1996).

THEOREM A.4. *Let C be a copula and T be a Markov operator. Then*

$$C_T(u,v):=\int_0^u T\mathbb{1}_{[0,v]}(t)\,\mathrm{d}t \;\text{ and }\; T_Cf(u):=\partial_u\int_0^1 \partial_2 C(u,v)f(v)\,\mathrm{d}v$$

define a copula C_T and a Markov operator T_C, respectively. The correspondence $C\mapsto T_C$ is bijective with $T_{C_T}=T$ and $C_{T_C}=C$. Moreover,

$$T_{C_1*C_2}=T_{C_1}\circ T_{C_2}$$

holds for all copulas C_1 and C_2.

The following definition of a concordance measure is adapted from Durante and Sempi (2016).

DEFINITION A.5. A function $\kappa : \mathcal{C} \to [-1,1]$ is called a measure of concordance if

1. $\kappa(C^-) = -1, \kappa(\Pi) = 0$ and $\kappa(C^+) = 1$
2. $\kappa(C^\top) = \kappa(C)$, where $C^\top(u,v) := C(v,u)$
3. $\kappa(C^- * C) = \kappa(C * C^-) = -\kappa(C)$
4. κ is monotone w.r.t. the pointwise order on the set of copulas
5. κ is continuous w.r.t. the pointwise[6] convergence of copulas.

For the decreasing rearrangement $f^* : [0,1] \to \mathbb{R}$ of a measurable function $f : [0,1] \to \mathbb{R}$ from (2.2), we state the following properties.

PROPOSITION A.6. *For any two measurable functions f and g, the following assertions hold:*

1. *f^* is decreasing and right-continuous on $[0,1]$.*
2. *$f \leq g$ implies $f^* \leq g^*$.*
3. *There exists a λ-preserving transformation $\sigma : [0,1] \to [0,1]$ such that $f = f^* \circ \sigma$.*
4. *The decreasing rearrangement is L^p-invariant for $1 \leq p \leq \infty$, i.e.*

$$\|f\|_p = \|f^*\|_p \ .$$

PROOF. Property (1) is stated in Theorem 4.2, properties (2) and (4) can be found in Proposition 4.3, and property (3) is stated in Theorem 6.2 of Chong and Rice (1971). □

Closely linked to the decreasing rearrangement of measurable functions is an ordering widely known as the majorization order, introduced by Hardy, Littlewood and Pólya for vectors, and by Ryff (1965) for functions.

DEFINITION A.7. Suppose $f, g \in L^1([0,1])$. Then f is majorized by g, denoted by $f \preceq g$, if

$$\int_0^t f^*(s)\, \mathrm{d}s \leq \int_0^t g^*(s)\, \mathrm{d}s$$

holds for all $t \in [0,1]$, as well as

$$\int_0^1 f^*(s)\, \mathrm{d}s = \int_0^1 g^*(s)\, \mathrm{d}s \ .$$

THEOREM A.8. *For $f, g \in L^1([0,1])$, the following statements are equivalent:*

1. *f is majorized by g, i.e. $f \preceq g$.*
2. *For every convex function $\phi : \mathbb{R} \to \mathbb{R}$ we have*

$$\int_0^1 \phi(f(s))\, \mathrm{d}s \leq \int_0^1 \phi(g(s))\, \mathrm{d}s \ .$$

[6] As copulas are continuous function on a compact set, pointwise and uniform convergence are equivalent.

3. *There exists a Markov operator T such that $f = Tg$.*

Furthermore, the following inequalities hold:

(a)

$$\int_0^1 |f^*(s)g^*(1-s)| \, \mathrm{d}s \leq \int_0^1 |f(s)g(s)| \, \mathrm{d}s \leq \int_0^1 |f^*(s)g^*(s)| \, \mathrm{d}s \, .$$

(b)

$$f^* - g^* \preceq f - g \, .$$

PROOF. The equivalence of (1) and (3) is shown in (Day, 1973, Thm. 4.9), while that of (1) and (2) is contained in (Chong, 1974, Thm. 2.5). The proofs of (a), called the Hardy-Littlewood inequality, and (b) can be found in (Day, 1972, (6.2) and (6.1)). □

APPENDIX B: PROOFS OF THE RESULTS IN SECTION 2

B.1. Proof of Theorem 2.3. In order to show that the stochastically increasing rearrangement, $C^{\uparrow}$ is a copula, we verify the properties (1) to (3) of Definition A.1.

