

Nonparametric Tests for Independence – A Review and Comparative Simulation Study with an Application to Malnutrition Data in India

Helmut Herwartz* Simone Maxand†

August 2018

Abstract

The detection of dependence structures within a set of random variables provides a valuable basis for a detailed subsequent investigation of their relationships. Nonparametric tests for independence require only basic assumptions on the marginal or joint distribution of the involved variables. In this paper, we review nonparametric tests of independence in bivariate as well as multivariate settings which are throughout ready-to-use, i.e., implemented in R packages. Highlighting their distinct empirical size and power properties in various small sample settings, our analysis supports an analyst in deciding for a most adequate test conditional on underlying distributional settings or data characteristics. Avoiding restrictive moment conditions, the copula based Cramér-von Mises distance of Genest & Rémillard (2004) is remarkably robust under the null hypothesis and powerful under diverse settings that are in line with the alternative hypothesis. Based on distinguished test outcomes in small samples, we detect nonlinear dependence structures between childhood malnutrition indices and possible determinants in an empirical application for India.

Keywords: Tests for independence; nonparametric methods; multivariate independence; spatial ranks; empirical copula; distance covariance.

Mathematics Subject Classification (2010): 62G10; 62H15; 62P10

*Department of Economics, University of Goettingen

†Corresponding author, Department of Political and Economic Studies, University of Helsinki, P. O. Box 17, FI-00014 University of Helsinki, Finland, e-mail: simone.maxand@helsinki.fi.

1 Introduction

Statistical analyses mostly target at the identification and quantification of dependence structures between the variables of interest. Yet, dependence between random variables can be present in various (e.g., linear or nonlinear) forms. Most commonly, analysts apply standard linear regression models presuming linear dependence structures. Whereas classical procedures, such as Pearson's correlation coefficient (e.g. Pearson, 1920) or Wilks' test (Wilks, 1935), diagnose linear dependence in a parametric framework, they might fail to detect nonlinear and nonmonotone dependence structures. Therefore, nonparametric tests aim at keeping prior assumptions on the variables' distribution under the null hypothesis and their relation under the alternative hypothesis at a minimum.

Classical nonparametric approaches have been developed to test for monotone, but not necessarily linear, bivariate dependence structures by means of ranks. Popular representatives for rank based dependence measures are Kendall's tau (Kendall, 1938) and Spearman's rho (Spearman, 1904). Such bivariate dependence tests, however, might lack consistency under several dependence structures (see Rémillard, 2014, for an example). Against this background, various tests for independence have been developed more recently. These tests are supposed to provide powerful tools to detect various forms of dependence especially if more than two random variables are considered.

As noticed by Josse & Holmes (2014), several test procedures are concurrently employed in distinct research communities. Suggestions of new tests are typically accompanied with comparative evidence gathered from stylized Monte Carlo experiments which use specific types of data (either under the null hypothesis or with regard to particular alternatives). For instance, Josse & Holmes (2014) compare a nonparametric approach based on distance covariances with a multivariate extension of Pearson's correlation coefficient for linear dependence. Similarly, Siqueira Santos et al. (2013) compare nonparametric tests with a focus on nonlinear dependence structures typically present in the gene expression literature. Noticing that such comparisons might miss important characteristics of various independence diagnostics under diverse frameworks of data generation, we provide a comprehensive overview on the diversity of nonparametric tests suggested in the recent literature. With particular attention on those procedures that are applicable in multivariate samples, we categorize the tests in regard to their underlying theoretical framework, and distinguish multivariate approaches based on

spatial signs and ranks, empirical copulas and distance covariances. Along these lines, we consider representative tests which are examined in more detail. Studying simplified, though representative, scenarios for the generation of bivariate and multivariate samples allows to trace the test performances (in finite samples) back to essential characteristics of the data. Since alternative nonparametric tests rely on distinct measures of dependence, our work (i) highlights the signaling content of rival dependence diagnostics under diverse dependence patterns, and (ii) points to the scope of alternative tests under more complex data structures.

In an application to data of childhood malnutrition in India we further illustrate the performance of the tests. We consider a standard regression scenario investigating the explanatory content of several variables on one (resp. two) outcome variables. Specifically, we examine the influence of certain characteristics of the child and it's mother on childhood malnutrition. By means of nonparametric independence tests we diagnose the dependence between malnutrition indices in a bivariate setting, and consider dependence between the bivariate malnutrition index and potential determinants by means of tests of groupwise (in)dependence. The nonparametric framework can be exploited to identify nonlinear and nonmonotone dependence structures as a cornerstone for further analysis of the explicit relation between child malnutrition and its possible determinants.

In the next section we describe distinguished dependence structures which might be present within a set of p random variables. In Section 3, we briefly characterize the considered test procedures along with some extensions and describe their theoretical background. Section 4 provides the simulation results, followed by the empirical example in Section 5. Section 6 concludes.

Throughout we use the following notation: Univariate continuous real valued random variables are denoted by $x_1, \dots, x_p \in \mathbb{R}$. A set of these random variables of size p_1 and p_2 is denoted by $\mathbf{x}_1 = (x_1, \dots, x_{p_1}) \in \mathbb{R}^{p_1}$ and $\mathbf{x}_2 = (x_1, \dots, x_{p_2}) \in \mathbb{R}^{p_2}$, respectively. The associated marginal distribution functions are F_{x_k} for $k = 1, \dots, p$, and $F_{\mathbf{x}_1}, F_{\mathbf{x}_2}$. Furthermore, the joint distribution functions are F_{x_1, \dots, x_p} (for the first two variables F_{x_1, x_2}) and $F_{\mathbf{x}_1, \mathbf{x}_2}$, respectively. Sample observations are indexed with $i = 1, 2, \dots, n$, such that n is the sample size. A random sample of, for instance, variable x_1 is $\{x_{1,1}, \dots, x_{1,n}\}$. Furthermore, a random sample of the set of variables \mathbf{x}_1 consists of observations $\mathbf{x}_{1,i} = (x_{11,i}, \dots, x_{1p_1,i})'$ for $i \in \{1, \dots, n\}$. The rank of observation $i = 1, \dots, n$ of variable x_k , $k = 1, \dots, p$, is denoted as $R_i^{(k)}$.

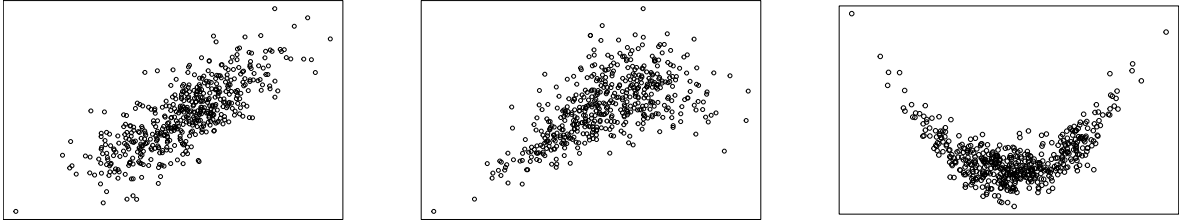


Figure 1: Bivariate standard normal distribution with $\rho = 0.8$ (left), normally distributed variables with Clayton copula with parameter $\theta = 1.5$ (middle) and the functional relationship $x_2 = x_1^2 + \varepsilon$ for $x_1 \sim \mathcal{N}(0, 0.5)$ and $\varepsilon \sim \mathcal{N}(0, 0.2)$ (right).

2 Dependence structures

Quantifying the relation between random variables often relies on the a-priori suggestion of a linear association (see, e.g., the linear positive linkage displayed in the left hand side panel of Figure 1). However, dependence between the variables can not only be characterized by a linear but by means of diverse functional forms. Besides the linear relationship two further examples of dependence structures between two random variables x_1 and x_2 are displayed in Figure 1. The second structure is characterized by dependence in the lower tail of the distributions. Such types of nonlinear dependence are commonly described by means of copulas, i.e. a function C which combines the two marginal distribution functions F_{x_1} and F_{x_2} to the joint distribution function $F_{x_1, x_2}(x_1, x_2) = C(F_{x_1}(x_1), F_{x_2}(x_2))$. One sided tail dependencies, as displayed in Figure 1, could be modeled by means of the Clayton copula. In general, the copula C can be uniquely determined following Sklar's theorem (Sklar, 1959). For a detailed description of dependence modelling by means of copulas see, e.g., Joe (1997). Furthermore, a functional nonlinear and nonmonotone association characterizes the relationship between the variables in the right hand side panel of Figure 1.

Although nonparametric tests of the null hypothesis of independence aim at performing adequately irrespective of the underlying distribution, they build upon certain (test specific) regularity assumptions. These might imply performance differences conditional on both the marginal distributions under the null hypothesis and the type of dependence under the alternative hypothesis. Starting from the examples of Figure 1, one might distinguish diverse nonmonotone and nonlinear dependence structures generated by copulas or based on functional associations. Additionally, for specific applications, e.g., economic data, modifications

of these structures might be of interest.¹ We consider several forms of dependence in subsets of a set $\{x_1, \dots, x_p\}$ of univariate random variables $x_1, \dots, x_p \in \mathbb{R}$. Besides pairwise (i.e., bivariate) dependencies the structures can become increasingly complicated in larger sets of random variables with $p > 2$. Next, we outline the null hypotheses of bivariate, groupwise and joint independence.

1. **Bivariate independence:** As illustrated in Figure 1, with $p = 2$, the considered test procedures assess dependence between two random variables x_1 and x_2 . The corresponding null hypothesis is $H_0 : F_{x_1, x_2}(x_1, x_2) = F_{x_1}(x_1)F_{x_2}(x_2)$ with joint distribution function F_{x_1, x_2} and marginals F_{x_1}, F_{x_2} .
2. **Groupwise independence:** Analyzing two sets of variables can be thought of as a generalization of bivariate dependence tests where two disjoint subsets of $\{x_1, \dots, x_p\}$ are subjected to testing, i.e., $\mathbf{x}_1 \in \mathbb{R}^{p_1}$ and $\mathbf{x}_2 \in \mathbb{R}^{p_2}$ such that $p_1 + p_2 = p$. The corresponding null hypothesis is $H_0 : F_{\mathbf{x}_1, \mathbf{x}_2}(\mathbf{x}_1, \mathbf{x}_2) = F_{\mathbf{x}_1}(\mathbf{x}_1)F_{\mathbf{x}_2}(\mathbf{x}_2)$ for multivariate distribution functions $F_{\mathbf{x}_1, \mathbf{x}_2}, F_{\mathbf{x}_1}$ and $F_{\mathbf{x}_2}$. Furthermore, some tests allow to diagnose the dependence between more than two disjoint subsets, where $p_1 + \dots + p_c = p$ and $c > 2$.
3. **Mutual independence:** To test for overall independence within a set of random variables $\{x_1, \dots, x_p\}$ the null hypothesis is formulated as $H_0 : F_{x_1, \dots, x_p}(x_1, \dots, x_p) = F_{x_1}(x_1) \cdots F_{x_p}(x_p)$. The tests exploit the fact that mutual independence is equivalent to independence within all subsets of $\{x_1, \dots, x_p\}$. This hypothesis is equivalent to stating groupwise independence and choosing subsets of size $p_1 = p_2 = \dots = p_c = 1$.

