MEASURING THE INFORMATIONAL EFFICIENCY IN THE STOCK MARKET AND ITS ECONOMIC EFFECTS

by

Wiston Adrián Risso Charquero

A Thesis submitted to the faculty of
University of Siena
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Economics

University of Siena

October 2008

ABSTRACT

The purpose of the present dissertation is to study the effect of the informational efficiency in the financial markets. As it is well known, the efficient market hypothesis (EMH) has been the central proposition in finance in the last 30 years. This hypothesis establishes that in an efficient market, the prices always fully reflect all the available information. However, behavioral economics and some empirical evidence challenge this hypothesis, sometimes rejecting it. In this study we highlight that financial markets are basically efficient, however they present long periods of inefficiency.

In the first chapter we present a definition of the EMH realizing a brief discussion about this hypothesis. In the second chapter we introduce a measure of the informational efficiency based on the symbolic dynamics and the Shannon entropy. The intuition is simply, if after symbolization the dynamic of the returns is recovered, then it is possible to apply the Shannon entropy in orden to measure the quantity of embodied information. We applied this measure to some US stock prices and test if randomness is an appropriated hypothesis for the asset returns. However, we find that at a daily frequency they are not completely efficient. Even though we ruled out the autocorrelation in the returns, the residuals suggest the existence of nonlinearity. Many proofs are realized to the statistic, we obtain the simulated distribution of the measure, and under certain assumptions we derive

the approximated distribution of the statistics for a small size sample. The power and size experiments suggest that the test is able to detect many different forms of nonlinearity, in particular it is able to detect the Non Linear Sign Model process when the BDS test cannot.

In the third chapter we study the difference in the informational efficiency levels between emerging and developed markets. We apply the measure introduced in the previous chapter based on symbolic time series analysis and Shannon entropy, in order to measure and rank the informational efficiency of 20 stock markets from July 1, 1997 to December 14, 2007. The results suggest that three Asian markets take the first position as the most efficient (Taiwan, Japan and Singapore). The last positions are taken by the ex-socialist countries, the most inefficient markets. This latter result could be due to the limited experience of these markets. In the fourth chapter, the evolution of the daily informational efficiency is measured for different stock market indices (Japanese, Malaysian, Russian, Mexican, and the US markets) by using the local entropy and the symbolic time series analysis. There is some evidence that for different stock markets, the probability of having a crash increases as the informational efficiency decreases. Further results suggest that this probability also increases for switching to a less efficient market. In addition, the US stock market seems to be the most structurally efficient and the Russian is the most inefficient, perhaps because it is a young market, only established in 1995. This seems to confirm the results obtained in chapter three. The fifth chapter tries to study the informational efficiency across a financial market. It introduces

a new methodology to construct Minimal Spanning Trees (MST) and Hierarchical Trees (HT) using information provided by more than one variable. The method is applied to the US and the Italian market and it detects clusters of companies belonging to the same branch of the economy. This fact provides some evidence of informational efficiency in the market, since the news arrivals in one company affect also the movements in the related companies. In addition, some Monte Carlo simulations of random markets suggest that the obtained trees are significant.

CONTENTS

[1] The models are taken from Lui et. al. (1992)

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

I would like to acknowledge Gabriel Brida, Doyne Farmer, Lionello Punzo, Martín Puchet, Ignacio Perrotini Hernández, Roberto Renò, Ahmad Jafari-Samimi, Murat Karagöz, Piotr Wdowinski, and David Matesanz Gómez. I would like to thank The XXI Economic Meeting of the Central Bank of Uruguay (Uruguay), FindEcon 2007 (Łodz, Poland), 8th Turkish Econometric and Statistical Congress (Malatya, Turkey), Econophysics Colloquium 2007 (Ancona, Italy), UNAM (City of Mexico, Mexico) for the opportunity of presenting part of my work. I would like to thank also to the Journals: *International Journal of Modern Physics C*, *Research in International Business and Finance*, *Physica A*, *Applied Financial Economics Letters* and *Expert Systems With Applications* for permitting to me to publish part of this dissertation. I would like to thank my parents for their love and support, without their help this work would not have been possible.

CHAPTER 1

Introduction

For more than twenty years the Efficient Market Hypothesis (EMH) has been the central proposition in Finance. It states that security prices in financial markets must equal the fundamental values, either because all investors are rational or because arbitrage eliminates pricing anomalies. Fama (1970) defined an efficient financial market as one in which security prices always fully reflect all available information. Even more, in 1978, Michael Jensen, a Chicago graduate and one of the creators of the EMH declared that there is no other proposition in economics which has more solid empirical evidence supporting it than the EMH (Jensen 1978, p. 95). The EMH in its weakly version, establishes that the best prediction of future prices we can make, is to use the present prices. Considering the latter proposition many authors have used the random walk as a stochastic model for asset prices.

Even though that hypothesis has many fundaments, it seems that markets sometimes behave in a inefficient way. The latter is more frequent when we consider stock markets in less developed countries. Actually, emerging markets would be less efficient than developed markets. Another reason to have inefficiency is the existence of anomalies in the stock markets; see Singal (2004) who reviews a series of anomalies in financial markets. On the other hand, recently the behavioral finance challenges EMH refuting the hypothesis that investors are fully rational.

In fact, behavioral finance proposes that agents in the market act on the basis of sentiments.

In the present dissertation my main hypothesis is that *financial markets are not always efficient and we can measure the level of efficiency.* Nowadays there are basically two positions about the efficieny of market. On the one hand, we find the old tradition supporting the EMH, on the other hand the behavioral finance rejects EMH. We try to highlight that sometimes the market is efficient and sometimes it does not. It means, in a market we can observe long periods of efficiency and other periods of high inefficiency. This has been present in the history of the stock makets, from the famous tulip mania in the 17th century, until the more recent bubbles in the technological and real state sectors.

Even more, it is possible to detect and analyze the effects of different levels of efficiency on the economy. For example, a measure of the informational efficiency can be useful for studying its effects in the crash of the stock markets and on monetary policy, or in measuring the efficiency of different markets throughout the world by analyzing the causes of their different levels of efficiency.

As an introduction, the next chapter presents briefly the Efficiency Market Hypothesis (EMH). Since the chapter is introductory, its aim is to define the main concepts in order to understand the next chapters. However, some references are given for who ever may want to study in-depth the topic. In the chapter the definition of the EMH is introduced, the fundaments of the hypothesis given and the challenges to the hypothesis also are provided. Once informational efficiency

is defined, the second chapter develops a measure of informational efficiency based on the concepts of symbolic dynamics and entropy. Actually, this chapter is quite methodological. The symbolic analysis is introduced, and the entropy is defined. The latter is basically the measure of efficiency that will be used in part of the dissertation. The measure is also used as a test of randomness, being able to detect nonlinearity in time series. The measure is applied to some US stock prices and indices[1], showing that they are not completely efficient if a daily frequency is considered. In chapter three, the hypothesis that the *emergent markets are more inefficient than the developed markets* is analyzed. The introduced measure is applied to different stock markets, and a kind of ranking is constructed for them. In particular, it will be studied if, in the last ten years, the new capitalist countries (having the youngest and least experienced stock markets) have achieved levels of efficiency comparable with those of the western European countries.

The fourth chapter will measure the effects of the efficiency in crash events for different stock markets, a further result will show that undeveloped countries have less efficient markets than the developed countries. The levels of efficiency are measured for different time-windows, obtaining time series which show that efficiency is not constant through the time. For example, if the market is developing a bubble, there is a short-run trend which will be detected by the measure. The question here is: *Is it possible to say something about the probability of developing a crash given that we know the levels of efficiency?.*

[1] It is supposed that the US Stock Markets are the most efficient in the world. Therefore, the EMH should take place here more than elsewhere.

In chapter five the structure of stock markets is analyzed (in particular, the U.S. and the Italian stock markets). To this end a new methodology is designed, based on symbolic analysis and graph theory (especially, the Minimal Spanning Tree (MST) and the Hierarchical Tree (HT) are applied). The hypothesis here is that, *if there are News about a particular branch of the economy and the market is efficient, all the firms within that branch should move in the same direction, at the same time.* Therefore, the structure of the stock market should exhibit clusters or groups of firms, all affected by the same information. In particular, the methodology developed is able to consider information from asset returns and volume trading, showing the structure of the financial markets in a graphical way.

CHAPTER 2

The Efficient Market Hypothesis

The Efficient Market Hypothesis (EMH) states that the securities prices in the financial markets must equal the fundamental values, either because all investors are rational or because arbitrage eliminates pricing anomalies. According to Shleifer (2000), the University of Chicago coined the term, becoming the central proposition in finance for nearly thirty years. As said above, Jensen (1978) even declared that there is no other proposition in economics having more solid empirical evidence supporting it than the EMH. Fama (1965) considered that an efficient market is characterized by a large number of rational profit-maximizers actively competing among them to try to predict future market values of assets, and the important current information, is almost freely available to all participants. As consequence of these actors interacting, actual prices of individuals securities already reflect the effects of information, being thus a good predictor of future prices. Samuelson (1965) proposed a mathematical proof about the EMH. He said that the asset prices in an efficient market should fluctuate randomly through time in response to the unanticipated components of news. Actually, Samuelson (1965) and Mandelbrot (1966) proved some of the earliest theorems showing how, in competitive markets with rational risk-neutral investors, the returns are unpredictable and prices follow random walks. Fama (1970) asserts the EMH is thus first and foremost a consequence of equilibrium in competitive markets with fully rational

investors.

2.1 The Three Versions of the EMH

As prices should reflect all available information, there exist three versions of the EMH, and they depend on what is considered as "all available information". In fact, Fama (1970) distinguishes three types of EMH: the weak, semi-strong, and strong forms.

1) *The weak form* affirms that the stock prices already reflect all the information derived by examining historical prices and trading volumes. Therefore, *past prices are useless* and they do not add more information in order to predict future prices. This form considers the random walk as a good model for stock prices.

This idea is diametrically opposed to the belief of chartists or technical analysts, whose views imply a sluggish response by prices to changes in the underlying "fundamental" so that any change in trend can be identified by the price tracing out one of the patterns.

2) *The semi-strong form* asserts that all the *publicly accessible information* about a firm is reflected in its stock prices. Such information includes, in addition to past prices, firms information such as patents held, balance sheet composition, accounting practices, earning predictions, and so on. This version rules out the possibility of stock prices being undervalued or overvalued. If investors were to have access to such information from publicly available sources, one would expect it to be in the prices. This form of EMH challenges the fundamental analysts which are an important group among the Wall Street financial analysts.

3) *The strong form* affirms that *all the information relevant* to the firm is considered, even information only obtainable by company insiders. This is an extreme version, and the consequence here is that even insider information is useless in order to predict prices because it is already included in actual prices.

The basic theoretical case for the EMH rests on three arguments which rely on progressively weaker assumptions. First, investors are assumed to be rational and hence to value securities rationally. Second, to the extent that some investors are not rational, their trades are random and therefore cancel each other out without affecting prices. Third, to the extent that investors are irrational in similar ways, they are met in the market by rational arbitrageurs who eliminate their influence on prices.

2.2 The EMH and the Random Walk

According to Mills (1992), the EMH is the essence of the argument according to which changes in stock prices will be random and unpredictable (i.e. prices follow a random walk). Therefore, considering the EMH in its weak version, many authors have considered the Brownian motion and the random walk as satisfactory models for financial variables as stock prices, the interest rate, and the exchange rate. However, a model that is appropriated, is one in which expected returns are constant, and where the returns sequence is uncorrelated. The model implies that

prices follow a martingale process[1], which is related to, but is rather more general than, a random walk, and this distinction can be important in a more detailed investigation into market efficiency.

In fact, many years before the EMH was defined, Bachelier (1900) who had derived the mathematical theory of Brownian Motion five years before Einstein, had just proposed that stock prices follow a random walk[2] process, but his work was ignored and forgotten for years. Actually, he proposed that prices changes are independent and identically distributed. He thought that fluctuations in prices depended on an infinite number of factors making impossible to aspire to mathematical prediction of them. On the other hand, King (1930) concluded that stock prices resemble accumulation of purely random changes even more strongly than do goods prices. Working (1934) noted that time series commonly possess in many aspects the characteristics of series of cumulated random numbers. For instance, he asserted that wheat prices resemble a random-difference series, in particular one that might be derived by cumulating random numbers drawn from a slightly skewed population of standard deviation varying rather systematically through time.

[1] A martingale is the mathematical model for a fair game, one in which the expected price change (or return) is constant. The term 'martingale' refers in addition to a gambling system, originally popular in the French viallage of Martigues, in which the stake is double after each losing bet.

[2] It is believed that the term was first used in an exchange of correspondence appearing in Nature in 1905 between Karl Pearson and Lord Rayleigh, which provided the answer to the following problem: If one leaves a drunk in an empty field in the dead of the night and wishes to find him some time later while it is still dark, what is the optimal search strategy? It is to start exactly where the drunk had been placed and to walk in a straight line away from that point in any direction you wish.

2.3 Challenges to the EMH

The challenge to the EMH is two-fold; theoretical and empirical. From the theoretical point of view, behavioral finance focused on the issues of limited arbitrage and investor sentiments, see Shleifer (2000) for an introduction to behavioral finance. As mentioned above, the EMH can be justified by the existence of rational agents, but it is difficult to argue that people in general and investors in particular, are fully rational. At the superficial level, many investors react to irrelevant information in forming their demand for securities. Furthermore, Black (1986), for example, asserted that investors trade on noise rather than information. The second line of defense considers that irrational investors may exist and trade randomly, and hence their trades would cancel each other out. However according to Shleifer (2000) the psychological evidence shows precisely that people do not deviate from rationality randomly, they deviate in similar way. Schiller (1984) shows that sometimes the noise traders behave socially and follow each others'mistakes by listening to rumors or imitating their neighbors. Investor sentimental reflects the common judgment errors made by a substantial number of investors, rather than uncorrelated random mistakes. The third line of defense maintains that even if sentiment is correlated across unsophisticated investors, the arbitrageurs should take other side of unsophisticated demand and bring prices back to fundamental values. Ultimately, the theoretical case for efficient markets depends on the effectiveness of such arbitrage. However Shleifer (2000) asserts that, in contrast to the efficient markets theory, the behavioral finance states that real-world arbitrage is

risky and therefore limited.

On the other hand, empirical evidence against the EMH started to appear. Niederhoffer and Osborne (1966) discover that accurate records of stock market ticker prices display striking properties of dependence. There is a general tendency for price reversal between trades. Despite positive evidence for the random walk model, Osborne (1967) affirmed that non-random properties of stock prices are primarily found in short interval data (daily and weekly) and in individual stock prices as opposed to an average. The assumption of normality also was criticized; Osborne (1967) plotted the density function of stock market returns, and labeled the returns "approximately normal" since there were extra observations in the tails of the distribution, a condition that statisticians call "kurtosis". Mandelbrot (1963) suggested that asset returns present a type of distribution belonging to the family of "stable paretian" distributions, which are characterized by undefined, or infinite variance. By that time Mandelbrot and Taylor (1967) recognized three schools of thought on the statistical distribution of stock price differences. First, the most popular approach is "technical analysis" or "Chartism" mentioned above. The other schools agree that sequences of prices describe random walks, where price changes are statistically independent of previous price history, but these schools disagree in their choice of the appropriate probability distributions. Some authors found price changes to be normal or Gaussian, while the other group found them to follow a stable Paretian law with infinite variance. The researchers have identified more ways to successfully predict security returns. For instance, Jegadeesh and

Titman (1993) show that movements in individual stock prices over the period of six to twelve months tend to predict future movements in the same direction. Even Fama (1991) admits that stock returns are predictable from past returns and that this represents a departure from the conclusions reached in earlier studies.

2.4 References

-Bachelier, L., (1900),"Theory of Speculation", (in Cootner, (Ed), *The Random Character of Stock Market Prices*, Cambridge, MA: MIT Press, 1964).

-Fama, E., (1991), "Efficient Capital Markets: II", *The Journal of Finance*, Vol. 46, pp. 1575-1617.

-Fama, E., (1970), "Efficient Capital Markets: A Review of Theory and Empirical Work", *The Journal of Finance*, Vol. 25, pp. 383-417.

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

-Black, F., (1986), "Noise", *The Journal of Finance*, Vol. 41, pp. 529-543.

-Jegadeesh, N., Titman, S., (1993), "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency", *The Journal of Finance*, Vol. 48, pp. 65-91.

-Jensen, M., (1978), "Some anomalous evidence regarding market efficiency", *The Journal of Financial Economics*, Vol. 6, pp. 95-101.

-King, W., (1930), *Index Numbers Elucidated*, Longmans, Green and Co.

-Mandelbrot, B., Taylor, H., (1967), "On the Distribution of Stock Price Differences", *Operations Research*, Vol. 15, no. 6, pp. 1057-1062.

-Mandelbrot, B., (1966), "Forecasts of Future prices, unbiased markets, and martingale models", *The Journal of Business*, Vol. 39, pp. 242-255.

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

-Mills, T., (1992), *Predicting the Unpredictable? Science and Guesswork in Financial Market Forecasting*, The Institute of Economic Affairs (Ed).

-Niederhoffer, V., Osborne, M., (1966), "Market Making and Reversal on the Stock Exchange", *Journal of the American Statistical Association*, Vol. 61, no. 316, pp. 897-916.

-Osborne, M., (1967), "Some Quantitative Tests for Stock Price Generating Models and Trading Folklore", *Journal of the American Association*, Vol. 62, no. 318, pp. 321-340.

-Samuelson, P., (1965), "Proof that properly anticipated prices fluctuate randomly", *Industrial Management Review*, Vol. 6, no. 2, pp. 41-49.

-Schiller, R., (1984), "Stock Prices and Social Dynamics", *Brookings Papers on Economic Activity*, Vol. 2, pp. 457-498.

-Shleifer (2000), *Inefficient Markets: An Introduction to Behavioral Finance*, ed. Oxford University Press Inc.

-Working, H., (1934), "A Random-Difference Series for Use in the Analysis of Time Series", *Journal of the American Statistical Association*, Vol. 29, no. 185, pp. 11-24.

CHAPTER 3

A Measure of Informational Efficiency

3.1 Introduction

The main purpose of the present chapter is to introduce a measure of the information efficiency, based on symbolic dynamics and information theory. This measure can also be considered as a test of independence in time series. Recent papers have tried to measure the informational efficiency by using the Hurst exponent, see Peter (1994), (1996), Grech and Mazur (2004), Coulliard and Davison (2005). However, the use of this measure has been criticized by some scholars, see Bassler et al. (2006) and McCauley et al. (2007). The present chapter basically proposes to apply the Shannon entropy after considering a symbolization of the time series. The intuition is simple, on the one hand, symbolic analysis is useful in detecting the very dynamics of a process when this is highly affected by noise, see Daw et al. (2003) for a review of symbolic analysis. This is the case of asset returns, which seem to be random processes since they are affected by noise, thus a proper symbolization could help to recover the dynamics of the process. On the other hand, Shannon Entropy has been used in Information Theory as a useful measure of dispersion of information, see Shannon (1948) and Cover and Thomas (1991). The idea here is that if after symbolization the dynamics of returns is recovered, then it is possible to use the Shannon entropy in order to measure the amount of embodied information. When the process is completely random no event

is more frequent than another, thus the entropy is maximal. However, if there are more frequent patterns, the entropy is low. As an extreme case of the latter, imagine a stock market where the returns or prices are fixed by the government (for instance, consider an exchange rate regime wherein the exchange rate is fixed to certain value), then the price will be always the same and entropy equal to zero.

Even though empirical evidence is presented, the chapter is quite methodological. It is organized as follows. In section 2 we briefly explain what symbolic analysis and Shannon entropy are. In section 3 we derive the simulated distribution of the statistics under the hypothesis of randomness for two symbols and an empirical application for some US asset returns is also presented. Section 4 derives the simulated distribution of the statistics when we consider four symbols, and some US asset returns are tested. In Section 5 the approximated distribution for the statistic is obtained and it is compared with the simulated distribution. Section 6 presents some experiments in size and power for different nonlinear model and the results are compared with the BDS test. Finally, Section 7 draws some conclusions.

3.2 The Symbolic Analysis and The Shannon Entropy

3.2.1 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 periodicity, 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 $\{x_t\}_{t=1}^{t=\infty}$ is an infinite sequence of continuous variables belonging to $\mathbb{R}$, selecting a partition in the continuous space, and thus an alphabet $A \equiv \{a_1, a_2, ..., a_n\}$ we can analyze the process in a discrete space S, where $\{s_t\}_{t=1}^{t=\infty}$ is an infinite discrete sequence. If the alphabet is well defined we can obtain rich dynamical information (qualitative) by analyzing the data 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.

3.2.2 The 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. However, 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 combinations of letters appear more frequently than others. In fact, English is not random but a complex process. Taking a page from an English Books we can note that the combinations of letters such as "THE" shall appear more frequently than "XCV"[1]. However, a random language should produce "THE" and "XCV" with the same probability. Hence the Shannon entropy will compute a value for English language less than the maximum. This idea is fundamental in the present work because if the symbolized time series behaves as a random process, it should produce also the maximum entropy otherwise the time series is not random.

Let us introduce the required properties of an entropy measure

1. 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.

2. 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 changes in H_n.

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

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

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

[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.

6. 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 $p_1 = p_2 = ... = P_n = 1/n$.

7. The maximum value of H_n should increase as n increases.

Shannon (1948) suggested the following measure:

$$H_n(p_1, p_2, ..., p_n) = -\sum p_i \log_2 p_i \tag{3.1}$$

Since the logarithms to base 2 are used, the 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 (3.2).

$$-\sum p_i \log_2 p_i - \lambda(\sum p_i = 1) \tag{3.2}$$

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 likely. Note also that when $p_i = 0$ then $0.log0 = 0$ which is proved by continuity since $xlogx \to 0$ as $x \to 0$. Thus adding zero probability terms does not change the entropy value.

In order to clarify the concept of Shannon, consider two possible events and their respective probabilities p and $q = 1 - p$. The Shannon entropy will be defined by (3.3)

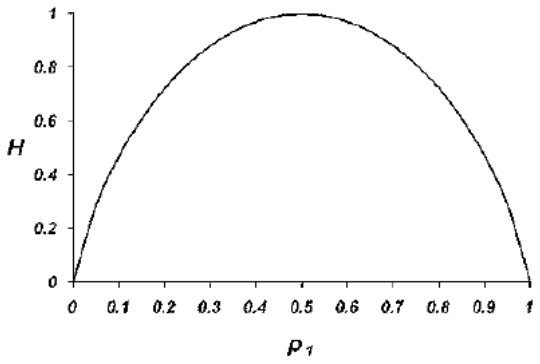


Figure 3.1: Shape of the Shannon entropy function. Note that maximum happens when the process is random (p=0.5)

$$H = -(p.\log(p) + q.\log(q)) \tag{3.3}$$

Figure (3.1) 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 fair coin), on the other hand, note that a certain event (when probability of one event is 1) will produce entropy equal to 0.

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

$$-k \sum p_i \log_2 p_i \tag{3.4}$$

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, since the maximum is always equal to 1.

3.3 Randomness Test using 2 symbols

3.3.1 Introduction

In this section test for independence is derived by using 2 symbols. Note that if the process is random, in a sequence of 2 events there are 4 possibilities and the probability should be 1/4 for each possible case. Reasoning in this manner, the probability of a combination of n events should be 2^{-n}.[2]

As mentioned, when we use 2 symbols, a random process should be Bernoulli with probability 1/2 for each result and normalized Shannnon entropy (H) equal to 1 (this is not discussed). However, when we consider finite samples the probability might be not exactly 1/2 and H can be less than 1. In order to derive the empirical distribution and obtain a critical value in finite samples the probabilty might be not exactly 1/2 and H can be less than 1. In order to derive the empirical distribution and obtain a critical value in finite sample for H we conduct Monte Carlo simulations. First, we simulate 10,000 random time series (of 0s and 1s) sized T by using the generator of pseudo random events provided by MatLab 7.0. Then for each time series we compute the frequency and the associated value of H. Therefore, we define the variable $R = 1 - H$, and the simulated distribution of R is obtained. The reason of defining R is only normalization, it is more manageable to have most of the probability in value 0 instead of 1. Finally, we will have a simulated distribution of R which will depend on T and with most of the proba-

[2] Actually, as will be shown this test for independence does not need the assumption of normality of the events, and permits the variance to follow different processes, like 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 the Run-test (when using 2 symbols), the advantages suggested by Moor and Wallis (1943) can be applied. It means, it can be useful when the magnitude of the time series is not so accurate as the time series sign.

bility concentrated on R=0 (The maximum H value). Note that no probability distribution is assumed, and so assumption about variance is considered. This is a general test for completely independent events. Once we obtain the simulated distribution for the associated T, the critical values are computed in order to compare the R-statistic from real data.

3.3.2 Obtaining the R-statistic from the Data Consider a time series of size T obtained for the continuous variable $r(t)$, for example, a time series of asset returns. Let μ be the mean and values above and below it have the same probability. Then it is possible to define the symbolic time series as in (3.5).

$$s(t) = \begin{cases} 0 & \text{if } r(t)<\mu \\ 1 & \text{if } r(t)>\mu \end{cases} \tag{3.5}$$

Once the symbolic sequence is obtained, different subsequences are defined and the R-statistic is computed. Finally, under the null hypothesis of randomness ($H_0) R = 0$) the R-statistic is compared with a critical value at 95%. If R-statistic is larger than the critical value from the simulated distribution, the null hypothesis is rejected and the process is not independent.

