Original Article

Journal of
Time Series Analysis

Final version received June 2008

Published online in Wiley InterScience: 14 January 2010

(www.interscience.wiley.com) DOI: 10.1111/j.1467-9892.2009.00645.x

A symbolic test for testing independence between time series

Mariano Matilla-García[a,*,†], José Miguel Rodríguez[b] and Manuel Ruiz Marín[b]

In this article we introduce a test for independence between two processes $\{X_t\}$ and $\{Y_t\}$. To this end we rely on symbolic dynamics and permutation entropy as a measure of dependence. As a result, a nonparametric (model-free) test for either linear or nonlinear processes is presented. The test is consistent for a broad range of dependent alternatives. Empirical simulations indicate and highlight the general utility of the test for time-series analysts.

Keywords: Independence test; symbolic analysis; entropy.

1. INTRODUCTION

The identification of a potential dynamic relationship between two time series is a relevant problem in science and it is particularly crucial in economics. The fact that several economic phenomena can be described by time series indicates why it is relevant and usual to test the existence of a relationship, mainly dependence between series, and then to investigate the dynamic nature of it. Elucidation of dependence (and, if possible, causal) relationships between time series is central to conduct forecasts.

In contrast to the general relevance of analysing serial dependence within bivariate time series, relatively few attempts have been made to test this kind of dependence between two time series. Haugh (1976) proposed a test for independence 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. Pierce (1977) found that when Haugh's test is used, the hypothesis of independence cannot be rejected for many monetary aggregates and measures of economic activity series whose relationship seemed to be well established. In this regard, Geweke (1981) found that Haugh's test often has low power, which hints, from an econometric perspective, that this low power might explain the influential results of Pierce (1977).

Haugh's procedure has been extended by Koch and Yang (1986) by introducing a modification that allows for a potential pattern among successive cross-correlation coefficients. In an influential article, Hong (1996) proposed a consistent test based on kernel estimation techniques which generalizes Haugh's statistic. Other extensions of Haugh's seminal work are available, see for instance Pham *et al.* (2003) and references therein. Common to all these articles is that a linear, basically autoregressive, representation of $\{X_t\}$ and $\{Y_t\}$ is required to the development of the tests. In general, these procedures follow the main skeleton given by Haugh, namely first obtaining white noise residuals by performing univariate autoregression, and then checking the sample cross-correlation of the residuals.

In this article, to avoid imposing restrictive parametric assumptions such as linearity and normality, we take a different way. To test dependence between $\{X_t\}$ and $\{Y_t\}$ we rely on symbolic dynamics and permutation entropy as a measure of dependence. In the last two decades, the use of entropy has played a leading role as a measure of the dependence present in a time series (see, for example, Joe 1989a,b who considered a smoothed nonparametric entropy measure of multi-variate dependence of an independent and identically distributed random vector). On the other hand, symbolic dynamics has been used for the investigation of nonlinear dynamical systems (for an overview, see Hao and Zheng, 1998) which has the important advantage, namely the essential features of the dynamics between the two series may be preserved. The absence of dependencies between the two unknown underlying data generating process is studied via symbolic dynamics. Recently, Matilla-García and Ruiz (2008), by means on the concept of permutation entropy, have proposed a powerful nonparametric test for serial independence in an univariate context.

Particularly in this article, given two time processes, we study the potential dependence between them by translating the problem (and the information set) into symbols by means of an easy symbolization technique and, then, we use the entropy measure associated with these symbols to construct a test for independence between the two series. As a result, we obtain a novel nonparametric test for cross-independence which is consistent for a broad range of dependent alternatives. The test is model free and does not require restrictive assumptions; even more, it avoids estimation of the true linear representation of the data set that, if misspecified or not consistently estimated, might invalidate asymptotic results of Haugh-type tests. In addition, we show that the test can be easily transformed into a multi-variate consistent test of independence.

[a] *UNED*
[b] *Universidad Politécnica, de Cartagena*
**Correspondence to: Mariano Matilla-García, Paseo Senda del Rey, 11, CP, 28040 Madrid, Spain.*
[†] *E-mail: mmatilla@cee.uned.es*

The rest of the article is structured as follows. In Section 2, we introduce the notation and several definitions to describe the symbolic dynamic representation methodology. The procedure is illustrated with a simple example. In Section 3 we give the construction of the bivariate independence test via permutation entropy. We prove that under the null of independence between two time series, the entropy-based test is asymptotically chi-square distributed. We also prove the consistency property of the test. This section includes some discussion about the selection of parameter m and we generalize these results to detect dependencies among n-dependent variables (multi-variate case). Finally, we introduce a straightforward modification of the test that allows dealing with any cross-lag-dependent models. Empirical size and power of the new test are studied by Monte Carlo methods in Section 4. As proofs for the multi-variate case follow basically the same steps as the bivariate one, they are not included in the article, but they are available upon request.

2. DEFINITIONS AND NOTATION

We give some definitions and we introduce the basic notation which are natural extensions of the formulation presented in Matilla-García and Ruiz (2008), where the basic concepts for symbolic analysis were established in terms of ordinal patterns. Notation and definitions are illustrated by means of a naive example.

Let $\{\mathbf{W}_t\}_{t\in I}$ be a two-dimensional time series with $\mathbf{W}_t = (X_t,Y_t)$ and where I is a time index set. For a positive integer $m \geq 2$ we denote by $S_m = \{\pi_1,\pi_2,\ldots,\pi_{m!}\}$ the symmetric group of order $m!$, that is the group formed by all the permutations of length m. Let $\pi = (i_1,i_2,\ldots,i_m) \in S_m$. We will call an element π in the symmetric group S_m a symbol.

Consider the direct product of the symmetric group $\Gamma_m = S_m \times S_m$. We now define an ordinal pattern for a symbol $\eta_{ij} = (\pi_i^x, \pi_j^y) \in \Gamma_m$, at a given time $t \in I$. To this end we consider that the time series is embedded in a $2m$-dimensional space as follows:

$$\mathbf{W}_m(t) = (X_{t+0}, X_{t+1}, \ldots, X_{t+m-1}, Y_{t+0}, Y_{t+1}, \ldots, Y_{t+m-1})$$

for $t \in I$.

