SFB 823

Regression dependence

Karl Friedrich Siburg, Pavel A. Stoimenov

Nr. 10/2009

Discussion Paper

REGRESSION DEPENDENCE

By Karl Friedrich Siburg, and Pavel A. Stoimenov*

Technische Universität Dortmund, Germany

This paper 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) and provide two examples of RDOs. In addition, we define a new nonparametric measure of regression dependence and study its properties. Beside being monotone in the new RDOs, the measure takes on its extreme values precisely at independence and almost sure functional dependence, respectively. Finally, a consistent nonparametric estimator of the new measure is constructed.

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 [13], Esary et al. [5] and Schriever [20]; see also Scarsini and Shaked [19] for a detailed survey. Axiomatic approaches to orders and measures of positive dependence were introduced by Kimeldorf and Sampson [11] and Scarsini [18], respectively. The abundance of notions of positive dependence contrasts, however, with the silence concerning regression dependence, with the exception of the work of Dąbrowska [1, 2] and the measure suggested by Hall [9].

This paper 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 random vector with marginal distribution functions F_X and F_Y, respectively, and joint distribution function $F_{X,Y}$. Since regression

*The financial support by the Deutsche Forschungsgemeinschaft (SFB 823, Statistik nichtlinearer dynamischer Prozesse) is gratefully acknowledged.

AMS 2000 subject classifications: Primary 62E10; secondary 62J02

Keywords and phrases: stochastic order, regression, measure of dependence, copula

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; see Lancaster [12]. 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 $(\widetilde{X}, \widetilde{Y})$ is another pair of random variables, then the general idea is to consider (X, Y) less regression dependent than $(\widetilde{X}, \widetilde{Y})$ if $F_{Y|X=x}$ is more variable than $F_{\widetilde{Y}|\widetilde{X}=x}$ 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 $\widetilde{X}$ have the same distribution. Moreover, it is even necessary that Y and $\widetilde{Y}$ are identically distributed because, otherwise, their different variability will affect the variability of $F_{Y|X}$ and $F_{\widetilde{Y}|\widetilde{X}}$ 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 $(\widetilde{U}, \widetilde{V}) = (F_{\widetilde{X}}(\widetilde{X}), F_{\widetilde{Y}}(\widetilde{Y}))$, which have uniform marginal distributions. In this case, we regard (X, Y) less regression dependent than $(\widetilde{X}, \widetilde{Y})$ if $F_{V|U=u}$ is more variable than $F_{\widetilde{V}|\widetilde{U}=u}$ 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 nonparametric measure of regression dependence, study its properties and demonstrate its advantages over the correlation ratio. Beside 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.

A sample version of the new dependence measure is addressed in the final Section 5 where we construct a consistent estimator.

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

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

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 $F_{V|U=u}$ is clear.

The first result describes the two extreme cases of regression dependence for (X,Y) in terms of (U,V).

PROPOSITION 2.1. *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.,

$$1 = \int_{I^2} \mathbf{1}_{\operatorname{gr} f}(u,v)\, dF_{U,V}(u,v) = \int_I \int_I \mathbf{1}_{\operatorname{gr} f}(u,v)\, dF_{V|U=u}(v)\, dF_U(u).$$

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 i-th variable. The following properties of copulas are easy consequences of the definition; for a proof see, e.g., [15].

Proposition 2.2. *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 2.3. (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; see [6]. Moreover, by [6, Thm. 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

$$F_{V|U=u}(v) = \partial_1 C_{X,Y}(u,v) \tag{2.1}$$

wherever the partial derivative exists; see [15]. Moreover, we have the following result related to Proposition 2.1.

PROPOSITION 2.4. *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 2.3 (i), while the second is a consequence of [3, Thm. 11.1] and [22, Thm. 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 [14] and [21] 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

$$\mathrm{E}[\phi(X)] \leq \mathrm{E}[\phi(Y)] \tag{2.2}$$

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 (2.2) holds, Y is more variable (or more dispersed) than X. The next result is a direct consequence of (2.2).

