A RANDOMNESS TEST FOR FINANCIAL TIME SERIES*

WISTON ADRIÁN RISSO**

ABSTRACT

A randomness test is generated using tools from symbolic dynamics, and the theory of communication. The new thing is that neither normal distribution nor symmetric probability distribution, nor variance process is necessary to be assumed. Even more, traditional independent identically normal white noise is nested. It also could be useful when signs of time series are more accurate than magnitude. The statistic is tested in stock asset returns rejecting randomness more times than Runs Test, Variance-Ratio Test and ADF.

Keywords: Random Walk Model, Finance, Shannon Entropy, Unit Root Test.
JEL classification: C12, C15, G12

---

* I would like to acknowledge J. Doyne Farmer, J. Gabriel Brida, Lionello Punzo, and Roberto Renò, and for their very helpful comments, and suggestions.

** Department of Economics - University of Siena, Italy. E-mail adress: risso@unisi.it, Tel.: +39 0577 235058, Fax: +39 0577 232661

INTRODUCTION

Based on Symbolic Dynamics and Information Theory, present work will introduce a manner of testing independence in time series trying to show its advantages. Such a test will be applied to financial time series. As it is well known Bachelier (1900) was the first proposing stock prices follow a Brownian motion. This conception presupposes that stock prices reflect all available information. In fact according to the efficient market hypothesis suggested by Fama (1965) the present prices are the best prediction about future prices. Since then, difference in stock prices has been modeled as white noise process, implying that stock prices are random walk processes. This assumption was criticized due to the existence of some well known stylized facts, Mandelbrot (1963) considers that financial returns have a long memory, then stock prices should be modeled using a fractal Brownian motion. Peters (1994) (1996) gives evidence of fractability in financial markets using Hurst exponent as a measure of persistence. Moreover, it seems that constant variance hypothesis in financial returns is not supported by empirical evidence. In fact Mandelbrot (1963) proposes an stable paretian distribution in order to model the asset returns which implies an infinite variance. Since then, different models have been proposed permitting the variance to change, an example is the ARCH model (see Engle (1982)), and the Markov switching models (see Hamilton (1989)). Lo and McKinlay (1988), using the variance ratio test, found that financial returns behavior would not be random. Singal (2004) suggests the existence of different anomalies, reviewing all the anomalies found until now. In this context, the present work has two principal objectives. At first, it will be shown that daily asset returns do not behave as any type of white noise process, and then random walk is a bad model for stock prices. The latter is done, taking the daily decreases and increases and seeing the behavior of the combination for 2, 3, 4, and 5 day decreases and increases. Since theory says that returns are completely random, combinations of different day decreases and increases in prices should have the same probability among them. It means it would not be expected to find combinations of decreases and increases more probable than others. For instance imagine that 0 means decrease in one day (negative returns) and 1 means increase in stock prices (positive returns). Imagine also that we have the following daily time series of codified returns:

01001101000101101010011000111011

There are 32 days of decreases and increases, if the process in completely random in 1 day the probability of decreases should be 1/2 =16/32 and the same for increasing. Moreover if the process is random, in 2 days we have 4 possibilities and the probability should be 1/4 for each possible case. Reasoning in this manner, the probability of an event composed by combination of *n* days should be 2^{-n}, in case of having a random process. In order to do this a test of randomness is developed based in the symbolic dynamic and the concept of entropy. The developing of a test of randomness not only for asset returns but also a general test of randomness will be the second objective. In fact the test of randomness does not need the assumption of normality, and it permits the variance to follow different processes, like a GARCH process, or even an infinite variance like in the case of the paretian distributions suggested by Mandelbrot (1963). Even more, since introduced test is similar to Run-test (when using 2 symbols) advantages suggested by Moore and Wallis (1943) are applied. They highlight that test based only on signs could be useful when time series magnitude is not such accurate as the time series sign. Section 2 explains what symbolic time series analysis and symbolic dynamic are, in section 3 the random walk model will be expressed in a 2 symbol dynamical model. Section 4 explains Shannon Entropy as a measure of uncertainty, section 5 proceeds to construct the randomness test using 2 symbols. As a further results, in section 6 introduced test is compared with others, and an AR(1) process is applied to the daily returns to check if residuals are random, and finally section 7 draws the conclusions and presents some future lines of studying.

SYMBOLIC DYNAMICS AND SYMBOLIC ANALYSIS

Models such as ARMA(p,q) do not have problems detecting linear dependence. When the observed dynamics are relatively simple, such as sinusoidal periodicities, traditional analytical tools such as Fourier transforms are easily used to characterize the patterns. More complex dynamics, such as bifurcation and chaotic oscillation, can require more sophisticated approaches.

