

# Non-random behavior in financial markets

Andrea Fenu

University of Cagliari

Department of Economics and Business Administration

Advisor: Prof. Pier Paolo Carrus

## Abstract

What is the nature of the price formation process? This thesis' project uses an ontological approach to analyze the time series of historical prices in financial markets, at different frequencies, and across multiple asset classes, in order to understand (i) the existence of any non-random pattern, (ii) the eventual information that they may contain, and (iii) what may be their implication at microstructural level. In the first (i) part we developed a pattern recognition algorithm to extract consecutive trend lines from prices, obtaining a robust statistical evidence of a significant and systematic memory in historical prices. In the second (ii) part we implemented an asset allocation model to test the value of the information contained in the memory of the process. The investment strategy significantly overperformed the market for every security and frequency analyzed. In the third (iii) part we studied the effect of the deterministic patterns at microstructural level demonstrating how their presence biases the well-documented relationship between trading frequency and volatility. Our results, in fact, are partially in contrast with the main literature, highlighting a non-stable relationship through the trading day.

**Keywords:** Prices, Returns, Randomness, Market Efficiency, Behavioral Finance, Market Behavior, Autocorrelation of Returns, Technical Analysis, Pattern Recognition, Asset Allocation, Trading, Fractals, Market Microstructure, Deterministic Patterns.

**Acknowledgements:** H.E. Stanley and all the Physics Department of the Boston University. Michael Salinger and all the Questrom School of Business of the Boston University. Ryan McKeon, Marko Svetina and all The University of San Diego School of Business. Shlomo Havlin. Didier Sornette at ETH and Dario Villani at Tudor Investments for their precious epistolary contribution. Long Feng Zhao and Tomas Gubiec for their quantitative support and invaluable exchange of ideas.

## General Introduction

What is the first thing that we do if we want to find out if prices in financial markets can be predicted? Probably we first select an information provider, we then download one or more time series of historical prices, and we try to see if we can extract any forecastable pattern out of them. Let us assume that we know nothing about financial markets, we are absolute profanes. After a certain research effort, we would obviously expect either to find one of those patterns or to find nothing. In the first case we would be sure about their presence, in the second we would not since the output would also depend on our skills. Let us focus for one moment on the first case, the one where we have been able to find a predictable pattern. We know that our time series belongs to a certain asset class and that its observations are recorded with a certain criteria and at a certain frequency. We also know that each observation represents a price level at which some transactions between willing counterparts occurred. We do not need to be financial experts to draw this conclusion. With this basic information set, if we are curious enough, we would ask ourselves where an eventual predictability may come from. At first, we may consider that an eventual property can be observable due to the type of the asset we analyze and its characteristics, as different assets may correspond to different market characteristics, and/or to different types of buyers and sellers. In this case we would obviously expect to find isolated patterns, specific to that precise asset, and so not observable by looking at other types of assets or markets. The same point can be made with respect to the frequency of the data contained in our time series. If we think that the properties and the predictable patterns that we can extract from the data are specific to the frequency, or the criteria with which the data has been collected, we would expect to find such properties to be isolated at that frequency and not observable at different ones. Of course, we may even expect multiple combinations of these two cases. If we look at the matrix in figure 1, we notice 4 possible combinations: a pattern can be found for a

		Frequency	
		Single	Multiple
Asset class	Multiple	1	2
	Single	3	4

Figure 1

specific frequency and spanning across multiple asset classes (1); or the same pattern can be

observed across multiple frequencies and multiple asset classes as well (2); on the other side, if we observe a pattern only on a specific asset class, the property can exist only at a specific frequency (3); or it can be valid across multiple frequencies (4).

The quadrants 1, 3 and 4 in figure 1 represent, to some extent, the possibilities of observing idiosyncratic patterns since they belong to either a specific asset or to a specific frequency. If we end up in one of these 3 possibilities, we can conclude that we discovered a certain degree of predictability and that this predictability can be linked to a specific characteristic of a determined market, a determined group of buyers and sellers, or to a certain trading frequency. What happens if we end up in quadrant 2? If we recall our basic information set outlined above, we cannot address the pattern neither versus a specific asset nor versus a specific frequency. We may feel to be in front of some sort of a universal property, which is valid in every circumstance. With our knowledge, if we try to explain our result, we may need to go back to the roots of a generic price for a generic frequency. What do all the recorded prices in every time series have in common? The answer may seem quite straightforward but it is: a transaction. Yes, every traded price in human history is, by definition, traded. In this context, the quadrant number 2 imposes an analysis on the very nature of a transaction, that is, to a larger extent, an analysis of the human nature and human interactions. Quadrant number 2 is the focus of this thesis. This work wants to explore the very nature of the price formation process without any prior argument or preconceptions due to existing research. Our extent is to investigate the object: price, what this object contains, and if what is contained is purely random or not, i.e. if prices contain any structural and/or predictable patterns. We use the term structural since our goal is not to investigate idiosyncratic characteristics of a certain asset or a certain frequency, but we intend to address the eventual overall nature of prices and their formation process which is, in our opinion, a quantitative expression of human behavior. We addressed the topic with two opposite and complementary approaches: bottom up and top down. In the former we started from the time series of prices, at any frequency, and we developed a new methodology to extract patterns, while in the latter we analyzed how well-known patterns may influence what we (supposedly) know about the price formation process at microstructural level.

In the main part of the Thesis, which covers the bottom up approach, we took various time series of historical prices (dividend adjusted), for multiple asset classes and for multiple frequencies (from tick-by-tick to monthly data), to find out what is inside of these prices in terms of structural and/or predictable patterns. Our aim is to try to understand the nature of the price formation process and its link to human transactions ranging from the microstructural level to the long-term cycles. Our aim is to find structural non-random behavior in financial markets. Now, let us try to contextualize our research question in the actual financial playground, specifying how we intend to address the topic, and what the main contribution that we intend to provide is.

In finance, if we say that we are able to extract non-random or predictable patterns from a time series of prices, we are basically implying that we are able, to some extent, to extract non-exploited information from historical prices, and so, that we are able to gain ‘extra returns’. The concept of extra returns is most likely the responsible for one of the most intense debate in finance. The research on this topic has always been split between the two main opposite factions: the academic and the financial professionals. While, according to the former, stable extra returns can be hardly (if not impossibly) achieved, whereas the latter ground the very nature of their existence on the opposite statement. This work wants to approach these two different worlds by orbiting around them without landing. On one side, we do not deny to sympathize with the approach of the so-called market technicians (the financial professionals who analyze the charts of historical prices to make investment decisions) since we approach prices as pure geometrical objects as they do. On the other side, considering the nature of this research, we must adopt a ‘scientific’ or ‘statistically valid’ approach, and, as far as we know, no financial professional or market technician has been able so far to provide a solid scientific evidence of his assumption, or justification of his (eventual) over performance. The way we decided to address the topic is also our main value proposition, or our main contribution to research. Orbiting around the two worlds, we try to build a bridge between them by introducing and exploring what we consider a crucial concept: a phenomenon (or its formalization, a time series) cannot be analyzed homogeneously. It is fundamental in fact, in order to fully understand the information contained in a time series, to identify its ‘crucial moments’. With the expression ‘crucial moments’ we mean that, in order to understand a process, we must consider not only its pure quantity (for example a price), but we must take into consideration also how this quantity evolves through time, and if its behavior differs through time. This statement may seem trivial, but it is not. Intuitively, someone may think that the analysis of a process always comprehends the observation of a quantity through time. However, the way we decompose the time may influence significantly our perception of a phenomenon. If we imagine to observe a certain event and put a determined number of flags in different moments in time, through the event itself, the decision where to put these flags may make a huge difference. For example, if we watch a basketball game and put a flag every minute, or every time a player passes the ball, or every time a player scores a point, the final picture on the time line will be extremely different than the original. If research is made without taking such time coordinates into consideration, what we would get is not just an approximation of the phenomenon, but it may be a complete blunder. On one side, there is a strong risk of a compensation effect, i.e. a time series analyzed homogeneously may result in just an average behavior, underweighting its tails but, the main risk, is to impose an artificial structure to the phenomenon itself. It’s easy to deepen this concept by using one of the most extensively addressed topics in finance. Let’s clarify. From a conceptual point of view, we strongly disagree with the statement that the absence of a statistically significant autocorrelation in a time series of consecutive returns means absence of useful information in the relative original

time series of prices. By computing the autocorrelation of returns, in fact, we are completely deleting the entire information set contained into the time dimension, since we impose it, and we (often) link it to the data generating process. This happens because we calculate the autocorrelation on a quantity, the returns, that are calculated over an arbitrary time interval between 2 prices, without considering if the length of this time interval has any relevance on the autocorrelation itself. In finance, this time interval is often linked to the frequency of the data generating process, getting rid, de facto, of any information that may be contained in the time distribution of the price formation process, which has nothing to do with the data generating one. This is only one of the many examples of the implications of our assumption, but it is one of the most relevant since it addresses one of the pillars of financial literature: The Efficient Market Hypothesis (EMH) (Fama, 1970). According to EMH (at least in its weak form, which consider only historical data), the absence of auto-correlation in the time series of returns can be considered a proof a market efficiency since non-auto-correlated means random, and so, the entire price formation process may be considered random as well. While we agree on the statement that non-auto-correlated returns means random returns, we absolutely do not agree to extend the statement to the price formation process, i.e. we do not agree that random returns implies randomness in the price formation process. The reason of our disagreement is grounded on the idea that the reduction of the entire price formation process to a time series of artificially spaced returns, may significantly bias the property we observe. With our statements, we absolutely do not want to imply that the eventual presence of non-random behavior means inefficiency since, in our opinion, there is no reason to link the concept of market efficiency to the absence of a stable extra performance.

Considering the amplitude of this statement, we must dedicate few lines to better specify our point. We said before that we want to approach the research with as little influence as possible from existing literature. In the specific, with respect to the assumptions made by the EMH, we do not intend to make any assumption about rational behavior of market agents. On the other side, by avoiding to get involved in this argument, we do not intend to make any assumption about irrational behavior either. We consider the topic to be not fully relevant (other than being not solvable in science) in order to draw conclusions about the structure of financial markets, being them efficient or not. The main point is that markets can be efficient in the sense of efficiently digesting and pricing information even allowing a portion of market agents to gain 'stable' extra profits. We must underline that with the term 'stable' we do not intend perpetual extra profits addressable to a specific investment fund or fund manager, but rather a performance that last a reasonable amount of time and that may correspond to the peak in a person or Institution career, or to a certain investment strategy. The reason we want to avoid any assumption about rationality, is not because we want to get involved in any irrational behavior model, but simply because we need space. The main drawback of rationality is, in fact, that it freezes the context implying an immediate

absorption of information, meaning basically no absorption process. Markets are not frozen, they are dynamic, and their dynamism is not only given, in our opinion, by the absorption and pricing of exogenous information. Being the market not an overall entity, but rather the sum of its participants, part of this dynamism is given by the simple action-reaction of its endogenous transactions. These transactions are made by participants with a certain degree of asymmetry in their information set and skills. While on one side, a transaction can be the consequence of the absorption process of the exogenous information, on the other side, it can also express the mentioned action-reaction process that constitutes by itself another, endogenous, set of information. The endogenous part can be considered the absorption process of the information. This absorption process may absorb both exogenous information but also, with some sort of a feedback effect, the endogenous set itself. Without the assumption of rationality, this absorption process may last a variable amount of time and it depends on the characteristics of market agents. Without the rationality assumption we also do not need to imply the concept of objectivity. While we consider market to be extremely efficient and agents to be on average, or at least asymptotically rational, in this context a stable extra return can be earned without implying inefficiency. With the concept of endogenous set of information we can add another level of possible knowledge of financial markets. Here, prices can evolve not only because of external inputs, but also because the market reacts to itself, i.e. agents react to each other. This Thesis focuses on this aspect since we consider it, if it exists, the expression of the real nature of human interaction, and so, the nature of the price formation process. The ability to understand this process may constitute a competitive advantage that may origin stable extra returns without implying inefficiency. In our opinion in fact, markets are efficient when they are able to incorporate all the available information. In this context, not all the information is available to everyone at the same time since part of the information can be deductible by the understanding of the absorption process, which is dynamic. Until the competitive advantage of a certain agent remains unknown in terms of saturation of the market, there is no reason to call for inefficiency. In our opinion this must be considered a discovered behavioral characteristic rather than an inefficiency and, considering that the market is able to incorporate all available information, this characteristic may disappear once it is made public and the rest of the market trades conditionally. This learning process, which is part of the endogenous action-reaction mechanism, is what makes the absorption process dynamic.

As previously said, we split this Thesis into two main parts where the bottom up approach covers the first and the top down the second. However, this Thesis is formally made by three separated parts where the first two are related to the bottom up approach and the third focuses on the top down. From now on we will refer as first, second and third part with respect to the formal division that has been decided for this Thesis. In the first we developed a new methodology for the extraction of the most elementary pattern from prices: trend lines. We analyzed the statistical properties of the extracted trends, and we compared the

results with the ones obtained from the implementation of the same methodology over randomly generated time series with the same statistical distribution of the analyzed prices. From a statistical point of view, the methodology extracts asset, and scale invariant trend lines from the data. The duration of the extracted trend lines is in fact a power law, at every frequency, and for any asset class. The same result holds when the methodology is applied to randomly generated prices, so we can conclude that the statistical distribution of the duration of the trend lines is either derived by the statistical distribution of the original times series (that is the same for real and random prices) or is generated by the methodology used. The same coherence does not appear when we look at the short and long term memory of the process. Here the real interesting part starts. As intuition may suggest, the results outlined that the slopes of consecutive trend lines presents negative autocorrelation. If the methodology is applied to both real and randomly generated prices, the ‘strength’ of the autocorrelation is significantly stronger in the real prices, implying the presence of a certain degree of memory in the process. The presence of this memory in the trend lines extracted coexists with absence of autocorrelation in the returns over the same time series, giving a confirmation that the autocorrelation of returns is not a reliable instrument to evaluate the absence of non-random behavior in the price formation process. To obtain our results we analyzed data belonging to all the asset classes and spanning from tick-by-tick to monthly frequency. With respect to low frequencies, we were able to analyze almost 90 years of data of the S&P 500 at monthly and weekly frequency. With such a long range database we were able to observe that the long term memory of market trend has an abrupt interruption after 20 years. If we look at the corresponding random time series, the decay in the autocorrelation function is extremely smooth and without abrupt changes, confirming the difference in the structure of the autocorrelation and the presence of a precise length in the long term memory of the trend lines. The difference in the properties between real and random prices underlines that the 20 years memory we observe is authentic and not biased by the methodology.

In the second part we propose an asset allocation model based on the pattern recognition algorithm developed in the first. The purpose is to find out if the memory that we are able to extract from the trend lines contains any useful information to gain stable extra returns. We applied the methodology to different asset classes, at different frequencies, and we used the autocorrelation of consecutive slopes to develop a long-short investment strategy. We built a portfolio of securities to be invested with that strategy, and we used the conditional probabilities up to 10 different memory steps to compute the portfolio weights. Our entire database is made by time series with execution prices. In order to have a reliable measure of trading costs, we introduced in the routine of our algorithm a negative return equal to what we consider a realistic trading fee for every trade operated by the asset allocation model. The results show a significant extra performance of the asset allocation model with respect to a comparable buy & hold strategy over the same security or basket of securities. The benefits of the methodology are stronger in the case of portfolio diversification where the conditional

probabilities allow to maximize the efficiency of the strategy, taking advantage of the strength of the trading signal across different securities. We applied the strategy to both daily and intraday data. We were able to observe what we can call time diversification, i.e. by increasing the frequency of the portfolio rebalancing, the model maximizes its efficiency. The maximum Sharpe ratios have been obtained at around 40 minutes' frequency. At higher frequencies the trading costs overcomes the time diversification benefits. The results are, of course, obtained through simulation, and without a real implementation into a trading platform. However, the strength and universality of the properties and of the performance obtained, suggest that the price formation process does contain a high degree of unexploited information.

The third part of the Thesis is what we considered the second in terms of approach used, i.e. we switched from the bottom up to the top down. While in the first, bottom up block, we started from the raw material, i.e. prices, in this third part and second block we start from already documented patterns in order to understand their impact. We are obviously not the first ones trying to explore non-random patterns in financial markets. Previous research has well documented the presence of patterns at intraday level. These patterns are mainly observable in the frequency of trades, in volume and in volatility. These three quantities present the same U shaped function with peaks in the mornings and evenings, and lower values in the central part of the day. In order to understand the price formation process, and especially how volatility evolves through time, past research has tried to analyze how trading frequency (time interval between consecutive trades) may be correlated with volatility. Various papers, especially Engle (2000) and Doufour and Engle (2000) demonstrated that these two quantities are positively correlated. However, as we will show in this paper, the presence of intraday patterns in both variables generates a high degree of non-stationarity that may compromise the observed relationship. The purpose of this section is to understand what is the effect of these patterns and the relative non-stationarity that they generate in the time series, in order to have a deeper understanding of the relationship between volatility and trading frequency. We decided to address the issue by standardizing the two time series of waiting times and volatility by using a detrending procedure with a polynomial at different orders. The methodology appears to be effective at reducing to almost zero-level the non-stationarity and the intraday patterns. The picture that emerges after the implementation of the detrending procedure is extremely different from what previously found in literature. In our study, the documented positive correlation between the two variables completely disappears in the first and last 20 minutes of each trading day, while being present (at a lower level) during the central part of the day. The explanation, in our opinion, must be found in the different type of trading operativity that takes place during different moments in the trading day. As pointed out by Admati and Pfleiderer (1988), the beginning and the end of each trading day, are characterized by a different proportion of informed versus uninformed traders with respect to the central part of the day. This difference, as we will further explain,



may generate a different degree of information that is priced by the market, generating a time varying correlation between volatility and trading frequency.

## PART I

# **A New methodology for the extraction of non-random patterns from prices in financial markets. A bottom up approach for the detection of market trends.**

### **1.1 Abstract**

What is the nature of the price formation process? Is it purely random or not? If so, does that mean inefficiency? In this first part we attempt to answer those questions by providing evidence of a structural, non-random, predictable behavior in financial markets, for all asset classes, at any frequency, proposing a theoretical framework in which such evidence can coexist with a broader efficient market hypothesis. We treat the price formation process as a mixture of patterns, or two-dimensional objects, made by a non-divisible combination of price changes over time changes, where both quantities are random variables. In this context, the analysis of returns becomes an overly simplistic measure to understand market behavior, since it alters the object by imposing a deterministic structure to its time dimension, linked to the data generating process. By developing a methodology that is able to extract trend lines from prices and to identify their breakout moments, we try to capture and analyze the full price-time dimensional structure of the most elementary price pattern: market trends. We show that strong memory in trends can coexist with absence of memory in returns over the same time series of prices, separating de facto the concept of efficiency from the concept of randomness. Consequently, the autocorrelation of returns may be considered a proof of market efficiency only and strictly under the agents' rationality assumption, but it cannot be considered a proof of market randomness or unpredictability.

### **1.2 Introduction**

We, humans, researchers, have always been fascinated by regularities and symmetries. Some of those phenomena are meaningful, others meaningless. Some of those are clearly observable, some others need effort to be discovered. The way the human brain engages in this effort is through assumptions and models, trying to translate the complexity of the real world in a language that can be understood. In finance, the main assumption so far has been rationality, while the models, at least the orthodox ones, have always been grounded on the notion of random walk. What happens if we clear the dashboard and we start from scratch? If we use an ontological approach to look at financial markets, what we see is an object. The price formation process is a succession of points in a two-dimensional space: price and time. In this first part we try to find out if the shapes that this object assumes are random or not. Being the price formation process a two-dimensional trajectory, we call these shapes price

patterns. We consider the analysis of the time series of returns to be overly simplistic in order to understand their behavior since it imposes an exogenous, deterministic time structure that may compromise the information set contained in the time dimension, being it linked to the data generating process. To address the issue, we propose a new approach for the detection of these patterns that takes into consideration their random and eventual non-homogenous duration. In this paper we decided to address what we consider the most elementary pattern observable: a straight line, i.e. market trends. By using a simple and orthodox statistical framework, we propose a scale invariant methodology to extract trend lines and their breakouts from any time series of regularly or irregularly spaced prices. By providing strong empirical evidence of a systematic non-random behavior in market trends for all asset classes, at any frequency, we give concrete proof of the presence of a universal, intrinsic, structural non-random behavior in financial markets. Moreover, we de facto separate the concept of efficiency from the concept of randomness by showing that the absence of autocorrelation in returns can coexist with the presence of non-random patterns in prices, implying that the lack of autocorrelation in returns may be considered a proof of market efficiency only and strictly under the Efficient Market Hypothesis (EMH) (Fama, 1970) assumptions, but it cannot be considered a proof of market randomness or unpredictability. In this way we confirm our assumption that the absence of a statistically significant autocorrelation in returns cannot be considered a proof of market randomness since we are able to extract structural non-random behavior from a time series of prices which has non-significant returns' autocorrelation.

The randomness notion in financial market has vexed generations of economists and it has always been dealt with by considering simple time series of consecutive prices, and the eventual memory in their returns. In order to assess if prices move according to a completely random process, as firstly suggested by Bachelier early in the last century (Bachelier, 1900), researchers have focused on the analysis of the autocorrelation of returns (Fama, 1965); Holbrook W. (1934); Cowles A. (1937); Granger e Morgenstern (1963); Samuelson (1965), and, although some anomalies have been observed, especially at lower frequencies, here we agree with the broad empirical evidence that returns do not show any significant and stable autocorrelation, confirming the weak form of the EMH. We agree with this view since those anomalies never appeared to be structural. We underline that in this paper we do not refer to the autocorrelation of absolute or squared returns, or any form of autocorrelation in volatility measures, but only to standard, directional returns calculated either in the form:

$$r_t = \frac{p_t - p_{t-1}}{p_{t-1}}$$

Where  $r_t$  is the return at time  $t$ ,  $p_t$  is the price at time  $t$  and  $p_{t-1}$  is the price at time  $t-1$ . Or in the corresponding logarithmic form:

$$r_t = \ln p_t - \ln p_{t-1}$$

Where  $\ln p_t$  is the natural logarithm of the price at time  $t$ , and  $\ln p_{t-1}$  is the natural logarithm of the price at time  $t-1$ . We consider important to stress this point as the purpose of this paper is to investigate the full nature of the price formation process, and this, in our opinion, must include directional behavior that is lost when considering volatility related measures. We do not consider the violations of the random walk model as the one proposed by Lo (A.W. Lo and A.C. MacKinlay, 1988) to be relevant for our purposes, as no directional predictive power can be extracted from such studies.

Under the EMH, at every moment in time the price reflects all available information, which is rationally treated by market agents. This means that there is only one, rational, and objective criteria of evaluating information, and that prices are the expression of a 'fair' evaluation of the entire information set available at every moment  $p_t$  in time. This is the most widely accepted form of market efficiency, which posits that prices express, at any time, the 'proper value' of a security (Arditti F. 1967). Efficiency, however, is not about constant 'correctness' of prices, rather it is about the tendency of prices to be coherent with fundamental values. (Lakonishok J., et al, 1994; Lintner J., 1965; Lintner J., 1971; Mehra R. and Prescott E., 1985; Russell T. And Thaler R., 1985; Shefrin H., 2000; Shefrin H. and Statman M., 1985; Shefrin H. and Thaler R., 1988; Shefrin H. and Statman M., 1997). If information is constantly and rationally priced, markets are efficient and it is not possible to extract any useful information from the time series of prices in order to predict their future behavior. This perspective should allow agents to use alternatively prices and information to explain one another, but only looking backward in time, while no forecast is possible, since no unknown information is foreseeable. Looking forward, prices are expected to appear, instead, as the result of a random process, with irregular, non-systematic deviations seldom called anomalies, as far as established theories pone (Fama, E.F. and French K.R. 2008; 2010). According to literature, these anomalies are either due to inefficiencies in the information pricing or to the irrationality of market agents, depending on the preferred stream of studies (De Bondt W.F.M., 1998; Benartzi S. and Thaler R., 1995; Black F., 1986; Blume M.E. and Friend I., 1975; Campbell J.Y. and Kyle A., 1993; De Bondt W.F.M. and Thaler R., 1985; De Bondt W.F.M. and Thaler R., 1987; De Bondt W.F.M., 1993; De Long B. et al, 1990; Granger C.W. and Morgenstern O., 1970; Grossman S. and Stiglitz J., 1980; Guijarati D.N., 2003; Haltiwanger J. and Waldman M., 1985; Shefrin H., 2000; Shiller R., 1981; Shiller R., 1984). If the analysis of historical prices does not help to forecast their future value, prices should behave as a random process with no memory, and a test of market efficiency is also a test of randomness in prices (Fama, E.F. 1965a; 1965b; 1970; 1975; 1976b; 1990b; 1991; 1998; Roll R., 1989; Scholes M., 1972). Under this framework, it is fully justified to use the autocorrelation of returns as a measure of market efficiency since returns, in an

efficient market, should be serially independent. If the entire information set is rationally priced at every moment in time, the return on the next 10 minutes will depend only on the new information available in the next 10 minutes. In general, any return over a certain time interval should be independent from any other return over another, previous or successive time interval, no matter if these time intervals have homogenous or non-homogenous length. The absence of autocorrelation in returns is then considered proof of both market efficiency and price randomness since successive returns show no memory.

