A non-parametric independence test using permutation entropy

Mariano Matilla-García, Manuel Ruiz Marín

Author's Accepted Manuscript

A non-parametric independence test using permutation entropy

Mariano Matilla-García, Manuel Ruiz Marín

www.elsevier.com/locate/jeconom

Cite this article as: Mariano Matilla-García and Manuel Ruiz Marín, A non-parametric independence test using permutation entropy, *Journal of Econometrics* (2008), doi:10.1016/j.jeconom.2007.12.005

A Non-Parametric Independence Test Using Permutation Entropy

Abstract

In the present paper we construct a new, simple, consistent and powerful test for independence by using symbolic dynamics and permutation entropy as a measure of serial dependence. We also give a standard asymptotic distribution of an affine transformation of the permutation entropy under the null hypothesis of independence. The test statistic and its standard limit distribution are invariant to any monotonuous transformation. The test applies to time series with discrete or continuous distributions. Eventhough the test is based on entropy measures, it avoids smoothed nonparametric estimation. An application to several daily financial time series illustrates our approach.

1 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 autoregresive 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

1

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[1].

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

---

[1] $X_t^* = \frac{1}{1 + exp^{-X_t}}$ where X_t is the observed data point and X_t^* its corresponding logistic transformation.

2

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 2 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 3 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 4. An empirical application for daily financial returns is reported in Section 5. Finally, we give the conclusions and final remarks in Section 6.

2 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 $\pi_i = (i_1, i_2, \ldots, i_m) \in S_m$. 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 $\pi_i = (i_1, i_2, \ldots, i_m) \in S_m$ 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:

$$X_m(t) = (X_{t+1}, X_{t+2}, \ldots, X_{t+m}) \text{ for } t \in I$$

3

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

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

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

Condition (b) guaranties uniqueness of the symbol π_i. 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 $t+0 < t+1 < \cdots < t+(m-1)$ 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

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

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 $\pi \in S_m$ by:

$$p(\pi) := p_\pi = \frac{\sharp \{t \in I \mid t \text{ is of } \pi - \text{type}\}}{|I| - m + 1} \tag{2}$$

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 $p((0, 1, 2)) = 0 = p((1, 2, 0))$, $p((0, 2, 1)) = \frac{1}{5}$, $p((1, 0, 2)) = \frac{1}{5}$, $p((2, 0, 1)) = \frac{1}{5}$ and $p((2, 1, 0)) = \frac{2}{5}$.

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:

$$h(m) = - \sum_{\pi \in S_m} p_\pi \ln(p_\pi). \tag{3}$$

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 $h(3) = -3\frac{1}{5}\ln(\frac{1}{5}) - \frac{2}{5}\ln(\frac{2}{5}) \approx 1.332179$.

4

3 Construction and Properties 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:

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

against any other alternative.

Now for a symbol $\pi_i = (i_1, i_2, \ldots, i_m) \in S_m$ we define the random variable $Z_{\pi_i t}$ as follows:

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

that is, we have that $Z_{\pi_i t} = 1$ if and only if t is of π_i-type, $Z_{\pi_i t} = 0$ otherwise.

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

$$\sum_{i=1}^{m!} p_{\pi_i} = 1 \tag{6}$$

Now assume that the set I is finite and of order T. Then we are interested in knowing how many t's are of π_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

$$Y_{\pi_i} = \sum_{t=1}^{K} Z_{\pi_i t} \tag{7}$$

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

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

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

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

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

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

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

The likelihood function of the distribution (10) is:

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

5

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

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

Then the logarithm of this likelihood function remains as

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

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

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

to get that

$$\widehat{p}_{\pi_i} = \frac{n_{\pi_i}}{K}. \tag{15}$$

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

$$\begin{aligned}
\lambda(Y) &= \frac{\frac{K!}{n_{\pi_1}! n_{\pi_2}! \cdot \ldots \cdot n_{\pi_{m!}}!} p_{\pi_1}^{n_{\pi_1}} p_{\pi_2}^{n_{\pi_2}} \cdots p_{\pi_{m!}}^{n_{\pi_{m!}}}}{\frac{K}{n_{\pi_1}! n_{\pi_2}! \cdot \ldots \cdot n_{\pi_{m!}}!} \widehat{p}_{\pi_1}^{n_{\pi_1}} \widehat{p}_{\pi_2}^{n_{\pi_2}} \cdots \widehat{p}_{\pi_{m!}}^{n_{\pi_{m!}}}} = \frac{\prod_{i=1}^{m!} p_{\pi_i}^{n_{\pi_i}}}{\prod_{i=1}^{m!} \left(\frac{n_{\pi_i}}{K}\right)^{n_{\pi_i}}} = \\
&= K^{\sum_{i=1}^{m!} n_{\pi_i}} \prod_{i=1}^{m!} \left(\frac{p_{\pi_i}}{n_{\pi_i}}\right)^{n_{\pi_i}} = K^K \prod_{i=1}^{m!} \left(\frac{p_{\pi_i}}{n_{\pi_i}}\right)^{n_{\pi_i}}.
\end{aligned} \tag{16}$$

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

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

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

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

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

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

Therefore we have proved the following theorem.

Theorem 3.1. *Let $\{X_t\}_{t\in I}$ be a real-valued time series with $|I| = T$. Denote by $h(m)$ the permutation entropy defined in (3) for a fixed embedding dimension $m > 2$, with $m \in \mathbb{N}$. 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 $0 \leq \alpha \leq 1$. 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 $100(1 - \alpha)\%$ confidence level is:

$$\begin{aligned} If \; 0 \leq G(m) \leq \chi^2_\alpha & \qquad Accept \; H_0 \\ Otherwise & \qquad Reject \; H_0 \end{aligned} \tag{21}$$

3.1 Consistency of the $G(m)-$test

Next we prove that the $G(m)$ test is consistent for a wide variety of serial dependence processes. This is a valuable property since the test will reject asymptotically serial independence whenever there is serial dependence within the $m-$history. We will denote by $\widehat{G}(m)$ the estimator of $G(m)$. The proof of the following Theorem can be found in Appendix 1.

Theorem 3.2. *Let $\{X_t\}_{t\in I}$ be a strictly stationary process, and $m > 2$ with $m \in \mathbb{N}$. Then $\lim_{T\to\infty} \Pr(\widehat{G}(m) > C) = 1$ under serial dependence of order $\leq m$ for all $0 < C < \infty, C \in \mathbb{R}$.*

Thus, the test based on $G(m)$ is consistent against all serial dependence of order $\leq m$ alternatives to the null of serial independence. Since Theorem 3.2 implies $G(m) \to +\infty$ with probability approaching 1 under serial dependence of order $\leq m$, then upper-tailed critical values are appropriated. We underline at this point that our conditions about the process are milder than those required in other test procedures (see for instance Hong and White (2005), Skaug and Tjøstheim (1996) and references therein).

3.2 Selection of the free parameter m

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) $\pi_i, i = 1, ..., m!$. When the χ^2 is applied in practice, and all the expected frequencies are $\geqslant 5$, the limiting tabulated χ^2 distribution gives, as a rule, the value χ^2_α with an approximation sufficient for ordinary purposes (see chapter 10 of Rohatgi (1976)). For this reason, we require to work with data sets containing *at least* five times the number of possible events (symbols). 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 (=5!) 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 $m = 2, 3, 4, \ldots$. For example, in case of $T = 500$, we then set $m = 4$. On the other hand, we 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 recommend to operate with $m = 6$ for $T > 3600$.

As indicated before, it is worth noting that the possible dependence detected by the $G(m)$ test has to be of order $\leq m$. This is due to the fact that if the dependence structure of the process is of order $> m$, then this dependence is not present in every m-history and therefore the symbols may not capture it.

4   Monte Carlo Evidence

In this section we examine the finite sample behavior of the $G(m)$ test and we also compare our test with other nonparametric tests of independence including those based on smoothing procedures and those based on the empirical distribution function and on the empirical characteristic function.

4.1   Finite sample behavior of $G(m)$

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[2]. 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_4^2

4. A Student's t-distribution with 4 degrees 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 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.

