Scandinavian Journal of Statistics Theory and Applications

Scandinavian Journal of Statistics
doi: 10.1111/j.1467-9469.2011.00767.x

A Copula-Based Non-parametric Measure of Regression Dependence

HOLGER DETTE

Fakultät für Mathematik, Ruhr-Universität Bochum

KARL F. SIBURG

Fakultät für Mathematik, Technische Universität Dortmund

PAVEL A. STOIMENOV

Fakultät Statistik, Technische Universität Dortmund

ABSTRACT. This article presents a framework for comparing bivariate distributions according to their degree of regression dependence. We introduce the general concept of a regression dependence order (RDO). In addition, we define a new non-parametric measure of regression dependence and study its properties. Besides being monotone in the new RDOs, the measure takes on its extreme values precisely at independence and almost sure functional dependence, respectively. A consistent non-parametric estimator of the new measure is constructed and its asymptotic properties are investigated. Finally, the finite sample properties of the estimate are studied by means of a small simulation study.

Key words: conditional distribution, copula, local linear estimation, measure of dependence, regression, stochastic order

1. Introduction and motivation

There is an extensive body of literature on the problem of ordering and measuring the dependence of two random variables. Almost all of the research in this area is concerned with the concept of positive dependence. Orders of positive dependence were considered by many authors, e.g. Lehmann (1966), Esary *et al.* (1967) and Schriever (1987); see also Scarsini & Shaked (1996) for a detailed survey. Axiomatic approaches to orders and measures of positive dependence were introduced by Schweizer & Wolff (1981), Scarsini (1984) and Kimeldorf & Sampson (1987). The abundance of notions of positive dependence contrasts, however, with the silence concerning regression dependence, with the exception of the work of Dabrowska (1981, 1985) and the measure suggested by Hall (1970).

This article presents a new approach to the problem of ordering and measuring regression dependence in the bivariate case. The terms 'order' and 'ordering' are used in the sense of a preorder, i.e. a reflexive and transitive relation. We drop the requirement of antisymmetry in order to allow for an arbitrary functional form of the regression. For convenience, an order for random variables and the corresponding relations for distributions and distribution functions are used synonymously. Also, we do not strictly discriminate between distribution functions and distributions; the notation is the same.

Let (X, Y) be a bivariate random vector with marginal distribution functions F_X and F_Y, respectively, and joint distribution function $F_{X,Y}$. Since regression dependence is a directional relationship, it is first necessary to specify the direction of interest. Without loss of generality, we study the dependence of Y on X. The fundamental idea behind regression is predictability – the more predictable Y is from X, the more regression dependent they are. It is straightforward to single out the two extreme cases: independence and almost sure functional dependence, when there exists a Borel measurable function g such that $Y = g(X)$ with probability one (Lancaster, 1963). In the former case, X provides no information about Y, whereas in the latter case there is perfect predictability of Y from X.

Apart from the two extreme cases, however, there exists a variety of intermediate ones with a certain degree of regression dependence in a sense yet to be specified. The essence of our approach is the fact that the predictability of Y from X is intrinsically related to the variability of the conditional distributions $F_{Y|X=x}$ of Y given $X = x$. More precisely, the less variable $F_{Y|X=x}$, the more predictable Y from X, and thus the more regression dependent (X, Y). For example, perfect predictability, i.e. almost sure functional dependence of Y on X, is equivalent to the degeneracy of $F_{Y|X=x}$ for almost all x. Unless otherwise stated 'almost' is used in the sense of the respective probability measure, which is clear from the context. It follows that, if $(\tilde{X}, \tilde{Y})$ is another pair of random variables, then the general idea is to consider (X, Y) less regression dependent than $(\tilde{X}, \tilde{Y})$ if $F_{Y|X=x}$ is more variable than $F_{\tilde{Y}|\tilde{X}=x}$ for almost all x. Therefore, a bivariate regression dependence order is associated to a univariate variability order, and different variability orders could lead, in general, to different regression orders.

This approach, however, is not applicable unless X and $\tilde{X}$ have the same distribution. Moreover, it is even necessary that Y and $\tilde{Y}$ are identically distributed because, otherwise, their different variability will affect the variability of $F_{Y|X}$ and $F_{\tilde{Y}|\tilde{X}}$ and, in this way, the degree of regression dependence. For this reason, a comparison of two bivariate random vectors with arbitrary marginals is possible only after their transformation to the same Fréchet class. If the marginals are continuous, it is natural to consider the probability integral transformations $(U, V) = (F_X(X), F_Y(Y))$ and $(\tilde{U}, \tilde{V}) = (F_{\tilde{X}}(\tilde{X}), F_{\tilde{Y}}(\tilde{Y}))$, which have uniform marginal distributions. In this case, we regard (X, Y) less regression dependent than $(\tilde{X}, \tilde{Y})$ if $F_{V|U=u}$ is more variable than $F_{\tilde{V}|\tilde{U}=u}$ for almost all u.

It should be noted, however, that while lower variability of the conditional distributions is a necessary condition for defining a regression dependence order, it is not sufficient. As the details will be given later in section 3, we only mention here that the choice of the variability order cannot be arbitrary, but should take into account the two extremes of regression dependence, namely, independence and almost sure functional dependence. We will show that the most common variability orders lead indeed to regression orders.

In section 4, we introduce a new non-parametric measure of regression dependence, study its properties and demonstrate its advantages over the correlation ratio. Besides being monotone in the new regression orders, the measure possesses several appealing properties. For instance, it takes on its minimum if and only if X and Y are independent, and its maximum if and only if Y is almost surely (a.s.) a Borel function of X.

Two estimates of the new dependence measure are introduced in section 5 and their asymptotic properties are investigated. Finally, section 6 contains a small simulation study which shows that the proposed estimates have a reasonable performance for moderate sample size.

2. Notation and preliminaries

This section introduces the notation and states some technical facts which will be needed in the sequel. Except for the results on univariate variability orders, attention is restricted to the set $\mathfrak{F}$ of all bivariate distribution functions with continuous marginal distribution functions, as well as the set $\mathfrak{X}$ of all bivariate random vectors with distribution functions in $\mathfrak{F}$. For $(X, Y) \in \mathfrak{X}$, $F_{X,Y} \in \mathfrak{F}$ denotes its joint distribution function with marginal distribution functions F_X and F_Y, respectively, while $F_{Y|X=x}$ denotes the conditional distribution function of Y given $X = x$. For the probability integral transformations of $(X, Y) \in \mathfrak{X}$, we shall write

$$U := F_X(X) \quad \text{and} \quad V := F_Y(Y).$$

Thus, U and V have uniform distributions on the closed unit interval $[0, 1]$, which will be denoted by I. The notation $F_{U,V}$ and $F_{V|U=u}$ will be used for joint and conditional distribution of (U, V) and V given $U=u$, respectively. The first result describes the two extreme cases of regression dependence for (X, Y) in terms of (U, V).

Proposition 1. *For any* $(X, Y) \in \mathfrak{X}$, *the following are true:*

(i) *X and Y are independent if and only if U and V are independent.*
(ii) *U and V are independent if and only if $F_{V|U=u}=F_V$ for almost all u.*
(iii) *Y is a.s. a Borel function of X if and only if V is a.s. a Borel function of U.*
(iv) *V is a.s. a Borel function of U if and only if $F_{V|U=u}$ is degenerate for almost all u.*

Proof. (i) and (ii) are obvious. As for (iii), since F_X is continuous, $Y=f\circ X$ a.s. implies $Y=f\circ F_X^{-1}\circ F_X\circ X$ a.s., so that $V=g\circ U$ a.s. with the measurable function $g:=F_Y\circ f\circ F_X^{-1}$; conversely, if $V=g\circ U$ we set $f:=F_Y^{-1}\circ g\circ F_X$. Finally, (iv) follows from the observation that $V=f(U)$ is equivalent to the fact that the graph of f is measurable and has probability one, i.e.

$$1=\int_{I^2}\mathbf{1}_{\mathrm{gr}f}(u, v)\,\mathrm{d}F_{U,V}(u, v)=\int_I\int_I\mathbf{1}_{\mathrm{gr}f}(u, v)\,\mathrm{d}F_{V|U=u}(v)\,\mathrm{d}F_U(u).$$

This is equivalent to $F_{V|U=u}$ being degenerate for almost all u.

Since we work with the probability integral transformations, the concept of copulas is tailored for our approach. Formally, a bivariate copula (or briefly, a copula) is the restriction to I^2 of a bivariate distribution function with uniform marginals on I. In fact, the unique copula $C_{X,Y}$ of $(X, Y)\in\mathfrak{X}$ coincides with $F_{U,V}$ on I^2. In particular, the copula corresponding to independent variables is the product copula $P(u, v)=uv$.

Denote by $\mathfrak{C}$ the set of all copulas, and by $\partial_i C$ the partial derivative of $C\in\mathfrak{C}$ with respect to the ith variable. The following properties of copulas are easy consequences of the definition; for a proof see, e.g. Nelson (2006).

Proposition 2. *For any* $C\in\mathfrak{C}$, *the following statements are true:*

(i) *C is Lipschitz continuous; more precisely, for all $(u_1, v_1), (u_2, v_2)\in I^2$ we have*

$$|C(u_2, v_2)-C(u_1, v_1)|\leq|u_2-u_1|+|v_2-v_1|.$$

(ii) *For each $v\in I$, $\partial_1 C(u, v)$ exists for almost all $u\in I$; similarly, for each $u\in I$, $\partial_2 C(u, v)$ exists for almost all $v\in I$. Moreover, the partial derivatives satisfy*

$$0\leq\partial_i C\leq 1$$

for $i=1, 2$ wherever they are defined.