PROPOSITION 2.5. *Let X and Y be two random variables. If $X \leq_{cx} Y$, then $E[X] = E[Y]$ and $Var[X] \leq Var[Y]$.*

As can be seen from Proposition 2.5, 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

$$X - \mathrm{E}[X] \leq_{\mathrm{cx}} Y - \mathrm{E}[Y]. \tag{2.3}$$

COROLLARY 2.6. *Let X and Y be two random variables. If $X \leq_{dil} Y$, then $Var[X] \leq Var[Y]$.*

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

$$F_X^{-1}(b) - F_X^{-1}(a) \leq F_Y^{-1}(b) - F_Y^{-1}(a) \tag{2.4}$$

for all $0 < a \leq b < 1$. As noted in [21], 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 $\leq_{\text{disp}}$ and $\leq_{\text{dil}}$; compare [21, Thm. 3.B.16].

PROPOSITION 2.7. *Let X and Y be two random variables with finite expectations. Then $X \leq_{disp} Y$ implies $X \leq_{dil} Y$.*

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 $(\widetilde{X}, \widetilde{Y}) \in \mathfrak{X}$ if $F_{\widetilde{V}|\widetilde{U}=u}$ 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 3.1. *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) $(X, Y) \preccurlyeq (\widetilde{X}, \widetilde{Y})$ implies $F_{\widetilde{V}|\widetilde{U}=u} \leq_\bullet F_{V|U=u}$ 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 $(X, Y) \preccurlyeq (\widetilde{X}, \widetilde{Y})$, then $\widetilde{Y}$ is a.s. a Borel function of $\widetilde{X}$;

(O3) If X and Y are independent, and if $(\widetilde{X}, \widetilde{Y}) \preccurlyeq (X, Y)$, then $\widetilde{X}$ and $\widetilde{Y}$ 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 $(X,Y) \preccurlyeq (\widetilde{X},\widetilde{Y})$ if and only if $F_{\widetilde{V}|\widetilde{U}=u} \leq_\bullet F_{V|U=u}$ 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 2.1, 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 $\leq_{\mathrm{cx}}$) 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 $\leq_{\mathrm{dil}}$ and the dispersive order $\leq_{\mathrm{disp}}$ are the most important and common ones. The next result states that the corresponding relations $\preccurlyeq_{\mathrm{dil}}$ and $\preccurlyeq_{\mathrm{disp}}$ are indeed RDOs.

THEOREM 3.2. *The relations $\preccurlyeq_{dil}$ and $\preccurlyeq_{disp}$ are RDOs.*

PROOF. In view of Proposition 2.7 we need only prove $(O2)$ and $(O3)$ for the relation $\preccurlyeq_{\mathrm{dil}}$. It is clear from Corollary 2.6 that $\preccurlyeq_{\mathrm{dil}}$ satisfies $(O2)$. In order to prove $(O3)$ we may, in view of Proposition 2.1, restrict to considering U and V instead of X and Y. Assuming that $(\widetilde{U},\widetilde{V}) \preccurlyeq_{\mathrm{dil}} (U,V)$ with independent U and V, we conclude from Corollary 2.6 that

$$\mathrm{Var}[\widetilde{V}|\widetilde{U}=u] \geq \frac{1}{12} \tag{3.1}$$

for almost all u. By the law of total variance, we obtain equality in (3.1), as well as

$$\mathrm{E}[\widetilde{V}|\widetilde{U}=u] = \mathrm{E}[\widetilde{V}] = \frac{1}{2} \tag{3.2}$$

for almost all u. From (3.2) and (3.1) it follows that, for almost all u, $F_{V|U=u} \leq_{\mathrm{cx}} F_{\widetilde{V}|\widetilde{U}=u}$ with equal variances. But then both distributions are the same; see [21, Thm. 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 4.1. *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) $(X,Y) \preccurlyeq (\widetilde{X},\widetilde{Y})$ implies $\mu(X,Y) \le \mu(\widetilde{X},\widetilde{Y})$;
(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 4.2. 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 [17], 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 4.1 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 nonparametric MRD. The following is the main result in this section.