Another view has risen in the last decades on how to look at financial markets and is called the Fractal Market Hypothesis (FMH). This field has its roots in the seminal work of Benoit Mandelbrot related to Fractal Geometry (Mandelbrot, 1983) and has an exhaustive review in the work of Peters (Peters, 1994). In financial terms, the fundamental claim of this field is that financial markets cannot be described by a Gaussian curve since the tails of the distribution of returns are fatter than normal, and appear to have a power law structure, implying self-similarity, or self-affinity, as stated by Mandelbrot (Mandelbrot, 2004). In such context extreme events are more likely to happen. The main conceptual contribution is, however, broader than that. We can imagine planting a tree. If we know that it is a pine, we can forecast its future shape with a high degree of certainty. Its trunk and its branches will have a certain, known shape. Every main branch will be similar to a smaller trunk with its branches and so on. As we get further into the details of the tree shape, we are less able to forecast the exact shape and the exact position of the branches, until we arrive at the leaves. At that point we are almost unable to forecast the exact position of the leaves' veins, even if we know they have certain proportions and that such proportions are similar to the main shape of the tree, with the veins being similar to the trunk and its branches. This is what self-similarity means in a fractal context, and how the concept of randomness, or unpredictability, dominates over certain scales. If we transpose this concept to financial markets, we can imagine self-similarity in time rather than in space. The price formation process evolves through time with a certain structure. Such evolution has a self-similar behavior in time and not in space. The degree of uncertainty of the tree shape, as the level of details increases, can be conceptually transposed to the volatility in financial markets. Under the Gaussian framework, the volatility increases as the square root of time. The empirical evidence shows that return's volatility increases through time at a faster pace than the one implied by the normal distribution (Peters, 1994; A.W. Lo and A.C. MacKinlay, 1988). Such pace is better described under the FMH framework, which can model with a higher degree of precision the tails of the returns' distribution, being fatter than Gaussian (Peters, 1994). One of the most interesting point of the FMH for this paper is the concept of the fractal object. In the FMH literature, this concept has been applied to the object 'return'. However, if we completely clear our mindset from the existing theoretical framework, and we think about the price formation process as an object, we can simply imagine a price chart. In this chart we have the price  $p$  on the Y axes and the time  $t$  on the X axes. Any object located in this

chart would have two coordinates:  $p$  and  $t$ . At this point the most natural object to imagine is something anyone would actually see in the chart. A trend, a series of consecutive waves, a double minimum, and so on, are the most natural objects any person would trace in a price chart. If we keep thinking under this framework, it would be worth it to analyze those patterns (or objects) and their characteristics, without being biased by the current financial literature that has always focused on the notion of returns. The natural implication that follows is also the real contribution of this paper, i.e. to re-think how we analyze events in the time dimension. Those patterns in fact, may have random duration and the time series of returns, that has a time dimension which is linked to the data generating process, may not be able to fully capture their information, if any. The field that most similarly approaches the study of financial markets in these terms is the same that comprehends the set of tools that have always been used by market practitioners to analyze prices. This field, known as Technical Analysis (TA), has its roots back to Babylon, and finds its formal modern definition in the early 18th century in Japan with the use of candlesticks (Lo and Hasanhodzic, 2010).

TA is the study of past market prices evolution in order to make prediction about future market behavior. In literature, the study of TA has never been particularly depth since it is extremely difficult to quantify its measures and methodologies, being it considered more an art than a science. One of the biggest paradoxes in finance is that the underlying concept behind TA, is exactly the same as the one behind the EMH. According to TA in fact, the only source of information that should be used to predict future prices are the historical prices themselves, since they incorporate all available information, expressing the equilibrium between demand and supply. The big conceptual difference between EMH and TA is that, according to the former, historical prices cannot be used to predict future ones since the price at a certain time  $t$  incorporates all available information up to time  $t$  itself, making previous information worthless. On the other side, according to TA, historical prices should be used to predict future ones properly because they incorporate all available information about market transactions. At first sight it may seem quite contradictory that two very different approaches use the same theoretical assumption as a proof of two opposite statements. Let us clarify how this is possible. As previously stated, according to EMH market agents tend to behave rationally. This assumption has important implication among which i) the idea that there is a rational (unique or fair) value for a certain security given a certain information set, ii) that different individuals have homogenous (rational) preferences, and iii) that the investment horizon does not influence investment choices. On the other side, TA makes no a priori assumptions about agents' behavior. Such information is considered part of the overall information set, without distinguishing between behavioral and security information, and is treated with a pure ontological approach. All the conclusions drawn by TA are based on the observation of historical prices, with the only assumption that they incorporate all available information about supply and demand. The fundamental

implication is that, in such framework, the concept of ‘rational’ or ‘fair’ price does not exist. We must underline that the focus of this research is not the presence or the absence of an assumption about rational behavior. Rather, the focus is the absence of an assumption about agent’s behavior, being such behavior rational or not. In fact, in TA, the absence of an a priori assumption about agent’s behavior does not imply an assumption about irrational behavior. In a framework that is free from the ‘fair rational value’ concept, another important difference between EMH and TA emerges. In EMH we have a static concept of information and price. This means that at every moment in time we can extract the entire information set from the corresponding price, i.e. the entire information set at time  $t$  is deductible from the price at time  $t$ . On the other side, how the information is incorporated and delivered in TA involves a dynamic approach. In fact, in TA, a single price point does not contain all the information available up to that moment. In the TA context, if we want to extract all the information set available up to time  $t$ , we have to analyze the whole time series of recorded prices up to time  $t$ , being the information contained in the full price formation process and not just in the single price point. If we exclude the concept of fair value in fact, preferences may be time varying (Barber and Odean, 2011) and non-transitive (Samuelson, 1950), leaving space to the possibility of a change in price even in the absence of new information. In such dynamic context, in which we assume that the information set is spread across the entire time series of prices, and it involves both information about the security itself and also about agent’s behavior in trading such security, the concept of behavior assumes a new form. In this paper we do not intend to develop a thesis to support TA since we agree with the tremendous difficulty in formalizing its various assumptions and characteristics, and we sympathize with the idea of TA being more an art than a real scientific methodology. However, we like the idea of using the theoretical framework behind it, in order to formulate a new approach to the concepts of market behavior and patterns in prices. In TA, the theoretical concept according to which it is possible to use historical prices to predict future market behavior is relative to the so called technical figures, i.e. particular price formations, or price patterns, that have very specific and observable graphic characteristics, and are supposed to replicate through time, at any frequency. These patterns are generated by the interactions between market Agents, and their self-replicating properties are the expression of a self-similar human behavior which leads financial markets to move (or behave) following actions and reactions, distortions, adjustments, and generally human emotions together with economical of financial valuations.

In this paper we are not even interested in evaluating market efficiency. As previously said, we rather intend to separate the concept of efficiency from the concept of randomness, or probably enlarging the concept of efficiency, becoming adaptive (A.W. Lo, 2005) to itself and not only to exogenous factors. With our methodology we do not need to make any assumption about agents’ behavior. We can include the possibility that the eventual non-random patterns observed in the time series of prices may be caused by the agent’s

interactions, which are the expression of a certain behavior that does not necessarily need to assume full rationality, or objective evaluation criteria. The main point then becomes the coexistence of a certain degree of heterogeneity in agents' behaviors. By supposing a context free from the 'fair value' or 'objective valuation' concepts, we do not intend to imply that human preferences embrace an entire universe of heterogeneity. We consider human preferences to be pretty much similar among individuals, at least from a statistical point of view. However, implying full rationality means to imply a static environment and we feel the need of formulating a framework where there is, at least, the conceptual space for movement. A full rational behavior is, in fact, not a behavior at all, since the concept of behavior and the concept of static cannot coexist. In a behaviorally dynamic context, people interact. A specific field of such interactions is the market, where interactions lead to transactions, i.e. prices. A price is then a result of a behavioral process which results in an objectively measurable quantity. In a static framework, every price at every moment in time includes an objective evaluation of the reality. This means that, in absence of new information to be rationally evaluated, two consecutive prices should be equal. If we get rid of the rationality assumption, we are able to include the concept of movement which is the consequence of the absence of objectivity. In such scenario, two consecutive prices can be different even in the absence of new exogenous information, just because of a change in preferences or because of the reaction of market agents to market transactions themselves. With this last point we can introduce the concept of circular information, and so, circular efficiency. If we imagine financial markets as a place where different players with heterogeneous preferences interact, we can imagine a complex system in which all of its single components do not necessarily know the characteristics of all the other components at every moment in time. Consequently we should not imagine the market as an entity with an overall conscious behavior, but rather with a certain behavior which is the consequence of the interactions among its single components' conscious behaviors. This assumption is possible only if we do not assume rationality. In this scenario, we can say that the market itself is too complex to be perfectly and entirely aware of its complexity. This, obviously, does not literally mean that the market has a certain consciousness and it is trying to be self-aware. Rather, it means that being such complexity originated by the interactions among its components and their behavior, the full understanding of such complexity cannot be known by all its components since, in the very exact moment of its existence, it would be priced by them, and their behavior would change according to the new information set, generating another different level of complexity. In this framework, we have the conceptual space to assume that a percentile of these components can have the knowledge, or the ability (or just the luck!) to understand this complexity better than the remaining percentiles, and that this competitive advantage can be used to gain extra profits at their expenses. Here, winner and losers are not distinguished by rationality or irrationality in their behavior, i.e. winners are not taking advantage of an irrational behavior of the losers. Both winners and losers can have a certain behavior, rational or irrational. We do not need to make any assumption about it. The point



here is the competitive use of information. We assume that useful information can be extracted from prices and that this information comes from the nature of the price formation process and its evolution through time, which derives also from agents' behavior. If we do not assume rationality, the information set contained in prices includes the heterogeneous preference set of individuals. As said before, this can change over time and even be non-transitive. If we suppose that heterogeneity in preferences may follow non-random behavior, information may be extracted from prices in the form of non-random patterns, and used to gain extra profits. If a certain number of market agents may develop a better understanding of such process, they may use the resulting competitive advantage to gain extra profit at the expense of the rest of the market, no matter who is rational and who is not, no matter if rationality exists or not. Once this information is extracted by the agents who (maybe) invested time and resources for that purpose, it can be used to obtain extra profits until a certain saturation point is reached, where the market prices the use of the extracted information itself, making it part of the overall information set. This circular process is the learning mechanism of the market in a non-static, fair-value-free context, in which stable extra profits can be obtained by the use of extracted information without implying market inefficiency. The market efficiency is here represented by the learning mechanism itself. Our main statement is that markets are efficient when they are able to be fully responsive to the circular information set and, in order for such circle to exist, it has to be admitted that useful information may be extracted from the market, implying the existence of non-random patterns. In this context, a market inefficiency can be claimed only if a strategy generating stable extra profits will never reach a saturation point and the strategy itself never becomes part of the overall market information set. In this scenario, stable profitable trading strategies can be justified without implying market inefficiency and we can also justify the well known empirical evidence about asset managers returns in which an over performance can last for a reasonable amount of time until its methodology is priced and its advantage disappears, or the strategy itself becomes ineffective because of changes in market conditions, or simply the appetite of the manager diminishes over his career.

In nature, any behavior is observable and evolves through time and space. In financial markets, the time evolution of prices is the expression of the supposed market behavior and, at least in the plain vanilla world, it can be described by only these two dimensions: price and time. If we imagine a basketball match in which we can only record the distance traveled by the ball at a certain frequency, let's suppose 1 second or 1 minute, it would be extremely unlikely to observe any deterministic pattern relative to the match evolution in such time series, even though we know that some deterministic aspects are present, like the change of the playground side at least every 24 seconds. By recording such quantity at homogenous time intervals it will be likely to collect evidence of a random process simply because the information regarding the behavior of the match has its main discriminant in a certain set of trigger events randomly distributed through time, such as a fault or a score. In this context,

if we can develop a model that is able to recognize these trigger events, and calculate the distance traveled by the ball between them, the statistical result may lead to strong statistical significance. We can extend the same concept to any natural behavior, or natural phenomenon, to draw the same conclusion, i.e. by just recording a quantity at homogenous time intervals, or by linking the intervals of our analysis to the data generating process, we risk to lose a significant degree of information, and to consider random what is not. If we transpose the same conceptual framework to financial markets, we can assume that the autocorrelation of returns on homogenous time intervals cannot be considered an exhaustive measure of randomness since we are compromising the information contained in the time dimension. We address the issue by proposing a new methodology that does not need neither rationality assumption nor any other kind of behavioral assumptions, and that treats the price formation process as a pure two-dimensional object where price changes cannot be separated by their duration in the formation of market trends. With this paper we would like to hopefully point versus an entire new approach to look at financial market research, in which we feel less constrained about the set of assumption we need to formalize to model the reality in which we live. Sometimes, if we completely clear the dashboard before starting to work, some new and exciting evidence may arise. We structured the paper as follows: section 1.3 describes the dataset used; section 1.4 describes the methodology used to derive market trends from prices; section 1.5 exposes the results obtained; section 1.6 is dedicated to the conclusions.

### **1.3 Dataset**

Considering the nature of this study, and its (lack of) assumptions about market behavior, we needed to test our hypothesis on a wide range of data, both in terms of different frequencies and also in terms of different asset classes. One of the main principles of TA, and the FMH, is that market behavior, i.e. the price patterns, are scale invariant. This means that any market behavior does not depend on the time scale or the frequency of the observations and that we should observe the same structures at any frequency. As a trader would say: “If you trade the chart, you should not care about what asset are you trading or at which frequency”. Apart from the frequency, if we want to address a ‘universal behavior’ in financial markets, it is necessary to evaluate eventual patterns in prices across all asset classes. We analyzed data at 5 different frequencies, both regularly and irregularly spaced: irregularly spaced tick-by-tick, regularly spaced 1 minute, 1 day, 1 week and 1 month. We covered all asset classes including: Equity, Fixed Income, Foreign Exchange and Commodity. For the tick-by-tick data, we analyzed 11 stocks of the German DAX30 index: RWE GY Equity, PSM GY Equity, LIN GY Equity, HEN3 GY Equity, DBK GY Equity, DB11 GY Equity, CBK GY Equity, BMW GY Equity, BAYN GY Equity, BAS GY Equity and ALV GY Equity; 9 stocks of the Japanese NIKKEI225 index: 7974 JP Equity, 9433 JP Equity, 6758 JP Equity, 7270 JP Equity, 7267 JP Equity, 9432 JP Equity, 2914 JP Equity, 6954 JP

Equity and 9983 JP Equity; and 4 stocks of the US SPX500 index: BAC UN Equity, GE UN Equity, HPE UN Equity and IBM UN Equity. In the case of the German stocks, the time horizon ranges between the 1<sup>st</sup> of November 2016 to the end of February 2017, for a total of 4 full months. For the Japanese and US stocks, the time horizon ranges between the 1<sup>st</sup> of November 2016 to the 9<sup>th</sup> of March 2017 for a total of 4 months and 9 days. The size of the database ranges between 90.000 and 900.000 observations for each stock, depending on the liquidity, i.e. the number of trades. Moving to regularly spaced data, we analyzed 1 minute frequency data of the entire German DAX30 index composed by the top 30 stocks of the German stock market: PSM GY Equity, DB11 GY Equity, ALV GY Equity, RWE GY Equity, BAYN GY Equity, BMW GY Equity, CBK GY Equity, DBK GY Equity, BAS GY Equity, HEN3 GY Equity, LIN GY Equity, LHA GY Equity, SIE GY Equity, VOW3 GY Equity, EOAN GY Equity, BEI GY Equity, HEI GY Equity, MUV2 GY Equity, FRE GY Equity, SAP GY Equity, MRK GY Equity, ADS GY Equity, DTE GY Equity, DPW GY Equity, FME GY Equity, DAI GY Equity, TKA GY Equity, IFX GY Equity, VNA GY Equity and CON GY Equity. The time horizon ranges between the 25<sup>th</sup> of August 2016 to the 9<sup>th</sup> of March 2017, for a total of 6.5 months. The size of the database in this case is slightly less than 70.000 observations for each stock. At 1 day frequency we analyzed 14 different securities corresponding to all asset classes, including equity and bond indexes, 2 commodities and 3 currencies: SX5E Index, UKX Index, CAC Index, DAX Index, IBEX Index, NKY Index, HSI Index, USDJPY Curncy, EURUSD Curncy, EURGBP Curncy, CL1 Comdty, GC1 Comdty, SPX Index, TR10Y. The time horizon in this case ranges between the beginnings of 1990 to early 2017, for a total of 27.4 years, apart for the SPX500 case which starts in 1950. At 1 week and 1 month frequency we analyzed the SPX500 ranging from March 1928 to September 2017. All the data analyzed and the relative ticker reported here have been obtained from Bloomberg. The composition of our database is mainly due to the availability of data. We intend to enlarge the database to confirm our findings, even though the evidence so far has shown extremely robust and homogenous results across all the time series analyzed.

## 1.4 Methodology

### *Conceptual framework*

We want to verify if it is possible to extract potential non-random patterns from prices. Our assumption is that the price formation process, in order to be fully understood, cannot be reduced to a time series of returns, where their time length is the consequence of the data generating process, being it regularly (as in the case of 1 day, 1 minute or other observation frequencies), or irregularly (as in the case of tick-by-tick data) spaced. We are convinced that any phenomenon, both social and natural, evolves through time and space, and if we pretend to reach a certain understanding of it, we need to know where in space, and when in time,

its key events are distributed. If we analyze the time series of returns we are imposing an artificial and exogenous time dimension, compromising the information set of the original price formation process. In this framework we will not be able to understand any eventual behavior that is more complex than a timely-linear persistency or anti-persistency. Useless to say, financial market can be way more complex than that. In order to gain a deeper understanding of them, we need to consider the price changes relatively to the moments in time where the price events, relevant for the behavior we address, happen. This means to study the phenomenon for how it exists in nature, without imposing any exogenous factor such as assumptions (rationality), or data analysis tools' influence (returns), that may distort (or bias) the phenomenon itself. This means, also, to analyze a non-divisible combination of price changes over time changes, where the two quantities represent the two dimensions, in space and time, of the phenomenon, or the object, called price formation process. With this approach it is obviously not possible to analyze markets straight from the data. We need first to understand where in time the events are distributed with respect to the pattern we want to address. It is crucial to clarify that we are not interested in identifying generic relevant events for financial markets but rather pattern endogenous ones, since we are talking about the intrinsic nature of a phenomenon. The events we intend to identify are strictly related to a determined pattern in the price formation process and they may change from pattern to pattern. It is then necessary to identify which pattern to analyze and develop a methodology to understand its multi-dimensional (including time) evolution.

In this paper we want to address market trends and, in order to do so, we developed a new methodology to identify trend lines and their relative breakout moments. The first thing to do is to define what a trend and a trend's breakout is for our purpose. When we observe a trend in prices, the first characteristic we may see in the chart is a direction (long or short) in the price formation process. The clearer the trend the smaller the fluctuations around the observable direction. The trend faces a breakout when its direction changes significantly, either because its slope becomes more or less steep, or because it changes sign. From this visual description to a formalization, where we specify what a trend is and when we observe its eventual breakout, the difference may cover an entire universe. The first problem is obviously related to the individualization of the beginning of a trend, then we have to evaluate the slope, and ultimately we have to be able to detect the breakout, distinguishing false and real ones. We consider TA and its notions of support and resistance to be a useful starting point for our purpose. In general, supports and resistances are pure chart object, where their domain is given by only two coordinates: price and time, and no exogenous time frequency imposition is made. Their purpose is to describe a price pattern with a pure ontological approach and their formalization can be easily linked to the concept of market trends. Moreover, the key aspect when traders analyze these TA patterns, is related to their breakout moment, and to the distinction between false and true breakout signals, as it is in our case. Overall, the TA approach is well representative of our conceptual framework,

where we do not formulate any assumption about agents' behavior, treating the information set contained into the price, without distinguishing between the eventual theoretical intrinsic value of the security, and the behavioral component of who trades it. Also, the scale invariance, which is a characteristic of our model, is consistent with TA, and the hypothesis that historical prices contain useful information, is both central in TA and a crucial aspect of our research question. However, our use of TA starts and ends with these general contributions, and with the conceptual specification of support and resistance. We do not intend to prove if its application has any scientific or statistical relevance since we agree with



Figure 2



Figure 3

the consideration of it being more a discretionary discipline rather than a science. Since we are dealing with chart objects, prior to try to provide a formal definition of what is a resistance

or a support, we should look at what is considered a resistance and a support by traders. Figure 2 and 3 show them respectively.

As we can see, a resistance (support) is a line on which the price approaches, bouncing down (up) without passing through, where these approaches are represented by the local maximum (minimum) of the price formation process. Although in TA resistances and supports may be



Figure 4



Figure 5

referred also as horizontal lines (as in figures 4 and 5 respectively), here we refer only to the cases in which they present either a long or short direction. We will try to isolate such cases when we will provide a definition of support and resistances later in this section.

The main implication of supports and resistances with respect to our research question is related to their breakout. With this term we define the moment in which the price formation

process violates the previously defined support or resistance line, passing through it. We can observe this phenomenon in figure 2 and 3, where the price passes respectively from below (above) to above (below) the resistance (support) line. The relevance of the breakout is related to the commonly accepted evidence among traders, that the end of this pattern often implies an ‘explosion’ in the price formation process, and that this explosion moves in the opposite direction with respect to the preceding trend, i.e. given a breakout, we should be able to predict, to a certain extent, the behavior in the price formation process, since there is supposed to be an inversion in the direction of the trend. The visual example is again given by figure 2 and 3, where we can observe a wide price movement immediately after the passage through the yellow lines. In this paper we investigate if this behavior has any statistical relevance and can be eventually predicted, or if what we think we observe in the chart are simply random events. Since we are using the notions of support and resistance in order to define market trends, and since there is no formal and unique definition of what is a price support and resistance in literature, we can try to give the first non-exhaustive one as follows:

*Given a Cartesian plane on which we have the time ( $t$ ) on the  $X$  axis and security price ( $p$ ) on the  $Y$  axis, a support (resistance) is one of the infinite combination of straight lines that pass through a specified neighborhood of at least two local minima (maxima) of  $p$ , starting at the first local minimum (maximum) and ending at the point  $t$  where the value of  $p$  falls (rises) below (above) the specified neighborhood of the defined support (resistance) itself.*

Three elements of the above definition need to be specified. First, the concept of Cartesian plan has been used to underline our ontological approach in treating the price formation process purely as a two-dimensional object with coordinates price  $p$  and time  $t$ . The second one is referred to the concept of specified neighborhood with respect to Cartesian coordinate  $p$  and  $t$ . In the attempt of specifying a definition for supports and resistances, it is crucial to incorporate a certain degree of subjectivity in the evaluation of the price patterns. By doing so we do not intend to insert any kind of manipulability which can be used to justify the validity of the defined concept. Rather, the aim here is to specify one very specific concept which allows the same methodology to hold in the entire scenario universe, being able to include and justify, under its specifications, both successful and unsuccessful points of views. This approach is consistent with what is observed every day in the market, where both successful and unsuccessful strategies are often run under the same assumptions and the same methodologies. To have a clearer idea about the concept of the specified neighborhood we can have a look at figures 6 and 7. In the figures we have the price on the  $Y$  axes, and the time on the  $X$  axes. The blue line represents the price formation process, while the red line represents the support of the price. The black circles represent the specified neighborhood of the local minimums, while the green lines represent the amplitude of such neighborhood. We can see in figure 6 that the amplitude of the neighborhood is larger with respect to figure 7. This means that in figure 7 it is possible to trace support lines with a higher precision since

the pattern presents a clearer behavior with the support lines traceable over a narrower range. We will deepen this point after having explained the third and last one as follows. As

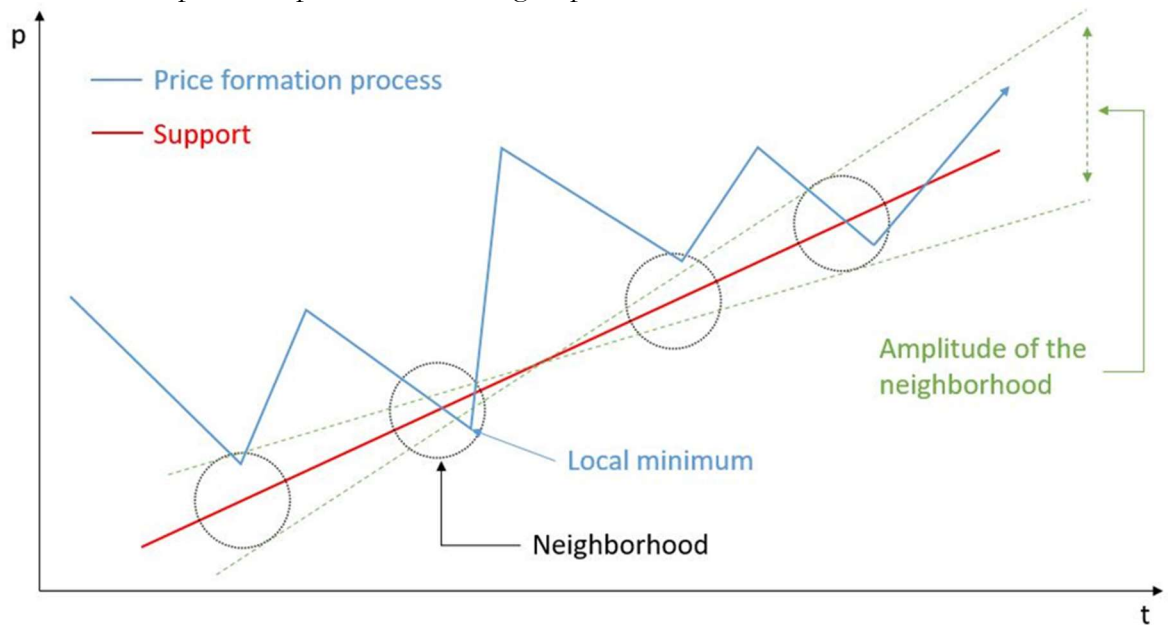


Figure 6

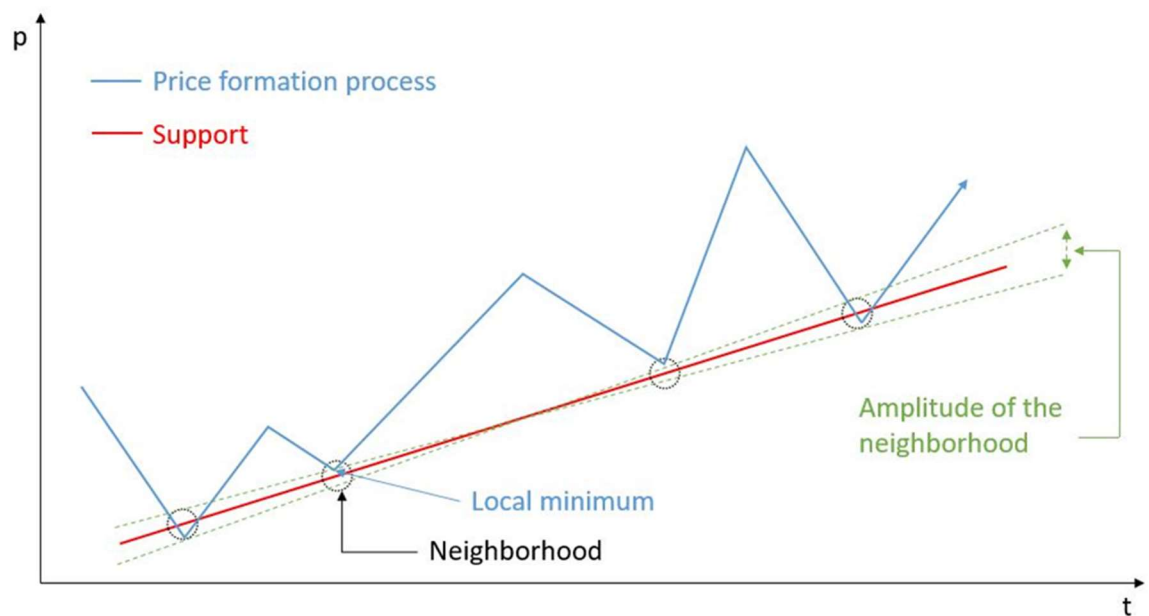


Figure 7

previously said, any isolated local minimum or local maximum may be considered a support or a resistance, depending on the trader's point of view. Here we do not consider such cases and we strictly define supports and resistances as defined lines with at least 2 passage points. This specification has the purpose of linking the concept of support and resistance to the concept of trend lines. In fact, while in TA supports and resistances may be trend lines as well as horizontal lines, here we intend to address only directional behavior, i.e. long or short. Whenever a support (resistance) is just a horizontal line in fact, its level coincides with an isolated local minimum (maximum). The discriminant in this case is the use of at least two points for the identification of the pattern, in order to give a direction (long or short) to the



pattern itself (the trend). Obviously, it can happen to have two (or even more) of these local minimums at the same level of  $p$ , generating horizontal trend lines but the number of occurrences is extremely limited, and the number of cases observed with this approach is statistically negligible.