Symbolic Dynamics as remarked by Williams (2004) have evolved as a tool for analyzing dynamical systems by discretizing spaces. In fact, Symbolic Dynamics is a method for studying nonlinear discrete-time systems by taking a previously codified trajectory using sequence of symbols from a finite set (alphabet). Consider $\{x_1,x_2,...,x_\infty\}$ is an infinite sequence of continuous variables belonging to R, selecting a partition in the continuous space and so an alphabet $A\equiv\{a_1,a_2,...,a_n\}$ we can analyze the process in a discrete space S where $\{s_1,s_2,...,s_\infty\}$ is an infinite discrete sequence. If the alphabet is well defined we can obtain rich dynamical information (qualitative) analyzing in the discrete space. Such analysis could be very difficult or even impossible in a continuous space.

Piccardi (2004) highlights that symbolic dynamics should be differentiated from symbolic analysis. The former denotes theoretical investigation on dynamical systems. The latter is suggested when data are characterized by low degree of precision. The idea in Symbolic Analysis is that discretizing the data with the right partition we obtain a symbolic sequence. This sequence is able to detect the very dynamic of the process when data are highly affected by noise. Again here the idea is to obtain rich qualitative information from data using statistical tools.

SYMBOLIZATION OF THE RANDOM WALK

In order to clarify how Symbolic Dynamics works and to apply the theory to Financial Analysis we shall try to express wellknown stock price model in terms of symbolic dynamic models. As it was mentioned, Bachelier (1900) and others proposed that stock market prices behaved as a random walk process. It means that prices follow equation (1)

$$P_t = P_{t-1} + \varepsilon_t \text{ where } \varepsilon_t \text{ dist. } i.i.d\ N(0,1) \quad \textbf{(1)}$$

This is a famous model which tries to capture the Efficient Market Hypothesis (EMH) in the weakly form proposed by Fama (1965). It means that in a perfect informed stock market it is impossible to predict future returns using past price information and the returns are independent random variables.

$$r_t \text{ dist. } i.i.d(0,\sigma^2) \quad \textbf{(2)}$$

Assuming that asset returns follows (2) and that $f(r_t)$ is the density function we obtain a stochastic model for financial returns. Using Symbolic Dynamics approach we can capture the qualitative essence of the process it means the independence. Let us take an alphabet $A\equiv\{0,1\}$ with 2 symbols we can discretized the continuous space in the following way:

$$\mathbf{s_t} = \begin{cases} \mathbf{0} \ \ \textbf{if} \ \ \mathbf{r_t} < \mathbf{0} \\ \mathbf{1} \ \ \textbf{if} \ \ \mathbf{r_t} \geq \mathbf{0} \end{cases} \tag{3}$$

Now the process is Bernoulli and the following is the probability function:

$$\mathbf{P(s)} = \begin{cases} \mathbf{1/2} \ \ \textbf{if} \ \ \mathbf{s} = \mathbf{0} \\ \mathbf{1/2} \ \ \textbf{if} \ \ \mathbf{s} = \mathbf{1} \\ \mathbf{0} \ \ \textbf{otherwise} \end{cases} \tag{4}$$

Hence *P(0)=P(1)=1/2*, no symbol is the most probable, and the process is completely random. In fact, since the process is independent history does not matter. In order to explain the latter let us consider a symbolic sequence $S_\ell \equiv \{s_1 s_2 s_3 \ldots s_\ell\}$ □ A^ℓ and define (for simplicity) a history $h_{\ell\text{-}1} \equiv \{s_1 s_2 \ldots s_{\ell\text{-}1}\}$ □ $A^{\ell\text{-}1}$, then consider the set of all the possible histories $\left\{h^i_{\ell-1}\right\}_{i=1}^{i=2^{\ell-1}}$. Since the process is independent $P(s_\ell / h^i_{\ell-1}) = P(s_\ell / h^j_{\ell-1}) = P(s_\ell) = 1/2 \ \forall i, j, s_\ell$. No matter what happened in the past probability of event remain the same. No word, no subsequence commands the dynamics. Taking all possible subsequence of length ℓ, $\left\{s^i_\ell\right\}_{i=1}^{i=2^\ell}$ then $P(s^i_\ell) = P(s^j_\ell) = 2^{-\ell} \ \ \forall i, j$. Computing Normalized Shannon Entropy (*H*) as a measure of randomness (as we shall explain later) this process will produce the maximum, $H(P(s^i_\ell) \ = 1$.

Note that we could modify the model in order to consider certain cycles or particular sequences appearing more frequently. For instance consider that we try to model daily dynamics of stock returns and the weekend effect is true. As remarked by Singal (2004) weekend effect refers to relatively large returns on Fridays compared to those on Mondays. Therefore 2 days sequences *(1,0)* should have a frequency larger than P(1)P(0)=1/4 because de effect *(Friday, Monday)=(1,0)*. More precisely when *ℓ=2* we have *P(1,0)>P(0,0)=P(1,0)=P(1,1)*. In fact the test we shall develop is able to tell us which patterns are causing the anomaly if the event is not random. Suppose that monthly data are used and the January effect is true as asserted by Singal (2004) then 2 month sequences *(0,1)* should be more frequent that the other sequences. It means since January returns are larger than December returns one should expect that sequence *(December, January)=(0,1)* will happen more frequently.