Remark 1. (i) Note that the Lipschitz continuity implies that a copula is absolutely continuous in each argument, so that it can be recovered from any of its partial derivatives by integration.

(ii) In fact, we have $0\leq\partial_i C\leq 1$ for $i=1, 2$ Lebesgue almost everywhere (a.e.) on I^2 since, as Lipschitz continuous functions, copulas are differentiable Lebesgue a.e. in view of Rademacher's Theorem (Evans, 1998). Moreover, by Evans (1998, theorem 5.8.4), we also have $\partial_i C\in L^p(I^2, \mathbb{R})$ with $p\geq 1$.

There is a relationship between the conditional distribution $F_{V|U=u}$ and the corresponding copula $C_{X,Y}$, which is given by

$$F_{V|U=u}(v)=\partial_1 C_{X,Y}(u, v) \tag{1}$$

wherever the partial derivative exists (Nelson, 2006). Moreover, we have the following result related to proposition 1.

Proposition 3. *For any* $(X, Y) \in \mathfrak{X}$, *the following are true:*

(i) *X and Y are independent if and only if* $\partial_1 C_{X,Y}(u,v) = v$ *for Lebesgue almost all* $(u,v) \in I^2$.
(ii) *Y is a.s. a Borel function of X if and only if* $\partial_1 C_{X,Y}(u,v) \in \{0,1\}$ *for Lebesgue almost all* $(u,v) \in I^2$.

Proof. The first statement follows from remark 1 (i), while the second is a consequence of Darsow *et al.* (1992, theorem. 11.1) and Siburg & Stoimenov (2010, theorem. 4.2).

Since our approach to ordering regression dependence employs the variability of the conditional distribution functions, the rest of this section deals with stochastic orders that compare the variability or dispersion of two arbitrary random variables X and Y (or their univariate distributions F_X and F_Y); we refer to Müller & Stoyan (2002) and Shaked & Shanthikumar (2007) for a detailed study of stochastic orders.

Probably, the most common variability order is the convex order. X is smaller than Y in the convex order (denoted as $X \leq_{\mathrm{cx}} Y$) if

$$\mathrm{E}[\phi(X)] \leq \mathrm{E}[\phi(Y)] \tag{2}$$

for all convex functions $\phi : \mathbb{R} \to \mathbb{R}$, provided the expectations exist. Depending on the context, i.e. whether working with random variables or distribution functions, we write $X \leq_{\mathrm{cx}} Y$ or $F_X \leq_{\mathrm{cx}} F_Y$. This order reflects the intuitive idea that convex functions take on their (relatively) larger values over regions of the form $(-\infty, a) \cup (b, \infty)$ for $a < b$. Therefore, if (2) holds, Y is more variable (or more dispersed) than X. The next result is a direct consequence of (2).

Proposition 4. *Let X and Y be two random variables. If* $X \leq_{\mathrm{cx}} Y$, *then* $\mathrm{E}[X] = \mathrm{E}[Y]$ *and* $\mathrm{Var}[X] \leq \mathrm{Var}[Y]$.

As can be seen from proposition 4, only random variables with the same expectations can be compared. When X and Y have finite expectations, we can use the convex order to define a location-free variability order. Namely, we call X smaller than Y in the dilation order (denoted as $X \leq_{\mathrm{dil}} Y$) if

$$X - \mathrm{E}[X] \leq_{\mathrm{cx}} Y - \mathrm{E}[Y]. \tag{3}$$

Corollary 1. *Let X and Y be two random variables.* If $X \leq_{\mathrm{dil}} Y$, *then* $\mathrm{Var}[X] \leq \mathrm{Var}[Y]$.

Another important location-free variability order is the dispersive order. F_X is smaller than F_Y in the dispersive order (denoted as $F_X \leq_{\mathrm{disp}} F_Y$) if

$$F_X^{-1}(b) - F_X^{-1}(a) \leq F_Y^{-1}(b) - F_Y^{-1}(a) \tag{4}$$

for all $0 < a \leq b < 1$. As noted in Shaked & Shanthikumar (2007), it is conceptually clear that this order compares the variability of F_X and F_Y because it requires the difference between any two quantiles of F_X to be smaller than the corresponding quantiles of F_Y.

The next result shows the relation between the orders $\leq_{\mathrm{disp}}$ and $\leq_{\mathrm{dil}}$; compare Shaked & Shanthikumar (2007, theorem. 3.B.16).

Proposition 5. *Let X and Y be two random variables with finite expectations. Then $X \leq_{\text{disp}} Y$ implies $X \leq_{\text{dil}} Y$.*

3. Regression dependence orders

The fundamental idea to introduce an order of regression dependence on $\mathfrak{X}$ (respectively $\mathfrak{F}$) is to compare the variability of the conditional distributions, since low and high dispersion is tantamount to high and low predictability, respectively. However, as discussed in the introduction, a comparison of two elements of $\mathfrak{X}$ with arbitrary marginals is possible only after their transformation to the same Fréchet class which can be accomplished using the probability integral transformations. Essentially, a random vector $(X, Y) \in \mathfrak{X}$ is less regression dependent than another random vector $(\tilde{X}, \tilde{Y}) \in \mathfrak{X}$ if $F_{\tilde{V}|\tilde{U}=u}$ is less variable (in some univariate variability order) than $F_{V|U=u}$ for almost all u. More precisely, we adopt the following definition.

Definition 1. *A relation $\preccurlyeq$ on $\mathfrak{X}$* (or $\mathfrak{F}$) *is a regression dependence order (RDO) if it is reflexive and transitive, and satisfies the following:*

- (O1) *$(X, Y) \preccurlyeq (\tilde{X}, \tilde{Y})$ implies $F_{\tilde{V}|\tilde{U}=u} \leq_{\bullet} F_{V|U=u}$ for almost all $u \in I$, where $\leq_{\bullet}$ is a univariate variability order.*
- (O2) *If Y is a.s. a Borel function of X, and if $(X, Y) \preccurlyeq (\tilde{X}, \tilde{Y})$,* then *$\tilde{Y}$ is a.s. a Borel function of $\tilde{X}$.*
- (O3) *If X and Y are independent, and if $(\tilde{X}, \tilde{Y}) \preccurlyeq (X, Y)$, then $\tilde{X}$ and $\tilde{Y}$ are independent.*

Property (O1) indicates that an RDO is always associated to a given variability order. Therefore, a relation $\preccurlyeq$ satisfying (O1) with respect to the univariate variability order $\leq_{\bullet}$ will be denoted by $\preccurlyeq_{\bullet}$.

Conditions (O2) and (O3) deal with the two extreme cases. Since almost sure functional dependence is equivalent to perfect predictability of Y from X, the corresponding distribution must have the greatest regression dependence possible. Consequently, any distribution which is more dependent must also correspond to almost sure functional dependence; hence (O2). Similarly, the least dependent situation is given when X and Y are independent. Hence, any less dependent distribution must be again the distribution of independent random variables, which is expressed in (O3).

In view of condition (O1), probably the easiest way to construct an RDO is to choose some variability order $\leq_{\bullet}$, define $(X, Y) \preccurlyeq (\tilde{X}, \tilde{Y})$ if and only if $F_{\tilde{V}|\tilde{U}=u} \leq_{\bullet} F_{V|U=u}$ for almost all $u \in I$, and check whether conditions (O2) and (O3) are satisfied. In fact, since no distribution is less dispersed than a degenerate one, (O2) should always be satisfied in view of proposition 1, and it remains to prove (O3).

It is important to note that an RDO corresponding to a variability order which is not location-free (e.g. the convex order $\leq_{\text{cx}}$) is unnecessarily restrictive, for then only distributions with the same regression function can be compared. However, since we want to compare the strength of regression dependence with respect to possibly different regression functions, we will consider location-free orders only. Amongst them, the dilation order $\leq_{\text{dil}}$ and the dispersive order $\leq_{\text{disp}}$ are the most important and common ones. The next result states that the corresponding relations $\preccurlyeq_{\text{dil}}$ and $\preccurlyeq_{\text{disp}}$ are indeed RDOs.

Theorem 1. The relations $\preccurlyeq_{\text{dil}}$ and $\preccurlyeq_{\text{disp}}$ are RDOs.

Proof. In view of proposition 5, we need only prove (O2) and (O3) for the relation $\preccurlyeq_{\text{dil}}$. It is clear from corollary 1 that $\preccurlyeq_{\text{dil}}$ satisfies (O2). In order to prove (O3), we may, in view of proposition 1, restrict to considering U and V instead of X and Y. Assuming that $(\tilde{U}, \tilde{V}) \preccurlyeq_{\text{dil}} (U, V)$ with independent U and V, we conclude from corollary 1 that

$$\text{Var}[\tilde{V} \mid \tilde{U} = u] \geq \frac{1}{12} \tag{5}$$

for almost all u. By the law of total variance, we obtain equality in (5), as well as

$$\text{E}[\tilde{V} \mid \tilde{U} = u] = \text{E}[\tilde{V}] = \frac{1}{2} \tag{6}$$

for almost all u. From (6) and (5) it follows that, for almost all u, $F_{V|U=u} \leq_{\text{cx}} F_{\tilde{V}|\tilde{U}=u}$ with equal variances. But then both distributions are the same (Shaked & Shanthikumar, 2007, theorem. 3.A.42). This proves (O3), and hence the theorem.

4. Measures of regression dependence

We now turn to the subject of how to measure the degree of regression dependence in the set $\mathfrak{X}$ (or $\mathfrak{F}$). It is clear that without specifying an RDO any discussion of measures of regression dependence is problematic. We adopt the following definition.