---

[2]Notice that according to subsection 3.2, the selected sample sizes (T) correspond to the lowest sample sizes for which, as a rule, we recomend 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_4^2	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 study the power of the $G(m)$ test, we have considered several data generating processes (DGPs) that have been previously analyzed in the relevant literature: Several processes studied in Granger and Lin (1994) have been chosen because of its rich nonlinear variety, and other selected models have been analyzed with different independence tests like those mainly found in Hong and White (2005), Skaug and Tjøstheim (1993a, 1996), Robinson (1991), Delgado (1996), Hong (1998) and Hong (2000). The models under scrutiny are the following:

$$\begin{aligned}
\text{DGP 1} \quad & X_t = \varepsilon_t + 0.8\varepsilon_{t-1}^2, \\
\text{DGP 2} \quad & X_t = \varepsilon_t + 0.8\varepsilon_{t-2}^2, \\
\text{DGP 3} \quad & X_t = \varepsilon_t + 0.8\varepsilon_{t-3}^2, \\
\text{DGP 4} \quad & X_t = \varepsilon_t + 0.8\varepsilon_{t-1}^2 + 0.8\varepsilon_{t-2}^2 + 0.8\varepsilon_{t-3}^2, \\
\text{DGP 5} \quad & X_t = 0.8|X_{t-1}|^{0.5} + \varepsilon_t, \\
\text{DGP 6} \quad & X_t = sign(X_{t-1}) + 0.43\varepsilon_t, \\
\text{DGP 7} \quad & X_t = 0.3X_{t-1} + \varepsilon_t, \\
\text{DGP 8} \quad & X_t = X_{t-1} + \varepsilon_t, \\
\text{DGP 9} \quad & X_t = 0.8\varepsilon_{t-1}X_{t-1} + \varepsilon_t, \\
\text{DGP 10} \quad & X_t = 4X_{t-1}(1 - X_t) \\
\text{DGP 11} \quad & X_t = \sqrt{h_t}\varepsilon_t, \quad h_t = (1 + 0.8X_{t-1}^2)
\end{aligned}$$

where $\varepsilon_t \sim iid$, $N(0,1)$.

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 nonparametric 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

9

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

| | $G(4)$ | | | $G(5)$ | | | $G(6)$ | | |
| | $T = 120$ | | | $T = 600$ | | | $T = 3600$ | | |
	1%	5%	10%	1%	5%	10%	1%	5%	10%
DGP 1	15.60	27.15	36.75	97.50	99.30	99.75	100	100	100
DGP 2	5.70	14.45	22.55	80.65	91.10	94.35	100	100	100
DGP 3	3.15	10.40	14.35	49.05	65.50	72.20	100	100	100
DGP 4	44.80	63.90	100	100	100	100	100	100	100
DGP 5	9.25	19.30	26.70	61.70	78.45	85.81	100	100	100
DGP 6	19.25	34.50	44.70	99.65	99.95	100	100	100	100
DGP 7	8.58	20.01	28.00	71.65	85.10	89.45	100	100	100
DGP 8	91.75	96.60	98.10	100	100	100	100	100	100
DGP 9	16.60	31.55	42.25	100	100	100	100	100	100
DGP 10	100	100	100	100	100	100	100	100	100
DGP 11	3.45	9.75	14.85	31.50	49.15	59.85	96.25	98.80	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 captured 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[3]. 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 merely by increasing the sample size (T). Therefore, even though the sample size is certainly important for $G(m)$ test's power, it seems to be more relevant the role played by the number of symbols used for detecting dependence.

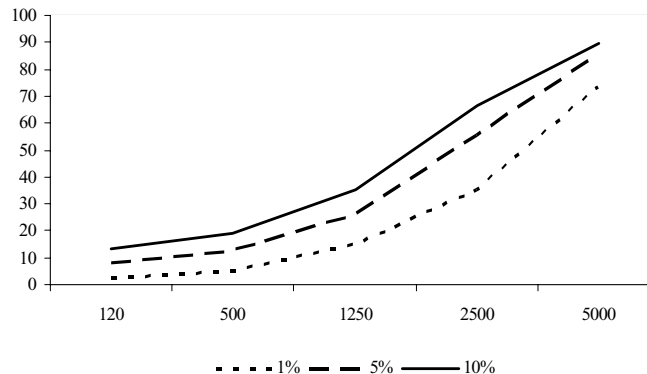
[3]The same occurs for $m = 5$.

