An Updated Literature Review of Distance Correlation and its Applications to Time Series

Dominic Edelmann[1], Konstantinos Fokianos[2], and Maria Pitsillou[3]

[1]*Department of Biostatistics, German Cancer Research Center, Heidelberg, Germany , E-mail: dominic.edelmann@dkfz-heidelberg.de*

[2]*Department of Mathematics & Statistics, Lancaster University, Lancaster, UK , E-mail: k.fokianos@lancaster.ac.uk*

[3]*Department of Mathematics & Statistics, University of Cyprus, Nicosia, Cyprus , E-mail: pitsillou.maria@ucy.ac.cy*

Summary

The concept of distance covariance/correlation was introduced recently to characterize dependence among vectors of random variables. We review some statistical aspects of distance covariance/correlation function and we demonstrate its applicability to time series analysis. We will see that the auto-distance covariance/correlation function is able to identify nonlinear relationships and can be employed for testing the i.i.d. hypothesis. Comparisons with other measures of dependence are included.

Key words: characteristic function; distance covariance; nonlinear time series; Portmanteau test statistics; spectral density.

1 Introduction

There has been a considerable recent interest in measuring dependence by employing the concept of *distance covariance*, a new and appealing measure of dependence for random variables, introduced by Székely et al. (2007). As examples for the importance of distance correlation in applications we mention the work by Martínez-Gómez et al. (2014); Richards et al. (2014) on astrophysical data; Guo et al. (2014); Piras et al. (2014); Oliveira et al. (2015) in molecular biology; Zhang et al. (2014) on return indices and stock market prices in the ship market; Donnelly et al. (2016) in hydrology; and Kong et al. (2012) on the association of lifestyle factors, diseases, familiar relationship and mortality. Furthermore, there has been an increasing number of works extending the concept of distance covariance in various scientific fields for independent data. In the first part of this paper, we review several of them.

Complementary to our work, we refer to the recent paper by Székely and Rizzo (2017) that gives a comprehensive overview of energy statistics and distance correlation methods.

Székely et al.'s (2007) distance covariance methodology is based on the assumption that the data are i.i.d. However, as this assumption is often violated in many practical problems, Rémillard (2009) proposed to extend the notion of distance covariance to the case of dependent data. Zhou (2012) defined the so-called *auto-distance covariance function* in a multivariate strictly stationary time series setting and Fokianos and Pitsillou (2018) introduced the *matrix auto-distance covariance function*. Compared to the i.i.d. case, there have been only few works regarding distance covariance in the context of time series. In the second part of this paper, we give a brief survey of these works and we review other measures of dependence, in the context of time series.

The article is divided as follows. In Section 2, we give the basic definitions of distance covariance and correlation functions and present their basic properties. In Section 3, we review the main works of distance covariance by presenting its extensions and modifications along with some related tests of independence as these appeared in the statistics literature, for the case of independent data. The concept of distance covariance for the case of dependent data, and especially time series data, is presented in Section 4. Tests of serial independence based on distance covariance and other measures of dependence are reviewed in Section 5. A comparison of these tests is discussed in Section 6, by giving some simulated and real data examples and testing the i.i.d. hypothesis.

2   The Distance Covariance Function

In what follows, X denotes a univariate random variable and $\mathbf{X}$ denotes a multivariate random vector.

Székely et al. (2007) introduced the distance covariance function as a new measure of dependence between random vectors, say $\mathbf{X}$ and $\mathbf{Y}$, of arbitrary, not necessarily equal, dimensions, p and q respectively. However, for the special case of $p = q = 1$, this definition can be also found in Feuerverger (1993, Sec. 4) who considered a rank test for bivariate dependence. The definition of distance covariance relies on the joint characteristic function of $\mathbf{X}$ and $\mathbf{Y}$, which is denoted by $\phi_{(\mathbf{X},\mathbf{Y})}(\mathbf{t},\mathbf{s}) = E\left[\exp\left(i(\mathbf{t}'\mathbf{X} + \mathbf{s}'\mathbf{Y})\right)\right]$, where $(\mathbf{t},\mathbf{s}) \in \mathbb{R}^{p+q}$ and $i^2 = -1$. The corresponding marginal characteristic function of $\mathbf{X}$ and $\mathbf{Y}$ are denoted by $\phi_{\mathbf{X}}(\mathbf{t}) = E\left[\exp\left(i\mathbf{t}'\mathbf{X}\right)\right]$ and $\phi_{\mathbf{Y}}(\mathbf{s}) = E\left[\exp\left(i\mathbf{s}'\mathbf{Y}\right)\right]$, respectively. The distance covariance function is defined as the nonnegative square root of a weighted L_2 distance between the joint and the product of the marginal characteristic functions of two random vectors, namely

$$V^2(\mathbf{X},\mathbf{Y}) \;=\; \int_{\mathbb{R}^{p+q}} \left|\phi_{(\mathbf{X},\mathbf{Y})}(\mathbf{t},\mathbf{s}) - \phi_{\mathbf{X}}(\mathbf{t})\phi_{\mathbf{Y}}(\mathbf{s})\right|^2 \omega(\mathbf{t},\mathbf{s}) d\mathbf{t} d\mathbf{s}, \tag{1}$$

where $\omega(\cdot, \cdot) : \mathbb{R}^{p+q} \to \mathbb{R}$ is a weight function for which the above integral exists and whose choice is discussed later on. Rescaling (1) leads to the definition of the distance correlation function between $\mathbf{X}$ and $\mathbf{Y}$, which is the positive square root of

$$R^2(\mathbf{X}, \mathbf{Y}) = \begin{cases} \dfrac{V^2(\mathbf{X}, \mathbf{Y})}{\sqrt{V^2(\mathbf{X}, \mathbf{X})V^2(\mathbf{Y}, \mathbf{Y})}}, & V^2(\mathbf{X}, \mathbf{X})V^2(\mathbf{Y}, \mathbf{Y}) > 0; \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

The previous display shows that the distance correlation function is a coefficient analogous to Pearson's correlation coefficient. But, unlike the classical coefficient which measures the linear relationship between $\mathbf{X}$ and $\mathbf{Y}$ and can be zero even when the variables are dependent, the distance correlation vanishes only in the case where $\mathbf{X}$ and $\mathbf{Y}$ are independent–see Fig 1. Further properties of (1) and (2) are established later. The choice of the weight function $\omega(\cdot, \cdot)$ is crucial. Choosing the non-integrable weight function of the form

$$\omega(\mathbf{t}, \mathbf{s}) \quad = \quad (c_p c_q \, |\mathbf{t}|_p^{1+p} \, |\mathbf{s}|_q^{1+q})^{-1}, \tag{3}$$

where $c_d = \pi^{(1+d)/2}/\Gamma((1+d)/2)$, with $\Gamma(\cdot)$ denoting the Gamma function, $R^2(\mathbf{X}, \mathbf{Y})$ given by (2) is scale and rotation invariant. In particular, (3) is (up to a constant) the only weight function such that $V^2(\mathbf{X}, \mathbf{Y})$ is rigid motion invariant and scale equivariant (Székely and Rizzo, 2012); moreover the resulting *distance standard deviation* defined as $V(X) := V(X, X)$ is then an axiomatic measure of spread (Edelmann et al., 2017). An early reference discussing the particular choice of (3) is that of Feuerverger (1993). For related work on characteristic functions, see also Meintanis and Iliopoulos (2008) and Hlávka et al. (2011), among others. Furthermore, Bakirov et al. (2006) proposed a weight function $\omega(\cdot, \cdot)$ different than (3). However, the choice of (3) yields to computational advantages. Denote by $|x|_p$ the Euclidean norm in $\mathbb{R}^p$. The main properties of (1) and (2), when (3) is used, are given by the following:

1. If $E\left(|\mathbf{X}|_p + |\mathbf{Y}|_q\right) < \infty$, then the distance correlation, $R(\mathbf{X}, \mathbf{Y})$, satisfies $0 \leq R(\mathbf{X}, \mathbf{Y}) \leq 1$ and $R(\mathbf{X}, \mathbf{Y}) = 0$ if and only if $\mathbf{X}$ and $\mathbf{Y}$ are independent.

2. $R(\mathbf{X}, \mathbf{Y})$ is invariant under orthogonal transformations of the form $(\mathbf{X}, \mathbf{Y}) \to \left(\alpha_1 + b_1 C_1 \mathbf{X}, \alpha_2 + b_2 C_2 \mathbf{Y}\right)$, where α_1, α_2 are arbitrary vectors, b_1, b_2 are arbitrary nonzero numbers and C_1, C_2 are arbitrary orthogonal matrices.

3. If $p = q = 1$ and X and Y have standard normal distributions with $r = \text{Cov}(X, Y)$, then $R(X, Y) \leq |r|$ and

$$R^2(X, Y) = \frac{r \arcsin r + \sqrt{1 - r^2} - r \arcsin r/2 - \sqrt{4 - r^2} + 1}{1 + \pi/3 - \sqrt{3}}. \tag{4}$$

4. When $E\big(|\mathbf{X}|_p^2 + |\mathbf{Y}|_q^2\big) < \infty$ then

$$V^2(\mathbf{X}, \mathbf{Y}) = E\left|\mathbf{X} - \mathbf{X}'\right|_p \left|\mathbf{Y} - \mathbf{Y}'\right|_q + E\left|\mathbf{X} - \mathbf{X}'\right|_p E\left|\mathbf{Y} - \mathbf{Y}''\right|_q - 2E\left|\mathbf{X} - \mathbf{X}'\right|_p \left|\mathbf{Y} - \mathbf{Y}''\right|_q. \tag{5}$$

with $(\mathbf{X}', \mathbf{Y}')$ and $(\mathbf{X}'', \mathbf{Y}'')$ being independent copies of $(\mathbf{X}, \mathbf{Y})$.

3   Estimation, Testing and Further Properties

3.1   *Estimation*

We consider the empirical counterparts of (1) and (2), when using the weight function (3). The empirical distance covariance and correlation measures are functions of the double centered distance matrices of the samples. Suppose that $(\mathbf{X}_i, \mathbf{Y}_i)$, $i = 1, \cdots, n$, is a random sample from the joint distribution of the random vectors $\mathbf{X}$ and $\mathbf{Y}$. Based on this sample, consider the $n \times n$ Euclidean pairwise distance matrices with elements $(a_{ij}) = \big(|\mathbf{X}_i - \mathbf{X}_j|_p\big)$ and $(b_{ij}) = \big(|\mathbf{Y}_i - \mathbf{Y}_j|_q\big)$. These matrices are double centered so that their row and column means are equal to zero. In other words, let

$$A_{ij} = a_{ij} - \bar{a}_{i.} - \bar{a}_{.j} + \bar{a}_{..}, \quad B_{ij} = b_{ij} - \bar{b}_{i.} - \bar{b}_{.j} + \bar{b}_{..},$$

where $\bar{a}_{i.} = \big(\sum_{j=1}^n a_{ij}\big)/n$, $\bar{a}_{.j} = \big(\sum_{i=1}^n a_{ij}\big)/n$, $\bar{a}_{..} = \big(\sum_{i,j=1}^n a_{ij}\big)/n^2$. Similarly, we define the quantities $b_{i.}, b_{.j}$ and $b_{..}$. The sample distance covariance is defined by the square root of the statistic

$$\widehat{V}^2(\mathbf{X}, \mathbf{Y}) \quad = \quad \frac{1}{n^2} \sum_{i,j=1}^n A_{ij} B_{ij}. \tag{6}$$

We compute the squared sample distance correlation in terms of (6) by (2) as

$$\widehat{R}^2(\mathbf{X}, \mathbf{Y}) = \begin{cases} \dfrac{\widehat{V}^2(\mathbf{X}, \mathbf{Y})}{\sqrt{\widehat{V}^2(\mathbf{X}, \mathbf{X})\widehat{V}^2(\mathbf{Y}, \mathbf{Y})}}, & \widehat{V}^2(\mathbf{X}, \mathbf{X})\widehat{V}^2(\mathbf{Y}, \mathbf{Y}) > 0; \\ \\ 0, & \text{otherwise.} \end{cases}$$

Furthermore, note that by multiplying out (6) we obtain the alternative expression

$$\widehat{V}^2(\mathbf{X}, \mathbf{Y}) = \frac{1}{n^2} \sum_{i,j=1}^{n} a_{ij} b_{ij} + \frac{1}{n^4} \sum_{i,j=1}^{n} a_{ij} \sum_{i,j=1}^{n} b_{ij} - \frac{2}{n^3} \sum_{i,j,k=1}^{n} a_{ij} \, b_{jk}. \tag{7}$$

The main properties of sample distance covariance and correlation functions are the following:

- The estimators $\widehat{V}^2(\mathbf{X}, \mathbf{Y})$ and $\widehat{R}^2(\mathbf{X}, \mathbf{Y})$ are both strongly consistent, that is they both converge almost surely to their population counterparts (1) and (2) respectively, as n tends to infinity, provided that $E[|\mathbf{X}|_p] < \infty$ and $E[|\mathbf{Y}|_q] < \infty$.

- $\widehat{V}^2(\mathbf{X}, \mathbf{Y}) \geq 0$ a.s., where the equality holds when $\mathbf{X}$ and $\mathbf{Y}$ are independent.

- $0 \leq \widehat{R}(\mathbf{X}, \mathbf{Y}) \leq 1$.