Definition 2. *Let $\preccurlyeq$ be an arbitrary RDO. A function $\mu : \mathfrak{X} \to [0, 1]$ is a measure of regression dependence (MRD) with respect to $\preccurlyeq$ if it satisfies the following conditions:*

- (M1) $(X, Y) \preccurlyeq (\tilde{X}, \tilde{Y})$ *implies* $\mu(X, Y) \leq \mu(\tilde{X}, \tilde{Y})$;
- (M2) $\mu(X, Y) = 1$ *if and only if Y is a.s. a Borel function of X*;
- (M3) $\mu(X, Y) = 0$ *if and only if X and Y are independent.*

Remark 2. Alternatively, μ can also be defined as a functional on $\mathfrak{F}$, and we sometimes write $\mu(F_{X,Y})$ instead of $\mu(X, Y)$.

Condition (M1) is the usual monotonicity property required by any measure of dependence. (M2) and (M3) concern the two extreme cases of regression dependence. We point out how strong both conditions are – in fact, a measure of dependence satisfying (M2) and (M3) has not yet been proposed in the literature. For instance, (M2) is much stronger than Rényi's corresponding postulate in Rényi (1959), according to which a measure of dependence should take on its maximal value 1 if one of X and Y is a.s. a function of the other. What is more, Rényi mentioned that it is natural to pose an 'only if' requirement, but since the condition was rather restrictive, it was better to leave it out. With respect to (M3), we point out that the well-known correlation ratio is not a MRD in the sense of Definition 2 because it attains its minimum at 0 not only when X and Y are independent; examples are presented later in this section.

We now turn to the construction of a non-parametric MRD. The following is the main result in this section.

Theorem 2. *The function $r : \mathfrak{X} \to [0, 1]$ defined by*

$$r(X, Y) = 6 \int_0^1 \int_0^1 F_{V|U=u}(v)^2 \, \mathrm{d}v \, \mathrm{d}u - 2 \tag{7}$$

is an MRD concurring with both $\preccurlyeq_{\text{dil}}$ and $\preccurlyeq_{\text{disp}}$.

Remark 3. Note that in view of (1), we have

$$r(X, Y) = 6 \|\partial_1 C_{X,Y}\|_2^2 - 2, \tag{8}$$

where $\|\cdot\|_2$ denotes the L^2-norm on I^2. By remark 1(ii), this shows that r is indeed well defined. Moreover, r can also be viewed as a functional on the set of copulas $\mathfrak{C}$, and we write $r(C_{X,Y})=r(X,Y)$.

In order to prove theorem 2, we make use of the following result.

Lemma 1. *For any* $C_{X,Y}\in\mathfrak{C}$, *we have* $\|\partial_1 C_{X,Y}\|_2^2\in[1/3,1/2]$. *Moreover, the following assertions hold:*

(i) $\|\partial_1 C_{X,Y}\|_2^2=1/3$ *if and only if X and Y are independent.*
(ii) $\|\partial_1 C_{X,Y}\|_2^2=1/2$ *if and only if Y is a.s. a Borel function of X.*

Proof.
(i) Consider the inequality

$$0\leq\int_0^1\int_0^1(\partial_1 C_{X,Y}(u,v)-v)^2\,\mathrm{d}u\,\mathrm{d}v=\int_0^1\int_0^1(\partial_1 C_{X,Y}(u,v))^2\,\mathrm{d}u\,\mathrm{d}v-\frac{1}{3}.$$

Hence, $\|\partial_1 C_{X,Y}\|_2^2\geq 1/3$ with equality if and only if $\partial_1 C_{X,Y}(u,v)=v$ Lebesgue a.e. on I^2, which by proposition 3(i) is equivalent to the independence of X and Y.

(ii) By theorem 2(ii), we have $0\leq\partial_1 C_{X,Y}\leq 1$ and thus $(\partial_1 C_{X,Y})^2\leq\partial_1 C_{X,Y}$, with equality if and only if $\partial_1 C_{X,Y}\in\{0,1\}$. Consequently,

$$\|\partial_1 C_{X,Y}\|_2^2\leq\int_0^1\int_0^1\partial_1 C_{X,Y}(u,v)\,\mathrm{d}u\,\mathrm{d}v=\frac{1}{2}$$

with equality if and only if $\partial_1 C_{X,Y}\in\{0,1\}$ Lebesgue a.e. in I^2, which by proposition 3(ii) is equivalent to Y being a.s. a Borel function of X.

We will also make use of the following representation formula for univariate distribution functions whose support is contained in I. The proof uses integration by parts for Lebesgue–Stieltjes integrals (Hewitt & Stromberg, 1975, theorem. 21.67) and is omitted.

Lemma 2. *Let F be a univariate distribution function with support in I. Then*

$$2\int_0^1\int_0^p F^{-1}(t)\,\mathrm{d}t\,\mathrm{d}p-\int_0^1 F^{-1}(t)\,\mathrm{d}t=\int_0^1 F(v)^2\,\mathrm{d}v-\int_0^1 F(v)\,\mathrm{d}v.$$

We now turn to the proof of the theorem.

Proof of theorem 2. The property $0\leq r(X,Y)\leq 1$, as well as the conditions (M2) and (M3), are immediately implied by lemma 1.

It remains to show the monotonicity condition (M1); in view of proposition 5, it suffices to prove it for the RDO $\preccurlyeq_{\mathrm{dil}}$. Ramos & Sordo (2003) showed that two univariate distribution functions F and G with finite expectations satisfy $F\leq_{\mathrm{dil}}G$ if and only if, for all $v\in[0,1]$,

$$\int_0^v F^{-1}(t)\,\mathrm{d}t-v\int_0^1 F^{-1}(t)\,\mathrm{d}t\geq\int_0^v G^{-1}(t)\,\mathrm{d}t-v\int_0^1 G^{-1}(t)\,\mathrm{d}t. \tag{9}$$

Now assume that $(X,Y)\preccurlyeq_{\mathrm{dil}}(\tilde{X},\tilde{Y})$ so that $F_{\tilde{V}|\tilde{U}=u}\leq_{\mathrm{dil}}F_{V|U=u}$ for almost all $u\in I$. Then, integrating (9) over v we obtain

$$\int_0^1\int_0^v F^{-1}_{\tilde{V}|\tilde{U}=u}(t)\,\mathrm{d}t\,\mathrm{d}v-\frac{1}{2}\int_0^1 F^{-1}_{\tilde{V}|\tilde{U}=u}(t)\,\mathrm{d}t\geq\int_0^1\int_0^v F^{-1}_{V|U=u}(t)\,\mathrm{d}t\,\mathrm{d}v-\frac{1}{2}\int_0^1 F^{-1}_{V|U=u}(t)\,\mathrm{d}t$$

for almost all $u \in I$. Applying lemma 2 we find that, for almost all u,

$$\int_0^1 F_{\tilde{V}|\tilde{U}=u}(v)^2\,\mathrm{d}v - \int_0^1 F_{\tilde{V}|\tilde{U}=u}(v)\,\mathrm{d}v \geq \int_0^1 F_{V|U=u}(v)^2\,\mathrm{d}v - \int_0^1 F_{V|U=u}(v)\,\mathrm{d}v.$$

Integrating this over $u \in I$, substituting $\partial_1 C_{X,Y}(u,v)$ for $F_{V|U=u}(v)$ by (1), and using $\int_0^1\int_0^1 \partial_1 C_{X,Y}(u,v)\,\mathrm{d}v\,\mathrm{d}u = 1/2$ for all $C_{X,Y} \in \mathfrak{C}$, we obtain

$$\|\partial_1 C_{\tilde{X},\tilde{Y}}\|_2^2 = \int_0^1\int_0^1 (\partial_1 C_{\tilde{X},\tilde{Y}}(u,v))^2\,\mathrm{d}v\,\mathrm{d}u \geq \int_0^1\int_0^1 (\partial_1 C_{X,Y}(u,v))^2\,\mathrm{d}v\,\mathrm{d}u = \|\partial_1 C_{X,Y}\|_2^2.$$

Since, by remark 3, $r(X, Y) = 6\|\partial_1 C_{X,Y}\|_2^2 - 2$, this proves (M1) and hence the theorem.

Proposition 6. *If $f, g: \mathbb{R} \to \mathbb{R}$ are strictly monotone functions then*

$$r(f(X), g(Y)) = r(X, Y).$$

Proof. We distinguish four different cases. If f and g are both increasing, it is well known (Nelsen, 2006, theorem 2.4.3) that

$$C_{f(X),g(Y)} = C_{X,Y},$$

which immediately implies $r(f(X), g(Y)) = r(X, Y)$. If f is increasing and g is decreasing, then (Nelsen, 2006, theorem 2.4.4)

$$C_{f(X),g(Y)}(u, v) = u - C_{X,Y}(u, 1 - v).$$

Therefore, we conclude $\|\partial_1 C_{f(X),g(Y)}\|_2^2 = \|\partial_1 C_{X,Y}\|_2^2$, which again implies $r(f(X), g(Y)) = r(X, Y)$. If f is decreasing and g is increasing, the result follows from interchanging f and g in the previous case. The final case when f and g are both decreasing can be shown similarly.

The following example illustrates the behaviour of the MRD r as a function of the copula parameter for some well-known one-parameter copula families.

Example 1. (a) Let C_a denote the Gaussian copula with parameter $a \in [-1, 1]$. Since the Gaussian copula is positively ordered with respect to a, i.e. it is monotone in the standard concordance order, and C_0 corresponds to independence, it is not surprising that $r(C_a)$ is an increasing function of $|a|$ as depicted in Fig. 1A. Moreover, for $a \in \{-1, 1\}$, we have almost sure functional dependence and, thus, $r = 1$.

