A Non-Parametric Independence Test Using Permutation Entropy

Mariano Matilla-García
Economía Aplicada Cuantitativa I.
UNED

Manuel Ruiz Marín
Universidad Politécnica de Cartagena

ABSTRACT

In the present paper we construct a new, simple and powerful test for independence by using symbolic dynamics and permutation entropy as a measure of serial dependence. We also give the asymptotic distribution of an affine transformation of the permutation entropy under the null hypothesis of independence. An application to several daily financial time series illustrates our approach.

Introduction

Independence is one of the most valuable notions in econometrics, time series analysis and statistics due to the fact that most tests boil down to checking some sort of independence assumption. As a result, an extensive literature on how to test independence has arisen: Correlation tests (see King (1987) for a survey) are widely used, but they are not consistent against alternatives with zero autocorrelation. Examples of serially dependent processes that exhibit zero autocorrelation include autoregressive conditional heteroskedastic (ARCH), bilinear, non-linear moving average processes and iterative logistic maps. The nonparametric literature also contains a large number of serial independence test (see Dufour et al. (1982) for a bibliographical survey on permutation, sign and rank tests for independence). These test procedures work well under commonly used dependence structures, like ARMA models, but they also fail to detect subtle nonlinear underlying dependence structures. Needless to say that other nonparemetric tests have emerged (Brock et al. (1996) and Pinkse (1998), among others) to cover these difficulties.

Serial independence has been increasingly studied by using entropy measures. These measures avoid restrictive parametric assumptions on the probability distribution generating the data, and they can capture the dependence present in a time series. As an outcome, establishing asymptotic distribution theory for smoothed nonparametric entropy measures of serial dependence has been so far challenging (see Hong and White (2005) and references therein). This line of research is narrowly connected with Information Theory: Jaynes (1957) introduced the maximum entropy principle (MEP) which determines the probability distribution of a random variable by maximizing the Shannon entropy, subject to certain moment conditions. This optimization principle is the same as the Kullback principle of minimizing the Kullback-Leibler relative entropy when one of the distributions is uniform. Jaynes' MEP was a turning point in the use of Shannon's entropy as a method of statistical inference.

The use of entropy has played a leading role as a measure of the dependence present in a time series in the last two decades. Joe (1989a, 1989b) considered a smoothed nonparametric entropy measure of multivariate dependence of an independent and identically distributed (i.i.d.) random vector. Granger and Lin (1994) proposed a normalized smoothed nonparametric entropy measure of serial dependence to identify important lags in time series. Robinson (1991) developed a test for serial dependence using a modified entropy measure. Skaug and Tjøstheim (1993b, 1996) also considered a general class of smoothed density-based tests for serial dependence, which includes a test based on an entropy measure modified with a weight function.

As Granger and Lin (1994) pointed out, there is no asymptotic distribution theory available for smoothed nonparametric entropy measures of serial dependence. Consistency and in some cases convergence rates have been established, but asymptotic distributions for these entropy estimators are not available. Robinson (1991) first provided an asymptotic distribution theory for a smoothed nonparametric modified entropy measure of serial dependence, using a sample-splitting device. Granger et al. (2004) introduced a transformed metric entropy of dependence. Recently, the relevant investigation of Hong and White (2005) have provided, under certain assumptions, an asymptotic theory for a class of kernel-based smoothed nonparametric entropy estimators of serial dependence. They also show that their theory yields the limit distribution of the Granger and Lin's normalized entropy measure, which was previously unknown in the literature. Moreover, they develop a test that is asymptotically locally more powerful than Robinson's test. Nevertheless, most of the methods used to test for independence via an entropy measure of serial dependence extrictly require a continuous distribution function of the unknown underlying data generating process and also need to estimate the density function with stochastic kernels. As a result, free-choice parameters are introduced. Another difficulty acknowledged by Hong and White (2005) is that the finite sample level of their own test (and in general of others entropy-based tests) differs from the asymptotic one; furthermore, asymptotic theory may not work well even for relatively large samples. This leads to implement, for each sample size, non-naive bootstrap procedures in order to correctly compute the test. Moreover, Hong and White need the time series X_t to have a compact support in the interval [0,1], although this is not necessarily a restriction whenever the test is invariant under monotonous transformations of the series. Obtaining a compact support can always be ensured by a continuous strictly monotonous transformation such as the logistic function.

On the other hand, there are a number of other nonparemetric tests for independence that avoid smoothed nonparemetric estimation (Skaug and Tjøstheim (1993a); Delgado (1996); Hong (1998); and Hong (2000), among others; see also Tjøstheim (1996) for an excellent complete survey). These procedures are based on the empirical distribution function or on the characteristic function. Importantly, some of these statistics are invariant to order preserving transformation; the distribution generating the data can be continuous or discrete; under certain conditions, the tests are distribution free. Unfortunately, some of these statistics have nonstandard limiting distribution. In these procedures, as in the case of tests based on smoothing estimation techniques, the test statistic is a distance between the joint density (or estimated joint distribution) and the marginal densities (or estimated marginal distributions).

