Bernoulli **24**(4A), 2018, 3087–3116
https://doi.org/10.3150/17-BEJ955

Applications of distance correlation to time series

RICHARD A. DAVIS[1,*], MUNEYA MATSUI[2], THOMAS MIKOSCH[3] and PHYLLIS WAN[1,**]

[1]*Department of Statistics, Columbia University, 1255 Amsterdam Ave, New York, NY 10027, USA. E-mail: *rdavis@stat.columbia.edu; **phyllis@stat.columbia.edu*
[2]*Department of Business Administration, Nanzan University, 18 Yamazato-cho, Showa-ku, Nagoya 466-8673, Japan. E-mail: mmuneya@gmail.com*
[3]*Department of Mathematics, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark. E-mail: mikosch@math.ku.dk*

The use of empirical characteristic functions for inference problems, including estimation in some special parametric settings and testing for goodness of fit, has a long history dating back to the 70s. More recently, there has been renewed interest in using empirical characteristic functions in other inference settings. The distance covariance and correlation, developed by Székely et al. (*Ann. Statist.* **35** (2007) 2769–2794) and Székely and Rizzo (*Ann. Appl. Stat.* **3** (2009) 1236–1265) for measuring dependence and testing independence between two random vectors, are perhaps the best known illustrations of this. We apply these ideas to stationary univariate and multivariate time series to measure lagged auto- and cross-dependence in a time series. Assuming strong mixing, we establish the relevant asymptotic theory for the sample auto- and cross-distance correlation functions. We also apply the auto-distance correlation function (ADCF) to the residuals of an autoregressive processes as a test of goodness of fit. Under the null that an autoregressive model is true, the limit distribution of the empirical ADCF can differ markedly from the corresponding one based on an i.i.d. sequence. We illustrate the use of the empirical auto- and cross-distance correlation functions for testing dependence and cross-dependence of time series in a variety of contexts.

Keywords: U-statistics; AR process; auto- and cross-distance correlation function; ergodicity; Fourier analysis; residuals; strong mixing; testing independence; time series

1. Introduction

In time series analysis, modeling serial dependence is typically the overriding objective. In order to achieve this goal, it is necessary to formulate a measure of dependence and this may depend on the features in the data that one is trying to capture. The autocorrelation function (ACF), which provides a measure of linear dependence, is perhaps the most used dependence measure in time series. It is closely linked with the class of ARMA models and provides guidance in both model selection and model confirmation. On the other hand, the ACF gives only a partial description of serial dependence. As seen with financial time series, data are typically uncorrelated but dependent so that the ACF is non-informative. In this case, the dependence becomes visible by examining the ACF applied to the absolute values or squares of the time series. In this paper, we consider the application of distance correlation in a time series setting, which can overcome some of the limitations of other dependence measures.

1350-7265 © 2018 ISI/BS

In recent years, the notions of distance covariance and correlation have become rather popular in applied statistics. Given vectors X and Y with values in $\mathbb{R}^p$ and $\mathbb{R}^q$, the *distance covariance* between X and Y with respect to a suitable measure μ on $\mathbb{R}^{p+q}$ is given by

$$T(X,Y;\mu)=\int_{\mathbb{R}^{p+q}}\left|\varphi_{X,Y}(s,t)-\varphi_X(s)\varphi_Y(t)\right|^2\mu(ds,dt), \tag{1.1}$$

where the characteristic function of any random vector $Z\in\mathbb{R}^d$ is denoted by $\varphi_Z(t)=\mathbb{E}[\mathrm{e}^{i\langle t,Z\rangle}]$, $t\in\mathbb{R}^d$. The *distance correlation* is the corresponding version of T standardized to values in $[0,1]$. The quantity $T(X,Y;\mu)$ is zero if and only if $\varphi_{X,Y}=\varphi_X\varphi_Y$ μ-a.e. In many situations, for example, when μ has a positive Lebesgue density on $\mathbb{R}^{p+q}$, we may conclude that X and Y are independent if and only if $T(X,Y;\mu)=0$. An empirical version $T_n(X,Y;\mu)$ of $T(X,Y;\mu)$ is obtained if the characteristic functions in (1.1) are replaced by their corresponding empirical versions. Then one can build a test for independence between X and Y based on the distribution of T_n under the null hypothesis that X and Y are independent.

The use of empirical characteristic functions for univariate and multivariate sequences for inference purposes has a long history. In the 1970s and 1980s, Feuerverger and Mureika [12], Csörgő [4–6] and many others proved fundamental asymptotic results for i.i.d. sequences, including Donsker-type theory for the empirical characteristic function. Statisticians have applied these methods for goodness-of-fit tests, changepoint detection, testing for independence, etc.; see, for example, Meintanis and coworkers [15,21,22], and the references therein. The latter authors employed the empirical distance covariance for finite measures μ. Feuerverger [11] was the first to apply statistics of the form (1.1) for general measures. In particular, he advocated the infinite measure

$$\mu(ds,dt)=|s|^{-2}|t|^{-2}\,ds\,dt$$

for testing independence of univariate data. Székely et al.[1]([27–29], see also the references therein) developed asymptotic techniques for the empirical distance covariance and correlation of i.i.d. sequences for the infinite measure μ given by

$$\mu(ds,dt)=c_{p,q}|s|^{-\alpha-p}|t|^{-\alpha-q}\,ds\,dt, \tag{1.2}$$

where $c_{p,q}$ is a constant (see (2.13)) and $\alpha\in(0,2)$. With this choice of μ, the distance correlation, $T(X,Y;\mu)/(T(X,X;\mu)T(Y,Y;\mu))^{1/2}$ is invariant relative to scale and orthogonal transformations, two desirable properties for measures of dependence. As a consequence, this choice of measure is perhaps the most common. However, there are other choices of measures for μ that are also useful depending on the context.

Dueck et al. [9] studied the affinely invariant distance covariance given by $\widetilde{T}(X,Y;\mu)=T(\Sigma_X^{-1}X,\Sigma_Y^{-1}Y)$, where Σ_X,Σ_Y are the respective covariance matrices of X and Y and μ is given by (1.2). They showed that the empirical version of $\widetilde{T}(X,Y;\mu)/(\widetilde{T}(X,X;\mu)\widetilde{T}(Y,Y;\mu))^{1/2}$, where Σ_X and Σ_Y are estimated by their empirical counterparts, is strongly consistent. In addition, they provide explicit expressions in terms of special functions of the limit in the case

[1]They appeared to have coined the terms distance covariance and correlation.

when X, Y are multivariate normal. Further progress on this topic has been achieved in [26] and [20], who generalized distance correlation to a metric space.

In this paper, we are interested in the empirical distance covariance and correlation applied to a stationary sequence $((X_t, Y_t))$ to study serial dependence, where X_t and Y_t assume values in $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively. We aim at an analog to the autocorrelation and autocovariance functions of classical time series analysis in terms of lagged distance correlation and distance covariance. Specifically, we consider the lagged-distance covariance function $T(X_0, Y_h; \mu)$, $h \in \mathbb{Z}$, and its standardized version that takes values in [0, 1]. We refer to these quantities as the *auto- and cross-distance covariance and correlation functions*. We provide asymptotic theory for the empirical auto- and cross-distance covariance and correlation functions under mild conditions. Under ergodicity, we prove consistency and under α-mixing, we derive the weak limits of the empirical auto- and cross-distance covariance functions for both cases when X_0 and Y_h are independent and dependent.

From a modeling perspective, distance correlation has limited value in providing a clear description of the nature of the dependence in the time series. To this end, it may be difficult to find a time series model that produces a desired distance correlation. In contrast, one could always find an autoregressive (or more generally ARMA) process that matches the ACF for an arbitrary number of lags. The theme in this paper will be to view the distance correlation more as a tool for testing independence rather than actually measuring dependence.

The literature on distance correlation for dependent sequences is sparse. To the best of our knowledge, Zhou [30] was the first to study the auto-distance covariance and its empirical analog for stationary sequences. In particular, he proved limit theory for $T_n(X_0, X_h; \mu)$ under so-called physical dependence measure conditions on (X_t) and independence of X_0 and X_h. Fokianos and Pitsillou [13] developed limit theory for a Ljung–Box-type statistic based on pairwise distance covariance $T_n(X_i, X_j; \mu)$ of a sample from a stationary sequence. In both papers, the measure μ is given by (1.2). The latter paper uses ideas from [16]. He applied the empirical characteristic function of a strongly mixing time series for testing various hypotheses on the dependence structure of a time series; he called it a generalized spectral approach. His test statistic bears some resemblance with the distance covariance: it is an integral of the weighted squared difference between the Fourier transform of the sequence $\mathrm{cov}(\mathrm{e}^{iuX_0}, \mathrm{e}^{ivX_h})$ and an empirical analog weighted by the density of a finite measure μ.

Typically, a crucial and final step in checking the quality of a fitted time series model is to examine the residuals for lack of serial dependence. The distance correlation can be used in this regard. However, as first pointed out in his discussion, Rémillard [24] indicated that the behavior of the distance correlation when applied to the residuals of a fitted AR(1) process need not have the same limit distribution as that of the distance correlation based on the corresponding i.i.d. noise. We provide a rigorous proof of this result for a general AR(p) process with finite variance under certain conditions on the measure μ. Interestingly, the conditions preclude the use of the standard weight function (1.2) used in [29]. In contrast, if the noise sequence is heavy-tailed and belongs to the domain of attraction of a stable distribution with index $\beta \in (0, 2)$, the distance correlation functions for both the residuals from the fitted model and the i.i.d. noise sequence coincide.

The paper is organized as follows. In Section 2, we commence with some basic results for distance covariance. We give conditions on the moments of X and Y and the measure μ, which

ensure that the integrals $T(X,Y;\mu)$ in (1.1) are well-defined. We provide alternative representations of $T(X,Y;\mu)$ and consider various examples of finite and infinite measures μ. Section 3 is devoted to the empirical auto- and cross-distance covariance and correlation functions. Our main results on the asymptotic theory of these functions are provided in Section 3.1. Among them are an a.s. consistency result (Theorem 3.1) under the assumption of ergodicity and asymptotic normality under a strong mixing condition (Theorem 3.2). Another main result (Theorem 4.2) is concerned with the asymptotic behavior of the empirical auto-distance covariance function of the residuals of an autoregressive process for both the finite and infinite variance cases. In Section 5, we provide a small study of the empirical auto-distance correlation functions derived from simulated and real-life dependent data of moderate sample size. The proofs of Lemma 4.1 and Theorem 4.2, which are significant but very technical, are relegated to the supplement [7].

2. Distance covariance for stationary time series

2.1. Conditions for existence

From (1.1), the distance covariance between two vectors X and Y is the squared L^2-distance between the joint characteristic function of (X,Y) and the product of the marginal characteristic functions of X and Y with respect to a measure μ on $\mathbb{R}^{p+q}$. Throughout we assume that μ is finite on sets bounded away from the origin, that is, on sets of the form

$$D_\delta^c = \big\{(s,t) : |s| \wedge |t| > \delta\big\}, \qquad \delta > 0. \tag{2.1}$$

In what follows, we interpret (s,t) as a concatenated vector in $\mathbb{R}^{p+q}$ equipped with the natural norm $|(s,t)|_{\mathbb{R}^p\times\mathbb{R}^q} = \sqrt{|s|^2+|t|^2}$. We suppress the dependence of the norm $|\cdot|$ on the dimension. The symbol c stands for any positive constant, whose value may change from line to line, but is not of particular interest. Clearly if X and Y are independent, $T(X,Y;\mu)=0$. On the other hand, if μ is an infinite measure, and X and Y are dependent, extra conditions are needed to ensure that $T(X,Y;\mu)$ is finite. This is the content of the following lemma.

Lemma 2.1. *Let X and Y be two possibly dependent random vectors and one of the following conditions is satisfied*:

1. *μ is a finite measure on $\mathbb{R}^{p+q}$.*
2. *μ is an infinite measure on $\mathbb{R}^{p+q}$, finite on the sets D_δ^c, $\delta>0$, such that*

$$\int_{\mathbb{R}^{p+q}} \big(1\wedge|s|^\alpha\big)\big(1\wedge|t|^\alpha\big)\mu(ds,dt) < \infty \tag{2.2}$$

and $\mathbb{E}[|X|^\alpha]+\mathbb{E}[|Y|^\alpha]<\infty$ for some $\alpha\in(0,2]$.

3. *μ is infinite in a neighborhood of the origin and for some $\alpha\in(0,2]$, $\mathbb{E}[|X|^\alpha]+\mathbb{E}[|Y|^\alpha]<\infty$ and*

$$\int_{\mathbb{R}^{p+q}} 1\wedge\big|(s,t)\big|^\alpha\mu(ds,dt) < \infty. \tag{2.3}$$

Then $T(X,Y;\mu)$ is finite.

Remark 2.2. If $\mu = \mu_1 \times \mu_2$ for some measures μ_1 and μ_2 on $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively, and if μ is finite on the sets D_δ^c then it suffices for (2.2) to verify that

$$\int_{|s|\le 1} |s|^\alpha \mu_1(ds) + \int_{|t|\le 1} |t|^\alpha \mu_2(dt) < \infty.$$

Proof. (1) Since the integrand in $T(X, Y; \mu)$ is uniformly bounded the statement is trivial.