- $\widehat{R}(\mathbf{X}, \alpha + b\mathbf{X}C) = 1$, where α is a vector, b is a nonzero real number and C is an orthogonal matrix.

Figure 1 illustrates a comparison between the empirical distance correlation $\widehat{R}$ and the empirical Pearson correlation $\widehat{Cor}$ for several examples of nonlinear association between two real-valued variables X and Y, where the sample size is fixed at $n = 1000$. While the Pearson correlation $\widehat{Cor}$ is very close to 0 in these examples, signifying that there are either no or very weak linear dependencies, the distance correlation coefficient $\widehat{R}$ is substantially larger than 0 for all cases.
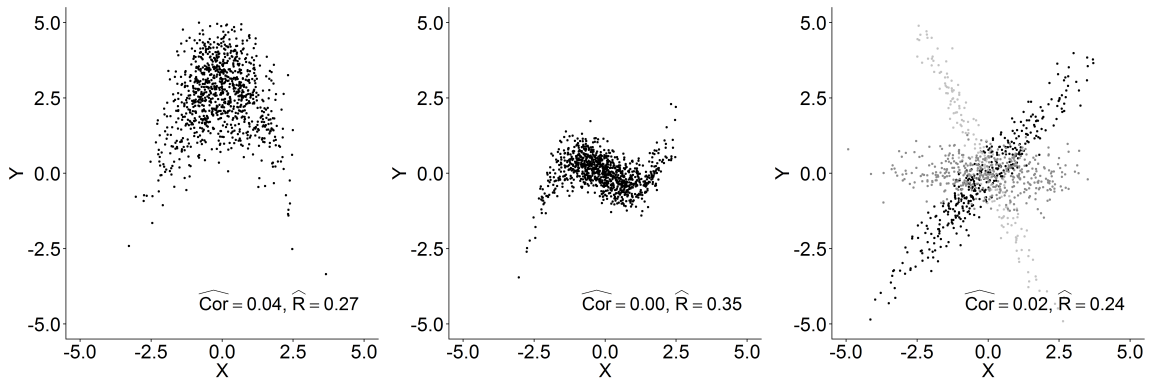

Figure 1: Comparison of empirical Pearson correlation $\widehat{Cor}$ and empirical distance correlation $\widehat{R}$ for different settings. The number of samples is fixed at $n = 1000$. The left graph illustrates a quadratic dependence between X and Y. The middle graph represents a cubic relationship between X and Y. In the right graph, the relationship between X and Y depends on three different subgroups. In the first subgroup (medium gray), X and Y are independent. In the second subgroup (black), there is a positive correlation between X and Y. In the second subgroup (light gray), there is a negative correlation between X and Y.

It turns out that $\widehat{V}^2(\mathbf{X}, \mathbf{Y})$ is not an unbiased estimator of $V^2(\mathbf{X}, \mathbf{Y})$, but it holds that

$$E[\widehat{V}^2(\mathbf{X}, \mathbf{Y})] \quad = \quad \frac{(n-1)(n-2)^2}{n^3} V^2(\mathbf{X}, \mathbf{Y}) + \frac{2(n-1)^2}{n^3}\gamma - \frac{(n-1)(n-2)}{n^3}\alpha\beta,$$

where $\alpha = E\left|\mathbf{X} - \mathbf{X}'\right|_p$, $\beta = E\left|\mathbf{Y} - \mathbf{Y}'\right|_q$ and $\gamma = E[\left|\mathbf{X} - \mathbf{X}'\right|_p \left|\mathbf{Y} - \mathbf{Y}'\right|_q]$. Here, $\mathbf{X}'$ (respectively $\mathbf{Y}'$) denotes an independent copy of $\mathbf{X}$ (respectively, $\mathbf{Y}$). An unbiased estimator of squared distance covariance proposed by Székely and Rizzo (2014) is given by

$$\widehat{V}_U(\mathbf{X}, \mathbf{Y}) \quad = \quad \frac{1}{n(n-3)} \sum_{i \neq j} \widetilde{A}_{ij} \widetilde{B}_{ij}$$

for $n > 3$, where the so-called $\mathcal{U}$-centered matrices, $\widetilde{A}_{ij}$ have the additional property that $E[\widetilde{A}_{ij}] = 0$ for all i,j and are defined by

$$\widetilde{A}_{ij} \quad = \quad \begin{cases} a_{ij} - \frac{1}{n-2}\sum_{l=1}^n a_{il} - \frac{1}{n-2}\sum_{k=1}^n a_{kj} + \frac{1}{(n-1)(n-2)}\sum_{k,l=1}^n a_{kl}, & i \neq j; \\ 0, & i = j. \end{cases}$$

For $p = q = 1$, Huo and Székely (2016) have shown that $\widehat{V}_U(X, Y)$ is a U-statistic, which is degenerate in the case where X and Y are independent. For further details concerning the properties of $\widehat{V}_U(\mathbf{X}, \mathbf{Y})$, we refer to the paper by Huang and Huo (2017). By using the U-statistic representation of $\widehat{V}_U(X, Y)$, Huo and Székely (2016) show that it can be computed by an $O(n \log n)$ algorithm. This algorithm considerably speeds up the calculation of the distance correlation coefficient for large sample sizes. For higher dimensions, Huang and Huo (2017) derive an alternative independence test based on random projections and distance correlation. Their method has computational complexity of order $O(Kn \log n)$, where K is the number of random projections.

3.2   *Asymptotic tests*

Following Székely et al. (2007), it can be shown that, under the null hypothesis of independence between the random vectors $\mathbf{X}$ and $\mathbf{Y}$,

$$n\widehat{V}^2(\mathbf{X}, \mathbf{Y}) \quad \overset{n \to \infty}{\Longrightarrow} \quad \sum_{j=1}^{\infty} \lambda_j Z_j, \tag{8}$$

in distribution, where $\{Z_j\}$ are independent standard normal variables and $\{\lambda_j\}$ are eigenvalues which depend on the joint distribution of the random vectors $(\mathbf{X}, \mathbf{Y})$, provided that $E[|\mathbf{X}|_p] < \infty$ and $E[|\mathbf{Y}|_q] < \infty$. Moreover the

quadratic form $Q := \sum_{j=1}^{\infty} \lambda_j Z_j$ satisfies $E[Q] = E\left|\mathbf{X} - \mathbf{X}'\right|_p E\left|\mathbf{Y} - \mathbf{Y}''\right|_q$.

On the other hand, if $\mathbf{X}$ and $\mathbf{Y}$ are not independent, then

$$n\widehat{V}^2(\mathbf{X}, \mathbf{Y}) \stackrel{n \to \infty}{\longrightarrow} \infty \tag{9}$$

in probability. Results (8) and (9) imply a test for independence based on the statistic $n\widehat{V}^2$. While Székely et al. (2007) derive an asymptotic test with asymptotic significance level at most α, they point out that this significance level can be quite conservative for many distributions. As an alternative, they propose a Monte-Carlo permutation test, which is now the standard method for testing independence based on the distance covariance. This permutation test is also implemented in the R package `energy` by Rizzo and Székely (2016). Recently, Huang and Huo (2017) have proposed an alternative distance covariance test based on a gamma approximation of the asymptotic distribution of the test statistic under the null hypothesis.

The idea of employing (1) for detecting independence was previously discussed by Feuerverger (1993) who considered measures of the form of (1). To develop a nonparametric test for independence, Feuerverger (1993) suggested to replace univariate sample points X_i and Y_i, $i = 1, \ldots, n$, by approximate normal score quantities X_i' and Y_i', where $X_i' = \Phi^{-1}\left(\left(\text{rank}(X_i) - 3/8\right)/\left(n + 1/4\right)\right)$. Then, he proposed the test statistic

$$\int_{\mathbb{R}^2} \left|\hat{\phi}_{(X', Y')}(t, s) - \hat{\phi}_{X'}(t)\hat{\phi}_{Y'}(s)\right|^2 |t|^{-2} |s|^{-2} \, dt ds, \tag{10}$$

where $\hat{\phi}_{(X', Y')}(t, s) = \sum_{j=1}^{n} e^{i(tX_j' + sY_j')}/n$ is the empirical joint characteristic function of (X, Y) and $\hat{\phi}_{X'}(t) := \hat{\phi}_{(X', Y')}(t, 0)$ and $\hat{\phi}_{Y'}(s) := \hat{\phi}_{(X', Y')}(0, s)$ are the empirical marginal characteristic functions. Clearly, (10) is the empirical version of (1) by using (3) and is identical to (8); the differences are the use of scores instead of observations and the restriction to the univariate case. More on the comparison between the distance covariance and the statistics proposed by Feuerverger (1993) can be found in Gretton et al. (2009). Furthermore, Feuerverger (1993) derived conditions for the weight function $\omega(\cdot, \cdot)$ under which (10) can be written analogously to the right-hand side of (7), where $|\cdot - \cdot|_1$ is replaced by some function $g(\cdot, \cdot)$. Kankainen (1995) gave an explicit example, showing that choosing a Gaussian weight function in (1) results in a statistic of the form (7), where a_{ij} and b_{ij} are replaced by $1 - \exp(-(X_i - X_j)^2/(2\sigma^2))$ and $1 - \exp(-(Y_i - Y_j)^2/(2\sigma^2))$, respectively. In a recent preprint, Böttcher and Schilling (2017a) enlighten the connection between the weight function ω and the corresponding distance $g(\cdot, \cdot)$. Furthermore, they give several examples including the one given by Kankainen (1995).

3.3   Generalizations and Modifications of the Distance Covariance

The α-distance covariance function, $V^{(\alpha)}(\mathbf{X}, \mathbf{Y})$, is a generalization of the distance covariance function (1) for a constant α that lies in the interval (0,2), when (3) is used. More precisely, it is the non-negative number given by the square root of (Székely et al., 2007)

$$V^{2(\alpha)}(\mathbf{X}, \mathbf{Y}) \quad = \quad \frac{1}{C(p, \alpha)C(q, \alpha)} \int_{\mathbb{R}^{p+q}} \frac{\left|\phi_{(\mathbf{X},\mathbf{Y})}(\mathbf{t}, \mathbf{s}) - \phi_{\mathbf{X}}(\mathbf{t})\phi_{\mathbf{Y}}(\mathbf{s})\right|^2}{|\mathbf{t}|_p^{\alpha+p} |\mathbf{s}|_q^{\alpha+q}} d\mathbf{t} d\mathbf{s}, \tag{11}$$

where $C(p, \alpha)$ and $C(q, \alpha)$ are given by

$$C(d, \alpha) \quad = \quad \frac{2\pi^{d/2}\Gamma((1 - \alpha/2))}{\alpha 2^\alpha \Gamma((d + \alpha)/2)}.$$

An estimator of (11) is computed analogously by defining a distance matrix with elements $a_{ij} = |\mathbf{X}_i - \mathbf{X}_j|_p^\alpha$ and $b_{ij} = |\mathbf{Y}_i - \mathbf{Y}_j|_q^\alpha$ and working analogously as in the case of (1). Note that (11) extends the theory of distance covariance since (1) is obtained as a special case when $\alpha = 1$. Possible extensions of α-distance covariance for values of the parameter α which do not lie in the interval $(0, 2)$ are discussed by Dueck et al. (2015).

On the other hand, Kosorok (2009) suggested the weight function

$$\omega(\mathbf{s}, \mathbf{t}) = c_p^{-1} c_q^{-1} (\sqrt{\mathbf{s}'M_1\mathbf{s}})^{-1-p} (\sqrt{\mathbf{t}'M_2\mathbf{t}})^{-1-q},$$

where M_1 and M_2 are positive-definite matrices. By choosing $M_1 = \Sigma_{\mathbf{X}}$ and $M_2 = \Sigma_{\mathbf{Y}}$, where $\Sigma_{\mathbf{X}}$ and $\Sigma_{\mathbf{Y}}$ are the covariance matrices of $\mathbf{X}$ and $\mathbf{Y}$, respectively we obtain the affinely invariant distance covariance and distance correlation (Dueck et al., 2014), which are given, respectively, by

$$\tilde{V}^2(\mathbf{X}, \mathbf{Y}) = V^2(\Sigma_{\mathbf{X}}^{-1/2}\mathbf{X}, \Sigma_{\mathbf{Y}}^{-1/2}\mathbf{Y}),$$
$$\tilde{R}^2(\mathbf{X}, \mathbf{Y}) = \frac{\tilde{V}^2(\mathbf{X}, \mathbf{Y})}{\sqrt{\tilde{V}^2(\mathbf{X}, \mathbf{X})\tilde{V}^2(\mathbf{Y}, \mathbf{Y})}} = R^2(\Sigma_{\mathbf{X}}^{-1/2}\mathbf{X}, \Sigma_{\mathbf{Y}}^{-1/2}\mathbf{Y}). \tag{12}$$

$\tilde{R}$ retains all the important properties of distance correlation. In particular $0 \leq \tilde{R}(\mathbf{X}, \mathbf{Y}) \leq 1$ and $\tilde{R}(\mathbf{X}, \mathbf{Y}) = 0$ if and only if $\mathbf{X}$ and $\mathbf{Y}$ are independent. Additionally it is invariant under general invertible affine transformation $\mathbf{X} \to A_1 \mathbf{X} + A_2$ and $\mathbf{Y} \to B_1 \mathbf{Y} + B_2$.