10

Figure 1: Powers of $G(4)$ test for increasing sample sizes $T : 120 - 5000$

4.2 Comparison with other tests for independence

So far, it has been shown that the $G(m)$ statistic has a standard limiting distribution under the null, it is distribution free, it is consistent against a broad class of alternatives. In addition the test seems to be well-behaved in finite samples sizes. We compare now the new statistic for independence with other existing tests that also deal with serial dependence (including smoothed and non-smoothed procedures). The comparison is made (i) taking into account some statistical properties of the tests under consideration and (ii) comparing the power of the tests under study against several alternatives.

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 select 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, then 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. (e) No moment condition is required; this is attractive for time series whose variances are infinite, as usually arises in economics and high frequency financial time series. In sum, advantages (a) − (e) make the $G(m)$ test not only more general and less dependent on free choice parameters than other entropy-based tests, but also easier to compute and shorter in computing running time.

On the other hand, we emphasize that a number of independence tests (Skaug and Tjøstheim, (1993a); Delgado (1996); Hong (1998); and Hong (2000), among others) which are based on the empirical distribution function (or on the characteristic function), avoid estimation by smoothing

11

procedures. Note that these tests rest on the fact that the null hypothesis holds if and only if the joint distribution equals the product of the marginals, while our test uses a new technique, namely, it rests on the uniformity of symbols' distribution under the null of independence. Some of these tests share in common with the $G(m)$ test *part but not all* of its advantages: for example, Skaug and Tjøstheim (1993a), Delgado (1996) and Hong (2000) test procedures have a non-standard limiting distribution; and the tests developed by Hong (1998) require stronger assumptions on the DGP to guarantee consistency; in addition, the free selection of kernel functions is necessitated in order to implement the tests.

Since the $G(m)$ test is strongly powerful when the sample size increases (due basically to the increase of the number of symbols under scrutiny), we have conducted a power comparison among different tests (Skaug and Tjøstheim, (1993a); Delgado (1996); Hong (1998); and Hong (2000)) for independence for a relatively small sample size, $T = 200$.

Skaug and Tjøstheim's (1993a) statistic and Hong's (1998) statistic are closely based on the Hoeffding (1948) dependence measure

$$\hat{\sigma}^2(j) \equiv (T - j) \sum_{t=j+1}^{T} \hat{\rho}_j^2 (X_t, X_{t-j}) \tag{22}$$

where

$$\hat{\rho}_j(x, y) \equiv \hat{F}_j(x, y) - \hat{F}_j(x, \infty)\hat{F}_j(\infty, y) \qquad j = 1, ..., T - 1$$

and

$$\hat{F}_j(x, y) \equiv (T - j) \sum_{t=j+1}^{T} \mathbf{1}\left(X_t \leq x\right) \mathbf{1}\left(X_{t-j} \leq y\right)$$

being $\mathbf{1}(\cdot)$ the indicator function.

The Skaug and Tjøstheim (1993a) statistic $T_{ST(p)} \equiv (T - 1) \sum_{j=1}^{p} \hat{\sigma}^2(j)$, under the null of independence, converges in distribution to

$$\sum_{j=1}^{\infty} \sum_{l=1}^{\infty} (j\pi)^{-2}(l\pi)^{-2}\chi_{jl}^2(p), \tag{23}$$

where $\chi_{jl}^2(p)$ are independent χ_p^2–variables. We generate 10000 realizations of (23) where $j = 1, ..., 200$ and $l = 1, ..., 200$, to obtain critical values.

Hong's (1998) statistic

$$T_{H(p)} \equiv 90 \frac{\sum_{j=1}^{T-1} k^2(j/p) \left\{ (T - j)\hat{\sigma}^2(j) - 36^{-1} \right\}}{\left\{ 2 \sum_{j=1}^{T-2} k^4(j/p) \right\}}$$

where we have selected, following Hong (2000), the Daniell kernel $k(z) = \sin(\pi z)/\pi z$, is asymptotically $N(0, 1)$.

Delgado's (1996) statistic is given by

$$T_{D(p)} \equiv \sum_{t=1}^{m} \left[m^{-1} \sum_{s=p+1}^{T} \prod_{j=0}^{p} \mathbf{1}\left(X_{t-j} \leq X_{s-j}\right) - \prod_{j=0}^{p} \left\{ m^{-1} \sum_{s=p+1}^{T} \mathbf{1}\left(X_{t-j} \leq X_{s-j}\right) \right\} \right]^2$$

and its limit distribution is non-standard. Its asymptotic critical values can be approximated in different ways (see Delgado, (1996)), however, since we focus only on $p = 1$, and since in this

12

case $T_{D(p=1)}$ has the same asymptotic distribution as the Blum et *al.* (1961) test, we then use the tabulated values given in Blum et *al.* Notice that statistics $T_{ST(p)}$ and $T_{H(p)}$ also involve the choice of lag order p.