3.3.3 Symbolic Model for the Asset Prices In order to clarify how the symbolic dynamics and the test work, we shall try to express the random walk model for stock prices, in terms of a symbolic dynamic model with 2 symbols. Bachelier (1900) and others proposed that stock market prices behaved as a random walk process. It means that prices follow equation (3.6)

$$P_t = P_{t-1} + \varepsilon_t \text{ where } \varepsilon_t \sim i.i.d\ N(0, \sigma^2) \tag{3.6}$$

It means that in an efficient stock market it is impossible to predict future returns by using the past prices. The present prices immediately incorporates the news and no trend is developed.

$$r_t \sim i.i.d(0, \sigma^2) \tag{3.7}$$

Assuming that asset returns follows (3.7) and that $f(r_t)$ is the density function we obtain a stochastic model for financial returns. Using the symbolic dynamics approach we can capture the qualitative essence of this process, its independence. Let us take an alphabet $A \equiv \{0, 1\}$ with 2 symbols, it is now possible to discretize the continuous space in the following way:

$$s_t = \begin{cases} 0 & \text{if } r_t < 0 \\ 1 & \text{if } r_t \geqslant 0 \end{cases} \tag{3.8}$$

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

$$P(s) = \begin{cases} 1/2 & \text{if } s = 0 \\ 1/2 & \text{if } s = 1 \\ 0 & \text{otherwise} \end{cases} \tag{3.9}$$

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 economics terms, it means that if some news arrive at the market,

the stock prices immediately embody the information. However, in an inefficient market the actual price does not embody all the new information at the moment, permitting the formation of a trend by adjustment. If a trend is developing in the market, some patterns will be more frequents, by instance when a bubble is forming, patterns with many 1s (increases in prices) are more frequent, reducing the probability of having 0s (decreases).

In order to show how history does not matter, be the following symbolic sequence $S_\ell \equiv \{s_1 s_2 s_3 ... s_\ell\} \in A^\ell$ and define (for the sake of simplicity) a history $h_{\ell-1} \equiv \{s_1 s_2 ... s_{\ell-1}\} \in A^{\ell-1}$, 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, the probability of the event remains the same. No word, no subsequence commands the dynamics. Taking all possible subsequences of length ℓ, $\{s^i_\ell\}_{i=1}^{i=2^\ell}$ then $P(s^i_\ell) = P(s^j_l) = 2^{-\ell}\ \forall i, j$. If the Normalized Shannon Entropy (H) as a measure of randomness is computed, this process will produce the maximum, $H(P(s^i_\ell)) = 1$.

3.3.4 Testing Independence in Asset Returns Different data series from the NYSE were used. A dataset of 10,500 days of asset returns starting on January 1962 was obtained[3], the symbolization is applied as in (3.5). Then we have two possibilities in one day, returns above or below the mean. If the random walk hypothesis is true, the probability of either event should be near 0.5 obtaining a maximum entropy or R-statistic=0. Of course, if the process is independent

[3] Data were obtained from finance.yahoo.com

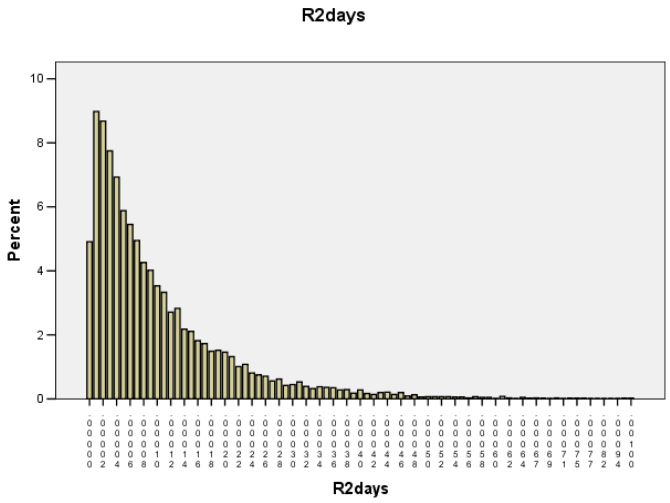


Figure 3.2: Empirical density function for 2 consecutive moments when T=10,500

combinations of 2, 3 or more days should produce maximum entropy as well, since all combinations are equally probable[4]. We simulated 10,000 random time series sized T=10,500 and H was computed (and $R = 1 - H$) for 1 day, 2, 3, 4, and 5 consecutive days. Figure 3.2 shows the simulated distribution of R, for combinations of 2 days. Note that most of the probability is accumulated near 0 which corresponds to $H = 1$ (a completely random process).

After obtaining the simulated distribution, the critical values were computed, Table 1 shows the critical values at 95% of the Monte Carlo simulations.

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, then we symbolize the data series obtaining the respective R-statistics.

[4] 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.

Table 2 presents the R-statistics for different asset returns. Note that all of the R-statistic values (Table 2) are greater than the critical values (Table 1), rejecting the null hypothesis that financial returns are completely random. Therefore after discounting the average returns, the process is still not random, a result that suggests evidence of inefficiency at the daily frequency (See Singal (2004) for a review of all the anomalies in the stock markets known until now).

Table 1

Critical Values 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

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 EI	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*

* Rejection of randomness hypothesis at 5%

Our results disagree with Coulliard and Davison (2005), who do not reject randomness for IBM, General Electric Co., and S&P 500, using daily data.

Studying the cause of the bias toward randomness, we note that 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 else, it suggests the existence of autocorrelation.

3.3.5 Residual of an AR(1) In this subsection an autorregressive process of order 1 is applied to the daily returns in order to eliminate possible autocorrelation, as suggested in the previous subsection. Equation (3.10) shows the AR(1) specification:

$$r_t = \alpha_0 + \alpha_1 r_{t-1} + \varepsilon_t \tag{3.10}$$

Where α_1 is expected to be less than 1 (in the case of asset returns, it should be around 0) and $\varepsilon_t \sim i.i.d(0, \sigma^2)$. The residuals are tested in order to study if they are random. Table 3 shows the R-statistics for the residuals of the AR(1) models, note that the values are smaller than correspondent in Table 2. However, comparing with critical values in Table 1, only the residuals for the Dow Jones seem to be random, and S&P 500 for sequences shorter than 4 days. This suggests that behavior of stock prices is less random than an index (i.e. a combination or mix of different stock prices)[5]. Note that, even when autocorrelation is considered, daily stock returns seem to retain a deterministic component.

[5] It is well known that a linear combination of variables produces an entropy greater than the entropy of the variables separately.

Table 3

2-symbols 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*

* Rejection of randomness hypothesis at 5%

3.3.6 Comparison with Other Tests Table 4 shows the performance of the test for 2 symbols, compared with other unit root tests (ADF, Variance Ratio Test, Runs Test, and BDS). With daily data, the R-statistic test is able to reject independence in all cases. However, the Runs test which seems to be similar to

the present test when taking 2 symbols, only rejects the hypothesis for 2 cases, IBM and Kodak. The Variance Ratio Test by Lo and MacKinlay (1988) rejects the hypothesis for 11, while the ADF does not reject stationarity in the series[6]. The most popular nonlinear test (the BDS test) rejects the null hypothesis of independence in all the cases as well. Thus, the 2-symbols randomness test seems to be as good as the BDS test.

Table 7

Different Tests applied to Daily Data

Financial Returns	*ADF*$_{(a)}$		*Variance Ratio Test*$_{(b)}$		*Run Test*$_{(c)}$		*R-statistic*$_{(d)}$		BDS Test$_{(e)}$
□	$t_{5\%}$	*p-val*	$VR_{q=16}$	*Sig.-Level*	*Z*	*Asymp. Sign*	*R*3	*CV at* 5%	*CV*
Alcoa Inc.	-96.71	0.0001	-2079.06	0.00000	-7.33	0.0000	0.0070	0.0004*	0.0000*
Boeing Co.	-98.81	0.0001	0.3657	0.71459*	-5.50	0.0000	0.0086	0.0004*	0.0000*
Caterpillar Inc.	-97.66	0.0001	-0.6664	0.50513*	-6.56	0.0000	0.0066	0.0004*	0.0000*
Coca Cola Co.	-103.31	0.0001	-1.8321	0.06693*	-2.48	0.0132	0.0031	0.0004*	0.0000*
Du Pont EI	-101.92	0.0001	0.3787	0.70491*	-5.49	0.0000	0.0046	0.0004*	0.0000*
Eastman Kodak Co.	-101.85	0.0001	-1.7386	0.08210*	-2.47	0.0134	0.0040	0.0004*	0.0000*
General Electric Co.	-102.21	0.0001	-2.0753	0.03795*	-2.29	0.0221	0.0025	0.0004*	0.0000*
General Motors Co.	-74.74	0.0001	-1.3572	0.17471*	-2.24	0.0252	0.0059	0.0004*	0.0000*
Hewlett Packard Co.	-102.14	0.0001	-1.6939	0.09028*	-3.15	0.0016	0.0027	0.0004*	0.0000*
IBM	-104.02	0.0001	-0.4748	0.63496*	-0.29	0.7713*	0.0011	0.0004*	0.0000*
Walt Disney Co.	-100.68	0.0001	-1.2775	0.20142*	-1.15	0.1447*	0.0053	0.0004*	0.0000*
S&P 500	-71.86	0.0001	0.3900	0.69655*	-12.18	0.0000	0.0030	0.0004*	0.0000*
Dow Jones	-100.97	0.0001	0.2311	0.81723*	-7.75	0.0000	0.0012	0.0004*	0.0000*
10 years treasure notes	-93.55	0.0001	-3.9651	0.00007	-12.56	0.0000	0.0200	0.0004*	0.0000*

(a). Augmented Dickey Fuller test using EViews4.0.

(b). Adjusted for the possible effect of heteroscedasticity. EViews4.0.

(c). Using SPSS13.0.

(d). Based on own calculations.

(e). Distance (m) is equal to 1.5 and epsilon (ε) is around 0.7.

*. Rejection of the randomness at 5%

[6] Remember the well known bias of this test to accept the unit root hypotesis.

3.4 Test for Independence using 4 symbols

The purpose of this section is to extend and confirm the results obtained in the previous section by using a different symbolization. The symbolic stock price model is extended by using 4 symbols and the previous time series are tested again.

3.4.1 4-Symbols Financial Symbolic Model

We extend the model introduced in section 3 to incorporate the fact that sometimes the absolute values of the returns tend to remain in regimes of high volatility or low volatility, for long time. We select a different alphabet and a different partition, representing the process in symbolic dynamics.

Define an alphabet $A \equiv \{1, 2, 3, 4\}$, where 1 and 4 represent high negative and positive returns, while 2 and 3 are low negative and positive returns. Defining $f(r_t)$ as the return empirical distribution, a transformation from the real space to symbolic space is defined by (3.11)

$$T(r_t) \begin{cases} s_t = 1 & \text{if } r_t \leq f^{-1}_{25\%}(r_t) \\ s_t = 2 & \text{if } f^{-1}_{25\%}(r_t) \leq r_t \leq f^{-1}_{50\%}(r_t) \\ s_t = 3 & \text{if } f^{-1}_{50\%}(r_t) \leq r_t \leq f^{-1}_{75\%}(r_t) \\ s_t = 4 & \text{if } r_t \geq f^{-1}_{75\%}(r_t) \end{cases} \tag{3.11}$$

If the process is completely random the density function is discrete uniform as is suggested in (3.12). Hence passing from one symbol to the other is independent, all the events having the same probability.

$$P(s)\begin{cases} 1/4 & \text{if } s=1 \\ 1/4 & \text{if } s=2 \\ 1/4 & \text{if } s=3 \\ 1/4 & \text{if } s=4 \\ 0 & \text{otherwise} \end{cases} \tag{3.12}$$

Note that by in the terminology introduced in the previous section, here also history does not matter since $P(s_\ell/h^i_{\ell-1}) = P(s_\ell/h^j_{\ell-1}) = P(s_\ell) = 1/4\ \forall i, j, s_\ell$. In order to embody the volatility clustering effect in symbolic dynamics, consider two sub-alphabets, $A_1 \equiv \{1, 4\}$ and $A_2 \equiv \{2, 3\}$, with $A = A_1 \cup A_2$, of course. Hence elements belonging to A_1 correspond to the space of "High Volatility" and the elements from A_2 to "Low Volatility". Define $h1_{\ell-1} \equiv \{s_{1,1}, s_{2,1}, s_{3,1}, ..., s_{\ell-1,1}\} \in A_1^{\ell-1}$, a history of "Low Volatility", and $h2_{\ell-1} \equiv \{s_{1,2}, s_{2,2}, s_{3,2}, ..., s_{\ell-1,2}\} \in A_2^{\ell-1}$, a history of "High Volatility".

Now the process is "History-Dependent", inequations (3.13) and (3.14) model the fact that volatility tends to accumulate in clusters.

$$P(s_{1,\ell}/h1^i_{\ell-1}) > P(s_{1,\ell}/h2^j_{\ell-1}) \quad \forall i, j, s_{1,\ell} \tag{3.13}$$

$$P(s_{2,\ell}/h2^i_{\ell-1}) > P(s_{2,\ell}/h2^j_{\ell-1}) \quad \forall i, j, s_{2,\ell} \tag{3.14}$$

Note that the probability of high (low) positive or negative returns is high when in the past we had high (low) positive or negative returns. However, no particular temporal pattern is more probable than other. For instance this may be the case

because inside the set $\left\{h1^i_{\ell-1}\right\}_{i=1}^{i=2^{\ell-1}}$ there is no $h1^k_{\ell-1}$ such that $P(s_{1,\ell}/h1^k_{\ell-1}) > P(s_{1,\ell}/h1^i_{\ell-1}) \quad \forall i,k,k \neq i$. I.e., there is no history that is better predicting $s_{i,\ell}$.

Then the test to be developed is able to detect the existence of "strange" patterns, if any, telling us "how they look like". For instance, we can test directly asset returns or their AR residuals in order to detect the existence of such strange patterns. In case of the existence of a strange patterns or unstable cycles, the inequations (3.15) and (3.16) would embody these patterns in our symbolic model.

$$P(s_{1,\ell}/h1^k_{\ell-1}) > P(s_{1,\ell}/h1^i_{\ell-1}) > P(s_{1,\ell}/h2^j_{\ell-1}) \quad \forall j,i,k \neq i, s_{1,\ell} \tag{3.15}$$

$$P(s_{2,\ell}/h2^k_{\ell-1}) > P(s_{2,\ell}/h2^i_{\ell-1}) > P(s_{2,\ell}/h2^j_{\ell-1}) \quad \forall i,j,k \neq i, s_{2,\ell} \tag{3.16}$$

Once more, we have a general model for stock prices which not only considers the particular case of a random walk but also allows for more complex processes as those with volatility clustering. The latter take place if higher probability is assigned to conditional events belonging to $A_1 \equiv \{1,4\}$ and $A_2 \equiv \{2,3\}$ separately, and lower probability to events combining elements of the two subsets. For instance, probability of being in 4 after having been in 1 in the past, should be greater than the probability of being in 4 after 2 in the past, $(P(4/1) > P(4/2))$. In words, if the returns are in a high volatility regime they are likely to remain in the same regime.

3.4.2 Construction of the Randomness Test (R) by using 4 symbols In this subsection the randomness test is constructed by using 4 symbols. The same method introduced in the section above is applied. We simulate random samples of size T and then the frequencies are computed and the entropy is calculated. This process is repeated 10,000 times, and an empirical distribution for R is obtained as explained in the section above. The test was computed by using 4 regions according to equation 3.11. For daily data we proceed to simulate 10,000 samples of size equal to 10,500, Table 5 shows the critical values at 5%.

Table 5

Critical Values at 95% for R-Statistic (T=10,500)			
R-1 days	R-2 days	R-3 days	R-4 days
0.0003	0.0005	0.0010	0.0026

We take asset returns netting the mean and then define three thresholds in the empirical distribution in order to compute the normalized entropy. Table 6 shows the R-statistic for different asset returns.

Table 6

Test of Randomness (R=1-H) Using 4 symbols (10,500 days)

Financial Returns	*R-1 day*	*R-2 days*	*R-3 days*	*R-4 days*
Alcoa Inc.	0.0047*	0.0077*	0.0098*	0.0126*
Boeing Co.	0.0062*	0.0084*	0.0108*	0.0137*
Caterpillar Inc.	0.0038*	0.0066*	0.0085*	0.0113*
Coca Cola Co.	0.0024*	0.0044*	0.0061*	0.0084*
Du Pont EI	0.0044*	0.0069*	0.0090*	0.0118*
Eastman Kodak Co.	0.0037*	0.0049*	0.0063*	0.0085*
General Electric Co.	0.0020*	0.0040*	0.0057*	0.0082*
General Motors Co.	0.0051*	0.0072*	0.0091*	0.0115*
Hewlett Packard Co.	0.0017*	0.0031*	0.0048*	0.0073*
IBM	0.0010*	0.0025*	0.0041*	0.0065*
Walt Disney Co.	0.0027*	0.0055*	0.0076*	0.0105*
S&P 500	0.0000	0.0040*	0.0072*	0.0108*
Dow Jones	0.0000	0.0022*	0.0043*	0.0072*
10 years treasure notes	0.0134*	0.0232*	0.0309*	0.0381*

* Rejection of randomness hypothesis at 5%

Table 6 shows that the randomness is rejected in all the cases as equal as just like when using 2 symbols and once more the random walk seems to be a bad model when using daily data. Analyzing the patterns causing this bias with respect to randomness, we observe the following: 1) The sequence [2,2] is the most frequent[7]

[7] In all the cases the frequence is around 0.09, but existing 16 possibilities a random process

in all the asset, the indices are the exception (in fact, DJIA and S&P500 shows [1,1] as the most frequent), in most of the cases the second and third most frequent patterns are [1,1] and [4,4]; 2) For sequences of length three the sequence [2,2,2] is the most frequent for assets and [1,1,1] for the indices.; 3) for four-length sequences, [2,2,2,2] is the most frequent and [1,1,1,1] for the two indices. These facts suggest the persistence in a regime of low loss [2] (below the mean) when considering asset returns, but persistence in a regime of high loss [1] with the indices.

3.4.3 Residual of an AR(1) As in the previous section, we applied an autorregressive process of order 1 to daily returns in order to remove linear components of the series. Residuals of these models are tested in order to see if they are random. Table 7 shows R-statistic for such residuals. Note that randomness is rejected in all cases. This result suggests that a linear model is not a good approach to modeling daily asset returns due to the presence of nonlinear components.

should present a frequence near 0.06.

Table 7

4 symbols 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*
Alcoa Inc.	0.0000	0.0017*	0.0034*	0.0062*
Boeing Co.	0.0000	0.0021*	0.0044*	0.0071*
Caterpillar Inc.	0.0000	0.0017*	0.0034*	0.0058*
Coca Cola Co.	0.0000	0.0021*	0.0036*	0.0059*
Du Pont EI	0.0000	0.0029*	0.0053*	0.0082*
Eastman Kodak Co.	0.0000	0.0021*	0.0038*	0.0061*
General Electric Co.	0.0000	0.0027*	0.0046*	0.0072*
General Motors Co.	0.0000	0.0029*	0.0051*	0.0078*
Hewlett Packard Co.	0.0000	0.0017*	0.0035*	0.0058*
IBM	0.0000	0.0016*	0.0032*	0.0057*
Walt Disney Co.	0.0000	0.0023*	0.0041*	0.0067*
S&P 500	0.0000	0.0025*	0.0055*	0.0089*
Dow Jones	0.0000	0.0013*	0.0031*	0.0059*
10 years treasure notes	0.0009*	0.0086*	0.0156*	0.0223*

* Rejection of randomness hypothesis at 5%

Notice that even when an AR(1) is applied to the returns, the test is able to detect determinism in the residuals. Table 7 shows that randomness is rejected for the residuals of an AR(1) model for the returns. The most frequent patters still show persistence in regimes of "high volatility" or "low volatility", [2,3],[3,2],[1,1],[4,4],

[3,2,3],[1,1,4],[3,2,3,2],[1,1,1,4].

Even eliminating autocorrelation, results are similar to the previous one.

3.5 The Approximated Distribution of the R-statistic

The objective of this subsection is to derive an approximation for the R-statistic distribution under the null hypothesis of randomness. This will be useful to analyze some properties of the introduced statistic.

At first we obtain the approximated distribution of R under the null hypothesis of independence when 2 symbols are considered. Assume that s can take 2 values $\{1, 2\}$ and it is distributed as follows:

$$p_x \begin{cases} \frac{1}{2} + \varepsilon_1 & \text{if} \quad s = 1 \\ \frac{1}{2} + \varepsilon_2 & \text{if} \quad s = 2 \\ 0 & \text{otherwise} \end{cases}$$

Suppose also that ε_i represents the sample size noise, and it is distributed as a normal $N_2(0, \sigma^2\mathbf{\Sigma})$, where σ^2 is less than $1/2$ and tending to zero as the sample size increases, note that $\sum_{i=1}^{i=2} \varepsilon_i = 0$, since the total noise should be cancelled in order to maintain the sum of probabilities equal to 1 for the density p_s. This simply tries to reflect the fact that for small sample of random events the frequency may not be exactly equal to $1/2$.

Consider $R = 1 - H$, as explained in the previous section and let us derive the approximated distribution of the R-statistic for length 1 in the following manner:

$$R_1 = 1 - \left(-\frac{1}{\log_2(2)}\right)\left\{\left(\frac{1}{2}+\varepsilon_1\right)\log_2\left(\frac{1}{2}+\varepsilon_1\right)+\left(\frac{1}{2}+\varepsilon_2\right)\log_2\left(\frac{1}{2}+\varepsilon_2\right)\right\} \tag{3.17}$$

Note that $\log_2(\frac{1}{2}+\varepsilon_i) = \log_2(1+2\varepsilon_i) - 1$ and since $|\varepsilon_1| \leq 1$ then $\log_2(\frac{1}{2}+\varepsilon_i) \simeq 2k\varepsilon_i - 1$, where $k = 1/\ln(2)$. Then $R_1 \simeq 1 - (-1)((\frac{1}{2}+\varepsilon_1)(2k\varepsilon_1 - 1) + (\frac{1}{2}+\varepsilon_2)(2k\varepsilon_2 + 1))$. After some calculations and because the theorem we will show later, it is obtained that:

$$R_1 \simeq 2k\sigma^2\left(\frac{\varepsilon_1^2+\varepsilon_2^2}{\sigma^2}\right) \sim 2k\sigma^2\chi^2 \tag{3.18}$$

Where the term in brackets is distributed as a chi-square with 1 degree of freedom. Note that R_1 positively depends on σ^2, the noise variance produced by the small sample effect. When the sample increases, the variance is reduced, determining a smaller critical value for R (in the limit, the variance is zero when the sample is infinite). In fact, note the different tables of critical values (see Appendix I), when the sample is smaller, the critical values increase.

Let us assume now that s can take 4 values $\{1, 2, 3, 4\}$ and is distributed as follows.

$$p_s \begin{cases} \frac{1}{4}+\varepsilon_1 & \text{if} \quad s=1 \\ \frac{1}{4}+\varepsilon_2 & \text{if} \quad s=2 \\ \frac{1}{4}+\varepsilon_3 & \text{if} \quad s=3 \\ \frac{1}{4}+\varepsilon_4 & \text{if} \quad s=4 \\ 0 & \text{otherwise} \end{cases} \tag{3.19}$$

Suppose a vector $\boldsymbol{\varepsilon} \equiv (\varepsilon_1, \varepsilon_2, \varepsilon_3, s_4)$ representing the sample size noise and consider it to be distributed as a multinormal $N_4(0, \sigma^2\mathbf{\Sigma})$, where σ^2 is less than $1/4$ and tending to zero and the matrix Σ is idempotent matrix as follows:

$$\Sigma \equiv \begin{bmatrix} 3/4 & -1/4 & -1/4 & -1/4 \\ -1/4 & 3/4 & -1/4 & -1/4 \\ -1/4 & -1/4 & 3/4 & -1/4 \\ -1/4 & -1/4 & -1/4 & 3/4 \end{bmatrix} \tag{3.20}$$

Of course, $\sum_{i=1}^{i=4} \varepsilon_i = 0$, since the total noise should cancel to maintain the sum of probabilities equal to 1 for the density p_s.

