# Testing independence for multivariate time series via the auto-distance correlation matrix

Konstantinos Fokianos     Maria Pitsillou

Department of Mathematics & Statistics, University of Cyprus

e-mail: {fokianos, pitsillou.maria}@ucy.ac.cy

**Abstract**

We introduce the matrix multivariate auto-distance covariance and correlation functions for time series, discuss their interpretation and develop consistent estimators for practical implementation. We also develop a test for testing the independent and identically distributed hypothesis for multivariate time series data and show that it performs better than the multivariate Ljung–Box test. We discuss computational aspects and present a data example to illustrate the methodology.

**Keywords** Characteristic Function; Correlation; Stationarity; $U$ – statistic; Wild Bootstrap.

# 1  Introduction

In applications from fields such as economics (e.g. Tsay, 2014), medicine (McLachlan et al., 2004) or environmetrics (Hipel and McLeod, 1994) we observe several time series evolving simultaneously. Analyzing each component separately might lead to wrong conclusions because of possible interrelationships among the series. Such relationships are usually identified by employing the autocovariance function. For a $d$-dimensional stationary time series $\{X_t, t \in \mathbb{Z}\}$ with mean $\mu$, the autocovariance function is defined by

$$\Gamma(j) = E\left\{(X_{t+j} - \mu)(X_t - \mu)^\intercal\right\} = \left\{\gamma_{rm}(j)\right\}_{r,m=1}^d \quad (j \in \mathbb{Z}).$$

A consistent estimator is the sample autocorrelation function (Brockwell and Davis, 1991, p. 397)

$$\hat{\Gamma}(j) = \begin{cases} n^{-1} \sum_{t=1}^{n-j} (X_{t+j} - \bar{X})(X_t - \bar{X})^\intercal, & 0 \leq j \leq n-1, \\ n^{-1} \sum_{t=-j+1}^{n} (X_{t+j} - \bar{X})(X_t - \bar{X})^\intercal, & -n+1 \leq j < 0, \end{cases}$$

which is often used to measure pairwise dependence. The multivariate Ljung–Box test statistic (Hosking, 1980; Li and McLeod, 1981) is formed in terms as

$$\mathrm{mLB} = n^2 \sum_{j=1}^{p} (n-j)^{-1} \mathrm{tr}\left\{\hat{\Gamma}^\intercal(j)\hat{\Gamma}^{-1}(0)\hat{\Gamma}(j)\hat{\Gamma}^{-1}(0)\right\}, \tag{1}$$

and it is widely applied for testing the hypotheses $\Gamma(1) = \cdots = \Gamma(p) = 0$. However, application of (1) should be done carefully because the number of lags included is held constant but, in practice, the dependence might be of higher order (Hong, 1998, 2000; Xiao and Wu, 2014). Furthermore, the autocovariance function cannot always detect serial dependence for purely non-Gaussian and non-linear models, though it is suitable for Gaussian models. Test statistics which are based on the autocovariance function for testing independence are not consistent against alternatives for models with zero autocovariance (Romano and Thombs, 1996; Shao, 2011), so alternative dependence measures should be studied (Tjøstheim, 1996; Lacal and Tjøstheim, 2017, 2018).

We study the auto-distance covariance function as a suitable statistic for detecting nonlinear relationships in multivariate time series. This function is based on the distance covariance (Székely et al., 2007). Feuerverger (1993)

gave an early treatment and Zhou (2012), Dueck et al. (2014), Fokianos and Pitsillou (2017) and Davis et al. (2016) extended it to time series. Work on the closely related notion of the Hilbert–Schmidt independence criterion includes Sejdinovic et al. (2013) and Gretton et al. (2008). In the present paper, we introduce the auto-distance covariance matrix to identify possible non-linear relationships among the components of a vector series $\{X_t\}$ and show that its sample version is a consistent estimator of the population auto-distance covariance matrix. The sample auto-distance covariance matrix may be used to construct tests for independence of multivariate time series. This is accomplished by following Hong (1999), who introduced the so-called generalized spectral density function. The generalized spectral density matrix captures all the forms of dependence because it is constructed by using the characteristic function. Hence, we can develop statistics for testing independence by considering an increasing number of lags.

The present paper extends Zhou (2012) and Fokianos and Pitsillou (2017), who consider univariate testing of independence based on auto-distance covariance and Székely et al. (2007), since some of our results can be transferred to independent data. Indeed, using the auto-distance covariance matrix for identification of possible dependencies among the components of a random vector could give rise to novel dimension-reduction methods. All methods are available in the R package `dCovTS` (Pitsillou and Fokianos, 2016).

## 2 Auto-Distance Covariance Matrix

### 2.1 Definitions

Suppose that $\{X_t, t \in \mathbb{Z}\}$ is a $d$-variate strictly stationary time series. Denote its cumulative distribution function by $F(x_1, \ldots, x_d)$ and assume that $E(|X_{t;r}|) < \infty$, for $r = 1, \ldots, d$. Let $F_r(\cdot)$ denote the marginal distribution function of $\{X_{t;r}\}$ and $F_{rm}(\cdot, \cdot)$ that of $(X_{t;r}, X_{t;m})$ with $r, m = 1, \ldots, d$. Let $\{X_t : t = 1, \ldots, n\}$ be a sample of size $n$. Zhou (2012), by extending the results of Székely et al. (2007), defines the distance covariance function for multivariate time series, but without taking into account possible cross-dependencies between all possible pairs of the component series of $\{X_t\}$. Here, we define the pairwise distance covariance function as the distance

between the joint characteristic function and the marginal characteristic functions of the pair $(X_{t;r}, X_{t+j;m})$, for $r, m = 1, \ldots, d$. Denote the joint characteristic function of $X_{t;r}$ and $X_{t+j;m}$ by

$$\phi_j^{(r,m)}(u, v) = E\left[\exp\left\{i(uX_{t;r} + vX_{t+j;m})\right\}\right] \quad (u, v \in \mathbb{R}; j \in \mathbb{Z}),$$

where $r, m = 1, \ldots, d$ and $i^2 = -1$. Let $\phi^{(r)}(u) = E\left[\exp\left\{i(uX_{t;r})\right\}\right]$ denote the marginal characteristic function of $X_{t;r}$ for $r = 1, \ldots, d$. Let

$$\Sigma_j(u, v) = \left\{\sigma_j^{(r,m)}(u, v)\right\} \quad (j \in \mathbb{Z}) \tag{2}$$

denote the $d \times d$ matrix whose $(r, m)$ element is

$$\sigma_j^{(r,m)}(u, v) = \mathrm{cov}\left\{\exp(iuX_{t;r}), \exp(ivX_{t+j;m})\right\} = \phi_j^{(r,m)}(u, v) - \phi^{(r)}(u)\phi^{(m)}(v). \tag{3}$$

If $\sigma_j^{(r,m)}(u, v) = 0$ for all $(u, v) \in \mathbb{R}^2$ then the random variables $X_{t;r}$ and $X_{t+j;m}$ are independent for all $j$. Let the $\|\cdot\|_{\mathcal{W}}$-norm of $\sigma_j^{(r,m)}(u, v)$ be defined by

$$\|\sigma_j^{(r,m)}(u, v)\|_{\mathcal{W}}^2 = \int_{\mathbb{R}^2} \left|\sigma_j^{(r,m)}(u, v)\right|^2 \mathcal{W}(du, dv) \quad (j \in \mathbb{Z}),$$

where $\mathcal{W}(\cdot, \cdot)$ is an arbitrary positive weight function such that $\|\sigma_j^{(r,m)}(u, v)\|_{\mathcal{W}}^2 < \infty$. Feuerverger (1993) and Székely et al. (2007) employ a non-integrable weight function,

$$\mathcal{W}(du, dv) = \frac{1}{\pi |u|^2} \frac{1}{\pi |v|^2} du dv. \tag{4}$$

The choice of $\mathcal{W}(\cdot, \cdot)$ is key in this work. Obviously, (A.4) is non-integrable in $\mathbb{R}^2$, but, choices with $\int d\mathcal{W} < \infty$ are possible. Following Hong (1999) and Chen and Hong (2012), suppose that $\mathcal{W}(\cdot, \cdot) : \mathbb{R}^2 \to \mathbb{R}^+$ is nondecreasing with bounded total variation. This obviously holds for $\mathcal{W}(du, dv) = d\Phi(u)d\Phi(v)$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function. In this case, $\|\sigma_j^{(r,m)}(u, v)\|_{\mathcal{W}}^2$ can be computed by Monte Carlo simulation. For related work, see also Meintanis and Iliopoulos (2008) and Hlávka et al. (2011). In what follows we use (A.4) throughout, taking into account the fact that integrable weight functions might miss potential dependence among observations (Székely et al., 2007, p. 2771).

**Definition 1** *The pairwise auto-distance covariance function between $X_{t;r}$ and $X_{t+j;m}$ is denoted by $V_{rm}(j)$ and defined as the positive square root of*

$$V_{rm}^2(j) \quad = \quad \|\sigma_j^{(r,m)}(u,v)\|_{\mathcal{W}}^2 \quad (r,m = 1,\ldots,d;\ j \in \mathbb{Z}), \tag{5}$$

*with $\mathcal{W}(\cdot,\cdot)$ given by (A.4). The auto-distance covariance matrix of $\{X_t\}$ at lag $j$ will be denoted by $V(j)$ and is the $d \times d$ matrix*

$$V(j) = \left\{V_{rm}(j)\right\}_{r,m=1}^d \quad (j \in \mathbb{Z}). \tag{6}$$

Clearly, $V_{rm}^2(j) \geq 0$, for all $j$ and $X_{t;r}$ and $X_{t+j;m}$ are independent if and only if $V_{rm}^2(j) = 0$. Furthermore, we define the $d \times d$ matrices

$$V^{(2)}(j) \quad = \quad \left\{V_{rm}^2(j)\right\}_{r,m=1}^d \quad (j \in \mathbb{Z}). \tag{7}$$

Definition 1 is valid for any weight function $\mathcal{W}(\cdot,\cdot)$ such that $V_{rm}^2(j) < \infty$; with (A.4), it is a pairwise auto-distance covariance function.

**Definition 2** *The pairwise auto-distance correlation function between $X_{t;r}$ and $X_{t+j;m}$ is denoted by $R_{rm}(j)$ and defined as the positive square root of*

$$R_{rm}^2(j) \quad = \quad \frac{V_{rm}^2(j)}{\left\{V_{rr}^2(0)V_{mm}^2(0)\right\}^{1/2}} \quad (r,m = 1,\ldots,d;\ j \in \mathbb{Z}),$$

*provided that $V_{rr}(0)V_{mm}(0) \neq 0$. The auto-distance correlation matrix of $\{X_t\}$ at lag $j$ is*

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

Similarly, define the $d \times d$ matrices

$$R^{(2)}(j) \quad = \quad \left\{R_{rm}^2(j)\right\}_{r,m=1}^d \quad (j \in \mathbb{Z}).$$

Then (7) shows that $R^{(2)}(j) = D^{-1}V^{(2)}(j)D^{-1}$, where $D = \text{diag}\left\{V_{rr}(0)\right\}$, $(r = 1,\ldots,d)$. All above population quantities exist and are well-defined because of standard properties of the characteristic function. Davis et al. (2016) gives existence results concerning more general weight functions.

When $j \neq 0$, $V_{rm}(j)$ measures the dependence of $X_{t;r}$ on $X_{t+j;m}$. For $j > 0$ and if $V_{rm}(j) > 0$, we say that the series $X_{t;m}$ leads the series $X_{t;r}$ at lag $j$. In general, $V_{rm}(j) \neq V_{mr}(j)$ for $r \neq m$, since they measure different types of dependence between the series $\{X_{t;r}\}$ and $\{X_{t;m}\}$ for all $r, m = 1, \ldots, d$. Thus, $V(j)$ and $R(j)$ are non-symmetric matrices, but by stationarity,

$$
\begin{aligned}
V_{rm}^2(-j) &= \|\mathrm{cov}\{\exp(iuX_{t;r}), \exp(ivX_{t-j;m})\}\|_{\mathcal{W}}^2 \\
&= \|\mathrm{cov}\{\exp(iuX_{t;m}), \exp(ivX_{t+j;r})\}\|_{\mathcal{W}}^2 = V_{mr}^2(j), \quad r, m = 1, \ldots, d.
\end{aligned}
$$

Consequently, $V(-j) = V^{\mathsf{T}}(j)$ and $R(-j) = R^{\mathsf{T}}(j)$, because the matrices $V(j)$ and $R(j)$ have as elements the positive square roots of the elements of $V^{(2)}(j)$ and $R^{(2)}(j)$.