Figure 8

With respect to the notion of neighborhood, it's worth to explain why this concept is crucial when talking about market trends and, in general, market behavior. When trading, any trader is looking at the price chart of the traded asset. The chart is the same for everyone, and the same holds for the information available. The TA framework and its set of tools are as well the same for everyone. It is obviously impossible for every trader to earn a profit but, since they all are pretty much using the same tools, the presence of winning and losing traders may seem controversial. The solution is the discretionary interpretation of the eventual patterns in the price formation process. In the definition of support and resistance provided before, we underlined the concept of specified neighborhood of a price point in time. We can imagine such neighborhood as the degree of interpretation of the pattern. The smaller the neighborhood, the clearer the pattern, the smaller the degree of interpretation and the smaller the degree of heterogeneity in the strategy of traders in the market. As we can see in Figure 8, from the same chart, multiple resistance lines may be identified depending on the personal point of view of the trader. There is no rule to specify which one is the correct one, and so there is a large space for interpretation. Obviously, ex post, some of those lines lead to successful breakout identification while others do not. However, ex ante, all of them are coherent with the TA general setting and rules, and no one of them could be considered wrong. The multitude of the possible traceable support and resistance lines represents the multitude of the strategies that can be implemented by the traders. In this context, the concept of neighborhood of points can be associated with the universe of different possible strategies, and the different possible trend lines traceable may represent a proxy for supply and demand. This framework is consistent with the concept outlined in the introduction

related to the complexity of the market, and with the fact that such complexity cannot be fully understood by all market participants at the same time. Since the market's behavior is made by the agents' conscious behavior, if those agents fully understood the functioning of the market, they will act accordingly, generating an entirely new market behavior which will incorporate the mentioned understanding in its prices, generating a new, different level of complexity. The application of this concept can be considered the reason of the different amplitudes of the neighborhood. The larger the neighborhood, the higher the degree of heterogeneity in the preferences of the agents, and probably in their degree of understanding of market behavior. This functioning can be translated in terms of supply and demand equilibrium, and its evolution process can give us a suggestion about the evolution of the agents' behavior itself. We can imagine the supply and demand functions of a certain asset that is being traded, as a function of the different strategies implemented by traders, and so, as a function of the size of the neighborhood in the price/time points. When the price pattern is extremely clear, we have a very low degree of interpretation. The consequence is a reduced level of heterogeneity in the supply and demand equilibrium that is spread in fewer hands, where these fewer hands represent the aggregation of the different strategies played by the traders. In such scenario in fact, the pattern is likely to vanish, or face a breakout, since there would likely be soon an imbalance in the supply and demand functions. In this case the immediate consequence is a peak in volatility immediately after the breakout (the 'explosion' previously mentioned), that correspond to the moment in which the supply and demand functions pass from a diversified to a bipolar state, in which there is no heterogeneity neither in the supply nor in the demand functions; then to the collapse, where supply and demand converge to a unique function; and finally to a diversified state again, where both supply and demand have a certain, new degree of heterogeneity. The implication of this assumption is the backbone of this paper, i.e. if the ability to identify the breakout moments in market trends, i.e. the collapses in the supply and demand equilibrium, can give us a predictive tool over market prices, since these breakouts often involve a change in the direction of the price formation process' trajectory. The identification of the breakout moments is obviously not trivial, and it is the other side of the coin with respect to the neighborhood concept. If we want to develop a model that is able to extract trend lines from prices, the key issue is the distinction of the real breakouts from false signals and noise. Such ability can be seen as the ability to understand the price formation process through time, and so, the dynamics between supply and demand that are reflected in the price formation process itself. In fact, once the pattern is established (with all the degrees of subjectivity explained before), the recognition of its real breakout point becomes the key element in distinguishing between successful and unsuccessful traders. If we look at figure 9, we can have a clear idea of a false breakout case in a resistance line. Apart from the various degrees of interpretation in the neighborhood in order to trace the resistance, in this plot we can claim that the pattern can be traced with sufficient clearness. Even though, also in presence of such clear pattern we observe two consecutive false breakouts. If we develop a methodology that will register a breakout in

those cases, its predictive power on one side, but also all the eventual properties that can be extracted by the model to describe the price formation process would be completely useless.

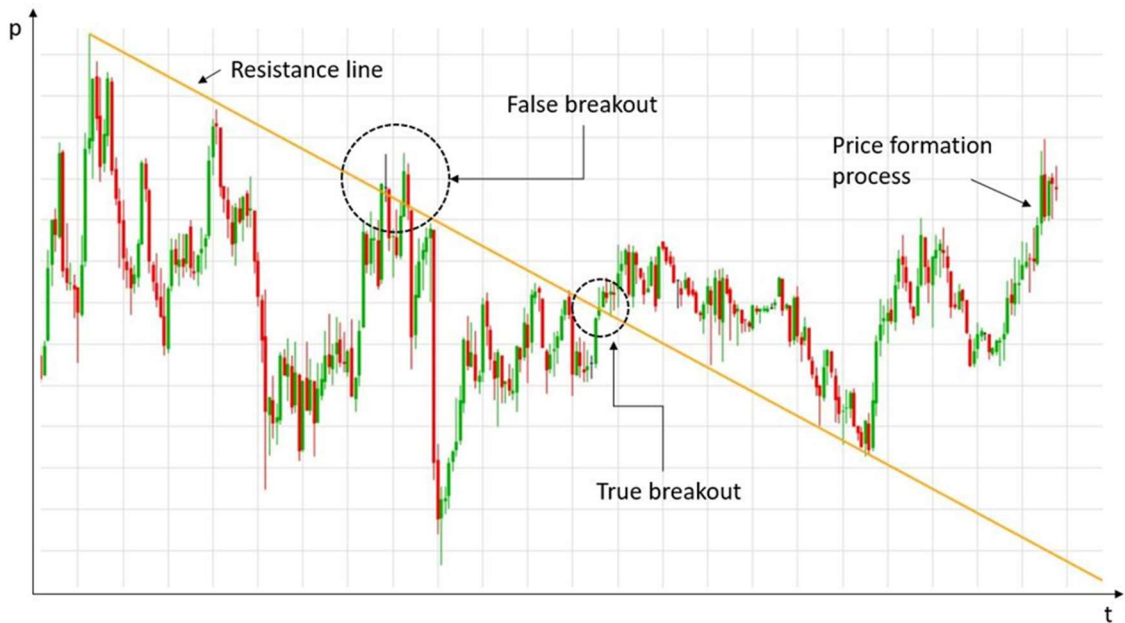


Figure 9

It is then necessary, for our methodology, to own the ability to recognize not only a market trend in place but also to distinguish between false and effective changes in market trends.

### *Model*

In order to extract trends from prices, we adopted the conceptual framework of supports and resistances of TA. We can now refer to a market trend as either a support or a resistance as formulated in the previous paragraph. The key goal, apart from the identification of the trend itself, is to distinguish between a false and an effective breakout. The methodology we present here works independently from the frequency of the data, making the methodology completely scale invariant. Every price point in time is treated as a pure geometrical object with coordinates  $X$  and  $Y$  in the Cartesian space, where the price  $p$  is located on the  $Y$  axis and the time  $t$  on the  $X$  axis. With this approach, the methodology can be applied to both regularly and irregularly spaced data, i.e. tick by tick, without any adaptation. In order to develop a methodology with predictive power, at every point in time  $t$ , only the information set up to time  $t$  can be used, i.e. the price points up to time  $t$ . The methodology outlined here does not use any future information and can be considered as a real time trader analyzing the price chart.

Starting from a certain  $t$ -zero, the model goes through the time series of price points looking for a stable trend. When a stable trend is identified, a trend line is traced until a breakout signal is observed. Once the trend is broken (breakout moment) the model starts the procedure again, looking for the next stable trend and then for its breakout. The model works

with 3 parameters: the time series of prices, their relative moments in time and the minimum number of price points. In this paper, all the results shown, are relative to 5 minimum price points. Every price point is treated as a two dimensional object with coordinates  $p$  and  $t$ , where  $p$  lies on the Y axes and  $t$  on the X axes of a Cartesian plan. We consider a hypothetical time series of prices that goes from  $p_1$  to  $p_T$ , where  $p_1$  is the first price recorded at time  $t=1$ , and  $p_T$  is the last price recorded at time  $t = T$ . Such time series can be considered also the historical time series ending at the present time  $T$ . With this methodology the time interval between consecutive observations (transactions or recorded prices) is irrelevant since the price points are treated as coordinates both in the price and in the time axes. This means that the methodology can be used for both regularly and irregularly spaced data (eg. tick-by-tick). The methodology we present in this paper extracts consecutive trend lines from the time series of prices. It can be used for both purely historical analysis, i.e. prices beginning and ending in past moments, and also for live trading strategies where the time series starts in the past and ends in the present 'live' price. From a purely technical point of view, we can set the starting point at any time in the time series, with the only condition of having at least the minimum number of historical observations that have been previously set (5 in this case). From a financial point of view, the minimum number of observations to be taken into considerations depends on how much data we need in order to obtain robust historical evidence. However, as we will extensively see in the upcoming sections, the statistical properties of the extracted trend lines are so homogenous over time, and across asset classes, that a behavior can be assumed even without running a specific test over a certain time series.

We define with  $P_t$  any price point in time chosen as the first historical observation for the extraction of the trend lines from the time series of prices. Being  $P_t$  part of the time series that goes from  $p_1$  to  $p_T$ , it is redundant to specify that:  $t \geq 1$ , and  $t \leq T - 4$ , considering that we set to 5 the minimum numbers of price points for a trend line. The model starts by running a standard linear OLS regression over the first 5 points, i.e. from  $P_t$  to  $P_{t+4}$ . A slope  $\Delta_{t+4}$  and an  $R^2_{t+4}$  value is then obtained and stored. At  $P_{t+5}$  another regression is run, including the points from  $P_t$  to  $P_{t+5}$ . As done in the first regression, a slope  $\Delta_{t+5}$  and an  $R^2_{t+5}$  value is then obtained and stored. If  $R^2_{t+5} > R^2_{t+4}$  then the trend line with slope  $\Delta_{t+5}$  is used for the price points from  $P_t$  to  $P_{t+5}$ . The procedure continues in the same way until this condition is satisfied for the following price points, i.e. if there is an increase in the  $R^2$  the new improved slope substitute the previous one. If  $0 < R^2_{t+5} < R^2_{t+4}$  the trend line with slope  $\Delta_{t+4}$  is extended to cover the price points from  $P_t$  to  $P_{t+5}$ . Also in this case, once a certain trendline is extended, the extension continues until the new recorded  $R^2$  stays positive. If  $R^2_{t+5} < 0$  a breakout in the trendline is recorded. In this case the trendline that we consider is the one with slope  $\Delta_{t+4}$  that covers the price points from  $P_t$  to  $P_{t+4}$ . Once the breakout is observed the procedure is started again. This time  $P_{t+4}$  is considered the first price point of the new regression as  $P_t$  was before. With this methodology, the link with the concept of neighborhood is given by the degree of sensitivity that we give to the  $R^2$ . In this version of the methodology we use the

maximum degree of sensitivity and for this reason we only consider the case where a new  $R^2$  is greater or smaller with respect to the previous one. We leave space for further studies that can consider changes in slopes and breakouts with respect to a more complex or refined changes in the  $R^2$ , for example, with respect to its percentile distribution. One of the key characteristics of the model is that, once a trend is identified, there is an underreaction to price changes up to a certain threshold, and an overreaction above such threshold, that is the source of the breakout. This threshold is dynamic and varies from trend to trend. However, it is not set by the user or by us, being directly linked to the amplitude of the neighborhood of the trend in place that is automatically detected by the model and, even in this case, not set by the user or by us. The result is a trending procedure where the smaller the neighborhood, the clearer the trend, the higher the sensitiveness to price changes, the easier the breakout.

Figures 10 to 19 show the results of the methodology applied to different time series of

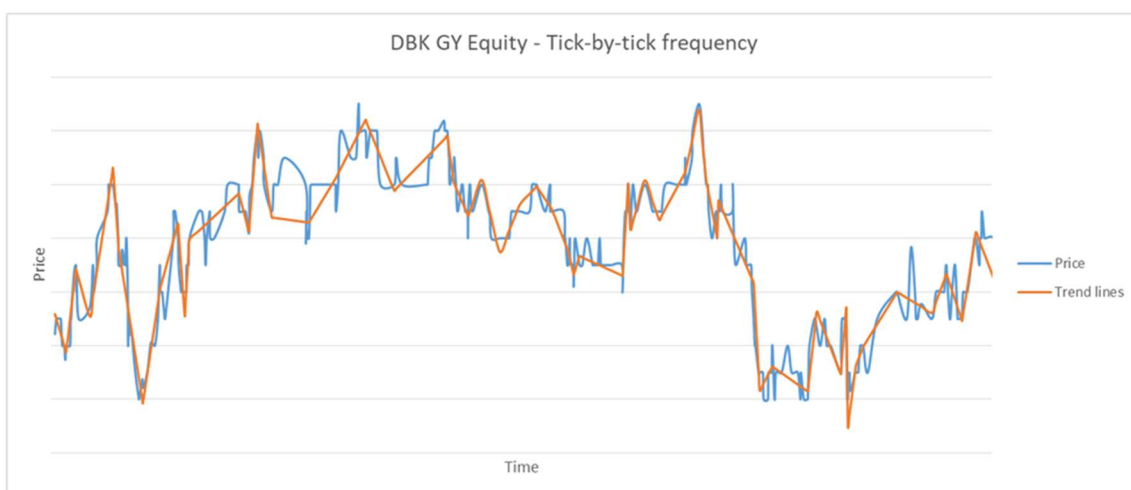


Figure 10

historical prices. In all the figures, the price is indicated by the blue line while the trend lines are represented by the orange line. Figures 10 to 12 show the results on tick-by-tick data for the German market: DBK GY Equity, PSM GY Equity and RWE GY Equity.

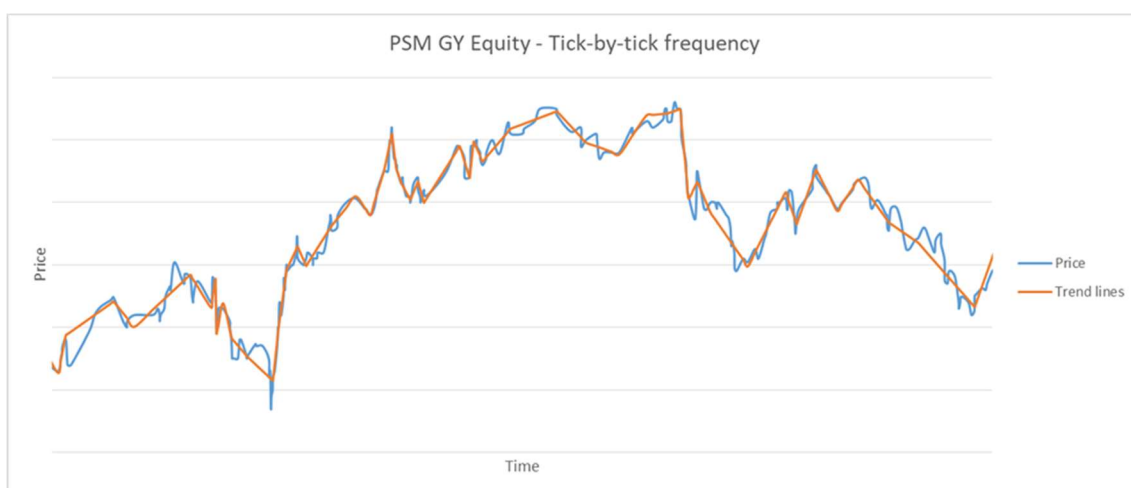


Figure 11

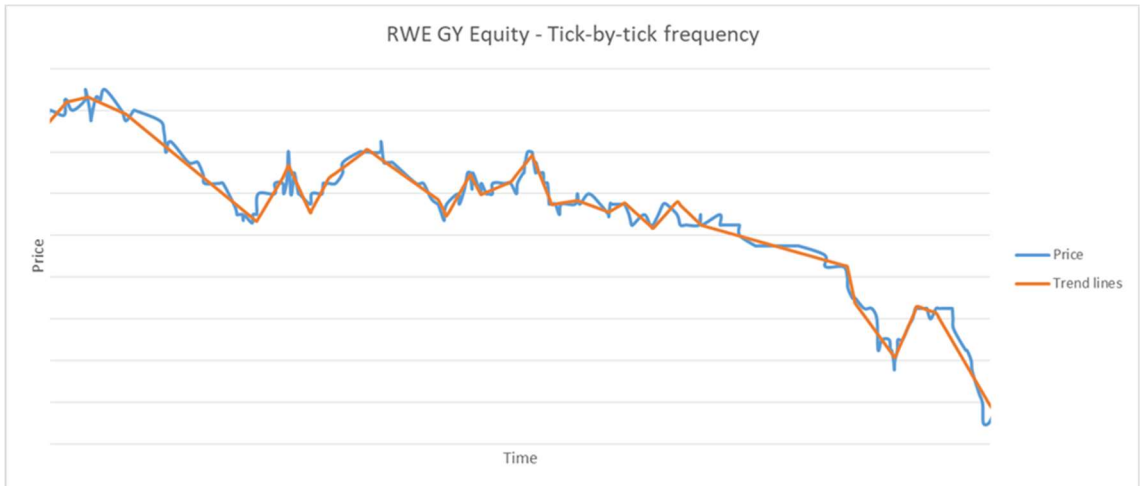


Figure 12

Figures 13 to 15 show the results on the same stocks at 1 minute frequency.

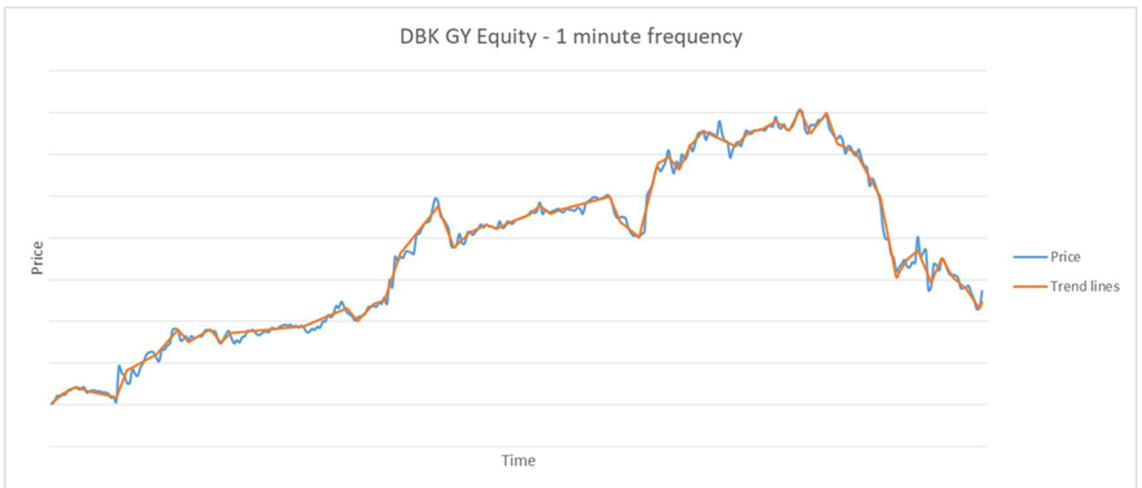


Figure 13

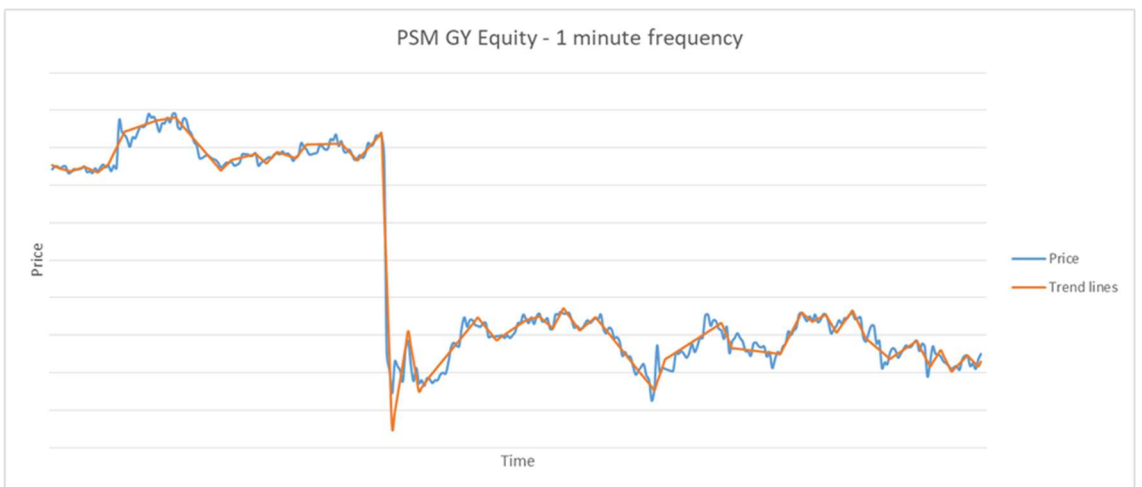


Figure 14

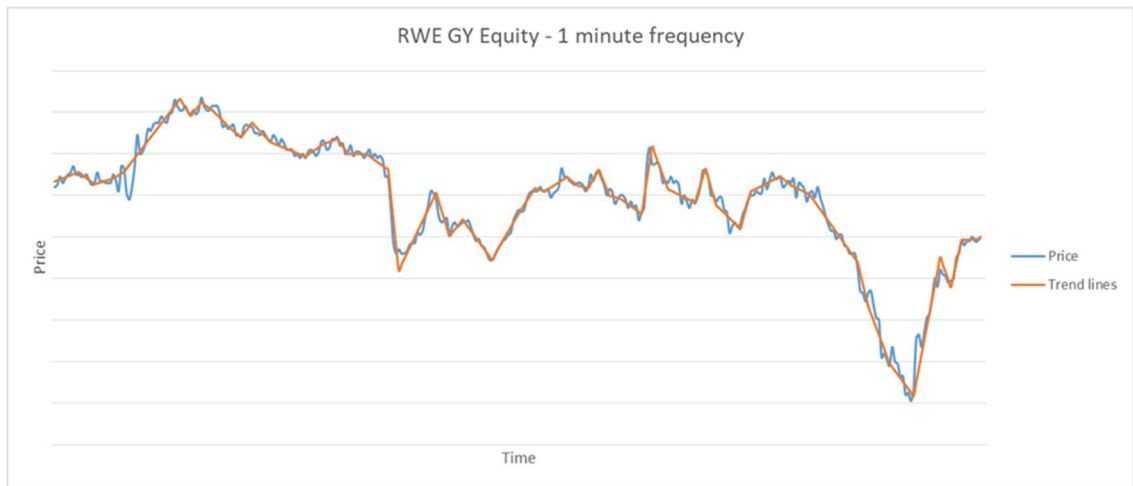


Figure 15

Figure 16 shows the result on an equity index, the S&P 500, at daily frequency. Figure 17 shows the results on a bond index, the 10y US treasury at daily frequency. Figure 18 shows the results on a commodity, gold, at daily frequency. Figure 19 shows the results on a currency, the Euro-Dollar exchange rate, at daily frequency.

