

Article

Two Tests for Dependence (of Unknown Form) between Time Series

M. Victoria Caballero-Pintado ¹, Mariano Matilla-García ^{2,*}, Jose M. Rodríguez ³ and Manuel Ruiz Marín ³ 

¹ Metodos Cuantitativos para la Economía y la Empresa, Universidad de Murcia, 30100 Murcia, Spain

² Facultad de Económicas y Empresariales, Universidad Nacional de Educación a Distancia (UNED), 28040 Madrid, Spain

³ Departamento Métodos Cuantitativos, Ciencias Jurídicas y Lenguas Modernas, Universidad Politecnica de Cartagena, 30201 Cartagena, Spain

* Correspondence: mmatilla@cee.uned.es

Received: 26 July 2019; Accepted: 5 September 2019; Published: 9 September 2019



Abstract: This paper proposes two new nonparametric tests for independence between time series. Both tests are based on symbolic analysis, specifically on symbolic correlation integral, in order to be robust to potential unknown nonlinearities. The first test is developed for a scenario in which each considered time series is independent and therefore the interest is to ascertain if two internally independent time series share a relationship of an unknown form. This is especially relevant as the test is nuisance parameter free, as proved in the paper. The second proposed statistic tests for independence among variables, allowing these time series to exhibit within-dependence. Monte Carlo experiments are conducted to show the empirical properties of the tests.

Keywords: dependence; permutation entropy; symbolic correlation integral

1. Introduction

In Science, either Natural or Social, a primary focus of attention has been the evaluation of whether two (or more) variables measured over time are related. If there is a relation, then further research into the nature and strength of the relation could be worthwhile in order to make new discoveries and/or gain predictive accuracy. Accordingly, testing for dependence between two univariate time series has been widespread, as we will show below. However, the vast majority of literature on the topic has relied on correlation as a main tool for developing other sophisticated statistical devices (tests). On the other hand, any correlation-based test is designed to detect linear relationships between time series, which limits the scope and utility of the test for dealing with (unknown) potential forms of dependence between the series beyond linear ones. The relatively few attempts to deal with nonlinear relationships have relied on nonparametric methods that generally require large data sets. In this age of massive data, this is not a limitation and, therefore, new statistical tests with fewer assumptions and with a wider range to detect potential relationships can be devised. The goal of this paper is to develop statistical tools to test dependence among time series robust to the form of the potential dependence that is established among the variables.

The classical Pearson cross-correlation coefficient assesses a linear relationship between two time series. However, it is well-known that it is not reliable when the time series under study are autocorrelated. In order to solve this caveat, scholars have developed several alternative statistical tests. Most of the work done is parametric and is based on the residuals of estimated models. The starting point was Haugh [1], who proposed a test for non-correlation between two jointly Gaussian covariance stationary time series, say $\{X_t\}$ and $\{Y_t\}$, by first prewhitening X_t and Y_t and then basing the test

on the residual cross-correlation function. However, Haugh's innovative test was rapidly criticized because (of problems) of its low power against popular alternative hypotheses (see [2] and [3] for the earliest documented critiques). Since then, Haugh's technique has been extended in several ways and the common denominator has been to improve the power of the test, (see: [4–11] for details). Many of the results are summarized by [12]. Basically, all these procedures follow the main skeleton given by Haugh, namely, first to obtain white noise residuals by performing univariate autoregression, and then to check the sample cross-correlation of the residuals.

There have also been several attempts to avoid parametric approaches: see [13–16]. The last of these proposed a nonparametric test of independence which was designed to avoid the autoregressive moving average (ARMA) pre-specification and to avoid kernel selection methods. Pre-specification and kernel selection were potential limitations in Haugh's [1] and Hong's tests [8]. In this vein, this paper proposes two new nonparametric tests for independence between time series. The first test is developed for a scenario similar to that in Pearson's original test, namely, for situations in which each considered time series is independent (within-independence or serially independent). Therefore, the interest is to investigate if two (or more) internally independent time series can be interrelated (either in a linear way, as in the case of correlation; or, more generally, in a nonlinear way). As mentioned earlier, there are many instances in several scientific domains where each time series is serially-dependent (auto-dependent). The second test proposed tests for independence between variables by allowing these time series to exhibit within-dependence. We use the concept of the symbolic correlation integral, as introduced in [17], to construct tests robust to nonlinearities.

Symbolic correlation integral is directly linked and influenced by the popular and well-known correlation integral definition, which can be understood as a measure of dependence. In this regard, dependence is not restricted to linear correlation (as in the case of Haugh-based tests) and therefore it will be robust to other complex forms of dependence that might occur between variables. Symbolic analysis is a field that has attracted the interest of many scholars and practitioners from several scientific disciplines (see [17]). By relying on these kinds of nonparametric concepts and techniques, we avoid imposing restrictive parametric assumptions, such as linearity and normality and, therefore, more generally applicable tests can be constructed. There might be other possibilities like entropy based statistics, as introduced in [18] that could avoid some of these restrictions.

The rest of the paper is structured as follows: in Section 2, we present the notation that will be used throughout the paper and the definition of symbolic correlation that serves as a common thread in the rest of the paper. Section 3 is devoted to presenting and defining the concept of joint symbolic correlation integral and its corresponding estimator. In Section 4, two tests for independence between series and their corresponding asymptotic treatment are described. In Section 5, empirical size and power are analyzed to better understand the finite sample behavior of the tests under several scenarios. We make a multi-level comparison with other tests available in the literature, and we provide some guidelines about how to fix the parameters of the new tests.

2. Definitions and Notation

Let $\{x_t\}_{t \in I}$ be a real-valued time series from a strictly stationary stochastic process of real random variables, where I is a set of time indexes. For a positive integer $m \geq 2$, we denote by S_m the symmetric group of order $m!$, that is, the group formed by all the permutations of length m . Let $\pi = (i_1, i_2, \dots, i_m) \in S_m$. We will call an element π in the symmetric group S_m a symbol. We consider that the time series is embedded in an m -dimensional space as follows:

$$\bar{x}_t = (x_t, x_{t+1}, \dots, x_{t+(m-1)}).$$

When the set of indexes I is finite and has cardinality T , $\{\bar{x}_t\}_{t=1}^n$ is a vectorial time series of length $n = T - m + 1$. Each \bar{x}_t is called m -history and the positive integer m is usually known as the *embedding dimension*.

We say that \bar{x}_t is of π -type if and only if $\pi = (i_1, i_2, \dots, i_m)$ is the unique symbol in the group S_m satisfying the two following conditions:

$$(a) \quad x_{t+i_1} \leq x_{t+i_2} \leq \dots \leq x_{t+i_m}, \text{ and}$$

$$(b) \quad i_{s-1} < i_s \text{ if } x_{t+i_{s-1}} = x_{t+i_s}.$$

Condition (b) guarantees the uniqueness of the symbol π . This is justified if the values of x_t have a continuous distribution, so equal values are very uncommon, with a theoretical probability of occurrence of 0. In the case of a discrete distribution, then condition (b) guarantees that the symbolization map is well defined.

Next, we define the symbolization map

$$s_x : \mathbb{R}^m \longrightarrow S_m$$

defined by $s(\bar{x}_t) = \pi$ if and only if \bar{x}_t is of π -type. Notice that the symbolization map s transforms the vectorial time series of m -histories into a sequence of symbols. Moreover, each element of \mathbb{R}^m is mapped to a symbol of S_m providing a partition of \mathbb{R}^m of size $m!$, called symbolic partition of \mathbb{R}^m . Given two time instants, t and s , and two ordinal patterns $\pi, \delta \in S_m$, we define $p_{ts}^{\pi\delta}$ as the probability that, in time t , \bar{x}_t is of π -type and, in time s , \bar{x}_s is of δ -type. Thus, we can construct the following $m! \times m!$ probability matrix

$$PM_x(|t - s|) = (p_{ts}^{\pi\delta})_{\pi\delta}.$$

Within this context, we define the indicator function

$$I(s_x(\bar{x}_t), s_x(\bar{x}_s)) = \begin{cases} 1, & \text{if } s_x(\bar{x}_t) = s_x(\bar{x}_s), \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

which always takes the value 1 when the ordinal patterns of the m -histories \bar{x}_t and \bar{x}_s are the same. The indicator variable, $I(s_x(\bar{x}_t), s_x(\bar{x}_s))$, is a Bernoulli random variable with probability of success

$$\mu_{ts}^x = \sum_{\pi \in S_m} p_{ts}^{\pi\pi}. \tag{2}$$

Moreover, if $\{x_t\}$ is i.i.d. and $|t - s| \geq m$, then $\mu_{ts}^x = \frac{1}{m!}$ (as shown in [17]).

Based on these concepts, Caballero et al. [17] defined the symbolic correlation integral of a time series $\{\bar{x}_t\}_{t \in I}$ for an embedding dimension $m \geq 2$ as

$$SC^m = \int \int I(s_x(\bar{x}), s_x(\bar{y})) d\mu(\bar{x}) d\mu(\bar{y}), \tag{3}$$

and they proved that under the null that $\{x_t\}_{t \in I}$ is i.i.d., the statistic

$$\widehat{SC}^m = \frac{2}{n(n-1)} \sum_{t>s} I(s_x(\bar{x}_t), s_x(\bar{x}_s))$$

is asymptotically distributed as an $N(\frac{1}{m!}, \sigma_m)$, where the standard deviation σ_m does not depend on the sample time series.