We will call m the *embedding dimension*. Let $\pi_i^x = (i_1, i_2, \ldots, i_m) \in S_m$ be a symbol. Then, we say that t is of π_i^x type for X if and only if $\pi_i^x = (i_1, i_2, \ldots, i_m)$ is the unique symbol in the group S_m satisfying the two following conditions:

(a) $X_{t+i_1} \leq X_{t+i_2} \leq \cdots \leq X_{t+i_m}$ and

(b) $i_{s-1} < i_s$, if $X_{t+i_{s-1}} = X_{t+i_s}$

Condition (b) guarantees uniqueness of the symbol η_{ij}. This is justified if the values of X_t have a continuous distribution so that equal values are very uncommon, with a theoretical probability of occurrence of 0. Similarly, for a symbol $\pi_j^y \in S_m$ we define that t is of π_j^y type for Y.

Now, let $\eta_{ij} = (\pi_i^x, \pi_j^y) \in \Gamma_m$. We will say that t is of η_{ij} type for $\mathbf{W} = (X,Y)$ or simply that t is of η_{ij} type, if and only if t is of π_i^x type for X and of π_j^y type for Y. Notice that for all t such that t is of η_{ij} type for $\mathbf{W}$ the $2m$-history $\mathbf{W}_m(t)$ is converted into a unique symbol η_{ij}. To see this, the following example will help the reader.

Take as embedding dimension $m = 3$. Thus, the symmetric group is

$$S_3 = \{(0,1,2), (0,2,1), (1,0,2), (1,2,0), (2,0,1), (2,1,0)\}.$$

Consider the finite time series of seven values

$$\{(X_1,Y_1) = (8,2), (X_2,Y_2) = (7,9), (X_3,Y_3) = (6,5), (X_4,Y_4) = (9,7), (X_5,Y_5) = (4,3), (X_6,Y_6) = (10,12), (X_7,Y_7) = (2,5)\}.$$

Then for $t = 2$ we have that $X_{t+1} = 6 < X_{t+0} = 7 < X_{t+2} = 9$ and $Y_{t+1} = 5 < Y_{t+2} = 7 < Y_{t+0} = 9$; therefore, we have that the period $t = 2$ is of ((1,0,2),(1,2,0)) type.

In addition, given a time series $\{\mathbf{W}_t\}_{t\in I}$ and an embedding dimension m, the relative frequency of a symbol $\pi_i^x \in S_m$ and $\pi_j^y \in S_m$ can be computed by:

$$p(\pi_i^x) := p_{\pi_i^x} = \frac{\sharp\{t \in I \,|\, t \text{ is of } \pi_i^x \text{ type for } X\}}{T - m + 1} = \frac{n_{\pi_i^x}}{T - m + 1} \tag{1}$$

$$p(\pi_j^y) := p_{\pi_j^y} = \frac{\sharp\{t \in I \,|\, t \text{ is of } \pi_j^y \text{ type for } Y\}}{T - m + 1} = \frac{n_{\pi_j^y}}{T - m + 1} \tag{2}$$

where T is the cardinality of set I. Similarly, for $\eta_{ij} \in \Gamma_m$ we obtain its relative frequency to be

$$p(\eta_{ij}) := p_{\eta_{ij}} = \frac{\sharp\{t \in I \,|\, t \text{ is of } \eta_{ij} \text{ type}\}}{T - m + 1} = \frac{n_{\eta_{ij}}}{T - m + 1} \tag{3}$$

Now, under this setting we can define the *permutation entropy* of a two-dimensional time series $\{\mathbf{W}_t\}_{t\in I}$ for an embedding dimension $m \geq 2$. This entropy is defined as the Shannon's entropy of the $m!^2$ distinct symbols as follows:

$$h_{\mathbf{W}}(m) = -\sum_{\eta\in\Gamma_m} p_\eta \ln(p_\eta). \tag{4}$$

Similarly, the marginal permutation entropy can be defined as

$$h_X(m) = -\sum_{\pi_i^x \in S_m} p_{\pi_i^x} \ln(p_{\pi_i^x}) \quad \text{and} \quad h_Y(m) = -\sum_{\pi_j^y \in S_m} p_{\pi_j^y} \ln(p_{\pi_j^y})$$

for $\{X_t\}$ and $\{Y_t\}$ respectively. Intuitively, permutation entropy can be thought of as the symbolic information contained in a time series. It is important to note that such information is provided by means of the order structure induced by the symbols utilized in the time series' symbolization procedure.

3. CONSTRUCTION AND PROPERTIES OF THE INDEPENDENCE TEST

We construct an independence test between two time series with all the machinery as defined in the previous section. We also extend the main theorems to a multi-variate independence test among series and some guidance to choose the free parameter is given. Finally, this section indicates how to deal with high-order cross-lag-dependent processes for a given m.

Let $\{\mathbf{W}_t\}_{t \in I}$ be a two-dimensional time series and m be a fixed embedding dimension. To test for independence between $\{X_t\}$ and $\{Y_t\}$, which is the aim of this article, we consider the following null hypothesis:

$$H_0 : \{X_t\}_{t \in I} \quad \text{and} \quad \{Y_t\}_{t \in I} \text{ are independent} \tag{5}$$

against any other alternative. We now provide a proposition whose proof can be found in the Appendix.

Proposition 1. *Let $\{\mathbf{W}_t = (X_t, Y_t)\}_{t \in I}$ be a two-dimensional time series and $m \geq 2$ with $m \in \mathbb{N}$ a fixed embedding dimension. Let $K = T - m + 1$. Then under the null hypothesis of independence between series, the statistic*

$$\Lambda(m) := -2\left[K \ln(K) + \sum_{i=1}^{m!}\sum_{j=1}^{m!} n_{\eta_{ij}} \ln\left(\frac{p_{\eta_{ij}}^{(0)}}{n_{\eta_{ij}}}\right)\right]$$

is asymptotically $\chi^2_{(m!-1)^2}$ distributed, where $p_{\eta_{ij}}^{(0)}$ refers to the probability of each symbol under the null.