SHANNON ENTROPY AS A MEASURE OF UNCERTAINTY

Clausius (1865) introduces the concept of entropy as a measure of the amount of energy in a thermodynamic system. Shannon (1948) considers entropy as a useful measure of uncertainty in the context of communication theory where a completely random process takes the maximum value. For instance, let us consider the English language as a nonlinear process. Some combination of letters appears more frequently than others. In fact English is not random but a complex process. Taking a page from an English Books combinations of letters such as "*THE*" shall appear more frequently than "*XCV*"[1]. Note that a random language should produce "*THE*" and "*XCV*" with the same probability. Hence Shannon Entropy will compute a value for English language less than the maximum. This idea is fundamental in the present work because if symbolized time series are random process should produce also the maximum entropy otherwise time series are not random.

Let us introduce the required properties of an entropy measure:

It should be a function of $P=(p_1,p_2,...,p_n)$ in this manner it is possible to write $H=H(p_1,p_2,...,p_n)=H(P)$, where P is probability distribution of the events.

It should be a continuous function of $p_1,p_2,...,p_n$. Small changes in $p_1,p_2,...,p_n$ should cause small change in H_n.

It should not change when the outcomes are rearranged among themselves.

It should not change if an impossible outcome is added to the probability scheme.

It should be minimum and possibly zero when there is no uncertainty.

It should be maximum when there is maximum uncertainty which arises when the outcomes are equally likely so that H_n should be maximum when $p_1=p_2=...=P_n=1/n$.

---

1 According to Shannon (1951) the English word "THE" has a probability of 0.071, the next more frequent word "OF" has a probability of 0.034.

The maximum value of H_n should increase as n increases.

Shannon (1948) suggested the following measure:

$$H_n(p_1,p_2,\ldots,p_n) = -\sum p_{i.} log_2(p_i) \quad \textbf{(5)}$$

Logarithms to base 2 are used then entropy is measured in bits. This measure satisfies all properties mentioned above and takes the maximum when all events are equally likely. The latter is easily to confirm by solving the Lagrange equation (6).

$$-\sum p_{i.} \log_2 p_i - \lambda(\sum p_i = 1) \quad (6)$$

Since the function is concave its local maximum is also a global maximum, this is consistent with Laplace's principle of insufficient reason that unless there is information to the contrary, all outcomes should be considered equally probably. Note also that when $p_i=0$ then $0.log0=0$ which is proved by continuity since $x.logx \to 0$ as $x \to 0$. Thus adding zero probability terms does not change the entropy value.

In order to clarify the concept of Shannon, let's take an event with two possibilities and their respective probabilities p and $q=1-p$. The Shannon Entropy will be defined by (7)

$$H=-(p.log(p)+q.log(q)) \quad \textbf{(7)}$$

Figure (1) shows graphically the function shape, note that the maximum is obtained when the probability is 0.5 for each event. This case corresponds to a random event (like flipping a coin), on the other hand, note that a certain event (when probability of one event is 1) will produce entropy equal to 1.

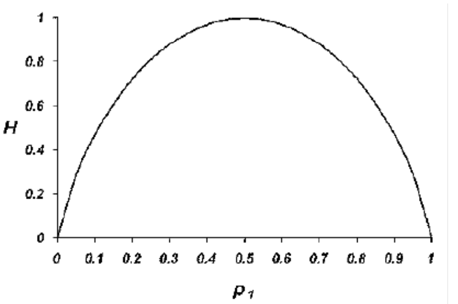

Figure 1. Shape of the Shannon entropy Function. Note that maximum happens when the process is random (p=0.5)

In general, Khinchin (1957) showed that any measure satisfying all the properties must take the following form:

$$-k\sum p_{i.} log_2 p_i \qquad \textbf{(8)}$$

Where k is an arbitrary constant. In particular it is possible to take $k=1/log_2(n)$, which will be useful comparing events of different lengths. This is also known as the Normalized Shannon Entropy.

CONSTRUCTION OF THE RANDOMNESS TEST (R) USING 2 SYMBOLS

Using 2 symbols a random process should be Bernoulli with probability 1/2 for each result as it was shown in section 3. Therefore Normalized Shannon Entropy (H) will be compute for small samples using Monte Carlo simulations.

In theory a completely random process should produce $H=1$, however since the sample is finite of size T, H will follow a distribution depending on T with most of the probability concentrated on 1 (the maximum H value). In the present study 10,000 time series (taking values 0 or 1 with the same probability) of size T are simulated. After the 10,000 H of size T are computed, it is defined variable $R=1-H$ and the simulated distribution of R is obtained. The reason of defining R is to obtain most of the probability in value 0 instead of 1.