In the present paper we take a different way, and we propose a new test for independence also based on Information Theory, but avoiding the potential disadvantage of depending on the choice of a smoothing number. More precisely, the absence of dependencies in the unknown underlying data generating process is studied via symbolic dynamics. Symbolic dynamics studies dynamical systems on the basis of the symbol sequences obtained for a suitable partition of the state space. The basic idea behind symbolic dynamics is to divide the phase space into a finite number of regions and label each region by an alphabetical letter. In this regard, symbolic dynamics is a coarse-grained description of dynamics. Some recent tests of independence are also a coarse-grained description of the underlying dynamic from which the data was generated. Even though coarse-grained methods lose a certain amount of detailed information, some essential features of the dynamics may be kept, e.g., periodicities and dependencies, among others. Symbolic dynamics has been used for investigation of non-linear dynamical systems (for an overview see Hao and Zheng, 1998). The process of symbolizing a time series is based upon the method of delay time coordinates, introduced by Takens (1981), in order to carry out the phase space reconstruction. Such a reconstruction is done from a scalar time series and all relevant components (relative to the underlying dynamics), such as dependencies, periodicity and complexity changes, have to be extracted from it.

Then, given a time series $\{X_t\}$, we study the dependence present in the series by translating the problem into symbolic dynamic and then, we use the entropy measure associated to these symbols to test the dependence present in the time series. More concretely, we study all m! permutations (symbols) π of length m in the symmetric group S_m which are considered here as possible order types of m different numbers. Afterwards, we give the distribution followed by the mentioned symbols and define the entropy measure associated to them. This entropy measure is called permutation entropy (see Bandt and Pompe (2002) and Section 2 for a detailed explanation). Moreover under the null of independence we prove that an affine transformation of the permutation entropy is asymptotically χ^2 distributed.

This allows us to construct a simple, consistent, easy to compute and powerful test for independence. The new test avoids restrictive assumptions on the probabilistic distribution generating the data. This fact allows the test to be of more general applicability. The distribution generating the data can be continuous or discrete. No moment is required; this is attractive for time series whose variances are infinite, as often arises in financial time series. It does not involve sample splitting (as Robinson (1991) requires) and thus, it does not need choosing tuning parameters that can lead to ambiguous conclusions when the test is used by two different practitioners. An interesting property of the proposed test is that it is invariant under monotonous (continuous or not) transformation of the data. Therefore, provided that $\{X_t\}$ is i.i.d. if and only if any series of its continuous monotonous transformation is i.i.d., the invariance property guaranties that no information is lost. Of important relevance for our test is that the finite sample level does not differ from the asymptotic level, and hence general applicability and reproducibility of the test is ensured.

The rest of the paper is structured as follows. In Section 1 we introduce the notation and several definitions in order to describe the symbolic dynamic representation methodology. The procedure is illustrated with an easy example. In Section 2 we give the construction of the independence test via permutation entropy and we prove that under the null of independence an affine transformation of the permutation entropy is asymptotically χ^2 distributed. Size and power of the new test are studied by Monte Carlo methods in Section 3. An empirical application for daily financial returns is reported in Section 5. Finally, we give the conclusions and final remarks in Section 4.

Definitions and Notation

In this section we give some definitions and we introduce the basic notation. We illustrate the definitions with a very easy example.

Let $\{X_t\}_{t\in I}$ be a real-valued time series. 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 $\pi_i = (i_1,...,i_m) \in S_m$. We will call an element π_i in the symmetric group S_m a symbol. The positive integer m is usually known as embedding dimension.

Now we define an ordinal pattern for a symbol $\pi_i = (i_1,...,i_m) \in S_m$ at a given time $t \in I$. To this end we consider that the time series is embedded in an *m*-dimensional space as follows:

$$X_m(t) = \left(X_{t+1}, X_{t+2},..., X_{t+m}\right), t \in I$$

Then we say that t is of π_i-type if and only if $\pi_i = (i_1, i_2,..., i_m)$ is the unique symbol in the group satisfying the two following conditions:

$$(a)\ X_{t+i_1} \le X_{t+i_2} \le ... \le X_{t+i_m}, \text{ and}$$
$$(b)\ i_{s-1} < i_s \ \ if \ \ X_{t+i_{s-1}} = X_{t+i_s}$$

Condition (b) guaranties uniqueness of the symbol $\pi_i = (i_1, i_2,..., i_m)$. 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.

Notice that for all *t* such that *t* is of π_i-type the *m*-history $X_m(t)$ is converted into a unique symbol π_i. This symbol π_i describes how the ordering of the dates *t+0<t+1<…<t+(m-1)* is converted into the ordering of the values in the time series under scrutiny. In order 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

$$\{X_1 = 2,\ X_2 = 8,\ X_3 = 6,\ X_4 = 5,\ X_5 = 4,\ X_6 = 9,\ X_7 = 3\} \qquad (1)$$

Then for $t = 2$ we have that $X_{t+2} = 5 < X_{t+1} = 6 < X_{t+0} = 8$ and therefore we have that the period $t=2$ is of (2,1,0)-type.

Also, given a time series $\{X_t\}_{t\in I}$ and an embedding dimension *m* one could easily compute the relative frequency of a symbol $\pi_i = (i_1,..., i_m) \in S_m$ by:

$$p(\pi) \equiv p_\pi = \frac{\#\{t \in I \,|\, t \text{ is of } \pi\text{-type}\}}{|I| - m + 1} \qquad (2)$$

where by |I| we denote the cardinal of the set I.