Auto-distance covariance matrices are interpreted as follows. 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 explain dependence among the pairs $(X_{t;r}, X_{t+j;r})$, $r = 1, \ldots, d$. The off-diagonal elements $\left\{V_{rm}(0)\right\}_{r,m=1}^{d}$ measure 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 $j \neq 0$, $\left\{V_{rm}(j)\right\}_{r,m=1}^{d}$ measures dependence between $\{X_{t;r}\}$ and $\{X_{t+j;m}\}$. If $V_{rm}(j) = 0$ for all $j \neq 0$, 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 $j \neq 0$, 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 $\{X_{t;m}\}$.

## 2.2 Estimation

To estimate (5) and (6), define, for $j \geq 0$,

$$
\hat{\sigma}_j^{(r,m)}(u, v) = \hat{\phi}_j^{(r,m)}(u, v) - \hat{\phi}^{(r)}(u)\hat{\phi}^{(m)}(v),
$$

with

$$
\hat{\phi}_j^{(r,m)}(u, v) = (n - j)^{-1} \sum_{t=1}^{n-j} \exp\left\{i(uX_{t;r} + vX_{t+j;m})\right\} \quad (u, v \in \mathbb{R}),
$$

and $\hat{\phi}^{(r)}(u) = \hat{\phi}_j^{(r,m)}(u,0)$, $\hat{\phi}^{(m)}(v) = \hat{\phi}_j^{(r,m)}(0,v)$. Then, the sample pairwise auto-distance covariance is defined by the positive square root of

$$\hat{V}_{rm}^2(j) \;\; = \;\; \pi^{-2} \int_{\mathbb{R}^2} \frac{\left|\hat{\sigma}_j^{(r,m)}(u,v)\right|^2}{|u|^2\,|v|^2} du dv.$$

Let $Y_{t;m} = X_{t+j;m}$. Then, based on the sample $\{(X_{t;r}, Y_{t;m}) : t = 1, \ldots, n-j\}$, we calculate the $(n-j) \times (n-j)$ Euclidean distance matrices $A^r = (A_{ts}^r)$ and $B^m = (B_{ts}^m)$ with elements

$$A_{ts}^r = a_{ts}^r - \bar{a}_{t\cdot}^r - \bar{a}_{\cdot s}^r + \bar{a}_{\cdot\cdot}^r,$$

where $\alpha_{ts}^r = |X_{t;r} - X_{s;r}|$, $\bar{\alpha}_{t\cdot}^r = \left(\sum_{s=1}^{n-j} a_{ts}^r\right)/(n-j)$, $\bar{\alpha}_{\cdot s}^r = \left(\sum_{t=1}^{n-j} a_{ts}^r\right)/(n-j)$, $\bar{\alpha}_{\cdot\cdot}^r = \left(\sum_{t=1}^{n-j}\sum_{s=1}^{n-j} a_{ts}^r\right)/(n-j)^2$. Similarly, define the quantities $b_{ts}^m = |Y_{t;m} - Y_{s;m}|$ to obtain $\bar{b}_{t\cdot}^m$, $\bar{b}_{\cdot s}^m$, $\bar{b}_{\cdot\cdot}^m$ and $B_{ts}^m$. The, by following Székely et al. (2007), we obtain that

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

If $j < 0$ we set $\hat{V}_{rm}^2(j) = \hat{V}_{mr}^2(-j)$. By (7) define the $d \times d$ matrices $\hat{V}^{(2)}(j) = \left\{\hat{V}_{rm}^2(j)\right\}$, $j \in \mathbb{Z}$. The sample distance covariance matrix is given by $\hat{V}(j) = \left\{\hat{V}_{rm}(j)\right\}$, $j \in \mathbb{Z}$. Similarly, define $\hat{R}^{(2)}(j)$ and $\hat{R}(j)$, $j \in \mathbb{Z}$.

## 2.3 Large sample properties of the sample distance covariance matrix

The assumption of stationarity is quite restrictive for applications and it is interesting to investigate the behavior of the distance covariance function when this assumption does not hold. Consider a simple random walk where $\{X_t\}$ is assumed to be a univariate Gaussian process with $E(X_t) = 0$, $\text{var}(X_t) = 1$ and $\text{cov}(X_t, X_{t+j}) = \rho(j)$. Then (Fokianos and Pitsillou, 2017)

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

The last display shows that the behavior of $\rho^2(j)$ determines that of $R^2(j)$, so we expect $R^2(j)$ to decay as slowly as $\rho^2(j)$ does for a random walk. We can test whether the increments $X_t - X_{t-1}$ form an independent and identically distributed sequence by employing the methods of Fokianos and Pitsillou (2017).

In the Supplementary Material we show that $\hat{V}_{rm}^2(j)$ is a $V$-statistic (Serfling, 1980, Section 5.1.2), which can be approximated by another $V$-statistic with a kernel of lower order. Indeed, suppose that $u, u' \in \mathbb{R}$ and let $X$ be a real-valued random variable. Define

$$m_X(u) = E(|X - u|), \quad \bar{m}_X = E\{m_X(X)\}, \quad d_X(u, u') = |u - u'| - m_X(u) - m_X(u') + \bar{m}_X.$$

With some abuse of notation and setting $X \equiv X_{t;r}$ and $Y \equiv X_{t+j;m}$ we obtain that

$$V_{rm}^2(j) = E\Big\{d_X(X, X')d_Y(Y, Y')\Big\},$$

where $(X', Y')$ is an independent and identically distributed copy of $(X, Y)$ (Székely and Rizzo, 2013, p. 1262). That is, there exists a kernel $h : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ given by

$$h(x, y; x', y') = d_X(x, x')d_Y(y, y'), \tag{8}$$

such that $V_{rm}^2(j) = \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} h(x, y; x', y')dF_{rm}(x, y)dF_{rm}(x', y')$. The kernel function is symmetric, continuous and positive semidefinite. Under independence, $\hat{V}_{rm}^2(\cdot)$ is a degenerate $V$-statistic. If $E(|X_{t;r}|^2) < \infty$, then Székely and Rizzo (2013, Lemma 1) and Fubini's theorem yield that

$$V_{rm}^2(j) \quad = \quad E(|X - X'|\,|Y - Y'|) + E(|X - X'|)E(|Y - Y''|) - 2E(|X - X'|\,|Y - Y''|),$$

where $(X', Y')$ and $(X'', Y'')$ are independent and identically distributed copies of $(X, Y)$. If $X_{t;r}$ is independent of $X_{t+j;m}$, then $E\{h(x, y; X, Y)\} = 0$, so $\hat{V}_{rm}^2(j)$ has a first order degeneracy. The following proposition shows the strong consistency of the estimator $\hat{V}(j)$.

**Proposition 1** *Let $\{X_t\}$ be a $d$-variate strictly stationary and ergodic process with $E(|X_{t;r}|^2) < \infty$ for $r = 1, \ldots, d$. Then, for all $j \in \mathbb{Z}$, $\hat{V}(j) \to V(j)$ almost surely as $n \to \infty$.*

Proposition 1 follows directly from the strong law of large numbers for $V$-statistics of ergodic and stationary sequences (Aaronson et al., 1996). Alternatively, it can be proved by the methods of Székely et al. (2007), Fokianos and Pitsillou (2017, Prop. 1) and Davis et al. (2016) by assuming $E(|X_{t;r}|) < \infty$ or by Zhou (2012) assuming $E(|X_{t;r}|^{1+\epsilon}) < \infty$ for some $\epsilon > 0$. In particular, by assuming strict stationarity, ergodicity and $E(|X_{t;r}|^2) < \infty$, for $r = 1, \ldots, d$, the strong consistency of the sample auto-distance covariance matrix is established by

individually considering each element of $\hat{V}^{(2)}(j)$ and then the corresponding element of $\hat{V}(j)$. Previous related work for strong laws of $V$-statistics requires stationarity, ergodicity, existence of second moments, almost surely $F_{rm}(\cdot, \cdot)$ continuity of the kernel function and uniform integrability. Under these assumptions, $\hat{V}(j)$ is a weakly consistent estimator of $V(j)$, see Borovcova et al. (1999, Theorem 1) and Aaronson et al. (1996, Proposition 2.8).

The following theorem is proved in the Supplementary Material and gives the limiting distribution of the sample pairwise auto-distance covariance function, $\hat{V}^2_{rm}(\cdot)$ when $V^2_{rm}(\cdot) \neq 0$ and when $\{X_t\}$ is a pairwise independent sequence. Related results are given by Zhou (2012) and Davis et al. (2016). We attack the problem by using results on $U$-statistics for $\beta$-mixing processes (Yoshihara, 1976). The first result shows that we can form asymptotic confidence intervals for $V^2_{rm}(\cdot)$; unfortunately this gives no further information on the dependence structure in the data.

**Theorem 1** *Suppose that $\{X_t\}$ is $\beta$-mixing and assume that there exists a positive number $\delta$ such that for $\nu = 2+\delta$: (i) $E(|X_{t;r}|^\nu) < \infty$, for $r = 1, \ldots, d$, (ii) for any integers $i_1$, $i_2$, $\sup_{i_1, i_2} E(|X_{i_1;r} - X_{i_2;r}|^\nu) < \infty$ for all $r = 1, \ldots, d$ and (iii) the mixing coefficients $\beta(k)$ satisfy $\beta(k) = O(k^{-(2+\delta')/\delta'})$, for some $\delta'$ such that $0 < \delta' < \delta$. Then, for fixed $j$:*

*1. if $V^2_{rm}(j) \neq 0$ then*

$$n^{1/2}\left\{\hat{V}^2_{rm}(j) - V^2_{rm}(j)\right\} \quad \to \quad N(0, 36\sigma^2), \quad n \to \infty,$$

*in distribution, where $\sigma^2$ is given in the Supplementary Material,*

*2. if $V^2_{rm}(j) = 0$ and $\delta' < \delta/(3 + \delta)$, then*

$$n\hat{V}^2_{rm}(j) \quad \to \quad Z = \sum_l \lambda_l Z_l^2, \quad n \to \infty, \tag{9}$$

*in distribution, where $(Z_l)$ is an independent and identically distributed sequence of standard normal random variables and $(\lambda_l)$ is a sequence of nonzero eigenvalues which satisfy the Hilbert–Schmidt equation $E\left\{h(x, y; X, Y)\Phi(X, Y)\right\} = \lambda\Phi(x, y)$. The kernel $h(\cdot)$ is defined by (8) and it is represented by $h(x, y; x', y') = \sum_{l=1}^{\infty} \lambda_l \Phi_l(x, y)\Phi_l(x', y')$, where $(\Phi_l)_l$ is the sequence of the corresponding orthonormal*

*eigenfunctions.*

The second part of Theorem 1 is proved by a Hoeffding decomposition (Hoeffding, 1948) by showing that $\hat{V}^2_{rm}(.)$ is approximated by a $V$-statistic of order two which is degenerate under independence. Then, applying Leucht and Neumann (2013a, Thm. 1) yields the result. If $\{X_t\}$ is an independent and identically distributed sequence then the second part of Theorem 1 holds; see Székely et al. (2007). In general, it is of interest to approximate the asymptotic distribution of the matrix variate $V$-statistic $\{V^{(2)}(j), j \in \mathbb{Z}\}$, whether or not independence holds. To the best of our knowledge, this problem has not been addressed in the literature, but see Chen (2016). Furthermore, it is of interest to determine simultaneous confidence intervals for the elements of $V(j)$ mimicking the methodology of the ordinary autocorrelation function, under the assumption of independence. But the asymptotic distribution given in (9) cannot be employed in applications, so a simulation-based method should be applied. We discuss this further in Sec. 4.2.

# 3 Testing the independent and identically distributed hypothesis for multivariate time series

In this section, we develop a test statistic for testing the null hypothesis $H_0$ : $\{X_t\}$ is an independent and identically distributed sequence. Recall that the Frobenius norm of an $m \times n$ matrix $A$ is defined as $\|A\|^2_F = \sum_{i=1}^m \sum_{j=1}^n |\alpha_{ij}|^2 = \text{tr}(A^*A)$, where $A^*$ denotes the conjugate transpose of $A$, and $\text{tr}(A)$ denotes the trace of the matrix $A$.

