

Article

# Canonical Correlations and Nonlinear Dependencies

Nicola Maria Rinaldo Loperfido

Dipartimento di Economia, Società e Politica, Università degli Studi di Urbino “Carlo Bo”, Via Saffi 42, 61029 Urbino (PU), Italy; nicola.loperfido@uniurb.it

**Abstract:** Canonical correlation analysis (CCA) is the default method for investigating the linear dependence structure between two random vectors, but it might not detect nonlinear dependencies. This paper models the nonlinear dependencies between two random vectors by the perturbed independence distribution, a multivariate semiparametric model where CCA provides an insight into their nonlinear dependence structure. The paper also investigates some of its probabilistic and inferential properties, including marginal and conditional distributions, nonlinear transformations, maximum likelihood estimation and independence testing. Perturbed independence distributions are closely related to skew-symmetric ones.

**Keywords:** canonical correlations; central symmetry; skew-symmetric distribution; sign symmetry

## 1. Introduction

Canonical correlation analysis is a multivariate statistical method purported to analyze the correlation structure between two random vectors

$$\mathbf{x} = \begin{pmatrix} X_1 \\ \dots \\ X_p \end{pmatrix} \text{ and } \mathbf{y} = \begin{pmatrix} Y_1 \\ \dots \\ Y_q \end{pmatrix}.$$

It obtains the linear transformations

$$\mathbf{z} = \mathbf{Ax} = \begin{pmatrix} Z_1 \\ \dots \\ Z_p \end{pmatrix} \text{ and } \mathbf{w} = \mathbf{By} = \begin{pmatrix} W_1 \\ \dots \\ W_q \end{pmatrix}.$$

where the only nonnull correlations are those between components of  $\mathbf{z}$  and  $\mathbf{w}$  with the same indices, that is

$$\begin{pmatrix} Z_1 \\ W_1 \end{pmatrix}, \dots, \begin{pmatrix} Z_r \\ W_r \end{pmatrix} \text{ with } r = \min(p, q).$$

The random vector  $(Z_1, W_1)^\top$  is the first canonical pair and the correlation between its components, that is, the first canonical correlation is the highest among all correlations between a projection of  $\mathbf{x}$  and a projection of  $\mathbf{y}$ . Similarly, the random vector  $(Z_i, W_i)^\top$  is the  $i$ -th canonical pair and the correlation between its components, that is, the  $i$ -th canonical correlation is the highest among all correlations between a projection of  $\mathbf{x}$  and a projection of  $\mathbf{y}$ , which are orthogonal to the previous canonical pairs, for  $i \in \{2, \dots, r\}$ .

Canonical correlation analysis is particularly appropriate when the joint distribution of the vectors  $\mathbf{x}$  and  $\mathbf{y}$  is multivariate normal but it often performs poorly when the data are nonnormal [1]. The problem has been addressed nonparametrically [2], semiparametrically [1] and parametrically [3]. In this paper we introduce a semiparametric model to investigate the nonlinear dependence structure by means of canonical correlations. Kernel canonical correlation analysis (KCCA) and distance canonical correlation analysis (DCCA) play a prominent role



**Citation:** Loperfido, N.M.R. Canonical Correlations and Nonlinear Dependencies. *Symmetry* **2021**, *13*, 1308. <https://doi.org/10.3390/sym13071308>

Academic Editor: Jinyu Li

Received: 6 July 2021  
Accepted: 19 July 2021  
Published: 20 July 2021

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

among nonparametric generalizations of CCA aimed at addressing nonlinear dependencies (see, e.g., [4,5]).

The main contributions of the paper are as follows. Firstly, it defines the perturbed independence distribution as a statistical model for the joint distribution of two random vectors. The proposed model is somewhat reminiscent of copula models, in that the parameters addressing the dependence structure between two random vectors do not appear in the marginal distributions of the vectors themselves; however, the generating mechanism of perturbed independence distributions is very different from those of ordinary copulas.

Secondly, the perturbed independence model allows for flexible and tractable modeling of the nonlinear dependence structure between two random vectors, since the conditional distribution of a random vector with respect to the other is skew-symmetric. The proposed model provides a parametric interpretation of KCCA and DCCA, which are commonly regarded as nonparametric multivariate methods.

Thirdly, some appealing properties of canonical correlation analysis that hold true in the normal case still hold true in the perturbed independence case. For example, the first (second) component of a canonical pair is independent from the second (first) component of any other canonical pair. Further, if the marginal distributions of the two given vectors are normal, any canonical pair is independent of any other canonical pair.

Fourthly, the paper investigates the bivariate perturbed independence models within the framework of positive and negative association. In particular, it shows that the canonical pairs obtained from a perturbed independence distribution have the desirable properties of being positive quadrant dependent, under mild assumptions on the perturbing function.

The rest of the paper is structured as follows. Section 2 defines perturbed independence distributions and states some of their probabilistic and inferential properties. Section 3 connects perturbed independence distributions, canonical correlation analysis, positive dependence orderings and ordinal measures of association. Section 4 uses both theoretical and empirical results to find nonlinear transformations that increase correlations. Appendix A contains all proofs.

## 2. Model

This section defines the perturbed independence model, states its invariance properties and the independence properties of its canonical pairs. The theoretical results are illustrated with the bivariate distribution  $2\phi(x)\phi(y)\Phi(\lambda xy)$  introduced by [6,7], where  $\phi(\cdot)$  and  $\Phi(\cdot)$  denote the probability and the cumulative density functions of a standard normal distribution, while  $\lambda$  is a real value. Ref. [8] thoroughly investigated its properties and proposed some generalizations.

A  $p$ -dimensional random vector  $\mathbf{x}$  is centrally symmetric (simply symmetric, henceforth) if there is a  $p$ -dimensional real vector  $\boldsymbol{\zeta}$  such that  $\mathbf{x} - \boldsymbol{\zeta}$  and  $\boldsymbol{\zeta} - \mathbf{x}$  are identically distributed ([9]). A real-valued function  $\pi(\cdot)$  is a skewing function (also known as perturbing function) if it satisfies the equality  $\pi(-\mathbf{a}) = 1 - \pi(\mathbf{a})$  and the inequalities  $0 \leq \pi(\mathbf{a}) \leq 1$  for any real vector  $\mathbf{a}$  [10]. The probability density function of a perturbed independence model is twice the product of two symmetric probability density functions and a skewing function evaluated at a bilinear function of the outcomes. A more formal definition follows.

**Definition 1.** Let the joint distribution of the random vectors  $\mathbf{x}$  and  $\mathbf{y}$  be

$$f(\mathbf{x}, \mathbf{y}) = 2h(\mathbf{x} - \boldsymbol{\mu})k(\mathbf{y} - \boldsymbol{\nu})\pi\left\{(\mathbf{y} - \boldsymbol{\nu})^\top \boldsymbol{\Psi}(\mathbf{x} - \boldsymbol{\mu})\right\},$$

where  $h(\cdot)$  is the pdf of a  $p$ -dimensional, centrally symmetric distribution,  $k(\cdot)$  is the pdf of a  $q$ -dimensional, centrally symmetric distribution,  $\boldsymbol{\Psi}$  is a  $q \times p$  matrix and  $\pi(\cdot)$  is a function satisfying  $0 \leq \pi(-a) = 1 - \pi(a) \leq 1$  for any real value  $a$ . We refer to this distribution as to a perturbed independence model, with components  $h(\cdot)$  and  $k(\cdot)$ , location vectors  $\boldsymbol{\mu}$  and  $\boldsymbol{\nu}$ , perturbing function  $\pi(\cdot)$  and association matrix  $\boldsymbol{\Psi}$ .

In the bivariate distribution  $2\phi(x)\phi(y)\Phi(\lambda xy)$ , both components coincide with the normal pdf, both location vectors coincide with the origin, the perturbing function is the standard normal cdf and the association matrix is the scalar parameter  $\lambda$ .

Random numbers having a perturbed independence distribution can be generated in a very simple way. For the sake of simplicity, we illustrate it in the simplified case where  $\boldsymbol{\mu}$  and  $\boldsymbol{\nu}$  are null vectors and  $\pi(\cdot)$  is a cumulative distribution function of a distribution symmetric at the origin. First, generate the vectors  $\mathbf{u}$  and  $\mathbf{v}$  from the densities  $h(\cdot)$  and  $k(\cdot)$ . Second, generate the scalar  $r$  from the distribution whose cumulative density function is  $\pi(\cdot)$ . Third, let the vector  $\mathbf{w}$  be  $(\mathbf{u}^\top, \mathbf{v}^\top)^\top$  if the bilinear form  $\mathbf{u}^\top \boldsymbol{\Psi} \mathbf{v}$  is greater than  $r$  and either  $(-\mathbf{u}^\top, \mathbf{v}^\top)^\top$  or  $(\mathbf{u}^\top, -\mathbf{v}^\top)^\top$  in the opposite case. Then, the distribution of  $\mathbf{w}$  is perturbed independence with components  $h(\cdot)$  and  $k(\cdot)$ , null location vectors, perturbing function  $\pi(\cdot)$  and association matrix  $\boldsymbol{\Psi}$ .

The bivariate distribution  $2\phi(x)\phi(y)\Phi(\lambda xy)$  might be generated as follows. First, generate three mutually independent, standard normal random numbers  $U, W$  and  $Z$ . Second, set  $X$  equal to  $U$  and  $Y$  equal to  $W$  if the product  $\lambda UW$  is greater than  $Z$ . Otherwise, set  $X$  equal to  $-U$  and  $Y$  equal to  $W$ . Then the joint distribution of  $X$  and  $Y$  is  $2\phi(x)\phi(y)\Phi(\lambda xy)$ .

A  $p$ -dimensional probability density function  $2g(\mathbf{a} - \boldsymbol{\zeta})\pi(\mathbf{a} - \boldsymbol{\zeta})$  is skew-symmetric with kernel  $g(\cdot)$  (i.e., a probability density function symmetric at the origin), location vector  $\boldsymbol{\zeta}$  and skewing function  $\pi(\cdot)$ . The function  $g(\cdot)$  would be more precisely denoted by  $g_p(\cdot)$ , since it depends on the dimension of the corresponding random vector. However, we use  $g(\cdot)$  instead of  $g_p(\cdot)$  to relieve the notational burden. Ref. [11] discuss hypothesis testing on  $g(\cdot)$  for any choice of function  $\pi(\cdot)$ . The most widely studied skew-symmetric distributions are the linearly skewed distributions, where the skewing function depends on  $\mathbf{a} - \boldsymbol{\zeta}$  only through its linear function  $\boldsymbol{\alpha}^\top (\mathbf{a} - \boldsymbol{\zeta})$ , as it happens for the multivariate skew-normal case. [12], as well as [13], investigated their inferential properties. Ref. [14] used them to motivate kurtosis-based projection pursuit.

In the notation of the above definition, the first part of the following theorem states that the marginal distributions of  $\mathbf{x}$  and  $\mathbf{y}$  are  $h(\mathbf{x} - \boldsymbol{\mu})$  and  $k(\mathbf{y} - \boldsymbol{\nu})$ . Thus, perturbed independent distributions separately model the marginal distributions and the association between two random vectors, and constitute an alternative to copulas. The second part of the following theorem states that the conditional distribution of a component with respect to the other is linearly skewed. Hence, the association between the two components has an analytical form, which has been thoroughly investigated.