Then for the time series given in (1) we have that the 3-history $X_3(1)=(X_1=2, X_2=8, X_3=6)$ is represented by the symbol (0,2,1); $X_3(2) = (8,6,5)$ and $X_3(3)=(6,5,4)$ are represented by the symbol (2,1,0); $X_3(4) = (5,4,9)$ is represented by the symbol (1,0,2) and finally $X_3(5) = (4,9,3)$ is represented by the symbol (2,0,1). Therefore we obtain that $p((0,1,2)) = 0 = p((1,2,0))$, $p((0,2,1)) = (1/5)$, $p((1,0,2)) = (1/5)$, $p((2,0,1)) = (1/5)$ and $p((2,1,0)) = (2/5)$.

Now under this setting we can define the permutation entropy of a time series $\{X_t\}_{t\in I}$ for an embedding dimension $m \geq 2$. This entropy is defined as the Shanon's entropy of the $m!$ distinct symbols as follows:

$$h(m) = -\sum_{\pi_i \in S_m} p_{\pi_i} \ln\left(p_{\pi_i}\right) \tag{3}$$

Permutation entropy, $h(m)$, is the information contained in comparing m consecutive values of the time series. It is clear that $0 \leq h(m) \leq \ln(m!)$ where the lower bound is attained for an increasing or decreasing sequence of values, and the upper bound for a completely random system (i.i.d. sequence) where all $m!$ possible permutations appear with the same probability. For the time series given in (1) we have that $h(3) = -3(1/5)\ln((1/5)) - (2/5)\ln((2/5)) \approx 1.332179$.

Construction of the independence test

In this section we construct an independence test with all the machinery defined in Section 2. We also prove that an affine transformation of the permutation entropy defined in (3) is asymptotically χ^2 distributed.

Let $\{X_t\}_{t\in I}$ be a time series and m be a fixed embedding dimension. In order to construct a test for serial independence in $\{X_t\}_{t\in I}$, which is the aim of this paper, we consider the following null hypothesis:

$$H_0 = \{X_t\}_{t\in I} \ \ i.i.d. \tag{4}$$

against any other alternative.

Now for a symbol $\pi_i = (i_1, ..., i_m) \in S_m$ we define the random variable $Z_{\pi_i t}$ as follows:

$$Z_{\pi_i t} = \begin{cases} 1 & \text{if } X_{t+i_1} \leq X_{t+i_2} \leq \cdots \leq X_{t+i_m} \\ 0 & \text{otherwise,} \end{cases} \qquad (5)$$

that is, we have that $Z_{\pi_i t} = 1$ if and only if t is of $\pi_i -$ type.

Then $Z_{\pi_i t}$ is a Bernoulli variable with probability of "success" p_{π_i}, where "success" means that *t* is of $\pi_i -$ type. It is straightforward to see that

$$\sum_{i=1}^{m!} p_{\pi_i} = 1 \qquad (6)$$

Now assume that the set *I* is finite and of order *T*. Then we are interested in knowing how many *t*'s are of $\pi_i -$ type for all symbol $\pi_i \in S_m$. Let us call *K=T-m+1*. In order to answer the question we construct the following variable

$$Y_{\pi_i} = \sum_{t=1}^{K} Z_{\pi_i t} \qquad (7)$$

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

$$Y_{\pi_i} \approx B(K, p_{\pi_i}). \qquad (8)$$

For each symbol $\pi_i \in S_m$ we are going to denote by

$$n_{\pi_i} = \sharp\{t \in I \mid t \text{ is of } \pi_i - \text{type}\}. \qquad (9)$$

for *i=1,2,…,m!*. Then under the null H_0, the joint probability density function of the *m*! variables $(Y_{\pi_1}, Y_{\pi_2}, ..., Y_{\pi_{m!}})$ is:

$$P(Y_{\pi_1} = a_1, Y_{\pi_2} = a_2, \ldots, Y_{\pi_{m!}} = a_{m!}) = \frac{(a_1 + a_2 + \cdots + a_{m!})!}{a_1! a_2! \cdot \cdots \cdot a_{m!}!} p_{\pi_1}^{a_1} p_{\pi_2}^{a_2} \cdots p_{\pi_{m!}}^{a_{m!}} \qquad (10)$$

where $a_1 + a_2 + ... + a_{m!} = K$. Consequently the joint distribution of the $m!$ variables $\left(Y_{\pi_1}, Y_{\pi_2}, ..., Y_{\pi_{m!}}\right)$ is a multinomial distribution.

The likelihood function of the distribution (10) is:

$$L(p_{\pi_1}, p_{\pi_2}, \ldots, p_{\pi_{m!}}) = \frac{K!}{n_{\pi_1}! n_{\pi_2}! \cdot \cdots \cdot n_{\pi_{m!}}!} p_{\pi_1}^{n_{\pi_1}} p_{\pi_2}^{n_{\pi_2}} \cdots p_{\pi_{m!}}^{n_{\pi_{m!}}} \qquad (11)$$

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

$$L(p_{\pi_1}, p_{\pi_2}, \ldots, p_{\pi_{m!}}) = \frac{K!}{n_{\pi_1}! n_{\pi_2}! \cdot \cdots \cdot n_{\pi_{m!}}!} p_{\pi_1}^{n_{\pi_1}} p_{\pi_2}^{n_{\pi_2}} \cdots (1 - p_{\pi_1} p_{\pi_2} \cdots p_{\pi_{m!-1}})^{n_{\pi_{m!}}} \qquad (12)$$