We obtain by (3) that

$$\sup_{(u,v)\in\mathbb{R}^2} \sum_{j=-\infty}^{\infty} \left| \sigma_j^{(r,m)}(u,v) \right| < \infty, \tag{10}$$

provided that $\{X_t\}$ is an $\beta$-mixing process with coefficients decaying to zero. Indeed, $\left|\text{cov}(e^{iuX_{t;r}}, e^{ivX_{t+j,m}})\right| = \left|\sigma_j^{(r,m)}(u,v)\right| \leq C\beta(j)$, for some constant $C$. Thus, the sequence of covariance matrices $\{\Sigma_j(u,v), j \in \mathbb{Z}\}$, defined by (2), has absolutely summable components for all $(u,v) \in \mathbb{R}^2$. Define the Fourier transform of

$\sigma_j^{(r,m)}(\cdot, \cdot),$

$$f^{(r,m)}(\omega, u, v) = (2\pi)^{-1} \sum_{j=-\infty}^{\infty} \sigma_j^{(r,m)}(u, v)e^{-ij\omega} \quad (\omega \in [-\pi, \pi]). \tag{11}$$

Because of (A.2), $f^{(r,m)}(\cdot, \cdot, \cdot)$ is bounded and uniformly continuous. If $r = m$, then $f^{(r,r)}(\omega, u, v)$ is called the generalized spectrum or generalized spectral density of $X_{t;r}$ at frequency $\omega$ for all $(u, v) \in \mathbb{R}^2$. If $r \neq m$, then $f^{(r,m)}(\omega, u, v)$ is called the generalized cross-spectrum or generalized cross spectral density of $X_{t;r}$ and $X_{t;m}$ at frequency $\omega$ for all $(u, v) \in \mathbb{R}^2$. Collecting all elements of (11) in a $d \times d$ matrix, we obtain the generalized spectral density matrix

$$F(\omega, u, v) = (2\pi)^{-1} \sum_{j=-\infty}^{\infty} \Sigma_j(u, v)e^{-ij\omega} = \left\{ f^{(r,m)}(\omega, u, v) \right\}_{r,m=1}^{d}.$$

Under the null hypothesis of independent and identically distributed data, $\Sigma_j(u, v) = 0$ for all $j \neq 0$. In this case denote $F(\cdot, \cdot, \cdot)$ by

$$F_0(\omega, u, v) = (2\pi)^{-1} \left\{ \sigma_0^{(r,m)}(u, v) \right\}_{r,m=1}^{d}.$$

In general, $F_0(\cdot, \cdot, \cdot)$ is not a diagonal matrix, but when $X_{t;r}$ and $X_{t;m}$ are independent for all $r, m = 1, \ldots, d$, then $F_0(\cdot, \cdot, \cdot)$ reduces to a diagonal matrix. Consider the following class of kernel-density estimators,

$$\hat{f}^{(r,m)}(\omega, u, v) = (2\pi)^{-1} \sum_{j=-(n-1)}^{(n-1)} (1 - |j|/n)^{1/2} K(j/p) \hat{\sigma}_j^{(r,m)}(u, v)e^{-ij\omega} \quad (\omega \in [-\pi, \pi]),$$

where $p$ is a bandwidth parameter and $K(\cdot)$ is a univariate kernel function satisfying

**Assumption 3.1** $K : \mathbb{R} \to [-1, 1]$ is symmetric and continuous at 0 and at 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 $b > 1/2$.

Next let

$$\hat{F}(\omega, u, v) = \left\{ \hat{f}^{(r,m)}(\omega, u, v) \right\}_{r,m=1}^{d}, \quad \hat{F}_0(\omega, u, v) = (2\pi)^{-1} \left\{ \hat{\sigma}_0^{(r,m)}(u, v) \right\}_{r,m=1}^{d}.$$

Then, it is shown in the Supplementary Material that for $\mathcal{W}(\cdot, \cdot)$ given by (A.4),

$$L_2^2 \left\{ \hat{F}(\omega, u, v), \hat{F}_0(\omega, u, v) \right\} = \int_{\mathbb{R}^2} \int_{-\pi}^{\pi} \|\hat{F}(\omega, u, v) - \hat{F}_0(\omega, u, v)\|_F^2 d\omega \mathcal{W}(du, dv)$$

$$= 2\pi^{-1} \sum_{j=1}^{n-1} (1 - j/n) K^2(j/p) \text{tr} \left\{ \hat{V}^*(j)\hat{V}(j) \right\}. \tag{12}$$

In terms of the distance correlation matrix, (12) becomes

$$L_2^2\Big\{\hat{G}(\omega,u,v),\hat{G}_0(\omega,u,v)\Big\} = 2\pi^{-1}\sum_{j=1}^{n-1}(1-j/n)K^2(j/p)\text{tr}\Big\{\hat{V}^*(j)\hat{D}^{-1}\hat{V}(j)\hat{D}^{-1}\Big\}, \qquad (13)$$

where $\hat{G}(\cdot,\cdot,\cdot)$, $\hat{G}_0(\cdot,\cdot,\cdot)$ are the estimators of the normalized multivariate generalized spectrums $G(\cdot,\cdot,\cdot)$ and $G_0(\cdot,\cdot,\cdot)$; details are given in the Supplementary Material. Equations (12) and (13) motivate our study of multivariate tests of independence. In particular, it is of interest to test whether the vector series $\{X_t\}$ is independent and identically distributed regardless of any possible dependence between time series components $\{X_{t;r}\}$ for $r = 1,\ldots,d$. Equation (13) can be viewed as a multivariate Ljung–Box type statistic based on the distance covariance matrix instead of ordinary autocovariance matrix. Indeed, by choosing $K(z) = 1$ for $|z| \leq 1$ and 0 otherwise, equation (13) becomes

$$L_2^2\Big\{\hat{G}(\omega,u,v),\hat{G}_0(\omega,u,v)\Big\} = 2\pi^{-1}\sum_{j=1}^{p}(1-j/n)\text{tr}\Big\{\hat{V}^*(j)\hat{D}^{-1}\hat{V}(j)\hat{D}^{-1}\Big\}.$$

Define $T_n^{(r,m)} = \sum_{j=1}^{n-1}(n-j)K^2(j/p)\hat{V}_{rm}^2(j)$. Then, the test statistic motivated by (12) is

$$T_n = \sum_{r,m}T_n^{(r,m)} = \sum_{j=1}^{n-1}(n-j)K^2(j/p)\text{tr}\Big\{\hat{V}^*(j)\hat{V}(j)\Big\}.$$

Similarly, by using (13), consider

$$\bar{T}_n = \sum_{r,m}\frac{T_n^{(r,m)}}{\Big\{\hat{V}_{rr}^2(0)\hat{V}_{mm}^2(0)\Big\}^{1/2}} = \sum_{j=1}^{n-1}(n-j)K^2(j/p)\text{tr}\Big\{\hat{V}^*(j)\hat{D}^{-1}\hat{V}(j)\hat{D}^{-1}\Big\}.$$

**Theorem 2** *Suppose that $E(|X_{t,r}|^2) < \infty$, $(r = 1,\ldots,d)$ and that Assumption A.1 holds. Let $p = cn^\lambda$, where $c > 0$ and $\lambda \in (0,1)$. If $\{X_t\}$ is an independent and identically distributed sequence, then*

$$M_n^{(r,m)} = \frac{T_n^{(r,m)} - \hat{C}_0^{(r,m)}p\int_0^\infty K^2(z)dz}{\Big\{\hat{D}_0^{(r,m)}p\int_0^\infty K^4(z)dz\Big\}^{1/2}} \rightarrow N(0,1), \quad n \to \infty$$

*in distribution, where*

$$C_0^{(r,m)} = \int_{\mathbb{R}^2}\sigma_0^{(r,r)}(u,-u)\sigma_0^{(m,m)}(v,-v)\mathcal{W}(du,dv),$$

$$D_0^{(r,m)} = 2\int_{\mathbb{R}^2}\Big|\sigma_0^{(r,r)}(u,u')\sigma_0^{(m,m)}(v,v')\Big|^2\mathcal{W}(du,dv)\mathcal{W}(du',dv') = 2V_{rr}^2(0)V_{mm}^2(0),$$

*and $\hat{C}_0^{(r,m)}$, $\hat{D}_0^{(r,m)}$ are their sample counterparts.*

Theorem 2 implies the following result.

**Corollary 1** *Suppose that $E(|X_{t,r}|^2) < \infty$, $(r = 1, \ldots, d)$ and that Assumption A.1 holds. Let $p = cn^\lambda$, where $c > 0$ and $\lambda \in (0,1)$. If $\{X_t\}$ is an independent and identically distributed sequence, then*

$$M_n \equiv \frac{T_n - \left(\sum_{r,m} \hat{C}_0^{(r,m)}\right) p \int_0^\infty K^2(z)dz}{\left\{\left(\sum_{r,m} \hat{D}_0^{(r,m)}\right) p \int_0^\infty K^4(z)dz\right\}^{1/2}} \to N(0,1), \quad \bar{M}_n \equiv \frac{\bar{T}_n - \left(\sum_{r,m} c_0^{(r,m)}\right) p \int_0^\infty K^2(z)dz}{d\left\{2p \int_0^\infty K^4(z)dz\right\}^{1/2}} \to N(0,1)$$

*in distribution, as $n \to \infty$, where $c_0^{(r,m)} = C_0^{(r,m)}/\{V_{rr}(0)V_{mm}(0)\}$, and $\hat{c}_0^{(r,m)}$ is the corresponding empirical analogue.*

The following result states the consistency of the test statistics.

**Theorem 3** *Let $\{X_t\}$ be a $\beta$-mixing strictly stationary, but not independent and identically distributed, process with mixing coefficients satisfying $\sum_k \beta(k) < \infty$. Suppose that $E(|X_{t,r}|^2) < \infty$, $(r = 1, \ldots, d)$ and that Assumption A.1 holds. Let $p = cn^\lambda$ for $c > 0$ and $\lambda \in (0,1)$. Then, as $n \to \infty$,*

$$\frac{p^{1/2}}{n} M_n \to \frac{\frac{\pi}{2} L_2^2\Big(F(\omega,u,v), F_0(\omega,u,v)\Big)}{\left\{\sum_{r,m} D_0^{(r,m)} \int_0^\infty K^4(z)dz\right\}^{1/2}}, \quad \frac{p^{1/2}}{n} \bar{M}_n \to \frac{\frac{\pi}{2} L_2^2\Big(G(\omega,u,v), G_0(\omega,u,v)\Big)}{d\left\{2 \int_0^\infty K^4(z)dz\right\}^{1/2}}$$

*in probability.*

When we deal with a non-stationary process, $\bar{T}_n$ converges to $\infty$ in probability, so the test will have asymptotic power one. Rejection of the null does not allow us to conclude that the process is stationary under the alternative hypothesis. Following the earlier discussion, we can test the independent and identically distributed hypothesis for the increments $X_t - X_{t-1}$. In general, the results depend on the bandwidth parameter $p$ and the sample size $n$. Choosing the bandwidth parameter is not considered further, but our limited experience is that choosing roughly $p \geq 15$ for a sample size of $n = 500$ yields a good asymptotic approximation. However, it is preferable to vary the value of $p$ and to examine the sensitivity of the results. We suggest the use of simulation-based techniques to approximate the distributions of $T_n$ or $\bar{T}_n$ for small $n$.

# 4 Computation of test statistic with applications

## 4.1 Bootstrap methodology

To approximate the asymptotic distribution of $T_n$ or $\bar{T}_n$ we resort to the work of Dehling and Mikosch (1994), who proposed a wild bootstrap for approximating the distribution of degenerate $U$-statistics for independent and identically distributed data. In recent contributions, Leucht and Neumann (2013a,b) and Chwialkowski et al. (2014) proposed a novel dependent wild bootstrap (Shao, 2010) for approximating the distribution of degenerate $U$-and $V$-statistics calculated from time series data. The method relies on generating auxiliary random variables $(W_{tn}^*)_{t=1}^{n-j}$ and on computing bootstrap realizations of $\hat{V}_{rm}^2(j)$ as

$$\hat{V}_{rm}^{2*}(j) \quad = \quad (n-j)^{-2} \sum_{t,s=1}^{n-j} W_{tn}^* h(X_{t;r}, Y_{t;m}; X_{s;r}, Y_{s;m}) W_{sn}^*,$$

where $h(\cdot)$ is defined by (8), for $r, m = 1, \ldots, d$ and $j = 1, \ldots, n-1$. A bootstrap realization of $T_n$ is computed as

$$T_n^* \quad = \quad \sum_{j=1}^{n-1} (n-j) K^2(j/p) \sum_{r,m} \hat{V}_{rm}^{2*}(j).$$

To test whether $\{X_t\}$ is an independent and identically distributed sequence, we repeat the above steps $B$ times to obtain $T_{n,1}^*, \ldots, T_{n,B}^*$ and then approximate the $p$-value of $T_n$ by $\left\{ \sum_{b=1}^{B} \mathbb{I}(T_{n,b}^* \geq T_n) \right\}/(B+1)$, where $\mathbb{I}(\cdot)$ denotes the indicator function. We work analogously for $\bar{T}_n$.

