Journal of Machine Learning Research 18 (2017) 1-40 Submitted 4/16; Revised 7/17; Published 8/17

Tests of Mutual or Serial Independence of Random Vectors with Applications

Martin Bilodeau BILODEAU@DMS.UMONTREAL.CA

Aurélien Guetsop Nangue GUETSOPN@DMS.UMONTREAL.CA
Département de mathématiques et de statistique
Université de Montréal
C.P. 6128, Succursale A
Montréal, Canada H3C 3J7

Editor: Arthur Gretton

Abstract

The problem of testing mutual independence between many random vectors is addressed. The closely related problem of testing serial independence of a multivariate stationary sequence is also considered. The Möbius transformation of characteristic functions is used to characterize independence. A generalization to p vectors of distance covariance and Hilbert-Schmidt independence criterion ($HSIC$) tests with the translation invariant kernel of a stable probability distribution is proposed. Both test statistics can be expressed in a simple form as a sum over all elements of a componentwise product of p doubly-centered matrices. It is shown that an $HSIC$ statistic with sufficiently small scale parameters is equivalent to a distance covariance statistic. Consistency and weak convergence of both types of statistics are established. Approximation of p-values is made by randomization tests without recomputing interpoint distances for each randomized sample. The dependogram is adapted to the proposed tests for the graphical identification of sources of dependencies. Empirical rejection rates obtained through extensive simulations confirm both the applicability of the testing procedures in small samples and the high level of competitiveness in terms of power. Applications to meteorological and financial data provide some interesting interpretations of dependencies revealed by dependograms.

Keywords: Distance covariance, Hilbert-Schmidt independence criterion, Möbius transformation, mutual independence, serial independence

1. Introduction

The problem of testing for independence between p components of a random vector has attracted considerable attention in statistics. Many nonparametric procedures exist in the literature. A natural approach is to consider a functional of the difference between the empirical joint distribution and the product of the empirical marginal distributions. This same approach can also use empirical characteristic functions. When the functional of the difference is above a certain threshold, the components are declared dependent. Csörgő (1985), Kankainen (1995), Sejdinovic et al. (2013b) and Fan et al. (2017) considered mutual tests of independence based on empirical characteristic functions. However, when dependence is declared, it is not possible to identify, with their proposed tests, subsets of

variables responsible for the dependence. This limitation is similar to that of a global F-test in an analysis of variance model with one fixed factor, as opposed to multiple comparisons procedures, or that of a global chi-square test of independence in a multi-way contingency table, as opposed to log-linear models with interaction terms. For tests of independence, a useful method is the Möbius transformation.

The Möbius transformation defined in (1) of Section 2 has a long history in statistics. The Möbius transformation of distribution functions was first proposed in Blum et al. (1961) for $p = 3$. The general case was treated in Deheuvels (1981), Ghoudi et al. (2001), Genest and Rémillard (2004), Kojadinovic and Holmes (2009), Kojadinovic and Yan (2011), and Duchesne et al. (2012). It can also be defined with characteristic functions as in Bilodeau and Lafaye de Micheaux (2005), with half-space probabilities as in Beran et al. (2007), or with cell probabilities in a contingency table as in Bilodeau and Lafaye de Micheaux (2009). The first appearance of a Möbius transformation, although not stated explicitly, goes back to the work of Lancaster (1951) on contingency tables as explained in Bilodeau and Lafaye de Micheaux (2009). The machine learning community (Sejdinovic et al., 2013a) proposed kernel nonparametric tests for Lancaster three-variable interaction. This test is in fact a test based on the empirical version of the Möbius transformation of the characteristic function when $p = 3$. The general Möbius transformation considered in this paper can be used to build tests for general interactions of any order, as well as tests of mutual and serial independence.

The paper is organized as follows. Section 2 introduces the Möbius transformation of characteristic functions. It presents a characterization of the mutual independence between p random vectors by the Möbius transformation. In Section 3, new tests based on the Möbius transformation of empirical characteristic functions are introduced. They generalize the Hilbert-Schmidt independence criterion ($HSIC$) test (Gretton et al., 2005, 2008) and the distance covariance test (Székely et al., 2007) to the case $p > 2$. This work addresses the case of finite-dimensional Euclidean spaces. $HSIC$ was originally defined more generally using any semimetric space of negative type (as in the distance covariance), or any Borel measurable space on which a kernel is defined (Sejdinovic et al., 2013b). For example, in Gretton et al. (2008), dependence was detected between text in different languages using kernels on strings. On the other hand, this manuscript proposes a criterion that is more general than $HSIC$ in a different respect, via subsets of components in the Möbius transformation, where the criterion coincides with $HSIC$ when the latter is specialized to Euclidean spaces and $p = 2$. The new test statistics have a common form as a sum of elements of a componentwise product of p doubly-centered matrices. An equivalence is established between an $HSIC$ statistic with infinitesimal scale parameters and a distance covariance statistic. The weak convergence of the empirical processes based on the Möbius transformation is proved. The consistency and weak convergence of the $HSIC$ and distance covariance functionals are also established. A difficulty encountered in establishing the asymptotic independence of the collection of distance covariance functionals, over all subsets of components, is described. Other competing nonparametric tests of independence are reviewed in Section 4.

Rather than relying on the limiting distribution to conduct tests, Section 5 describes randomization tests as an approximation to permutation tests. In Section 6, p-values obtained by randomization tests for all possible subsets of components are combined using

the methods of Fisher (1950, pp. 99-101) or Tippett (1952) to obtain a global test of mutual independence. Section 7 describes the dependogram, a graphical device of Genest and Rémillard (2004), in the context of $HSIC$ and distance covariance tests. Section 8 adapts all the results described for the mutual independence situation to the problem of testing for the serial independence of a multivariate stationary sequence. Computational costs are given in Section 9 together with a short description of `R` (R Core Team, 2015) packages for nonparametric independence tests. Simulated models are considered in Section 10 to verify that the proposed tests have empirical significance levels close to the nominal level in small samples and comparable or higher powers in many situations when compared to existing tests such as those of Csörgő (1985), Kojadinovic and Holmes (2009) or Kojadinovic and Yan (2011).

Finally, Section 11 contains an application to real data on variables related to air temperature, soil temperature, humidity, wind, and evaporation. $HSIC$ or distance covariance tests should be preferred to the Gaussian likelihood ratio test since a multivariate Gaussian model is rejected by the test of Henze and Zirkler (1990). According to these tests, wind does not exhibit any dependence with all other variables considered. Another application finds significant serial dependencies in the S&P/TSX composite, DOW JONES, and S&P500 daily percent increasing rates ranging from January 2, 2014 to March 2, 2016. The strongest dependency observed at a lag of 4 days by a distance covariance test is interpreted using a broken line regression model as the tendency of stock markets to recover in the days following a sharp decline.

2. Möbius Transformation

The Möbius transformation of characteristic functions is a powerful tool for the characterization of mutual independence between p random vectors $Z^{(1)}, \ldots, Z^{(p)}$. The dimension of the vector $Z^{(j)}$ is d_j, for $j = 1, \ldots, p$. Let f be the joint characteristic function of these p vectors, and let $f^{(j)}$ be the marginal characteristic function of $Z^{(j)}$. Mutual independence is characterized by the factorization

$$f(t^{(1)}, \ldots, t^{(p)}) = \prod_{j=1}^{p} f^{(j)}(t^{(j)}),$$

for all $t^{(1)}, \ldots, t^{(p)}$. It may also be characterized by the Möbius transformation which is defined as follows. Let $\mathcal{I}_p$ be the family of subsets B of $\{1, \ldots, p\}$ of cardinality $|B| > 1$. The set $\mathcal{I}_p$ has $2^p - p - 1$ elements since the empty set is excluded, as well as all p singletons. For any $B \in \mathcal{I}_p$ and any $t^{(1)}, \ldots, t^{(p)}$, define $t^{(B)} = (t^{(j)} : \ j \in B)$. Similarly, $Z^{(B)} = (Z^{(j)} : j \in B)$, and $f^{(B)}$ is the joint characteristic function of $Z^{(B)}$. The Möbius transformation of the characteristic function f for the set $B \in \mathcal{I}_p$ is given by

$$\mu_B(t^{(B)}) = \sum_{C \subseteq B} (-1)^{|B \setminus C|} f^{(C)}(t^{(C)}) \prod_{j \in B \setminus C} f^{(j)}(t^{(j)}). \tag{1}$$

By convention, for $C = \emptyset$, both $f^{(C)}(t^{(C)})$ and $\prod_{j \in C} f^{(j)}(t^{(j)})$ are equal to one. The following characterization holds: $Z^{(1)}, \ldots, Z^{(p)}$ are mutually independent if and only if, $\mu_B(t^{(B)}) = 0$,

for all $B \in \mathcal{I}_p$, and all vectors $t^{(B)}$. A proof by induction of this characterization using distribution functions is given in Ghoudi et al. (2001) and is immediately applicable to characteristic functions.

3. Dependence Statistics

Consider $Z_k = (Z_k^{(1)}, \ldots, Z_k^{(p)})$, $k = 1, \ldots, n$, an independent and identically distributed sample of size n. The Möbius processes corresponding to μ_B, $B \in \mathcal{I}_p$, are defined as

$$R_{nB}(t^{(B)}) = \sqrt{n} \sum_{C \subseteq B} (-1)^{|B \setminus C|} f_n^{(C)}(t^{(C)}) \prod_{j \in B \setminus C} f_n^{(j)}(t^{(j)}), \tag{2}$$

where

$$f_n^{(C)}(t^{(C)}) = \frac{1}{n} \sum_{k=1}^{n} e^{i\langle t^{(C)}, Z_k^{(C)} \rangle} \tag{3}$$

is the empirical characteristic function. When $C = \{j\}$ is a singleton, the notation used for $f_n^{(C)}$ is simply $f_n^{(j)}$. The empirical processes R_{nB}, for all $B \in \mathcal{I}_p$, are illustrated when $p = 3$ in Table 1. The processes R_{nB} are identical to the empirical characteristic independence processes

$$S_{nB}(t^{(B)}) = \sqrt{n} \left[f_n^{(B)}(t^{(B)}) - \prod_{j \in B} f_n^{(j)}(t^{(j)}) \right]$$

in Csörgő (1985) only for subsets B of cardinality 2. The process S_{nB} appears later in the test statistic $\mathcal{J}_n^2$ in (17) used for testing the hypothesis of mutual independence. For $B = \{1, 2, 3\}$, the process

$$S_{nB}(t^{(B)}) = \sqrt{n} \left[f_n^{(1,2,3)}(t^{(1)}, t^{(2)}, t^{(3)}) - f_n^{(1)}(t^{(1)}) f_n^{(2)}(t^{(2)}) f_n^{(3)}(t^{(3)}) \right]$$

can be contrasted with the process R_{nB} in Table 1. Although, at first sight, R_{nB} may look more complicated than S_{nB} and both processes converge weakly to Gaussian processes, the processes R_{nB} have major advantages which are enunciated in Section 4.

B	$R_{nB}(t^{(B)})/\sqrt{n}$
$\{1,2\}$	$f_n^{(1,2)}(t^{(1)}, t^{(2)}) - f_n^{(1)}(t^{(1)}) f_n^{(2)}(t^{(2)})$
$\{1,3\}$	$f_n^{(1,3)}(t^{(1)}, t^{(3)}) - f_n^{(1)}(t^{(1)}) f_n^{(3)}(t^{(3)})$
$\{2,3\}$	$f_n^{(2,3)}(t^{(2)}, t^{(3)}) - f_n^{(2)}(t^{(2)}) f_n^{(3)}(t^{(3)})$
$\{1,2,3\}$	$-f_n^{(1,2,3)}(t^{(1)}, t^{(2)}, t^{(3)}) + f_n^{(1,2)}(t^{(1)}, t^{(2)}) f_n^{(3)}(t^{(3)}) + f_n^{(1,3)}(t^{(1)}, t^{(3)}) f_n^{(2)}(t^{(2)})$ $+f_n^{(2,3)}(t^{(2)}, t^{(3)}) f_n^{(1)}(t^{(1)}) - 2 f_n^{(1)}(t^{(1)}) f_n^{(2)}(t^{(2)}) f_n^{(3)}(t^{(3)})$

Table 1: Empirical processes R_{nB}, $B \in \mathcal{I}_3$.

The dependence statistic for the subset B is now defined as the Cramér-von Mises functional

$$W_{nB} = \frac{1}{n} \int |R_{nB}(t^{(B)})|^2 dw_B(t^{(B)}), \tag{4}$$

where $dw_B(t^{(B)}) = \prod_{j\in B} dw^{(j)}(t^{(j)})$ is a product measure. The evaluation of this integral is facilitated using another representation of the process. First, recall the multinomial formula (Ghoudi et al., 2001)

$$\sum_{C\subseteq B}\left(\prod_{i\in C} u^{(i)}\right)\left(\prod_{j\in B\setminus C} v^{(j)}\right) = \prod_{i\in B}\left(u^{(i)} + v^{(i)}\right). \tag{5}$$

Then, the empirical process (2) can be written as

$$R_{nB}(t^{(B)}) = \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\prod_{j\in B}\left[e^{i\langle t^{(j)}, Z_k^{(j)}\rangle} - f_n^{(j)}(t^{(j)})\right]. \tag{6}$$

The representation (6) is obtained after replacing the expression (3) for the empirical characteristic function in (2) and by applying the multinomial formula (5). The representation given by (6) allows the integral (4) to be evaluated explicitly in some cases and simplifies the proof of theorems to come. Two important cases are now presented.

3.1 Hilbert-Schmidt Independence Criterion

Assume that the measure $dw_B(t^{(B)}) = \prod_{j\in B} dG^{(j)}(t^{(j)})$ is a product of symmetric (around the origin) probability measures. The population Hilbert-Schmidt independence criterion is

$$\mathcal{H}_B^2 = \int |\mu_B(t^{(B)})|^2 \prod_{j\in B} dG^{(j)}(t^{(j)}) \tag{7}$$

which is well defined since the function μ_B is bounded. For the sample version, let $\varphi^{(j)}$ be the (real) characteristic function of $G^{(j)}$. The sample version of the Hilbert-Schmidt independence criterion (7) is denoted $\mathcal{H}_{nB}^2$ and has the following explicit expression.

Theorem 1 *For any $B \in \mathcal{I}_p$, the dependence statistic $\mathcal{H}_{nB}^2$ is given by*

$$\mathcal{H}_{nB}^2 = \frac{1}{n^2}\sum_{k=1}^{n}\sum_{l=1}^{n}\prod_{j\in B} A_{kl}^{(j)}, \tag{8}$$

where

$$\begin{aligned} a_{kl}^{(j)} &= \varphi^{(j)}(Z_k^{(j)} - Z_l^{(j)}), \\ A_{kl}^{(j)} &= a_{kl}^{(j)} - \bar{a}_{k.}^{(j)} - \bar{a}_{.l}^{(j)} + \bar{a}_{..}^{(j)}, \\ \bar{a}_{k.}^{(j)} &= \frac{1}{n}\sum_{l=1}^{n} a_{kl}^{(j)},\ \bar{a}_{.l}^{(j)} = \frac{1}{n}\sum_{k=1}^{n} a_{kl}^{(j)},\ \bar{a}_{..}^{(j)} = \frac{1}{n^2}\sum_{k,l=1}^{n} a_{kl}^{(j)}. \end{aligned} \tag{9}$$

By definition, matrices $A^{(j)} = (A_{kl}^{(j)})$, $j = 1, \ldots, p$, are doubly-centered, *i.e.* rows and columns of these matrices sum up to zero. In the special case $p = 2$ of testing the independence between two vectors, the statistic $\mathcal{H}_{nB}^2$, for $B = \{1, 2\}$, is the Hilbert-Schmidt independence criterion, or $HSIC$, with translation invariant kernels $\varphi^{(j)}(Z_k^{(j)} - Z_l^{(j)})$, $j = 1, 2$. A proof of Theorem 1 for $p = 3$ was provided by Sejdinovic et al. (2013a).

An important special case is when $G^{(j)}$ is the stable distribution of index $\alpha \in (0,2]$ with scale parameter $\beta_j > 0$ first studied by Lévy (1925). Then, the translation invariant kernel is

$$a_{kl}^{(j)} = e^{-\beta_j^\alpha |Z_k^{(j)} - Z_l^{(j)}|_{d_j}^\alpha}, \tag{10}$$

where $|\cdot|_{d_j}$ is the Euclidean norm in dimension d_j. The corresponding dependence statistic is then denoted $\mathcal{H}_{nB}^{2(\alpha)}$. The case $\alpha = 2$ is the Gaussian kernel, and $\alpha = 1$ is the Cauchy kernel, often referred to as the Laplace kernel in machine learning (Gretton et al., 2005, 2009) because of its similarity to a Laplace density in dimension one.

The following result establishes the consistency of the Hilbert-Schmidt independence criterion $\mathcal{H}_{nB}^2$.

Theorem 2

(i) $\mathcal{H}_{nB}^2 \overset{a.s.}{\to} \mathcal{H}_B^2$, *as* $n \to \infty$.

(ii) *If* $\mu_B(t^{(B)}) \neq 0$ *for some vector* $t^{(B)}$, *then* $n\mathcal{H}_{nB}^2 \overset{a.s.}{\to} \infty$, *as* $n \to \infty$.

3.2 Distance Covariance

Assume that the measure $dw_B(t^{(B)}) = \prod_{j \in B} dw^{(j)}(t^{(j)})$ is a product of non integrable measures. The measure $dw^{(j)}$ of index $0 < \alpha < 2$ is defined as

$$dw^{(j)}(t^{(j)}) = \left[C(d_j, \alpha) |t^{(j)}|_{d_j}^{d_j + \alpha} \right]^{-1} dt^{(j)},$$

with the normalizing constant

$$C(d, \alpha) = 2\pi^{d/2} \Gamma(1 - \alpha/2) / \left[\alpha 2^\alpha \Gamma((d + \alpha)/2) \right].$$

A similar representation as in (8) also holds. The corresponding dependence statistic is denoted $\mathcal{V}_{nB}^{2(\alpha)}$, which is the usual notation for distance covariance of index α.