An interesting relation between \widehat{SC}^m with a well-known index can be obtained. If $N(\pi)$ denotes the absolute frequency of $\pi \in S_m$ in the symbolic time series (of length n), then symbolic correlation integral can be rewritten as:

$$\widehat{SC}^m = \frac{1}{n(n-1)} \left(\sum_{\pi} N(\pi)^2 - n \right)$$

which is closely related to normalized Gini index or Index of Qualitative Variation (see [19] for details), which is given by

$$\widehat{IQV} = \frac{m!}{m! - 1} \left(1 - \sum_{\pi} \frac{N(\pi)^2}{n^2} \right).$$

3. Joint Symbolic Correlation Integral

There is a natural extension of the symbolic correlation integral to a multivariate scenario. Similar to Symbolic Correlation Integral, Joint Symbolic Correlation Integral will measure the probability that all univariate time series forming the multivariate time series have the same ordinal pattern at different time periods (where the patterns could potentially vary along the time series). As we are therefore interested in analyzing a k -dimensional time series, first, we will define the Joint Symbolic Correlation Integral (JSC) of a set $X \subset \mathbb{R}^{m_1} \times \dots \times \mathbb{R}^{m_k}$, and then we will focus on multivariate time series.

Definition 1 (Joint Symbolic Correlation Integral). Let $X_j \subset \mathbb{R}^{m_j}$ with $j = 1, 2, \dots, k$ be distributed according to invariant measures $\nu_1, \nu_2, \dots, \nu_k$, respectively, and $\nu = (\nu_1, \dots, \nu_k)$ a invariant measure on $X = X_1 \times X_2 \times \dots \times X_k$. The joint symbolic correlation integral of the set $X = X_1 \times X_2 \times \dots \times X_k$ is defined as

$$JSC^{\mathbf{m}}(X) = \int \underbrace{\dots}_{2k} \int \prod_{j=1}^k I(s_{x_j}(\bar{x}_j), s_{x_j}(\bar{y}_j)) d\nu(\bar{x}_1, \dots, \bar{x}_k) d\nu(\bar{y}_1, \dots, \bar{y}_k). \tag{4}$$

Notice that the symbolization map for the multivariate time series is component-wise, and we allow for each component of the multivariate time series to have a different embedding dimension.

In the case in which the invariant measures are independent, that is, $\nu = (\nu_1, \dots, \nu_k) = \prod_{j=1}^k \nu_j$, it follows that

$$\begin{aligned} JSC^{\mathbf{m}}(X) &= \int \underbrace{\dots}_{2k} \int \prod_{j=1}^k I(s_{x_j}(\bar{x}_j), s_{x_j}(\bar{y}_j)) d\nu(\bar{x}_1, \dots, \bar{x}_k) d\nu(\bar{y}_1, \dots, \bar{y}_k) = \\ &= \prod_{j=1}^k \int \int I(s_{x_j}(\bar{x}_j), s_{x_j}(\bar{y}_j)) d\nu_j = \prod_{j=1}^k SC^{m_j}(X_j), \end{aligned} \tag{5}$$

where $\mathbf{m} = (m_1, m_2, \dots, m_k)$.

Notice that, only under the null of independence among the time series, $JSC = \prod_{j=1}^k SC^{m_j} = \prod_{j=1}^k \left(1 - \frac{m_j! - 1}{m_j!} IQV_j \right)$.

Based on this definition, we also have a natural extension of JSC for multivariate time series. To this end, given a multivariate time series $\{w_t = (x_{1t}, x_{2t}, \dots, x_{kt})\}_{t \in I}$, each of the time series $\{x_{jt}\}$ is embedded in \mathbb{R}^{m_j} , as in the previous section, to construct the embedded multivariate time series $\bar{w}_t = (\bar{x}_{1t}, \bar{x}_{2t}, \dots, \bar{x}_{kt})$. Then, the symbol space for $\{\bar{w}_t\}_{t \in I}$ is defined as $\Gamma_k = \prod_{j=1}^k S_{m_j}$, the Cartesian product of symmetric groups S_{m_j} with $j = 1, 2, \dots, k$, and the symbolization map is defined as $s_w = (s_{x_1}, s_{x_2}, \dots, s_{x_k})$. Thus, the symbolization of the multivariate time series is component-wise.

Next, we are interested in computing the estimator, $\widehat{JSC}^{\mathbf{m}}$, of the joint symbolic correlation integral for $\{w_t = (x_{1t}, x_{2t}, \dots, x_{kt})\}_{t=1}^T$. To this end, we define the indicator function

$$JSts = \prod_{j=1}^k I(s_{x_j}(\bar{x}_{jt}), s_{x_j}(\bar{x}_{js})) \tag{6}$$

and then

$$\widehat{JSC}^m = \frac{2}{n(n-1)} \sum_{t=1}^{n-1} \sum_{s=t+1}^n JS_{ts}, \tag{7}$$

where $n = \min\{n_1, n_2, \dots, n_k\}$ with $n_j = T - m_j + 1$ for $j = 1, 2, \dots, k$. In addition, the indicator function defined by Equation (6) is a Bernoulli random variable with probability of success $\tilde{\mu}_{ts}$. When the time series that form the vectorial time series $\{w_t\}$ are independent between themselves, we have that

$$\tilde{\mu}_{ts} = \prod_{j=1}^k \mu_{ts}^{x_j}. \tag{8}$$

We should highlight that, when every $\{x_{jt}\}_{t=1}^T$ is a stationary time series, from [20], the estimator of joint symbolic correlation integral is asymptotically unbiased:

$$\lim_{n \rightarrow \infty} E \left[\widehat{JSC}^m \right] = JSC^m. \tag{9}$$

4. Testing Independence between Time Series with JSC

In this section, we construct a test for independence between time series based on JSC. We will distinguish between two null hypotheses. In the first, we will test independence between time series when they are i.i.d. In this case, we will provide the asymptotic distribution of the test statistic and will show the conditions under which this test is not affected by the intermediate step of estimating the parameters of a given model. Such tests are called nuisance-parameter-free tests. In the second, we will relax the null hypothesis to allow for the time series under consideration not to be i.i.d. In this second case, we will not give the asymptotic distribution but use bootstrapping to test for significance. Proofs can be found in Appendix A.

4.1. H_0 : Independent between Themselves While Being I.I.D.

We are going to consider a vectorial time series, $\{w_t = (x_{1t}, x_{2t}, \dots, x_{kt})\}_{t \in I}$, where each $\{x_{jt}\}$ is i.i.d. and they are independent between themselves.

Since the time series $\{x_{jt}\}$ are i.i.d., from [17], we have that $SC^{m_j} = \frac{1}{m_j!}$ for $j = 1, 2, \dots, k$. Moreover, since the time series are independent between themselves, from Equations (8) and (9), we have that

$$\lim_{n \rightarrow \infty} E \left[\widehat{JSC}^m \right] = JSC^m = \prod_{j=1}^k SC^{m_j} = \prod_{j=1}^k \frac{1}{m_j!}.$$

As the time series $\{x_{jt}\}$ are i.i.d. for all $j = 1, 2, \dots, k$, using Equation (8), it follows that

$$E \left[\widehat{JSC}^m \right] = \frac{2}{n(n-1)} \left(\sum_{t=1}^{n-m+1} \sum_{s=t+1}^{t+m-1} \tilde{\mu}_{ts} + \sum_{t=n-m+2}^{n-1} \sum_{s=t+1}^n \tilde{\mu}_{ts} + \sum_{t=1}^{n-m} \sum_{s=t+m}^n \prod_{j=1}^k \frac{1}{m_j!} \right), \tag{10}$$

where $m = \max\{m_1, m_2, \dots, m_k\}$ and then

$$\lim_{n \rightarrow \infty} E \left[\widehat{JSC}^m \right] = \prod_{j=1}^k \frac{1}{m_j!}.$$

Under these conditions, the following result provides the asymptotic distribution of \widehat{JSC}^m .

Theorem 1. Let $\{x_{jt}\}_{t \in I}$ with $j = 1, 2, \dots, k$ be k i.i.d. times series that are independent between themselves. Given embedding dimensions $m_1, m_2, \dots, m_k \geq 2$ and $m = \max\{m_1, m_2, \dots, m_k\}$, it follows that

$$\sqrt{\frac{n(n-1)}{2}} (\widehat{JSC}^m - JSC^m) \tag{11}$$

is asymptotically $\mathcal{N}(0, \sigma_m)$ distributed, where

$$JSC^m = \prod_{j=1}^k \frac{1}{m_j!},$$

$$\sigma_m^2 = \sum_{h_1=-m}^m \sum_{h_2=-m}^m E[(JS_{ts} - \tilde{\mu}_{ts})(JS_{t+h_1s+h_2} - \tilde{\mu}_{t+h_1s+h_2})].$$

In many cases, the researcher wants to apply statistics, like statistic Equation (11), to estimated errors (residuals) of a model fitted to the raw data. The asymptotic distribution of Equation (11) is derived under the assumption that true errors are considered. Even in the case that true errors are iid, the estimated errors might exhibit some form of dependence, and therefore the asymptotic distribution might be affected by the estimation process. We now wonder if the statistic Equation (11) can be safely applied after the estimation of some intermediate parameters.