In spite of assessing the same null hypothesis, the considered nonparametric independence tests differ in their theoretical derivation. To identify sources of performance differences, we review the theoretical background of the test procedures in the next section and consider their performance under specific marginal distributions and dependence structures by means of a simulation study in Section 4.

¹Tests for serial dependence in time series are not explicitly considered here. An overview of corresponding approaches is given in Diks (2009).

3 Tests for independence

Independence diagnostics might be classified into four distinct categories according to their theoretical background. Recently, copula, spatial sign and rank and kernel based methods have been developed to test nonparametrically for independence in a multivariate framework. For benchmarking purposes we compare these approaches with classical test procedures, namely Hoeffding's D and diagnostics going back to Wilks (1935) in bivariate and multivariate designs, respectively. A direct comparison between the described tests is hardly feasible, as the approaches define multivariate dependence differently, namely groupwise (spatial sign and rank based, kernel based methods) or mutual (copula based method). Accounting for these disparities, we describe in this section how the tests and their modifications each assess all types of possible dependencies, i.e. bivariate, groupwise and mutual. Table 1 at the end of this section is supposed to provide further guidance for assigning distinct independence diagnostics to their corresponding implementations in R packages and functions. Throughout the section, we describe the framework of the tests, the test statistics and their empirical formulation.

3.1 Classical tests for independence

The category of classical independence tests consists of widely applied approaches that are frequently implemented in statistical software. Several nonparametric tests for bivariate dependence and one parametric test for multivariate dependence are shortly described in the following.

Pearson's correlation coefficient (e.g., Pearson, 1920) was one of the first measures of linear correlation between two random variables. Shortly after, rank correlation methods such as Kendall's tau (Kendall, 1938) and Spearman's rho (Spearman, 1904) were developed to test nonparametrically for independence in bivariate settings. The nonparametric procedure introduced in Hoeffding (1948) was further extended by Blum et al. (1961) who tabulate the distribution of Hoeffding's D under the null hypothesis of independence. For two random variables x_1 and x_2 , Hoeffding's D builds on the theoretical statistic

$$\Delta_{x_1, x_2} = \int [F_{x_1, x_2} - F_{x_1} F_{x_2}]^2 dF_{x_1, x_2}, \quad (1)$$

which measures the distance between the joint distribution and the product of marginal distri-

butions in a Cramér-von Mises (CvM) sense. For two random samples of size n , $x_{1,1}, \dots, x_{1,n}$ and $x_{2,1}, \dots, x_{2,n}$ the empirical counterpart of Δ_{x_1, x_2} reads as

$$\mathcal{T}_d = \frac{\alpha - 2(n-2)\beta + (n-2)(n-3)\gamma}{n(n-1)(n-2)(n-3)(n-4)}, \quad (2)$$

where

$$\alpha = \sum_{i=1}^n (R_i^{(1)} - 1)(R_i^{(1)} - 2)(R_i^{(2)} - 1)(R_i^{(2)} - 2), \quad (3)$$

$$\beta = \sum_{i=1}^n (R_i^{(1)} - 2)(R_i^{(2)} - 2)Q_i \quad \text{and} \quad \gamma = \sum_{i=1}^n Q_i(Q_i - 1). \quad (4)$$

Here, $R_i^{(1)}$ and $R_i^{(2)}$ are the ranks of observations $x_{1,i}$ and $x_{2,i}$, respectively. Furthermore, Q_i denotes the number of observation pairs $(x_{1,j}, x_{2,j})$ for which the ranks of $x_{1,j}$ and $x_{2,j}$ are both smaller than the ranks of $x_{1,i}$ and $x_{2,i}$, respectively, i.e. $Q_i = \sum_{j=1}^n \mathbb{I}\{R_j^{(1)} < R_i^{(1)}\} \mathbb{I}\{R_j^{(2)} < R_i^{(2)}\}$.

The statistic in (2) evaluates dependence between two univariate random variables. Wilks' test (Wilks, 1935) can serve as a benchmark diagnostic in a multivariate set of random variables under the assumption of Gaussianity. For p variables x_1, \dots, x_p , mutual dependence is assessed by means of Wilks' Lambda, i.e.,

$$\mathcal{T}_{Lm} = -n \cdot \log \left(\frac{\det(\text{cov}(x_1, \dots, x_p))}{\text{var}(x_1) \cdot \dots \cdot \text{var}(x_p)} \right). \quad (5)$$

The covariance and the variances in (5) are estimated on the basis of a random sample of x_1, \dots, x_p . Similarly, the statistic

$$\mathcal{T}_{Lg} = -n \cdot \log \left(\frac{\det(\text{cov}(x_1, \dots, x_p))}{\det(\text{cov}(\mathbf{x}_1)) \cdot \det(\text{cov}(\mathbf{x}_2))} \right) \quad (6)$$

is suitable to test for independence between two groups of variables $\mathbf{x}_1 \in \mathbb{R}^{p_1}$ and $\mathbf{x}_2 \in \mathbb{R}^{p_2}$. The empirical versions of the test statistics in (5) and (6) are asymptotically χ^2 -distributed with p and 2 degrees of freedom, respectively.

For the simulation study in Section 4 we use the test statistic (2) in bivariate settings. This statistic is implemented in the function `hoeffd` of the R package *Hmisc* (Harrell, 2015). To test for mutual and groupwise independence in multivariate settings we use the statistics in (5) and (6), respectively.

3.2 Tests based on spatial signs and spatial ranks

In the following, we consider two nonparametric analogs to Wilks' test in (6) based on standardized spatial signs and ranks. These dependence measures were introduced in Taskinen

et al. (2005) and extend the method of Puri & Sen (1971). More precisely, Kendall's tau and Spearman's rho are formulated in the multivariate setting by means of spatial signs and ranks. In this sense, the test statistics serve to test for groupwise independence and reduce to Kendall's tau and Spearman's rho in the bivariate case. To test for mutual independence we use a Fisher combination of p -values of these statistics.

Two sets of random variables \mathbf{x}_1 and \mathbf{x}_2 are assumed to follow an elliptically symmetric marginal distribution. Accordingly, the multivariate marginal densities of \mathbf{x}_k , $k = 1, 2$, can be given as

$$f_{\mathbf{x}_k}(\mathbf{x}_k) = \det(\Sigma_k)^{-1/2} \exp\left(-\Psi(\|\Sigma_k^{-1/2}(\mathbf{x}_k - \boldsymbol{\mu}_k)\|)\right) \quad (7)$$

for some function $\Psi(\cdot)$, shape matrix Σ_k and location vector $\boldsymbol{\mu}_k$. For common choices of Ψ the density $f_{\mathbf{x}_k}$ corresponds to the multivariate normal distribution, t -distribution or power exponential function. The shape matrix Σ_k is a positive definite, symmetric and affine invariant matrix, and $\|\cdot\|$ is any permutation and sign change invariant metric. Furthermore, let $\mathbf{z}_{1,i}$ denote a vector of standardized data points of observation i , i.e. $\mathbf{z}_{1,i} = \widehat{V}_1^{-1/2}(\mathbf{x}_{1,i} - \widehat{\boldsymbol{\mu}}_1)$ with $\widehat{\boldsymbol{\mu}}_1$ being an affine-equivariant location estimator and \widehat{V}_1 denoting an estimator of the shape matrix.²

Then, for the standardized data points $\mathbf{z}_{1,i}$ and $\mathbf{z}_{1,j}$, $i, j = 1, \dots, n$, the vector of *standardized spatial signs* reads as

$$\widehat{\mathbf{S}}_{ij}^{(1)} = \begin{cases} \frac{\mathbf{z}_{1,i} - \mathbf{z}_{1,j}}{((\mathbf{z}_{1,i} - \mathbf{z}_{1,j})'(\mathbf{z}_{1,i} - \mathbf{z}_{1,j}))^{1/2}} & \text{if } \mathbf{z}_{1,i} - \mathbf{z}_{1,j} \neq \mathbf{0} \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

Analogously, the standardized spatial sign vector for the second set $\mathbf{x}_{2,i} = (x_{21,i}, \dots, x_{2p_2,i})'$ is defined by $\widehat{\mathbf{S}}_{ij}^{(2)}$. The vector of the *standardized spatial ranks* of observation i then results as the average of these signs: $\widehat{\mathbf{R}}_i^{(k)} = \frac{1}{n} \sum_{j=1}^n \widehat{\mathbf{S}}_{ij}^{(k)}$, where $k = 1, 2$.

Using these definitions the multivariate extensions of Kendall's tau and Spearman's rho are

$$\tau^2 = \left\| \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \widehat{\mathbf{S}}_{ij}^{(1)} \widehat{\mathbf{S}}_{ij}^{(2)'} \right\|^2 \quad \text{and} \quad \rho^2 = \left\| \frac{1}{n} \sum_{i=1}^n \widehat{\mathbf{R}}_i^{(1)} \widehat{\mathbf{R}}_i^{(2)'} \right\|^2, \quad (9)$$

²More detailed descriptions of covered distributions and the theoretical background are given in Oja (2010). For the multivariate extension of Spearman's rho we estimate the shape matrix by means of a rank based covariance matrix (see also the documentation of the corresponding R package *SpatialNP* (Sirchia et al., 2018)).

respectively, and the corresponding test statistics are

$$\mathcal{T}_s = \frac{np_1p_2}{4c_1^2c_2^2}\tau^2 \quad \text{and} \quad \mathcal{T}_{sr} = \frac{np_1p_2}{c_1^2c_2^2}\rho^2. \quad (10)$$

The constants c_1 and c_2 in (10) depend on the marginals $F_{\mathbf{x}_1}$ and $F_{\mathbf{x}_2}$. Respective estimates are

$$\hat{c}_1^2 = \frac{1}{n} \sum_{i=1}^n \left(\widehat{\mathbf{R}}_i^{(1)'} \widehat{\mathbf{R}}_i^{(1)} \right) \quad \text{and} \quad \hat{c}_2^2 = \frac{1}{n} \sum_{i=1}^n \left(\widehat{\mathbf{R}}_i^{(2)'} \widehat{\mathbf{R}}_i^{(2)} \right). \quad (11)$$

The test statistics in (10) are χ^2 -distributed with p_1p_2 degrees of freedom under the null hypothesis of no dependence between \mathbf{x}_1 and \mathbf{x}_2 . Furthermore, the tests are efficient for alternatives that are contiguous to an elliptical null distribution. Under these alternatives the test statistics in (10) follow a noncentral χ^2 -distribution where the noncentrality parameter depends on the constants c_1^2 and c_2^2 and the assumptions on the distribution under the alternative hypothesis (for more details on its exact definition, see Taskinen et al., 2005).

For the simulation study in Section 4 we apply the test statistic based on spatial ranks which is implemented in the function `sr.indep.test` of the package *SpatialNP* (Sirkiä et al., 2018) with option `rank`. Furthermore, we use Fisher combined p -values to test for mutual dependence.

3.3 Tests based on the empirical copula

By means of copulas the null hypothesis of mutual independence within a set of random variables $\{x_1, \dots, x_p\}$ is $H_0 : C(F_{x_1}(x_1), \dots, F_{x_p}(x_p)) = F_{x_1}(x_1) \cdot \dots \cdot F_{x_p}(x_p)$, where the function C refers to the corresponding unique copula (Sklar, 1959).