Theorem 3 *Let* $0 < \alpha < 2$. *Then, for any* $B \in \mathcal{I}_p$, *the dependence statistic* $\mathcal{V}_{nB}^{2(\alpha)}$ *has the same form as in (8) of Theorem 1 with*

$$a_{kl}^{(j)} = -|Z_k^{(j)} - Z_l^{(j)}|_{d_j}^\alpha. \tag{11}$$

In the special case $|B| = 2$, the statistic $\mathcal{V}_{nB}^{2(\alpha)}$ reduces to the distance covariance of index α in Feuerverger (1993) for the case $d_1 = d_2 = 1$, and later generalized by Székely et al. (2007) to the case $d_1 \geq 1$, $d_2 \geq 1$. A very special case requiring a separate analysis is when $|B| = 2$ and $\alpha = 2$. In this case, $\mathcal{V}_{nB}^{2(2)}$ is the numerator of the RV coefficient of Escoufier (1973) as noticed by Josse and Holmes (2016) when $d_1 \geq 1$ and $d_2 \geq 1$, and earlier by Székely et al. (2007) only when $d_1 = d_2 = 1$. It should be noted that the case $\alpha = 2$ leads only to a test of non correlation but not of independence, unless the joint distribution is Gaussian. For this reason, the value $\alpha = 2$ will not be considered for distance covariance in the sequel.

Székely et al. (2007) showed the consistency of distance covariance for $0 < \alpha < 2$ and $|B| = 2$. Consistency of distance covariance for $|B| > 2$ is now established. Only when necessary, the notation $\mathbb{E}_3$ is for the expectation with respect to Z_3, treating the other variable as a constant to avoid using conditional expectations.

Theorem 4 *Let* $0 < \alpha < 2$. *Assume*

$$\mathbb{E} \prod_{j \in B} |Z_1^{(j)} - Z_2^{(j)}|_{d_j}^{\alpha} < \infty. \tag{12}$$

Define

$$\mathcal{V}_B^{2(\alpha)} = \mathbb{E} \prod_{j \in B} \left[-|Z_1^{(j)} - Z_2^{(j)}|_{d_j}^{\alpha} + \mathbb{E}_3 |Z_1^{(j)} - Z_3^{(j)}|_{d_j}^{\alpha} + \mathbb{E}_3 |Z_2^{(j)} - Z_3^{(j)}|_{d_j}^{\alpha} - \mathbb{E}|Z_3^{(j)} - Z_4^{(j)}|_{d_j}^{\alpha} \right].$$

Then,

(i) $\mathcal{V}_B^{2(\alpha)} = \int |\mu_B(t^{(B)})|^2 dw_B(t^{(B)}) < \infty$.

(ii) *If* $\mu_B(t^{(B)}) \neq 0$ *for some vector* $t^{(B)}$, *then* $n\mathcal{V}_{nB}^{2(\alpha)} \stackrel{a.s.}{\to} \infty$, *as* $n \to \infty$.

For $0 < \alpha < 2$, the following limit establishes that $\mathcal{V}_{nB}^{2(\alpha)}$ is, for all practical purpose, equivalent to $\mathcal{H}_{nB}^{2(\alpha)}$ when scale parameters β_j, $j \in B$, are sufficiently small:

$$\lim_{\beta_j \to 0, \forall j \in B} \mathcal{H}_{nB}^{2(\alpha)} / \prod_{j \in B} \beta_j^{\alpha} = \mathcal{V}_{nB}^{2(\alpha)}. \tag{13}$$

This result, proved in the appendix, implies that $\mathcal{H}_{nB}^{2(\alpha)}$, properly normalized with sufficiently small scale parameters, can be as close as desired to $\mathcal{V}_{nB}^{2(\alpha)}$ and thus, $\mathcal{H}_{nB}^{2(\alpha)}$ will have a power function indistinguishable from that of $\mathcal{V}_{nB}^{2(\alpha)}$. For semimetrics generated by kernels, Sejdinovic et al. (2013b, Theorem 24) established an equivalence between distance covariance and $HSIC$. However, for distance covariance defined in terms of a weighted distance between characteristic functions, one can not find a continuous translation invariant kernel for which $HSIC$ coincide with distance covariance (Sejdinovic et al., 2013b, Section 5.3). Nevertheless, (13) provides an equivalence of a different nature: for α-stable distributions, appropriately normalized $HSIC$ and distance covariance are equivalent in the limit, as the scale parameters converge to zero.

As simple as it may seem, this equivalence, for the simplest case $|B| = 2$, has gone unnoticed in the discussions of distance covariance (Székely and Rizzo, 2009; Gretton et al., 2009). In Section 8.2 of Sejdinovic et al. (2013b), for $|B| = 2$, the $HSIC$ test based on $\mathcal{H}_{nB}^{2(2)}$ with Gaussian kernels with scale parameters set at the inverse of median of interpoint distances is compared to distance covariance tests of varying index α. It was found in the independent component analysis benchmark example that $\mathcal{V}_{nB}^{2(1/3)}$ is more powerful than $\mathcal{H}_{nB}^{2(2)}$. From (13), the $HSIC$ test based on $\mathcal{H}_{nB}^{2(1/3)}$, with translation invariant kernels of the stable distribution of index $\alpha = 1/3$ and very small scale parameters, would have a power function indistinguishable from that of $\mathcal{V}_{nB}^{2(1/3)}$. In another example with sinusoidally dependent data, the $HSIC$ test based on $\mathcal{H}_{nB}^{2(2)}$ has a very poor power function compared to $\mathcal{V}_{nB}^{2(1/6)}$. Sejdinovic et al. (2013b) explained: "the exponent in the distance-induced kernel plays a similar role as the bandwidth of the Gaussian kernel, and smaller exponents are able to detect dependencies at smaller lengthscales. Poor performance of the Gaussian kernel with median bandwidth in this example is a consequence of the mismatch between the overall

lengthscale of the marginal distributions (captured by the median interpoint distances) and the lengthscales at which dependencies are present". In fact, the exponent in the distance-induced kernel of a distance covariance test plays the same role as the index of the translation invariant kernel of a stable distribution in an $HSIC$ test. Indistinguishable power functions can be obtained by choosing sufficiently small scale parameters in the translation invariant kernel. This all means that $HSIC$ with sufficiently small scale parameters always match distance covariance in terms of power. But $HSIC$ with scale parameters appropriately selected may, in some cases, improve on distance covariance.

3.3 Asymptotic Distribution

Empirical processes as in (2) have been recently very useful at tackling problems related to mutual independence because of the simplicity of the asymptotic distribution. Let $d_B = \sum_{j\in B} d_j$. Each process R_{nB} is defined on the space $C(\mathbb{R}^{d_B}, \mathbb{C})$ of complex-valued continuous functions defined on $\mathbb{R}^{d_B}$. Let $d = \sum_{j=1}^{p} d_j$ and $t = (t_1, \ldots, t_d) \in \mathbb{R}^d$. The following mild tail condition (Csörgő, 1981, 1985) is assumed

$$\int_0^1 \frac{\bar{\psi}(h)}{h\left(\log \frac{1}{h}\right)^{1/2}} dh < \infty, \tag{14}$$

where

$$\bar{\psi}(h) = \sup\left\{ y:\ 0 \le y \le 1,\ \lambda_d\left\{ t:\ |t|_\infty < \frac{1}{2},\ \psi(t) < y \right\} < h \right\},$$

with λ_d standing for the Lebesgue measure in $\mathbb{R}^d$ and $|t|_\infty = \max(|t_1|, \ldots, |t_d|)$, is the nondecreasing rearrangement of the function $\psi(t) = [1 - Ref(t)]^{1/2}$. Weak convergence of the collection of processes R_{nB} is now established. For details concerning the metrics on the spaces in Theorem 5, the reader is referred to the appendix. The symbol $\Rightarrow$ stands for weak convergence.

Theorem 5 *If $Z^{(1)}, \ldots, Z^{(p)}$ are mutually independent, then the process $R_{nB} \Rightarrow R_B$ in $C(\mathbb{R}^{d_B}, \mathbb{C})$, where R_B is a zero mean complex Gaussian process with $\bar{R}_B(t^{(B)}) = R_B(-t^{(B)})$ and complex covariance function*

$$\mathbb{E}\left[R_B(t^{(B)}) \bar{R}_B(s^{(B)}) \right] = \prod_{j\in B} [f^{(j)}(t^{(j)} - s^{(j)}) - f^{(j)}(t^{(j)}) f^{(j)}(-s^{(j)})].$$

Moreover, the collection of processes $(R_{nB}:\ B \in \mathcal{I}_p) \Rightarrow (R_B:\ B \in \mathcal{I}_p)$ on the product of spaces $\underset{B\in\mathcal{I}_p}{\times} C(\mathbb{R}^{d_B}, \mathbb{C})$ to a zero mean Gaussian process such that the marginal processes R_B, $B \in \mathcal{I}_p$, are mutually independent.

The convergence of functionals (4) also holds even though they are not defined on the whole space $C(\mathbb{R}^{d_B}, \mathbb{C})$, but only on the space of squared integrable functions. In the next theorem, the asymptotic distribution of $n\mathcal{H}_{nB}^2$ is described. In particular, it provides the asymptotic distribution of $n\mathcal{H}_{nB}^{2(\alpha)}$, for any $\alpha \in (0, 2]$.

Theorem 6 *Let $W_{nB} = \mathcal{H}_{nB}^2$. If $Z^{(1)}, \ldots, Z^{(p)}$ are mutually independent, then $nW_{nB} \Rightarrow W_B$ for each $B \in \mathcal{I}_p$, where $W_B = \int |R_B(t^{(B)})|^2 \prod_{j\in B} dG^{(j)}(t^{(j)})$. Moreover, the collection*

of variables $(nW_{nB} : \ B \in \mathcal{I}_p) \Rightarrow (W_B : \ B \in \mathcal{I}_p)$*, where the variables* W_B*,* $B \in \mathcal{I}_p$*, are mutually independent.*

Theorem 6 holds without moment conditions for $\mathcal{H}_{nB}^2$ since in this case dw_B is a probability measure. Theorems 5 and 6 were proved when $p = 2$ by Zhang et al. (2011). The case for $p = 3$ was covered in Appendix E of Sejdinovic et al. (2013a). The distribution of W_B can be represented using the Karhunen-Loève expansion. Without loss of generality, let $B = \{1, \ldots, k\}$. Then, W_B is distributed as

$$\sum_{i_1=1}^{\infty} \cdots \sum_{i_k=1}^{\infty} \lambda_{i_1}^{(1)} \ldots \lambda_{i_k}^{(k)} Z_{i_1 \ldots i_k}^2, \tag{15}$$

where $Z_{i_1 \ldots i_k}$ are independent standard normal variables, and for $j = 1, \ldots, k$, $\lambda_1^{(j)}, \lambda_2^{(j)}, \ldots$ are eigenvalues depending only on the probability measure $P^{(j)}$ of $Z^{(j)}$ and the weighting probability measure $dG^{(j)}$. From arguments as in Sejdinovic et al. (2013a) and Sejdinovic et al. (2013b), the eigenvalues $\lambda_1^{(j)}, \lambda_2^{(j)}, \ldots$ are those of the integral operator

$$S_{\tilde{k}^{(j)}} g(s^{(j)}) = \int_{\mathbb{R}^{d_j}} \tilde{k}^{(j)}(s^{(j)}, t^{(j)}) g(t^{(j)}) dP^{(j)}(t^{(j)}),$$

where $k^{(j)}(s^{(j)}, t^{(j)}) = \varphi^{(j)}(t^{(j)} - s^{(j)})$ is the translation invariant kernel (9) and

$$\tilde{k}^{(j)}(s^{(j)}, t^{(j)}) = k^{(j)}(s^{(j)}, t^{(j)}) - \mathbb{E}k^{(j)}(Z_1^{(j)}, t^{(j)}) - \mathbb{E}k^{(j)}(s^{(j)}, Z_2^{(j)}) + \mathbb{E}k^{(j)}(Z_1^{(j)}, Z_2^{(j)}), \tag{16}$$

for $Z_1^{(j)}$, $Z_2^{(j)}$ independent and distributed according to $P^{(j)}$, is the corresponding doubly-centered kernel.

The same type of results for $|B| = 2$ have been obtained by Székely et al. (2007) for distance covariance, see also Lyons (2013, Corollary 2.8 and Remark 2.9) for the product structure of eigenvalues. For $|B| > 2$, provided that $\mathbb{E} \prod_{j \in B} |Z_1^{(j)} - Z_2^{(j)}|_{d_j}^{2\alpha}$ is finite, the V-statistic structure of $W_{nB} = \mathcal{V}_{nB}^{2(\alpha)}$ can be used as in Lyons (2013, Theorem 2.7) or Sejdinovic et al. (2013a) to show that $nW_{nB} \Rightarrow W_B$, where W_B is of the same form as in (15). However, this V-statistic argument does not establish the asymptotic independence of the collection of variables $(W_B : \ B \in \mathcal{I}_p)$. It can not be proven either as in Theorem 6 since the generalization of a result of Kellermeier (1980) used in the proof no longer holds since it is based on Jensen's inequality which is valid only for probability measures. The asymptotic independence of the collection could be concluded if the collection of processes $(R_{nB} : \ B \in \mathcal{I}_p)$ were independent for each n, unfortunately this is not the case. For distance covariance, the asymptotic independence of the collection remains unanswered.

In Section 6, p-values of global tests, such as tests of Fisher in (18) or Tippett in (19), computed by combining individual p-values for each $B \in \mathcal{I}_p$ must be approximated. Simple approximations assume that individual p-values are mutually independent. However, in finite samples with sample size as small as 100, an approximation relying on randomization tests which does not rely on the independence of individual p-values was found to yield better conformity of global p-values to the nominal significance level of 0.05.

4. Other Functionals

The test of Kojadinovic and Holmes (2009, Proposition 13) is based on the Möbius decomposition of the independence empirical copula process and is defined as the Cramér-von Mises functional (4) in which the process R_{nB} in (2) defined with empirical characteristic functions is replaced by an analogous process defined with empirical copulas. Also, the integrating measure dw_B is replaced by the uniform distribution over the hypercube. The dependence statistics W_{nB} thus obtained will be denoted KH^2_{nB}. The statistics KH^2_{nB} can also be represented in the form (8) with appropriately defined terms $a^{(j)}_{kl}$.

The test of Beran et al. (2007) uses the Möbius decomposition of the independence empirical half-space process and is defined as the Kolmogorov statistic obtained by taking the supnorm of the processes. The Kolmogorov statistics do not have an explicit form as in (8). They must be computed by solving a costly optimization problem over discretized unit spheres of dimensions d_j and p-values are approximated by the bootstrap. Distance covariance and $HSIC$ tests are only orthogonally invariant, whereas the test of Beran et al. (2007) is invariant to the general linear group. The heavy computational cost of this test makes it unsuitable for large scale simulations of power functions. For this reason, it will not be considered in the simulations of Section 10.

A global test of mutual independence can also be constructed directly from the empirical characteristic independence process (Csörgő, 1985; Kankainen, 1995; Sejdinovic et al., 2013a; Fan et al., 2017),

$$\begin{aligned}\mathcal{J}^2_n &= \int \left| f_n(t^{(1)},\ldots,t^{(p)}) - \prod_{j=1}^{p} f^{(j)}_n(t^{(j)}) \right|^2 \prod_{j=1}^{p} dG^{(j)}(t^{(j)}) \\ &= \frac{1}{n^2}\sum_{k=1}^{n}\sum_{l=1}^{n}\prod_{j=1}^{p} a^{(j)}_{kl} - \frac{2}{n^{p+1}}\sum_{k=1}^{n}\prod_{j=1}^{p}\sum_{l=1}^{n} a^{(j)}_{kl} + \frac{1}{n^{2p}}\prod_{j=1}^{p}\sum_{k=1}^{n}\sum_{l=1}^{n} a^{(j)}_{kl}, \qquad (17)\end{aligned}$$

where $dG^{(j)}$, $j = 1,\ldots,p$, are probability measures and $a^{(j)}_{kl}$ is defined in (9). The special case of the stable distribution of index α with $a^{(j)}_{kl}$ defined in (10) yields a statistic denoted as $\mathcal{J}^{2(\alpha)}_n$. This choice of weight function, except for $\alpha = 2$, seems to have been overlooked in the literature and no reference to this choice appears in the recent paper by (Fan et al., 2017). In the more general context of random variables taking values in separable metric spaces, Pfister et al. (2017) embedded the joint distribution and the product of the marginals in a reproducing kernel Hilbert space and defined the p-component Hilbert-Schmidt independence criterion as the squared distance between the embeddings. This framework provides a global test of mutual independence which contains (17) as a special case. Also, similar tests to (17) based on empirical distribution functions have been proposed by Blum et al. (1961), Cotterill and Csörgő (1982), and Cotterill and Csörgő (1985) in the univariate case, *i.e.* $d_j = 1$ for $j = 1,\ldots,p$. Kojadinovic and Holmes (2009, Proposition 10) also considered in the multivariate case a test similar to (17) based on empirical copulas.

Simulations in Section 10 will compare $HSIC$ tests $\mathcal{H}^{2(\alpha)}_{nB}$ and distance covariance tests $\mathcal{V}^{2(\alpha)}_{nB}$ to the tests KH^2_{nB} and $\mathcal{J}^{2(2)}_n$. Dependence statistics resulting from the Möbius decomposition have the following advantages over statistics of the type $\mathcal{J}^2_n$ in (17).

1. The statistic W_{nB} in (8) has a simpler structure than that of $\mathcal{J}_n^2$ in (17). The integral operator defining the eigenvalues associated with $\mathcal{J}_n^2$ is an integral of dimension $d = \sum_{j=1}^{p} d_j$ which can not be written as a product of integral operators as soon as $p > 2$. The integral operator for W_{nB} is always a product of $|B|$ integral operators in smaller dimensions d_j, $j \in B$.