Substituting 3.19 in $R = 1 - H$, we obtain equation 3.21.

$$R_1 = 1 - \left(-\frac{1}{\log_2(4)}\right) \left\{ \sum_{i=1}^{i=4} \left(\frac{1}{4}+\varepsilon_i\right) \log_2 \left(\frac{1}{4}+\varepsilon_i\right) \right\} \tag{3.21}$$

Note that $\log_2(\frac{1}{4}+\varepsilon_i) = \log_2(1+4\varepsilon_i) - 2$ and since $|\varepsilon_i| \leq 1$ then $\log_2(\frac{1}{4}+\varepsilon_i) \simeq +4k\varepsilon_i - 2$, where $k = 1/\ln(2)$. Then $R_1 \simeq 1 - \left(-\frac{1}{\log_2(4)}\right) \left\{ \sum_{i=1}^{i=4} \left(\frac{1}{4}+\varepsilon_i\right)(4k\varepsilon_i - 2) \right\} = 1 + \frac{1}{2} \left\{ -2 + \sum_{i=1}^{i=4} 4k\varepsilon_i^2 + (k-2) \sum_{i=1}^{i=4} \varepsilon_i \right\}$. Since $\sum_{i=1}^{i=4} \varepsilon_i = 0$, the following expression is obtained:

$$R_1 \simeq 2k\sigma^2 \left\{ \frac{\varepsilon_1^2 + \varepsilon_2^2 + \varepsilon_3^2 + \varepsilon_4^2}{\sigma^2} \right\} \tag{3.22}$$

Where the term in brackets is a quadratic form in random normal variables. As Mathai and Provost (1992) show, the distribution of quadratic forms in normal variables has been extensively studied by many authors. Various representations of the distribution function have been derived and several different procedures have been given for computing the distribution and preparing appropriate tables. Approximated distributions have been proposed by Patnaik (1949), Pearson (1959), Siddiqui (1965), Solomon and Stephens (1978) and Oman and Zacks (1981). However, in the present case we can apply the following theorem (Mathai and Provost (1992) p. 197):

Necessary and sufficient conditions for a quadratic form $\mathbf{X}$'$A\mathbf{X}$ to be distributed as a chi-square variates with r degrees of freedom when X has a multivariate normal distribution with mean vector 0 and possibly singular covariance matrix Σ, are:

(i) $(A\Sigma)^2 = (A\Sigma)^3$ and $tr(A\Sigma) = r$

(ii) $tr(A\Sigma)^2 = tr(A\Sigma) = r$ and $\rho(\Sigma A\Sigma) = r$

Theorem can in fact be applied in the present case. The quadratic form obtained is $Q = \left\{ \frac{\varepsilon_1^2+\varepsilon_2^2+\varepsilon_3^2+\varepsilon_4^2}{\sigma^2} \right\} = \mathbf{X}$'$A\mathbf{X}$, where $\mathbf{X} \equiv \left(\frac{\varepsilon_1}{\sigma}, \frac{\varepsilon_2}{\sigma}, \frac{\varepsilon_3}{\sigma}, \frac{\varepsilon_4}{\sigma} \right)$, $\mathbf{X}$ is distributed $N_4(0, \mathbf{\Sigma})$, $A = A$' symmetric matrix, and $\mathbf{\Sigma}$ is symmetric, singular, and idempotent. Since $tr(A\Sigma) = 3$, then $\mathbf{X}$'$A\mathbf{X} \sim \chi_3^2$ Therefore, under these assumptions the approximated distribution of the R-statistic for 1 length is the following:.

$$R_1 \text{ approximately distributes as } \quad 2k\sigma^2\chi_3^2 \tag{3.23}$$

Generalizing for a given number of random events n, $A = I$ and the covariance matrix has $(n-1)/n$ in the principal diagonal and $-1/n$ elsewhere as follows:

$$\Sigma_{nxn} \equiv \begin{bmatrix} (n-1)/n & -1/n & ... & -1/n \\ -1/n & (n-1)/n & ... & -1/n \\ ... & ... & ... & ... \\ -1/n & -1/n & ... & (n-1)/n \end{bmatrix} \tag{3.24}$$

Therefore, since $tr(A\Sigma) = (n-1)$, it is possible to assert that R is approximately distributed as $\frac{n}{\log_2(n)} k\sigma^2\chi_{n-1}^2$, where $k = 1/\ln(2)$.

Considering σ^2 as the variance due to sample size error, notice that when $\sigma^2 = 0$ (there is no error) R is equal 0 and, as it was mentioned, it is a complete random process. However, σ^2 increases as the sample size T is reducing, then it is possible to establish that $\lim \sigma_T^2 = 0$ as $T \to \infty$, and that R increases as σ_T^2 and T decreases, see Figure 3.3. As σ^2 increases R-statistic also increases, this being the reason why the critical values increase for small samples (an so, σ^2 increases).

Note in Figure 3.4 as the length increases, the maximum moves to the right for values larger than 0, and values far from 0 have greater probability, this is the reason why the simulated critical values in Table 12, 14, 16 also increase when the length increases.

Table 8 compares the critical values from the simulation density (SD) and the critical values from the approximated distribution (AD). The two effects are

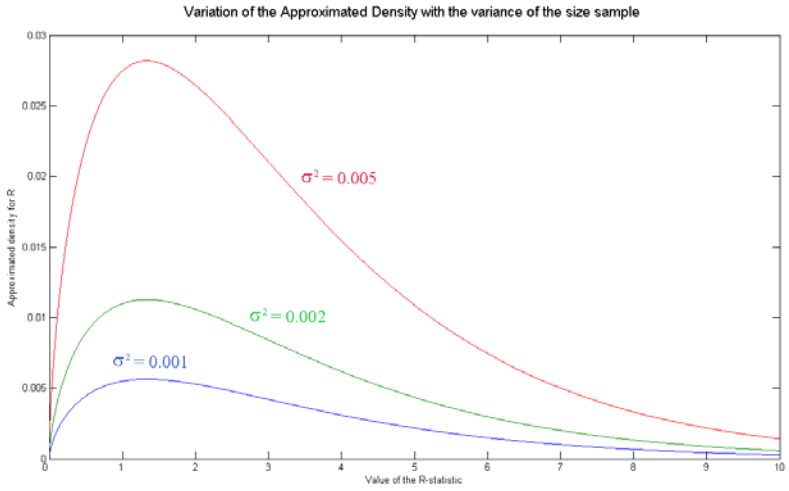


Figure 3.3: Variation of the Approximated Density with the variance of the size sample

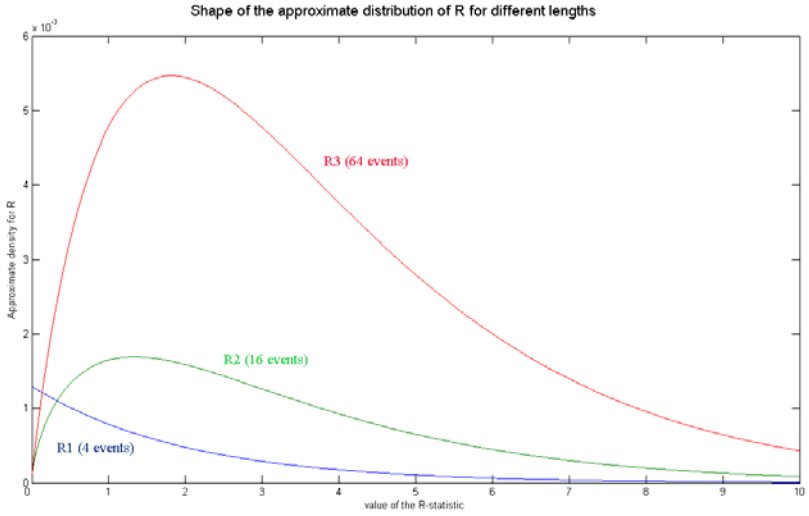


Figure 3.4: Shape of the approximated distribution of R for different lenghts

shown, the critical values increase as the sample size decreases and longer lengths are considered. The critical values from the AD are more conservative than from the SD, however they get closer, as the sample increases. Notice in Table 8 that for T=10,500 the difference between SD and AD is not so large.

Table 8. CV at 95% from the S. Den. (SD) and the A. Den. (AD)

	T=500		**T=2,000**		**T=10,500**	
Length	SD	AD	SD	AD	SD	AD
R-1	*0.0056*	*0.0084*	*0.0014*	*0.0021*	*0.0003*	*0.0004*
R-2	*0.0097*	*0.0168*	*0.0024*	*0.0042*	*0.0005*	*0.0008*
R-3	*0.0214*	*0.0388*	*0.0057*	*0.0098*	*0.0010*	*0.0019*
R-4	*0.0573*	*0.1045*	*0.0146*	*0.0263*	*0.0026*	*0.0050*

In all the cases the mean of σ^2 from the simulations was considered

In Figure 3.5 the simulated and approximated densities are compared for R of different lengths, and the approximated distributions follow similar shapes. Note also that as the length increases the shape tends to be normal. Actually, Mathai and Provost (1992) assert that this kind of quadratic form converges to normal distribution as the degree of freedom increases.

In summary, the approximated distribution of R was obtained to study some properties of the R-statistic density under the null hypothesis. Some facts are confirmed: at first for small sample the variance increases, increasing the critical value at 5%. In the second place, as longer lengths are considered the critical value also increases. The approximated distribution is similar to the empirical

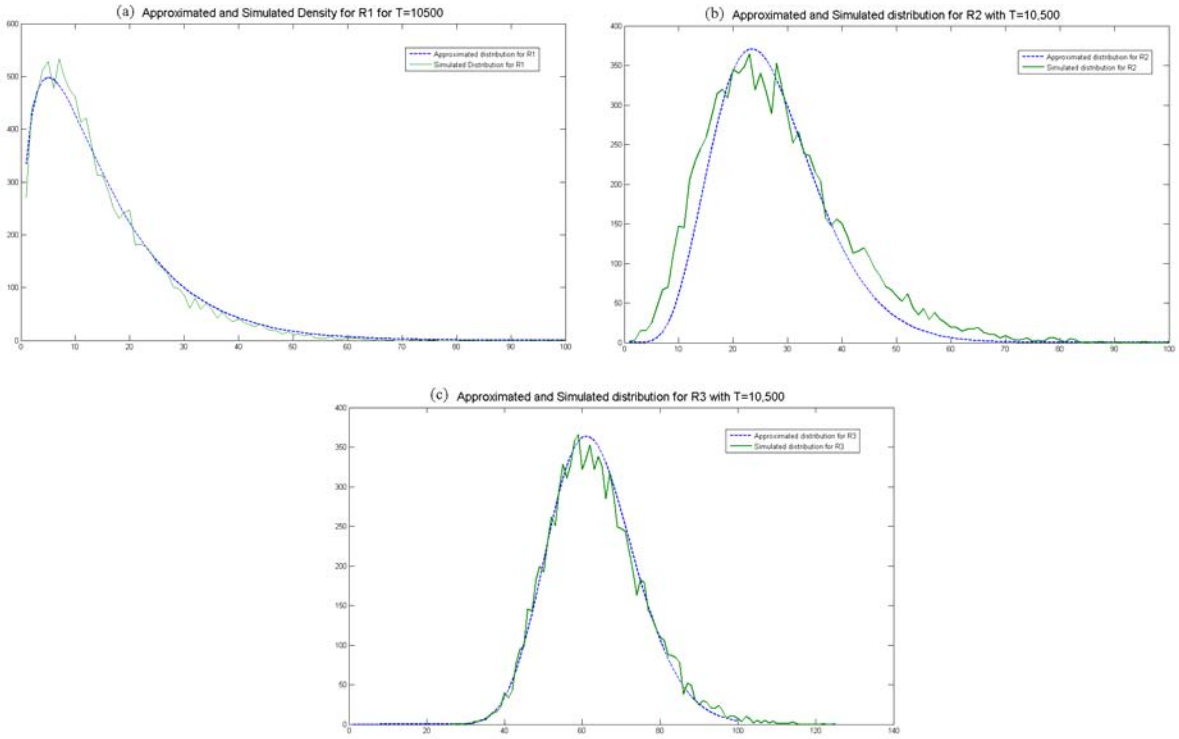


Figure 3.5: Simulated (dashed line) and Empirical (solid line) densities for (a) R_1, (b) R_2, (c) R_3

distribution in shape and the critical values are close. Even if approximated critical values are more conservative than empirical ones, they tend to each other as the sample size increases.

3.6 Power and Size of the 4-symbol Randomness Test

At first, we conduct an experiment aiming to check if the critical values used in the test are unbiased. A time series of length 500 is generated by a Gaussian distribution, the test is applied and the null hypothesis is rejected or not, this procedure being repeated 1,000 times. Table 9 shows the size of the randomness test for various significance levels, actually the results are the percentage of the null hypothesis rejection over 1,000 simulations. For instance, for columns with significance level $\alpha = 0.05$, the proportion of times the null hypothesis is rejected should be 5 per cent of the time.

Table 9: Size of the 4-symbols Randomness

length	$\alpha = 1\%$	$\alpha = 2.5\%$	$\alpha = 5\%$	$\alpha = 10\%$
2	0.0000	0.0000	0.0010	0.0060
3	0.0010	0.0020	*0.0100*	0.0270
4	0.0000	0.0000	0.0000	0.0000

Sample Size T=500

Table 9 also suggests that the test tends to accept the null hypothesis more times than the critical values in small samples. However, the results alleviate when considering a length of 3 consecutive events, here the best result is obtained, note

that for $\alpha = 5\%$, the independence is rejected 1% of the time.

Following Liu et al. (1992) we try to check the power and size of the 4-symbols randomness test comparing the results with the popular BDS test. Using Monte Carlo Simulation we simulate 2,000 times, samples of 500 and 2,000, and for different models (see appendix II). Then, an AR(1) process is applied to all the time series and the residuals are tested applying our test (for different lengths). In addition, the BDS test is also applied in order to study its performance compared to our test in detecting nonlinearity. As remarked by Liu et al. (1992), since the tests are applied as tests for stochastic or deterministic nonlinearity, it is necessary to remove linear components of the series before applying them. To do this, in practice an AR(p) model is built for x_t, using some criteria such as AIC or BIC. Then the test is applied to the residuals of the linear fitting procedure. The BDS test is applied with a distance (m) equal to 1.5 and an epsilon (ε) around 0.7, and the Randomness Test is applied for a length of 2 and 3 consecutive events.

Table 10. Size and Power of the 4-symbolic Randomness Test for residuals

	T=500	T=2,000	T=500	T=2,000	T=500	T=2,000	T=500	T=2,000
	AR(1)		**MA(2)**		**NLSIGN**		**Bilinear**	
BDS	*0.0610*	*0.0490*	*0.0430*	*0.0485*	*0.0570*	*0.0495*	*1.0000*	*1.0000*
R-1	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.3820*	*0.2830*
R-2	*0.0020*	*0.0020*	*0.0030*	*0.0020*	*0.0070*	*0.0385*	*0.8890*	*1.0000*
R-3	*0.0070*	*0.0050*	*0.0050*	*0.0060*	*0.1620*	0.9765	*1.0000*	*1.0000*
	Logistic Map		**Tent Map**		**NLMA1**		**BLMA**	
BDS	*0.8880*	*0.9965*	*1.0000*	*1.0000*	*0.0980*	*0.1360*	*1.0000*	*1.0000*
R-1	*0.0000*	*0.0000*	*1.0000*	*1.0000*	*0.1350*	*0.4360*	*0.2290*	*0.3210*
R-2	*0.9720*	*0.9900*	*1.0000*	*1.0000*	*0.1800*	*0.7620*	*0.9990*	*1.0000*
R-3	*0.9720*	*0.9900*	*1.0000*	*1.0000*	*0.9840*	*1.0000*	*0.9990*	*1.0000*

Note: The residuals from first-order autoregressive regression for AR(1), NLSIGN, Bilinear. For MA(2), BLMA, NLMA1 models, residuals are deriver from a second-order autorregressive regression. In case of Chaos, tests are applied to the original series. The numbers show the percentage rejection in 2000 replications with 5 percent significance level

Note the performance of the test of nonlinearity proposed in the present chapter has a high power respect to the BDS. At first, note that both tests have good performance recognizing nonlinearity and chaos, when testing chaotic processes such as the Logistic Map and the Tent map both tests reject independence hypothesis more than 90% of the times. Residuals of AR(1) and MA(2) models are also recognized as independent, note that both tests reject null hypothesis less

than 5% of the times. Bilinear and bilinear moving average (BLMA) models are also recognized by the two test, note that both BDS and R-statistic reject the null hypothesis more than 90%. According to Liu et al. (1992) the BDS has the greatest power on the Bilinear model, in fact the hypothesis is rejected 100%, but also in the R-statistic test the hypothesis is rejected. The most important, note that Liu et al. (1992) remark that BDS has the least power on the nonlinear sign model (NLSIGN), the residuals seem to be i.i.d. by the BDS. Actually, note in Table 21 that the hypothesis is rejected around 5% of the time as in a random process, however R-statistic rejects the hypothesis more than 90% of the time for a sample of 2,000 and for a length of 3 events. The NLMA1 is another case where independence is rejected few times by using BDS test but more than 90% for the present R-statistic. Therefore it is possible to conclude two things: 1) At first, introduced test seems to have greater power than the BDS recognizing these kind of nonlinearity; 2) the test has the best performance when considering a length of 3 consecutive events and when the sample is large.

3.7 Conclusions

The main purpose of the chapter was to introduce a statistic in order to measure the informational efficiency in the stock markets. We used symbolic dynamics to rule out the noise that typically affects asset returns. On the other hand, we applied the Shannon entropy widely used in information theory, to recover the quantity of information in the data. Even though the present chapter is a bit technical, it is important because the introduced statistic will be central in the

rest of the dissertation.

We constructed a test to check if the EMH, at least in its weak version is present in some assets and indices. Models for 2 and 4 symbols are constructed to compare the results. An some expertiments were realized to check the performance of the statistics.

We obtained that the randomness hypothesis is rejected for the daily asset returns and indices at levels and when the deterministic linear components are eliminated through an AR(1) model. The results are similar whether using 2 or 4 symbols.

An approximate distribution of the test was obtained in order to derive certain results. It was shown that the critical value increases when taking smaller samples and longer lengths, on the other hand critical values between asymptotic and simulated density seem to be similar when taking larger samples. Some experiments were done in order to check the power and size of the test. At first, simulation of normal random process was produced the test presented the best performance for a length of 3 consecutive events. However, the test seems to be conservative, accepting the null hypothesis more time than the critical value probability. Performance was compared with the famous BDS test, some nonlinear models were tested and both tests presented good performance. In special R-statistic with 3 length detected NLSIGN model, while BDS did not detected this nonlinearity as it was highlighted by Liu et al. (1992). Note that also NLMA1 was detected by R-statistic (rejecting null hypothesis more than 90% of the time) while BDS rejected

the null hypothesis only 14% percent of the time.

The next 2 chapters will use this measure for informational efficiency. On the one hand, we shall measure the efficiency for different stocks trying to study if developed markets are more efficient than emerging ones. On the other hand, in the chapter 5 we shall study the relationship between efficiency and the probability of a crash.

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

-Cover, T., Thomas J., (1991), *Elements of Information Theory*, Wiley ed.

-Daw, C., Finney, C., Tracy, E., (2003), "A review of symbolic analysis of experimental data", *Review of Scientific Instruments*, Vol. 74, No. 2, pp. 915-930.

-Khinchin, A., (1957), *Mathematical Foundations of Information Theory,* Courier Dover Publications.

-Lo, A.W., 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.

-Lui, T., Granger, C., Heller, W., (1992), "Using the Correlation Ex-

ponent to Decide Whether an Economic Series is Chaotic", *Journal of Applied Econometrics*, Vol. 7, Supplement: Special Issues on Nonlinear Dynamics and Econometrics, pp. S25-S39.

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

-Mathai, A., Provost, S., (1992), *Quadratic Forms in Random Variables: Theory and Applications*, Marcel Dekker, Inc.

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

-Oman, S., Zacks, S., (1981), "A mixture approximation to the distribution of weighted sum of chi-squared variables", *Journal of Statistical Computation and Simulation*, Vol. 13, pp. 215-224

-Patnaik, P., (1949), "The non-central Chi-square and F-distributions and their applications", *Biometrika*, Vol. 36, pp. 128-131.

-Pearson, E., (1959), "Note on an approximation to the distribution of noncentral χ^2", *Biometrika*, Vol. 46, p. 364.

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

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

-Shannon, C., (1948), "A Mathematical Theory of Communication", *Bell*

System Technical Journal, 27: 379-423., 623-656.

-Siddiqui, M., (1965), "Approximations to the distribution of quadratic forms", *The Annals of Mathematical Statistics*, Vol. 36, pp.677-682.

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

-Solomon, H., Stephens, M., (1978), "Approximations to density functions using Pearson curves", *Journal of the American Statistical Association*, Vol. 73, pp. 153-160.

-Williams, S., (2004), *Symbolic Dynamics and its Applications,* Proceeding of Symposia in Applied Mathematics, Vol.60, 150 pp.

3.9 APPENDIX I: Critical Values for different samples (Test for 2-symbols)

Table 9

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 10

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 11

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

3.10 APPENDIX II: Models applied in the size and power experiment[8]

1) The autorregressive model, AR(1):

$$r_t = 0.45.r_{t-1} + \varepsilon_t$$

2) The moving average process, MA(2):

$$r_t = \varepsilon_t - 0.1.\varepsilon_{t-1} + 0.2.\varepsilon_{t-2}$$

3) The Non Linear Sign model, NLSIGN:

$$r_t = SIGN(x_{t-1}) + \varepsilon_t, \quad SIGN(x) = 1, 0, \text{or } -1, \quad \text{if } x < 0, = 0, > 0$$

4) T$\int$he Bilinear model, BL:

$$r_t = 0.7.r_{t-1}\varepsilon_{t-2} + \varepsilon_t$$

5) The Logistic Map:

$$r_t = 0.4.r_{t-1}.(1 - r_{t-1})$$

6) The Tent Map:

$$\begin{cases} r_t = 0.49.r_{t-1} & \text{if} \quad 0 \leq r_{t-1} < 0.49 \\ r_t = (1 - 0.49)^{-1}.(1 - r_{t-1}) & \text{if} \quad 0.49 \leq r_{t-1} \leq 1 \end{cases}$$

7) The Nonlinear Moving Average, NLMA1:

$$r_t = \varepsilon_t - 0.4.\varepsilon_{t-1} + 0.3.\varepsilon_{t-2} + 0.5.\varepsilon_t.\varepsilon_{t-2}$$

8) The Bilinear Moving Average, BLMA:

$$r_t = 0.4.r_{t-1} - 0.3.r_{t-2} + 0.5.r_{t-1}\varepsilon_{t-1} + 0.8.\varepsilon_{t-1} + \varepsilon_t$$

CHAPTER 4

The Informational Efficiency: Emerging vs Developed Markets

4.1 Introduction

The Stock Exchange Markets around the world are governed by different rules. Some markets are more liberal, another markets have more constraints. In particular, the Emerging Markets have been accused of being inefficient in contrast to developed markets where the information is quickly assimilated. As far as we know, there are few studies measuring the efficiency for different stock markets. In particular, Cajueiro and Tabak (2004) (2005) use the Hurst exponent as a efficiency measure concluding that Asian markets are more efficient than those in Latin America (with the exception of Mexico).

The present chapter aims to measure the informational efficiency of different stock markets around the world in order to analyze if the emerging markets are more inefficient than the developed ones. In words, news affecting the emerging market encounter greater difficulty to be incorporated in the prices compared to the developed markets. The following section will introduce the methodology, which is basically the Shannon entropy and the symbolic analysis used and analyzed in the previous chapter. In section three a ranking of efficiency is constructed for different countries taking the last teen years as period of study. Finally, some conclusions are drawn.

4.2 Methodology

As explained above, the measure of efficiency is the Shannon entropy applied to coded time series. In the present case 2 symbols are used taking, as threshold, the average returns of the indices. After taking the coded time series we compute the modified Shannon entropy for different lengths as shown in equation 4.1. Actually, note that in this equation H_ℓ is 3.4 when $k = 1/log_2(N_O)$.

$$H_\ell = \frac{-1}{\log_2(N_O)} \sum_{i=1}^{i=2^\ell} p_{i,\ell} \log_2 p_{i,\ell} \tag{4.1}$$

H_ℓ is the modified Shannon entropy for a length ℓ, $p_{i,\ell}$ is the probability for the event i of the length ℓ, note that N_O is the number of observed sequences with non-zero frequency as propossed by Finney et al. (1998). Therefore, for some sequences the entropy will arrive to a minimum which will be the level of efficiency.

4.3 Ranking of Informational Efficiency

In this section, a ranking of the informational efficiency is constructed for different stock markets around the world. The dataset was basically obtained from finance.yahoo.com, where daily data was obtained for the stock indices from July 1, 1997 to December 14, 2007. Data for Russian, Slovenia and Czech Republic were obtained from their respective stock markets sites, www.rts.rs, www.ljse.si, and www.pse.cz.

The entropies were applied to the coded time series, in most cases the minimum entropy is obtained for a length of 10 days. Table 1 shows the results for the 20

stock markets.

Table I: Ranking of Informational Efficiency for different stock indexes

Rank	Index	Country	Entropy	Rank	Index	Country	Entropy
1	TSEC	Taiwan	0.9833	11	MERVAL	Argentina	0.9791
2	NIKKEI	Japan	0.9806	12	ATX	Austria	0.9791
3	Straits Times	Singapore	0.9805	13	JKSE	Indonesia	0.9787
4	IPC	Mexico	0.9804	14	SSMI	Switzerland	0.9783
5	TASE	Israel	0.9801	15	Hang Seng	Hong Kong	0.9777
6	DJIA	USA	0.9800	16	KLSE	Malaysia	0.9769
7	AEX	Holland	0.9799	17	PX GLOB	Czech Rep.	0.9768
8	KOSPI	South Korea	0.9795	18	RTS	Russia	0.9751
9	DAX	Germany	0.9794	19	CMA	Egypt	0.9669
10	FTSE	UK	0.9794	20	SBI 20	Slovenia	0.9481

Based on own calculations.