In this power comparison, we also consider the two generalized statistics presented in Hong (2000). Importantly both tests, as happens with $G(m) - test$, do not depend on lag (p) selection. These tests are the following,

$$T_{GCM} \equiv \sum_{j=1}^{T-1} \frac{(T-j)\tilde{\sigma}^2(j)}{(j\pi)^2}$$

where now $\tilde{\sigma}^2(j) \equiv T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} \hat{\rho}_j^2 (X_t, X_s)$, and

$$T_{GKS} \equiv \max_{1 \le t,s \le T} \sup_{\lambda \in [0,1]} \left| \sum_{j=1}^{T-1} (T-j)^{1/2} \hat{\rho}_j (X_t, X_s) \frac{\sqrt{2}\sin(j\pi\lambda)}{j\pi} \right|.$$

Both statistics have a non-standard asymptotic distribution. In consequence, empirical critical values have to be used. In this paper, given the rapid convergence of empirical critical values for these tests, we have utilized critical values given in Hong (2000).

Lastly, we have also included the power of three smooth tests given in Hong and White (2005), namely, Robinson (1991), $R(p)$; Skaug and Tjøstheim (1996), $J(p)$, and Hong and White's test, $HW(p)$. To save space, in Table 3 we only report power at the 1% level based on 2000 iterations of each of the following alternatives: DGP1, DGP5, DGP7, DGP9 and DGP11.

Table 3. Power Comparison

DGPs	Permutation Entropy Test $G(4)$	Smoothed Tests $HW(1)$	$R(1)$	$J(1)$	Non-smoothed Tests $T_{ST(1)}$	$T_{D(1)}$	$T_{H(1)}$	T_{GCM}	T_{GKS}
DGP1	28.0	49.8	37.2	43.2	36.9	36.8	36.9	37.1	22.2
DGP5	13.1	10.0	8.4	7.4	78.8	78.9	79.1	69.0	66.5
DGP7	14.0	7.4	7.6	5.8	80.1	79.6	79.6	72.5	48.3
DGP9	31.7	86.8	75.4	92.4	83.1	83.1	79.7	72.1	48.9
DGP11	4.1	41.0	25.2	68.0	8.9	8.7	8.7	4.2	3.9

Notes: (i) DGP1, $X_t = \varepsilon_t + 0.8\varepsilon_{t-1}^2$;DGP5, $X_t = 0.8|X_{t-1}|^{0.5} + \varepsilon_t$;DGP7, $X_t = 0.3X_{t-1} + \varepsilon_t$; DGP9, $X_t = 0.8\varepsilon_{t-1}X_{t-1} + \varepsilon_t$;DGP11, $X_t = \sqrt{h_t}\varepsilon_t$, $h_t = (1 + 0.8X_{t-1}^2)$, where $\varepsilon_t \sim N(0,1)$.
(ii) For each DGP, 2000 simulation iterations for each test have been computed.

We observe the following:

(a) Kernel-based smoothed nonparametric entropy tests dominates other tests against the bilinear, ARCH and nonlinear moving average models, while the permutation entropy new test and clearly the rest of non-smoothed tests dominate against AR(1) and fractional AR(1) models.

(b) Those tests that do not depend on selecting lag-parameter have less power than others.

(c) Comparing with non-smoothed tests (ie., $G(4)$, $T_{ST}, T_D, T_H, T_{GCM}$ and T_{GKS}), the G(4)-test has less power than other non-smoothed tests when the lag parameter is correctly chosen. Note that in this power comparison $p = 1$ and models of Table 3 are all first-order dependence alternatives. However, in practice, p is unknown and hence it is possible that two different lag orders may give conflicting conclusions. It is natural then to question what is the behavior of such tests when a wrong lag order is chosen to compute non-smoothed corresponding statistic.

To answer the question indicated in (c) we have firstly compute the power of $G(4)$, $T_{ST}, T_D, T_H, T_{GCM}$ and T_{GKS} against the nonlinear model (DGP3) for $T = 200$ and $T = 400$. Results at 1% level are presented in Table 4.