THEOREM 4.3. *The function $r : \mathfrak{X} \to [0,1]$ defined by*

$$r(X,Y) = 6 \int_0^1 \int_0^1 F_{V|U=u}(v)^2 \, dv \, du - 2 \tag{4.1}$$

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

REMARK 4.4. Note that in view of (2.1), we have

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

where $\|\cdot\|_2$ denotes the L^2-norm on I^2. By Remark 2.3 (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 4.3 we make use of the following result.

LEMMA 4.5. *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

$$0 \le \int_0^1 \int_0^1 (\partial_1 C_{X,Y}(u,v) - v)^2 \, du \, dv = \int_0^1 \int_0^1 (\partial_1 C_{X,Y}(u,v))^2 \, du \, dv - \frac{1}{3}.$$

Hence, $\|\partial_1 C_{X,Y}\|_2^2 \ge 1/3$ with equality if and only if $\partial_1 C_{X,Y}(u,v) = v$ Lebesgue a.e. on I^2, which by Proposition 2.4 (i) is equivalent to the independence of X and Y.

(ii) By Theorem 2.2 (ii) we have $0 \le \partial_1 C_{X,Y} \le 1$ and thus $(\partial_1 C_{X,Y})^2 \le \partial_1 C_{X,Y}$, with equality if and only if $\partial_1 C_{X,Y} \in \{0, 1\}$. Consequently,

$$\|\partial_1 C_{X,Y}\|_2^2 \le \int_0^1 \int_0^1 \partial_1 C_{X,Y}(u,v) \, du \, dv = \frac{1}{2}$$

with equality if and only if $\partial_1 C_{X,Y} \in \{0, 1\}$ Lebesgue a.e. in I^2, which by Proposition 2.4 (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.

LEMMA 4.6. *Let F be a univariate distribution function with support in I. Then*

$$2 \int_0^1 \int_0^p F^{-1}(t) \, dt \, dp - \int_0^1 F^{-1}(t) \, dt = \int_0^1 F(v)^2 \, dv - \int_0^1 F(v) \, dv.$$

PROOF. Using integration by parts for Lebesgue-Stieltjes integrals (see, e.g., [10, Thm. 21.67]) we obtain

$$\begin{aligned}
\int_0^1 & F(v)^2\,dv - \int_0^1 F(v)\,dv \\
&= \int_0^1 (t^2 - t)\,dF^{-1}(t) \\
&= -2\int_0^1 tF^{-1}(t)\,dt + \int_0^1 F^{-1}(t)\,dt \\
&= -2\Big(\int_0^1 F^{-1}(t)\,dt - \int_0^1\int_0^t F^{-1}(s)\,ds\,dt\Big) + \int_0^1 F^{-1}(t)\,dt \\
&= 2\int_0^1\int_0^p F^{-1}(t)\,dt\,dp - \int_0^1 F^{-1}(t)\,dt.
\end{aligned}$$

□

We now turn to the proof of the theorem.

PROOF OF THEOREM 4.3. The property $0 \le r(X,Y) \le 1$, as well as the conditions $(M2)$ and $(M3)$, are immediately implied by Lemma 4.5.

It remains to show the monotonicity condition $(M1)$; in view of Proposition 2.7, it suffices to prove it for the RDO $\preccurlyeq_{\text{dil}}$. Ramos and Sordo showed in [16, Thm. 2.1] that two univariate distribution functions F and G with finite expectations satisfy $F \le_{\text{dil}} G$ if and only if, for all $v \in [0,1]$,

$$\int_0^v F^{-1}(t)\,dt - v\int_0^1 F^{-1}(t)\,dt \ge \int_0^v G^{-1}(t)\,dt - v\int_0^1 G^{-1}(t)\,dt\,. \tag{4.2}$$