Notice that three Asian markets take the first positions as the most efficient. Of course, the stock markets of Taiwan, Japan, and Singapore are important financial centers in the world. On the other hand, Mexico takes the third position. These results confirm those obtained by Cajueiro and Tabak (2004) (2005). In the last positions are the ex-socialist countries as Slovenia, being the most inefficient in the group. The unique African stock market (Egypt) is also in the position 19.

The last results seem to support the hypothesis that the emerging stock markets are more inefficient than developed ones. Note in particular that the ex-socialist countries have not achieved levels of efficiency similar to the more devel-

oped European markets. The average efficiency of the developed markets (0.9795), greater that the average efficiency of the emerging markets (0.9756).

When the European stock markets are analyzed, the difference among the Western and Eastern markets is clear. While the average efficiency among UK, Germany, Austria, Holland, and Switzerland is 0.9792, for Czech Republic, Russia and Slovenia this is 0.9666.

4.4 Conclusions

The symbolic analysis and the Shannon entropy seem to be useful as a measure of informational efficiency. The purpose of this chapter was to study the hypothesis of emerging markets as more inefficient than developed markets. The ranking constructed by using different stock indices gives support to such hypothesis. Note, that while more developed stock markets tend to remain in positions with high inefficiency, the emerging markets are in the lowest positions. In particular, the ex-socialist countries are in the last positions suggesting that there may lack experience in managing stock markets.

4.5 References

-Cajueiro, D., Tabak, B., (2004), "Ranking Efficiency for Emerging Markets", *Chaos, Solitions and Fractals*, Vol. 22, pp. 349-352.

-Cajueiro, D., Tabak, B., (2005), "Ranking Efficiency for Emerging Markets II", *Chaos, Solitions and Fractals*, Vol. 23, pp. 671-675.

-**Finney, C.E.A., Daw, C.S., and Green, J.B. (1998)**: "Symbolic Time-

Series Analysis of Engine Combustion measurements", SAE paper No. 980624.

CHAPTER 5

The Role of Efficiency in Predicting Crashes

5.1 Introduction

In Chapter 2 we mentioned that the Efficient Market Hypothesis (EMH), at least in its weakly version, assumes that all information provided by past prices is already embodied in present prices. Therefore price prediction is impossible in an efficient market. The most used and common framework for stock prices has been the random walk model. Recently, numerous research works have shown that sometimes the stock prices present a deviation from an idealized efficient behavior, see Lo and MacKinlay (1988), and Singal (2004).

The present Chapter applies the measure of the informational efficiency of stock exchange market introduced in Chapter 3. Studying the dynamics of this measure we are able to detect the formation of trends. In fact, if the market is efficient the new information should be immediately embodied in the prices, however if the market is inefficient, due to mispricing or anomalies (see Singal (2004)) maybe originated in the cost of information, the cost of trading, or the limits of arbitrages, the actual price may not reflect the news, permitting the formation of trends. As an example, let the price overshoot after some news; the market can follow a decreasing trend until it returns to the fundamental price.

On the other hand, we study how inefficiency affects the probability of having crashes. The idea is simple, an inefficient market may present patterns because the

available information is not immediately fully embodied in prices, however when the information is understood by agents, prices may adjust with abrupt movements, and the probability of having a crash by readjusting the prices to levels of efficiency, should increase. In order to detect the relation between efficiency with a crash probability, a binary model is applied.

Ever since Bachelier (1900) had proposed the Brownian motion as a model for stock prices suggesting that the difference in prices is a random process, academic world used Brownian motion and the random walk as models for stock prices. However, empirical evidence such as the famous "stylized facts" (fat tails and volatility clustering) and critical events like the 1987 crisis brought some scholars to study the possibility of nonlinearity in the evolution of prices, see Hsieh (1991) (1995). In particular, some scholars proposed the possibility of chaos, see Peters (1994) (1996), and LeBaron (1994). A chaotic dynamics is a deterministic process that looks like a random process, a main characteristic being that it is highly sensitive to changes in initial conditions, see Alligood et al. (1997) for an introduction to chaos.

One defender of the chaotic dynamics in finance is Peters (1994) (1996), who points out that people take decisions reacting to information in different ways and some do not react until a trend is confirmed clearly. The amount of information necessary to validate a trend varies, but its uneven assimilation may cause a biased random walk, extensively studied by Hurst in the 1940s and later in the 1960s and 1970s, Mandelbrot called it fractional Brownian motion. In more recent years the

local Hurst exponent has been proposed as a measure of efficiency. Grech and Mazur (2004) it to measure efficiency in the Dow Jones index, and they argue that even if we cannot predict the detailed evolution scenario, we might be able to say something else about the process. For example, the Hurst exponent seems to be able to detect crisis. In fact, it has been widely used in detecting long-time correlation in finance. However authors, such as Bassler et al. (2006) and McCauley et al. (2007) criticize this measure asserting that a value different from 1/2 (the number corresponding to a random walk process) does not necessarily imply long time correlations like those to be found in fractional Brownian motion.

Instead of the Hurst exponent, this chapter proposes the Shannon entropy. As it will be explained in the next section, the idea is simple: since symbolic analysis is useful detecting the very dynamics of highly noisy time series as the asset returns are, the use of entropy recovers the information in the series detecting the emergence of patterns. Once the measure is constructed, a logit model is applied in order to study the relationship between efficiency and probability of a financial crash.

The chapter is organized as follows. In section 5.2 we present the methodology, the efficiency measure is explained and the logit model is described. In section 5.3 we present the results for different stock markets (Japan, Malaysia, Russia, Mexico, and USA), showing that in all cases the relationship between the probability of a crash and the efficiency is negative. In the section 5.4 further results are presented on the global market, showing that US market is the most efficient. In section 5.5

we develop a model trying to explain the relationship between efficiency and news arrivals. Finally, in section 5.6 some conclusions are drawn.

5.2 Methodology

5.2.1 Measuring Informational Efficiency

At first, we will assume that in an efficient market the current price P_t reflects all the available information, and thus the past prices are useless in order to predict the prices. Therefore, we will suppose that the returns in a perfectly efficient market are unpredictable. Under the martingale measure or assuming risk neutral agents, it is allowed to assume that the returns are independently distributed.

The measure of efficiency is computed in two steps, first the symbolization of the returns is applied in order to detect the very dynamics of the process, the Shannon entropy is applied next in order to measure the quantity of information embodied in the series.

Using the STSA we can obtain richer information from a time series if we transform data series of many possible values (real-valued for instance) into a time series of only a few distinct symbols. According to Daw et al. (2003) this coarse-graining has the practical effect of producing low-resolution data from high-resolution one. As far as we know, few works are done applying symbolic analysis, Lawrence et al. (1998) predict the direction of change for next day in foreign exchange rates with an error of 47.1%. On the other hand, Schittenkopf et al. (2002) predict the daily change in volatility of two major stock indices, they assert that symbolic information processing is a promising approach to financial prediction tasks undermining

the hypothesis of efficient capital markets.

The problem of the symbolic analysis is that there is no formal way to define the time series partitions. However in our case we are interested in combinations of negative and positive stock returns. Therefore we take zero as separative value in these returns, and two symbols: we call "0" for negative returns, and "1" for positive ones. In fact, we can define two extreme symbolic dynamics for the asset returns as follows. On the one hand, in a completely efficient market the returns instantaneously reflect the news appearing as shocks and, in such case the process is a sequence of Bernoulli processes (0s and 1s). On the other hand, if the market is inefficient the returns will not reflect the news immediately, instead, they may present sequences of increments (decrements) if returns under-(over-) shoot as reaction to good news. Therefore, assume we have a time series of size T defined as $\{r_1, r_2, r_3, \ldots, r_T\}$, r_t being the asset returns (defined as the log price difference) at time t, for $t = 1, 2, \ldots, T$. Then, we proceed to transform the time series into a symbolic one, according to rule 5.1.

$$if \begin{cases} r_t < 0 & s_t = 0 \\ r_t > 0 & s_t = 1 \end{cases} \tag{5.1}$$

Thus symbolic time series $\{s_1, s_2, s_3, \ldots, s_T\}$ is obtained. It is a time series expressed in sequences of 0s and 1s, representing the decreases and increases in prices, respectively.

On the other hand, we have a measure of uncertainty which will be our measure

of efficiency. In fact, this measure is the normalized Shannon entropy (H) achieving its maximum value at 1 when the process is completely random (see subsection 3.2.2), and its minimum at 0, when the process is a completely certain event. The theoretical expression for H is the equation 5.2, a particular case of the equation 3.4 propossed by Khinchin (1957)

$$H = -(1/\log_2(n)) \sum p_i \log_2 p_i \tag{5.2}$$

where n is the total number of sequences and p_i the probability of sequencec i, with $i = 1, 2, \ldots, n$.

Note that the entropy is maximum when the n sequences or events are equally probable, $p_1 = p_2 = p_3 = \ldots = p_n$. However, the entropy is at minimum at 0 when one event cumulates all the probability (it means this is a completely certain event). For instance, imagine that we are computing our measure only for two events which can happen in a day. Assume that p is the probability of having a decrease in prices (symbol 0) and $(1-p)$ is the probability of having an increase in prices (symbol 1). According to the Shannon entropy the efficiency will be given by:

$$H = -\left(\frac{1}{\log_2(2)}\right)(p.\log_2(p) + (1-p).\log_2(1-p)) \tag{5.3}$$

Note that H is a concave function and the maximum is obtained for $p = 1/2$ (maximum uncertainty) and the minimum is for $p = 1$ and $p = 0$. Intuitively, in our case if the market is efficient and no trend is developing, the probability of 0

and 1 is the same. However in a "bear market" (or a "bull market") probability of 0 (1) is greater than 1/2, highlighting the formation of a trend and thus, a lower efficiency level.

In practice, the probability of prices decreasing (p) is calculated counting the quantity of 0s over the whole period, while $(1-p)$ is the number of 1s over the total period.

In order to study the evolution in time of efficiency, a time-window is selected. In fact, a sub-period $v < T$ is taken and shifted across time. Then, the Shannon Entropy is computed for each time-window from moment 1 to T. As is shown in the next section different sizes of windows are considered, 100, 240, 350, and 420 days. However as suggested by Grech and Mazur (2004), the time-window should not be too large in order to capture the locality. We will proceed to select the time-window that best fits the best the logit model.

5.2.2 The Logit Model The logit and probit models are two famous models for binary endogenous variables[1]. Assume there is a variable y that takes on one of two values, 0 and 1, in the present case, financial crash (1) and no-crash (0). Define a latent variable y^* such that:

$$y_i^* = \alpha + \beta H_i + \varepsilon_i \tag{5.4}$$

where H is the efficiency measure and ε_i follows what is called an extreme value distribution, see McFadden (1984).

[1] See Johnston and DiNardo (1997) for an introductory discussion of the binary models.

We do not observe y^*, but rather y, which takes on values of 0 or 1 according to the following rule:

$$y_i = \begin{cases} 1 & \text{if} \quad y_i^* > 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.5}$$

In the present work, the variable y takes value 1 when the empirical distribution of the index returns cumulates the 1% (the negative tail), and 0 otherwise.

According to Johnston and DiNardo (1997) the present logit model is given by:

$$p(y_i = 1) = \frac{\exp(\alpha + \beta H_i)}{1 - \exp(\alpha + \beta H_i)} \tag{5.6}$$

The equation 5.6 says that the probability of the financial crash in one day, $P(y_i = 1)$ depends on the efficiency level (H).

The formulation of the model (5.6) ensures that the predicted probabilities lie between 0 and 1. The model is estimated by maximum likelihood methods.

Note that the sign pattern of the coefficients is the same one is observed for the linear model, however, calculating the change in the probability is not so simple as it was in a linear model. The derivative of the probability of crash with respect to the efficiency (H) varies with H:

$$\frac{\partial E(y)}{\partial H} = \frac{\exp(\alpha + \beta H_i)}{(1 - \exp(\alpha + \beta H_i))^2}\beta \tag{5.7}$$

The logit model can be expressed in odds-ratio as in equation 5.8 which is usually more intuitive:

$$\frac{p(y_i = 1)}{1 - p(y_i = 1)} = \exp(\alpha + \beta H_i) \tag{5.8}$$

$exp(\alpha+\beta H)$ is the effect of the independent variable (our measure of efficiency) in the "odds ratio".

5.3 Empirical Results for Different Stock Markets

The previous methodology is used in this section, studying five markets in order to find some facts about the relationship between the efficiency and the different crashes occurred in the past. Time series data for all the above indices were obtained with daily frequency[2], as it was mentioned before we take time-windows (v) for 100, 240, 350, and 420 days. Then, we proceed to symbolize the returns as it was shown in 5.1, after that we compute the normalized Shannon entropies for each v.

5.3.1 The Japanese stock market index (Nikkei 225)

Nikkei 225 is a stock market index for the Tokyo Stock Exchange (TSE , the second largest stock exchange market in the world by monetary volume). The Nikkei average is the most watched index of Asian stocks. It has been calculated daily by the Nihon Keizai Shimbun (Nikkei) newspaper since 1971. It is a price-weighted average (the unit is Yen), and the components are reviewed once a year.

The Japanese case was studied by Shiller et al. (1996) among others. Between 1982 and 1992 the Nikkei Index lost most of its value, after rising dramatically

[2] Daily data for different stock market indices (Nikkei 225, KLSE, IPC, DJIA) where obtained from finance.yahoo.com, RTS index was obtained from www.rts.ru

through the 1980s, fell from 38,915.9 on December 29, 1989 to 14,309.4 on August 18, 1992 (a decline of 63,2%).

Malkiel (2003) highlights that the prices of land and buildings started to increase until levels over the fundamental values, when the agents realized about their fundamental prices the bubble suddenly exploded.

Defining crash as the first percentile of the empirical density function for the Nikkei returns from September 17, 1985 to January 26, 2007 and considering the evolutions of the entropy for different time-windows, a logit model is applied. The pseudo-R^2 is taken as a measure of fit in order to select an appropriated time-window. Table 1 shows that a time-window of 240 days and words of 5 days provides the best fit.

Note in Table 1 that the optimal time-window seems to be 240, actually using a different methodology based on the Hurst exponent for DJIA, Grech and Mazur (2004) found that 240, a year is a good option. Note also that when we take a larger time-window the fit is worse. The latter happens because as the time-window increases in size, the entropy loses its locality.

Note in Table 2 that probability of crash depends negatively on our measure of informational efficiency. The latter suggests that when the market increases its efficiency, prices tend to reflect better the news, thus patterns are less frequent and the crash by adjustment is less likely.

Figure 5.3.1 shows the Nikkei 225 efficiency evolution, it presents the minimum efficiency for the whole period on October 13, 1987, just 7 days before famous 1987

crash. A local minimum is obtained on September 3, 1986, just 3 days before a crash, and September 12, 2001 (in this case, the crash is produced the same day).

Table 1: The Pseudo R^2 in different Logit Model specifications for Nikkei 225

Sequences	$v = 100$	$v = 240$	$v = 350$	$v = 420$
2-days	0.0054	0.0039	0.0011	0.0012
3-days	0.0029	0.0039	0.0011	0.0011
4-days	0.0037	0.0052	0.0014	0.0016
5-days	0.0040	**0.0089***	0.0013	0.0013

The logit model was estimated using the efficiency measure as independent variable, taking different sequence or combination of days (2, 3, 4, 5) and different time-windows (100, 240, 350, and 420 days). 16 models were estimated in total.

* The maximum Pseudo R^2 is shown in darker characters and highlights the logit model specification (time-window and sequence) which fits the best.

Source: Own Calculations

Table 2: The Logit Model for Nikkei 225 (Japan)

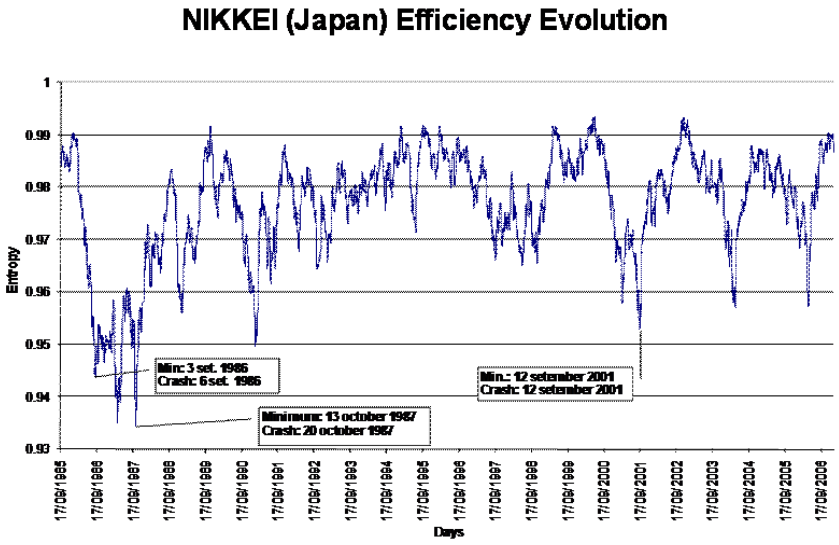
$p(y_t = 1) = \frac{\exp(\alpha + \beta H_t)}{1 - \exp(\alpha + \beta H_t)}$			Observations: 5256	
			LR Chi^2(1): 5.25	
			Prob>Chi^2 = 0.022$^{(a)}$	
Log Likelihood =-293.74			Pseudo-R^2 = 0.0089	
Crash Prob.$^{(c)}$	Coefficients	Stand. Error.(σ)	t=coeff./σ	p-value>$\lvert t \rvert$
Entropy (β)	-27.11	11.29	-2.40	0.02$^{(b)}$
Constant (α)	21.85	10.99	1.99	0.05$^{(b)}$
Crash Prob.$^{(c)}$	Odds-Ratio	Stand. Error.(σ)	t=coeff./σ	p-value>$\lvert t \rvert$
Entropy	1.68E-12	1.9E-11	-2.4	0.016$^{(b)}$

The results were obtained with STATA program.

(a) Indicates that the model is significative at 5%. (b) Indicates that the coefficients are significative at 5%. (c) Is the estimation of equation 5.6.

(d) The model expressed in odds-ratio as in equation 5.8

Source: Own Calculations



Daily Efficiency Evolution for Nikkei

5.3.2 The Malasyan stock market index (KLCI) The Kuala Lumpur Composite Index (KLCI) is the main stock market index containing 100 companies from the main board, and is now one of the three primary indices for the Malaysian stock market the "Bursa Malaysia" previously known as Kuala Lumpur Stock Exchange (KLSE).

According to Chowdhry and Goyal (2000) the financial crisis of East Asia in 1997 was unanticipated and made asset prices and currency values fall in several countries simultaneously.

Five countries were mainly affected (Indonesia, Malaysia, Philippines, South Korea, and Thailand). The event that triggered the crisis in East Asia was the announcement on July 2, 1997 that Thai Baht would be allowed to float.

Using the same procedure, Table 3 shows the optimal time-window, and opti-

mal sequence, according to the Pseudo-R^2 of the logit model. Note in this case, the best model fits for a time-window of 100 days, this is not a problem, as Grech and Mazur (2004) assert, sometimes different markets can have different time-window lengths[3]. This fact could indicate that this market assimilates the information in a faster way respect to the Japanese market.

As in the Japanese case, the efficiency affects negatively the probability of having a crash in the Malaysian stock market.

Note in Figure 5.3.2 that minimum is obtained on August 11, 1998 and a crash is produced on September 8, 1998.

[3] Actually, they suggest that the Warsaw market has a time length different from the US market.

Table 3: The Pseudo R^2 in different Logit Model specifications for KLCI

Sequences	$v = 100$	$v = 240$	$v = 350$	$v = 420$
2-days	**0.1202***	0.1062	0.0985	0.0645
3-days	0.1178	0.1133	0.1046	0.0694
4-days	0.1040	0.1145	0.1080	0.0717
5-days	0.0954	0.1011	0.1058	0.0664

The logit model was estimated using the efficiency measure as independent variable, taking different sequence or combination of days (2, 3, 4, 5) and different time-windows (100, 240, 350, and 420 days). 16 models were estimated in total.

* The maximum Pseudo R^2 is shown in darker characters and highlights the logit model specification (time-window and sequence) which fits the best.

Source: Own Calculations

Table 4: The Logit Model for KLCI (Malaysia)

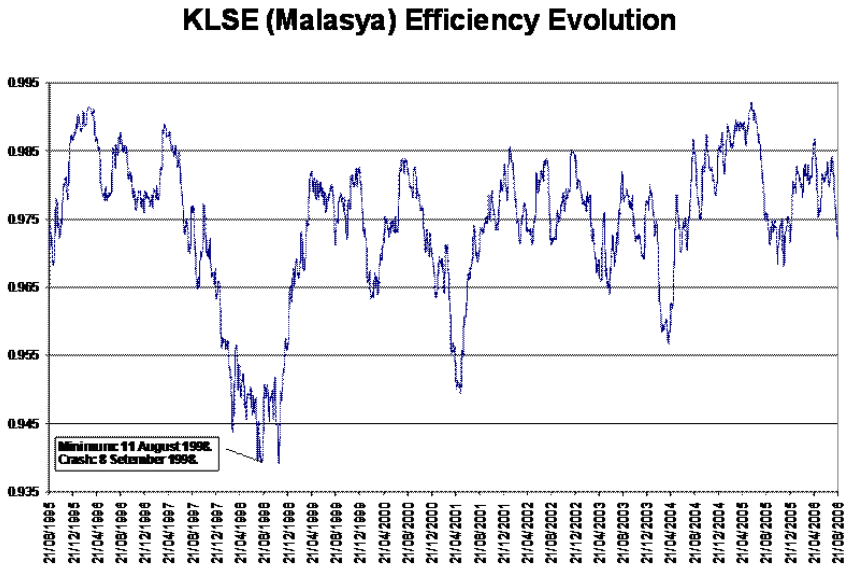
$p(y_t = 1) = \frac{\exp(\alpha + \beta H_t)}{1 - \exp(\alpha + \beta H_t)}$			Observations: 2716	
			LR Chi^2(1): 37.49	
			Prob>Chi^2 = 0.000$^{(a)}$	
Log Likelihood =-137.2			Pseudo-R^2 = 0.1202	
Crash Prob.$^{(c)}$	Coefficients	Stand. Error.(σ)	t=coeff./σ	p-value>$\lvert t \rvert$
Entropy (β)	-43.03	6.41	-6.71	0.00$^{(b)}$
Constant (α)	37.19	6.13	6.06	0.00$^{(b)}$
Crash Prob.$^{(c)}$	Odds-Ratio	Stand. Error.(σ)	t=coeff./σ	p-value>$\lvert t \rvert$
Entropy	2.05E-19	1.31E-18	-6.71	0.00$^{(b)}$

The results were obtained with STATA program.

(a) Indicates that the model is significative at 5%. (b) Indicates that the coefficients are significative at 5%. (c) Is the estimation of equation 5.6.

(d) The model expressed in odds-ratio as in equation 5.8

Source: Own Calculations



Daily Efficiency Evolution for KLSE

5.3.3 The Russian Stock Index, The Russian Trading System index (RTS) The Russian Trading System is a stock market established in 1995 in Moscow, consolidating various regional trading floors into one exchange. At the moment RTS is in the process of reorganization, it is being transformed into a joint-stock company.

According to Sutela (2000) the Russian crisis was connected with the earlier Asian crisis, and sent shock waves across global financial markets. In August 1998, Russia experienced a currency crisis combined with banking crisis and debt crisis. In August short-term capital started to leave the country and the exchange rate RUR/USD passed from 6 to 20-25. However, the Moscow Stock Exchange started to decline since October 1997. In fact, before the crashes in August 27, 1998 (-18.78%) and May 12, 1999 (-17,66%) there was another important crash on October 28, 1997 (-21.10%). We take daily data from September 1, 1995 to

August 23, 2006 and apply the same method.

In the Russian crash the model seems to have a high significance. It shows that efficiency affects negatively the probability of crashes (Tables 5 and 6).

Figure 5.3.3 shows that a minimum is produced on August 1997 and then crashes are produced on October 1997, January 1997 and August 1998. Minimum for all the period is produced on May 15, 2006 and crashes are produced on May 20, 2006 and June 13, 2006.

Table 5: The Pseudo R^2 in different Logit Model specifications for the RTS

Sequences	$v = 100$	$v = 240$	$v = 350$	$v = 420$
2-days	0.0549	0.0538	0.0181	0.0179
3-days	0.0635	0.0863	0.0463	0.0442
4-days	0.0662	0.1057	0.0659	0.0611
5-days	0.0523	**0.1126***	0.0745	0.0743

The logit model was estimated using the efficiency measure as independent variable, taking different sequence or combination of days (2, 3, 4, 5) and different time-windows (100, 240, 350, and 420 days). 16 models were estimated in total.

* The maximum Pseudo R^2 is shown in darker characters and highlights the logit model specification (time-window and sequence) which fits the best.