Then the logarithm of this likelihood function remains as

$$\begin{aligned} Ln(L(p_{\pi_1}, p_{\pi_2}, \ldots, p_{\pi_{m!}})) &= Ln(\frac{K!}{n_{\pi_1}! n_{\pi_2}! \cdot \cdots \cdot n_{\pi_{m!}}!}) + \sum_{i=i}^{m!-1} n_{\pi_i} Ln(p_{\pi_i}) \\ &\quad + n_{\pi_{m!}} Ln(1 - p_{\pi_1} p_{\pi_2} \cdots p_{\pi_{m!-1}}). \qquad (13) \end{aligned}$$

In order to obtain the maximum likelihood estimators $\hat{p}_{\pi_i}$ of p_{π_i} for all $i = 1,2,\ldots,m!$, we solve the following equation

$$\frac{\partial Ln(L(p_{\pi_1}, p_{\pi_2}, \ldots, p_{\pi_{m!}}))}{p_{\pi_i}} = 0 \qquad (14)$$

to get that

$$\widehat{p}_{\pi_i} = \frac{n_{\pi_i}}{K}. \qquad (15)$$

Then the likelihood ratio statistic is (see for example Lehmann, 1986):

$$\lambda(Y) = \frac{\frac{K!}{n_{\pi_1}!n_{\pi_2}!\cdots n_{\pi_{m!}}!} p_{\pi_1}^{n_{\pi_1}} p_{\pi_2}^{n_{\pi_2}} \cdots p_{\pi_{m!}}^{n_{\pi_{m!}}}}{\frac{K}{n_{\pi_1}!n_{\pi_2}!\cdots n_{\pi_{m!}}!} \widehat{p}_{\pi_1}^{n_{\pi_1}} \widehat{p}_{\pi_2}^{n_{\pi_2}} \cdots \widehat{p}_{\pi_{m!}}^{n_{\pi_{m!}}}} = \frac{\prod_{i=1}^{m!} p_{\pi_i}^{n_{\pi_i}}}{\prod_{i=1}^{m!} (\frac{n_{\pi_i}}{K})^{n_{\pi_i}}} =$$

$$= K^{\sum_{i=1}^{m!} n_{\pi_i}} \prod_{i=1}^{m!} (\frac{p_{\pi_i}}{n_{\pi_i}})^{n_{\pi_i}} = K^K \prod_{i=1}^{m!} (\frac{p_{\pi_i}}{n_{\pi_i}})^{n_{\pi_i}}. \quad (16)$$

On the other hand, $G(m) = -2Ln(\lambda(Y))$ asymptotically follows a Chi-squared distribution with $m!$-1 degrees of freedom (see for instance Lehmann, 1986). Hence

$$G(m) = -2Ln(\lambda(Y)) = -2[KLn(K) + \sum_{i=1}^{m!} n_{\pi_i} Ln(\frac{p_{\pi_i}}{n_{\pi_i}})] \sim \chi^2_{m!-1} \quad (17)$$

Now under the null H_0 it is clear that $p_{\pi_i} = \frac{1}{m!}$ for all i=1,2,…,m!. Then it follows that

$$\begin{aligned} G(m) &= -2K[Ln(K) + \sum_{i=1}^{m!} \frac{n_{\pi_i}}{K} Ln(\frac{p_{\pi_i}}{n_{\pi_i}})] \\ &= -2K[Ln(K) + \sum_{i=1}^{m!} \frac{n_{\pi_i}}{K} (Ln(\frac{1}{m!}) - Ln(n_{\pi_i}))] \\ &= -2K[Ln(K) + \sum_{i=1}^{m!} \frac{n_{\pi_i}}{K} (Ln(\frac{1}{m!}) - Ln(\frac{n_{\pi_i}}{K}) - Ln(K))] \quad (18) \end{aligned}$$

Now taking into account that $h(m) = -\sum_{\pi_i \in S_m} p_{\pi_i} \ln\left(p_{\pi_i}\right) = -\sum_{i=1}^{m!} \frac{n_{\pi i}}{K} \ln\left(\frac{n_{\pi i}}{K}\right)$ we have that

$$G(m) = -2K[Ln(\frac{1}{m!}) + h(m)] = -2(K)[h(m) - Ln(m!)] = 2K[Ln(m!) - h(m)]. \quad (19)$$

Therefore we have proved the following result.

Theorem 3.1. *Let $\{X_t\}_{t\in I}$ be a time series with |I|=T. Denote by h(m) the permutation entropy defined in (3) for a fixed embedding dimension m>2, with m∈N. If the time series $\{X_t\}_{t\in I}$ is i.i.d., then the affine transformation of the permutation entropy*

$$G(m) = 2(T - m + 1)[Ln(m!) - h(m)] \tag{20}$$

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

Let α be a real number with 0≤α≤1. Let χ^2_α be such that

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

Then to test

$$H_0 : \{X_t\}_{t\in I} \quad i.i.d.$$

the decision rule in the application of the G(m) test at a 100(1-α)% confidence level is:

$$\begin{array}{ll} If\ \chi^2_{1-\frac{\alpha}{2}} \leq G(m) \leq \chi^2_{\frac{\alpha}{2}} & Accept\ H_0 \\ \qquad\qquad Otherwise & Reject\ H_0 \end{array} \tag{21}$$

