Spectroscopy of Fat Tails

Complex Correlations in Financial Time Series

Valentina Alfi

Ph.D. Thesis in Physics

University of Rome
"Roma Tre"

Advisors: Prof. L. Pietronero and Prof. M. Rovere

Contents

Introduction

Since a long time concepts and methods of Statistical Physics are used within the framework of Economics and Finance. Historically, the most famous example, in which, however, the inverse process had happened, is that of Random Walk. In fact, the RW was originally introduced at the beginning of 1900 by L. Bachelier [14] as a model for describing the market price dynamics. Only a few years later, with the famous works of E. Einstein and J. Perrin, the RW model was formalized mathematically and was applied in Physics as the representation of the Brownian motion. Another renowned person who was interested in Physics applied to socio-economic sciences was E. Majorana. In his last paper "Il Valore delle Leggi Statistiche nella Fisica e nelle Scienze Sociali", Majorana discussed about the possible applications of Physics to social sciences and he described in a premonitory way some of the developments which then have actually happened.

Going on in more recent times, we can cite the Black and Scholes model that was drawn up in 1973 [22] (thanks to this model Scholes was also winner of the Nobel prize in Economics in 1997). This model is based on the assumption that the price fluctuations follow a geometrical Brownian motion and is used to estimate the price of options and other financial derivate products. The idea is that, analyzing the price fluctuations for short period, one can predict the price fluctuations in a much longer period, with the assumptions of considering a stationary process and the absence of temporal correlations. The Black and Scholes model is a beautiful theoretical work but, because of its assumptions, it does not correspond to an exact description of real markets. In fact, even though the Black and Scholes model is commonly used in market analysis, it is often complemented with additional modifications inspired by empirical observations. Furthermore, the fundamental assumptions of considering the price dynamics described by a geometrical Brownian motion and the absence of temporal correlations, are not satisfied in real markets. These observations pose some conceptual problems that have important practical implications because the risk estimated from the Black and Scholes model, results much smaller than the real one. With these considerations one can realize that the socio-economic dynamics are much more complex than the simple Random Walk and the study of such complications is extremely important and actual.

Physics and Economics are also very different in many aspects. In Physics it is assumed that some Law of Nature do exist and one tries to discover them. In Economics the question if such laws with general validity exist or not is an open question and it is still object of debate. Even it is not clear if Economics can be considered a Science in the Popper's view [106], that is where an experiment can falsify a theory. Also the reproducibility of experiments, that is one of the pillars of Physics, is indeed not an evident fact in Economics. This last aspect is due to many factors. The first is that one can never deal with identical situations and the dynamics in not stationary. In fact, the laws of the dynamics are not necessarily the same at different times. The second aspect is that Economics depends on human behavior and, in general, on the whole society. This situation implies elements of adaptive and evolutive nature which, in general, are not present elementary physical phenomena and which indeed recall biological themes.

Physics of Complex Systems represents a step towards the quantitative study of these interdisciplinary situations. In this perspective, during the last years, there has been a great interest to analyze, and when possible explain, the socio-economic behaviors in the perspective of Complex Systems. Let us discuss some some fundamental aspects that cannot be disregarded when studying economic systems.

1 An important characteristic of Economics is that the system self organizes themselves in a way that excludes the possibility of exploiting deterministic previsions on the price behavior. This is the Arbitrage principle that justify the use of Random Walk as the basic model for price dynamics.

2 Many of the probability density functions of economic quantity disclose a power law shape unlike the usual Gaussian probability density function of the elementary stochastic processes. The presence of power laws with anomalous exponents recalls the properties of fractal structures and of self-organized critical phenomena [84, 85].

3 The main deviation from the Random Walk model is observed in price fluctuations. These fluctuations are much more larger than whose expected for a Gaussian dynamics or, in general terms, for a stationary dynamics which satisfy the Central Limit Theorem. The nature itself of these fluctuations is anomalous in the sense that their distribution

is characterized by a power law. This is the famous problem of fat tails [84, 88, 30].

4 The Arbitrage principle avoids the presence of elementary time correlations for the sign (direction) of price fluctuations. Nevertheless, if one considers the absolute values of price fluctuations, strong time correlations are present. Also these correlations are well described by power law functions. This features is called volatility clustering [88, 30].

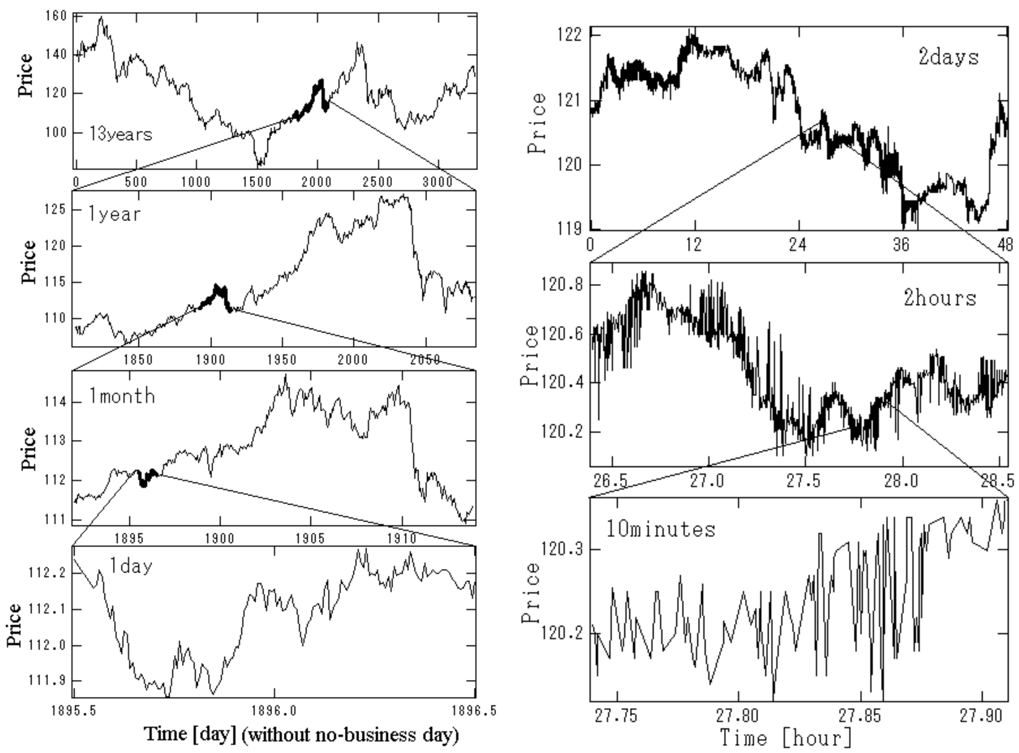
These features, called *stylized facts*, are considered as peculiar and unavoidable to have a bit realistic description of economic phenomena.

Another field where the methods of Complex Systems have been used to describe economic dynamics is the research line on Agents Based Models [13]. These models consider an heterogeneous structure with many different agents and investigate the development of large fluctuations and of evolutive behaviors. Models like these deviate in a qualitative way from the simple description of phenomena in terms of differential equations for variables that correspond to mean values. This situation is analogous to that of Critical Phenomena in Physics where one moves from a Mean Field description to Monte Carlo and Renormalization Group approaches. At the moment, the comparison between these models and the experimental data is however complicated because the stylized facts are too few and as a consequence many different models can reproduce the four features described above.

Hence, an essential point for the progress in this field is to identify further general properties and to formulate more specific models which will permit a critical comparison. In this thesis we intend to move toward both these directions.

In Fig. 1 we show an excellent example of self-similarity for the dynamics of Yen/USD rates on a wide interval of time scales. A way to characterize this property is through the scaling exponent of the so called *roughness* [73, 113]. What is commonly done to study the roughness of a given variable, is to consider its fluctuations as a function of the length of the considered time interval.

Roughness is usually considered as an important element to identify long range correlations in the data. In this thesis we are going to challenge this point of view by showing that a number of other effects, different from long range correlations, can lead to non trivial roughness exponents. An important result of our work will then be the introduction of new concepts and



(a) Scale Invariance

(b) Breakdown

Figure 1: This is an overview of a 13-years period for the time evolutions of Yen/USD rates. The fractal property of price can be confirmed for large time scale. This is a general property of open markets for time scale larger than a few hours. Scale invariance breaks down at short time ($<$ 1 hour). At tick level there is more zig-zag and signal looks discrete.

methods which permit to distinguish the origin of non trivial roughness exponent in terms of correlations or other effects. An important example of this problematic is given by the role of finite size effects in the data and their implications in terms of effective roughness exponent. The problem of finite size effects is actually very relevant in many situations. For many reasons, one may be interested in estimating the roughness properties of a given variable from a relatively limited set of data. For real time analysis, the smaller is the dataset necessary for a reasonable signal, the more reactive and useful is the result. In some cases this necessity is due to the nature itself of the data. For example, this is the case if we consider high frequency, tick by tick data from the market of New York (NYSE). During the night this market is close and the price undergoes a large discontinuity in its value. This implies that our data are systematically homogeneous only if we consider the transaction prices within a single trading day.

Therefore, in this thesis we have first considered the importance of finite size effects in the estimate of the roughness. This analysis has been performed both with analytical results on a finite size Random Walk and with suitable simulations [4, 6]. The results show that, even for the simple case of a Random Walk, the finite size effects in estimating the roughness exponent are intrinsic and unavoidable. In particular, we identify a systematic enhancement of the effective roughness exponent, even for the simple case of a Random Walk. In addition to the Random Walk behavior, we have also studied more complex cases in which the dynamics presents also fat tails and correlations. Simulations of these modified Random Walk models have led to a further bias in the estimation of the roughness exponent. Further more, for non stationary processes, our results can also be relevant for infinite time series. Given this situation, we have introduced a new tool to estimate roughness by means the analysis of the fluctuations of a given variable from a suitable moving average [9]. This new method appears much more useful than the standard one in order to characterize the fluctuations behavior of a given process.