The test procedure considered in the following was introduced in Genest & Rémillard (2004), and further analyzed and extended in subsequent works by Genest et al. (2006), Genest et al. (2007) and Kojadinovic & Holmes (2009). The test statistic is formulated as a Cramér-von Mises (CvM) distance and moreover, applies the decomposition techniques for empirical copulas introduced in Deheuvels (1981). In a first step, a set $\{x_1, \dots, x_p\}$ of univariate random variables $x_1 \in \mathbb{R}, \dots, x_p \in \mathbb{R}$ is partitioned into all possible decompositions. The global coefficient for mutual dependence in $\{x_1, \dots, x_p\}$ then consists of the dependence measures within all decompositions. Let $A \subset S_p = \{1, \dots, p\}$ denote a possible subset of indices. For instance, in the bivariate case only one single subset $A = \{1, 2\}$ has to be considered.

For subsets of indices $A, B \subset S_p$, the joint copula of x_1, \dots, x_p is expressed by means of a Möbius decomposition \mathcal{M} which decomposes the copula C as

$$\mathcal{M}_A(C) \equiv \sum_{B \subset A} (-1)^{|A \setminus B|} C(u^B) \prod_{k \in A \setminus B} u_k \quad (12)$$

for $u_1, \dots, u_p \in [0, 1]$ and $u^B \in [0, 1]^p$ such that

$$u_k^B = \begin{cases} u_k & \text{if } k \in B \\ 1 & \text{if } k \notin B. \end{cases}$$

Mutual independence, i.e. the independence copula, is characterized by the copula C for which $\mathcal{M}_A(C) \equiv 0$ for all $A \subset S_p$. To test for independence based on a sample of observations, the empirical version of the decomposition in (12) reads as

$$\mathcal{M}_A(C_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \prod_{k \in A} \left[\mathbb{I}\{R_i^{(k)} \leq (n+1)u_k\} - U_n(u_k) \right], \quad (13)$$

where C_n corresponds to the empirical copula, $R_i^{(k)}$ is the rank of $x_{k,i}$ and U_n is the distribution function of a random variable uniformly distributed on $\{1/(n+1), 2/(n+1), \dots, n/(n+1)\}$. The resulting $2^p - p - 1$ CvM statistics (one for each possible decomposition of $A \subset \{1, \dots, p\}$) consist of the decomposition in (13),

$$T_A = \int_{[0,1]^p} \{\mathcal{M}_A(C_n)\}^2 du, \quad (14)$$

which is calculated as

$$T_A = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \prod_{k \in A} \left[\frac{2n+1}{6n} + \frac{R_i^{(k)}(R_i^{(k)} - 1)}{2n(n+1)} + \frac{R_j^{(k)}(R_j^{(k)} - 1)}{2n(n+1)} - \frac{\max(R_i^{(k)}, R_j^{(k)})}{n+1} \right]. \quad (15)$$

It is worth to notice that the CvM statistic in (14) forms a multivariate measure of dependence similar to Hoeffding's D in the bivariate case. The expression in (13) defines the distance between the empirical copula, instead of the bivariate empirical joint distribution function, and the distribution under independence. Mutual dependence in a subset $A \subset S_p$ is then measured by combining these distances in the CvM statistic in (14). Genest & Rémillard (2004) discuss several methods to obtain a global test statistic for mutual independence in $\{x_1, \dots, x_p\}$. On the one hand, various combination methods of the p -values of T_A are considered. For instance, the Fisher combination of p -values is defined as

$$\mathcal{T}_W = -2 \sum_{|A| > 1} \log(p_{T_A}), \quad (16)$$

where p_{T_A} is the p -value of T_A . Under the null hypothesis, \mathcal{T}_W converges to a χ^2 -distribution with $2(2^p - p - 1)$ degrees of freedom. As an alternative, a measure of mutual dependence can be defined by means of a global CvM functional

$$\mathcal{T}_B = \int_{(0,1)^d} \left\{ \sqrt{n} \left(C_n(u) - \prod_{k=1}^p U_n(u_k) \right) \right\} du, \quad (17)$$

with cumulative distribution function U_n of a uniformly distributed variable on $\{1/n, \dots, n/n\}$ and the empirical copula C_n . It is worth mentioning that the combination of p -values has been shown to yield more powerful test procedures than measuring overall dependence based on the test statistic in (17) (Genest & Rémillard, 2004).

The described procedures apply to test for bivariate or mutual independence within a set of univariate random variables. Additionally, the tests based on the empirical copula can be extended to the multivariate case by means of a bootstrap procedure. Kojadinovic & Holmes (2009) derive a bootstrap version to test for mutual independence between vectors of random variables $\{\mathbf{x}_1, \dots, \mathbf{x}_p\}$. Under the null hypothesis of independence, Kojadinovic & Holmes (2009) build the bootstrap samples by sampling independently from the empirical marginal distribution function of each vector $\mathbf{x}_i, i = 1, \dots, p$. Furthermore, Beran et al. (2007) use a similar approach by applying the theory of so-called half-spaces and a CvM statistic to diagnose dependence between random vectors \mathbf{x}_1 and \mathbf{x}_2 .

For the simulation study in Section 4 we apply \mathcal{T}_B in (17) to test for independence in bivariate settings. To test for mutual dependence we compare the statistics \mathcal{T}_W in (16) and \mathcal{T}_B as implemented in `indepTest` from the package *copula* (Hofert et al., 2015). For groupwise dependence in multivariate settings we apply the bootstrap version implemented in `multIndepTest` of the *copula* package.

3.4 Tests based on distance covariances and kernel based distances

The two last subsections covered multivariate extensions of Kendall’s tau, Spearman’s rho and Hoeffding’s D with $\mathcal{T}_s, \mathcal{T}_{sr}$ in (10) and \mathcal{T}_B in (17), respectively. In the following, we point out that the dependence coefficient proposed in Székely et al. (2007) pursues an alternative way by processing interpoint distances. For sets of random variables $\mathbf{x}_1 \in \mathbb{R}^{p_1}$ and $\mathbf{x}_2 \in \mathbb{R}^{p_2}$ with finite moments, let $\varphi_{x_1}, \varphi_{x_2}$ and φ_{x_1, x_2} denote the marginal and joint characteristic functions, respectively. Székely et al. (2007) introduce the test for independence between \mathbf{x}_1 and \mathbf{x}_2 in

two versions: On the one hand, based on the *distance covariance*

$$\mathcal{V}^2(\mathbf{x}_1, \mathbf{x}_2) = \|\varphi_{x_1, x_2}(t, s) - \varphi_{x_1}(t)\varphi_{x_2}(s)\|^2 \geq 0, \quad (18)$$

and alternatively, based on the *distance correlation*

$$\mathcal{R}^2(\mathbf{x}_1, \mathbf{x}_2) = \begin{cases} \frac{\mathcal{V}^2(\mathbf{x}_1, \mathbf{x}_2)}{\sqrt{\mathcal{V}^2(\mathbf{x}_1, \mathbf{x}_1)\mathcal{V}^2(\mathbf{x}_2, \mathbf{x}_2)}} & \text{if } \mathcal{V}^2(\mathbf{x}_1, \mathbf{x}_1)\mathcal{V}^2(\mathbf{x}_2, \mathbf{x}_2) > 0, \\ 0 & \text{otherwise.} \end{cases} \quad (19)$$

In (18), $\|\cdot\|$ corresponds to the norm in a weighted L_2 -space of functions on $\mathbb{R}^{p_1+p_2}$. More precisely,

$$\mathcal{V}^2(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{c_p c_q} \int_{\mathbb{R}^{p+q}} \frac{|\varphi_{x_1, x_2}(t, s) - \varphi_{x_1}(t)\varphi_{x_2}(s)|^2}{|t|_p^{1+p}|s|_q^{1+q}} dt ds, \quad (20)$$

for constant $c_d = \frac{\pi^{(1+d)/2}}{\Gamma((1+d)/2)}$, with $\Gamma(\cdot)$ denoting the complete gamma function (see Székely et al., 2007, for more details). Hence, the dependence measures \mathcal{V}^2 and \mathcal{R}^2 are zero if and only if the two considered sets \mathbf{x}_1 and \mathbf{x}_2 are independent. For two random samples, consisting of the vectors $\mathbf{x}_{1,i} = (x_{11,i}, \dots, x_{1p_1,i})'$ and $\mathbf{x}_{2,i} = (x_{21,i}, \dots, x_{2p_2,i})'$, $i = 1, \dots, n$, the corresponding test statistics are calculated from the sample covariances

$$\begin{aligned} \mathcal{T}_{dCov} = \mathcal{V}_n^2(\mathbf{x}_1, \mathbf{x}_2) = \\ \frac{1}{n^2} \sum_{i,j=1}^n \left(|\mathbf{x}_{1,i} - \mathbf{x}_{1,j}|_{p_1} - \frac{1}{n} \sum_{j=1}^n |\mathbf{x}_{1,i} - \mathbf{x}_{1,j}|_{p_1} - \frac{1}{n} \sum_{j=1}^n |\mathbf{x}_{1,i} - \mathbf{x}_{1,j}|_{p_1} + \frac{1}{n^2} \sum_{i,j=1}^n |\mathbf{x}_{1,i} - \mathbf{x}_{1,j}|_{p_1} \right) \\ \times \left(|\mathbf{x}_{2,i} - \mathbf{x}_{2,j}|_{p_2} - \frac{1}{n} \sum_{j=1}^n |\mathbf{x}_{2,i} - \mathbf{x}_{2,j}|_{p_2} - \frac{1}{n} \sum_{j=1}^n |\mathbf{x}_{2,i} - \mathbf{x}_{2,j}|_{p_2} + \frac{1}{n^2} \sum_{i,j=1}^n |\mathbf{x}_{2,i} - \mathbf{x}_{2,j}|_{p_2} \right), \end{aligned} \quad (21)$$

where $|\cdot|_{p_1}$ and $|\cdot|_{p_2}$ denote interpoint Euclidean distances in \mathbb{R}^{p_1} and \mathbb{R}^{p_2} , respectively. The empirical version of \mathcal{R}_n^2 obtains from inserting $\mathcal{V}_n^2(\mathbf{x}_1, \mathbf{x}_2)$ into (19). Restricting \mathbf{x}_1 and \mathbf{x}_2 to have finite moments, the test is consistent for any type of dependence. Under the null hypothesis, $n\mathcal{V}_n^2/S \xrightarrow{d} Q$ for $n \rightarrow \infty$, where $S = (\frac{1}{n^2} \sum_{i,j=1}^n |\mathbf{x}_{1,i} - \mathbf{x}_{1,j}|_{p_1})(\frac{1}{n^2} \sum_{i,j=1}^n |\mathbf{x}_{2,i} - \mathbf{x}_{2,j}|_{p_2})$ and Q is a nonnegative quadratic form of centered Gaussian random variables. More specifically, $Q \stackrel{d}{=} \sum_{j=1}^{\infty} \lambda_j Z_j^2$ for independent standard normal distributed random variables Z_j and nonnegative constants λ_j which depend on the distributions of \mathbf{x}_1 and \mathbf{x}_2 .