(2) By (2.1), $\mu(D_\delta^c) < \infty$ for any $\delta > 0$. Therefore, it remains to verify the integrability of $|\varphi_{X,Y}(s,t) - \varphi_X(s)\varphi_Y(t)|^2$ on one of the sets D_δ. We consider only the case $|s| \vee |t| \le 1$; the cases when $|s| \le 1$, $|t| > 1$ and $|s| > 1$, $|t| \le 1$ are similar. An application of the Cauchy–Schwarz inequality yields

$$\big|\varphi_{X,Y}(s,t) - \varphi_X(s)\varphi_Y(t)\big|^2 \le \big(1 - \big|\varphi_X(s)\big|^2\big)\big(1 - \big|\varphi_Y(t)\big|^2\big). \tag{2.4}$$

Since $1 - |\varphi_X(s)|^2 = \int_{\mathbb{R}^p} (1 - \cos\langle s, x\rangle)\mathbb{P}(X - X' \in dx)$ for an independent copy X' of X, a Taylor expansion and the fact that X, X' have finite αth moments yield for $\alpha \in (0, 2]$ and some constant $c > 0$,

$$\begin{aligned} 1 - \big|\varphi_X(s)\big|^2 &\le \int_{\mathbb{R}^p} \big(2 \wedge \big|\langle s, x\rangle\big|^2\big)\mathbb{P}\big(X - X' \in dx\big) \\ &\le 2\int_{|\langle s,x\rangle|\le\sqrt{2}} \big|\langle s, x\rangle/\sqrt{2}\big|^\alpha \mathbb{P}\big(X - X' \in dx\big) + 2\mathbb{P}\big(\big|\langle s, X - X'\rangle\big| > \sqrt{2}\big) \\ &\le c|s|^\alpha \mathbb{E}\big[\big|X - X'\big|^\alpha\big] < \infty. \end{aligned} \tag{2.5}$$

In the last step we used Markov's inequality and the fact that $|\langle s, x\rangle| \le |s||x|$. A corresponding bound holds for $1 - |\varphi_Y(t)|^2$. Now, $T(X, Y; \mu) < \infty$ follows from (2.2) and (2.4).

(3) By (2.3), $\mu(\{(s,t) : |(s,t)| > 1\})$ is finite. Therefore, we need to show integrability of $|\varphi_{X,Y}(s,t) - \varphi_X(s)\varphi_Y(t)|^2$ only for $|(s,t)| \le 1$. Using the arguments from part (2) and the finiteness of the αth moments, we have

$$\big|\varphi_{X,Y}(s,t) - \varphi_X(s)\varphi_Y(t)\big|^2 \le c\big(|s|^\alpha + |t|^\alpha\big) \le c\big|(s,t)\big|^\alpha.$$

Now integrability of the left-hand side at the origin with respect to μ is ensured by (2.3). □

2.2. Alternative representations and examples

If $\mu = \mu_1 \times \mu_2$ for measures μ_1 and μ_2 on $\mathbb{R}^p$ and $\mathbb{R}^q$ we write for $x \in \mathbb{R}^p$ and $y \in \mathbb{R}^q$,

$$\hat{\mu}(x, y) = \int_{\mathbb{R}^{p+q}} \cos\big(\langle s, x\rangle + \langle t, y\rangle\big)\mu(ds, dt),$$

$$\hat{\mu}_1(x) = \int_{\mathbb{R}^p} \cos\langle s, x\rangle \mu_1(ds), \qquad \hat{\mu}_2(y) = \int_{\mathbb{R}^q} \cos\langle t, y\rangle \mu_2(dt),$$

for the real parts of the Fourier transforms with respect to μ, μ_1, μ_2, respectively. We assume that these transforms are well-defined. Let (X', Y') be an independent copy of (X, Y), and let Y'' and Y''' be independent copies of Y which are also independent of (X, Y), (X', Y'). We have

$$\begin{aligned} T(X, Y; \mu) = \int_{\mathbb{R}^{p+q}} \mathbb{E}\big[&\mathrm{e}^{i\langle s, X-X'\rangle + i\langle t, Y-Y'\rangle} + \mathrm{e}^{i\langle s, X-X'\rangle}\mathrm{e}^{i\langle t, Y''-Y'''\rangle} \\ &- \mathrm{e}^{i\langle s, X-X'\rangle + i\langle t, Y-Y''\rangle} - \mathrm{e}^{-i\langle s, X-X'\rangle - i\langle t, Y-Y''\rangle}\big]\mu(ds, dt). \end{aligned} \tag{2.6}$$

Notice that the complex-valued trigonometric functions under the expected value may be replaced by their real parts. We intend to interchange the integral with respect to μ and the expectation.

2.2.1. *Finite μ*

For a finite measure on $\mathbb{R}^{p+q}$, we may apply Fubini's theorem directly and interchange integration with expectation to obtain

$$\begin{aligned} T(X, Y; \mu) = \mathbb{E}\big[\hat{\mu}\big(X - X', Y - Y'\big)\big] + \mathbb{E}\big[\hat{\mu}\big(X - X', Y'' - Y'''\big)\big] \\ - 2\mathbb{E}\big[\hat{\mu}\big(X - X', Y - Y''\big)\big]. \end{aligned} \tag{2.7}$$

If $\mu = \mu_1 \times \mu_2$ we also have

$$\begin{aligned} T(X, Y; \mu) = \mathbb{E}\big[\hat{\mu}_1\big(X - X'\big)\hat{\mu}_2\big(Y - Y'\big)\big] + \mathbb{E}\big[\hat{\mu}_1\big(X - X'\big)\big]\mathbb{E}\big[\hat{\mu}_2\big(Y - Y'\big)\big] \\ - 2\mathbb{E}\big[\hat{\mu}_1\big(X - X'\big)\hat{\mu}_2\big(Y - Y''\big)\big]. \end{aligned}$$

2.2.2. *The case of an infinite measure μ*

We consider an infinite measure μ on $\mathbb{R}^{p+q}$ which is finite on D_δ^c for any $\delta > 0$. We assume that $T(X, Y; \mu)$ is finite and $\mu = \mu_1 \times \mu_2$. In this case, we cannot pass from (2.6) to (2.7) because the Fourier transform $\hat{\mu}$ is not defined as a Lebesgue integral. We have

$$T(X, Y; \mu) = \int_{\mathbb{R}^{p+q}} \big(\mathbb{E}\big[\mathrm{COS}(s, t)\big] + \mathbb{E}\big[\mathrm{SIN}(s, t)\big]\big)\mu(ds, dt), \tag{2.8}$$

where

$$\begin{aligned} \mathrm{COS}(s, t) &= \cos\big(\langle s, X - X'\rangle\big)\cos\big(\langle t, Y - Y'\rangle\big) + \cos\big(\langle s, X - X'\rangle\big)\cos\big(\langle t, Y'' - Y'''\rangle\big) \\ &\quad - 2\cos\big(\langle t, X - X'\rangle\big)\cos\big(\langle s, Y - Y''\rangle\big), \\ \mathrm{SIN}(s, t) &= -\sin\big(\langle s, X - X'\rangle\big)\sin\big(\langle t, Y - Y'\rangle\big) - \sin\big(\langle s, X - X'\rangle\big)\sin\big(\langle t, Y'' - Y'''\rangle\big) \\ &\quad + 2\sin\big(\langle t, X - X'\rangle\big)\sin\big(\langle s, Y - Y''\rangle\big). \end{aligned}$$

Using the fact that

$$\cos u \cos v = 1 - (1 - \cos u) - (1 - \cos v) + (1 - \cos u)(1 - \cos v),$$

calculation shows that

$$\begin{aligned}\mathbb{E}\big[\mathrm{COS}(s,t)\big] = \mathbb{E}\big[&\big(1-\cos\big(\langle s, X-X'\rangle\big)\big)\big(1-\cos\big(\langle t, Y-Y'\rangle\big)\big)\\ &+\big(1-\cos\big(\langle s, X-X'\rangle\big)\big)\big(1-\cos\big(\langle t, Y''-Y'''\rangle\big)\big)\\ &-2\big(1-\cos\big(\langle t, X-X'\rangle\big)\big)\big(1-\cos\big(\langle s, Y-Y''\rangle\big)\big)\big].\end{aligned}$$

A Taylor series argument shows that for $\alpha \in (0,2]$,

$$\begin{aligned}\mathbb{E}\big[\big|\mathrm{COS}(s,t)\big|\big] \le c(&\mathbb{E}\big[\big(1\wedge\big|\langle s, X-X'\rangle/\sqrt{2}\big|^{\alpha}\big)\big(1\wedge\big|\langle t, Y-Y'\rangle/\sqrt{2}\big|^{\alpha}\big)\big]\\ &+\mathbb{E}\big[1\wedge\big|\langle s, X-X'\rangle/\sqrt{2}\big|^{\alpha}\big]\mathbb{E}\big[1\wedge\big|\langle t, Y-Y'\rangle/\sqrt{2}\big|^{\alpha}\big]\\ &+\mathbb{E}\big[\big(1\wedge\big|\langle t, X-X'\rangle/\sqrt{2}\big|^{\alpha}\big)\big(1\wedge\big|\langle s, Y-Y''\rangle/\sqrt{2}\big|^{\alpha}\big)\big]\big).\end{aligned}$$

Under condition (2.2) the right-hand side is integrable with respect to μ if

$$\mathbb{E}\big[|X|^{\alpha}+|Y|^{\alpha}+|X|^{\alpha}|Y|^{\alpha}\big] < \infty. \tag{2.9}$$

An application of Fubini's theorem yields

$$\begin{aligned}&\int_{\mathbb{R}^{p+q}} \mathbb{E}\big[\mathrm{COS}(s,t)\big]\mu(ds,dt)\\ &\quad= \mathbb{E}\bigg[\int_{\mathbb{R}^{p+q}} \big(\big(1-\cos\big(\langle s, X-X'\rangle\big)\big)\big(1-\cos\big(\langle t, Y-Y'\rangle\big)\big)\\ &\qquad+\big(1-\cos\big(\langle s, X-X'\rangle\big)\big)\big(1-\cos\big(\langle t, Y''-Y'''\rangle\big)\big)\\ &\qquad-2\big(1-\cos\big(\langle t, X-X'\rangle\big)\big)\big(1-\cos\big(\langle s, Y-Y''\rangle\big)\big)\big)\mu(ds,dt)\bigg].\end{aligned}$$

If we assume that the restrictions μ_1, μ_2 of μ to $\mathbb{R}^p$ and $\mathbb{R}^q$ are symmetric about the origin then we have $\mathbb{E}[\mathrm{SIN}(s,t)] = -\mathbb{E}[\mathrm{SIN}(-s,t)] = -\mathbb{E}[\mathrm{SIN}(s,-t)]$. Together with the symmetry property of μ this implies that $\int_{\mathbb{R}^{p+q}} \mathbb{E}[\mathrm{SIN}(s,t)]\mu(ds,dt) = 0$.

We summarize these arguments. For any measure ν on $\mathbb{R}^d$ we write

$$\tilde{\nu}(s) = \int_{\mathbb{R}^d} \big(1-\cos\langle s, x\rangle\big)\nu(dx), \qquad s \in \mathbb{R}^d.$$

Lemma 2.3. *Assume* (2.2) *and* (2.9) *for some* $\alpha \in (0,2]$. *If* μ_1, μ_2 *are symmetric about the origin and* $\mu = \mu_1 \times \mu_2$ *then*

$$\begin{aligned}T(X,Y;\mu) = \mathbb{E}\big[\tilde{\mu}_1\big(X-X'\big)\tilde{\mu}_2\big(Y-Y'\big)\big] &+ \mathbb{E}\big[\tilde{\mu}_1\big(X-X'\big)\big]\mathbb{E}\big[\tilde{\mu}_2\big(Y-Y'\big)\big]\\ &- 2\mathbb{E}\big[\tilde{\mu}_1\big(X-X'\big)\tilde{\mu}_2\big(Y-Y''\big)\big].\end{aligned} \tag{2.10}$$

Remark 2.4. For further use, we mention the alternative representation of (2.10):

$$\begin{aligned} T(X,Y;\mu) &= \operatorname{cov}\big(\tilde{\mu}_1\big(X-X'\big), \tilde{\mu}_2\big(Y-Y'\big)\big) \\ &\quad - 2\operatorname{cov}\big(\mathbb{E}\big[\tilde{\mu}_1\big(X-X'\big)|X\big], \mathbb{E}\big[\tilde{\mu}_2\big(Y-Y'\big)|Y\big]\big). \end{aligned} \tag{2.11}$$

2.2.3. *Examples*

Example 2.5. Assume that μ has density w on $\mathbb{R}^{p+q}$ given by

$$w(s,t) = c_{p,q}|s|^{-\alpha-p}|t|^{-\alpha-q}, \qquad s \in \mathbb{R}^p, t \in \mathbb{R}^q, \tag{2.12}$$

for some positive constant $c_{p,q} = c_p c_q$. For any $d \geq 1$ and $\alpha \in (0,2)$, one can choose c_d such that

$$\int_{\mathbb{R}^d} \big(1 - \cos\langle s, x\rangle\big) c_d |s|^{-\alpha-d}\, ds = |x|^\alpha. \tag{2.13}$$

Under the additional moment assumption (2.9), we obtain from (2.10)

$$\begin{aligned} T(X,Y;\mu) &= \mathbb{E}\big[\big|X-X'\big|^\alpha \big|Y-Y'\big|^\alpha\big] + \mathbb{E}\big[\big|X-X'\big|^\alpha\big]\mathbb{E}\big[\big|Y-Y'\big|^\alpha\big] \\ &\quad - 2\mathbb{E}\big[\big|X-X'\big|^\alpha \big|Y-Y''\big|^\alpha\big]. \end{aligned} \tag{2.14}$$

This is the *distance covariance* introduced by [29].