Now assume that $(X,Y) \preccurlyeq_{\text{dil}} (\widetilde{X},\widetilde{Y})$ so that $F_{\widetilde{V}|\widetilde{U}=u} \le_{\text{dil}} F_{V|U=u}$ for almost all $u \in I$. Then, integrating (4.2) over v we obtain

$$\begin{aligned}
\int_0^1\int_0^v F^{-1}_{\widetilde{V}|\widetilde{U}=u}(t)\,dt\,dv - \frac{1}{2}\int_0^1 F^{-1}_{\widetilde{V}|\widetilde{U}=u}(t)\,dt \ge \\
\int_0^1\int_0^v F^{-1}_{V|U=u}(t)\,dt\,dv - \frac{1}{2}\int_0^1 F^{-1}_{V|U=u}(t)\,dt
\end{aligned}$$

for almost all $u \in I$. Applying Lemma 4.6 we find that, for almost all u,

$$\begin{aligned}
\int_0^1 F_{\widetilde{V}|\widetilde{U}=u}(v)^2\,dv - \int_0^1 F_{\widetilde{V}|\widetilde{U}=u}(v)\,dv \ge \\
\int_0^1 F_{V|U=u}(v)^2 dv - \int_0^1 F_{V|U=u}(v)\,dv\,.
\end{aligned}$$

Integrating this over $u \in I$, substituting $\partial_1 C_{X,Y}(u,v)$ for $F_{V|U=u}(v)$ by (2.1), and using $\int_0^1\int_0^1 \partial_1 C_{X,Y}(u,v)\,dv\,du = 1/2$ for all $C_{X,Y} \in \mathfrak{C}$, we obtain

$$\|\partial_1 C_{\widetilde{X},\widetilde{Y}}\|_2^2 = \int_0^1\int_0^1 (\partial_1 C_{\widetilde{X},\widetilde{Y}}(u,v))^2\,dv\,du \geq \int_0^1\int_0^1 (\partial_1 C_{X,Y}(u,v))^2\,dv\,du = \|\partial_1 C_{X,Y}\|_2^2.$$

Since, by Remark 4.4, $r(X,Y) = 6\|\partial_1 C_{X,Y}\|_2^2 - 2$, this proves $(M1)$ and hence the theorem. □

Proposition 4.7. *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 [15, 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

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

see [15, Theorem 2.4.4]. 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. □

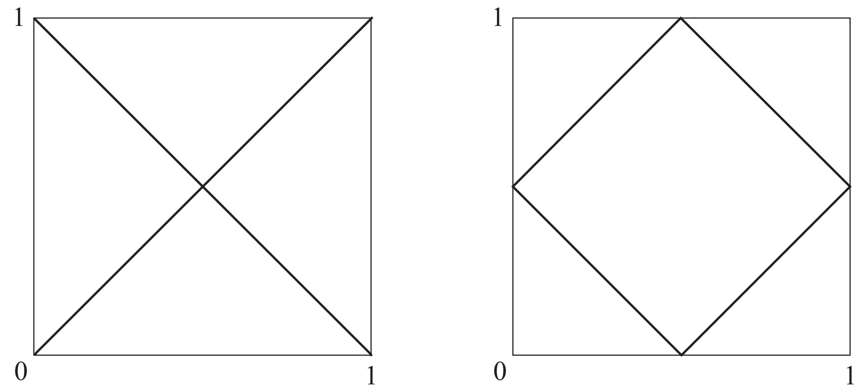
We now turn attention to another quantity that might seem a natural choice for an MRD, namely the correlation ratio of the probability integral transformations. Define the functional $\widetilde{\eta} : \mathfrak{X} \to \mathbb{R}$ by

$$\widetilde{\eta}(X,Y)^2 := \eta(U,V)^2 = \frac{\mathrm{Var}[\mathrm{E}[V|U]]}{\mathrm{Var}[V]} = 1 - \frac{\mathrm{E}[\mathrm{Var}[V|U]]}{\mathrm{Var}[V]}\,. \tag{4.3}$$

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

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

In fact, the ordering of regression dependence suggested in [1, Sec. 3.1] is an ordering by correlation ratios and therefore is not consistent with



(a) Support of the distribution of (U, V) in Example 4.8

(b) Support of the distribution of (U, V) in Example 4.9

FIG 1. *Examples of $\widetilde{\eta}(X, Y) = 0$ where X and Y are not independent*

our approach to RDOs. Moreover, neither the correlation ratio of Y on X nor the related measure $\widetilde{\eta}(X, Y)^2$ are MRDs in the sense of Definition 4.1, because $(M3)$ will not be satisfied. Indeed, it follows from Propositions 2.7 and 2.5 that $\widetilde{\eta}$ is monotone with respect to both $\preccurlyeq_{\text{disp}}$ and $\preccurlyeq_{\text{dil}}$; in addition, $\widetilde{\eta}(X, Y) = 1$ if and only if Y is a.s. a Borel function of X. However, $\widetilde{\eta}$ does not satisfy condition $(M3)$ because there are random variables X and Y with $\widetilde{\eta}(X, Y) = 0$, which are not independent; we give two such examples.