Definition 2. Let $S(\hat{\theta})$ be a statistic that depends upon some consistently estimated parameter, $\hat{\theta}$. Assume that, at a true value θ ,

$$\sqrt{n} \left(\frac{S_n(\theta) - \mu(\theta)}{\sigma(\theta)} \right) \rightarrow^d \mathcal{N}(0, 1),$$

then $S(\hat{\theta})$ is a nuisance-parameter free statistic if

$$\sqrt{n} (S_n(\theta) - S_n(\hat{\theta})) \rightarrow^P 0,$$

where $\mu(\theta) = \lim_{n \rightarrow \infty} E[S_n(\theta)]$ and $\sigma^2(\theta) = \lim_{n \rightarrow \infty} E[(S_n(\theta) - \mu(\theta))^2]$.

Theorem 2 (Nuisance-free parameter property). Assume that the data-generating process for each time series $\{x_{jt}\}_{t \in I}$ is given by

$$x_{jt} = G_j(I_{t-1}^j, \theta_j) + u_{jt}, \tag{12}$$

where I_{t-1}^j is a finite set of regressors, θ_j is a vector of parameters, $\{u_{jt}\}$ are i.i.d., and G_j is a continuous function defined on a compact set for $j = 1, 2, \dots, k$. Then, given embedding dimensions $m_1, m_2, \dots, m_k \leq 2$ and $m = \max\{m_1, m_2, \dots, m_k\}$, the statistic \widehat{JSC}^m applied to the residuals $\{(u_{1t}, u_{2t}, \dots, u_{kt})\}_t$ is nuisance-parameter free.

4.2. H_0 : Time Series Independent between Themselves

When the time series under consideration, $\{x_{jt}\}_{t \in I}$, are not i.i.d., we consider the following statistic:

$$\hat{\delta}(\mathbf{m}) = \widehat{JSC}^{\mathbf{m}} - \prod_{j=1}^k \widehat{SC}^{m_j}. \tag{13}$$

Notice that, when the time series are independent between themselves, $\hat{\delta}(\mathbf{m}) = 0$ and $\hat{\delta}(\mathbf{m}) \neq 0$ otherwise. Thus, to test for significance, we use a bootstrap test based on the block-bootstrap proposed by Politis and White [21] and corrected in Patton et al. [22]. In order to be under the null of independence among the time series, we resample each time series independently rather than jointly.

Let us illustrate the procedure. Given the original time-series $\{x_{jt}\}_{t=1}^T$, we compute $\hat{\delta}(\mathbf{m})$ and evaluate the same for each of the $B - 1$ bootstrap realizations of the k time-series, namely $\hat{\delta}^{(b)}(\mathbf{m})$. Next, we compute the upper and lower p -values as

$$\text{upper} - p = \frac{1}{B} \sum_{i=1}^{B-1} \mathbf{1}(\hat{\delta}^{(b)}(\mathbf{m}) > \hat{\delta}(\mathbf{m})), \quad (14)$$

$$\text{lower} - p = \frac{1}{B} \sum_{i=1}^{B-1} \mathbf{1}(\hat{\delta}^{(b)}(\mathbf{m}) < \hat{\delta}(\mathbf{m})), \quad (15)$$

where $\mathbf{1}$ is the Heaviside function. Based on these values, the null hypothesis of independence among time series is rejected at a significance level α if $\text{upper} - p < \frac{\alpha}{2}$ or $\text{lower} - p < \frac{\alpha}{2}$.

5. Simulations

5.1. Empirical Size and Power

We present some evidence of the performance of the proposed tests when they are applied to systems that might (or might not) exhibit within-dependence (i.e., autodependent processes, whose easiest form is autocorrelation) and/or dependence between processes (between-dependence or cross-dependence). We refer to the statistics depicted in Equations (11) and (13) by Test1 and Test2. For simplicity, the statistics have been computed assuming that the correct time-lag has been obtained. To this end, we have considered the following systems of relations:

$$\begin{array}{ll} S1 : Y_t = \varepsilon_t & X_t = v_t \\ S2 : Y_t = 0.5Y_{t-1} + \varepsilon_t & X_t = 0.5X_{t-1} + v_t \\ S3 : Y_t = 0.5\varepsilon_{t-1}^2 + \varepsilon_t & X_t = 0.5v_{t-1}^2 + v_t \\ S4 : Y_t = 0.8Y_{t-1} + X_t + \varepsilon_t & X_t = 0.8X_{t-1} + v_t \\ S5 : Y_t = 0.3Y_{t-1} + 0.5Y_{t-2}X_{t-1} + \varepsilon_t & X_t = 0.7X_{t-1} + v_t \\ S6 : Y_t = \varepsilon_t & X_t = 0.8Y_{t-1}^2 + v_t \\ S7 : Y_t = 0.6v_{t-1}^2 + \varepsilon_t & X_t = 0.6X_{t-1} + v_t \\ S8 : Y_t = X_{t-1} & X_t = 4X_{t-1}(1 - X_{t-1}). \end{array}$$

We have selected these systems (with these parameters, sample sizes and relationships) because they have been studied in previous works, so comparability is easy. In particular, system S1 specifies two within-independent processes that are independent between themselves, while S2 specifies two between-independent Gaussian autoregressive AR(1) processes. System S3 represents two between-independent processes that exhibit a nonlinear form of within-dependence. On the other hand, systems S4, S5, S6, and S7 are between-dependent processes. S4 is a linear within-dependent model that entails ideal conditions (normality and linearity) for the application of parametric statistics. System S5 has a stable bivariate nonlinear autoregressive within-dependence. S6 model shows within-dependence in one of the variables, but not in the other. System S7 shows first a nontrivial nonlinear relation between the processes and, second, it considers two different forms of within-dependence. Finally, system S8 is very interesting as it is a nonlinear deterministic process formed from the chaotic logistic equation.

Each system was simulated 2000 times and the following tables collect the proportion of rejection the null hypothesis considered at the 5% nominal level. The tests were performed for $m = 3$. The selection of parameter m is open to the practitioner, and we will deal with this regard later on this section.

Table 1 shows the power and size results of Test1 for the eight models S1–S8 for the null X_t and Y_t being i.i.d. and independent between themselves.

Table 1. Size and Power of Test1.

	S1	S2	S3	S4	S5	S6	S7	S8
T = 200	0.045	0.288	0.175	1	0.389	0.190	0.271	1
T = 500	0.048	0.838	0.528	1	0.938	0.670	0.849	1
T = 1000	0.048	1	0.943	1	1	0.990	0.999	1
T = 3000	0.049	1	1	1	1	1	1	1

We observe first that, when the stochastic system is generated under the null (i.e., two independent processes that are both auto-independent), the test rejects according at the fixed 5% level. In this regard, the new test Test1 behaves as expected under the null, regardless of the sample size. The results for systems S2 and S3 suggest that Test1 test easily detects a departure from one of the conditions of the null (namely, within-dependence), even when they are between-independent processes. The results for S4, S5 and S7 indicate that, when the departure from the null is larger (in this case both null conditions are violated), then Test1 behaves powerfully, even in the case of nonlinear dependencies, as in systems S5 and S7. As regards system S6, notice that, despite the fact that process Y is i.i.d., process X depends on Y . Even in the case that one process has no within-independence but there is between dependence, the test exhibits power to detect these departures. Finally, the results for system S8 are especially interesting because the processes involved are purely deterministic (there is no random term) and the dependence between them is evident. Despite this peculiar dynamic structure, it is noticeable that the Pearson cross-correlation test has a rejection rate of the null of 5%; in other words, Pearson's test behaves at the nominal level of the statistic, which implies that it systematically suggests not to reject the null of independence between processes when there is an obvious deterministic dependence. In contrast, Test1 test rejects the null with great power. The explanation for this performance is that Pearson's test is limited to detecting linear relationships between variables, while Test1 considers any form of potential relationship.

On the other hand, it can be also concluded from the above comments on simulations that Test1 is not capable of distinguishing between the forms of dependences (between and/or within). In other words, if the final user obtains a rejection of the null, she does not know the reason for the rejection. Given the simplicity of the test and its power with complex dependence forms, it is advisable to use Test1 as a first step. However, to complete the process, it is necessary either to use Test2 or to apply some sort pre-whitening process. We now explore the first solution (Test2) and, later in the paper, we consider the behavior of the test in the case of pre-whitening in a multivariate scenario.

Table 2 shows the power and size results of Test2 for the eight models S1–S8 for the null X_t and Y_t being independent just between themselves.

Table 2. Size and Power of Test2.