1. It follows from $(\partial_1 C)^*(u,0) = 0^* = 0$ that $C^{\uparrow}(u,0) = 0$. The identity $C^{\uparrow}(0,v) = 0$ is trivial by Definition 2.2.
2. By definition, we have

$$C^{\uparrow}(u,1) = \int_0^u (\partial_1 C)^*(s,1) \, \mathrm{d}s = \int_0^u 1^* \, \mathrm{d}s = u \, .$$

   In view of Proposition A.6(3), we further obtain that

$$C^{\uparrow}(1,v) = \int_0^1 (\partial_1 C)^*(s,v) \, \mathrm{d}s = \int_0^1 (\partial_1 C)^*(\sigma_v(s),v) \, \mathrm{d}s = \int_0^1 \partial_1 C(t,v) \, \mathrm{d}t = v \, .$$

3. From Definition A.1(3) we see that $0 \leq \partial_1 C(\cdot, v_1) \leq \partial_1 C(\cdot, v_2)$ whenever $v_1 \leq v_2$. Combining this with Proposition A.6(2) yields $(\partial_1 C)^*(\cdot, v_1) \leq (\partial_1 C)^*(\cdot, v_2)$. Thus, the $C^{\uparrow}$-volume of a rectangle $[u_1,u_2) \times [v_1,v_2)$ satisfies

$$\begin{aligned} V_{C^{\uparrow}}([u_1,u_2) \times [v_1,v_2)) &= C^{\uparrow}(u_2,v_2) - C^{\uparrow}(u_1,v_2) - C^{\uparrow}(u_2,v_1) + C^{\uparrow}(u_1,v_1) \\ &= \int_{u_1}^{u_2} (\partial_1 C)^*(s,v_2) - (\partial_1 C)^*(s,v_1) \, \mathrm{d}s \geq 0. \end{aligned}$$

Finally, we show that C is stochastically increasing if and only if $C = C^{\uparrow}$. If $C = C^{\uparrow}$, of course, C is stochastically increasing because $C^{\uparrow}$ is. Conversely, suppose C is stochastically increasing, i.e., each $u \mapsto C(u,v)$ is concave. Then the right-hand derivative $u \mapsto \partial_1^+ C(u,v)$ is a decreasing and right-continuous function, and (Chong and Rice, 1971, Thm. 4.2) guarantees that $\partial_1^+ C(u,v) = (\partial_1 C)^*(u,v)$. This implies

$$C(u,v) = \int_0^u \partial_1^+ C(t,v) \, \mathrm{d}t = \int_0^u (\partial_1 C)^*(t,v) \, \mathrm{d}t = C^{\uparrow}(u,v) \, .$$

B.2. Proof of Theorem 2.4. We will require a preliminary result. For this, we first note that the so-called (SD)-rearrangement of C defined by

$$C^{\downarrow}(u,v) := \int_0^u (\partial_1 C)^*(1-s,v)\,\mathrm{d}s = v - C^{\uparrow}(1-u,v) = (C^- * C^{\uparrow})(u,v)$$

is a stochastically decreasing copula.

LEMMA B.1. *For any copula C, we have*

$$C^{\downarrow}(u,v) \le C(u,v) \le C^{\uparrow}(u,v)\ .$$

PROOF. By Theorem A.8(a) we obtain the upper estimate

$$C(u,v) = \int_0^1 \mathbb{1}_{[0,u]}(t)\partial_1 C(t,v)\,\mathrm{d}t \le \int_0^1 \mathbb{1}_{[0,u]}(t)(\partial_1 C)^*(t,v)\,\mathrm{d}t = C^{\uparrow}(u,v)\ .$$

The lower estimate follows analogously. □

We will now prove properties (1.1)–(1.3) for $R_\mu(C) = \mu(C^{\uparrow})$. For this, we say that the copula C is *completely dependent* if there exists a measurable function f such that $V = f(U)$. It is proven in Darsow, Nguyen and Olsen (1992) that C is completely dependent if, and only if,

$$\partial_1 C(u,v) \in \{0,1\} \tag{3.17}$$

for almost all $u \in [0,1]$ and all $v \in [0,1]$.