Székely et al. (2007) and Székely & Rizzo (2009) modify and extend the tests based on (18) and (19) in several ways. For instance, the norm used in (18) is generalized to $\|\cdot\|_{\alpha}$

which implies a more general weight function and results in α -distance dependence measures. A further specification of the distance covariance is obtained by choosing the covariance with respect to a certain stochastic process. The Brownian motion, for instance, obtains the *Brownian distance covariance*.

Rémillard (2009) argues that the performance of the tests based on (18) and (19) depends on the marginal distributions and further, the statistic in (21) is only applicable to test for independence between two sets of random variables. To address these concerns, Matteson & Tsay (2013) suggest probability integral transformations to avoid the dependence on the marginal distributions. In addition, they provide a test for mutual independence using the fact that the null hypothesis of mutual independence within a set of random variables $\{x_1, \dots, x_p\}$ is equivalent to $H_0 : \varphi_{x_k, x_{k+}} = \varphi_{x_k} \varphi_{x_{k+}}$ for all $k = 1, \dots, p-1$ and $k+ = k+1, \dots, p$. Then, the corresponding test statistics is a combination of the bivariate statistics

$$\mathcal{T}_{dCov} = n \cdot \sum_{k=1}^{p-1} \mathcal{V}_n^2(x_k, x_{k+}). \quad (22)$$

Sejdinovic et al. (2013) embed the distance covariance within a more general group of dependence measures which has originated from machine learning. The kernel based so-called Hilbert-Schmidt independence criterion (HSIC) measures the distance between embeddings of distributions into reproducing kernel Hilbert spaces (RKHS). Choosing specific distance induced kernels, the distance covariance is then equivalent to the HSIC based on the RKHS. By linking these two classes of statistics \mathcal{T}_{dCov} might be considered as a representative for HSICs. In the following, we only consider \mathcal{T}_{dCov} and refer to Sejdinovic et al. (2013) for performance comparisons of further HSICs with alternative kernel choices.

For the simulation study in Section 4 we apply \mathcal{T}_{dCov} in (21) to test for independence in bivariate and groupwise settings. We use the function `indep.test` of the R package *energy* (Rizzo & Szekely, 2014). For mutual independence we apply the function `permTest` of the *steadyICA* package (Risk et al., 2015).

4 Performance under specific dependence structures

Although all considered tests have been proposed to evaluate the null hypothesis of independence nonparametrically, their underlying distributional assumptions are more or less restrictive. Especially in small samples this might lead to size and power differentials under

Table 1: R packages and functions corresponding to the procedures described in Section 3 and applied within the simulation study. Corresponding independence diagnostics are given in parentheses.

	classical	spatial rank	empirical copula	distance covariance
R package	<i>Hmisc</i> (Harrell, 2015)	<i>SpatialNP</i> (Sirchia et al., 2018)	<i>copula</i> (Hofert et al., 2015)	<i>energy</i> (Rizzo & Szekely, 2014) <i>steadyICA</i> (Risk et al., 2015)
R function for distinct dependence levels				
bivariate	<code>hoeffd</code> (\mathcal{T}_d)	<code>sr.indep.test</code> (\mathcal{T}_{sr})	<code>indepTest</code> (\mathcal{T}_B)	<code>indep.test</code> (\mathcal{T}_{dCov})
mutual	Wilks' Lambda (\mathcal{T}_{Lm})	Fisher comb. of \mathcal{T}_{sr}	<code>indepTest</code> ($\mathcal{T}_W, \mathcal{T}_B$)	<code>permTest</code> (\mathcal{T}_{dCov})
groupwise	Wilks' Lambda (\mathcal{T}_{Lg})	<code>sr.indep.test</code> (\mathcal{T}_{sr})	<code>multIndepTest</code> (\mathcal{T}_B)	<code>indep.test</code> (\mathcal{T}_{dCov})

the null hypothesis and certain dependence alternatives, respectively.³ The following simulation study is supposed to identify such performance differentials. We describe the simulation design first and discuss the results afterwards.

4.1 Simulation setting

As outlined in Section 2, the considered tests diagnose dependence between two continuous random variables, two or more vectors of variables, or mutual dependence in a set of more than two variables. Within these settings we compare the size and power of the tests either with respect to the implied correlation ρ or the sample size n . The underlying distributional settings are summarized in Table 2. We consider representative distributions and dependence structures which are supposed to unravel differences and similarities of the tests. In more diverse settings, for instance, alternative choices of sample sizes, copulas and marginals, the results are in line with those discussed in the following.

4.1.1 Bivariate sets of random variables

The two random samples $(x_{1,1}, \dots, x_{1,n})$ and $(x_{2,1}, \dots, x_{2,n})$ are generated under the null hypothesis (independence) and under the alternative hypothesis (dependence). Two elliptical

³For instance, Shih & Emura (2016) study the properties of Spearman's rho and Kendall's tau under specific copula structures.

copulas and one representative of Archimedean copulas determine the dependence structure alternatively. In addition, we study a direct association by means of a nonlinear and non-monotonic function with noise. Finally, we investigate robustness of the tests to modifications of these dependence structures.

For correlation levels $\rho = 0, 0.1, \dots, 0.8$, we generate bivariate sets of random variables, and, hence, focus on the correlation moving from the null hypothesis of independence to larger degrees of dependence. More precisely, for elliptical copulas a correlation matrix $V_{x_1, x_2} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ determines the dependence structure. Archimedean copulas, e.g. the Clayton copula, can be formulated with respect to a coefficient determining the strength of correlation by means of a generator function $\psi(\rho)$, $\rho \in [0, \infty)$. For the explicit definition of the copula as a function of correlation ρ we refer to the documentation of the respective R functions (Hofert et al., 2015). The generator function of the Clayton copula is, for instance, $\psi(\rho) = (1 + \rho)^{-1/\theta}$ with $\theta \in (0, \infty)$. We calculate four test statistics, namely Hoeffding's \mathcal{T}_d , the Cramér-von Mises statistic \mathcal{T}_B , the multivariate extension of Spearman's rho \mathcal{T}_{sr} , and the distance covariance \mathcal{T}_{dCov} (see Table 1). The estimated power of the tests is the share of $R = 1000$ test statistics $\mathcal{T}_d, \mathcal{T}_B, \mathcal{T}_{sr}$ and \mathcal{T}_{dCov} with p -value below the nominal significance level of $\alpha = 0.05$.⁴ We provide the size adjusted power with respect to the empirical level $\hat{\alpha}$, and compare the size and power of the tests for sample sizes $n = 10, 50, 100$.

Dependence modeling by means of copulas: Three distinct marginal distributions and a dependence structure determined by three copulas specify the bivariate distribution structure. Regarding the univariate marginal distributions we choose the standard normal, the exponential and the Cauchy distribution. Monotonic and linear dependence is covered by means of the bivariate Gaussian distribution. Moreover, the Student's t - and Clayton copula allow for tail dependencies and thus, represent nonlinear dependence structures. In order to generate respective random samples of size n , we apply the R functions `mvdc` and `rMvdc` from the R package *copula* (Hofert et al., 2015).

Functional dependence structure: From a distinct perspective, dependence can be seen as an information structure characterizing the data. Increasing the level of noise in a bivariate set of random variables changes the structure from a deterministic relationship to independence.

⁴In this study, the considered nominal significance level is $\alpha = 0.05$. Similar results obtain with respect to other conventional levels, for instance $\alpha = 0.1$.

We relate two random variables x_1 and x_2 directly by means of a function, i.e. $x_2 = f(x_1)$, to allow for nonlinear and nonmonotonic types of dependence. As an example, we consider a quadratic structure $x_2 = x_1^2 + \varepsilon$, where $x_1 \sim \mathcal{N}(0, 0.5)$,⁵ and $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ is a Gaussian noise term with increasing standard deviation $\sigma = 0, 0.1, \dots, 1.5$. The variables x_1 and ε are independently drawn in every Monte Carlo iteration indexed by $r = 1, \dots, R$. Perfect dependence corresponds to $\sigma = 0$, whereas a low level of association is present if $\sigma = 1.5$. A generated sample with $\sigma = 0.2$ has been shown as an example in the right hand side panel of Figure 1.

Modifications of the dependence structures: In practice, the assumption of a homogeneous distribution within the entire sample might be not appropriate for an actual data set. For instance, in economic data a varying dependence structure might be present. Furthermore, not only dependence between the marginals might exist but the marginals themselves might incorporate dependence in their variances (see, for instance, Manner & Reznikova, 2012). Allowing for modifications of the distributional settings we consider varying degrees of dependence first. For this purpose, we generate a bivariate normally distributed sample with two distinct levels of correlation, i.e. $\rho_1 = 0.2$ in the first half and $\rho_2 = 0.4$ in the second half of the sample. As a second modification, we formalize dependence among the marginals as implied by a bivariate GARCH process. More explicitly, we sample data from a so-called Constant Conditional Correlation GARCH(1,1) process (CCC-GARCH(1,1), see Bollerslev, 1990). Accordingly, observations $x_{k,i}$, $i = 1, \dots, n$, $k = 1, 2$, are drawn as

$$x_{k,i} = h_{k,i}^{1/2} z_{k,i} \quad \text{with} \quad h_{k,i} = a_{k0} + a_{kk} x_{k,i-1}^2 + b_{kk} h_{k,i-1} \quad \text{and} \quad z_i \sim \mathcal{N}(0, P), \quad (23)$$

with Gaussian innovations $z_{i,k}$ and $a_{k0} = 1$, $a_{kk} = b_{kk} = 0.4$. Dependence between the univariate GARCH processes is modeled by means of an unconditional covariance matrix P with $p_{11} = p_{22} = 1$ and off diagonal elements $p_{12} = p_{21} = \rho = 0.4$. For a more detailed description of CCC-GARCH sampling we refer to the manual of the R package *cggarch* (Nakatani, 2010).

⁵Siqueira Santos et al. (2013) consider alternative choices for the distribution of x_1 as, for instance, equidistant points or the uniform distribution. Additionally, they study further nonmonotonic and nonlinear dependence structures, i.e., alternative choices of the function f .