An alternative way to generalize the distance covariance is based on (5) (or (6) if second moments do not exist). Lyons (2013) extends distance covariance to random variables $\mathbf{X}$ and $\mathbf{Y}$ on general metric spaces of negative type.

If $d_{\mathcal{X}}$ (respectively, $d_{\mathcal{Y}}$) is a distance on the space where $\mathbf{X}$ is defined (respectively $\mathbf{Y}$) and if second moments exist then the generalization is given by

$$V_{d_{\mathcal{X}},d_{\mathcal{Y}}}^2(\mathbf{X}, \mathbf{Y}) = E\, d_{\mathcal{X}}(\mathbf{X}, \mathbf{X}')d_{\mathcal{Y}}(\mathbf{Y}, \mathbf{Y}') + E\, d_{\mathcal{X}}(\mathbf{X}, \mathbf{X}')\, E\, d_{\mathcal{Y}}(\mathbf{Y}, \mathbf{Y}') - 2E\, d_{\mathcal{X}}(\mathbf{X}, \mathbf{X}')d_{\mathcal{Y}}(\mathbf{Y}, \mathbf{Y}'), \qquad (13)$$

where $\mathbf{X}'$, $\mathbf{Y}'$ are independent copies of $\mathbf{X}$ and $\mathbf{Y}$. Lyons (2013) showed that this measure characterizes independence if the underlying metric spaces are of *strong* negative type; in particular, this class of sets includes all separable Hilbert spaces. Sejdinovic et al. (2013) extend (13) to semimetric spaces of negative type. They further establish the equivalence of distance covariance measures on semimetric spaces of negative type (as in (13)) and the Hilbert-Schmidt independence criterion (HSIC), see Gretton et al. (2005, 2008). In particular, the standard distance covariance can be written as a special case of the HSIC with kernel functions given by

$$k_1(\mathbf{x}_1, \mathbf{x}_2) = |\mathbf{x}_1|_p + |\mathbf{x}_2|_p - |\mathbf{x}_1 - \mathbf{x}_2|_p, \qquad k_2(\mathbf{y}_1, \mathbf{y}_2) = |\mathbf{y}_1|_q + |\mathbf{y}_2|_q - |\mathbf{y}_1 - \mathbf{y}_2|_q.$$

Another extension of distance covariance was proposed by Székely and Rizzo (2009) who considered the notion of Brownian distance covariance. Suppose that $\mathbf{X}$ and $\mathbf{Y}$ are real-valued random variables in $\mathbb{R}^p$ and $\mathbb{R}^q$ respectively, and $\mathbf{W}(\mathbf{s})$ and $\mathbf{W}'(\mathbf{t})$ are independent Brownian motions for all $\mathbf{s} \in \mathbb{R}^p$ and $\mathbf{t} \in \mathbb{R}^q$. Then the Brownian covariance function of $\mathbf{X}$ and $\mathbf{Y}$ is defined as the positive square root of

$$W^2(\mathbf{X}, \mathbf{Y}) \quad = \quad E[\mathbf{X}_{\mathbf{W}}\mathbf{X}'_{\mathbf{W}'}\mathbf{Y}_{\mathbf{W}}\mathbf{Y}'_{\mathbf{W}'}], \qquad\qquad (14)$$

where

$$\mathbf{X}_{\mathbf{W}} \quad = \quad \mathbf{W}(\mathbf{X}) - E[\mathbf{W}(\mathbf{X})|\mathbf{W}]$$

and $(\mathbf{W}, \mathbf{W}')$ are independent of $(\mathbf{X}, \mathbf{Y}, \mathbf{X}', \mathbf{Y}')$. The authors proved the surprising result that the Brownian covariance coincides with the squared population distance covariance (1) (Székely and Rizzo, 2009, Thm. 8), that is

$$W^2(\mathbf{X}, \mathbf{Y}) = V^2(\mathbf{X}, \mathbf{Y}).$$

Further extensions of this idea to other stochastic processes, such as fractional Brownian motions and to more general Gaussian processes, like the Laplace-Gaussian process are given by Székely and Rizzo (2013). When

(X, Y) is a bivariate vector and the processes W and W' are replaced by the identity function id, that is $X_{id} = X - E[X]$ and $Y_{id} = Y - E[Y]$, then we obtain the absolute value of the classical Pearson covariance.

Furthermore, Székely and Rizzo (2013) extended this idea to a general class of Gaussian processes, showing that for any stochastic processes of this form, (14) can be expressed similarly as the right-hand side of (13) (where the (semi)-metrics are replaced by conditional negative definite continuous symmetric functions generating the covariance functions of the processes). Fiedler (2016) developed new dependence measures based on the distance correlation function for spatial stochastic processes. Moreover, a recent contribution of the appropriate definitions of distance covariance for measuring dependence in the context of stochastic processes is that of Matsui et al. (2017).

Rémillard (2009) suggested to extend the notion of distance covariance by replacing the samples $\{\mathbf{X}_i\}$, $\{\mathbf{Y}_i\}$ by their normalized ranks, that is $\{R_{\mathbf{X},i}/n\}$ and $\{R_{\mathbf{Y},i}/n\}$ respectively, where $R_{X,ij}$ is the rank of X_{ij} among $X_{1j}, \ldots, X_{nj}, j = 1, \ldots, p$. He further suggested to generalize the distance covariance to measure dependence among more than two random vectors. For random vectors $\mathbf{X}_1, \ldots, \mathbf{X}_d$, taking values in $\mathbb{R}^{p_1}, \ldots, \mathbb{R}^{p_d}$, Belu (2012) defined the distance covariance as the weighted difference between the d-dimensional joint characteristic function,

$$\phi(\mathbf{t}_1, \ldots, \mathbf{t}_d) = E\Big[\exp\Big(i(\mathbf{t}_1'\mathbf{X}_1 + \cdots + \mathbf{t}_d'\mathbf{X}_d)\Big)\Big],$$

and the product of the marginals

$$\phi^{(j)}(\mathbf{t}_j) = E\Big[\exp\Big(i(\mathbf{t}_j'\mathbf{X}_j)\Big)\Big], \quad j = 1, \cdots, d.$$

The extended version of distance covariance (compare (1) and (11)) is defined as the positive square root of

$$V^{2(\alpha)}(\mathbf{X}_1, \ldots, \mathbf{X}_d) \quad = \quad \frac{1}{c_{p_1} \ldots c_{p_d}} \int_{\mathbb{R}^{p_1 + \cdots + p_d}} \frac{\big|\phi(\mathbf{t}_1, \ldots, \mathbf{t}_d) - \phi^{(1)}(\mathbf{t}_1) \cdots \phi^{(d)}(\mathbf{t}_d)\big|^2}{|\mathbf{t}_1|_{p_1}^{\alpha + p_1} \ldots |\mathbf{t}_d|_{p_d}^{\alpha + p_d}},$$

where $c_{p_j}, j = 1, \ldots, d$ is given as before and α is a positive constant that lies in the interval (0,2). The above quantity can be employed for testing independence in the case of more than two random vectors but Bakirov and Székely (2011, p. 471) point out that this integral can diverge. For novel approaches of multivariate version of distance covariance, see Yao et al. (2018), Böttcher and Schilling (2017b) and Chakraborty and Zhang (2018). A related line of research is based on HSIC; see Pfister et al. (2018).

Wang et al. (2015) introduced the *conditional distance covariance* as the square root of

$$\text{cdCov}^2(\mathbf{X}, \mathbf{Y}|\mathbf{Z}) = \frac{1}{c_p\, c_q} \int_{\mathbb{R}^{p+q}} \frac{\left|\phi_{(\mathbf{X},\mathbf{Y}|\mathbf{Z})}(\mathbf{t},\mathbf{s}) - \phi_{\mathbf{X}|\mathbf{Z}}(\mathbf{t})\phi_{\mathbf{Y}|\mathbf{Z}}(\mathbf{s})\right|^2}{|\mathbf{t}|_p^{p+1}|\mathbf{s}|_q^{q+1}} d\mathbf{t}d\mathbf{s},$$

where $\phi_{(\mathbf{X},\mathbf{Y})|\mathbf{Z}}$ denotes the conditional joint characteristic function of $(\mathbf{X},\mathbf{Y})$ given $\mathbf{Z}$ and $\phi_{\mathbf{X}|\mathbf{Z}}$ and $\phi_{\mathbf{Y}|\mathbf{Z}}$ denote the respective conditional marginal characteristic functions. Analogously the *conditional distance correlation* is defined as the square root of

$$\text{cdCor}^2(\mathbf{X}, \mathbf{Y}|\mathbf{Z}) = \frac{\text{cdCov}^2(\mathbf{X}, \mathbf{Y}|\mathbf{Z})}{\sqrt{\text{cdCov}^2(\mathbf{X}, \mathbf{X}|\mathbf{Z})\text{cdCov}^2(\mathbf{Y}, \mathbf{Y}|\mathbf{Z})}},$$

provided that $\text{cdCov}^2(\mathbf{X}, \mathbf{X})|\mathbf{Z})\text{cdCov}^2(\mathbf{Y}, \mathbf{Y}|\mathbf{Z}) > 0$, and 0, otherwise. It is easy to see that $\text{cdCov}^2(X, Y|Z)$ equals 0 if and only if X and Y are independent given Z. In addition, Wang et al. (2015) derive two consistent, ready-to-use sample measures for the conditional distance covariance, which do not involve integration of the empirical characteristic functions.

It is important to point out that the conditional distance correlation is not equivalent to the *partial distance correlation* introduced by Székely and Rizzo (2014), which can be defined via

$$R^*(X, Y; Z) = \frac{R^2(X, Y) - R^2(X, Z)R^2(Y, Z)}{\sqrt{(1 - R^4(X, Z))}\sqrt{(1 - R^4(Y, Z))}},$$

if $R(X, Z) \neq 1$ and $R(Y, Z) \neq 1$ and 0 otherwise. In particular, Székely and Rizzo (2014) show that $R^*(X, Y; Z) = 0$ is not equivalent to conditional independence of X and Y given Z.

Another related measure is the *martingale difference correlation* (Shao and Zhang, 2014), which is defined for $\mathbf{X} \in \mathbb{R}^p$ and $Y \in \mathbb{R}$ by

$$\text{MDC} = \frac{\text{MDD}(Y|\mathbf{X})}{\sqrt{\text{var}^2(Y)V^2(\mathbf{X}, \mathbf{X})}}.$$

Here $\text{MDD}(U, V)$ denotes the *martingale difference divergence* given by

$$\text{MDD}(Y|\mathbf{X}) = \frac{1}{c_p} \int_{\mathbb{R}^{p+q}} \frac{\left|E\left[Y \exp\left(i\mathbf{t}'\mathbf{X}\right)\right] - E\left[\exp\left(i\mathbf{t}'\mathbf{X}\right)\right]E[Y]\right|^2}{|\mathbf{t}|_p^{p+1}} d\mathbf{t}d\mathbf{s}.$$

The martingale correlation satisfies the property that $\text{MCC}(V|U) = 0$ if and only if $E[V|U] = E[V]$ almost surely. In Park et al. (2015), a conditional version of the martingale difference correlation is introduced; it is defined analogously to the partial distance correlation of Székely and Rizzo (2014).

3.4 Variable selection

Since the distance correlation coefficient captures any kind of dependence including nonlinear or non monotone dependencies, it appears natural to apply this coefficient for variable selection in nonlinear models. For this purpose, Li et al. (2012) have developed the distance correlation sure independence screening method (DC-SIS) for variable screening. As the usual SIS (Fan and Lv, 2008) selects the variables with the highest Pearson correlation, DC-SIS selects the variables with the highest distance correlation. Unlike SIS, which is applicable to linear models only, DC-SIS can detect arbitrary associations of the predictors with the response. Li et al. (2012) show the sure screening property for DC-SIS, under moderate assumptions, and show that this method accommodates linear models as a special case. Two iterative modifications for DC-SIS have been developed recently (see Yenigün and Rizzo (2015); Zhong and Zhu (2015)). However there are no known theoretical properties of these methods as of yet, to the best of our knowledge. Kong et al. (2015) have derived an automatic stopping rule for DC-SIS, which determines the number of predictors to select. Recently, Liu and Wang (2017) have proposed a modified version of DC-SIS based on the concept of conditional distance correlation (Wang et al., 2015), which allows for feature screening conditional on some prespecified variables. Other notable approaches are the methods by Shao and Zhang (2014), who propose a variable screening algorithm based on the martingale difference correlation and Kong et al. (2017) who derive the sure screening property for a two-stage screening procedure that applies distance correlation on transformed variables.

Considering the equivalence of the generalized distance covariance and the Hilbert Schmidt Independence Criterion (Sejdinovic et al., 2013), more general distance correlation based variable selection methods emerge as special cases of HSIC based methods. As examples, see Song et al. (2012), Balasubramanian et al. (2013) and Tonde (2016).