(b) Consider the Farlie–Gumbel–Morgenstern (FGM) family of copulas, defined by $C_a(u, v) = uv + auv(1-u)(1-v)$ with $a \in [-1, 1]$. As in the previous example, the FGM family is positively ordered and C_0 characterizes independent random variables. In particular, we obtain $r(C_a) = a^2/15$; see Fig. 1B. However, as mentioned in Nelsen (2006), FGM copulas can only model relatively weak positive dependence, which explains intuitively the low values of r.

(c) A plot of r as a function of the parameter of the Frank copula

$$C_a = -\frac{1}{a}\ln\left(1 + \frac{(\mathrm{e}^{-au} - 1)(\mathrm{e}^{-av} - 1)}{\mathrm{e}^{-a} - 1}\right)$$

with $a \in (-\infty, \infty)\setminus\{0\}$ is presented in Fig. 1C.

(d) Finally, consider the Gumbel copula, given by

$$C_a = uv\,\mathrm{e}^{-a\ln u \ln v}$$

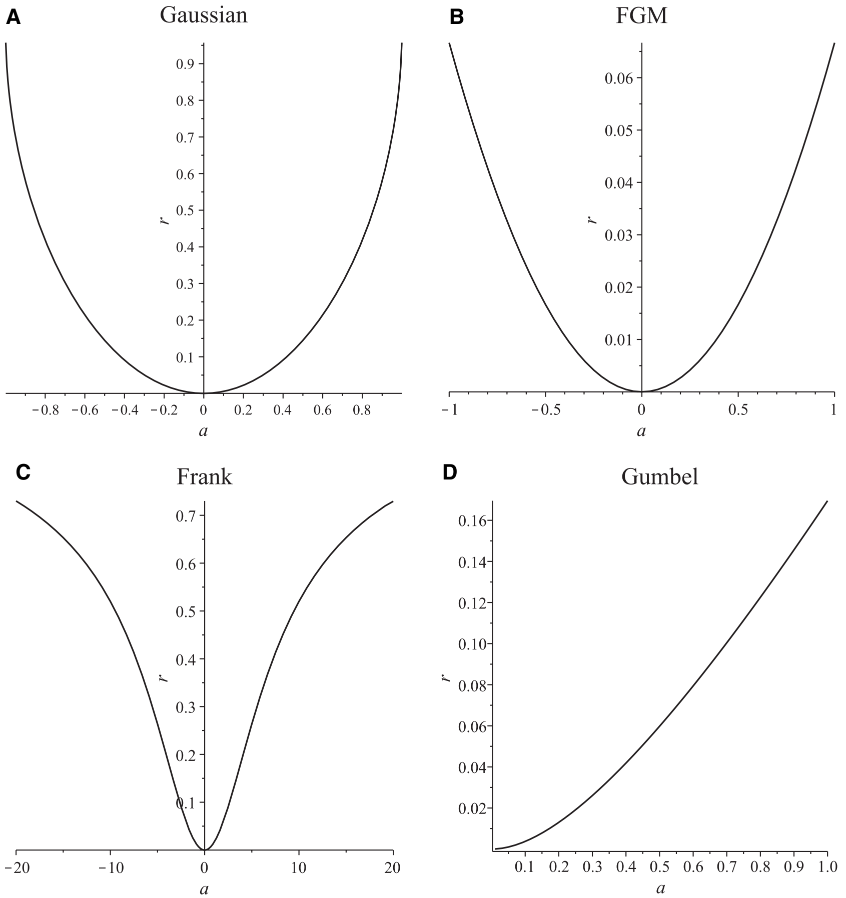


Fig. 1. The measure of regression dependence (MRD) r as a function of the parameter a of the four parametric copula families in example 1.

with $a \in (0, 1]$. By direct calculation we obtain

$$r(C_a) = \frac{3}{4a} e^{\frac{3}{2a}} E_1\left(\frac{3}{2a}\right) + \frac{a}{3} - \frac{1}{2},$$

where $E_1(x) := \int_1^\infty e^{-xs}/s\, ds$ is the exponential integral; see Fig. 1D.

We now turn attention to another quantity that might seem a natural choice for an MRD, namely the correlation ratio of the probability integral transformations. Define the functional $\tilde{\eta} : \mathfrak{X} \to \mathbb{R}$ by

$$\tilde{\eta}(X, Y)^2 := \eta(U, V)^2 = \frac{\mathrm{Var}[E[V \mid U]]}{\mathrm{Var}[V]} = 1 - \frac{E[\mathrm{Var}[V \mid U]]}{\mathrm{Var}[V]}. \tag{10}$$

Since $\mathrm{Var}[V] = 1/12$, it follows that

$$\tilde{\eta}(X, Y)^2 = 12\, \mathrm{Var}[E[V \mid U]].$$

In fact, the ordering of regression dependence suggested in Dabrowska (1981, section 3.1) is an ordering by correlation ratios and therefore is not consistent with our approach to RDOs. Moreover, neither the correlation ratio of Y on X nor the related measure $\tilde{\eta}(X, Y)^2$

are MRDs in the sense of definition 2, because (M3) will not be satisfied. Indeed, it follows from propositions 5 and 4 that $\tilde{\eta}$ is monotone with respect to both $\preccurlyeq_{\text{disp}}$ and $\preccurlyeq_{\text{dil}}$; in addition, $\tilde{\eta}(X, Y)=1$ if and only if Y is a.s. a Borel function of X. However, $\tilde{\eta}$ does not satisfy condition (M3) because there are random variables X and Y with $\tilde{\eta}(X, Y)=0$, which are not independent; we give two such examples.

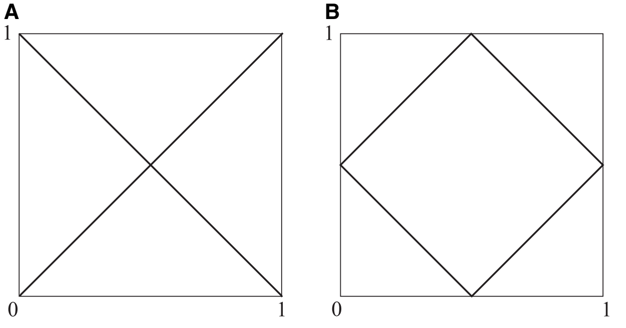
Example 2. Consider X and Y whose probability integral transformations U and V have the singular distribution with the support depicted in Fig. 2A. The support is the union of the main and secondary diagonal in I^2, so that probability mass 1/2 is uniformly distributed on each line segment. For every $u \in I$, the resulting conditional distribution $F_{V|U=u}$ is a two-point distribution at $v=u$ and $v=1-u$ and, thus, $E[V \mid U=u]=1/2$. Consequently, the conditional expectation $E[V \mid U]$ is degenerate and its variance Var[E[$V|U$]] vanishes, which means that $\eta(U, V)=\tilde{\eta}(X, Y)=0$. However, U and V and, thus, X and Y are not independent.

Example 3. Another situation where $\tilde{\eta}(X, Y)=0$ but X and Y are not independent is given when $F_{X,Y}$ is the circular uniform distribution. It is well known that in this case the ordinary correlation ratio $\eta(X, Y)$ vanishes. The same is true for the related measure $\tilde{\eta}(X, Y)$ since in this case $F_{U,V}$ is a degenerate distribution whose support is given in Fig. 2B (Nelsen, 2006, section 3.1.2). The arguments are analogous to those in the previous example.

5. Non-parametric estimation of *r*

In this section, we present a sample version of the MRD defined in (8). As pointed out in remark 3, r is a function of the copula $C_{X,Y}$ alone. $C_{X,Y}$ can be consistently estimated by the empirical copula (Deheuvels, 1979; Fermanian *et al.*, 2004). However, the empirical copula is locally constant and, thus, the estimation of r is more involved since it requires the estimation of the copula's partial derivative. The need for differentiability calls for a s smooth (differentiable) estimation of the copula, e.g. with a kernel-based technique.

For this purpose let $(X_1, Y_1), \ldots (X_n, Y_n)$ denote i.i.d. random variables with distribution function F and copula C, let K denote a symmetric kernel with compact support, say $[-1, 1]$, with corresponding cumulative distribution function



Support of the distribution of (U,V) in Example 2

Support of the distribution of (U,V) in Example 3

Fig. 2. Examples of $\tilde{\eta}(X, Y)=0$ where X and Y are not independent.

$$\bar{K}(x) = \int_{-\infty}^{x} K(t)\,\mathrm{d}t.$$

As an estimate for the partial derivative of the copula $\tau(u,v) = \partial_1 C(u,v)$ we use

$$\hat{\tau}_n(u,v) = \frac{1}{nh_1}\sum_{i=1}^{n}\omega\left(\frac{u-\hat{F}_{n1}(X_i)}{h_1}, \frac{v-\hat{F}_{n2}(Y_i)}{h_2}\right), \tag{11}$$

where $\hat{F}_{n1}$ and $\hat{F}_{n2}$ denote the empirical distribution functions of $X_1,\ldots,X_n$ and $Y_1,\ldots,Y_n$, respectively, h_1, h_2 denote bandwidths converging to 0 with increasing sample size and $\omega(x,y) = K(x)\bar{K}(y)$. Note that $\hat{\tau}_n$ is an integrated version of the estimate for the copula density considered in Fermanian (2005). Intuitively, we have for large sample size

$$\begin{aligned}
E[\hat{\tau}_n(u,v)] &\approx \frac{1}{h_1}\int \omega\left(\frac{u-F_X(x)}{h_1}, \frac{v-F_Y(y)}{h_2}\right)\mathrm{d}F(x,y) \\
&= \frac{1}{h_1}\int_0^1\int_0^1 \omega\left(\frac{u-s}{h_1}, \frac{v-t}{h_2}\right) c(s,t)\,\mathrm{d}s\,\mathrm{d}t \\
&= \int_0^1 \bar{K}\left(\frac{v-t}{h_2}\right) c(u,t)\,\mathrm{d}t\cdot(1+o(1)) \\
&= \int_0^v c(u,t)\,\mathrm{d}t\cdot(1+o(1)) = \partial_1 C(u,v)(1+o(1)),
\end{aligned}$$

where F_X, F_Y denote the marginal distributions of (X_1, Y_1) and $c(s,t)$ is the copula density. The following result makes these heuristic arguments more precise and gives a corresponding statement for the integrated version of $\hat{\tau}_n(u,v)$

$$\hat{\tau}_n^2 = \int_0^1\int_0^1 \hat{\tau}_n^2(u,v)\,\mathrm{d}u\,\mathrm{d}v, \tag{12}$$

which will serve as an estimate for the quantity

$$\tau^2 = \int_0^1\int_0^1 \|\,\partial_1 C(u,v)\,\|_2^2\,\mathrm{d}u\,\mathrm{d}v.$$

The estimate of the measure r defined in (8) is finally given by

$$\hat{r}_n = 6\hat{\tau}_n^2 - 2. \tag{13}$$

The next results show that $\hat{\tau}_n^2$ and $\hat{r}_n$ are asymptotically normal distributed.