**Theorem 1.** *Let the random vectors  $\mathbf{x}$  and  $\mathbf{y}$  have a perturbed independence distribution with components  $h(\cdot)$ ,  $k(\cdot)$  and location vectors  $\boldsymbol{\mu}$ ,  $\boldsymbol{\nu}$ . Then the following statements hold true.*

- *The marginal probability density functions of  $\mathbf{x}$  and  $\mathbf{y}$  are  $h(\mathbf{x} - \boldsymbol{\mu})$  and  $k(\mathbf{y} - \boldsymbol{\nu})$ .*
- *The conditional probability density functions of  $\mathbf{x}$  given  $\mathbf{y}$  and  $\mathbf{y}$  given  $\mathbf{x}$  are skew-symmetric with kernels  $h(\cdot)$  and  $k(\cdot)$ , while the associated location vectors are  $\boldsymbol{\mu}$  and  $\boldsymbol{\nu}$ .*

The marginal distributions of  $2\phi(x)\phi(y)\Phi(\lambda xy)$  are standard normal:  $X \sim N(0, 1)$  and  $Y \sim N(0, 1)$ . The conditional distributions are skew-normal: the probability density functions of  $X|Y = y \sim \mathcal{SN}(\lambda y)$  and of  $Y|X = x \sim \mathcal{SN}(\lambda x)$  are  $2\phi(x)\Phi(\lambda xy)$  and  $2\phi(y)\Phi(\lambda xy)$ . The sign of the correlation between  $X$  and  $Y$  is the same as the sign of  $\lambda$  but the two random variables are nonlinearly dependent [7]:

$$E(X|Y = y) = \sqrt{\frac{2}{\pi}} \frac{\lambda y}{\sqrt{1 + \lambda^2 y^2}}, \quad E(Y|X = x) = \sqrt{\frac{2}{\pi}} \frac{\lambda x}{\sqrt{1 + \lambda^2 x^2}}.$$

There is a close connection between order statistics and either skew-normal distributions or their generalizations. For example, any linear combination of the minimum and the maximum of a bivariate, exchangeable and elliptical random vector is skew-elliptical [15]. In particular, any skew-normal distribution might be represented as the maximum or the minimum of a bivariate, normal and exchangeable random vector. At present, it is not

clear whether there exists a meaningful connection between order statistics and perturbed independence distributions, which would ease both the interpretation and the application of these distributions.

The mean vector  $\mathbf{m}$  and the covariance matrix  $\mathbf{S}$  of the  $n \times d$  data matrix  $\mathbf{X}$  are statistically independent, if the rows of  $\mathbf{X}$  are a random sample from a multivariate normal distribution. As a direct consequence, the components of the pairs  $(\mathbf{m}_1, \mathbf{S}_2)$  and  $(\mathbf{m}_2, \mathbf{S}_1)$  are statistically independent, too, where  $\mathbf{m}_1$  and  $\mathbf{S}_1$  ( $\mathbf{m}_2$  and  $\mathbf{S}_2$ ) are the mean vector and the covariance matrix of  $\mathbf{X}_1$  ( $\mathbf{X}_2$ ), that is the data matrix whose columns coincide with the first  $0 < p < d$  (the last  $d - p$ ) columns of  $\mathbf{X}$ . The same property holds true for perturbed independence models, as a corollary of the following theorem.

**Theorem 2.** *Let the random vectors  $\mathbf{x}$  and  $\mathbf{y}$  have the perturbed independence distribution with location vectors  $\boldsymbol{\mu}$  and  $\boldsymbol{\nu}$ . Then any even function of  $\mathbf{x} - \boldsymbol{\mu}$  is independent of  $\mathbf{y}$ . Similarly, any even function of  $\mathbf{y} - \boldsymbol{\nu}$  is independent of  $\mathbf{x}$ .*

Let the joint distribution of the random variables  $X$  and  $Y$  be  $2\phi(x)\phi(y)\Phi(\lambda xy)$ . Then  $Y$  and  $X^2$  are mutually independent. Similarly,  $X$  and  $Y^2$  are mutually independent.

The components of the canonical covariates  $\mathbf{z} = (Z_1, \dots, Z_p)^\top$  and  $\mathbf{w} = (W_1, \dots, W_q)^\top$  are uncorrelated when their indices differ:

$$\rho(Z_i, Z_j) = \rho(W_i, W_j) = \rho(Z_i, W_j) = 0 \text{ when } i \neq j.$$

A  $p$ -dimensional random vector  $\mathbf{v}$  is said to be sign-symmetric if there is a  $p$ -dimensional real vector  $\mathbf{u}$  such that  $\mathbf{v} - \mathbf{u}$  and  $\mathbf{U}(\mathbf{v} - \mathbf{u})$  are identically distributed, where  $\mathbf{U}$  is any  $p \times p$  diagonal matrix whose diagonal elements are either 1 or  $-1$  [9]. For example, spherical random vectors are sign-symmetric. The following theorem shows that the canonical covariates belonging to different canonical vectors and with different indices are independent, if the joint distribution of the original variables is perturbed independence with sign-symmetric components.

**Theorem 3.** *Let the random vectors  $\mathbf{x} \in \mathbb{R}^p$  and  $\mathbf{y} \in \mathbb{R}^q$  have a perturbed independence distribution with sign-symmetric components. Further, let  $\mathbf{z} = (Z_1, \dots, Z_p)^\top$  and  $\mathbf{w} = (W_1, \dots, W_q)^\top$  be the canonical covariates of  $\mathbf{x}$  and  $\mathbf{y}$ . Then  $Z_i$  and  $W_j$  are independent when  $i \neq j$ .*

Under normal sampling, the components of different canonical pairs are statistically independent. The following corollary of the above theorem shows that the same property still holds true when the original variables have a perturbed independence distribution with normal components.

**Corollary 1.** *Let the random vectors  $\mathbf{x} \in \mathbb{R}^p$  and  $\mathbf{y} \in \mathbb{R}^q$  have a perturbed independence distribution with normal components. Further, let  $\mathbf{z} = (Z_1, \dots, Z_p)^\top$  and  $\mathbf{w} = (W_1, \dots, W_q)^\top$  the canonical covariates of  $\mathbf{x}$  and  $\mathbf{y}$ . Then the variables  $Z_i$ ,  $Z_j$ ,  $W_i$  and  $W_j$  are pairwise independent when  $i \neq j$ .*

As remarked by [16], the default measures of multivariate skewness and kurtosis are those introduced by [17]. Mardia's skewness is the sum of all squared, third-order, standardized moments, while Mardia's kurtosis is the fourth moment of the Mahalanobis distance of the random vector from its mean.

Mardia's kurtosis of  $2\phi(x_1)\phi(x_2)\Phi(\lambda x_1 x_2)$  is

$$\frac{8 + 4E^2(XY)}{\{1 - E^2(XY)\}^2},$$

so that it increases with the squared correlation between  $X$  and  $Y$  ([18]).

It is tempting to generalize  $2\phi(x)\phi(y)\Phi(\lambda xy)$  by letting

$$2\Phi\left(\lambda\prod_{i=1}^p x_i\right)\prod_{i=1}^p \phi(x_i),$$

as performed in [6,7]. Unfortunately, this model does not preserve the nonlinear associations between pairs of its components. For example, the joint bivariate marginals of the trivariate distribution  $2\phi(x)\phi(y)\phi(z)\Phi(\lambda xyz)$  are bivariate, standard normal random vectors [19]. Other generalizations of  $2\phi(x)\phi(y)\Phi(\lambda xy)$  have been proposed by [8].

Let  $f_{\mathbf{xy}}$  be the joint probability density function of the  $p$ -dimensional random vector  $\mathbf{x}$  and of the  $q$ -dimensional random vector  $\mathbf{y}$ . Further, let  $f_{\mathbf{x}}$  and  $f_{\mathbf{y}}$  be the marginal probability density functions of  $\mathbf{x}$  and  $\mathbf{y}$ . The distance covariance between  $\mathbf{x}$  and  $\mathbf{y}$  with respect to the weight function  $w$  is

$$\mathcal{V}^2(\mathbf{x}, \mathbf{y}, w) = \int_{\mathbb{R}^{p+q}} |f_{\mathbf{xy}}(\mathbf{t}, \mathbf{s}) - f_{\mathbf{x}}(\mathbf{t})f_{\mathbf{y}}(\mathbf{s})|^2 w(\mathbf{t}, \mathbf{s}) d\mathbf{t}d\mathbf{s},$$

where  $\mathbf{t} \in \mathbb{R}^p, \mathbf{s} \in \mathbb{R}^q$  and  $w(\mathbf{t}, \mathbf{s}) \geq 0$  [20]. If the joint distribution of  $\mathbf{x}$  and  $\mathbf{y}$  is a perturbed independence model with components  $h(\cdot)$  and  $k(\cdot)$ , location vectors  $\boldsymbol{\mu}$  and  $\boldsymbol{\nu}$ , perturbing function  $\pi(\cdot)$  and association matrix  $\boldsymbol{\Psi}$  we have

$$f_{\mathbf{xy}}(\mathbf{t}, \mathbf{s}; \boldsymbol{\mu}, \boldsymbol{\nu}, \boldsymbol{\Psi}) = 2h(\mathbf{t} - \boldsymbol{\mu})k(\mathbf{s} - \boldsymbol{\nu})\pi\left\{(\mathbf{s} - \boldsymbol{\nu})^\top \boldsymbol{\Psi}(\mathbf{t} - \boldsymbol{\mu})\right\},$$

$$f_{\mathbf{x}}(\mathbf{t}; \boldsymbol{\mu}) = h(\mathbf{t} - \boldsymbol{\mu}), f_{\mathbf{y}}(\mathbf{s}; \boldsymbol{\nu}) = k(\mathbf{s} - \boldsymbol{\nu}), \pi(-a) = 1 - \pi(a).$$

A little algebra leads to the identities

$$f_{\mathbf{xy}}(\mathbf{t}, \mathbf{s}) - f_{\mathbf{x}}(\mathbf{t})f_{\mathbf{y}}(\mathbf{s}) = 2h(\mathbf{t} - \boldsymbol{\mu})k(\mathbf{s} - \boldsymbol{\nu})\pi\left\{(\mathbf{s} - \boldsymbol{\nu})^\top \boldsymbol{\Psi}(\mathbf{t} - \boldsymbol{\mu})\right\} - h(\mathbf{t} - \boldsymbol{\mu})k(\mathbf{s} - \boldsymbol{\nu}) =$$

$$h(\mathbf{t} - \boldsymbol{\mu})k(\mathbf{s} - \boldsymbol{\nu})\left[2\pi\left\{(\mathbf{s} - \boldsymbol{\nu})^\top \boldsymbol{\Psi}(\mathbf{t} - \boldsymbol{\mu})\right\} - 1\right] = \frac{f_{\mathbf{xy}}(\mathbf{t}, \mathbf{s}; \boldsymbol{\mu}, \boldsymbol{\nu}, \boldsymbol{\Psi}) - f_{\mathbf{xy}}(\mathbf{t}, \mathbf{s}; \boldsymbol{\mu}, \boldsymbol{\nu}, -\boldsymbol{\Psi})}{2}.$$

Hence, for perturbed independence models, the distance covariance is just half the difference between  $f_{\mathbf{xy}}(\mathbf{t}, \mathbf{s}; \boldsymbol{\mu}, \boldsymbol{\nu}, \boldsymbol{\Psi})$  and  $f_{\mathbf{xy}}(\mathbf{t}, \mathbf{s}; \boldsymbol{\mu}, \boldsymbol{\nu}, -\boldsymbol{\Psi})$ , which is the probability density functions of

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \text{ and } \begin{pmatrix} \mathbf{x} \\ -\mathbf{y} \end{pmatrix}.$$

In particular, if the joint distribution of the random variables  $X$  and  $Y$  is  $2\phi(x)\phi(y)\Phi(\lambda xy)$  we have

$$f_{XY}(t, s) - f_X(t)f_Y(s) = \phi(x)\phi(y)\{\Phi(\lambda xy) - \Phi(-\lambda xy)\}.$$

### 3. Concordance

This section investigates the bivariate perturbed independence models within the framework of positive and negative association. In particular, it shows that the canonical pairs obtained from a perturbed independence distribution have the desirable properties of being positive quadrant dependent, under mild assumptions on the perturbing function. The seminal paper by [21] started a vast literature on dependence orderings and their connections with ordinal measures of association. For the sake of brevity, here we mention only some thorough reviews of the concepts in this section: [22–27].