Importantly, notice that $p_{\eta_{ij}}^{(0)}$ is unknown and hence $\Lambda(m)$ is not useful in terms of Proposition 1. Fortunately, under the H_0 it has to be satisfied that

$$p_{\eta_{ij}}^{(0)} = p_{\pi_i^x} p_{\pi_j^y} \tag{6}$$

In addition, note that

$$\sum_{i=1}^{m!}\sum_{j=1}^{m!} \frac{n_{\eta_{ij}}}{K} = 1$$

and that

$$p_{\pi_i^x} p_{\pi_j^y} = \frac{n_{\pi_i^x}}{K}\frac{n_{\pi_j^y}}{K},$$

and thus it follows from Proposition 1 that

$$\begin{aligned}
\hat{\Lambda}(m) &= -2K\left[\ln(K) + \sum_{i=1}^{m!}\sum_{j=1}^{m!} \frac{n_{\eta_{ij}}}{K} \ln\left(\frac{p_{\eta_{ij}}^{(0)}}{n_{\eta_{ij}}}\right)\right] \\
&= -2K\left[\ln(K) + \sum_{i=1}^{m!}\sum_{j=1}^{m!} \frac{n_{\eta_{ij}}}{K} \ln\left(\frac{\frac{n_{\pi_i^x}}{K}\frac{n_{\pi_j^y}}{K}}{n_{\eta_{ij}}}\right)\right] \\
&= -2K\left[\ln(K) + \sum_{i=1}^{m!}\sum_{j=1}^{m!} \frac{n_{\eta_{ij}}}{K}\left(\ln\left(\frac{n_{\pi_i^x}}{K}\right) + \ln\left(\frac{n_{\pi_j^y}}{K}\right)\right) - \sum_{i=1}^{m!}\sum_{j=1}^{m!} \frac{n_{\eta_{ij}}}{K}\ln\left(\frac{n_{\eta_{ij}}}{K}\right) - \sum_{i=1}^{m!}\sum_{j=1}^{m!} \frac{n_{\eta_{ij}}}{K}\ln(K)\right] \\
&= -2K\left[\sum_{i=1}^{m!}\sum_{j=1}^{m!} \frac{n_{\eta_{ij}}}{K}\left(\ln\left(\frac{n_{\pi_i^x}}{K}\right) + \ln\left(\frac{n_{\pi_j^y}}{K}\right)\right) - \sum_{i=1}^{m!}\sum_{j=1}^{m!} \frac{n_{\eta_{ij}}}{K}\ln\left(\frac{n_{\eta_{ij}}}{K}\right)\right] \\
&= -2K\left[\sum_{i=1}^{m!}\sum_{j=1}^{m!} \frac{n_{\pi_i^x}}{K}\ln\left(\frac{n_{\pi_i^x}}{K}\right)\frac{n_{\pi_j^y}}{K} + \sum_{i=1}^{m!}\sum_{j=1}^{m!} \frac{n_{\pi_j^y}}{K}\ln\left(\frac{n_{\pi_j^y}}{K}\right)\frac{n_{\pi_i^x}}{K} - \sum_{i=1}^{m!}\sum_{j=1}^{m!} \frac{n_{\eta_{ij}}}{K}\ln\left(\frac{n_{\eta_{ij}}}{K}\right)\right]
\end{aligned} \tag{7}$$

Notice also that

$$\sum_{i=1}^{m!} \frac{n_{\pi_i^x}}{K} = \sum_{i=j}^{m!} \frac{n_{\pi_j^y}}{K} = 1.$$

Then, the statistic $\hat{\Lambda}(m)$ remains as:

$$\hat{\Lambda}(m) = -2K\left[\frac{n_{\pi_1^x}}{K}\ln\left(\frac{n_{\pi_1^x}}{K}\right)\sum_{j=1}^{m!}\frac{n_{\pi_j^y}}{K} + \frac{n_{\pi_2^x}}{K}\ln\left(\frac{n_{\pi_2^x}}{K}\right)\sum_{j=1}^{m!}\frac{n_{\pi_j^y}}{K} + \cdots + \frac{n_{\pi_{m!}^x}}{K}\ln\left(\frac{n_{\pi_{m!}^x}}{K}\right)\sum_{j=1}^{m!}\frac{n_{\pi_j^y}}{K}\right.$$
$$+\frac{n_{\pi_1^y}}{K}\ln\left(\frac{n_{\pi_1^y}}{K}\right)\sum_{i=1}^{m!}\frac{n_{\pi_i^x}}{K} + \frac{n_{\pi_2^y}}{K}\ln\left(\frac{n_{\pi_2^y}}{K}\right)\sum_{i=1}^{m!}\frac{n_{\pi_i^x}}{K} + \ldots + \frac{n_{\pi_{m!}^y}}{K}\ln\left(\frac{n_{\pi_{m!}^y}}{K}\right)\sum_{i=1}^{m!}\frac{n_{\pi_i^x}}{K}$$
$$\left.-\sum_{i=1}^{m!}\sum_{j=1}^{m!}\frac{n_{\eta_{ij}}}{K}\ln\left(\frac{n_{\eta_{ij}}}{K}\right)\right] = -2K[h_{\mathbf{W}}(m) - h_X(m) - h_Y(m)]$$

Therefore, we have proved Theorem 1.

THEOREM 1. *Let $\{\mathbf{W}_t = (X_t, Y_t)\}_{t\in I}$ be a two-dimensional time series. Denote by $h_{\mathbf{W}}(m)$, $h_X(m)$ and $h_Y(m)$ the permutation entropy defined in eqn* (4) *for a fixed embedding dimension $m \geq 2$, with $m \in \mathbb{N}$. If the time series $\{X_t\}_{t\in I}$ and $\{Y_t\}_{t\in I}$ are independent then the statistic*

$$\hat{\Lambda}(\mathrm{m}) = 2(\mathrm{T} - \mathrm{m} + 1)[\mathrm{h_X}(\mathrm{m}) + \mathrm{h_Y}(\mathrm{m}) - \mathrm{h}_{\mathbf{W}}(\mathrm{m})] \tag{8}$$

is asymptotically $\chi^2_{(m!-1)^2}$ distributed.

We emphasize that Theorem 1 does not require any stationary assumption for model $\mathbf{W}_t$ under the null of independence between processes. This is remarkable, as it makes the test more generally applicable, in comparison with other available tests.

Let α be a real number with $0 \leq \alpha \leq 1$. Let χ^2_α be such that

$$P(\chi^2_{(m!-1)^2} > \chi^2_\alpha) = \alpha.$$

Then to test

$$H_0 : \{X_t\}_{t\in I} \text{ and } \{Y_t\}_{t\in I} \text{ are independent}$$

the decision rule in the application of the $\Lambda(m)$ test at a $100(1 - \alpha)\%$ confidence level is:

$$\begin{array}{ll} \text{If } 0 \leq \hat{\Lambda}(m) \leq \chi^2_\alpha, & \text{Accept } H_0, \\ \text{Otherwise}, & \text{Reject } H_0. \end{array} \tag{9}$$

The $\Lambda(m)$ test can be generalized to the case of n variables, $\mathbf{W} = (X_1, X_2, \ldots, X_n)$, as it can be seen in Corollary 1. The proof is straightforward following the steps of the proof of Theorem 1.