It is important to note, from a practical point of view, that the researcher has to decide upon the embedding dimension m in order to compute permutation entropy and therefore to calculate the G(m) statistic. Fortunately, this decision can be easily conducted. Note that T should be larger than the number of permutation symbols m! in order to have at least the same number of m-histories as possible symbols (events) $\pi_i \in S_m$. When the χ^2 is applied in practice, and all the expected frequencies are ≥5, the limiting tabulated χ^2 distribution gives, as a rule, the value χ^2_α with an approximation sufficient for ordinary purposes. For this reason, we require to work with data sets containing at least five times the number of possible events. For instance, a data set of 200 observations is enough for computing G(4) because 24 symbols are obtained for *m*

= 4; similarly, 600 observations is the smallest data set that can be considered for an embedding dimension of *m*=5 since in this case 120 symbols might be found. Beyond embedding dimension of *m*=6, data requirements are unrealistic for real economic time series, so we do not use such dimensions. Conversely, for *m*=3 only six possible symbols are analyzed, and then the degree of information capture by these symbols is very limited and therefore we do not suggest the use of *m*=3. Through this paper we compute permutation entropy in a manner that the researcher has not to choose the embedding dimension: For a given data set of T observations, the embedding dimension will be the largest m that satisfies $5m! \leq T$ with m=2,3,4,.... For example, for case of T = 500, we set[1] *m*=4.

Finite-Sample Behavior

Various time series were generated in order to test the size and power of the *G(m)* test. We have studied the new test for three embedding dimensions: *m*=4,5 and 6, accordingly, sample sizes of *T=120,T=600* and *T=3600* have been considered, respectively. In order to conduct size experiments the analyzed models have been the following:

1. A Gaussian distribution, zero mean and unit variance, N(0,1)
2. A Uniform distribution on the (0,1) interval, U(0,1)
3. A Chi-square distribution with 4 degrees of freedom, χ^2
4. A Student's t-distribution with 4degrees of freedom, t_4

Each process was repeated 2000 times and the proportion of rejections of the i.i.d. null was calculated using a nominal size of 1, 5 and 10 per cent.

Table 1 reports the empirical size of our test under the four i.i.d. models. As it can be seen, even for the smallest considered data set (that is, for m=4 where T=120=5m!), the test is reasonably well sized, with rejection frequencies occurring at approximately their nominal rates. The finite sample level does not differ from the asymptotic level. Furthermore, test's size improves as T increases (T>5m!). The same behavior is

[1] Note that if T is too large (T>25.200), then the selected m will be too large as well (indeed m>7), and hence the procedure will be too expensive in terms of computational time. For this reason, and because of the usual length of economic time series, we recomend to operate with m=6 for T>3600

obtained (although not reported here for sake of space) for m=5 and m=6. These observations are relevant because as Hong and White (2005) note the finite sample level of their own test (and in general of the previous entropy-based tests) differs from the asymptotic one, not only for small samples but also for large samples. As a result, the G(m) test does not need computing non-naive bootstrap procedures in order to compute the test. These results also show that the new test behaves well when facing several random distributions with different shaped probability density functions such as fat-tailed, either symmetric or not, and infinitely fat-tailed like the uniform distribution.

Table 1. Size of the $G(m)$ test

	$G(4)$								
	$T=120$			$T=240$			$T=580$		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
$N(0,1)$	1,90	5,45	10,49	1.65	5.40	10.40	1.60	4.49	10.22
$U(0,1)$	2,00	6,15	10,32	1.40	4.45	10.21	1.81	5.10	10.15
χ^2_4	2,05	6,14	10,43	1.85	4.95	10.10	1.51	4.67	10.03
t_4	1,85	5,60	10,16	1.62	4.30	9.51	1.12	5.00	9.63

Notes: For each distribution, 2000 simulation iterations for $G(m)$ test have been computed.

On the other hand, in order to test for the power of the G(m) test, we have considered the same data generating processes (DGPs) as in Granger and Lin (1994) because of its rich nonlinear variety. The models are the following:

DGP 1 $X_t = \varepsilon_t + 0.8\varepsilon_{t-1}^2$,
DGP 2 $X_t = \varepsilon_t + 0.8\varepsilon_{t-2}^2$,
DGP 3 $X_t = \varepsilon_t + 0.8\varepsilon_{t-3}^2$,
DGP 4 $X_t = \varepsilon_t + 0.8\varepsilon_{t-1}^2 + 0.8\varepsilon_{t-2}^2 + 0.8\varepsilon_{t-3}^2$,
DGP 5 $X_t = |X_{t-1}|^{0.8} + \varepsilon_t$,
DGP 6 $X_t = sign(X_{t-1}) + \varepsilon_t$,
DGP 7 $X_t = 0.8X_{t-1} + \varepsilon_t$,
DGP 8 $X_t = X_{t-1} + \varepsilon_t$,
DGP 9 $X_t = 0.6\varepsilon_{t-1}X_{t-2} + \varepsilon_t$,
DGP 10 $X_t = 4X_{t-1}(1 - X_t)$

where x_t □ $i.i.d.\ N(0,1)$. For the simulations that follow, we increase Granger and Lin's (1994) DGPs with the following process:

DGP 11 $X_t = \sqrt{h_t} \varepsilon_t, h_t = (1 + 0.5 h_{t-1}^2)$

For each DGP, we first generate T+200 observations and then discard the first 200 to mitigate the impact of initial values. DGP's 1-4 are MA processes of order 1,2,3 and 3 respectively. DGP's 5-7 are AR(1) autoregressions with various decaying memory properties. DGP 8 is a simple I(1) with persistent memory and DGP 9 is bilinear with white noise characteristics. DGP 10 is the logistic function generating chaotic dynamics. DGP 11 is a ARCH(1) process commonly employed in financial applications.
We shall use these models to evaluate the power of our non-parametric dependence test. A minimum of 2000 Monte Carlo replications from each model are computed. Code was written in Mathematica 5.2 programming language.