	S1	S2	S3	S4	S5	S6	S7	S8
T = 200	0.049	0.056	0.056	0.998	0.068	0.247	0.104	1
T = 500	0.053	0.054	0.057	1	0.091	0.742	0.254	1
T = 1000	0.058	0.055	0.057	1	0.200	0.992	0.636	1
T = 3000	0.050	0.049	0.059	1	0.816	1	1	1

The empirical behavior of other available tests on the same systems and sample sizes are reported in the Appendix. Based on the results for these models, we make the following remarks:

(i) The output for systems S1, S2 and S3 hints that Test2 can correctly deal with models that exhibit several forms of within-dependence, and this internal dependence does not contaminate the ability of Test2 to indicate, at a nominal level of 0.05 that both processes are independent. In this regard, it is noteworthy to observe that, a Haugh-type test could not have been used with system S3 because these tests report confident results only for systems of linear and Gaussian dependence, as is the case of S2.

(ii) When within-dependence is linear, as in S4, the power of the test is impressive regardless of the sample size. This empirical fact implies that Test2 can be used to detect simple cases of linear relationships between variables. As reported in [14], Haugh-based tests also have extremely good power for this type of linear dependence. Accordingly, the final user of the tests could safely use the nonparametric Test2 test or parametric Haugh-based tests.

(iii) However, when there is within-dependence of nonlinear nature, as in S5, S6, S7 and S8, it is well known that Haugh based tests are unable to detect dependence, regardless of the sample size. As can be observed from this simulation, Test2 detects dependence when the sample is large enough.

From (i)–(iii), it can be concluded that Test2 can be used to effectively detect dependence between variables with fewer restrictions than other available tests in the literature, and, from a practical point of view, the larger the sample size, the more reliable the results are.

As mentioned earlier, an interesting advantage of both tests is that they can be used in a multivariate setting. We now consider two new sets of multivariate systems. The first set is formed of S9 and S10 systems. Each system is a three-variable stochastic linear system that is used in this paper to show that Test1 can be satisfactorily used for pre-whitened data as proved in the previous section:

$$\begin{aligned} \text{S9} \quad & Y_t = 0.6Y_{t-1} + \varepsilon_{yt} \quad X_t = 0.5X_{t-1} + \varepsilon_{xt} \quad Z_t = 0.7Z_{t-1} + \varepsilon_{zt}, \\ \text{S10} \quad & Y_t = 0.6X_{t-1} + \varepsilon_{yt} \quad X_t = 0.5X_{t-1} + \varepsilon_{xt} \quad Z_t = 0.7Y_{t-1} + \varepsilon_{zt}. \end{aligned}$$

These systems were generated and estimated by Ordinary Least Squares (OLS) to obtain the linear structure of each variable and residuals were then tested with Test1. The results are in Table 3.

Table 3. Test1 on Residuals.

	S9	S10
$T = 200$	0.053	0.051
$T = 500$	0.050	0.050
$T = 1000$	0.053	0.055
$T = 3000$	0.053	0.054

After removing the linear structure, independent estimated errors are obtained. Provided that errors are simulated independently, it is expected that Test1 will not reject the null at the nominal level, as shown in the table above.

The second set of systems is conducted to study the behavior of Test2 in a multivariate system of complex relationships. To this end, we have considered three systems:

$$\begin{aligned} \text{S11} \quad & Y_t = 0.5Y_{t-1} + Z_t & X_t = 0.5X_{t-1} + Z_t & Z_t = \varepsilon_t \\ \text{S12} \quad & R_{Y_t} = Y_{t-1}(2.9(1 - Y_{t-1}))\exp(-0.36Z_t) & R_{X_t} = X_{t-1}(3.1(1 - X_{t-1}))\exp(-0.3Z_t) & Z_t = \varepsilon_t \\ & Y_t = 0.35Y_{t-1} + \max(R_{Y_{t-4}}, 0) & X_t = 0.4X_{t-1} + \max(R_{X_{t-4}}, 0) & \\ \text{S13} \quad & Y_t = 0.5Y_{t-1} + \varepsilon_{yt} & X_t = 0.5X_{t-1} + 0.5Y_{t-1} + \varepsilon_{xt} & Z_t = \varepsilon_{zt}. \end{aligned}$$

S11 considers the case where two variables, X and Y, do not have between-dependence (cross-dependence) but are both driven by a common variable Z. One can think of Z as an environmental variable that determines (explains) X and Y, and therefore both are related by this external variable. Here, we expect Test2 to reject the null. Thus, S12 is a nonlinear and more complex model than S11, but, in essence, is similar. This system has two non-interacting variables, X and Y, that share common environmental forcing. Finally, S13 considers the case of three-variables where one of them has no dynamic structure, and the other two only have one-side dependence. We expect Test2 to be able to detect this dependence. The results are given in Table 4.

Table 4. Power of Test2 on Multivariate Systems.

	S11	S12	S13
$T = 200$	1	0.855	0.121
$T = 500$	1	0.935	0.415
$T = 1000$	1	0.974	0.879
$T = 3000$	1	0.998	1

The results suggest that Test2 is able to clearly detect departures from the null in a multivariate context. Even in the case of hidden common variables, the test unveils the indirect relationship, despite its linear or nonlinear nature and despite the sample size. It is also concluded that Test2 needs more observations (larger sample sizes) to detect dependences when considering scenarios like S13, where of six potential relationships between variables, only one exists (from Y to X).

5.2. Comparison with Other Tests

As we indicated in the introductory section, the technical literature on this topic has produced several statistics that test for independence between time series. This subsection aims to compare among the most relevant tests.

A comparison among tests can be conducted at several levels. We compare at the level of: the assumptions required to derive and implement the test, the parameters that the final user has to fix to conduct the test, and the empirical power of the test.

According to the literature, the improvements have occurred around some criteria that to some extent are related to the required assumptions for deriving the statistic and implementing the test(s). On this regard, scholars have mainly focused on the following criteria:

- (i) stationarity (or not) of the system generating process,
- (ii) linearity (or not) of the system, and
- (iii) robustness (or not) to the presence of outliers.

On the other hand, all available statistics require the final user of these statistical tests to make certain decisions on some aspects that will necessarily affect the final result of the test. Provided the test is used in the residuals of the model, one of the most important decisions is the fact that a correct model needs to be estimated. Obviously, pre-estimation (or not) of an autoregressive model before using the test is a critical decision. Another important choice for using the test is due to the fact that some of the tests relied on the use of kernels. Throughout the literature, there has been some controversy regarding how to choose the kernel and to what extent empirical behavior of the test changes because of the selected kernel. Along with the kernel, a selection is also required for truncation parameters. Finally, all the tests have to choose the number of observations in the lag vector, which is equivalent to parameter m (embedding) of our tests.

These observations lead us to complete the previous list with the following items:

- (iv) Pre-estimation,
- (v) Kernel selection,
- (vi) embedding selection.

Table 5 allows us to compare the tests considered in terms of the robustness to processes that might be nonstationary and nonlinear, and to the presence of outliers (criteria (i)–(iii)). The table also facilitates comparisons in line with the choices that the user has to make before using the test(s), (criteria (iv)–(vi)).

Table 5. Comparison among Tests’ Properties.

Tests Properties	Robust to			Choices		
	Nonstationarity	Nonlinearity	Outliers	Pre-Estimation	Kernel	Embedding
Haugh	No	No	No	Yes	No	Yes
Hong	No	No	No	Yes	Yes	Yes
Koch & Yang	No	No	No	Yes	No	Yes
Li & Hui	No	No	Yes	Yes	No	Yes
Hallin & Saidi [23]	No	No	No	Yes	No	Yes
Pham et al.	Allows cointegration	No	No	Yes	No	Yes
Test 2 (Test 1)	No (No)	Yes (Yes)	Yes (Yes)	No (No)	No (No)	Yes (Yes)

According to the previous table, the tests presented in this work have a greater range of applicability. The data that can be analyzed can be compatible with an ample number of models. In other words, other tests are less generally applicable. From a practical point of view, the new tests facilitate user work, since she has to select a smaller number of parameters. This is especially relevant since we alleviate the burden of modeling a (correct) autoregressive process. Any of our techniques only require selecting parameter m , which is a necessary parameter in all the available tests.

As explained in the introductory section, mainly all the available tests are derived from a seminal Haugh’s test, which is best known along with Hong’s test, which is the test with better behavior in terms of power. To complete the comparison, we now compare the results in terms of power. To do so, we consider these two well-known tests, namely Haugh and Hong tests, and compare it with Test2, which is the most general one. To make a fair comparison, it is only conducted on models to which both tests can be applied.

We firstly describe these competitive tests and then show their results on the corresponding systems.

The Haugh’s (1976) [1] procedure considers the following portmanteau statistic given by

$$S_M = n \sum_{j=-M}^M r_{\hat{u}\hat{v}}^2(j),$$

where $r_{\hat{u}\hat{v}}(j) = \sum_{t=j+1}^n \hat{u}_t \hat{v}_{t-j} / (\sum_{t=1}^n \hat{u}_t^2 \sum_{t=1}^n \hat{v}_t^2)^{1/2}$ are the residual cross-correlations for $0 \leq j \leq n - 1$, $r_{\hat{u}\hat{v}}(j) = r_{\hat{u}\hat{v}}(-j)$ for $1 - n \leq j < 0$, and $\hat{u}_t, \hat{v}_t, t = 1, \dots, n$ are the two residual series of length n , obtained by fitting univariate models to each of the series. The constant $M \leq n - 1$ is a fixed integer and must be chosen a priori. The asymptotic distribution of S_M is chi-square under the null hypothesis of independence and the hypothesis is rejected for large values of the test statistic.