Table 4. Power of Non-smoothed Tests for $X_t = \varepsilon_t + 0.8\varepsilon_{t-3}^2$.

	$G(4)$	$T_{ST(1)}$	$T_{D(1)}$	$T_{H(1)}$	T_{GCM}	T_{GKS}
$T = 200$	4.5	2.0	1.9	1.9	1.9	1.8
$T = 400$	46.1	2.1	2.0	1.9	2.8	2.6

Note: Rejection frequencies at 1 per cent level for 2000 iterations of DPG 3.

From Table 4 it seems that the G(4) test dominates other non-smoothed statistics for independence when these tests are computed based on a wrong lag order choice (p) regarding the unknown underlying process. Note further that when the sample size increases, statistics that do not depend on lag order (ie., $G(4)$, T_{GCM} and T_{GKS}) notably improve their power performance (this is particularly relevant for the G(4) test). On the contrary, for the remainder tests (T_{ST}, T_D and T_H), power almost does not vary by changing the sample size.

In sum, the new test has desirable theoretical statistical properties. Some of these properties are shared by other tests for independence, either smoothed or non-smoothed tests. For relatively small sample sizes, our test dominates other tests when prior information on the dependence structure is not available (as it usually occurs), while the opposite happens if the lag dependence structure is known beforehand.

5   Empirical Applications

This section illustrates our test by using the $G(m)$ statistic to explore possible serial dependences in the following well studied daily financial returns: Dow Jones Industrial Average (DJIA), S&P 500 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. S&P 500 ranges from January 1, 1992 to December 31, 2003. Finally 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[4] 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).

Table 5. $G(m)$ Test for Independence for several financial returns

	DJIA	S&P 500	French Franc	German Mark	Canadian Dollar
G(m)	2151.01	152.94	3597.55	3480.00	3980.14
p-value	($<$0.0001)	(0.019)	($<$0.0001)	($<$0.0001)	($<$0.0001)

Note: The G(5) has been used for the S&P500 data set, while the G(6) has been used otherwise.

Provided that $T > 3600$ for all data sets except for the S&P 500 data set, we have computed the $G(m)$ test for $m = 6$ and $m = 5$, respectively[5]. Results for the five data sets are reported in Table 5. As can be observed, the null of independence is rejected at 1 per cent significance level for all studied returns. These results indicate that the DJIA daily stock returns, the S&P 500

---

[4]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). Hong and White (2005) have explored for possible nonlinear serial dependence in the daily S&P 500 stock price index.

[5]See subsection 3.2.

return index, and the three exchange rate time series do not follow a random walk[6]. However such rejection of the random walk hypothesis might be due to potential strong volatility clustering (see Ding et *al.*, 1993).

Given that volatility clustering might be present in stock returns (and then may have contributed to the rejection of the random walk hypothesis), we fit an ARMA(p,q)-GARCH(1,1) model to each time series under consideration. To this end the quasi-maximum likelihood estimation (QMLE) method has been used, and the lags have been selected according to the optimizer of Schwarz's Bayesian Criterion. The corresponding significant estimated models are presented in Appendix 2. We now apply the $G(m)$ test to the standardized residuals $\{\hat{\varepsilon}_t\}_{t=1}^T$ of those models that, according to our estimations, have significant estimated parameters. We note that like test procedures described in section 4, our theory is based on observed raw data rather than estimated residuals, however it is plausible that parameter estimation uncertainty has no impact asymptotically, as the symbols' hypothetical distribution does not contain additional unknowns parameters. Moreover, to counter the problem of potential changes in the finite sample distribution due to test over standardized residuals, we simulate an empirical distribution of $G(m)$ statistic by first generating the true ARMA(p,q)-GARCH(1,1) process, then estimating a correctly specified model on the generated data, and then we calculate the $G(m)$ statistic of the standardized residuals. Finally, we compute the p-value (denoted by p^*-value) as the fraction of the simulations runs that gave a $G(m)$ value greater than that obtained by the ARMA(p,q)-GARCH(1,1) standardized residuals, $\{\hat{\varepsilon}_t\}_{t=1}^T$ (from the original returns under study). Results are reported in Table 6, where p^*-values of the $G(m)$ test are given in parenthesis.