Two random variables are said to be either concordant, positively associated or positively dependent if larger (smaller) outcomes of one of them often occur together with larger (smaller) outcomes of the other random variable. Conversely, two random variables are said to be either discordant, negatively associated or negatively dependent if larger (smaller) outcomes of one of them often occur together with smaller (larger) outcomes of

the other random variable. For example, financial returns from different markets are known to be positively dependent (see, e.g., [28–30]). The degree of concordance or discordance is assessed with ordinal measures of association, of which the most commonly used are Pearson's correlation (simply correlation, for short), Spearman's rho and Kendall's tau.

The correlation is the best known measure of ordinal association. The correlation between two random variables  $X$  and  $Y$  is

$$\rho(X, Y) = \frac{E\{(X - \mu_X)(Y - \mu_Y)\}}{\sqrt{E\{(X - \mu_X)^2\} \cdot E\{(Y - \mu_Y)^2\}}},$$

where  $\mu_X$  and  $\mu_Y$  are the expectations of  $X$  and  $Y$ . The ordinal association between two random variables might be decomposed into a linear component and a nonlinear component. The linear component refers to the tendency of the random variables to deviate from their means in a proportional way. The correlation only detects and measures the linear component of the ordinal association. When the nonlinear component is not negligible, the information conveyed by the correlation needs to be integrated with information from other measures of ordinal association.

Spearman's rho, also known as Spearman's correlation, between the random variables  $X$  and  $Y$  is the correlation between the two variables after being transformed according to their marginal cumulative distribution functions:

$$\rho_S(X, Y) = \rho\{F_X(X), F_Y(Y)\},$$

where  $F_X(\cdot)$  and  $F_Y(\cdot)$  are the marginal cumulative distribution functions of  $X$  and  $Y$ . Its sample counterpart is the correlation between the observed ranks. Spearman's rho is a measure of ordinal association detecting both linear and nonlinear dependence. It is also more robust to outliers than the Pearson's correlation.

Kendall's tau, also known as Kendall's correlation, between two random variables is the difference between their probability of concordance and their probability of discordance. The former (latter) is the probability that the difference between the first components of two independent outcomes from a bivariate distribution have the same sign of (a different sign than) the difference between the second components of the same pairs. More formally, Kendall's tau between the random variables  $X$  and  $Y$  is

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

where  $(X_1, Y_1)^\top$  and  $(X_2, Y_2)^\top$  are two independent outcomes from the bivariate random vector  $(X, Y)^\top$ . Just like Spearman's rho, Kendall's tau is an ordinal measure of association detecting linear as well as nonlinear dependence and is more robust to outliers than Pearson's correlation.

Unfortunately, Pearson's correlation, Spearman's rho and Kendall's tau might take different signs, thus making it difficult to measure ordinal association. In order to prevent this from happening, it is convenient to impose some constraints on the bivariate distribution. The distribution of a bivariate random vector  $(X, Y)^\top$  is said to be positively quadrant dependent (PQD) if its joint cdf is greater or equal than the product of the marginal cdf:

$$F_{X,Y}(x, y) \geq F_X(x) \cdot F_Y(y)$$

for any two real values  $x$  and  $y$ . Similarly, the distribution of a bivariate random vector  $(X, Y)^\top$  is said to be negatively quadrant dependent (NQD) if its joint cdf is either smaller or equal than the product of the marginal cdf:

$$F_{X,Y}(x, y) \leq F_X(x) \cdot F_Y(y)$$



for any two real values  $x$  and  $y$ . Pearson's correlation, Spearman's rho and Kendall's tau of PQD (NQD) distributions are either null or have positive (negative) signs.

Independent random variables are special cases of PQD and NQD random variables. In order to rule this case out, the PQD and NQD condition can be made more restrictive that the above inequalities needs to be strict for measurable sets of  $x$  and  $y$  values. For example, a strictly positive quadrant dependent pair of random variables satisfies the inequality

$$F_{X,Y}(x,y) > F_X(x) \cdot F_Y(y)$$

for any two real values  $x$  and  $y$  belonging to given interval of positive length. Pearson's correlation, Spearman's rho and Kendall's tau of strictly positive (negative) quadrant dependent distributions have positive (negative) signs. As shown in the following theorem, a bivariate perturbed independence model is strictly positive (negative) quadrant dependent if the perturbing function is a cumulative density function and the association parameter is a positive (negative) scalar.

**Theorem 4.** *Let the joint distribution of the random variables  $X$  and  $Y$  be perturbed independent with components  $h(\cdot)$  and  $k(\cdot)$ , perturbing function  $\pi(\cdot)$  and association parameter  $\lambda$ :  $f(x,y) = 2h(x)k(y)\pi(\lambda xy)$ . Further, let  $\pi(\cdot)$  be the cumulative density function of a symmetric distribution. Then the random variables  $X$  and  $Y$  are strictly positive (negative) quadrant dependent when  $\lambda$  is positive (negative).*

The joint distribution  $2\phi(x)\phi(y)\Phi(\lambda xy)$  of the bivariate random vector  $(X, Y)^\top$  introduced in the previous section fulfills the assumptions in Theorem 5. In particular, if the association parameter  $\lambda$  is positive, the random variables  $X$  and  $Y$  are strictly positive quadrant dependent:

$$F_{X,Y}(a,b) = \int_{-\infty}^b \int_{-\infty}^a 2\phi(x)\phi(y)\Phi(\lambda xy) dx dy > \Phi(a)\Phi(b) = F_X(a)F_Y(b),$$

for any two real values  $a$  and  $b$ . As a direct consequence, their Pearson's correlation  $\rho(X, Y)$ , their Spearman's rho  $\rho_S(X, Y)$  and their Kendall's tau  $\tau(X, Y)$  are positive.

Pearson's correlation between the components of a canonical pair is nonnegative. However, within a nonparametric framework, their Spearman's rho and their Kendall's tau can take any sign. When Pearson's correlation between the components of a canonical pair is positive but their Spearman's rho and their Kendall's tau are negative, the former ordinal association measure becomes quite unreliable and canonical correlation analysis provides little insight into the dependence structure. This problem does not occur under a perturbed independence model satisfying the assumptions stated in the following theorem.

**Theorem 5.** *Let  $(Z_1, W_1)^\top, \dots, (Z_r, W_r)^\top$ , with  $r = \min(p, q)$  the canonical pairs obtained from a perturbed independence distribution, and let their density be*

$$2\pi \left( \sum_{i=1}^r \lambda_i z_i w_i \right) \left\{ \prod_{i=1}^r h_i(z_i) \right\} \left\{ \prod_{i=1}^r k_i(w_i) \right\},$$

where  $\pi(\cdot)$  is a strictly increasing perturbing function. Then the joint distribution of the  $i$ -th canonical pair is a bivariate perturbed independence model:

$$2h_i(z_i)k_i(w_i)G(\lambda_i z_i w_i),$$

where  $G(\cdot)$  is a strictly increasing perturbing function.

We illustrate the above theorem with the perturbed independence distribution

$$2\phi_p(\mathbf{x}; \Sigma_x)\phi_p(\mathbf{y}; \Sigma_y)F(\mathbf{y}^\top \Psi \mathbf{x}),$$

where  $\phi_q(\cdot; \Omega)$  is the  $q$ -dimensional normal density with null mean vector and covariance matrix  $\Sigma$ ,  $F(\cdot)$  is the cdf of a continuous distribution symmetric at the origin and  $\Psi$  is a symmetric  $p \times p$  matrix. The distribution of the canonical variates  $\mathbf{z} = \mathbf{Ax} = (Z_1, \dots, Z_p)^\top$  and  $\mathbf{w} = \mathbf{By} = (W_1, \dots, W_q)^\top$  is

$$2\phi_{2p}(\mathbf{z}, \mathbf{w}; \mathbf{I}_{2p})F\left(\sum_{i=1}^p \lambda_i z_i w_i\right),$$

which fulfills the assumptions in Theorem 6. Then the joint distribution of the  $i$ -th canonical pair  $(Z_i, W_i)^\top$  is

$$2\phi(z_i)\phi(w_i)G_i(\lambda_i z_i w_i),$$

where  $\phi(\cdot)$  is the pdf of a univariate, standard normal distribution and  $G_i(\cdot)$  is the cdf of a continuous distribution symmetric at the origin. By Theorem 5 and since the  $i$ -th canonical correlation  $\rho(Z_i, W_i)$  is nonnegative, the association parameter  $\lambda_i$ , Kendall's tau  $\tau(Z_i, W_i)$  and Spearman's rho  $\rho_S(Z_i, W_i)$  are nonnegative, too. Moreover, if  $\rho(Z_i, W_i)$  is positive, the association parameter  $\lambda_i$ , Kendall's tau  $\tau(Z_i, W_i)$  and Spearman's rho  $\rho_S(Z_i, W_i)$  are positive, too.

#### 4. Nonlinearity

As a desirable property, CCA decomposes the covariance matrix between the  $p$ -dimensional random vector  $\mathbf{x}$  and the  $q$ -dimensional random vector  $\mathbf{y}$  into linear combinations of the covariances between uncorrelated linear functions of  $\mathbf{x}$  and  $\mathbf{y}$ . Ref. [31] thoroughly investigate the interpretation of CCA within the framework of linear dependence. The first output of CCA are the linear combinations of  $\mathbf{x}$  and  $\mathbf{y}$ , which are maximally correlated:

$$\max_{\mathbf{a} \in \mathbb{R}_0^p, \mathbf{b} \in \mathbb{R}_0^q} \text{corr}(\mathbf{a}^\top \mathbf{x}, \mathbf{b}^\top \mathbf{y}),$$

where  $\mathbb{R}_0^p$  and  $\mathbb{R}_0^q$  are the sets of  $p$ -dimensional and  $q$ -dimensional nonnull, real vectors.

As mentioned in the Introduction and in the previous section, both the interpretability and the usefulness of CCA are severely diminished by nonlinear dependencies between  $\mathbf{x}$  and  $\mathbf{y}$ . A solution would be looking for the linear and nonlinear transformations of  $\mathbf{x}$  and  $\mathbf{y}$ , which are maximally correlated:

$$\max_{g_1, g_2 \in \mathbb{G}} \max_{\mathbf{a} \in \mathbb{R}_0^p, \mathbf{b} \in \mathbb{R}_0^q} \text{corr}\{g_1(\mathbf{a}^\top \mathbf{x}), g_2(\mathbf{b}^\top \mathbf{y})\},$$

where  $\mathbb{G}$  is the set of all real valued monotonic functions. In the general case, the maximization needs to be performed simultaneously with respect to the nonlinear functions  $g_1(\cdot)$ ,  $g_2(\cdot)$  and the real vectors  $\mathbf{a}$ ,  $\mathbf{b}$ , thus being difficult to compute and difficult to interpret. Ref. [1] addressed the problem by proposing the Gaussian copula model, where the components of  $\mathbf{x}$  and  $\mathbf{y}$  have a joint distribution that is multivariate normal, after being transformed according to monotonic and nonlinear functions. However, these monotonic transformations do not have a clear interpretation and they are not guaranteed to increase the correlations.