Hong (1996) [8] generalizes Haugh’s statistic. In fact, Hong’s test is a weighted sum of residual cross-correlations of the form

$$Q_n = \frac{n \sum_{j=1-n}^{n-1} k^2(j/d) r_{\hat{u}\hat{v}}^2(j) - M_n(k)}{[2V_n(k)]^{1/2}},$$

where $M_n(k) = \sum_{j=1-n}^{n-1} (1 - |j|/n) k^2(j/d)$ and $V_n(k) = \sum_{j=2-n}^{n-2} (1 - |j|/n) (1 - (|j| + 1)/n) k^4(j/d)$. The weighting depends on a kernel function k and a smoothing parameter d (both have to be selected a priori). Under the null hypothesis, the test statistic Q_n is asymptotically $N(0, 1)$ and it rejects the null for large values of Q_n .

The empirical power results on systems are collected in Table 6.

Table 6. Power Comparison among Tests.

	Test	S4	S5	S6	S7
T = 200	S_M	1	0.21	0.11	0.11
	Q_m	1	0.23	0.20	0.29
	Test2	0.99	0.07	0.25	0.10
T = 500	S_M	1	0.20	0.10	0.14
	Q_m	1	0.22	0.23	0.31
	Test2	1	0.09	0.74	0.25
T = 1000	S_M	1	0.31	0.11	0.15
	Q_m	1	0.24	0.28	0.34
	Test2	1	0.20	0.99	0.64
T = 3000	S_M	1	0.29	0.37	0.17
	Q_m	1	0.24	0.27	0.33
	Test2	1	0.82	1	1

Results point to several observations: (a) all tests have maximum power for the simplest system (S4), so all are highly competitive, and therefore the following comments will only apply to systems S5, S6 and S7. (b) None of the tests compared is competitive for (small) sample sizes: 200 and 500. (c) Haugh and Hong tests do not improve power by increasing the sample size; however, Test2 not only improves, but also reaches levels close to full power.

5.3. Selecting Parameter m

As mentioned earlier, all tests involve the selection of a parameter, m that conform the basic units of analysis. This parameter is an integer which stands for the fix length of the vectors that are formed to be introduced in the tests. These vectors are generally the first m consecutive observations from a time series (raw data or residuals). We have referred throughout this paper to this parameter as embedding dimension or m -history. This terminology is very frequent in the field related to entropy and nonlinear chaotic dynamical systems.

None of the cited tests provide advice for how to select this parameter. In this section, we reflect on how to select the parameter m and we analyze the empirical effect of increasing m .

An obvious observation is that, if the system generating process is constructed by two (or more) cross dependent equations where dependence is in lags larger than m , then no test will capture such a dependence and will consider that they are independent time series. One natural solution is to construct the same m -histories with some fixed time-delay, namely, $\bar{x}_{t,\tau} = (x_t, x_{t+\tau}, x_{t+2\tau}, \dots, x_{t+(m-1)\tau})$.

This problem and also this solution is rarely found in Economics and Finance; however, it is more frequent in Physics and subfields related with nonlinear and chaotic behavior. Indeed, the modeling and prediction of chaotic time series require proper reconstruction of the state space from the available data in order to successfully estimate invariant properties of the embedded attractor. Thus, one must choose appropriate delay time τ and embedding dimension m for phase space reconstruction. For the aim of the presented tests, there is no need to go beyond what other tests do for analyzing dependences.

Provided that the new tests rely on symbolic measures, it is worth considering in this regard that previous research on these measures (see [24]) has provided rules for selecting m . Each m induces a number of symbols. The number of symbols has to be large enough to capture departures from the null. For $m = 2$, only $(2!)^2 = 4$ symbols are being evaluated and are expected to be too few to detect departures from the null(s). For the next integer, $m = 3$, 36 symbols are evaluated, and in the case of increasing to an embedding dimension of 4, 576 symbols will then be used.

According to the authors, gains in terms of power can be obtained by increasing m according to the following rule: given a data set of T observations, the embedding dimension will be the largest m that satisfies $5 \times \text{number of symbols} < T$ with $m = 2, 3, 4, \dots$. In our case, the rule is $5(m!)^2 < T$. The intuition beyond this rule is clear; on the one hand, the larger the number of symbols, the larger the

sensitivity for detecting departures from the null. On the other hand, $m!$ grows very fast and statistical devices require enough samples to behave normally. Note that the larger the m , the finer the search for dependences, but at the cost of increasing sample size required to satisfy the rule.

Finally, the study is completed by empirically analyzing the effect (in terms of power) of a change in the embedding parameter. We consider the basic models (S4–S8) and we evaluate for $m = 2, 3, 4$. It makes no sense to use $m = 5$ because, in this case, 14,400 symbols will be used while (at most) only 3000 observations are available. Results are collected in Table 7.

Table 7. Behavior of Test2 for Several m .

m	T	S4	S5	S6	S7	S8
2	200	0.996	0.07	0.052	0.064	0.927
3	200	0.999	0.046	0.209	0.082	1
4	200	0.983	0.054	0.195	0.087	1
2	500	1	0.057	0.053	0.066	0.998
3	500	1	0.098	0.732	0.271	1
4	500	1	0.112	0.699	0.289	1
2	1000	1	0.06	0.052	0.059	1
3	1000	1	0.226	0.995	0.64	1
4	1000	1	0.321	0.994	0.749	1
2	3000	1	0.062	0.037	0.052	1
3	3000	1	0.799	1	0.998	1
4	3000	1	0.98	1	1	1

From these results, we firstly observe that, for a few number of symbols (i.e., for $m = 2$), the test has no power (as expected), except for the deterministic system and the linear one. In addition, secondly, the power of the test tends to increase with m . For this reason, we recommend the potential users of the test to adhere to the automatic rule for choosing parameter m .

6. Conclusions

In this work, we have extended the concept symbolic correlation integral to the multivariate domain with the intention of studying, in a non-parametric way, the dependency relationships that might occur between observed variables. Taking as our starting point the multivariate correlation integral, we have developed two statistical tests that can be used to detect dependence between series, even in the case when the relations between them are complex. Each new test is characterized by its corresponding null hypothesis. In both cases, these tests improve on the existing one in terms of power and usability, which are mostly designed to capture linear relationships. As a consequence of the nonparametric character of both tests, each test can be used with guarantees and practically without restrictions as long as the series are stationary and the sample size is not excessively small. This means that both tests are appropriate for massive data analysis. However, for small data sets (say, below 200 observations), users should be aware that nonlinearity is highly difficult to detect and our advice is to rely on other statistics if linear relations can provide a plausible link among series.

Author Contributions: Conceptualization, M.V.C.-P., M.M.-G., J.M.R. and M.R.M.; methodology, M.V.C.-P., M.M.-G., J.M.R. and M.R.M.; software, M.V.C.-P., M.M.-G., J.M.R. and M.R.M.; writing—original draft preparation, M.V.C.-P., M.M.-G., J.M.R. and M.R.M.

Funding: This research is the result of the activity performed under the program Groups of Excellence of the Region of Murcia, the Fundación Seneca, Science and Technology Agency of the region of Murcia project under grant 19884/GERM/15. All remaining errors are our responsibility.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proofs

This section is devoted to the proof of Theorem 1. To this end, we need to prove the following proposition where $J_{ts} = JS_{ts} - \tilde{\mu}_{ts}$, which is a random variable with zero mean by Equations (6) and (8).

Proposition A1. Let $\{w_t = (x_{1t}, x_{2t}, \dots, x_{kt})\}_{t=1}^T$ be a k -dimensional time series, where $\{x_{jt}\}$ is i.i.d. for all $t = 1, 2, \dots, k$ and they are independent between themselves. Given embedding dimensions $m_1, m_2, \dots, m_k \geq 2$ and $m = \max\{m_1, m_2, \dots, m_k\}$, it follows that

$$\lim_{n \rightarrow \infty} \frac{n(n-1)}{2} \text{Var}(\widehat{JSC}^m - JSC^m) = \sigma_m^2,$$

where

$$JSC^m = \prod_{j=1}^k \frac{1}{m_j!},$$

$$\sigma_m^2 = \sum_{h_1=-m}^m \sum_{h_2=-m}^m E[J_{ts} J_{t+h_1s+h_2}],$$

with $s \geq t + 3m - 2$.

In order to prove Proposition A1, we need the following two technical lemmas.

Lemma A1. Let $\{w_t = (x_{1t}, x_{2t}, \dots, x_{kt})\}_{t=1}^T$ be a k -dimensional time series where $\{x_{jt}\}$ is i.i.d. for all $t = 1, 2, \dots, k$ and they are independent between themselves. Given embedding dimensions $m_1, m_2, \dots, m_k \geq 2$ and $m = \max\{m_1, m_2, \dots, m_k\}$, it follows that

$$E[J_{ts} J_{t's'}] = 0$$

provided that $|t' - t| > m - 1$ or $|s - s'| > m - 1$.