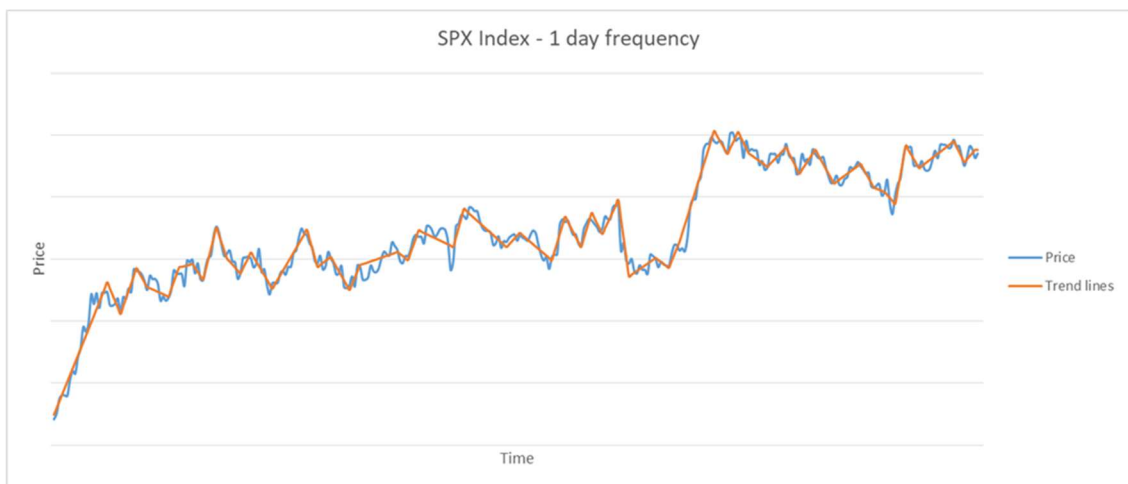


Figure 16

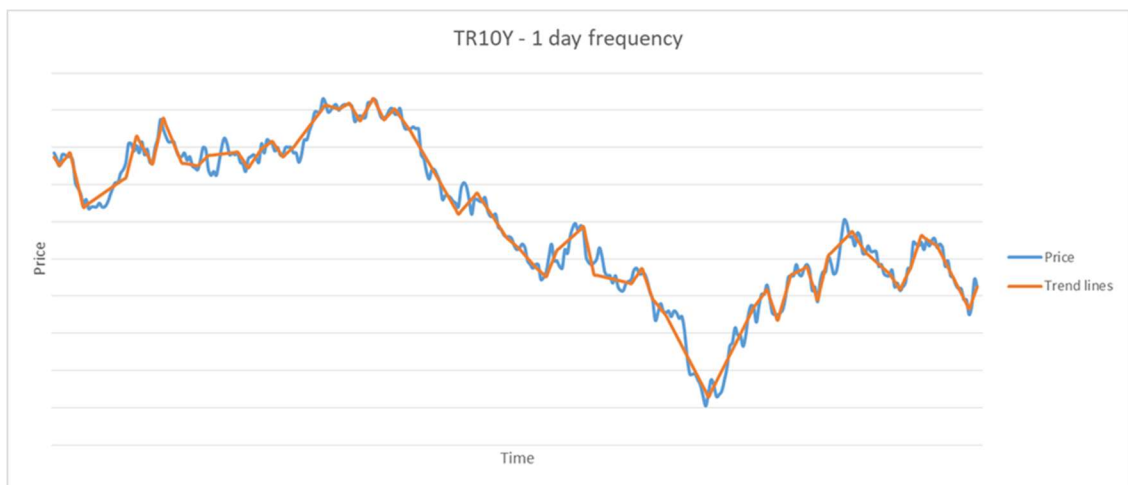


Figure 17

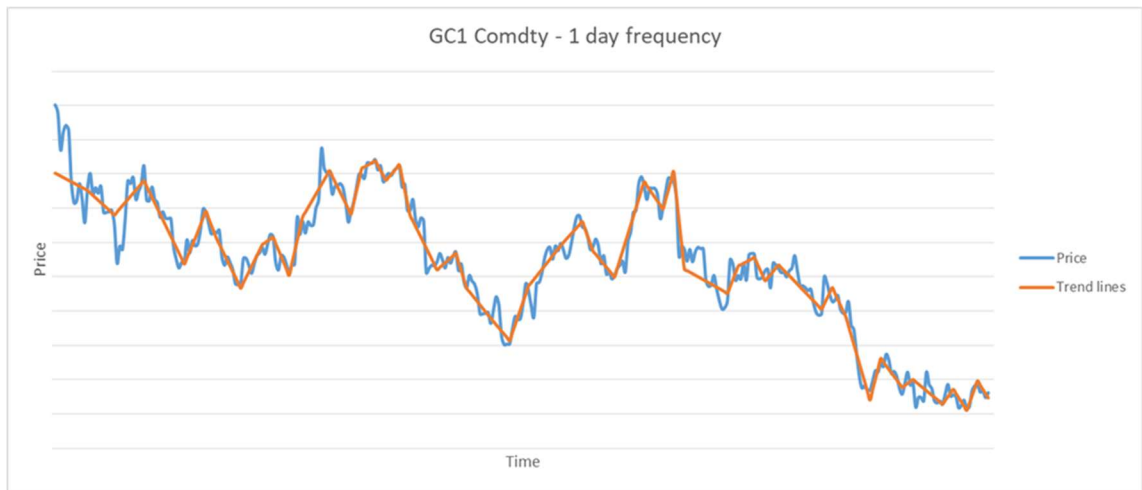


Figure 18

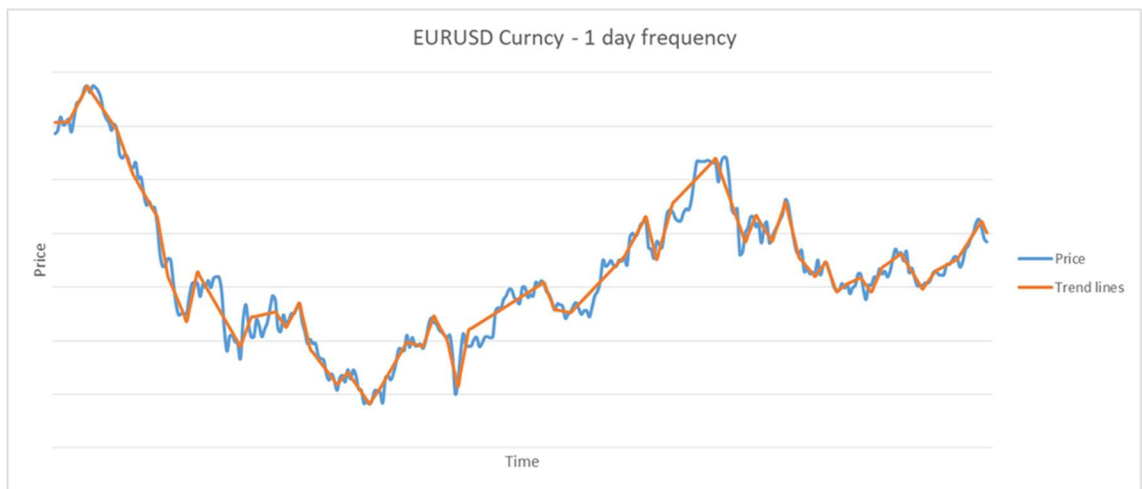


Figure 19

The figures give an idea about how the model works and about its ability to track the price formation process, extracting the relevant movements in terms of trend lines. It is extremely difficult to provide a statistical evaluation of a pattern recognition methodology, being it a hot topic even in computer science. In this paper, and generally in finance, however, we are not so much interested in such evaluation, since the measure of goodness comes in another perspective, i.e. how much information the methodology is able to extract from the price formation process, no matter of its ability to visually track the process itself. This means how much money can we make with that information. In this terms, even if just from an eyeball evaluation, the trend lines and their relative breakout extracted from prices seem to describe extremely well the evolution of the time series of prices, we have to evaluate the properties of these trend lines, and if they allow us to extract useful information to understand and predict market behavior. We will do this in the second part of the Thesis.



## 1.5 Results

Our intention is to find out if there is any non-exploited information that can be extracted from the time series of historical prices, which cannot be directly observed from the standard time series of returns. Our assumption is that the autocorrelation of the standard, regularly spaced returns, is not a reliable measure of price randomness since it compromises the time evolution of the price formation process, imposing an exogenous, deterministic time structure. Since the time evolution of a phenomenon has to be considered as important as the magnitude of the phenomenon itself, we started to talk about price patterns, i.e. chart objects in the two-dimensional space with coordinates price  $p$ , and time  $t$ . In this paper we developed a methodology to extract the most elementary price patterns from the time series of historical prices, i.e. market trends. Once the trend lines have been extracted from the corresponding prices, we need to establish if our assumption was right, i.e. if such patterns contain information not directly deductible from the standard time series of returns. In order to have a reference for the reliability of our results, and to prove that the patterns we claim to observe are not the result of a purely random coincidence, we built another time series of randomly generated prices with the same distribution of the original one. We obtained the random series by calculating the standard returns of the prices, shuffling them and calculating their cumulative sum in order to obtain a new time series of prices with the same distribution of the original one but killing any eventual memory in the price formation process. From now on we will refer to this new, random time series of prices as either random or shuffled prices, and we will compare the trend lines extracted from this time series to the ones extracted from the one of the actual (or real) prices.

### *Consecutive slopes*

The first and most intuitive step is to calculate the slopes of all the trend lines extracted, and check if there is a relationship between consecutive slopes. Trend slopes are calculated by taking the price difference, i.e. the return, from the start to the end of the trend, with respect to the length (or duration) of the trend itself:

$$m_{t,t+\Delta t} = \frac{p_{t+\Delta t} - p_t}{\Delta t}$$

Where  $m_{t,t+\Delta t}$  is the slope of the trend that starts in  $t$  and ends in  $t + \Delta t$ , and  $\Delta$  is the length of the trend which is a random variable. We underline that here we refer to the slopes calculated with respect to the combination of  $p$  and  $t$ , and not with respect to the combination of fitted  $p$  and  $t$ , where with fitted  $p$  we refer to the orange line in the figures from 10 to 19. To better clarify, fitted  $p$ , i.e. the trend lines represented by the orange line in the figures 10 to 19, are the trend lines extracted with our methodology. From these trend lines we are able to identify the points in the time dimension where the breakouts occur.

These points correspond to a change in the slope of the orange line. However, if we want to identify the real slope of market trends, we should not consider the slope of the fitted  $p$  since the price movements are represented by  $p$ , i.e. the blue line. Consequently, the slopes are calculated by taking the moment in time where we observe the change in the slope of the orange line with respect to the corresponding level of  $p$  in the blue line. The real purpose of the methodology is, in fact, to identify the duration of the trend and its breakout moment in time. We proceed in this way since we want to evaluate the eventual properties of the

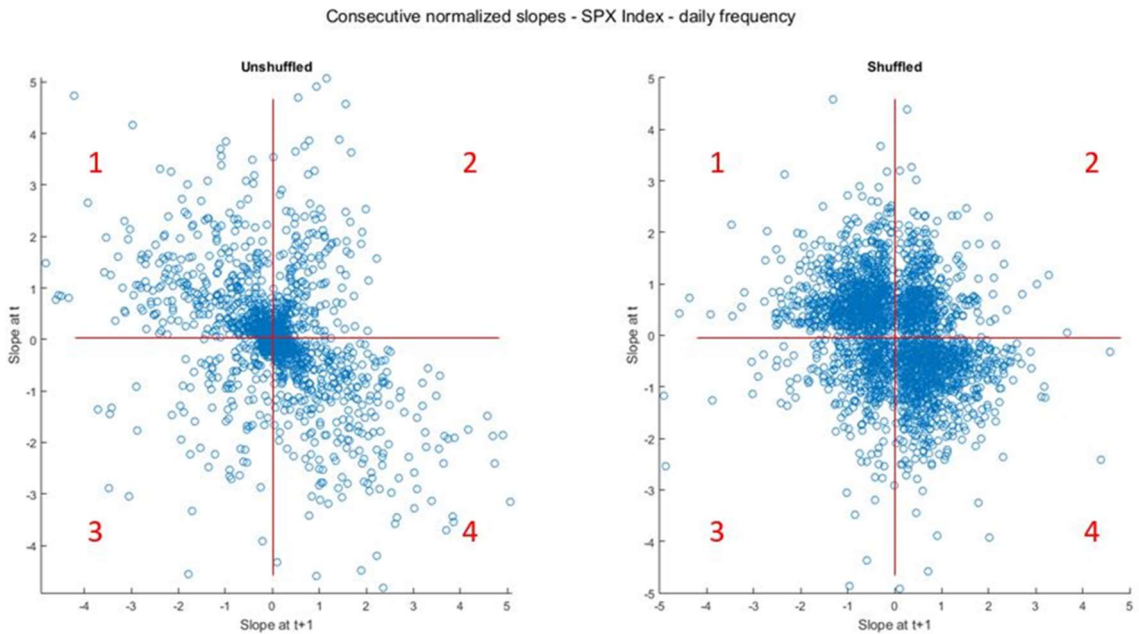


Figure 20

combination: pattern return over pattern duration and considering the fitted values of  $p$  may bias the effective slopes.

The first evidence can be visualized, even just by the naked eye, through the scatter plot in figure 20, showing two consecutive normalized slopes on the two axes.

The slopes have been normalized by dividing their value by the standard deviation of the time series of slopes. We did so since we want to compare the slopes on both real and random data at similar scales. The figure shows and compares the slopes of consecutive trends on i) real prices on the left side of the plot, and ii) shuffled prices on the right side. On the X axes we have the vector of slopes from the second to the last observation, and on the Y axes the same vector from the first to the penultimate observation. This means that on the Y axes we have the slope of the trend observed at a certain time, while on the X axes the slope of the successive trend. If we divide the two plots in 4 quadrants, we can observe respectively the probability of having:

- Quadrant 1 – Negative trend at time  $t+1$  after a positive trend at time  $t$ .
- Quadrant 2 – Positive trend at time  $t+1$  after a positive trend at time  $t$ .

- Quadrant 3 – Negative trend at time  $t+1$  after a negative trend at time  $t$ .
- Quadrant 4 – Positive trend at time  $t+1$  after a negative trend at time  $t$ .

By looking at the figure we can observe that there seems to be a negative relationship between consecutive slopes on real prices, while this relationship seems less pronounced on shuffled data. Quadrants 1 and 4 on the left plot in fact (real prices), appear to be denser with respect to quadrant 2 and 3, while in the right plot (random prices), the difference is less clear. This visual result gives the first suggestion that the behavior of market trends may be far from a random, or memoryless process. On the other side, the fact that such an apparent negative relationship seems to be present, even if in a smaller proportion, also in random data, is due to the nature of the methodology itself and should be treated carefully. The model tries to find a stable trend with its relative breakout. This means to find a similar structure in any process upon which the model is applied, even a random process. Whenever we try to look for patterns in prices, we may find them even if they are totally meaningless. In fact, if you really want to find a precise pattern, you can find it even in a randomly generated time series. This means that finding a pattern does not mean that such pattern has any useful information at all (Mandelbrot and Hudson, 2004). In the case of our analysis however, although it is encouraging for our purpose to find a clear visual difference by applying the methodology to real and random data. As a first observation, we can deduct that the difference in the two plots may represent the amount of information contained in the pure memory of the price formation process.

Obviously, a simple scatter plot cannot be used to draw such a big conclusion, so we need to get deeper into the analysis. If we claim that consecutive slopes have a relationship, we are claiming that there exists a significant autocorrelation in the time series of slopes. In the previous sections we made a strong position against the autocorrelation of returns as a reliable measure of market randomness. However, we must underline that our position is not against the autocorrelation as a statistical tool to measure randomness. In fact, the problem when using this tool on standard returns arises from the time series of returns itself, and not from the autocorrelation function. We intend to evaluate the randomness of a process which mutually evolves through two dimensions, i.e. price and time. For this reason we do not use the autocorrelation of returns for such purpose, since the imposition of a frequency, which is linked to the data generating process, especially at homogenous frequencies, may destroy the information contained in the time dimension. Obviously, the autocorrelation as a statistical tool to evaluate persistency and anti-persistency still holds in our framework, and we intend to use it, not on the time series of returns, but on the time series of slopes. In this way we can measure the randomness of the price formation process leaving intact the information contained in the time dimension, if present, but still using a commonly accepted statistical methodology for process-memory evaluation.

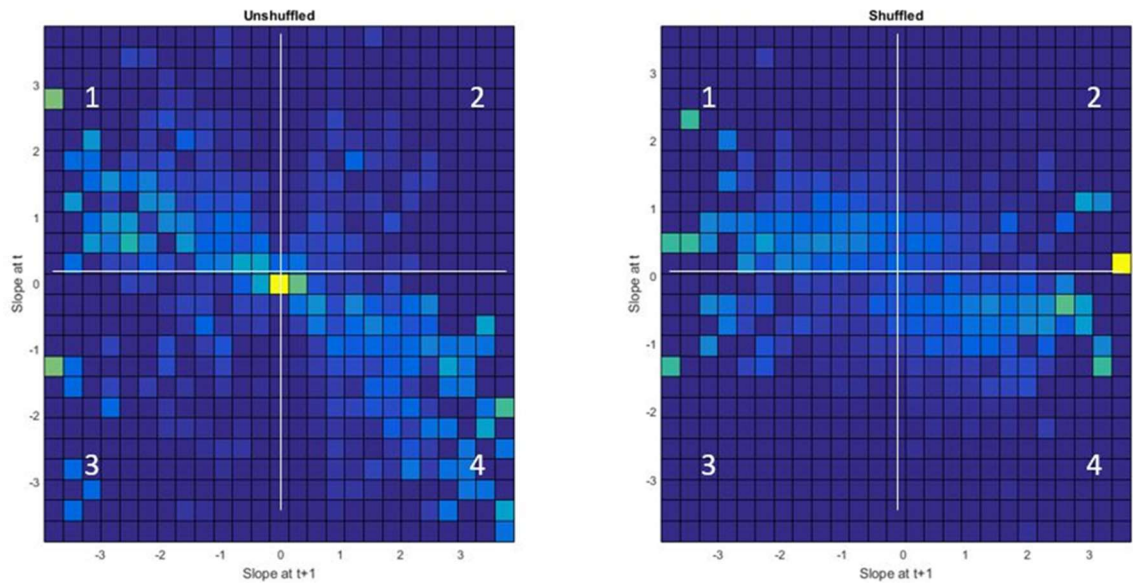


Figure 21

### *Conditional probabilities*

If the relationship observed in the scatter plot exists, we should have a confirmation by looking at the conditional probabilities of consecutive slopes. To do so, we created a heat map shown in figure 21, dividing the scatter plot in a 25x25 bins space.

The intensity of the color of each bin indicates the amount of conditional slopes falling into the relative bin, with the dark blue representing low values and yellow to white representing high values. For simplicity of exposition, we cut at  $\pm 4$  the values of the normalized trend slopes. Every column in the heat map correspond to a full probability distribution summing to 1, meaning that for each slope bin interval at time  $t+1$  we consider the value of the previous slope at time  $t$ . Also in this case the plot shows the comparison between slopes on real, unshuffled prices on the left, and shuffled on the right. We can clearly observe that the conditional probabilities of consecutive slopes derived from real prices are very different from the ones derived from the random ones. As in the scatter, quadrants 1 and 4 in the left plot present a much higher density with respect to quadrants 2 and 3. Such difference in the points' density is present also in the right plot but the proportions are less significant and less pronounced. We have then a confirmation that the information set contained in the time series of slopes is remarkably different if such slopes are extracted from real or random prices. As previously said, the magnitude of this difference represents the portion of the information contained in the pure memory of the price formation process. To have a better comparison between the conditional probabilities on real and random prices, it is useful to look at figure 22, where we report the probability density functions divided in three thirds: two belonging to the tails of the distribution and one belonging to its center. In order to create the conditional probability curves, we used the same dataset used for the previous

25x25 bins conditional probability plot, but this time with only 9x9 bins (not displayed here for simplicity). We divided the 9 columns of the plot in 3 sections of 3 columns each, corresponding to the 2 tails and to the center. We then averaged the values of each of the three groups in order to have a smoother and more readable plot. As it can be seen from the plots, the upper and lower ones show the 2 tails of the distribution, where the curves derived from the real prices present much fatter tails with respect to the one derived from random

Consecutive slopes conditional probability curves - SPX Index - daily frequency

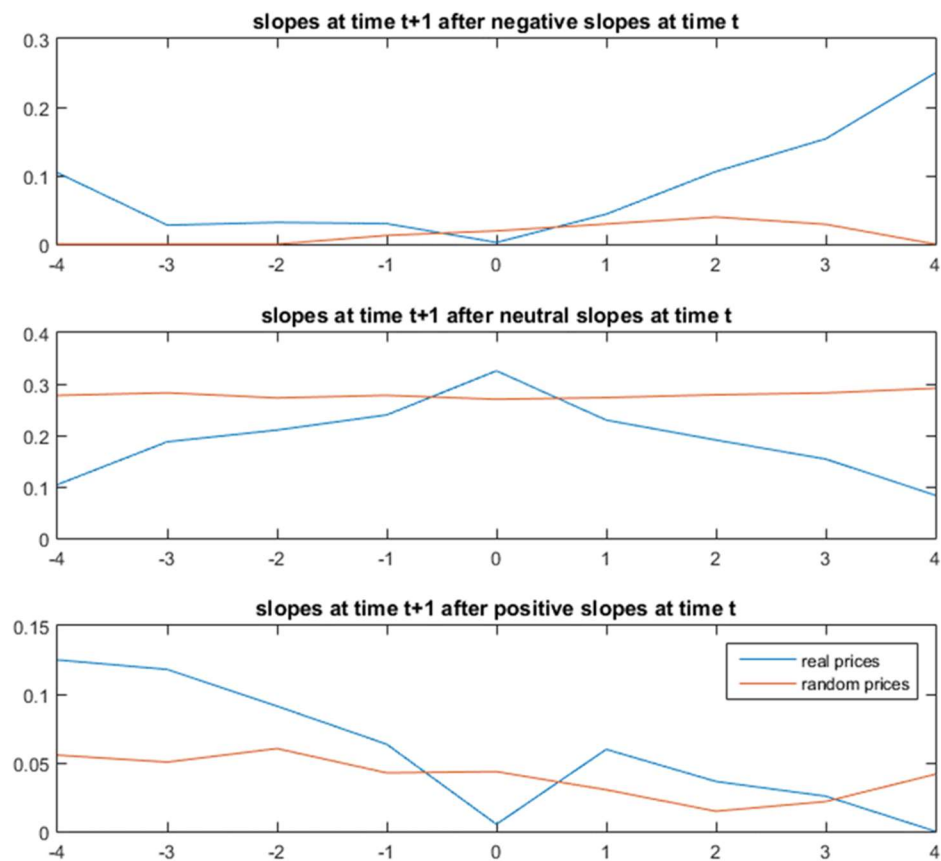


Figure 22

prices. On the other side, the center of the distribution is obviously denser in the case of random prices, with the exception of a peak in the very central part of the distribution. This last plot gives an effective idea of the difference between the conditional probabilities calculated on consecutive slopes of real prices with respect to random ones, confirming the evidence obtained so far.

#### *Autocorrelation analysis*

The conditional probabilities observed in the previous pictures suggest the presence of a significant negative autocorrelation in the time series of trend slopes. Figures 23 to 30

confirm this result. The analysis of the autocorrelation function on all the dataset, including different asset classes, at different frequencies, confirms entirely what we observed in the conditional probabilities. If the model is applied to real prices, with respect to randomly generated ones, the results present a significant difference in the autocorrelation values, confirming that there is a significant degree of information in the pure memory of the price formation process. Since the model is looking for a stable trend and its eventual breakout, it is reasonable to assume that it will be able to find trends even in a random process, as our shuffled time series. It is then necessary, as in the previous sections, to compare the autocorrelation extracted from the trends' slopes of real prices from the random ones. Prior to get into the details outlined by the autocorrelation functions, we underline that such autocorrelation is significant in any time series of slopes analyzed, both from random and not random prices. The autocorrelation values (apart from a single case) are always larger, in absolute terms, in the cases of the real prices. However, if we extract trend lines from random prices, we observe a significant negative autocorrelation in their slopes as well, in any time series analyzed. We are now going to report the result from the autocorrelation analysis and to explain the approach we used. We will then go back to try to give an explanation as clear as possible about the evidence we found. In order to distinguish between the results generated on real prices, and on random ones, we can analyze two aspects of the autocorrelation function:

1. The first 10 lags, in order to evaluate the short term pure memory of the process, and the difference between real prices and a random walk.
2. The full length of the autocorrelation function, in order to evaluate the autocorrelation structure itself of the two processes, and eventual long pure memory behavior.