The roughness exponent gives an absolute measure of fluctuations. In some sense it is similar to volatility, but it adds to it the properties of scaling as a function of the time interval considered. In addition to this information on the absolute value of the fluctuations, one may also consider directional ones. These bring us naturally to the realm of trading strategies. For example, if the price shows a marked positive deviation from some type of reference value, this can lead to different reactions of the traders which depend on their strategic attitude. The "trend following" strategy would imply that this deviation is the beginning of a strong upward trend. On the other hand, a "trend adverse" strategy would instead expects that the fluctuations will be soon absorbed by an opposite tendency.

The question we intend to consider is whether from a time series data set one can identify the nature of the prevailing strategy. We would like to introduce a new type of statistical analysis which is focused on the identification of these "hidden forces" of effective attraction or repulsion with respect to a certain reference value.

Together with a group of Japanese researchers (H. and M. Takayasu, Sony Research and Tokyo Univ.) we have considered the possibility that the price dynamics can be affected by its distance from a given reference value [114, 5]. For example we have chosen as reference value a moving average performed on the values of the price in the previous time steps.

Starting from this idea, we introduce a model for the price dynamics based on a Random Walk with the addition of force terms which depends on the history of the walker in terms of its previous time steps. The aim is to identify the tendency of the price dynamics to be attractive or repulsive with respect its moving average. For example, if the force is linear means that the price is attracted or repelled by its moving average proportionally to its distance from it. The idea is that different kind of forces (attractive or repulsive) can reflect various kind of market agents (trend followers or trend adverse). An analysis on real data has led to the identification and the visualization of such "hidden" forces, indicating that this new method is indeed suitable to detect complex correlations in real stock price fluctuations [7]. We retain that the development of this original method could add a new element to the stylized facts, connected with the kind of strategy that prevails between the agents in a certain time interval. This element could have important implications both from the conceptual and practical point of view.

In addiction to the analysis of real data, the above concepts can suggest a new framework for the modeling of the strategies in an Agent Based Models. In these models the rules of trading are defined at a microscopic level and then one observes the collective macroscopic behavior. Starting from the previous concepts of a mathematical interpretation of strategies, one can define a new kind of agent based models in which the strategies are based on the fluctuations moving averages, as it usually happens in real behavior of traders. The various agents can consider a variety of moving averages calculated at different time scales and, as well, can be trend adverse trend follower. What one expects to find is that in a phase of prevalence of trend adverse agents the market will be stable and the price dynamics is a Random Walk with an attractive force. Instead, if the trend followers dominate one expects an instable market with the formation of speculative bubbles and trends, The real dynamics should correspond to a complex equilibrium between this two tendencies which can be described in a quantitative way. At the moment the model is in a preliminary stage, [8] and it will to be conveniently developed in the near future.

Now we will describe in detail as this thesis is organized.

First Chapter In this first chapter we discuss the possible applications of Physics in Economics and Finance. The relationship between Physics and Social Science has a long and intriguing story. Many physicists and mathematicians gave a great contribution in the filed of Economics and Finance and vice versa. To cite some famous names we can mention Bernoulli, Newton and Majorana who were physicists interested in the social and economic field. On the other side the Italian economist Pareto was the first to introduce a power law probability distribution (also called Pareto distribution), now widely used in Physics, to describe a large number of real-world situations. Also Bechelier, in his phD thesis dealing with the pricing of options in speculative market, was the first to model the Brownian motion, one of the most important model in Physics.

In spite of these historical links, Physics and Economics are also very different. The approach of Physics in studying a given process is to look for universal laws and fundamental principles. Furthermore, Physics attempts to describe the natural world by the application of the scientific method in which one proposes specific hypotheses as explanations of natural phenomena, and design experimental studies that test these predictions. In Economics, not only it is not clear if such universal laws exists, but also the scientific approach to make an experiment to disprove or not a theory is very difficult to carry out. This is mainly due to the fact that in Economics one deals with human persons and hence the reproducibility of experiments is not an easy task. In Economics one deals with human person and not particles. Economic theories are mainly based on a series of assumptions that allow the development of models and theories. The main assumptions are that the market is in equilibrium, that the market is efficient, i.e. possibilities of speculation are not allowed, and that the market agents are perfectly rational in their choices. Evidently these assumptions correspond to an idealized market that not always matches with the reality.

In our view, one of the main goal of physicists in this field is to investigate the features of real markets and to propose model and explanations that include the real characteristics of economic systems. In this perspective one of the most important line of approach of physicists is to find general features for a given observable. These general features are called stylized facts and are very important to construct realistic models and theories. Up to now, the main stylized facts are the fat tailed shape of the distribution of price increments and the volatility clustering phenomenon that is the presence of long range correlation in the time series of absolute price fluctuations. The other main line of research in this field is the implementation of models that can reproduce these stylized facts. Models have been proposed to describe the price dynamics, such as the Lévy model or the ARCH model, but also to explain some collective behavior of agents. These last are called Agent Based Models in which the price formation is the results of a non linear interaction between a large number of agents which can cooperate and evolve.

Second Chapter In this chapter we perform a statistical analysis of financial data. In these perspective we have done an effort to systematically analyze our data to verify the presence of the stylized facts common to all financial time series. Our data are high frequency data from the New York Stock Exchange (NYSE) and are a collection of 100 stock-index for a period of nearly one year in 2005-2006. In a high frequency dataset are reported all the transaction and the quotes, and the relative trading volumes, of a given stock index on a tick-by-tick daily basis. We have verified that the financial time series are highly inhomogeneous within a trading day. In fact the trading activity, i.e. the the number of transaction in a time interval, has a peak both at the open and at the close of the market, with a minimum around lunch time. Similarly, the activity is not uniform across days of the week or months of the year. This phenomenon can be avoid if one consider the transaction time instead of the physical time, as we have done. Therefore, considering our tick by tick dataset we have analyzed the price fluctuations and the trading volume fluctuations. We have studied the empirical probability density function for high frequency price differences (~ 1 min) which displays a leptokurtotic nature. This "fat tails" in the probability distributions decrease with increasing time difference between successive observations. Analyzing the correlation properties of price fluctuations we have found that, while the autocorrelation function of price increments is a delta function, the autocorrelation of absolute price increments discloses long range correlation. This phenomenon is called "volatility clustering". The autocorrelation function of absolute price fluctuations is well described by a power law with scaling exponent between 0.3 and 0.4 as is in literature. Also the volume fluctuations display a similar behavior. The distributions of volume increments are fat tailed disclosing a non Gaussian behavior and the absolute volume fluctuations are long range correlated. In the last part of the chapter we have analyzed the quotations. Quotations are offers to buy or sell a given quantity of shares of a given stock. When the two part reach a mutual agreement a transactions occurs at a price that usually is an intermediate value between the price proposed by the seller (bid price) and that of the buyer (ask price). The transaction price is often different from the exact mean value between bid and ask prices. In this perspective we have analyzed the statistic of the mean price showing that it has the same characteristic of the real transaction price (fat tail, volatility clustering) but has a smoother profile.

Third Chapter In this chapter we have considered an overview of theoretical and empirical models to describe price dynamics. The most common stochastic model of stock price dynamics is the Gaussian behavior that assumes a geometric Brownian motion for the stock price. The Gaussian probability distribution is a universal consequence of the central limit theorem in the limit of long time series on the condition that the financial market is in a stationary states. This hypothesis is not always true, in fact empirical distributions disclose non Gaussian shapes. An improvement in describing financial price fluctuations can be obtained considering other models. In particular we describe the truncated Lévy process because of its impact and acceptance in the scientific community [88]. A Lévy process is characterized by probability density function that discloses a power law like shape for large values of the price fluctuations. This features match with the observed fat tails of the empirical distributions of real price increments. Nevertheless, the Lévy model can not reproduce the volatility clustering feature of real data. To overcome this problem Engle in 1982 proposed a stochastic process with a time dependent variance (ARCH and GARCH models) [53]. For these models, it is possible to show analytically that the kurtosis of the probability density function is larger than 3 (that is the value for a Gaussian distribution). This leptokurtic probability density function can fit very well high frequency real data. Also the correlation of absolute price increments for an ARCH (or GARCH) model is non zero, such as real data. The difference is that, while the ARCH model displays an exponentially decaying correlation, real data has a long range correlation.

An important theoretical result obtained assuming the that the price follows a geometrical Brownian Motion, is the Black and Scholes equation for the option pricing problem [22]. Options are a kind of financial derivatives, that are financial products whose price depends on the price of another financial product. Assuming a geometrical Brownian motion for the price of the underlying product, Black and Scholes have found a way to estimate the price of an option. This theory is based o some hypothesis that are the efficiency of the market, the absence of transactions cost, the existence of a constant market interest rate and other assumptions. Real markets do not satisfy all these hypothesis, therefore the Black and Scholes equation is only a first approximation in solving the option pricing problem. More realistic models for option pricing have been developed thanks to the knowledge of the statistical properties of real price dynamics.

At last, we have described the Minority Game Theory [39]. This is an alternative approach in describing economic systems in which one constructs a simplified model of the market where a large number of agents interact and it is possible to study the collective behaviors in financial market. The Minority Game is the formalization of the so called “El Farol Bar” problem [10, 11]. In this game each day agents decide whether or not to go to the bar; if there is room in the bar they are happy, and if it is too crowded they are disappointed. By definition only a minority of the people can be happy. With model like this it is possible to show that players, in the absence of communication, cooperate with each other, with the aim of maximize the collective utility.