Proof.

$$E[J_{ts} J_{t's'}] = E[(JS_{ts} - \tilde{\mu}_{ts})(JS_{t's'} - \tilde{\mu}_{t's'})],$$

$$E[J_{ts} J_{t's'}] = E \left[\prod_{j=1}^k (I(s(\bar{x}_{jt}), s(\bar{x}_{js})) - \tilde{\mu}_{ts}) \prod_{j=1}^k (I(s(\bar{x}_{jt'}), s(\bar{x}_{js'})) - \tilde{\mu}_{t's'}) \right].$$

As time series are independent between themselves, if we call $I_{ts}^{x_j} = I(s(\bar{x}_{jt}), s(\bar{x}_{js})) - \mu_{ts}^{x_j}$, we have

$$E[J_{ts} J_{t's'}] = \prod_{j=1}^k E \left[\left(I(s(\bar{x}_{jt}), s(\bar{x}_{js})) - \mu_{ts}^{x_j} \right) \left(I(s(\bar{x}_{jt'}), s(\bar{x}_{js'})) - \mu_{t's'}^{x_j} \right) \right] =$$

$$= E \left[\prod_{j=1}^k I_{ts}^{x_j} \cdot I_{t's'}^{x_j} \right].$$

As $m = \max\{m_1, m_2, \dots, m_k\}$, we have that $I_{ts}^{x_j}$ and $I_{t's'}^{x_j}$ are independent, then $E[I_{ts}^{x_j} I_{t's'}^{x_j}] = 0$ ([17], p. 549) provided that $|t' - t| > m - 1$ or $|s - s'| > m - 1$ and $E[J_{ts} J_{t's'}] = 0$ as desired. \square

Lemma A2. Let $\{w_t = (x_{1t}, x_{2t}, \dots, x_{kt})\}_{t=1}^T$ be a k -dimensional time series, where $\{x_{jt}\}$ are i.i.d. for $t = 1, 2, \dots, k$, and are independent between themselves. Let $m = \max\{m_1, m_2, \dots, m_k\}$. Then, for $s \geq t + 3m - 2$, the covariance of the variable J_{ts} with any other $J_{t's'}$ depends only on $h_1 = |t' - t|$ and $h_2 = |s' - s|$ and not on the time periods themselves, that is,

$$\delta(h_1, h_2) = E[J_{ts} J_{t's'}]$$

for all t, s, t', s' .

Proof. Recall that, for $\pi, \delta \in S_{m_j}$, $(p_{tt'}^{\pi\delta})_{\pi\delta} = PM_{x_j}(|t - t'|)$ and $(p_{ss'}^{\pi\delta})_{\pi\delta} = PM_{x_j}(|s - s'|)$ for every $j = 1, 2, \dots, k$. Fix t and s such that $s \geq t + 3m - 2$. Then, it follows that

1. If $h_1 \geq m$ or $h_2 \geq m$, then by Lemma A1, we have that $\delta(h_1, h_2) = 0$.
2. If $h_1 < m$ or $h_2 < m$, we have the following cases:

- If $t' > t$ and $s' > s$

$$\begin{aligned} E [JS_{ts}JS_{t's'}] &= E \left[\prod_{j=1}^k I_j(s(\bar{x}_{jt}), s(\bar{x}_{js})) I_j(s(\bar{x}_{jt'}), s(\bar{x}_{js'})) \right] = \\ &= \prod_{j=1}^k \sum_{\pi \in S_{m_j}} \left(\sum_{\delta \in S_{m_j}} p_{tt'}^{\pi\delta} p_{ss'}^{\pi\delta} \right) \end{aligned} \tag{A1}$$

$$\delta(h_1, h_2) = E [J_{ts}J_{t's'}] = \prod_{j=1}^k \sum_{\pi \in S_{m_j}} \sum_{\delta \in S_{m_j}} p_{tt'}^{\pi\delta} p_{ss'}^{\pi\delta} - \prod_{j=1}^k \left(\frac{1}{m_j!} \right)^2.$$

- If $t' > t$ and $s > s'$

$$\begin{aligned} E [JS_{ts}JS_{t's'}] &= E \left[\prod_{j=1}^k I_j(s(\bar{x}_{jt}), s(\bar{x}_{js})) I_j(s(\bar{x}_{jt'}), s(\bar{x}_{js'})) \right] = \\ &= \prod_{j=1}^k \sum_{\pi \in S_{m_j}} \left(\sum_{\delta \in S_{m_j}} p_{tt'}^{\pi\delta} p_{s's}^{\pi\delta} \right) \end{aligned} \tag{A2}$$

$$\delta(h_1, h_2) = E [J_{ts}J_{t's'}] = \prod_{j=1}^k \sum_{\pi \in S_{m_j}} \sum_{\delta} p_{tt'}^{\pi\delta} p_{s's}^{\pi\delta} - \prod_{j=1}^k \left(\frac{1}{m_j!} \right)^2.$$

- If $t' = t$ and $s' \neq s$

$$\begin{aligned} E [JS_{ts}JS_{ts'}] &= E \left[\prod_{j=1}^k I_j(s(\bar{x}_{jt}), s(\bar{x}_{js})) I_j(s(\bar{x}_{jt}), s(\bar{x}_{js'})) \right] = \\ &= \prod_{j=1}^k \left(\sum_{\pi \in S_{m_j}} p_t^{\pi} p_{ss'}^{\pi\pi} \right) \end{aligned} \tag{A3}$$

$$\delta(h_1, h_2) = E [J_{ts}J_{ts'}] = \prod_{j=1}^k \left(\sum_{\pi \in S_{m_j}} p_t^{\pi} p_{ss'}^{\pi\pi} \right) - \prod_{j=1}^k \left(\frac{1}{m_j!} \right)^2.$$

The remaining cases $t' < t, s < s', t' < t, s' < s, t \neq t'$, and $s = s'$ are symmetric to the previous ones. Since the probabilities used in Equations (A1)–(A3) depend only on h_1 and h_2 , the proof follows. \square

Proof of Proposition A1. We consider

$$\begin{aligned} \frac{n(n-1)}{2} \text{Var}(\widehat{JSC}^m - JSC^m) &= \frac{2}{n(n-1)} \text{Var} \left(\sum_{t=1}^{n-1} \sum_{s=t+1}^n J_{ts} \right) \\ &= \frac{2}{n(n-1)} \sum_{t=1}^{n-1} \sum_{t'=1}^{n-1} \sum_{s=t+1}^n \sum_{s'=t'+1}^n E [J_{ts}J_{t's'}]. \end{aligned} \tag{A4}$$

From Lemma A1, the variance Equation (A4) can be rewritten as

$$\begin{aligned} \frac{2}{n(n-1)} \sum_{t=1}^{n-1} \sum_{t'=1}^{n-1} \sum_{s=t+1}^n \sum_{s'=t'+1}^n E[J_{ts}J_{t's'}] &= \frac{2}{n(n-1)} \sum_{t=1}^{n-1} \sum_{t' \in N(t)} \sum_{s \in N(t)} \sum_{s' \in N(s)} E[J_{ts}J_{t's'}] \quad (A5) \\ &+ \frac{2}{n(n-1)} \sum_{t=1}^{n-1} \sum_{t' \in N(t)} \sum_{s \notin N(t)} \sum_{s' \in N(s)} E[J_{ts}J_{t's'}]. \end{aligned}$$

Since $N(t)$ and $N(s)$ are finite sets, the first summand in Equation (A5) converges to zero as n approaches infinity. The second summand in Equation (A5) can be written as

$$\begin{aligned} \frac{2}{n(n-1)} \sum_{t=1}^{n-1} \sum_{t' \in N(t)} \sum_{s \notin N(t)} \sum_{s' \in N(s)} E[J_{ts}J_{t's'}] &= \frac{2}{n(n-1)} \sum_{t=1}^{n-1} \sum_{t' \in N(t)} \sum_{s \notin N(t)} \sum_{s' \in N(s) \cap N(t')} E[J_{ts}J_{t's'}] \quad (A6) \\ &+ \frac{2}{n(n-1)} \sum_{t=1}^{n-1} \sum_{t' \in N(t)} \sum_{s \notin N(t)} \sum_{s' \in N(s) \cap \overline{N(t')}} E[J_{ts}J_{t's'}], \end{aligned}$$

where we denote by $\overline{N(t)}$ the complementary set of $N(t)$.

In the first summand in Equation (A6), since $t' \in N(t)$ and $s \notin N(t)$, we have that, for a fixed t , the number of time indexes s that make $s' \in N(s) \cap N(t')$ possible is finite, and therefore this summand converges to zero as n goes to infinity.