Note that no probability distribution is assumed, and no assumption about variance are considered. This is a general test for completely random events.

OBTAINING THE R-STATISTIC FROM THE DATA

This is the most important part because Symbolic Analysis matters. Consider a time series of size T is obtained for the continuous random variable *r(t)*. Let us consider that μ is the mean, then values above and below this threshold should have the same probability. It is possible to define symbolic time series as in (9).

$$\mathbf{s(t)} = \begin{cases} \mathbf{0} & \textbf{if} \quad \mathbf{r(t) < \mu} \\ \mathbf{1} & \textbf{if} \quad \mathbf{r(t) > \mu} \end{cases} \tag{9}$$

Once the symbolic sequence is obtained, different subsequences are defined and H and respective R-statistics are computed. Finally under null hypothesis of randomness, H_0)R=0 the R-statistic is compared with critical value at 95%, if R-statistic is larger than critical value from the Simulated Distribution, null hypothesis is rejected.

TESTING INDEPENDENCE IN ASSET RETURNS

Daily Data

Different data series from NYSE were obtained for more than 10,500 days of asset returns starting on January 1962[2], symbolization is made as in (9). Then we have two possibilities in one day, returns above or bellow the mean. If random walk hypothesis is true, probability of each event should be near 0.5 obtaining a maximum entropy or *R-statistic=0*. Of course is the process is independent combinations of 2, 3 or more days should produce maximum entropy as well since all combinations are equally probable[3]. In this manner 10,000 H were obtained for 1 day, 2, 3, 4, and 5 consecutive

2 Data were obtained from *finance.yahoo.com*

3 In general, taking n consecutive days of independent events the possibilities increase at the rate of 2^n and probability for each possibility is 2^{-n} always producing a maximum entropy.

days (where $T=10,500$) and then 10,000 $R=1-H$ were computed. Figure 2) shows the simulated distribution for combinations of 2 days. Note that most of the probability is accumulated near 0 which correspond to H=1 (a completely random process).

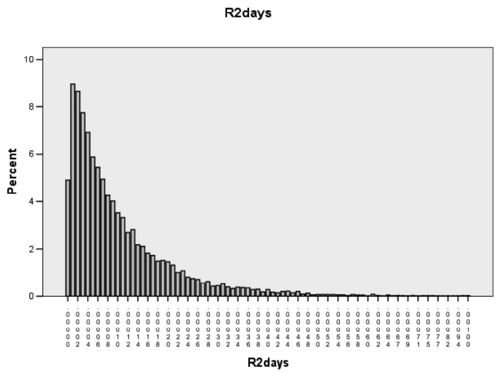

Figure 2. Empirical density function for 2 consecutive moments when T=10,500

Data series from the S&P 500, Dow Jones, and the 10 year treasure notes interest rate were obtained. Taking 10,500 daily data for 11 asset returns, the 10 years treasure note interest rate difference, the Dow Jones, and the S&P 500 index differences, and codifying the series are obtained.

Table 1 shows the critical value a 95% of the Monte Carlo Simulations and Table 2 present the R-statistic for the different asset returns. Note that all the R-statistic values are greater than critical values rejecting the null hypothesis that financial returns are completely random, for instance after discounting the mean the process is still no random.

TABLE 1: Critical Value at 95% for R-Statistic (T=10,500)

R-1 day	R-2 days	R-3 days	R-4 days	R-5 days
0.00026	0.00032	0.00040	0.00054	0.00075

Source: Based on the obtained results

TABLE 2: Test of Randomness (R=1-H) Using the Mean as Partition (10,500 days)					
Financial Returns	**R-1 day**	**R-2 days**	**R-3 days**	**R-4 days**	**R-5 days**
Alcoa Inc.	0.0047	0.0064	0.0070	0.0074	0.0079
Boeing Co.	0.0063	0.0076	0.0086	0.0092	0.0099
Caterpillar Inc.	0.0039	0.0058	0.0066	0.0070	0.0073
Coca Cola Co.	0.0025	0.0029	0.0031	0.0032	0.0033
Du Pont El	0.0044	0.0045	0.0046	0.0047	0.0048
Eastman Kodak Co.	0.0036	0.0038	0.0040	0.0042	0.0045
General Electric Co.	0.0021	0.0022	0.0025	0.0028	0.0030
General Motors Co.	0.0051	0.0054	0.0059	0.0063	0.0068
Hewlett Packard Co.	0.0017	0.0022	0.0027	0.0030	0.0035
IBM	0.0010	0.0010	0.0011	0.0012	0.0014
Walt Disney Co.	0.0027	0.0044	0.0053	0.0061	0.0067
S&P 500	0.0001	0.0021	0.0030	0.0036	0.0041
Dow Jones	0.0000	0.0008	0.0012	0.0016	0.0020
10 years treasure notes	0.0133	0.0182	0.0200	0.0208	0.0215
Source: Based on the obtained results. Asset prices adjusted by splits were considered					

Results disagree with Coulliard and Davison (2005) who do not reject randomness for IBM, General Electric Co., and S&P 500 using daily data. The most frequent sequences are *[0,0], [0,0,0], [0,0,0,0],* and *[0,0,0,0,0]* in almost all the cases (S&P 500 is the exception presenting *[1,1], [1,1,1], [1,1,1,1],* and *[0,0,1,1,1]* as the most frequent patterns). This reflects persistence in remaining at the same regime or it could suggest the existence of autocorrelation.