(1.1) Since μ only takes values between 0 and 1, we obtain the first assertion.
(1.2) If $C = \Pi$, we have $\mu(C^{\uparrow}) = \mu(\Pi^{\uparrow}) = \mu(\Pi) = 0$. If, on the other hand, $\mu(C^{\uparrow}) = 0$, we conclude $C^{\uparrow} = \Pi$ by the properties of μ. But then $C^{\downarrow} = C^- * \Pi = \Pi$, and Lemma B.1 yields $\Pi = C^{\downarrow} \le C \le C^{\uparrow} = \Pi$, hence $C = \Pi$.
(1.3) If C is completely dependent, then $C^{\uparrow} = C^+$ and $\mu(C^{\uparrow}) = \mu(C^+) = 1$ by definition. On the other hand, $\mu(C^{\uparrow}) = 1$ implies $C^{\uparrow} = C^+$ by the properties of μ. Thus, $\partial_1 C(u,v) = (\partial_1 C)^*(\sigma_v(u),v) \in \{0,1\}$, so C is completely dependent by (3.17).

B.3. Proof of Equation (2.5). The statement is an immediate consequence of the fact that the decreasing rearrangement of $g_v(u) := \partial_1 C(u,v) - v$ is $g_v^*(u) = \partial_1 C^{\uparrow}(u,v) - v$. As the decreasing rearrangement leaves all L^p-norms invariant, we conclude

$$\int_0^1 |\partial_1 C(u,v) - v|^p\ \mathrm{d}u = \int_0^1 \left|\partial_1 C^{\uparrow}(u,v) - v\right|^p\ \mathrm{d}u\ .$$

Integrating with respect to v yields the desired result with $p = 1,2$.

B.4. Proof of the statements in Example 2.6. In this section we show that the Schweizer-Wolff measure σ_p in (2.6) for $1 \le p < \infty$ satisfies the properties (1.1) to (1.3) on the set $\mathcal{C}^{\uparrow}$.

(1.1) σ_p takes values only between 0 and 1, since $C^{\uparrow}$ is stochastically increasing and fulfils

$$0 \le C^{\uparrow} - \Pi \le C^+ - \Pi\ .$$

(1.2) $\sigma_p(C) = 0$ holds if and only if $C = \Pi$.
(1.3) Suppose $C = C^{\uparrow}$ is completely dependent. Then $\partial_1 C^{\uparrow}(u,v) \in \{0,1\}$ by (3.17) and $\partial_1 C^{\uparrow}(u,v) = \mathbb{1}_{[0,v]}(u)$ by Definition A.1(2). Thus, $C^{\uparrow} = C^+$ which yields $\sigma_p(C) = 1$. On the other hand, if C is not completely dependent, then an analogous argument shows that $C^{\uparrow} < C^+$ on a set of positive measure such that

$$\sigma_p(C) = \frac{\left\|C^{\uparrow} - \Pi\right\|_p}{\left\|C^+ - \Pi\right\|_p} < \frac{\left\|C^+ - \Pi\right\|_p}{\left\|C^+ - \Pi\right\|_p} = 1 \,.$$

B.5. Proof of the statements in Example 2.7. We introduce the concordance functional

$$Q(C_1, C_2) := 4 \int\limits_{[0,1]^2} C_1(u,v)\, \mathrm{d}C_2(u,v) - 1$$

and point out for later reference that Q is symmetric and fulfils

$$Q(C_1, C_2) \le Q(C_1', C_2) \tag{3.18}$$

whenever $C_1 \le C_1'$.