3.5 Closed form expressions for the distance covariance of multivariate distributions

Distance correlation possesses an astonishingly simple sample quantity but the analytical evaluation of its population quantity for known parametric distribution often represents computational problems. For distributions with finite support, it is straightforward (but cumbersome) to evaluate (2) to obtain closed form expressions for the distance covariance. For example, for two Bernoulli-distributed random variables X and Y with $P\big((X,Y) = (0,0)\big) = P\big((X,Y) = (1,1)\big) = p/2$ and $P\big((X,Y) = (0,1)\big) = P\big((X,Y) = (1,0)\big) = (1-p)/2$, we have that

$$R^2(X,Y) = (2p-1)^2 = \rho^2(X,Y),$$

where ρ denotes the ordinary Pearson correlation coefficient. For distributions with non-finite support, the usual approach is to directly calculate the integral (1). For normally distributed X and Y with correlation r and $EX = EY = 0$, $\mathrm{Var}(X) = \mathrm{Var}(Y) = 1$ we have already seen identity (4). Dueck et al. (2014) extended this result to the affinely invariant distance correlation (recall (12)) of two multivariate normal vectors $\mathbf{X}$ and $\mathbf{Y}$. When $Cov(\mathbf{X}) = I_p$ and $Cov(\mathbf{Y}) = I_q$ this coincides with the regular distance correlation. In this particular case, the squared affinely invariant distance covariance is given as

$$\tilde{V}^2(\mathbf{X}, \mathbf{Y}) = 4\pi \frac{c_{p-1}}{c_p} \frac{c_{q-1}}{c_q} \times \left({}_3F_2\left(\tfrac{1}{2}, -\tfrac{1}{2}, -\tfrac{1}{2}; \tfrac{p}{2}, \tfrac{q}{2}; \Lambda\right) - 2\, {}_3F_2\left(\tfrac{1}{2}, -\tfrac{1}{2}, -\tfrac{1}{2}; \tfrac{p}{2}, \tfrac{q}{2}; \tfrac{1}{4}\Lambda\right) + 1 \right),$$

where c_p has been defined immediately after (3), Λ denotes the squared cross covariance matrix of $\mathbf{X}$ and $\mathbf{Y}$ and ${}_3F_2$ denotes a generalized hypergeometric function of matrix argument (see Dueck et al. (2014) for the details). Dueck et al. (2017) outlined a general approach for calculating the distance correlation for the so-called Lancaster class of distributions. The authors illustrate this result by calculating the distance correlation for the bivariate gamma, the bivariate Poisson and the bivariate negative binomial since all of those distributions belong to the Lancaster class. Moreover, they derive an alternative expression for the affinely invariant distance correlation of the multivariate normal distribution and derive the distance correlation for generalizations of the bivariate normal.

3.6 The distance correlation coefficient in high dimensions

As stated in subsection 3.2, distance covariance provides a test for independence for random vectors $\mathbf{X}$ and $\mathbf{Y}$ of arbitrary dimension p and q, respectively. However, practitioners often use distance correlation rather as an ad-hoc measure for independence than applying a formal test. While this appears to work well for univariate random variables, some authors have reported that the direct interpretation of distance correlation is questionable in high dimensions. Dueck et al. (2014) showed that for fixed q and standard multivariate normal random vectors $(\mathbf{X}, \mathbf{Y}) \in \mathbb{R}^{p+q}$

$$\lim_{p \to \infty} R(\mathbf{X}, \mathbf{Y}) = 0, \tag{15}$$

irrespectively of the dependence structure between $\mathbf{X}$ and $\mathbf{Y}$. On the other hand, Székely et al. (2013) proved that for fixed n

$$\lim_{p,q \to \infty} \hat{R}(\mathbf{X}, \mathbf{Y}) = 1, \tag{16}$$

provided that the coordinates of $\mathbf{X}$ and $\mathbf{Y}$ are i.i.d. and the second moments of $\mathbf{X}$ and $\mathbf{Y}$ exist. Comparison of (15) and (16) reveals that there are actually two different effects, which complicate the interpretation of distance

correlation in high dimension. Equation (15) shows that the *population* distance correlation is close to 0 when comparing a low-dimensional and a high-dimensional vector. Equation (16 shows that the *sample* distance correlation is close to 1 when comparing two high-dimensional vectors. These two problems may occur simultaneously, when comparing two high-dimensional vectors with substantially distinct dimensions. In particular, (15) shows that we can find sequences $(p_k)_{k \in \mathbb{N}}, (q_k)_{k \in \mathbb{N}}$, such that $p_k \to \infty$, $q_k \to \infty$ and

$$\lim_{k \to \infty} R(\mathbf{X}, \mathbf{Y}) = 0, \qquad \lim_{k \to \infty} \hat{R}(\mathbf{X}, \mathbf{Y}) = 1,$$

where for any k, $\mathbf{X}$ and $\mathbf{Y}$ are p_k and q_k-dimensional standard normal vectors and n is fixed. When $p = q$, the population distance correlation $R(\mathbf{X}, \mathbf{Y})$ can attain both the values 0 ($\mathbf{X}$ and $\mathbf{Y}$ independent) and 1 ($\mathbf{Y} = \alpha + b\,C\,\mathbf{X}$ with $\alpha \in \mathbb{R}^p$, $b \in \mathbb{R}$, C orthogonal). In view of (16) we can apply an unbiased estimator for the distance covariance as described in Székely et al. (2013). This has the further advantage that transformation of this estimator converges to a Student-t distribution with $n(n-3)/2$ degrees of freedom, which enables a distance correlation t-test for independence in high dimensions. When p and q differ (but are both high-dimensional), the t-test may still be applied, however a direct interpretability is critical (recall equation (15)).

4 Auto - Distance Covariance Function

4.1 Basic definitions

The distance covariance methodology discussed in Section 2 is based on the assumption that the observations are i.i.d. However, in many practical problems this assumption is violated. Rémillard (2009) proposed an extension of the distance covariance methodology to non-i.i.d. observations, especially time series data, for measuring serial dependence. In this section, we review and compare a number of related dependence measures that have appeared in the time series literature. Throughout the remainder of the paper, we denote by bold $\{\mathbf{X}_t\}$ a d-dimensional stationary time series process, with components $\{X_{t;i}\}_{i=1}^d$. Define further the joint characteristic function of $\mathbf{X}_t$ and $\mathbf{X}_{t+j}$ by $\phi_j(\mathbf{u}, \mathbf{v}) = E\left[\exp\left(i(\mathbf{u}'\mathbf{X}_t + \mathbf{v}'\mathbf{X}_{t+j})\right)\right]$ and the marginal characteristic function of $\mathbf{X}_t$ by $\phi(\mathbf{u}) = E\left[\exp\left(i\mathbf{u}'\mathbf{X}_t\right)\right]$, with $(\mathbf{u}, \mathbf{v}) \in \mathbb{R}^{2d}$, $i^2 = -1$ and $j \in \mathbb{Z}$. The above definitions include the case of univariate time series; that is $d = 1$ and $(u, v) \in \mathbb{R}^2$.

Zhou (2012) introduced the definition of distance covariance in time series literature. In particular, extending Székely et al.'s (2007) definition, Zhou (2012) defined the auto-distance covariance function (ADCV) as the posi-

tive square root of (1), between the joint and the product of the marginal characteristic functions of $\mathbf{X}_t$ and $\mathbf{X}_{t+j}$. However, Hong (1999) had already defined a similar measure of dependence that captures all types of pairwise dependencies. In what follows, we explain how this work is related to Székely et al.'s (2007) work for the definition of distance covariance in time series.

For a strictly stationary α-mixing univariate time series $\{X_t\}$, Hong (1999) defined a new measure of dependence between the joint characteristic function of X_t and its lagged observation X_{t+j} and the product of their marginals, namely

$$\sigma_j(u,v) = \text{Cov}\big(e^{iuX_t}, e^{ivX_{t+j}}\big) = \phi_j(u,v) - \phi(u)\phi(v), \tag{17}$$

where $j \in \mathbb{Z}$. Given that the joint characteristic function factorizes into the product of its marginals under independence of X_t and X_{t+j}, $\sigma_j(u,v)$ equals 0 if and only if X_t and X_{t+j} are independent. Thus, compared to the classical autocorrelation function (ACF), (17) can capture all pairwise dependencies including those with zero autocorrelation. Note that various measures of dependence can be based on (17) (recall (1)) by considering symmetric weighted integrals of the form (see also Davis et al. (2018))

$$\int |\sigma_j(u,v)|^2 \, \omega(u,v) du dv,$$

for a suitable function $\omega(\cdot,\cdot)$. For instance, consider $\omega(u,v) = \phi(u)\phi(v)$ where $\phi(\cdot)$ is the density of the standard normal. Motivated by the work of Székely et al. (2007), Zhou (2012) also defined the so-called auto-distance covariance function of $\{X_t\}$ as the positive square root of

$$V_X^2(j) \;=\; \frac{1}{\pi^2} \int_{\mathbb{R}^2} \frac{|\sigma_j(u,v)|^2}{|u|^2 \, |v|^2} du dv, \quad j \in \mathbb{Z}. \tag{18}$$

From the above definition, we can observe that $V_X^2(j)$ equals 0 if and only if X_t and X_{t+j} are independent. Rescaling (18), we define the auto-distance correlation function (ADCF) as the positive square root of

$$R_X^2(j) = \begin{cases} \dfrac{V_X^2(j)}{V_X^2(0)}, & V_X^2(0) \neq 0; \\ 0, & V_X^2(0) = 0. \end{cases} \tag{19}$$

Clearly, $R_X^2(j)$ equals zero if and only if X_t and X_{t+j} are independent. Thus, the ADCF is suitable for exploring possible nonlinear dependence structures in time series that are not detected by the ACF. This is better understood
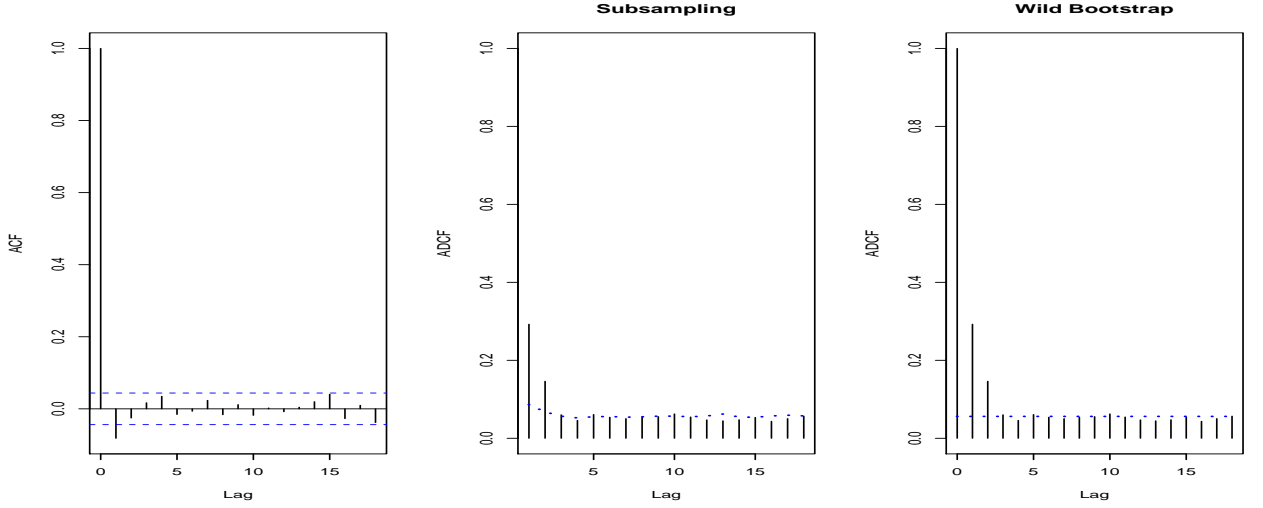
Figure 2: Comparison of the sample ADCF and sample ACF. The data are generated by the NMA(2) model given by (20). Results are based on sample size $n = 2000$.

by examining Figure 2. The plot presents the ACF and ADCF of a second order nonlinear moving average process (NMA(2)) defined by

$$X_t = \epsilon_t \epsilon_{t-1} \epsilon_{t-2}, \tag{20}$$

where $\{\epsilon_t\}$ is a sequence of i.i.d. standard normal random variables. It is well known that the process $\{X_t\}$ consists of uncorrelated but two-dependent random variables. The white noise property is clearly discovered by the ACF, whereas the dependence structure is only detected by the ADCF. We note that the shown critical values of the ADCF plots (second column) are the pairwise 95% critical values that correspond to the pairwise independence test. They are computed via the subsampling approach suggested by Zhou (2012, Section 5.1), where the choice of the block size is based on the minimum volatility method proposed by Politis et al. (1999, Section 9.4.2). Moreover, the 95% simultaneous critical values, for testing whether the data form an iid sequence, are also shown in the sample ADCF plots (last column). They are computed via the wild bootstrap method (Dehling and Mikosch, 1994; Shao, 2010; Leucht and Neumann, 2013) following the algorithm presented in Fokianos and Pitsillou (2018).