Theorem 3. *Assume that the copula $C(u,v)$ is three and two times continuously differentiable with respect to the variable u and v, respectively. If the kernel K is symmetric, two times continuously differentiable with compact support and the bandwidths h_1 and h_2 satisfy*

$$nh_1^3 \longrightarrow \infty; \quad nh_1h_2 \longrightarrow \infty; \quad nh_1^4 \longrightarrow 0; \quad nh_2^4 \longrightarrow 0 \tag{14}$$

then

$$\sqrt{n}(\hat{r}_n - r) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 144\sigma^2),$$

where

$$\sigma^2 = \int_{[0,1]^3} \tau(s, v \wedge w)\tau(s, v)\tau(s, w)\, ds\, dv\, dw + \frac{1}{2}\left(\int_{[0,1]^2} \tau^2(u, v)\, du\, dv\right)^2$$
$$+ \frac{1}{4}\int_{[0,1]^4} \partial_1\tau^2(x_1, y_1)\partial_1\tau^2(x_2, y_2)(x_1 \wedge x_2 - x_1x_2)\, dx_1\, dy_1\, dx_2\, dy_2$$
$$+ \frac{1}{4}\int_{[0,1]^4} \partial_2\tau^2(x_1, y_1)\partial_2\tau^2(x_2, y_2)(y_1 \wedge y_2 - y_1y_2)\, dx_1\, dy_1\, dx_2\, dy_2$$
$$- \int_{[0,1]^3} \tau^2(x_1, v)\tau^2(x_1, w)\, dx_1\, dv\, dw$$
$$- \int_{[0,1]^4} I\{y_1 \le w\}\tau(x_1, w)\tau^2(u, y_1)\partial_2\tau(x_1, y_1)\, du\, dw\, dx_1\, dy_1$$
$$+ \frac{1}{2}\int_{[0,1]^4} \tau^2(x_1, v)\tau^2(u, y_1)\partial_2\tau(x_1, y_1)\, du\, dv\, dx_1\, dy_1. \tag{15}$$

Proof. The assertion follows from (13) and the weak convergence

$$\sqrt{n}(\hat{\tau}_n^2 - \tau^2) \xrightarrow{\mathscr{D}} \mathscr{N}(0, 4\sigma^2). \tag{16}$$

Recalling the definition of $\hat{\tau}_n(u, v)$ and $\hat{\tau}_n^2$ in (11) and (12) and using the notation

$$(\omega_{h_1,h_2} * c)(u, v) = \frac{1}{h_1}\int_{[0,1]^2} \omega\left(\frac{u - u_1}{h_1}, \frac{v - v_1}{h_2}\right) c(u_1, v_1)\, du_1\, dv_1$$

we obtain the decomposition

$$\hat{\tau}_n^2 = B_{1n} + 2B_{2n} + B_{3n}, \tag{17}$$

where

$$B_{1n} = \int_{[0,1]^2} (\hat{\tau}_n - \omega_{h_1,h_2} * c)^2(u, v)\, du\, dv,$$
$$B_{2n} = \int_{[0,1]^2} (\hat{\tau}_n - \omega_{h_1,h_2} * c)(u, v) \cdot (\omega_{h_1,h_2} * c)(u, v)\, du\, dv,$$
$$B_{3n} = \int_{[0,1]^2} (\omega_{h_1,h_2} * c)^2(u, v)\, du\, dv.$$

Arguments similar to the ones in Fermanian (2005) show that

$$B_{1n} = O_p\left(\frac{1}{n\sqrt{h_1}}\right) = o_p\left(\frac{1}{\sqrt{n}}\right), \tag{18}$$

while standard arguments (using the differentiability of the copula) yield

$$B_{3n} = \tau^2 + O(h_1^2), \tag{19}$$

$$B_{2n} = \hat{B}_{2n}(1 + o_p(1)), \tag{20}$$

where the quantity $\hat{B}_{2n}$ is defined by

$$\hat{B}_{2n} = \int_{[0,1]^2} (\hat{\tau}_n - \omega_{h_1,h_2} * c)(u, v) \cdot \tau(u, v)\, du\, dv.$$

In addition, we obtain

$$\hat{B}_{2n} - \tilde{B}_{2n} - C_{1n} - C_{2n} = o_P\left(\frac{1}{\sqrt{n}}\right), \tag{21}$$

where the random variables C_{1n} and C_{2n} are defined by

$$C_{1n} = \frac{1}{nh_1^2} \int_0^1 \int_0^1 \sum_{i=1}^n K'\left(\frac{u - F_X(X_i)}{h_1}\right) \bar{K}\left(\frac{v - F_Y(Y_i)}{h_2}\right) \tau(u, v)\, \mathrm{d}u\, \mathrm{d}v \\ \times (\hat{F}_{n1}(X_i) - F_X(X_i)),$$

$$C_{2n} = \frac{1}{nh_1h_2} \int_0^1 \int_0^1 \sum_{i=1}^n K\left(\frac{u - F_X(X_i)}{h_1}\right) K\left(\frac{v - F_Y(Y_i)}{h_2}\right) \tau(u, v)\, \mathrm{d}u\, \mathrm{d}v \\ \times (\hat{F}_{n2}(Y_i) - F_Y(Y_i))$$

and where the statistic $\tilde{B}_{2n}$ is obtained from $\hat{B}_{2n}$ by replacing the empirical distribution function $\hat{F}_{n1}$ and $\hat{F}_{n2}$ by their theoretical counterparts F_X and F_Y, respectively, that is

$$\tilde{B}_{2n} = \int_{[0,1]^2} (\tilde{\tau}_n - \omega_{h_1,h_2} * c)(u, v) \cdot \tau(u, v)\, \mathrm{d}u\, \mathrm{d}v$$

with

$$\tilde{\tau}_n(u, v) = \frac{1}{nh_1} \sum_{i=1}^n \omega\left(\frac{u - F_X(X_i)}{h_1}, \frac{v - F_Y(Y_i)}{h_2}\right). \tag{22}$$

A standard calculation shows that

$$\begin{aligned}\int_{[0,1]^2} \tilde{\tau}_n(u, v)\tau(u, v)\, \mathrm{d}u\, \mathrm{d}v &= \frac{1}{nh_1} \sum_{i=1}^n \int_{[0,1]^2} \omega\left(\frac{u - F_X(X_i)}{h_i}, \frac{v - F_Y(Y_i)}{h_2}\right) \tau(u, v)\, \mathrm{d}u\, \mathrm{d}v \\ &= \frac{1}{n} \sum_{i=1}^n \int_0^1 \tau(F_X(X_i), w) I\{w \geq F_Y(Y_i)\}\, \mathrm{d}w \cdot (1 + o_p(1)) \\ &= \tilde{C}_{0n} \cdot (1 + o_p(1)),\end{aligned} \tag{23}$$

where we have used the assumption (14) and the last line defines the random variable $\tilde{C}_{0n}$ in an obvious manner. By an approximation of a sum of conditional expectations, we obtain

$$\begin{aligned}C_{1n} &= \frac{(1 + o_p(1))}{n^2h_1^2} \sum_{i \neq k} \int_{[0,1]^2} K'\left(\frac{u - F_X(X_i)}{h_1}\right) \bar{K}\left(\frac{v - F_Y(Y_i)}{h_2}\right) \\ &\quad \times (I\{X_k \leq X_i\} - F_X(X_i))\tau(u, v)\, \mathrm{d}u\, \mathrm{d}v \\ &= \frac{(1 + o_p(1))}{nh_1^2} \sum_{k=1}^n \int_{[0,1]^2} E\left[K'\left(\frac{u - F_X(X)}{h_1}\right) \bar{K}\left(\frac{v - F_Y(Y)}{h_2}\right)\right. \\ &\quad \left. \times (I\{X_k \leq X\} - F_X(X)) \,|\, X_k\right] \tau(u, v)\, \mathrm{d}u\, \mathrm{d}v \\ &= \frac{(1 + o_p(1))}{nh_1^2} \sum_{k=1}^n \int_{[0,1]^4} (I\{F_X(X_k) \leq x_1\} - x_1) K'\left(\frac{u - x_1}{h_1}\right) \bar{K}\left(\frac{v - y_1}{h_2}\right) \\ &\quad \times \tau(u, v) c(x_1, y_1)\, \mathrm{d}u\, \mathrm{d}v\, \mathrm{d}x_1\, \mathrm{d}y_1 \\ &= -\frac{(1 + o_p(1))}{n} \sum_{k=1}^n \int_{[0,1]^3} (I\{F_X(X_k) \leq x_1\} - x_1) \bar{K}\left(\frac{v - y_1}{n_2}\right) \\ &\quad \times \partial_1 \tau(x_1, v) c(x_1, y_1)\, \mathrm{d}v\, \mathrm{d}x_1\, \mathrm{d}y_1 \\ &= -\frac{(1 + o_p(1))}{n} \sum_{k=1}^n \int_{[0,1]^2} (I\{F_X(X_k) \leq x_1\} - x_1)\, \partial_1 \tau(x_1, w) \tau(x_1, w)\, \mathrm{d}x_1\, \mathrm{d}w \\ &= (1 + o_p(1)) \cdot \tilde{C}_{1n},\end{aligned}$$

where $(X, Y)\sim F$ is independent of $(X_i, Y_i)_{i=1}^n$ and the last identity defines $\tilde{C}_{1n}$ in an obvious manner. Similar arguments yield

$$C_{2n}=(1+o_p(1))\tilde{C}_{2n}$$

where

$$\tilde{C}_{2n}=\frac{1}{n}\sum_{k=1}^{n}\int\int(I\{F_Y(Y_k)\leq y_1\}-y_1)\partial_2\tau(x_1,y_1)\tau(x_1,y_1)\,\mathrm{d}x_1\,\mathrm{d}y_1.$$