As we can observe in table 1, by looking at the DAX30 basket, 1 minute frequency, 30 out of 30 securities have stronger lag 1 autocorrelation in real prices with respect to randomly generated ones. The average difference is -0.068 while the minimum difference is slightly above -0.035, and the maximum one is above -0.106. Figure 23 shows the example of PSM GY Equity with the autocorrelation function up to lag 10. We can observe that the autocorrelation is significant at lag 1 and lag 2 in both real and random prices, with a stronger negative value in the real case. If we look at figure 24 we observe the autocorrelation structure up to 9000 lags. We notice that the behavior appears to be noisy and to gradually decay over time. The structure of the noise in real prices appears different, and wider, with respect to the random process but the decaying structure is quite similar. As we will see, this characteristics makes an interesting difference with lower frequencies. Moving to the daily frequency, in figure 25 and 26, if we analyze the SPX Index time series from 1928, we observe the same behavior in the short part of the pure memory, i.e. the first two lags of the autocorrelation function present a stronger significance in real prices with respect to random



Slopes' Autocorrelation - DAX 30 Basket - 1 Minute frequency

Securities 1 to 8

Real prices	PSM GY Equity	DB11 GY Equity	ALV GY Equity	RWE GY Equity	BAYN GY Equity	BMW GY Equity	CBK GY Equity	DBK GY Equity
lag 1	- 0,2429	- 0,2894	- 0,2597	- 0,2865	- 0,2570	- 0,3122	- 0,2776	- 0,2928
lag 2	0,0638	0,0269	0,0476	0,0246	0,0500	0,0804	0,0764	0,0725
bounds +-	0,0196	0,0196	0,0195	0,0196	0,0194	0,0195	0,0194	0,0195

Random prices	PSM GY Equity	DB11 GY Equity	ALV GY Equity	RWE GY Equity	BAYN GY Equity	BMW GY Equity	CBK GY Equity	DBK GY Equity
lag 1	- 0,1566	- 0,2026	- 0,2244	- 0,1804	- 0,2149	- 0,2145	- 0,2334	- 0,1960
lag 2	0,0363	0,0591	0,0612	0,0466	0,0448	0,0474	0,0431	0,0485
bounds +-	0,0198	0,0200	0,0197	0,0199	0,0197	0,0197	0,0197	0,0197

Differences	PSM GY Equity	DB11 GY Equity	ALV GY Equity	RWE GY Equity	BAYN GY Equity	BMW GY Equity	CBK GY Equity	DBK GY Equity
lag 1	- 0,0863	- 0,0868	- 0,0353	- 0,1061	- 0,0421	- 0,0977	- 0,0442	- 0,0968
lag 2	0,0275	- 0,0322	- 0,0136	- 0,0219	0,0053	0,0331	0,0333	0,0240

Securities 9 to 16

Real prices	BEI GY Equity	HEI GY Equity	MUV2 GY Equity	FRE GY Equity	SAP GY Equity	MRK GY Equity	ADS GY Equity	DTE GY Equity
lag 1	- 0,2825	- 0,3147	- 0,2343	- 0,2308	- 0,2795	- 0,3280	- 0,3166	- 0,2730
lag 2	0,0826	0,0573	0,0566	0,0526	0,0897	0,1132	0,0754	0,0713
bounds +-	0,0194	0,0196	0,0195	0,0195	0,0195	0,0194	0,0194	0,0195

Random prices	BEI GY Equity	HEI GY Equity	MUV2 GY Equity	FRE GY Equity	SAP GY Equity	MRK GY Equity	ADS GY Equity	DTE GY Equity
lag 1	- 0,2336	- 0,2341	- 0,1611	- 0,1864	- 0,2026	- 0,2552	- 0,2324	- 0,2005
lag 2	0,0545	0,0564	0,0218	0,0436	0,0429	0,0635	0,0660	0,0457
bounds +-	0,0196	0,0196	0,0198	0,0198	0,0197	0,0197	0,0198	0,0199

Differences	BEI GY Equity	HEI GY Equity	MUV2 GY Equity	FRE GY Equity	SAP GY Equity	MRK GY Equity	ADS GY Equity	DTE GY Equity
lag 1	- 0,0489	- 0,0806	- 0,0731	- 0,0444	- 0,0769	- 0,0729	- 0,0841	- 0,0725
lag 2	0,0281	0,0009	0,0348	0,0090	0,0468	0,0497	0,0094	0,0257

Securities 17 to 24

Real prices	LIN GY Equity	LHA GY Equity	SIE GY Equity	VOW3 GY Equity	EOAN GY Equity	DAI GY Equity	TKA GY Equity	IFX GY Equity
lag 1	- 0,2572	- 0,2780	- 0,2981	- 0,3087	- 0,2840	- 0,2234	- 0,2840	- 0,3131
lag 2	0,0715	0,0536	0,0840	0,0966	0,0628	0,0552	0,0607	0,0729
bounds +-	0,0196	0,0194	0,0195	0,0195	0,0195	0,0196	0,0195	0,0195

Random prices	LIN GY Equity	LHA GY Equity	SIE GY Equity	VOW3 GY Equity	EOAN GY Equity	DAI GY Equity	TKA GY Equity	IFX GY Equity
lag 1	- 0,2149	- 0,2094	- 0,1986	- 0,2042	- 0,2280	- 0,1744	- 0,2190	- 0,2347
lag 2	0,0552	0,0583	0,0569	0,0557	0,0524	0,0365	0,0532	0,0441
bounds +-	0,0198	0,0197	0,0197	0,0197	0,0198	0,0197	0,0196	0,0197

Differences	LIN GY Equity	LHA GY Equity	SIE GY Equity	VOW3 GY Equity	EOAN GY Equity	DAI GY Equity	TKA GY Equity	IFX GY Equity
lag 1	- 0,0423	- 0,0685	- 0,0995	- 0,1045	- 0,0561	- 0,0490	- 0,0650	- 0,0784
lag 2	0,0163	- 0,0046	0,0272	0,0409	0,0104	0,0187	0,0075	0,0287

Securities 25 to 30

Real prices	BAS GY Equity	HEN3 GY Equity	VNA GY Equity	CON GY Equity	DPW GY Equity	FME GY Equity
lag 1	- 0,2606	- 0,2856	- 0,2459	- 0,2646	- 0,3025	- 0,2853
lag 2	0,0474	0,0899	0,0578	0,0519	0,0601	0,0862
bounds +-	0,0195	0,0194	0,0195	0,0196	0,0195	0,0195

Random prices	BAS GY Equity	HEN3 GY Equity	VNA GY Equity	CON GY Equity	DPW GY Equity	FME GY Equity
lag 1	- 0,2134	- 0,2079	- 0,2102	- 0,2055	- 0,2447	- 0,2271
lag 2	0,0437	0,0484	0,0485	0,0515	0,0674	0,0597
bounds +-	0,0198	0,0197	0,0196	0,0197	0,0198	0,0197

Differences	BAS GY Equity	HEN3 GY Equity	VNA GY Equity	CON GY Equity	DPW GY Equity	FME GY Equity
lag 1	- 0,0472	- 0,0777	- 0,0357	- 0,0591	- 0,0578	- 0,0582
lag 2	0,0037	0,0415	0,0093	0,0004	0,0073	0,0265

Table 1

ones. However, if we look at longer lags, an interesting fact emerges. Contrary to the intraday case, while the autocorrelation on random prices has a stable, non-significant noise in its whole length, behaving very similarly to its 1 minute frequency counterpart, the structure on real prices presents a much wider and often significant noise until lags between 600 and 800 (meaning between 18 and 24 years of time). After this point the function suddenly,

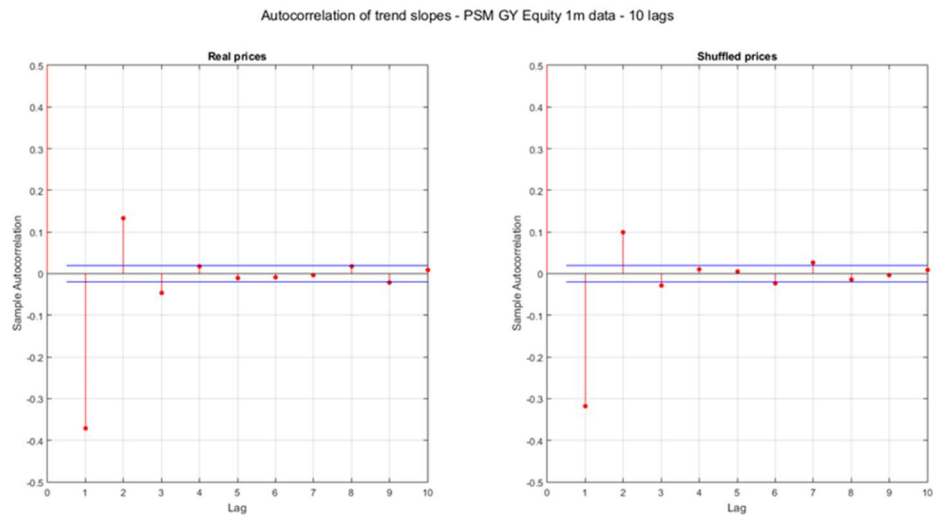


Figure 23

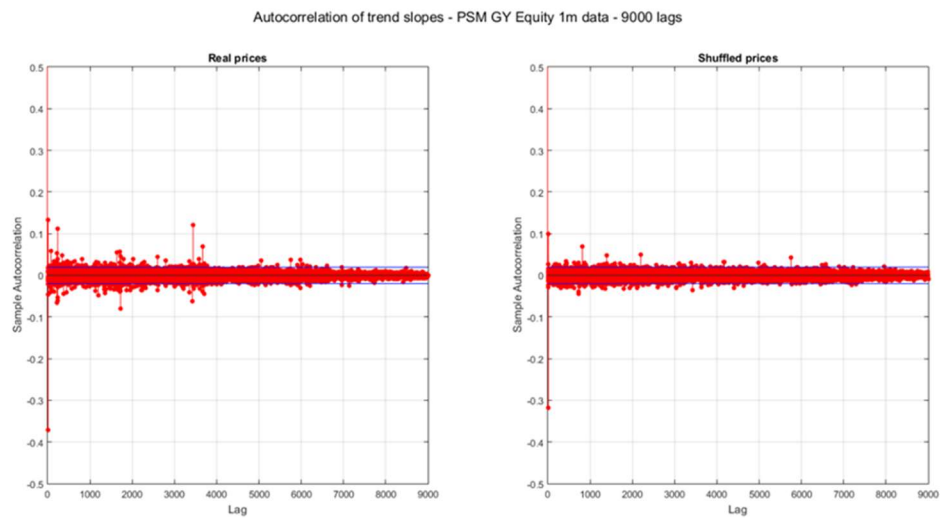


Figure 24

completely flattens, suggesting a possible long term pure memory in the process of market trends, which remains constant (while noisy) for around 20 year and then it completely disappears. It has to be mentioned that the magnitude of the autocorrelation we are observing here is not extremely large in any case.

However, this paper aims at opening up a new research approach, and both the methodology, and the framework in which we are operating, are far from being perfect, or fully tested. This work must be considered an opening line, its methodology still a work in progress, and more a spark for further research and improvements. The same 20 years effect in the long part of the pure memory can be observed also by analyzing the SPX Index at weekly, and monthly frequencies, as showed in figures 27 to 30. The time series, also in these cases, start in 1928. Obviously, for weekly and monthly data, the number of lags at which we observe the collapse of the autocorrelation function is lower with respect to the 600-800 of the daily time series.



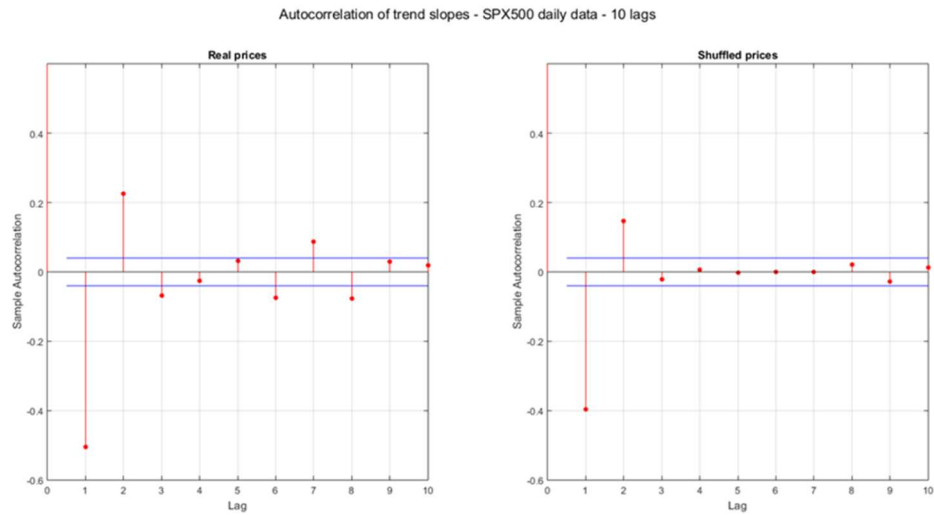


Figure 25

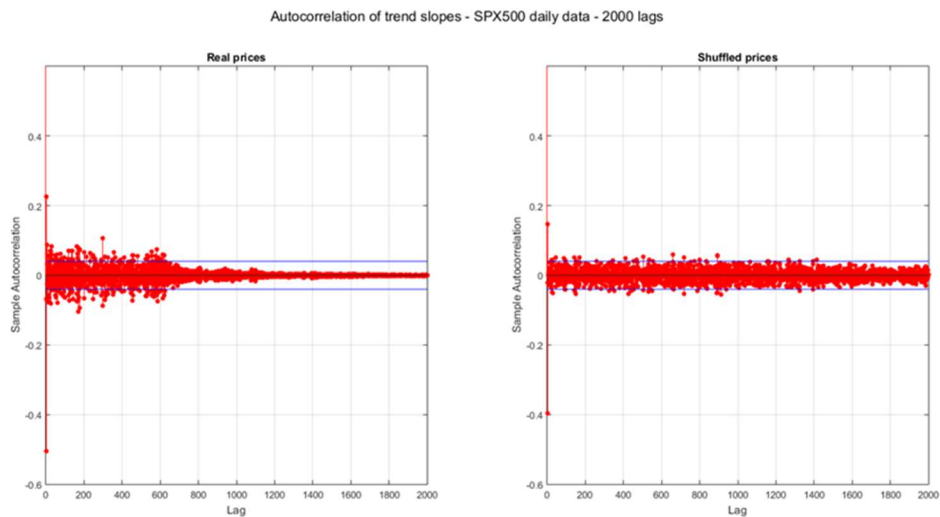


Figure 26

This happens because the structure of the pure memory in the price formation process is not linked to a certain number of observations but strongly linked by the time itself. We must underline that at monthly frequency the first lag is totally not significant and the lower point density of the time series does not allow to draw the same reliable conclusions of the higher frequency cases, having also much wider confidence intervals for the autocorrelation function. We started this paper with the assumption that there exist a non-random structural market behavior, and that such behavior may have scale invariant characteristics, as suggested by the FMH and by TA. The evidence we present from our analysis suggests both that such behavior exists and also that it possess scale invariant characteristics. However, although inside this 20 years space we do observe scale invariance among different frequencies, we must underline that this 20 years maximum length in the pure memory process of market trends raises interesting aspects. One of the solutions can be to assume not the presence of a non-random structural market behavior, but rather a more complex entire set of non-random market behaviors that coexist. In such scenario some of them may be scale invariant

Autocorrelation of trend slopes - SPX500 weekly data - 10 lags

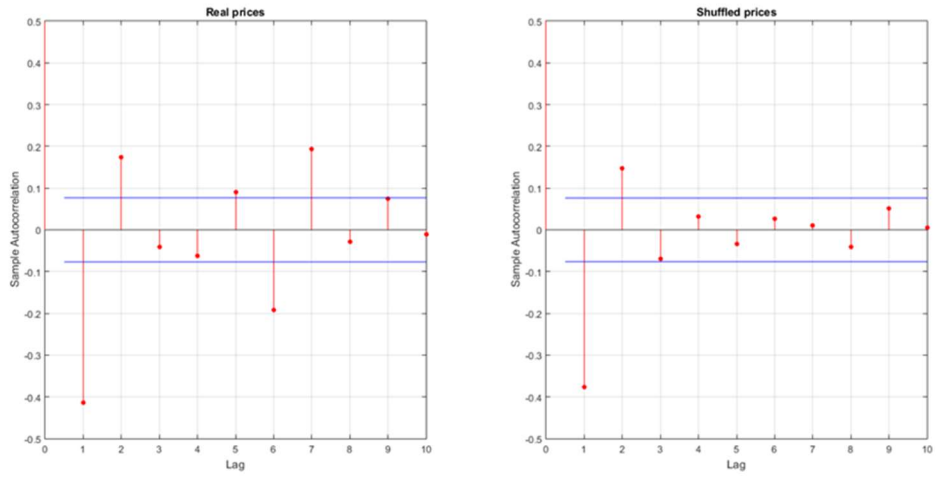


Figure 27

Autocorrelation of trend slopes - SPX500 weekly data - 650 lags

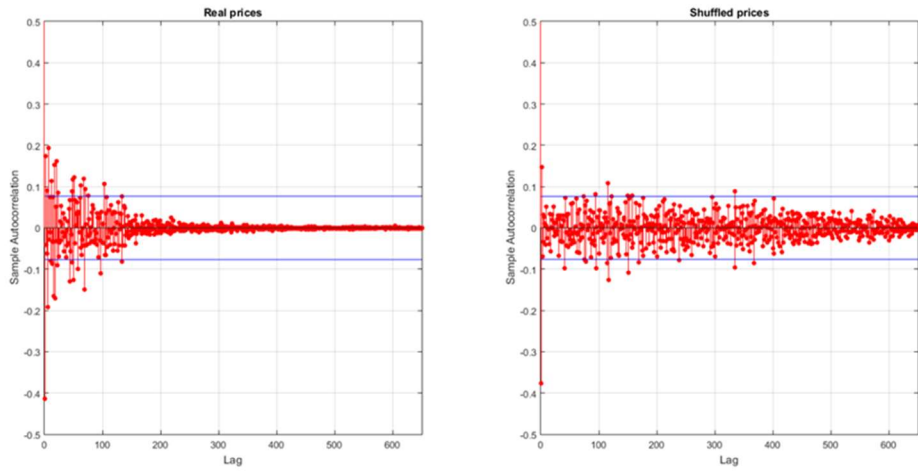


Figure 28

Autocorrelation of trend slopes - SPX500 monthly data - 10 lags

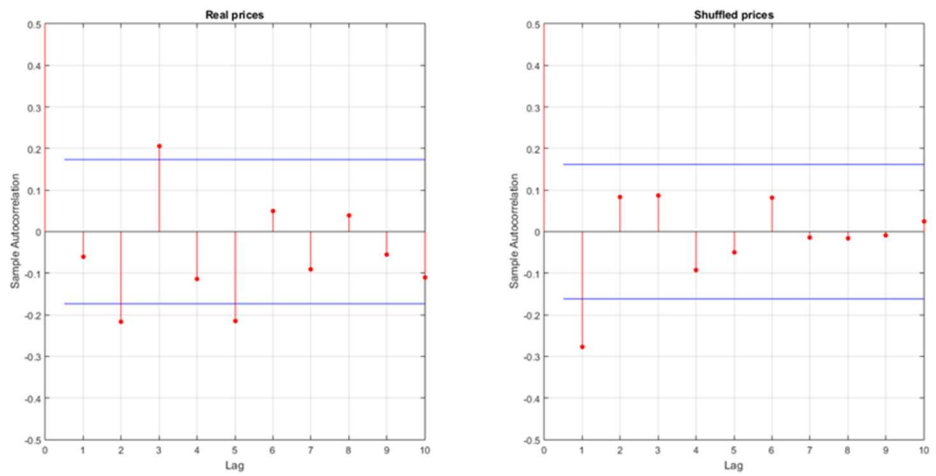


Figure 29

and some others may be strongly linked to the natural time evolution, as the 20 years effect

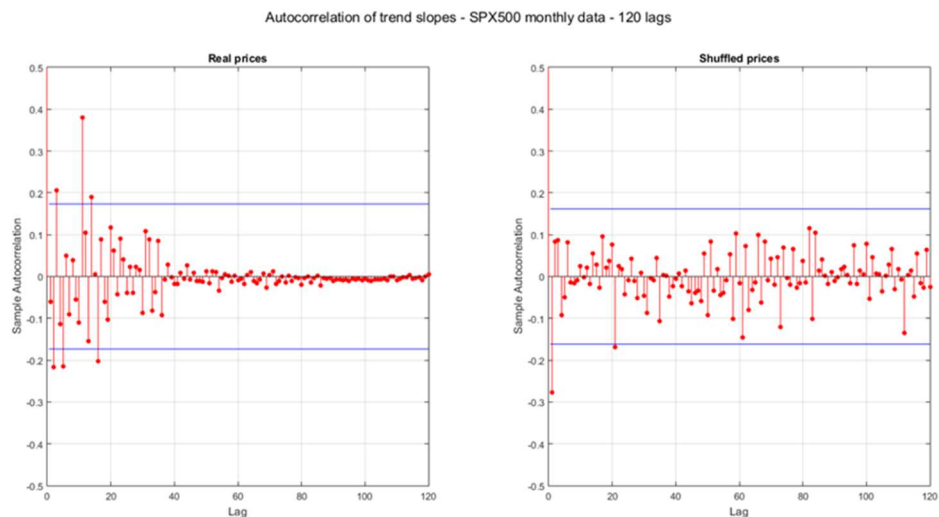


Figure 30

we outlined here. However, such digression would be out of focus for this paper and we would like to leave it for future research. We underline that we did not have the possibility to study this long term pure memory in other assets apart from the SPX Index, since their time length was not sufficient to work on a 20 years process.

In table 2, we show the results of the analysis at daily frequency for a basket of 14 securities, including the SPX Index, other equity indexes, commodities, fixed income and FX. 13 out of 14 securities present results that are coherent with the ones previously observed. In fact, apart from the EURUSD Curncy, all the securities have a lag 1 autocorrelation value that is stronger in real prices with respect to random ones, suggesting the presence of pure memory in the price formation process. The fact that we are able to observe the same autocorrelation structure at different frequencies, and with different securities and asset classes, suggest that there may be a non-random structural behavior in the price formation process in financial markets, no matter the time horizon, no matter the security traded. The obvious implication is that such behavior may be linked to the intrinsic behavior of market agents rather than the type of market or the context in which these agents trade. We report in table 3 the same results reported in table 1 and 2 but, in this case, they are relative to the tick-by-tick data. For simplicity, we decided not to deeply discuss the results obtained at this frequency since, even if the statistical properties are coherent with the other frequencies, it is not possible to evaluate the quality of the information extracted. In fact, as we will show in the next part, the trading costs absorb entirely the profit under a certain frequency threshold, making it impossible to evaluate the goodness of the statistical findings. However, being these findings in line with the other frequencies, we feel confident to claim that the information contained in the trend lines at tick-by-tick frequency has the same quality. We underline, however, that at this frequency market microstructure factors may play a significant role.

Table 2

**Slopes' Autocorrelation - 1 day frequency***Securities 1 to 7*

Real prices	SX5E Index	UKX Index	CAC Index	DAX Index	IBEX Index	NKY Index	HSI Index
lag 1	- 0,3790	- 0,3660	- 0,3473	- 0,3033	- 0,3774	- 0,3078	- 0,3350
lag 2	0,1138	0,0968	0,0778	0,0395	0,1180	0,0781	0,0904
bounds +-	0,0617	0,0609	0,0612	0,0616	0,0618	0,0617	0,0620
Random prices	SX5E Index	UKX Index	CAC Index	DAX Index	IBEX Index	NKY Index	HSI Index
lag 1	- 0,2998	- 0,3223	- 0,3068	- 0,2951	- 0,3220	- 0,2939	- 0,2480
lag 2	0,0754	0,0888	0,0880	0,0461	0,0784	0,0291	0,0893
bounds +-	0,0630	0,0620	0,0619	0,0617	0,0611	0,0621	0,0616
Differences	SX5E Index	UKX Index	CAC Index	DAX Index	IBEX Index	NKY Index	HSI Index
lag 1	- 0,0792	- 0,0437	- 0,0404	- 0,0082	- 0,0555	- 0,0139	- 0,0870
lag 2	0,0383	0,0080	- 0,0102	- 0,0066	0,0396	0,0490	0,0011

*Securities 8 to 14*

Real prices	USDJPY Curncy	EURUSD Curncy	EURGBP Curncy	CL1 Comdty	GC1 Comdty	SPX	TR10Y
lag 1	- 0,2994	- 0,2816	- 0,3600	- 0,3429	- 0,2935	- 0,4245	- 0,3439
lag 2	0,0975	0,0415	0,1243	0,0507	0,0954	0,1530	0,1592
bounds +-	0,0609	0,0617	0,0609	0,0618	0,0617	0,0612	0,0607
Random prices	USDJPY Curncy	EURUSD Curncy	EURGBP Curncy	CL1 Comdty	GC1 Comdty	SPX	TR10Y
lag 1	- 0,2890	- 0,3458	- 0,2956	- 0,2904	- 0,2787	- 0,2991	- 0,3057
lag 2	0,0604	0,0497	0,0841	0,0422	0,0534	0,0703	0,0707
bounds +-	0,0611	0,0616	0,0630	0,0617	0,0625	0,0623	0,0625
Differences	USDJPY Curncy	EURUSD Curncy	EURGBP Curncy	CL1 Comdty	GC1 Comdty	SPX	TR10Y
lag 1	- 0,0104	0,0642	- 0,0644	- 0,0525	- 0,0148	- 0,1254	- 0,0382
lag 2	0,0371	- 0,0083	0,0402	0,0085	0,0420	0,0828	0,0885

Table 3

## Slopes' Autocorrelation - Germany, USA, Japan - Tick-by-tick data

## Securities 1 to 8

Real prices	DBK GY Equity	PSM GY Equity	RWE GY Equity	BAYN GY Equity	BAS GY Equity	BMW GY Equity	DB11 GY Equity	HEN3 GY Equity
lag 1	- 0,2883	- 0,2078	- 0,1748	- 0,2248	- 0,3170	- 0,2873	- 0,1693	- 0,2265
lag 2	0,0759	0,0485	0,0271	0,0963	0,1098	0,0709	- 0,0261	0,0313
bounds +-	0,0094	0,0126	0,0139	0,0103	0,0097	0,0098	0,0175	0,0179

Random prices	DBK GY Equity	PSM GY Equity	RWE GY Equity	BAYN GY Equity	BAS GY Equity	BMW GY Equity	DB11 GY Equity	HEN3 GY Equity
lag 1	- 0,1959	- 0,1485	- 0,1272	- 0,1518	- 0,1687	- 0,1827	- 0,1107	- 0,1386
lag 2	0,0450	0,0425	0,0265	0,0313	0,0378	0,0397	0,0221	0,0282
bounds +-	0,0095	0,0128	0,0141	0,0106	0,0099	0,0099	0,0179	0,0183

Differences	DBK GY Equity	PSM GY Equity	RWE GY Equity	BAYN GY Equity	BAS GY Equity	BMW GY Equity	DB11 GY Equity	HEN3 GY Equity
lag 1	- 0,0925	- 0,0593	- 0,0476	- 0,0730	- 0,1483	- 0,1046	- 0,0586	- 0,0879
lag 2	0,0309	0,0060	0,0006	0,0650	0,0720	0,0312	- 0,0482	0,0031

## Securities 9 to 16

Real prices	ALV GY Equity	CBK GY Equity	LIN GY Equity	BAC UN Equity	IBM UN Equity	HPE UN Equity	GE UN Equity	7974 JP Equity
lag 1	- 0,3065	- 0,2464	- 0,2191	- 0,2372	- 0,2060	- 0,3000	- 0,2395	- 0,2728
lag 2	0,0886	0,0999	0,0562	0,0441	0,0369	0,0710	0,0535	0,0913
bounds +-	0,0113	0,0096	0,0118	0,0086	0,0131	0,0128	0,0111	0,0083

Random prices	ALV GY Equity	CBK GY Equity	LIN GY Equity	BAC UN Equity	IBM UN Equity	HPE UN Equity	GE UN Equity	7974 JP Equity
lag 1	- 0,1823	- 0,1817	- 0,0901	- 0,1097	- 0,1437	- 0,1598	- 0,1441	- 0,1406
lag 2	0,0527	0,0408	0,0133	0,0316	0,0326	0,0383	0,0314	0,0397
bounds +-	0,0115	0,0097	0,0120	0,0087	0,0132	0,0129	0,0113	0,0084

Differences	ALV GY Equity	CBK GY Equity	LIN GY Equity	BAC UN Equity	IBM UN Equity	HPE UN Equity	GE UN Equity	7974 JP Equity
lag 1	- 0,1242	- 0,0647	- 0,1291	- 0,1274	- 0,0623	- 0,1402	- 0,0954	- 0,1322
lag 2	0,0359	0,0591	0,0428	0,0125	0,0042	0,0328	0,0220	0,0516

## Securities 17 to 24

Real prices	9433 JP Equity	6758 JP Equity	7270 JP Equity	7267 JP Equity	9432 JP Equity	2914 JP Equity	6954 JP Equity	9983 JP Equity
lag 1	- 0,3097	- 0,3018	- 0,3092	- 0,3091	- 0,2663	- 0,2582	- 0,2295	- 0,2003
lag 2	0,0841	0,1116	0,1048	0,1020	0,0808	0,0635	0,0801	0,0584
bounds +-	0,0093	0,0090	0,0097	0,0101	0,0103	0,0101	0,0143	0,0142

Random prices	9433 JP Equity	6758 JP Equity	7270 JP Equity	7267 JP Equity	9432 JP Equity	2914 JP Equity	6954 JP Equity	9983 JP Equity
lag 1	- 0,2048	- 0,1519	- 0,1524	- 0,1660	- 0,2065	- 0,1354	- 0,0821	- 0,1672
lag 2	0,0428	0,0385	0,0441	0,0460	0,0538	0,0344	0,0396	0,0357
bounds +-	0,0096	0,0092	0,0098	0,0104	0,0105	0,0104	0,0146	0,0144

Differences	9433 JP Equity	6758 JP Equity	7270 JP Equity	7267 JP Equity	9432 JP Equity	2914 JP Equity	6954 JP Equity	9983 JP Equity
lag 1	- 0,1049	- 0,1499	- 0,1569	- 0,1431	- 0,0598	- 0,1228	- 0,1474	- 0,0332
lag 2	0,0413	0,0730	0,0607	0,0560	0,0270	0,0291	0,0404	0,0227

## 1.6. Conclusions

With this paper we try to propose a new approach to look at financial markets, where we analyze the price formation process with an ontological approach, without making any a priori assumptions about agent's behavior. The main consequence is a context free from the 'fair value' concept. We assume that the price formation process evolves through a non-divisible combination of price changes over time changes, i.e. two-dimensional objects, or patterns, where both their return and their duration are a random variable. We also assume that the evolution of these patterns may contain unexploited information, due to the structure of the market itself. In fact, if we move in a context where prices cannot express an objective measure, their time evolution may be the expression of both aspects related to the security but also to aspects related to the behavior of the agents who trade the security, rationally or not. We expressly do not distinguish between these two components of the

information set. If we want to capture the full information set contained in the time series of prices and evaluate if it contains any non-random pattern, we should not look at the autocorrelation of returns since it imposes an artificial, deterministic structure to the time dimension that is linked to the data generating process, compromising the structure of the information contained in the price formation process. By implementing a methodology which extracts trend lines from prices, with their relative breakout moments, we try to extract the full information set by taking into consideration the eventual random and non-homogenous duration of the price patterns. We applied our methodology to all asset classes at different frequencies, from tick-by-tick data to 1 minute, daily, weekly and monthly observations. From our results we observe that there is a significant degree of unexploited information in terms of non-random behavior in financial markets. In particular, we were able to extract a significant negative autocorrelation in market trends. From our results we observe that the same time series of prices can have both non-autocorrelated, i.e. random returns, and also negatively auto-correlated trends, implying that the autocorrelation of returns can be used as a proof of market efficiency only and strictly under the EMH assumptions (rationality of market agents), but it cannot be used as a proof of market randomness or unpredictability. We demonstrated that the commonly used time series of returns is an overly simplistic measure to analyze the complexity of financial market since it misses a crucial characteristic which is the time location of price events. Financial markets in fact, just like any other observable phenomenon, have a certain behavior, rational or not, random or not, but one cannot disregard that the moment in which an event happens is as important as the magnitude of the event itself. This applies to financial markets as well as any other field of research. In developing our research, we encountered several other evidences, some of them were in line to our initial assumption, some others offered others, interesting views. By analyzing the autocorrelation of market trends we observed an interesting property. While the overall structure seems to be scale invariant, confirming our initial assumption, there seems to be a precise duration in the pure memory of market trends. This duration is around 20 years and can be observed in daily, weekly and monthly data. We haven't deepened that finding enough in this paper, leaving space for future research.