4.2   Estimation of ADCV

The sample ADCV, denoted by $\widehat{V}_X(\cdot)$, is defined analogously as in Section 3.1, with the second sample replaced by the lagged observation of the time series, that is by letting $Y_t = X_{t+j}$ for $j \geq 0$. In particular, by defining the

$(n-j) \times (n-j)$ pairwise Euclidean distance matrices by $(a_{rl}) = |X_r - X_l|$ and $(b_{rl}) = |Y_r - Y_l|$, the sample ADCV is given by

$$\widehat{V}_X^2(j) \quad = \quad \frac{1}{(n-j)^2} \sum_{r,l=1}^{n-j} A_{rl} B_{rl}. \tag{21}$$

where $A_{rl} = a_{rl} - \bar{a}_{r.} - \bar{a}_{.l} + \bar{a}_{..}$ and $B_{rl} = b_{rl} - \bar{b}_{r.} - \bar{b}_{.l} + \bar{b}_{..}$, with $\bar{a}_{r.} = \left(\sum_{l=1}^{n-j} a_{rl}\right)/(n-j)$, $\bar{a}_{.l} = \left(\sum_{r=1}^{n-j} a_{rl}\right)/(n-j)$ and $\bar{a}_{..} = \left(\sum_{r,l=1}^{n-j} a_{rl}\right)/(n-j)^2$. Similarly, we define the quantities $b_{r.}, b_{.l}$ and $b_{..}$. If $j < 0$, then $\widehat{V}_X^2(j) = \widehat{V}_X^2(-j)$. Replacing (21) into (19), the sample ADCF is the nonnegative number defined by the square root of

$$\widehat{R}_X^2(j) = \begin{cases} \dfrac{\widehat{V}_X^2(j)}{\widehat{V}_X^2(0)}, & \widehat{V}_X^2(0) \neq 0; \\ 0, & \widehat{V}_X^2(0) = 0. \end{cases}$$

The main properties of ADCV and ADCF and their empirical counterparts are summarized below:

1. Under the minimal conditions of a strictly stationary α-mixing process and existence of first moment, Fokianos and Pitsillou (2017) obtained the strong consistency of $\widehat{V}_X^2(\cdot)$ to its population counterpart for a fixed lag j. Zhou (2012) obtained the weak consistency of $\widehat{V}_X(\cdot)$ requiring $(1 + r_0)$;th moment of the process for some $r_0 > 0$ by assuming short range dependence condition. Davis et al. (2018) also proved strong consistency for ADCV, under various choices of finite and infinite weight functions $\omega(\cdot, \cdot)$. In addition, Fokianos and Pitsillou (2018) show strong consistency of $\widehat{V}_X^2(\cdot)$ by assuming existence of second order moments of X_t. The proof is simpler when compared to other proofs but at the cost of assuming $\mathrm{E}|X_t|^2 < \infty$.

2. $\widehat{V}_X^2(j) \geq 0$ a.s., for $j \in \mathbb{Z}$ where the equality holds under independence.

3. If X_t has finite first moments, then $0 \leq R_X(j) \leq 1$. $R_X(j) = 0$ if and only if X_t and X_{t+j} are independent and $R_X(j) = 1$ if X_t and X_{t+j} are linearly related by an orthogonal matrix.

4. If $\{X_t\}$ is a univariate Gaussian zero-mean stationary time series process with $\mathrm{Var}(X_t) = 1$ and $\rho(j) = \mathrm{Cov}(X_t, X_{t+j})$ then $R_X(j) \leq |\rho(j)|$ and

$$R_X^2(j) = \frac{\rho(j) \arcsin\rho(j) + \sqrt{1 - \rho^2(j)} - \rho(j) \arcsin\rho(j)/2 - \sqrt{4 - \rho^2(j)} + 1}{1 + \pi/3 - \sqrt{3}}, \quad j \in \mathbb{Z}.$$

4.3 *Extensions of ADCV*

Zhou (2012) defined the distance correlation coefficient to explore temporal nonlinear dependence structure in multivariate time series. However, his approach is missing the interrelationships between the various time series components. For a strictly stationary d-dimensional time series $\{\mathbf{X}_t\}$, Zhou (2012) defined

$$V_{\mathbf{X}}^2(j) \quad = \quad \frac{1}{c_d^2} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{|\sigma_j(\mathbf{u}, \mathbf{v})|^2}{|\mathbf{u}|^{d+1} |\mathbf{v}|^{d+1}} d\mathbf{u} d\mathbf{v}, \quad j \in \mathbb{Z}. \tag{22}$$

When testing for independence for multivariate time series, a natural extension consists of defining the distance correlation appropriately for examining whether there exists some inherent nonlinear interdependence between the component series. Fokianos and Pitsillou (2018) investigated this problem by defining the *pairwise auto-distance covariance and correlation matrices*, and extended the theory described in Section 5.2 to multivariate processes. In particular, define an analogous dependence measure to (17) by

$$\sigma_j^{(r,m)}(u, v) \quad = \quad \phi_j^{(r,m)}(u, v) - \phi^{(r)}(u)\phi^{(m)}(v), \quad j \in \mathbb{Z},$$

where $\phi_j^{(r,m)}(u, v)$, $\phi^{(r)}(u)$ and $\phi^{(m)}(v)$ are the joint characteristic function and the marginal characteristic functions of $X_{t;r}$ and $X_{t+j;m}$ respectively, for $r, m = 1, \ldots, d$. Then the pairwise auto-distance covariance function between $X_{t;r}$ and $X_{t+j;m}$ as the positive square root of

$$V_{rm}^2(j) \quad = \quad \int_{\mathbb{R}^2} \left| \sigma_j^{(r,m)}(u, v) \right|^2 \omega(u, v) du dv, \quad j \in \mathbb{Z}, \tag{23}$$

where $\omega(\cdot, \cdot)$ is given in (3). The auto-distance covariance matrix of $\{\mathbf{X}_t\}$ at a specific lag j, denoted by $V(j)$, is the $d \times d$ matrix

$$V(j) = \left[V_{rm}(j) \right]_{r,m=1}^d, \quad j \in \mathbb{Z}.$$

The pairwise auto-distance correlation function between $X_{t;r}$ and $X_{t+j;m}$, say $R_{rm}(j)$, is then defined by rescaling (23)

$$R_{rm}^2(j) \quad = \quad \frac{V_{rm}^2(j)}{\sqrt{V_{rr}^2(0)}\sqrt{V_{mm}^2(0)}},$$

for $V_{rr}(0)V_{mm}(0) \neq 0$ and zero otherwise. The auto-distance correlation matrix at lag j, denoted by $R(j)$, is the

$d \times d$ matrix

$$R(j) = \left[R_{rm}(j) \right]_{r,m=1}^{d}, \quad j \in \mathbb{Z}.$$

The pairwise distance correlation coefficient can be employed for exploring complex nonlinear interdependence structures that can not be discovered by the classical cross-correlation coefficient.

The sample quantity of (23) is obtained analogously as in Section 4.2. In particular, for $j \geq 0$ consider a sample $(X,Y) = \{(X_{t;r}, X_{t+j;m})\}$ for $r, m = 1, 2, \ldots, d$. Define the Euclidean distance matrices by $(a_{ts}^r) = |X_t - X_s|$ and $(b_{ts}^m) = |Y_t - Y_s|$ and the double centered distance matrices by

$$A_{ts}^r = a_{ts}^r - \bar{a}_{t.}^r - \bar{a}_{.s}^r + \bar{a}_{..}^r, \quad B_{ts}^m = b_{ts}^m - \bar{b}_{t.}^m - \bar{b}_{.s}^m + \bar{b}_{..}^m,$$

where the quantities in the right hand side are defined analogously as those defined in Section 4.2. Thus, the sample counterpart of $V_{rm}(\cdot)$ is the positive square root of

$$\widehat{V}_{rm}^2(j) \quad = \quad \frac{1}{(n-j)^2} \sum_{t,s=1}^{n-j} A_{ts}^r B_{ts}^m.$$

If $j < 0$, set $\widehat{V}_{rm}^2(j) = \widehat{V}_{mr}^2(-j)$. The sample distance covariance matrix is then given by

$$\widehat{V}(j) = \left[\widehat{V}_{rm}(j) \right]_{r,m=1}^{d}, \quad j \in \mathbb{Z}.$$

The main properties of $V_{rm}(\cdot)$ and $R_{rm}(\cdot)$ and their sample counterparts are outlined below:

- For all $j \in \mathbb{Z}$, the diagonal elements $\left(V_{rr}(j) \right)_{r=1}^{d}$ correspond to the auto-distance covariance function of $\{X_{t;r}\}$ and they express the dependence among the pairs $\left(X_{t;r}, X_{t+j;r} \right)$, $r = 1, \ldots, d$.

- The off-diagonal elements $\left(V_{rm}(0) \right)_{r,m=1}^{d}$ measure the concurrent dependence between $\{X_{t;r}\}$ and $\{X_{t;m}\}$. If $V_{rm}(0) > 0$, $\{X_{t;r}\}$ and $\{X_{t;m}\}$ are concurrently dependent.

- For $j \in \mathbb{Z} - \{0\}$, $\left(V_{rm}(j) \right)_{r,m=1}^{d}$ measures the dependence between $\{X_{t;r}\}$ and $\{X_{t+j;m}\}$. If $V_{rm}(j) = 0$ for all $j \in \mathbb{Z} - \{0\}$, then $\{X_{t+j;m}\}$ does not depend on $\{X_{t;r}\}$.

- For all $j \in \mathbb{Z}$, $V_{rm}(j) = V_{mr}(j) = 0$ implies that $\{X_{t;r}\}$ and $\{X_{t+j;m}\}$ are independent. Moreover, for all $j \in \mathbb{Z} - \{0\}$, if $V_{rm}(j) = 0$ and $V_{mr}(j) = 0$ then $\{X_{t;r}\}$ and $\{X_{t;m}\}$ have no lead-lag relationship.

- If for all $j > 0$, $V_{rm}(j) = 0$ but there exists some $j > 0$ such that $V_{mr}(j) > 0$, then $\{X_{t;m}\}$ does not depend on any past values of $\{X_{t;r}\}$, but $\{X_{t;r}\}$ depends on some past values of $\{X_{t;m}\}$.

As a closing remark, Davis et al. (2018) applied distance covariance methodology to stationary univariate and multivariate time series to study serial dependence under various choices of the weight function $\omega(\cdot, \cdot)$. One of their main results was to show that the asymptotic distribution of the empirical ADCV when applied to the residuals of a fitted autoregressive process converges to a random variable whose distribution differs from that of ordinary ADCV.

5   Test Statistics for Pairwise Dependence in Time Series

5.1   Introduction

There are many available tests for testing serial independence in time series literature using both time domain and frequency domain methodologies, see Tjøstheim (1996). For a univariate time series, the most well known testing procedures are mainly based on the serial correlation coefficient including those by Box and Pierce (1970)

$$\text{BP} = n \sum_{j=1}^{p} \hat{\rho}^2(j),$$

and Ljung and Box (1978)

$$\text{LB} = n(n+2) \sum_{j=1}^{p} (n-j)^{-1} \hat{\rho}^2(j),$$

among others, where $\rho(j) = \text{Cor}(X_t, X_{t+j})$ for $j \in \mathbb{Z}$ and $\hat{\rho}(j)$ denotes its empirical analogue (Brockwell and Davis, 1991, pg.29). Most of these procedures were extended to the multivariate case; see Mahdi (2011) for an overview and a newly proposed correlation-based test. The multivariate Ljung-Box test statistic Hosking (1980),Li and McLeod (1981), Li (2004)

$$\text{mLB} = n^2 \sum_{j=1}^{p} (n-j)^{-1} \text{trace}\{\widehat{\Gamma}'(j)\widehat{\Gamma}^{-1}(0)\widehat{\Gamma}(j)\widehat{\Gamma}^{-1}(0)\},$$

is widely used for testing $\Gamma(1) = \cdots = \Gamma(p) = 0$, where $\Gamma(j) = E\left[(\mathbf{X}_{t+j} - \boldsymbol{\mu})(\mathbf{X}_t - \boldsymbol{\mu})'\right]$ is the auto-covariance matrix of $\{\mathbf{X}_t\}$ and $\widehat{\Gamma}(j)$ denotes the sample autocovariance matrix (see Brockwell and Davis (1991, pg. 397)). Although these tests perform well under linear and Gaussian processes, they behave poorly against general types of nonlinear dependencies including those with zero autocorrelation (ARCH, bilinear, nonlinear moving average processes); see Romano and Thombs (1996) and Shao (2011). Tjøstheim (1996) pointed out that

a standard procedure for increasing the power of the correlation based tests in ARCH type models is to compute the correlation of the squared observations. However, this strategy leads to a loss of power compared to the ordinary correlation based tests. Robinson (1991) gave a variety of test statistics for testing serial correlation under the presence of conditional heteroskedasticity, whereas Escanciano and Lobato (2009) further proposed an automatic Portmanteau test that allows for nonlinear dependencies and automatic selection of lag order p.