COROLLARY 1. *Let $\{\mathbf{W}_t = (X_{1t}, X_{2t}, \ldots, X_{nt})\}_{t\in I}$ be an n-dimensional time series. Denote by $h_{\mathbf{W}}(m)$ and $h_{X_i}(m)$ the permutation entropy defined in eqn* (4) *for a fixed embedding dimension $m \geq 2$, with $m \in \mathbb{N}$. If the time series $\{X_{it}\}_{t\in I}$ are independent for $i = 1, 2, \ldots, n$, then the statistic*

$$\hat{\Lambda}(m) = 2(T - m + 1)\left[\sum_{i=1}^{n} h_{X_i}(m) - h_{\mathbf{W}}(m)\right] \tag{10}$$

is asymptotically $\chi^2_{m!^n-1-n(m!-1)}$ distributed.

3.1. Consistency

We next prove that $\Lambda(m)$ test is consistent for those processes where dependence structure is within the embedding dimension. This is a valuable property since the test will asymptotically reject independence between $\{X_t\}_{t\in I}$ and $\{Y_t\}_{t\in I}$ whenever there is dependence of order $\leq m$.

Theorem 2 establishes consistency property under dependence of order $\leq m$ between two time series. The proof is given in the Appendix.

THEOREM 2. *Let $\{X_t\}_{t\in I}$ and $\{Y_t\}_{t\in I}$ be a jointly strictly stationary processes, and $m \geq 2$ with $m \in \mathbb{N}$. Then, under dependence of order $\leq m$ between $\{X_t\}_{t\in I}$ and $\{Y_t\}$,*

$$\lim_{T\to\infty} \Pr(\hat{\Lambda}(m) > C) = 1$$

for all $0 < C < \infty, C \in \mathbb{R}$.

Since $\hat{\Lambda}(m) \to +\infty$ under the alternative of dependence between the processes, the upper tailed critical values are used. The consistency property of the test can be generalized for the case of fixed alternatives that consider n cross-dependent processes. Corollary 2 applies in this situation.

COROLLARY 2. *Let* $\{\mathbf{W}_t = (X_{1t}, X_{2t}, \ldots, X_{nt})\}_{t \in I}$ *be an n-strictly stationary dimensional time series where* $m \geq 2, m \in \mathbb{N}$. *Then, under cross-dependence of order* $\leq m$ *among* $\{X_{it}\}_{t \in I}$ *for* $i = 1,2,\ldots,n$,

$$\lim_{T \to \infty} \Pr(\hat{\Lambda}(m) > C) = 1$$

for all $0 < C < \infty, C \in \mathbb{R}$.

Notice that because of Theorem 2, stationarity is now required to rule out pathological processes that might appear under the alternative hypothesis. For instance, take the case of two cross-dependent trend-time series with an increasing trend. For these series the trend will force the symbol $(0,1,2,\ldots,m-1)$ to appear yielding to a very low (very close to 0) value of permutation entropy and therefore the Λ-test will not reject the null of independence when it is false.

3.2. The parameter *m*

From an application point of view, the utilization of the new test requires selection of the *m*-dimensional parameter. The number of symbols (events) increases powerly with *m*, and then the selection has to be in relation to the available sample size. For $m = 3$, the number of symbols used for computing $\Lambda(3)$ is 36, while for $m = 4$ the test evaluates over 576 possible symbols. Following Rohatgi (1976), for chi-squared tests, the sample size of the data sets should be well above the number of possible events (symbols), and hence the general recommendation is to work with samples that contain approximately 10 times the number of symbols. In any case, the number of symbols should not be less than five times, otherwise test's conclusions might be misleading.

Parameter *m* (or embedding dimension) is fixed when the test $\Lambda(m)$ is applied. Notice that by increasing *m*, the test develops a finer search (in comparison with lower *m*s) for potential dependencies among series. This is so because the test evaluates over an increasing number symbols and because the order of potential dependence between series augments with *m*. On the other hand, as noted before, higher *m*s required higher sample sizes. Therefore, the selection of parameter *m* will depend on researcher's interest and on the available number of observations. Needless to say that sample size requirements are stronger if the multi-variate ($n \geq 3$) version of the test is applied. The availability of data will basically depend on user's domain of research. Obviously, the new test can be used in all branches of science where time-series analysis is required. Apart from financial time series where data are not a limitation, social sciences are relatively scarce on data. This limitation sharply contrasts with data abundance in life and technical sciences. In this article we have studied the test under the restriction of relatively small data sets, and according to previous observations we have set $m = 3$ which evaluates over 36 symbols and so we can consider sample sizes above 180 observations.

3.3. Detecting cross-dependence in the case of high-order lags

Notice that when one wants to test cross-dependence at a high-order lag, say $d \geq 5$, then one should take the embedding dimension *m* greater than or equal to *d*. This fact would lead us to consider time series of length $T \geq 5m!^2$ ($T \geq 72{,}000$ for $m = 5$) for the Λ-test not to be misleading (see previous section for a detailed discussion).

Nevertheless, this problem can be easily overcome. To do so, we will use the following modified permutation entropy. Let $\{X_t\}_{t \in I}$ be a time series and fix a positive integer *d*. Then, we say that *t* is of (π_i^x, d) type if and only if $\pi_i^x = (i_1, i_2, \ldots, i_m)$ is the unique symbol satisfying the following two conditions:

(a) $X_{t+di_1} \leq X_{t+di_2} \leq \cdots \leq X_{t+di_m}$ and
(b) $i_{s-1} < i_s$ if $X_{t+di_{s-1}} = X_{t+di_s}$

Then within this frame we define the *d*-permutation entropy with embedding dimension *m* as

$$h_X(m,d) = -\sum_{\pi_i^x \in S_m} p((\pi_i^x, d)) \log(p((\pi_i^x, d))) \tag{11}$$

where

$$p((\pi^x, d)) = \frac{\sharp\{t \in I \mid t \text{ is of } (\pi^x, d) \text{ type}\}}{T - dm + 1}$$

Similarly, for the time series $\{Y_t\}_{t \in I}$ and for the two-dimensional time series $\{\mathbf{W}_t = (X_t, Y_t)\}_{t \in I}$ we can define

$$\begin{aligned} h_Y(m,d) &= -\sum_{\pi_j^y \in S_m} p(\pi_j^y, d) \ln(p(\pi_j^y, d)) \\ h_{\mathbf{W}}(m,d) &= -\sum_{\eta \in \Gamma_m} p(\eta, d) \ln(p(\eta, d)). \end{aligned} \tag{12}$$

Then under this setting we obtain the following result whose proof is straightforward from the proof of Theorem 1.