Obviously, $\tilde{C}_{0n}, \tilde{C}_{1n}$ and $\tilde{C}_{2n}$ are of order $O_p(1/\sqrt{n})$ and observing (17), (18) and (19)–(23) now yields

$$\sqrt{n}(\hat{\tau}_n^2-\tau^2)=2\sqrt{n}(\tilde{C}_{0n}-\tau^2+\tilde{C}_{1n}+\tilde{C}_{2n})+o_P(1). \tag{24}$$

The assertion can now be proved by showing the asymptotic normality of

$$\sqrt{n}(\tilde{C}_{0n}-\tau^2+\tilde{C}_{1n}+\tilde{C}_{2n})=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_{ni}-E[X_{ni}]) \tag{25}$$

with

$$\begin{aligned}X_{ni}&=\int_0^1 I\{w\geq F_Y(Y_i)\}\tau(F_X(X_i),w)\,\mathrm{d}w\\&\quad+\int_{[0,1]^2} I\{F_X(X_i)\leq u\}\partial_1\tau(u,v)\tau(u,v)\,\mathrm{d}u\,\mathrm{d}v\\&\quad+\int_{[0,1]^2} I\{F_Y(Y_i)\leq v\}\partial_2\tau(u,v)\tau(u,v)\,\mathrm{d}u\,\mathrm{d}v\\&=X_{ni}^{(1)}+X_{ni}^{(2)}+X_{ni}^{(3)}\end{aligned} \tag{26}$$

$(i=1,\ldots,n)$, where the last identity defines the random variables in an obvious manner. The expectation of $X_{ni}^{(1)}$ is given by

$$E[X_{ni}^{(1)}]=\int_{[0,1]^3} I\{w\geq y_1\}\tau(x_1,w)c(x_1,y_1)\,\mathrm{d}y_1\,\mathrm{d}x_1\,\mathrm{d}w=\int_{[0,1]^2}\tau^2(u,v)\,\mathrm{d}u\,\mathrm{d}v,$$

where we have used the fact that $\int_0^1\int_0^w c(x_1,y_1)\,\mathrm{d}y_1\,\mathrm{d}x_1=\int_0^1\tau(x_1,w)\,\mathrm{d}x_1$. Similarly, the second moment is obtained as

$$E[(X_{n1}^{(1)})^2]=\int_{[0,1]^3}\tau(s,v\wedge\tilde{v})\tau(s,v)\tau(s,\tilde{v})\,\mathrm{d}s\,\mathrm{d}v\,\mathrm{d}\tilde{v},$$

which yields

$$\mathrm{Var}(X_{n1}^{(1)})=\int_{[0,1]^3}\tau(s,v\wedge w)\tau(s,v)\tau(s,w)\,\mathrm{d}s\,\mathrm{d}v\,\mathrm{d}w-\left(\int_{[0,1]^2}\tau^2(u,v)\,\mathrm{d}u\,\mathrm{d}v\right)^2.$$

By an analogous calculation, we have

$$E[X_{ni}^{(2)}]=\frac{1}{2}\int_{[0,1]^2}(\tau^2(1,y_1)-\tau^2(x_1,y_1))\,\mathrm{d}x_1\,\mathrm{d}y_1=\frac{1}{2}\int_{[0,1]^2}y_1\partial_2\tau^2(x_1,y_1)\,\mathrm{d}x_1\,\mathrm{d}y_1,$$

$$\mathrm{Var}(X_{ni}^{(2)})=\frac{1}{4}\int_{[0,1]^4}(x_1\wedge x_2-x_1x_2)\partial_1\tau^2(x_1,y_1)\partial_1\tau^2(x_2,y_2)\,\mathrm{d}x_1\,\mathrm{d}y_1\,\mathrm{d}x_2\,\mathrm{d}y_2$$

and for the covariance it follows by a similar calculation that

$$2\ \mathrm{Cov}(X_{ni}^{(1)},X_{ni}^{(2)})=-\int_{[0,1]^3}\tau^2(x_1,v)\tau^2(x_1,w)\,\mathrm{d}x_1\,\mathrm{d}v\,\mathrm{d}w+\left(\int_{[0,1]^2}\tau^2(x_1,y_1)\,\mathrm{d}x_1\,\mathrm{d}y_1\right)^2.$$

For the remaining variances and covariances, we obtain

$$\begin{aligned}
\mathrm{Var}(X_{ni}^{(3)}) &= \frac{1}{4}\int_{[0,1]^4} (y_1 \wedge y_2 - y_1 y_2)\partial_2\tau^2(x_1,y_1)\partial_2\tau^2(x_2,y_2)\,\mathrm{d}x_1\,\mathrm{d}y_1\,\mathrm{d}x_2\,\mathrm{d}y_2,\\
2\,\mathrm{Cov}(X_{ni}^{(1)}, X_{ni}^{(3)}) &= -\int_{[0,1]^4} \tau^2(u,y_1)\tau(x_1,w)\partial_2\tau(x_1,y_1)I\{y_1 \le \omega\}\,\mathrm{d}x_1\,\mathrm{d}y_1\,\mathrm{d}u\,\mathrm{d}w\\
&\quad + \left(\int_{[0,1]^2} \tau^2(x_1,y_1)\,\mathrm{d}x_1\,\mathrm{d}y_1\right)^4,\\
4\,\mathrm{Cov}(X_{ni}^{(2)}, X_{ni}^{(3)}) &= \int_{[0,1]^4} \tau^2(x_3,y_1)\tau^2(x_2,y_3)\,\mathrm{d}y_1\,\mathrm{d}x_2\,\mathrm{d}x_3\,\mathrm{d}y_3\\
&\quad - \left(\int_{[0,1]^2} \tau^2(x_1,y_1)\,\mathrm{d}x_1\,\mathrm{d}y_1\right)^4.
\end{aligned}$$

This gives for the variance of $X_{ni}\mathrm{Var}(X_{ni}) = \sigma^2$, where σ^2 is defined in (15). The asymptotic normality in (16) now follows from (24) and Ljapunoff's Theorem, which yields

$$\sqrt{n}(\tilde{C}_{0n} + \tilde{C}_{1n} + \tilde{C}_{2n}) \xrightarrow{\mathscr{D}} \mathscr{N}(0, \sigma^2).$$

Remark 4. Note that it follows from lemma 1 that the random variables X and Y are independent if and only if $r = 0$. In this case, we have $\tau(u,v) = v$ and the asymptotic variance in theorem 3 simplifies, substantially. More precisely, one obtains by a straightforward calculation that $\sigma^2 = 52/5$ and by theorem 3 an asymptotic level α test for the hypothesis of independence is obtained by rejecting the null hypothesis whenever

$$|\sqrt{n}\hat{r}_n| > u_{1-\alpha/2}\sqrt{52/5},$$

where $u_{1-\alpha/2}$ denotes the $1-\alpha/2$ quantile of the standard normal distribution.

Remark 5. Theorem 3 can be generalized to dependent data under suitable mixing properties of the data generating process (Fermanian & Scaillet, 2003, assumptions 3 and 4). The details are omitted for the sake of brevity.

Remark 6. Theorem 3 remains correct on subsets of the form $[\delta, 1-\delta]^2 \subset [0,1]^2$. This observation is of importance, because some of the commonly used copula models do not satisfy the assumptions of theorem 3 on the full square $[0,1]^2$. From a practical point of view, the calculation of the measure τ_n^2 on $[\delta, 1-\delta]^2$ for sufficiently small $\delta > 0$ is obviously sufficient.