Source: Own Calculations

Table 6: The Logit Model for the RTS (Russia)

$p(y_t = 1) = \frac{\exp(\alpha + \beta H_t)}{1 - \exp(\alpha + \beta H_t)}$	Observations: 2322
	LR Chi^2(1): 29.05
	Prob>Chi^2 = 0.000$^{(a)}$
Log Likelihood =-114.5	Pseudo-R^2 = 0.113

Crash Prob.$^{(c)}$	Coefficients	Stand. Error.(σ)	t=coeff./σ	p-value>$\lvert t \rvert$
Entropy (β)	-83.30	16.56	-5.03	0.00$^{(b)}$
Constant (α)	75.35	15.77	4.78	0.00$^{(b)}$
Crash Prob.$^{(c)}$	Odds-Ratio	Stand. Error.(σ)	t=coeff./σ	p-value>$\lvert t \rvert$
Entropy	6.63E-37	1.10E-35	-5.03	0.00$^{(b)}$

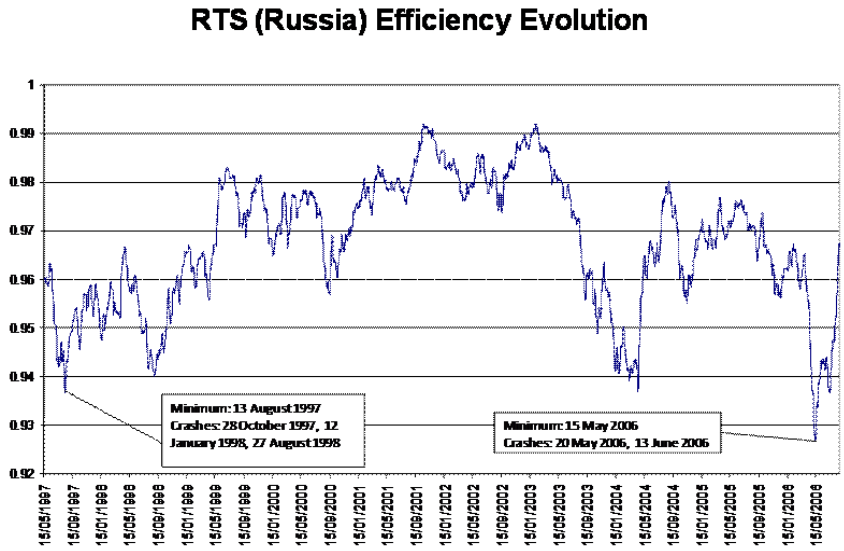
The results were obtained with STATA program.

(a) Indicates that the model is significative at 5%. (b) Indicates that the coefficients are significative at 5%. (c) Is the estimation of equation 5.6.

(d) The model expressed in odds-ratio as in equation 5.8

Source: Own Calculations

In the Russian crash the model seems to have a high significance. It shows that efficiency affects negatively the probability of crashes.



Daily Efficiency Evolution for RTS

5.3.4 The Mexican Stock Market Index, Indice de Precios y Cotizaciones (IPC)

The Indice de Precios y Cotizaciones (IPC) is computed by The Bolsa Mexicana de Valores or BMV, the Mexico's only stock exchange. It is the second largest Stock Exchange in Latin America, behind the São Paulo Stock Exchange.

Here it is interesting to analyze if the crisis in 1994 was originated in a high period of inefficiency. Mexico had suffered other crisis in 1982, however the Tequila Crisis is famous for being the first global crisis affecting other stock markets.

According to Calvo (1996), by the end of 1994 the country was susceptible to speculative attacks. In fact, he marks that the crisis started to develop in February/March 1994. The curious thing is just a few month, if not days before the collapse, there was a strong consensus that Mexico had finally graduated into the first world.

Note in Figure 5.3.4 a minimum is obtained on February 10, 1994 after crashes on 4 and 20 April, 1994, January and February 1995. Note also that this is a very low level of efficiency (around 0.8, even the lowest comparing with the other markets) indicating the formation of strong patterns.

The result is always the same, Table 7 shows an optimal time-window at 100 days and the Table 8 shows that the probability of having a crash depends negatively on informational efficiency.

Table 7: The Pseudo R^2 in different Logit Model specifications for the IPC

Sequences	$v = 100$	$v = 240$	$v = 350$	$v = 420$
2-days	0.0050	0.0004	0.0006	0.0009
3-days	0.0066	0.0017	0.0029	0.0042
4-days	0.0069	0.0034	0.0057	0.0066
5-days	**0.0159***	0.0089	0.0079	0.0066

The logit model was estimated using the efficiency measure as independent variable, taking different sequence or combination of days (2, 3, 4, 5) and different time-windows (100, 240, 350, and 420 days). 16 models were estimated in total.

* The maximum Pseudo R^2 is shown in darker characters and highlights the logit model specification (time-window and sequence) which fits the best.

Source: Own Calculations

Table 8: The Logit Model for the IPC (Mexico)

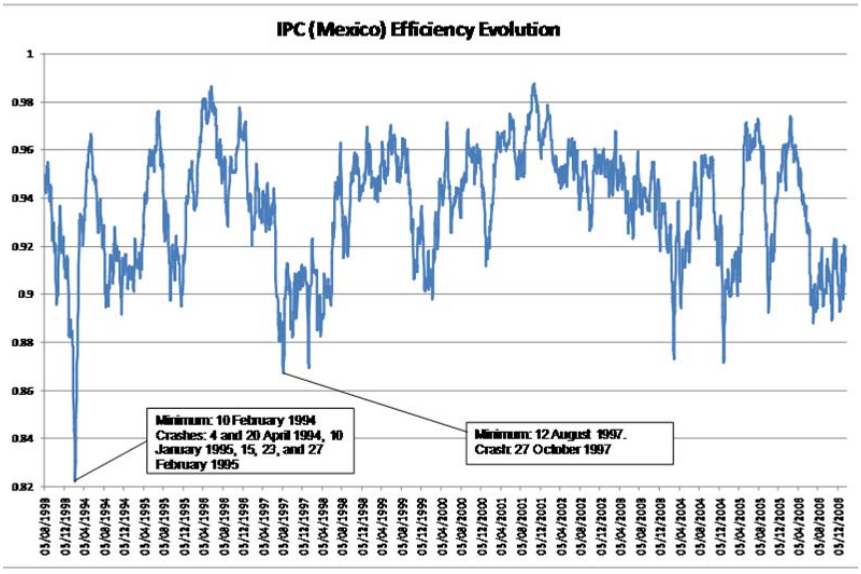
$p(y_t = 1) = \frac{\exp(\alpha + \beta H_t)}{1 - \exp(\alpha + \beta H_t)}$			Observations: 3376	
			LR Chi^2(1): 6.06	
			Prob>Chi^2 = 0.014$^{(a)}$	
Log Likelihood =-187.13			Pseudo-R^2 = 0.0159	
Crash Prob.$^{(c)}$	Coefficients	Stand. Error.(σ)	t=coeff./σ	p-value>$\lvert t \rvert$
Entropy (β)	-16.06	6.33	-2.54	0.01$^{(b)}$
Constant (α)	10.36	5.86	1.77	0.08$^{(b)}$
Crash Prob.$^{(c)}$	Odds-Ratio	Stand. Error.(σ)	t=coeff./σ	p-value>$\lvert t \rvert$
Entropy	1.06E-07	6.72E-07	-2.54	0.01$^{(b)}$

The results were obtained with STATA program.

(a) Indicates that the model is significative at 5%. (b) Indicates that the coefficients are significative at 5%. (c) Is the estimation of equation 5.6.

(d) The model expressed in odds-ratio as in equation 5.8

Source: Own Calculations



Daily Efficiency Evolution for IPC

5.3.5 The US Stock Market index, Dow Jones Industrial Average (DJIA) The Dow Jones Industrial Average (DJIA) was created by Charles Dow, the Wall Street Journal editor, and the Dow Jones & Company co-founder. Dow compiled the index as a way to gauge the performance of the industrial component of America's stock markets. It is the oldest continuing US market index, and consists of 30 of the largest and most widely held public companies in the US.

Table 9: The Pseudo R^2 in different Logit Model specifications for the DJIA

Sequences	$v = 100$	$v = 240$	$v = 350$	$v = 420$
2-days	0.0002	0.0004	0.0012	0.0004
3-days	0.0000	0.0001	0.0010	0.0002
4-days	0.0001	0.0000	0.0000	0.0000
5-days	**0.0020***	0.0002	0.0003	0.0008

The logit model was estimated using the efficiency measure as independent variable, taking different sequence or combination of days (2, 3, 4, 5) and different time-windows (100, 240, 350, and 420 days). 16 models were estimated in total.

* The maximum Pseudo R^2 is shown in darker characters and highlights the logit model specification (time-window and sequence) which fits the best.

Source: Own Calculations

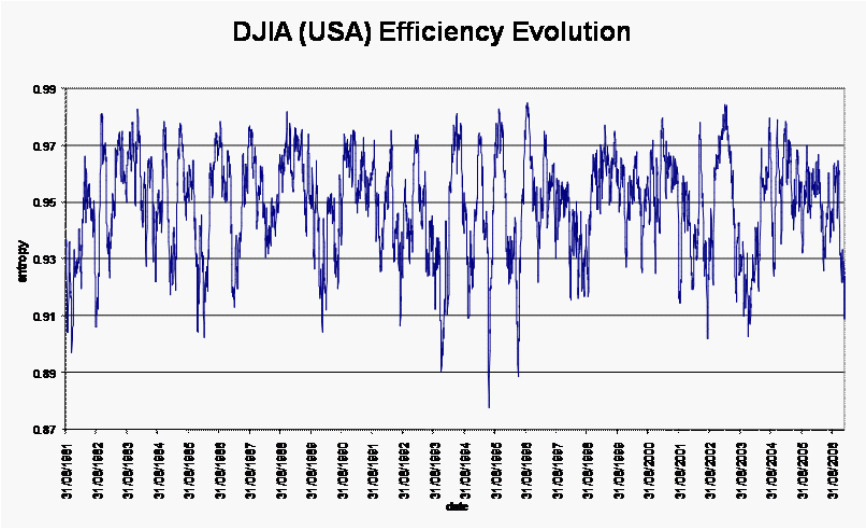
The results are similar to the previous one (Tables 9 and 10), an increase in the informational efficiency produces a reduction in the probability of having a crash. However, the coefficients of the model are not so significative, maybe because the US market is very efficient and then the econometric model hardly captures the relation between inefficiency and crash (Figure 5.3.5).

Table 10: The Logit Model for the DJIA (USA)

$p(y_t = 1) = \frac{\exp(\alpha + \beta H_t)}{1 - \exp(\alpha + \beta H_t)}$			Observations: 6412	
			LR Chi^2(1): 1.44	
			Prob>Chi^2 = 0.2296	
Log Likelihood =-357.82			Pseudo-R^2 = 0.0020	
Crash Prob.$^{(a)}$	Coefficients	Stand. Error.(σ)	t=coeff./σ	p-value>$\|t\|$
Entropy (β)	-8.31	6.91	-1.20	0.230
Constant (α)	3.28	6.55	0.50	0.617
Crash Prob.$^{(b)}$	Odds-Ratio	Stand. Error.(σ)	t=coeff./σ	p-value>$\|t\|$
Entropy	0.0002467	0.001706	-1.20	0.230

The results were obtained with STATA program.

(a) Is the estimation of equation 5.6. (b) The model expressed in odds-ratio as in equation 5.8. Source: Own Calculations



Daily Efficiency Evolution for DJIA

5.4 Global Effect

The purpose of this section is to study the relationship between the crash probability and the informational efficiency in the global market by introducing a logit model. This model considers the five markets in the same periods as shown in equation (5.9). Four dummy variables are introduced for the countries (MEX, MAL, JAP, RUS) in order to recover the structural differences among the countries. The crash is defined by taking all the returns in the markets and defining a common limit where the returns cumulate 1% of the frequency, it happens at -0.0441. On the other hand, we consider time-windows of 100, 240, 350, and 420 days and words of 2, 3, 4, and 5 days. The minimum Akaike was obtained for a window of 350 days and 5 days sequence, the Table 11 shows the results.

$$p(y=1) = \frac{\exp(\alpha+\beta_0 H+\beta_1 MEX+\beta_2 MAL+\beta_3 JAP+\beta_4 RUS)}{1-\exp(\alpha+\beta_0 H+\beta_1 MEX+\beta_2 MAL+\beta_3 JAP+\beta_4 RUS)} \tag{5.9}$$

Note that the sign of the efficiency is the correct, an increase in the informational efficiency reduces the global probability of having a crash. On the other hand, since USA is the variable of control, the other signs are compared with this country.

Table 11: Logit Model for the Five Stock Markets

$p(y=1) = \frac{\exp(\alpha+\beta_0 H+\beta_1 MEX+\beta_2 MAL+\beta_3 JAP+\beta_4 RUS)}{1-\exp(\alpha+\beta_0 H+\beta_1 MEX+\beta_2 MAL+\beta_3 JAP+\beta_4 RUS)}$			Observations =11552	
			LR Chi^2(5) = 235.27	
			Prob>Chi^2	0.0000$^{(a)}$
Log Likelihood = -801.75			Pseudo-R^2 = 0.1280	
Crash Prob.$^{(c)}$	Coefficients	Stand. Error (σ)	t=coeff./σ	p-value>$\lvert t \rvert$
Efficiency (β_0)	-55.07	8.56	-6.44	0.000$^{(b)}$
MEX (β_1)	0.67	0.48	1.41	0.160
MAL (β_2)	1.01	0.46	2.18	0.029$^{(b)}$
JAP (β_3)	0.79	0.50	1.58	0.113
RUS (β_4)	2.41	0.43	5.56	0.000$^{(b)}$
Constant (α)	48.60	8.48	5.73	0.000$^{(b)}$

The results were obtained with STATA program.

(a) Indicates that the model is significative at 5%. (b) Indicates that the coefficient are significative at 5%. (c) is the estimation of equation 5.9.

Source: Own Calculations

From Table 11 we can deduce that our hypothesis in the section before seems to be true, note that the US market is the most structural efficient market, jumping from the US market to a different market increases the probability of having a crash, being Russia (RUS) the most structurally inefficient stock market in the period. The latter result can be explained by the fact that the Russian stock

market is the youngest, established in the 1995, as it was mentioned in subsection 5.3.3. This also agrees with the results obtained in the Chapter 4 where eastern European stock markets presented the lowest efficiency.

5.5 Theoretical relation between Efficiency and News arrival

In this section we will try to give an explanation about the relation between news arrival and information efficiency and how it might be determining the crashes. As mentioned above, an informational efficient market should immediately assimilate the news arrivals, hence no pattern in prices should be formed. Many scholars has considered the random walk process in order to modelling asset prices. We will assume that if the market is completely efficient the probability of having a positive or negative return tomorrow should be the same, it means 1/2. In terms of our entropic measure of efficiency should produce the maximum entropy as in the following equation:

$$H(t) = -\left(\left(\frac{1}{2}\right)\log_2\left(\frac{1}{2}\right) + \left(\frac{1}{2}\right)\log_2\left(\frac{1}{2}\right)\right) = -\log_2\left(\frac{1}{2}\right) = 1$$

It means that efficiency always will be the maximum (1) because there is no reason for predicting a negative or positive return tomorrow. Consider now, a process ε_t, representing the news arrivals in the market. If they are good news then probability of having positive returns tomorrow should increase, on the other hand if they are bad news the probability of having negative returns tomorrow should increase. However, in an efficient market this effect should be assimilated

the following day. It means, the news arrive and the market can "forecast" the returns at least that day, of course in our model the efficiency decreases for that day because probability of having positive or negative returns increases, nonetheless the news should be assimilated and the following day the efficiency come back to the maximum value (1).

Consider that news arrivals are independent and identically distributed as a Poisson process with parameter λ.

$$u(t) \text{ is } \text{ i.d.d.} Poisson(\lambda)$$

This news will affect the efficiency as a noise normalized by a parameter, reducing the efficiency for once and then the efficiency recovers its maximum value. Note that $u(t)$ takes value 0 or 1, however the effect of the news in the efficiency should not be larger than 1/2, then we define $\varepsilon(t) = \delta u(t)$, where δ is the impact in the efficiency. Therefore we can redefine our efficiency measure under news arrivals:

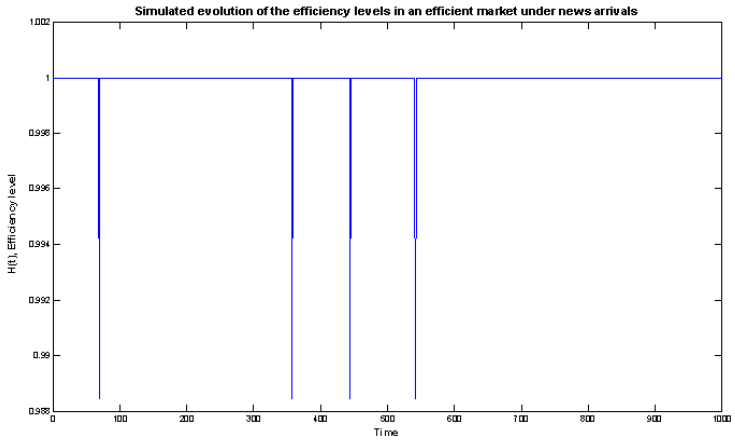
$$H(t) = -\left(\left(\frac{1}{2}+\varepsilon(t)\right)\log_2\left(\frac{1}{2}+\varepsilon(t)\right)+\left(\frac{1}{2}-\varepsilon(t)\right)\log_2\left(\frac{1}{2}-\varepsilon(t)\right)\right)$$

Note that $\log_2\left(\frac{1}{2}+\varepsilon(t)\right) = \log_2(1+2\varepsilon(t)) - 1 \simeq 2log_2(e)\varepsilon(t) - 1$, similarly $\log_2\left(\frac{1}{2}-\varepsilon(t)\right) \simeq -2log_2(e)\varepsilon(t) - 1$. Then, our measure can be approximated as the following equation:

$$H(t) \simeq H^*(t) = 1 - \gamma\varepsilon(t)^2$$

where $\gamma = 4log_2(e)$, now the relation is clear, in an efficient market the efficiency should be always equal to 1 (the maximum), however when news arrive, they affect the efficiency for once and the efficiency comes back to its maximum level because new information is immediately assimilated.

The following Figure is a simulation of this process for $T = 1000$, $\delta = 0.002$, and $\lambda = 0.005$.



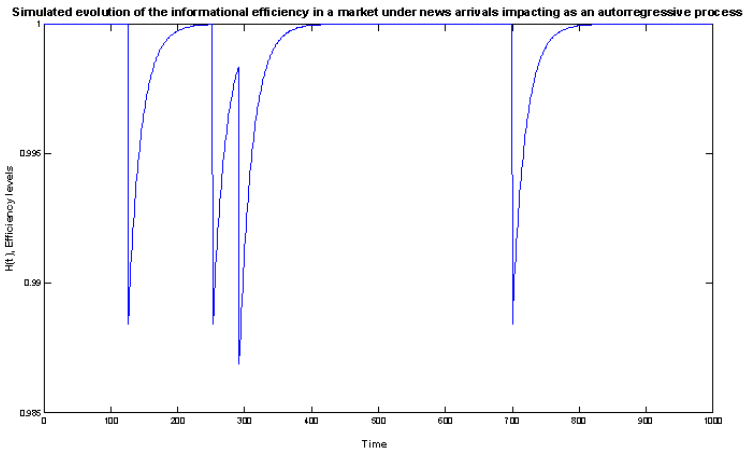
In the example of 1000 days, only four news arrived to the market. Note, that the efficiency is impacted by the news only for once, and then it recovers achieving its maximum level.

Assume now that news arrivals are not well understood and so they arrive in the form of an autorregressive process:

$$\varepsilon(t) = \alpha\varepsilon(t-1) + \delta u(t)$$

where $\delta u(t)$ is again the Poisson process of information arrival corrected by the impact factor δ. On the other hand, α is the autorregressive coefficient less than 1

which defines the memory of the process.



Note that now the information is not well embodied and then inefficiency remains for more time. Note also that the efficiency was recovering from the second news impact and before arriving to the maximum levels some news arrive impacting a third time the efficiency. In order to obtain a more realistic approach to the efficiency evolution, assume that the efficiency is also affected by the absolute value of random and exogenous factor $\psi(t)$, normal distributed with mean equal to 0 and variance σ_ψ, now our stochastic system is the following:

$$H(t) \simeq H^*(t) = 1 - \gamma\varepsilon(t)^2 - |\psi(t)|$$

$$\varepsilon(t) = \alpha\varepsilon(t-1) + \delta u(t)$$

Simulating the model for $T = 1000$, $\lambda = 0.005$, $\delta = 0.002$, $\sigma_\psi = 0.002$, $\alpha = 0.95$, we obtain the following figure.

Note that now, the evolution of the efficiency is similar to the results obtained for real data. Therefore, we can say that a great part of the inefficiency can arrive from the arrival of not well understood news.

5.6 Conclusions

For long time the EMH has been the central proposition in Finance. Jensen (1978) affirmed that there was no other proposition in economics which had more solid empirical evidence supporting it than the EMH. However, many scholars started to show that the market not always behaves in this manner, sometimes presenting mispricing or anomalies, see Singal (2004).

Considering the possibility of finding deterministic chaos in the financial markets, or at least fractal Brownian motion some authors proposed the Hurst exponent as a measure of the efficiency, see Peters (1994) (1996). Even more, Grech and Mazur (2004) propose to use the Hurst exponent for measuring the evolution of the efficiency through the time. Some other authors, such as Bassler et al. (2006)

and McCauley et al. (2007) criticize this measure asserting that Hurst exponent does not necessarily detect long time correlations like those found in fractional Brownian motion.

In the present chapter we applied the measure introduced in chapter 3 in order to study the evolution of the informational efficiency through the time. As explained in chapter 3, the intuitive idea is simple, using STSA we try to recover the very dynamic of the process and later by applying the entropy we try to measure the quantity of information embodied in the sample. In order to study the evolution of the previous measure we take a time-window, the local entropy for the symbolic series is computed obtaining a daily time series of efficiency. A logit model is applied in order to study the relationship between the informational efficiency and the probability of having a crash. Using data for the Japanese, Malaysian, Russian, Mexican, and US markets, the informational measure is computed and a logit model is applied. The appropriated time-windows seem to be different among country but going from half a year to a trading year. This fact could be reflecting how much fast the market reacts or assimilates the new information. On the other hand, the results of the logit models are always the same, the lower the informational efficiency, the higher the probability of having a crash. It means that a market presenting a short-time trend ("bull" or "bear") will produce a reduced local entropy (because some patterns will be more frequents). The results also seem to confirm that US market is the most efficient, being the Russian market the most inefficient, because it is also the youngest, starting to work in 1995. This

result agrees with conclusion obtained in chapter 4 about the Russian market as the most inefficient. This measure could be useful as a tool in controlling the daily evolution of informational efficiency not only in the stock markets but also in the exchange market, for instance trying to forecast the proximity of a devaluation.

5.7 References

-Alligood, K., Sauer, T., Yorke, J., (1997), *Chaos: An Introduction to Dynamical Systems*, Spring-Verlag, New York.

-Bassler, K., Gunaratne, G., McCauley, J., (2006), "Markov processes, Hurst exponents, and nonlinear diffusion equations: With application to finance", *Physica A*, Vol. 369, No. 2, pp. 343-353.

-Calvo, G., (1996), "Capital Flows and Macroeconomic Management: Tequila Lessons", *International Journal of Finance and Economics*, John Wiley & Sons, Ltd., Vol. 1, pp. 207-223.

-Chowdhry, B., Goyal, A., (2000), "Understanding the financial crisis in Asia", *Pacific-Basin Finance Journal*, Vol. 8, pp. 135-152.

-Daw, C., Finney, C., Tracy, E., (2003), "A review of symbolic analysis of experimental data", *Review of Scientific Instruments*, Vol. 74, No. 2, pp. 915-930.

-Grech, D., Mazur, Z., (2004), "Can one make any crash prediction in finance using the local Hurst exponent idea?", *Physica A*, Vol. 336, pp. 133-145.

-Hsieh, D., (1995), "Non-linear dynamic in financial markets: evidence and implications", *Financial Analysis Journal*, Vol. 51, pp. 55-62.

-Hsieh, D., (1991), "Chaos and non-linear dynamics: application to financial

markets", *Journal of Finance*, Vol. 46, pp. 1833-1877.

-Jensen, M., (1978), "Some anomalous evidence regarding market efficiency", *Journal of Financial Economics*, Vol. 6, pp. 95-101.

-Johnston, J., DiNardo, J., (1997), *Econometric Methods*, McGraw Hill, fourth edition.

-Lawrence, S., Ah Chung, T., Lee Giles, C., (1998), *Noisy Time Series Prediction using Symbolic Representation and Recurrent Neural Network Grammatical Inference*. NEC Research Institute, Princeton.

-Khinchin, A., (1957), *Mathematical Foundations of Information Theory,* Courier Dover Publications.

-LeBaron, B., (1994), "Chaos and Nonlinear Forecastability in Economics and Finances", *Philosophical Transactions: Physical Sciences and Engineering*, Vol. 348, No.1688, Chaos and Forecasting (Sep. 15, 1994), pp. 397-404.

-**Lo, A.W., 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.

-Malkiel, B., (2003), *A Random Walk Down Wall Street*, W. W. Norton & Company, Inc.

-McCauley, J., Gunaratne, G., Bassler, K., (2007), "Hurst Exponents, Markov Processes, and Fractional Brownian Motion", *Physica A*, Vol. 379, No. 1, pp. 1-9.