Perturbed independence models do not suffer from these limitations. Firstly, the monotonic transformations have a simple interpretation, being the expectations of one variable conditioned with respect to the other. Secondly, the same transformations are guaranteed to increase the correlations, under mild assumptions. These statements are made more precise in the following theorem.



**Theorem 6.** *Let the joint distribution of the random variables  $X$  and  $Y$  be perturbed independent with null location parameters, nonnull association parameter and increasing perturbing function. Finally, let  $X$  and  $Y$  have finite second moments. Then the conditional expectation  $g(X) = E(Y|X)$  is a monotone, odd and nonlinear function, while the correlation between  $Y$  and  $X$  is smaller than the correlation between  $Y$  and  $g(X)$ .*

We illustrate the above theorem with the distribution  $2\phi(x)\phi(y)\Phi(\lambda xy)$  of the bivariate random vector  $(X, Y)^T$  introduced in Section 1. The conditional expectations of  $Y$  and  $X$  with respect to the outcomes  $x$  of  $X$  and  $y$  of  $Y$  are

$$E(Y|X = x) = \frac{\lambda x\sqrt{2/\pi}}{\sqrt{1 + \lambda^2 x^2}} \text{ and } E(X|Y = y) = \frac{\lambda y\sqrt{2/\pi}}{\sqrt{1 + \lambda^2 y^2}},$$

so that the nonlinear function of  $X$  and  $Y$  maximally correlated with  $Y$  and  $X$  are

$$g_1(X) = E(Y|X) = \frac{\lambda X\sqrt{2/\pi}}{\sqrt{1 + \lambda^2 X^2}} \text{ and } g_2(Y) = E(X|Y) = \frac{\lambda Y\sqrt{2/\pi}}{\sqrt{1 + \lambda^2 Y^2}}.$$

The above theorem does not guarantee that  $E(Y|X)$  is the nonlinear transformation of one component that is maximally correlated with  $Y$ , nor that such correlation is smaller than the correlation between  $E(X|Y)$  and  $E(Y|X)$ . We empirically address this point by simulating  $n = 10,000$  bivariate data from  $2\phi(x)\phi(y)\Phi(\lambda xy)$ , where  $\lambda \in \{1, 2, 3, 4, 5, 6\}$ . The left-hand scatterplots in Figure 1 clearly hint at positive dependence: more points lie in the first and in the third quadrants as the association parameter increases, despite the absence of the ellipsoidal shapes associated with bivariate normality. For each simulated sample, we computed Kendall’s tau, Spearman’s rho and Pearson’s correlation and report their values in Table 1. The three measures of ordinal association are positive and they increase with the association parameter, consistently with the theoretical results in Section 2. More surprisingly, Spearman’s rho is always greater than Kendall’s tau and Pearson’s correlation, unlike the bivariate normal distribution, where Pearson’s correlation is always greater than Kendall’s tau and Spearman’s rho.

**Table 1.** Mean values of Pearson’s correlation, Kendall’s tau and Spearman’s rho for 100,00 outcomes from  $2\phi(x)\phi(y)\Phi(\lambda xy)$ , where  $\lambda \in \{1, 2, 3, 4, 5, 6\}$ . The column “Transformed” contains Pearson’s correlations between the transformed components.

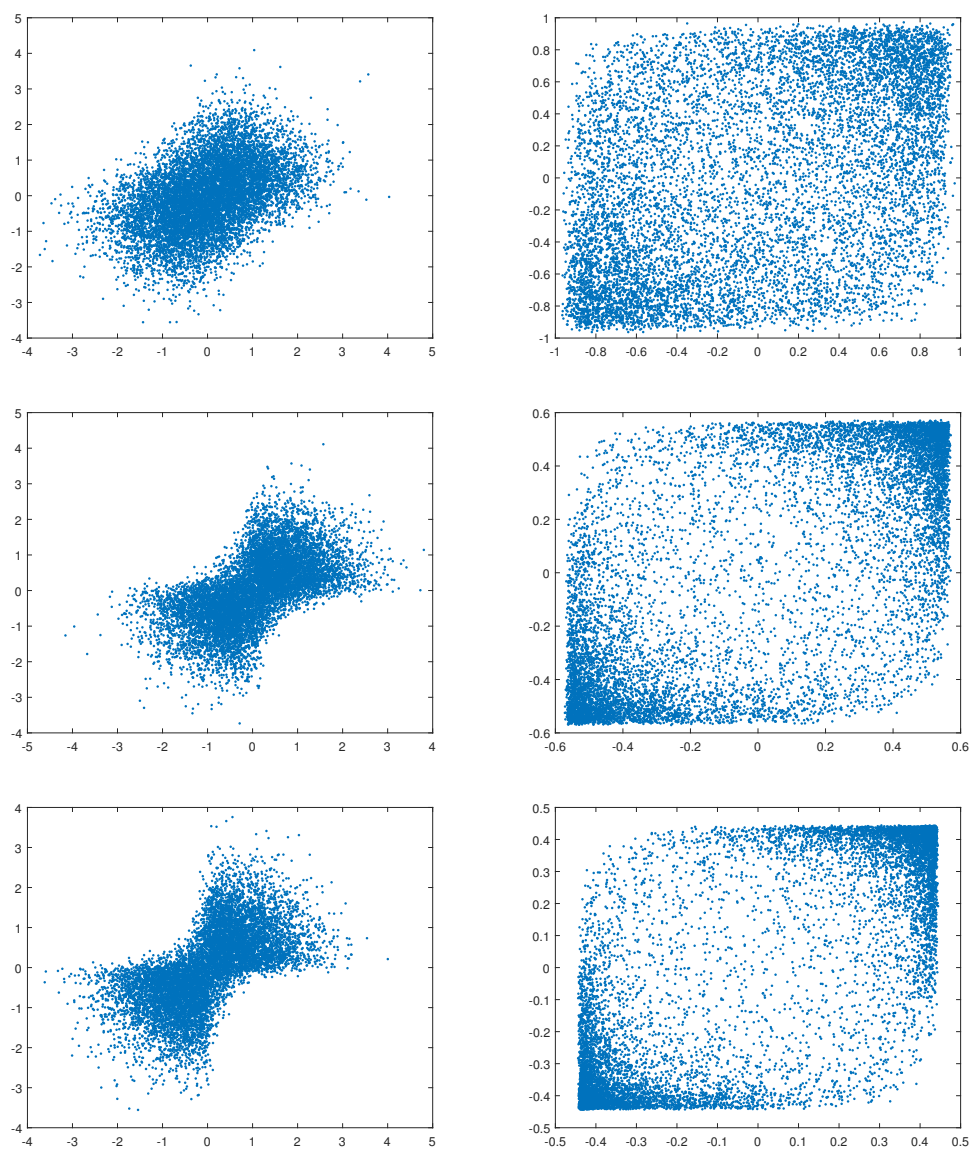
Lambda	Kendall	Pearson	Spearman	Transformed
1	0.311	0.439	0.468	0.469
2	0.412	0.552	0.619	0.642
3	0.444	0.587	0.667	0.713
4	0.458	0.598	0.689	0.752
5	0.471	0.612	0.709	0.791
6	0.480	0.618	0.720	0.810

Finally, for each simulated sample  $(X_1, Y_1), \dots, (X_n, Y_n)$ , we computed Pearson’s correlation between  $Z_1, \dots, Z_n$  and  $W_1, \dots, W_n$ , where

$$Z_i = \frac{X_i}{\sqrt{1 + \lambda^2 X_i^2}} \text{ and } W_i = \frac{Y_i}{\sqrt{1 + \lambda^2 Y_i^2}}$$

are proportional to the sample counterpart of the expectation of  $Y$  given  $X = x$  and  $X$  given  $Y = y$  under the model  $2\phi(x)\phi(y)\Phi(\lambda xy)$ . For each simulated sample, these correlations are always greater than the correlations between the original data, consistently with Theorem 7. Moreover, Pearson’s correlations between  $W_1, \dots, W_n$  and  $Z_1, \dots, Z_n$  are always greater their Spearman’s correlations. As shown in the right-hand scatterplots

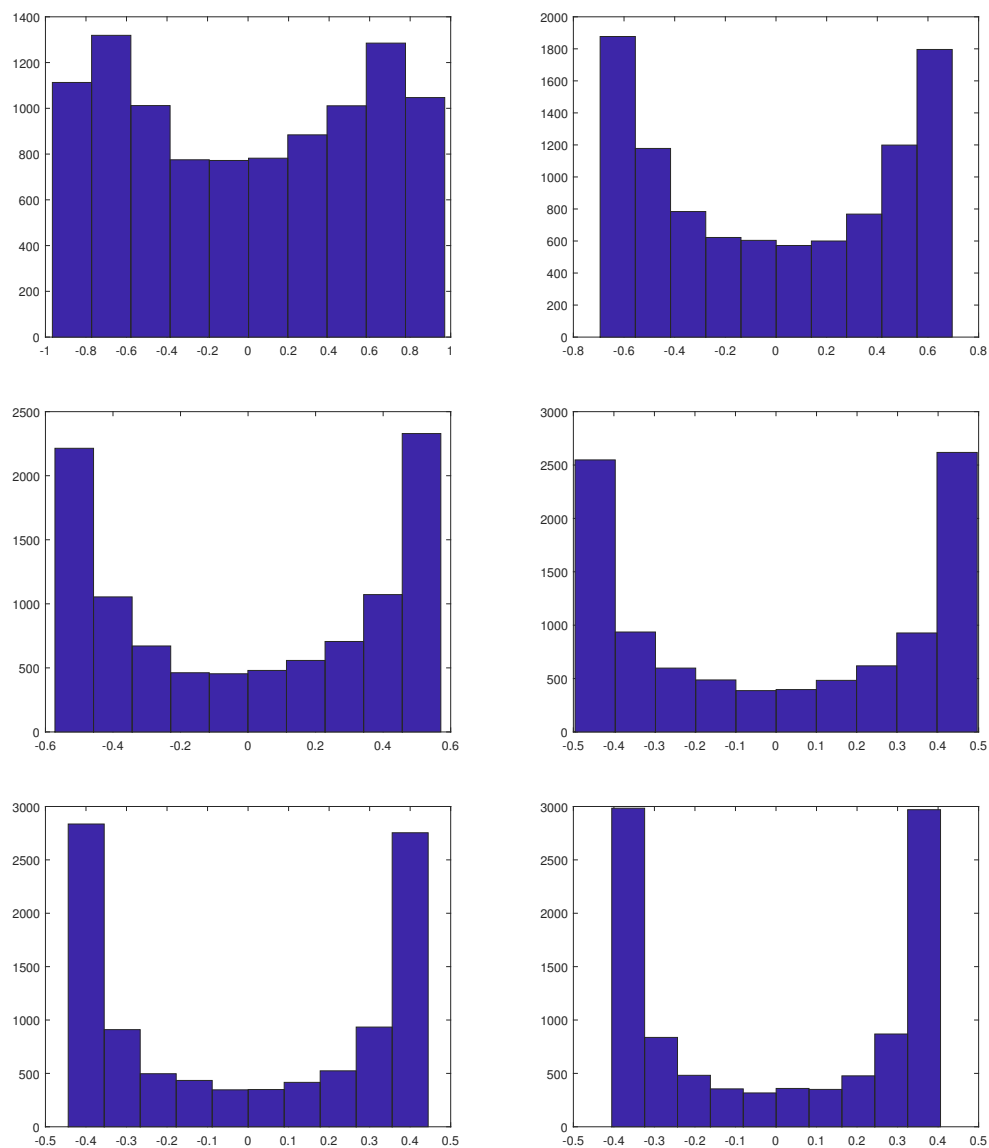
of Figures 1 and 2, the transformed data lie at the lower left corner and at the upper right corner of a square. This pattern becomes more evident as the association parameter increases. The histograms of  $Z_1, \dots, Z_n$  in Figure 2 are symmetric and bimodal, with both modes at the ends of the observed range. Bimodality becomes more evident as the association parameter increases. The behavior of the transformed data  $W_1, \dots, W_n$  is virtually identical and therefore is not reported.