The distance covariance $T(X,Y;\mu)$ introduced in (2.14) has several good properties. It is homogeneous under positive scaling and is also invariant under orthonormal transformations of X and Y. Some of these properties are shared with other distance covariances when μ is infinite. We illustrate this for a Lévy measure μ on $\mathbb{R}^{p+q}$, that is, it satisfies (2.3) for $\alpha = 2$. In particular, μ is finite on sets bounded away from zero. Via the Lévy–Khintchine formula, a Lévy measure μ corresponds to an $\mathbb{R}^{p+q}$-valued infinitely divisible random vector (Z_1, Z_2) (with Z_1 assuming values in $\mathbb{R}^p$ and Z_2 in $\mathbb{R}^q$) and characteristic function

$$\begin{aligned} \varphi_{Z_1,Z_2}(x,y) = \exp\bigg\{ -\int_{\mathbb{R}^{p+q}} \big(\mathrm{e}^{i\langle s,x\rangle + i\langle t,y\rangle} - 1 \\ - \big(i\langle x,s\rangle + i\langle y,t\rangle\big)\mathbf{1}\big(\big|(s,t)\big| \leq 1\big)\big)\mu(ds,dt)\bigg\}. \end{aligned} \tag{2.15}$$

Lemma 2.6. *Assume that there exists an $\alpha \in (0,2]$ such that $\mathbb{E}[|X|^\alpha] + \mathbb{E}[|Y|^\alpha] < \infty$ and μ is a symmetric Lévy measure corresponding to* (2.15) *such that* (2.3) *holds. Then*

$$\begin{aligned} T(X,Y;\mu) &= \operatorname{Re}\mathbb{E}\big[-\log\varphi_{Z_1,Z_2}\big(X-X', Y-Y'\big) - \log\varphi_{Z_1,Z_2}\big(X-X', Y''-Y'''\big) \\ &\quad + 2\log\varphi_{Z_1,Z_2}\big(X-X', Y-Y''\big)\big]. \end{aligned} \tag{2.16}$$

Remark 2.7. We observe that (2.16) always vanishes if Z_1 and Z_2 are independent.

Proof. By the symmetry of the random vectors in (2.6) and the measure μ, we have

$$\begin{aligned}
&\operatorname{Re}\int_{\mathbb{R}^{p+q}} \mathbb{E}\big[\mathrm{e}^{i\langle s,X-X'\rangle+i\langle t,Y-Y'\rangle}-1\big]\mu(ds,dt)\\
&\quad=\operatorname{Re}\int_{\mathbb{R}^{p+q}} \mathbb{E}\big[\mathrm{e}^{i\langle s,X-X'\rangle+i\langle t,Y-Y'\rangle}-1\\
&\qquad-\big(i\langle s,X-X'\rangle+i\langle t,Y-Y'\rangle\big)\mathbf{1}\big(\big|(s,t)\big|\le 1\big)\big]\mu(ds,dt)\\
&\quad=\operatorname{Re}\mathbb{E}\big[-\log\varphi_{Z_1,Z_2}\big(X-X',Y-Y'\big)\big].
\end{aligned}$$

The last step is justified if we can interchange the integral and the expected value. Therefore, we have to verify that the following integral is finite:

$$\int_{\mathbb{R}^{p+q}} \mathbb{E}\big[\big|\mathrm{e}^{i\langle s,X-X'\rangle+i\langle t,Y-Y'\rangle}-1-\big(i\langle s,X-X'\rangle+i\langle t,Y-Y'\rangle\big)\mathbf{1}\big(\big|(s,t)\big|\le 1\big)\big|\big]\mu(ds,dt).$$

The integrals over the disjoint sets $\{(s,t):|(s,t)|\le 1\}$ and $\{(s,t):|(s,t)|>1\}$ are denoted by I_1 and I_2, respectively. The quantity I_2 is bounded since the integrand is bounded and μ is finite on sets bounded away from zero. A Taylor expansion shows for $\alpha\in(0,2]$,

$$\begin{aligned}
I_1&\le c\int_{|(s,t)|\le 1}\mathbb{E}\big[2\wedge\big(\big|\langle s,X-X'\rangle\big|+\big|\langle t,Y-Y'\rangle\big|\big)^2\big]\mu(ds,dt)\\
&\le c\big(\mathbb{E}\big[|X|^\alpha\big]+\mathbb{E}\big[|Y|^\alpha\big]\big)\int_{|(s,t)|\le 1}\big(1\wedge\big|(s,t)\big|^\alpha\big)\mu(ds,dt)
\end{aligned}$$

and the right-hand side is finite by assumption.

Proceeding in the same way as above for the remaining expressions in (2.6), the lemma is proved. □

Example 2.8. Assume that μ is a probability measure of a random vector (Z_1,Z_2) in $\mathbb{R}^{p+q}$ and that Z_1 and Z_2 are independent. Then

$$\begin{aligned}
T(X,Y;\mu)&=\mathbb{E}\big[\varphi_{Z_1}\big(X-X'\big)\varphi_{Z_2}\big(Y-Y'\big)\big]+\mathbb{E}\big[\varphi_{Z_1}\big(X-X'\big)\big]\mathbb{E}\big[\varphi_{Z_2}\big(Y''-Y'''\big)\big]\\
&\quad-2\mathbb{E}\big[\varphi_{Z_1}\big(X-X'\big)\varphi_{Z_2}\big(Y-Y''\big)\big].
\end{aligned}$$

For example, consider independent symmetric Z_1 and Z_2 with multivariate β-stable distributions in $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively, for some $\beta\in(0,2]$. They have joint characteristic function given by $\varphi_{Z_1,Z_2}(x,y)=\mathrm{e}^{-(|x|^\beta+|y|^\beta)}$. Therefore

$$\begin{aligned}
T(X,Y;\mu)&=\mathbb{E}\big[\mathrm{e}^{-(|X-X'|^\beta+|Y-Y'|^\beta)}\big]+\mathbb{E}\big[\mathrm{e}^{-|X-X'|^\beta}\big]\mathbb{E}\big[\mathrm{e}^{-|Y-Y'|^\beta}\big]\\
&\quad-2\mathbb{E}\big[\mathrm{e}^{-(|X-X'|^\beta+|Y-Y''|^\beta)}\big].
\end{aligned}\tag{2.17}$$

Example 2.9. Assume that X and Y are integer-valued. Consider the spectral densities w_1 and w_2 on $[-\pi,\pi]$ of two real-valued second-order stationary processes and assume $\mu(s,t)=$

$w_1(s)w_2(t)$. Denote the covariance functions on the integers corresponding to w_1 and w_2 by γ_1 and γ_2, respectively. We have the well-known relation

$$\int_{-\pi}^{\pi} \mathrm{e}^{itk} w_i(t)\,dt = \int_{-\pi}^{\pi} \cos(tk) w_i(t)\,dt = \gamma_i(k), \qquad k \in \mathbb{Z},$$

where we also exploit the symmetry of the functions w_i. If we restrict integration in (2.6) to $[-\pi, \pi]^2$ we obtain, abusing notation,

$$\begin{aligned} T(X, Y; \mu) = {} & \mathbb{E}\big[\gamma_1\big(X - X'\big)\gamma_2\big(Y - Y'\big)\big] + \mathbb{E}\big[\gamma_1\big(X - X'\big)\big]\mathbb{E}\big[\gamma_2\big(Y - Y'\big)\big] \\ & - 2\mathbb{E}\big[\gamma_1\big(X - X'\big)\gamma_2\big(Y - Y''\big)\big]. \end{aligned}$$

The spectral density of a stationary process may have singularities (e.g. for fractional ARMA processes) but this density is integrable on $[-\pi, \pi]$. If w_1, w_2 are positive Lebesgue a.e. on $[0, \pi]$ then $T(X, Y; \mu) = 0$ if and only if X, Y are independent. Indeed, the characteristic function of an integer-valued random variable is periodic with period 2π.

Example 2.10. To illustrate (2.16), we consider a symmetric α-stable vector (Z_1, Z_2) for $\alpha \in (0, 2)$ with log-characteristic function

$$-\log \varphi_{Z_1, Z_2}(x, y) = \int_{\mathbb{S}^{p+q-1}} \big|\langle s, x\rangle + \langle t, y\rangle\big|^{\alpha} m(ds, dt)$$

and m is a finite symmetric measure on the unit sphere $\mathbb{S}^{p+q-1}$ of $\mathbb{R}^{p+q}$. Then we have

$$\begin{aligned} T(X, Y; \mu) = {} & \int_{\mathbb{S}^{p+q-1}} \mathbb{E}\big[\big|\big\langle s, X - X'\big\rangle + \big\langle t, Y - Y'\big\rangle\big|^{\alpha} + \big|\big\langle s, X - X'\big\rangle + \big\langle t, Y'' - Y'''\big\rangle\big|^{\alpha} \\ & - 2\big|\big\langle s, X - X'\big\rangle + \big\langle t, Y' - Y''\big\rangle\big|^{\alpha}\big] m(ds, dt). \end{aligned}$$

A special case is the sub-Gaussian $\alpha/2$-stable random vectors with characteristic function $-\log \varphi_{Z_1, Z_2}(x, y) = |(x, y)'\Sigma(x, y)|^{\alpha/2}$, where Σ is the covariance matrix of an $\mathbb{R}^{p+q}$-valued random vector and we write (x, y) for the concatenation of any $x \in \mathbb{R}^p$ and $y \in \mathbb{R}^q$. Then

$$\begin{aligned} T(X, Y; \mu) = {} & \mathbb{E}\big[\big|\big(X - X', Y - Y'\big)'\Sigma\big(X - X', Y - Y'\big)\big|^{\alpha/2} \\ & + \big|\big(X - X', Y'' - Y'''\big)'\Sigma\big(X - X', Y'' - Y'''\big)\big|^{\alpha/2} \\ & - 2\big|\big(X - X', Y - Y''\big)'\Sigma\big(X - X', Y - Y''\big)\big|^{\alpha/2}\big]. \end{aligned}$$

In particular, if Σ is block-diagonal with Σ_1 a $p \times p$ covariance matrix and Σ_2 a $q \times q$ covariance matrix, we have

$$\begin{aligned} T(X, Y; \mu) = {} & \mathbb{E}\big[\big|\big(X - X'\big)'\Sigma_1\big(X - X'\big) + \big(Y - Y'\big)'\Sigma_2\big(Y - Y'\big)\big|^{\alpha/2} \\ & + \big|\big(X - X'\big)'\Sigma_1\big(X - X'\big) + \big(Y'' - Y'''\big)'\Sigma_2\big(Y'' - Y'''\big)\big|^{\alpha/2} \\ & - 2\big|\big(X - X'\big)'\Sigma_1\big(X - X'\big) + \big(Y - Y''\big)'\Sigma_2\big(Y - Y''\big)\big|^{\alpha/2}\big], \end{aligned}$$

and if Σ is the identity matrix,

$$T(X,Y;\mu) = \mathbb{E}\big[\big|\big|X-X'\big|^2 + \big|Y-Y'\big|^2\big|^{\alpha/2} + \big|\big|X-X'\big|^2 + \big|Y''-Y'''\big|^2\big|^{\alpha/2} - 2\big|\big|X-X'\big|^2 + \big|Y-Y''\big|^2\big|^{\alpha/2}\big]. \tag{2.18}$$

We notice that for these examples, $T(X,Y;\mu)$ is scale homogeneous ($T(cX,cY;\mu) = |c|^\alpha T(X,Y;\mu)$) and (2.18) is invariant under orthonormal transformations ($T(RX,SY;\mu) = T(X,Y;\mu)$ for orthonormal matrices R and S), properties also enjoyed by the weight function in Example 2.5.

3. The empirical distance covariance function of a stationary sequence

In this section, we consider the empirical distance covariance for a stationary time series $((X_t,Y_t))$ with generic element (X,Y) where X and Y assume values in $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively. The empirical distance covariance is given by

$$T_n(X,Y;\mu) = \int_{\mathbb{R}^{p+q}} \big|\varphi^n_{X,Y}(s,t) - \varphi^n_X(s)\varphi^n_Y(t)\big|^2 \mu(ds,dt),$$

where the empirical characteristic function is given by $\varphi^n_{X,Y}(s,t) = \frac{1}{n}\sum_{j=1}^n \mathrm{e}^{i\langle s,X_j\rangle + i\langle t,Y_j\rangle}$, $n \ge 1$, and $\varphi^n_X(s) = \varphi^n_{X,Y}(s,0)$ and $\varphi^n_Y(s) = \varphi^n_{X,Y}(0,t)$.

3.1. Asymptotic results for the empirical distance correlation

Under the conditions of Lemma 2.1 that ensure the finiteness of $T(X,Y;\mu)$, we show that T_n is consistent for stationary ergodic time series; see [25], Chapter 2, for a definition of ergodicity.