Fourth Chapter In this chapter we study the stock price dynamics by mean of its roughness properties. The roughness properties of a given variable are the way in which the fluctuations of the variable depend from the time interval in which these fluctuations are calculated. In 1951 Hurst found that, for long time series, this function is empirically well described by a power law, the scaling exponent is called the Hurst exponent. For a simple random walk the Hurst exponent measured on long scales is $H = 0.5$. The usual interpretation of an exponent larger than 0.5 is that the process studied show long range correlations and persistence. In estimating the roughness properties of a given variable, for many reasons one can be interested in considering finite size datasets. In our case this is due to the fact that financial time series are homogenous only within a single trading day because in the night the price undergoes a large jump. In this perspective, we have performed a detailed analysis of the problem of estimating the roughness of a finite size process [6, 4]. We have started studying the properties of a finite size Random Walk. With an analytical calculation we have shown that considering finite size samples has the effect to enhance the Hurst exponent even for a simple Random Walk. Including the effects of correlations and fat tails, by means of simulations, we have observed that this effect is further increased. Hence, we introduce an alternative way of characterize the roughness properties of a process looking to the fluctuations of a variable from a suitable moving average. In this way be obtain an automatic detrendization of the signal that leads to some interesting results that could be useful also in a technical analysis of the market. At last, we introduce a way of detecting complex correlations. Linear correlation can be studied by means of the usual statistical tools such as the correlation function, but higher order correlations are not so easy to analyze. Our tool consist in using an analytical equation that we have derived for an uncorrelated process, and see if this equation is still valid for another process. We have tested our tool for an ARCH process and we show that it is suitable to detect the complex correlations present in this process. Also an analysis on real financial data have lead to good results.

Fifth Chapter While in the previous chapter we have studied the properties of the price fluctuations, now we analyze also the properties of the sign (direction) of these fluctuations. In this perspective, we have considered a new tool to analyze stock price dynamics that we have introduced together with a group of Japanese researchers [114, 5]. This model assumes the possibility that the price dynamics can be affected by its distance from a moving average calculated on the previous steps. This model corresponds to a Random Walk with also a term of force (the derivative of a potential), and this force depends on the distance of the price from the moving average. The idea is that price can be attracted or repelled by its moving average. This aspect can be interpreted as due to the agents' behavior. In a phase in which the market is dominated by trend followers agents, the market is unstable and hence the price is repelled by its moving average, while if fundamentalists dominate the market is stable. The simplest case is to assume a linear force and hence a quadratic potential. An analysis on Yen/USD rates shows that the real data indeed are well described by this model of Random Walk in a quadratic potential that can be either attractive or repulsive. We have also performed an analysis of this method on a minority-majority game as a test. This minority-majority game is a mixed game in which there is an alternation of phases in which the market is unstable and dominated by trend followers and other phases in which the market is stable. This method is able to identifying the different phases of the game disclosing attractive potentials when agents play a minority game and hence the market is stable and a repulsive potential when agents are trend followers. We have also performed an analysis on NYSE stock price one day long datasets. Here the situation is much more complicated because this analysis shows some asymmetric, non quadratic potentials. Only averaging over many trading days one can recover a quadratic shape of the potential.

Sixth Chapter In this last chapter we will show some very preliminary results obtained in simulating an agent based model in which the agents have trading strategies based on some moving average. Agents can be both fundamentalist or trend followers and perform their moving average on different time scales. The collective behavior (price) is obtained averaging over all the possible strategies. We have performed the potential analysis described in the previous chapter on such a model. When the market is dominated by trend followers we indeed recover a repulsive potential while when the market is dominated by fundamentalist the potential is attractive. If the relative number of trend followers exceed a given threshold the strength of the contribution of trend followers is not balanced by the noise and the collective price increases exponentially. This results represents only a phase of test study of such a model. The effects of evolution must still be included to observe a more realistic and complex behavior.

Chapter 1

Physics and Economics

In this first chapter we will explain the possible connections between Physics and Economics. In the past decades the interest of physicists in the field of Economics is increased and a lot of persons have began to do academic research on the analysis and the modeling of economic systems.

In accordance with a research made by the *UK Engineering and Physical Sciences Research Council* in 1997, the 48% of the PhD students in Mathematical Physics and in Statistical Physics have continued their career working in the financial field [63]. An important question may be whether the growth of physicist doing researching in Economics is just a temporary phenomenon or whether a background in Physics is a is a real advantage to understanding Economics [58].

Actually the relationship between Physics and Social Science has a long and interesting story [45] and many of the persons who carried an important contribution to the development of Economics, had a physical or a mathematical background.

1.1 The possible relevance of Statistical Physics to Economics and Finance

Maybe the first famous example of a physicist who held an important position in a financial institution was Isaac Newton [119]. In 1669 he was designated director of the Royal Mint of London and he quickly began the terror of the English falsifiers. Another illustrious example is Daniel Bernoulli that, in 1738, introduced the idea of utility to describe people's preferences [58].

But the first who presciently outlined both the opportunities and fitfalls in applying statistical Physics to social sciences was Ettore Majorana. In 1938 he wrote a paper called *The Value of the Statistical Laws in Physics* *and in the Social Sciences* [82]. In his pioneering paper he underlined the importance that the basic principles of quantum mechanics have pointed out the discovery that the laws of elementary processes have a statistical character. This conclusion, in the view of Majorana, make clear the analogy between Physics and Social Sciences. In the field of Economics and Social Sciences, the analysis of financial markets is the discipline that better can be suitable for a rigorous mathematical treatment.

In the last decade, the market operators have made use of the calculators and this has led to the electronic registration of events which describe detailed economic phenomena. For example financial markets are continously monitored, down to scales of seconds. Many of these informations are avaliable in the form of electronic data for studies and researches. This informations permits to perform empirical analysis of financial dynamics with a very accurate statistical resolution. Nowadays the scientific community can dispose of time series of million of data coming from the financial markets. In this way is emerged the possibility to compare with extreme accuracy the information coming from the real data with models, theories and hypothesis. These hypothesis can also be refuted in the proper spirit of the scientific method.

Moreover, from a theoretical point of view, the financial time series can be schematized as the result of a nonlinear interaction between many agents [81]. In this sense financial markets exhibit several of the properties that characterize complex systems. Therefore we are in the conceptual scheme of the Statistical Physics and physicists can contribute to the modeling of financial and economic systems using tools and methodologies developed in Statistical Mechanics and Theoretical Physics.

These considerations can explain why, starting from '80, an increasing number of scientists devoted oneself to the application of theories belonging to the Physics of Complex Systems to the understanding of the subtle mechanism of the socio-economic interactions. In the nineties the elimination of the barriers for the free movements of the capitals and the increase of the volume of the negotiations, have made essential the use of quantitative methods in the analysis of financial markets. In these years, in fact, the major financial institutions have recruited a lot of people coming from mathematical and physical backgrounds.

Statistical Physics is the general study of any composite system formed of a large number of similar components; particularly the study of the collective behaviors of this system. It applies just as well to an assembly of something, when that "something"is people, rocks, cue balls, particles, cars, bits on a network, cells, etc. It doesn't matter what the system is, or at what level the abstraction is made (particles to physical objects, individuals to societies, cells to bodies, stars to galaxies; galaxies to universes; etc.). The same general principles will come to bear; and the same math and methods will apply.

In this context, Economics is classified as a manifestation of Complex Social Systems. To clarify this statement, we have to discuss what we understand by Complex Systems. Unfortunately, an exact definition of a complex systems is still an open problem. In a heuristic manner, we may describe them as systems composed of many particles, or objects, or elements which may be of the same or of different kinds. The elements may be interact in a more or less complicated fashion by more or less nonlinear couplings [110]. It has been found that in spite of the vast differences among various stocks, commodities, currencies, and stock indices, there are several stylized facts common to all of them.

1.2 Interdisciplinarity: a critical discussion

One may wonder how such different subjects as Physics and Economics relate to each other. Even though, the vast amount of research conducted during the last decades proved that the marriage between these subjects is a success. This recently developed interdisciplinary area is coined as *econophysics*. The approach of these two disciplines to the field is quite different.

In Physics, it is assumed that some universals Laws exist, and one try to discover them. In Economics it is not clear if such universal laws do exist or not. Another open question is if Economics can be considered a Science in the definition of Popper. In Popper's view, any hypothesis that does not make testable predictions is simply not Science. Such a hypothesis may be useful or valuable, but it cannot be said to be Science. He take falsification as his criterion of demarcation between what is and is not genuinely scientific: a theory should be considered scientific if and only if it is refutable. So a real Science must be disprovable with experiments. Following the Scientific Methods, one of the major points is the reproducibility of experiments. This last aspect is not so clear in Economics. This is due to many reasons. The first is that in Economics one deals with human persons that could not to behave at the same way at different times. The second is that one can never deal with identical situations because the market is not always stationary and the laws of the dynamics are not necessarily the same at different times. Further more, the economists view, according to standard Economics, is that the stock price variation is a mere reflection of external information: a political change, environmental disaster, etc. Therefore, no one should be able to earn huge profits just by analyzing past prices. Several researcher have shown that this is far from the truth. If markets are so "efficient" in reflecting external information, it would be not possible to explain how can many investors gain profits. In other words, human agents are not completely rational.

Physicists approach in financial market is fundamentally different from economists. In Physics one try to find the underlying concepts behind the real word phenomena. The physical way offers, however, the possibility of describing financial phenomena in a universal theoretical framework. There is the hope that the physical progress in investigation of complex systems with other apparent behavior allows a deeper insight also into the dynamics of financial markets. Results that would be derived from discoveries in other field of Physics. The financial market is a complex system from a physical point of view.

Actually, the acceptance of econophysicists by the economic community seems a particularly long way off [51]. Even fully fledged economists are ostracized by the mainstream if they do not embrace the tenets of "neo-classical" economic theory, no matter how untenable its principles (identical, utility-maximizing economic agents operating in an equilibrium market) now seem.

In the last year there has been an open discussion about the real merit of econophysicists in Economics. Following [17] , some economists had hoped that physicists might shake up the rigid theories typical of mainstream Economics. But so far, they're unimpressed by physicist's handling of the markets. Another group of economists penned a paper entitled "Worrying trends in Econophysics" [62]. In this article the authors recognize that econophysics has already made a number of important empirical contributions to the understanding of the social and economic world. These fall mainly into the areas of Finance and industrial Economics, where in each case there is a large amount of reasonably well defined data. The critics are that econophysicists are often in the dark about the work of economists, do not use robust statistical methodology and believe that universal empirical regularities can be found in many area of Economics. In [107] there is an advise on how can econophysicists contribute to Economics. The idea is that the natural strength of Physics is dealing with empirical data so they should give highest priority to solve concrete problems from economic practice rather than from economic theory. To conclude the review, an optimistic picture of the union between Physics and Economics comes from [58]. The authors says that collaborations between physicists and economists can add value to the science of Economics.