Table 2 reports the empirical rejection rates of G(4), G(5), G(6) under DGPs 1-11, for T=120, T=600 and T=3600, respectively. As we can see, the power of our test against dependent models (either linear or non-linear) is near 100% when T=3600. For T=600 and regarding the nonlinear moving average processes (DGPs 1-4), the power of the G(m) test is significantly high at all levels, except, perhaps, for the DGP 3 at 1%. This test's performance does not hold for the smallest data set. Concerning the AR(1) processes (DGPs 5-7) the power of the test for T=600 is very close to 100% at all levels, while for T=120 the results are not so optimistic. The behavior of the test, in terms of power, in presence of an I(1) process with persistent memory (DGP 8) is near 100%, regardless the sample size. The performance of the test for the bilinear with white noise (DGP 9) is very high at all levels, except for the case of T=120. As expected, the chaotic nature of process (DGP 10) is always capture regardless the sample size. Finally, for the ARCH model (DGP 11), our test only performs well for the largest sample size.

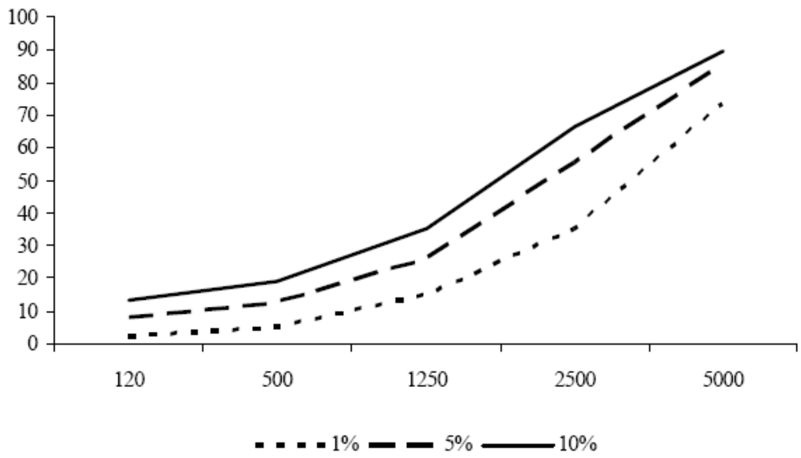
Table 2. Power of the $G(m)$ test

	$G(4)$ $T=120$			$G(5)$ $T=600$			$G(6)$ $T=3600$		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
DGP 1	15,65	29,5	38,5	97,50	99,30	99,49	100	100	100
DGP 2	5,55	16,10	22,10	78,95	91,50	94,15	100	100	100
DGP 3	3,15	10,40	15,7	47,80	67,05	76,15	100	100	100
DGP 4	6,00	18,05	24,10	84,50	92,55	96,05	100	100	100
DGP 5	38,55	59,65	65,30	100	100	100	100	100	100
DGP 6	19,65	35,25	44,25	99,55	99,80	99,90	100	100	100
DGP 7	74,20	86,00	89,75	100	100	100	100	100	100
DGP 8	94,85	98,05	98,15	100	100	100	100	100	100
DGP 9	3,65	10,55	16,85	77,00	89,00	100	100	100	100
DGP 10	100	100	100	100	100	100	100	100	100
DGP 11	2,60	8,20	13,15	11,50	26,15	35,20	93,11	97.81	100

Notes: For each DGP, 2000 simulation iterations for $G(m)$ test have been computed

Two compatible reasons can explain why DGP 11 is not capture by G(4) and G(5). One relies on the fact that we are computing the tests over small sample sizes. The other one is that the number of analyzed symbols (24 symbols for m=4, and 120 symbols for m=5) is not enough to capture the complexity of ARCH processes. Figure 1 shows the power of G(4) test for increasing sample sizes (T goes from 120 to 5000). By inspection of Figure 1 it is evident that valuable powers are attained for large sample sizes at all nominal levels[2]. Sample sizes greater than 5000 are needed for G(4) to obtain, at all nominal levels, powers larger than 75%. This fact underlines the central role played by the number of analyzed symbols for a given data set. In this regard, observe from Table 2 that the power against DGP 11 is almost 100 per cent for T=3600 and m=6. Note that this power's gain is obtained via rising the number of symbols with which the data set is analyzed, and not by increasing the sample size (T). Therefore, even though the sample size is certainly important for G(m) test's power, it is more relevant the role played by the number of symbols used for detecting dependence.

[2] The same occurs for m=5.