We generate $W_{tn}^*$ as independent and identically distributed standard normal variables because we operate under the null hypothesis. Theorem 2 and Corollary 1 show that we can employ the test statistic with a normal approximation but experience shows that a rather large sample size is required to achieve the nominal significance level. Furthermore a standard non-parametric bootstrap provides an alternative test of the null hypothesis. This is implemented in the R package `dCovTS` (Pitsillou and Fokianos, 2016). The wild bootstrap saves computational time, as simulation is done in a separate loop. The results of Leucht and Neumann (2013b) imply the wild bootstrap validity of $\hat{V}_{rm}^{2*}(j)$ as a proxy to $\hat{V}_{rm}^2(j)$ for a fixed lag $j$ and for $\tau$-dependent processes. In this contribution we restrict ourselves to $\beta$-mixing processes. Interesting time series models, such as autoregressive moving average

models with continuous innovations or generalized autoregressive conditional heteroscedastic models, belong to both classes of processes. The approximation of $\hat{V}_{rm}^{2*}(j)$ as a proxy to $\hat{V}_{rm}^2(j)$ for a fixed lag $j$ is guaranteed to hold under suitable conditions on the mixing coefficients and for processes that lie in both classes. To gain insight about the behavior of $T_n$ and its wild bootstrap counterpart we will need to study the joint distribution of $\{\hat{V}_{rm}^{2*}(j)\}$ as a proxy to the joint distribution of $\{\hat{V}_{rm}^2(j)\}$. Although we do not address this problem theoretically, empirical evidence supports that distribution of $T_n^*$ approximates that of $T_n$ adequately.

## 4.2 Obtaining simultaneous critical values for the auto-distance correlation plots

It is customary to check the white noise assumption by plotting the sample autocorrelation function with simultaneous confidence intervals. The critical values employed for obtaining confidence intervals are computed by using the asymptotic normality of the first $q$ sample autocorrelations for white noise (Brockwell and Davis, 1991, Thm. 7.2.1). Here we use a similar plot to check independence using the auto–distance correlation function. This task is complicated because the vector comprising $R_{rm}(j)$, $(j = 1, \ldots, q)$, is a function of degenerate $V$-statistics under the null hypothesis. To overcome this difficulty we resort to Monte Carlo simulation.

Critical values chosen by the wild bootstrap asymptotically maintain the nominal size of a test statistic. Given $B$ bootstrap realizations of $\hat{R}_{rm}(j)$, say $\{\hat{R}_{rm,b}^*(j), b = 1, \ldots, B\}$, we compute the $p$-value

$$p_{rm}(j) \quad = \quad (B+1)^{-1} \sum_{b=1}^{B} \mathbb{I}\Big\{ \hat{R}_{rm,b}^*(j) \geq \hat{R}_{rm}(j) \Big\}.$$

But the $p$-values $\{p_{rm}(j), j = 1, \ldots, q\}$ correspond to testing the hypotheses $R_{rm}(j) = 0$, $(j = 1, \ldots, q)$. Because this is a multiple testing situation, we adjust them at some prespecified level $\alpha$, to obtain a new set of $p$-values, say $\{\tilde{p}_{rm}(j), j = 1, \ldots, q\}$, by using the false discovery rate of Benjamini and Hochberg (1995). Using the adjusted $p$-values, we get critical points $\{c_{rm}(j), j = 1, \ldots, q\}$ for which

$$\tilde{p}_{rm}(j) \quad = \quad \frac{\#\Big\{ \hat{R}_{rm}^*(j) \geq c_{rm}(j) \Big\}}{B} \quad (j = 1, \ldots, q).$$

The horizontal line in the plots corresponds to $c = \max_{r,m,j} c_{rm}(j)$, a conservative approach that guarantees that

all simultaneous confidence intervals are at a given level $\alpha$. In the Supplementary Material we show that these critical values depend only on the length and not on the dimension of the series.

## 4.3 Results

We now describe a limited simulation study concerning the test statistic $\bar{T}_n$ computed using a Lipschitz continuous univariate kernel function $K(\cdot)$. That is for any $z_1, z_2 \in \mathbb{R}$, $|K(z_1) - K(z_2)| \leq C |z_1 - z_2|$, for some constant $C$. We use the Daniell, the Parzen and Bartlett kernels. The Lipschitz condition rules out the truncated kernel but results based on it are also given. We also compare the performance of $\bar{T}_n$ to that of the multivariate Ljung–Box statistic (1). We first investigate the size of the test by considering bivariate standard normal data. Table 1 shows the achieved levels of all test statistics calculated at $5\%$ and $10\%$ nominal levels, and indicates that the proposed test statistics approximate the nominal levels adequately. In order to compare the power of both test statistics, we

Table 1: Empirical type I error ($\%$) of statistics for testing the hypothesis that the data are independent and identically distributed. The data are generated from the bivariate standard normal distribution. The bandwidth $p$ is $[3n^\lambda]$, $\lambda = 0.1, 0.2$ and $0.3$. Results are based on $B = 499$ bootstrap replications for each of 1000 simulations.

| $n$ | | 500 | | | | | | 1000 | | | | | |
|-----|---|-----|---|----|---|----|---|-----|---|----|---|----|---|
| $p$ | | 6 | | 11 | | 20 | | 6 | | 12 | | 24 | |
| | | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% |
| $\bar{T}_n$ | BAR | 11.5 | 5.5 | 10.2 | 5.2 | 10.8 | 4.7 | 9.7 | 4.8 | 8.3 | 3.7 | 12 | 6 |
| | TRUNC | 8.7 | 3.8 | 8.3 | 3.4 | 8.6 | 4.1 | 8 | 3 | 8 | 5 | 5 | 3 |
| | PAR | 9.7 | 5.1 | 9.3 | 4.3 | 8.4 | 4.0 | 8 | 3 | 11 | 6 | 12 | 6 |
| | DAN | 8.9 | 3.9 | 10.7 | 5.7 | 10.2 | 4.6 | 10.2 | 5.0 | 11.7 | 5.5 | 9.2 | 4.2 |
| mLB | | 11.4 | 4.8 | 10.1 | 4.9 | 11.1 | 5.6 | 10.5 | 5.5 | 9.1 | 4.9 | 9.6 | 5.4 |

BAR, Bartlett kernel; TRUNC, truncated kernel; PAR, Parzen kernel; DAN, Daniell kernel; mLB, multivariate Ljung–Box statistic.

consider a bivariate nonlinear moving average model of order 2,

$$X_{t;i} = \epsilon_{t;i}\epsilon_{t-1;i}\epsilon_{t-2;i} \quad (i = 1, 2), \tag{14}$$

where $\{\epsilon_{t;i}, i = 1, 2\}$ is an independent and identically distributed sequence of standard normal random variables,

a bivariate generalized autoregressive conditional heteroscedastic model of order $(1, 1)$

$$X_{t;i} = h_{t;i}^{1/2}\epsilon_{t;i} \quad (i = 1, 2), \tag{15}$$

where

$$\begin{bmatrix} h_{t;1} \\ h_{t;2} \end{bmatrix} = \begin{bmatrix} 0.003 \\ 0.005 \end{bmatrix} + \begin{bmatrix} 0.2 & 0.1 \\ 0.1 & 0.3 \end{bmatrix} \begin{bmatrix} X_{t-1;1}^2 \\ X_{t-1;2}^2 \end{bmatrix} - \begin{bmatrix} 0.4 & 0.05 \\ 0.05 & 0.5 \end{bmatrix} \begin{bmatrix} h_{t-1;1} \\ h_{t-1;2} \end{bmatrix}$$

and $\{\epsilon_{t;i}, i = 1, 2\}$ is a sample from a bivariate normal with correlation $\rho = 0.4$, and a bivariate autoregressive

model of order 1

$$\begin{bmatrix} X_{t;1} \\ X_{t;2} \end{bmatrix} = \begin{bmatrix} 0.04 & -0.10 \\ 0.11 & 0.50 \end{bmatrix} \begin{bmatrix} X_{t-1;1} \\ X_{t-1;2} \end{bmatrix} + \begin{bmatrix} \epsilon_{t;1} \\ \epsilon_{t;2} \end{bmatrix}, \tag{16}$$

and the error as in the previous model but with $\rho = 0$ and $0.4$. Table 2 shows that $\bar{T}_n$ attains larger power than (1)

when data are generated by the non-linear models (14) and (15). The Ljung–Box test statistic performs better than

$\bar{T}_n$ when data are generated by (16) with $\rho = 0$ and for small values of $p$ and large sample size. When $p$ is large,

$\bar{T}_n$ performs generally better than (1).

## 4.4 Application

We analyze monthly unemployment rates of the 50 US states from January 1976 to September 2011, seasonally ad-

justed and available from Tsay's (2014) book site `http://faculty.chicagobooth.edu/ruey.tsay/`

`teaching/mtsbk/`. We consider first the 416 differenced monthly rates of Alaska, Arkansas, Colorado and

Delaware. Figure 1 displays the sample auto-distance correlation matrices at each lag, and reveals the dependence

structure for lags 1 to 12. After the sixth lag, the auto-distance correlation function suggests that there is no depen-

dence among the four series. Assuming that the four-dimensional series follows a vector autoregressive model and

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 statistic $\bar{T}_n$ is calculated using the Bartlett kernel.

| $n$ | | 500 | | | 800 | | | 1000 | |
|---|---|---|---|---|---|---|---|---|---|
| $p$ | 6 | 11 | 20 | 6 | 12 | 23 | 6 | 12 | 24 |
| | | | | Model (14) | | | | | |
| $\bar{T}_n$ | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| mLB | 40 | 32.5 | 26.6 | 47.1 | 35.9 | 28.5 | 46.6 | 36.6 | 26.8 |
| | | | | Model (15) | | | | | |
| $\bar{T}_n$ | 76.4 | 79.5 | 78.3 | 97.2 | 98 | 95.4 | 99.4 | 98 | 99 |
| mLB | 56.9 | 54.6 | 48.1 | 65.1 | 60.4 | 53.9 | 64.9 | 65.2 | 54.9 |
| | | | | Model (16) with $\rho = 0$ | | | | | |
| $\bar{T}_n$ | 66.6 | 58.4 | 44.8 | 41.3 | 41.1 | 69.1 | 30.4 | 27 | 80.4 |
| mLB | 48.5 | 36.1 | 25.6 | 75.7 | 59.4 | 44.2 | 86.2 | 71.8 | 55.8 |
| | | | | Model (16) with $\rho = 0.4$ | | | | | |
| $\bar{T}_n$ | 71.4 | 62.5 | 46.9 | 92.7 | 89.7 | 75.7 | 96.8 | 95.6 | 88 |
| mLB | 50.8 | 37.1 | 28.3 | 77.5 | 58.9 | 44.6 | 86.9 | 72.6 | 54.1 |