Then the four measures of concordance (see Definition A.5) Spearman's ρ, Kendall's τ, Gini's γ and Blomqvist's β are given by (see, e.g., Chapter 5 in Nelsen (2006))

$$\rho(C) = 3Q(C,\Pi) = 12 \int\limits_{[0,1]^2} C(u,v)\, \mathrm{d}\lambda(u,v) - 3$$

$$\tau(C) = Q(C,C) = 4 \int\limits_{[0,1]^2} C(u,v)\, \mathrm{d}C(u,v) - 1$$

$$\gamma(C) = Q(C,C^-) + Q(C,C^+) = 2 \int\limits_{[0,1]^2} |u+v-1| - |u-v|\, \mathrm{d}C(u,v)$$

$$\beta(C) = 4C\left(\frac{1}{2},\frac{1}{2}\right) - 1 \,.$$

First of all, β does not satisfy (1.3) on $\mathcal{C}^{\uparrow}$ because the copula[7]

$$C(u,v) = \begin{cases} 2\Pi(u,v) & \text{if } (u,v) \in [0,1/2]^2 \\ C^+ & \text{else} \end{cases}$$

is stochastically increasing with $C \neq C^+$, yet $\beta(C) = 4C(1/2,1/2) - 1 = 1 = \beta(C^+)$.

We now show that ρ, τ and γ all satisfy the properties (1.1)–(1.3) on $\mathcal{C}^{\uparrow}$. Since any concave function $f : [0,1] \to [0,v]$ with $f(0) = 0$ and $f(1) = v$ satisfies $f(u) \ge uv = \Pi(u,v)$, any stochastically increasing copula C satisfies

$$\Pi \le C = C^{\uparrow} \le C^+ \,. \tag{3.19}$$

Hence we conclude from Definition A.5(4) that $0 = \kappa(\Pi) \le \kappa(C^{\uparrow}) \le \kappa(C^+) = 1$. It remains to verify properties (1.2) and (1.3) for ρ, τ and γ.

First, we look at Spearman's ρ. By Proposition 2.8, R_ρ coincides with R_{σ_1} so that, in view of Example 2.6 with $p = 1$, the properties (1.2) and (1.3) hold.

[7] C is a so-called ordinal sum; see (Nelsen, 2006, Sect. 3.2.2).

Next, consider Kendalls's τ. In order to prove (1.2), we assume $\tau(C) = \tau(\Pi)$, i.e. $Q(C,C) = Q(\Pi,\Pi)$, for some $C \in \mathcal{C}^{\uparrow}$. In view of (3.18) and (3.19) we obtain $Q(\Pi,\Pi) \le Q(C,\Pi) \le Q(C,C) = Q(\Pi,\Pi)$ so that

$$0 \le 4 \int_{[0,1]^2} |C(u,v) - \Pi(u,v)| \, \mathrm{d}\lambda(u,v)$$

$$= 4 \int_{[0,1]^2} C(u,v) - \Pi(u,v) \, \mathrm{d}\lambda(u,v) = Q(C,\Pi) - Q(\Pi,\Pi) = 0$$

which indeed implies $C = \Pi$. For the proof of (1.3), we suppose $\tau(C) = \tau(C^+)$, i.e. $Q(C,C) = Q(C^+,C^+)$. In view of (3.18) and (3.19) we obtain $Q(C,C) \le Q(C,C^+) \le Q(C^+,C^+) = Q(C,C)$ so that

$$0 \le 4 \int_0^1 |u - C(u,u)| \, \mathrm{d}u = 4 \int_0^1 u - C(u,u) \, \mathrm{d}u$$

$$= 4 \int_{[0,1]^2} C^+(u,v) - C(u,v) \, \mathrm{d}C^+(u,v) = Q(C^+,C^+) - Q(C,C^+) = 0\,.$$

Therefore $C(u,u) = u$ for all $u \in [0,1]$ so that $C = C^+$ (see (Durante and Sempi, 2016, Ex 2.6.4)).[8]

Finally, we turn to Gini's γ. In order to prove (1.2), we assume $\gamma(C) = \gamma(\Pi)$, i.e.

$$Q(C,C^+) + Q(C,C^-) = Q(\Pi,C^+) + Q(\Pi,C^-)\,,$$

for some $C \in \mathcal{C}^{\uparrow}$. In view of (3.18) and (3.19) we obtain

$$\begin{aligned} Q(\Pi,C^+) + Q(\Pi,C^-) &\le Q(C,C^+) + Q(\Pi,C^-) \\ &\le Q(C,C^+) + Q(C,C^-) \\ &= Q(\Pi,C^+) + Q(\Pi,C^-) \end{aligned}$$

so that

$$0 \le 4 \int_0^1 |C(u,u) - \Pi(u,u)| \, \mathrm{d}u = 4 \int_0^1 C(u,u) - \Pi(u,u) \, \mathrm{d}u = Q(C,C^+) - Q(\Pi,C^+) = 0\,.$$