6. Finite sample properties

6.1. Simulation results

In this section, we present a simulation study of the finite sample properties of the proposed estimate and illustrate its performance in a data example. We begin with the study of the bias, variance and mean squared error of the estimate in the case, where the underlying copula is the Gaussian and the Clayton copula, that is

$$C(u,v) = \Phi_{\theta,2}(\Phi^{-1}(u), \Phi^{-1}(u)); \quad \theta \in [-1,1] \tag{27}$$

$$C(u,v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}; \quad \theta \ge 0, \tag{28}$$

where Φ and $\Phi_{\theta,2}$ denote the cdf of a one-dimensional standard normal and a centred two-dimensional normal distribution function with correlation θ (and variances equal to 1). In

Table 1. *Simulated mean squared error of the estimate* (13), *when the underlying copula is the Clayton copula defined in* (28) *with parameter* θ

$n\backslash\theta$	0.0	0.5	1.0	2.0
50	6.961×10^{-3}	6.981×10^{-3}	8.327×10^{-3}	1.011×10^{-2}
100	2.926×10^{-3}	3.352×10^{-3}	3.774×10^{-3}	5.331×10^{-3}
200	1.341×10^{-3}	1.605×10^{-3}	1.841×10^{-3}	2.401×10^{-3}

Table 2. *Simulated bias of the estimate* (13), *when the underlying copula is the Clayton copula defined in* (28) *with parameter* θ

$n\backslash\theta$	0.0	0.5	1.0	2.0
50	4.071×10^{-2}	4.056×10^{-2}	3.703×10^{-2}	3.805×10^{-2}
100	2.518×10^{-2}	2.497×10^{-2}	2.553×10^{-2}	2.238×10^{-2}
200	1.598×10^{-2}	2.098×10^{-2}	1.439×10^{-2}	1.151×10^{-2}

Table 3. *Simulated variances of the estimate* (13), *when the underlying copula is the Clayton copula defined in* (28) *with parameter* θ

$n\backslash\theta$	0.0	0.5	1.0	2.0
50	5.303×10^{-3}	5.344×10^{-3}	6.956×10^{-3}	9.967×10^{-3}
100	2.294×10^{-3}	2.769×10^{-3}	3.123×10^{-3}	4.475×10^{-3}
200	1.093×10^{-3}	1.348×10^{-3}	1.633×10^{-3}	2.269×10^{-3}

order to address the problem of boundary effects in the statistic $\hat{\tau}_n$, we have adapted the estimate investigated recently in Chen & Huang (2007) and Omelka *et al.* (2009) to our problem. To be precise, we have used the statistic

$$\hat{\tau}_n^{(\mathrm{LLS})}(u,v)=\frac{1}{b(u)h_1n}\sum_{i=1}^{n}K_{u,h_1}\left(\frac{u-\hat{F}_{m1}(X_i)}{b(u)h_1}\right)\bar{K}_{v,h_2}\left(\frac{v-\hat{F}_{n2}(Y_i)}{b(v)h_2}\right),$$

as an estimate of $\tau(u,v)=\partial_1C(u,v)$ where $b(w)=\min\{\sqrt{w},\sqrt{1-w}\}$ and the kernel K_{u,h_1} is defined by

$$K_{u,h_1}(x)=\frac{K(x)\{a_2(u,h_1)-a_1(u,h_1)x\}}{a_0(u,h_1)a_2(u,h_1)-a_1^2(u,h_1)}I\left\{\frac{u-1}{h_1}<x<\frac{u}{h_1}\right\}$$

with

$$a_\ell(u,h_1)=\int_{(u-1)/h_1}^{u/h_1}t^\ell K(t)\,\mathrm{d}t;\quad \ell=0,1,2.$$

Note that $\hat{\tau}_n^{(\mathrm{LLS})}$ is a local linear estimate, where the bandwidth function 'shrinks' the value of the bandwidth close to zero at the corners of the unit square. The estimate $\hat{\tau}_n^2$ is constructed by (12) replacing $\hat{\tau}_n$ by $\hat{\tau}_n^{(\mathrm{LLS})}$, and similar arguments as in Chen & Huang (2007) and Omelka *et al.* (2009) show that theorem 5.1(a) remains valid.

In Tables 1, 2 and 3 we present the simulated mean squared error, bias and variances of the estimate for the sample sizes 50, 100 and 200 for the Clayton copula. These results are based on 25,000 simulation runs. The bandwidth is chosen as $h=n^{-3/10}$ and the integral in the definition of the estimate is calculated over a grid of 49×49 points. The random variables distributed according to the Clayton copula are generated by the method presented in Cook & Johnson (1981). We observe that in all cases the measure r is estimated with reasonable precision. It is worthwhile mentioning that the mean squared error is increasing with the parameter θ and that the estimate is less accurate if $\theta=2.0$ (see Table 1). Investigating the bias and variance, we observe that the contribution of the variance to the mean squared error is larger than the contribution of the squared bias and this effect is increasing with θ. For example, if $n=200$ the variance contributes 81% to the mean squared error in the case $\theta=0$, while its contribution is 95% in the case $\theta=2$. It is also notable that the bias is relatively stable with respect to θ, while the variance is increasing with θ.

The corresponding simulation results for the Gaussian copula in (27) are displayed in Tables 4, 5 and 6 and show a similar picture. If the correlation coefficient is smaller than 0.9, the main part of the mean squared error can be explained by the variance (between 65% and 90%). On the other hand, if the correlation is given by $\theta=0.9$ the contribution of the bias is larger (up to 75%). Note also that the estimates are less accurate if θ is increasing.

In the second part of our numerical study we investigate the approximation by the normal distribution for moderate sample sizes. In order to obtain a good approximation by the limit distribution, it is important to have a precise estimate of the limiting variance. For this purpose, we propose an estimate which is motivated by a careful inspection of the proof of theorem 3. To be precise, note that by (24) the statistic $\sqrt{n}(\hat{\tau}_n^2-\tau^2)$ is asymptotically equivalent to a sum of i.i.d. random variables defined by (25), in particular

$$\mathrm{Var}(\sqrt{n}\hat{\tau}_n)\approx 4\,\mathrm{Var}(X_{ni}),$$

where X_{ni} is defined in (26). Therefore, we use the empirical variance of the random variables

$$\begin{aligned}
V_{ni}=&\frac{1}{h_1}\int_{[0,1]^2}\omega\left(\frac{u-\hat{F}_{n1}(X_i)}{h_1},\frac{v-\hat{F}_{n2}(Y_i)}{h_2}\right)\hat{\tau}_n(u,v)\,\mathrm{d}u\,\mathrm{d}v\\
&+\frac{1}{nh_1^2}\sum_{k\neq i}\int_{[0,1]^2}K'\left(\frac{u-\hat{F}_{n1}(X_k)}{h_1}\right)\bar{K}\left(\frac{v-\hat{F}_{n2}(Y_i)}{h_2}\right)\\
&\times(I\{X_i\leq X_k\}-\hat{F}_{n1}(X_k))\hat{\tau}_n(u,v)\,\mathrm{d}u\,\mathrm{d}v\\
&+\frac{1}{nh_1h_2}\sum_{k\neq i}\int_{[0,1]^2}K\left(\frac{u-\hat{F}_{n1}(X_k)}{h_1}\right)K\left(\frac{v-\hat{F}_{n2}(Y_k)}{h_2}\right)\\
&\times(I\{X_i\leq X_k\}-\hat{F}_{n2}(Y_k))\hat{\tau}_n(u,v)\,\mathrm{d}u\,\mathrm{d}v
\end{aligned}\tag{29}$$

as an estimate for the asymptotic variance of $\hat{\tau}_n^{(\mathrm{LSS})}$, that is

$$\hat{\sigma}_n^2=\frac{4}{n}\sum_{i=1}^{n}(V_{ni}-\overline{V}_{n\cdot})^2.\tag{30}$$

In Table 7, we show the simulated variances of the statistic $\sqrt{n}\tau_n$ and the corresponding values for the estimate $\hat{\sigma}_n^2$ based on 1000 simulation runs. The underlying copula is the Clayton copula and we observe a reasonable performance of the estimate (30) if the sample size is larger than 100.

Table 4. *Simulated mean squared error of the estimate* (13), *when the underlying copula is a Gaussian copula defined in* (27) *with correlation* θ

$n\backslash\theta$	0.0	0.3	0.6	0.9
50	5.406×10^{-3}	5.855×10^{-3}	8.443×10^{-3}	1.089×10^{-2}
100	2.326×10^{-3}	2.603×10^{-3}	3.959×10^{-3}	5.554×10^{-3}
200	1.183×10^{-3}	1.214×10^{-3}	1.692×10^{-3}	2.471×10^{-3}

Table 5. *Simulated bias of the estimate* (13), *when the underlying copula is a Gaussian copula defined in* (27) *with correlation* θ

$n\backslash\theta$	0.0	0.3	0.6	0.9
50	4.296×10^{-2}	4.281×10^{-2}	4.205×10^{-2}	9.318×10^{-2}
100	2.431×10^{-2}	2.477×10^{-2}	2.441×10^{-2}	6.425×10^{-2}
200	1.513×10^{-2}	1.305×10^{-2}	1.389×10^{-2}	4.813×10^{-2}

Table 6. *Simulated variance of the estimate* (13), *when the underlying copula is a Gaussian copula defined in* (27) *with correlation* θ

$n\backslash\theta$	0.0	0.3	0.6	0.9
50	3.561×10^{-3}	4.031×10^{-3}	6.675×10^{-3}	2.207×10^{-3}
100	1.735×10^{-3}	1.989×10^{-3}	3.363×10^{-3}	1.298×10^{-3}
200	9.537×10^{-4}	1.044×10^{-3}	1.499×10^{-3}	1.545×10^{-4}

Table 7. *Simulated variances of the statistics* $\sqrt{n}\hat{\tau}_n$ *and* $\hat{\sigma}_n^2$ *defined in* (12) *and* (30), *where the underlying copula is the Clayton copula*

	$\theta=0.0$		$\theta=0.5$		$\theta=1.0$		$\theta=2.0$	
n	$\sqrt{n}\hat{\tau}_n$	$\hat{\sigma}_n^2$	$\sqrt{n}\hat{\tau}_n$	$\hat{\sigma}_n^2$	$\sqrt{n}\hat{\tau}_n$	$\hat{\sigma}_n^2$	$\sqrt{n}\hat{\tau}_n$	$\hat{\sigma}_n^2$
50	0.616	0.333	0.523	0.361	0.611	0.483	0.847	0.626
100	0.301	0.279	0.351	0.317	0.417	0.424	0.536	0.527
200	0.289	0.243	0.314	0.287	0.349	0.357	0.514	0.525

In Tables 8 and 9, we show the simulated probabilities

$$P\left(\frac{\sqrt{n}(\hat{\tau}_n-\tau)}{6\hat{\sigma}_n}\leq u_{1-\alpha}\right)\approx 1-\alpha, \tag{31}$$

where $u_{1-\alpha}$ denotes the $(1-\alpha)$-quantile of the standard normal distribution. The sample is $n=100$ and $n=200$, the bandwidth is again chosen as $h=n^{-3/10}$ and the underlying copulas are the Clayton and Gaussian copula defined in (28) and (27). For the Clayton copula, we observe a reasonable approximation by the limit distribution in all cases under consideration (see Table 8). For the Gaussian copula the results are similar, but the approximation is less accurate in the case $\theta=0.9$ of strong correlation (see Table 9).