employing the Akaike information criterion, a fifth-order model fits the data well. The auto-distance correlation plots of the resulting residual series in Figure 2 shows no serial dependence. Tests of independence for the residual series using the Bartlett kernel for the computation of $\bar{T}_n$ and mLB yield $p$-values 0.428, 0.228, 0.158 and 1, 0.999, 0.999 when $p =$ 6, 11 and 19, confirming the adequacy of the model fit.
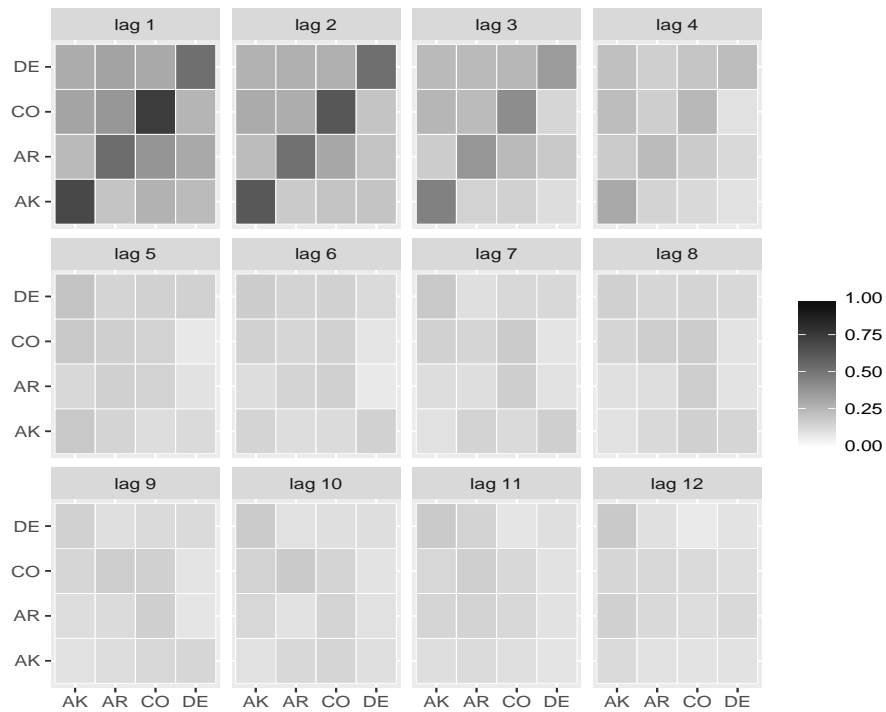
Figure 1: Visualizing the sample auto-distance correlation matrices of the four-dimensional unemployment series of Alaska, Arkansas, Colorado and Delaware, starting on top left at lag $j = 1$. The darker rectangles correspond to higher distance correlations between series at a specific lag.
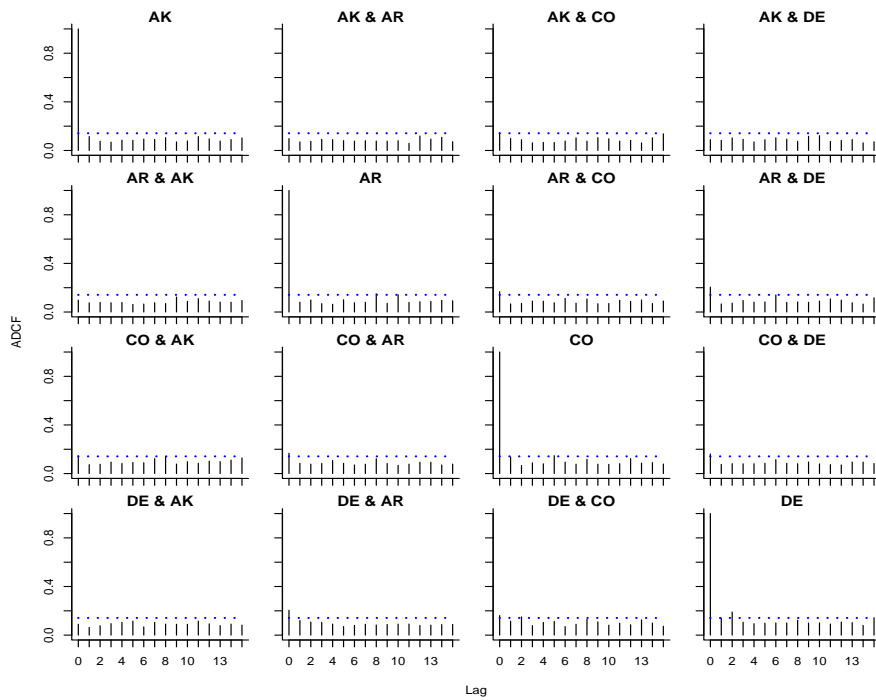
Figure 2: Auto-distance correlation plot of the residuals vector process after fitting a fifth-order vector autoregressive model to the unemployment data. The dotted horizontal line is drawn following the methodology described in Section 4.2.

## ACKNOWLEDGMENT

## A Appendix

### A.1 $\hat{V}_{rm}^2(\cdot)$ as a $V$-statistic

For ease of notation let $X \equiv X_{t;r}$ and $Y \equiv X_{t+j;m}$. Suppose that $Z^i = (X^i, Y^i)$, $i = 1, \ldots, 6$, are independent and identically distributed copies of the vector $Z = (X, Y)$. By Prop. 2.6 of Lyons (2013), define the function

$$f(u_1, u_2, u_3, u_4) = |u_1 - u_2| - |u_1 - u_3| - |u_2 - u_4| + |u_3 - u_4|,$$

where $u_i \in \mathbb{R}$, $i = 1, \ldots, 4$. Now, set

$$h(Z^1, \ldots, Z^6) = f(X^1, X^2, X^3, X^4) f(Y^1, Y^2, Y^5, Y^6). \tag{A.1}$$

Provided that $E(|X_{t;r}|^2) < \infty$, $\forall\, r = 1, \ldots, d$, we obtain that

$$
\begin{aligned}
E\left\{ h(Z^1, \ldots, Z^6) \right\} &= E(|X^1 - X^2|\,|Y^1 - Y^2|) + E(|X^1 - X^2|) E(|Y^5 - Y^6|) \\
&\quad - 2E(|X^1 - X^2|\,|Y^1 - Y^5|) \\
&= V_{rm}^2(j),
\end{aligned}
$$

where the last equality follows from Székely et al. (2007, Remark 3). We consider the symmetrized version of $h(\cdot)$ given by (A.1), defined as

$$\tilde{h}(Z^1, \ldots, Z^6) = \frac{1}{6!} \sum_{\sigma \in \{1, \ldots, 6\}} h(Z^{\sigma(1)}, \ldots, Z^{\sigma(6)}), \tag{A.2}$$

where $\sigma$ is a permutation of $\{1, \ldots, 6\}$. Then, we observe that

$$E\left\{\tilde{h}(Z^1, \ldots, Z^6)\right\} = \frac{1}{6!} \sum_{\sigma \in \{1, \ldots, 6\}} E\left\{h(Z^{\sigma(1)}, \ldots, Z^{\sigma(6)})\right\} = V_{rm}^2(j).$$

Thus, we conclude that the function $V_{rm}^2(j)$ can be expressed as

$$V_{rm}^2(j) = \int_{\mathbb{R}^2} \cdots \int_{\mathbb{R}^2} \tilde{h}(z^1, \ldots, z^6) dF_{rm}(z^1) \ldots dF_{rm}(z^6).$$

Because of symmetry and following v. Mises (1947), a biased estimator of $V_{rm}^2(j)$ is given by

$$
\begin{aligned}
T_{rm} &= \frac{1}{(n-j)^6} \sum_{i_1=1}^{n-j} \cdots \sum_{i_6=1}^{n-j} h\left\{(X_{i_1;r}, X_{i_1+j;m}), \ldots, (X_{i_6;r}, X_{i_6+j;m})\right\} \\
&= \frac{1}{(n-j)^6} \sum_{i_1=1}^{n-j} \cdots \sum_{i_6=1}^{n-j} \left(|X_{i_1;r} - X_{i_2;r}| - |X_{i_1;r} - X_{i_3;r}| - |X_{i_2;r} - X_{i_4;r}| + |X_{i_3;r} - X_{i_4;r}|\right) \\
&\quad \times \left(|X_{i_1+j;m} - X_{i_2+j;m}| - |X_{i_1+j;m} - X_{i_3+j;m}| - |X_{i_2+j;m} - X_{i_4+j;m}| + |X_{i_5+j;m} - X_{i_6+j;m}|\right).
\end{aligned}
$$

After some calculations, we obtain that

$$
\begin{aligned}
T_{rm} &= \frac{1}{(n-j)^2} \sum_{i_1, i_2=1}^{n-j} |X_{i_1;r} - X_{i_2;r}| \, |X_{i_1+j;m} - X_{i_2+j;m}| \\
&\quad + \frac{1}{(n-j)^4} \sum_{i_1, i_2=1}^{n-j} |X_{i_1;r} - X_{i_2;r}| \sum_{i_1, i_2=1}^{n-j} |X_{i_1+j;m} - X_{i_2+j;m}| \\
&\quad - \frac{2}{(n-j)^3} \sum_{i_1, i_2, i_3=1}^{n-j} |X_{i_1;r} - X_{i_2;r}| \, |X_{i_1+j;m} - X_{i_3+j;m}| \\
&= \frac{1}{(n-j)^2} \sum_{t,s=1}^{n-j} A_{ts}^r B_{ts}^m \\
&= \hat{V}_{rm}^2(j),
\end{aligned}
$$

where the second equality is proved in Székely et al. (2007, Appendix). Because of symmetry we have that

$$\hat{V}_{rm}^2(j) = \int_{\mathbb{R}^2} \cdots \int_{\mathbb{R}^2} \tilde{h}(z^1, \ldots, z^6) d\hat{F}_{rm}(z^1) \ldots d\hat{F}_{rm}(z^6),$$

where $\hat{F}_{rm}(\cdot)$ denotes the empirical distribution function defined as

$$\hat{F}_{rm}(z) = \frac{1}{(n-j)} \sum_{t=1}^{n-j} \mathbb{I}(X_{t;r} \leq x, X_{t+j;m} \leq y).$$

Hence, we have shown that $\hat{V}_{rm}^2(j)$ is a $V$–statistic of order 6 with kernel function $\tilde{h}(\cdot)$ given by (A.2).

## A.2 Derivation of test statistics

Recall that the kernel $K(\cdot)$ satisfies

**Assumption A.1** $K : \mathbb{R} \to [-1, 1]$ 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 $|K(z)| \le C |z|^{-b}$ for large $z$ and $b > 1/2$.

Following the notation of the paper, we have that

$$\hat{F}(\omega, u, v) = \left\{ \hat{f}^{(r,m)}(\omega, u, v) \right\}_{r,m=1}^{d}, \quad \hat{F}_0(\omega, u, v) = \frac{1}{2\pi} \left\{ \hat{\sigma}_0^{(r,m)}(u, v) \right\}_{r,m=1}^{d},$$

where

$$\hat{f}^{(r,m)}(\omega, u, v) = \frac{1}{2\pi} \sum_{j=-(n-1)}^{(n-1)} (1 - |j|/n)^{1/2} K(j/p) \hat{\sigma}_j^{(r,m)}(u, v) e^{-ij\omega} \quad (\omega \in [-\pi, \pi]), \qquad (A.3)$$

with $K(\cdot)$ satisfying Assumption A.1 and $p$ is a bandwidth parameter. Consider the squared $L_2$-distance between $\hat{F}(\cdot, \cdot, \cdot)$ and $\hat{F}_0(\cdot, \cdot, \cdot)$

$$L_2^2 \left\{ \hat{F}(\omega, u, v), \hat{F}_0(\omega, u, v) \right\} = \int_{\mathbb{R}^2} \int_{-\pi}^{\pi} \|\hat{F}(\omega, u, v) - \hat{F}_0(\omega, u, v)\|_F^2 d\omega \mathcal{W}(du, dv)$$

$$= \int_{\mathbb{R}^2} \int_{-\pi}^{\pi} \mathrm{tr}\left[ \left\{ \hat{F}(\omega, u, v) - \hat{F}_0(\omega, u, v) \right\}^* \times \left\{ \hat{F}(\omega, u, v) - \hat{F}_0(\omega, u, v) \right\} \right] d\omega \mathcal{W}(du, dv).$$

But,

$$L_2^2 \left\{ \hat{F}(\omega, u, v), \hat{F}_0(\omega, u, v) \right\} = \frac{2}{\pi} \sum_{r,m} \sum_{j=1}^{n-1} (1 - j/n) K^2(j/p) \int_{\mathbb{R}^2} \left| \hat{\sigma}_j^{(r,m)}(u, v) \right|^2 \mathcal{W}(du, dv),$$

for any suitably weighting function $\mathcal{W}(\cdot, \cdot)$. In particular, employing

$$\mathcal{W}(du, dv) = \frac{1}{\pi |u|^2} \frac{1}{\pi |v|^2} du dv, \quad ((u, v) \in \mathbb{R}^2),$$

yields to

$$L_2^2 \left\{ \hat{F}(\omega, u, v), \hat{F}_0(\omega, u, v) \right\} = \frac{2}{\pi} \sum_{r,m} \sum_{j=1}^{n-1} (1 - j/n) K^2(j/p) \hat{V}_{rm}^2(j)$$

$$= \frac{2}{\pi} \sum_{j=1}^{n-1} (1 - j/n) K^2(j/p) \mathrm{tr}\{\hat{V}^*(j)\hat{V}(j)\}.$$

In terms of correlation matrices, recall that $D = \text{diag}\{V_{rr}(0), r = 1, \ldots, d\}$ and define the $d \times d$ matrix

$$R_j(u, v) = D^{-1/2}\Sigma_j(u, v)D^{-1/2}$$

with elements

$$\rho_j^{(r,m)}(u, v) = \frac{\sigma_j^{(r,m)}(u, v)}{\{V_{rr}(0)V_{mm}(0)\}^{1/2}}.$$

By recalling that

$$\sup_{(u,v)\in\mathbb{R}^2} \sum_{j=-\infty}^{\infty} \left|\sigma_j^{(r,m)}(u, v)\right| < \infty,$$

we can define the Fourier transform of $\rho_j^{(r,m)}(\cdot, \cdot)$ by

$$g^{(r,m)}(\omega, u, v) \quad = \quad \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \rho_j^{(r,m)}(u, v)e^{-ij\omega} \quad (\omega \in [-\pi, \pi]).$$

Define the $d \times d$ matrix

$$G(\omega, u, v) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} R_j(u, v)e^{-ij\omega} = \left\{g^{(r,m)}(\omega, u, v)\right\}_{r,m=1}^{d}.$$