COROLLARY 1. *Let* $\{(X_t, Y_t)\}_{t \in I}$ *be a two-dimensional time series with* $|I| = T$. *Let d be a fixed positive integer number. Denote by* $h_{\mathbf{W}}(m,d)$, $h_X(m,d)$ *and* $h_Y(m,d)$ *the d-permutation entropy defined in eqn* (12) *for a fixed embedding dimension* $m \geq 2$, *with* $m \in \mathbb{N}$. *If the time series* $\{X_t\}_{t \in I}$ *and* $\{Y_t\}_{t \in I}$ *are independent then the statistic*

$$\hat{\Lambda}(m,d) = 2(T - dm + 1)[h_X(m,d) + h_Y(m,d) - h_{\mathbf{W}}(m,d)] \tag{13}$$

is asymptotically $\chi^2_{(m!-1)^2}$ *distributed.*

Notice that the same discussion as in the previous section applies here for the parameter m because the same number of symbols ($m!^2$) is used. Therefore, Corollary 3 allows the researcher to test for cross-dependence at order lags d greater than m, avoiding the problem of the necessity of a huge sample size. The usefulness of this corollary will be shown in the following section.

4. FINITE-SAMPLE PERFORMANCE

From a practical point of view, it is natural to enquire about the finite-sample properties of the proposed test and about the behaviour of the new test in comparison with other available tests. To this end, Monte Carlo methods are used. In addition to the Λ-statistic, the Haugh (1976) and Hong (1996) statistics are also included in the simulation process.

Haugh's (1976) procedure considers the following portmanteau statistic given by

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

where

$$r_{\hat{u}\hat{v}}(j) = \frac{\sum_{t=j+1}^{n} \hat{u}_t \hat{v}_{t-j}}{\left(\sum_{t=1}^{n} \hat{u}_t^2 \sum_{t=1}^{n} \hat{v}_t^2\right)^{1/2}}$$

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 $\hat{u}_t, \hat{v}_t, t = 1, \ldots, n$ 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) generalizes Haugh's statistic. In fact, Hong's test is a weighted sum of residual cross-correlations of the form

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

where

$$M_n(k) = \sum_{j=1-n}^{n-1} (1 - |j|/n) k^2(j/d) \quad \text{and} \quad V_n(k) = \sum_{j=2-n}^{n-2} (1 - |j|/n)(1 - (|j|+1)/n) k^4(j/d).$$

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. In general, some authors (see Fan and Linton, 1999) have indicated that asymptotic behaviour of kernel-based tests are of little value in finite samples and that their results heavily depend on the bandwidth (Skaug and Tjostheim, 1993). As several kernels are available, we have selected the Daniell kernel as it seems to outperform, in terms of power, other smooth ones.

To conduct size and power experiments we have used the following models:

$$\begin{aligned}
&\text{DGP1: } X_t = 0.5X_{t-1} + u_t, \;\; Y_t = 0.5Y_{t-1} + v_t.\\
&\text{DGP2: } X_t = X_{t-1} + u_t, \;\; Y_t = Y_{t-1} + v_t.\\
&\text{DGP3: } X_t = u_t + 0.5u^2_{t-1}, \;\; Y_t = v_t + 0.5v^2_{t-1}\\
&\text{DGP4: } X_t = 0.8X_{t-1} + u_t, \;\; Y_t = 0.8Y_{t-1} + X_t + v_t.\\
&\text{DGP5: } X_t = 0.7X_{t-1} + u_t, \;\; Y_t = 0.3Y_{t-1} + 0.5Y_{t-2}X_{t-1} + v_t.\\
&\text{DGP6: } X_t = u_t, \;\; Y_t = 0.8X^2_{t-1} + v_t.\\
&\text{DGP7: } X_t = 0.6X_{t-1} + u_t, \;\; Y_t = 0.6u^2_{t-1} + v_t.
\end{aligned}$$

where both u_t and v_t are independent and identically distributed (i.i.d.) $N(0,1)$ and independent of each other.

We use DGP1, which specifies two independent Gaussian autoregressive AR(1) processes, to provide a size experiment on the three statistic tests. DGP2 also allows us to study the level of $\Lambda(m)$ test under departures from the stationary assumption by considering two random walks. Notice that for DGP2, we do not report results for Haugh and Hong tests as both require an stationary assumption to control the level of the tests. DGP3 represents a nonlinear independent model that can only be contrasted with the $\Lambda(m)$ test.

On the other hand, DGP4 is a linear dependent model that entails ideal conditions (normality and linearity) for the application of Haugh and Hong statistics. Finally, to compare the power of the considered tests against nonlinear alternatives, we have studied in

Table 1. Rejection rates (%) out of 2000 replications under DGP1 for $\hat{\Lambda}(m), S_M$ and Q_m

	$\hat{\Lambda}(m)$	S_M	Q_n	$\hat{\Lambda}(m)$	S_M	Q_n
	$T=200$			$T=500$		
5%	5.8	4.3	6.4	5.3	4.5	6.3
10%	11.1	10.6	11.5	9.4	9.1	11.1
	$T=1000$			$T=3000$		
5%	5.1	4.8	5.4	4.9	5.3	5.4
10%	10.1	9.9	9.8	9.7	10.0	10.1

DGP1: $X_t = 0.5X_{t-1} + u_t$; $Y_t = 0.5Y_{t-1} + v_t$. $\hat{\Lambda}(m)$, entropy-based test; S_M, Haugh's classical test; Q_m, Hong's test.

Table 2. Rejection rates (%) out of 2000 replications under the Nulls (DGP2 and DGP3) for $\hat{\Lambda}(m)$

	DGP2				DGP3			
	$T=200$	$T=500$	$T=1000$	$T=3000$	$T=200$	$T=500$	$T=1000$	$T=3000$
5%	6.5	6.4	5.0	5.0	6.7	6.2	5.2	5.0
10%	12.1	11.9	11.1	10.0	12.3	11.6	11.4	10.1

DGP2: $X_t = X_{t-1} + u_t$, $Y_t = Y_{t-1} + v_t$; DPG3: $X_t = u_t + 0.5u_{t-1}^2, Y_t = v_t + 0.5v_{t-1}^2$.