It follows that $C(u,u) = \Pi(u,u)$, and Proposition 2.1 in Durante and Papini (2009) yields $C = \Pi$. For the proof of (1.3), we suppose $\gamma(C) = \gamma(C^+)$, i.e.

$$Q(C,C^+) + Q(C,C^-) = Q(C^+,C^+) + Q(C^+,C^-).$$

In view of (3.18) and (3.19) we obtain

$$\begin{aligned} Q(C,C^+) + Q(C,C^-) &\le Q(C^+,C^+) + Q(C,C^-) \\ &\le Q(C^+,C^+) + Q(C^+,C^-) \\ &= Q(C,C^+) + Q(C,C^-)\,, \end{aligned}$$

[8] The observation that $\tau(C) = \tau(C^+)$ implies $C = C^+$ also in the multivariate case is contained in (Fuchs, McCord and Schmidt, 2018, Thm. 3.2).

which implies

$$0 \le 4 \int_0^1 |u - C(u,u)| \, \mathrm{d}u = 4 \int_0^1 u - C(u,u) \, \mathrm{d}u$$

$$= 4 \int_{[0,1]^2} C^+(u,v) - C(u,v) \, \mathrm{d}C^+(u,v) = Q(C^+, C^+) - Q(C, C^+) = 0 \, .$$

Therefore $C(u,u) = u$ for all $u \in [0,1]$ so that $C = C^+$ (Durante and Sempi, 2016, Ex. 2.6.4).

B.6. Proof of Proposition 2.8. This follows readily from the fact that $C^\uparrow \ge \Pi$ since

$$R_{\sigma_1}(C) = \frac{\|C^\uparrow - \Pi\|_1}{\|C^+ - \Pi\|_1} = 12 \int_{[0,1]^2} C^\uparrow(u,v) - uv \, \mathrm{d}\lambda(u,v)$$

$$= 12 \int_{[0,1]^2} C^\uparrow(u,v) \, \mathrm{d}\lambda(u,v) - 3 = R_\rho(C) \, .$$

B.7. Proof of Theorem 2.9. In view of Definition A.5, we have $\kappa(C^\downarrow) = \kappa(C^- * C^\uparrow) = -\kappa(C^\uparrow)$. Consequently, we know from Lemma B.1 and the monotonicity of κ with respect to the pointwise ordering that

$$-\kappa(C^\uparrow) = \kappa(C^\downarrow) \le \kappa(C) \le \kappa(C^\uparrow) \, ,$$

which implies $|\kappa(C)| \le \kappa(C^\uparrow) = R_\kappa(C)$. Moreover, if C is stochastically monotone we have $C = C^\downarrow$ or $C = C^\uparrow$ and, therefore, $|\kappa(C)| = \kappa(C^\uparrow)$.

B.8. Proof of Proposition 2.10. First, we point out that that the Markov product of two copulas C and D satisfies

$$\partial_1 (C * D)(\cdot, v) = \partial_u \int_0^1 \partial_2 C(\cdot, t) \cdot \partial_1 D(t,v) \, \mathrm{d}t \preceq \partial_1 D(\cdot, v) \tag{3.20}$$

for all $v \in [0,1]$, where "$\preceq$" denotes the majorization order introduced in Definition A.7. This follows from Theorem A.8(3) and the fact that $\partial_1(C * D)(u,v) = T_C \partial_1 D(\cdot, v)(u)$. In particular,

$$(C * D)^\uparrow(u,v) \le D^\uparrow(u,v) \, .$$

Now suppose X, Y and Z are continuous random variables such that Y and Z are conditionally independent given X. Then $C_{ZY} = C_{ZX} * C_{XY}$ in view of Theorem 3.1 in Darsow, Nguyen and Olsen (1992), and (3.20) yields

$$C_{ZY}^\uparrow = (C_{ZX} * C_{XY})^\uparrow \le C_{XY}^\uparrow \, .$$

Thus, the data processing inequality $R_\mu(C_{ZY}) = \mu(C_{ZY}^\uparrow) \le \mu(C_{XY}^\uparrow) = R_\mu(C_{XY})$ follows from the monotonicity of μ.