EXAMPLE 4.8. Consider X and Y whose probability integral transformations U and V have the singular distribution with the support depicted in Figure 1(a). 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, $E[V|U = u] = 1/2$. Consequently, the conditional expectation $E[V|U]$ is degenerate and its variance $\text{Var}[\text{E}[V|U]]$ vanishes, which means that $\eta(U, V) = \widetilde{\eta}(X, Y) = 0$. However, U and V and, thus, X and Y are not independent.

EXAMPLE 4.9. Another situation where $\widetilde{\eta}(X, Y) = 0$ 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 $\widetilde{\eta}(X, Y)$ since in this case $F_{U,V}$ is a degenerate distribution whose support is given in Figure 1(b); see [15, Sec. 3.1.2]. The arguments are analogous to those in the previous example.

5. Nonparametric estimation of r. In this final section, we present a sample version of the MRD

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

As pointed out in Remark 4.4, r is a function of the copula $C_{X,Y}$ alone. $C_{X,Y}$ can be consistently estimated by the empirical copula; see Deheuvels [4] and Fermanian et al. [7]. 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 smooth (differentiable) estimation of the copula, e.g., with a kernel-based technique.

In the following we present the nonparametric approach given in Fermanian and Scaillet [8], who introduced a kernel estimator of the partial derivative of a copula, which provides a basis for developing a sample analogue of r. Assuming that the joint distribution $F_{X,Y}$ of $(X,Y) \in \mathfrak{X}$ is absolutely continuous, the marginal density f_X at x will be estimated from a sample of size N by

$$\widehat{f_X}(x) = \frac{1}{Nh_X}\sum_{i=1}^{N} k_X\left(\frac{x - X_i}{h_X}\right)\,,$$

where the kernel $k_X : \mathbb{R} \to \mathbb{R}$ is symmetric and bounded with $\int_{-\infty}^{\infty} k_X(x)\,dx = 1$ and the bandwidth h_X is a positive function of N such that $h_X \to 0$ when $N \to \infty$. The estimator $\widehat{f_Y}$ of f_Y with corresponding kernel k_Y and bandwidth h_Y is defined analogously. In order to estimate the joint density $f_{X,Y}$ we use a two-dimensional kernel which, for the sake of simplicity, is the product of k_X and k_Y. Thus we have

$$\widehat{f_{X,Y}}(x,y) = \frac{1}{Nh_Xh_Y}\sum_{i=1}^{N} k_X\left(\frac{x - X_i}{h_X}\right) k_Y\left(\frac{y - Y_i}{h_Y}\right)\,.$$

Hence, an estimator of F_X at some point x is obtained as

$$\widehat{F_X}(x) = \int_{-\infty}^{x} \widehat{f_X}(s)\,ds\,,$$

while an estimator of the joint distribution function $F_{X,Y}$ is obtained as

$$\widehat{F_{X,Y}}(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} \widehat{f_{X,Y}}(s,t)\,ds\,dt\,.$$

Fermanian and Scaillet [8] suggest to estimate $C_{X,Y}$ by a simple plug-in method, which in view of the identity

$$C_{X,Y}(u,v) = F_{X,Y}(F_X^{-1}(u), F_Y^{-1}(v))\,,$$

yields

$$\widehat{C_{X,Y}}(u,v) = \widehat{F_{X,Y}}(\widehat{F_X}^{-1}(u), \widehat{F_Y}^{-1}(v))\,, \tag{5.1}$$

where $\widehat{F_X}^{-1}$ and $\widehat{F_Y}^{-1}$ stand for the respective quantile functions of $\widehat{F_X}$ and $\widehat{F_Y}$. Under some technical conditions, the authors establish the asymptotic normality of the estimator defined in (5.1).