Theorem 3.1. *Consider a stationary ergodic time series $((X_j,Y_j))_{j=1,2,\dots}$ with values in $\mathbb{R}^{p+q}$ and assume one of the three conditions in Lemma 2.1 are satisfied. Then*

$$T_n(X,Y;\mu) \overset{\text{a.s.}}{\to} T(X,Y;\mu) \qquad \text{as } n \to \infty.$$

Proof. For $(s,t) \in \mathbb{R}^{p+q}$ the difference between the joint characteristic function with the product characteristic function and the empirical analog are given by

$$C(s,t) = \varphi_{X,Y}(s,t) - \varphi_X(s)\varphi_Y(t) \quad \text{and} \quad C_n(s,t) = \varphi^n_{X,Y}(s,t) - \varphi^n_X(s)\varphi^n_Y(t).$$

Each of the processes $\varphi^n_{X,Y}$, φ^n_X, φ^n_Y is a sample mean of i.i.d. bounded continuous processes defined on $\mathbb{R}^{p+q}$. Consider the compact set

$$K_\delta = \big\{(s,t) \in \mathbb{R}^{p+q} : \delta \le |s| \wedge |t|, |s| \vee |t| \le 1/\delta\big\} \tag{3.1}$$

for small $\delta > 0$. By the ergodic theorem on $\mathcal{C}(K_\delta)$, the space of continuous functions on K_δ, $\varphi^n_{X,Y} \overset{\text{a.s.}}{\to} \varphi_{X,Y}$ as $n \to \infty$; see [18]. Hence,

$$\int_{K_\delta} \left|C_n(s,t)\right|^2 \mu(ds,dt) \overset{\text{a.s.}}{\to} \int_{K_\delta} \left|C(s,t)\right|^2 \mu(ds,dt), \qquad n \to \infty.$$

It remains to show that

$$\lim_{\delta \downarrow 0} \limsup_{n\to\infty} \int_{K_\delta^c} \left|C_n(s,t)\right|^2 \mu(ds,dt) = 0 \qquad \text{a.s.}$$

If μ is a finite measure, we have

$$\lim_{\delta \downarrow 0} \limsup_{n\to\infty} \int_{K_\delta^c} \left|C_n(s,t)\right|^2 \mu(ds,dt) \le c \lim_{\delta \downarrow 0} \mu\left(K_\delta^c\right) = 0.$$

Now assume that μ is infinite on the axes or at zero and (2.2) holds. We apply inequality (2.4) under the assumption that (X,Y) has the empirical probability measure of the sample (X_j, Y_j), $j = 1, \dots, n$. Since the empirical measure has all moments finite we obtain from (2.5) that for $\alpha \in (0,2]$,

$$1 - \left|\varphi^n_X(s)\right|^2 \le c|s|^\alpha \mathbb{E}_{n,X}\left[\left|X - X'\right|^\alpha\right] = c|s|^\alpha n^{-2} \sum_{1\le k,l\le n} |X_k - X_l|^\alpha,$$

where X, X' are independent and each of them has the empirical distribution of the X-sample. The right-hand side is a U-statistic which converges a.s. to $\mathbb{E}[|X - X'|^\alpha]$ as $n \to \infty$ provided this moment is finite. This follows from the ergodic theorem for U-statistics; see [1]. The same argument as for part (2) of Lemma 2.1 implies that on K_δ^c,

$$\left|C_n(s,t)\right|^2 \le c\mathbb{E}_{n,X}\left[\left|X - X'\right|^\alpha\right]\mathbb{E}_{n,Y}\left[\left|Y - Y'\right|^\alpha\right]\left(1 \wedge |s|^\alpha\right)\left(1 \wedge |t|^\alpha\right).$$

By the ergodic theorem,

$$\begin{aligned}
&\limsup_{n\to\infty} \int_{K_\delta^c} \left|C_n(s,t)\right|^2 \mu(ds,dt) \\
&\quad \le c\mathbb{E}\left[\left|X - X'\right|^\alpha\right]\mathbb{E}\left[\left|Y - Y'\right|^\alpha\right] \int_{K_\delta^c} \left(1 \wedge |s|^\alpha\right)\left(1 \wedge |t|^\alpha\right)\mu(ds,dt)
\end{aligned}$$

a.s. and the latter integral converges to zero as $\delta \downarrow 0$ by assumption.

If the measure μ is infinite at zero and (2.3) holds the proof is analogous. □

In order to prove weak convergence of T_n we assume that the sequence $((X_i, Y_i))$ with values in $\mathbb{R}^{p+q}$ is α-mixing with rate function (α_h); see [8], p. 18 and [17], p. 305, for the definition. We have the following result.

Theorem 3.2. *Assume that $((X_j, Y_j))$ is a strictly stationary sequence with values in $\mathbb{R}^{p+q}$ such that $\sum_h \alpha_h^{1/r} < \infty$ for some $r > 1$. Set $u = 2r/(r-1)$ and write $X = (X^{(1)}, \dots, X^{(p)})$ and $Y = (Y^{(1)}, \dots, Y^{(q)})$.*

1. *Assume that X_0 and Y_0 are independent and for some $\alpha \in (u/2, u]$, $\epsilon \in [0, 1/2)$ and $\alpha' \le \min(2, \alpha)$, the following hold*:

$$\mathbb{E}\big[|X|^\alpha + |Y|^\alpha\big] < \infty, \qquad \mathbb{E}\left[\prod_{l=1}^{p} \big|X^{(l)}\big|^\alpha\right] < \infty, \qquad \mathbb{E}\left[\prod_{l=1}^{q} \big|Y^{(l)}\big|^\alpha\right] < \infty, \tag{3.2}$$

and

$$\int_{\mathbb{R}^{p+q}} \big(1 \wedge |s|^{\alpha'(1+\epsilon)/u}\big)\big(1 \wedge |t|^{\alpha'(1+\epsilon)/u}\big)\mu(ds, dt) < \infty. \tag{3.3}$$

Then

$$nT_n(X, Y; \mu) \xrightarrow{d} \|G\|_\mu^2 = \int_{\mathbb{R}^{p+q}} \big|G(s,t)\big|^2 \mu(ds, dt), \tag{3.4}$$

where G is a complex-valued mean-zero Gaussian process whose covariance structure is given in (3.9) with $h = 0$ and depends on the dependence structure of $((X_t, Y_t))$.

2. *Assume that X_0 and Y_0 are dependent and for some $\alpha \in (u/2, u]$, $\epsilon \in [0, 1/2)$ and for $\alpha' \le \min(2, \alpha)$ the following hold*:

$$\mathbb{E}\big[|X|^{2\alpha} + |Y|^{2\alpha}\big] < \infty, \qquad \mathbb{E}\left[\left(1 \vee \prod_{l=1}^{p} \big|X^{(l)}\big|^\alpha\right)\left(1 \vee \prod_{k=1}^{q} \big|Y^{(k)}\big|^\alpha\right)\right] < \infty, \tag{3.5}$$

and

$$\int_{\mathbb{R}^{p+q}} \big(1 \wedge |s|^{\alpha'(1+\epsilon)/u}\big)\big(1 \wedge |t|^{\alpha'(1+\epsilon)/u}\big)\mu(ds, dt) < \infty. \tag{3.6}$$

Then

$$\sqrt{n}\big(T_n(X, Y; \mu) - T(X, Y; \mu)\big) \xrightarrow{d} G'_\mu = \int_{\mathbb{R}^{p+q}} G'(s,t)\mu(ds, dt), \tag{3.7}$$

where $G'(s,t) = 2\,\mathrm{Re}\{G(s,t)C(s,t)\}$ is a mean-zero Gaussian process.

The proof of Theorem 3.2 is given in the Appendix.

Remark 3.3. We notice that (3.3) and (3.6) are always satisfied if μ is a finite measure.

Remark 3.4. If (X_i) and (Y_i) are two independent i.i.d. sequences then the statement of Theorem 3.2(1) remains valid if for some $\alpha \in (0, 2]$, $\mathbb{E}[|X|^\alpha] + \mathbb{E}[|Y|^\alpha] < \infty$ and

$$\int_{\mathbb{R}^{p+q}} \big(1 \wedge |s|^\alpha\big)\big(1 \wedge |t|^\alpha\big)\mu(ds, dt) < \infty. \tag{3.8}$$

Remark 3.5. The distribution of the limit variable in (3.4) is generally not tractable. Therefore one must use numerical or resampling methods for determining quantiles of $nT_n(X, Y; \mu)$. On the other hand, the limit distribution in (3.7) is normally distributed with mean 0 and variance σ_μ^2 that can be easily calculated from the covariance function of $G(s,t)$ and $C(s,t)$. Notice that if $C(s,t) = 0$, the limit random variable in (3.7) is 0 and part (1) of the theorem applies. Again resampling or subsampling methods must be employed to determine quantiles of nT_n.

3.2. Testing serial dependence for multivariate time series

Define the *cross-distance covariance function* (CDCVF) of a strictly stationary sequence $((X_t, Y_t))$ by

$$T_\mu^{X,Y}(h) = T(X_0, Y_h; \mu), \qquad h \in \mathbb{Z},$$

and the *auto-distance covariance function* (ADCVF) of a stationary sequence (X_t) by

$$T_\mu^{X}(h) = T_\mu^{X,X}(h), \qquad h \in \mathbb{Z}.$$

Here and in what follows, we assume that $\mu = \mu_1 \times \mu_2$ for suitable measures μ_1 on $\mathbb{R}^p$ and μ_2 on $\mathbb{R}^q$. In the case of an ADCVF we also assume $\mu_1 = \mu_2$. The empirical versions $T_{n,\mu}^X$ and $T_{n,\mu}^{X,Y}$ are defined correspondingly. For example, for integer $h \geq 0$, one needs to replace $\varphi_{X,Y}^n(s,t)$ in the definition of $T_n(X, Y; \mu)$ by

$$\varphi_{X_0,Y_h}^n(s,t) = \frac{1}{n}\sum_{j=1}^{n-h} \mathrm{e}^{i\langle s, X_j\rangle + i\langle t, Y_{j+h}\rangle}, \qquad s \in \mathbb{R}^p, t \in \mathbb{R}^q, n \geq h+1,$$

with the corresponding modifications for the marginal empirical characteristic functions. For finite h, the change from the upper summation limit n to $n-h$ has no influence on the asymptotic theory.

We also introduce the corresponding *cross-distance correlation function* (CDCF) and *auto-distance correlation function* (ADCF) respectively;

$$R_\mu^{X,Y}(h) = \frac{T_\mu^{X,Y}(h)}{\sqrt{T_\mu^X(0)T_\mu^Y(0)}} \quad \text{and} \quad R_\mu^X(h) = \frac{T_\mu^X(h)}{T_\mu^X(0)}, \qquad h \in \mathbb{Z}.$$

The quantities $R_\mu^{X,Y}(h)$ assume values in $[0,1]$, with the two endpoints representing independence and complete dependence. The empirical CDCF $R_{n,\mu}^{X,Y}$ and ADCF $R_{n,\mu}^X$ are defined by replacing the distance covariances $T_\mu^{X,Y}(h)$ by the corresponding empirical versions $T_{n,\mu}^{X,Y}(h)$.

The empirical ADCV was examined in [30] and [13] as an alternative tool for testing serial dependence, in the way that it also captures non-linear dependence. They always choose the measure $\mu = \mu_1 \times \mu_1$ with density (2.12).

In contrast to the autocorrelation and cross-correlation functions of standard stationary time series models (such as ARMA, GARCH) it is in general complicated (or impossible) to provide

explicit (and tractable) expressions for $T_\mu^X(h)$ and $T_\mu^{X,Y}(h)$ or even to say anything about the rate of decay of these quantities when $h \to \infty$. However, in view of (2.11), we observe that

$$\begin{aligned} T_\mu^X(h) &= \operatorname{cov}\big(\tilde{\mu}_1\big(X_0 - X_0'\big), \tilde{\mu}_1\big(X_h - X_h'\big)\big) \\ &\quad - 2\operatorname{cov}\big(\mathbb{E}\big[\tilde{\mu}_1\big(X_0 - X_0'\big)|X_0\big], \mathbb{E}\big[\tilde{\mu}_1\big(X_h - X_0'\big)|X_h\big]\big). \end{aligned}$$

While this is not the autocovariance function of a stationary process, it is possible to bound each of the terms in case (X_t) is α-mixing with rate function (α_h). In this case, one may use bounds for the autocovariance functions of the stationary series $(\tilde{\mu}_1(X_t - X_t'))$ and $(\mathbb{E}[\tilde{\mu}_1(X_t - X_0')|X_t])$ which inherit α-mixing from (X_t) with the same rate function. For example, a standard inequality ([8], Section 1.2.2, Theorem 3(a)) yields that $T_\mu^X(h) \le c\alpha_h^{1/r}(\mathbb{E}[(\tilde{\mu}_1(X_0 - X_0'))^u])^{2/u}$ for positive c and $r > 0$ such that $r^{-1} + 2u^{-1} = 1$. If $\tilde{\mu}_1$ is bounded we also have $T_\mu^X(h) \le c\alpha_h$ for some positive constant. Similar bounds can be found for $T_\mu^{X,Y}(h)$ provided $((X_t, Y_t))$ is α-mixing.

Next, we give an example where the ADCVF can be calculated explicitly.

Example 3.6. Consider a univariate strictly stationary Gaussian time series (X_t) with mean zero, variance σ^2 and autocovariance function γ_X. We choose a Gaussian probability measure μ which leads to the relation (2.17). Choose N_1, N_2, N_3 i.i.d. $N(0, 2)$-distributed independent of the independent quantities (X_0, X_h), (X_0', X_h'), X_h''. Then for $h \ge 0$,

$$\begin{aligned} T_\mu^X(h) &= \mathbb{E}\big[\mathrm{e}^{iN_1(X_0 - X_0') + iN_2(X_h - X_h')}\big] + \big(\mathbb{E}\big[\mathrm{e}^{iN_1(X_0 - X_0')}\big]\big)^2 \\ &\quad - 2\mathbb{E}\big[\mathrm{e}^{iN_1(X_0 - X_0') + iN_2(X_h - X_h'')}\big] \\ &= \mathbb{E}\big[\mathrm{e}^{i(N_1X_0 + N_2X_h) - i(N_1X_0' + N_2X_h')}\big] + \big(\mathbb{E}\big[\mathrm{e}^{iN_1(X_0 - X_0')}\big]\big)^2 \\ &\quad - 2\mathbb{E}\big[\mathrm{e}^{i(N_1X_0 + N_2X_h) - i(N_1X_0' + N_2X_h'')}\big] \\ &= \mathbb{E}\big[\mathrm{e}^{iN_3(N_1^2\sigma^2 + N_2^2\sigma^2 + 2\gamma_X(h)N_1N_2)^{1/2}}\big] + \big(\mathbb{E}\big[\mathrm{e}^{iN_3(N_1^2\sigma^2)^{1/2}}\big]\big)^2 \\ &\quad - 2\mathbb{E}\big[\mathrm{e}^{iN_3(N_1^2\sigma^2 + N_2^2\sigma^2 + \gamma_X(h)N_1N_2)^{1/2}}\big] \\ &= \mathbb{E}\big[\mathrm{e}^{-(N_1^2\sigma^2 + N_2^2\sigma^2 + 2\gamma_X(h)N_1N_2)}\big] + \big(\mathbb{E}\big[\mathrm{e}^{-N_1^2\sigma^2}\big]\big)^2 \\ &\quad - 2\mathbb{E}\big[\mathrm{e}^{-(N_1^2\sigma^2 + N_2^2\sigma^2 + \gamma_X(h)N_1N_2)}\big]. \end{aligned}$$