Powers of $G(4)$ test for increasing sample sizes
$T : 120 - 5000$

In comparison with other tests for independence based also on entropy concepts (Robinson 1991; Skaug and Tjøstheim 1996; Hong and White 2005), the main advantages of our test are the following: (a) It does not require to selected many free parameters, in fact, the unique free parameter is the number of symbols (m!) upon which permutation entropy is computed. (b) The test is well defined for both continuous and discrete processes. (c) The test is invariant under monotonous (either linear or nonlinear) transformation of the data set. Invariance is important since otherwise inadvertent transformations would produce different levels of dependence. Notice that this property is essential when testing independence in time series. For instance in Hong and White's test, in order to ensure that the support of the time series belongs to the compact interval [0,1] they make a logistic transformation (monotonous) of the data. Therefore if the test is invariant under monotonous transformations, the condition on the support is not a restriction. (d) G(m) test does not need any estimation of the density function nor the use of stochastic kernels for its computation, something that does not occur with most of the entropy-based tests for serial dependence.

Advantages (a)-(d) make the G(m) test not only more general and less dependent on free choice parameters, but also easier to compute and shorter in computing running time.

Empirical Application

This section illustrates our test by using the G(m) statistic to explore possible dependences in the following well studied daily financial returns: Dow Jones Industrial Average (DJIA) and three exchange rate time series, namely, the French franc, the German mark and the Canadian dollar, all against the U.S. dollar. Daily returns of Dow Jones Industrial Average (DJIA) ranges from January 3, 1928 to October 18, 2000 while daily exchange rate's returns go from January 4, 1971 to December 31, 1998. Returns, R_t, are defined as the difference of logarithm of the stock price index (or of the corresponding exchange rate) ($R_t = \Delta \ln P_t$), where P_t is the daily closing price index (or the closing exchange rate). We are interested in testing independence[3] of the R_t as a way of examining the correctness of the random walk hypothesis for the logged prices (or for the logged the three exchange rates).

We also test for independence in various power transformed absolute returns series, $|R_t|^d$,d=1/2,1,1,5 and 2, since it is well known (see Ding et al., 1993) that proxy volatility measures, as power transformations of financial returns, have higher autocorrelation compared to the return series.

Given that T>3600 for all data sets, we have compute the G(m) test for m=6. Results for the four data sets are reported in Table 3. As can be observed, as regards the DJIA returns, the null of independence is rejected at 1 per cent significance level; however, independence is not rejected for the transformed absolute returns. Exchange rate's returns behave similarly, independence is rejected at 1 per cent; however, now the test rejects the null at 1% level, for all absolute return's power transformations. All these results indicate that the DJIA daily stock prices and the three exchange rate time series do not follow a random walk. As regarding volatility, measured via the proxy variables $|R_t|^d$, these results point out that, while rejection of the random walk hypothesis for the three exchange rate returns might be due to potential strong volatility clustering, this is not the case for the DJIA since independence cannot be rejected for absolute power

[3] The exchange rate time series under study in this section have been recently analyzed looking for chaotic behavior in Fernández et al. (2005); the same happens for the DJIA data set, see Shintani and Linton (2004).

transformations. It is also interesting to comment that the same conclusions are obtained by fixing m=4 and m=5([4]).

Table 3. $G(6)$ Test for Independence for several financial returns

DJIA	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$	$G(6)$
R_t	Reject	Reject	Reject	2151.01
$\lvert R_t \rvert^{0.5}$	Accept	Accept	Accept	715.445
$\lvert R_t \rvert$	Accept	Accept	Accept	715.445
$\lvert R_t \rvert^{1.5}$	Accept	Accept	Accept	715.445
$\lvert R_t \rvert^{2}$	Accept	Accept	Accept	715.445
French frank	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$	$G(6)$
R_t	Reject	Reject	Reject	3597.55
$\lvert R_t \rvert^{0.5}$	Reject	Reject	Reject	1434.57
$\lvert R_t \rvert$	Reject	Reject	Reject	1434.57
$\lvert R_t \rvert^{1.5}$	Reject	Reject	Reject	1434.57
$\lvert R_t \rvert^{2}$	Reject	Reject	Reject	1434.57
German mark	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$	$G(6)$
R_t	Reject	Reject	Reject	3480
$\lvert R_t \rvert^{0.5}$	Reject	Reject	Reject	1427.62
$\lvert R_t \rvert$	Reject	Reject	Reject	1427.62
$\lvert R_t \rvert^{1.5}$	Reject	Reject	Reject	1427.62
$\lvert R_t \rvert^{2}$	Reject	Reject	Reject	1427.62
Canadian dollar	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$	$G(6)$
R_t	Reject	Reject	Reject	3980.14
$\lvert R_t \rvert^{0.5}$	Reject	Reject	Reject	2994.45
$\lvert R_t \rvert$	Reject	Reject	Reject	2994.45
$\lvert R_t \rvert^{1.5}$	Reject	Reject	Reject	2994.45
$\lvert R_t \rvert^{2}$	Reject	Reject	Reject	2994.45

Note:The critical values for the $G(6)$ test at an $\alpha = 1\%, 5\%$ and 10% are:
$\chi^2_{0.005} = 782.491$; $\chi^2_{0.95} = 657.783$; $\chi^2_{0.025} = 795.2$; $\chi^2_{0.975} = 646.588$; $\chi^2_{0.005} = 820.431$; $\chi^2_{0.995} = 625.08$

4 Although not reported in the paper, these results are available from the authors.

Conclusions

From a general and wide perspective, this paper expands the interrelationship between Information Theory, statistics and inference, and the research line based on entropy concepts. Particularly, the present paper attempts to analyze limited and noisy data using minimal assumptions. In this fashion, we have proposed a new test for independence which relies on the concept of entropy. This concept, as presented here, is formulated in terms of symbols obtained from ordinal patterns found in a time series. In other words, we do not work with the actual observed values which are real numbers; rather we take the number of order patterns in the observed series as a measure of its complexity. Although this methodology loses a certain amount of detailed information, some essential features of the dynamics are kept, among others, dependence or independence of the data generating process.