Moreover, the kernel-based approach has the advantage of providing a smooth (differentiable) reconstruction of $C_{X,Y}$. Thus, it is natural to consider an estimator of $\partial_1 C_{X,Y}$ based on the differentiation of $\widehat{C_{X,Y}}$ with respect to the first variable:

$$\widehat{\partial_1 C_{X,Y}}(u,v) = \partial_1 \widehat{C_{X,Y}}(u,v)\,. \tag{5.2}$$

In view of (5.2), we suggest to estimate the MRD r by replacing the unknown partial derivative $\partial_1 C_{X,Y}$ by $\widehat{\partial_1 C_{X,Y}}$. Thus, a kernel estimator of r is given by

$$\widehat{r}(X,Y) = \widehat{r}(C_{X,Y}) = 6\|\widehat{\partial_1 C_{X,Y}}\|_2^2 - 2\,. \tag{5.3}$$

Under suitable assumptions on the kernel, the asymptotic behaviour of the bandwith, the regularity of the densities, and some mixing property of the data generating process [8, Assumptions 3 and 4], the estimator $\widehat{r}$ is consistent since, in this case, for any $(u_1,v_1),\ldots,(u_d,v_d) \in (0,1)^2$, the random vector

$$(Nh_X)^{1/2}((\widehat{\partial_1 C_{X,Y}} - \partial_1 C_{X,Y})(u_1,v_1),\ldots,(\widehat{\partial_1 C_{X,Y}} - \partial_1 C_{X,Y})(u_d,v_d))$$

tends weakly to a centered Gaussian vector; see [8, Thm. 4].

References.

[1] Dąbrowska, D. (1981). Regression-based orderings and measures of stochastic dependence. *Math. Operationsforsch. Statist. Ser. Statist.* **12**, 3, 317–325.

[2] Dąbrowska, D. (1985). Descriptive parameters of location, dispersion and stochastic dependence. *Statistics* **16**, 1, 63–88.

[3] Darsow, W. F., Nguyen, B., and Olsen, E. T. (1992). Copulas and Markov processes. *Illinois J. Math.* **36**, 4, 600–642.

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

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

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

[7] Fermanian, J. D., Radulović, D., and Wegkamp, M. (2004). Weak convergence of empirical copula processes. *Bernoulli* **10**, 5, 847–860.

[8] Fermanian, J. D. and Scaillet, O. (2003). Nonparametric estimation of copulas for time series. *Journal of Risk* **5**, 4, 25–54.

[9] Hall, W. J. (1970). On characterizing dependence in joint distributions. In *Essays in Probability and Statistics*, R. C. Bose, I. M. Chakravarti, P. C. Mahalanobis, C. R. Rao, and K. J. C. Smith, Eds. Univ. of North Carolina Press, Chapel Hill, N.C., 339–376.

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

[11] Kimeldorf, G. and Sampson, A. R. (1987). Positive dependence orderings. *Ann. Inst. Statist. Math.* **39**, 1, 113–128.

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

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

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

[15] Nelsen, R. B. (2006). *An introduction to copulas*, second ed. Springer, New York.

[16] Ramos, H. M. and Sordo, M. A. (2003). Dispersion measures and dispersive orderings. *Statist. Probab. Lett.* **61**, 2, 123–131.

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

[18] Scarsini, M. (1984). On measures of concordance. *Stochastica* **8**, 3, 201–218.

[19] Scarsini, M. and Shaked, M. (1996). Positive dependence orders: a survey. In *Athens Conference on Applied Probability and Time Series Analysis, Vol. I (1995)*, C. C. Heyde, Y. V. Prohorov, R. Pyke, and S. T. Rachev, Eds. Springer, New York, 70–91.

[20] Schriever, B. F. (1987). An ordering for positive dependence. *Ann. Statist.* **15**, 3, 1208–1214.

[21] Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic Orders.* Springer, New York.

[22] Siburg, K. F. and Stoimenov, P. A. (2009). A measure of mutual complete dependence. *Metrika*.

Fakultät für Mathematik
Technische Universität Dortmund
Vogelpothsweg 87
44227 Dortmund
Germany
E-mail: karl.f.siburg@math.tu-dortmund.de

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