Table 2: Simulation settings.

	dependence structure	power performance wrt parameter
bivariate	normal, t , Clayton copula with	$n = 10, 50, 100$
x_1, x_2	normal, exponential, Cauchy marginals	(size, size adj power for $\rho = 0.4$) $\rho = 0, 0.1, \dots, 0.8$ (power)
	$x_2 = x_1^2 + \varepsilon, x_1 \sim \mathcal{N}(0, 0.5), \varepsilon \sim \mathcal{N}(0, \sigma^2)$	$\sigma = 0, 0.1, \dots, 1.5$
	varying dependence, $\rho_1 = 0.2, \rho_2 = 0.4$	$n = 20, 50, 100$
	CCC-GARCH(1,1), $\rho = 0.4$	$n = 20, 50, 100$
mutual	normal copula with normal and	$\rho = 0, 0.1, \dots, 0.8$
x_1, x_2, x_3	Cauchy marginals, $n = 100$	
groupwise	normal copula and marginals,	$\rho_{inter} = 0, 0.1, \dots, 0.8$
$\{x_{11}, x_{12}\}, \{x_{22}\}$	$n = 100, \rho_{intra} = 0$ and 0.8	

4.1.2 Multivariate sets of random variables

As described in Section 2, a set of more than two random variables might exhibit groupwise or mutual dependence. To uncover differences and similarities between tests for mutual independence, we consider a most simple framework, i.e. a set of three univariate random variables. Within such sets $\{x_1, x_2, x_3\}$ we formalize the dependence structure under the alternative hypothesis by means of equal correlation ρ in bivariate tuples $\{x_1, x_2\}, \{x_1, x_3\}, \{x_2, x_3\}$. Accordingly, the correlation matrix of $\{x_1, x_2, x_3\}$ reads as

$$V_{x_1, x_2, x_3} = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}. \quad (24)$$

Similar to the bivariate case, we consider several marginal distributions and three dimensional copulas with increasing levels of correlation $\rho = 0, 0.1, \dots, 0.8$. We study the performance of two versions of the copula based procedures, namely the global CvM statistic \mathcal{T}_B and the Fisher combination of p -values in subsamples \mathcal{T}_W . Moreover, we consider the mutual version of the distance covariance \mathcal{T}_{dCov} , a Fisher combination of p -values of the bivariate \mathcal{T}_{sr} statistics and Wilks' lambda \mathcal{T}_{Lm} .

Furthermore, we compare tests for groupwise dependence between two sets of variables

$\mathbf{x}_1 = \{x_1, x_2\}$ and $\mathbf{x}_2 = \{x_3\}$. In (24), we have only considered one single correlation level ρ such that independence implies zero correlation globally (i.e., $\rho = 0$). Borrowing from the simulation study in Kojadinovic & Holmes (2009), the correlation matrix employed to formalize groupwise dependence of $\{\mathbf{x}_1, \mathbf{x}_2\}$ reads as

$$V_{\mathbf{x}_1, \mathbf{x}_2} = \begin{pmatrix} 1 & \rho_{intra} & \rho_{inter} \\ \rho_{intra} & 1 & \rho_{inter} \\ \rho_{inter} & \rho_{inter} & 1 \end{pmatrix}. \quad (25)$$

The null hypothesis of groupwise independence corresponds to absence of inter group correlation ($\rho_{inter} = 0$). Accordingly, we study power properties with respect to increasing inter group correlation, i.e. $\rho_{inter} = 0.1, \dots, 0.8$. Apart from inter group dependence, ρ_{intra} in (25) denotes the strength of intra group correlation. Intra group correlation ρ_{intra} might differ from zero even under the null hypothesis of groupwise independence. To account for distinct degrees of intra group dependence in the simulation study, we select two distinct levels of correlation within \mathbf{x}_1 , namely, $\rho_{intra} = 0$ (no correlation) and $\rho_{intra} = 0.8$ (strong correlation).

To be more precise on the performance of the considered multivariate tests in a general multivariate setting, we also study a four dimensional case and test for groupwise dependence between two sets $\mathbf{x}_1 = \{x_1, x_2\}$ and $\mathbf{x}_2 = \{x_3, x_4\}$. Dependence is generated in analogy to the trivariate case using a correlation matrix

$$V_{\mathbf{x}_1, \mathbf{x}_2} = \begin{pmatrix} 1 & \rho_{intra} & \rho_{inter} & \rho_{inter} \\ \rho_{intra} & 1 & \rho_{inter} & \rho_{inter} \\ \rho_{inter} & \rho_{inter} & 1 & \rho_{intra} \\ \rho_{inter} & \rho_{inter} & \rho_{intra} & 1 \end{pmatrix}. \quad (26)$$

We compare four test statistics under the null hypothesis and the alternative hypothesis of groupwise dependence: the Cramér-von Mises statistic \mathcal{T}_B , the statistic \mathcal{T}_{sr} based on spatial ranks, the distance covariance \mathcal{T}_{dCov} and the parametric test based on Wilks' lambda \mathcal{T}_{Lg} .

4.2 Simulation results

In the following discussion, the results for size and size adjusted power provide a baseline comparison of the tests. In the subsequent investigation we consider power properties with

respect to increasing correlation for copulas and decreasing noise for functional dependence. Furthermore, we address robustness of the tests under modifications of the stylized dependence structures. In multivariate sets of random variables we study the performance of the considered tests under mutual and groupwise dependence alternatives.

4.2.1 Bivariate sets of random variables

Empirical size and size adjusted power

Table 3 documents the estimated size and the size adjusted power of the test statistics $\mathcal{T}_d, \mathcal{T}_B, \mathcal{T}_{sr}$ and \mathcal{T}_{dCov} with respect to three distinct copulas for sample sizes $n = 10, 50, 100$ and correlation levels $\rho = 0$ (size, in columns 4-7) and 0.4 (power, in columns 8-11). Under respective regularity conditions the test statistics are supposed to converge to the asymptotic distribution for increasing sample sizes (see Section 3). Consequently, the empirical size $\hat{\alpha}$ converges to the true level $\alpha = 0.05$ under these regularity conditions. Deviations from the true level might reflect, on the one hand, violations of the conditions. On the other hand, they contrast the small sample performance of the tests with asymptotic approximations and, in particular, are informative on the speed of convergence.

We generate bivariate samples under the null hypothesis by means of the respective copula with zero correlation. Although all samples comprise independently drawn marginals, the independence tests perform differently under distinct choices of copulas and marginals. Over all generated samples, the size distortions of the CvM statistic \mathcal{T}_B appear smaller compared with those of the other test procedures. Furthermore, the empirical size of \mathcal{T}_B changes only slightly with respect to the chosen marginal distributions. In contrast, size distortions of \mathcal{T}_{sr} are slightly larger and those of \mathcal{T}_d are much larger. The statistic \mathcal{T}_d shows oversizing in nearly all considered samples. The statistic \mathcal{T}_{dCov} holds adequate size properties for a sample generated by means of the Gaussian or the Clayton copula. However, we can observe oversizing of this test for the Student's t -copula in combination with all marginals. Under Cauchy marginals, \mathcal{T}_{dCov} shows an empirical level as large as $\hat{\alpha} = 0.461$.

The power estimates displayed in Table 3 are adjusted with respect to the empirical size of the tests. The size adjustment lowers (increases) the rejection frequencies of oversized (undersized) tests to enable a direct comparison of the power of alternative test procedures. Overall, \mathcal{T}_d shows a slight lead in terms of size adjusted power in small samples ($n = 10$)

and \mathcal{T}_{sr} in larger samples ($n = 100$). Under Gaussian copula dependence, the size adjusted power of all tests converges with a similar rate and almost approaches unity for $n = 100$. As the only exception, the size adjusted power of \mathcal{T}_{dCov} fails to converge within the considered sample sizes under the Gaussian copula with Cauchy marginals. For dependence generated by means of a Student's t -copula, the results are similar to the Gaussian dependence structure. Moreover, the tests (except \mathcal{T}_{dCov} for Cauchy marginals) are consistent under dependence modeled by means of a Clayton copula while showing slower convergence rates as under a Gaussian and Student's t -copula.

Theoretically, the inferior performance of the distance covariance \mathcal{T}_{dCov} for specific marginals is in line with its dependence on the marginal distribution (see Section 3). In particular, the moments of the Cauchy distribution are not finite and thus, the regularity conditions that underlie \mathcal{T}_{dCov} do not hold. Overall, the size distortions in small samples indicate which tests might not be appropriate given the underlying distributional setting. The baseline comparison of size and size adjusted power displays comparable test performances in the standard setting, i.e., the Gaussian copula with normal and exponential marginals. However, under a t -copula we can observe notable differences as described above. Consequently, under tail dependencies the choice of the test appears more crucial for the test decision. In terms of empirical size, the test statistic \mathcal{T}_B seems to perform best irrespective of the underlying distribution. Additionally, we observe that the considered tests show inferior power under dependence governed by the Clayton copula. Moreover, the distance covariance performs weakly for specific marginals.

In order to compare the tests by means of the size adjusted power one has to have in mind that the size adjustments are substantial in case of large size distortions. The size adjusted power of a test serves to compare the test in simulated settings but is, however, not applicable in practice since the empirical level $\hat{\alpha}$ is typically unknown. Therefore, in empirical research one might rather be interested in the comparison of unadjusted power of tests for which it can be safely presumed that their empirical size is close to the nominal level.

Power curves

Figure 2 displays unadjusted power curves with respect to varying levels of correlation $\rho = 0, 0.1, \dots, 0.8$ and fixed sample size $n = 100$. In particular, for $\rho = 0.4$ the size adjusted counterparts of these empirical power estimates are displayed in Table 3. Studying the power

Table 3: Empirical size $\hat{\alpha}$ and size adjusted power for correlation $\rho = 0.4$, $R = 1000$ and $\alpha = 0.05$ with respect to alternative bivariate copulas and marginals. Empirical sizes deviating from the true level by more than 0.014 ($\approx 1.96\sqrt{0.05 \cdot 0.95/1000}$) are marked in bold.

copula	marginal	n	size				size adjusted power			
			d	B	sr	$dCov$	d	B	sr	$dCov$
normal	normal	10	0.099	0.052	0.057	0.050	0.334	0.184	0.240	0.205
		50	0.066	0.050	0.048	0.048	0.795	0.748	0.809	0.790
		100	0.063	0.055	0.053	0.055	0.970	0.967	0.979	0.978
	exp.	10	0.100	0.046	0.045	0.048	0.361	0.157	0.195	0.217
		50	0.062	0.048	0.050	0.049	0.765	0.720	0.780	0.728
		100	0.058	0.050	0.045	0.043	0.968	0.964	0.971	0.954
Student's t	normal	10	0.135	0.058	0.071	0.083	0.390	0.197	0.263	0.296
		50	0.074	0.053	0.057	0.085	0.720	0.687	0.739	0.805
		100	0.067	0.065	0.075	0.105	0.969	0.968	0.978	0.986
	exp.	10	0.131	0.054	0.070	0.066	0.331	0.167	0.232	0.268
		50	0.057	0.047	0.054	0.100	0.745	0.693	0.754	0.837
		100	0.069	0.057	0.062	0.137	0.959	0.956	0.961	0.983
Clayton	normal	10	0.104	0.046	0.051	0.062	0.248	0.104	0.126	0.151
		50	0.061	0.045	0.042	0.047	0.429	0.346	0.369	0.406
		100	0.063	0.059	0.055	0.062	0.727	0.714	0.716	0.772
	exp.	10	0.120	0.058	0.061	0.044	0.245	0.091	0.123	0.083
		50	0.050	0.040	0.048	0.046	0.372	0.333	0.396	0.228
		100	0.044	0.048	0.055	0.057	0.663	0.677	0.722	0.481
Cauchy	10	0.120	0.051	0.050	0.053	0.126	0.080	0.086	0.111	
	50	0.071	0.056	0.058	0.051	0.460	0.407	0.450	0.251	
	100	0.065	0.048	0.055	0.057	0.714	0.660	0.712	0.355	

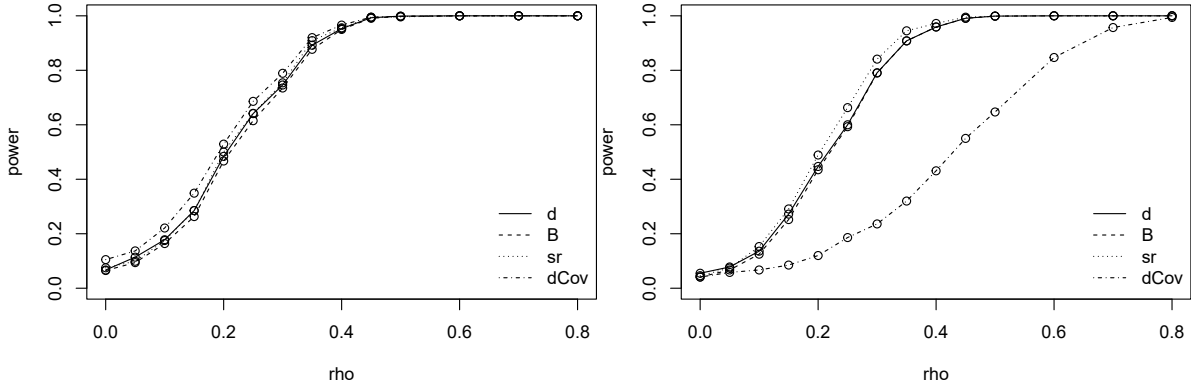


Figure 2: Power curves for bivariate dependence modeled by means of copulas: t -copula with Gaussian marginals (left) and Gaussian copula with Cauchy marginals (right).

curves, our interest is in their overall shape or, more specifically, in the degree of dependence that yields a power of unity such that the tests detect dependence with probability one.