Weekly Data

Weekly data is also considered for T=2,000 weeks. Hence 10,000 Monte Carlo simulations were made in order to obtain the critical values at 95% as presented in Table 3.

TABLE 3: Critical Value at 95% for R-Statistic (T=2,000)				
R-1 week	**R-2 weeks**	**R-3 weeks**	**R-4 weeks**	**R-5 weeks**
0.00133	0.00167	0.00214	0.00283	0.00397
Source: Based on the obtained results				

Symbolizing and computing *R-statistic* for the asset returns randomness null hypothesis is tested. Table 4 shows the *R-statistics* for the weekly data.

TABLE 4: Test of Randomness (R=1-H) Using the Mean as Partition (2,000 weeks)					
Financial Returns	**R-1 week**	**R-2 weeks**	**R-3 weeks**	**R-4 weeks**	**R-5 weeks**
Alcoa Inc.	0.0012	0.0013	0.0015	0.0021	0.0028
Boeing Co.	0.0025*	0.0027*	0.0031*	0.0035*	0.0039*
Caterpillar Inc.	0.0027*	0.0036*	0.0041*	0.0046*	0.0053*
Coca Cola Co.	0.0042*	0.0062*	0.0070*	0.0075*	0.0080*
Du Pont EI	0.0008	0.0010	0.0011	0.0013	0.0020
Eastman Kodak Co.	0.0008	0.0011	0.0017	0.0024	0.0033*
General Electric Co.	0.0001	0.0006	0.0011	0.0027	0.0042*
General Motors Co.	0.0004	0.0006	0.0010	0.0016	0.0025
Hewlett Packard Co.	0.0003	0.0005	0.0007	0.0015	0.0023
IBM	0.0023*	0.0028*	0.0047*	0.0060*	0.0074*
Walt Disney Co.	0.0005	0.0006	0.0016	0.0026	0.0035
S&P 500	0.0038*	0.0039*	0.0039*	0.0041*	0.0047*
Dow Jones	0.0010	0.0010	0.0013	0.0016	0.0022
10 years treasure notes	0.0004	0.0024*	0.0041*	0.0055*	0.0069*
Source: Based on the obtained results. Asset prices adjusted by splits were considered					

Note that independence is rejected for index S&P 500 and 10 years treasure notes, but Dow Jones. Boeing Co., Caterpillar Inc., Coca Cola Co., General Electric Co., and IBM also reject null hypothesis. These assets are independent according to test developed by Coulliard and Davison (2005). In this case persistence also seems to be the cause of not being independent. Here the most frequent patterns are *[1,1], [1,1,1], [1,1,1,1],* and *[1,1,1,1,1].*

Monthly Data

Obtaining data of 500 months the same procedure is applied. Critical values and R-statistics are computed as shown in Tables 5 and 6.

TABLE 5: Critical Value at 95% for R-Statistic (T=500)				
R-1 month	**R-2 months**	**R-3 months**	**R-4 months**	**R-5 months**
0.0051	0.00674	0.00859	0.01148	0.01623
Source: Based on the obtained results				

TABLE 6: Test of Randomness (R=1-H) Using the Mean as Partition (500 days)					
Financial Returns	**R-1 month**	**R-2 months**	**R-3 months**	**R-4 months**	**R-5 months**
Alcoa Inc.	0.0046	0.0044	0.0045	0.0061	0.0085
Boeing Co.	0.0033	0.0032	0.0041	0.0084	0.0156
Caterpillar Inc.	0.0037	0.0038	0.0039	0.0062	0.0089
Coca Cola Co.	0.0195*	0.0199*	0.0207*	0.0244*	0.0293*
Du Pont EI	0.0014	0.0017	0.0020	0.0026	0.0049
Eastman Kodak Co.	0.0056*	0.0056	0.0058	0.0071	0.0114
General Electric Co.	0.0014	0.0013	0.0015	0.0037	0.0064
General Motors Co.	0.0004	0.0015	0.0024	0.0046	0.0090
Hewlett Packard Co.	0.0017	0.0017	0.0026	0.0036	0.0063
IBM	0.0014	0.0013	0.0017	0.0021	0.0033
Walt Disney Co.	0.0176*	0.0181*	0.0187*	0.0202*	0.0226*
S&P 500	0.0026	0.0031	0.0038	0.0058	0.0088
Dow Jones	0.0012	0.0013	0.0016	0.0038	0.0094
10 years treasure notes	0.0006	0.0020	0.0035	0.0045	0.0077
Source: Based on the obtained results. Asset prices adjusted by splits were considered					

Note that only Coca Cola Co. and Walt Disney Co. reject independence hypothesis, Eastman Kodak would be not independent when considering 1 month sequences. Also here, persistence is the cause of determinism, the sequences *[1,1], [1,1,1], [1,1,1,1],* and *[1,1,1,1,1]* are the most frequents in Coca Cola Co. and Walt Disney Co. Even if expected value of sequence *[1,1,1,1,1]* is 1/32=0.03125 their frequencies are 0.09 and 0.06 respectively.