Under independence, $G(\cdot, \cdot, \cdot)$ reduces to

$$G_0(\omega, u, v) \quad = \quad \frac{1}{2\pi}\left\{\rho_0^{(r,m)}(u, v)\right\}_{r,m=1}^{d}.$$

An analogous to (A.3) kernel-density estimator of $g^{(r,m)}(\cdot, \cdot)$ is given by

$$\hat{g}^{(r,m)}(\omega, u, v) \quad = \quad \frac{1}{2\pi} \sum_{j=-(n-1)}^{(n-1)} (1 - |j|/n)^{1/2}K(j/p)\hat{\rho}_j^{(r,m)}(u, v)e^{-ij\omega} \quad (\omega \in [-\pi, \pi]).$$

We can then define the estimators of $G(\cdot, \cdot, \cdot)$ and $G_0(\cdot, \cdot, \cdot)$ by

$$\hat{G}(\omega, u, v) \quad = \quad \left\{\hat{g}^{(r,m)}(\omega, u, v)\right\}_{r,m=1}^{d}$$

and

$$\hat{G}_0(\omega, u, v) \quad = \quad \frac{1}{2\pi}\left\{\hat{\rho}_0^{(r,m)}(u, v)\right\}_{r,m=1}^{d}$$

respectively. Considering now the squared $L_2$-distance between $G(\cdot, \cdot, \cdot)$ and $G_0(\cdot, \cdot, \cdot)$ we get

$$L_2^2\Big(\hat{G}(\omega, u, v), \hat{G}_0(\omega, u, v)\Big) = \int_{\mathbb{R}^2} \int_{-\pi}^{\pi} \|\hat{G}(\omega, u, v) - \hat{G}_0(\omega, u, v)\|_F^2 d\omega \mathcal{W}(du, dv)$$

$$= \int_{\mathbb{R}^2} \int_{-\pi}^{\pi} \text{tr}\left[\left\{\hat{G}(\omega, u, v) - \hat{G}_0(\omega, u, v)\right\}^* \times \left\{\hat{G}(\omega, u, v) - \hat{G}_0(\omega, u, v)\right\}\right] d\omega \mathcal{W}(du, dv).$$

After some calculations and choosing the weighting function defined in (A.4) of the main article, we find that

$$
\begin{aligned}
L_2^2\Big(\hat{G}(\omega, u, v), \hat{G}_0(\omega, u, v)\Big) &= \frac{2}{\pi} \sum_{r,m} \sum_{j=1}^{n-1} (1 - j/n) K^2(j/p) \int_{\mathbb{R}^2} \left|\hat{\rho}_j^{(r,m)}(u, v)\right|^2 \mathcal{W}(du, dv) \\
&= \frac{2}{\pi} \sum_{r,m} \sum_{j=1}^{n-1} (1 - j/n) K^2(j/p) \frac{\hat{V}_{rm}^2(j)}{\{\hat{V}_{rr}^2(0)\hat{V}_{mm}^2(0)\}^{1/2}} \\
&= \frac{2}{\pi} \sum_{r,m} \sum_{j=1}^{n-1} (1 - j/n) K^2(j/p) \hat{R}_{rm}^2(j) \\
&= \frac{2}{\pi} \sum_{j=1}^{n-1} (1 - j/n) K^2(j/p) \text{tr}\left\{\hat{R}^*(j)\hat{R}(j)\right\} \\
&= \frac{2}{\pi} \sum_{j=1}^{n-1} (1 - j/n) K^2(j/p) \text{tr}\left[\{\hat{D}^{-1/2}\hat{V}(j)\hat{D}^{-1/2}\}^* \hat{D}^{-1/2}\hat{V}(j)\hat{D}^{-1/2}\right] \\
&= \frac{2}{\pi} \sum_{j=1}^{n-1} (1 - j/n) K^2(j/p) \text{tr}\left\{\hat{V}^*(j)\hat{D}^{-1}\hat{V}(j)\hat{D}^{-1}\right\}.
\end{aligned}
$$

## A.3    Simultaneous critical values

Table 3 illustrates that critical values obtained under independence are not sensitive to the choice of response distribution or the dimension of a series. They depend on the sample size, as it should be expected. The first six columns of Table 3 have been obtained by considering univariate independent and identically distributed random variables. The last two columns correspond to independent samples drawn from a $d$-dimensional normal distribution with zero mean and equicorrelation matrix with non-diagonal elements $\rho_{ij} = \rho$, $i \neq j$. We set $\rho = 0.9$ and $0.4$ respectively. To further support our argument we include Table 4 which gives critical values for levels $\alpha = 0.025$ and $0.975$. Those values can be used for forming confidence intervals.

Table 3: Simultaneous empirical critical values at level $\alpha = 0.05$, for different sample sizes and dimensions. Results are based on $B = 499$ bootstrap replications for each of simulations.

| | N(0,1) | Pois(4) | Gamma(1,1) | Beta(2,3) | U(1,1) | $X_4^2$ | $\rho = 0.9$ | $\rho = 0.4$ |
|---|---|---|---|---|---|---|---|---|
| $n = 500$ | | | | | | | | |
| $d = 2$ | 0.116 | 0.117 | 0.118 | 0.113 | 0.109 | 0.117 | 0.117 | 0.116 |
| $d = 3$ | 0.117 | 0.116 | 0.130 | 0.115 | 0.111 | 0.119 | 0.119 | 0.117 |
| $d = 4$ | 0.118 | 0.117 | 0.132 | 0.118 | 0.112 | 0.121 | 0.119 | 0.119 |
| $d = 5$ | 0.121 | 0.116 | 0.129 | 0.114 | 0.113 | 0.122 | 0.116 | 0.117 |
| $n = 600$ | | | | | | | | |
| $d = 2$ | 0.106 | 0.105 | 0.102 | 0.103 | 0.100 | 0.105 | 0.106 | 0.106 |
| $d = 3$ | 0.107 | 0.105 | 0.106 | 0.106 | 0.102 | 0.112 | 0.108 | 0.106 |
| $d = 4$ | 0.108 | 0.106 | 0.104 | 0.104 | 0.104 | 0.108 | 0.105 | 0.107 |
| $d = 5$ | 0.109 | 0.107 | 0.106 | 0.102 | 0.106 | 0.110 | 0.110 | 0.107 |
| $n = 700$ | | | | | | | | |
| $d = 2$ | 0.098 | 0.095 | 0.099 | 0.096 | 0.092 | 0.098 | 0.097 | 0.098 |
| $d = 3$ | 0.098 | 0.097 | 0.098 | 0.097 | 0.092 | 0.103 | 0.099 | 0.099 |
| $d = 4$ | 0.100 | 0.099 | 0.096 | 0.097 | 0.095 | 0.101 | 0.099 | 0.099 |
| $d = 5$ | 0.099 | 0.097 | 0.098 | 0.098 | 0.093 | 0.100 | 0.099 | 0.098 |
| $n = 800$ | | | | | | | | |
| $d = 2$ | 0.091 | 0.090 | 0.092 | 0.091 | 0.088 | 0.092 | 0.095 | 0.092 |
| $d = 3$ | 0.091 | 0.093 | 0.093 | 0.090 | 0.089 | 0.095 | 0.094 | 0.094 |
| $d = 4$ | 0.093 | 0.091 | 0.095 | 0.090 | 0.087 | 0.093 | 0.092 | 0.094 |
| $d = 5$ | 0.092 | 0.092 | 0.095 | 0.090 | 0.089 | 0.097 | 0.092 | 0.092 |
| $n = 900$ | | | | | | | | |
| $d = 2$ | 0.087 | 0.084 | 0.088 | 0.085 | 0.080 | 0.086 | 0.086 | 0.084 |
| $d = 3$ | 0.086 | 0.085 | 0.089 | 0.084 | 0.081 | 0.088 | 0.087 | 0.088 |
| $d = 4$ | 0.087 | 0.086 | 0.090 | 0.085 | 0.083 | 0.089 | 0.088 | 0.088 |
| $d = 5$ | 0.087 | 0.086 | 0.093 | 0.085 | 0.084 | 0.087 | 0.087 | 0.087 |
| $n = 1000$ | | | | | | | | |
| $d = 2$ | 0.081 | 0.080 | 0.082 | 0.081 | 0.079 | 0.081 | 0.083 | 0.082 |
| $d = 3$ | 0.081 | 0.081 | 0.084 | 0.082 | 0.080 | 0.083 | 0.083 | 0.084 |
| $d = 4$ | 0.083 | 0.081 | 0.084 | 0.081 | 0.081 | 0.082 | 0.082 | 0.082 |
| $d = 5$ | 0.083 | 0.081 | 0.085 | 0.083 | 0.079 | 0.084 | 0.084 | 0.083 |

Table 4: Simultaneous empirical critical values at two different s levels $\alpha = 0.025$ and $0.975$, for different sample sizes and dimensions. Results are based on $B = 499$ bootstrap replications for each of 1000 simulations.

| | $\alpha = 0.025$ | | | $\alpha = 0.975$ | | |
|---|---|---|---|---|---|---|
| | $X_4^2$ | $\rho = 0.9$ | $\rho = 0.4$ | $X_4^2$ | $\rho = 0.9$ | $\rho = 0.4$ |
| $n = 500$ | | | | | | |
| $d = 2$ | 0.056 | 0.057 | 0.057 | 0.125 | 0.123 | 0.124 |
| $d = 3$ | 0.057 | 0.057 | 0.058 | 0.129 | 0.123 | 0.127 |
| $d = 4$ | 0.057 | 0.058 | 0.059 | 0.127 | 0.128 | 0.129 |
| $d = 5$ | 0.058 | 0.059 | 0.059 | 0.127 | 0.129 | 0.129 |
| $n = 800$ | | | | | | |
| $d = 2$ | 0.044 | 0.045 | 0.047 | 0.098 | 0.101 | 0.103 |
| $d = 3$ | 0.045 | 0.046 | 0.048 | 0.102 | 0.097 | 0.103 |
| $d = 4$ | 0.046 | 0.046 | 0.045 | 0.103 | 0.101 | 0.101 |
| $d = 5$ | 0.045 | 0.046 | 0.046 | 0.100 | 0.101 | 0.102 |

## A.4 Proofs

**of Theorem 1**     1. Suppose that $V_{rm}^2(j) \neq 0$ for a fixed $j$. Recall the notation of Section A.1. The stated assumptions imply that there exists a positive number $\delta$ such that for $\nu = 2 + \delta$ the following statements are true:

i. $E\left\{\left|\tilde{h}(Z^1, \ldots, Z^6)\right|^\nu\right\} < \infty,$

ii. $\sup_{i_1 < \cdots < i_6} E\left\{\left|\tilde{h}(Z^{i_1}, \ldots, Z^{i_6})\right|^\nu\right\} < \infty.$

Their verification is based on the proof of (Lyons, 2013, Prop. 2.6) and the form of the kernel given by (A.2). Hence, we conclude that

$$n^{1/2}\left\{\hat{V}_{rm}^2(j) - V_{rm}^2(j)\right\} \quad \to \quad N(0, 36\sigma^2),$$

with

$$\sigma^2 = \left[E\{\tilde{h}_1^2(Z^1)\} - V_{rm}^4(j)\right] + 2\left[\sum_{k=1}^{\infty} E\{\tilde{h}_1^2(Z^1)\tilde{h}_1^2(Z^{k+1})\} - V_{rm}^4(j)\right], \tag{A.4}$$

where $\tilde{h}_c(z^1, \ldots, z^c)$ denotes the conditional expectation of $\tilde{h}(\cdot)$ defined as:

$$\tilde{h}_c(z^1, \ldots, z^c) = E\left\{\tilde{h}(z^1, \ldots, z^c, Z^{c+1}, \ldots, Z^6)\right\} \quad (c = 1, \ldots, 5).$$

The above result follows directly from Yoshihara (1976, Thm. 1).