-McFadden, D., (1984), "Econometric Analysis of qualitative Choice Mod-

els", Chapter 24, *Handbook of Econometrics*, eds. Giliches and Intriligator, North-Holland.

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

-Peters, E., (1994), *Fractal Market Analysis*, first edition, Wiley Finances Edition.

-Schittenkopf, C., Tiňo, P., Dorffner, G., (2002), "The benefit of information reduction for trading strategies", *Applied Economics*, Vol. 34, pp. 917-930.

-Shiller, R., Kon-Ya, F., Tsutsui, Y., (1996), "Why Did the Nikkei Crash? Expanding the Scope of Expectations Data Collection". *The Review of Economic and Statistics*, Vol. 78, No. 1, pp.156-164.

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

-Sutela, P., (2000), "The Financial Crisis in Russia", in Bisignano, J., Hunter, W., and Kaufman, G., eds.: *Global Financial Crisis: Lessons from Recent Events*, Kluwer, Boston, 2000, pp. 63-73.

CHAPTER 6

Efficiency Across the Stock Market

6.1 Introduction

The purpose of this chapter is to study how the information expands through a given stock market. In an efficient market, the news affecting a particular company should be immediately embodied in all the related companies, moving the firms in the same direction. Suppose that Intel announces a powerful new processor, the news should make the stock prices increase in Intel but also in companies producing computers (IBM, Hewlett-Packard), and companies producing software (i.e. Microsoft). However, if the market is inefficient the companies may not understand the news, and information could be incorporated at different moments.

In order to study the efficiency across the market a methodology is developed. The method is based on the symbolic analysis explained before and the graph theory (in especial the Minimal Spanning Tree (MST) and the Hierarchical Tree (HT) are modified). We try to study the structure and dynamics of the stock market, if the market is efficient it should present clusters of related companies. These groups of firms react to the information moving in the same way. However, an inefficient market should present no cluster inside, since information is embody at different times.

The structure of the chapter is as follows. In the section 6.2 the methodology is introduced, the new things are two; On the one hand, we are able to study

the structure of the market using information from more than one variable, on the other hand we can study different scenarios, for instance we can analyze the structure in a normal situation and in a extreme situation. The section 6.3 tries to establish the importance of the relation between stock prices and volume trade, we will use these variables later, in the study of the stock markets. Section 6.4 is an empirical application to the US stock market, and section 6.5 is an application to the Italian Stock market. Finally, section 6.6 draws some conclusions.

6.2 Multidimensional Symbolic Minimal Spanning Tree (MSMST)

6.2.1 Introduction to Taxonomy

The first systematic classification of objects and things comes from Biology, and more precisely the Zoology. The classification of living beings in different classes according to the "natural system" is due to Linneo. All the classifications try to separate the individuals in classes or groups in such a way that individuals belonging to the same group are "similar" among them and "different". The concepts of "similarity" and "different" are not defined in a precise form, but everybody understand what they refer. It is clear that a frog is "similar" to a toad, and "different" from a rabbit, even when we know that also frogs and toads are different. This science of classifying living organisms is called Taxonomy. The present Chapter will introduce a methodology in order to determine the taxonomy of a set of elements by using symbolic analysis. The flexibility advantage of symbolic method will permit us to determine the structure of a set of elements in different situations and use many variables.

6.2.2 Minimal Spanning Tree and Hierarchical Tree Mantegna (1999) proposed to study the structure and taxonomy of the stock markets by constructing the Minimal Spanning Tree (MST) and Hierarchical Tree (HT). These trees come from the Graph Theory, a mathematical study of the properties of the formal mathematical structures called graphs. A graph, denoted by G, is a mathematical object composed of points, known as vertices or nodes, and lines connecting them, known as edges. Trees are one of the most important types of graphs with many applications (i.e. family trees, organization charts, electrical networks, and often railway lines). Note that such a graph is a tree if and only if there is a unique simple path between any two of its vertices. A spanning tree is a tree containing, or spanning, all the N vertices of the graph, and therefore it must have $N-1$ edges. The MST is the smallest such tree in a connected weighted graph, constructed so that the sum of all edge weights is at minimum. There exist two popular algorithms for the MST problem, the Kruskal's and the Prim's algorithms. Onnela (2002) describes a pseudo-code for the Kruskal's algorithm, it is shown in Table 1

Table 1: Kruskal's algorithm for Minimum Spanning Tree

begin Kruskal;

sort edges so that $\omega(e_1) \leq ... \leq \omega(e_K)$;

LIST=$\emptyset$;

while $|\text{LIST}| < N - 1$ do

begin

if the next edge e_i does not create a cycle then add it to LIST

else discard it

end;

end;

Where N are the number of elements, e_i is for edge i, and $\omega(e_i)$ is the weight of edge i

These trees present a net among the different elements, highlighting the relevant connections among them, and the most important clusters. A metric distance is necessary in order to obtain this taxonomic representation; i.e., a function d defined for each pair of time series that takes values in $\mathbb{R}$ such that:

1. $d(i,j) \geq 0 \ \forall i,j$

2. $d(i,j) = 0$ if and only if $i = j$

3. $d(i,j) = d(j,i) \ \forall i,j$

4. $d(i,j) \leq d(i,k) + d(k,j) \ \forall i,j,k$

Computing all the distances between elements permits to construct the distance matrix D. This symmetric matrix determines the minimal spanning tree connecting the n elements of a set showing the most relevant connections. The methodology proposed by Mantegna (1999) for studying the financial markets, is based on the Pearson correlation coefficient and the distance function proposed by Gower (1966). Equation 6.1 is the Pearson correlation coefficient.

$$\rho_{ij} = \frac{\langle r_i r_j \rangle - \langle r_i \rangle \langle r_j \rangle}{\sqrt{\left(\langle r_i^2 \rangle - \langle r_i \rangle^2\right)\left(\langle r_j^2 \rangle - \langle r_j \rangle^2\right)}} \tag{6.1}$$

where r_i and r_j are the asset returns of firms i and j. This coefficient is a temporal average performed on all the trading days of the investigated time period. By definition, ρ_{ij} can vary from -1 (completely anti-correlation) to 1 (completely correlation). Gower (1966) proposes to compute the distances between companies by using the correlation coefficient presented in 6.1. Equation 6.2 is the distance propossed by Gower.

$$d(i,j) = \sqrt{2(1-\rho_{ij})} \tag{6.2}$$

The next step is to use these distances to construct the distance matrix D, a symmetric matrix which shows all the distances among the different companies or elements of the set. This matrix permits to construct the Minimal Spanning Tree (MST) connecting the set of time series. The MST is progressively constructed by linking all the time series together in a graph characterized by a minimal distance between time series, starting with the shortest distance. This method is the Kruskal's algorithm also called single linkage (nearest neighbor) presented in Table

1.

In the first step, we choose the pair of time series with the nearest distance and we connect them. In the second step we also connect a pair with the second nearest distance with a line proportional to the distance. In the third step we also connect the nearest pair that is not connected by the same tree. We repeat the third step until all the given companies are connected in a unique tree. MST is attractive because provides an arrangement of asset returns which selects the most relevant connections of each element of the set.

The MST permits to obtain the subdominant ultrametric distance matrix $d^<$. This matrix, as Mantegna (1999) asserts can be constructed from the ultrametric distance $d^<(i,j)$. According to Mantegna (1999) and Mantegna and Stanley (2000) the subdominant ultrametric distance $d^<(i,j)$ between i and j is the maximum value of any Euclidean distance $d_k(l;m)$ detected by moving in single steps from i to j through the shortest path connecting i and j in the MST. The ultrametric distance $d^<$ is used to construct the Hierarchical Tree (HT). One method to obtain $d^<(i,j)$ directly from the distance matrix $d_k(i,j)$ is through the MST method as described in Ramal et al. (1986). From the MST, the distance $d^<(i,j)$ between two companies i and j is given by

$$d^<(i,j) = Max\left\{d_k(w_i;w_{i+1}); 1 \leq i \leq n-1\right\} \tag{6.3}$$

where $\{(w_1;w_2);(w_2;w_3);\ldots;(w_{n-1},w_n)\}$ denotes the unique path in the MST connecting i and j, where $w_1 = i$ and $w_n = j$. Note the following property is

Links	Dist.
D(i,k)	1
D(k,h)	2
D(h,j)	3

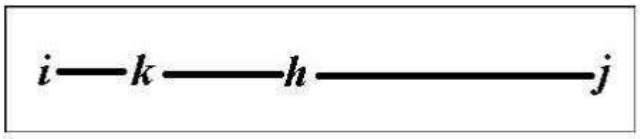


Figure 6.1: Example of MST and distances

satisfied in the ultrametric distances.

$$d^{<}(i,j) \leq \max\{d_k(i,l), d_k(l,j)\} \tag{6.4}$$

For instance, assume that Figure 6.1 is a MST between 4 elements where links are the minimal distance among elements as explained before. As Onnela (2002) asserts that an optimal portfolio can be constructed by choosing the points that are as far from each other as possible, which is equivalent to choosing those points that span the maximal volume in this space. In this case, note that ultrametric distance $d^{<}(i,j) = d(h,j) = 3$. From these trees we can obtain both geometrical (throughout the MST) and taxonomic (throughout the hierarchical tree) information of the correlation present between the elements of the set. This methodology has demonstrated useful insights on the global structure, taxonomy and hierarchy in the dynamics of the financial data, specially on the stock markets, but also in the exchange markets. (see Ortega and Matesanz (1999), Mantegna (1999), Kaski et al. (2003), Mizuno et al. (2006), Bonanno et al. (2001), Bonanno et al. (2004)). In the next section we modified this methodology in order to embody information from more than one variable, and study different scenarios.

6.2.3 Multidimensional Symbolic Minimal Spanning Tree (MSMST) As explained above, the original methodology was proposed by Mantegna (1999) in order to study structure and hierarchy in the financial markets. The present section is an extension of the latter method, trying to give more flexibility to this methodology by using Symbolic Analysis. As far as we know, until now the method suggested by Mantegna (1999) uses only one variable in order to obtain the structure and taxonomy of the stock markets and analyses only one scenario. Actually, every work using MST in financial markets basically focuses on financial returns. However, we may loss information if more than one variable were important in the construction of the financial market structure. No method has been applied in order to derive the MST incorporating information from more than one variable. Nonetheless, as it will be explained later, many works show that there exists a relationship between returns and trading volume. In the Wall Street tradition is well known that it takes volume in order to move the prices, highlighting the existence of a positive correlation between trading volume and absolute value of returns. Even more, it seems that in "bull market" the volume is heavy and it is light in "bear markets" suggesting a positive correlation between returns and volume trading. Since volume trading seems to give important information to the market, the introduced methodology aims to embody information providing not only from returns but also from volume trading. With this purpose, this section generalizes the Minimal Spanning Tree introduced in the previous section, into a Multidimensional Symbolic Minimal Spanning Tree (MSMST).

Basically, we modified the distance used in order to construct the MST. As will be explained later, the euclidean distance is applied after an appropriated symbolization of the dataset.

Assume that in the construction of certain structure is important to consider the following multidimensional time series for each element i:

$$\{\mathbf{X}_i\}_{t=1}^{t-T} = \left\{ \begin{pmatrix} x_{i1} \\ y_{i1} \\ . \\ z_{i1} \end{pmatrix}, \begin{pmatrix} x_{i2} \\ y_{i2} \\ . \\ z_{i2} \end{pmatrix}, ..., \begin{pmatrix} x_{iT} \\ y_{iT} \\ . \\ z_{iT} \end{pmatrix} \right\} \tag{6.5}$$

Of course, time series 6.5 is continuous measure but we can pass to one-dimensional symbolic space $\mathbf{S}$, by defining a determined partition in the multidimensional space $\mathbb{R}^n$, obtaining the following symbolic time series for each element i:

$$\{s_{i1}, s_{i2}, ..., s_{iT}\} \tag{6.6}$$

The key step in applying symbolization to time series measurements involves transforming the original values into a sequence of symbols. Daw et al. (2003) highlight that symbolizations permits to reduce noise in highly contaminated time series, and of course, asset returns are not the exception. Therefore, we have to select the partition which will define the regions assigning a symbol to each measurement according to the region on which it falls into. In words, if we start with a given set of measurements $\{x_1, x_2, \ldots, x_t, \ldots, x_T\}$ made up of vectors $x_t \in D \subset \mathbb{R}^q$,

for $t = 1, 2, \ldots, T$ and the state space $\mathbb{R}^q$ is endowed with a suitable partition, then, we transform the sequence of data $\{x_1, x_2, \ldots, x_t, \ldots, x_T\}$ into the sequence of symbols $s_1 s_2 \ldots s_t \ldots s_T$, where $s_t = s$ if and only if x_t belongs to the regime region labeled by s. This converts the original signal into a symbolic sequence, from which the symbol sequence statistics can be estimated. This process of transformation of data into a symbolic sequence is called *symbolization* in the Symbolic Time Series Analysis (STSA) literature and can be done in several ways. For example, the simplest scheme is to assign values of 0 and 1 to each observation depending on whether it is above or below some critical value (binary partition). In some applications, we can define discretization partitions such that 1) the occurrence frequency of any particular symbol is equiprobable with all others (see Tang and Tracy (1997)), or 2) the measurement rage covered by each region is equal (see Tang and Tracy (1997)). In some cases the context of the problem or the underlying economic interpretation dictates a natural choice for partitions.

Once the symbolic time series is obtained for each element, procedure introduced in Brida and Risso (2007) is applied in order to derive Minimal Spanning Tree and Hierarchical Tree. It means, after symbolization, it is possible to define a simple distance as follows:

$$d_0(s_i, s_j) = \sqrt{\sum_{t=1}^{t=T} (s_{it} - s_{jt})^2} \tag{6.7}$$

Note that $\{s_{it}\}_{t=1}^{t=T}$ and $\{s_{jt}\}_{t=1}^{t=T}$ are two symbolic sequences for companies i and j respectively. Once the distance are computed the distance matrix D is constructed.

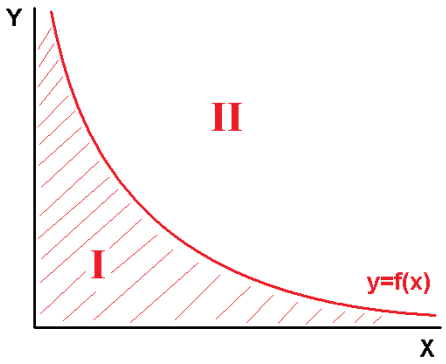


Figure 6.2: Relation defining a partition in a bidimensional space.

It is a symmetric matrix which shows all the distances among the different elements of the set. This matrix permits to construct the Minimal Spanning Tree (MST) connecting the set of time series. The MST is progressively constructed by linking all the time series together in a graph characterized by a minimal distance between time series, starting with the shortest distance, as explained in subsection 6.2.3.

Consider now a bidimensional problem in order to clarify the method, X and Y are relevant in order to explain the structure of a set of n elements. Assume for instance, that it is known a relationship between X and Y given by function $y = f(x)$, this function maybe useful as a proper partition in order to symbolize the bidimensional space. In Figure 6.2 is presented an example of a possible symbolization by using this relation as partition of the space.

It is possible to obtain an unidimensional time series by putting a symbol when (x, y) are below the curve $y = f(x)$ and another symbol when (x, y) are above the function, defining two regimes **I** and **II**. Transformation can be done by using the following expression:

$$s_i = \begin{cases} 0 \quad \text{if} \quad (x_i, y_i) \in \mathbf{I} \\ 1 \quad \text{if} \quad (x_i, y_i) \in \mathbf{II} \end{cases} \tag{6.8}$$

Note that also here different scenarios can be analyzed, for instance in the latter example it is possible to put the partition for a curve that takes less values of X and Y weighting more the small values in the regime **I**.

6.2.4 Concepts characterizing the Trees Some concepts used to characterize the dynamic asset trees are introduced in this subsection:

Total and Normalized tree length One important measure is the Total tree length, which is calculated by summing up the weights on all edges and is a measure of concentration and expansion of the structure. However, in order to compare lengths between different portfolios or economies with non-equal number of stock, it is useful to normalize the quantity by $N-1$, the number of edges. $L(t)$ is the normalized tree length at the moment t.

$$L(t) = \frac{1}{N-1} \sum_{d(i,j) \in T^t} d\,(i,j) \tag{6.9}$$

where $d(i,j)$ is the distance between i and j belonging to the MST (T^t) at the moment t. In the present work this measure will be used in order to study the concentration of the market structure through the time. The more contracted the MST the smaller will be the normalized tree length, in this case, it is likely that all the companies react to the new information in the same way. Note that as the $L(t)$ decreases the stock markets will tend to move together in the same way, i.e.

$L(t) = 0$ indicates that all the elements move in the same direction. Onnela (2002) noted that this measure is related with risk market. Actually, the length of the tree decreases as the stocks become more closely packed together and, consequently, the diversification potential reduces, meaning an increased risk for the minimum risk portfolio. The behavior is reversed when the stock are more spread out on the tree.

Single-step survival ratio The robustness of Dynamic asset trees may be studied by examining the short term persistence or survival of edge connections between two consecutive frames, a concept that is known as the single-step survival ratio. This means taking two consecutive trees T^t and T^{t-1}, physically separated by the step length δT, investigating which connections are found in both trees. In practice, it is calculated as the fraction of connections found in both trees. Mathematically it may be expressed as follows:

$$\sigma(t) = \frac{1}{N-1} \left| E^t \cap E^{t-1} \right| \tag{6.10}$$

Where E_t is the set of edges that make up the graph at time t, the operator $\cap$ gives the intersect of two or more sets and the $|...|$ operator gives the number of elements in the set of edges.

Multi-step survival ratio A natural extension of the latter measure is to investigate the survival ratio of connections over time periods longer than one time step. Note that as more time steps are taken, more births and deaths occur and the graphs should become more dissimilar. The difference with the latter indicator

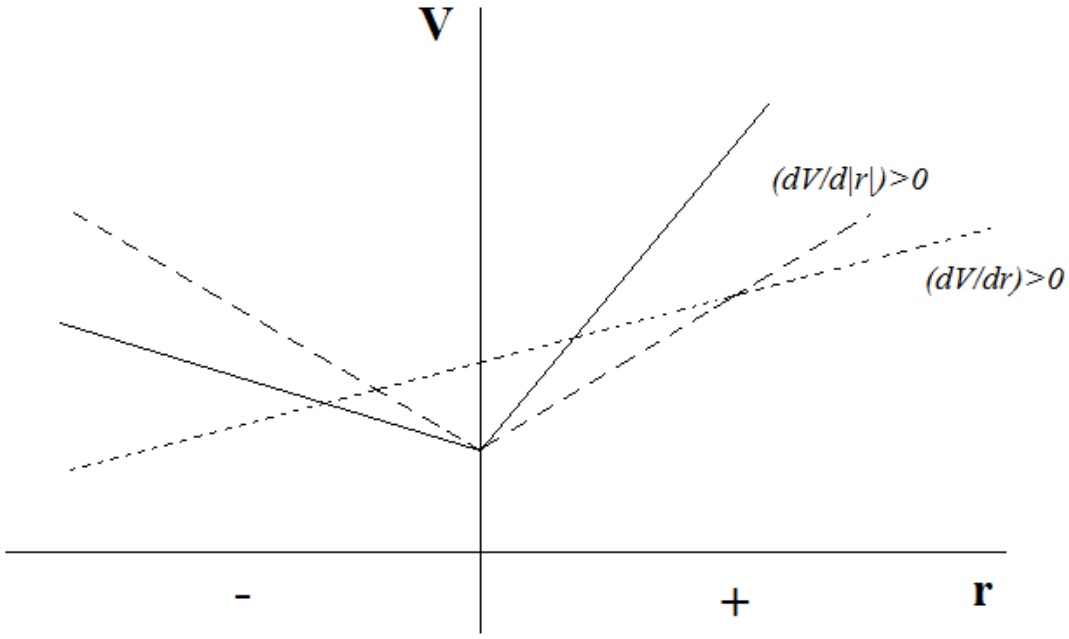
is that the single-step survival ratio can be used to study the robustness of graphs, the multi-step survival ratio describes the long-term behavior of connections, and is used to study graph evolution.

$$\sigma(t,k) = \frac{1}{N-1}\left|E^t \cap E^{t-1} \cap ... \cap E^{t-k}\right| \tag{6.11}$$

6.3 Importance of Volume and Price Change

Karpoff (1987) reviews previous research on the relation between price changes and trading volume in financial markets. In general, the article shows two empirical relations. At first, an old Wall Street adage says "*It takes Volume to make prices move*", Karpoff (1987) asserts that numerous empirical findings support what is called positive volume-absolute price change correlation (see Crouch (1970), Clark (1973), Morgan (1976), Westerfield (1977), Cornell (1981), and Harris (1983), among others). Figure 2 shows the symmetric correlation between the volume (V) and the price change ($|\Delta p|$) in dashed lines. This stylized fact is present in both the future and equity markets. On the other hand, according to Karpoff (1987), despite this positive correlation is almost universally found, some tests indicate that the correlation is weak. Another familiar adage says that "*Volume is relatively heavy in bull markets and light in bear markets*". In this sense Karpoff (1987) remarks that in equity markets, there is evidence of positive relation between volume and price change (see Jain and Joh (1986), Rogalski (1978), Morgan (1976), Harris (1986) among others). One point is that correlation such as the dotted line in Figure 2 has been found only in equity markets. Karpoff concludes that what seems to be

a contradiction maybe explained by an asymmetric volume-price change relation as shown by the solid asymmetric lines in Figure 2, indicating that the relation is fundamentally different for positive and negative price changes. This asymmetric relation explains the two empirical findings reported in the Karpoff (1987) surveys.



Relationship between returns and volume. The dotted line represents positive relation, the dashed line is the symmetric relationship. Finally, the solid line is the asymmetric relationship propossed by Karpoff.

Since there is evidence of a relationship between volume trading and price changes, it seems adequate to consider the two variables in order to construct the stock market topology, studying the informational efficiency accross the market.

6.3.1 Defining a bidimensional partition for the Stock Markets The previous subsection suggested the importance of considering not only price changes but

also trading volume in financial market. This subsection defines a simple partition recovering information carried by volume and price changes.

At first, let us consider a kind of gross return for each company i given by the product between returns and volume trading at moment t: $R_i(t) = r_i(t).V_i(t)$. We construct the empirical distribution $f^*(R_i)$ for the series of size T. According to Molgedey and Ebeling (2000), for statistical reasons one would like to work with small partitions, obtaining a small alphabet. However, taking only two symbols will not consider the fact that dynamics can be different in high negative and positive returns respect to normal return. For this reason, Molgedey and Ebeling (2000) suggest a partition with three pieces. Applying the maximum entropy principle, at first three equally probable regions are defined selecting two partitions where empirical density cumulates 1/3 and 2/3 of the distribution. The latter is done in order to describe the market structure in a normal situation, when analyzing the market in a extreme situation threshold are defined where empirical distribution cumulates 15% and 85% of the distribution weighting more that distribution tails. Once the two thresholds have been obtained we go to the space of returns and volume trading.

Therefore, each pair (returns and volume trading) takes an unique symbol according to the region they are in, as seen in Figure 6.3. Note that, even when we define the global returns only in order to obtain the thresholds, the symbolization is obtained from a bidimensional space where each bidimensional region has 1/3 of the probability.

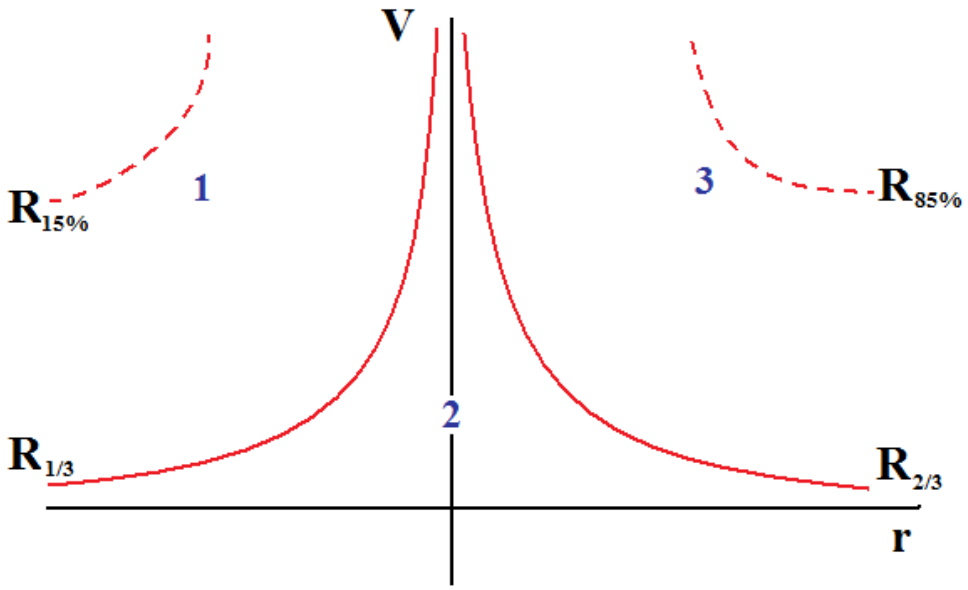


Figure 6.3: Symbolization with 3 symbols in a bidimensional space. Normal situation in full line and critical situation in dashed line.