Table 8. *Simulated probabilities of the form* (31) *for the Clayton copula*

$1-\alpha$	$n\backslash\theta$	0.0	0.5	1.0	2.0
90%	100	0.843	0.876	0.895	0.904
	200	0.881	0.879	0.889	0.905
95%	100	0.932	0.948	0.961	0.959
	200	0.931	0.942	0.951	0.960

Table 9. *Simulated probabilities of the form* (31) *for the Gaussian copula*

$1-\alpha$	$n\backslash\theta$	0.0	0.3	0.6	0.9
90%	100	0.864	0.871	0.881	0.825
	200	0.881	0.885	0.889	0.831
95%	100	0.931	0.937	0.947	0.891
	200	0.939	0.941	0.955	0.921

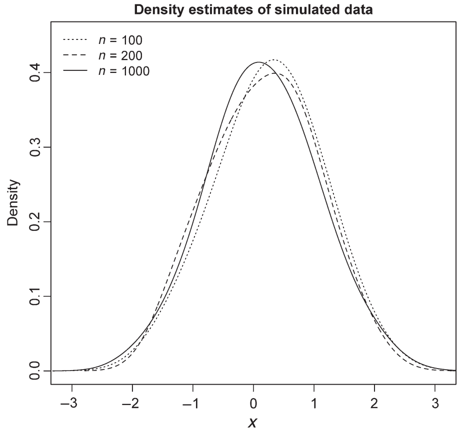


Fig. 3. Kernel density estimates of 1000 simulated replications of the statistic $\sqrt{n}(\hat{\tau}_n - \tau)/(6\hat{\sigma}_n)$ for the Clayton copula with parameter $\theta = 1$. The sample size is $n = 100$ (dotted line), $n = 200$ (dashed line) and $n = 1000$ (solid line).

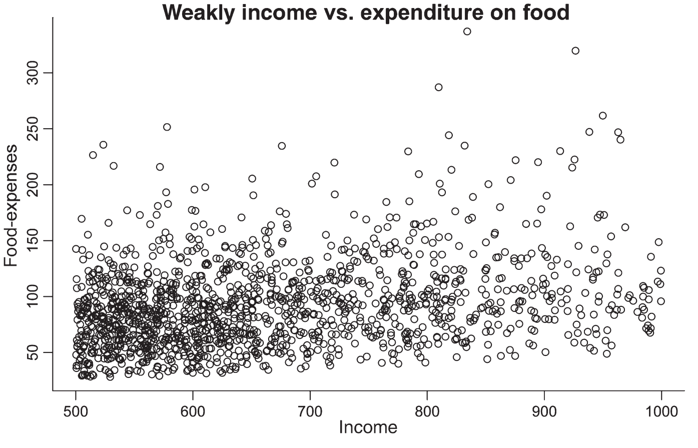


Fig. 4. Scatterplot of the household data.

Kernel density estimates of 1000 simulated values of the statistic $\sqrt{n}(\hat{\tau}_n - \tau)/6\hat{\sigma}_n$ are shown in Fig. 3 for sample sizes $n = 100, 200$ and 1000 where the underlying copula is the Clayton copula with parameter $\theta = 1$.

6.2. *Data example*

In the following paragraph, we provide a brief empirical example investigating the relationship between disposable income of households and the expenditure on food. It is widely accepted that this relation is highly nonlinear and the usual regression model with the usual coefficient of determination cannot be applied. In particular, we analyze UK data from the Family Expenditure Survey, 2000–2001provided by the Office for National Statistics and distributed by the UK Data Archive. The data consists of pairs of weekly disposable household income and expenditure on food. In order to concentrate on a homogeneous region, we consider data with a weekly disposable household income between 500 and 1000 British pounds. After eliminating obviously incorrect or incomplete records, our data base covers a total number of 1438 observations. A scatter plot of the normal weekly disposable household income and weekly expenditure on food is shown in Fig. 4. The Pearson correlation coefficient, Kendall's tau and Spearman's rho for these data are given by 0.278, 0.179 and 0.267, respectively. The new measure of regression was calculated with a bandwidth $h = 0.023$ (chosen by cross-validation) and was given by $\hat{r} = 0.142$. There results indicate a rather weak dependence between the two variables for the data set under consideration.

Acknowledgements

Financial support by the Deutsche Forschungsgemeinschaft (SFB 823, Teilprojekt A1) is gratefully acknowledged. We thank two anonymous referees for their constructive comments, which improved the article. We are also grateful to the Office for National Statistics and the UK Data Archive for providing the data (Family Expenditure Survey, 2000–2001) used in section 6.2.

References

Chen, S. X. & Huang, T. M. (2007). Nonparametric estimation of copula functions for dependence modelling. *Canad. J. Statist.* **35**, 265–282.

Cook, D. & Johnson, E. (1981). A family of distributions for modelling non-elliptically symmetric multivariate data. *J. Roy. Statist. Soc. B* **43**, 210–218.

Dabrowska, D. (1981). Regression-based orderings and measures of stochastic dependence. *Math. Operationsforsch. Statist. Ser. Statist.* **12**, 317–325.

Dąbrowska, D. (1985). Descriptive parameters of location, dispersion and stochastic dependence. *Statistics* **16**, 63–88.

Darsow, W. F., Nguyen, B. & Olsen, E. T. (1992). Copulas and Markov processes. *Illinois J. Math.* **36**, 600–642.

Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés. *Acad. R. Belg., Bull. Cl. Sci.* **65**, 274–292.

Esary, J., Proschan, F. & Walkup, D. (1967). Association of random variables with applications. *Ann. Math. Statist.* **38**, 1466–1474.

Evans, L. C. (1998). *Partial differential equations*. American Mathematical Society, Providence, RI.

Fermanian, J. D. (2005). Goodness-of-fit-tests for copulas. *J. Multivariate Anal.* **95**, 119–152.

Fermanian, J. D. & Scaillet, O. (2003). Nonparametric estimation of copulas for time series. *J. Risk* **5**, 25–54.

Fermanian, J. D., Radulović, D. & Wegkamp, M. (2004). Weak convergence of empirical copula processes. *Bernoulli* **10**, 847–860.

Hall, W. J. (1970). On characterizing dependence in joint distributions. In *Essays in probability and statistics* (eds R. C. Bose, I. M. Chakravarti, P. C. Mahalanobis, C. R. Rao & K. J. C. Smith), 339–376. Univ. of North Carolina Press, Chapel Hill, NC.

Hewitt, E. & Stromberg, K. (1975). *Real and abstract analysis*. Springer-Verlag, New York.

Kimeldorf, G. & Sampson, A. R. (1987). Positive dependence orderings. *Ann. Inst. Statist. Math.* **39**, 113–128.

Lancaster, H. O. (1963). Correlation and complete dependence of random variables. *Ann. Math. Statist.* **34**, 1315–1321.

Lehmann, E. L. (1966). Some concepts of dependence. *Ann. Math. Statist.* **37**, 1137–1153.

Müller, A. & Stoyan, D. (2002). *Comparison methods for stochastic models and risks*. John Wiley & Sons Ltd., Chichester.

Nelsen, R. B. (2006). *An introduction to copulas*, 2nd edn. Springer, New York.

Omelka, M., Gijbels, I. & Veraverbeke, N. (2009). Improved kernel estimation of copulas: weak convergence and goodness-of-fit testing. *Ann. Statist.* **37**, 3023–3058.

Ramos, H. M. & Sordo, M. A. (2003). Dispersion measures and dispersive orderings. *Statist. Probab. Lett.* **61**, 123–131.

Rényi, A. (1959). On measures of dependence. *Acta Math. Acad. Sci. Hungar.* **10**, 441–451.

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

Scarsini, M. & Shaked, M. (1996). Positive dependence orders: a survey. In *Athens conference on applied probability and time series analysis*, Vol. I (1995) (eds C. C. Heyde, Y. V. Prohorov, R. Pyke & S. T. Rachev), 70–91. Springer, New York.

Schriever, B. F. (1987). An ordering for positive dependence. *Ann. Statist.* **15**, 1208–1214.

Schweizer, B. & Wolff, E. F. (1981). On nonparametric measures of dependence for random variables. *Ann. Statist.* **9**, 879–885.

Shaked, M. & Shanthikumar, J. G. (2007). *Stochastic orders*. Springer, New York.

Siburg, K. F. & Stoimenov, P. A. (2010). A measure of mutual complete dependence. *Metrika* **71**, 239–251.

Received 28 February 2011, in final form 21 September 2011

Pavel A. Stoimenov, Fakultät Statistik, Technische Universität Dortmund, Vogelpothsweg 78, 44227 Dortmund, Germany.
E-mail: pavel.stoimenov@gmail.com