FURTHER RESULTS

Comparison With Other Tests

Performance of the test for 2 symbols is compared with other unit root test (ADF, Variance Ratio Test, and Runs Test). Considering daily data introduced R-statistic test is able to reject independence in all cases. However Runs test which could be similar to the present test when taking 2 symbols, only rejects the hypothesis for 2 cases, IBM and Kodak. Variance Ratio Test by Lo and MacKinlay (1988) rejects hypothesis for 11 confirming its great performance, while ADF does not reject stationarity in the series. Then Shannon Entropy seems to be powerful detecting nonlinearities and complexities in time series that are not detected by other statistics.

TABLE 7: Different Tests applied to daily data								
Asset Returns	ADF(a)		Variance Ratio Test (b)		Run Test (c)		R-statistic (d)	
	$t_{5\%}$	p-value	$VR_{q=16}$	Sign-Level	Z	Asymp. Sign	R3	CV at 5%
Alcoa Inc.	-96.7085	0.0001	-2079.0615	0.00000	-7.3258	0.0000	0.0070	0.0004*
Boeing Co.	-98.8113	0.0001	0.3657	0.71459*	-5.4980	0.0000	0.0086	0.0004*
Caterpillar Inc.	-97.6613	0.0001	-0.6664	0.50513*	-6.5561	0.0000	0.0066	0.0004*
Coca Cola Co.	-103.3078	0.0001	-1.8321	0.06693*	-2.4787	0.0132	0.0031	0.0004*
Du Pont EI	-101.9246	0.0001	0.3787	0.70491*	-5.4873	0.0000	0.0046	0.0004*
Eastman Kodak Co.	-101.8526	0.0001	-1.7386	0.08210*	-2.4732	0.0134	0.0040	0.0004*
General Electric Co.	-102.2102	0.0001	-2.0753	0.03795*	-2.2889	0.0221	0.0025	0.0004*
General Motors Co.	-74.7423	0.0001	-1.3572	0.17471*	-2.2387	0.0252	0.0059	0.0004*
Hewlett Packard Co.	-102.1406	0.0001	-1.6939	0.09028*	-3.1500	0.0016	0.0027	0.0004*
IBM	-104.0243	0.0001	-0.4748	0.63496*	-0.2907	0.77131*	0.0011	0.0004*
Walt Disney Co.	-100.6847	0.0001	-1.2775	0.20142*	-1.14584	0.14472*	0.0053	0.0004*
S&P 500	-71.8586	0.0001	0.3900	0.69655*	-12.1798	0.0000	0.0030	0.0004*
Dow Jones	-100.9704	0.0001	0.2311	0.81723*	-7.7490	0.0000	0.0012	0.0004*
10 years treasure notes	-93.5496	0.0001	-3.9651	0.00007	-12.5589	0.0000	0.0200	0.0004*

(a) Augmented Dickey Fuller test using Eviews 4.0.

(b) Adjusted for the possible effect of heteroscedasticity. Eviews 4.0

(c) Using SPSS 13.0

(d) Based on obtained results using MatLab 6.0

RESIDUAL OF AN AR(1)

In subsection 5.2 randomness was rejected when taking daily data. Persistence in a regime or autocorrelations could be the cause. In this part an autoregressive process of order 1 is applied to the daily returns. Equation (10) shows specification of an AR(1).

$$\mathbf{r_t = \alpha_0 + \alpha_1 r_{t-1} + \varepsilon_t} \qquad \textbf{(10)}$$

Where α_1 is expected to be less than 1 (in the case of asset returns, it should be almost 0) and ε_t dist. i.i.d$(0,\sigma^2)$.Residuals of these models are tested in order to see if they are random. Table 8 shows R-statistics for the residuals of the AR(1) models, note that the values are smaller than correspondent in Table 2. Comparing these values with critical values in Table 1, note that only Dow Jones residuals seems to be random, and S&P 500 for sequences smaller than 4 days. This suggests that behavior of stock prices are less random than an index (it means a combination or mix of different stock prices). Note that, even when autocorrelation is considered daily stock returns seems to have still a deterministic component.