2. The p-values of W_{nB}, $B \in \mathcal{I}_p$, can be combined (see Section 6) to get a global test of mutual independence with a predetermined global significance level.

3. When a global test in item 2 rejects the mutual independence hypothesis, subsets $B \in \mathcal{I}_p$ yielding small p-values can be identified as the possible source of dependence using a dependogram described in Section 7.

5. Approximate p-values

Ways to approximate the null distribution of test statistics are now discussed.

5.1 Spectral Approach

The asymptotic distribution of $n\mathcal{J}_n^2$ in (17) is also an infinite linear combination of chi-squared variables with one degree of freedom with coefficients which are eigenvalues of an integral operator (Fan et al., 2017). Using estimated eigenvalues, they resorted to the algorithm of Imhof (1961), see Duchesne and Lafaye De Micheaux (2010), to compute the cumulative distribution function by a numerical inversion of the characteristic function. For $p = 2$, the spectral approach is also proposed by Zhang et al. (2017). After estimation of eigenvalues, rather than using the algorithm of Imhof (1961), they simulated a large number of values of (15) by generating independent random $N(0, 1)$ variables. The spectral approach to approximate (15) is appropriate for large sample sizes and kernels with an eigenspectrum which decays very rapidly such as the Gaussian kernel. For slowly decaying kernels, the number $(N_\lambda)^k$ of terms in (15), where N_λ is the number of eigenvalues considered, may be too large to apply such methods. The spectral approach for the very simple RV coefficient of Escoufier (1973) is appropriate only in large samples (Josse and Holmes, 2016).

5.2 Resampling Techniques

Another approximation is the *permutation test* which recomputes the statistics for all $(n!)^{p-1}$ permutations. From a theoretical point of view (Hoeffding, 1952), permutation tests are well known to guarantee a non asymptotic control of the significance level (by permutation invariance of the test statistic under the null hypothesis, that is mutual independence here). Since this is not feasible, even for moderate sample sizes, the strategy is to rely on resampling techniques. As an approximation to the permutation test, the *randomization test* simulates the null distribution by permuting (resample without replacement) the observations $Z_1^{(j)}, \ldots, Z_n^{(j)}$, independently for each component, a large number of, say, 1000 permutations. Sejdinovic et al. (2013a) proposed to approximate the null distribution of $n\mathcal{J}_n^2$ in (17) by a randomization test. Another technique is to resample with replacement, independently for each component, a large number of times which is a *bootstrap test*. Bootstrap tests and randomization tests are asymptotically equivalent in the sense that

the resulting critical values and power functions are appropriately close (Romano, 1989; van der Vaart and Wellner, 1996). General references for permutation tests are Efron and Tibshirani (1993), Good (2000), and Pesarin and Salmaso (2010).

5.3 Methods of Moments

Other approximations are based on the method of moments. The distribution obtained by recomputing a statistic for all $(n!)^{p-1}$ permutations is called the permutation distribution of the statistic. The exact first three moments of the permutation distribution of statistics of the general form (8) were obtained by Kazi-Aoual et al. (1995) when $p = 2$ and generalized by Guetsop Nangue (2016) to the case $p \geq 2$. The Pearson type III approximation is a shifted gamma distribution with the same first three moments as the permutation distribution. The Pearson type III distribution is part of the original system of distributions devised by Pearson (1895) in an effort to model visibly skewed observations. The first published paper in which a Pearson type III distribution is used as an approximation to a permutation test is Mielke et al. (1981). For an historical account of the Pearson type III distribution in the context of permutation tests, the reader is referred to Berry et al. (2016, Section 1.2.2). For $p = 2$, Gretton et al. (2008) proposed the approximation by a gamma distribution with the same first two moments as those of the asymptotic distribution, not the permutation distribution, in (15). These first two moments depend on unknown eigenvalues but can be estimated using traces of matrices involving the doubly-centered matrices $A^{(j)} = (A^{(j)}_{kl})$ without having to compute eigenvalues. An empirical failure mode of the gamma approximation was demonstrated in Gretton and Györfi (2010, Figure 1). More recently, Guetsop Nangue (2016, Table 2.9) simulated empirical significance levels of $HSIC$ for meta-Gaussian distributions in the case $p = 2$. The kernel was Gaussian with scaling set at the median of distances. For sample sizes $n = 15$, 25, 50, and 100, the Pearson type III approximation gave rates close to the nominal rate of 0.05 for dimensions up to $d_1 = d_2 = 50$, whereas the gamma approximation gave rates close or equal to 0 as the dimensions were increased. Although the exact permutation test guarantees a non asymptotic control of the significance level, it should be stressed that neither the Pearson type III nor the gamma approximations guarantee an asymptotic control of the significance level. Although p-values of individual statistics W_{nB} can be accurately approximated by the Pearson type III approximation for small samples, the independence of these p-values, guaranteed by the asymptotic distribution, holds to a satisfying degree only in large samples. This is particularly the case as soon as $p > 4$ and $d_j > 1$. Solely for this reason, p-values of global tests which combine individual p-values of tests for sub-hypotheses will be assessed by randomization tests.

5.4 Randomization Tests

Randomization tests are now described to approximate p-values of individual test statistics. Randomization tests for global tests which combine individual p-values are deferred to Section 6. Let Q_n be any test statistic such as $\mathcal{J}^2_n$, $\mathcal{V}^{2(\alpha)}_{nB}$ or $\mathcal{H}^{2(\alpha)}_{nB}$, $B \in \mathcal{I}_p$. Denote a permutation of $\{1, \ldots, n\}$ by $\sigma = (\sigma(1), \ldots, \sigma(n))$. Consider independent random permutations $\sigma_1, \ldots, \sigma_p$ of $\{1, \ldots, n\}$. For $j = 1, \ldots, p$, the permutation σ_j permutes the observations

$(Z_1^{(j)}, \ldots, Z_n^{(j)})$ to yield the permuted data $(Z_{\sigma_j(1)}^{(j)}, \ldots, Z_{\sigma_j(n)}^{(j)})$ of the component j. The statistic Q_n is then recomputed on the permuted data. To this end, note that one need not recompute the elements $a_{kl}^{(j)}$ in (10) or (11). For $\mathcal{J}_n^2$ in (17), one simply permutes according to σ_j the rows and corresponding columns of the $n \times n$ matrix $a^{(j)} = (a_{kl}^{(j)})$. For $\mathcal{V}_{nB}^{2(\alpha)}$ or $\mathcal{H}_{nB}^{2(\alpha)}$, this same argument applies to the matrix $A^{(j)} = (A_{kl}^{(j)})$ in (1).

Approximate p-values are now obtained by recomputing the test statistic N times. Generate Np independent random permutations $\sigma_{i,j}$, $i = 1, \ldots, N$, $j = 1, \ldots, p$.[1] For $i = 1, \ldots, N$, let $Q_{n,i}$ be the test statistic computed from the permuted data $(Z_{\sigma_{i,j}(1)}^{(j)}, \ldots, Z_{\sigma_{i,j}(n)}^{(j)})$, $j = 1, \ldots, p$. An approximate p-value is then obtained as follows.

(1) Let $Q_{n,0}$ be the test statistic computed from the original data.

(2) Generate N randomized samples from the original data and compute $Q_{n,i}$ for $i = 1, \ldots, N$.

(3) An approximate p-value is then given by

$$\frac{1}{N+1}\left[1 + \sum_{i=1}^{N} \mathbb{I}\left\{Q_{n,i} \geq Q_{n,0}\right\}\right].$$

This approximate p-value of the randomization test guaranties (by virtue of exchangeability of the variables $Q_{n,i}$, $i = 0, \ldots, N$) a non-asymptotic control of type I error rates for each individual test (Romano and Wolf, 2005, Lemma 1). The N randomized samples can also be used to compute an approximate quantile of the test statistic Q_n. Let $Q_{n,(1)}, \ldots, Q_{n,(N)}$ be the order statistics of $Q_{n,1}, \ldots, Q_{n,N}$, an approximate π-quantile of Q_n is the order statistic $Q_{n,(\lfloor N\pi \rfloor)}$. The finite sample control of the significance level of randomization tests will be assessed by simulations in Section 10.

6. Combining p-values

If the exact distribution function F_{nB} of W_{nB} was known, then $1 - F_{nB}(W_{nB})$ would be the exact p-value. Since F_{nB} is continuous, this p-value would be exactly uniformly distributed over the interval $(0, 1)$. Moreover, exact p-values obtained by varying B would also be approximately independent in large samples. This holds due to the asymptotic mutual independence of statistics W_{nB} in Theorem 6.

As an approximation, a p-value $\hat{p}_{nB}$ is computed for every dependence statistic W_{nB}, $B \in \mathcal{I}_p$, using randomization tests as in Section 5. Under the mutual independence hypothesis, these $2^p - p - 1$ p-values are, for large samples, approximately independent and also approximately uniformly distributed on $(0, 1)$. The quantity $-2 \log \hat{p}_{nB}$ thus has approximately a χ_2^2 null distribution. Fisher's global test statistic rejects the mutual independence when $-2 \sum_{B \in \mathcal{I}_p} \log \hat{p}_{nB}$ is large. It reports a global p-value of

$$\hat{p}_n = P\left(\chi_\nu^2 > -2 \sum_{B \in \mathcal{I}_p} \log \hat{p}_{nB}\right), \text{ where } \nu = 2(2^p - p - 1). \tag{18}$$

1. Since observations in one component can be held fixed, $N(p-1)$ permutations would suffice. However, treating all components the same way is convenient for purposes of notation and software development.

The number of subsets may be too large for some p or only low order interaction terms may be of interest. In this case, p-values of a test of mutual independence of order q, where $2 \leq q \leq p$, are computed as

$$\hat{p}_n = P\left(\chi^2_\nu > -2 \sum_{B \in \mathcal{I}_p, |B| \leq q} \log \hat{p}_{nB}\right), \text{ where } \nu = 2\sum_{i=2}^{q} \binom{p}{i}.$$

Another way of combining independent p-values leads to Tippett's global test with a rejection region consisting of small values of $\min_{B \in \mathcal{I}_p} \hat{p}_{nB}$. Tippett's test of mutual independence of order q reports a combined p-value of

$$\hat{p}_n = 1 - \left(1 - \min_{B \in \mathcal{I}_p, |B| \leq q} \hat{p}_{nB}\right)^r, \text{ where } r = \sum_{i=2}^{q} \binom{p}{i}. \tag{19}$$

In the simulations of Section 10, although each individual p-value is well approximated by a uniform variable on $(0, 1)$, the global p-values obtained by Fisher's (18) or Tippett's (19) method sometimes lead to a global test of significance level which exceeds the nominal level. This was observed especially for $p > 4$ and some $d_j > 1$ and can be attributed to the lack of independence between statistics W_{nB} in finite samples. For a test of mutual independence of order q, $2 \leq q \leq p$, it is preferable to obtain the global p-values using randomized samples as follows (Kojadinovic and Holmes, 2009).

(1) Compute the statistics $W_{nB,0}$, $B \in \mathcal{I}_p$ and $|B| \leq q$, from the original data.

(2) Generate N randomized samples from the original data and let $W_{nB,i}$ be the statistics from the ith randomized sample.

(3) An approximate p-value for the statistic $W_{nB,j}$ based on Romano and Wolf (2005) [2] is

$$\psi(W_{nB,j}) = \frac{1}{N+1}\left[1 + \sum_{i \in \{0,\ldots,N\} \setminus \{j\}} \mathbb{I}\left\{W_{nB,i} \geq W_{nB,j}\right\}\right], \ j = 0, 1, \ldots, N.$$

(4) For $i = 0, 1, \ldots, N$, compute Fisher's and Tippett's statistics

$$F_{n,i} = -2 \sum_{B \in \mathcal{I}_p, |B| \leq q} \log[\psi(W_{nB,i})] \text{ and } T_{n,i} = \min_{B \in \mathcal{I}_p, |B| \leq q} \psi(W_{nB,i}).$$

(5) Approximate p-values for the global tests of Fisher and Tippett are given respectively by

$$\frac{1}{N+1}\left[1 + \sum_{i=1}^{N} \mathbb{I}\left\{F_{n,i} \geq F_{n,0}\right\}\right] \text{ and } \frac{1}{N+1}\left[1 + \sum_{i=1}^{N} \mathbb{I}\left\{T_{n,i} \leq T_{n,0}\right\}\right].$$

2. The non-asymptotic control of type I error rates can not be concluded from the expression $\psi(W_{nB,j}) = \frac{1}{N+1}\left[\frac{1}{2} + \sum_{i=1}^{N} \mathbb{I}\left\{W_{nB,i} \geq W_{nB,j}\right\}\right]$ in Kojadinovic and Holmes (2009, p. 1152).

Note that variables $F_{n,i}$ (and $T_{n,i}$), $i = 0, \ldots, N$, are exchangeable since this is the case of variables $W_{nB,j}$, and thus also $\psi(W_{nB,j})$, $j = 0, \ldots, N$, for each fixed B. The p-values of the global tests of Fisher and Tippett in item (5) thus lead to a non-asymptotic control of type I global error rates (Romano and Wolf, 2005, Lemma 1).

Neither of the two methods is generally more powerful than the other. When only a few of the individual hypotheses are strongly false, Tippett's method is preferable. To detect alternatives for which many of the individual hypotheses are equally false, Fisher's method is likely to be preferable (Westberg, 1985; Loughin, 2004).

7. Dependogram

The dependogram is described for test statistics W_{nB}, $B \in \mathcal{I}_p$, which can be either $\mathcal{V}_{nB}^{2(\alpha)}$ or $\mathcal{H}_{nB}^{2(\alpha)}$. Genest and Rémillard (2004) introduced the dependogram which is a graphical tool, with subsets ordered by size on the horizontal axis, and corresponding values W_{nB} represented by vertical bars on the vertical axis. For each subset $B \in \mathcal{I}_p$, statistics W_{nB} are computed together with corresponding critical values given by the π-quantile, $c_{\pi B}$, obtained from N randomized samples as in Section 5. Critical values are represented by dashes. Dependence in a subset is declared when the vertical bar extends beyond the dash. If the number of subsets is large, a dependogram of order q, $2 \leq q \leq p$, can be constructed by considering only subsets of maximum cardinality q. Assuming statistics W_{nB} are independent, the choice $\pi = (1 - \alpha')^{1/r}$ with $r = \sum_{i=2}^{q} \binom{p}{i}$ leads to a global significance level α'. Indeed, under the mutual independence hypothesis

$$P(W_{nB} \leq c_{\pi B} : \ |B| \leq q) = \prod_{|B| \leq q} P(W_{nB} \leq c_{\pi B}) = \pi^r = 1 - \alpha'.$$

The dependogram should be seen as an exploratory tool that can be used when a global test rejects the mutual independence hypothesis. It helps to identify the dependency structures present in the data as in Figure 4. A dependogram built from $HSIC$ statistics will be normalized as in (13) so that, for small scale parameters, it has roughly the same order of magnitude as the dependogram based on distance covariance statistics.

A difficulty in interpreting a dependogram is now illustrated with an example. Consider a model with three dependent variables $Z^{(1)}$, $Z^{(2)}$ and $Z^{(3)}$, where $Z^{(1)}$ and $Z^{(2)}$ are dependent, but the pair $(Z^{(1)}, Z^{(2)})$ is independent of $Z^{(3)}$. Looking at Table 1 with the true characteristic functions, one can verify that $\mu_{12} \not\equiv 0$, $\mu_{13} \equiv 0$, $\mu_{23} \equiv 0$, and $\mu_{123} \equiv 0$. Therefore, only the test for the subset $\{1, 2\}$ is expected to be significant. The test for the subset $\{1, 2, 3\}$ is not expected to be significant even though the three variables are dependent. Of course, the global test is expected to be significant as it will combine tests for all subsets.

8. Tests of Serial Independence

The problem of testing for serial independence of a multivariate stationary ergodic sequence is now addressed. The test statistic in the serial context is very similar. Consider a stationary ergodic sequence $Y_1, Y_2, \ldots$ in $\mathbb{R}^{d_1}$, where Y_1 is distributed according to the characteristic function $f^{(1)}$. Let $p \geq 2$ be a fixed integer. For a sequence of length m, let $n = m - p + 1$

and

$$Z_k = (Z_k^{(1)}, \ldots, Z_k^{(p)}) = (Y_k, Y_{k+1}, \ldots, Y_{k+p-1}),\ k = 1, \ldots, n, \tag{20}$$

where $Z_k^{(j)} = Y_{k+j-1}$, $j = 1, \ldots, p$. Let $f(t^{(1)}, \ldots, t^{(p)})$ be the joint characteristic function of $Z_1 = (Y_1, \ldots, Y_p)$ and $f^{(B)}$ be the joint characteristic function of $Z_1^{(B)} = (Y_j :\ j \in B)$. For example, if $p = 4$ and the indices in the subset $B = \{1, 2, 3\}$ are translated by one to yield $C = B + 1 = \{2, 3, 4\}$ then, $f^{(B)}$ and $f^{(C)}$ are the same characteristic function since the process is stationary. In the serial context, a subset B and its translate, say $B + k$, can be treated as the same subset. The set $\mathcal{I}_p$ is thus reduced to $\mathcal{B}_p = \{B \in \mathcal{I}_p : 1 \in B\}$ and has now cardinality $2^{p-1} - 1$. For a given $B \in \mathcal{B}_p$, the Möbius transformation of the characteristic function f for the set B is given by

$$\mu_{B,s}(t^{(B)}) = \sum_{C \subseteq B} (-1)^{|B \setminus C|} f^{(C)}(t^{(C)}) \prod_{j \in B \setminus C} f^{(1)}(t^{(j)}).$$

The subscript s stands for serial. The Möbius transformation characterizes the serial independence: $Y_1, \ldots, Y_p$ are mutually independent if and only if, $\mu_{B,s}(t^{(B)}) = 0$, for all $B \in \mathcal{B}_p$, and all vectors $t^{(B)}$.