5.2   *Asymptotic tests based on ADCV*

Under the null hypothesis of pairwise independence between $\mathbf{X}_t$ and $\mathbf{X}_{t+j}$, Zhou (2012) obtained the asymptotic distribution of the sample ADCV at fixed lag j and shows that it has identical limiting distribution as that obtained by Székely et al. (2007) for the case of independent data (recall (8)). Under the alternative hypothesis that $\mathbf{X}_t$ and $\mathbf{X}_{t+j}$ are dependent, $n\widehat{V}_{\mathbf{X}}^2(j)$ tends to infinity. However, Zhou (2012) studied the behavior of ADCV in multivariate time series at a fixed lag. Fokianos and Pitsillou (2017) relaxed this assumption by defining the ADCV in the framework of generalized spectral domain following Hong's (1999) methodology. It can be shown that serial dependence in time series can be characterized by the generalized spectral density function. In fact, it is well known that any deviation of the generalized spectral density from uniformity is a strong evidence of serial dependence. Hong (1999) confirmed these facts by showing that the standard spectral density approach becomes inappropriate for non-Gaussian and nonlinear processes with zero autocorrelation. To bypass such problems he proposes a generalized spectral density that is defined based on (17), that is

$$f(\omega, u, v) \quad = \quad \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \sigma_j(u, v) e^{-ij\omega}, \quad \omega \in [-\pi, \pi],$$

provided that $\sup_{(u,v) \in \mathbb{R}^2} \sum_j \left| \sigma_j^{(r,m)}(u, v) \right| < \infty$. This 'generalized' spectral density can be defined for both linear and nonlinear processes and captures all pairwise dependencies while preserving the nice properties of the standard spectrum. The null hypothesis of pairwise independence can be interpreted in terms of the generalized spectral density; i.e. testing whether

$$f(\omega, u, v) = f_0(\omega, u, v) = \frac{\sigma_0(u, v)}{2\pi}.$$

To test for serial dependence, Hong (1999) compared the Parzen's (1957) kernel-type estimators of $\hat{f}(\omega, u, v)$ and $\hat{f}_0(\omega, u, v)$ by the L_2-norm. The final test statistic takes the form

$$\text{H}_{99} \quad = \quad \int_{\mathbb{R}^2} \sum_{j=1}^{n-1} (n-j)k^2(j/p) \, |\hat{\sigma}_j(u, v)|^2 \omega(u, v) du dv, \tag{24}$$

where p is a bandwidth or lag order and $k(\cdot)$ is a Lipschitz-continuous kernel function with the following properties: $k : \mathbb{R} \to [-1, 1]$ is symmetric and is continuous at 0 and all but a finite number of points, with $k(0) = 1$, $\int_{-\infty}^{\infty} k^2(z)dz < \infty$ and $| k(z) | \leq C \, | z |^{-b}$ for large z and $b > 1/2$.

Distance measures obtained by characteristics functions have been also considered by Pinkse (1998) and Hong (1999). Hong (1999, pg.1206) explained how tests based on the empirical characteristic function may have omnibus power against tests based on the empirical distribution function. In addition to the test of serial dependence, Hong (1999) proposed testing other aspects of dependence after substituting $\hat{\sigma}_j(u, v)$ in (24) by its derivatives $\hat{\sigma}_j^{(m,l)}(u, v) = \partial^{m+l}\hat{\sigma}_j(u, v)/\partial^m u \partial^l v$. Different choices of the indices (m, l) and the weight function $\omega(u, v) = \omega_1(u)\omega_2(v)$ deliver a wide range of various hypotheses tests. For instance, choosing $(m, l) = (1, 1)$ and both $\omega_1(u), \omega_2(v)$ having the Dirac delta function, yields a test statistic for serial correlation. To test for the martingale hypothesis which implies that the best predictor of future values of a time series, in the sense of least mean squared error, is simply the current value of the time series, put $(m, l) = (1, 0)$ and assume that $\omega_1(u)$ has a Dirac density and $\omega_2(v)$ is an arbitrary increasing function that is integrable. For these latter choices of $\omega_1(\cdot)$ and $\omega_2(\cdot)$ and for the choices of $(m, l) = (3, 0)$ and $(m, l) = (4, 0)$, we can test for conditional symmetry and conditional heterokurtosis respectively; see Hong (1999) for more details.

Under the null hypothesis of independence, Hong (1999) proved that a standardized version of H_{99} given in (24) is asymptotically standard normally distributed, provided that $p = cn^\lambda$, for $c > 0$ and $\lambda \in (0, 1)$. He further proved that the asymptotic power of the derived test statistic approaches unity; hence the test statistic is consistent. Recalling (22), one can observe that the statistic H_{99} proposed by Hong (1999) is similar to that of Székely et al. (2007) with the main differences being the restriction to the univariate case and the choice of the weight function. Moreover, the number of lags used in the construction of H_{99} is not fixed but increases with the sample size of the process. Fokianos and Pitsillou (2017) built a bridge between these two methodologies. They have chosen $\omega(\cdot, \cdot)$ of the form of (3), resulting in a test statistic based on the ADCV

$$\text{FP} \quad = \quad \sum_{j=1}^{n-1} (n-j)k^2(j/p)\widehat{V}_X^2(j). \tag{25}$$

Clearly, the statistic FP is a portmanteau type statistic, with each squared distance covariance being multiplied by a weight determined by a kernel function $k(\cdot)$. Under the null hypothesis that the process X_t forms an i.i.d. sequence, a suitable standardization of FP was proved to follow asymptotically a standard normal distribution, provided that $p = cn^\lambda$, for $c > 0$ and $\lambda \in (0, 1)$. In addition, considering any type of pairwise dependence for the process, the authors showed that the proposed test statistic is asymptotically consistent. The latter test statistic is faster to compute than the one proposed by Hong (1999) since the authors avoid a two dimensional integration. Although both test statistics, FP and H$_{99}$ given in (25) and (24) respectively, are defined for univariate α-mixing processes, the proposed methodologies can be extended for multivariate processes as explained in the following.

Following a similar methodology for multivariate time series and considering the generalized spectral density matrix, Fokianos and Pitsillou (2018) proposed a test statistic based on the pairwise auto-distance covariance function

$$\widetilde{\text{FP}} = \sum_{r,m} \sum_{j=1}^{n-1} (n-j) k^2(j/p) \widehat{V}_{rm}^2(j) = \sum_{j=1}^{n-1} (n-j) k^2(j/p) \text{trace}\{\widehat{V}^*(j)\widehat{V}(j)\},$$

where $\widehat{V}^*(\cdot)$ denotes the conjugate transpose of the sample distance covariance matrix $\widehat{V}(\cdot)$. Under the null hypothesis of independence, a standardized version of $\widetilde{\text{FP}}$ follows asymptotically a standard normal distribution, provided that the numbers of lags employed in the construction of $\widetilde{\text{FP}}$ increases with the sample sizes, whereas under any type of alternative hypothesis the proposed test statistic has asymptotic power one.

Chen and Hong (2012) developed a nonparametric test for the Markov property of a multivariate time series based on the conditional characteristic function. Based on the fact that the correlation coefficient is suitable for Gaussian data but it fails for nonlinear cases, many authors defined the correlation function in a local sense, including Tjøstheim and Hufthammer (2013), Berentsen and Tjøstheim (2014), Støve et al. (2014) and Støve and Tjøstheim (2014) among others. In the context of time series, Lacal and Tjøstheim (2017) defined a new measure of dependence, the so-called local Gaussian autocorrelation that works well for nonlinear models. The authors compared the proposed test statistic to distance covariance function and found that they both work in a similar way.

5.3 Other test statistics

Many studies in the literature have considered the problem of measuring dependence between X_t and $X_{t+j} \equiv Y_t$, in terms of the distance between the bivariate distribution of (X_t, Y_t), $F_{X;Y}(x, y) = P(X_t \leq x, Y_t \leq y)$, and the product of their marginal distribution functions, $F_X(x) = P(X_t \leq x)$ and $F_Y(y) = P(Y_t \leq y)$. Well known

distance measures for distribution functions are the Kolmogorov-Smirnov distance

$$D_1(j) = \sup_{(x,y)\in\mathbb{R}^2} |F_{X;Y}(x,y) - F_X(x)F_Y(y)| \tag{26}$$

and the Cramer-von Mises type distance

$$D_2(j) = \int_{\mathbb{R}^2} \left\{ F_{X;Y}(x,y) - F_X(x)F_Y(y) \right\}^2 dF_{X;Y}(x,y). \tag{27}$$

Replacing the theoretical distribution functions with their empirical analogues

$$\widehat{F}_{X;Y}(x,y) = \frac{1}{(n-j)} \sum_{t=1}^{n-j} \mathbb{I}(X_t \le x, Y_t \le y),$$

$$\widehat{F}_X(x) = \frac{1}{(n-j)} \sum_{t=1}^{n-j} \mathbb{I}(X_t \le x), \tag{28}$$

where $\mathbb{I}(\cdot)$ is the indicator function, we form the corresponding empirical estimates. Similar distance measures are obtained by employing density functions instead of distribution functions (see Skaug and Tjøstheim (1996) and Bagnato et al. (2014) for an overview).

Skaug and Tjøstheim (1993) extended the work by Blum et al. (1961) and considered the asymptotic behavior of the Cramer-von Mises type statistic (27) at lag j, under ergodicity of $\{X_t\}$. Moreover, they constructed a test for pairwise independence among pairs $(X_t, X_{t+1}), (X_t, X_{t+2}), \ldots, (X_t, X_{t+p})$ using the statistic

$$\mathrm{ST} = \sum_{j=1}^{p} \widehat{D}_2(j)$$

where p is a fixed constant denoting the maximum lag order employed for the test. Under the null hypothesis that $\{X_t\}$ consists of i.i.d. random variables, $n\,\mathrm{ST}$ converges in distribution (as $n \to \infty$) to an infinite mixture of chi-squared random variables. However, the latter test statistic is consistent against serial dependencies up to a finite order p. From a practical point of view, this may be restrictive since the actual serial dependence may be detected at lags larger than p. Motivated by the approach introduced by Skaug and Tjøstheim (1993), Hong (1998) proposed a statistic where the number of lags increases with the sample size and different weights are given to different lags; in other words let

$$\mathrm{H}_{98} = \sum_{j=1}^{n-1} (n-j)k^2(j/p)\widehat{D}_2^2(j)$$

where $k(\cdot)$ is a kernel function with properties similar to the kernel function of (24). He further proved that after proper standardization and for large p, provided that $p = cn^\lambda$, for $c > 0$ and $\lambda \in (0, 1)$, the statistic is asymptotically standard normally distributed.

In general, incorporating a large number of lags in the asymptotic theory is an issue nicely addressed by the frequency domain framework. Hong (1996) proposed three test statistics for testing serial independence in univariate time series by considering a kernel-based standardized spectral density

$$\hat{f}(\omega) = \frac{1}{2\pi} \sum_{j=-(n-1)}^{(n-1)} k(j/p)\hat{\rho}(j)\cos(j\omega), \quad \omega \in [-\pi, \pi]$$

where the notation follows the notation of (24) and $\hat{\rho}(j)$ is the sample autocorrelation function of $\{X_t\}$ at lag j. Comparison of $\hat{f}(\omega)$ to $f_0(\omega) = 1/2\pi$ (spectral density of white noise process) via divergence measures yields test statistics that can be employed for testing that the process $\{X_t\}$ is white noise. In particular, he employs the quadratic norm, the Hellinger metric and the Kullback-Leibler information criterion, i.e.

$$
\begin{aligned}
T_{1n} &= \left[2\pi \int_{-\pi}^{\pi} \left\{ \hat{f}(\omega) - f_0(\omega) \right\}^2 d\omega \right]^{1/2}, \\
T_{2n} &= \left[\int_{-\pi}^{\pi} \left\{ \hat{f}^{1/2}(\omega) - f_0^{1/2}(\omega) \right\}^2 d\omega \right]^{1/2}, \\
T_{3n} &= -\int_{\omega:\hat{f}(\omega)>0} \ln\left(\hat{f}(\omega)/f_0(\omega) \right) f_0(\omega) d\omega,
\end{aligned}
$$

respectively. A closed form expression for T_{1n} is given by

$$\mathrm{H}_{96} = n \sum_{j=1}^{n-1} k^2(j/p)\hat{\rho}^2(j).$$

Compared to the classical portmanteau statistics, like BP and LB, all these suggested statistics, properly standardized, are asymptotically standard normally distributed and can be developed without imposing a specific alternative model. Xiao and Wu (2014) proved that the asymptotic null distribution of H_{96} in its standardized form, remains intact under the presence of serial correlation, whereas Shao (2011) proved its robustness to conditional heteroscedasticity. Although Hong's (1996) tests incorporate an increasing number of lags, they are based on the classical correlation coefficient and are not consistent against all pairwise dependencies of unknown form. To avoid this problem, Hong (2000) generalizes both the Cramer-von Mises and Kolmogorov-Smirnov type test statistics

by employing the generalized spectral theory. Define a dependence measure analogous to (17) by

$$\rho_j^*(x,y) = F_{X;Y}(x,y) - F_X(x)F_Y(y). \tag{29}$$

Clearly, the distance dependence measure $\rho^*(\cdot,\cdot)$ defined in (29) vanishes only in the case where X_t and $Y_t \equiv X_{t+j}$ are independent, leading to the observation that $\rho_j^*(\cdot,\cdot)$ can capture all pairwise dependencies. Replacing the theoretical distribution functions with the empirical ones given in (28), one can find the empirical analogue of (29), $\hat{\rho}_j^*(\cdot,\cdot)$. Clearly, Skaug and Tjøstheim's (1993) approach is based on (29).

In a recent work, Dette et al. (2015) introduced a "new" spectrum as the Fourier transform of the so-called copula cross-covariance kernel

$$\rho_j^U(\tau_1,\tau_2) := \text{Cov}\Big(\mathbb{I}(U_t \leq \tau_1), \mathbb{I}(U_{t+j} \leq \tau_1)\Big), \tag{30}$$

where $(\tau_1, \tau_2) \in (0,1)^2$ and $U_t := F(X_t)$. They proposed to estimate the corresponding spectral densities associated with measures (29) and (30) via Laplace periodogram (Li, 2008). In addition, they highlighted that in the case of using (30), replacing the original observations with their ranks, potentially preserves the invariance property with respect to transformations of the marginal distributions.