## **PART II**

### **Information hidden in historical prices. An Asset allocation model based on market trends.**

#### **2.1 Abstract**

In this section of the Thesis we implement an asset allocation model based on the trend detection methodology developed in the first part. We apply a long-only and a long-short strategy to verify if the information contained in the autocorrelation function of the trends' slopes can be exploited to obtain stable extra returns from an investment in financial markets. We examined all asset classes at daily frequency and all frequencies for equities, including both single stocks and indexes. Our results show that there is a high degree of non-exploited information in historical prices that allows to gain stable extra profits. The results outlined in this section are relative to investment strategies on both single assets and also portfolio of securities. The main advantages of the extracted information are relative to the reduction of volatility and drawdown, especially when a basket of securities is taken into consideration.

#### **2.2 Introduction**

If we claim that a time series of historical prices contains hidden information, the only way to prove our point is by showing that we can use this information to gain extra profits. In literature there is abundance of papers observing time-varying predictable components in financial markets (Solnik, 1993), or portions of predictability linked to the type of return or asset class (Bekaert and Hodrick, 1992) but there is no evidence of structural predictability. Even the well-known momentum strategies, that appear to be the only ones sufficiently spread across frequencies and securities to be called universal, presents strong limitation to be accounted as structural. Specifically, as pointed out in literature (N. Jegadeesh and S. Titman, 2001), these strategies have significant differences in performances between small and large cap and depending on the asset class. In this Thesis we deliberately do not want to recall a wide range of literature to ground our assumptions since our approach is to try to start from scratch, with the benefits and the limitations that this implies. In literature, every time the topic of excess returns has been addressed, it has been done not with the purpose of investigating the nature of the price formation process but rather to prove wrong the EMH (Fama, 1970). Notwithstanding the importance of this theory, it seems to us that the whole research playground has been biased by its existence, where researchers are looping around a circle without reaching solid conclusions and without exploring what is around that or other circles.

As previously and widely stated in the first part of this Thesis, we do not intend to move against the concepts of efficiency or rationality. We rather want to investigate the nature of the price formation process, providing a solid evidence that it is far from being random. Deviation from randomness can be considered inefficiencies sometimes, especially when they involve predictable deterministic components, but if we consider the dynamic information absorption process formulated in the first part, we can admit structural and predictable behavior without implying inefficiency. What we consider the main value of this paper is the quality of the information that we are testing. When in literature (N. Jegadeesh and S. Titman, 2001) certain overperformance are proved or explained, most of the times they are tested against the concept of efficiency and with the mere parameter of excess return. In this Thesis the main focus is the effectiveness and stability (structural) of the exploitable information extracted with the trend detection methodology. In order to test the quality of this information, we designed and implemented an asset allocation strategy based on the evidence extracted from the negative autocorrelation of the trend slopes, and their conditional probabilities outlined in the first part of this Thesis. To test the robustness of our result we mainly focus on the stability of a time varying Sharpe ratio and maximum drawdown measure over time, rather than purely excess returns. We applied the investment strategy to different securities and, for each case, we compared our results to the performance of a corresponding buy and hold portfolio with the same underlying assets. We show first the results of the strategy based on the SPX500 at daily frequency. We then show the benefits of a diversified portfolio at daily frequency by expanding the investment strategy to a basket of 14 underlying. We finally consider the case of intraday 1 minute frequency for the DAX30 basket. This section is organized as follows: 2.3 briefly explains the asset allocation strategy; 2.4 outlines firstly the application to the case of a single security, secondly it shows the benefits of both diversification and increased trading frequency; 2.5 concludes. We skip the section of the dataset since we use the same one previously outlined in Part I.

### **2.3 Asset allocation model**

The investment strategy we implemented to test the quality of the information extracted from the price formation process in the first section is very simple and works as follow. The model extracts the trend lines from the prices, and it produces a trading signal every time a breakout in the trend line is observed. For each signal the strategy considers the slope of the previous trend and implements a reverting strategy until the next breakout is observed, i.e. if a breakout is observed during a trend with a positive (negative) slope, the strategy produces a short (long) signal to take advantage of the supposed negative autocorrelation in the trend slopes at lag 1.



## 2.4 Application of the investment strategy

### *Daily frequency – single asset*

In the case of the SPX500 at daily frequency (figure 31 and table 4), we consider a long only strategy where the signal is either ‘long’ or ‘stay liquid’, and we compare the result with a buy and hold portfolio on the same underlying.

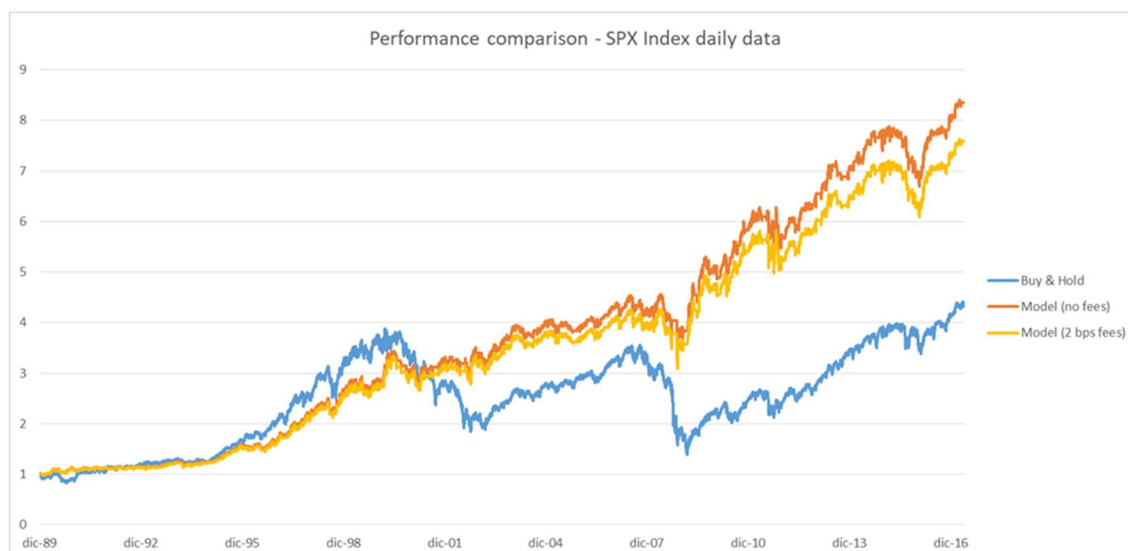


Figure 31 – SPX Index daily data

Table 4 – SPX Index daily data

	Sum of Returns	Standard Deviation	Sharpe Ratio	Max Drawdown
Buy & Hold	5.36%	17.8%	0.3	65%
Model	7.39%	11.60%	0.64	27%

We consider a time frame of 27.4 years, ranging from the beginning of the 1990 to the early 2017. During such period, the average sum of returns of the SPX500 has been around 5.36% per year (we do not consider compounded returns in our analysis), with a yearly volatility slightly above 17.8%. The Sharpe ratio is then calculated by summing all the daily returns, annualizing the result by dividing for 27.4 years of the investment period, and dividing such average annual return by the averaged, annualized standard deviation. The resulting average annual Sharpe ratio is 0.3. The performance of the model applied to the SPX500 at daily frequency, without considering transaction costs, produces a Sharpe ratio that is more than the double than the buy and hold portfolio, with a value of 0.67. The implementation of the strategy involves transactions every week and a half or two on average. Considering retail transaction costs of 2bps per trade (fee charged by the main retail brokers to trade the SPX500) the resulting Sharpe ratio drops at 0.64, still more than double with respect to the

buy and hold portfolio, and only 0.03 less than a transaction-free portfolio, meaning that transaction costs impact is extremely low with such trading strategy. However, the value of the information extracted is not limited to the pure absolute performance. Since the model is trying to extract a systematic market behavior, its main advantage should be the regularity and stability of the competitive advantage of the strategy with respect to a buy and hold portfolio. By looking at the figure, we can clearly observe that the profitability of an investment in the buy and hold portfolio can vary significantly upon the choice of the subscription date. Buying the SPX500 at the peak of the .com bubble for example, would have involved a negative inflation adjusted P&L for more than 15 years. If we move the subscription date from the beginning of 1990 to early 2000, the Sharpe ratio of a buy and hold portfolio drops to a mere 0.04, while the model stays slightly below 0.5, considering 2bps transaction fees. The same evidence obviously appears if we move the subscription date to any peak in the index, including the pre-2008 crisis one. By investing in the model, we can move the subscription date to any other period without influencing dramatically the performance. This evidence is confirmed if we look at the drawdown of the portfolios through their lifetime. The buy and hold one faces a maximum drawdown of almost 65% while the model is less than a half of it, being slightly above 27.4%. This advantage, as we will see in the next sections, is much more pronounced in a diversified portfolio. In this first section, we underline that this stability is reached without relying on a long-short strategy, while in the next sections we will show the impact and the benefits of a long-short implementation together with diversification.

#### *Daily frequency – multiple assets*

The trading strategy defined before uses the information extracted from the autocorrelation function of the trend slopes and, consequently, their conditional probabilities. With such approach it is possible to specify the intensity of the trading signal conditional, not just on the previous slope, but on  $n$  previous slopes. The natural consequence of developing a trading strategy with  $n$  levels of signal intensity, is to mix such signals across different securities to increase the benefit of diversification. In this case the value of the diversification is double since i) there is the standard benefit of risk/return optimization according to the standard portfolio theory, and ii) there is the benefit of signal diversification allowing to increase the weights of the portfolio where we observe a higher conditional probability. The results we present here refers to a portfolio of 14 securities including all asset classes: Equity, Bond, Foreign Exchange and Commodity. We used Bloomberg as data provider and the securities are as follows: SX5E Index, UKX Index, CAC Index, DAX Index, IBEX Index, NKY Index, HSI Index, USDJPY Curncy, EURUSD Curncy, EURGBP Curncy, CL1 Comdty, GC1 Comdty, SPX Index and TR10Y. The results are showed in figure 32 and table 5. As in the single SPX500 case, the 14 time series have the same time horizon of 27.4 years. If we build an equally weighted buy and hold portfolio on the basket of the 14 different

assets, during the 27.4 years investment period, the average yearly sum of returns is 2.04%, with an annualized volatility of 11.03%, resulting in a Sharpe ratio of 0.18. If we use the investment strategy based on autocorrelation of the trend slopes at lag 1 to build an equally weighted portfolio, and if we keep assuming 2bps transaction costs, the resulting Sharpe ratio is 1.28, almost 7 times higher than the buy and hold portfolio, showing a significant increase in the performance spread from the single security SPX500 case. As previously said, one of the benefits of having a diversified portfolio, with the methodology we are proposing, is the possibility of specifying the portfolio weights depending on the conditional probabilities. In its simplest form, the model uses the negative autocorrelation at lag 1 to implement a reverting strategy, going short if the previous trend was long and vice versa. Obviously, dealing with probabilities, not every bet is a success, and there can be positive trends following other positive trends, as well as negative. In order to increase the efficiency of the strategy and better exploit the information extracted from prices, we compute the weights based on the conditional probabilities, considering up to 10 previous slopes (values above 10 do not make a significant difference). We assign a value of 1 to a long (short) strategy to be implemented at time  $t$ , and lasting until the next signal at time  $t+1$ , if the slope between time  $t$  and time  $t-1$  was negative (positive) and the slope between time  $t-1$  and time  $t-2$  was positive (negative). We assign a value of 2 if both the slopes between time  $t$  and time  $t-1$ , and between time  $t-1$  and time  $t-2$  were negative (positive). We assign a value of 3 if the 3 preceding slopes were negative (positive), and so on up to a case in which the 10 preceding slopes were negative (positive), where we assign a value of 10 to the signal power. The result is a vector of signal intensities, with each value ranging from 1 to 10. We then normalize these values by dividing them for their total sum, in order to have portfolio weights summing to 1. If we implement these conditional probabilities to compute the portfolio weights, the resulting Sharpe ratio rises at 1.89, 10.25 times higher than the buy and hold portfolio. In the SPX500 case previously analyzed, we underlined that the main benefit of the model is the stability of the information provided, i.e. the stability of the performance during the investment period. This property is much more pronounced in the case of a diversified portfolio in which, no matter the starting date of the investment period, the resulting Sharpe ratio remains stable, while in the buy and hold case, the choice of the portfolio subscription date makes a tremendous difference in the final performance. The confirmation of these statements comes from the evidence from the maximum drawdown. Already in the single asset case, its value in the model was less than a half with respect to the buy and hold case. Here, with a 14-underlying diversified portfolio, such difference is more than 1 over 20. The Buy and hold portfolio in fact, has a maximum drawdown, during its lifetime, of more than 44%. The model has less than 3%, confirming the fact that the information extracted by the trend lines is so stable to be called structural, and that there is a massive degree of non-exploited information in financial market. In order to establish the source of the predictive power from the autocorrelation, we tried to apply our trading strategy to two different portfolios over the same 14-underlying basket. In the first case we excluded the only case in

which the lag 1 negative autocorrelation in the real prices was weaker than the random prices, i.e. the EURUSD Currency. In the second case we excluded the security with the strongest difference, i.e. the SPX Index. The resulting portfolio, obviously, in both cases was a 13-underlying basket. While we observed that the basket with the exclusion of the currency showed a better performance, the difference was not particularly strong. The consequence is that we are not able to fully evaluate if the diminished performance was due to the negative difference in the absolute level of autocorrelation with respect to the random prices, or if it was due to the general lower level of the autocorrelation absolute value.

Table 5 – Basket of securities daily data

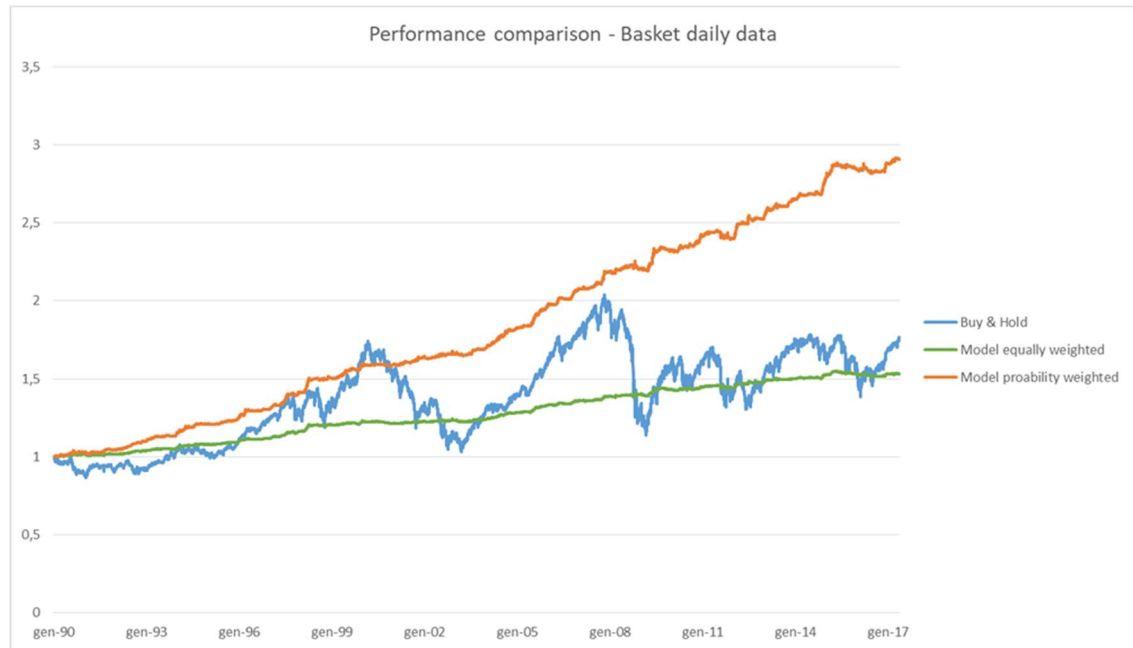


Figure 32 – Basket of securities daily data

	Sum of Returns	Standard Deviation	Sharpe Ratio	Max Drawdown
Buy & Hold	2.04%	11.03%	0.18	44%
Model	3.9%	2.06%	1.89	3%

*Intraday frequency – multiple assets*

Moving to the intraday space frequency, being the methodology focused both on price and time, we would expect to increase the efficiency of the trading strategy with increased frequencies, since we are able to get deeper inside the price fluctuations in what we can call as time diversification, i.e. we are able to increase the efficiency of the trading strategy by taking advantage of a higher number of trends in a more capillary way, with respect to the daily frequency. By increasing the frequency, we are able to capture a higher number of movements but these movements have smaller proportions while involving the same level

of transactions costs. The crucial point is so the individuation of the right trade-off between the benefits of the increased frequency and the minimization of transaction costs. We applied the same trading strategy outlined before for the basket of 14 assets at daily frequency, to the DAX30 basket with data observed at 1 minute frequency. The dataset ranges from august 2016 to march 2017, and the data have been downloaded from Bloomberg. The list of securities is as follows: PSM GY Equity, DB11 GY Equity, ALV GY Equity, RWE GY Equity, BAYN GY Equity, BMW GY Equity, CBK GY Equity, DBK GY Equity, BAS GY Equity, HEN3 GY Equity, LIN GY Equity, LHA GY Equity, SIE GY Equity, VOW3 GY Equity, EOAN GY Equity, BEI GY Equity, HEI GY Equity, MUV2 GY Equity, FRE GY Equity, SAP GY Equity, MRK GY Equity, ADS GY Equity, DTE GY Equity, DPW GY Equity, FME GY Equity, DAI GY Equity, TKA GY Equity, IFX GY Equity, VNA GY Equity, CON GY Equity. In this case we do not use 2bps as a measure of trading costs since we are dealing with single stocks and a more reliable measure for retail fees is approximately 4bps for the German market. Applying the methodology to 1-minute data, we observe that transaction costs completely overcomes any possible profit. By using the 1-minute dataset we build up lower frequencies time series by filtering the original one, finding that the minimum frequency on which it is possible to obtain a profit after transaction costs is 6 minutes. However, the trade-off between transaction costs and the benefit of time diversification at higher frequencies finds its optimal balance at 45-minute frequency. We compare at such frequency an equally weighted buy and hold portfolio composed by the 30 securities of the German DAX30 Index, with the trading strategy with portfolio weights calculated with the conditional probabilities, as in the previous case with the 14 securities at daily frequency. The results are shown in figure 33 and table 6. The buy and hold portfolio, during the 6 months investment period, had a sum of returns of around 8.9% (17.79% annualized), with an annualized volatility of 13.97%, resulting in a Sharpe ratio of 1.27. Volatility is calculated as the square root of the number of 45 minutes interval during 1 year. Volatility measurement at intraday level is a non-trivial issue. For such reason we limit our job to the creation of a common playground for the two strategies to be compared, without the aim of obtaining a reliable measure of annual volatility from intraday data. If we look at the performance of the trading model, the sum of returns is slightly lower at 8.38% (16.75% annualized), while the volatility is incredibly reduced at just 1.93%, resulting in a Sharpe ratio of 8.67, 6.8 times higher than the buy and hold portfolio, confirming the result obtained in the case of the portfolio at daily frequency. Again, even at intraday frequency, the most interesting and valuable property is the stability of the performance. In fact, if we take subsets of the 6 months investment horizon, the performance of the buy and hold portfolio is extremely instable and most of its performance has been made in just 1/10 of its total time horizon, making it extremely crucial (ad risky) the decision of the subscription moment. In the case of the trading strategy instead, the performance is incredibly stable, and the decision on which is the best moment to invest in the strategy during the 6 months period is totally irrelevant. This point confirms the quality of the information extracted from the prices with

our methodology, underlying its stability, and its ability to take advantage of a systematic behavior both through time, and securities. As a confirmation, the maximum drawdown of the buy and hold portfolio is above 18% while the model stays at around 0.5%.

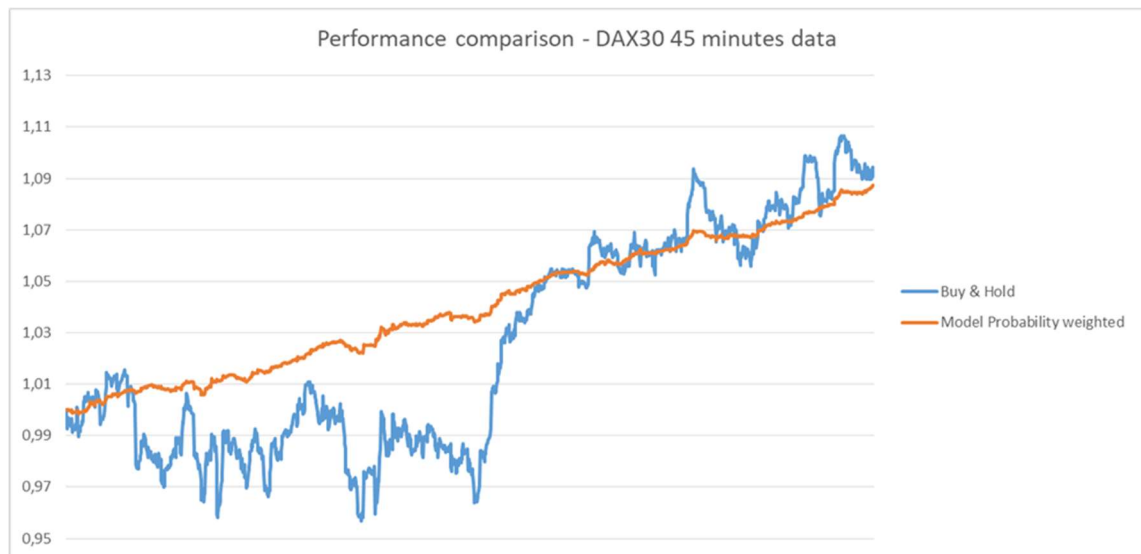


Figure 33 – Intraday frequency DAX 30

Table 6 – Intraday frequency DAX 30

	Sum of Returns	Standard Deviation	Sharpe Ratio	Max Drawdown
Buy & Hold	17.79%	13.97%	1.27	18%
Model	16.75%	1.93%	8.67	0.5%

## 2.5 Conclusions

In this second section of the Thesis we show the result of an asset allocation strategy based on the trend detection methodology outlined in the first part. The purpose is to find out if the information extracted from the time series of prices in the form of a negative autocorrelation in market trends can be exploited to gain stable extra returns. We addressed the topic from a purely operational point of view, with the only intent of testing the eventual presence of structural, predictable behavior in financial market. We intentionally do not want to get involved in any discussion or test about market efficiency or rational versus irrational behavior since we consider the topic to be not fully relevant for our purpose. Usually, when financial researchers try to examine excess returns, they approach the topic as an argument pro or against the Efficient Market Hypothesis, and they reduce their analysis to a mere returns-based argument that can be compatible or not with the concept of efficiency as widely discussed in literature. In our opinion, if we approach the concept of market predictability towards the EMH side, we can be subject to a significant bias that can

compromise the very nature of the analysis, i.e. understand market behavior and not to test previous theories. As we stated in the first part of this Thesis, our purpose is to demonstrate the presence of structural non-random behavior in financial market, and to show that historical prices structurally contains non-exploited information. From our point of view, the presence of predictable structural behavior can exist together with the concept of efficiency, being the two topics deeply separated. Our results in fact, provide evidence that with an appropriate investment strategy we can successfully use the information extracted in the first part of the Thesis to gain stable extra returns. By providing solid evidence of a significant and stable overperformance, which is obtained on time series with non-significant returns autocorrelation, we provide further proof that efficiency and predictability must be analyzed separately. In literature, various studies show the presence of time-varying predictable component in returns, or the possibility of obtaining extra returns for certain assets and/or frequencies under precise circumstances or time frames. However, as far as we know, there is no track in literature of a universal or structural predictability in returns. With this paper we want to demonstrate that a structural and predictable component in market behavior can exist. To support this point, we show not only that with the presented methodology we are systematically able to obtain excess returns with respect to buy and hold portfolios but also that this over performance involves a dramatic reduction in volatility and drawdown, confirming the stability of the information extracted from historical prices. We tested our asset allocation model on a single asset at daily frequency, on a basket of securities covering all asset classes at daily frequency, and on a basket of single stocks at intraday frequency. In all cases we were able to obtain a significant overperformance with respect to a buy and hold portfolio on the same underlying in terms of return, Sharpe ratio and inferior drawdowns. Our results have been obtained through simulation, as it is commonly done in research, so there no implementation of the model in a real trading platform has been done. However, the universality of the results suggests that a high degree of structural and non-exploited information in financial market does exist.