TABLE 8: Test of Randomness (R=1-H) on AR(1)-residuals (T=10,499)					
Financial Returns	**R-1 day**	**R-2 days**	**R-3 days**	**R-4 days**	**R-5 days**
Alcoa Inc.	0.00060	0.00060	0.00070	0.00080	0.00110
Boeing Co.	0.00240	0.00260	0.00320	0.00360	0.00410
Caterpillar Inc.	0.00028	0.00032	0.00053	0.00065	0.00086
Coca Cola Co.	0.00090	0.00090	0.00090	0.00100	0.00110
Du Pont EI	0.00170	0.00170	0.00170	0.00180	0.00190
Eastman Kodak Co.	0.00320	0.00330	0.00350	0.00370	0.00400
General Electric Co.	0.00120	0.00130	0.00150	0.00170	0.00200
General Motors Co.	0.00500	0.00530	0.00580	0.00620	0.00670
Hewlett Packard Co.	0.00130	0.00160	0.00210	0.00240	0.00280
IBM	0.00070	0.00070	0.00080	0.00100	0.00120
Walt Disney Co.	0.00060	0.00080	0.00120	0.00170	0.00210
S&P 500	0.00000	0.00007	0.00037	0.00064	0.00090
Dow Jones	0.00001	0.00004	0.00015	0.00034	0.00052
10 years treasure notes	0.00090	0.00090	0.00100	0.00100	0.00120
Source: Based on the obtained results. Asset prices adjusted by splits were considered					

CONCLUSIONS

The present work developed a methodology to test independence in finite sample time series. The test is based on symbolic dynamics and information theory. When taking 2 symbols (negative and positive returns) the test seems to be similar to an old idea of taking signs of a time series an looking at the runs (sequence of price changes of the same sign), if they are not far from a random process it will not be possible to reject randomness hypothesis. However section 5.3 showed that Runs test are different from the one presented here.

There are some advantages of using signs instead of the time series. First, it is not necessary to assume a probability distribution of the variable as it is common in other unit root test as Dickey-Fuller or Variance-Ratio Test. Second, it is not necessary to assume a variance process in the series which is important in some economic data such as asset returns where it have been discussed if they are constant as in Fama (1965), autoregressive as deduced by Engle (1982) or infinite as suggested by Mandelbrot (1963). More importantly, as it is suggested by Moore and Wallis (1943), tests based only on signs could be useful when time series magnitude is not such accurate as time series sign. They remark that problem with runs test is that ideal distribution function is not known using an asymptotic approach based on normality distribution when the sample increases. A different approach is considered in the present work. As it is well known

Normalized Shannon Entropy is interpreted as a measure of randomness in Information Theory. Then using Monte Carlo Simulations it is possible to find simulated distribution in small samples when a process is random and different critical values for different samples can be obtained (see Appendix).

The *R-statistic* was tested for stock asset returns because as it is known they were assumed white noise process. This idea was present in Bachelier (1900) and is connected with the Efficient Market Hypothesis, if the market is efficient there is not place for forecasting. Despite Fama (1965) using Runs test is not able to reject the randomness hypothesis, Fama and French (1988) marked that there is evidence of predictability in returns. Introduced test rejects randomness in all cases when considering daily data showing a better performance than other tests. Randomness is even rejected when considering autocorrelation of first order.

Next step in developing the test will be to assume a different symbolization. In the case of asset returns a 4 symbols test could consider stylized facts such as volatility clustering.

REFERENCES

Bachelier, L., 1900. *"Theory of Speculation",* (in Cootner, P., ed.), The Random Character of Stock Market Prices, Cambridge.

Clausius, R., 1865. *"The nature of the motion we call heat",* (translated in Sthephen G. Brush ed.), Kinetic Theory, 1965.

Coulliard, M., and Davison, M., 2005. *"A Comment on Measuring the Hurst Exponent of Financial Time Series",* Physica A, No. 348, pp. 404-418.

Engle, R. F., 1982. *"Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of the United Kingdom Inflation",* Econometrica, Vol. 50, No. 4, pp. 987-1008.

Fama, E., 1965. *"The Behavior of Stock Market Prices",* Journal of Business, Vol. 38, pp. 34-105.

Fama, E.; French, K., 1988. *"Dividend Yields and Expected Stock Returns",* Journal of Financial Economics, Vol. 22, pp. 3-25.

Hamilton, 1989. *"A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle",* Econometrica, 57, pp. 357-384.

Khinchin, 1957. *"Mathematical Foundations of Information Theory",* Courier Dover Publications.

Lo, A.W., and MacKinlay, A.C., 1988. *"Stock Market Prices do not Follow Random Walks: Evidence from a Simple Specification Test",* The Review of Financial Studies, Vol. 1, No. 1, pp. 41-66.

Mandelbrot, B., 1963. *"The Variation of Certain Speculative Prices",* The Journal of Business, Vol. 36, No. 4, pp. 394-419.