1.3 Significant contributions of physicists in Finance

In this section we expose some of the related problems in Finance in which physicists have made significant contributions over the last decades. The recent body of work done by physicists and others have produce convincing evidences that the *Standard Model of Finance* is not fully capable of describing real markets, and hence new ideas and models are called for, some of which have come straight from Physics [117]. Almost all the models used in finance is based on hypothesis that guarantee for a mathematical analysis. The main hypothesis are that the market is “efficient” and “arbitrage-free”. With these hypothesis the dynamic of the the market-prices is assumed to be a stochastic Gaussian process. The work of physicists has pointed out two important statistical properties of the price fluctuations that shows the limits of the classical model of the market. These properties are the non-Gaussian nature of the stochastic process of the financial prices and the correlations of the price fluctuations amplitude. These statistical properties are now accepted as stylized facts in the analysis of financial time series. First we will describe what we mean for efficient and arbitrage-free market an then we will expose the statistical stylizes facts of the real markets.

1.3.1 Efficient Market

The concept of efficiency is central in Finance and was introduced by E. Fama in 1970 in his famous article “Efficient Capital Markets” [55]. In Finance markets are “efficient” when the price on traded assets, e.g. stocks, bonds, or property, already reflects all known information and therefore is unbiased in the sense that reflects the collective beliefs of all investors about future prospects. The efficient market hypothesis implies that it is not possible to consistently outperform the market by using any information that the market already knows, except through luck or obtaining and trading on inside information. Information or *news* in an efficient market are defined as anything that may affect stock prices that is unknowable in the present and thus appears randomly in the future. This random information will be the cause of future stock price changes. So one of the consequences of market efficiency is that prices should not be predictable. In efficient markets, prices become not predictable but random, so no investment pattern can be discerned. A planned approach to investment, therefore, cannot be successful. This ”random walk”of prices results in the failure of any investment strategy that aims to beat the market consistently.

The concept of market efficiency had been anticipated at the beginning of the century in the dissertation submitted by Bachelier to the Sorbonne for his PhD in Mathematics [14]. In his opening paragraph, Bachelier recognize that *past, present and even discounted future events are reflected in market price, but often show no apparent relation to price changes.* This recognition of the informational efficiency of the market leads Bachelier to conclude that *if the market, in effect, does not predict its fluctuations, it does assess them as being more or less likely, and this likelihood can be evaluated mathematically.* This leads to a brilliant analysis where he developed, essentially correctly, the theory of Random Walk that anticipates the Albert Einstein's subsequent derivation of the Brownian motion. In his thesis, in fact, Bachelier proposed the Random Walk as the fundamental model for financial time series many decades before this idea became the basis for modern theoretical Finance.

The efficient market hypothesis is simple in principle, but remains elusive. Evolving from an initial puzzling set of observations about the random character of security prices, it became the dominant paradigm in finance during the 1970s [48, 83]. Although many in finance now believe that markets are efficient, it is not unanimous. One group of these people are technical analysts. These people believe they can predict price movements based on historical prices (a clear violation of efficiency). Although the vast majority of academic papers finds no benefit to technical analysis, it has long been difficult to explain why there are technical analysts around (how can we say the market is inefficient in keeping technical analysts while say it is efficient in other things). Several recent academic papers have at least suggested that technical analysis might not be worthless.

1.3.2 Arbitrage

In Economics, arbitrage is the practice of taking advantage of a state of imbalance between two or more markets: a combination of matching deals are struck that capitalize upon the imbalance, the profit being the difference between the market prices.

As an example, suppose that the exchange rate in London are $L5 = \$10 = Y1000$, and that the exchange rate in Tokyo are $L6 = \$10 = Y1000$. Converting \$10 to $L6$ in Tokyo and converting that $L6$ into \$12 in London, for a profit of \$2, would be arbitrage. In reality, this triangle arbitrage is so simple that it almost never occurs. But more complicated foreign exchange arbitrage are much more common.

If the market price does not allow for profitable arbitrage, the prices are said to constitute an arbitrage equilibrium or arbitrage free market. An arbitrage equilibrium is a precondition for a general economic equilibrium.

If we are in an economic equilibrium, as soon as an arbitrage opportunity begins to be exploited, the system moves in the direction that gradually eliminate the arbitrage opportunity.

1.3.3 Fat tails

Now we are going to analyze some of the empirical evidence that real markets deviate from the hypothesis of efficient and arbitrage free market.

Following the arguments of Bachelier and in the hypothesis of efficient market, the most diffuse model for the price dynamic in a financial market is the geometric Brownian motion. In a geometrical Brownian motions the differences of the logarithms of prices are Gaussian distributed. In Finance the differences of the logarithms of prices are called *returns*. It is thus important to ask whether real markets actually fit into this model.

Mandelbrot in 1963, [84], was perhaps the first person to challenge the paradigm that returns are normally distributed. He analyzed cotton prices on various exchanges in the United States and found evidences that their distribution of returns decay as a power law and hence much slower than a Gaussian. Hence, there is an evidence that the empirical distribution of returns has "fatter tails" when compared with a Gaussian distribution. In his studies, Mandelbrot proposed to describe the distribution of returns with a Lévy stable distribution.

1.3.4 Volatility clustering

Empirical studies have shown that, while returns themselves are uncorrelated, absolute returns or their squares display a positive, significant and slowly decaying autocorrelation function. This phenomenon is called clustered volatility. As noted by Mandelbrot, large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes. The correlation decays as a power law of the form $\tau^{-\gamma}$. Since $0 < \gamma < 1$, the size of price changes is a long memory process and thus displays anomalous diffusion and a very slow convergence of statistical averages.

Many model for prices dynamic have been proposed to explain phenomenon like volatility clustering and fat tails. Between them one of the most famous and widely used in Finance and Economics for modeling conditional heteroscedasticity and volatility clustering is the autoregressive conditional heteroskedastic (ARCH) processes first proposed by Engle in 1982 [53].

1.4 The main lines of approaches

1.4.1 Statistical analysis of financial time series

One of the main lines of research in this field is represented by the effort to obtain a complete statistical characterization of the price dynamics and of other quantities. The main effort is to find empirical regularities in financial data. These features are called "stylized fact" and are very important to test and discriminate theoretical models. The most important stylized fact is perhaps the non Gaussian distribution of price fluctuations. A lot of work as been done to identify the shape of the distribution of price changes [27, 67, 86, 113]. Another important stylized fact is the clustering of volatility. This phenomenon means that the absolute price fluctuations are long ranged correlated variables. In this perspective many works address to the detection of complex higher order correlations in price fluctuations [41, 100]. Other statistical analysis regard the scaling properties of the price fluctuations. A way to characterize this property is through the scaling exponent of the so called roughness [85]. This is a measurement of the small-scale variations in the height of a physical variable. Another relevant aspect is the analysis of the correlations between stocks. This kind of analysis important to obtain a hierarchical organization of a group of stocks [26]. An interesting way to study the correlations between stock is by means of financial networks [25].

1.4.2 Theoretical models to describe the features of financial markets

Another area concerns the development of theoretical models which could reproduce the stylized facts of financial data. The most common stochastic model used to describe the price dynamics is the Random Walk model [14]. This model can reproduce the price dynamics only in a first approximation. In fact this model can not reproduce the fat tailed distribution of real stock-price distribution. From this situations a lot of model have been proposed to explain the non Gaussian shape of the price increments distribution. One of the most accepted model which can explain the empirical evidence that the price fluctuations distribution has power law shaped tails is the Lévy flight model [87]. Another famous model which can encompass the empirical features of financial markets is the ARCH model (and is generalization GARCH) [52]. This model assumes that the variance is not constant but it is a time depend stochastic variable.

A theoretical effort is also done in the problem to estimate the price of derivative financial options. The Black & Scholes model represents the most renowned work in this field [22]. Nevertheless, this model represents only a first approximation of what happens in real markets and other models have been proposed [28].

1.4.3 Agent based models

Another way is studying financial systems is to consider a bottom-up approach analyzing the market macroscopic variables as the results of an interaction between many microscopic units (agents) [13]. The idea is to construct the computational devices (known as agents with some properties) and then, simulate them in parallel to model the real phenomena. The process is one of emergence from the lower (micro) level of the social system to the higher level (macro). In Physics applied to financial systems, the the most successfully agent based model is the so called Minority Game and its developments [39]. In the Minority Game, an odd number of players have to choose one of two choices independently at each turn. The players who end up on the minority side win. From this simple model, which also can be solved analytically [37, 92], and from some of its developments, can be recovered the stylized facts of real markets.