B.9. Proof of Corollary 2.11. The data processing inequality in Proposition 2.10 states that $R_\mu(f(X), Y) \le R_\mu(X, Y)$ for all measurable functions f. If, in addition, X and Y are independent given $f(X)$, a second application of Proposition 2.10 yields $R_\mu(X,Y) \le R_\mu(f(X), Y)$, and equality holds.

APPENDIX C: PROOFS OF THE RESULTS IN SECTION 3

C.1. Proof of Theorem 3.1. The equality $C^{\#}_{N_1,N_2}(A)^{\uparrow} = C^{\#}_{N_1,N_2}(A^{\uparrow})$ follows directly from the definition of Algorithm 1 and the characterization (3.6). It remains to show that the matrix $A^{\uparrow}$ satisfies indeed the properties in (3.2). To do so, we calculate

$$\sum_{\ell=1}^{N_2} a^{\uparrow}_{k\ell} = \sum_{\ell=1}^{N_2} \widetilde{B}^{\ell}_k - \widetilde{B}^{\ell-1}_k = \widetilde{B}^{N_2}_k - \widetilde{B}^{0}_k = \widetilde{B}^{N_2}_k = \sum_{\ell=1}^{N_2} a_{k\ell} = N_2$$

as well as

$$\sum_{k=1}^{N_1} a_{k\ell} = \sum_{k=1}^{N_1} \widetilde{B}^{\ell}_k - \widetilde{B}^{\ell-1}_k = \sum_{k=1}^{N_1} B^{\ell}_k - B^{\ell-1}_k$$
$$= \sum_{j=1}^{\ell}\sum_{k=1}^{N_1} a_{kj} - \sum_{j=1}^{\ell-1}\sum_{k=1}^{N_1} a_{kj} = \ell N_1 - (\ell-1)N_1 = N_1 \, .$$

The nonnegativity of $a^{\uparrow}_{k\ell}$ follows by construction.

C.2. Proof of Theorem 3.2. We will start by showing a contraction property of the (SI)-rearrangement with respect to D_p. For all copulas C and D, it holds by Theorem A.8(b)

$$\partial_1 C^{\uparrow}(\cdot, v) - \partial_1 D^{\uparrow}(\cdot, v) \preceq \partial_1 C(\cdot, v) - \partial_1 D(\cdot, v)$$

for all v in $[0,1]$, where "$\preceq$" denotes the majorization order introduced in Definition A.7. Thus, due to Theorem A.8, we have for all $v \in [0,1]$ and any $1 \le p < \infty$

$$\int_0^1 \left|\partial_1 C^{\uparrow}(u,v) - \partial_1 D^{\uparrow}(u,v)\right|^p \, \mathrm{d}u \le \int_0^1 \left|\partial_1 C(u,v) - \partial_1 D(u,v)\right|^p \, \mathrm{d}u \, .$$

and integrating with respect to v yields

$$D_p(C^{\uparrow}, D^{\uparrow}) \le D_p(C, D) \, .$$

Now it follows by similar arguments as in the proof of Theorem 4.5.8 in Durante and Sempi (2016) (these authors considered the case $N_1 = N_2$) that

$$0 \le D_p(C^{\#}_{N_1,N_2}(C)^{\uparrow}, C^{\uparrow}) \le D_p(C^{\#}_{N_1,N_2}(C), C) \to 0 \, .$$

C.3. Proof of Theorem 3.4. The almost sure convergence of $D_1(\hat{C}^{\#}_{N_1,N_2,n}, C) \to 0$ follows from Theorem 3.12 in Junker, Griessenberger and Trutschnig (2021), where $\hat{C}^{\#}_{N_1,N_2,n}$ is a genuine copula. Thus, an application of the continuity property given in Theorem 3.2 implies

$$0 \le D_1((\hat{C}^{\#}_{N_1,N_2,n})^{\uparrow}, C^{\uparrow}) \le D_1(\hat{C}^{\#}_{N_1,N_2,n}, C) \to 0 \, .$$

and therefore $\hat{R}_{\mu} \to R_{\mu}(C)$ almost surely.