Table 3. Rejections rates (%) out of 2000 replications at the 5% level under four alternative hypotheses

	$T=200$	$T=500$	$T=1000$	$T=3000$	$T=200$	$T=500$	$T=1000$	$T=3000$
	DGP4				DGP5			
S_M	100	100	100	100	21	20	31	29
Q_m	100	100	100	100	23	22	24	24
$\Lambda(m)$	100	100	100	100	19	28	51	97
	DGP6				DGP7			
S_M	11	10	11	37	11	14	15	17
Q_m	20	23	28	27	29	31	34	33
$\Lambda(m)$	21	39	80	100	30	50	76	100

DGP4: $X_t = 0.8\,X_{t-1} + u_t$, $Y_t = 0.8\,Y_{t-1} + X_t + v_t$; DGP5: $X_t = 0.7\,X_{t-1} + u_t$, $Y_t = 0.3\,Y_{t-1} + 0.5\,Y_{t-2}\,X_{t-1} + v_t$; DGP6: $X_t = u_t, Y_t = 0.6X_{t-1}^2 + v_t$; DGP7: $X_t = 0.6X_{t-1} + u_t, Y_t = 0.6u_{t-1}^2 + v_t$, where $u_t, v_t \sim N(0,1)$.

Table 4. Rejections rates (%) out of 2000 replications at the 5% level under a nonlinear dependent model (DGP8) for $\hat{\Lambda}(3,d)$

	$T=200$	$T=500$	$T=1000$	$T=3000$
$d=1$	6.5	6.1	5.0	5.0
$d=2$	6.4	6.4	5.0	5.1
$d=5$	40	92	99	100

DGP8: $X_t = 0.5X_{t-1} + u_t, Y_t = v_t + 0.8u_{t-5}^2$.

DGP5 a stable bivariate nonlinear autoregressive model and we have completed the study by analysing two nontrivial nonlinear models, DGP6 and DGP7.

Four sample sizes have been contemplated $T = 200$, 500, 1000 and 3000. For each T, we generate $T + 200$ observations and then discard the first 200 to reduce the effects of initial values. As noticed earlier, in this simulation we have utilized the Daniell kernel, namely

$$k(z) = \frac{\sin(\pi z)}{\pi z}, \quad z \in \mathbb{R},$$

to calculate Hong's test. In addition, it is also required to compute Haugh[1] and Hong tests to choose the rate d. In this respect, we have considered $d = \lfloor \log n \rfloor$ ($\lfloor a \rfloor$ denotes the integer part of a). As already commented, we have fixed $m = 3$ when using the new test.[2]

Tables 1 and 2 report size performances under distinct nulls (DGP1--3) based on 2000 replications. The main observation is that the three tests have reasonably good size behaviour under stationary and linear independent processes. Table 2 points out the good performance of the new entropy-based test for nonstationary and nonlinear departures under the null. However, Table 3 shows the empirical performance (power) of the tests under several alternative hypotheses (either linear and nonlinear). As for the linear model, Haugh-type statistics and $\Lambda(m)$ behave really well independent of the sample size. This also highlights that the three tests present such a competitive performance for linear models that none of the three tests dominates in terms of power. This observation motivates the study of tests' performance under nonlinear dependencies. Under three nonlinear models (DGP5−7) the $\Lambda(m)$ test outperforms Haugh and Hong tests under that alternative. In fact, as sample size increases, $\Lambda(m)$ test rapidly augments power, while the Haugh-type statistics do not better off. When data availability is not a restriction (say $T > 500$) the new test shows an unbeatable behaviour.

Finally, we study the capability of the new test to detect dependence at lags (d) when $d > m$. To this end, we use the modified statistic $\hat{\Lambda}(3, d)$ as described in eqn (13) and we study its performance under the following nonlinear jointly dependent model where nonlinear dependence is at lag 5:

$$\text{DGP 8}: X_t = 0.5X_{t-1} + u_t, \quad Y_t = v_t + 0.8\,u_{t-5}^2$$

We have computed the statistic for the same sample sizes as before and for several ds. Table 4 reports the results.

The entropy-based test clearly detects nonlinear dependencies even when $d > m$ (i.e. $5 > 3$) and it does at the correct lag. As a result, the utility of the new model-free test is emphasized, among others because the lag parameter d allows testing for independence without increasing the embedding dimension m.

In summary, we have presented a new model-free test for independence between two stationary time series, that, in contrast to Haugh-type tests, avoids autoregressive moving average (ARMA) prespecification and kernel selection. Besides, it is suitable for either linear or nonlinear processes and it is consistent against any stationary-dependent process of order $\leq m$. The test and its theoretical properties are easily generalized to detect cross-dependencies among n different variables ($n \geq 3$). Empirical simulations indicate and highlight the general utility of the test.

APPENDIX : PROOFS

Proof of Proposition 1. Following the main lines of the proof given in Matilla-García and Ruiz (2008), for a given symbol $\eta \in \Gamma_m$, we define the random variable $Z_{\eta t}$ as follows:

$$Z_{\eta t} = \begin{cases} 1, & \text{if } t \text{ is of } \eta \text{ type,} \\ 0, & \text{otherwise,} \end{cases} \tag{14}$$

Then $Z_{\eta t}$ is a Bernoulli variable with probability of 'success' p_η, where 'success' means that t is of η type. It is straightforward to see that

$$\sum_{\eta \in \Gamma_m} p_\eta = 1 \tag{15}$$

Now assume that the set I is finite and of order T. Then we are interested in knowing how many t's are of η type for each symbol $\eta \in \Gamma_m$. Recall that $K = T - m + 1$. To answer the question we construct the following variable

$$Y_\eta = \sum_{t=1}^{K} Z_{\eta t} \tag{16}$$

The variable Y_η can take the values {0, 1, 2,..., K}. Then it follows that the variable Y_η is the Binomial random variable

$$Y_\eta \approx B(K, p_\eta). \tag{17}$$

As a result, the joint probability density function of the $m!^2$ variables ($Y_{\eta_1}, Y_{\eta_2}, \ldots, Y_{\eta_{m!2}}$) is:

$$P(Y_{\eta_{11}} = a_1, Y_{\eta_{12}} = a_2, \ldots, Y_{\eta_{m!m!}} = a_{m!^2}) = \frac{(a_1 + a_2 + \ldots + a_{m!^2})!}{a_1!a_2!\cdots a_{m!^2}!} p_{\eta_{11}}^{a_1} p_{\eta_{12}}^{a_2} \cdots p_{\eta_{m!m!}}^{a_{m!^2}} \tag{18}$$

where $a_1 + a_2 + \cdots + a_{m!^2} = K$. Consequently, the joint distribution of the $m!^2$ variables ($Y_{\eta_{11}}, Y_{\eta_{12}}, \ldots, Y_{\eta_{m!m!}}$) is a multinomial distribution.