Chapter 2

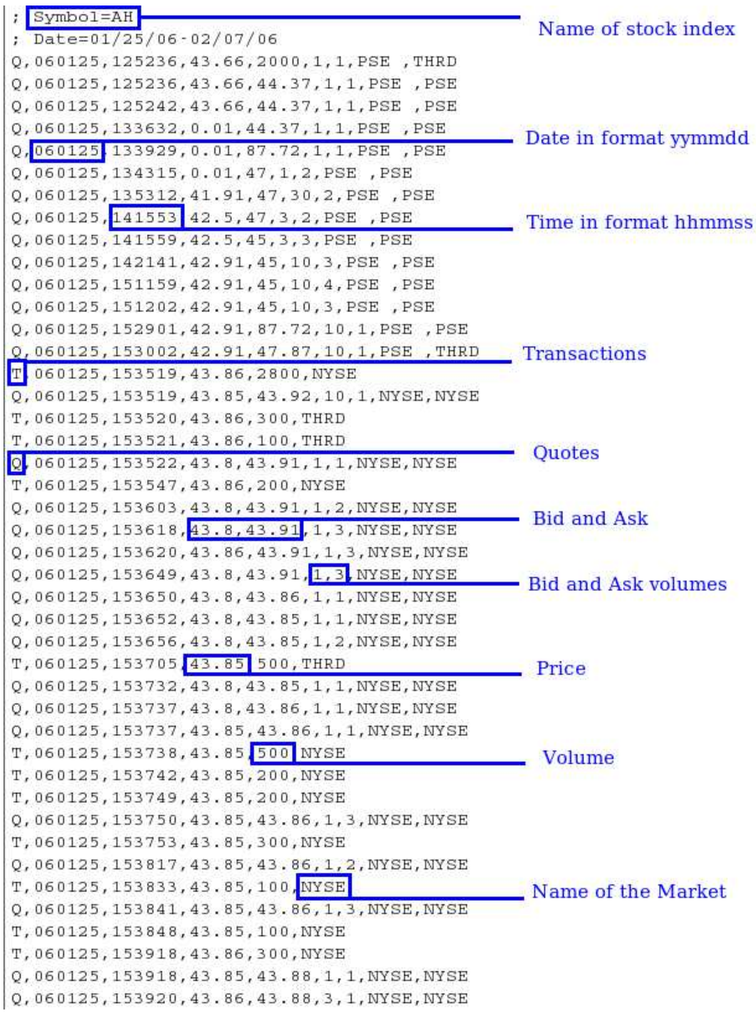
Statistical Analysis of Market Data

When economists, finance mathematicians, and physicists are dealing with a financial problem, a wide variety of ideas and techniques are avaliable to generate quantitative answers. The input to solve this problem may be obtained from an empirical analysis of financial data or from other quantitative economic investigations. Actually, this part is an intermediate field between financial mathematics and financial management because it requires economic experience to decide which data are important in the context of the problem and what is the order of their significance. Therefore, in this chapter, we will analyze some statistical properties of the financial variables which we consider important for our studies. In particular we will analyze the statistical properties of financial time series from our dataset. Actually these properties are recognized as "stylized facts" which seem to be common for a wide variety of markets [43, 49, 102, 103].

2.1 Dataset properties

There are thousands of possible price series we could look at, and hundreds of different markets. One of the significant contributions of Econophysics has been to establish that the price statistics across a wide range of supposedly different financial markets worldwide, exhibit certain universal properties. It seems that despite the differences in detail between these markets in terms of how trades are registered, trading hours, etc., there is something that they hold in common which is driving the market price-dynamics. Given the limited space, we cannot carry out a thorough statistical analysis over many different markets. Instead the results presented here provide a flavor of the ones already present in the Econophysics literature. In particular we refer to Mantegna and Stanley, [88], and Bouchaud and Potters, [30].

In this chapter we will illustrate some common features and stylized facts present in financial markets, focusing on a specific market dataset. We will use this dataset also to perform all the analysis present in this thesis. This dataset is the New York Stock Exchange (NYSE) composite index recorded on a tick-by-tick daily basis in a period of nearly one year in 2005-2006. The NYSE composite index reflects the long-running, well-established and highly liquid US stock market.



```
; Symbol=AH
; Date=01/25/06-02/07/06
Q,060125,125236,43.66,2000,1,1,PSE ,THRD
Q,060125,125236,43.66,44.37,1,1,PSE ,PSE
Q,060125,125242,43.66,44.37,1,1,PSE ,PSE
Q,060125,133632,0.01,44.37,1,1,PSE ,PSE
Q,060125,133929,0.01,87.72,1,1,PSE ,PSE
Q,060125,134315,0.01,47,1,2,PSE ,PSE
Q,060125,135312,41.91,47,30,2,PSE ,PSE
Q,060125,141553,42.5,47,3,2,PSE ,PSE
Q,060125,141559,42.5,45,3,3,PSE ,PSE
Q,060125,142141,42.91,45,10,3,PSE ,PSE
Q,060125,151159,42.91,45,10,4,PSE ,PSE
Q,060125,151202,42.91,45,10,3,PSE ,PSE
Q,060125,152901,42.91,87.72,10,1,PSE ,PSE
Q,060125,153002,42.91,47.87,10,1,PSE ,THRD
T,060125,153519,43.86,2800,NYSE
Q,060125,153519,43.85,43.92,10,1,NYSE,NYSE
T,060125,153520,43.86,300,THRD
T,060125,153521,43.86,100,THRD
Q,060125,153522,43.8,43.91,1,1,NYSE,NYSE
T,060125,153547,43.86,200,NYSE
Q,060125,153603,43.8,43.91,1,2,NYSE,NYSE
Q,060125,153618,43.8,43.91,1,3,NYSE,NYSE
Q,060125,153620,43.86,43.91,1,3,NYSE,NYSE
Q,060125,153649,43.8,43.91,1,3,NYSE,NYSE
Q,060125,153650,43.8,43.86,1,1,NYSE,NYSE
Q,060125,153652,43.8,43.85,1,1,NYSE,NYSE
Q,060125,153656,43.8,43.85,1,2,NYSE,NYSE
T,060125,153705,43.85,500,THRD
Q,060125,153732,43.8,43.85,1,1,NYSE,NYSE
Q,060125,153737,43.8,43.86,1,1,NYSE,NYSE
Q,060125,153737,43.85,43.86,1,1,NYSE,NYSE
T,060125,153738,43.85,500,NYSE
T,060125,153742,43.85,200,NYSE
T,060125,153749,43.85,200,NYSE
Q,060125,153750,43.85,43.86,1,3,NYSE,NYSE
T,060125,153753,43.85,300,NYSE
Q,060125,153817,43.85,43.86,1,2,NYSE,NYSE
T,060125,153833,43.85,100,NYSE
Q,060125,153841,43.85,43.86,1,3,NYSE,NYSE
T,060125,153848,43.85,100,NYSE
T,060125,153918,43.86,300,NYSE
Q,060125,153918,43.85,43.88,1,1,NYSE,NYSE
Q,060125,153920,43.86,43.88,3,1,NYSE,NYSE
```


Figure 2.1: High frequency tick-by-tick records of AH index. The dataset contains information about all the transactions and the quotes of the index at each time.

Our dataset contains the price time series of all the transactions of a selection of 100 NYSE stocks. These have been selected to be representative and with intermediate volatility in comparison with the other NYSE stock indexes.

In Fig. 2.1 is shown a typical dataset with many details stored. In particular are reported all the transactions for the index AH (Armor Holdings) in a given day. In the dataset we can see date and time of each transaction. The time interval between two subsequent transactions is of a few seconds, due to the high liquidity of the stock considered. The symbol "T" at the beginning of a line means that the data refer to a transaction, that is an effective change in the status of the price of the given stock. So we have the price of the transaction and the volume of shares traded during the transaction. Instead, the lines with the symbol "Q" refers to a quote. The term quote specifically refers to the "bid" or "ask" prices of an asset. The bid and ask (or offer) of a security are the prices at which buyers and sellers are willing to purchase or sell. The bid shows the current price at which a buyer is willing to purchase shares, while the ask shows the current price at which they are willing to sell. The quantities at which these trades are placed are referred to as "bid size"and "ask size". For example, suppose you own 100 shares of ABC Corporation, and you want to buy or sell these shares. You obtain a quote which indicates that the current bid price (the price at which you can sell this stock) is \$100 per share, and the current ask price (the price at which you could buy the shares) is \$101. These quotes mean that someone is willing to buy at least 100 shares of ABC at \$100 and that another person is willing to sell at least 100 shares of the stock at \$101. The difference between these bid and ask prices is referred to as the "bid-ask spread". In the above example, the spread is one dollar. If you were to simultaneously buy and sell 100 shares of this stock in the market you would lose the \$1.00 spread. So the bid-ask spread is a form of transaction cost. In our dataset bid and ask prices and the relative sizes are reported.

At last, in Fig. 2.1 appears the specific market in which the stocks are traded. For example we can see the symbols NYSE that is New York Stock Exchange and PSE that is Pacific Stock Exchange.

2.2 Time scales in financial data

In Natural Sciences, especially in Physics, the problem of reference units is considered basic to all experimental and theoretical works. The problem with the financial data is that the scales used here are often given in units that are themselves fluctuating in time.

The main and natural alternatives that can be used for the time scale are physical time, number of transactions and trading time. The physical time is well-defined, but one has to consider that stock markets are not open all day long. Traditional stock markets close overnight, on weekends and more or less randomly on holidays. Therefore, news that become known when the market is closed cannot affect the price immediately, but rather accumulate until the market open again. This lead to a gap in the price during the period in which the market has been closed, that appears like a discontinuity. For example, the overnight gap, i.e. the difference between the closing price of a given day and the new opening price, has the same order of magnitude of the fluctuation of price in a whole trading day. In Fig. 2.2 we can see a typical overnight gap of the price for a stock index.

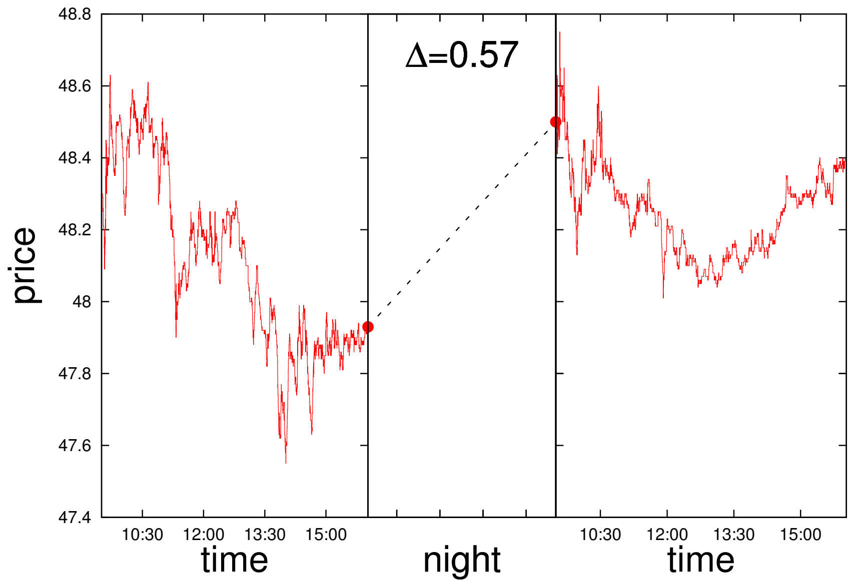


Figure 2.2: Overnight jump of price between two days. This gap is very large, typically of the order of the total daily fluctuation (see Tab.2.1).