Once the symbolization is done, we can construct the matrix distance D as explained in subsection 6.2.2.

6.4 Bidimensional Structure for the Main U.S. Companies

The purpose of this section is to study the informational efficiency across the U.S. stock market, since it is considered one of the most efficient in the world. Hence, at first we apply the methodology suggested by Mantegna (1999), then we compute the multidimensional symbolic MSTs and HTs, in a normal and extreme situation. In the whole study a dataset of companies included in Dow Jones Industrial Average[1] is used. The returns are obtained from the stock prices for 30

[1] Data is obtained from database available on-line (http://finance.yahoo.com) and coincides with the daily data from July 10th, 1986 to January 26th, 2007.

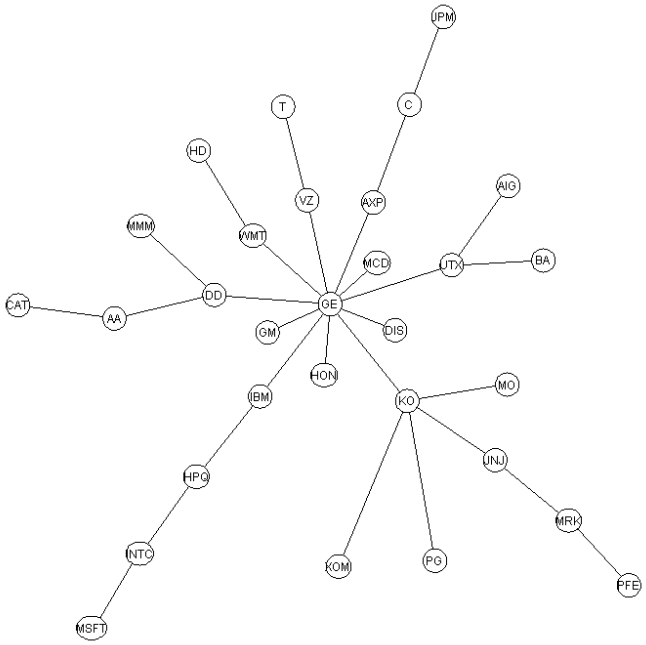


Figure 6.4: The US MST using the distance based on the Pearson correlation coefficient.

companies composing the Dow Jones Industrial.

6.4.1 The US stock market by using de Pearson Correlation Coefficient The methodology suggested by Mantegna (1999) was applied in order to compare the results with those obtained by using the symbolic methodology. Therefore, after computing the Pearson coefficient correlations and the respective distance, company-by-company, MST and HT in Figure 6.4 and 6.5 were obtained.

Analyzing the MST and HT we can check that companies working in the same branch of production tend to clusters. The closest distance is composed by Verizon (VZ) and AT&T (T), two telecommunication companies. Note that there is a clear group of companies working in the informatics sector composed by Hewlett Packard (HPQ), Intel ($INTC$), Microsoft ($MSFT$), and IBM (IBM).

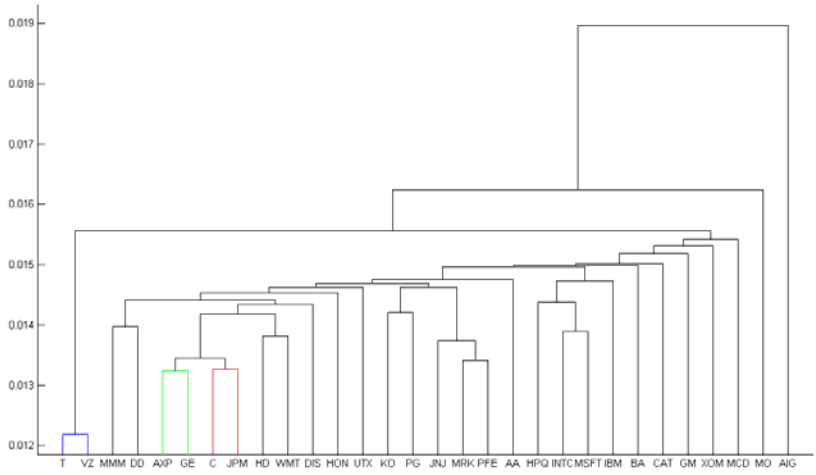


Figure 6.5: The US HT using the distance based on the Pearson correlation coeffcient.

The financial sector also tends to cluster, note that American Express (AXP), Citygroup (C), and J P Morgan (JPM) are in the same group. Another group is composed by pharmaceutical companies such as Pfeizer (PFE), Merck (MRK), and Johnson & Johnson (JNJ). Two retailer companies also clusters, Home Depot (HD) and Wal-Mart (WMT).

Note that MST shows General Electric (GE) as the most linked company, and Coke as the second most linked. On the other hand, note in the HT that AIG is the furthest company in the set of 30 companies. The results suggest that the news affecting a determined sector company are undersood by the companies of the same branch, in the same way. As first approach this is an encouraging result showing that there are some evidence of efficiency across the market.

6.4.2 The US Stock Market in Normal Situation In this subsection we consider the information from returns and volume trading as suggested in the fourth

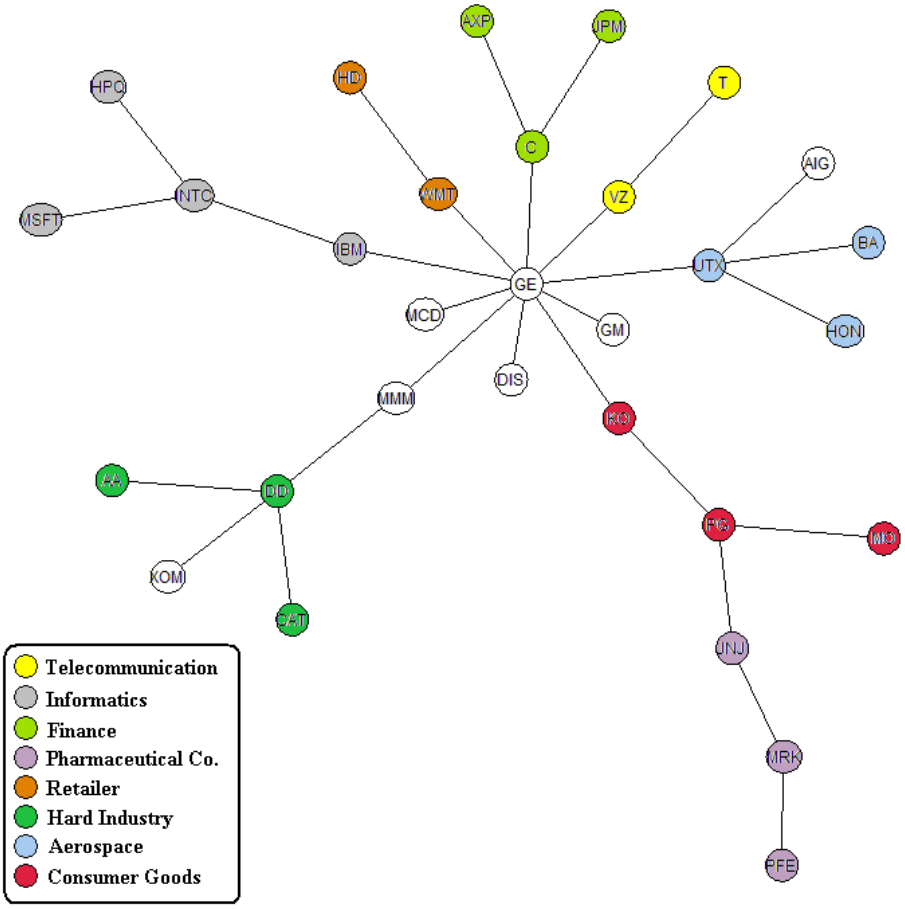


Figure 6.6: MST for the US stock market in a normal situation considering returns and volume trading.

section. Now, we try to analyze if the market is efficient embodying the new information provided by the two variables. Figure 6.6 shows the multidimensional symbolic MST (MSMST) in a normal situation.

Note that it is possible to identify eigth different groups of companies in the MSMST. This tree presents General Electric (GE) as the central node, linking most of the groups. Analysing the groups, at the top left side of the tree, there are companies working in information and computer technology, IBM, Hewell-

Packard (HPQ), Microsoft ($MSFT$) and Intel ($INTC$) belong to the same cluster. There are high complementarity among these industries, and then news make the companies move in the same direction, in an efficienct market. There is a cluster formed by retailers services, Wal Mart (WMT) and Home Depot (HD). Note that financial sector composed by the large financial companies American Express (AXP), Citigroup (C), and J.P. Morgan (JPM), appears connected by GE (as is well known, finance is part of the GE conglomerate). The telecommunication companies (Telephone, Television, Internet) AT&T (T) and Verizon (V) appear forming another cluster. The companies belonging to this group present the least distance among them. On the rigth side there is a cluster formed by companies specialized in aerospace and defense, such as Boeing (BA), United Technology (UTX), and Honeywell (HON). In the south part of the tree, at left we find a group of hard industry companies. This group is composed by Alcoa (AA), Du Pont (DD) and Caterpillar (CAT). The latter could be an example of social embeddedness as suggested by Halinen and Tornroos (1998). Note that DD has a director in common with CAT (John T. Dillon), AA (Alain J.P. Delda). These links are strong considering that James W. Owens is director in AA and CAT. Then, it is possible that they control the same new information, leading to similar dynamics between the two companies. In fact, Ahrne (1994) explains that members of an organization interact with individuals from the same company and from other companies, creating social networks both inside and between organizations. At the south and rigth, we find the consumption branch composed by Coca Cola Co.

(KO) well known beverage industry, Procter and Gamble (PG), the important consumer goods company, and Altria (MO), the cigarettes company. Finally, note that there is a cluster linked to the latter group. This group is composed by pharmaceutical companies such as Johnson and Johnson (JNJ), Merck (MRK) and Pfizer (PFE). As we can appreciate, this topology could be a very useful visual tool in order to identify companies which have presented similar dynamics. In the present case, the cluster suggests the existence of some kind of efficiency inside the US stock market. Note that an inefficient market should present few or no cluster in the tree.

Note that HT (Figure 6.7) spresents similar results. Here we can note that the telecommunication cluster (T, VZ) as the closest group. Of course, they are the best interpreting in the same way the arrival of new information. The pharmaceutical industry (MRK, JNJ, PFE) is the second cluster with the nearest distance among them. In addition, the financial cluster (AXP, C, JPM), the retailers group (HD and WMT) and the cluster composed by the informatics sector ($HPQ, INTC, MSFT$, and IBM) aslo appear.

Note in the HT, that the furthest companies are MMM, AIG, MO, and XOM, while T and VZ are the closets, hence in a portfolio it is well worth to put together companies such as T or VZ with one of the former one, like XOM.

6.4.3 The US Stock Market in an Extreme Situation Defining the thresholds for extreme situations as in the later section, the MSMST is constructed. Note in Figure 6.8 that the structure is similar to the latter, in particular GE still appears

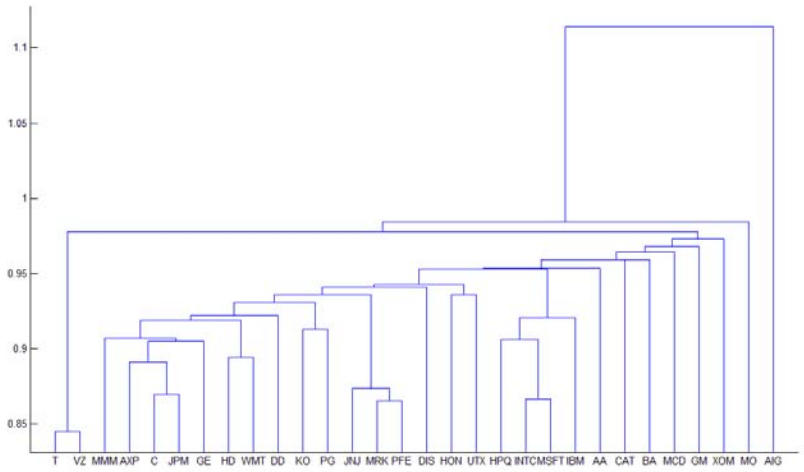


Figure 6.7: The HT in a normal situation for the main US companies.

as the central node and the clusters remain the same. This showns a kind a kind of stability in the structure of the US stock market in a normal situation but also in an extreme scenario.

Note in Figure 6.9 that the HT shows the same clusters as in a normal situation, the same companies are the closest and the same companies are the furthest.

6.4.4 Stability of the MST respect to the partition One question is if the symbolic MST is sensible to the partition. As we now, when we select a equally probably partition (1/3 of probability in each bidimensional region) with threshold at 33.33%-66.66%, and a partition giving 15% to the extremes and 70% in the middle of the space (thresholds at 15%-85%), the main results do not change. It means, we obtain the same company as the central node in the market (GE) and the same eight clusters. We realized a sensibility analysis by changing the partitions and constructing the MST each time. We defined 9 different partitions presenting

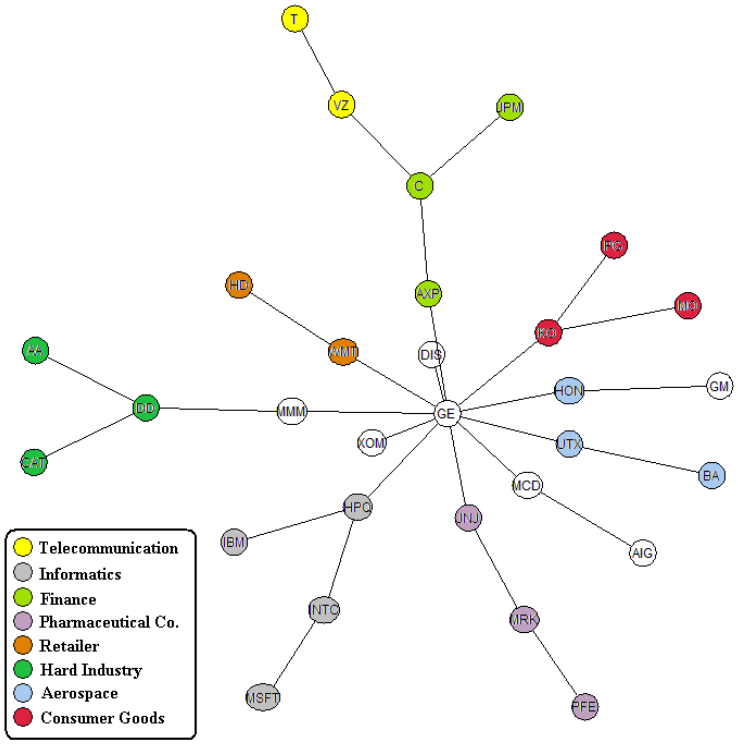


Figure 6.8: MST for the US stock market in a extreme situation considering returns and volume trading.

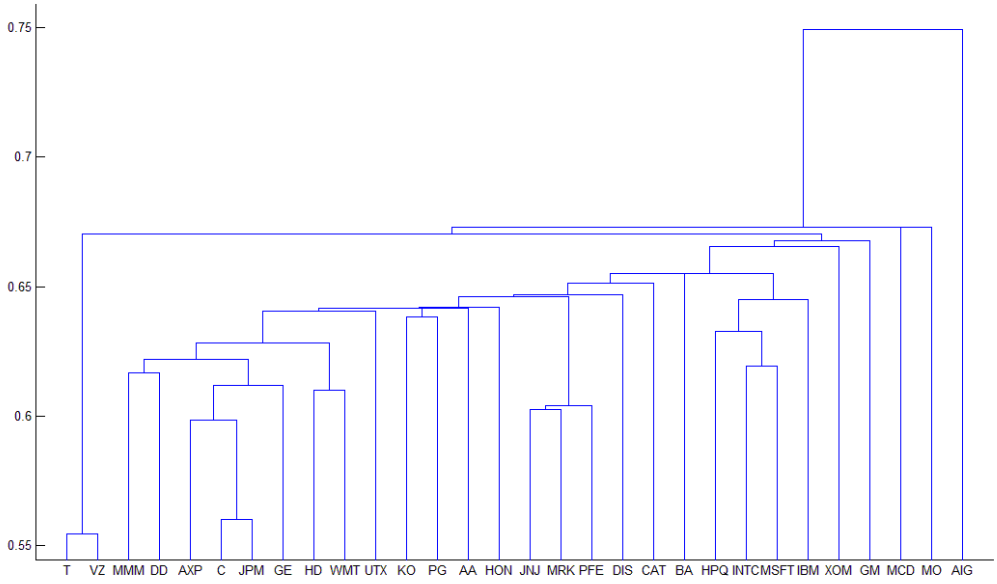


Figure 6.9: The HT for the main US Companies in an extreme situation.

thresholds at: 5%-95%, 10%-90%, 15%-85%, 20%-80%, 25%-75%, 30%-70%, 35%-65%, 40%-60%, and 45%-55%. For all these cases the fundamental structure does not change, GE remains as the central node and the eight clusters are present in the tree.

A second question is whether our measure is significative or the links in the MMST are random. To study this we realized 1,000 Monte Carlo simulations for 30 random companies for a period T=5,184 days (which is the size of our dataset). Then, we compute 1,000 random MST, obtaining the simulated distribution of the distances (or links) belonging to the MST. We define the confidence intervals at 5% and 95% where a random link should enter in, for a sample sized T=5,184 with 30 companies (29 links). The results are shown in Table 1, note that if we have the smallest distance between 80.76 and 81.57 we cannot reject the hypothesis that

this link is random, note also that for the link 29, the interval is 81.98 and 82.32.

Table 1: Confidence Intervals at 5% and 95% for random links

Links	$(5\% - 95\%)$	Links	$(5\% - 95\%)$	Links	$(5\% - 95\%)$
link 1	$(80.76 - 81.57)$	link 11	$(81.67 - 82.05)$	link 21	$(81.96 - 82.29)$
link 2	$(81.08 - 81.68)$	link 12	$(81.70 - 82.08)$	link 22	$(81.98 - 82.32)$
link 3	$(81.25 - 81.76)$	link 13	$(81.73 - 82.11)$	link 23	$(82.01 - 82.34)$
link 4	$(81.35 - 81.82)$	link 14	$(81.76 - 82.13)$	link 24	$(82.05 - 82.38)$
link 5	$(81.41 - 81.86)$	link 15	$(81.79 - 82.14)$	link 25	$(82.08 - 82.43)$
link 6	$(81.47 - 81.90)$	link 16	$(81.82 - 82.17)$	link 26	$(82.12 - 82.43)$
link 7	$(81.52 - 81.95)$	link 17	$(81.85 - 82.19)$	link 27	$(82.16 - 82.56)$
link 8	$(81.56 - 81.97)$	link 18	$(81.87 - 82.22)$	link 28	$(82.21 - 82.66)$
link 9	$(81.60 - 82.01)$	link 19	$(81.90 - 82.24)$	link 29	$(82.31 - 82.85)$
link 10	$(81.64 - 82.03)$	link 20	$(81.93 - 82.26)$		

Based on the 1,000 random simulations of 30 random companies for 5,184 days

However, our tree is far from the random intervals showing the high significance of the links (see Table 2). Note, that the smallest distance is $d(VZ, T)$=60.83 and the link 29 is the distance $d(AIG, UTX)$=80.22.

Table 2: Links of the Main US companies in the MST

Link	Firms	Dist.	Link	Firms	Dist.	Links	Firms	Dist.
1	VZ-T	60.83	11	PG-KO	65.71	21	AA-DD	68.67
2	MRK-PFE	62.32	12	WMT-GE	66.16	22	CAT-DD	69.06
3	INTC-MSFT	62.38	13	INTC-IBM	66.30	23	BA-UTX	69.06
4	C-JPM	62.61	14	MMM-DD	66.41	24	GE-MCD	69.43
5	MRK-JNJ	62.89	15	GE-KO	67.01	25	GM-GE	69.43
6	C-AXP	64.16	16	PG-JNJ	67.38	26	XOM-DD	70.06
7	HD-WMT	64.40	17	HON-UTX	67.41	27	VZ-GE	70.40
8	C-GE	65.16	18	DIS-GE	67.75	28	PG-MO	70.89
9	INTC-HPQ	65.25	19	UTX-GE	67.87	29	AIG-UTX	80.22
10	MMM-GE	65.33	20	GE-IBM	68.60			

Based on the obtained results for a partition at 1/3 and 2/3

6.4.5 Total Tree length and Survival Ratio We mention in subsection 6.2.4 that two important measures are the total tree length and the survival ratio. The former is calculated by summing up the weights on all edges and is a measure of concentration and expansion of the structure. The latter measures the robustness of dynamic asset trees, examining the short term persistence or survival of edge connections between two consecutive frames. To compute the Survival Ratio the total period is divided in two subperiods of 2592 days. Computing the respective MSTs and HTs, it is observed that the total tree length goes from 27.5886 to 39.5364 in the second period. It means that the US stock market expanded 43.31%

between the two periods, according to Onnela (2002) this is a sign that the market decreased its risk.

According to the single-step survival ratio 41.38% of the connections survived between the periods in the MST. The HT shows that clusters such as Telecommunication (T and VZ), Pharmaceutical (PFE, JNJ and MRK), Retailers (HD and WMT), Informatics (HPQ, $INTC$, $MSFT$, and IBM), and Financial sector (JPM, C, AXP) survive between the two periods. Even more, the closest companies are VZ and T. On the other hand, the furthest firms are MMM, XOM, AIG and MO, confirming that a portfolio could be composed by one of the former companies with the one of the latter.

6.4.6 Further analysis: Cointegration and Granger Causality between GE and AIG Another question which rises from the results is what the reason for GE to take a central position in the tree is? One hypothesis could be that the largest companies leads the movements of the other firms. Therefore, we collected three proxies for the size of our companies (number of employees, market capitalization and total revenues). Note in Table 2 that GE ranks the 2nd according the Market Capitalization, 3rd according the number of employees and 4th if we consider the total revenues. However, note that Exxon is the largest company in the group but it does not take the central position.

We also applied econometric methods in order to study the relation among the companies. In particular, we apply a Vector Autorregressive Model (VAR) for the 30 variables and applied the Granger causality test in order to study the prece-

Table 2: Ranking of the largest companies in the Dow Jones Industrial Average Index

Company	Rank	Market Cap [1]	Rank	Total Revenue [2]	Rank	Full time employees
Exxon	1	459.46	1	377,635,000		N/A
General Electric	**2**	**345.35**	**4**	**163,391,000**	**3**	**300,000.00**
Microsoft	3	264.50	16	51,122,000	17	79,000.00
AT & T	4	228.95	13	63,055,000	2	309,050.00
Procter and Gambler	5	204.04	12	76,476,000	10	138,000.00
Wal Mart	6	198.00	2	348,650,000	1	1,900,000.00
Johnson & Johnson	7	181.73	15	53,324,000	12	119,000.00
Altria Group	8	152.60	7	101,407,000		N/A
Pfeizer	9	151.26	17	48,371,000		N/A
IBM	10	147.56	9	91,424,000		N/A
JP Morgan	11	145.64	8	99,845,000	8	180,667.00
Coca Cola	12	135.79	26	24,088,000	19	71,000.00
Citigroup	13	127.26	5	146,558,000		N/A
Intel	14	117.46	21	35,382,000	16	86,300.00
American International Group	15	116.95	6	113,194,000	14	106,000.00
Verizon	16	108.84	11	88,144,000	6	234,971.00
Merck	17	103.51	28	22,636,000		N/A
United Tecnologies	18	70.21	18	47,829,000	7	225,600.00
MacDonald	19	66.18	29	21,586,400		N/A
Boeing	20	65.42	14	61,530,000	9	159,300.00
Walt Disney	21	61.18	20	35,510,000	11	137,000.00
3M	22	56.70	27	22,923,000	18	75,000.00
American Express	23	52.24	25	27,136,000	20	64,800.00
Home Depot	24	46.44	10	90,837,000	5	247,520.00
Caterpillar	25	43.65	19	41,517,000	15	101,333.00
Honeywell	26	41.86	22	31,367,000		N/A
DuPont	27	40.93	24	28,982,000		N/A
Alcoa	28	29.55	23	30,379,000	13	107,000.00
General Motors	29	14.79	3	207,349,000	4	267,000.00
Hewlett Packard	30	4.35	30	1,629,658	21	6,444.00

(1) In billions of dollars

(2) year 2006

dence of one variable respect to the other. In order to select the best lag length we apply the minimum AIC (Akaike Information Criterio), one length was selected. Some empirical results were found, when we applied the Granger causality test, GE caused 12 variables ($AIG, AXP, BA, C, CAT, DIS, GM, HD, HON, MCD, MMM, UTX$), there was a bidirectional causality between PFE and GE and four variables caused GE (WMT, JPM, MO, MRK). Note again, the importance of GE preceding the movement of 40% of the largest companies. On the other hand, I discovered that the movement of AIG, one of the furthest companies in the trees was caused by all the companies. This result suggests that we could forecast its movement by knowing the movement of the other companies the day before. To study this idea I conduct an econometric analysis between GE and AIG using the cointegration technique propossed by Johansen (1995). We use the log of the respective prices from 7 september 1984 to 31 December 2007.

We estimate a VAR with 5 lags (according to the minimum AIC) and test the cointegration relationship. Table 3 indicates that 1 co-integrating relationship is obtained.