Notice that, for fixed t and s with $s \geq t + 2m$, we have that $E[J_{ts}J_{t+h_1s+h_2}] = E[J_{t's'}J_{t'+h_1s'+h_2}]$ always and $s' \notin N(t')$. Furthermore, from Lemma A1, it follows that

$$\sum_{t'} \sum_{s' \geq t'+m} E[J_{ts}J_{t's'}] = \sum_{t'=t-m+1}^{t+m-1} \sum_{s'=s-m+1}^{s+m-1} E[J_{ts}J_{t's'}] = \sigma_m^2. \quad (A7)$$

Then, from Equation (A7), we have that

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{n(n-1)}{2} \text{Var}(\widehat{JSC}^m - JSC^m) &= \lim_{n \rightarrow \infty} \frac{2}{n(n-1)} \sum_{t=1}^{n-1} \sum_{t' \in N(t)} \sum_{s \notin N(t)} \sum_{s' \in N(s) \cap \overline{N(t')}} E[J_{ts}J_{t's'}] \\ &= \lim_{n \rightarrow \infty} \frac{2}{n(n-1)} \sum_{t=1}^{n-1} \sum_{s=t+2m}^n \sum_{t'=t-m+1}^{t+m-1} \sum_{s'=s-m+1}^{s+m-1} E[J_{ts}J_{t's'}] \\ &= \lim_{n \rightarrow \infty} \frac{2}{n(n-1)} \sum_{t=1}^{n-1} \sum_{s=t+2m}^n \sigma_m^2. \end{aligned}$$

Since the number of summands in $\sum_{t=1}^{n-1} \sum_{s=t+2m}^n$ is $\frac{1}{2}(-2 - 4m(-1 + n) + n + n^2)$, then

$$\lim_{n \rightarrow \infty} \frac{n(n-1)}{2} \text{Var}(\widehat{JSC}^m - \prod_{j=1}^k \frac{1}{m_j!}) = \lim_{n \rightarrow \infty} \frac{2}{n(n-1)} \sum_{t=1}^{n-1} \sum_{s=t+2m}^n \sigma_m^2 = \sigma_m^2 \quad (A8)$$

as desired. \square

Proof of Theorem 1. For each integer $k > 4(m - 1)$, we define

$$Y_n^k = \sqrt{\frac{2}{n(n-1)}} \left(\sum_{j_1=1}^r \sum_{j_2=j_1+1}^r S_{j_1j_2} + \sum_{j=1}^r T_j \right),$$

where

$$\begin{aligned} S_{j_1j_2} &= \sum_{t=(j_1-1)k+1}^{j_1k-2m+2} \sum_{s=(j_2-1)k+1}^{j_2k-2m+2} J_{ts}, \\ T_j &= \sum_{t=(j-1)k+1}^{jk-2m+1} \sum_{s=t+1}^{jk-2m+2} J_{ts}, \end{aligned}$$

with $r = [(n - 1)/k]$ the integer part of $(n - 1)/k$.

It is interesting to note that the variables $S_{j_1 j_2}$ are i.i.d. of mean zero and variance

$$v = \sum_{t=1}^{k-2m+2} \sum_{s=k+1}^{2k-2m+2} \sum_{t'=1}^{k-2m+2} \sum_{s'=k+1}^{2k-2m+2} E[J_{ts} J_{t's'}]. \tag{A9}$$

Then, from Lemma A2, it follows that

$$v = \sum_{|h_1| < m} \sum_{|h_2| < m} (k - 2m + 2 - |h_1|)(k - 2m + 2 - |h_2|) E[J_{ts} J_{t+h_1 s+h_2}]. \tag{A10}$$

Therefore, from the Central Limit Theorem for i.i.d random variables, we have that

$$\sqrt{\left(\frac{2}{n(n-1)}\right)} \left(\sum_{j_1=1}^r \sum_{j_2=j_1+1}^r S_{j_1 j_2}\right)$$

is asymptotically $\mathcal{N}\left(0, \frac{v}{k^2}\right)$ as $n \rightarrow \infty$.

On the other hand, the variables T_j are i.i.d with zero mean, independent of $\sum_{j_1=1}^r \sum_{j_2=j_1+1}^r S_{j_1 j_2}$ and $\lim_{n \rightarrow \infty} \text{Var}\left(\sqrt{\frac{2}{n(n-1)}} \sum_{j=1}^r T_j\right) = 0$. Moreover,

$$\lim_{k \rightarrow \infty} \frac{v}{k^2} = \sigma_m^2.$$

Then,

$$\lim_{k \rightarrow \infty} \lim_{n \rightarrow \infty} Y_n^k = Y, \tag{A11}$$

where $Y \sim \mathcal{N}(0, \sigma_m^2)$.

It only remains to show that

$$\lim_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} P\left(\left|\sqrt{\frac{n(n-1)}{2}} \left(\widehat{JSC}^m - E[\widehat{JSC}^m]\right) - Y_n^k\right| > \varepsilon\right) = 0 \tag{A12}$$

for all $\varepsilon > 0$, since Equation (11) will then follow from Proposition 6.3.9 in Brockwell and Davis (1987).

To establish Equation (A12), we write

$$\left(\sqrt{\frac{n(n-1)}{2}} \left(\widehat{JSC}^m - E[\widehat{JSC}^m]\right) - Y_n^k\right) = \sqrt{\frac{2}{n(n-1)}} \left(US_n^k + U_n^k + UT_n^k\right),$$

where

$$\begin{aligned} US_n^k &= \sum_{j_1=1}^{r-1} \sum_{t=(j_1-1)k+1}^{j_1 k-2m+2} \sum_{j_2=j_1+1}^{r-1} \sum_{s=j_2 k-2m+3}^{j_2 k} J_{ts} + \sum_{j_1=1}^{r-1} \sum_{t=j_1 k-2m+3}^{j_1 k} \sum_{s=j_1 k+1}^{rk-2m+2} J_{ts}, \\ U_n^k &= \sum_{t=1}^{rk-2m+2} \sum_{s=rk-2m+3}^n J_{ts} + \sum_{t=rk-2m+3}^{n-1} \sum_{s=t+1}^n J_{ts}, \\ UT_n^k &= \sum_{j_1=1}^{r-1} \sum_{t=(j_1-1)k+1}^{j_1 k-2m+2} \sum_{s=j_1 k-2m+3}^{j_1 k} J_{ts} + \sum_{j_1=1}^{r-1} \sum_{t=j_1 k-2m+3}^{j_1 k-1} \sum_{s=t+1}^{j_1 k} J_{ts}. \end{aligned}$$

Notice that, to prove Equation (A12), it is sufficient to prove the following equation by Chebyshev inequality:

$$\lim_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} \text{Var} \left(\sqrt{\frac{n(n-1)}{2}} \left(\widehat{JSC}^m - E[\widehat{JSC}^m] \right) - Y_n^k \right) = 0. \tag{A13}$$

Taking into account that $\text{Var}(J_{ts}) \leq 1$ for all t, s that $|\text{Cov}(X, Y)| \leq \sqrt{\text{Var}(X)\text{Var}(Y)}$ and Lemma A1, we have that

$$\text{Var} \left(\sum_{t=\alpha_1}^{\alpha_2} \sum_{s=\beta_1}^{\beta_2} J_{ts} \right) \leq (\alpha_2 - \alpha_1 + 1)(\beta_2 - \beta_1 + 1)(1 + (2m - 1)^2), \tag{A14}$$

$$\begin{aligned} \text{Var} \left(\sqrt{\frac{2}{n(n-1)}} US_n^k \right) &\leq \frac{2}{n(n-1)} \left(\frac{(r-1)(r-2)}{2} (k-2m+2)(2m-2)(1+(2m-1)^2) \right. \\ &+ \frac{(r-1)((r-1)k-2m+2)(1+(2m-1)^2)}{2} \\ &\left. + 2(r-1)(1+(2m-1)^2) \sqrt{\frac{r-2}{2} (k-2m+2)(2m-2)((r-1)k-2m+2)} \right), \end{aligned}$$

$$\begin{aligned} \text{Var} \left(\sqrt{\frac{2}{n(n-1)}} UT_n^k \right) &\leq \frac{2}{n(n-1)} \left((rk-2m+2)m(1+(2m-1)^2) + \right. \\ &+ m^2(1+(2m-1)^2) + \\ &\left. + 2m(1+(2m-1)^2) \sqrt{(rk-2m+2)m} \right), \end{aligned} \tag{A15}$$

$$\begin{aligned} \text{Var} \left(\sqrt{\frac{2}{n(n-1)}} UT_n^k \right) &\leq \frac{2}{n(n-1)} \left((r-1)(k-2m+2)(2m-2)(1+(2m-1)^2) + \right. \\ &\frac{(r-1)(2m-3)^2(1+(2m-1)^2)}{2} + \\ &\left. + 2(r-1)(2m-3)(1+(2m-1)^2) \sqrt{(k-2m+2)(2m-2)} \right). \end{aligned}$$

Then, it follows that

$$\begin{aligned} \lim_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} \text{Var} \left(\sqrt{\frac{2}{n(n-1)}} US_n^k \right) &= 0, \\ \lim_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} \text{Var} \left(\sqrt{\frac{2}{n(n-1)}} UT_n^k \right) &= 0, \\ \lim_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} \text{Var} \left(\sqrt{\frac{2}{n(n-1)}} UT_n^k \right) &= 0. \end{aligned}$$