These jumps pose serious problem in linking the data of one day to those of the next day to obtain a longer time series. This means that the data are reasonably homogeneous from the time scale of a few seconds to a few hours but going to longer times can be rather arbitrary due to these large overnight jumps [6]. In Tab. 2.1 we present a detailed analysis of this phenomenon.

For each stock, which is labeled in the first column, we have, in the second column the average over 80 days of the absolute value of the gap between

stock	$<\|P_{op}-P_{cl}\|>$	$<\|\Delta\|>$	σ_Δ	σ_P	$\sigma_{\delta P}$
AH	0.73494	0.36950	0.59100	0.28539	0.02152
AVO	0.47561	0.18862	0.56508	0.23161	0.01698
BA	0.41926	0.21437	0.42131	0.19530	0.01056
BRO	0.40877	0.15750	0.37375	0.19091	0.01607
CAI	0.81284	0.39238	0.86836	0.31750	0.02323
DRI	0.30753	0.09850	0.23245	0.11490	0.01065
GE	0.22691	0.11688	0.17154	0.10304	0.00652
GLK	0.28272	0.10212	0.23420	0.12998	0.01054
GM	0.35593	0.15725	0.25058	0.14597	0.00833
JWN	0.44531	0.23325	0.45625	0.20444	0.01249
KSS	0.57759	0.29628	0.48844	0.22275	0.01355
MCD	0.24457	0.13850	0.20268	0.10288	0.00758
MHS	0.43605	0.20437	0.40161	0.17267	0.01126
MIK	0.34531	0.62375	3.12751	0.14479	0.01320
MLS	0.55309	0.17287	0.27860	0.21948	0.02045
PG	0.40321	0.24462	0.46493	0.17056	0.00906
TXI	0.79704	0.22362	0.62309	0.33799	0.02964
UDI	0.44679	0.22375	0.80100	0.19003	0.01469
VNO	0.65864	0.21950	0.36921	0.26285	0.02443
WGR	0.40877	0.16937	0.36687	0.17846	0.01681

Table 2.1: Properties of the night jumps with respect to the daily fluctuations of various stocks. The data refer to the average values over 80 trading days. $<|P_{op}-P_{cl}|>$: average of absolute value of the gap between opening and closing price for each day. $<|\Delta|>$: average of the absolute value of the night jumps. σ_Δ: variance of the absolute value of the night jumps. σ_P: total daily variance of the price value. $\sigma_{\delta P}$: variance of the price fluctuations between two transactions.

opening and closing price, ($<|P_{op}-P_{cl}|>$), indicated in US \$. In the third column we indicate with $<|\Delta|>$ the average of the absolute values of the night jumps. One can see immediately that they are of the same order of magnitude. In the fourth column σ_Δ indicates the variance of the night jumps. These values are really very large and clearly show that there is a strong discontinuity from the closing price to the next day opening. In the fifth column we show the variance of price fluctuation within one day averaged over the 80 days (average single day volatility). Finally in the sixth column we show the variance of the price fluctuations between two transactions. One can see therefore that the night jumps are more than one order of magnitude larger than the typical price change between two transactions. This leads to a very serious problem if one tries to extend these time series beyond the time scale of a single day. In fact, if one simply continues to the next day, one has anomalous jumps for the night which cannot be treated as a standard price change. An alternative possibility could be to artificially eliminate the night jumps and rescale the price correspondingly. This would produce a homogeneous data set which, however, does not correspond to the original data.

The alternative is to use the number of transactions. This scale eliminates the effect of the randomly distributed time intervals elapsing between transactions. Another source of randomness, the volume of transactions, still remains.

At last, the trading time is the time that elapses during open market hours. This timescale depends on the local stock exchange. Furthermore, price changes and the release of relevant information during the night lead to jumps at the opening. So arises the problem that in high-frequency analysis, overnight price changes are treated as short time price changes. The second problem is that, using this scale, one assumes that market activity is uniform during the market hours. This last assumption is not verified by empirical analysis, as we will see in the next section. However the trading time is the most common choice in many research studies.

2.3 Trading activity

Trading activity is not uniform during opening hours, either in terms of volume or in number of trades. The trading activity has a peak both at the open and at the close of the market, with a minimum around lunch time.

In our dataset the number of transactions per day ranges from 500 to 5000 implying a typical time interval between transactions of a few seconds. The density of operations within a day is characterized by a concave shape which is rather universal as shown in Fig. 2.3. In this figure each point is an average value performed over 80 trading days for 20 stock indexes from NYSE. The result shown in Fig. 2.3 means that, with respect to the physical time there are systematic density fluctuations up to a factor of two with a minimum around the center. This effect is obviously eliminated using the tick by tick time, in which the physical time is not considered.

Similarly, the activity is not uniform across days of the week, or months of the year. This phenomenon is called the “Calendar Effects”, that is the tendency of stocks to perform differently at different times, including such anomalies as the January effect, month-of-the-year effect, day-of-the-week

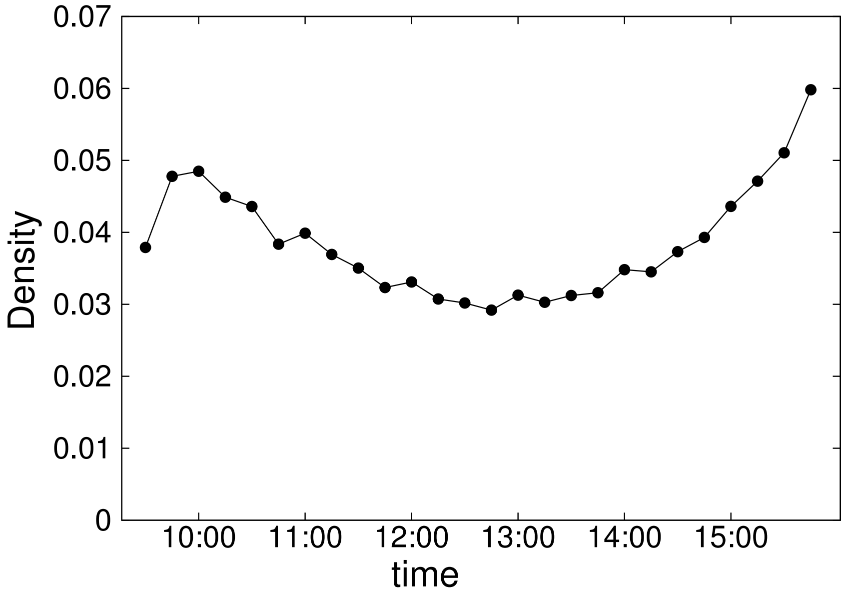


Figure 2.3: Behavior of the density of transactions within a day. This concave behavior with a maximum fluctuation up to a factor of two is a general feature for all stocks.

effect, and holiday effect.

Finally not all markets are simultaneously open, and different time zones behave differently. A well-known effect is for example an increased activity at 2:30 p.m. European time, due to the fact that major U.S. economic figures are announced at 8:30 east-coast time.

2.4 Price fluctuations

Let us define $X(t)$ as the price of a given financial asset at time t. Then, we may ask which is the appropriate variable describing the stochastic behavior of the price fluctuations. The simplest choice is the introduction of the price changes:

$$\Delta X(t, \Delta t) = X(t + \Delta t) - X(t) \tag{2.1}$$

where δt is a well-defined interval of the time series. The merit of this approach is that Eq. 2.1 is a simple linear relation. This means in particular that the price change are additive:

$$\Delta X(t, \Delta t_1 + \Delta t_2) = \Delta X(t + \Delta t_1, \Delta t_2) + \Delta X(t, \Delta t_1) \tag{2.2}$$

Unfortunately, the definition in Eq. 2.1 is seriously affected by possible change in money scales due to possible fluctuations in the global currency markets or inflation effects. Furthermore, the strength of the fluctuations $\Delta X(t)$ depends seriously on the order of magnitude of the actual price of the asset $X(t)$. A more appropriate choice is to use the *returns*, defined as:

$$R(t) = \frac{\Delta X(t)}{X(t)} \tag{2.3}$$

The advantage of this definition is that returns provide a direct percentage of gain or loss in a given time period. Therefore, the return is a very natural measure describing the price fluctuations of share. The disadvantage is that the returns are nonlinearly coupled:

$$R(t, \Delta t_1 + \Delta t_2) = [R(t, \Delta t_1) + 1][R(t + \Delta t_1, \Delta t_2) + 1] - 1 \tag{2.4}$$

To overcome this problem, we introduce the difference of the natural logarithm of the price (*log-returns*):

$$\eta(t) = \ln X(t + \Delta t) - \ln X(t) = \ln \frac{X(t + \Delta t)}{X(t)} \tag{2.5}$$

This quantity is additive,

$$\eta(t, \Delta t_1 + \Delta t_2) = \eta(t, \Delta t_1) + \eta(t + \Delta t_1, \Delta t_2) \tag{2.6}$$

and the corrections of scale changes is incorporated without requiring deflator or discounting factors. The problem is that a nonlinear transformation is used, and nonlinearity strongly affects the statistical properties of a stochastic process.

In the whole of modern financial literature, it is postulated that the relevant variable to study is not the increments $\Delta X(t)$ itself, but rather the log-return $\eta(t)$. Different stocks can have completely different prices, and therefore unrelated absolute daily price changes, but rather similar daily returns. Moreover, on long time scales, price changes tend to be proportional to prices themselves. However, on shorter time scales, the choice of returns rather than price increments is much less justified.