5.3.1   Multivariate Extension

All univariate testing methodologies presented in this section can be extended to the multivariate case by considering the distances between $X_{t;r}$ and $Y_{t;m} \equiv X_{t+j;m}$ for $r, m = 1, \ldots, d$. The idea is to compare the pairwise bivariate distribution function of $(X_{t;r}, Y_{t;m})$, $F_{X_r;Y_m}(x,y) = P(X_{t;r} \leq x, Y_{t;m} \leq y)$, and the product of their marginal distribution functions, $F_{X_r}(x) = P(X_{t;r} \leq x)$ and $F_{Y_m}(y) = P(Y_{t;m} \leq y)$ via the dependence measures (26) and (27). The final test statistics can then be defined by summing over $r, m = 1, \ldots, d$. The multivariate extensions of ST and H$_{98}$ are derived as

$$\widetilde{\text{ST}} = \sum_{r,m=1}^{d} \sum_{j=1}^{p} \widehat{D}_2^{(r,m)}(j)$$

and

$$\widetilde{\text{H}_{98}} = \sum_{r,m=1}^{d} \sum_{j=1}^{n-1} (n-j)k^2(j/p)\Big(\widehat{D}_2^{(r,m)}(j)\Big)^2,$$

Table 1: Achieved type I error of the test statistics for testing the hypothesis that the data are i.i.d. The data are generated by the standard normal distribution. Significance level is set to $\alpha = 0.05$. Achieved significance levels are given in percentages. The results are based on $b = 499$ bootstrap replications and 1000 simulations.

n	p	BP	LB	H_{96}	H_{98}	H_{99}	ST	FP
100	5	4.5	5.0	4.9	4.6	4.2	6.0	4.8
	8	4.9	5.2	5.1	5.8	4.3	5.9	4.7
	12	4.0	5.6	5.6	3.9	4.1	3.4	4.0
500	6	3.5	5.4	5.6	5.8	4.5	4.8	4.3
	11	4.1	4.4	5.0	5.2	5.6	6.1	5.0
	20	3.3	5.5	4.1	5.1	5.7	5.1	4.8

respectively, where $\widehat{D}_2^{(r,m)}(\cdot)$ is the Cramer-von Mises type distance given by

$$D_2^{(r,m)}(j) = \int_{\mathbb{R}^2} \left\{ F_{X_r;Y_m}(x,y) - F_{X_r}(x)F_{Y_m}(y) \right\}^2 dF_{X_r;Y_m}(x,y).$$

All other univariate test statistics mentioned in this section can be developed along the same lines. Although the development of the asymptotic theory of these test statistics is not in the scope of this review paper, we think that this is a challenging topic for further research.

6   Applications

6.1   Univariate Case

We investigate the finite sample performance of the statistics BP, LB, H_{96}, H_{98}, H_{99}, ST and FP presented in Section 5.2. All simulations performed correspond to different sample sizes ($n = 100, 500$). We employ ordinary bootstrap (with size $b = 499$) to obtain critical values in order to study the size and the power of the statistics. We first examine the size of the tests. Suppose that $\{X_t\}$ is an i.i.d. sequence of standard normal random variables. The spectral density based test statistics, H_{96}, H_{98}, H_{99} and FP, are all computed based on the Bartlett kernel with three different choices of bandwidth $p = [3n^\lambda]$, where $\lambda = 0.1, 0.2$ and 0.3. Table 1 shows the achieved type I error rates at 5% nominal level. All test statistics achieve nominal sizes quite adequately especially for large bandwidth parameters.

For investigating the power of the tests we consider the NMA(2) model given in (20) and the following additional models:

- AR(1) -model

$$X_t = 0.4X_{t-1} + \epsilon_t, \tag{31}$$

- ARCH(2)-model

$$X_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = 0.5 + 0.8 X_{t-1}^2 + 0.1 X_{t-2}^2 \tag{32}$$

where $\{\epsilon_t\}$ is a sequence of i.i.d. standard normal random variables. Note that (31) corresponds to a first-order autoregressive process, (32) corresponds to an autoregressive conditional heteroscedastic model of order two (see Engle (1982)). It is well known that the process $\{X_t\}$ generated by (20) consists of a sequence of dependent but uncorrelated random variables. Table 2 show the power of all test statistics considered for various sample sizes and bandwidth parameters by models (20), (31) and (32) respectively. It is clear that when the data are generated by the nonlinear models NMA(2) and ARCH(2) the test statistics FP and H_{99} perform better, whereas in the case where the data are generated by the linear AR(1) model all test statistics exhibit similar power.

Table 2: Empirical power (%) of all test statistics of size 5%. The bandwidth p is $[3n^\lambda]$, $\lambda = 0.1$, 0.2 and 0.3. The results are based on $B = 499$ bootstrap replications for each of 1000 simulations. The test statistics H_{96}, H_{99}, FP and ST are calculated by employing the Bartlett kernel.

| n		100			200			500	
p	5	8	12	6	9	15	6	11	20
					Model (20)				
BP	23.5	20.5	17.8	27.9	22.7	20.5	35.8	33.3	25.1
LB	22.8	21.8	16.9	27.9	25.7	21.2	35.4	30.2	24.6
H_{96}	36.9	34.7	31.4	42.3	40	36.2	47.5	42.6	37.6
H_{98}	47.4	43.1	37.7	93.9	85.8	75.5	100	100	100
H_{99}	97	82	59	100	100	100	100	100	100
ST	26.1	20	17.5	53.7	43.9	31.2	100	98.3	84.4
FP	89.2	72.8	49.5	99.8	99.6	95	100	100	100
					Model (31)				
BP	88.3	81.2	71.6	99.6	99.6	98.1	100	100	100
LB	87.4	84.2	72	99.3	98.8	98	100	100	100
H_{96}	94.4	94.2	93.2	99.9	100	100	100	100	100
H_{98}	95	93.3	92.5	100	100	100	100	99.8	100
H_{99}	81.5	75.7	68.8	98.6	98	96.1	100	100	100
ST	80.5	76.1	67.1	98.2	97.2	93.3	100	100	100
FP	93.5	87.7	84.8	100	99.8	99.4	100	100	100
					Model (32)				
BP	23.4	19.4	17.3	30.1	24.8	23.5	40.6	34.2	27.1
LB	21.4	18.7	15.9	28.9	24.6	20.7	38.2	35.9	26.4
H_{96}	37.1	35.2	35.6	48	48.2	47.9	64.7	63.1	63.6
H_{98}	24.9	22.9	19.3	40.1	39.7	35.4	89	83.8	72.2
H_{99}	71	74	60	99	96	86	100	100	100
ST	20.8	14.3	14.3	29.1	27.3	20.8	67.3	54.5	39
FP	59.8	54	45.3	89	81.3	76	99.9	100	100
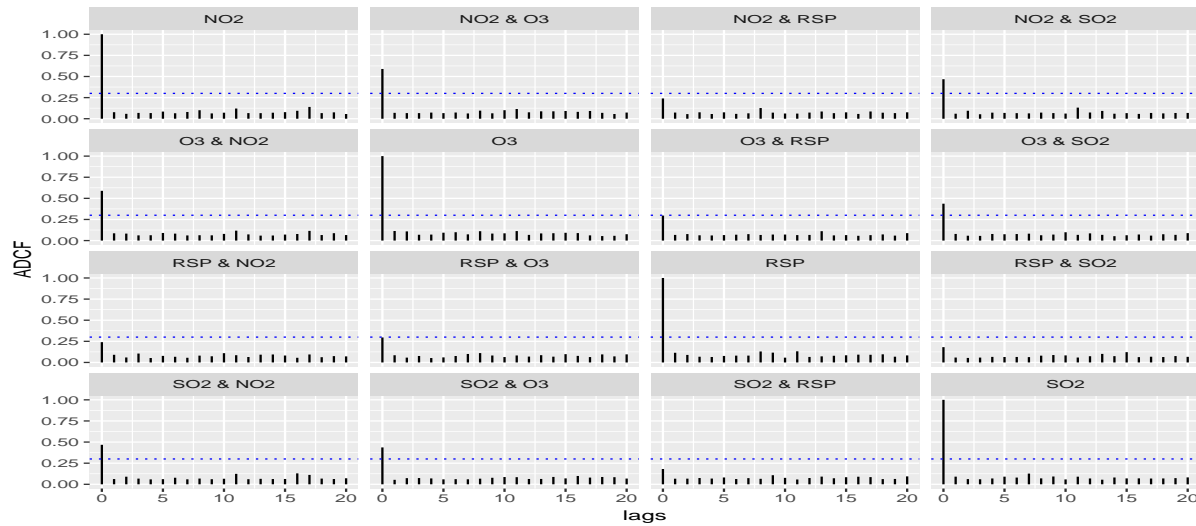
Figure 3: The ADCF plots of the residuals of the VAR(7) model for the four-dimensional air pollution time series data.

6.2 A Data Example

We analyze air pollution data on four pollutants, the sulfur dioxide (SO_2), ozone (O_3), nitrogen dioxide (NO_2) and respirable suspended particulates (RSP) measured daily from the monitoring station at Tsuen Wan in Hong Kong from February 2015 to December 2016. The time series data can be obtained from the Environmental Protection Department of Hong Kong (http://epic.epd.gov.hk/EPICDI/air/station/?lang=en). Each time series consists of 700 observations. We use a log-transformation to the data and applied a first-order differencing on the series to remove seasonal patterns. Assuming that the four dimensional series follow a vector autoregressive model (VAR) and employing the AIC to choose its order, we obtain that a seventh order model would be appropriate for the data. After a VAR(7) model fit, we examine the ADCF plots of the residuals (Figure 3). The plot does not show any evidence of serial dependence. Cnstructing multivariate tests of independence among the residuals, all test statistics used, $\widetilde{\text{FP}}$, mLB, $\widetilde{\text{ST}}$ and $\widetilde{\text{H}_{98}}$, yield large p-values (Table 3) suggesting the sufficiency of this model fit. Some of the methods are implemented in the R package dCovTS–see Pitsillou and Fokianos (2016).

References

Bagnato, L., L. De Capitani, and A. Punzo (2014). Testing serial independence via density-based measures of divergence. *Methodol. Comput. Appl. Probab.* **16**, 627–641.

Table 3: P-values of tests of independence among the residuals after fitting a VAR(7) model to the four dimensional time series. All results are based on $B = 499$ bootstrap replications. Both statistics $\widetilde{FP}$ and $\widetilde{H}_{98}$ are calculated based on the Bartlett kernel.

p	$\widetilde{FP}$	mLB	$\widetilde{ST}$	$\widetilde{H}_{98}$
6	0.332	0.999	0.528	0.584
11	0.368	0.973	0.512	0.546
22	0.104	0.576	0.554	0.128

Bakirov, N. K., M. L. Rizzo, and G. J. Székely (2006). A multivariate nonparametric test of independence. *J. Multivariate Anal.* **97**, 1742–1756.

Bakirov, N. K. and G. J. Székely (2011). Brownian covariance and central limit theorem for stationary sequences. *Theory Probab. Appl. 55*, 371–394.

Balasubramanian, K., B. K. Sriperumbudur, and G. Lebanon (2013). Ultrahigh dimensional feature screening via RKHS embeddings. In *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*, pp. 126–134.

Belu, A. C. (2012). *Multivariate measures of dependence for random variables and Lévy processes*. Ph. D. thesis, Department of Statistics. Case Western Reserve University.

Berentsen, G. D. and D. Tjøstheim (2014). Recognizing and visualizing departures from independence in bivariate data using local Gaussian correlation. *Stat. Comput.* **24**, 785–801.

Blum, J. R., J. Kiefer, and M. Rosenblatt (1961). Distribution free tests of independence based on the sample distribution function. *Ann. Math. Stat.* **32**, 485–498.

Böttcher, B.and Keller-Ressel, M. and R. L. Schilling (2017a). Detecting independence of random vectors i. generalized distance covariance and gaussian covariance. *https://arxiv.org/abs/1711.07778*.

Böttcher, B.and Keller-Ressel, M. and R. L. Schilling (2017b). Detecting independence of random vectors ii. distance multivariance and gaussian multivariance. *https://arxiv.org/abs/1711.07775*.

Box, G. E. P. and D. A. Pierce (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *J. Amer. Statist. Assoc.* **65**, 1509–1526.

Brockwell, P. J. and R. A. Davis (1991). *Time series: theory and methods*. New York: Springer-Verlag. Second Edition.

Chakraborty, S. and X. Zhang (2018). Distance metrics for measuring joint dependence with application to causal inference. https://arxiv.org/abs/1711.09179.

Chen, B. and Y. Hong (2012). Testing for the Markov property in time series. *Econom. Theory* **28**, 130–178.

Davis, R. A., M. Matsui, T. Mikosch, and P. Wan (2018). Applications of distance correlation to time series. *Bernoulli* **4A**, 3087–3116.

Dehling, H. and T. Mikosch (1994). Random quadratic forms and the bootstrap for U-statistics. *J. Multivariate Anal.* **51**, 392–413.