2. Considering now the case where the data are pairwise independent, we observe that

$$h_1(z^1) = E\left\{h(z^1, Z^2, \ldots, Z^6)\right\} = 0,$$

which implies that $\tilde{h}_1(z^1) = 0$, almost sure. The latter shows that the $V$–statistic $\hat{V}_{rm}^2(j)$ has a degeneracy of order 1. The statistic $\hat{V}_{rm}^2(\cdot)$ can be decomposed as (Hoeffding, 1948; Sen, 1972; Yoshihara, 1976)

$$\hat{V}_{rm}^2(j) = \binom{6}{2} V_n^{(2)} + R_n, \tag{A.5}$$

where $R_n = O_P(n^{1+\gamma})$, $\gamma > 2$. $V_n^{(2)}$ is a $V$–statistic of order 2 with kernel function $\tilde{h}^{(2)}(z^1, z^2)$ which is

the 'centered' version of $\tilde{h}_2(z^1, z^2)$ given by (Serfling, 1980, p. 222)

$$
\begin{aligned}
\tilde{h}^{(2)}(z^1, z^2) &= \tilde{h}_2(z^1, z^2) - \int_{\mathbb{R}^2} \tilde{h}_2(z^1, z^2) dF_{rm}(z^1) - \int_{\mathbb{R}^2} \tilde{h}_2(z^1, z^2) dF_{rm}(z^2) \\
&+ \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \tilde{h}_2(z^1, z^2) dF_{rm}(z^1) dF_{rm}(z^2),
\end{aligned}
$$

such that

$$
V_n^{(2)} = \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \tilde{h}^{(2)}(z^1, z^2) d\hat{F}_{rm}(z^1) d\hat{F}_{rm}(z^2). \tag{A.6}
$$

Under pairwise independence, $\tilde{h}^{(2)}(z^1, z^2) = \tilde{h}_2(z^1, z^2)$ almost surely.

We further observe that $h_2(z^1, z^2) = E\left\{ h(z^1, z^2, Z^3, \ldots, Z^6) \right\} = d_X(x^1, x^2) d_Y(y^1, y^2)$ by recalling that for $u, u' \in \mathbb{R}$ and $X$ a real valued random variable

$$
m_X(u) = E(|X - u|), \quad \bar{m}_X = E\{m_X(X)\}, \quad d_X(u, u') = |u - u'| - m_X(u) - m_X(u') + \bar{m}_X.
$$

In addition,

$$
\tilde{h}_2(z^1, z^2) = E\left\{ \tilde{h}(z^1, z^2, Z^3, \ldots, Z^6) \right\} = \binom{6}{2}^{-1} h_2(z^1, z^2).
$$

Thus, (A.5) and (A.6) show that

$$
\begin{aligned}
\hat{V}_{rm}^2(j) &= \binom{6}{2} \binom{6}{2}^{-1} \frac{1}{(n-j)^2} \sum_{t,s=1}^{n-j} h_2(z^t, z^s) + R_n \\
&= \frac{1}{(n-j)^2} \sum_{t,s=1}^{n-j} d_X(X_{t;r}, X_{s;r}) d_Y(Y_{t;m}, Y_{s;m}) + R_n.
\end{aligned}
$$

But $nR_n \to 0$, in probability, as $n \to \infty$. Therefore,

$$
n\hat{V}_{rm}^2(j) - \frac{1}{n} \sum_{t,s=1}^{n-j} d_X(X_{t;r}, X_{s;r}) d_Y(Y_{t;m}, Y_{s;m}) \to 0,
$$

in probability, as $n \to \infty$. In addition, under pairwise independence, we have that $E\left\{ \left( h_2(z, Z^s) \mid Z^1, \ldots Z^{s-1} \right) \right\} = 0$ almost surely. Therefore applying Theorem 1 of Leucht and Neumann (2013a) shows that

$$
\frac{1}{n} \sum_{t,s=1}^{n-j} d_X(X_{t;r}, X_{s;r}) d_Y(Y_{t;m}, Y_{s;m}) \to Z = \sum_l \lambda_l Z_l^2,
$$

in distribution, as $n \to \infty$, and so the proof is now completed.

Define

$$\bar{f}^{(r,m)}(\omega, u, v) \quad = \quad \frac{1}{2\pi} \sum_{j=-(n-1)}^{(n-1)} K(j/p)(1-|j|/n)^{1/2} \tilde{\sigma}_j^{(r,m)}(u,v) e^{-ij\omega},$$

where

$$\tilde{\sigma}_j^{(r,m)}(u,v) \quad = \quad \frac{1}{n-|j|} \sum_{t=|j|+1}^{n} \psi_{t;r}(u) \psi_{t-|j|;m}(v) \tag{A.7}$$

and

$$\psi_{t;r}(u) \quad \equiv \quad e^{iuX_{t;r}} - \phi^{(r)}(u). \tag{A.8}$$

The corresponding pseudoestimator of the generalized spectral density matrix is defined as

$$\bar{F}(\omega, u, v) \quad = \quad \frac{1}{2\pi} \sum_{j=-(n-1)}^{(n-1)} K(j/p)(1-|j|/n)^{1/2} \tilde{\Sigma}_{|j|}(u,v) e^{-ij\omega},$$

where $\tilde{\Sigma}_{|j|}(\cdot, \cdot)$ is the covariance matrix of $e^{iuX_t}$ with elements given by (A.7). For the proof of Theorem 2, we will need the following two lemmas whose proof is omitted as it follows closely the arguments given in the supplementary material of Fokianos and Pitsillou (2017) and the fact that $\beta$-mixing implies $\alpha$-mixing.

**Lemma A.4.1** *Let $\{X_t\}$ be a $\beta$-mixing strictly stationary process with mixing coefficients satisfying $\beta(k) = O(k^{-2})$. Suppose that $E|X_{t,r}|^2 < \infty$, $r = 1, \dots, d$. Then we have that*
$(n-j)^2 E\left|\hat{\sigma}_j^{(r,m)}(u,v) - \tilde{\sigma}_j^{(r,m)}(u,v)\right|^2 \leq C$ *and* $(n-j)E\left|\tilde{\sigma}_j^{(r,m)}(u,v)\right|^2 \leq C$ *uniformly in $(u,v) \in \mathbb{R}^2$ for*
$r, m = 1, \dots, p$.

**Lemma A.4.2** *Let $\{X_t\}$ be a $\beta$-mixing strictly stationary process with mixing coefficients satisfying $\beta(k) = O(k^{-2})$. Suppose that $E|X_{t,r}|^2 < \infty$, $r = 1, \dots, d$. For each $\gamma > 0$, let $D(\gamma) = \{(u,v) : \gamma \leq |u| \leq 1/\gamma, \gamma \leq |v| \leq 1/\gamma\}$. Then*

$$\int_{D(\gamma)} \sum_{j=1}^{n-1} K^2(j/p)(n-j)\left(\left|\hat{\sigma}_j^{(r,m)}(u,v)\right|^2 - \left|\tilde{\sigma}_j^{(r,m)}(u,v)\right|^2\right)\mathcal{W}(du,dv) \quad = \quad O_P(p/\sqrt{n})$$
$$= \quad o_P(\sqrt{p})$$

*for $r, m = 1, \dots, p$ as $p/n \to 0$.*

**of Theorem 2** It can be shown that (Hong (1999))

$$\sum_{j=1}^{n-1} K^2(j/p)(n-j) \left| \tilde{\sigma}_j^{(r,m)}(u,v) \right|^2 \;\; = \;\; \hat{C}^{rm}(u,v) + \hat{V}^{rm}(u,v) \tag{A.9}$$

where

$$\hat{C}^{rm}(u,v) \;\; = \;\; \sum_{j=1}^{n-1} \frac{K^2(j/p)}{n-j} \left\{ \sum_{t=j+1}^{n} C_{ttj}^{rm}(u,v) \right\},$$

$$\hat{V}^{rm}(u,v) \;\; = \;\; \sum_{j=1}^{n-1} \frac{K^2(j/p)}{n-j} \left\{ \sum_{t=j+2}^{n} \sum_{s=j+1}^{t-1} V_{tsj}^{rm}(u,v) \right\},$$

with

$$V_{tsj}^{rm}(u,v) \;\; = \;\; C_{tsj}^{rm}(u,v) + C_{stj}^{rm}(u,v)^*$$

and

$$C_{tsj}^{rm}(u,v) \;\; = \;\; \psi_{t;r}(u)\psi_{s;r}(u)^*\psi_{t-j;m}(v)\psi_{s-j;m}(v)^*,$$

where $^*$ denotes complex conjugate.

For the first summand of (A.9), it holds that $\int_{D(\gamma)} C_{ttj}^{rm}(u,v)\mathcal{W}(du,dv)$ and $\int_{D(\gamma)} C_{ssj}^{rm}(u,v)\mathcal{W}(du,dv)$ are independent integrals unless $t = s$ or $s \pm j$ . In addition,

$$E \int_{D(\gamma)} C_{ttj}^{rm}(u,v)\mathcal{W}(du,dv) \;\; = \;\; C_0^{rm\gamma} \equiv \int_{D(\gamma)} \sigma_0^{(r,r)}(u,-u)\sigma_0^{(m,m)}(v,-v)\mathcal{W}(du,dv) < \infty,$$

shows that $E\left[\sum_{t=j+1}^{n}\left\{ \int_{D(\gamma)} C_{ttj}^{rm}(u,v)d\mathcal{W} - C_0^{rm\gamma} \right\}\right]^2 \leq C(n-j)$. Hence, by Markov's inequality, Cauchy-Schwarz inequality and the properties of the kernel function, we obtain that

$$\int_{D(\gamma)} \hat{C}^{rm}(u,v)d\mathcal{W} \;\; = \;\; O_P(p/\sqrt{n}) + C_0^{rm\gamma} \sum_{j=1}^{n-1} K^2(j/p). \tag{A.10}$$

So, using Lemma A.4.2, equations (A.9) and (A.10) we have the following:

$$\int_{D(\gamma)} \left\{ \sum_{j=1}^{n-1} K^2(j/p)(n-j) \left| \hat{\sigma}_j^{(r,m)}(u,v) \right|^2 \right\} \mathcal{W}(du,dv)$$

$$= \int_{D(\gamma)} \left\{ \sum_{j=1}^{n-1} K^2(j/p)(n-j) \left| \tilde{\sigma}_j^{(r,m)}(u,v) \right|^2 \right\} \mathcal{W}(du,dv) + O_P(p/\sqrt{n})$$

$$= \int_{D(\gamma)} \hat{C}^{rm}(u,v) \mathcal{W}(du,dv) + \int_{D(\gamma)} \hat{V}^{rm}(u,v) \mathcal{W}(du,dv) + O_P(p/\sqrt{n})$$

$$= C_0^{rm\gamma} \sum_{j=1}^{n-1} K^2(j/p) + \hat{V}_n^{rm\gamma} + O_P(p/\sqrt{n}),$$

where $\hat{V}_n^{rm\gamma} \equiv \int_{D(\gamma)} \hat{V}^{rm}(u,v) \mathcal{W}(du,dv)$. Therefore

$$
\begin{aligned}
T_{n;\gamma}^{(r,m)} &\equiv \int_{D(\gamma)} \left\{ \sum_{j=1}^{n-1} K^2(j/p)(n-j) \left| \hat{\sigma}_j^{(r,m)}(u,v) \right|^2 \right\} \mathcal{W}(du,dv) \\
&= C_0^{rm\gamma} \sum_{j=1}^{n-1} K^2(j/p) + \hat{V}_n^{rm\gamma} + O_P(p/\sqrt{n}).
\end{aligned}
\tag{A.11}
$$

Given Assumption A.1 and by applying Hong (1999, Thm. A.3) on $D(\gamma)$, we obtain

$$\hat{V}_n^{rm\gamma} = \hat{V}_{ng}^{rm\gamma} + o_P(\sqrt{p}) \tag{A.12}$$

where

$$\hat{V}_{ng}^{rm\gamma} = \sum_{t=g+2}^{n} \sum_{s=1}^{t-g-1} \sum_{j=1}^{g} \frac{K^2(j/p)}{n-j} \int_{D(\gamma)} V_{tsj}^{rm}(u,v) \mathcal{W}(du,dv)$$

and $g \equiv g(n)$ such that $g/p \to 0, g/n \to 0$. Now, by applying Hong (1999, Thm. A.4) on $D(\gamma)$ we get the following:

$$\left\{ p D_0^{rm\gamma} \int_0^{\infty} K^4(z) dz \right\}^{-1/2} \hat{V}_{ng}^{rm\gamma} \to N(0,1) \tag{A.13}$$

as $n \to \infty$ in distribution, where

$$D_0^{rm\gamma} = 2 \int_{D(\gamma)} \left| \sigma_0^{(r,r)}(u,u') \sigma_0^{(m,m)}(v,v') \right|^2 \mathcal{W}(du,dv) \mathcal{W}(du',dv').$$

Equations (A.11), (A.12) and (A.13) yield to

$$
\frac{\int_{D(\gamma)}\left\{\sum_{j=1}^{n-1}K^2(j/p)(n-j)\left|\hat{\sigma}_j^{(r,m)}(u,v)\right|^2\right\}\mathcal{W}(du,dv)-C_0^{rm\gamma}\sum_{j=1}^{n-1}K^2(j/p)}{\left\{pD_0^{rm\gamma}\int_0^\infty K^4(z)dz\right\}^{1/2}} \to N(0,1), \quad \text{(A.14)}
$$

as $n \to \infty$, in distribution.