Table 6. G(m) test for Independence for Standardized Residuals

	S&P 500	French Frank	German Mark	Canadian Dollar
G(m) test	153.84	725.22	784.75	711.65
p^*-value	(0.016)	(0.429)	(0.047)	(0.601)

Note: The G(5) has been used for S&P500 residuals, while the G(6) has been used otherwise.

Our results show that the ARMA(1,1)-GARCH(1,1) model cannot (at the 5 per cent level) fully capture the dynamics of the S&P 500 daily returns. This is consistent with Hong and White's (2005) results which also included a similar conclusion by using Robinson's (1991) test. This empirical finding suggests that either higher-order conditional moments are time-varying, or conditional mean and variance are misspecified. The same occurs for the German Mark daily returns. On the other hand, it seems that rejection of the random walk hypothesis for the French Franc and the Canadian Dollar daily returns was due to clustering volatility[7].

6 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

---

[6]It is also interesting to comment that the same conclusions are obtained by fixing $m = 4$ and $m = 5$. Although not reported in the paper, these results are available from the authors.

[7]Interestingly, the same conclusions are obtained by using standard p-values, i.e. considering that parameter estimation uncertainty has no impact on the distribution of the G(m) statistic, given the sample sizes here considered.

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 tests 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 (standard) of an affine transformation of the permutation entropy under the null of independence. The theoretical distributions allows us to construct a test for independence which is consistent against a broad class of serial dependences (including those with zero autocorrelation). 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 monotonous transformations of data. Invariance makes our procedure very attractive in practice. Most importantly, our test makes no assumptions about the continuous or discrete nature of the data generating process and of its marginal densities. In sum, it is more general than other entropy-based tests. Two final advantages are its computational simplicity and hence its short running computational times.

An empirical application to daily Dow Jones Industrial Average price index; to S&P 500 index; and to three daily exchange rate returns has illustrated our approach by testing the random walk hypothesis on these returns.

7 Appendix 1: Proofs

Proof of Theorem 3.2

Proof. First notice that the estimator $\widehat{h}(m) = - \sum_{\pi \in S_m} \widehat{p}_\pi \ln(\widehat{p}_\pi)$ of $h(m)$, where $\widehat{p}_\pi = \frac{n_\pi}{T-m+1}$, is consistent because $p \lim_{T \to \infty} \widehat{p}_\pi = p_\pi$ for every stationary process (see Bandt and Pompe (2002)), and hence

$$p \lim_{T \to \infty} \widehat{h}(m) = h(m). \tag{24}$$

Recall that $G(m) = 2(T - m + 1)[\ln(m!) - h(m)]$ and $0 \leq h(m) \leq \ln(m!)$. Denote by $H(m) = \ln(m!) - h(m)$, so $G(m) = 2(T - m + 1)H(m)$. Then by (24) it follows that

$$p \lim_{T \to \infty} \widehat{H}(m) = H(m) \tag{25}$$

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

$$\frac{C}{2(T - m + 1)} < H(m). \tag{26}$$

16

Then, under serial dependence of order $\leq m$, it follows that $H(m) \neq 0$ and thus

$$\Pr[\widehat{G} > C] = \Pr[2(T - m + 1)\widehat{H}(m) > C] =$$
$$\Pr[2(T - m + 1)(\widehat{H}(m) - H(m)) > C - 2(T - m + 1)H(m)] =$$
$$\Pr[2(T - m + 1)(H(m) - \widehat{H}(m)) < 2(T - m + 1)H(m) - C] =$$
$$\Pr[H(m) - \widehat{H}(m) < H(m) - \frac{C}{2(T - m + 1)}]. \tag{27}$$

Therefore, by (25), (26) and (27) we have that

$$\lim_{T \to \infty} \Pr(\widehat{G} > C) = 1$$

as desired. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

8   Appendix 2: ARMA(p,q)-GARCH(1,1) Models

S&P 500 estimated model:

$$X_t = \underset{(0.0001)}{0.0003} + \underset{(0.1043)}{0.8222}X_{t-1} - \underset{(0.0960)}{0.8516}\varepsilon_{t-1} + \hat{\varepsilon}_t\hat{h}_t^{1/2}$$
$$\hat{h}_t = \underset{(1.52E-07)}{5.86E - 07} + \underset{(0.0108)}{0.0662}\hat{h}_{t-1}\hat{\varepsilon}_{t-1}^2 + \underset{(0.0089)}{0.9305}\hat{h}_{t-1}$$