First, we study the results for a representative copula structure in the left hand side panel of Figure 2, namely the t -copula combined with Gaussian marginals. The resulting empirical sizes for $\rho = 0$ are comparable with those displayed in Table 3. In this setting the power curves of the considered tests have a similar shape. Power of unity is attained for a similar level of correlation for the t -copula ($\rho \approx 0.5$). Comparing these results with those generated, e.g., by the Clayton-copula, the power of all four tests converges to unity faster in case of dependence generated from the t -copula than for dependence emerging from the Clayton-copula. The procedures are consistent against both alternatives. Overall, \mathcal{T}_{dCov} and \mathcal{T}_{sr} slightly outperform the independence diagnostics based on the CvM statistic in these scenarios.

Furthermore, the right hand side panel of Figure 2 displays power curves for a Gaussian copula with Cauchy marginals. In line with the results shown in Table 3, the power curve of \mathcal{T}_{dCov} stays throughout remarkably below the other curves. Especially, under Cauchy marginals with non existing moments \mathcal{T}_{dCov} suffers from power weakness. In addition, having also in mind the size distortions under the t -copula with Cauchy marginals, \mathcal{T}_{dCov} might not be appropriate under these specific marginal distributions. Nevertheless, for alternatives far away from the null hypothesis of independence ($\rho = 0.8$) all tests show power of unity.

Besides copula dependencies we relate the variables x_1 and x_2 in a functional manner $x_2 = x_1^2 + \varepsilon$ to represent a nonlinear and nonmonotone dependence alternative. Rejection

frequencies for samples of size $n = 100$ are depicted in the left hand side panel of Figure 3. Starting with the deterministic relationship and modeling dependence up to a certain level of noise, we are interested if this type of dependence is detected, rather than in the test behavior close to/under the null hypothesis. Except the spatial rank based procedure \mathcal{T}_{sr} the power of all tests is unity for the deterministic relationship $x_2 = x_1^2$ (i.e. $\sigma = 0$) up to moderate levels of uncertainty ($\sigma = 0.3$). Power estimates for both the CvM statistic \mathcal{T}_B and Hoeffding's \mathcal{T}_d increase with a decreasing level of uncertainty, but are throughout smaller in comparison with the power of \mathcal{T}_{dCov} . In contrast, the spatial rank based procedure \mathcal{T}_{sr} indicates the deterministic association $x_2 = x_1^2$ in only 30% of the cases. For the convergence of \mathcal{T}_{sr} , Taskinen et al. (2005) assume an elliptical distribution so that the procedure is not necessarily consistent against the nonmonotone alternative. Furthermore, Ding & Li (2014) argue that dependence structures formalized as functional relationships might correspond to singular copulas. A singular copula violates the assumption of absolutely continuous copulas imposed by Genest & Rémillard (2004). Thus, for such a dependence structure \mathcal{T}_{dCov} might be preferred over \mathcal{T}_d and \mathcal{T}_B while \mathcal{T}_{sr} suffers from prohibitive power loss.

In summary, the results for the power curves allow similar conclusions as those documented for empirical size and size adjusted power of the tests. Standard distributional settings lead to comparable performances of all tests. In particular, oversizing under the t -copula and the inferior performance of \mathcal{T}_{dCov} are notable in this respect. Furthermore, one would rank the tests differently based on their performance under a nonmonotone dependence structure.

Robustness to modifications

Results documented in Table 4 address the robustness of the tests to non standard data structures for samples of size $n = 20, 50, 100$. In heterogeneous random samples a varying dependence structure (compared with constant dependence in the entire sample) might be present. The results for a bivariate normal distribution with $\rho_1 = 0.2$ and $\rho_2 = 0.4$ indicate that all considered test procedures remain consistent. The power for $n = 100$ is, in fact, comparable with rejection frequencies in a sample with homogeneous correlation $\rho = 0.3$. For larger samples or stronger levels of correlation all procedures show satisfactory power properties. Nevertheless, the power estimates of \mathcal{T}_d and \mathcal{T}_B converge slower than their counterparts obtained from \mathcal{T}_{sr} and \mathcal{T}_{dCov} .

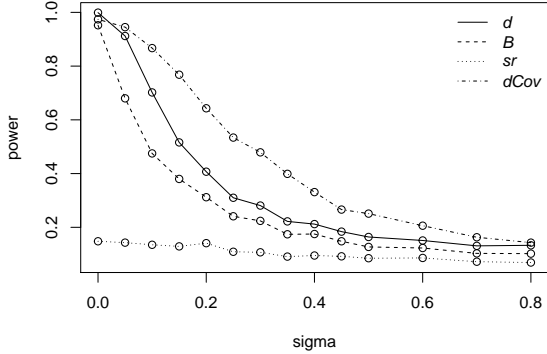


Figure 3: Power curve for quadratic dependence with respect to the standard deviation of the noise term ($\sigma = 0$ refers to perfect dependence and $\sigma = 1.5$ to weak association).

Type of modification	n	power			
		d	B	sr	dCov
Varying dependence, $\rho_1 = 0.2, \rho_2 = 0.4$	20	0.269	0.210	0.247	0.259
	50	0.520	0.497	0.531	0.568
	100	0.781	0.775	0.804	0.835
CCC-GARCH(1,1)	20	0.363	0.323	0.404	0.342
	50	0.714	0.704	0.743	0.709
	100	0.938	0.942	0.956	0.930

Table 4: Power with respect to different modifications of the dependence structure and marginals for a bivariate normal distribution.

Table 4 documents the results for a normally distributed CCC-GARCH(1,1) process with unconditional correlation of $\rho = 0.5$. As it turns out, all considered tests are consistent against this type of dependence with comparable speed of convergence and power of unity for about $n = 120$. In particular, \mathcal{T}_{sr} shows slight power leads in small samples. Overall, the tests are robust under this data structure.

Summarizing the results for the bivariate case, we cannot identify a single nonparametric test which is most powerful against all alternatives. Instead, the size and power performance differs for distinct types of data. We have discovered dependence structures where slight differences between the tests are identifiable, as well as structures where the test decision might depend more strongly on the choice of the test. Based on its empirical size properties, the CvM statistic \mathcal{T}_B might be preferred as it shows the most stable results. Irrespective of distinct dependence structures the empirical level of \mathcal{T}_B is close to the nominal level of $\alpha = 0.05$. The other tests show oversizing especially under a t -copula, and the distance covariance \mathcal{T}_{dCov} performs worst under the considered copula dependence structures. Nevertheless, \mathcal{T}_{dCov} outperforms the other tests under nonmonotone dependence structures where, in contrast, \mathcal{T}_{sr} shows severe lacks of power.

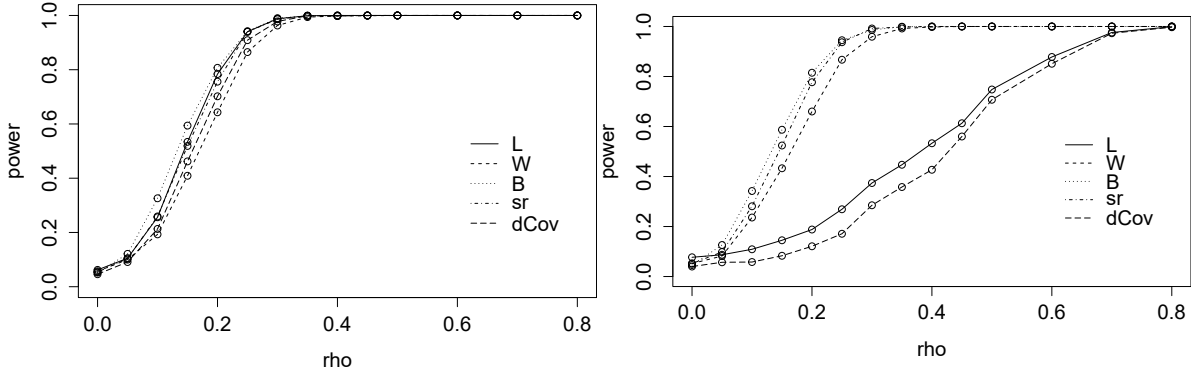


Figure 4: Power curves for mutual dependence as implied by a Gaussian copula with Gaussian (left hand side) and Cauchy (right hand side) marginals: Wilks' Lambda \mathcal{T}_{Lm} (labeled L in the figure), the Fisher combination of p -values \mathcal{T}_W and \mathcal{T}_{sr} , the global CvM statistic \mathcal{T}_B and the distance covariance \mathcal{T}_{dCov} .

4.2.2 Multivariate sets of random variables

Multivariate nonparametric independence tests are supposed to have power against alternative hypotheses of mutual and groupwise dependence. In the following, we consider mutual dependence first. Being representative for diverse copula structures the power curves under a Gaussian copula with Gaussian and Cauchy marginals are displayed in Figure 4 for increasing levels of correlation ρ among all pairs of variables.⁶ For $\rho = 0$, the considered tests exhibit an empirical level close to $\alpha = 0.05$. Similar to the bivariate scenarios, the shape of all power curves shows comparable characteristics for a multivariate Gaussian distribution. The curves displayed in the left hand side panel of Figure 4 uncover slight power differences between the distinct test procedures for correlation levels between $\rho = 0$ and 0.4. In particular, the CvM distance \mathcal{T}_B appears to outperform the other tests.

As displayed in the right hand side panel of Figure 4, both the distance covariance \mathcal{T}_{dCov} and Wilks' Lambda \mathcal{T}_{Lm} perform poorly under Cauchy marginals in terms of power. The distance covariance \mathcal{T}_{dCov} might suffer from power losses under a distribution lacking finite moments (cf. Section 4.2.1), while the parametric test \mathcal{T}_{Lm} relies on the assumption of Gaussian distributed variables (see Wilks, 1935). Furthermore, Wilks' Lambda \mathcal{T}_{Lm} exceeds the nominal significance level of $\alpha = 0.05$ under Cauchy marginals.