The corresponding process in the serial case is denoted $R_{nB,s}$ and is defined exactly as in (2). Here, the index set of the process $R_{nB,s}$ is the Euclidean space $\mathbb{R}^{d_B}$, where $d_B = d_1|B|$. Note that all the empirical characteristic functions $f_n^{(j)}$, $j = 1, \ldots, p$, are now essentially the same estimate of the unknown characteristic function $f^{(1)}$. They are not replaced by a single estimate based on all n observations to preserve the representation of the functional (8) in terms of doubly-centered matrices. The dependence statistic for the subset B is now defined as the functional

$$W_{nB,s} = \frac{1}{n} \int |R_{nB,s}(t^{(B)})|^2 \prod_{j \in B} dw(t^{(j)}).$$

It can be computed as before in (8). Due to the stationarity, it should be noted that the same kernel is used to define the elements $a_{kl}^{(j)}$, for all $j \in B$. Two types of weighting measures are considered.

1. For $HSIC$, $dw(t^{(j)}) = dG(t^{(j)})$ is a probability measure with characteristic function φ in which case

$$a_{kl}^{(j)} = \varphi(Z_k^{(j)} - Z_l^{(j)}). \tag{21}$$

The dependence statistic is denoted $W_{nB,s} = \mathcal{H}_{nB,s}^2$. As in Theorem 2, the population Hilbert-Schmidt independence criterion

$$\mathcal{H}_{B,s}^2 = \int |\mu_{B,s}(t^{(B)})|^2 \prod_{j \in B} dG(t^{(j)})$$

is consistently estimated by $\mathcal{H}_{nB,s}^2$. The proof is omitted since it can be proven as Theorem 2 using the ergodic theorem. The special case when $dG(t^{(j)})$ is the probability measure of the stable distribution of index α is denoted $\mathcal{H}_{nB,s}^{2(\alpha)}$, for $\alpha \in (0, 2]$.

2. For distance covariance, $dw(t^{(j)}) = \left[C(d_1,\alpha)|t^{(j)}|_{d_1}^{d_1+\alpha}\right]^{-1}$, in which case $a_{kl}^{(j)} = -|Z_k^{(j)} - Z_l^{(j)}|_{d_1}^{\alpha}$. Under an appropriate moment condition as in Theorem 4, the consistency and weak convergence of the functional, which is denoted $W_{nB,s} = \mathcal{V}_{nB,s}^{2(\alpha)}$ for $\alpha \in (0,2)$, should hold.

Pinkse (1998) proposed a non parametric test based on characteristic functions of serial independence against serial dependence of fixed lag k that is consistent against all such alternatives. It is based on a consistent estimator for an upper bound of $\mathcal{H}_{B,s}^2$ for the subset $B = \{1, 1+k\}$. The control of the global significance level when multiple tests are done for several values of k is not addressed by this author. Diks and Panchenko (2007) constructed a test based on divergence measures between distributions using kernel-based quadratic forms to detect dependence at lag k using the subset $B = \{1, 1+k, \ldots, 1+(c-1)k\}$ of given cardinality c. Their criterion is written as $Q = Q_{11} - 2Q_{12} + Q_{22}$. They used U-statistics as estimates for each term. Unfortunately, an error in the estimates of the last two terms on p. 85 leads to an erroneous statistic. If V-statistics are used instead, it is not hard to verify that their test statistic is of the exact form (17).

Under the hypothesis of serial independence, the vectors $Z_k = (Y_k, Y_{k+1}, \ldots, Y_{k+p-1})$, $k = 1, \ldots, n$, are dependent due to some overlapping Y's. The proof of Theorem 5 is not directly applicable. Nevertheless, a very similar weak convergence theorem still holds.

Theorem 7 *Let $Y_1, Y_2, \ldots$ be independent and identically distributed. Then, for any fixed p, the process $R_{nB,s} \Rightarrow S_B$ in $C(\mathbb{R}^{d_B}, \mathbb{C})$ to a zero mean complex Gaussian processes S_B with complex covariance function*

$$\mathbb{E}\left[S_B(t^{(B)})\bar{S}_B(s^{(B)})\right] = \prod_{j\in B}\left[f^{(1)}(t^{(j)} - s^{(j)}) - f^{(1)}(t^{(j)})f^{(1)}(-s^{(j)})\right].$$

Moreover, the collection of processes $(R_{nB,s} : B \in \mathcal{B}_p) \Rightarrow (S_B : B \in \mathcal{B}_p)$ on the product of spaces $\underset{B\in\mathcal{B}_p}{\times} C(\mathbb{R}^{d_B}, \mathbb{C})$ to a zero mean Gaussian process such that the marginal processes S_B, $B \in \mathcal{B}_p$, are mutually independent.

In the next theorem, the asymptotic distribution of $n\mathcal{H}_{nB,s}^2$ is described. The proof is omitted since it is the same as for Theorem 6.

Theorem 8 *Let $W_{nB,s} = \mathcal{H}_{nB,s}^2$. If $Y_1, Y_2, \ldots$ are independent and identically distributed, then $nW_{nB,s} \Rightarrow W_{B,s}$ for each $B \in \mathcal{B}_p$, where $W_{B,s} = \int |S_B(t^{(B)})|^2 \prod_{j\in B} dG(t^{(j)})$. Moreover, the collection of variables $(nW_{nB,s} : B \in \mathcal{B}_p) \Rightarrow (W_{B,s} : B \in \mathcal{B}_p)$, where the variables $W_{B,s}$, $B \in \mathcal{B}_p$, are mutually independent.*

The distribution of $W_{B,s}$ can be also represented using the Karhunen-Loève expansion. Without loss of generality, let $B = \{1, \ldots, k\}$. Then, $W_{B,s}$ is distributed as in (15), where the eigenvalues $\lambda_1^{(j)}, \lambda_2^{(j)}, \ldots$ no longer depend on j, but only on the probability measure $P^{(1)}$ of Y_1 and the weighting probability measure $dG(t^{(1)})$. The eigenvalues $\lambda_1, \lambda_2, \ldots$ are those of the integral operator

$$S_{\tilde{k}}g(s^{(1)}) = \int_{\mathbb{R}^{d_1}} \tilde{k}(s^{(1)}, t^{(1)})g(t^{(1)})dP(t^{(1)}),$$

where $k(s^{(1)}, t^{(1)}) = \varphi(t^{(1)} - s^{(1)})$ is the translation invariant kernel (21) and

$$\tilde{k}(s^{(1)}, t^{(1)}) = k(s^{(1)}, t^{(1)}) - \mathbb{E}k(Y_1, t^{(1)}) - \mathbb{E}k(s^{(1)}, Y_2) + \mathbb{E}k(Y_1, Y_2),$$

for Y_1, Y_2 independent and distributed according to $P^{(1)}$, is the corresponding doubly-centered kernel.

Computations of p-values of individual tests and global tests in Sections 5 and 6 can also be used for testing for serial independence. Subsets B are now restricted to $B \in \mathcal{B}_p$. A modification to the dependogram is necessary since the null distributions of statistics of the same cardinality are now identical. The critical values $c_{\pi B}$ of Section 7 for all B of the same cardinality are replaced by a single critical value as in Figure 6. Since there are $\binom{p-1}{\omega-1}$ subsets B containing 1 and of cardinality ω, the single critical value is taken as the π-quantile of the amalgamated $N\binom{p-1}{\omega-1}$ statistics $W_{nB,i}$, for $i = 1, \ldots, N$ and $|B| = \omega$.

It has been shown how distance covariance and $HSIC$ tests can be adapted for testing for serial independence. The global test $\mathcal{J}_n^2$ in (17) can also be used for serial independence by defining subsequences of length p as in (20).

9. Computational Aspects

The complexity cost of all randomization tests based on the Möbius decomposition in this paper are of the order $O(n^2 r N)$, where n is the sample size, $r = \sum_{i=2}^{q} \binom{p}{i}$ is the number of subsets considered in a global test of order q, and N is the number of randomized samples. Even for the smallest value $q = 2$, the number of subsets of the order $r = O(p^2)$ increases quadratically with p. These tests thus become rapidly infeasible as n or p increases. The serial statistics in Section 8 have a lower complexity cost since in this case $r = \sum_{i=2}^{q} \binom{p-1}{i-1}$ yields a number of subsets of the order $r = O(p)$ increasing linearly with p when $q = 2$.

The statistics $\mathcal{H}_{nB}^{2(\alpha)}$ and $\mathcal{V}_{nB}^{2(\alpha)}$ are combined as in Section 6 to yield global statistics $\mathcal{H}_{n}^{2(\alpha)}$ and $\mathcal{V}_{n}^{2(\alpha)}$, respectively. Scale parameters β_j in (10) must be selected for $\mathcal{H}_{nB}^{2(\alpha)}$. They are set to

$$\beta_j = c_j / \text{med}_{k<l} |Z_k^{(j)} - Z_l^{(j)}|_{d_j}, \ j = 1, \ldots, p. \tag{22}$$

The choice $c_j = 1$ leads to the so-called heuristic method suggested by Gretton et al. (2008). Selection of very small constants c_j allows to check the equivalence between distance covariance and $HSIC$ tests as described in (13). Unless stated otherwise, the heuristic method will be used for $HSIC$ tests. Distance covariance and $HSIC$ tests will be compared to two other tests: the test of Kojadinovic and Holmes (2009) denoted KH_n^2 for the mutual independence case, or Kojadinovic and Yan (2011) denoted KY_n^2 for the serial case, and the test $\mathcal{J}_n^{2(2)}$ in (17) with heuristic scale parameters (22) as in Sejdinovic et al. (2013a).

When $d_j = 1$ for all $j = 1, \ldots, p$, the statistics KH_n^2 (or KY_n^2) are replaced by the equivalent statistic of Genest and Rémillard (2004) denoted GR_n^2. KH_n^2 differs from GR_n^2 only in the approximation used for p-values. The former uses a randomization test, whereas the latter takes advantage of the fact that for $d_j = 1$, the test is distribution free implying that critical values can be obtained by simulating the null distribution for given values of n and p. Thus, a single set of critical values can be used for all replicates.

The `copula R` package (Kojadinovic and Yan, 2010) functions `multIndepTest` and `multSerialIndepTest` were used for tests based on KH_n^2 and KY_n^2, respectively. Inci-

dentally, the implementation of KH_n^2 unnecessarily recomputes the doubly-centered matrices $A^{(j)}$ for every permutation in their `C` subroutine `bootstrap_MA_I`. The same package also contains the functions `indepTest` and `serialIndepTest` for GR_n^2. An `R` function for distance covariance and $HSIC$ tests proposed in this paper for $p \geq 2$ is available at `dms.umontreal.ca/~bilodeau`. For $p = 2$, it produces, up to a factor of $1/n$ depending on the definitions of statistics, the same result as the function `dcov.test` of the `energy` package (Rizzo and Szekely, 2016). Computations in Section 10 were done on an Intel(R) Core(TM) i7 with a CPU of 3.20 GHz.

10. Simulated Models

All empirical significance levels and empirical powers in this section are evaluated with 1000 tests (replicates) conducted at a global significance level $\alpha' = 0.05$. The p-values of all randomization tests are based on 1000 permutations (the default). Power results are summarized by graphics with a tick mark at 0.05 on the power axis to check for the conformity of empirical significance levels to the nominal level. Six tests are compared: distance covariance tests $\mathcal{V}_n^{2(1/2)}$ and $\mathcal{V}_n^{2(1)}$, $HSIC$ tests $\mathcal{H}_n^{2(1/2)}$ and $\mathcal{H}_n^{2(1)}$, the test KH_n^2 of Kojadinovic and Holmes (2009) and the test $\mathcal{J}_n^{2(2)}$ in (17). When all $d_j = 1$, the test KH_n^2 is replaced by GR_n^2 as explained in Section 9. Also, for the serial case, the test KH_n^2 is replaced by the test KY_n^2 of Kojadinovic and Yan (2011). Unless stated otherwise, p-values of tests based on the Möbius decomposition are combined using the method of Fisher.

10.1 Copula Models

Power comparisons are made for the Gaussian, Student, Frank and Clayton copulas. A general reference for copulas is the book by Nelsen (2006).

10.1.1 Gaussian and Student Copulas with Bivariate Marginals

As in Kojadinovic and Holmes (2009), a correlation matrix is constructed of the form

$$R = \begin{pmatrix} (1-\rho_w)I_2 + \rho_w J_2 J_2' & \rho_b J_2 J_2' & \rho_b J_2 J_2' \\ \rho_b J_2 J_2' & (1-\rho_w)I_2 + \rho_w J_2 J_2' & \rho_b J_2 J_2' \\ \rho_b J_2 J_2' & \rho_b J_2 J_2' & (1-\rho_w)I_2 + \rho_w J_2 J_2' \end{pmatrix}, \tag{23}$$

where I_2 is the identity matrix of dimension 2 and J_2 is the vector of ones of dimension 2. Notations w and b stand for within and between, respectively. Samples of size $n = 100$ are generated from a multivariate distribution of dimension 6 (Gaussian or Student with 2 degrees of freedom) with correlation matrix R in (23). Probability transforms are then applied so that all 6 variables are uniformly distributed on the interval $(0, 1)$. The resulting vector is partitioned into three two-dimensional vectors. The value of ρ_w is set to 0.5. For the Gaussian model, Figure 1 shows that the best test is KH_n^2 followed closely by $\mathcal{V}_n^{2(1)}$. However, for the Student model, KH_n^2 performs poorly compared to all the other tests, the most powerful test being $\mathcal{H}_n^{2(1/2)}$. For the Student model, the three components are always dependent even when $\rho_b = 0$ which explains the power at $\rho_b = 0$ in the right panel of Figure 1. In fact, the only elliptical distribution for which uncorrelatedness implies independence is the Gaussian distribution (Bilodeau and Brenner, 1999, Proposition 4.11).

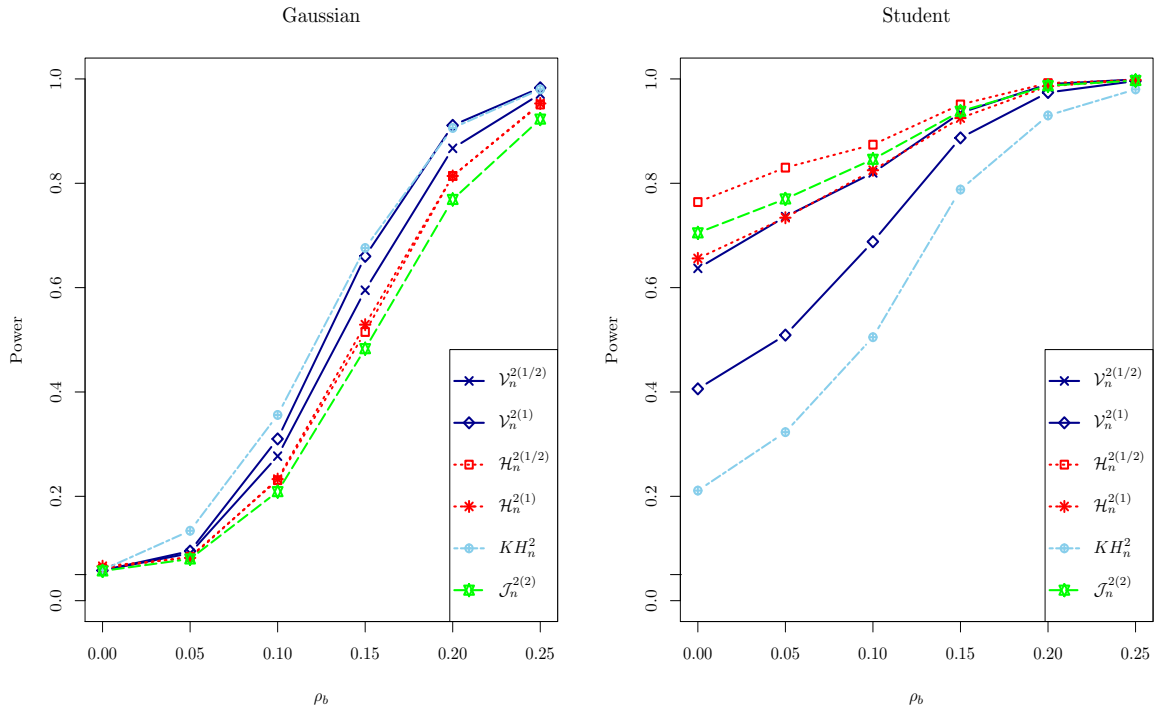

Figure 1: Empirical powers for the Gaussian (left panel) and Student (right panel) copulas with bivariate marginals of Section 10.1.1.

10.1.2 Frank and Clayton Copulas with Univariate Marginals

The Frank and Clayton copulas are now considered for $n = 100$, $p = 3$ and $d_j = 1$ for $j = 1, 2, 3$. These two copulas have a parameter θ with the value $\theta = 0$ corresponding to the independence copula. For the Frank copula, independence is obtained as the limiting case $\theta \to 0$. Figure 2 shows that for both copulas, the tests GR_n^2 and $\mathcal{V}_n^{2(1)}$ have very similar powers and are the most powerful. The least powerful test is $\mathcal{J}_n^{2(2)}$. The $HSIC$ tests are less powerful than their distance covariance counterparts, but their powers could be increased to those of distance covariance by selecting smaller scale parameters as predicted by (13).

10.1.3 Frank Copula with Bivariate Marginals

The Frank copula model is now considered for $p = 3$ and $d_j = 2$ for $j = 1, 2, 3$. The sample size is still 100. A random vector is generated from the Frank copula of dimension six with parameter θ and it is partitioned into three vectors of dimension two. The most powerful test in Figure 3 is $\mathcal{V}_n^{2(1)}$ and the least powerful is KH_n^2.