**Figure 1.** The upper, medium and lower right-hand (left-hand) panels contain the scatterplots of 10,000 (transformed) data from  $2\phi(x)\phi(y)\Phi(\lambda xy)$ , where  $\lambda \in \{1, 3, 5\}$ .

We conclude that perturbed independence distributions, by modeling the nonlinear association between random variables, might help in finding the nonlinear transformations that are maximally correlated to each other. A positive Pearson's correlation much lower than Spearman's rho and Kendall's tau hints for the presence of nonlinear association, whose analytical form might be estimated by looking for the maximally correlated nonlinear transformations of the random variables. This approach is particularly appropriate for the single index regression model  $Y = g(X) + \varepsilon$ , where the response variable  $Y$  is the sum of a smooth function  $g(\cdot)$  of the predictor  $X$  and the error term  $\varepsilon$ . When  $g(\cdot)$  is monotone,

its analytical form might be estimated by looking for the transformation  $g(X)$  that is maximally correlated with  $Y$ .



**Figure 2.** Histograms of the first transformed components of 10,000 data from  $2\phi(x)\phi(y)\Phi(\lambda xy)$ , where  $\lambda \in \{1, 2, 3, 4, 5, 6\}$ .

As remarked in the Introduction, kernel canonical correlation analysis (KCCA) and distance canonical correlation analysis (DCCA) are the two most popular generalizations of CCA aimed at dealing with nonlinear dependencies. A formal description of KCCA, based on Hilbert spaces and their inner products, might be found in the seminal papers by [32,33]. For most practical purposes, KCCA might be defined as the statistical method searching for linear projections of nonlinear functions of a random vector that are maximally correlated with linear projections of nonlinear functions of another random vector. Let  $\mathbb{F}$  be a class of  $p$ -dimensional random vectors whose  $i$ -th components are nonlinear functions of the  $p$ -dimensional random vector  $\mathbf{x}$ . Similarly, let  $\mathbb{G}$  be a class of  $q$ -dimensional random vectors whose  $i$ -th components are nonlinear functions of the  $q$ -dimensional random vector  $\mathbf{y}$ . Then KCCA looks for the random vectors  $\mathbf{f} \in \mathbb{F}$ ,  $\mathbf{g} \in \mathbb{G}$  and for the real vectors  $\mathbf{a} \in \mathbb{R}^p$ ,  $\mathbf{b} \in \mathbb{R}^q$  such that  $\mathbf{a}^\top \mathbf{f}$  and  $\mathbf{b}^\top \mathbf{g}$  are maximally correlated with each other.

In a nonparametric framework, the choice of the nonlinear functions may not be straightforward. On the other hand, in the perturbed independence framework, the theoretical and empirical results in this section suggest to set them equal to the conditional expectations:  $\mathbf{f} = E(\mathbf{x}|\mathbf{y})$  and  $\mathbf{g} = E(\mathbf{y}|\mathbf{x})$ . In particular, for the perturbed independence model

$$2\phi_p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Omega})\phi_q(\mathbf{y}; \boldsymbol{\nu}, \boldsymbol{\Gamma})\Phi\left\{(\mathbf{y} - \boldsymbol{\nu})^\top \boldsymbol{\Psi}(\mathbf{x} - \boldsymbol{\mu})\right\}$$

the suggested nonlinear functions of  $\mathbf{x}$  and  $\mathbf{y}$  are

$$\mathbf{f} = \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}} \frac{\boldsymbol{\Omega} \boldsymbol{\Psi}^\top (\mathbf{y} - \boldsymbol{\nu})}{\sqrt{1 + (\mathbf{y} - \boldsymbol{\nu})^\top \boldsymbol{\Psi} \boldsymbol{\Omega} \boldsymbol{\Psi}^\top (\mathbf{y} - \boldsymbol{\nu})}} \quad \text{and} \quad \mathbf{g} = \boldsymbol{\nu} + \sqrt{\frac{2}{\pi}} \frac{\boldsymbol{\Gamma} \boldsymbol{\Psi} (\mathbf{x} - \boldsymbol{\mu})}{\sqrt{1 + (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Psi}^\top \boldsymbol{\Gamma} \boldsymbol{\Psi} (\mathbf{x} - \boldsymbol{\mu})}}.$$

DCCA looks for two projections whose joint distribution differs the most from the product of their marginal distributions, where difference is measured by distance correlation. The distance correlation between the random variables  $X$  and  $Y$  with respect to the weight function  $w$  is

$$\mathcal{R}(X, Y, w) = \frac{\mathcal{V}(X, Y, w)}{\sqrt{\mathcal{V}(X, X, w)\mathcal{V}(Y, Y, w)}},$$

where  $\mathcal{V}^2(X, Y, w)$  is the distance covariance between  $X$  and  $Y$  with respect to  $w$ , as defined in the previous section. Hence the first canonical correlation between the  $p$ -dimensional random vector  $\mathbf{x}$  and the  $q$ -dimensional random vector  $\mathbf{y}$  is

$$\max_{\mathbf{a} \in \mathbb{R}^p} \max_{\mathbf{b} \in \mathbb{R}^q} \mathcal{R}(\mathbf{a}^\top \mathbf{x}, \mathbf{b}^\top \mathbf{y}, w).$$

For other distance canonical correlations, the distance canonical pairs and the distance canonical transformations are defined similarly to their CCA analogues.

A natural question to ask is whether CCA and DCCA lead to identical projections, under the assumption of perturbed independence. At present, we are unable to either prove or disprove this statement, which we conjecture to be true, under the assumptions of Theorem 6: increasing perturbing functions that increase more steeply are more likely to imply both higher Pearson and distance correlations. We plan to investigate this conjecture by means of both theoretical arguments and simulation studies.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The author would like to thank two anonymous reviewers for their very insightful comments. They greatly helped in improving the quality of the present paper.

**Conflicts of Interest:** The author declares no conflict of interest.

## Appendix A

**Proof of Theorem 1.** Let the  $p$ -dimensional random vector  $\mathbf{x}$  and the  $q$ -dimensional random vector  $\mathbf{y}$  have a perturbed independence distribution with components  $h(\cdot)$  and  $k(\cdot)$ , location vectors  $\boldsymbol{\mu}$  and  $\boldsymbol{\nu}$ , perturbing function  $\pi(\cdot)$  and association matrix  $\boldsymbol{\Psi}$ . For the sake of simplicity and without loss of generality we assume that the location vectors are null vectors and that the joint probability density function of  $\mathbf{x}$  and  $\mathbf{y}$  is absolutely continuous. The marginal probability density function of  $\mathbf{x}$  is

$$f(\mathbf{x}) = \int_{\mathbb{R}^q} 2h(\mathbf{x})k(\mathbf{y})\pi(\mathbf{y}^\top \boldsymbol{\Psi} \mathbf{x}) d\mathbf{y}.$$

Let  $\mathbb{A}^+$  and  $\mathbb{A}^-$  be the sets of  $q$ -dimensional real vectors whose first nonnull component are nonnegative and negative, so that  $\mathbb{A}^+ \cup \mathbb{A}^- = \mathbb{R}^q$ ,  $\mathbb{A}^+ \cap \mathbb{A}^- = \emptyset$  and

$$f(\mathbf{x}) = \int_{\mathbb{A}^+} 2h(\mathbf{x})k(\mathbf{y})\pi(\mathbf{y}^\top \mathbf{\Psi}\mathbf{x})d\mathbf{y} + \int_{\mathbb{A}^-} 2h(\mathbf{x})k(\mathbf{y})\pi(\mathbf{y}^\top \mathbf{\Psi}\mathbf{x})d\mathbf{y}.$$

If a nonnull  $q$ -dimensional real vector  $\mathbf{y}$  belongs to  $\mathbb{A}^+$  then  $-\mathbf{y}$  belongs to  $\mathbb{A}^-$ . By making the change of variable  $\mathbf{u} = -\mathbf{y}$  in the second integral we have

$$\int_{\mathbb{A}^-} 2h(\mathbf{x})k(\mathbf{y})\pi(\mathbf{y}^\top \mathbf{\Psi}\mathbf{x})d\mathbf{y} = \int_{\mathbb{A}^+} 2h(\mathbf{x})k(-\mathbf{u})\pi(-\mathbf{u}^\top \mathbf{\Psi}\mathbf{x})d\mathbf{u}.$$

By assumption, the functions  $k(\cdot)$  and  $\pi(\cdot)$  satisfy the identities  $k(\mathbf{v}) = k(-\mathbf{v})$  and  $\pi(a) = 1 - \pi(-a)$ :

$$\int_{\mathbb{A}^-} 2h(\mathbf{x})k(\mathbf{y})\pi(\mathbf{y}^\top \mathbf{\Psi}\mathbf{x})d\mathbf{y} = \int_{\mathbb{A}^+} 2h(\mathbf{x})k(\mathbf{u})\{1 - \pi(\mathbf{u}^\top \mathbf{\Psi}\mathbf{x})\}d\mathbf{u}.$$

The marginal density of  $\mathbf{x}$  is then

$$f(\mathbf{x}) = \int_{\mathbb{A}^+} 2h(\mathbf{x})k(\mathbf{y})\pi(\mathbf{y}^\top \mathbf{\Psi}\mathbf{x})d\mathbf{y} + \int_{\mathbb{A}^+} 2h(\mathbf{x})k(\mathbf{u})\{1 - \pi(\mathbf{u}^\top \mathbf{\Psi}\mathbf{x})\}d\mathbf{u} = h(\mathbf{x}).$$

The last identity follows from  $k(\cdot)$  being a symmetric probability density function. In a similar way it can be proved that the marginal probability density function of  $\mathbf{y}$  is  $k(\cdot)$ . The conditional probability density function of  $\mathbf{x}$  given  $\mathbf{y}$  is

$$f(\mathbf{x}|\mathbf{y}) = \frac{f(\mathbf{x}, \mathbf{y})}{f(\mathbf{y})} = \frac{2h(\mathbf{x})k(\mathbf{y})\pi(\mathbf{y}^\top \mathbf{\Psi}\mathbf{x})}{k(\mathbf{y})} = 2h(\mathbf{x})\pi(\mathbf{y}^\top \mathbf{\Psi}\mathbf{x}),$$

that is skew-symmetric with symmetric kernel  $h(\cdot)$ , null location vector, skewing function  $\pi(\cdot)$  and shape parameter  $\mathbf{\Psi}^\top \mathbf{y}$ . In a similar way it is possible to prove that the conditional probability density function of  $\mathbf{y}$  given  $\mathbf{x}$  is skew-symmetric with symmetric kernel  $k(\cdot)$ , null location vector, skewing function  $\pi(\cdot)$  and shape parameter  $\mathbf{x}^\top \mathbf{\Psi}^\top$ .  $\square$