For the evaluation of this expression, we focus on the first term, the other cases being similar. Observing that $\sigma^2 \pm \gamma_X(h)$ are the eigenvalues of the covariance matrix

$$\begin{pmatrix} \sigma^2 & \gamma_X(h) \\ \gamma_X(h) & \sigma^2 \end{pmatrix},$$

calculation shows that

$$N_1^2\sigma^2 + N_2^2\sigma^2 + 2\gamma_X(h)N_1N_2 \stackrel{d}{=} N_1^2\big(\sigma^2 - \gamma_X(h)\big) + N_2^2\big(\sigma^2 + \gamma_X(h)\big).$$

Now the moment generating function of a χ^2-distributed random variable yields

$$\mathbb{E}\big[\mathrm{e}^{-(N_1^2\sigma^2+N_2^2\sigma^2+2\gamma_X(h)N_1N_2)}\big] = \big(1+4\big(\sigma^2-\gamma_X(h)\big)\big)^{-1/2}\big(1+4\big(\sigma^2+\gamma_X(h)\big)\big)^{-1/2}.$$

Proceeding in a similar fashion, we obtain

$$\begin{aligned} T_\mu^X(h) &= \big(1+4\big(\sigma^2-\gamma_X(h)\big)\big)^{-1/2}\big(1+4\big(\sigma^2+\gamma_X(h)\big)\big)^{-1/2} + \big(1+4\sigma^2\big)^{-1} \\ &\quad - 2\big(1+4\big(\sigma^2-\gamma_X(h)/2\big)\big)^{-1/2}\big(1+4\big(\sigma^2+\gamma_X(h)/2\big)\big)^{-1/2}. \end{aligned}$$

If $\gamma_X(h)\to 0$ as $h\to\infty$ Taylor expansions yield $T_\mu^X(h)\sim 4\gamma_X^2(h)/(1+4\sigma^2)^3$. A similar result was given in [13], where they derived an explicit expression for $T_\mu^X(h)$ for a stationary Gaussian process (X_t) with weight function (1.2).

If $((X_t,Y_t))$ is strictly stationary and ergodic then $((X_t,Y_{t+h}))$ is a strictly stationary ergodic sequence for every integer h. Then Theorem 3.1 applies.

Corollary 3.7. *Under the conditions of Theorem 3.1, for $h\ge 0$,*

$$T_{n,\mu}^{X,Y}(h) \overset{\text{a.s.}}{\to} T_\mu^{X,Y}(h) \quad \textit{and} \quad T_{n,\mu}^{X}(h) \overset{\text{a.s.}}{\to} T_\mu^{X}(h),$$

and

$$R_{n,\mu}^{X,Y}(h) \overset{\text{a.s.}}{\to} R_\mu^{X,Y}(h) \quad \textit{and} \quad R_{n,\mu}^{X}(h) \overset{\text{a.s.}}{\to} R_\mu^{X}(h).$$

Applying Theorem 3.2 and Theorem 3.1, we also have the following weak dependence result under α-mixing. Zhou [30] proved the corresponding result under conditions on the so-called *physical dependence measure*.

Corollary 3.8. *Assume that X_0 and Y_h are independent for some $h\ge 0$ and the sequence $((X_t,Y_t))$ satisfies the conditions of Theorem 3.2. Then*

$$nT_{n,\mu}^{X,Y}(h) \overset{d}{\to} \|G_h\|_\mu^2 \quad \textit{and} \quad nR_{n,\mu}^{X,Y}(h) \overset{d}{\to} \frac{\|G_h\|_\mu^2}{\sqrt{T_\mu^X(0)T_\mu^Y(0)}},$$

where G_h is a centered Gaussian process on $\mathbb{R}^{p+q}$.

Remark 3.9. From the proof of Theorem 3.2 (the central limit theorem for the multivariate empirical characteristic function) it follows that G_h has covariance function

$$\begin{aligned} \Gamma\big((s,t),(s',t')\big) &= \operatorname{cov}\big(G_h(s,t),G_h(s',t')\big) \\ &= \sum_{j\in\mathbb{Z}} \mathbb{E}\big[\big(\mathrm{e}^{i\langle s,X_0\rangle}-\varphi_X(s)\big)\big(\mathrm{e}^{i\langle t,Y_h\rangle}-\varphi_Y(t)\big) \\ &\qquad \times \big(\mathrm{e}^{-i\langle s',X_j\rangle}-\varphi_X(-s')\big)\big(\mathrm{e}^{-i\langle t',Y_{j+h}\rangle}-\varphi_Y(-t')\big)\big]. \end{aligned} \tag{3.9}$$

In the special case when (X_t) and (Y_t) are independent sequences G_h is the same across all h with covariance function

$$\Gamma\big((s,t),(s',t')\big) = \big(\varphi_X(s-s') - \varphi_X(s)\varphi_X(s')\big)\big(\varphi_Y(t-t') - \varphi_Y(t)\varphi_Y(t')\big).$$

Since G_h is centered Gaussian its squared L^2-norm $\|G_h\|_\mu^2$ has a weighted χ^2-distribution; see [19], Chapter 1. The distribution of $\|G_h\|_\mu^2$ is not tractable and therefore one needs resampling methods for determining its quantiles.

Remark 3.10. Corollary 3.8 can be extended to the joint convergence of the function $nT_{n,\mu}^{X,Y}(h)$ at finitely many lags h, provided X_0 and Y_h are independent for these lags.

Remark 3.11. Corollary 3.8 does not apply when X_0 and Y_h are dependent. Then $nT_{n,\mu}^{X,Y}(h) \to \infty$ a.s. and $nR_{n,\mu}^{X,Y}(h) \to \infty$ a.s.

4. Auto-distance covariance of fitted residuals from AR(p) process

An often important problem in time series is to assess the goodness-of-fit of a particular model. As an illustration, consider a causal autoregressive process of order p (AR(p)) given by the difference equations,

$$X_t = \sum_{k=1}^{p} \phi_k X_{t-k} + Z_t, \qquad t = 0, \pm 1, \dots,$$

where (Z_t) is an i.i.d. sequence with a finite moment $\mathbb{E}[|Z|^\kappa] < \infty$ for some $\kappa > 0$. It is further assumed Z_t has mean 0 if $\kappa \geq 1$. It is often convenient to write the AR(p) process in the form, $Z_t = X_t - \boldsymbol{\phi}^T \mathbf{X}_{t-1}$, where $\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)^T$, $p \geq 1$ and $\mathbf{X}_t = (X_t, \dots, X_{t-p+1})^T$. Since the process is assumed causal, we can write $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ for absolutely summable constants (ψ_j); see [3], p. 85. For convenience, we also write $\psi_j = 0$ for $j < 0$ and $\psi_0 = 1$.

The least-squares estimator $\hat{\boldsymbol{\phi}}$ of $\boldsymbol{\phi}$ satisfies the relation

$$\hat{\boldsymbol{\phi}} - \boldsymbol{\phi} = \Gamma_{n,p}^{-1} \frac{1}{n} \sum_{t=p+1}^{n} \mathbf{X}_{t-1} Z_t, \qquad \text{where } \Gamma_{n,p} = \frac{1}{n} \sum_{t=p+1}^{n} \mathbf{X}_{t-1}^T \mathbf{X}_{t-1}.$$

If $\sigma^2 = \mathrm{var}(Z_t) < \infty$, we have by the ergodic theorem,

$$\Gamma_{n,p} \xrightarrow{\text{a.s.}} \Gamma_p = \big(\gamma_X(j-k)\big)_{1 \leq j,k \leq p}, \qquad \text{where } \gamma_X(h) = \mathrm{cov}(X_0, X_h), h \in \mathbb{Z}. \tag{4.1}$$

Causality of the process implies that the partial sum $\sum_{t=p+1}^{n} \mathbf{X}_{t-1} Z_t$ is a martingale and applying the martingale central limit theorem yields

$$\sqrt{n}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}) \xrightarrow{d} \mathbf{Q}, \tag{4.2}$$

where $\mathbf{Q}$ is $N(\mathbf{0}, \sigma^2 \Gamma_p^{-1})$ distributed.

The residuals of the fitted model are given by

$$\widehat{Z}_t = X_t - \hat{\boldsymbol{\phi}}^T \mathbf{X}_{t-1} = (\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})^T \mathbf{X}_{t-1} + Z_t, \qquad t = p+1, \dots, n. \tag{4.3}$$

For convenience, we set $\widehat{Z}_t = 0$, $t = 1, \dots, p$ since this choice does not influence the asymptotic theory. Each of the residuals $\widehat{Z}_t$ depends on the estimated parameters and hence the residual process exhibits serial dependence. Nevertheless, we might expect the test statistic based on the distance covariance function of the residuals given by

$$T_{n,\mu}^{\widehat{Z}}(h) = \int_{\mathbb{R}} \left| C_n^{\widehat{Z}}(s,t) \right|^2 \mu(ds, dt)$$

to behave in a similar fashion for the true noise sequence (Z_t). If the model is a good fit, then we would not expect $T_{n,\mu}^{\widehat{Z}}(h)$ to be extraordinarily large. As observed by [24], the limit distributions for $T_{n,\mu}^{\widehat{Z}}(h)$ and $T_{n,\mu}^{Z}(h)$ are not the same. As might be expected, the residuals, which are fitted to the actual data, tend to have smaller distance covariance than the true noise terms for lags less than p, if the model is correct. As a result, one can fashion a goodness-of-fit test based on applying the distance covariance statistics to the residuals. In the following theorem, we show that the distance covariance based on the residuals has a different limit than the distance covariance based on the actual noise, if the process has a finite variance. So in applying a goodness-of-fit test, one must make an adjustment to the limit distribution. Interestingly, if the noise has heavy-tails, the limits based on the residuals and the noise terms are the same and no adjustment is necessary.

For the formulation of the next result, we need some auxiliary limit theory; the proofs are given in the supplementary material in [7].

Lemma 4.1. *Consider an i.i.d. sequence (Z_t) with finite variance.*

1. *For every $h \geq 0$,*

$$\sqrt{n}\big(C_n^Z, \hat{\boldsymbol{\phi}} - \boldsymbol{\phi}\big) \xrightarrow{d} (G_h, \mathbf{Q}),$$

where the convergence is in $\mathcal{C}(K) \times \mathbb{R}^p$, $K \subset \mathbb{R}^2$ is a compact set, G_h is the limit process of C_n^Z with covariance structure specified in Remark 3.9 for the sequence $((Z_t, Z_{t+h}))$, $\mathbf{Q}$ is the limit in (4.2), $(G_h, \mathbf{Q})$ are mean-zero and jointly Gaussian with covariance matrix

$$\operatorname{cov}\big(G_h(s,t), \mathbf{Q}\big) = -\varphi_Z'(s)\varphi_Z'(t)\Gamma_p^{-1}\Psi_h, \qquad s, t \in \mathbb{R}, \tag{4.4}$$

where $\Psi_h = (\psi_{h-j})_{j=1,\dots,p}$ and φ_Z' is the first derivative of φ_Z.

2. *For every $h \geq 0$,*

$$\sqrt{n}\big(C_n^Z, C_n^{\widehat{Z}} - C_n^Z\big) \xrightarrow{d} (G_h, \xi_h),$$

where $(G_h, \mathbf{Q})$ are specified in (4.4) and

$$\xi_h(s,t) = t\varphi_Z(t)\varphi_Z'(s)\Psi_h^T \mathbf{Q}, \qquad (s,t) \in K, \tag{4.5}$$

the convergence is in $\mathcal{C}(K, \mathbb{R}^2)$, $K \subset \mathbb{R}^2$ *is a compact set. In particular, we have*

$$\sqrt{n} C_n^{\widehat{Z}} \xrightarrow{d} G_h + \xi_h. \tag{4.6}$$

Now we can formulate the following result; the proof is given in the supplementary material in [7].

Theorem 4.2. *Consider a causal* AR(p) *process with i.i.d. noise* (Z_t). *Assume*

$$\int_{\mathbb{R}^2} \big[\big(1 \wedge |s|^2\big)\big(1 \wedge |t|^2\big) + \big(s^2 + t^2\big)\mathbf{1}\big(|s| \wedge |t| > 1\big)\big]\mu(ds, dt) < \infty. \tag{4.7}$$

1. *If* $\sigma^2 = \mathrm{Var}(Z) < \infty$, *then*

$$nT_{n,\mu}^{\widehat{Z}}(h) \xrightarrow{d} \|G_h + \xi_h\|_\mu^2 \quad \textit{and} \quad nR_{n,\mu}^{\widehat{Z}}(h) \xrightarrow{d} \frac{\|G_h + \xi_h\|_\mu^2}{T_\mu^Z(0)}, \tag{4.8}$$

where (G_h, ξ_h) *are jointly Gaussian limit random fields on* $\mathbb{R}^2$. *The covariance structure of* G_h *is specified in Remark 3.9 for the sequence* $((Z_t, Z_{t+h}))$, ξ_h *and the joint limit structure of* (G_h, ξ_h) *are given in Lemma 4.1.*

2. *Assume that* Z *is in the domain of attraction of a stable law of index* $\alpha \in (0, 2)$, *that is,* $\mathbb{P}(|Z| > x) = x^{-\alpha} L(x)$ *for* $x > 0$, $L(\cdot)$ *is a slowly varying function at* ∞, *and*

$$\frac{\mathbb{P}(Z > x)}{\mathbb{P}(|Z| > x)} \to p \quad \textit{and} \quad \frac{\mathbb{P}(Z < -x)}{\mathbb{P}(|Z| > x)} \to 1 - p$$

as $x \to \infty$ *for some* $p \in [0, 1]$ ([10], *p.* 313). *Then we have*

$$nT_{n,\mu}^{\widehat{Z}}(h) \xrightarrow{d} \|G_h\|_\mu^2 \quad \textit{and} \quad nR_{n,\mu}^{\widehat{Z}}(h) \xrightarrow{d} \frac{\|G_h\|_\mu^2}{T_\mu^Z(0)}, \tag{4.9}$$

where G_h *is a Gaussian limit random field on* $\mathbb{R}^2$. *The covariance structure of* G_h *is specified in Remark 3.9 for the sequence* $((Z_t, Z_{t+h}))$.