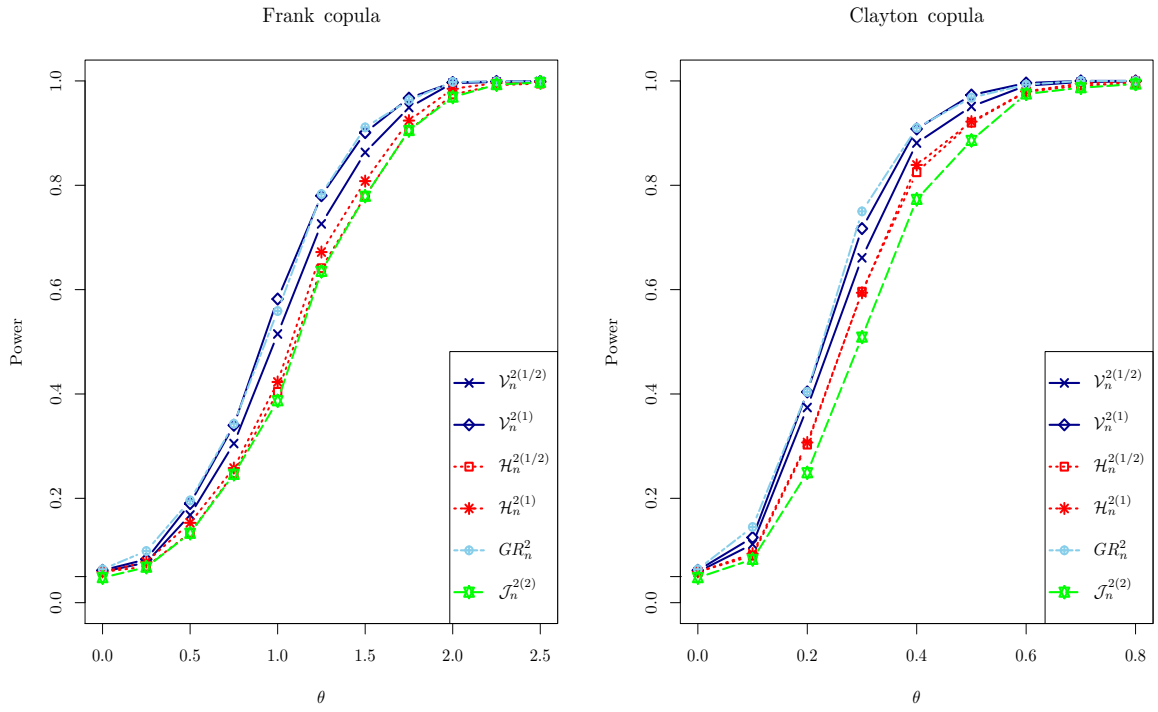

Figure 2: Empirical powers for the Frank (left panel) and Clayton (right panel) copulas with univariate marginals of Section 10.1.2.

10.2 Model of Romano and Siegel

The original model is taken from Kojadinovic and Holmes (2009) which extends an example in Genest and Rémillard (2004). Samples of size $n = 100$ are generated from the distribution of a 12-dimensional random vector as follows.

1. Generate a two-dimensional Gaussian vector $X^{(1)} = (X_1^{(1)}, X_2^{(1)})$ with means 0, variances 1, and covariance 0.5.

2. Generate two independent copies $Z^{(2)}$ and $Z^{(3)}$ of $X^{(1)}$.

3. Define $Z^{(1)} = (Z_1^{(1)}, Z_2^{(1)})$ by $Z_i^{(1)} = |X_i^{(1)}|\text{sign}(Z_1^{(2)} Z_1^{(3)})$, $i = 1, 2$.

4. Generate a three-dimensional Gaussian vector $Z^{(4)}$ with means 0, variances 1, and covariances 0.3.

5. Generate an independent copy $X^{(5)}$ of $Z^{(4)}$.

6. Define $Z^{(5)} = Z^{(4)} + X^{(5)}$.

Following Romano and Siegel (1986), the three two-dimensional vectors $(Z^{(1)}, Z^{(2)}, Z^{(3)})$ are pairwise independent, but not jointly independent. This vector is independent of $(Z^{(4)}, Z^{(5)})$

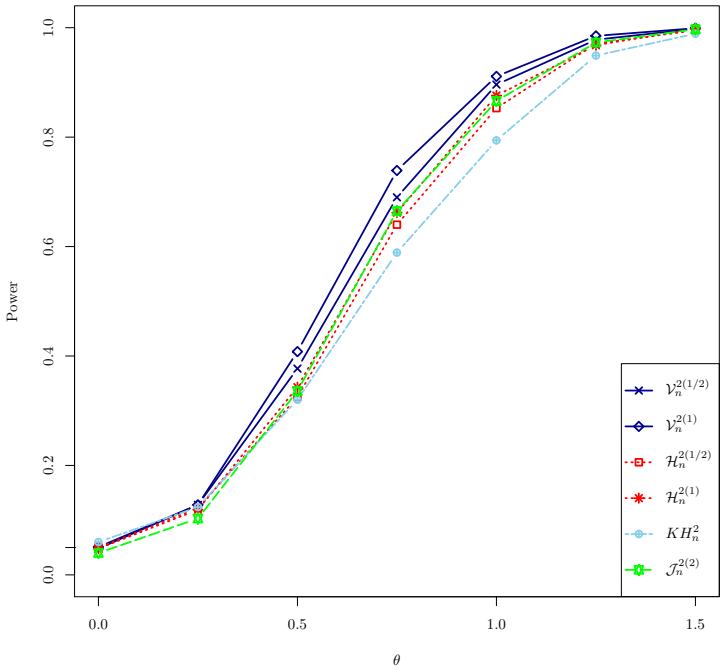

Figure 3: Empirical powers for the Frank copula with bivariate marginals of Section 10.1.3.

in which the two three-dimensional vectors $Z^{(4)}$ and $Z^{(5)}$ are dependent. One can check that the only non null terms μ_B are for the subsets $\{4,5\}$, $\{1,2,3\}$, and $\{1,2,3,4,5\}$.

In order to compare the powers of various tests, a modified model with weaker dependence is introduced. Items 3 and 6 are modified to

3'. Define $Z^{(1)} = (Z_1^{(1)}, Z_2^{(1)})$ by $Z_i^{(1)} = (1-\theta)|X_i^{(1)}| + \theta|X_i^{(1)}|\text{sign}(Z_1^{(2)}Z_1^{(3)})$, $i = 1, 2$,

6'. Define $Z^{(5)} = \theta Z^{(4)} + X^{(5)}$,

for some $\theta \in [0, 0.4]$. Mutual independence now holds among the 5 components for $\theta = 0$ and the dependence increases with θ. The value $\theta = 1$ leading to the original model is not considered since it yields a dependence too easily detected.

Figure 4 is the dependogram based on one simulated sample with $\theta = 0.4$. The statistics computed are $\mathcal{V}_{nB}^{2(1)}$ and $\mathcal{H}_{nB}^{2(1)}$ with very small constants $c_j = .0001$ in (22). It illustrates the equivalence for small scale parameters between $\mathcal{V}_{nB}^{2(\alpha)}$ and $\mathcal{H}_{nB}^{2(\alpha)}$ described precisely in (13). Both dependograms are identical apart from small variations between critical values due to the permutations generated by the randomization tests. The dependence among $Z^{(4)}$ and $Z^{(5)}$, represented by the subset $\{4,5\}$, is significant. Moreover, the third order dependence between $Z^{(1)}$, $Z^{(2)}$ and $Z^{(3)}$, represented by the subset $\{1,2,3\}$, is also significant. The dependence for the subset $\{1,2,3,4,5\}$ is not significant. This can be explained by the powers of tests for false sub-hypotheses corresponding to subsets of small cardinality which

are generally higher because the presence of noise is less important than in subsets of large cardinality. Tests based on the Möbius decomposition combine p-values according to the

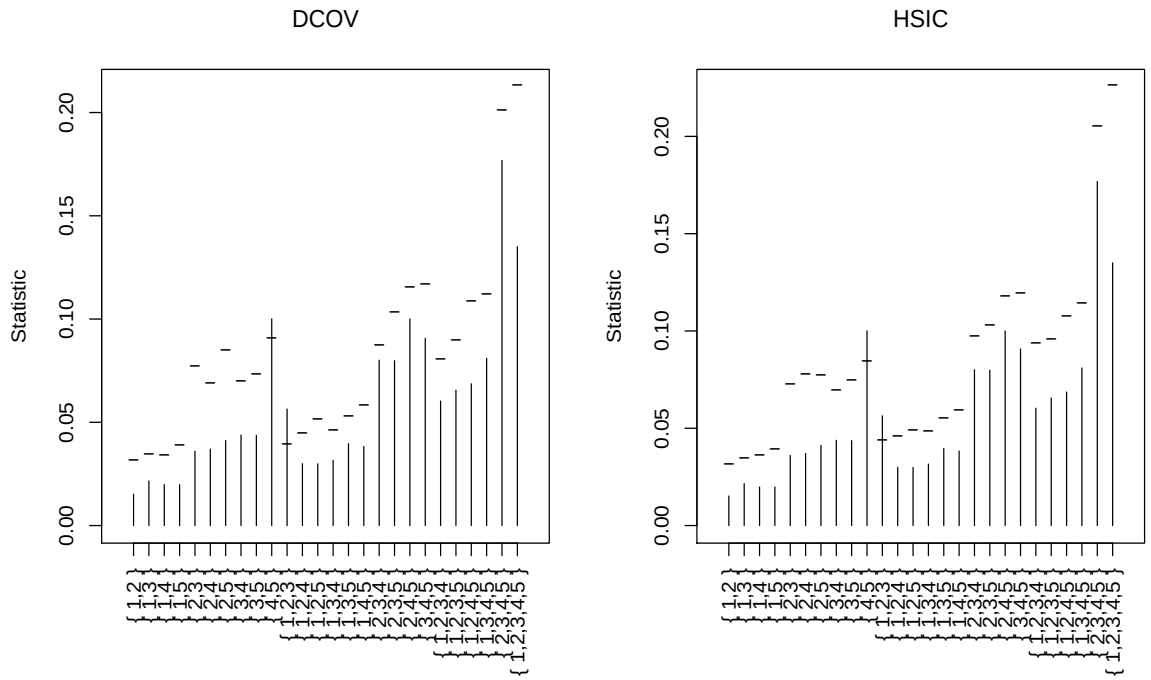

Figure 4: Dependograms for one sample of size $n = 100$ from the modified model of Romano and Siegel with $\theta = 0.4$ in Section 10.2 based on $\mathcal{V}_{nB}^{2(1)}$ (left panel) and $\mathcal{H}_{nB}^{2(1)}$ (right panel) with small constants $c_j = .0001$ in (22).

methods of Tippett and Fisher. Figure 5 shows that KH_n^2 is more powerful than distance covariance and $HSIC$ tests when using the method of Fisher. However, the method of Tippett is markedly more powerful than that of Fisher. This finding should not come as a surprise since only 3 of the 26 sub-hypotheses are false. Using the method of Tippett the most powerful test $\mathcal{V}_n^{2(1)}$ is markedly better than $\mathcal{J}_n^{2(2)}$ and KH_n^2. The popular belief that the method of Fisher is more powerful than that of Tippett should not be given too much consideration. The preferred test depends on the model under consideration.

10.3 Bivariate AR(1) Model

The model considered is the bivariate AR(1) model defined by $Y_k = AY_{k-1} + \epsilon_k$, where the innovations ϵ_k are independently distributed as bivariate Gaussian with mean vector 0 and covariance matrix C. The final specification is made by defining

$$A = \begin{pmatrix} 0 & \theta \\ \theta & 0 \end{pmatrix} \text{ and } C = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}.$$

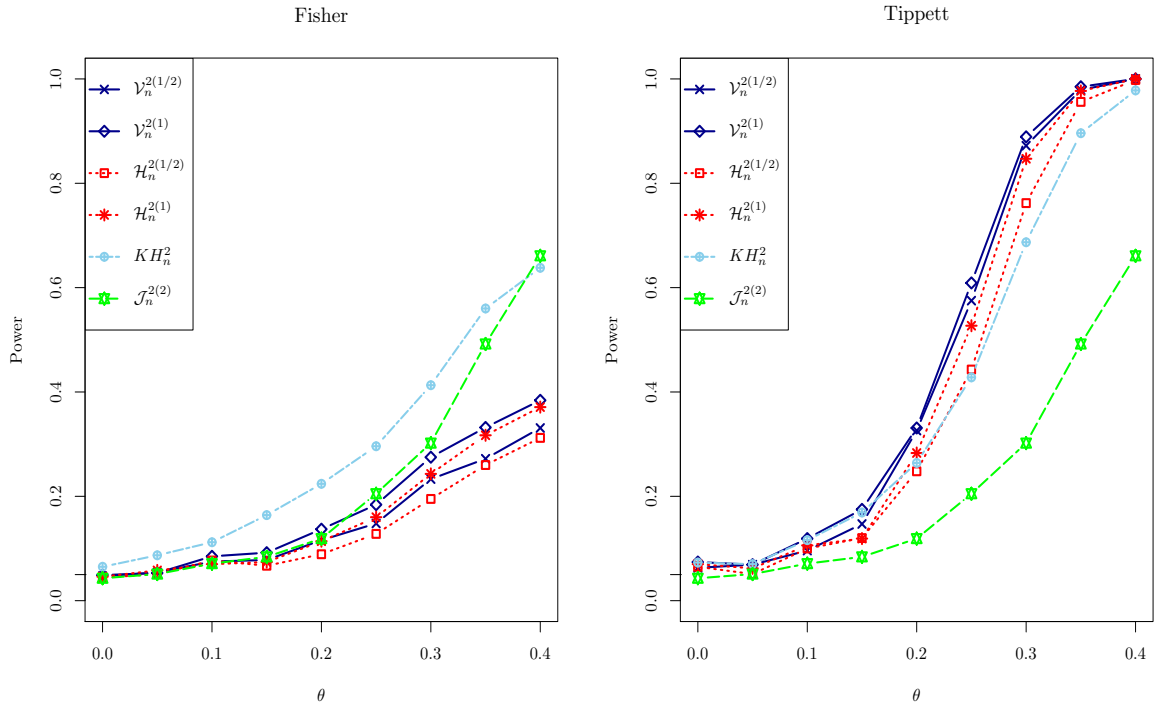

Figure 5: Empirical powers for the modified model of Romano and Siegel in Section 10.2. The methods of Fisher (left panel) and Tippett (right panel) are used to combine p-values.

The serial dependence of the sequence increases with θ and serial independence holds for $\theta = 0$. Tests of Tippett and Fisher are compared. The value $p = 3$ chosen arbitrarily can detect dependencies among three successive observations. In particular, it can detect dependencies at lags one or two. Sequences of length $m = 100$ are generated using the `mAr.sim` function of the `R` package `mAr` (Barbosa, 2012). The dependogram in Figure 6 shows a significant dependence at lag one. It shortly fails to detect the weaker dependence at lag 2 with the short sequence length of 100. Power functions in Figure 7 reveal $\mathcal{V}_n^{2(1)}$ as the most powerful test. Comparisons between Fisher and Tippett tests show comparable powers with a slight advantage for Tippett. Higher powers of $\mathcal{J}_n^{2(2)}$ locally around $\theta = 0$ can be attributed to the higher empirical significance level of this test.

11. Applications

Applications to meteorological and financial data are now provided in the following two sections.

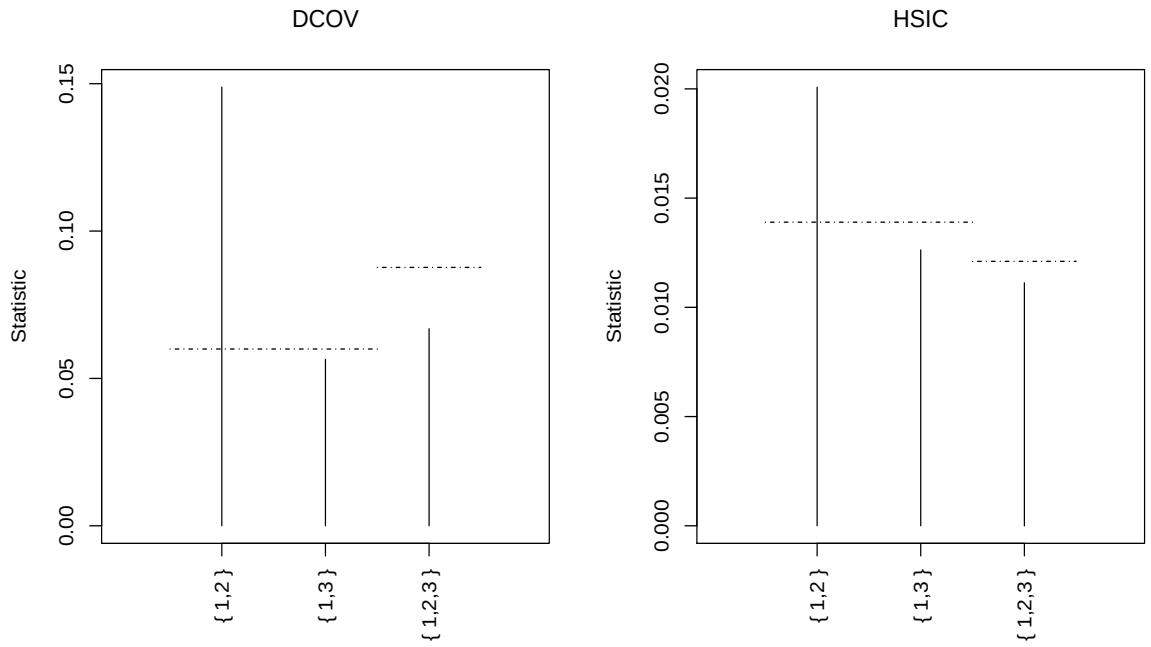

Figure 6: Dependograms for one sequence of length $m = 100$ for the AR(1) model with $\theta = 0.4$ of Section 10.3 based on $\mathcal{V}_{nB}^{2(1)}$ (left panel) and $\mathcal{H}_{nB}^{2(1)}$ (right panel).