Independence is one of the most valuable notions in statistics and econometrics, therefore testing for serial independence is crucial. In this regard and in connection with entropy econometrics, certain amount of significant research has tested for independence by using smoothed nonparametric entropy measures. Robinson (1991), Skaug and Tjøstheim (1996), and recently Hong and White (2005) have provided an asymptotic distribution theory for certain entropy measures, and as a result they have obtained some test for independence. These tests rely on kernel-based estimation techniques, and hence kernels and bandwidths have to be freely selected by the researcher. Most importantly, the finite sample level of these tests differs from the asymptotic one; furthermore, as Hong and White point out, asymptotic theory may not work well even for relatively large samples. Also of relevant importance is that all known entropy-based tests for independence make several assumptions about the data generating process that restrict the general applicability of the test.

As has been shown, this paper provides the asymptotic distribution of an affine transformation of the permutation entropy under the null of independence. The theoretical distributions allows us to construct a test for independence. Importantly for our test, the finite sample level does not differ from the asymptotic level, which is an interesting property that guaranties general applicability and reproducibility of the test. Moreover, the test is invariant under monotonuous transformations of data. Invariance makes our procedure very attractive in practice. Most importantly, our test makes no assumptions about the continuos or discrete nature of the data generating process and of its marginal densities. In sum, the G(m) test only needs X_t be a real-valued time series, and hence it is more general than other entropy-based tests. Two final advantages are its computational simplicity and its short running computational times.

An empirical application to daily Dow Jones Industrial Average price index and to three daily exchange rate returns has illustrated our approach by testing the random walk hypothesis on these prices.

References

Barnett, W.A., Gallant, A.R, Hinich, M.J., Jungeils, J.A, Kaplan, D.T. and Jensen, M.J. 1997. A single blind controlled competition among tests for nonlinearity and chaos. Journal of Econometrics, 82, 157-192.

Brock, W.A., Dechert, W., Scheinkman, J.A and LeBaron, B, 1996. A test for independence based on the Correlation Dimension. Econometric Reviews, 15, 197-235.

Cameron, A Colin & Trivedi, Pravin K. 1993. Tests of Independence in Parametric Models with Applications and Illustrations, Journal of Business & Economic Statistics, 11(1), 29-43.

Ding Z, Granger CWJ, Engle RF. 1993. A long memory property of stock market returns and a new model. Journal of Empirical Finance 1: 83--106.

Fernández-Rodríguez F., Sosvilla-Rivero, S. and Andrada-Félix, J. 2005. Testing chaotic dynamics via Lyapunov Exponents. Journal of Applied Econometrics, 20: 901-930.

Granger, C.W.J. and Lin J.L. 1994. Using the mutual information coefficient to identify lags in nonlinear models. Journal of Time Series analysis, 1, 157-165.

Hao, B. and Zheng, W. 1998. Applied Symbolic Dynamics and Chaos. World Scientific, Singapore.

Hoeffding, W. 1948. A non-parametric test of independence. Annals of Mathematical Statistics 19, 546-557

Hong, Y. and White, H. 2005. Asymptotic distribution theory for nonparametric entropy measures of serial dependence. Econometrica 73, 837--901.

Jaynes, E.T. (1957). Information theory and statistical mechanics. Phys. Rev. 106, 620-630.

Joe, H. (1989a): Relative Entropy Measures of Multivariate Dependence. Journal of the American Statistical Association, 84, 157--164.

Joe, H.(1989b): Estimation of Entropy and Other Functionals of a Multivariate Density, Annals of the Institute of Statistical Mathematics. 41, 683--697.

Johnson, D. and McClelland, R. 1998, A general dependence test and applications. Journal of Applied Econometrics, 13, 627-644.

Lehmann, E.L. 1986. Testing Statistical hypothesis. John Wiley & Sons, New York.
Pinkse, J. 1998. A consistent nonparametric test for serial independence. Journal of Econometrics, 84, 205-231.

Robinson, P.M. 1991. Consistent nonparametric Entropy-Based testing. Review of Economic Studies, 58, 437-453.

Shintani, M. and Linton, O. 2004. Nonparametric neural network estimation of Lyapunov exponents and a direct test for chaos. Journal of Econometrics, 120, 1-33.

Skaug H. J. and Tjøstheim D. (1993) Nonparametric Tests of Serial Independence. Developments in Time Series Analysis: The Priestley Birthday Volume, ed. by T. Subba Rao. London: Chapman & Hall, 207--229.

Skaug H. J. and Tjøstheim D. (1996): Measures of Distance Between Densities with Application to Testing for Serial Independence. Time Series Analysis in Memory of E. J. Hannan, ed. by P. Robinson and M. Rosenblatt. New York Springer-Verlag, 363--377.

Takens F. 1981. Detecting strange attractors in turbulence. In Dynamical Systems and Turbulence, Rand D,   Young L. (eds). Springer-Verlag: Berlin.

Tjøstheim, D. 1996, Measures of dependence and tests for independence, Statistics, 28, 249-284.