The likelihood function of the distribution (18) is:

$$L(p_{\eta_{11}}, p_{\eta_{12}}, \ldots, p_{\eta_{m!^2}}) = \frac{K!}{n_{\eta_{11}}!n_{\eta_{12}}!\cdots n_{\eta_{m!m!}}!} p_{\eta_{11}}^{n_{\eta_{11}}} p_{\eta_{12}}^{n_{\eta_{12}}} \cdots p_{\eta_{m!m!}}^{n_{\eta_{m!m!}}} \tag{19}$$

and since $\sum_{i=1}^{m!^2} p_{\eta_i} = 1$ it follows that

$$L(p_{\eta_{11}}, p_{\eta_{12}}, \ldots, p_{\eta_{m!m!}}) = \frac{K!}{n_{\eta_{11}}!n_{\eta_{12}}!\cdots n_{\eta_{m!m!}}!} p_{\eta_{11}}^{n_{\eta_{11}}} p_{\eta_{12}}^{n_{\eta_{12}}} \cdots (1 - p_{\eta_{11}} - p_{\eta_{12}} - \cdots - p_{\eta_{m!(m!-1)}})^{n_{\eta_{m!m!}}} \tag{20}$$

Then the logarithm of this likelihood function remains as

$$\begin{aligned} \ln(L(p_{\eta_{11}}, p_{\eta_{12}}, \ldots, p_{\eta_{m!m!}})) = \ln\left(\frac{K!}{n_{\eta_{11}}!n_{\eta_{12}}!\cdots n_{\eta_{m!m!}}!}\right) + \sum_{i=1}^{m!} \sum_{j=1}^{m!-1} n_{\eta_{ij}} \ln(p_{\eta_{ij}}) \\ + n_{\eta_{m!m!}} \ln(1 - p_{\eta_{11}} - p_{\eta_{12}} - \cdots - p_{\eta_{m!(m!-1)}}). \end{aligned} \tag{21}$$

To obtain the maximum likelihood estimators $\hat{p}_{\eta_{ij}}$ of $p_{\eta_{ij}}$ for all $i,j = 1, 2, \ldots, m!$, we solve the following equation

$$\frac{\partial \ln(L(p_{\eta_{11}}, p_{\eta_{12}}, \ldots, p_{\eta_{m!m!}}))}{\partial p_{\eta_{ij}}} = 0 \tag{22}$$

to obtain

$$\hat{p}_{\eta_{ij}} = \frac{n_{\eta_{ij}}}{K}. \tag{23}$$

Let us consider that $p^{(0)}_{\eta_{ij}}$ refers to the probability of each symbol under the null. Then the likelihood ratio statistic is (see, for example, Lehmann, 1986):

$$\begin{aligned}\lambda(Y) &= \frac{\frac{K!}{n_{\eta_{11}}!n_{\eta_{12}}!\cdots n_{\eta_{m!m!}}!} p^{(0)n_{\eta_{11}}}_{\eta_{11}} p^{(0)n_{\eta_{12}}}_{\eta_{12}} \cdots p^{(0)n_{\eta_{m!n}}}_{\eta_{m!m!}}}{\frac{K!}{n_{\eta_{11}}!n_{\eta_{12}}!\ldots n_{\eta_{m!m!}}!} \hat{p}^{n_{\eta_{11}}}_{\eta_{11}} \hat{p}^{n_{\eta_{12}}}_{\eta_{12}} \cdots \hat{p}^{n_{\eta_{m!m!}}}_{\eta_{m!m!}}} = \frac{\prod_{i=1}^{m!}\prod_{j=1}^{m!} p^{(0)n_{\eta_{ij}}}_{\eta_{ij}}}{\prod_{i=1}^{m!}\prod_{j=1}^{m!}\left(\frac{n_{\eta_{ij}}}{K}\right)^{n_{\eta_{ij}}}} \\ &= K^{\sum_{i=1}^{m!}\sum_{j=1}^{m!} n_{\eta_{ij}}} \prod_{i=1}^{m!}\prod_{j=1}^{m!}\left(\frac{p^{(0)}_{\eta_{ij}}}{n_{\eta_{ij}}}\right)^{n_{\eta_{ij}}} = K^K \prod_{i=1}^{m!}\prod_{j=1}^{m!}\left(\frac{p^{(0)}_{\eta_{ij}}}{n_{\eta_{ij}}}\right)^{n_{\eta_{ij}}}.\end{aligned} \tag{24}$$

Moreover, $\Lambda(m) = -2 \ln (\lambda(Y))$ asymptotically follows a chi-squared distribution with $(m! - 1)^2$ degrees of freedom (see, for instance, Lehmann, 1986). Hence,

$$\Lambda(m) = -2\ln(\lambda(Y)) = -2\left[K\ln(K) + \sum_{i=1}^{m!}\sum_{j=1}^{m!} n_{\eta_{ij}} \ln\left(\frac{p^{(0)}_{\eta_{ij}}}{n_{\eta_{ij}}}\right)\right] \sim \chi^2_{(m!-1)^2} \tag{25}$$

Proof of Theorem 2. First notice that the limits $\lim_{T\to\infty} \hat{p}_\pi = p_\pi$ exists for every stationary process (see Bandt and Pompe, 2002), where $\hat{p}_\pi = n_\pi/(T - m + 1)$. Therefore, by continuity of the logarithm function, it follows that

$$\lim_{T\to\infty} \hat{h}(m) = h(m). \tag{26}$$

Recall that

$$\hat{\Lambda}(m) = 2(T - m + 1)[\hat{h}_X(m) + \hat{h}_Y(m) - \hat{h}_{\mathbf{W}}(m)].$$

Notice that

$$\begin{aligned}-\sum_{i=1}^{m!}\frac{n_{\pi_i^x}}{K}\ln\left(\frac{n_{\pi_i^x}}{K}\right) - \sum_{j=1}^{m!}\frac{n_{\pi_j^y}}{K}\ln\left(\frac{n_{\pi_j^y}}{K}\right) &= -\sum_{i=1}^{m!}\left(\frac{n_{\eta_{i1}}}{K}\ln\left(\frac{n_{\pi_i^x}}{K}\right) + \cdots + \frac{n_{\eta_{im!}}}{K}\ln\left(\frac{n_{\pi_i^x}}{K}\right)\right) \\ &\quad -\sum_{j=1}^{m!}\left(\frac{n_{\eta_{1j}}}{K}\ln\left(\frac{n_{\pi_j^y}}{K}\right) + \cdots + \frac{n_{\eta_{m!j}}}{K}\ln\left(\frac{n_{\pi_i^x}}{K}\right)\right) \\ &= -\sum_{i=1}^{m!}\sum_{j=1}^{m!}\frac{n_{\eta_{ij}}}{K}\left(\ln\left(\frac{n_{\pi_i^x}}{K}\right) + \ln\left(\frac{n_{\pi_j^y}}{K}\right)\right) \\ &= -\sum_{i=1}^{m!}\sum_{j=1}^{m!}\frac{n_{\eta_{ij}}}{K}\ln\left(\frac{n_{\pi_i^x}}{K}\frac{n_{\pi_j^y}}{K}\right)\end{aligned}$$

and hence

$$\hat{h}_X(m) + \hat{h}_Y(m) = -\sum_{i=1}^{m!}\sum_{j=1}^{m!}\frac{n_{\eta_{ij}}}{K}\ln\left(\frac{n_{\pi_i^x}}{K}\frac{n_{\pi_j^y}}{K}\right) \tag{27}$$