Remark 4.3. Rémillard [24] mentioned that $T_{n,\mu}^Z(h)$ and $T_{n,\mu}^{\widehat{Z}}(h)$ for an AR(1) process have distinct limit processes and he also suggested the limiting structure in (4.8).

Remark 4.4. The proof of Theorem 4.2 requires knowledge of the precise form of the AR parameter estimates. We believe that this result, especially the limit in (4.8) can be extended to cover ARMA processes and some non-linear processes that are invertible. This is the subject of ongoing research.

The structure of the limit process in (4.8) is rather implicit. In applications, one needs to rely on resampling methods. Relation (4.8) can be extended to a joint convergence result for finitely

many lags h but the dependence structure of the limiting vectors is even more involved. Condition (4.7) holds for probability measures $\mu = \mu_1 \times \mu_1$ on $\mathbb{R}^2$ with finite second moment but it does not hold for the benchmark measure $\mu = \mu_1 \times \mu_1$ described in (2.12). A reason for this is that $\|\xi_h\|_\mu^2$ is in general not well defined in this case. If Z_t has characteristic function φ_Z then by virtue of (4.5), $\|\xi_h\|_\mu^2$ is finite a.s. if and only if

$$\int_{-\infty}^{\infty} \left|t\varphi_Z(t)\right|^2 \mu_1(dt) \int_{-\infty}^{\infty} \left|\varphi_Z'(s)\right|^2 \mu_1(ds) < \infty.$$

Now assume that Z_t has a density function f and choose $\mu_1(dt) = c_1 t^{-2}\,dt$. Then by Plancherel's identity, the first integral becomes

$$\int_{-\infty}^{\infty} \left|\varphi_Z(t)\right|^2 dt = c \int_{-\infty}^{\infty} f^2(t)\,dt.$$

If one chooses f to be a symmetric gamma distribution with shape parameter $\delta \in (0, 1/2)$, that is, $f(z) = 0.5\beta^\delta |z|^{\delta-1} \mathrm{e}^{-|z|\beta}/\Gamma(\delta)$, then the integral $\int_{-\infty}^{\infty} f^2(t)\,dt$ is infinity and hence the limit random variable in (4.8) cannot be finite.

AR simulation

We illustrate the results of Theorem 4.2. First, we generate independent replications of a time series $(X_t)_{t=1,\dots,1000}$ from a causal AR(10) model with $Z_t \sim N(0, 1)$ and

$$\boldsymbol{\phi} = (-0.140, 0.038, 0.304, 0.078, 0.069, 0.013, 0.019, 0.039, 0.148, -0.062).$$

In this and the following examples, we choose the weight measure $\mu = \mu_1 \times \mu_2$, where μ_i is the $N(0, 0.5)$-distribution and hence (4.7) is satisfied. From the independent replications of the simulated residuals, we approximate the limit distribution $\|G_h + \xi_h\|_\mu^2 / T_\mu^Z(0)$ of $nR_{n,\mu}^{\widehat{Z}}(h)$ by the corresponding empirical distribution.

The left graph in Figure 1 shows the box-plots for $nR_{n,\mu}^{\widehat{Z}}(h)$ based on 1000 replications from the AR(10) model, each with sample size $n = 1000$. As seen from the plots, the distribution at each lag is heavily skewed. In the right panel of Figure 1, we compare the empirical 5%, 50%, 95% quantiles of $nR_{n,\mu}^{\widehat{Z}}(h)$ to those of $nR_{n,\mu}^{Z}(h)$, the scaled ADCF of i.i.d. noise, all of which have the same limit, $\|G_h\|_\mu^2 / T_\mu^Z(0)$. The asymptotic variance of the ADCF of the residuals is smaller than that of i.i.d. noise at initial lags, and gradually increases at larger lags to the values in the i.i.d. case. This behavior is similar to that of the ACF of the residuals of an AR process; see for example Chapter 9.4 of [3].

Theorem 4.2 provides a visual tool for testing the goodness-of-fit of an AR(p) model, by examining the serial dependence of the residuals after model fitting. Under the null hypothesis, we expect $nR_{n,\mu}^{\widehat{Z}}(h)$ to be well bounded by the 95% quantiles of the limit distribution $\|G_h + \xi_h\|_\mu^2 / T_\mu^Z(0)$. For a single time series, this quantity can be approximated using a parametric bootstrap (generating an AR(10) process from the estimated parameters and residuals); see for

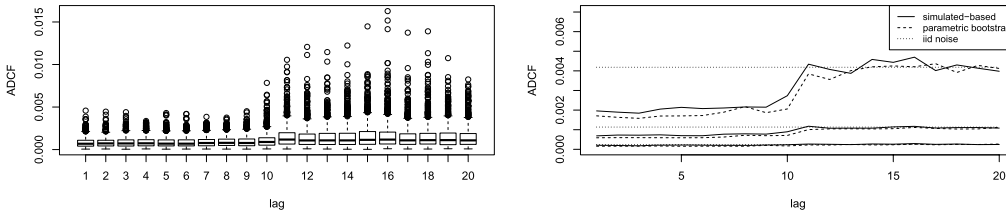

Figure 1. Distribution of $nR^{\widehat{Z}}_{n,\mu}(h)$, $n = 1000$ for the residuals of an AR(10) process with $N(0, 1)$ innovations. Left: Box-plots from 1000 independent replications. Right: 5%, 50%, 95% empirical quantiles of $nR^{\widehat{Z}}_{n,\mu}(h)$ based on simulated residuals, on resampled residuals and on i.i.d. noise, respectively. The weight measure is $\mu = \mu_1 \times \mu_2$, with each $\mu_i \sim N(0, 0.5)$.

example [23]. In the right graph of Figure 1, we overlay the empirical 5%, 50%, 95% quantiles of $nR^{\widehat{Z}}_{n,\mu}(h)$ estimated from one particular realization of the time series. As can be seen in the graph, the parametric bootstrap provides a good approximation to the actual quantiles found via simulation. On the other hand, the quantiles found by simply bootstrapping the residuals provides a rather poor approximation, at least for the first 10 lags.

We now consider the same AR(10) model as before, but with noise having a t-distribution with 1.5 degrees of freedom. (Here the noise is in the domain of attraction of a stable distribution with index 1.5.) The left graph of Figure 2 shows the box-plots of $nR^{\widehat{Z}}_{n,\mu}(h)$ based on 1000 replications, and the right graph shows the 5%, 50%, 95% quantiles of $nR^{\widehat{Z}}_{n,\mu}(h)$ and $nR^{Z}_{n,\mu}(h)$, both of which have the same limit distribution $\|G_h\|^2_\mu / T^Z_\mu(0)$. In this case, the quantiles of $\|G_h\|^2_\mu / T^Z_\mu(0)$ can be approximated naively by bootstrapping the fitted residuals $(\widehat{Z}_t)$ of the AR model. The left graph of Figure 2 overlays the 5%, 50%, 95% quantiles from bootstrapping with those from the simulations. The agreement is reasonably good.

We next provide an empirical example illustrating the limitation of using the measure in (2.12). Again, we use the same AR(10) model as before, but with noise now generated from the symmetric gamma distribution with $\delta = 0.2$, $\beta = 0.5$. The corresponding pair of graphs with boxplots and quantiles for $nR^{\widehat{Z}}_{n,\mu}(h)$ is displayed in Figure 3. Notice now that the box plots for the sam-

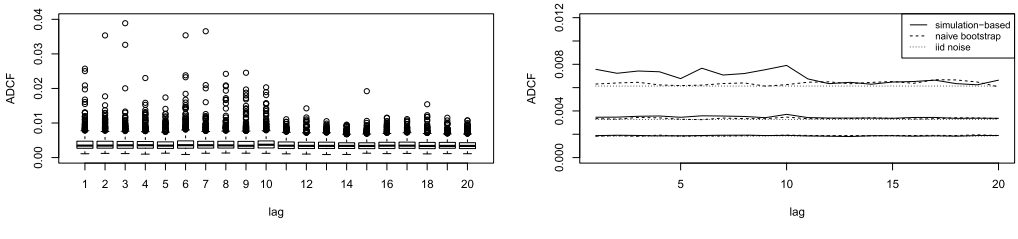

Figure 2. Distribution of $nR^{\widehat{Z}}_{n,\mu}(h)$ for residuals of AR process with $t_{1.5}$ innovations. Left: lag-wise box–plots. Right panel: empirical 5%, 50%, 95% quantiles from simulated residuals, empirical quantiles from resampled residuals, and empirical quantiles from i.i.d. noise. The weight measure is $\mu = \mu_1 \times \mu_2$, with each $\mu_i \sim N(0, 0.5)$.

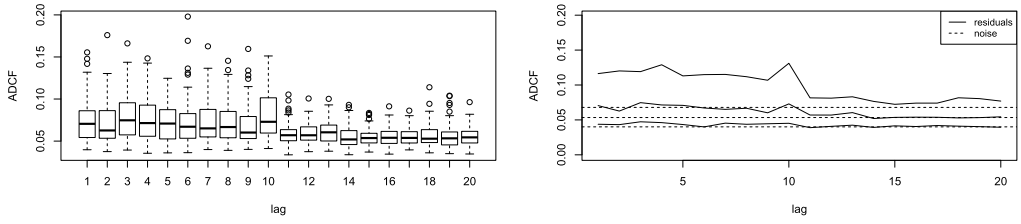

Figure 3. Distribution of $nR^{\widehat{Z}}_{n,\mu}(h)$, $n = 1000$ for residuals of AR process with a symmetric Gamma(0.2, 0.5) noise. Left: box-plots from 500 independent replications. Right panel: empirical 5%, 50%, 95% quantiles from simulated residuals and from i.i.d. noise. The measure μ is given by (2.12).

pling distribution of the distance correlation for the first 10 lags are rather spread out compared to those at lags greater than 10. In particular, the sampling behavior of these distance correlations is directly opposite of what we observed in Figure 1 where a finite measure was used. To further illustrate this disparity, the plot on the right in Figure 3 displays the 95%, 50%, 5% quantiles for the companion box plots (the dotted lines are the corresponding quantiles for i.i.d. noise with the Gamma(0.2,0.4) distribution). Now, compared to quantiles of distance correlation based on the i.i.d. noise, we see a stark difference. The median for the estimates based on the residuals using the weight function in (2.12) is nearly the same as the 95% quantile for the noise at lags 1–10. This illustrates the problem with using (2.12) as a weight function applied to the residuals.

5. Data examples

5.1. Amazon daily returns

In this example, we consider the daily stock returns of Amazon from 05/16/1997 to 06/16/2004. Denoting the series by (X_t), Figure 4 shows the ACF of (X_t), (X_t^2), $(|X_t|)$ and ADCF of (X_t) with weight measure $\mu(ds, dt) = s^{-2}t^{-2}\,ds\,dt$. In the right panel, we compare the ADCF with the 5%, 50%, 95% confidence bounds of the ADCF for i.i.d. data, approximated by the corresponding empirical quantiles from 1000 random permutations. With most financial time series, which are typically uncorrelated, serial dependence can be detected by examining the ACF of the absolute values and squares. Interestingly for the Amazon data, the ACF of the squared data also fails to pick up any signal. On the other hand, the ADCF has no trouble detecting serial dependence without having to resort to applying any transformation.

5.2. Wind speed data

For the next example, we consider the daily averages of wind speeds at Kilkenny's synoptic meteorological station in Ireland. The time series consists of 6226 observations from 1/1/1961 to 1/17/1978, after which a square root transformation has been applied to stabilize the variance. This transformation has also been suggested in previous studies (see, for example, [14]). The

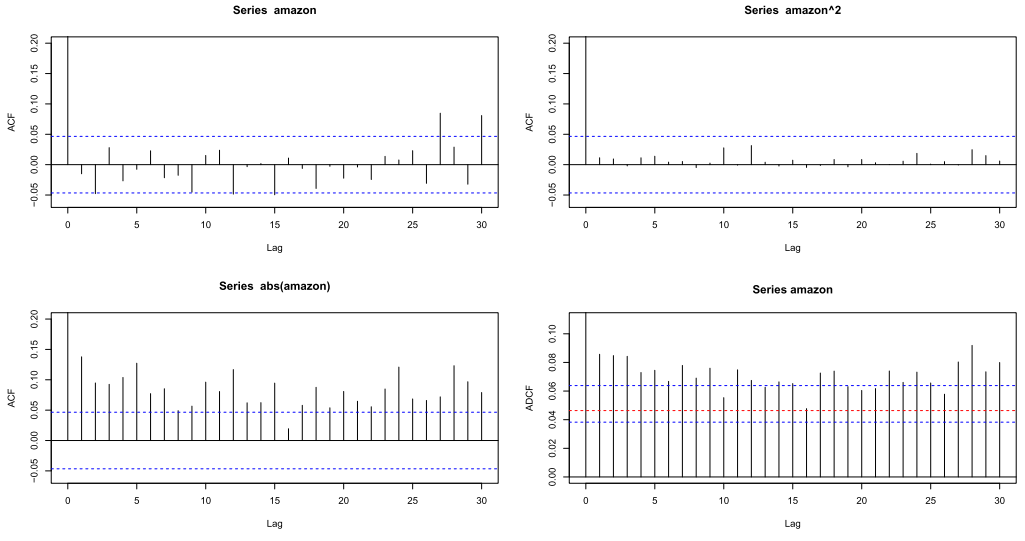

Figure 4. ACF and ADCF of daily stock returns of Amazon (X_t) from 05/16/1997 to 06/16/2004. Upper left: ACF of (X_t); Upper right: ACF of (X_t^2); Lower left: ACF of $(|X_t|)$; Lower right: ADCF of (X_t), the 5%, 50%, 95% confidence bounds of ADCF from randomly permuting the data.