11.1 Testing Mutual Independence Between Air Temperature, Soil Temperature, Humidity, Wind and Evaporation

These meteorological data are from Rencher (1995, p. 294). Table 2 describes the data with 46 observations on 11 variables. When a data set consists of variables measured on different scales, the scaling of variables often helps to enhance the appearance of the dependogram. In this application, variables were scaled to zero mean and unit variance. The `R` package `MVN` (Korkmaz et al., 2014) contains the function `hzTest` to perform the test for multivariate normality of Henze and Zirkler (1990). This test applied to the joint distribution of all 11 variables rejected a multivariate Gaussian model with a p-value of 0. Now, five groups of variables are considered as in Table 2. The Gaussian likelihood ratio test found a significant mutual dependence between the five groups. However, this test should not be relied on since it was found that data are not jointly Gaussian and it is well known that Gaussian likelihood ratio tests are not robust (Tyler, 1983; Bilodeau and Brenner, 1999). Mutual independence between these five groups is tested with distance covariance and $HSIC$ tests, both of index 1. Figure 8 contains dependograms only for subsets B of order up to 3. Both dependograms are very similar and lead to the same conclusions. It reveals that air temperature, soil temperature, relative humidity and evaporation are pairwise dependent.

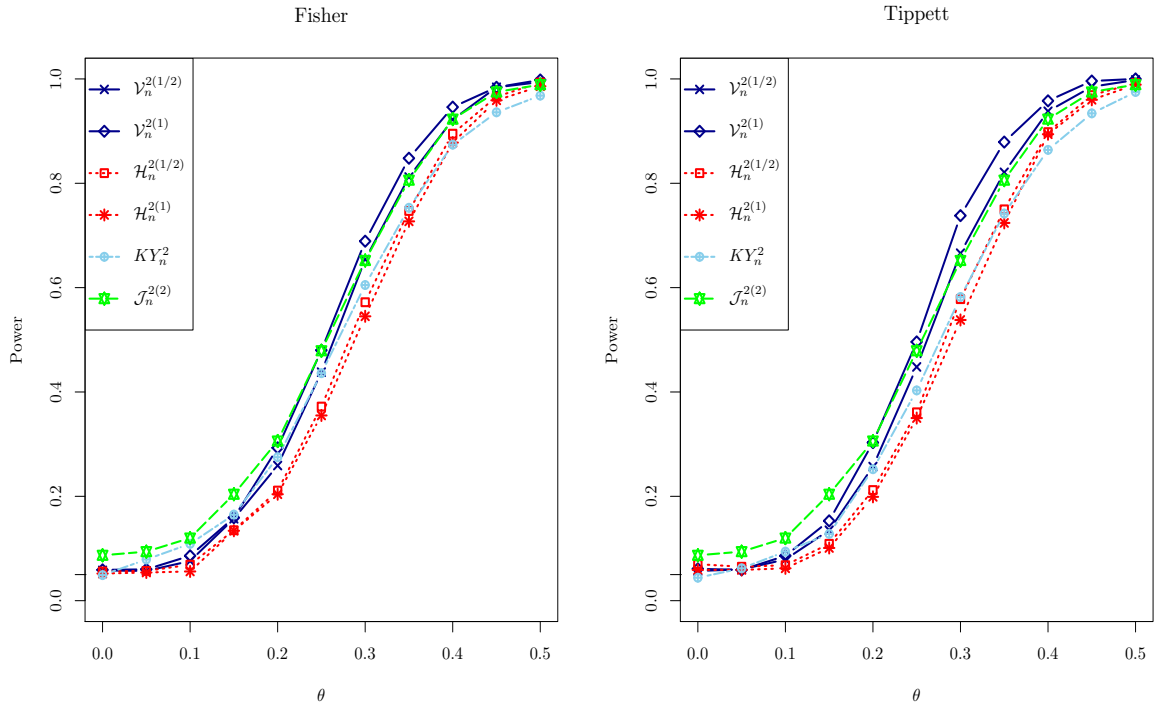

Figure 7: Empirical powers for the AR(1) model of Section 10.3. The methods of Fisher (left panel) and Tippett (right panel) are used to combine p-values.

However, wind does not exhibit any dependence of order 2 or 3 with any of the other 4 groups of variables.

11.2 Testing Serial Independence for Financial Data

Tests of serial independence of the three dimensional sequence formed by the daily percent increasing rates (DPIR) of indices from three stock markets: S&P/TSX composite (TSX), DOW JONES and S&P500. The values of indices are taken at closure. The series of length 534 range from January 2, 2014 to March 2, 2016. Note that five index values are observed weekly since the stock exchanges are not opened on weekends. The top row of Figure 9 shows that the financial series considered are not stationary. It is more appropriate to consider DPIR values. In the bottom row, one may see that DPIR values are more stationary, although still not perfectly stationary.

Tests of serial independence KY_{nB}^2, $\mathcal{V}_{nB}^{2(1)}$, and the test of non serial correlation $\mathcal{V}_{nB}^{2(2)}$ are conducted on the 3 joint series. The value of $p = 10$ allows a maximum lag of 9 days. Figure 10 reveals dependencies at small lags of 1, 2, and 4 in the dependogram of $\mathcal{V}_{nB}^{2(1)}$. The dependogram of KY_{nB}^2 was produced with the `copula` package and does not include critical values. Nevertheless, KY_{nB}^2 and $\mathcal{V}_{nB}^{2(1)}$ agree on the strongest dependency observed

Variables	Labels
$Z^{(1)}$	maximum daily air temperature minimum daily air temperature integrated area under daily air temperature curve
$Z^{(2)}$	maximum daily soil temperature minimum daily soil temperature integrated area under soil temperature
$Z^{(3)}$	maximum daily relative humidity minimum daily relative humidity integrated area under daily humidity curve
$Z^{(4)}$	total wind, measured in miles per day
$Z^{(5)}$	evaporation

Table 2: Variables related to air temperature, soil temperature, humidity, wind and evaporation.

at lag 4. The distance covariance $\mathcal{V}_{nB}^{2(2)}$ of index 2 was also performed on the sequence. One should recall that $\mathcal{V}_{nB}^{2(2)}$ is no longer a test of serial independence, but merely of non serial correlation. Interestingly, this test did not reveal any significant serial correlation. In an attempt to unravel the most significant dependency at lag 4 declared by $\mathcal{V}_{nB}^{2(1)}$, Figure 11 is a scatterplot of DPIR values observed on day k and $k+4$ for the TSX market. The Pearson (p-value of 0.11) and the Kendall (p-value of 0.14) correlation tests applied to this scatterplot are not significant. The test $\mathcal{V}_{nB}^{2(1)}$ for $B = \{1, 5\}$ on the single TSX market is very significant (p-value less than 0.001). A broken line regression with a change point at the origin was fitted to this scatterplot to account for different regimes according to whether DPIR is negative or positive. This regression model has three parameters for the intercept and slopes at the left and right of the origin. The left slope is very significant (p-value of 0.000015) contrary to the right slope (p-value of 0.021). The very significant left slope could be interpreted by the tendency of the TSX market to recover in the days following a decline. The sharper the market declines, the stronger it recovers. Among days such that $\text{DPIR}_k < \xi$, the percentage of days with $\text{DPIR}_{k+4} > 0$ is 60.7% for $\xi = -0.5$. This percentage goes up to 63.4% for $\xi = -1$ and to 68.2% for $\xi = -1.5$. Similar conclusions were found for the DOW JONES and S&P500 stock markets.

12. Conclusion

Generalizations of distance covariance and $HSIC$ tests were done successfully. For both mutual and serial independence hypotheses, the dependence statistics related to distance covariance and $HSIC$ were defined using the Möbius transformation. Simple and explicit expressions for dependence statistics were derived in the explicit form (8) as a sum over all elements of a componentwise product of doubly-centered matrices $A^{(j)} = (A_{kl}^{(j)})$. Computationally efficient approximation of p-values by randomization tests is made possible by this

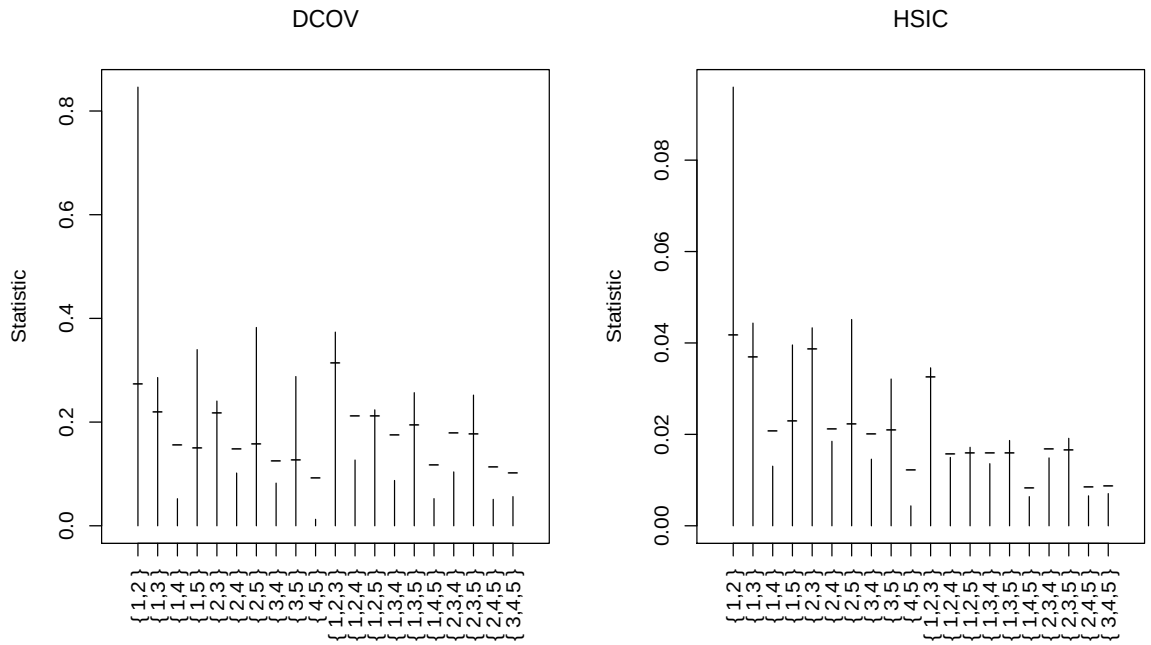

Figure 8: Dependograms of order $q = 3$ of air temperature, soil temperature, relative humidity, wind and evaporation for the meteorological data in Section 11.1 based on $\mathcal{V}_{nB}^{2(1)}$ (left panel) and $\mathcal{H}_{nB}^{2(1)}$ (right panel).

explicit form. Indeed, distances in $A^{(j)}$ do not have to be recomputed since it suffices to permute rows and corresponding columns of $A^{(j)}$ for every randomized sample. The method of combining individual p-values was put forward to construct global tests whose p-values evaluated by randomization tests yielded global significance levels close to the nominal level of 0.05 in all simulated models considered.

Distance covariance tests yielded powers generally very competitive with other tests considered. This paper has presented some advances to the problem of testing independence but some questions remain unanswered. The index α of distance covariance tests has a major influence on power functions. The adaptive selection of this index is a major difficulty which should be the object of future investigations. At the same time, it offers more flexibility and possibilities than tests based on copulas for which the only integrating measure is uniform. $HSIC$ tests can always achieve the same power function as distance covariance tests with the same index simply by selecting very small scale parameters. The additional adaptive selection of scale parameters could possibly yield tests which in certain cases would be more powerful than distance covariance tests. For $p = 2$, Guetsop Nangue (2016) selected the scale parameters which maximize the variance of the permutation distribution. This yielded $HSIC$ tests more powerful than distance covariance tests in two

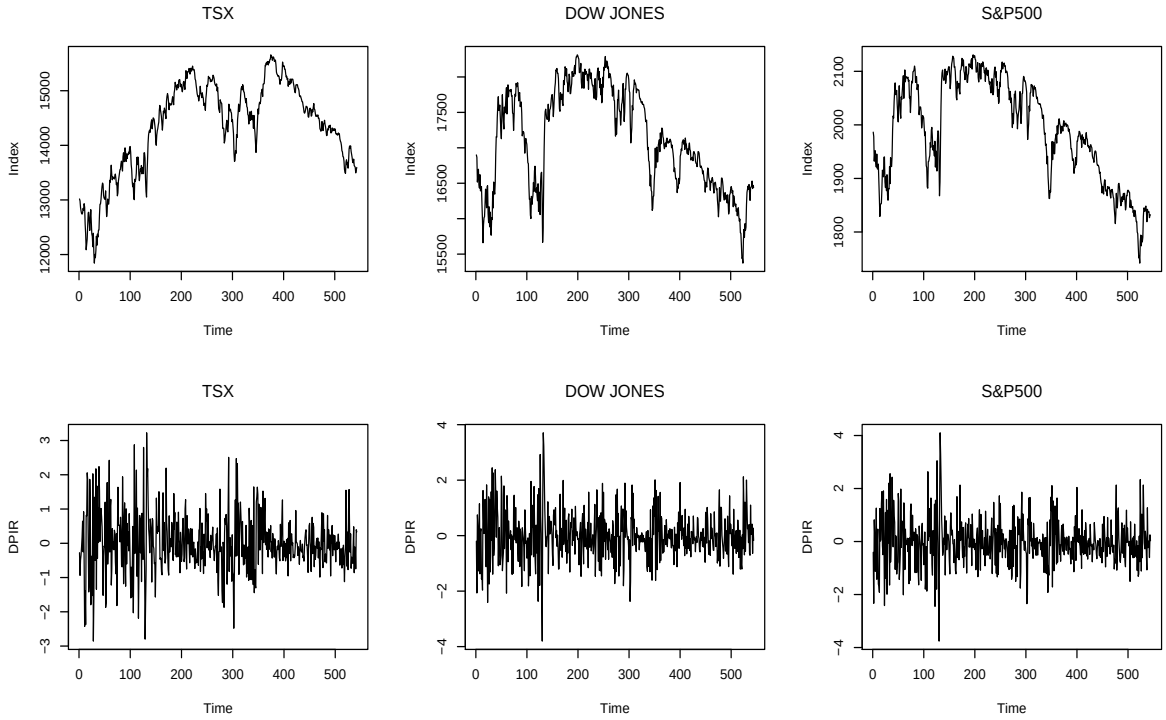

Figure 9: Daily evolution of the S&P/TSX composite (TSX), DOW JONES and S&P500 stock markets. The period of observation ranges from January 2, 2014 to March 2, 2016.

examples form the machine learning community considered in Sejdinovic et al. (2013b): the Independent Component Analysis (ICA) benchmark densities of Bach and Jordan (2002) and the sinusoidally dependent data. The adaptive selection of scale parameters for $p > 2$ is more challenging and is worthy of future research.

Acknowledgments

The authors would like to thank Arthur Gretton and three referees for their careful reading and constructive comments and suggestions which enhanced the quality of the paper. Financial support by the Faculté des arts et des sciences of Université de Montréal during the graduate thesis of the second author was greatly appreciated. This research was done without the financial support from the Natural Sciences and Engineering Research Council of Canada.

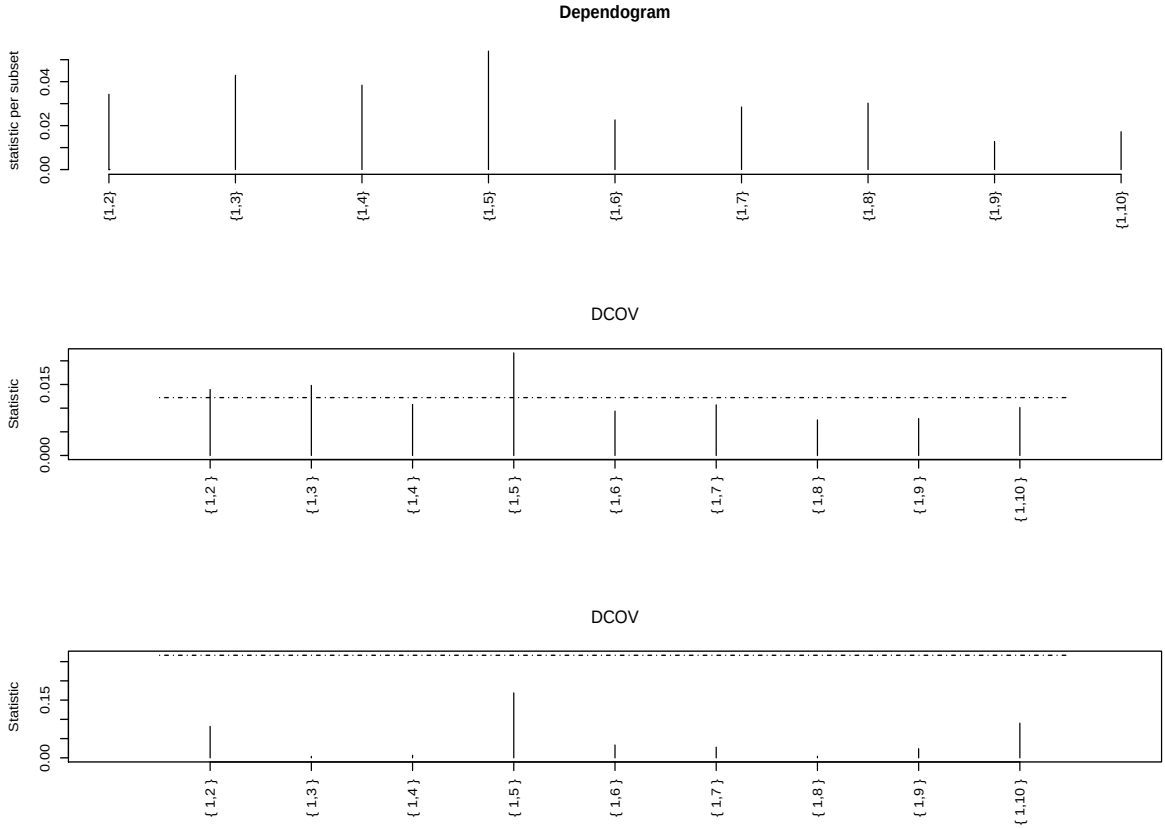

Figure 10: Dependograms for joint DPIR of S&P/TSX composite, DOW JONES, and S&P500. For $p = 10$, the maximum lag is 9. From top to bottom, three tests are considered: KY^2_{nB}, $\mathcal{V}^{2(1)}_{nB}$, and $\mathcal{V}^{2(2)}_{nB}$.