In addition, for every positive real number $x \in \mathbb{R}$ we have $-\ln(x) \geq 1 - x$. Indeed, for $x \neq 1$ we have $-\ln(x) > 1 - x$. Then, using this fact and eqn (27) and since the processes are jointly stationary, we have that under H_1

$$\begin{aligned}\hat{h}_X(m) + \hat{h}_Y(m) - \hat{h}_{\mathbf{W}}(m) &= -\sum_{i=1}^{m!}\sum_{j=1}^{m!}\frac{n_{\eta_{ij}}}{K}\ln\left(\frac{n_{\pi_i^x}}{K}\frac{n_{\pi_j^y}}{K}\right) + \sum_{i=1}^{m!}\sum_{j=1}^{m!}\frac{n_{\eta_{ij}}}{K}\ln\left(\frac{n_{\eta_{ij}}}{K}\right) \\ &= -\sum_{i=1}^{m!}\sum_{j=1}^{m!}\frac{n_{\eta_{ij}}}{K}\ln\left(\frac{\frac{n_{\pi_i^x}}{K}\frac{n_{\pi_j^y}}{K}}{\frac{n_{\eta_{ij}}}{K}}\right) > \sum_{i=1}^{m!}\sum_{j=1}^{m!}\frac{n_{\eta_{ij}}}{K}\left(1 - \frac{\frac{n_{\pi_i^x}}{K}\frac{n_{\pi_j^y}}{K}}{\frac{n_{\eta_{ij}}}{K}}\right) = \sum_{i=1}^{m!}\sum_{j=1}^{m!}\frac{n_{\eta_{ij}}}{K} - \sum_{i=1}^{m!}\sum_{j=1}^{m!}\frac{n_{\pi_i^x}}{K}\frac{n_{\pi_j^y}}{K} = 1 - 1 = 0\end{aligned}$$

Therefore, under H_1, we obtain

$$\hat{h}_X(m) + \hat{h}_Y(m) - \hat{h}_{\mathbf{W}}(m) > 0 \tag{28}$$

Let $0 < C < \infty$ with $C \in \mathbb{R}$ and take T large enough such that

$$\frac{C}{2(T-m+1)} < \hat{h}_X(m) + \hat{h}_Y(m) - \hat{h}_{\mathbf{W}}(m). \tag{29}$$

Then

$$\begin{aligned}\Pr[\hat{\Lambda}(m) > C] &= \Pr[2(T-m+1)(\hat{h}_X(m) + \hat{h}_Y(m) - \hat{h}_{\mathbf{W}}(m)) > C] \\ &= \Pr\left[\hat{h}_X(m) + \hat{h}_Y(m) - \hat{h}_{\mathbf{W}}(m) > \frac{C}{2(T-m+1)}\right]\end{aligned}$$

Therefore, by eqn (29) we have

$$\lim_{T\to\infty} \Pr(\hat{\Lambda}(m) > C) = 1$$

as desired.

Acknowledgements

Mariano Matilla-García was partially supported by MCI (Ministerio de Ciencia e Innovación) and FEDER (Fondo Europeo de Desarrollo Regional) by grant MTM2008-03679/MTM. Manuel Ruiz Marín was partially supported by MEC (Ministerio de Educación y Ciencia), grant MTM2009-07373 and Fundación Séneca of Región de Murcia.

NOTES

1. That is $M = d$.
2. Results for $m = 4$ and 5, for a fixed $T = 200{,}000$ are available upon request.

REFERENCES

Bandt, C. and Pompe, B. (2002) Permutation entropy: a natural complexity measure for time series. *Physical Review Letters* **88**, 174102.

Fan, Y. and Linton, O. (1999) Some higher order theory for a consistent nonparametric model specification test. *The Journal of Statistical Planning and Inference* **109**, 125–54.

Geweke, J. (1981) A comparison of tests of independence of two covariance stationary time series. *Journal of the American Statistical Association* **76**, 363–73.

Hao, B. and Zheng, W. (1998) *Applied Symbolic Dynamics and Chaos*. Singapore: World Scientific.

Haugh, L. D. (1976) Checking the independence of two covariance-stationary time series: a univariate residual cross correlation approach. *Journal of the American Statistical Association* **71**, 378–85.

Hong, Y. (1996) Testing for independence between two covariance stationary time series. *Biometrika* **83**, 615–25.

Joe, H. (1989a) Relative entropy measures of multivariate dependence. *Journal of the American Statistical Association* **84**, 157–64.

Joe, H. (1989b) Estimation of entropy and other functionals of a multivariate density. *Annals of the Institute of Statistical Mathematics* **41**, 683–97.

Koch, P. D. and Yang, S. S. (1986) A method for testing the independence of two time series that accounts for a potential pattern in the cross-correlation function. *Journal of the American Statistical Association* **81**, 533–44.

Lehmann, E. L. (1986) *Testing Statistical Hypothesis*. New York: John Wiley & Sons.

Matilla-García, M. and Ruiz, M. (2008) A nonparametric independence test using permutation entropy. *Journal of Econometrics* **144**, 139–55.

Pham, D. T., Roy, R. and Cedras, L. (2003) Tests for non-correlation of two cointegrated ARMA time series. *Journal of Time Series Analysis* **24**, 553–77.

Pierce, A. (1977) Lack of dependence among economic variables. *Journal of the American Statistical Association* **72**, 11–22.

Rohatgi, V. K. (1976) *An Introduction to Probability Theory and Mathematical Statistics*. New York: John Wiley & Sons.

Skaug, H. J. and Tjøstheim, D. (1993) Nonparametric tests of serial independence. In *Developments in Time Series Analysis: The Priestley Birthday Volume* (ed. T. Subba Rao). London: Chapman & Hall, pp. 207–29.