where the number inside the parentheses are robust QMLE standard errors.
French Franc estimated model:

$$X_t = \underset{(7.77\ E-05)}{1.66E - 05} + \underset{(0.0119)}{0.0348}X_{t-1} + \hat{\varepsilon}_t\hat{h}_t^{1/2}$$
$$\hat{h}_t = \underset{(2.4E-09)}{1.47E - 08} + \underset{(0.0201)}{0.1652}\hat{h}_{t-1}\hat{\varepsilon}_{t-1}^2 + \underset{(0.0187)}{0.8742}\hat{h}_{t-1}$$

German Mark estimated model:

$$X_t = \underset{(8.02\ E-05)}{-0.0001} + \underset{(0.0119)}{0.0328}X_{t-1} + \hat{\varepsilon}_t\hat{h}_t^{1/2}$$
$$\hat{h}_t = \underset{(3.9E-07)}{7.37E - 07} + \underset{(0.0201)}{0.0909}\hat{h}_{t-1}\hat{\varepsilon}_{t-1}^2 + \underset{(0.0287)}{0.8942}\hat{h}_{t-1}$$

Canadian Dolar estimated model:

$$X_t = \underset{(3.18\ E-03)}{3.79E - 03} + \underset{(0.0110)}{0.0526}X_{t-1} + \hat{\varepsilon}_t\hat{h}_t^{1/2}$$
$$\hat{h}_t = \underset{(3.9E-07)}{4.13E - 08} + \underset{(0.0083)}{0.0986}\hat{h}_{t-1}\hat{\varepsilon}_{t-1}^2 + \underset{(0.0075)}{0.9003}\hat{h}_{t-1}$$

References

[1] Bandt, C. and Pompe, B. 2002. Permutation entropy - a natural complexity measure for time series. *Physics Review Letters*, 88, 174102.

[2] Blum, J.R., Kiefer, J. and Rosenblatt, M. 1961. Distribution free tests of independence based on the sample distribution function. *Annals of Mathematical Statistics*, 32, 485-498.

[3] 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.

[4] Delgado, M.A. 1996. Testing serial independence using the sample distribution function. *Journal of Time Series Analysis*, 17, 271-285.

[5] 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.

[6] Dufour, J.M., Lepage, Y. and Zeidan, H. 1982. Nonparamentric testing for time series: a bibliography. *The Canadian Journal of Statistics* 10: 1-38.

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

[8] 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*, 15, 371-384.

[9] Granger, C.W.J., Maasoumi, E. and Racine, J. 2004. A dependence metric for possibly nonlinear processes. Journal of Time Series Analysis, 25, 649-669.

[10] Hao, B. and Zheng, W. 1998. *Applied Symbolic Dynamics and Chaos*. World Scientific, Singapore.

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

[12] Hong, Y. 1998. Testing for pairwise serial independence via the empirical distribution function. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 60, 429-453.

[13] Hong, Y. 2000. Testing serial independence via the empirical characteristic function. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 62, 557-574.

[14] Hong, Y. and White, H. 2005. Asymptotic distribution theory for nonparametric entropy measures of serial dependence. *Econometrica* 73, 837–901.

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

[16] Joe, H. (1989a). Relative Entropy Measures of Multivariate Dependence. *Journal of the American Statistical Association*, 84, 157–164.

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

[18] King, M.L., 1987. Testing for autocorrelation in linear regression models: a survey. In : King, M.L. Giles, D.E.A. (Eds), Specification Analysis in the Linear Regression Model. Rutledge and Kegan Paul, London.

18

[19] Lehmann, E.L. 1986. *Testing Statistical hypothesis.* John Wiley & Sons, New York.

[20] Pinkse, J. 1998. A consistent nonparametric test for serial independence. *Journal of Econometrics*, 84, 205-231.

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

[22] Rohatgi, V.K. 1976. *An Introduction to Probability Theory and Mathematical Statistics.* John Wiley & Sons, New York.

[23] 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.

[24] Skaug H. J. and Tjøstheim D. (1993a). Nonparametric test of serial independence based on the empirical distribution function. *Biometrika* 80, 591-602.

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

[26] 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.

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

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