## PART III

### **Rush hours: the real impact of trading activity on volatility.**

#### **Evidences from the Dow Jones Industrial 30.**

##### **3.1 Abstract**

In this third and last section of the Thesis we approach the concept of patterns in financial market with a top down approach. We start from well known intraday patterns in trading frequency and volatility to show their effect in the price formation process. In literature it has been widely shown that trading frequency is positively correlated with volatility. However, deterministic intraday patterns in both variable may bias this relationship. In this paper we show that if we adopt a detrending methodology with a polynomial to reduce these patterns and the relative non-stationarity that they generate, that relationship is strongly compromised. In fact, if we analyze the detrended time series, the correlation disappears in the first and last 20 minute of each trading session, and it appears to be significantly reduced during the central part of the day.

##### **3.2 Introduction**

This section of the Thesis examines the influence that trading frequency has over volatility at intraday level. We analyzed the correlation between the waiting times between consecutive trades and the amplitude of price movements, partially confirming the results found by Engle (2000), and Doufur and Engle (2000), according to which volatility is positively correlated with the frequency of the trades, and so, negatively correlated with the waiting times. However, our study demonstrates that these findings are strongly influenced by the presence of deterministic patterns in both waiting times and volatility. In fact, both time series are highly non-stationary and autocorrelated. In order to analyze an unbiased correlation between the two quantities, we adopted a detrending procedure by normalizing the time series with a polynomial at different orders. The results show that once the stationarity is almost eliminated with an appropriate detrending order of the polynomial, the correlation between trading activity and volatility is reduced during the central part of the day, and it completely disappears at the beginning and at the end of the trading session. The explanation may lie in the activity of informed vs uninformed traders. As Admati and Pfleiderer (1988) pointed out, in the central part of the day the liquidity traders are staying away leaving a high proportion of informed traders, which may translate in a direct information impact from the trades to the price formation process. On the other side, the higher activity of noise traders in the first and in the last 20 minutes of the trading session may destroy the relationship between transactions' frequency and price movements.



Historically, financial research has focused on regularly spaced data, paying not much attention to the distribution of the trades over time, and how this distribution may impact the price formation process. Nowadays, thanks to the availability of high frequency data, the flow is changing, and a new stream of research is trying to model and incorporate the moment in time when a trade occurs, in order to model price behavior. In literature, various studies such as Engle (2000), Doufour and Engle (2000), O'Hara (1995), and Admati and Pfleiderer (1988), investigated the relationship between trading frequency and the price formation process. The main results so far show that trading frequency, i.e. the time interval between consecutive trades, has a significant impact over volatility, and so, over the price formation process. Such relationship, as documented in the cited papers, consist in a positive relationship between trading frequency and volatility, and so, a negative relationship between volatility and the waiting times between consecutive trades. However, as pointed out especially by Admati and Pfleiderer (1988), both trading frequency and volatility present strong intraday patterns. According also to several other studies, mainly related on statistical mechanics (Scalas and Mainardi, 2002; Bollerslev T and Ole Mikkelsen H, 1996), both quantities show a high degree of non-stationarity and a significant autocorrelation with long memory. In this paper we intend to investigate if what has been previously found in literature, i.e. the positive relationship between trading frequency and volatility, still holds if we implement a methodology to correct the non-stationarity and the autocorrelation from the two time series, and we measure their relationship after reducing as much as possible any possible deterministic patterns in place. To address the issue, we implemented a detrending procedure by normalizing the time series with a polynomial at various orders, and we measured the correlation between the two variables through their detrended values. Our findings, if on one side they confirm the literature for the central part of the trading day, they clearly show that the relationship between trading frequency and volatility completely disappears at the beginning and at the end of the trading session. We attempt to give an explanation to our results by referring to the activity of informed versus uninformed traders, and how their different activity during the trading day may impact, or not, the price formation process. The paper is organized as follows: section 3.3 outlines the dataset used; section 3.4 outlines the methodology and the results; section 3.5 elaborates the conclusions.

### **3.3 Dataset**

The dataset is made by tick-by-tick data of all the component of the Dow Jones Industrial 30. The dataset has been downloaded from Bloomberg and is outlined in table 7. The number of days for each stock varies, as from a computational point of view the data provider allows to download only a certain amount of data points and, depending on the level of liquidity for each stock, the time length may vary. The most liquid stock appears to be Apple with over a million transactions in 139 days, while the less liquid is Travelers with slightly over 130,000

observations in 196 days. The labels in the second column of table 7 are relative to the Bloomberg ticker.

Table 7

#	Stock	first observation	last observation	n° of observations	n° of days
1	AAPL	November 14, 2017	April 2, 2018	1.007.866	139
2	INTC	October 16, 2017	April 27, 2018	669.941	193
3	CSCO	October 16, 2017	April 27, 2018	464.697	193
4	MSFT	October 16, 2017	April 26, 2018	865.714	192
5	IBM	September 1, 2017	March 2, 2018	258.878	182
6	AXP	October 10, 2017	April 24, 2018	236.507	196
7	BA	October 10, 2017	April 24, 2018	227.802	196
8	CAT	October 10, 2017	April 24, 2018	301.312	196
9	CVX	October 10, 2017	April 24, 2018	410.468	196
10	DIS	October 11, 2017	April 25, 2018	461.330	196
11	DWDP	October 11, 2017	April 25, 2018	368.150	196
12	HD	October 10, 2017	April 24, 2018	321.447	196
13	JNJ	October 10, 2017	April 24, 2018	404.356	196
14	JPM	October 16, 2017	April 27, 2018	872.608	193
15	KO	October 11, 2017	April 25, 2018	341.347	196
16	MMM	October 10, 2017	April 24, 2018	192.042	196
17	MRK	October 11, 2017	April 25, 2018	432.540	196
18	NKE	October 11, 2017	April 25, 2018	389.636	196
19	PFE	October 11, 2017	April 25, 2018	423.382	196
20	PG	December 11, 2017	April 25, 2018	466.247	135
21	TRV	October 10, 2017	April 24, 2018	132.934	196
22	UNH	October 10, 2017	April 24, 2018	236.605	196
23	UTX	October 10, 2017	April 24, 2018	244.235	196
24	V	October 10, 2017	April 24, 2018	399.722	196
25	VZ	October 11, 2017	April 25, 2018	477.924	196
26	WMT	October 11, 2017	April 25, 2018	473.339	196
27	XOM	October 11, 2017	April 25, 2018	573.346	196
28	GE	November 1, 2016	March 9, 2017	236.628	128
29	GS	September 1, 2017	March 2, 2018	186.187	182
30	MCD	September 1, 2017	March 2, 2018	225.764	182

### 3.4 Methodology and results

In order to evaluate if the frequency of the trades has a significant impact on volatility, we measured the correlation between the waiting times between consecutive trades and the amplitude of the price movements over a certain period of time. Formally, we define with  $t_i$  the moment in time when the  $i^{\text{th}}$  trade takes place, and with  $\tau = t_i - t_{i-1}$  the time interval between two consecutive trades, or waiting time. To perform the study, we built two  $n \times m$  matrixes from the original time series of tick-by-tick transactions, where  $n$  is the number of trading days taken into account, and  $m$  is the number of 5-minutes time intervals inside a trading day. For example 9:00am to 9:04am is the first 5-minutes time interval, 9:05am to 9:09am is the second one, and so on.  $\mathbf{T}$  in (3.1) contains, for each value, the corresponding

average waiting time  $\tau$  between consecutive trades with respect to the  $m^{\text{th}}$  5-minute time interval of the  $n^{\text{th}}$  trading day.  $\Sigma$  in (1) contains the volatilities  $\sigma$  of the  $m^{\text{th}}$  5-minute time interval of the  $n^{\text{th}}$  trading day. Volatility is calculated as the simple difference between the highest minus the lowest price recorded in that certain time interval.

$$\mathbf{T} = \begin{bmatrix} \tau_{1,1} & \cdots & \tau_{1,m} \\ \vdots & \ddots & \vdots \\ \tau_{n,1} & \cdots & \tau_{n,m} \end{bmatrix}; \quad \Sigma = \begin{bmatrix} \sigma_{1,1} & \cdots & \sigma_{1,m} \\ \vdots & \ddots & \vdots \\ \sigma_{n,1} & \cdots & \sigma_{n,m} \end{bmatrix}; \quad (3.1)$$

In this paper we intend to analyze how these two quantities ( $\mathbf{T}$  and  $\Sigma$ ) mutually behave during a trading session, i.e. from the opening in the morning to the closing in the evening. We did not cut the time series deleting any time interval at the beginning or at the end of the trading session as it has been done in Engle (2000), where the first 30 minutes had not been considered in the analysis. We consider all the trades from the opening to the closing of the trading session, for every day. The joint behavior of  $\mathbf{T}$  and  $\Sigma$  has been analyzed by observing how the two quantities evolve during the same  $m^{\text{th}}$  5-minute time interval inside the trading day, across multiple trading days, for each stock, using the simple Pearson correlation coefficient. The result (2) is a vector  $\mathbf{P}$  of correlations  $\rho$  for every stock:

$$\mathbf{P} = \begin{bmatrix} \rho_1 \\ \cdots \\ \rho_m \end{bmatrix} \quad (3.2)$$

Where every element  $\rho_i$  is the correlation between the  $m^{\text{th}}$  column of  $\mathbf{T}$  and the  $m^{\text{th}}$  column of  $\Sigma$  as in (3), and every correlation value refers to a precise 5-minute time interval:

$$\rho_i = \text{correlation}([\tau_{1,i}; \dots; \tau_{n,i}], [\sigma_{1,i}; \dots; \sigma_{n,i}]) \quad (3.3)$$

This approach has been adopted since the aim is to measure how the two quantities behave in the same, precise, and definite moments in time during a trading day, i.e. to measure how an eventual correlation can differ from the morning to the noon, the evening, or other moments inside the trading day. It would have been impossible to extract this kind of information by modelling the two time series of waiting times and volatility in a ‘standard way’, i.e. analyzing their co-movements from the first to the last observation in the orthodox time dimension. Even using a rolling window, considering that the minimum tick data frequency is 1 second, would have implied that the extracted correlations would have referred to, at least, a time interval of few hours inside a trading day. This would have made the results irrelevant for the purpose of this paper, in which the aim is to analyze precise points in time during the day. In this analysis such points in time are defined by 5-minutes time intervals but the same results are obtainable by using any time interval from 1 to 10 minutes. Using longer time intervals compromises the results for the same reason outlined above, i.e. if the time interval is too long the correlations is not a descriptive measure of a precise moment in

time, considering that it is well known in literature (Admati and Pfleiderer, 1988) that both waiting times and volatility present significant intraday patterns, with peaks in trading frequency and volatility in the opening and closing of the market, and more calm moments in the central part of the day. In figure 34 we show the vectors of  $P$  for all the 30 stocks composing the Dow Jones Industrial 30 on the Y axes, while on the X axes we have the 78 5-minutes time intervals composing the DJI30 trading day:

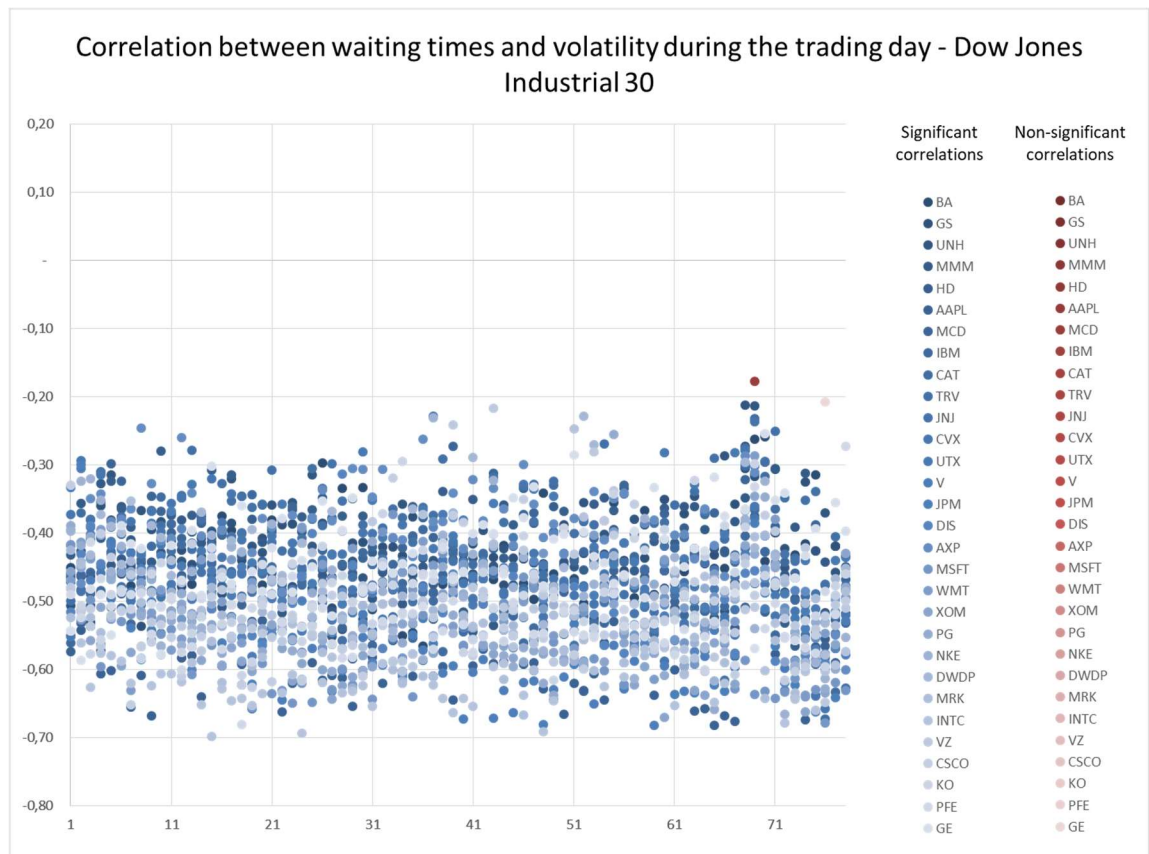


Figure 34

Figure 34 clearly confirms what has been already found in literature (Engle, 2000; Doufour and Engle, 2000), where the volatility is positively correlated with trading frequency. In this case, the negative values in the correlation are due to the fact that they measure the relationship between volatility and waiting times between trades that are the inverse of the trading frequency. The blue points in figure 1 represent the correlations that are considered statistically significant at 95% confidence level. In the entire dataset, only 2 correlations are statistically non-significant, as represented by the red dots. The research question that this paper wants to investigate is if the observed relationship between waiting times and volatility is stable over time and authentic, i.e. if it is not influenced by the deterministic patterns already observed in literature and also found in the dataset used in this paper. In fact, such relationship may be influenced by the fact that both time series, inside a trading day, present a clear U shape, with peaks in the opening and closing of the trading session. We decided to test the authenticity of this relationship by measuring the same correlations after detrending

the two time series with a polynomial of a certain order, in order to eliminate any possible deterministic pattern. The polynomials used range from order 1 to order 4. Two polynomials are calculated for every trading day, one for the waiting times, and one for the volatility. The least-squares methodology is used to fit a polynomial of order  $k$  on every  $n$ th row of  $\mathbf{T}$ , and every  $n$ th row of  $\mathbf{\Sigma}$ , meaning that a dedicated polynomial is used starting with the first transaction of every trading day, and ending with the last transaction of every trading day. The resulting matrixes with fitted values are outlined in (4), where  $fit-k$  indicates the order of the polynomial  $k$  used to obtain the fitted values:

$$\mathbf{T}^{fit-k} = \begin{bmatrix} \tau_{1,1}^{fit-k} & \dots & \tau_{1,m}^{fit-k} \\ \vdots & \ddots & \vdots \\ \tau_{n,1}^{fit-k} & \dots & \tau_{n,m}^{fit-k} \end{bmatrix}; \mathbf{\Sigma}^{fit-k} = \begin{bmatrix} \sigma_{1,1}^{fit-k} & \dots & \sigma_{1,m}^{fit-k} \\ \vdots & \ddots & \vdots \\ \sigma_{n,1}^{fit-k} & \dots & \sigma_{n,m}^{fit-k} \end{bmatrix}; \quad (3.4)$$

The detrended values for waiting times and volatility are then obtained through the Hadamard division, i.e. dividing value by value the original  $\mathbf{T}$  and  $\mathbf{\Sigma}$  matrixes by their fitted counterparts  $\mathbf{T}^{fit-k}$  and  $\mathbf{\Sigma}^{fit-k}$  as respectively in (5) and (6), where  $det-k$  indicates the order of the polynomial  $k$  used to obtain the detrended values:

$$\mathbf{T}_{n,m}^{det-k} = \mathbf{T}_{n,m} / \mathbf{T}_{n,m}^{fit-k} \quad (3.5)$$

$$\mathbf{\Sigma}_{n,m}^{det-k} = \mathbf{\Sigma}_{n,m} / \mathbf{\Sigma}_{n,m}^{fit-k} \quad (3.6)$$

In the detrending process of both waiting times and volatility, due to the parabolic shape of both time series at intraday level, it happens to encounter negative fitted values at the beginning and at the end of some trading days. The resulting detrended time series will then present negative values in the correspondence of the negative fits, since the detrending is obtained with a normalization (division) of the original value with respect to the fitted value as in (5) and (6). While this does not make any significant difference from a statistical point of view, the concept of a negative waiting time or negative volatility does not hold in the real world. For this reason, we adopted two different procedures to ‘correct’ the negative values on both fitted time series to check if that would have any impact on the final results.

The first procedure consists in the simple shifting of the polynomial by a certain value  $S$ . Whenever we encounter a polynomial with one or more negative values, its lowest value is multiplied by a negative constant  $-C$  and the resulting quantity it’s added to every value of the fitted polynomial. In this way, for any absolute value of  $C > 1$ , the resulting shifted polynomial will have strictly positive values. We applied this procedure with values of  $C = 1.1, 2$  and  $4$ , and we did not observe any significant difference in the results. This procedure is separately applied polynomial by polynomial (meaning day by day), only in presence of at least one negative fit. The second procedure operates only on the eventual negative fits, value

by value, without making any changes in the remaining values of the polynomial which presents the negative fits. If we refer to (4), column by column of  $T^{fit-k}$  and  $\Sigma^{fit-k}$ , any negative column value is eliminated, and the column average is used to replace the previously eliminated column values. In such a way, for any negative values encountered, we are replacing it with the average waiting times, or volatility, of that  $m^{th}$  5-minutes time interval. As far as we observed, there is no significant difference in the results switching from the first to the second procedure, as the negative fitted values are extremely limited. The results presented in this paper refer to the second procedure.

As previously said, both waiting times and volatility present clear patterns at intraday level. As a confirmation, both time series are strongly non-stationary and autocorrelated. Figure 35 below shows the results of the KPSS stationarity test at 95% confidence level. The test has been performed on every DJI30 stock, after separating their time series into single days. The Y axes shows the percentage of trading days that are non-stationary for both waiting times and volatility, with respect to their standard ( $T$  and  $\Sigma$ ) and detrended values ( $T_{n,m}^{det-k}$  and  $\Sigma_{n,m}^{det}$ ).

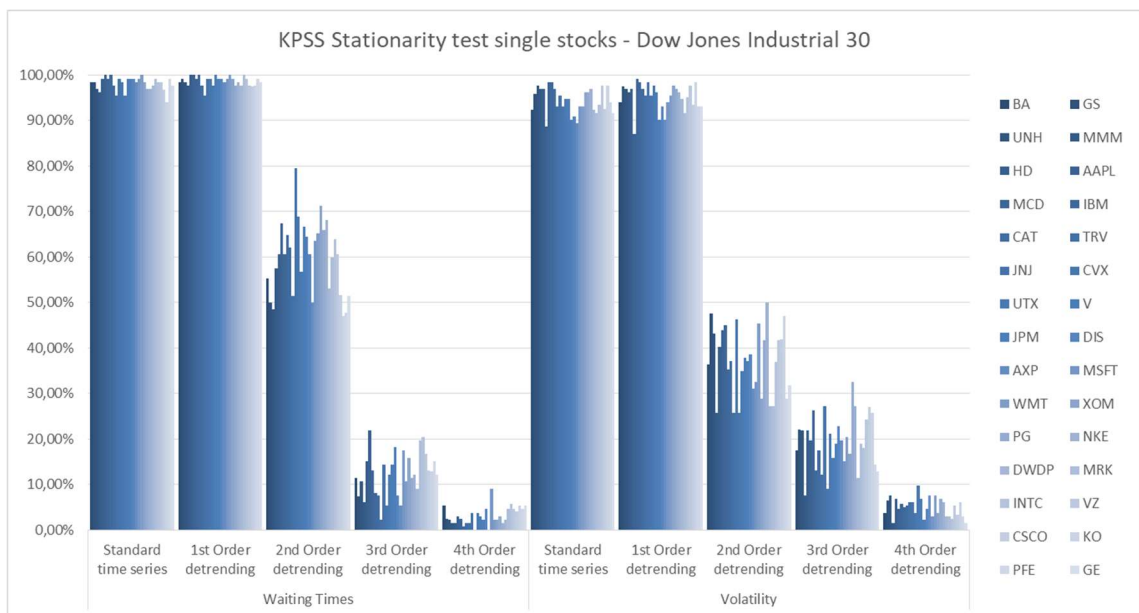


Figure 35

The results in figure 35 suggests what can be observed also with normal eyeball by looking at the intraday patterns in waiting times and volatility, i.e. since both time series have a parabolic shape, using a linear detrending does not produce any significant improvement. In fact, the first order  $k$  of detrending does not improve at all the KPSS stationarity test, leaving the results almost unchanged. The biggest improvements, in relative terms, happens when we pass from the 2<sup>nd</sup> to the 3<sup>rd</sup> order in the waiting times, and from the 1<sup>st</sup> to the 2<sup>nd</sup> order in the volatility. The same result is confirmed if we look at the autocorrelation of the same time series in figure 3 and 4 where, respectively for waiting times and volatility, the reduction to

almost noise level happens at the 3<sup>rd</sup> and 2<sup>nd</sup> detrending order. In the two figures the red lines indicate the 95% confidence level of the autocorrelation function while the blue (figure 36) and purple (figure 37) lines indicate the autocorrelation function for, respectively, waiting times and volatility. Both figures comprehend all the 30 stocks of the DJI30 that, as it can be clearly seen, present very similar autocorrelation functions, since they almost completely overlap. The wave form (top-bottom-top) that is observable in the figures has a duration of one day and it's expressive of the intraday pattern of both waiting times and volatility. Such pattern is clearly eliminated as we proceed with higher detrending orders.

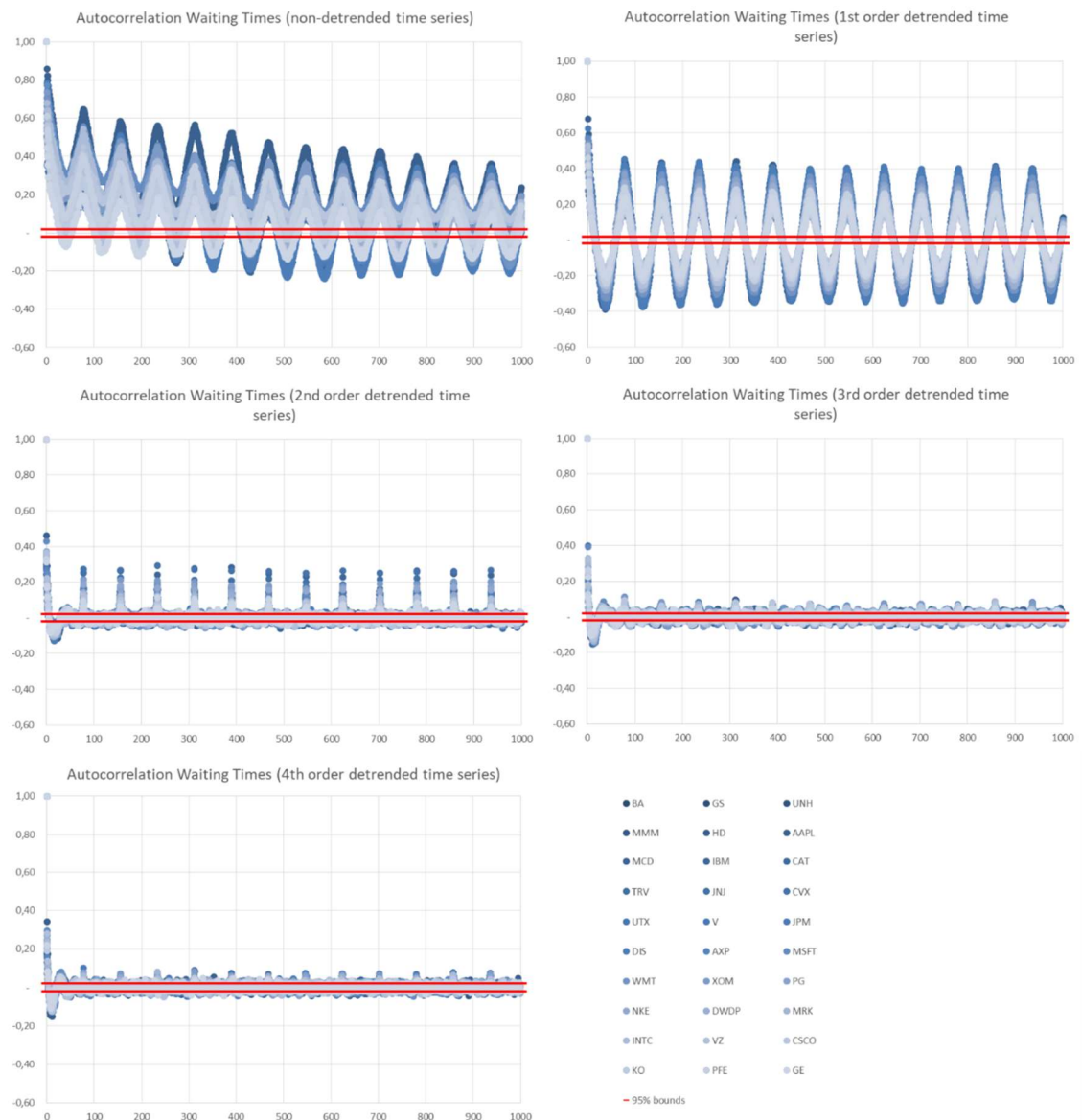


Figure 36

Deciding which detrending order is optimal to test the correlation between  $T_{n,m}^{det}$  and  $\Sigma_{n,m}^{det-k}$ , and to analyze if the joint behavior of  $T$  and  $\Sigma$  is influenced by any deterministic pattern, is not trivial. As far as the detrending order increases in fact, there is a higher probability of observing artificially inducted properties in the data. This happens because



with high detrending orders we are basically overweighting the small fluctuations that may be not descriptive of the underlying process. In this context it is then crucial to choose a certain detrending order that minimizes such risks while reducing as much as possible any deterministic patterns in the time series. Considering the results of the KPSS stationarity tests, and the autocorrelation functions, we decided to opt for the 3<sup>rd</sup> order detrending for the waiting times and the 2<sup>nd</sup> order for the volatility.