Appendix A: Proofs

Proof [Theorem 1] Upon using the representation (6) of the process,

$$\begin{aligned}\int |R_{nB}(t^{(B)})|^2 \prod_{j\in B} dG^{(j)}(t^{(j)}) &= \frac{1}{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\prod_{j\in B}\int \left[e^{i\langle t^{(j)}, Z_k^{(j)} - Z_l^{(j)}\rangle} - \frac{1}{n}\sum_{v=1}^{n} e^{i\langle t^{(j)}, Z_k^{(j)} - Z_v^{(j)}\rangle}\right. \quad (24)\\ &- \left.\frac{1}{n}\sum_{u=1}^{n} e^{i\langle t^{(j)}, Z_u^{(j)} - Z_l^{(j)}\rangle} + \frac{1}{n^2}\sum_{u=1}^{n}\sum_{v=1}^{n} e^{i\langle t^{(j)}, Z_u^{(j)} - Z_v^{(j)}\rangle}\right] dG^{(j)}(t^{(j)})\\ &= \frac{1}{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\prod_{j\in B}[a_{kl}^{(j)} - \bar{a}_{k.}^{(j)} - \bar{a}_{.l}^{(j)} + \bar{a}_{..}^{(j)}],\end{aligned}$$

where $a_{kl}^{(j)} = \varphi^{(j)}(Z_k^{(j)} - Z_l^{(j)})$ and $\varphi^{(j)}$ is the characteristic function of $G^{(j)}$. ■

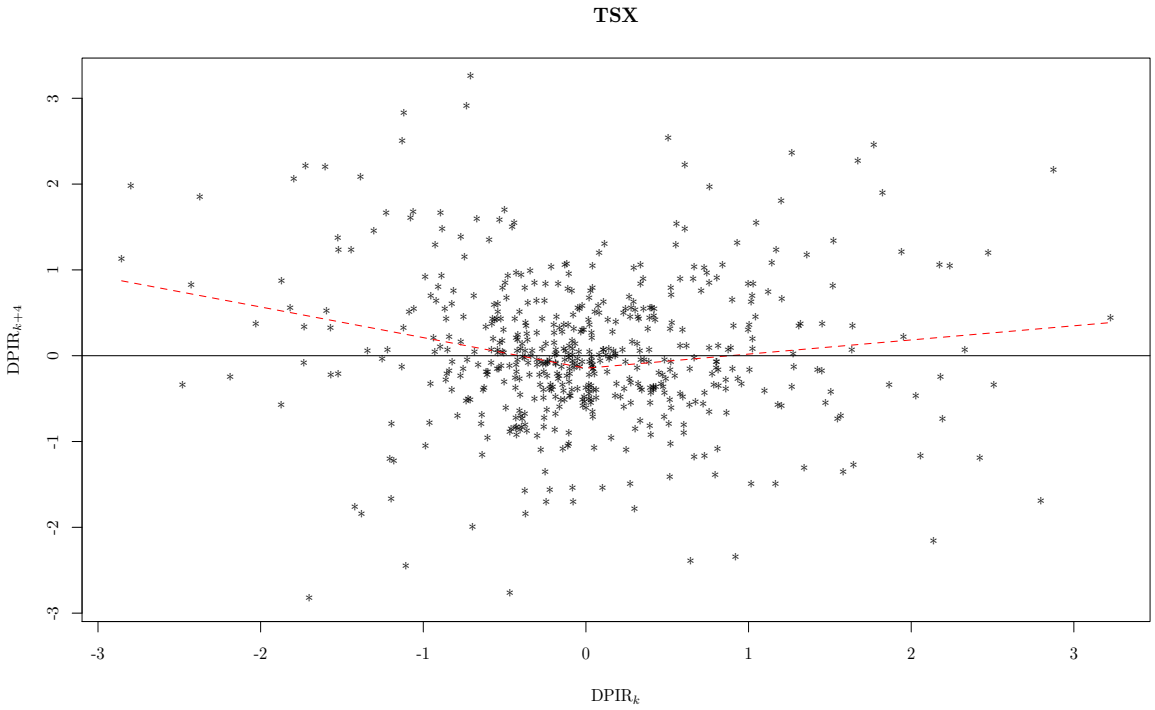

Figure 11: For the TSX market, broken line regression of DPIR on a given day on DPIR four days earlier with a change point at the origin.

Proof [Theorem 2] From (2) and the strong law of large numbers, $R_{nB}(t^{(B)})/\sqrt{n} \overset{a.s.}{\to} \mu_B(t^{(B)})$. Since the number of subsets of B is $2^{|B|}$, it is also clear that $|R_{nB}(t^{(B)})/\sqrt{n}| \leq 2^{|B|}$. Any constant being integrable with respect to the probability measure $\prod_{j\in B} dG^{(j)}(t^{(j)})$, it follows from the dominated convergence theorem that

$$\frac{1}{n}\int |R_{nB}(t^{(B)})|^2 \prod_{j\in B} dG^{(j)}(t^{(j)}) \overset{a.s.}{\to} \int |\mu_B(t^{(B)})|^2 \prod_{j\in B} dG^{(j)}(t^{(j)}),$$

i.e. $\mathcal{H}^2_{nB} \overset{a.s.}{\to} \mathcal{H}^2_B$. This proves (i). To prove (ii), since $\mu_B(t^{(B)}) \neq 0$ and the function μ_B is continuous, it follows that $\mathcal{H}^2_B > 0$. Therefore, $n\mathcal{H}^2_{nB} \overset{a.s.}{\to} \infty$. ■

Proof [Theorem 3] Because of double-centering, the expression between brackets in (24) is unchanged if one is subtracted from all four exponential functions. Then, it suffices in the proof of Theorem 1, for the weight function defining $\mathcal{V}^{2(\alpha)}_{nB}$, to evaluate

$$\int e^{i\langle t^{(j)}, Z^{(j)}\rangle} - 1 dw^{(j)}(t^{(j)}) = -|Z^{(j)}|^{\alpha}_{d_j} \tag{25}$$

using Lemma 1 of Székely et al. (2007, p. 2771). ■

The representation (6) of the process was unexploited in Székely et al. (2007). It simplifies greatly their derivations.

Proof [Theorem 4] The notation $\mathbb{E}_3$ is for the expectation with respect to Z_3, treating the other variable as a constant to avoid using for conditional expectations. The notation $\mathbb{E}_{12}$ and $\mathbb{E}_{34}$ are defined similarly. Condition (12) implies that

$$\mathbb{E}_{12}\prod_{j\in B}\left||Z_1^{(j)}-Z_2^{(j)}|_{d_j}^{\alpha}-\mathbb{E}_3|Z_1^{(j)}-Z_3^{(j)}|_{d_j}^{\alpha}-\mathbb{E}_3|Z_2^{(j)}-Z_3^{(j)}|_{d_j}^{\alpha}+\mathbb{E}_{34}|Z_3^{(j)}-Z_4^{(j)}|_{d_j}^{\alpha}\right|<\infty.$$

and that $\mathcal{V}_B^{2(\alpha)}$ is well defined. The integral (25) yields

$$\begin{aligned}
\mathcal{V}_B^{2(\alpha)} &= \mathbb{E}_{12}\prod_{j\in B}\left[\int e^{i\langle t^{(j)},Z_1^{(j)}-Z_2^{(j)}\rangle}-1dw^{(j)}(t^{(j)})+\mathbb{E}_3\int 1-e^{i\langle t^{(j)},Z_1^{(j)}-Z_3^{(j)}\rangle}dw^{(j)}(t^{(j)})\right.\\
&\quad\left.+\mathbb{E}_3\int 1-e^{i\langle t^{(j)},Z_3^{(j)}-Z_2^{(j)}\rangle}dw^{(j)}(t^{(j)})+\mathbb{E}_{34}\int e^{i\langle t^{(j)},Z_3^{(j)}-Z_4^{(j)}\rangle}-1dw^{(j)}(t^{(j)})\right]\\
&= \mathbb{E}_{12}\prod_{j\in B}\mathbb{E}_{34}\int\left[e^{i\langle t^{(j)},Z_1^{(j)}-Z_2^{(j)}\rangle}-e^{i\langle t^{(j)},Z_1^{(j)}-Z_3^{(j)}\rangle}\right.\\
&\quad\left.-e^{i\langle t^{(j)},Z_3^{(j)}-Z_2^{(j)}\rangle}+e^{i\langle t^{(j)},Z_3^{(j)}-Z_4^{(j)}\rangle}\right]dw^{(j)}(t^{(j)}).
\end{aligned}$$

The theorem of Fubini yields

$$\begin{aligned}
\mathcal{V}_B^{2(\alpha)} &= \mathbb{E}_{12}\prod_{j\in B}\int\mathbb{E}_{34}\left[e^{i\langle t^{(j)},Z_1^{(j)}-Z_2^{(j)}\rangle}-e^{i\langle t^{(j)},Z_1^{(j)}-Z_3^{(j)}\rangle}\right.\\
&\quad\left.-e^{i\langle t^{(j)},Z_3^{(j)}-Z_2^{(j)}\rangle}+e^{i\langle t^{(j)},Z_3^{(j)}-Z_4^{(j)}\rangle}\right]dw^{(j)}(t^{(j)})\\
&= \mathbb{E}_{12}\prod_{j\in B}\int\left[e^{i\langle t^{(j)},Z_1^{(j)}-Z_2^{(j)}\rangle}-e^{i\langle t^{(j)},Z_1^{(j)}\rangle}f^{(j)}(-t^{(j)})\right.\\
&\quad\left.-e^{-i\langle t^{(j)},Z_2^{(j)}\rangle}f^{(j)}(t^{(j)})+f^{(j)}(t^{(j)})f^{(j)}(-t^{(j)})\right]dw^{(j)}(t^{(j)})\\
&= \mathbb{E}_{12}\int\prod_{j\in B}\left[e^{i\langle t^{(j)},Z_1^{(j)}\rangle}-f^{(j)}(t^{(j)})\right]\left[e^{-i\langle t^{(j)},Z_2^{(j)}\rangle}-f^{(j)}(-t^{(j)})\right]dw_B(t^{(B)})\\
&= \int\mathbb{E}_1\prod_{j\in B}\left[e^{i\langle t^{(j)},Z_1^{(j)}\rangle}-f^{(j)}(t^{(j)})\right]\mathbb{E}_2\prod_{j\in B}\left[e^{-i\langle t^{(j)},Z_2^{(j)}\rangle}-f^{(j)}(-t^{(j)})\right]dw_B(t^{(B)}).
\end{aligned}$$

By a similar development leading to (6),

$$\mu_B(t^{(B)})=\mathbb{E}_1\prod_{j\in B}\left[e^{i\langle t^{(j)},Z_1^{(j)}\rangle}-f^{(j)}(t^{(j)})\right].$$

Hence, $\mathcal{V}_B^{2(\alpha)}=\int|\mu_B(t^{(B)})|^2dw_B(t^{(B)})$. This proves (i). For (ii), if $\mu_B(t^{(B)})\neq 0$, then necessarily $t^{(B)}\neq 0$. Consider a compact ball $\mathfrak{B}$ not containing 0 and but containing the point $t^{(B)}$. The measure dw_B is integrable on $\mathfrak{B}$. Then, arguing as in the proof of Theorem 2

$$\mathcal{V}_{nB}^{2(\alpha)}\geq\frac{1}{n}\int_{\mathfrak{B}}|R_n(t^{(B)})|^2dw_B(t^{(B)})\overset{a.s.}{\to}\int_{\mathfrak{B}}|\mu_B(t^{(B)})|^2dw_B(t^{(B)})>0$$

and therefore, $n\mathcal{V}_{nB}^{2(\alpha)} \overset{a.s.}{\to} \infty$. ■

Proof [Equation (13)] The result follows using the invariance by translation, $a_{kl}^{(j)} \mapsto a_{kl}^{(j)} - 1$, and the following limit,

$$\lim_{\beta_j \to 0} \frac{e^{-\beta_j^\alpha |Z_k^{(j)} - Z_l^{(j)}|_{d_j}^\alpha} - 1}{\beta_j^\alpha} = -|Z_k^{(j)} - Z_l^{(j)}|_{d_j}^\alpha.$$

■

Let $d_B = \sum_{j \in B} d_j$. Define the metric (Whitt, 1970)

$$\rho_B(x, y) = \sum_{s=1}^{\infty} 2^{-s} \frac{\rho_{sB}(x, y)}{1 + \rho_{sB}(x, y)},$$

where

$$\rho_{sB}(x, y) = \sup_{|t^{(B)}|_{d_B} \leq s} |x(t^{(B)}) - y(t^{(B)})|,$$

on the linear complete metric space of continuous functions $\mathcal{C}(\mathbb{R}^{d_B}, \mathbb{C})$. The Borel σ-field in $\mathcal{C}(\mathbb{R}^{d_B}, \mathbb{C})$ is generated by the coordinate projections, *i.e.* it is the smallest σ-field with respect to which all coordinate projections are measurable. Weak convergence of random variables in $\mathcal{C}(\mathbb{R}^{d_B}, \mathbb{C})$ is equivalent to weak convergence on any compact subset; see Whitt (1970, Theorem 5) or Kallenberg (2002, Proposition 16.6). Moreover, weak converge of a sequence on a compact subset is equivalent to finite dimensional weak convergence and tightness of that sequence. The metric defined on the product of spaces $\underset{B \in \mathcal{I}_p}{\times} C(\mathbb{R}^{d_B}, \mathbb{C})$ (Whitt, 1970) is

$$\rho\left((x_B,\ B \in \mathcal{I}_p), (y_B,\ B \in \mathcal{I}_p)\right) = \max_{B \in \mathcal{I}_p} \rho_B(x_B, y_B).$$

From Whitt (1970, Corollary 7), weak convergence on this product of spaces is equivalent to finite dimensional weak convergence of the joint process and tightness on compacta of each individual process.

Proof [Theorem 5] The process $R_{nB}(t)$ in (6) is closely related to the process

$$\breve{R}_{nB}(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \prod_{j \in B} \left[e^{i\langle t^{(j)}, Z_k^{(j)} \rangle} - f^{(j)}(t^{(j)}) \right]. \tag{26}$$

in which marginal characteristic functions are not estimated. The process $\breve{R}_{nB}$ in (26) is a sum of independent and identically distributed random variables. Bilodeau and Lafaye de Micheaux (2005, Theorem 2.1) proved that the collection of processes $\breve{R}_{nB}$ converges as stated in Theorem 5 under the weak condition (14). The independence of the asymptotic processes for $B \neq C$ is verified

$$\mathbb{E}\left[R_B(t^{(B)}) \bar{R}_C(s^{(C)}) \right] \quad = \quad \mathbb{E}\left\{ \prod_{j \in B} \left[e^{i\langle t^{(j)}, Z_k^{(j)} \rangle} - f^{(j)}(t^{(j)}) \right] \prod_{j \in C} \left[e^{i\langle s^{(j)}, -Z_k^{(j)} \rangle} - f^{(j)}(-s^{(j)}) \right] \right\}$$

$$= \quad 0,$$

because there is an index j in B, but not in C, or the converse, for which the corresponding term has expectation zero. Then, it suffices to show $\rho_{sB}(R_{nB}, \breve{R}_{nB}) \xrightarrow{P} 0$, for all $s \geq 1$ and $B \in \mathcal{I}_p$. The representation in Ghoudi et al. (2001, p. 212) holds for characteristic functions

$$R_{nB}(t) = \sum_{C \subseteq B} (-1)^{|C|} \prod_{j \in C} [f_n^{(j)}(t^{(j)}) - f^{(j)}(t^{(j)})] \breve{R}_{n,B\setminus C}(t^{(B\setminus C)}). \tag{27}$$

From (27), it follows that

$$|R_{nB}(t^{(B)}) - \breve{R}_{nB}(t^{(B)})| \leq \sum_{C \subseteq B, C \neq \emptyset} \prod_{j \in C} |f_n^{(j)}(t^{(j)}) - f^{(j)}(t^{(j)})| |\breve{R}_{n,B\setminus C}(t^{(B\setminus C)})|, \tag{28}$$

where the sum has only a finite number of terms. Using the Glivenko-Cantelli convergence in Csörgő (1981, Theorem 2.1) and the fact that the processes $\breve{R}_{n,B\setminus C}$ are tight, it follows that, for any $s \geq 1$, $\rho_s(R_{nB}, \breve{R}_{nB}) \xrightarrow{P} 0$. ■

Proof [Theorem 6] Let

$$|[R_{nB}(t^{(B)})]|^2 = \int |R_{nB}(t^{(B)})|^2 \prod_{j \in B} dG^{(j)}(t^{(j)})$$

be the squared L_2 norm, on the space of squared integrable functions, of the process R_{nB}. Then, $nW_{nB} = |[R_{nB}]|^2$ and $n\breve{T}_{nB} = |[\breve{R}_{nB}]|^2$. Use $\int \tilde{k}(t^{(B)}, t^{(B)}) \prod_{j \in B} dG^{(j)}(t^{(j)}) < \infty$, where $\tilde{k}$ is defined in (16), and Tonelli's theorem to conclude that $W_B := |[R_B]|^2$ is finite almost surely. The proof consists in showing the following two results (Henze and Wagner, 1997): (i) $|[\breve{R}_{nB}]|^2 \Rightarrow |[R_B]|^2$ and (ii) $|[R_{nB} - \breve{R}_{nB}]|^2 \xrightarrow{P} 0$. Note that from (i) and the continuous mapping theorem, $|[\breve{R}_{nB}]| \Rightarrow |[R_B]|$, which, with the triangle inequality $\Big|\, |[R_{nB}]| - |[\breve{R}_{nB}]| \,\Big| \leq |[R_{nB} - \breve{R}_{nB}]|$ and (ii), implies $|[R_{nB}]| \Rightarrow |[R_B]|$ and, thus $|[R_{nB}]|^2 \Rightarrow |[R_B]|^2$. Using a slight generalization of Kellermeier (1980, Theorem 3.3), it suffices for (i) to show the following uniform integrability condition:

$$\lim_{N \to \infty} \limsup_{n \to \infty} \int_{\{|t^{(j)}|_{d_j} > N, \forall j \in B\}} \mathbb{E}|\breve{R}_{nB}(t^{(B)})|^2 \prod_{j \in B} dG^{(j)}(t^{(j)}) = 0. \tag{29}$$