ACF of the data, displayed in Figure 5, suggests a possible AR model for the data. An AR(9) model was found to provide the best fit (in terms of minimizing AICC among all AR models) to the data. The ACF of the residuals (see upper right panel in Figure 5) shows that the serial correlation has been successfully removed. The ACF of the squared residuals and ADCF of the residuals are also plotted in the bottom panels Figure 5. For computation of the ADCF, we used the N(0,0.5) distribution for the weight measure, which satisfies the condition (4.7). The ADCF of the residuals is well bounded by the confidence bounds for the ADCF of iid noise, shown by the dotted line in the plot. Without adjusting these bounds for the residuals, one would be tempted to conclude that the AR model is a good fit. However, the adjusted bounds for the ADCF of residuals, represented by the solid line in the plot and computed using a parametric bootstrap, suggest that some ADCF values among the first 8 lags are in fact larger than expected. Hence, this sheds some doubt on the validity of an AR(9) model with iid noise for this data. A similar conclusion can be reached by inspecting the ACF of the squares of the residuals (see lower left panel in Figure 5).

One potential remedy for the lack of fit of the AR(9) model, is to consider a GARCH(1, 1) model applied to the residuals. The GARCH model performs well in devolatilizing the AR-fitted residuals and no trace of a signal could be detected through the ACF of the GARCH-residuals applied to the squares and absolute values. The ADCF of the devolatilized residuals, seen in Figure 6, still presents some evidence of dependence. Here the confidence bounds plotted are for i.i.d. observations, obtained from 1000 random permutations of the GARCH-residuals and as such do not include an adjustment factor. Ultimately, a periodic AR model, which allows for periodicity in both the AR parameters and white noise variance might be a more desirable model.

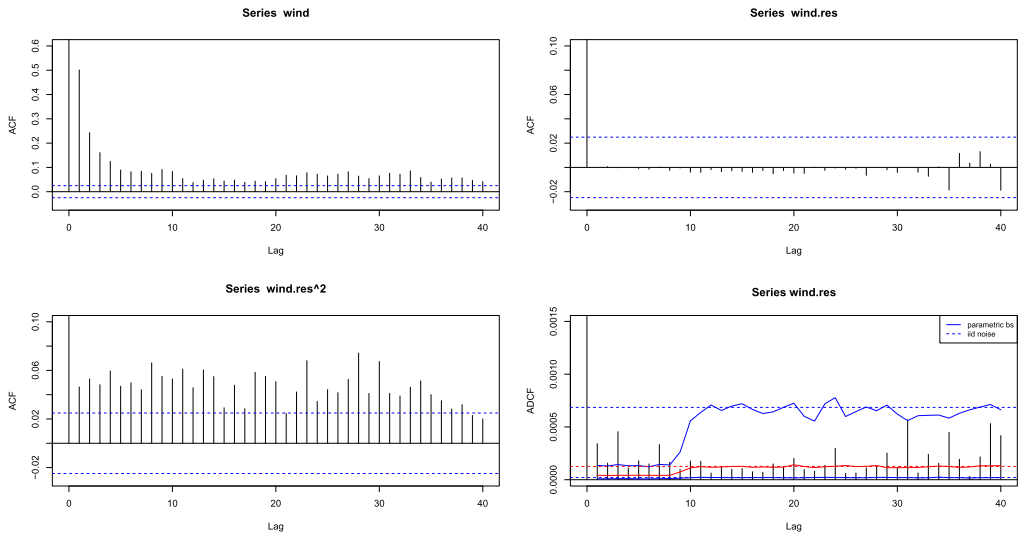

Figure 5. ACF and ADCF of Kilkenny wind speed time series and AR(9) fitted residuals. Upper left: ACF of the series. Upper right: ACF of the residuals. Lower left: ACF of the residual squares. Lower right: ADCF of the residuals, the 5%, 50%, 95% confidence bounds of ADCF for fitted residuals from 1000 parametric bootstraps, and that for iid noise from 1000 random permutations.

Appendix: Proof of Theorem 3.2

The proof follows from the following lemma.

Lemma A.1. *Assume that* $\sum_h \alpha_h^{1/r} < \infty$ *for some* $r > 1$ *and set* $u = 2r/(r-1)$. *We also assume the moment conditions* (3.2) (*or* (3.5)) *for some* $\alpha > 0$ *if* X_0 *and* Y_0 *are independent* (*dependent*).

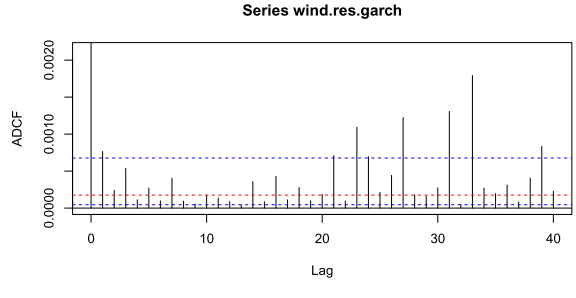

Figure 6. ADCF of the residuals of Kilkenny wind speed time series from AR(9)-GARCH fitting and the 5%, 50%, 95% confidence bounds of ADCF for i.i.d. noise from 1000 random permutations.

1. *For $\alpha \le 2$ there exists a constant $c > 0$ such that for $\epsilon \in [0, 1/2)$,*

$$n\mathbb{E}\big[\big|C_n(s,t) - C(s,t)\big|^2\big] \le c\big(1 \wedge |s|^{\alpha(1+\epsilon)/u}\big)\big(1 \wedge |t|^{\alpha(1+\epsilon)/u}\big), \qquad n \ge 1. \quad \text{(A.1)}$$

2. *If $\alpha \in (u/2, u]$ then $\sqrt{n}(\varphi^n_{X,Y} - \varphi_{X,Y}) \xrightarrow{d} G$ on compact sets $K \subset \mathbb{R}^{p+q}$ for some complex-valued mean-zero Gaussian field G.*

Remark A.2. Notice that $C(s,t) = 0$ when X_0 and Y_0 are independent.

Proof. (1) We focus on the proof under the assumption of independence. At the end, we indicate the changes necessary when X_0 and Y_0 are dependent.

We write

$$U_k = \mathrm{e}^{i\langle s, X_k\rangle} - \varphi_X(s), \qquad V_k = \mathrm{e}^{i\langle t, Y_k\rangle} - \varphi_Y(t), \qquad k \ge 1,$$

where we suppress the dependence of U_k and V_k on s and t, respectively. Then

$$\begin{aligned} n\mathbb{E}\big[\big|C_n(s,t)\big|^2\big] &= n\mathbb{E}\left|\frac{1}{n}\sum_{k=1}^n U_k V_k - \frac{1}{n}\sum_{k=1}^n U_k \frac{1}{n}\sum_{l=1}^n V_l\right|^2 \\ &\le 2n\mathbb{E}\left[\left|\frac{1}{n}\sum_{k=1}^n U_k V_k\right|^2\right] + 2n\mathbb{E}\left[\left|\frac{1}{n}\sum_{k=1}^n U_k \frac{1}{n}\sum_{l=1}^n V_l\right|^2\right] =: 2(I_1 + I_2). \end{aligned}$$

We have by stationarity

$$I_1 = \mathbb{E}\big[|U_0 V_0|^2\big] + 2\sum_{h=1}^{n-1}(1 - h/n)\,\mathrm{Re}\,\mathbb{E}[U_0 V_0 \overline{U_h V_h}].$$

Since U_0 and V_0 are independent $\mathbb{E}[U_0 V_0] = 0$. In view of the α-mixing condition (see [8], Section 1.2.2, Theorem 3(a)) we have

$$\begin{aligned} \big|\mathrm{Re}\,\mathbb{E}[U_0 V_0 \overline{U_h V_h}]\big| &\le c\alpha_h^{1/r}\big(\mathbb{E}\big[|U_0 V_0|^u\big]\big)^{2/u} \\ &= c\alpha_h^{1/r}\big(\mathbb{E}\big[|U_0|^u\big]\big)^{2/u}\big(\mathbb{E}\big[|V_0|^u\big]\big)^{2/u} \\ &\le c\alpha_h^{1/r}\big(\mathbb{E}\big[|U_0|^2\big]\big)^{2/u}\big(\mathbb{E}\big[|V_0|^2\big]\big)^{2/u}. \end{aligned} \quad \text{(A.2)}$$

In the last step we used that $u = 2r/(r-1) > 2$ and that $\max(|U_0|, |V_0|) \le 2$. We have for $\alpha \in (0, 2]$

$$\mathbb{E}\big[|U_0|^2\big] = 1 - \big|\varphi_X(s)\big|^2 \le \mathbb{E}\big[1 \wedge \big|\langle s, X - X'\rangle\big|^\alpha\big] \le c\big(1 \wedge |s|^\alpha\big).$$

Therefore and since $\sum_h \alpha_h^{1/r} < \infty$ we have $I_1 \le c(1 \wedge |s|^\alpha)^{2/u}(1 \wedge |t|^\alpha)^{2/u}$.

Now we turn to I_2. By the Cauchy–Schwarz inequality and since $|\frac{1}{n}\sum_{k=1}^n U_k|$ and $|\frac{1}{n}\sum_{k=1}^n V_k|$ are bounded by 2 we have

$$\begin{aligned} I_2 &\le 2n\left(\mathbb{E}\left|\frac{1}{n}\sum_{k=1}^n U_k\right|^4\right)^{1/2}\left(\mathbb{E}\left|\frac{1}{n}\sum_{k=1}^n V_k\right|^4\right)^{1/2} \\ &\le c\left(n\mathbb{E}\left|\frac{1}{n}\sum_{k=1}^n U_k\right|^{2+\delta}\right)^{1/2}\left(n\mathbb{E}\left|\frac{1}{n}\sum_{k=1}^n V_k\right|^{2+\delta}\right)^{1/2}, \end{aligned}$$

for any $\delta \in [0, 2]$. In view of Lemma 18.5.1 in [17], we have for $\delta \in [0, 1)$,

$$I_2 \le c\left(n\mathbb{E}\left|\frac{1}{n}\sum_{k=1}^n U_k\right|^2\right)^{(2+\delta)/4}\left(n\mathbb{E}\left|\frac{1}{n}\sum_{k=1}^n V_k\right|^2\right)^{(2+\delta)/4}.$$

Similar arguments as for I_1 show that

$$I_2 \le c\big(1 \wedge |s|^{\alpha(2+\delta)/4}\big)^{2/u}\big(1 \wedge |t|^{\alpha(2+\delta)/4}\big)^{2/u}.$$

Combining the bounds for I_1 and I_2, we arrive at (A.1).

Now we indicate the changes necessary when X_0 and Y_0 are dependent. We use the notation above and, additionally, write $\widetilde{W}_k = U_k V_k - C(s,t)$. We have

$$C_n(s,t) - C(s,t) = \frac{1}{n}\sum_{k=1}^n \widetilde{W}_k - \frac{1}{n}\sum_{k=1}^n U_k \frac{1}{n}\sum_{l=1}^n V_l.$$

Then

$$n\mathbb{E}\big[\big|C_n(s,t) - C(s,t)\big|^2\big] \le 2n\mathbb{E}\left[\left|\frac{1}{n}\sum_{k=1}^n \widetilde{W}_k\right|^2\right] + 2n\mathbb{E}\left[\left|\frac{1}{n}\sum_{k=1}^n U_k \frac{1}{n}\sum_{l=1}^n V_l\right|^2\right] = 2\big(I_1' + I_2\big).$$

Since $\mathbb{E}[\widetilde{W}_0] = 0$, we have by stationarity

$$I_1' = \mathbb{E}\big[|\widetilde{W}_0|^2\big] + 2\sum_{h=1}^{n-1}(1 - h/n)\operatorname{Re}\mathbb{E}[\widetilde{W}_0\overline{\widetilde{W}_h}].$$

Observe that $\mathbb{E}[|\widetilde{W}_0|^2] \le 2(\mathbb{E}|U_0|^4\mathbb{E}|V_0|^4)^{1/2} + 2|C(s,t)|^2$ and

$$\begin{aligned} |U_0|^2 &\le \big(\big|\mathrm{e}^{i\langle s, X_0\rangle} - 1\big| + \mathbb{E}\big[\big|1 - \mathrm{e}^{i\langle s, X_0\rangle}\big|\big]\big)^2 \\ &\le c\big(1 \wedge \big(|s||X_0|\big)^{\alpha/2}\big)^2 + c\big(1 \wedge \big(|s|^{\alpha/2}\mathbb{E}|X_0|^{\alpha/2}\big)\big)^2. \end{aligned}$$

Since $\mathbb{E}[|X_0|^{2\alpha}] < \infty$ we have $\mathbb{E}[|U_0|^4] \le c(1 \wedge |s|^{2\alpha})$ and in a similar manner, $\mathbb{E}|V_0|^4 \le c(1 \wedge |t|^{2\alpha})$. We also have $|C(s,t)|^2 \le c(1 \wedge |s|^\alpha)(1 \wedge |t|^\alpha)$. Finally, we conclude that

$$\mathbb{E}\big[|\widetilde{W}_0|^2\big] \le c\big(1 \wedge |s|^\alpha\big)\big(1 \wedge |t|^\alpha\big).$$

With the α-mixing condition we obtain

$$\big|\operatorname{Re}\mathbb{E}[\widetilde{W}_0\overline{\widetilde{W}_h}]\big| \le c\alpha_h^{1/r}\big(\mathbb{E}\big[|\widetilde{W}_0|^u\big]\big)^{2/u} \le c\alpha_h^{1/r}\big(\mathbb{E}\big[|\widetilde{W}_0|^2\big]\big)^{2/u}.$$

This together with $\sum_h \alpha_h^{1/r} < \infty$ yields $I_1' \le c(1 \wedge |s|^\alpha)^{2/u}(1 \wedge |t|^\alpha)^{2/u}$. The remaining term I_2 can be treated in the same way as in the independent case. Combining the bounds for I_1' and I_2, we arrive at (A.1).