Dette, H., M. Hallin, T. Kley, and S. Volgushev (2015). Of copulas, quantiles, ranks and spectra: An L_1-approach to spectral analysis. *Bernoulli* **21**, 781–831.

Donnelly, C., J. C. M. Andersson, and B. Arheimer (2016). Using flow signatures and catchment similarities to evaluate the E-HYPE multi-basin model across europe. *Hydrological Sciences Journal* **61**, 255–273.

Dueck, J., D. Edelmann, T. Gneiting, and D. Richards (2014). The affinely invariant distance correlation. *Bernoulli* **20**, 2305–2330.

Dueck, J., D. Edelmann, and D. Richards (2015). A generalization of an integral arising in the theory of distance correlation. *Statist. Probab. Lett.* **97**, 116–119.

Dueck, J., D. Edelmann, and D. Richards (2017). Distance correlation coefficients for Lancaster distributions. *J. Multivariate Anal.* **154**, 19–39.

Edelmann, D., D. Richards, and D. Vogel (2017). The distance standard deviation. *https://arxiv.org/abs/1705.05777*.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflations. *Econometrica* **50**, 987–1007.

Escanciano, J. C. and I. N. Lobato (2009). An automatic Portmanteau test for serial correlation. *J. Econometrics* **151**, 140–149.

Fan, J. and J. Lv (2008). Sure independence screening for ultrahigh dimensional feature space. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70**, 849–911.

Feuerverger, A. (1993). A consistent test for bivariate dependence. *Int. Stat. Rev.* **61**, 419–433.

Fiedler, J. (2016). *Distances, Gegenbauer expansions, curls, and dimples: On dependence measures for random fields*. Ph. D. thesis, Department of Applied Mathematics. University of Heidelberg.

Fokianos, K. and M. Pitsillou (2017). Consistent testing for pairwise dependence in time series. *Technometrics* **59**, 262–270.

Fokianos, K. and M. Pitsillou (2018). Testing independence for multivariate time series via the auto-distance correlation matrix. *Biometrika* **105**, 337–352.

Gretton, A., K. Fukumizu, and B. K. Sriperumbudur (2009). Discussion of: Brownian distance covariance. *Ann. Appl. Stat.* **3**, 1285–1294.

Gretton, A., K. Fukumizu, C. H. Teo, L. Song, B. Schölkopf, and A. Smola (2008). A kernel statistical test of independence. *Advances in neural information processing systems* **20**, 585–592.

Gretton, A., R. Herbrich, A. Smola, O. Bousque, and B. Schölkopf (2005). Kernel methods for measuring independence. *J. Mach. Learn. Res.* **6**, 2075–2129.

Guo, X., Y. Zhang, W. Hu, H. Tan, and X. Wang (2014). Inferring nonlinear gene regulatory networks from gene expression data based on distance correlation. *PloS one* **9**, e87446.

Hlávka, Z., M. Hušková, and S. G. Meintanis (2011). Tests for independence in non-parametric heteroscedastic regression models. *J. Multivariate Anal.* **102**, 816–827.

Hong, Y. (1996). Consistent testing for serial correlation of unknown form. *Econometrica* **64**, 837–864.

Hong, Y. (1998). Testing for pairwise serial independence via the empirical distribution function. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60**, 429–453.

Hong, Y. (1999). Hypothesis testing in time series via the empirical characteristic function: A generalized spectral density approach. *J. Amer. Statist. Assoc.* **94**, 1201–1220.

Hong, Y. (2000). Generalized spectral tests for serial dependence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **62**, 557–574.

Hosking, J. R. M. (1980). The multivariate Portmanteau statistic. *J. Amer. Statist. Assoc.* **75**, 602–608.

Huang, C. and X. Huo (2017). A statistically and numerically efficient independence test based on random projections and distance covariance. https://arxiv.org/abs/1701.06054.

Huo, X. and G. J. Székely (2016). Fast computing for distance covariance. *Technometrics* **58**, 435–447.

Kankainen, A. (1995). *Consistent testing of total independence based on the empirical characteristic function*. Ph. D. thesis, University of Jyväskylä.

Kong, J., B. E. K. Klein, R. Klein, K. E. Lee, and G. Wahba (2012). Using distance correlation and SS-ANOVA to assess associations of familial relationships, lifestyle factors, diseases, and mortality. *Proc. Natl. Acad. Sci. USA* **109**, 20352–20357.

Kong, J., S. Wang, and G. Wahba (2015). Using distance covariance for improved variable selection with application to learning genetic risk models. *Stat. Med.* **34**, 1708–1720.

Kong, Y., D. Li, Y. Fan, and J. Lv (2017). Interaction pursuit in high-dimensional multi-response regression via distance correlation. *Ann. Statist.* **45**, 897–922.

Kosorok, M. R. (2009). Discussion of: Brownian distance covariance. *Ann. Appl. Stat.* **3**, 1270–1278.

Lacal, V. and D. Tjøstheim (2017). Local Gaussian autocorrelation and tests for serial independence. *J. Time Series Anal.* **38**, 51–71.

Leucht, A. and M. H. Neumann (2013). Dependent wild bootstrap for degenerate U- and V-statistics. *J. Multivariate Anal.* **117**, 257–280.

Li, R., W. Zhong, and L. Zhu (2012). Feature screening via distance correlation learning. *J. Amer. Statist. Assoc.* **107**, 1129–1139.

Li, T. H. (2008). Laplace periodogram for time series analysis. *J. Amer. Statist. Assoc.* **103**, 757–768.

Li, W.-K. (2004). *Diagnostic Checks in Time Series*. Boca Raton, FL: Chapman & Hall/CRC.

Li, W.-K. and A. I. McLeod (1981). Distribution of the residual autocorrelations in multivariate ARMA time series models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **43**, 231–239.

Liu, Y. and Q. Wang (2017). Model-free feature screening for ultrahigh-dimensional data conditional on some variables. *Ann. Inst. Statist. Math.*.

Ljung, G. M. and G. E. P. Box (1978). On a measure of lack of fit in time series models. *Biometrika* **65**, 297–303.

Lyons, R. (2013). Distance covariance in metric spaces. *Ann. Probab.* **41**, 3284–3305.

Mahdi, E. (2011). *Diagnostic Checking, Time Series and Regression*. Ph. D. thesis, The University of Western Ontario.

Martínez-Gómez, E., M. T. Richards, and D. S. P. Richards (2014). Distance correlation methods for discovering associations in large astrophysical databases. *The Astrophysical Journal* **781**, 39 (11 pp.).

Matsui, M., T. Mikosch, and G. Samorodnitsky (2017). Distance covariance for stochastic processes. *Probab. Math. Stat. Pol.* **37**, 355–372.

Meintanis, S. G. and G. Iliopoulos (2008). Fourier methods for testing multivariate independence. *Comput. Statist. Data Anal.* **52**, 1884–1895.

Oliveira, A. P., S. Dimopoulos, A. G. Busetto, S. Christen, R. Dechant, L. Falter, M. H. Chehreghani, S. Jozefczuk, C. Ludwig, F. Rudroff, J. C. Schulz, A. Gonz'alez, A. Soulard, D. Stracka, R. Aebersold, J. M. Buhmann, M. N. Hall, M. Peter, U. Sauer, and J. Stelling (2015). Inferring causal metabolic signals that regulate the dynamic TORC1-dependent transcriptome. *Molecular Systems Biology* **11**, 802.

Park, T., X. Shao, and S. Yao (2015). Partial martingale difference correlation. *Electron. J. Stat.* **9**, 1492–1517.

Parzen, E. (1957). On consistent estimates of the spectrum of a stationary time series. *Ann. Math. Stat.* **28**, 329–348.

Pfister, N., P. Bühlmann, B. Schölkopf, and J. Peters (2018). Kernel-based tests for joint independence. *J. R. Stat. Soc. B. 80*, 5–31.

Pinkse, J. (1998). Consistent nonparametric testing for serial independence. *J. Econometrics* **84**, 205–231.

Piras, V., M. Tomita, and K. Selvarajoo (2014). Transcriptome-wide variability in single embryonic development cells. *Scientific reports* **4**, Article number 7137.

Pitsillou, M. and K. Fokianos (2016). dCovTS: Distance covariance and correlation for time series analysis. *The R Journal* **8**, 324–340.

Politis, N. P., J. P. Romano, and M. Wolf (1999). *Subsampling*. New York: Springer-Verlag.

Rémillard, B. (2009). Discussion of: Brownian distance covariance. *Ann. Appl. Stat.* **3**, 1295–1298.

Richards, M. T., D. S. P. Richards, and E. Martínez-Gómez (2014). Interpreting the distance correlation results for the COMBO-17 survey. *The Astrophysical Journal Letters* **784**, L34 (5 pp.).

Rizzo, M. L. and G. J. Székely (2016). *energy: E-Statistics: Multivariate Inference via the Energy of Data*. R package version 1.7-0.

Robinson, P. M. (1991). Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *J. Econometrics* **47**, 67–84.

Romano, P. and L. A. Thombs (1996). Inference for autocorrelations under weak assumptions. *J. Amer. Statist. Assoc.* **91**, 590–600.

Sejdinovic, D., B. Sriperumbudur, A. Gretton, and K. Fukumizu (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann. Statist.* **41**, 2263–2291.

Shao, X. (2010). The dependent wild bootstrap. *J. Amer. Statist. Assoc.* **105**, 218–235.

Shao, X. (2011). Testing for white noise under unknown dependence and its applications to diagnostic checking for time series models. *Econom. Theory* **27**, 312–343.

Shao, X. and J. Zhang (2014). Martingale difference correlation and its use in high-dimensional variable screening. *J. Amer. Statist. Assoc.* **109**, 1302–1318.

Skaug, H. J. and D. Tjøstheim (1993). A nonparametric test of serial independence based on the empirical distribution function. *Biometrika* **80**, 591–602.

Skaug, H. J. and D. Tjøstheim (1996). Measures of distance between densities with application to testing for serial independence. *Time Series Analysis in Memory of E. J. Hannan (eds. P. Robinson and M. Rosenblatt, New York)*, 363–377.

Song, L., A. Smola, A. Gretton, J. Bedo, and K. Borgwardt (2012). Feature selection via dependence maximization. *J. Mach. Learn. Res.* **13**, 1393–1434.

Støve, B. and D. Tjøstheim (2014). Asymmetric dependence patterns in financial returns: An empirical investigation using local Gaussian correlation. In N. Haldrup, M. Meitz, and P. Saikkonen (Eds.), *Essays in Nonlinear Time Series Econometrics*, pp. 307–329. Oxford: Oxford University Press.

Støve, B., D. Tjøstheim, and K. O. Hufthammer (2014). Using local Gaussian correlation in a nonlinear re-examination of financial contagion. *Journal of Empirical Finance* **25**, 62–82.

Székely, G. J. and M. L. Rizzo (2009). Brownian distance covariance. *Ann. Appl. Stat.* **3**, 1236–1265.

Székely, G. J. and M. L. Rizzo (2012). On the uniqueness of distance covariance. *Statist. Probab. Lett.* **82**, 2278–2282.

Székely, G. J. and M. L. Rizzo (2013). Energy statistics: A class of statistics based on distances. *J. Statist. Plann. Inference* **143**, 1249–1272.

Székely, G. J. and M. L. Rizzo (2014). Partial distance correlation with methods for dissimilarities. *Ann. Statist.* **42**, 2382–2412.

Székely, G. J. and M. L. Rizzo (2017). The energy of data. *Annu. Rev. Stat. Appl.* **4**, 447–479.

Székely, G. J., M. L. Rizzo, and N. K. Bakirov (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35**, 2769–2794.

Székely, G. J., M. L. Rizzo, and N. K. Bakirov (2013). The distance correlation t-test of independence in high dimension. *J. Multivariate Anal.* **117**, 193–213.

Tjøstheim, D. (1996). Measures of dependence and tests of independence. *Statistics* **28**, 249–284.

Tjøstheim, D. and K. O. Hufthammer (2013). Local Gaussian correlation: A new measure of dependence. *J. Econometrics* **172**, 33–48.

Tonde, C. J. (2016). *Supervised feature learning via dependency maximization*. Ph. D. thesis, Rutgers University, New Brunswick.

Wang, X., W. Pan, W. Hu, Y. Tian, and H. Zhang (2015). Conditional distance correlation. *J. Amer. Statist. Assoc.* **110**, 1726–1734.

Xiao, H. and W. B. Wu (2014). Portmanteau test and simultaneous inference for serial covariances. *Statist. Sinica* **24**, 577–599.

Yao, S., X. Zhang, and X. Shao (2018). Testing mutual independence in high dimension via distance covariance. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80**, 455–480.

Yenigün, C. D. and M. L. Rizzo (2015). Variable selection in regression using maximal correlation and distance correlation. *J. Stat. Comput. Simul.* **85**, 1692–1705.

Zhang, X., B. Podobnik, D. Y. Kenett, and H. E. Stanley (2014). Systemic risk and causality dynamics of the world international shipping market. *Phys. A* **415**, 43–53.

Zhong, W. and L. Zhu (2015). An iterative approach to distance correlation-based sure independence screening. *J. Stat. Comput. Simul.* **85**, 2331–2345.

Zhou, Z. (2012). Measuring nonlinear dependence in time-series, a distance correlation approach. *J. Time Series Anal.* **33**, 438–457.