**Proof of Theorem 2.** Let the joint distribution of the  $p$ -dimensional random vector  $\mathbf{x}$  and the  $q$ -dimensional random vector  $\mathbf{y}$  be

$$f(\mathbf{x}, \mathbf{y}) = 2h(\mathbf{x} - \boldsymbol{\mu})k(\mathbf{y} - \mathbf{v})\pi\{(\mathbf{y} - \mathbf{v})^\top \mathbf{\Psi}(\mathbf{x} - \boldsymbol{\mu})\}.$$

Further, let  $w(\mathbf{u}) = w(-\mathbf{u})$  be an even function of the  $p$ -dimensional real vector  $\mathbf{u}$ . We prove the theorem for an even function of  $\mathbf{x} - \boldsymbol{\mu}$  only, the proof for an even function of  $\mathbf{y} - \mathbf{v}$  being very similar. Without loss of generality we assume that  $\boldsymbol{\mu}$  and  $\mathbf{v}$  are null real vectors, that  $f(\mathbf{x}, \mathbf{y})$  is an absolutely continuous probability density function and that  $\mathbf{w} = w(\mathbf{x})$  is a  $k$ -dimensional random vector. The characteristic function of  $\mathbf{w}$  in the  $k$ -dimensional real vector  $\mathbf{t}$  is

$$\varphi_{\mathbf{w}}(\mathbf{t}) = E\{\exp(\mathbf{w}^\top \mathbf{t})\} = \int_{\mathbb{R}^p} \int_{\mathbb{R}^q} \exp\{w^\top(\mathbf{x})\mathbf{t}\}2h(\mathbf{x})k(\mathbf{y})\pi(\mathbf{y}^\top \mathbf{\Psi}\mathbf{x})d\mathbf{x}d\mathbf{y}.$$

Let  $\mathbb{S}^+$  and  $\mathbb{S}^-$  be the sets of  $p$ -dimensional real vectors whose first nonnull component are nonnegative and negative, so that  $\mathbb{S}^+ \cup \mathbb{S}^- = \mathbb{R}^p$ ,  $\mathbb{S}^+ \cap \mathbb{S}^- = \emptyset$  and

$$\begin{aligned} \varphi_{\mathbf{w}}(\mathbf{t}) &= \int_{\mathbb{S}^+} \int_{\mathbb{R}^q} \exp\{w^\top(\mathbf{x})\mathbf{t}\}2h(\mathbf{x})k(\mathbf{y})\pi(\mathbf{y}^\top \mathbf{\Psi}\mathbf{x})d\mathbf{x}d\mathbf{y} \\ &+ \int_{\mathbb{S}^-} \int_{\mathbb{R}^q} \exp\{w^\top(\mathbf{x})\mathbf{t}\}2h(\mathbf{x})k(\mathbf{y})\pi(\mathbf{y}^\top \mathbf{\Psi}\mathbf{x})d\mathbf{x}d\mathbf{y}. \end{aligned}$$

If a nonnull vector  $\mathbf{x}$  belongs to  $\mathbb{S}^+$  then  $-\mathbf{x}$  belongs to  $\mathbb{S}^-$ . By making the change of variable  $\mathbf{u} = -\mathbf{x}$  in the second integral we have

$$\int_{\mathbb{S}^-} \int_{\mathbb{R}^q} \exp\{w^\top(\mathbf{x})\mathbf{t}\} 2h(\mathbf{x})k(\mathbf{y})\pi(\mathbf{y}^\top \mathbf{\Psi} \mathbf{x}) d\mathbf{x}d\mathbf{y} =$$

$$\int_{\mathbb{S}^+} \int_{\mathbb{R}^q} \exp\{w^\top(-\mathbf{x})\mathbf{t}\} 2h(-\mathbf{u})k(\mathbf{y})\pi(-\mathbf{y}^\top \mathbf{\Psi} \mathbf{u}) d\mathbf{u}d\mathbf{y}.$$

By assumption, the functions  $h(\cdot)$ ,  $\pi(\cdot)$  and  $w(\cdot)$  satisfy the identities  $h(\mathbf{v}) = h(-\mathbf{v})$ ,  $\pi(a) = 1 - \pi(-a)$  and  $w(\mathbf{u}) = w(-\mathbf{u})$ , so that

$$\int_{\mathbb{S}^-} \int_{\mathbb{R}^q} \exp\{w^\top(\mathbf{x})\mathbf{t}\} 2h(\mathbf{x})k(\mathbf{y})\pi(\mathbf{y}^\top \mathbf{\Psi} \mathbf{x}) d\mathbf{x}d\mathbf{y} =$$

$$\int_{\mathbb{S}^+} \int_{\mathbb{R}^q} \exp\{w^\top(\mathbf{u})\mathbf{t}\} 2h(\mathbf{u})k(\mathbf{y})\{1 - \pi(\mathbf{y}^\top \mathbf{\Psi} \mathbf{u})\} d\mathbf{u}d\mathbf{y}.$$

The characteristic function of  $\mathbf{w}$  is then

$$\varphi_{\mathbf{w}}(\mathbf{t}) = \int_{\mathbb{S}^+} \int_{\mathbb{R}^q} \exp\{w^\top(\mathbf{x})\mathbf{t}\} 2h(\mathbf{x})k(\mathbf{y})\pi(\mathbf{y}^\top \mathbf{\Psi} \mathbf{x}) d\mathbf{x}d\mathbf{y}$$

$$+ \int_{\mathbb{S}^+} \int_{\mathbb{R}^q} \exp\{w^\top(\mathbf{x})\mathbf{t}\} 2h(\mathbf{u})k(\mathbf{y})\{1 - \pi(\mathbf{y}^\top \mathbf{\Psi} \mathbf{u})\} d\mathbf{u}d\mathbf{y} =$$

$$2 \int_{\mathbb{S}^+} \exp\{w^\top(\mathbf{x})\mathbf{t}\} h(\mathbf{x})k(\mathbf{y}) d\mathbf{x}d\mathbf{y} = 2 \int_{\mathbb{S}^+} \exp\{w^\top(\mathbf{x})\mathbf{t}\} h(\mathbf{x}) d\mathbf{x}.$$

The definitions of  $\mathbb{S}^+$  and  $\mathbb{S}^-$ , together with the identity  $h(\mathbf{v}) = h(-\mathbf{v})$  yield

$$\int_{\mathbb{S}^+} \exp\{w^\top(\mathbf{x})\mathbf{t}\} h(\mathbf{x}) d\mathbf{x} = \int_{\mathbb{S}^-} \exp\{w^\top(\mathbf{x})\mathbf{t}\} h(\mathbf{x}) d\mathbf{x}.$$

The characteristic function of  $\mathbf{w}$  is then

$$\varphi_{\mathbf{w}}(\mathbf{t}) = \int_{\mathbb{R}^p} \exp\{w^\top(\mathbf{x})\mathbf{t}\} h(\mathbf{x}) d\mathbf{x}.$$

which does not depend neither on the association matrix  $\mathbf{\Psi}$  nor on the perturbing function  $\pi(\cdot)$ . In order to prove that  $\mathbf{w}$  and  $\mathbf{y}$  are stochastically independent we consider their joint characteristic function

$$\varphi_{\mathbf{w},\mathbf{y}}(\mathbf{t}, \mathbf{r}) = E\left(\exp\left[\{w^\top(\mathbf{x})\mathbf{t} + \mathbf{y}^\top \mathbf{r}\}\right]\right) =$$

$$\int_{\mathbb{R}^p} \int_{\mathbb{R}^q} \exp\{w^\top(\mathbf{x})\mathbf{t} + \mathbf{y}^\top \mathbf{r}\} 2h(\mathbf{x})k(\mathbf{y})\pi(\mathbf{y}^\top \mathbf{\Psi} \mathbf{x}) d\mathbf{x}d\mathbf{y},$$

where  $\mathbf{r}$  is a  $q$ -dimensional real vector. An argument very similar to the one in the first part of the proof yields

$$\varphi_{\mathbf{w},\mathbf{y}}(\mathbf{t}, \mathbf{r}) = E\left[\exp\{w^\top(\mathbf{x})\mathbf{t}\}\right] \cdot E\left\{\exp(\mathbf{y}^\top \mathbf{r})\right\} = \varphi_{\mathbf{w}}(\mathbf{t}) \cdot \varphi_{\mathbf{y}}(\mathbf{r}),$$

thus implying that  $\mathbf{w}$  and  $\mathbf{y}$  are stochastically independent.  $\square$

**Proof of Theorem 3.** Without loss of generality we assume that the location vectors are null vectors. We prove the theorem in the special case of both  $\mathbf{x}$  and  $\mathbf{y}$  being bivariate



vectors, the proof for the general case being very similar but much more cumbersome in the notation. Let the joint distribution of  $\mathbf{z}$  and  $\mathbf{w}$  be

$$f(z_1, z_2, w_1, w_2) = 2h(z_1, z_2)k(w_1, w_2)\pi(\lambda_1 z_1 w_1 + \lambda_2 z_2 w_2).$$

The marginal distribution  $f(z_1, w_2)$  of  $z_1$  and  $w_2$  is

$$\begin{aligned} & \int_{-\infty}^{+\infty} \left\{ \int_{-\infty}^{+\infty} 2h(z_1, z_2)k(w_1, w_2)\pi(\lambda_1 z_1 w_1 + \lambda_2 z_2 w_2) dz_1 \right\} dw_2 = \\ & \int_0^{+\infty} \left\{ \int_0^{+\infty} 2h(z_1, z_2)k(w_1, w_2)\pi(\lambda_1 z_1 w_1 + \lambda_2 z_2 w_2) dz_1 \right\} dw_2 + \\ & \int_0^{+\infty} \left\{ \int_{-\infty}^0 2h(z_1, z_2)k(w_1, w_2)\pi(\lambda_1 z_1 w_1 + \lambda_2 z_2 w_2) dz_1 \right\} dw_2 + \\ & \int_{-\infty}^0 \left\{ \int_0^{+\infty} 2h(z_1, z_2)k(w_1, w_2)\pi(\lambda_1 z_1 w_1 + \lambda_2 z_2 w_2) dz_1 \right\} dw_2 + \\ & \int_{-\infty}^0 \left\{ \int_{-\infty}^0 2h(z_1, z_2)k(w_1, w_2)\pi(\lambda_1 z_1 w_1 + \lambda_2 z_2 w_2) dz_1 \right\} dw_2. \end{aligned}$$

By assumption, both  $h(\cdot)$  and  $k(\cdot)$  are sign-symmetric distributions, that is, they satisfy the equalities

$$h(z_1, z_2) = h(-z_1, z_2) = h(z_1, -z_2) = h(-z_1, -z_2) \text{ and}$$

$$k(w_1, w_2) = k(-w_1, w_2) = k(w_1, -w_2) = k(-w_1, -w_2).$$

The change of variables  $u_1 = -z_1$  and  $u_2 = -w_2$  yields

$$\begin{aligned} & \int_{-\infty}^0 \left\{ \int_{-\infty}^0 h(z_1, z_2)k(w_1, w_2)\pi(\lambda_1 z_1 w_1 + \lambda_2 z_2 w_2) dz_1 \right\} dw_2 = \\ & \int_0^{+\infty} \left[ \int_0^{+\infty} h(z_1, z_2)k(w_1, w_2)\{1 - \pi(\lambda_1 z_1 w_1 + \lambda_2 z_2 w_2)\} dz_1 \right] dw_2. \end{aligned}$$