A lot of different models have been proposed to study the dynamic of price fluctuations. The most common stochastic model stock price dynamics is the Gaussian behavior that assumes a geometric Brownian diffusion of the asset prices and a corresponding arithmetic Brownian motion for the log-returns. This model provides a first approximation of the behavior observed in empirical data. However, the Gaussian probability distribution is a universal consequence of the central limit theorem in the limit of long times on the condition that the financial market is in a stationary state. Indeed, the Gaussian law can possibly deviate considerably from the portability distribution function determined empirically for short timescales.

Serious systematic deviations from the Gaussian model predictions are observed, which indicate that the empirically determined probability distributions exhibit a pronounced leptokurtic behavior. A highly leptokurtic distribution is characterized by a narrower and larger maximum and by fatter tails than in the Gaussian case. The degree of leptokurtosis increases with decreasing time difference between successive observations.

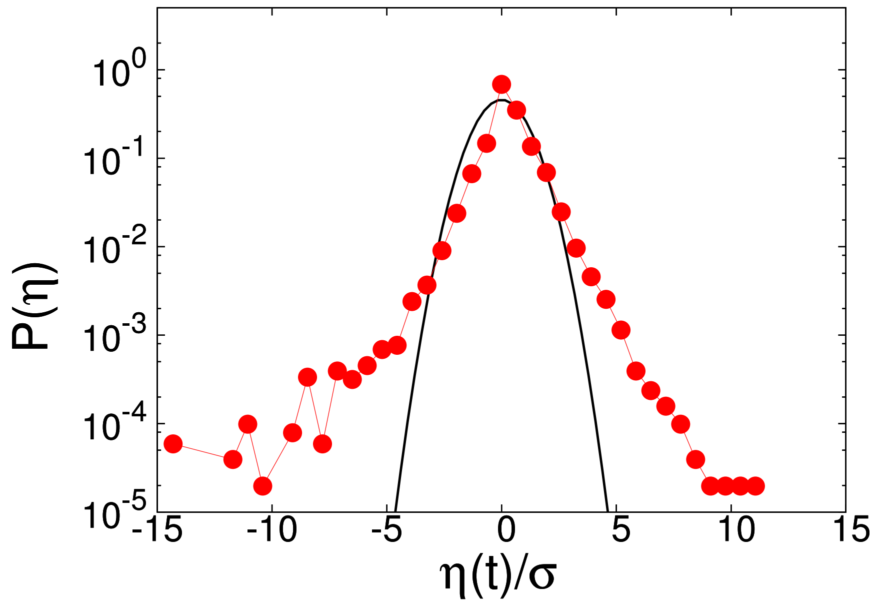


Figure 2.4: Empirical probability density function for high frequencies (~ 1 min) price differences of the WGR stock traded in the NYSE during a 80-days period. This semi-logarithmic plot shows the leptokurtic nature observed in empirical investigation. For comparison a Gaussian with the measured standard deviation is also shown.

The pioneering study of empirical data was performed by Mantegna and Stanley [113] who studied minute by minute data for the S$P500 index. Our study is similar to their original analysis. We form a series of high frequencies log-returns for an 80-days period for a given stock from NYSE.

$$\eta(t) = \ln X(t + \Delta t) - \ln X(t) \qquad \Delta t \sim 1 \text{ min} \tag{2.7}$$

In Fig. 2.4 we can see the leptokurtic nature of the log-returns distribution. In Chapter 3 we will discuss a set of different models, present in literature, to describe the dynamics of price fluctuations. Another way to study empirical data is to analyze the correlation properties. The direct method to study the correlations of some data is by mean of the autocorrelation function. The autocorrelation function of a discrete time series or a process Y(t) describes the correlation between the process at different points in time. If Y(t) has mean μ and variance σ^2, the definition of the linear autocorrelation function is:

$$C(\tau) = \frac{\mathbb{E}[(Y(t)-\mu)(Y(t+\tau)-\mu)]}{\sigma^2} \tag{2.8}$$

where t is a discrete time and $\mathbb{E}(\cdot)$ denotes the expectation value. For a sample time series of length n, $Y(1), Y(2)...Y(n)$ with known mean and variance, an estimate of the autocorrelation function may be obtained from:

$$\hat{C}(\tau) = \frac{1}{(n-\tau)\,\sigma^2}\sum_{t=1}^{n-\tau}[(Y(t)-\mu)(Y(t+\tau)-\mu)] \tag{2.9}$$

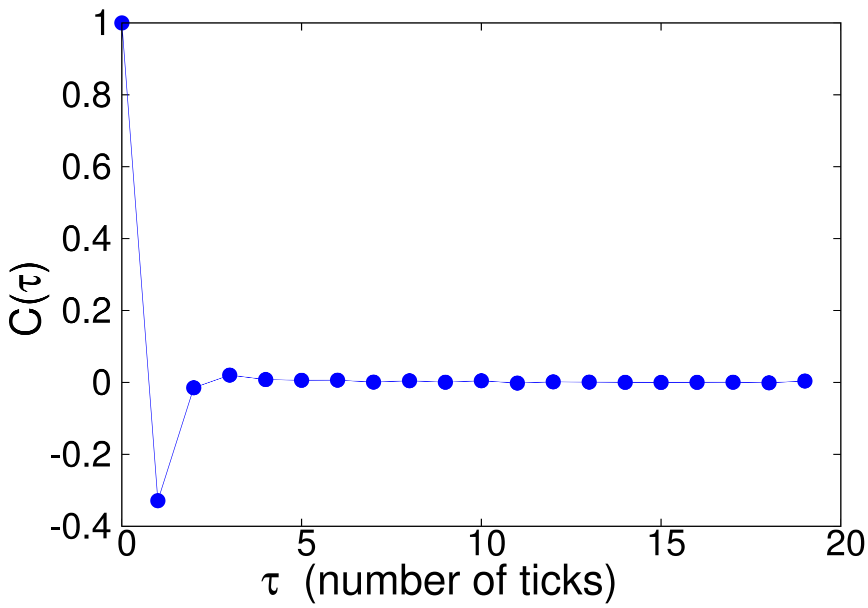


Figure 2.5: Linear autocorrelation function of tick by tick returns on GE shares traded on NYSE. The time scale is the number of transactions. We observe a negative autocorrelation for one trade.

Now we will study the correlation properties of the log-returns variable, $\eta(t)$. Here we consider the price variations for a trade.

$$\eta(t) = \ln X(t + \Delta t) - \ln X(t) \qquad \Delta t = 1 \text{ tick} \tag{2.10}$$

It is a well known fact that price movements in liquid markets do not exhibit any significant autocorrelation: the autocorrelation function of the log-returns, $C(\tau)$, rapidly decays to zero in few minutes. The absence of significant linear correlations in price increments has been widely studied [55, 103] and is often cited as support for the "efficient market hypothesis" [56]. In high-frequency log-return series of transactions prices, one actually observes a negative autocorrelation at a very short lags (typically one or a few trades) [43]. This is traditionally attributed to the bid-ask bounce [31]: transactions prices may take place either close to the ask or closer to the bid price and tend to bounce between these two limits. This feature can be attributed to the action of a market maker [66].

In Fig. 2.5 we show the linear autocorrelation function for a time series of tick by tick log-returns on GE index traded on the NYSE. The length of the time series is around 80 trading days which corresponds nearly to 450000 transactions. The absence of autocorrelation does not seem to hold systematically when the time scale Δt for calculating the log-returns is increased [43]. Anyway the available data-sets are inversely proportional to Δt, so the statistical evidences are less conclusive and more variable from sample to sample. The absence of autocorrelation in log-return gave some empirical support for "random walk" models of prices in which the returns are considered to be independent random variables. However it is well known that the absence of serial correlation does not imply the independence of the increments: independence implies that any nonlinear function of log-returns will also have no-autocorrelation [43]. For the real data this property does not hold.

Simple nonlinear functions of the log-returns, such as absolute log-returns, squared log-returns, exhibit significant positive autocorrelations. Unlike price changes that are correlated only on very short time scales, the absolute values of price log-returns show long-range power-law correlations on time scales up to a year or more [103, 104]. This is a quantitative signature of the well-known phenomenon of the *volatility clustering*. This means that large price variations are more likely to be followed by large price variations and vice versa. Fig. 2.6 illustrates this phenomenon for the same time series shown in Fig. 2.5, i.e. a tick by tick series of the index GE traded in the NYSE, for a period of 80 trading days. We can see that the autocorrelation function, $C(\tau)$, remain different from zero for all the time lags considered,

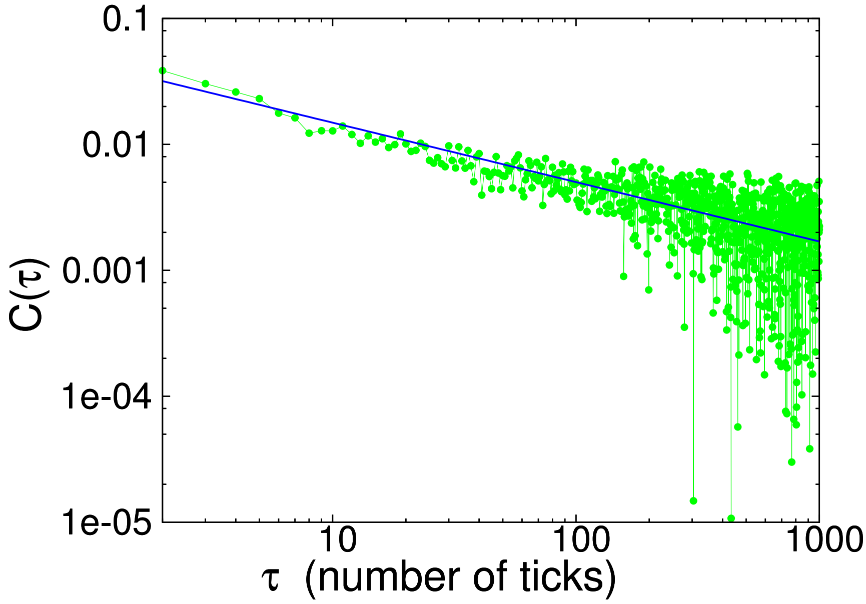


Figure 2.6: Nonlinear autocorrelation function for log-returns of the index GE averaged. The function plotted is the autocorrelation function of the absolute value of the the log returns. We can observe that this function remains different from zero for the entire time window considered. This phenomenon is called *volatility clustering*.

and slowly goes to zero. A power law fit on the data, disclose a scaling exponent $\gamma = -0.472 \pm 0.006$.