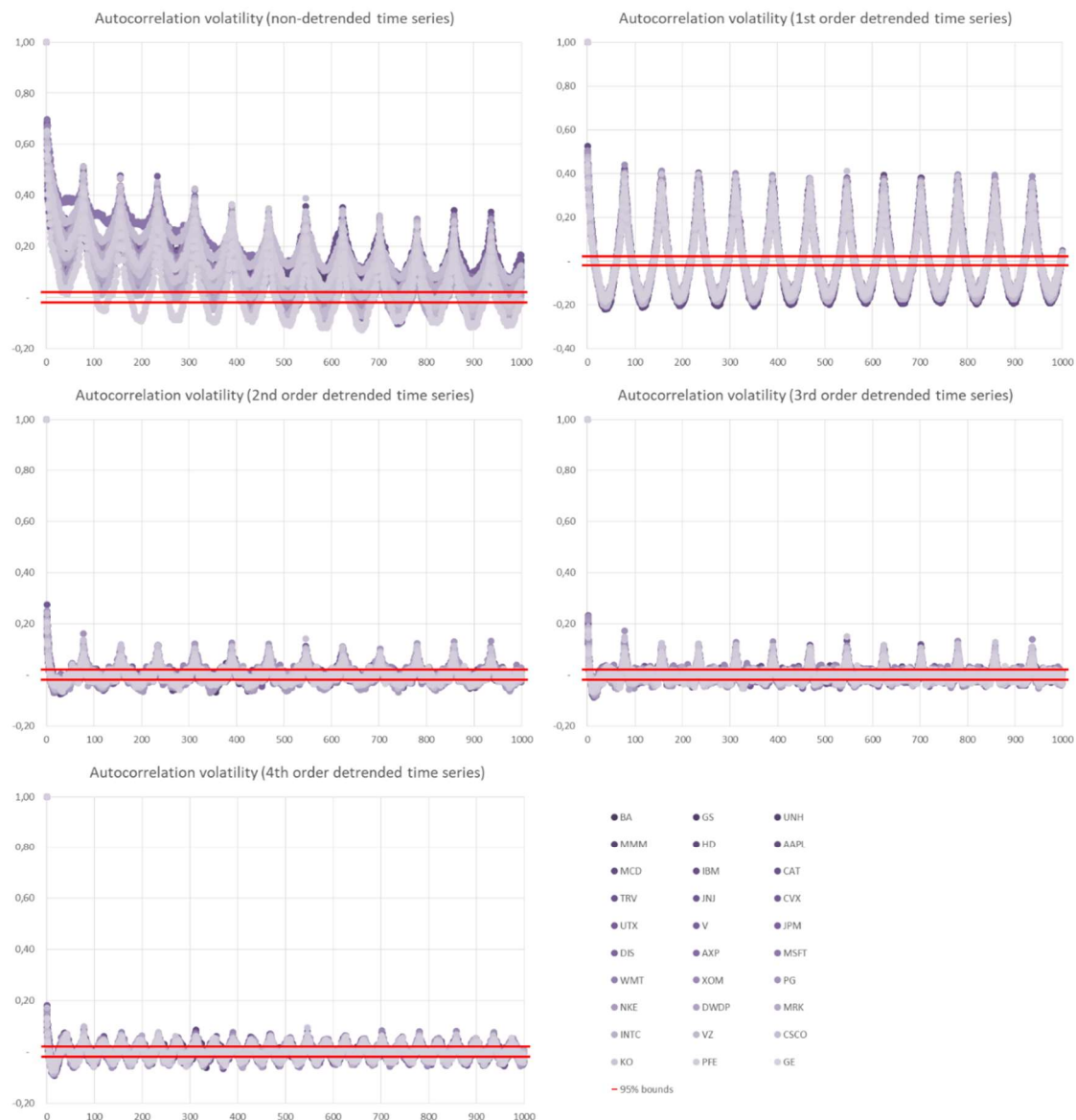


Figure 37

However, as a confirmation to the robustness of the results, the same evidence holds at any combination of 2<sup>nd</sup>, 3<sup>rd</sup>, and 4<sup>th</sup> order detrending, even if in the latter case all the correlation structure starts to be significantly compromised. Figure 38 below shows the same result



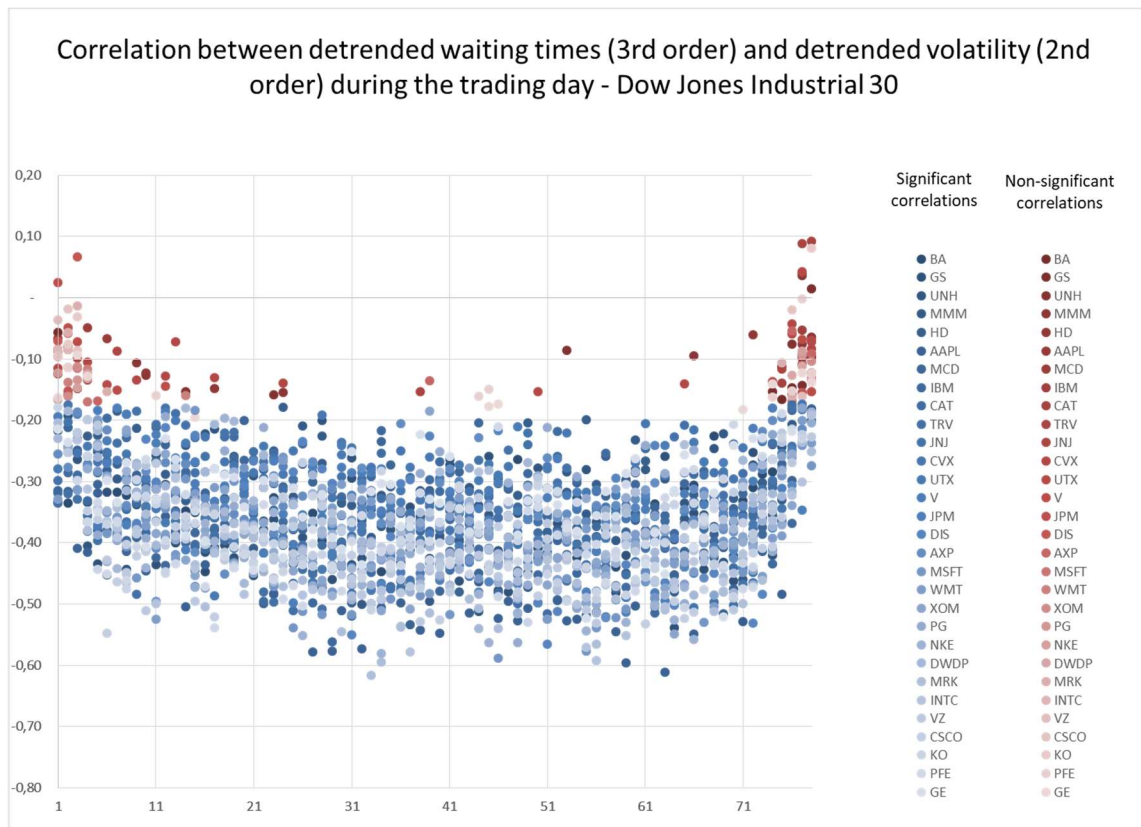


Figure 38

shown in figure 34 but in this case the correlation is applied to  $T_{n,m}^{det-3}$  and  $\Sigma_{n,m}^{det-2}$  rather than to  $T$  and  $\Sigma$ :

Figure 38 clearly shows that, once we try to eliminate any form of deterministic patterns, non-stationarity and autocorrelation in the time series of waiting times and volatility, a very different picture emerges with respect to figure 1, and also with respect with what has been previously found in literature. The correlation between trading frequency and volatility seems to hold, even if with a lower intensity, in the central part of the trading day while it completely disappears for almost every stock in the first and last 20 minutes of the trading day.

### 3.5 Conclusions

While the beginning and the end of the trading day present both a higher frequency in trading, and a higher volatility with respect to the other parts of the day, our results suggest that these joint behaviors do not have any statistically significant relationship and are independent. On the other hand this relationship becomes statistically significant after (before) the first (last) ‘trading rush’ of the trading day, i.e. the first and last 20 minutes of trading. According to Admati and Pfleiderer (1988), the moments after the opening and before the closing of the market are characterized by a high concentration of both informed and uninformed traders. Engle (2000) writes also, with respect to Admati and Pfleiderer (1988) paper, that in the central part of the day there is a higher proportion of informed

traders. If we assume that these statements hold, we can assume that the correlation between the trading activity and the volatility is, to some extent, the proof of the direct impact of the information carried by the transactions on the price formation process. Since information, of any sort, is what influences the price formation process, we can also assume that, during the trading day, we have different degree of impact of the trading activity over the price formation process itself. These different degrees can be influenced by the different types of traders operating in the market. In the central part of the day, where the majority of the operating traders are informed ones, we have a clear flow of information from the traders to the market through their activity. This phenomenon has a direct impact on the price formation process in terms of a significant relationship between trading frequency and volatility (or amplitude of price fluctuations). With this statement we are not saying that the central part of the day is characterized by a certain level of volatility or trading activity. We are saying that the two quantities analyzed are significantly correlated and this happens since the majority of the trades in place are carrying information that impacts the price formation process. On the other side, the first and last 20 minutes of the trading day are characterized by a higher trading activity. The absolute number of both informed and uninformed traders is higher. The difference relies in the proportion of the two categories since, with respect to the central part of the day, the opening and the closing of the market are characterized by a higher percentage of liquidity (uninformed) traders. In this context, in order to give an explanation to our findings, we can hypothesize that the trading activity does not influence the price formation process in these precise moments, and that volatility cannot be explained through trading frequency. Since we consider that information is what causes an impact in the price formation process, uninformed traders, de facto, do not influence the price formation process. On the other side, informed traders, when trading in these particular moments of the day, mainly trade to take advantage of uninformed traders without any desire of disclosing their competitive advantage, resulting in a compensative effect which destroys any trace of information carried by their trades. As the activity of uninformed trading decreases, both in absolute but mainly in percentage terms, the activity of informed traders becomes predominant and their ability to hide their competitive advantage decreases. In this scenario, every informed trade starts having a direct impact on the price formation process, measured by a statistically significant relationship between trading activity and volatility. Such relationship holds as long as the proportion of informed traders remains higher, and it start to disappear again when, at the end of the trading day, liquidity traders start to trade again before the market closes.

## General Conclusions

With this Thesis we tried to approach financial market research with a new, ontological point of view, trying to understand the nature of the price formation process, but also trying to reduce as much as possible the influence from existing research. We treated the price as a pure geometrical object in a Cartesian space where we put the quantity price  $p$  on the Y axes and the its relative moment in time  $t$  on the X axes, without making any assumption about agents' behavior, rationality or the concept of market efficiency. Formally, we analyzed times series of historical prices for all asset classes and for various frequencies, from tick by tick to monthly data, to investigate the presence of deterministic or stochastic non-random patterns. The thesis is divided into three parts: In the first two we adopted a bottom up approach while in the last we reverted to a top down.

In the first section we developed a pattern recognition methodology to extract the most elementary pattern from prices: market trends. The results show that the price formation process is far from being a memory less process. In fact, non-random and negatively autocorrelated consecutive trendlines are observable in the same time series that present random and non-autocorrelated returns. This fact, de facto, separates the concept of efficiency from the concept of randomness since random returns coexist with non-random behavior. This point proves the main contribution that this paper intends to provide, i.e. phenomenon (or its formalization, a time series) cannot be analyzed homogenously. By treating the price formation process as a geometrical object we want to underline that the imposition of a deterministic structure to the time dimension of the process may bias the results we observe. The proof is given by the fact that using artificially spaced returns compromises the properties of the memory contained in the price formation process.

In the second section we developed an asset allocation model to verify if the information contained in the memory of market trends could be exploited to gain stable extra returns. The results show that the methodology allows to obtain a significant over performance both if it is applied to a single asset but mainly when the strategy is implemented on a portfolio of securities. The benefits appear to be applicable to all asset classes, at any frequency, and they are not only related to e measure of excess returns. As a confirmation of the fact that the information extracted can be called structural, we are able to observe a stable excess returns altogether with a significant reduction of both volatility and drawdowns. The results have been obtained with a simulation and so, as almost always in literature, there is no implementation of the strategy on a real trading platform. However, considering the strength of our results, we feel confident to state that the price formation process contains a significant amount of non-exploited information in terms of hidden memory, and that this information is structural.

In the third and last part we adopted a reverse approach to our analysis, i.e. top down. We started from the presence of well-known intraday patterns in trading frequency and volatility in order to understand their effect on the price formation process. In literature the positive correlation between trading frequency and volatility has been widely shown but, considering that both time series are highly non-stationary and they contain strong deterministic patterns, we tried to understand what would be the effect if we tried to get rid of them. We adopted a simple detrending procedure with the use of a polynomial at different orders. This methodology appeared to be extremely effective in eliminating both the intraday patterns and also the non-stationarity they generate. Once the same correlation is calculated over the detrended time series, the relationship completely disappears in the first and last 20 minutes of each trading day, becoming non-significant, while being reduced during the central part of the day. In our opinion this is due to the different type of trading that takes place during the trading day. While in the morning and evenings most of the trading is non-informed, in the central part of the day there is a higher concentration of informed traders, which translates in a direct relationship between the frequency of the trades and the price formation process, in terms of volatility impact.

The aim of this Thesis is to propose a new approach to look at financial market research. In our opinion, sometimes it is useful to clear the dashboard and start from scratch, getting rid of preconceptions that may bias our opinions and results. In our approach we did not consider the main pillars of financial research, such as the concepts of efficiency, rationality and well-established relationship in market microstructure. The results we obtained with this approach are significant. We were able to demonstrate the presence of a structural non-random behavior in financial market, at any frequency and for every asset class. We were also able to demonstrate how this information can be used to obtain stable extra profits that are widely higher than anything previously documented in literature. We also demonstrated how the patterns we observed in this paper, if not taken into consideration by orthodox research, can lead to a misinterpretation of well-documented relationships. Methodologies and criteria used in this Thesis are far from being perfect or well-tested. However, our wish is to stimulate further research with an eye opened to what sometimes is considered non-orthodox but, as science always shows, is the non-orthodox that leads to the most interesting discoveries.

## References

Admati, Anat R., and Paul Pfleiderer, 1988, A theory of intraday trading patterns: Volume and price variability, *Review of Financial Studies* 1, 3-40.

Andersen, Torben G. (1996) Return volatility and trading volume: An information flow interpretation of stochastic volatility, *The Journal of Finance* 51, 169–204.

Arditti F. (1967), Risk and the Required Return on equity, in *Journal of Finance*, March.

Arrow K.J. and Debreu G. (1954), Existence of An Equilibrium For Competitive Economy, In *Econometrica*, Vol. 22

Bachelier (1900), *Théorie de la speculation*;

Barber B. M., Odean, T. And Zhu, Ning, Do Noise Traders Move Markets? (September 2006). Efa 2006 Zurich Meetings Paper. Available At Ssrn: <https://ssrn.com/abstract=869827>

Barber, Brad M. And Odean, Terrance, The Behavior Of Individual Investors (September 7, 2011). Available Atssrn: <https://ssrn.com/abstract=1872211>

Barber, B. M. And Odean, T., The Courage of Misguided Convictions: The Trading Behavior Of Individual Investors. Available At Ssrn: <https://ssrn.com/abstract=219175>

Bekaert, G. and R.J. Hodrick, 1992, Characterizing predictable components in excess returns on equity and foreign exchange markets, *Journal of Finance* 47.

Benartzi S. E Thaler R. (1995), Myopic Loss Aversion and The Equity Premium Puzzle, In *Quarterly Journal Of Economics*, Vol 110;

Black F. (1986), Noise, In *Journal of Finance*, Vol.41;

Blume M.E., And Friend I. (1975), The Asset Structure Of Individual Portfolios And Some Implications For Utility Functions, In *Journal Of Finance*, Vol. 30, Pp. 585-603;

Bollerslev T Ole Mikkelsen H, Modeling and pricing long memory in stock market volatility (1996), *Journal of Econometrics* vol: 73 (1) pp: 151-184

Bradley, D., Jordan, B., And Ritter, J. (2003) “The Quiet Period Goes Out With A Bang”, *Journal Of Finance*, 58, Pp. 1-36. <sup>[1]</sup><sub>[SEP]</sub>

- Campbell J.Y. E Kyle A. (1993), Smart Money, Noise Trading And Stock Price Behavior, In Review Of Economic Studies, Vol. 60;
- Clark, Peter K. (1973) A subordinated stochastic process model with finite variance for speculative prices, *Econometrica* 41, 135–155.
- Cowles A. (1937);
- De Bondt W. F.M. E Thaler R. (1985), Does The Market Overreact ?, In *Journal Of Finance*, Vol. 40;
- De Bondt W.F.M. (1993), Betting On Trends: Intuitive Forecasts Of Financial Risk And Return, In *International Journal Of Forecasting*, Vol. 9, Pp. 355-371;)
- De Bondt W.F.M. (1998), A Portrait Of The Individual Investor, In *American Economic Review*, Vol. 85, Pp. 181-185;
- De Bondt W.F.M. E Thaler R. (1987), Further Evidence On Investor Overreaction And Stock Market Seasonality, In *Journal Of Finance*, Vol. 42;
- De Long B., Shleifer A., Summers L. And Waldman R.(1990), Noise Trader Risk In Financial Markets, In *Journal Of Political Economy*, Vol. 98;
- Doufour and Engle (2000) Time and the Price Impact of a Trade, *The Journal of Finance*, vol 55 issue 6
- Engle (2000) The Econometrics of Ultra-high-frequency Data, *Econometrica* vol 68 issue 1, 1-22.
- Epps, Thomas W. and Epps, Mary Lee. (1976) The stochastic dependence of security price changes and transaction volumes: Implications for the mixture-of-distributions hypothesis, *Econometrica* 44, 305–21.
- Fama, E F. 1965a.; “The Behavior of Stock Market Prices.” *Journal of Business* 38:34–105
- Fama, E F 1965b; “Random Walks in Stock Market Prices.” *Financial Analysts Journal* September/October: 55–59
- Fama, E.F., Fisher I., Jensen M.C, Roll R. 1969. “The Augment of Stock Prices to New Information.” *International Economic Review* 10: 1–21
- Fama, E. F. 1970. “Efficient Capital Markets: A Review of Theory and Empirical Work.” *Journal of Finance* 25:383–17

- Fama, E F 1975; “Short-Term Interest Rates as Predictors of Inflation.” *American Economic Review* 65:269–82
- Fama, E F 1976b; “Inflation Uncertainty and Expected Returns on Treasury Bills.” *Journal of Political Economy* 84: 427–48.
- Fama (1981), “Stock Returns, Real Activity, Inflation, and Money.” *American Economic Review* 71: 545–65
- Fama, E F 1990b; “Stock Returns, Expected Returns, and Real Activity.” *Journal of Finance* 45:1089–1108.
- Fama, E F 1991; “Efficient Markets II.” *Journal of Finance* 46:1575–1617
- Fama, E F 1998; “Market Efficiency, Long-Term Returns, and Behavioral Finance.” *Journal of Financial Economics* 49: 283–06
- Fama, E.F., And French K.R. 2008; “Dissecting Anomalies.” *Journal of Finance* 63:1653–1678
- Fama, E.F., And French K.R. 2010, “Luck Versus Skill in the Cross-Section of Mutual Fund Returns.” *Journal of Finance* 65: 1915–1947
- Fama E. F., Two Pillars Of Asset Pricing, Nobel Prize Lecture, Dec. 8, 2013, [Nobelfundation.Org](http://Nobelfundation.Org), P.365
- Gallant, A. Ronald, Rossi, Peter E. and Tauchen, George. (1992) Stock prices and volume, *Review of Financial Studies* 5, 199–242.
- Godfrey, M.D., Granger, C.W.J., Morgenstern, O., (1964). The random-walk hypothesis of stock market behavior. *Kyklos*, 17 (1), 1–30.
- Granger C.W., Morgenstern O. (1970), *Predictability Stock Market Prices*, D.C. Heath & Co. Lexington, MA
- Grossman S. And Stiglitz J. (1980), On The Impossibility Of Informationally Efficient Markets, In *American Economic Review*, Vol. 70, Pp. 393-408;
- Guajarati D.N. (2003), *Basic Econometrics*, Mcgraw-Hill, Fourth Edition;
- Haltiwanger J. And Waldman M. (1985), Rational Expectations And The Limits Of Rationality: An Analysis Of Heterogeneity, In *American Economic Review*, Vol. 75, June;

Holbrook W. (1934);

Jegadeesh N. and Titman S., 2001, Profitability of Momentum Strategies: An Evaluation of Alternative Explanations *The Journal of Finance* Vol. LVI, N. 2 (2001).

Jones, Charles M., Kaul, Gautam and Lipson, Marc L. (1994) Transactions, volume and volatility, *The Review of Financial Studies* 7, 631–651.

Karpoff, Jonathan M. (1987) The relation between price changes and trading volume: A survey, *Journal of Financial and Quantitative Analysis* 22, 109–126.

Lakonishok J., Shleifer A. E Vishny R. (1994), Contrarian Investment, Extrapolation And Risk, In *Journal Of Finance*, Vol. 49;

Lintner J. (1965), The Valuation Of Risk Assets And The Selection Of Risk Investments In Stock Portfolios And Capital Budgets, In *The Review Of Economics And Statistics*, Febbraio;

Lintner J. (1971), The Effects Of Short Selling And Margin Requirements In Perfect Capital Markets, In *Journal Of Financial And Quantitative Analysis*, Dicembre;

Lo And Hasanhodzic (2010), The evolution of technical analysis;

Mandelbrot (1983), The fractal geometry of nature;

Mandelbrot and Hudson (2004), The misbehavior of markets – A Fractal view of risk ruin and reward;

Mehra R., Prescott E. (1985), The Equity Premium: A Puzzle, In *Journal Of Monetary Economics*, Vol. 15;

O'Hara, Maureen, 1995, *Market Microstructure Theory* (Blackwell Publishers, Malden, Mass.).

Peters (1994), *Fractal Market Analysis - Applying Chaos Theory to Investment and Economics*.

Raberto, Marco; Scalas, Enrico; Mainardi, Francesco Waiting-times and returns in high-frequency financial data: An empirical study (2002), *Physica A: Statistical Mechanics and its Applications*

Roll R. (1989), R2, In *Journal Of Finance*, Vol. 43;



- Ross, Stephen A. (1989) Information and volatility: The no-arbitrage martingale approach to timing and resolution irrelevancy, *Journal of Finance* 44, 1–17.
- Russell T. And Thaler R. (1985), The Relevance of Quasi Rationality in Competitive Markets, In *American Economic Review*, Vol. 75, December;
- Samuelson, P. A. (1950), The Problem of Integrability in Utility Theory, In *Economica*, New Series, 17(68), 355–385;
- Samuelson P. A. (1965), Proof That Properly Anticipate Prices Fluctuate Randomly, In *Industrial Management Review*, Vol. 6;
- Scholes M. (1972), The Market For Securities: Substitution Versus Price Pressure And Information On Share Prices, In *Journal Of Business*, Vol. 45
- Shefrin H. And Statman M. (1985), The Disposition to Sell Winners Too Early And Ride Losers Too Long: Theory And Evidence, In *Journal Of Finance*, Vol. 40, Pp. 777-792;
- Shefrin H. M. E Thaler R.(1988), The Behavioral Life-Cycle Hypothesis, In *Economic Inquiry*, Vol. 26, P. 609-643;
- Shefrin H., Statman M. (1997), Comparing Expectations About Stock Returns To Realize Returns, In *Working Papers*, Leavey School Of Business, Santa Clara University, February
- Shefrin H. (2000), *Beyond Greed And Fear: Understanding Behavioral Finance And The Psychology Of Investing*, Harvard Business School Press, Boston, Usa;
- Shiller (2015);
- Shiller R. (1981), Do Stock Prices Move Too Much To Be Justified By Subsequent Changes In Dividends?, In *American Economic Review*, Vol. 71;
- Shiller R. (1984), Stock Prices and Social Dynamics, *Brooking Papers On Economic Activity*, Vol. 2;
- Solnik B., 1993, The performance of international asset allocation strategies using conditioning information, *Journal of Empirical Finance* 1 (1993) 33-55.
- Tauchen, George E., and Pitts, Mark. (1983) The price variability-volume relationship on speculative markets, *Econometrica* 51, 485–505.