It can be verified that $\mathbb{E}|\breve{R}_{nB}(t^{(B)})|^2 = \prod_{j \in B} \left[1 - |f^{(j)}(t^{(j)})|^2\right]$ does not depend on n. Now, for each $j \in B$,

$$\int 1 - |f^{(j)}(t^{(j)})|^2 dG^{(j)}(t^{(j)}) < \infty$$

since the integrand is bounded. This proves (29). For (ii), since all processes $\breve{R}_{n,B\setminus C}$ are tight, it suffices from (28) to show that for all $j \in B$,

$$\int |f_n^{(j)}(t^{(j)}) - f^{(j)}(t^{(j)})|^2 dG^{(j)}(t^{(j)}) \xrightarrow{P} 0.$$

This follows from the dominated convergence theorem, since $f_n^{(j)}(t^{(j)}) \stackrel{a.s.}{\to} f^{(j)}(t^{(j)})$ and the integrand is bounded by 4. The weak convergence of the collection $(|[R_{nB}]|^2,\ B \in \mathcal{I}_p)$ follows by the same argument above from (i') $(|[\breve{R}_{nB}]|^2,\ B \in \mathcal{I}_p) \Rightarrow (|[R_B]|^2,\ B \in \mathcal{I}_p)$ and (ii) $|[R_{nB} - \breve{R}_{nB}]|^2 \stackrel{P}{\to} 0$, for each $B \in \mathcal{I}_p$. Thus, it remains to establish (i'). From Theorem 5 and the continuous mapping theorem, $\sum_{B\in\mathcal{I}_p} c_B|\breve{R}_{nB}(t^{(B)})|^2 \Rightarrow \sum_{B\in\mathcal{I}_p} c_B|R_B(t^{(B)})|^2$, for all constants c_B, $B \in \mathcal{I}_p$. Also, $\sum_{B\in\mathcal{I}_p} c_B|[\breve{R}_{nB}]|^2 \Rightarrow \sum_{B\in\mathcal{I}_p} c_B|[R_B]|^2$ because the following uniform integrability condition

$$\begin{aligned}
&\lim_{N\to\infty} \limsup_{n\to\infty} \int_{\{|t^{(j)}|_{d_j}>N, j=1,\ldots,p\}} \mathbb{E}\left|\sum_{B\in\mathcal{I}_p} c_B|\breve{R}_{nB}(t^{(B)})|^2\right| \prod_{j=1}^{p} dG^{(j)}(t^{(j)}) \\
\leq\ & \sum_{B\in\mathcal{I}_p} |c_B| \lim_{N\to\infty} \limsup_{n\to\infty} \int_{\{|t^{(j)}|_{d_j}>N, \forall j\in B\}} \mathbb{E}|\breve{R}_{nB}(t^{(B)})|^2 \prod_{j\in B} dG^{(j)}(t^{(j)}) = 0
\end{aligned}$$

is satisfied from (29). Finally, (i') follows from the Cramér-Wold theorem. The mutual independence of $(|[R_B]|^2,\ B \in \mathcal{I}_p)$ follows from that of $(R_B,\ B \in \mathcal{I}_p)$ in Theorem 5. ■

Proof [Theorem 7] Consider the processes

$$\breve{R}_{nB,s}(t^{(B)}) = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \prod_{j\in B} \left[e^{i\langle t^{(j)}, Z_k^{(j)}\rangle} - f^{(1)}(t^{(j)})\right],\ B \in \mathcal{B}_p.$$

Finite dimensional weak convergence of the processes is proved. Because of overlapping of Y's in consecutive Z_k's, the Z_k's form an $(p-1)$-dependent sequence, see Ferguson (1996, p. 69). Thus, the central limit theorem for such dependent sequences establishes that $\breve{R}_{nB,s}(t^{(B)})$ and $\breve{R}_{nC,s}(s^{(C)})$ are asymptotically and jointly normal with asymptotic covariance $\sigma_{0,0} + 2\sigma_{0,1} + \cdots + 2\sigma_{0,p-1}$, where

$$\sigma_{0,u} = \mathbb{E}\left\{\prod_{j\in B}\left[e^{i\langle t^{(j)}, Z_k^{(j)}\rangle} - f^{(1)}(t^{(j)})\right] \prod_{j\in C}\left[e^{i\langle s^{(j)}, -Z_{k+u}^{(j)}\rangle} - f^{(1)}(-s^{(j)})\right]\right\}.$$

All of the above expectations are null unless $B = C$ (both in $\mathcal{B}_p$) and $u = 0$. Next, to establish weak convergence of the process on any compact, assume without loss of generality that n is a multiple of p, say $n = rp$. This amounts to neglecting at most $p-1$ terms in the sequence. Rewrite the sequence $Z_1, Z_2, \ldots$ as an array with p rows, each consisting of r independent and identically distributed vectors,

$$\begin{array}{cccc}
Z_1 & Z_{1+p} & \cdots & Z_{1+(r-1)p} \\
Z_2 & Z_{2+p} & \cdots & Z_{2+(r-1)p} \\
\vdots & \vdots & \ddots & \vdots \\
Z_p & Z_{p+p} & \cdots & Z_{p+(r-1)p}.
\end{array}$$

Then, the expression

$$\breve{R}_{nB,s}(t^{(B)}) = \frac{1}{\sqrt{p}} \sum_{h=1}^{p} \sum_{C\subseteq B} (-1)^{|B\setminus C|} \prod_{j\in B\setminus C} f^{(1)}(t^{(j)}) \cdot \frac{1}{\sqrt{r}} \sum_{i=0}^{r-1} \left[e^{i\langle t^{(C)}, Z_{pi+h}^{(C)}\rangle} - \prod_{j\in C} f^{(1)}(t^{(j)})\right]$$

establishes weak convergence since for each pair (h, C) in finite number, the last sum over i is an empirical characteristic function process over a compact. Finally, $R_{nB,s}$ and $\breve{R}_{nB,s}$ are equivalent processes follows from the inequality

$$|R_{nB,s}(t^{(B)}) - \breve{R}_{nB,s}(t^{(B)})| \leq \sum_{C \subseteq B, C \neq \emptyset} \prod_{j \in C} |f_n^{(j)}(t^{(j)}) - f^{(1)}(t^{(j)})| |\breve{R}_{nB \setminus C,s}(t^{(B \setminus C)})|,$$

and the same arguments following (28). ■

References

F. R. Bach and M. I. Jordan. Kernel independent component analysis. *Journal of Machine Learning Research*, 3:1–48, 2002.

S. M. Barbosa. *mAr: Multivariate AutoRegressive analysis*, 2012. URL `https://CRAN.R-project.org/package=mAr`. R package version 1.1-2.

R. Beran, M. Bilodeau, and P. Lafaye de Micheaux. Nonparametric tests of independence between random vectors. *Journal of Multivariate Analysis*, 98(9):1805–1824, 2007.

K. J. Berry, P. W. Mielke, Jr., and J. E. Johnston. *Permutation Statistical Methods*. Springer International Publishing, 2016.

M. Bilodeau and D. Brenner. *Theory of Multivariate Statistics*. Springer Texts in Statistics. Springer-Verlag, New York, 1999.

M. Bilodeau and P. Lafaye de Micheaux. A multivariate empirical characteristic function test of independence with normal marginals. *Journal of Multivariate Analysis*, 95(2): 345–369, 2005.

M. Bilodeau and P. Lafaye de Micheaux. A-dependence statistics for mutual and serial independence of categorical variables. *Journal of Statistical Planning and Inference*, 139 (7):2407–2419, 2009.

J. R. Blum, J. Kiefer, and M. Rosenblatt. Distribution free tests of independence based on the sample distribution function. *Annals of Mathematical Statistics*, 32(2):485–498, 1961.

D. S. Cotterill and M. Csörgő. On the limiting distribution of and critical values for the multivariate Cramér-von Mises statistic. *The Annals of Statistics*, 10(1):233–244, 1982.

D. S. Cotterill and M. Csörgő. On the limiting distribution of and critical values for the Hoeffding, Blum, Kiefer, Rosenblatt independence criterion. *Statistics & Decisions. International Journal for Statistical Theory and Related Fields*, 3(1-2):1–48, 1985.

S. Csörgő. Multivariate empirical characteristic functions. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 55(2):203–229, 1981.

S. Csörgő. Testing for independence by the empirical characteristic function. *Journal of Multivariate Analysis*, 16(3):290–299, 1985.

P. Deheuvels. An asymptotic decomposition for multivariate distribution-free tests of independence. *Journal of Multivariate Analysis*, 11(1):102–113, 1981.

C. Diks and V. Panchenko. Nonparametric tests for serial independence based on quadratic forms. *Statistica Sinica*, 17:81–98, 2007.

P. Duchesne and P. Lafaye De Micheaux. Computing the distribution of quadratic forms: Further comparisons between the Liu-Tang-Zhang approximation and exact methods. *Computational Statistics and Data Analysis*, 54(4):858–862, 2010.

P. Duchesne, K. Ghoudi, and B. Rémillard. On testing for independence between the innovations of several time series. *The Canadian Journal of Statistics*, 40(3):447–479, 2012.

B. Efron and R. J. Tibshirani. *An Introduction to the Bootstrap*, volume 57 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, New York, 1993.

Y. Escoufier. Le traitement des variables vectorielles. *Biometrics*, 29(4):751–760, 1973.

Y. Fan, P. Lafaye de Micheaux, S. Penev, and D. Salopek. Multivariate nonparametric test of independence. *Journal of Multivariate Analysis*, 153:189–210, 2017.

T. S. Ferguson. *A Course in Large Sample Theory*. Chapman & Hall Texts in Statistical Science Series. Taylor & Francis, 1996.

A. Feuerverger. A consistent test for bivariate dependence. *International Statistical Review*, 61(3):419–433, 1993.

R. A. Fisher. *Statistical Methods for Research Workers*. Oliver and Boyd, London, 11th edition, 1950.

C. Genest and B. Rémillard. Tests of independence or randomness based on the empirical copula process. *Test*, 13(2):335–370, 2004.

K. Ghoudi, R. J. Kulperger, and B. Rémillard. A nonparametric test of serial independence for time series and residuals. *Journal of Multivariate Analysis*, 79(2):191–218, 2001.

P. Good. *A Practical Guide to Resampling Methods for Testing Hypotheses*. Springer Series in Statistics. Springer-Verlag, New York, second edition, 2000.

A. Gretton and L. Györfi. Consistent nonparametric tests of independence. *Journal of Machine Learning Research*, 11:1391–1423, 2010.

A. Gretton, O. Bousquet, A. Smola, and B. Scholköpf. Measuring statistical dependence with Hilbert-Schmidt norms. In S. Jain, H. Simon, and E. Tomita, editors, *Algorithmic Learning Theory*, volume 3734 of *Lecture Notes in Computer Science*, pages 63–77. Springer Berlin Heidelberg, 2005.

A. Gretton, K. Fukumizu, C.-H. Teo, L. Song, B. Schölkopf, and A. Smola. A kernel statistical test of independence. In *Advances in Neural Information Processing Systems 20*, pages 585–592. MIT Press, Cambridge, MA, 2008.

A. Gretton, K. Fukumizu, and B. K. Sriperumbudur. Discussion of: Brownian distance covariance. *The Annals of Applied Statistics*, 3(4):1285–1294, 2009.

A. Guetsop Nangue. *Tests de permutation d'indépendance en analyse multivariée*. PhD thesis, Université de Montréal, 2016.

N. Henze and T. Wagner. A new approach to the BHEP tests for multivariate normality. *Journal of Multivariate Analysis*, 62(1):1–23, 1997.

N. Henze and B. Zirkler. A class of invariant consistent tests for multivariate normality. *Communications in Statistics-Theory and Methods*, 19(10):3595–3617, 1990.

W. Hoeffding. The large-sample power of tests based on permutations of observations. *Annals of Mathematical Statistics*, 23(2):169–192, 1952.

J. P. Imhof. Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48(3-4):419–426, 1961.

J. Josse and S. Holmes. Measuring multivariate association and beyond. *Statistics Surveys*, 10:132–167, 2016.

O. Kallenberg. *Foundations of Modern Probability*. Probability and its applications. Springer, New York, 2002.

A. Kankainen. *Consistent Testing of Total Independence Based on the Empirical Characteristic Function*. PhD thesis, University of Jyväskylä, 1995.

F. Kazi-Aoual, S. Hitier, R. Sabatier, and J.-D. Lebreton. Refined approximations to permutation tests for multivariate inference. *Computational Statistics & Data Analysis*, 20(6):643–656, 1995.

J. Kellermeier. The empirical characteristic function and large sample hypothesis testing. *Journal of Multivariate Analysis*, 10(1):78–87, 1980.

I. Kojadinovic and M. Holmes. Tests of independence among continuous random vectors based on Cramér-von Mises functionals of the empirical copula process. *Journal of Multivariate Analysis*, 100(6):1137–1154, 2009.

I. Kojadinovic and J. Yan. Modeling multivariate distributions with continuous margins using the copula R package. *Journal of Statistical Software*, 34(9):1–20, 2010.

I. Kojadinovic and J. Yan. Tests of serial independence for continuous multivariate time series based on a Möbius decomposition of the independence empirical copula process. *Annals of the Institute of Statistical Mathematics*, 63(2):347–373, 2011.

S. Korkmaz, D. Goksuluk, and G. Zararsiz. MVN: An R package for assessing multivariate normality. *The R Journal*, 6(2):151–162, 2014.

H. O. Lancaster. Complex contingency tables treated by the partition of χ^2. *Journal of the Royal Statistical Society. Series B. Methodological*, 13(2):242–249, 1951.

P. Lévy. *Calcul des probabilités*. Gauthier-Villars, Paris, 1925.

T. M. Loughin. A systematic comparison of methods for combining p-values from independent tests. *Computational Statistics & Data Analysis*, 47(3):467–485, 2004.

R. Lyons. Distance covariance in metric spaces. *The Annals of Probability*, 41(5):3284–3305, 2013.

P. W. Mielke, Jr, K. J. Berry, and G. W. Brier. Application of multi-response permutation procedures for examining seasonal changes in monthly mean sea-level pressure patterns. *Monthly Weather Review*, 109:120–126, 1981.

R. B. Nelsen. *An Introduction to Copulas*. Springer Series in Statistics. Springer, New York, second edition, 2006.

K. Pearson. Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 186:343–414, 1895.

F. Pesarin and L. Salmaso. *Permutation Tests for Complex Data. Theory, Applications and Software*. Wiley Series in Probability and Statistics. John Wiley & Sons, Chichester, UK, 2010.

N. Pfister, P. Bühlmann, B. Schölkopf, and J. Peters. Kernel-based tests for joint independence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2017. URL `http://dx.doi.org/10.1111/rssb.12235`.

J. Pinkse. A consistent nonparametric test for serial independence. *Journal of Econometrics*, 84(2):205–231, 1998.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015. URL `https://www.R-project.org/`.

A. C. Rencher. *Methods of Multivariate Analysis*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York, 1995.

M. L. Rizzo and G. J. Szekely. *E-Statistics: Multivariate Inference via the Energy of Data*, 2016. URL `https://CRAN.R-project.org/package=energy`. R package version 1.7-0.

J. Romano and A. Siegel. *Counterexamples in Probability and Statistics*. Wadsworth, London, 1986.

J. P. Romano. Bootstrap and randomization tests of some nonparametric hypotheses. *The Annals of Statistics*, 17(1):141–159, 1989.

J. P. Romano and M. Wolf. Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association*, 100(469):94–108, 2005.

D. Sejdinovic, A. Gretton, and W. Bergsma. A kernel test for three-variable interactions. In *Advances in Neural Information Processing Systems 26*, pages 1126–1134. Curran Associates, Inc., Red Hook, NY, 2013a.

D. Sejdinovic, B. Sriperumbudur, A. Gretton, and K. Fukumizu. Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *The Annals of Statistics*, 41(5): 2263–2291, 2013b.

G. J. Székely and M. L. Rizzo. Rejoinder: Brownian distance covariance. *The Annals of Applied Statistics*, 3(4):1303–1308, 2009.

G. J. Székely, M. L. Rizzo, and N. K. Bakirov. Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, 35(6):2769–2794, 2007.

L. H. C. Tippett. *The Methods of Statistics*. John Wiley & Sons, Inc., New York, N. Y.; Williams & Norgate, Ltd., London, 4th edition, 1952.

D. E. Tyler. Robustness and efficiency properties of scatter matrices. *Biometrika*, 70(2): 411–420, 1983.

A. W. van der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes*. Springer Series in Statistics. Springer-Verlag, New York, 1996.

M. Westberg. Combining independent statistical tests. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 34(3):287–296, 1985.

W. Whitt. Weak convergence of probability measures on the function space $C[0, \infty)$. *The Annals of Mathematical Statistics*, 41(3):939–944, 1970.

K. Zhang, J. Peters, D. Janzing, and B. Scholköpf. Kernel-based conditional independence test and application in causal discovery. In *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence*, pages 804–813. AUAI Press, Corvallis, Oregon, 2011.

Q. Zhang, S. Filippi, A. Gretton, and D. Sejdinovic. Large-scale kernel methods for independence testing. *Statistics and Computing*, pages 1–18, 2017.