Peters, E., 1994. "*Fractal Market Analysis"*, first edition, Wiley Finances Edition.

Peters, E., 1996. "*Chaos and order in the capital markets"*, second edition, Wiley Finances Edition.

Moore, G.H, and Wallis, W.A., 1943. *"Time Series Significance Tests Based on Signs of Differences",* Journal of the American Statistical Association, Vol. 38, No. 222, pp. 153-164.

Piccardi, C., 2004. *"On the Control of Chaotic System via Symbolic Time Series Analysis",* Chaos, Vol. 14, No. 4, pp. 1026-1034.

Shannon, C., 1948. *"A Mathematical Theory of Communication",* Bell System Technical Journal, Vol. 27: 379-423., pp. 623-656.

Shannon, C., 1951. *"Prediction and Entropy of Printed English",* Bell System Technical Journal, 30,50.

Singal, V., 2004. "*Beyond the Random Walk: A Guide to Stock Market Anomalies and Low-Risk Investing"*, Oxford University Press

Williams, S., 2004. "*Symbolic Dynamics and its Applications"*, Prooceding of Symposia in Applied Mathematics, Vol.60,150 pp.

APPENDIX: Critical Values for different samples

TABLE I: Critical Values at 1%					
Sample Size	**R1**	**R2**	**R3**	**R4**	**R5**
30	0.16340	0.17650	0.14640	0.28170	0.33680
60	0.08170	0.08910	0.09890	0.14130	0.18520
90	0.05190	0.05970	0.06980	0.08920	0.12270
100	0.04930	0.05240	0.06200	0.08040	0.11030
200	0.02350	0.02620	0.03100	0.03940	0.05270
300	0.01560	0.01800	0.02140	0.02660	0.03440
500	0.00910	0.01060	0.01270	0.01580	0.02090
600	0.00770	0.00890	0.01070	0.01330	0.01740
900	0.00510	0.00600	0.00700	0.00870	0.01150
1,000	0.00490	0.00550	0.00630	0.00780	0.01040
2,000	0.00240	0.00260	0.00310	0.00380	0.00510
3,000	0.00160	0.00180	0.00210	0.00260	0.00350
5,000	0.00100	0.00110	0.00130	0.00160	0.00200
6,000	0.00080	0.00090	0.00100	0.00130	0.00170
9,000	0.00060	0.00060	0.00070	0.00090	0.00110
10,500	0.00045	0.00050	0.00059	0.00073	0.00095

TABLE II: Critical Values at 5%					
Sample Size	**R1**	**R2**	**R3**	**R4**	**R5**
30	0.08170	0.11970	0.10620	0.21340	0.28110
60	0.05190	0.05650	0.06960	0.10380	0.14980
90	0.02900	0.03710	0.04820	0.06680	0.09860
100	0.02900	0.03440	0.04360	0.05950	0.08740
200	0.01420	0.01670	0.02120	0.02910	0.04210
300	0.00930	0.01140	0.01470	0.01930	0.02740
500	0.00560	0.00680	0.00860	0.01150	0.01630
600	0.00460	0.00550	0.00720	0.00970	0.01360
900	0.00300	0.00370	0.00470	0.00640	0.00890
1,000	0.00280	0.00330	0.00420	0.00560	0.00800
2,000	0.00140	0.00170	0.00210	0.00280	0.00400
3,000	0.00090	0.00110	0.00140	0.00190	0.00270
5,000	0.00050	0.00070	0.00090	0.00110	0.00160
6,000	0.00050	0.00060	0.00070	0.00090	0.00130
9,000	0.00031	0.00037	0.00047	0.00063	0.00088
10,500	0.00026	0.00032	0.00040	0.00054	0.00076

TABLE III: Critical Values at 10%					
Sample Size	**R1**	**R2**	**R3**	**R4**	**R5**
30	0.05190	0.08920	0.08800	0.18500	0.25610
60	0.02900	0.04250	0.05710	0.08820	0.13310
90	0.02290	0.02840	0.03880	0.05690	0.08680
100	0.01850	0.02600	0.03490	0.05050	0.07670
200	0.01040	0.01240	0.01720	0.02430	0.03660
300	0.00630	0.00860	0.01150	0.01620	0.02400
500	0.00420	0.00520	0.00690	0.00960	0.01410
600	0.00320	0.00410	0.00570	0.00790	0.01170
900	0.00220	0.00280	0.00380	0.00530	0.00780
1,000	0.00200	0.00260	0.00340	0.00480	0.00700
2,000	0.00100	0.00130	0.00170	0.00240	0.00350
3,000	0.00060	0.00080	0.00110	0.00160	0.00230
5,000	0.00040	0.00050	0.00070	0.00090	0.00140
6,000	0.00030	0.00040	0.00060	0.00080	0.00120
9,000	0.00022	0.00028	0.00037	0.00052	0.00076
10,500	0.00018	0.00024	0.00032	0.00045	0.00066