Hence,

$$\lim_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} \text{Var} \left(\sqrt{\frac{n(n-1)}{2}} \left(\widehat{JSC}^m - E[\widehat{JSC}^m] \right) - Y_n^k \right) = 0.$$

From Chebyshev’s inequality, we have that

$$\lim_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} P \left(\left| \sqrt{\frac{n(n-1)}{2}} \left(\widehat{JSC}^m - E[\widehat{JSC}^m] \right) - Y_n^k \right| > \varepsilon \right) = 0, \tag{A16}$$

which finishes the proof of the theorem. □

Proof of Theorem 2. Let $\hat{\theta}_j$ be a consistently estimated parameter of θ_j and $m_j \geq 2$ for all $j = 1, 2, \dots, k$ and let $m = \max\{m_1, m_2, \dots, m_k\}$. Now, we want to prove that, for the multidimensional time series given by $\{(u_{1t}, u_{2t}, \dots, u_{kt})\}_t$, it follows that

$$\sqrt{\frac{n(n-1)}{2}} \left(\widehat{JSC}^m(\theta_1, \theta_2, \dots, \theta_k) - \widehat{JSC}^m(\hat{\theta}_1, \hat{\theta}_1, \dots, \hat{\theta}_k) \right) \xrightarrow{P} 0.$$

For all $j = 1, 2, \dots, k$, the residual function is given by

$$u_{jt}(\hat{\theta}_j) = x_{jt} - G_j(I_{t-1}^j, \hat{\theta}_j) = u_{jt}(\theta_j) + G_j(I_{t-1}^j, \theta_j) - G_j(I_{t-1}^j, \hat{\theta}_j). \quad (\text{A17})$$

Then,

$$|u_{jt}(\theta_j) - u_{jt}(\hat{\theta}_j)| = |G_j(I_{t-1}^j, \theta_j) - G_j(I_{t-1}^j, \hat{\theta}_j)|. \quad (\text{A18})$$

The $m_j!$ symbols provide a partition of \mathbb{R}^{m_j} in $m_j!$ subsets $\{A_{\pi_1}^j, A_{\pi_2}^j, \dots, A_{\pi_{m_j!}}^j\}$, such that the probability $P(\bar{u}_{jt}(\theta_j) \in A_{\pi_i}^j \cap A_{\pi_l}^j) = 0$ for all $i \neq l$ if the values of u_{jt} come from a continuous distribution because then equal values are very uncommon, with a theoretical probability of occurrence of 0. Therefore, for a given m_j history $\bar{u}_{jt}(\theta_j)$, we may assume without loss of generality that it belongs to a ball of radius $\varepsilon_j > 0$ satisfying $B(\bar{u}_{jt}(\theta_j), \varepsilon_j) \subset A_{\pi}^j$ for some $\pi \in S_{m_j}$.

For each ε_j and given that G_j is a continuous map in a compact set (and hence uniformly continuous) for all $j = 1, 2, \dots, k$, there exists a $\delta_j > 0$ independent of I_{t-1}^j, θ_j and $\hat{\theta}_j$ such that, if $\|(I_{t-1}^j, \theta_j) - (I_{t-1}^j, \hat{\theta}_j)\| < \delta_j$, then by Equation (A18) $|u_{jt}(\theta_j) - u_{jt}(\hat{\theta}_j)| < \varepsilon_j$ and hence $\bar{u}_{jt}(\hat{\theta}_j) \in A_{\pi}^j$.

Thus, it follows that $I(s_{x_j}(\bar{u}_{jt}(\theta_j)), s_{x_j}(\bar{u}_{js}(\theta_j))) = I(s_{x_j}(\bar{u}_{jt}(\hat{\theta}_j)), s_{x_j}(\bar{u}_{js}(\hat{\theta}_j)))$ for almost all t, s , and hence

$$JS_{ts}(\theta_1, \theta_2, \dots, \theta_k) = JS_{ts}(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k)$$

and therefore

$$\lim_{n \rightarrow \infty} P \left(\sqrt{\frac{n(n-1)}{2}} \left| \widehat{JSC}^m(\theta_1, \theta_2, \dots, \theta_k) - \widehat{JSC}^m(\hat{\theta}_1, \hat{\theta}_1, \dots, \hat{\theta}_k) \right| > 0 \right) = 0,$$

which finishes the proof of the theorem. \square

Remark A1. Notice that we have used only the fact that the map G_j with $j = 1, 2, \dots, k$, are uniformly continuous. Therefore, we could state only this condition in Theorem 2 and avoid the compact domain of every G_j .

References

1. Haugh, L.D. Checking the independence of two covariance-stationary time series: A univariate residual cross-correlation approach. *J. Am. Stat. Assoc.* **1976**, *71*, 378–385. [[CrossRef](#)]
2. Pierce, A. Lack of dependence among economic variables. *J. Am. Stat. Assoc.* **1977**, *72*, 11–22.
3. Geweke, J. A comparison of tests of independence of two covariance stationary time series. *J. Am. Stat. Assoc.* **1981**, *76*, 363–373. [[CrossRef](#)]
4. Bouhaddioui, C.; Dufour, J.-M. Tests for non-correlation of two infinite-order cointegrated vector autoregressive series. *J. Appl. Probab. Stat.* **2008**, *3*, 78–94.
5. Bouhaddioui, C.; Roy, R. A generalized portmanteau test for independence of two infinite-order vector autoregressive series. *J. Time Ser. Anal.* **2006**, *27*, 505–544. [[CrossRef](#)]
6. Duchesne, P.; Roy, R. Robust tests for independence of two time series. *Stat. Sin.* **2003**, *13*, 827–852.
7. El Himdi, K.; Roy, R. Tests for noncorrelation of two multivariate ARMA time series. *Can. J. Stat.* **1997**, *25*, 233–256. [[CrossRef](#)]
8. Hong, Y. Testing for independence between two covariance stationary time series. *Biometrika* **1996**, *83*, 615–625. [[CrossRef](#)]

9. Koch, P.D.; Yang, S.S. A method for testing the independence of two time series that accounts for a potential pattern in the crosscorrelation function. *J. Am. Stat. Assoc.* **1986**, *81*, 533–544. [[CrossRef](#)]
10. Li, W.K.; Hui, Y.V. Robust residual cross correlation tests for lagged relations in time series. *J. Stat. Comput. Simul.* **1994**, *49*, 103–109. [[CrossRef](#)]
11. Pham, D.T.; Roch, R.; Cedras, L. Tests for non-correlation of two cointegrated ARMA time series. *J. Time Ser. Anal.* **2003**, *24*, 553–577. [[CrossRef](#)]
12. Muhammad K.W.; Islam K.A. Most stringent test of independence for time series. *Commun. Stat. Simul. Comput.* **2018**. [[CrossRef](#)]
13. Chan, N.H.; Tran, L.T. Nonparametric tests for serial dependence. *J. Time Ser. Anal.* **1992**, *13*, 19–28. [[CrossRef](#)]
14. Matilla-Garcia, M.; Rodriguez, J.M.; Ruiz, M. A symbolic test for testing independence between time series. *J. Time Ser. Anal.* **2010**, *31*, 76–85. [[CrossRef](#)]
15. Robinson, P.M. Consistent nonparametric entropy-based testing. *Rev. Econ. Stud.* **1991**, *58*, 437–453. [[CrossRef](#)]
16. Skaug, H.J.; Tjøstheim, D. A nonparametric test of serial independence based on the empirical distribution function. *Biometrika* **1993**, *80*, 591–602. [[CrossRef](#)]
17. Caballero-Pintado, M.V.; Matilla-Garcia, M.; Marin, M.R. Symbolic correlation integral. *Econom. Rev.* **2019**, *38*, 533–556. [[CrossRef](#)]
18. Granger, C.W.; Maasoumi, E.; Racine, J. A dependence metric for possibly nonlinear processes. *J. Time Ser. Anal.* **2004**, *25*, 649–669. [[CrossRef](#)]
19. Agresti, A.; Agresti, B.F. Statistical Analysis of Qualitative Variation. *Sociol. Methodol.* **1978**, *9*, 204. [[CrossRef](#)]
20. Aaronson, J.; Burton, R.; Dehling, H.; Gilat, D.; Hill, T.; Weiss, B. Strong laws for L - and u -statistics. *Trans. Am. Math. Soc.* **1996**, *348*, 2845–2866. [[CrossRef](#)]
21. Politis, D.N.; White, H. Automatic block-length selection for the dependent bootstrap. *Econom. Rev.* **2004**, *23*, 53–70. [[CrossRef](#)]
22. Patton, A.J.; Politis, D.N.; White, H. Correction: automatic block-length selection for the dependent bootstrap. *Econom. Rev.* **2009**, *28*, 372–375. [[CrossRef](#)]
23. Hallin, M.; Saidi, A. Testing noncorrelation and noncausality between multivariate ARMA time series. *J. Time Ser. Anal.* **2005**, *26*, 83–105. [[CrossRef](#)]
24. Matilla-Garcia, M.; Ruiz, M. A non-parametric independence test using permutation entropy. *J. Econom.* **2008**, *144*, 139–155. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).