Table 3: Johansen Cointegration Test

Hypothesis	**Trace Statistic**	**C.V. at 0.05**	**p-value**
None*	*23.70*	*15.49*	*0.002*
At most 1	*2.32*	*3.84*	*0.127*
Hypothesis	**Max-Eigen Statistic**	**C.V. at 0.05**	**p-value**
None*	*21.38*	*14.26*	*0.003*
At most 1	*2.32*	*3.84*	*0.127*

Source: Own calculations. * Indicates rejection of the null hypothesis at 5%

To do inference we need to conduct the weakly exogeneity on AIG. The Chi^2 statistic is 1.503 producing a p-value of 0.22; therefore we cannot reject the hypothesis that the price of AIG is exogenous to the model. The following equation shows the long-run estimated relationship.

$$AIG(t) = 0.884 + 0.94GE(t)$$

$$.....[-41.78]$$

Cointegration by itself does not indicate the direction of the causal relationship. Granger (1988) proposed a test to study causality. However, this is not causality in a philosophical sense. It should be understood as a kind of predetermination among variables.

The dynamic Granger causality can be captured from the VAR model. However, since the variables are integrated, application of the standard Granger causality test is invalid. Toda and Yamamoto (1995) suggest an alternative procedure. When the variables are integrated, they propose to estimate a VAR model with (k+$dmax$) lags, where k is the standard optimal number of lags and $dmax$ is the

maximal order of integration that we suspect might occur in the process. Once the VAR is estimated, we test Granger causality only using the first k lags. For instance, if we consider the following equation from a VAR model:

$$AIG(t) = \gamma_0 + \gamma_1 GE(t-1) + ... + \gamma_6 GE(t-6) + \gamma_7 AIG(t-1) + ... + \gamma_{12} AIG(t-6) + \epsilon(t)$$

where k = 5 was selected according the minimum AIC and dmax = 1, the null hypothesis of non-causality from GE to AIG should be:

$$H0 : \gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = \gamma_5 = 0$$

It means,

$$H0 : GE \text{ does not Granger-cause } AIG$$

The hypothesis is tested using the Wald test. However, Toda and Yamamoto (1995) assert that Wald and LR tests are asymptotically equivalent in the present situation. Table 4 shows the results for all the variables.

Table 4: Granger Causality Test (by Toda & Yamamoto)

Null Hypothesis	*Wald-statistic*	*p-value*
GE does not cause AIG	14.95	0.0106*
AIG does not cause GE	6.93	0.2262

We used a VAR with k+dmax = 5+1. p-values correspond to the Chi-square distribution with 1 degree of freedom.

* indicates rejection of the hypothesis at 5%

Notice that we reject the first hypothesis but not the second. Therefore, GE causes the movement of AIG.

6.5 Bidimensional Structure for the Main Italian Companies

The Milan Stock Exchange (MSE) concentrates more than 90 percent of the transaction volume of the Italian stock market. It was founded in 1808, privatized in 1997 and acquired by the London Stock Exchange Group in 2007. The most important index is the S&P/Mib, embodying the highest capitalized companies (more than 1000 million euros). An important characteristic is that the 30% of these companies work in the financial sector (insurance and bank firms) representing the 48% of the market capitalization.

Few papers have studied the Italian Stock market. We can refer to Barone (1990) analyzing the efficiency and the anomalies in this market, Michaely and Murgia (1995), studying the effects of tax heterogeneity on price and volume in the MSE and Brida and Risso (2007) studying the structure of the market considering the asset returns.

It is well known that Italian market is basically characterized by the prevalence of small and medium-sized companies with a small number of large companies due to increased concentration. This section is an extension of the unidimensional symbolic approach introduced in Brida and Risso (2007). The data from the S&P/Mib is used, collecting data of trading volume and asset returns for 32 companies[2].

[2] Daily data (from December 7th, 2001 to September 12th, 2007) was obtained from database available on-line (http://finance.yahoo.com).

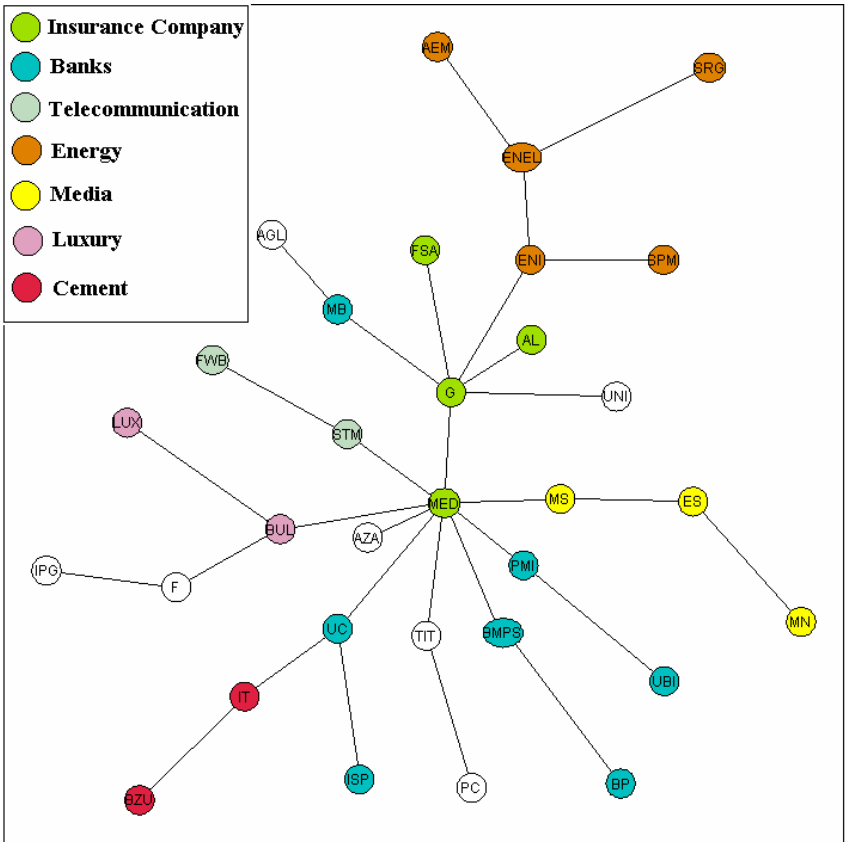


Figure 6.10: MST in a normal situation for the Italian market considering trading volume and asset returns.

6.5.1 The Italian Stock Market in a Normal Situation From the matrix of distances D a ranking of distances is considered from the closest to the furthest. Therefore the MST is constructed by connecting the most relevant distances. The MST obtained for the Italian case is shown in Figure 6.10.

Note that companies working in the same branch tend to form groups of clusters. Therefore, we obtain some evidence of efficiency across the market. In the north section of the tree it is possible to identify a cluster of companies working in the sector of energy, AEM working in liquid gas, $SRG, ENEL, ENI$ and SPM

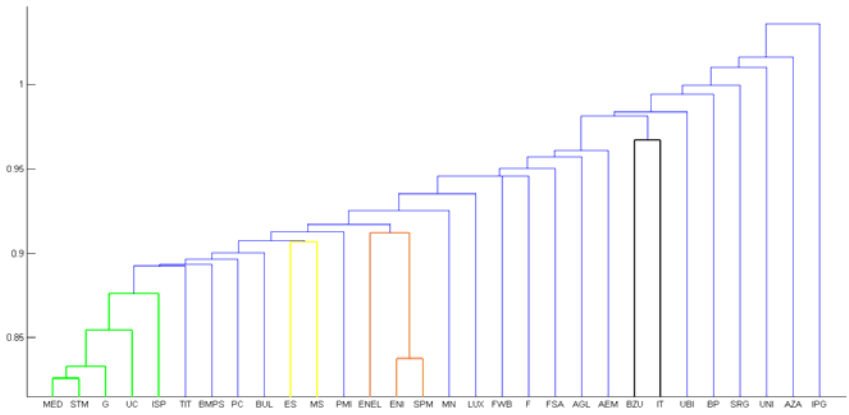


Figure 6.11: HT for the Italian Stock Market in a normal situation

working in Gas and Electricity. The center of the tree is composed by companies working in Banks and Insurance, actually Mediolanum (MED) and Generali (G) have a central position with more connections than the other companies. At the left we have a cluster composed by luxury firms (Luxottica and Bulgari) is found and a group of Telecommunication formed by Fastweb (FWB) and STMicroelectronics (STM) also appears. At the we have the media sector is identified composed by Mediaset (MS), L'Espresso (ES) and Mondadori (MN). Finally, at the south three groups appear, one formed by two companies working in cement, Buzzi Unicem (BZU) and Italcementi (IT). Note that Unicredit (UC) and Intesa-San Paolo (ISP) have a close distance, they are the main banks in Italy. A third group composed by Telecom (TIT) and Pirelli (PC) can also be identified, as it is well known, Pirelli is the major shareholder of the Olimpia group, and this group is the principal shareholder of Telecom.

Figure 6.11 shows the existence of three clusters: 1) composed by Generali (G)

and Mediolanum (MED) two insurance companies having the central position in the tree, but also Unicredit (UC) and Intesa San Paolo (ISP), the most important banks. Note that also STMicroelectronic (STM) is included. This is the group at left with the closet distances; 2) a second group is formed by L'Espresso (ES) and Mediaset (MS) both working in the media sector; 3) $ENEL$, ENI and Saipem (SPM) form a group working in energy sector, note the close distance between ENI and SPM. Actually, they have a strong relationship, SPM (plant design and installation) was part of ENI (Petroil subsector) until 1969, government has the 42.9% of the shares of the first and 20,32% of the second; 4) The group formed by cement companies includes Buzzi Unicem (BZU) and Italcemento (IT).

6.5.2 The Italian Stock Market in an Extreme Situation In figure 6.12 we have constructed the MST for the extreme situation. Note that the fundamental structure of the tree remains and this can be interpreted as some kind of stability of the tree from the normal to the extreme situation. Basic groups remain, financial sector (Banks and insurance companies) are the center of the tree with especial focus on Generali (G) and Mediolanum (MED).

Figure 6.13 shows the HT in the extreme situation. Note that a cluster composed by Pirelli (PC) and (TIT) Telecom is highlighted and this could reflect the problem about the decision of selling part of the shares of Telecom which happened in the analyzed period.

6.5.3 Total Tree length and Survival Ratio As in the US market, we proceed to compute the total tree length and the survival ratio. At first, the period T is

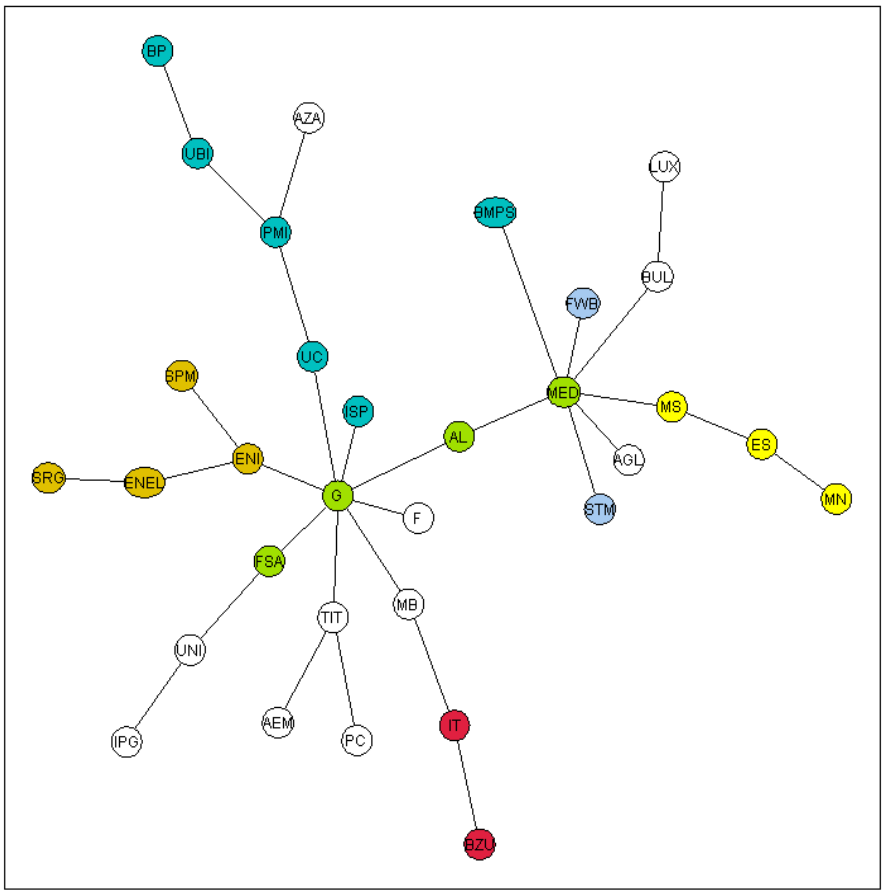


Figure 6.12: MST in an extreme situation for the Italian Stock Market

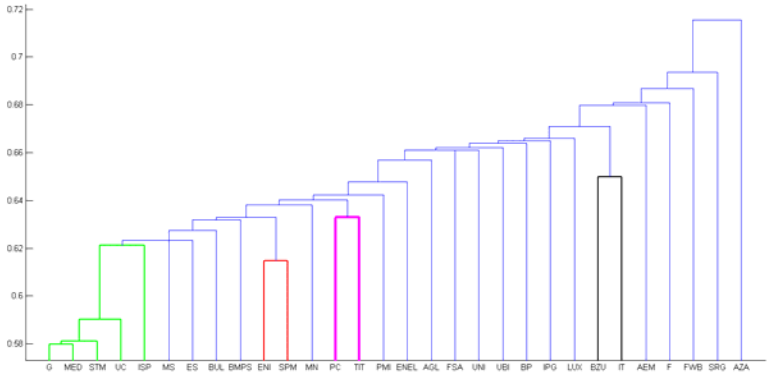


Figure 6.13: HT in extreme situation for the Italian Market

divided in 2 sub-periods of 746 days. The respective MSTs and HTs are computed and the differences are studied. It is observed that the total tree length goes from 27.3589 to 28.5724 in the second period, it means that the Italian market expanded 4.44% between the two periods, according to Onnela (2002) this is a sign that the market decreased its risk.

According to the single-step survival ratio only 26% of the single connections survived between the periods. However, the clusters seem to be more stable than the particular single connections. The HT shows that Generali and Alleanza are in the group of the 4 closest companies between the two periods, whereas Alitalia is among the 4 furthest companies between the two periods.

Figure 6.14 shows the total tree length for time-windows of 120, 240, and 480 days. Note that the Market shows a larger expansion through the time.

6.6 Conclusions

This Chapter aimed to analyze the informational efficiency across a determined market. In order to study this kind of efficiency a methodology was developed. In fact, minimal spanning tree and hierarchical tree introduced by Mantegna (1999) was modified. We introduced a multidimensional symbolic method which gives more flexibility, on the one hand, it permits to analyze more than one variable, on the other hand, it permits to study different scenarios. The method basically detect the formation of clusters in the market which have similar behavior. We interpret the formation of clusters as evidence of efficiency inside a market. In fact, a cluster represents a group of companies reacting at the same time, in the same

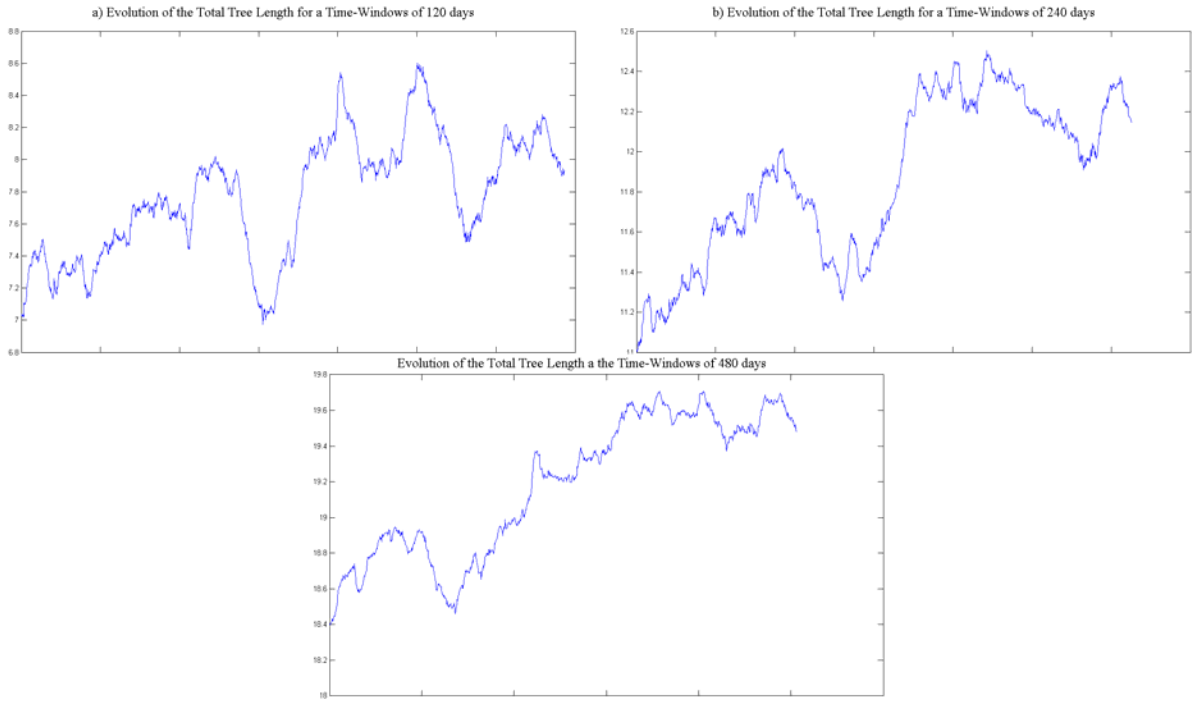


Figure 6.14: Evolution of the Total Tree Length for different time-windows. (a) 120 days, (b) 240 days, (c) 480 days.

way to the new information arriving in the market. An inefficient market should present no clusters, since all the companies would react in different manners to the news.

We applied a bidimensional methodology (considering not only asset returns but also trading volume, since they provide important information as suggested by Karpoff (1987)) to the US market and the Italian market, in order to see if they can be considered as efficient markets. In both cases we detected the formation of clusters which make sense from an economic point of view, companies working in the same branch tend to cluster.

The US companies present a structure which makes sense from an economic point of view, showing eight different clusters of firms working in the same branch. Interesting thing is that both in normal and extreme situation the US structure does not change. We always find eight clusters where GE takes a central position in the tree. The structure is also stable, in particular it does not change when we define different partitions. When we divide the period in two and computing the total tree length, it goes from 27.5886 to 39.5364 in the second period, while the survival ratio shows that the 43,41% of the connections survive. However, the clusters are stables, note that T and VZ are the closest companies whereas MMM, XOM, AIG and MO are the furthest, this is important for the construction of a portfolio where maybe is logical to put companies such as T and MO together. We studied if the central position of GE was due to the fact that is a Hugh company, however, the largest company is Exxon-Mobile taking a not im-

portant position. In addition, we discover that the prices of the furthest company (AIG) seem to be caused by the other companies, in particular a long run relationship is obtained between GE and AIG. The results show that knowing the returns of GE the day before helps predicting the price of AIG the next day.

The Italian stock market also presents a structure with economic meaning. However, the structure does not significantly change from a normal situation to the extreme situation, it means that in a critical situation the links among the companies remain the same. Financial companies take a central position in the structure, where Generali and Mediolanum are the most connected companies. When dividing the total period in two parts the single connections seem to change but the group are stable, companies such as Generali and Alleanza are the closest whereas Alitalia is one of the furthest companies. When considering two variables, the evolution of the total tree length suggests that the market has evolved to a more expansive position, as in the US case.

The results suggest that there is some evidence of informational efficiency across the US and Italian markets. It also suggests that the links among the companies are not so strong. However, the clusters seem to be stable. Further results show that both market have tended to a more expansive situation and then less risky, according to Onnela (2002).

6.7 References

-Ahrne, G., (1994), *Social Organizations. Interaction inside, Outside and Between Organizations.* Sage Publications, London.

-Barone, E. (1990), "The Italian Stock Market: Efficiency and Calendar Anomalies", *Journal of Banking and Finance*, Vol. 14, pp. 483-510.

-**Bonanno, G., Lillo, F., Mantegna, R., (2001),** "Level of complexity in Financial markets", *Physica A*, Vol. 299, pp. 16-27.

-Bonanno, G., Calderelli, G., Lillo, F., Micciché, S., Vandewalle, N., Mantegna, R., (2004), "Networks of equities in financial markets", *The European Physical Journal B*, Vol. 38, pp. 363-371.

-Brida, J.G., and Risso, W. A. (2007): "Dynamic and Structure of the Main Italian Companies", *International Journal of Modern Physics C*, Vol. 18 (11), pp. 1-11.

-**Clark, P., (1973)**, "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices", *Econometrica*, Vol. 41, pp. 135-155.

-**Cornell, B., (1981)**, "The Relationship between Volume and Price Variability in Future Markets", *The Journal of Future Markets*, Vol. 1, pp. 303-316.

-**Crouch, R., (1970)**, "The Volume of Transactions and Price Changes on the New York Stock Exchange", *Financial Analysis Journal*, Vol. 26, pp. 104-109.

-Daw C., Finney, C., Tracy, E., (2003), "A review of symbolic analysis of experimental data", *Review of Scientific Instruments,* Vol. 74, pp. 916-930.

-Gower, J., (1966), "Some distance properties of latent root and vector methods used in multivariate analysis", *Biometrika,* Vol. 53, no. 3-4, pp. 325-338.

-Granger, C., (1988), "Some recent developments in a concept of causality", *Journal of Econometrics*, vol 39, pp. 199-211.

-**Halinen, A., Tornroos, J., (1998)**, "The Role of Embeddedness in the Evolution of Business Networks", *Scandinavian Journal of Management*, Vol. 14, No. 3, 187-205.

-**Harris, L., (1983)**, "The Joint Distribution of Speculative Prices and of Daily Trading Volume", *Working Paper*, Univ. of Southern CA.

-**Harris, L., (1986)**, "Cross-Security Tets of the Mixture of Distribution Hypothesis", *Journal of Financial and Quantitative Analysis*, Vol. 21, pp. 39-46.

-Jain, P. , Joh, G., (1986), "The Dependence between Hourly Prices and Triding Volume", *Working Paper*, The Wharton School, Univ. of PA.

-**Johansen, S. (1988),** "Statistical Analysis of cointegration vectors", *Journal of Economic Dynamics and Control*, vol. 12, pp. 231-254.

-Kaski, K., Onnela, J., Chakraborti, A., (2003), "Dynamics of Market Correlations: Taxonomy and Portfolio Analysis", *Physical Review E,* Vol. 68, 056110.

-**Karpoff, J., (1987),** "The Relation Between Price Change and Trading Volume: A Survey", *The Journal of Financial and Quantitative Analysis,* Vol. 22, no. 1, pp. 109-126

-Mantegna, R., (1999), "Hierarchical Structure in Financial Markets", *The European Physical Journal B,* Vol. 11, pp. 193-197.

-Mantegna, R., Stanley, H., (2000), *An introduction to Econophysics: Correlations and Complexity in Finance*, Cambridge University Press, UK.

-Michaely, R., Murgia, M., (1995), "The Effect of tax heterogeneity on

prices and volume around the Ex-dividend day: Evidence from the Milan Stock Exchange", *The Review of Financial Studies*, Vol. 8 (2), pp. 369-399.

-Mizuno, T., Takayasu, H., Takayasu, M., (2006). "Correlation networks among currencies", Physica A, Vol. 364, pp. 336-342.

-**Molgedey, L., Ebeling, W., (2000)**, "Local Order, Entropy and Predictability of Financial Time Series", *The European Physical Journal B*, Vol. 15, pp. 733-737.

-Morgan, I., (1976), "Stock Prices and Heteroskedasticity", *Journal of Business*, Vol. 49, pp. 496-508.

-Onnela,J., (2002), *Taxonomy of Financial Assets*, Thesis for the degree of Master of Science in Engineering, Dep. of Electrical and Communications Engineering, Helsinki University of Technology.

-Ortega, G., Matesanz, D., (2005). "Cross-country Hierarchical Structure and Currency Crises", *International Journal of Modern Physics C,* Vol. 17, Issue 03, 333-341.

-**Ramal, R., Toulouse, G., Virasoro, M., (1986)**, "Ultrametricity for Physicists", *Review of Modern Physics*, Vol. 58, no. 3, pp. 765-788.

-Rogalski, R., (1978), "The Dependence of Price and Volume", *The Review of Economics and Statistics*, Vol. 36, pp. 268-274.

-Tang, X., Tracy, E., (1997), "Data Compression and Information Retrieval via Symbolization" *Chaos*, Vol. 8, no. 3, pp. 688-696.

-**Toda, H., and Yamamoto, T., (1995),** "Statistical inference in vector

autorregressions with possibly integrated processes", *Journal of Econometrics*, vol. 66, pp. 225-250.

-Westerfield, R., (1977), "The Distribution of Common Stock Price Changes: An Application of Transactions Time and Subordinated Stochastic Models", *Journal of Financial and Quantitative Analysis*, Vol. 12, pp. 743-765.