The change of variable  $u_1 = -z_1$  yields

$$\begin{aligned} & \int_0^{+\infty} \left\{ \int_{-\infty}^0 2h(z_1, z_2)k(w_1, w_2)\pi(\lambda_1 z_1 w_1 + \lambda_2 z_2 w_2) dz_1 \right\} dw_2 = \\ & \int_0^{+\infty} \left[ \int_0^{+\infty} 2h(u_1, z_2)k(w_1, w_2)\pi(-\lambda_1 u_1 w_1 + \lambda_2 z_2 w_2) dz_1 \right] du_2 = \\ & \int_0^{+\infty} \left[ \int_0^{+\infty} 2h(u_1, z_2)k(w_1, w_2)\{1 - \pi(\lambda_1 u_1 w_1 - \lambda_2 z_2 w_2)\} dz_1 \right] du_2. \end{aligned}$$

The change of variable  $u_2 = -w_2$  yields

$$\int_{-\infty}^0 \left\{ \int_0^{+\infty} 2h(z_1, z_2)k(w_1, w_2)\pi(\lambda_1 z_1 w_1 + \lambda_2 z_2 w_2) dz_1 \right\} dw_2$$

$$\int_0^{+\infty} \left\{ \int_0^{+\infty} 2h(z_1, z_2)k(w_1, u_2)\pi(\lambda_1 z_1 w_1 - \lambda_2 z_2 u_2) dz_1 \right\} dw_2.$$

We can show that the random variables  $Z_1$  and  $W_2$  are independent:

$$f(z_1, w_2) = \int_{-\infty}^{+\infty} \left\{ \int_{-\infty}^{+\infty} 2h(z_1, z_2)k(w_1, w_2)\pi(\lambda_1 z_1 w_1 + \lambda_2 z_2 w_2) dz_1 \right\} dw_2 =$$

$$2 \int_0^{+\infty} \left\{ \int_0^{+\infty} 2h(z_1, z_2)k(w_1, w_2) dz_1 \right\} dw_2 = \int_{-\infty}^{+\infty} h(z_1, z_2) dz_1 \int_{-\infty}^{+\infty} k(w_1, w_2) dw_2.$$

□

**Proof of Theorem 4.** For the sake of simplicity and without loss of generality we assume that the perturbing function  $\pi(\cdot)$  is the cumulative distribution function of a continuous distribution. The random variables  $X$  and  $Y$  are either positively or negatively dependent if they satisfy the inequalities

$$\Pr(X < a, Y < b) \geq \Pr(X < a) \cdot \Pr(Y < b) \text{ or } \Pr(X < a, Y < b) \leq \Pr(X < a) \cdot \Pr(Y < b)$$

for all real values  $a$  and  $b$ , with strict inequalities holding for at least some  $a$  and  $b$ . We only prove the theorem for positive values of  $a, b$  and  $\lambda$ : the proofs of the remaining cases are very similar. The integral representation of the joint probability that  $X$  and  $Y$  are smaller than  $a$  and  $b$  is

$$\Pr(X < a, Y < b) = \int_{-\infty}^b \int_{-\infty}^a 2h(x)k(y)\pi(\lambda xy) dx dy.$$

Convenient partitioning of the integration region leads to the identity

$$\int_{-\infty}^b \int_{-\infty}^a 2h(x)k(y)\pi(\lambda xy) dx dy = \int_{-\infty}^{-b} \int_{-\infty}^{-a} 2h(x)k(y)\pi(\lambda xy) dx dy$$

$$+ \int_0^b \int_{-\infty}^0 2h(x)k(y)\pi(\lambda xy) dx dy + \int_{-b}^0 \int_{-\infty}^0 2h(x)k(y)\pi(\lambda xy) dx dy$$

$$+ \int_{-\infty}^0 \int_0^a 2h(x)k(y)\pi(\lambda xy) dx dy + \int_{-\infty}^0 \int_{-a}^0 2h(x)k(y)\pi(\lambda xy) dx dy$$

$$+ \int_0^b \int_0^a 2h(x)k(y)\pi(\lambda xy) dx dy + \int_{-b}^0 \int_{-a}^0 2h(x)k(y)\pi(\lambda xy) dx dy.$$

By making the transformations  $z = -x$  and  $w = -y$  we obtain

$$\begin{aligned} \int_{-\infty}^b \int_{-\infty}^a 2h(x)k(y)\pi(\lambda xy) dx dy &= \int_{-\infty}^{-b} \int_{-\infty}^{-a} 2h(-z)k(-w)\pi(\lambda zw) dz dw \\ &+ \int_0^b \int_{-\infty}^0 2h(x)k(y)\pi(\lambda xy) dx dy - \int_b^0 \int_{-\infty}^0 2h(x)k(-w)\pi(-\lambda xw) dx dw \\ &+ \int_{-\infty}^0 \int_0^a 2h(x)k(y)\pi(\lambda xy) dx dy - \int_{-\infty}^0 \int_a^0 2h(-z)k(y)\pi(-\lambda zy) dz dy \\ &+ \int_0^b \int_0^a 2h(x)k(y)\pi(\lambda xy) dx dy - \int_b^0 \int_a^0 2h(-z)k(-w)\pi(\lambda zw) dz dw. \end{aligned}$$

We use the identities  $h(x) = h(-x)$ ,  $h(x) = h(-x)$ ,  $\pi(-\lambda xy) = 1 - \pi(\lambda xy)$  to obtain

$$\begin{aligned} \int_{-\infty}^b \int_{-\infty}^a 2h(x)k(y)\pi(\lambda xy) dx dy &= \int_{-\infty}^{-b} \int_{-\infty}^{-a} 2h(z)k(w)\pi(\lambda zw) dz dw \\ &+ \int_0^b \int_{-\infty}^0 2h(x)k(y)\pi(\lambda xy) dx dy - \int_b^0 \int_{-\infty}^0 2h(x)k(w)[1 - \pi(\lambda xw)] dx dw \\ &+ \int_{-\infty}^0 \int_0^a 2h(x)k(y)\pi(\lambda xy) dx dy - \int_{-\infty}^0 \int_a^0 2h(z)k(y)[1 - \pi(\lambda zy)] dz dy \\ &+ \int_0^b \int_0^a 2h(x)k(y)\pi(\lambda xy) dx dy - \int_b^0 \int_a^0 2h(z)k(w)\pi(\lambda zw) dz dw. \end{aligned}$$

Apply now basic properties of integrals:

$$\begin{aligned} \int_{-\infty}^b \int_{-\infty}^a 2h(x)k(y)\pi(\lambda xy) dx dy &= \int_b^{\infty} \int_a^{\infty} 2h(x)k(y)\pi(\lambda xy) dx dy \\ &+ \int_0^b \int_{-\infty}^0 2h(x)k(y)\pi(\lambda xy) dx dy + \int_0^b \int_{-\infty}^0 2h(x)k(y)[1 - \pi(\lambda xy)] dx dy \\ &+ \int_{-\infty}^0 \int_0^a 2h(x)k(y)\pi(\lambda xy) dx dy + \int_{-\infty}^0 \int_0^a 2h(x)k(y)[1 - \pi(\lambda xy)] dx dy \\ &+ \int_0^b \int_0^a 2h(x)k(y)\pi(\lambda xy) dx dy - \int_0^b \int_0^a 2h(x)k(y)\pi(\lambda xy) dx dy. \end{aligned}$$

A little algebra simplifies the above integrals:

$$\begin{aligned} \int_{-\infty}^b \int_{-\infty}^a 2h(x)k(y)\pi(\lambda xy) dx dy &= \int_b^{\infty} \int_a^{\infty} 2h(x)k(y)\pi(\lambda xy) dx dy \\ &= 2 \left\{ \int_{-\infty}^0 h(x) dx \right\} \left\{ \int_0^b k(y) dy \right\} + 2 \left\{ \int_{-\infty}^0 k(y) dy \right\} \left\{ \int_0^a h(x) dx \right\}. \end{aligned}$$

The above integral identity leads to the probabilistic identity

$$\Pr(X < a, Y < b) = \Pr(X > a, Y > b) + 2\Pr(X < 0)P(0 < Y < b) + 2\Pr(0 < X < a)P(Y < 0).$$

The probabilities  $\Pr(X < 0)$ ,  $\Pr(0 < X < a)$ ,  $\Pr(Y < 0)$  and  $\Pr(0 < Y < b)$  do not depend on the perturbing function  $\pi(\cdot)$ , while the probability  $\Pr(X > a, Y > b)$  does. By assumption,  $\pi(\cdot)$  is the cdf of a symmetric, continuous distribution, so that  $\pi(q) > 0.5$  for any positive real value  $q$ . It follows that

$$\begin{aligned} \Pr(X > a, Y > b) &= \int_b^\infty \int_a^\infty 2h(x)k(y)\pi(\lambda xy)dx dy \\ &> \int_b^\infty \int_a^\infty h(x)k(y)dx dy = \Pr(X > a) \Pr(Y > b). \end{aligned}$$

Logical steps similar to the above ones lead to the inequality

$$\Pr(X > -a, Y > -b) > \Pr(X > -a) \cdot \Pr(Y > -b)$$

for any two negative values  $-a$  and  $-b$ . The joint distribution of  $X$  and  $Y$  is centrally symmetric, so that the random variables  $-X$  and  $-Y$  have the same joint distribution of  $X$  and  $Y$ :

$$\Pr(-X > -a, -Y > -b) = \Pr(X < a, Y < b) >$$

$$\Pr(-X > -a) \cdot \Pr(-Y > -b) = \Pr(X < a) \cdot \Pr(Y < b).$$

□

**Proof of Theorem 5.** Let  $h_i(\cdot)$  and  $k_i(\cdot)$  be the probability density functions of the random variables  $U_i$  and  $V_i$ , for  $i \in \{1, \dots, n\}$ . Further, let the random variables  $U_1, \dots, U_n$  and  $V_1, \dots, V_n$  be mutually independent, so that their joint probability density function is

$$f(u_1, \dots, u_n, v_1, \dots, v_n) = \left\{ \prod_{i=1}^r h_i(u_i) \right\} \left\{ \prod_{i=1}^r k_i(v_i) \right\}.$$

By assumption  $\pi(\cdot)$  is a strictly increasing perturbing function, so that we can define a random variable  $Q$ , which is symmetric at the origin and whose cumulative density function is  $\pi(\cdot)$ :

$$\Pr(Q \leq a) = F_Q(a) = \pi(a) = 1 - \pi(a), \text{ where } a \in \mathbb{R}.$$

By an argument virtually identical to the one used in Section 2 to generate a bivariate perturbed independence distribution we obtain

$$(Z_1, \dots, Z_n, W_1, \dots, W_n) \sim (U_1, \dots, U_n, V_1, \dots, V_n) | Q \leq \sum_{i=1}^r \lambda_i U_i V_i.$$

Simple algebra leads to the distributional identity

$$(Z_j, W_j) \sim (U_j, V_j) | Q - \sum_{i \neq j} \lambda_i U_i V_i \leq \lambda_j U_j V_j.$$

Since the random variables  $Q, U_1, \dots, U_n, V_1, \dots, V_n$  are continuous, symmetric and mutually independent, their linear combination

$$Q_j = Q - \sum_{i \neq j} \lambda_i U_i V_i$$

is continuous and symmetric, too. Let  $\pi_j(\cdot)$  be the cumulative density function of  $Q_j$ , so that  $\pi_j(\cdot)$  is a strictly increasing perturbing function, too:

$$\pi_j(a) < \pi_j(b) \text{ if } a < b \text{ and } 0 \leq \pi_j(-a) = 1 - \pi_j(a) \leq 1.$$

By further application of the generating argument described in Section 2 we conclude that the joint distribution of  $Z_j$  and  $W_j$  have a joint distribution, which is perturbed independent:

$$f_j(z, w) = 2h_j(z)k_j(w)\pi_j(\lambda_j zw).$$

□

**Proof of Theorem 6.** By assumption, the association parameter is nonnull and the perturbing function is increasing. Therefore, by Theorem 5,  $X$  and  $Y$  are either positively dependent (if the association parameter is positive) or negatively dependent (if the association parameter is negative). By ordinary properties of monotone association, the conditional expectation  $g(x) = E(Y|X = x)$  of  $Y|X = x$  is an increasing and nonconstant function of  $x$  if the association parameter is positive and a decreasing function of  $x$  if the association parameter is negative. In either case,  $g(\cdot)$  is a monotone and nonconstant function.