Observe that $\hat{C}_0^{rm\gamma} - C_0^{rm\gamma} = O_P(1/\sqrt{n})$ and that $\hat{D}_0^{rm\gamma} \to D_0^{rm\gamma}$ in probability. Moreover, under Assumption A.1, $p \to \infty$ and $p/n \to 0$, $p^{-1}\sum_{j=1}^{n-1}K^2(j/p) = \int_0^\infty K^2(z)dz + O(p^{-1/2})$. Thus, one can replace $C_0^{rm\gamma}\sum_{j=1}^{n-1}K^2(j/p)$ by $\hat{C}_0^{rm\gamma}p\int K^2(z)dz$. Summarizing, (A.14) becomes

$$
\frac{T_{n;\gamma}^{(r,m)} - \hat{C}_0^{rm\gamma}p\int_0^\infty K^2(z)dz}{\left\{p\hat{D}_0^{rm\gamma}\int_0^\infty K^4(z)dz\right\}^{1/2}} \to N(0,1).
$$

The rest of the proof follows by similar arguments given in Fokianos and Pitsillou (2017), by showing that

$$
\limsup_{\gamma\to 0}\limsup_{n\to\infty}\left|T_n^{(r,m)} - T_{n;\gamma}^{(r,m)}\right| = 0. \quad \text{(A.15)}
$$

**of Corollary 1** We only show the proof of the first result. From Theorem 2 and under the null hypothesis of independence, the random variables $T_n^{(r,m)}$ satisfy

$$
\frac{T_n^{(r,m)} - \hat{C}_0^{rm\gamma}p\int_0^\infty K^2(z)dz}{\left\{p\hat{D}_0^{rm\gamma}\int_0^\infty K^4(z)dz\right\}^{1/2}} \to N(0,1) \quad \text{(A.16)}
$$

in distribution, as $n \to \infty$, for $r,m = 1,\ldots,d$. Following similar arguments analogous to the proof of Theorem 2, it can be shown that

$$
\sum_{r,m}\int_{D(\gamma)}\left\{\sum_{j=1}^{n-1}K^2(j/p)(n-j)\left|\hat{\sigma}_j^{(r,m)}(u,v)\right|^2\right\}\mathcal{W}(du,dv) = \sum_{r,m}C_0^{rm\gamma}\sum_{j=1}^{n-1}K^2(j/p)
$$
$$
+ \sum_{r,m}\hat{V}_n^{rm\gamma} + O_P(p/\sqrt{n}).
$$

Given Assumption A.1 and employing similar arguments to those of Hong (1999, Proof of Thm. A3), we obtain that

$$
\sum_{r,m}\hat{V}_n^{rm\gamma} = \sum_{r,m}\hat{V}_{ng}^{rm\gamma} + o_P(\sqrt{p})
$$

where

$$\hat{V}_{ng}^{rm\gamma} = \sum_{t=g+2}^{n} \sum_{s=1}^{t-g-1} \sum_{j=1}^{g} \frac{K^2(j/p)}{n-j} \int_{D(\gamma)} V_{tsj}^{rm}(u,v)\mathcal{W}(du,dv)$$

and $g \equiv g(n)$ such that $g/p \to 0, g/n \to 0$. Considering the definition of $V_{tsj}^{rm}(u,v)$ we obtain that

$$\hat{V}_{ng}^{\gamma} = \sum_{t=g+2}^{n} \left[ \sum_{j=1}^{g} \sum_{s=1}^{t-g} \frac{K^2(j/p)}{(n-j)} \int_{D(\gamma)} \sum_{r,m} \left\{ \psi_{t;r}(u)\psi_{t+j;m}(v)\psi_{s;r}(-u)\psi_{s+j;r}(-v) \right\} \mathcal{W}(du,dv) \right.$$

$$+ \left. \sum_{j=1}^{g} \sum_{s=1}^{t-g} \frac{K^2(j/p)}{(n-j)} \int_{D(\gamma)} \sum_{r,m} \left\{ \psi_{t;r}(-u)\psi_{t+j;m}(-v)\psi_{s;r}(u)\psi_{s+j;r}(v) \right\} \mathcal{W}(du,dv) \right]$$

From (A.8), we observe that $\left\{ \sum_{r,m} \psi_{t;r}(u)\psi_{t+j;m}(v) \right\}$ and $\left\{ \sum_{r,m} \psi_{s;r}(u)\psi_{s+j;m}(v) \right\}$ are independent for $t - s > g$ and $1 \le j \le g$. Thus, by applying Hong (1999, Thm. A4) we obtain that

$$\left\{ \sum_{r,m} pD_0^{rm\gamma} \int_0^{\infty} K^4(z)dz \right\}^{-1/2} \sum_{r,m} \hat{V}_{ng}^{rm\gamma} \quad \to \quad N(0,1)$$

as $n \to \infty$ in distribution. Thus, we get the required result on $D(\gamma)$. Then, following the same arguments as in the proof of Theorem 2, we finally get the following result

$$\tilde{M}_n = \frac{\sum_{r,m} T_n^{(r,m)} - \left(\sum_{r,m} \hat{C}_0^{rm\gamma}\right) p \int_0^{\infty} K^2(z)dz}{\left\{ \left(\sum_{r,m} \hat{D}_0^{rm\gamma}\right) p \int_0^{\infty} K^4(z)dz \right\}^{1/2}} \quad \to \quad N(0,1),$$

in distribution, as $n \to \infty$.

For the second result, recall that $\bar{T}_n^{(r,m)}$ may be written as

$$\bar{T}_n^{(r,m)} = \frac{1}{\hat{V}_{rr}(0)\hat{V}_{mm}(0)} T_n^{(r,m)}.$$

By recalling result (A.16), we get

$$\frac{\bar{T}_n^{(r,m)} - \hat{c}_0^{(r,m)} p \int_0^{\infty} K^2(z)dz}{\left\{ \hat{d}_0^{(r,m)} p \int_0^{\infty} K^4(z)dz \right\}^{1/2}} \to N(0,1),$$

in distribution, as $n \to \infty$, for $r, m = 1, 2, \ldots, d$, where

$$\hat{c}_0^{(r,m)} = \frac{\hat{C}_0^{(r,m)}}{\hat{V}_{rr}(0)\hat{V}_{mm}(0)}, \quad \hat{d}_0^{(r,m)} = \frac{\hat{D}_0^{(r,m)}}{\hat{V}_{rr}^2(0)\hat{V}_{mm}^2(0)}.$$

But $\hat{d}_0^{(r,m)} \to 2$ almost surely. Following the same methodology as before, we get that

$$\bar{M}_n = \frac{\bar{T}_n - \left(\sum_{r,m} \hat{c}_0^{(r,m)}\right) p \int_0^\infty K^2(z)dz}{d\left\{2p \int_0^\infty K^4(z)dz\right\}^{1/2}} \quad \to \quad N(0,1),$$

in distribution, as $n \to \infty$ and the proof is now completed.

**of Theorem 3** We prove the first result of the theorem. Recall $D(\gamma)$ defined in Lemma 2. For the proof we show the following: (i) $E \int_{D(\gamma)} \int_{-\pi}^{\pi} \left|\hat{f}^{(r,m)}(\omega, u, v) - f^{(r,m)}(\omega, u, v)\right|^2 d\omega \mathcal{W}(du, dv) \to 0$ which is proved similarly to the proof of Hong (1999, Proof of Thm. 2, p. 1213) on $D(\gamma)$ for all $r, m = 1, \ldots, d$ given that $\{X_t\}$ is a $\beta$-mixing strictly stationary, but not independent and identically distributed process, with mixing coefficients satisfying $\sum_k \beta(k) < \infty$ and the kernel function satisfies assumption A.1. Additionally, by applying Markov's inequality we get (ii) $\hat{C}_0^{rm\gamma} = O_P(1)$ and (iii) $\hat{D}_0^{rm\gamma} \to D_0^{rm\gamma}$ in probability.

The last remark together with assumption A.1 shows that

$$\frac{1}{p^{1/2}}\left\{\sum_{r,m} \hat{D}_0^{rm\gamma} p \int_0^\infty K^4(z)dz\right\}^{1/2} \quad \to \quad \left\{\sum_{r,m} D_0^{rm\gamma} \int_0^\infty K^4(z)dz\right\}^{1/2},$$

in probability. Using remarks (i) and (ii) and after some calculations we get that

$$\frac{1}{n}\left\{T_{n;\gamma} - \sum_{r,m} \hat{C}_0^{rm\gamma} \int_0^\infty K^2(z)dz\right\} \quad \to \quad \frac{\pi}{2} L_{2;\gamma}^2\left\{F(\omega, u, v), F_0(\omega, u, v)\right\},$$

in probability, where

$$T_{n;\gamma} = \sum_{r,m} T_{n;\gamma}^{(r,m)},$$

and

$$L_{2;\gamma}^2\left\{F(\omega, u, v), F_0(\omega, u, v)\right\} = \int_{D(\gamma)} \int_{-\pi}^{\pi} \text{tr}\left[\left\{F(\omega, u, v) - F_0(\omega, u, v)\right\}^*\right.$$

$$\left. \times \left\{F(\omega, u, v) - F_0(\omega, u, v)\right\}\right] d\omega \mathcal{W}(du, dv).$$

However, considering $\hat{C}_0^{rm\gamma} \to \hat{C}_0^{rm}$ and $\hat{D}_0^{rm\gamma} \to \hat{D}_0^{rm}$ as $\gamma \to 0$ and (A.15), the first result is now proved on $\mathbb{R}^2$. The second result follows with analogous arguments.

# References

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

Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: A practical and powerful approach. *J. R. Stat. Soc. Series B* **57**, 289–300.

Borovcova, S., R. Burton, and H. Dehling (1999). Consistency of the Takens estimator for the correlation dimension. *Ann. Appl. Probab.* **9**, 376–390.

Brockwell, P. J. and R. A. Davis (1991). *Time Series : Theory and Methods*. New York: Springer-Verlag. Second Edition.

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

Chen, X. (2016). Gaussian approximation for the sup-norm of high-dimensional matrix-variate U-statistics and its applications. availabe at http://arxiv.org/abs/1602.00199.

Chwialkowski, K. P., D. Sejdinovic, and A. Gretton (2014). A wild bootstrap for degenerate kernel tests. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Weinberger (Eds.), *Advances in Neural Information Processing Systems 27*, pp. 3608–3616.

Davis, R. A., M. Matsui, T. Mikosch, and P. Wan (2016). Applications of distance correlation to time series. *Bernoulli*. to appear.

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

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

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

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

Gretton, A., K. Fukumizu, C. H. Teo, L. Song, B. Schölkopf, and A. Smola (2008). A kernel statistical test of independence. *Adv. Neural. Inf. Process. Syst.* **20**, 585–592.

Hipel, K. W. and A. I. McLeod (1994). *Time Series Modelling of Water Resources and Environmental Systems*. Amsterdam: Elsevier Science Pub Co.

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

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Stat.* **19**, 293–325.

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

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

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

Hosking, J. R. M. (1980). Multivariate Portmanteau statistic. *J. Am. Stat. Assoc.* **75**, 349–386.

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

Lacal, V. and D. Tjøstheim (2018). Estimating and testing nonlinear local dependence between two time series. *J. Bus. Econ. Stat.*. to appear.

Leucht, A. and M. H. Neumann (2013a). Degenerate U- and V-statistics under ergodicity: asymptotics, bootstrap and applications in statistics. *Ann. Inst. Stat. Math.* **65**, 349–386.

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

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

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

McLachlan, G. J., K. Do, and C. Ambroise (2004). *Analyzing Microarray Gene Expression Data*. Hoboken, New Jersey: Wiley.

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

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

Romano, P. and L. A. Thombs (1996). Inference for autocorrelations under weak assumptions. *J. Am. Stat. 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. Stat.* **41**, 2263–2291.

Sen, P. K. (1972). Limiting behavior of regular functionals of empirical distributions for stationary *-mixing processes. *Z. Wahrsch. Verw. Gebiete* **25**, 71–82.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. New York: Wiley.

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

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

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

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

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

Tsay, R. S. (2014). *Multivariate Time Series Analysis With R and Financial Applications*. Hoboken, NJ: Wiley.

v. Mises, R. (1947). On the asymptotic distribution of differentiable statistical functions. *Ann. Math. Stat.* **18**, 309–348.

Xiao, H. and W. B. Wu (2014). Portmanteau test and simultaneous inference for serial covariances. *Stat. Sin.* **24**, 577–600.

Yoshihara, K. (1976). Limiting behavior of U-statistics for stationary absolutely regular processes. *Z. Wahrsch. Verw. Gebiete* **35**, 237–252.

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