In summary, in the considered multivariate sets the tests perform in analogy to the bivari-

⁶The asymptotic properties of Wilks' Lambda have been shown under the multivariate Gaussian distribution. Thus, the comparison with the nonparametric tests is informative on the trade-off between efficient dependence detection within the Gaussian model, and robustness under more general distributional conditions.

ate case under distinct marginals and copulas, nonlinear nonmonotone dependence structures and the further modifications. Nevertheless, it is worth mentioning that, e.g., performance differences between \mathcal{T}_B and \mathcal{T}_W might reflect distinct combinations of p -values. Given the results in Genest et al. (2007), heterogeneous power properties (more precisely, power leads of \mathcal{T}_W) could result in higher dimensions for which the combination method might become more important.

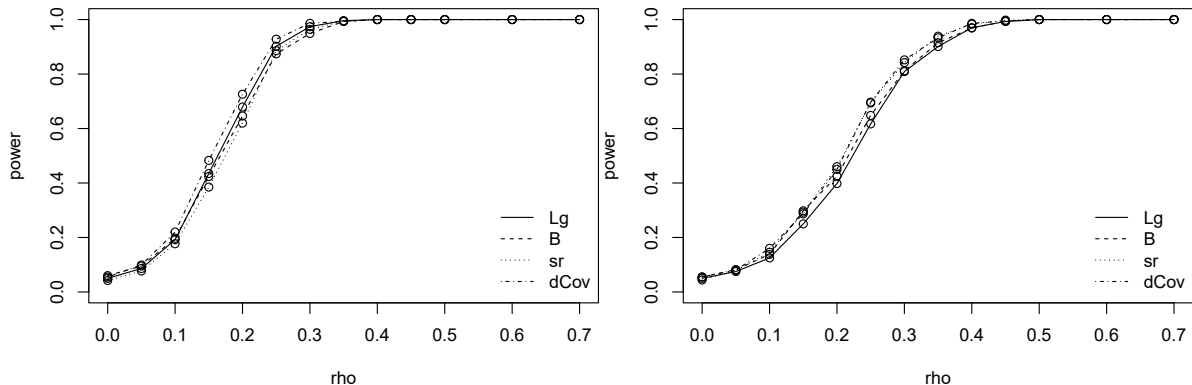


Figure 5: Power curves for multivariate groupwise dependence modeled by means of the Gaussian copula, intra group dependence modeled as in (25) with $\rho_{intra} = 0$ (left) and $\rho_{intra} = 0.8$ (right).

If the random variables can be aggregated to groups, it might be more interesting to analyze the strength of dependence between the groups of variables (and not within the groups). In Figure 5, power curves are shown for a trivariate set of Gaussian variables with dependence between the marginals determined by means of the covariance matrix in (25). Intra group dependence is fixed whereas inter group dependence varies between 0 and 0.7. The power curves in Figure 5, for $\rho_{intra} = 0$ (left) and $\rho_{intra} = 0.8$ (right), show characteristics which are comparable with the results for alternatives of mutual dependence. In both cases the power of all test statistics equals unity for levels of inter group correlation in excess of $\rho_{inter} = 0.5$. However, the higher ρ_{intra} the slower is the convergence to a power of unity. The performance differences between the tests are relatively small for the standard copula structures. The resulting power properties for variations of marginals and copulas are not displayed here for space considerations but show qualitatively identical characteristics as discussed above for the multivariate dependence structures.

We lastly study a four dimensional setting which is supposed to indicate how the results further generalize to higher dimensions. While only displaying exemplary results, we refer

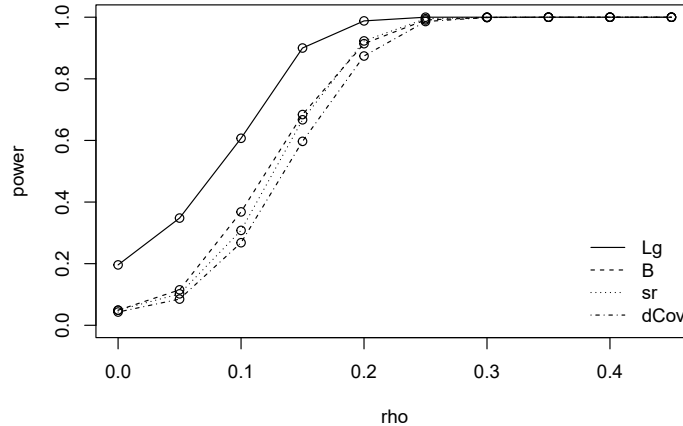


Figure 6: Power curves for multivariate groupwise dependence modeled by means of the Gaussian copula with correlation modeled as in (26) with $\rho_{intra} = 0$.

the reader to Kojadinovic & Holmes (2009) and Taskinen et al. (2005), who compare their suggested tests (namely, \mathcal{T}_B and \mathcal{T}_{sr}) with Wilks' Lambda \mathcal{T}_{Lg} in higher dimensions. The power curves in Figure 6 show the behavior of the test statistics \mathcal{T}_{Lg} , \mathcal{T}_B , \mathcal{T}_{sr} and \mathcal{T}_{dCov} under the Gaussian copula with Gaussian marginals. The correlation structure between the variables is generated following the matrix in (26) without intra group dependence, i.e., $\rho_{intra} = 0$. We find similar power properties for all tests but Wilks' Lambda which shows strong oversizing under the null hypothesis of independence. Overall, the detected properties of the nonparametric tests for independence can largely be generalized in higher dimensions.

5 Diagnosing dependence patterns for childhood undernutrition

After analyzing diverse pairwise, groupwise and mutual dependence structures by means of a simulation study, this section illustrates the performance of the tests by means of an application to empirical data. We consider data for childhood undernutrition, one of the most urgent public health challenges in developing and transition countries. In studying these data, we are interested in the relation between distinguished measures of undernutrition and a set of child's and mother's characteristics recorded in 1998/99 in the state Uttar Pradesh in Northern India (provided by Demographic and Health Surveys, DHS, www.measuredhs.com). In the following, we apply independence tests to subsamples of $n = 87$ and $n = 55$ children

at the age of 3 month ($age = 3$) and 9 month ($age = 9$), respectively.

The impact of certain characteristics on undernutrition might be quantified by means of regression models (see, for instance, Kandala et al., 2001; Klein & Kneib, 2016). As a prerequisite, the considered nonparametric independence tests provide guidance for such subsequent analysis. Childhood undernutrition itself can be measured by means of three distinct criteria. First, acute undernutrition (*wasting*) measures insufficient weight for given height. Second, chronic undernutrition (*stunting*) measures insufficient height given age. Third, both forms of undernutrition are captured by means of measuring insufficient weight given age (*underweight*). We apply bivariate and groupwise independence tests to study the relationship between two undernutrition measures, namely *wasting* and *underweight*, and their relationship to two of their possible determinants, namely the mother’s age and the mother’s body mass index. It might be noted that further results for the remaining bivariate combinations of indices, e.g. $\{underweight, stunting\}$, are similar to those displayed below.

First, we apply the bivariate independence tests Wilks’ Lambda \mathcal{T}_L , the CvM statistic \mathcal{T}_B , the spatial rank based statistic \mathcal{T}_{sr} and the distance covariance \mathcal{T}_{dCov} to the set $\{underweight, wasting\}$. The test statistics and the corresponding p -values are displayed in Table 5. The test results indicate dependence between these two indices of malnutrition at a significance level of $\alpha = 0.1$ for both samples with $age = 3$ and $age = 9$. Furthermore, all tests except for Wilks’ Lambda \mathcal{T}_L indicate significant dependence at level $\alpha = 0.05$ in the sample of $age = 3$. The distinguished outcomes of \mathcal{T}_L and the nonparametric tests for $age = 3$ might result from an underlying dependence structure that differs from the bivariate normal model, for instance, including tail dependence (subsequent investigations of the precise dependence structure could, for instance, follow Rosco & Joe, 2013). Accordingly, the level of dependence between the indices might be stronger for more extreme levels of undernutrition. Additionally, the dependence between *underweight* and *wasting* is indicated to be stronger in the second sample ($age = 9$), since the corresponding p -values are throughout below 0.005. For older children ($age = 9$) it might be more likely that both forms of undernutrition, rather than only one, are observed jointly.

Moreover, we investigate the dependence between the two dimensional set of $\{underweight, wasting\}$ and two of the mother’s characteristics, namely, the mother’s age at birth (*mage*)

Table 5: Independence test results for the set $\{underweight, wasting\}$ with respect to the child’s age ($age = 3$ and $age = 9$ month) and based on samples of sizes $n = 87$ and $n = 55$, respectively.

$\{underweight, wasting\}$		\mathcal{T}_L	\mathcal{T}_B	\mathcal{T}_{sr}	\mathcal{T}_{dCov}
$n = 87, age = 3$	statistic	5.163	0.090	7.956	20.946
	p -value	0.076	0.009	0.005	0.01
$n = 55, age = 9$	statistic	22.232	0.219	23.657	40.938
	p -value	0.000	0.001	0.000	0.005

and the mother’s body mass index ($mbmi$).⁷ We apply the same tests as in the bivariate setting in their multivariate form (studied in Section 4.1.2) to the two dimensional set of malnutrition indices $\mathbf{x}_1 = \{underweight, wasting\}$ and one further characteristic being either $\mathbf{x}_2 = \{mage\}$ or $\mathbf{x}_2 = \{mbmi\}$, respectively.

The test results are documented in Table 6. We can diagnose marked differences between the test outcomes. Studying the dependence between malnutrition and the mother’s age, i.e. $\mathbf{x}_2 = \{mage\}$, none of the considered tests except for Wilks’ Lambda \mathcal{T}_L for $age = 3$ leads to a rejection of the independence hypothesis with 10% significance. For $age = 9$ the p -values of the CvM statistic \mathcal{T}_B and the distance covariance \mathcal{T}_{dCov} are smaller but still do not indicate dependence in the second sample ($age = 9$) with significance of 10%. In contrast, the p -values of Wilks’ Lambda \mathcal{T}_L and the spatial rank based statistic \mathcal{T}_{sr} are larger in the sample of nine month old children in comparison with three month old children. In light of the simulation results discussed in Section 4.2 this discrepancy could, on the one hand, reflect a nonlinear relationship that differs from an elliptical distribution. On the other hand, the smaller sample size and the stronger dependence within \mathbf{x}_1 , i.e. between *underweight* and *wasting*, could explain performance weaknesses in the sample with $age = 9$ (see Section 4.2.2).

In contrast, the test results partly indicate dependence between the mothers’s body mass index and the two dimensional undernutrition index of their children. For instance, the null

⁷The whole set of characteristics, i.e. possible covariates in a regression model, are listed in Klein & Kneib (2016) and the references therein. Klein & Kneib (2016) further describe the nonlinear effects of the covariates on the bivariate distribution using nutrition data from all over India.

Table 6: Independence test results for the sets $\mathbf{x}_1 = \{\textit{underweight}, \textit{wasting}\}$ and $\mathbf{x}_2 = \{\textit{mage}\}$ or $\mathbf{x}_2 = \{\textit{mbmi}\}$ with respect to the child’s age ($\textit{cage} = 3$ and $\textit{cage} = 9$ month) and based on samples of sizes $n = 87$ and $n = 55$, respectively.