By definition, a real valued function  $q(\cdot)$  is odd if it satisfies the identity  $q(-a) = -q(a)$  for any real value  $a$  belonging to the domain of  $q(\cdot)$ . As remarked at the end of Section 2, a perturbed independence distribution is centrally symmetric: the bivariate random vectors  $(X, Y)^\top$  and  $(-X, -Y)^\top$  are identically distributed, so that

$$E(Y|X = x) = g(x) = E(-Y| -X = x) = -E(Y|X = -x) = -g(-x).$$

As a direct consequence, the identity  $-g(x) = g(-x)$  holds true for any real value  $x$  belonging to the support of  $X$ .

By assumption, both  $X$  and  $Y$  have finite second moments and their probability density functions are symmetric at the origin. Hence their expectations equal zero:  $E(X) = E(Y) = 0$ . Without loss of generality we can assume that their variances equal one:  $E(X^2) = E(Y^2) = 1$ . Finally, let  $\rho$  be the correlation between  $X$  and  $Y$ :  $E(XY) = \rho$ . Were  $g(x)$  a nonlinear function, it would coincide with  $\rho x$ , that is the linear regression of  $Y$  on  $X$ . By Theorem 1, the conditional distribution of  $Y$  given  $X = x$  is skew-symmetric, thus implying the identity  $E(Y^2|X = x) = 1$ . However,  $\rho x$  is an unbounded function but  $g(x) = E(Y|X = x)$  is not, since it must satisfy the inequality

$$1 = E(Y^2|X = x) \geq E^2(Y|X = x) = g^2(x).$$

We therefore proved by contradiction that  $g(\cdot)$  is a nonlinear function.

By the minimizing property of the expected value we have

$$g(x) = E(Y|X = x) = \arg \min_{a \in \mathbb{A}} E \left[ \{Y - a(X)\}^2 \middle| X = x \right].$$

where  $\mathbb{A}$  is the set of all real valued functions defined on the support of  $X$ . As a direct consequence, we have

$$E \left[ \{Y - g(X)\}^2 \middle| X = x \right] \leq E \left\{ \left[ (Y - X)^2 \middle| X = x \right] \right\}.$$

By taking expectation with respect to  $X$  we have

$$E \left[ \{Y - g(X)\}^2 \right] \leq E \left\{ \left[ (Y - X)^2 \right] \right\}.$$

By recalling that  $E(X) = E(Y) = 0$  and  $E(X^2) = E(Y^2) = 1$  we therefore obtain the inequality

$$E\left[\{Y - g(X)\}^2\right] \leq 2 - 2\rho.$$

By expanding the squares we obtain

$$E\left[\{Y - g(X)\}^2\right] = E\left(Y^2\right) - 2E\{Yg(X)\} + E\left\{g^2(X)\right\} \leq 2 - 2\rho,$$

which in turn leads to the inequality  $E\{Yg(X)\} \geq \rho$  by noticing that  $E\{g^2(X)\} < 1 = E(Y^2)$ , by ordinary properties of variance decomposition. We further apply the same inequality to obtain

$$\text{corr}\{g(X), Y\} = \frac{E\{Yg(X)\}}{E(Y^2)E\{g^2(X)\}} > E\{Yg(X)\} \geq \rho,$$

which implies the strict inequality  $\text{corr}\{g(X), Y\} > \rho$ , thus completing the proof.  $\square$

## References

1. Yoon, G.; Carroll, R.; Gayanova, I. Sparse semiparametric canonical correlation analysis for data of mixed types. *Biometrika* **2020**, *107*, 609–625. [\[CrossRef\]](#)
2. Hardoon, D.; Szedmak, S.; Shawe-Taylor, J. Canonical correlation analysis: An overview with application to learning methods. *Neural Comp.* **2004**, *16*, 2639–2664. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Zoh, R.; Mallick, B.; Ivanov, I.; Baladandayuthapani, V.; Manyam, G.; Chapkin, R.; Lampe, J.; Carroll, R.J. PCAN: Probabilistic correlation analysis of two non-normal data sets. *Biometrics* **2016**, *72*, 1358–1368. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Fukumizu, K.; Bach, F.; Gretton, A. Statistical Consistency of Kernel Canonical Correlation Analysis. *J. Mach. Learn. Res.* **2007**, *8*, 361–383.
5. Hu, W.; Zhang, A.; Cai, B.; Calhoun, V.; Wang, Y.P. Distance canonical correlation analysis with application to an imaging-genetic study. *J. Med. Imag* **2019**, *6*, 026501. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Arnold, B.; Castillo, E.; Sarabia, J. Conditionally specified distributions: An introduction. *Stat. Sci.* **2001**, *16*, 249–274. [\[CrossRef\]](#)
7. Arnold, B.; Castillo, E.; Sarabia, J. Conditionally specified multivariate skewed distributions. *Sankhya Ser. A* **2002**, *64*, 206–226.
8. Adcock, C. Copulaesque Versions of the Skew-Normal and Skew-Student Distributions. *Symmetry* **2021**, *13*, 815. [\[CrossRef\]](#)
9. Serfling, R.J. Multivariate symmetry and asymmetry. In *Encyclopedia of Statistical Sciences*, 2nd ed.; Kotz, S., Read, C.B., Balakrishnan, N., Vidakovic, B., Eds.; Wiley: New York, NY, USA, 2006.
10. Genton, M.; Loperfido, N. Generalized Skew-Elliptical Distributions and their Quadratic Forms. *Ann. Inst. Statist. Math.* **2005**, *57*, 389–401. [\[CrossRef\]](#)
11. Jiménez-Gamero, M.; Alba-Fernández, M.; Jodrá, P.; Chalco-Cano, Y. Testing for the symmetric component in skew distributions. *Math. Methods Appl. Sci.* **2016**, *39*, 4713–4722. [\[CrossRef\]](#)
12. Ley, C.; Paindaveine, D. On the singularity of skew-symmetric models. *J. Multiv. Anal.* **2010**, *101*, 1434–1444. [\[CrossRef\]](#)
13. Hallin, M.; Ley, C. Skew-symmetric distributions and Fisher information: a tale of two densities. *Bernoulli* **2012**, *18*, 747–763. [\[CrossRef\]](#)
14. Loperfido, N. A New Kurtosis Matrix, with Statistical Applications. *Lin. Alg. Appl.* **2017**, *512*, 1–17. [\[CrossRef\]](#)
15. Loperfido, N. A Note on Skew-Elliptical Distributions and Linear Functions of Order Statistics. *Stat. Probab. Lett.* **2008**, *78*, 3184–3186. [\[CrossRef\]](#)
16. Kollo, T. Multivariate skewness and kurtosis measures with an application in ICA. *J. Multivar. Anal.* **2008**, *99*, 2328–2338. [\[CrossRef\]](#)
17. Mardia, K. Measures of multivariate skewness and kurtosis with applications. *Biometrika* **1970**, *57*, 519–530. [\[CrossRef\]](#)
18. Loperfido, N. Some theoretical properties of two kurtosis matrices, with application to invariant coordinate selection. *J. Multivar. Anal.* **2021**, in press.
19. Loperfido, N. Skewness-Based Projection Pursuit: A Computational Approach. *Comp. Statist. Data Anal.* **2018**, *120*, 42–57. [\[CrossRef\]](#)
20. Székely, G.; Rizzo, M.; Bakirov, N. Measuring and testing dependence by correlation of distances. *Ann. Statist.* **2007**, *35*, 2769–2794. [\[CrossRef\]](#)
21. Lehmann, E. Some concepts of dependence. *Ann. Math. Statist.* **1966**, *37*, 1137–1153. [\[CrossRef\]](#)
22. Kendall, M. *Rank Correlation Methods*; Griffin: London, UK, 1962.
23. Kimeldorf, G.; Sampson, A. Positive dependence orderings. *Ann. Inst. Statist. Math.* **1987**, *39*, 113–128. [\[CrossRef\]](#)
24. Kimeldorf, G.; Sampson, A. A framework for positive dependence. *Ann. Inst. Statist. Math.* **1989**, *41*, 31–45.
25. Kruskal, W. Ordinal measures of association. *J. Am. Statist. Assoc.* **1958**, *53*, 814–861. [\[CrossRef\]](#)



26. Nelsen, R. On measures of association as measures of positive dependence. *Statist. Prob. Lett.* **1992**, *14*, 269–274. [[CrossRef](#)]
27. Tchen, A. Inequalities for distribution functions with given marginals. *Ann. Prob.* **1980**, *8*, 814–827. [[CrossRef](#)]
28. De Luca, G.; Loperfido, N. A Skew-in-Mean GARCH Model for Financial Returns. In *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*; Genton, M.G., Ed.; Chapman & Hall, CRC: Boca Raton, FL, USA, 2004; pp. 205–222.
29. De Luca, G.; Genton, M.; Loperfido, N. A Multivariate Skew-Garch Model. In *Advances in Econometrics: Econometric Analysis of Economic and Financial Time Series, Part A (Special volume in Honor of Robert Engle and Clive Granger, the 2003 Winners of the Nobel Prize in Economics)*; Terrell, D., Ed.; Elsevier: Oxford, UK, 2006; Volume 20, pp. 33–56.
30. De Luca, G.; Loperfido, N. Modelling Multivariate Skewness in Financial Returns: A SGARCH Approach. *Eur. J. Financ.* **2015**, *21*, 1113–1131. [[CrossRef](#)]
31. Cox, D.; Wermuth, N. On the Calculation of Derived Variables in the Analysis of Multivariate Responses. *J. Multivar. Anal.* **1992**, *42*, 162–170. [[CrossRef](#)]
32. Lai, P.; Fyfe, C. Kernel and nonlinear canonical correlation analysis. *Int. J. Neural Syst.* **2000**, *10*, 365–377. [[CrossRef](#)]
33. Akaho, S. A kernel method for canonical correlation analysis. In Proceedings of the International Meeting on Psychometric Society (IMPS2001), Osaka, Japan, 15–19 July 2001.