(2) We need an analog of S. Csörgő's central limit theorem (Csörgő [4–6]) for the empirical characteristic function of an i.i.d. multivariate sequence with Gaussian limit. For ease of notation we focus on the X-sequence; the proof for the (X,Y)-sequence is analogous and therefore omitted. The convergence of the finite-dimensional distributions of $\sqrt{n}(\varphi_X^n - \varphi_X)$ follows from Theorem 18.5.2 in [17] combined with the Cramér–Wold device. We need to show tightness of the normalized empirical characteristic function on compact sets. We use the sufficient condition of Theorem 3 in [2] for multiparameter processes. We evaluate the process on cubes $(s,t] = \prod_{k=1}^p (s_k, t_k]$, where $s = (s_1, \dots, s_p)$ and $t = (t_1, \dots, t_p)$ and $s_i < t_i$, $i = 1, \dots, p$. The increment of the normalized empirical characteristic function on $(s,t]$ is given by

$$\begin{aligned} I_n(s,t] &= \sqrt{n}\big(\varphi_X^n(s,t] - \varphi_X(s,t]\big) \\ &= \frac{\sqrt{n}}{n}\sum_{r=1}^n \Bigg\{ \sum_{k_1=0,1} \cdots \sum_{k_p=0,1} (-1)^{p-\sum_j k_j} \Bigg(\prod_{l=1}^p \mathrm{e}^{i(s_l + k_l(t_l - s_l))X_r^{(l)}} \\ &\qquad - \mathbb{E}\Bigg[\prod_{l=1}^p \mathrm{e}^{i(s_l + k_l(t_l - s_l))X_r^{(l)}}\Bigg]\Bigg)\Bigg\} =: \frac{1}{\sqrt{n}}\sum_{r=1}^n W_r, \end{aligned} \tag{A.3}$$

where $X_r = (X_r^{(1)}, \dots, X_r^{(p)})$ and

$$W_r = \prod_{l=1}^p \big(\mathrm{e}^{it_l X_r^{(l)}} - \mathrm{e}^{is_l X_r^{(l)}}\big) - \mathbb{E}\Bigg[\prod_{l=1}^p \big(\mathrm{e}^{it_l X_r^{(l)}} - \mathrm{e}^{is_l X_r^{(l)}}\big)\Bigg].$$

We apply the sums $\sum_{k_j=0,1}$ inductively to derive (A.3). Observe that

$$\mathbb{E}\big[\big|I_n(s,t]\big|^2\big] = \mathbb{E}\big[|W_0|^2\big] + 2\sum_{h=1}^{n-1}(1 - h/n)\operatorname{Re}\mathbb{E}[W_0\overline{W}_h].$$

By the Lipschitz property of trigonometric functions we have for some constant $c > 0$ and $\alpha \in (0,2]$,

$$\big|\mathrm{e}^{is_l X_r^{(l)}} - \mathrm{e}^{it_l X_r^{(l)}}\big|^2 \le c\big(1 \wedge |t_l - s_l|^2\big(X_r^{(l)}\big)^2/4\big) \le c\big(1 \wedge |s_l - t_l|^\alpha\big|X_r^{(l)}\big|^\alpha/4^\alpha\big).$$

Proceeding as for (A.2) and noticing that $\alpha \le 2 \le u$, we have

$$\begin{aligned}\big|\mathbb{E}[W_0\overline{W}_h]\big| &\le c\alpha_h^{1/r}\big(\mathbb{E}\big[|W_0|^u\big]\big)^{2/u}\\ &\le \alpha_h^{1/r}\prod_{l=1}^{p}|s_l-t_l|^{2\alpha/u}\left(\mathbb{E}\left[\prod_{l=1}^{p}\big|X_0^{(l)}\big|^{\alpha}\right]\right)^{2/u}.\end{aligned}$$

Using the summability of $(\alpha_h^{1/r})$ and the moment condition on X_0, we may conclude that

$$\mathbb{E}\big[\big|I_n(s,t]\big|^2\big] \le c\prod_{l=1}^{p}|s_l-t_l|^{2\alpha/u}.$$

If $2\alpha/u > 1$ the condition of Theorem 3 in [2] yields that the processes $(\sqrt{n}(\varphi_X^n - \varphi_X))$ are tight on compact sets. □

Proof of Theorem 3.2(1). Recall the definition of K_δ from (3.1) and that X_0 and Y_0 are independent. From Lemma A.1 and the continuous mapping theorem, we have

$$\int_{K_\delta}\big|\sqrt{n}C_n(s,t)\big|^2\mu(ds,dt) \xrightarrow{d} \int_{K_\delta}\big|G(s,t)\big|^2\mu(ds,dt), \qquad n\to\infty.$$

From (3.3), (A.1) and the dominated convergence theorem, for any $\varepsilon > 0$, some $\epsilon \in (0, 1/2]$ and $\alpha' \le \min(2,\alpha)$,

$$\begin{aligned}&\lim_{\delta\downarrow 0}\limsup_{n\to\infty}\mathbb{P}\left(\int_{K_\delta^c}\big|\sqrt{n}C_n(s,t)\big|^2\mu(ds,dt) > \varepsilon\right)\\ &\quad\le \varepsilon^{-1}\lim_{\delta\downarrow 0}\limsup_{n\to\infty}\int_{K_\delta^c}\mathbb{E}\big[\big|\sqrt{n}C_n(s,t)\big|^2\big]\mu(ds,dt)\\ &\quad\le \lim_{\delta\downarrow 0}\int_{K_\delta^c}c\big(1\wedge|s|^{\alpha'(1+\epsilon)/u}\big)\big(1\wedge|t|^{\alpha'(1+\epsilon)/u}\big)\mu(ds,dt) = 0.\end{aligned}$$

□

Proof of Theorem 3.2(2). Now we assume that X_0 and Y_0 are dependent. We observe that

$$\sqrt{n}\big(T_n(s,t;\mu) - T(s,t;\mu)\big) = \int_{\mathbb{R}^{p+q}}\sqrt{n}\big(\big|C_n(s,t)\big|^2 - \big|C(s,t)\big|^2\big)\mu(ds,dt).$$

In view of Lemma A.1(2) and the a.s. convergence of C_n on compact sets the continuous mapping theorem implies that for some Gaussian mean-zero process G',

$$\begin{aligned}&\int_{K_\delta}\sqrt{n}\big\{\big(C_n(s,t)-C(s,t)\big)\overline{C}_n(s,t) + C(s,t)\big(\overline{C}_n(s,t)-\overline{C}(s,t)\big)\big\}\mu(ds,dt)\\ &\quad\xrightarrow{d}\int_{K_\delta}G'(s,t)\mu(ds,dt), \qquad n\to\infty,\end{aligned}$$

where $G'X(s,t) = 2\operatorname{Re}\{G(s,t)C(s,t)\}$. We have

$$\big||C_n|^2 - |C|^2\big| = \big||C_n - C|^2 + 2\operatorname{Re}\big(\overline{C}(C_n - C)\big)\big| \le c|C_n - C|.$$

By Markov's inequality, (A.1) and (3.3),

$$\begin{aligned}
&\lim_{\delta\downarrow 0}\limsup_{n\to\infty}\mathbb{P}\left(\int_{K_\delta^c}\sqrt{n}\big|\big|C_n(s,t)\big|^2 - \big|C(s,t)\big|^2\big|\mu(ds,dt) > \varepsilon\right)\\
&\quad\le c\lim_{\delta\downarrow 0}\limsup_{n\to\infty}\int_{K_\delta^c}\big(n\mathbb{E}\big[|C_n - C|^2\big]\big)^{1/2}\mu(ds,dt)\\
&\quad\le \lim_{\delta\downarrow 0}\int_{K_\delta^c} c\big(1\wedge|s|^{\alpha'(1+\epsilon)/u}\big)\big(1\wedge|t|^{\alpha'(1+\epsilon)/u}\big)\mu(ds,dt) = 0.
\end{aligned}$$

□

Acknowledgements

We would like to thank two anonymous referees for careful reading and constructive comments. This research by Davis and Wan was supported in part by Army MURI grant W911NF-12-1-0385. Muneya Matsui's research is partly supported by JSPS Grant-in-Aid for Young Scientists B (16K16023) and Nanzan University Pache Research Subsidy I-A-2 for the 2017 academic year. Thomas Mikosch's research is partly supported by the Danish Research Council Grant DFF-4002-00435.

Supplementary Material

Supplement to "Applications of distance correlation to time series" (DOI: 10.3150/17-BEJ955SUPP; .pdf). Complementary results to the proofs of Theorem 4.2 and Lemma 4.1 are provided in the supplement.

References

[1] Aaronson, J., Burton, R., Dehling, H., Gilat, D., Hill, T. and Weiss, B. (1996). Strong laws for L- and U-statistics. *Trans. Amer. Math. Soc.* **348** 2845–2866. MR1363941

[2] Bickel, P.J. and Wichura, M.J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Math. Stat.* **42** 1656–1670. MR0383482

[3] Brockwell, P.J. and Davis, R.A. (1991). *Time Series*: *Theory and Methods*, 2nd ed. *Springer Series in Statistics*. New York: Springer. MR1093459

[4] Csörgő, S. (1981). Limit behaviour of the empirical characteristic function. *Ann. Probab.* **9** 130–144. MR0606802

[5] Csörgő, S. (1981). Multivariate characteristic functions and tail behaviour. *Z. Wahrsch. Verw. Gebiete* **55** 197–202. MR0608016

[6] Csörgő, S. (1981). Multivariate empirical characteristic functions. *Z. Wahrsch. Verw. Gebiete* **55** 203–229. MR0608017

[7] Davis, R.A., Matsui, M., Mikosch, T. and Wan, P. (2017). Supplement to "Applications of distance correlation to time series." DOI:10.3150/17-BEJ955SUPP.
[8] Doukhan, P. (1994). *Mixing*: *Properties and Examples*. *Lecture Notes in Statistics* **85**. New York: Springer. MR1312160
[9] Dueck, J., Edelmann, D., Gneiting, T. and Richards, D. (2014). The affinely invariant distance correlation. *Bernoulli* **20** 2305–2330. MR3263106
[10] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*. *Vol. II*, 2nd ed. New York: Wiley. MR0270403
[11] Feuerverger, A. (1993). A consistent test for bivariate dependence. *Int. Stat. Rev.* **61** 419–433.
[12] Feuerverger, A. and Mureika, R.A. (1977). The empirical characteristic function and its applications. *Ann. Statist.* **5** 88–97. MR0428584
[13] Fokianos, K. and Pitsillou, M. (2017). Consistent testing for pairwise dependence in time series. *Technometrics* **59** 262–270. MR3635048
[14] Haslett, J. and Raftery, A.E. (1989). Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **38** 1–50.
[15] Hlávka, Z., Hušková, M. and Meintanis, S.G. (2011). Tests for independence in non-parametric heteroscedastic regression models. *J. Multivariate Anal.* **102** 816–827. MR2772338
[16] 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. MR1731483
[17] Ibragimov, I.A. and Linnik, Y.V. (1971). *Independent and Stationary Sequences of Random Variables*. Groningen: Wolters-Noordhoff Publishing. MR0322926
[18] Krengel, U. (1985). *Ergodic Theorems*. *De Gruyter Studies in Mathematics* **6**. Berlin: de Gruyter. MR0797411
[19] Kuo, H.H. (1975). *Gaussian Measures in Banach Spaces*. *Lecture Notes in Mathematics* **463**. Berlin: Springer. MR0461643
[20] Lyons, R. (2013). Distance covariance in metric spaces. *Ann. Probab.* **41** 3284–3305. MR3127883
[21] Meintanis, S.G. and Iliopoulos, G. (2008). Fourier methods for testing multivariate independence. *Comput. Statist. Data Anal.* **52** 1884–1895. MR2418478
[22] Meintanis, S.G., Ngatchou-Wandji, J. and Taufer, E. (2015). Goodness-of-fit tests for multivariate stable distributions based on the empirical characteristic function. *J. Multivariate Anal.* **140** 171–192. MR3372561
[23] Politis, D.N., Romano, J.P. and Wolf, M. (1999). *Subsampling*. *Springer Series in Statistics*. New York: Springer. MR1707286
[24] Rémillard, B. (2009). Discussion of: Brownian distance covariance [MR2752127]. *Ann. Appl. Stat.* **3** 1295–1298. MR2752133
[25] Samorodnitsky, G. (2016). *Stochastic Processes and Long Range Dependence*. *Springer Series in Operations Research and Financial Engineering*. Cham: Springer. MR3561100
[26] Sejdinovic, D., Sriperumbudur, B., Gretton, A. and Fukumizu, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann. Statist.* **41** 2263–2291. MR3127866
[27] Székely, G.J. and Rizzo, M.L. (2009). Brownian distance covariance. *Ann. Appl. Stat.* **3** 1236–1265. MR2752127
[28] Székely, G.J. and Rizzo, M.L. (2014). Partial distance correlation with methods for dissimilarities. *Ann. Statist.* **42** 2382–2412. MR3269983
[29] Székely, G.J., Rizzo, M.L. and Bakirov, N.K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35** 2769–2794. MR2382665
[30] Zhou, Z. (2012). Measuring nonlinear dependence in time-series, a distance correlation approach. *J. Time Series Anal.* **33** 438–457. MR2915095

Received July 2016 and revised February 2017