$\mathbf{x}_1 = \{\textit{underweight}, \textit{wasting}\}$			\mathcal{T}_L	\mathcal{T}_B	\mathcal{T}_{sr}	\mathcal{T}_{dCov}
$\mathbf{x}_2 = \{\textit{mage}\}$	$\textit{cage} = 3$	statistic	5.542	0.015	2.027	3.476
		p -value	0.063	0.840	0.363	0.585
	$\textit{cage} = 9$	statistic	1.356	0.026	1.527	4.642
		p -value	0.508	0.363	0.466	0.420
$\mathbf{x}_2 = \{\textit{mbmi}\}$	$\textit{cage} = 3$	statistic	3.490	0.053	4.150	2.920
		p -value	0.175	0.041	0.126	0.055
	$\textit{cage} = 9$	statistic	2.620	0.082	6.229	3.568
		p -value	0.270	0.007	0.044	0.070

hypothesis of independence is rejected with 10% significance by means of the CvM statistic \mathcal{T}_B and the distance covariance \mathcal{T}_{dCov} in both samples ($\textit{cage} = 3, 9$). Based on the sign rank based statistic \mathcal{T}_{sr} we can only diagnose dependence for $\textit{cage} = 9$. In contrast, by means of Wilks’ Lambda \mathcal{T}_L independence cannot be rejected and throughout, the p -values are even larger for $\textit{cage} = 9$ in comparison with $\textit{cage} = 3$. These distinct test results point to a nonlinear, possibly nonmonotone, and at least non Gaussian dependence structure.

Overall, the test results are in line with the results of Klein & Kneib (2016) who study the dependence between childhood undernutrition and a set of the child’s and their mother’s characteristics by means of copula regressions for data from all over India. Our results show that the dependence for $\textit{cage} = 9$ is stronger as it is for $\textit{cage} = 3$, and might exhibit a non elliptical distribution in both samples. In line with our dependence diagnosis Klein & Kneib (2016) characterize the dependence between *wasting* and *underweight* by means of a bivariate Clayton copula obtaining a larger dependence coefficient in the sample of children aged 9 months.

Applying the multivariate tests we have detected dependence between the mother’s body mass index (*mbmi*) and the bivariate set of undernutrition measures, and we are led to expect a nonlinear form of dependence. Furthermore, the relation between the mother’s age (*mage*) and the undernutrition indices $\{\textit{underweight}, \textit{wasting}\}$ lacks significance. Testing for

independence between distinct combinations of possible covariates and the bivariate response variable might serve to select covariates with significant explanatory content.

In summary, performance differences between the considered tests show up in most of the samples of the nutrition data. The independence tests benefit from satisfactory power even for samples of small size. Moreover, the multivariate tests are applicable to large sets of variables to diagnose between or within dependence in a flexible way. Thereby, dependence between the undernutrition measures and the set of all determining characteristics, as well as mutual dependence within the set of indices could be assessed in further investigations.

6 Conclusions

Nonparametric tests for independence provide a useful basis to decide if the multivariate distribution of random variables merely relies on their marginal distributions, or if it is worth to undertake the specification of a dependence structure. Meeting basic distributional assumptions, nonparametric independence tests have been developed to detect various forms of dependence between two or more random variables. We have described several dependence structures fundamentally, and provided a comprehensive overview of the theoretical background of multivariate nonparametric independence tests. Our review comprises traditional tests, as well as more recently suggested approaches based on spatial signs and ranks, the empirical copula and the distance covariance.

In a comparative simulation study we consider diverse distributional settings, such as (non)linear copula dependencies, nonmonotone structures and some modifications which point at diverse potential applications. A simulation study unravels distinguished size and power properties under the null hypothesis and specific dependence alternatives, respectively. As a general conclusion, our results do not indicate one overall most powerful test. Rather, the form of dependence appears crucial for the tests to perform preferably. Whereas under multivariate normality the tests show almost equivalent performance, the choice of the tests should be made more cautiously under non Gaussian distributional settings. In particular, the distance covariance performs poorly under distributions which lack finite moments. Furthermore, one might not be able to detect a nonmonotone nonlinear dependence structure by means of spatial rank based tests whereas the distance covariance performs best under this dependence alternative. The test based on the Cramér-von Mises (CvM) statistic seems to be most robust

to the diversity of dependence structures. Generally, merits and drawbacks of the alternative tests found in bivariate settings are confirmed for trivariate and four dimensional tests on mutual and groupwise dependence.

In an application to malnutrition data we find that distinguished test outcomes are informative for diverse forms of dependence between the variables and its strength even in samples of small size. Consequently, their nonlinear relation might be subjected to further analysis, for instance, by means of a semiparametric regressions.

The literature on nonparametric independence tests is growing, and already provides refinements of the methods discussed in this work. For instance, Ding & Li (2014) combine the distance covariance and copula based measures which might lead to power gains in the case of a singular copula. Similarly, the set of Hilbert-Schmidt independence criteria (Sejdinovic et al., 2013) promise improvements of dependence diagnosis over the stylized nonparametric approaches compared here. While our results hint at test specific performance patterns, it appears a fruitful avenue of future research to characterize merits and risks of most recent dependence diagnostics under diverse distributional settings and higher dimensionality by means of simulation studies.

Acknowledgements

We thank two anonymous referees and the Associate Editor Christine H. Müller for helpful comments and suggestions. Financial support from the Academy of Finland (308628) is gratefully acknowledged.

References

- Beran, R., Bilodeau, M., & Lafaye de Micheaux, P. (2007). Nonparametric tests of independence between random vectors. *J. Multivariate Anal.*, 98(9), 1805–1824.
- Blum, J. R., Kiefer, J., & Rosenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function. *Ann. Math. Statist.*, 32, 485–498.
- Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized arch model. *Review of Economics and Statistics*, 72, 498–505.

- Deheuvels, P. (1981). An asymptotic decomposition for multivariate distribution-free tests of independence. *J. Multivariate Anal.*, 11(1), 102–113.
- Diks, C. (2009). Nonparametric tests for independence. In R. A. Meyers (Ed.), *Encyclopedia of Complexity and Systems* (pp. 6252–6271). Springer New York.
- Ding, A. A. & Li, Y. (2014). Copula correlation : An equitable dependence measure and extension of pearson’s correlation. arXiv:1312.7214v3.
- Genest, C., Quessy, J.-F., & Rémillard, B. (2006). Local efficiency of a Cramér-von Mises test of independence. *J. Multivariate Anal.*, 97(1), 274–294.
- Genest, C., Quessy, J.-F., & Remillard, B. (2007). Asymptotic local efficiency of Cramér-von Mises tests for multivariate independence. *Ann. Statist.*, 35(1), 166–191.
- Genest, C. & Rémillard, B. (2004). Tests of independence and randomness based on the empirical copula process. *Test*, 13(2), 335–370.
- Harrell, F. E. (2015). *Hmisc*. R package version 3.15-0.
- Hoeffding, W. (1948). A non-parametric test of independence. *Ann. Math. Statistics*, 19, 546–557.
- Hofert, M., Kojadinovic, I., Maechler, M., & Yan, J. (2015). *copula: Multivariate Dependence with Copulas*. R package version 0.999-13.
- Joe, H. (1997). *Multivariate models and dependence concepts*, volume 73 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.
- Josse, J. & Holmes, S. (2014). Measures of dependence between random vectors and tests of independence. literature review. arXiv:1307.7383v3.
- Kandala, N., Lang, S., Klasen, S., & Fahrmeir, L. (2001). Semiparametric analysis of the socio-demographic and spatial determinants of undernutrition in two african countries. *Research in Official Statistics*, 1, 81–100.
- Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, 30(1/2), 81–93.

- Klein, N. & Kneib, T. (2016). Simultaneous inference in structured additive conditional copula regression models: a unifying bayesian approach. *Stat Comput*, 26(4), 841–860.
- Kojadinovic, I. & Holmes, M. (2009). Tests of independence among continuous random vectors based on Cramér-von Mises functionals of the empirical copula process. *J. Multivariate Anal.*, 100(6), 1137–1154.
- Manner, H. & Reznikova, O. (2012). A survey on time-varying copulas: Specification, simulations, and application. *Econometric Reviews*, 31(6), 654–687.
- Matteson, D. S. & Tsay, R. S. (2013). Independent component analysis via distance covariance. *pre-print*. <http://arxiv.org/abs/1306.4911>.
- Nakatani, T. (2010). *Conditional Correlation GARCH models*. R package version 0.2.3.
- Oja, H. (2010). *Multivariate nonparametric methods with R*, volume 199 of *Lecture Notes in Statistics*. Springer, New York. An approach based on spatial signs and ranks.
- Pearson, K. (1920). Notes on the history of correlation. *Biometrika*, (13).
- Puri, M. L. & Sen, P. K. (1971). *Nonparametric methods in multivariate analysis*. John Wiley & Sons, Inc., New York-London-Sydney.
- Rémillard, B. (2009). Discussion of: Brownian distance covariance. *Ann. Appl. Stat.*, 3(4), 1295–1298.
- Rémillard, B. (2014). Tests of independence. In M. Lovric (Ed.), *International Encyclopedia of Statistical Science* (pp. 1598–1601). Springer Berlin Heidelberg.
- Risk, B. B., James, N. A., & Matteson, D. S. (2015). *steadyICA: ICA and Tests of Independence via Multivariate Distance Covariance*. R package version 1.0.
- Rizzo, M. L. & Szekely, G. J. (2014). *energy: E-statistics (energy statistics)*. R package version 1.6.2.
- Rosco, J. & Joe, H. (2013). Measures of tail asymmetry for bivariate copulas. *Statistical Papers*, 54(3), 709–726.

- Sejdinovic, D., Sriperumbudur, B., Gretton, A., & Fukumizu, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann. Statist.*, *41*(5), 2263–2291.
- Shih, J.-H. & Emura, T. (2016). Bivariate dependence measures and bivariate competing risks models under the generalized fgm copula. *Statistical Papers*.
- Siqueira Santos, S. d., Takahashi, D. Y., Nakata, A., & Fujita, A. (2013). A comparative study of statistical methods used to identify dependencies between gene expression signals. *Briefings in bioinformatics*, *15*(6), 906–18.
- Sirkia, S., Miettinen, J., Nordhausen, K., Oja, H., & Taskinen, S. (2018). *SpatialNP: Multivariate nonparametric methods based on spatial signs and ranks*. R package version 1.1-3.
- Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, *8*, 229–231.
- Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, *15*, 72–101.
- Székely, G. J. & Rizzo, M. L. (2009). Brownian distance covariance. *Ann. Appl. Stat.*, *3*(4), 1236–1265.
- Székely, G. J., Rizzo, M. L., & Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.*, *35*(6), 2769–2794.
- Taskinen, S., Oja, H., & Randles, R. H. (2005). Multivariate nonparametric tests of independence. *J. Amer. Statist. Assoc.*, *100*(471), 916–925.
- Wilks, S. (1935). On the independence of k sets of normally distributed statistical variables. *Econometrica*, *3*(3), 309–326.