The phenomenon of volatility clustering has intrigued many researchers and oriented in a major way the development of stochastic models in finance. In Chapter 3 we will analyze ARCH and GARCH models that are intended to model the volatility clustering phenomenon.

2.5 Trading Volume

The trading volume represents the total number of shares traded for a given time frame. Volume is a measure of liquidity in a stock or index. The higher the volume, the more liquidity is present and the more competitive the market will be. Similarly to the trading activity, also the trading volume is not uniform during the trading hours. In Fig. 2.7 we can see the typical U-shaped intra-daily pattern of the trading volume. The trading volume is not constant during a trading day but displays a concave shape with a minimum at lunch-

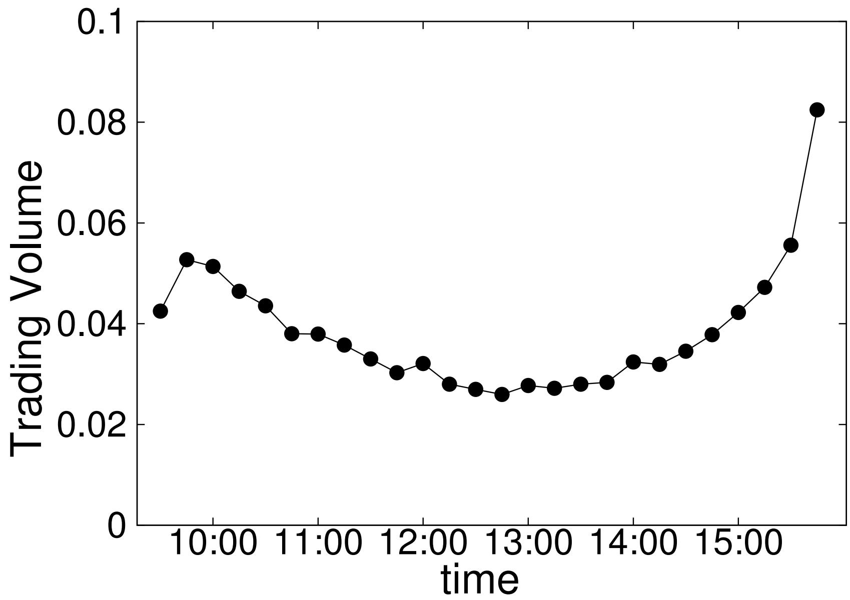


Figure 2.7: Trading volume during a trading day. We can observe that the volume is not uniform, having a minimum at lunch-time. The data correspond to an average over 20 stocks from NYSE for a period of 80 trading days.

time. Although trading volume is, like price, an important quantity that characterizes the activities of financial markets, only a few attempts have so far been made to understand the statistical properties of trading volume [68]. If we look to the statistical properties of the fluctuations of trading volume, we can observe some properties that are very similar to those of stock-prices fluctuations: non-Gaussian probability density function, absence of linear correlations and presence of non-linear long-range correlations. This suggests a correlation between trading volume and price fluctuations. In [61] is argued that large price movements are associated with higher subsequent volumes. So price changes lead to volume movements.

Now we investigate the statistical properties of the data of trading volume using a 80-trading days database from NYSE. We define $V(t)$ to be the traded volume of the transaction occurred at time t. We have analyzed the increments of volume:

$$\Delta V(t, \Delta t) = V(t + \Delta t) - V(t) \qquad \Delta t = 1 \text{ tick} \tag{2.11}$$

In Fig. 2.8 are plotted the trading volume fluctuations for the stock GE over a period of 80 trading days. We can observe the non Gaussian and non

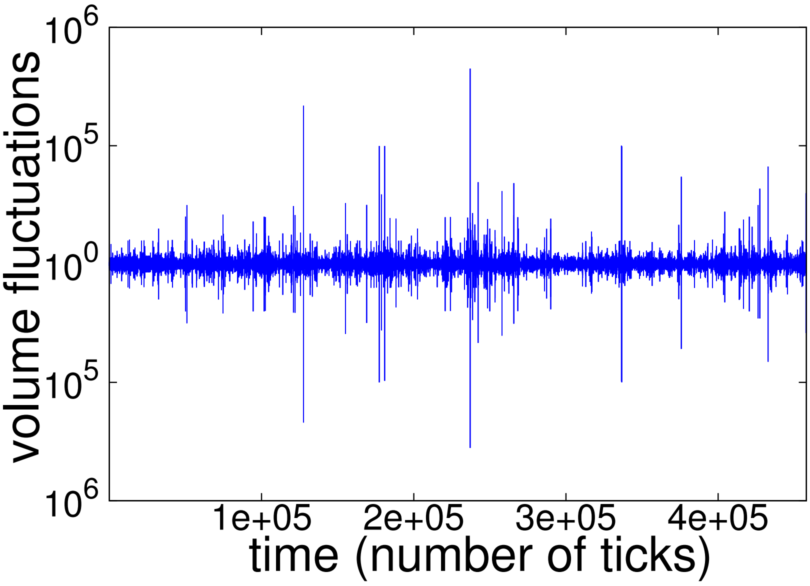


Figure 2.8: In this figure is plotted the time series of trading volume fluctuations of the stock GE for a period of 80 trading days without overnight, weekend and holidays closing. The plot shows the presence of large fluctuations that are strikingly non-Gaussian.

stationary nature of the trading volume fluctuations. The times series of trading volume fluctuations shows large fluctuations that are strikingly non-Gaussian. In Fig. 2.9 is shown the probability distribution of the increments of trading volume ΔV for the stock GE. At a glance, the difference with a Gaussian distribution can be appreciated. The mean value is close to zero and the distribution shows a positive kurtosis. The probability distribution of trading volume fluctuations exhibits apparently greater probability mass in the tails than in the center.

Regarding the correlation properties of trading volume fluctuations, we can see in Fig. 2.10 that the linear autocorrelation function is approximately zero. Instead absolute trading volume fluctuations are correlated over a long time period. The autocorrelation function seems to be well fitted by a power law function, disclosing a long range correlation in the trading volume absolute fluctuations. We can also observe, comparing Fig. 2.10(b) and Fig. 2.6, that trading volume and volatility show the same type of "long memory" behavior [79]. The scaling exponent, estimate by a power law fit, is $\gamma = -0.431 \pm 0.003$. In his article [42], Clark proposed the idea that the leptokurtic nature of prices log-returns, can be explained by the fact that

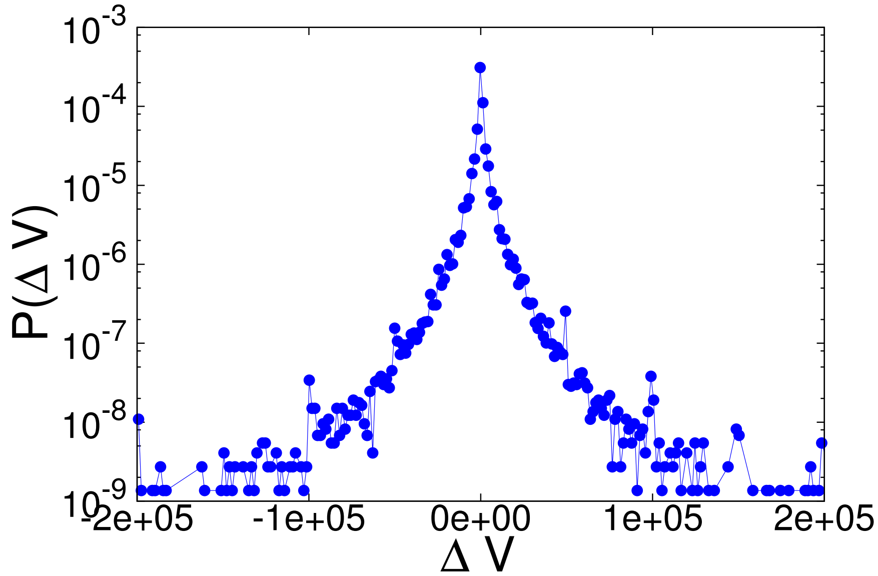
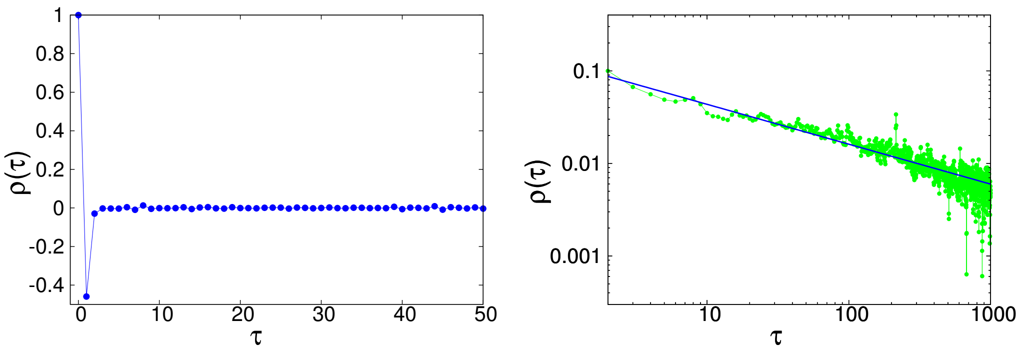


Figure 2.9: Probability density distribution of trading volume tick-by-tick increments for a 80-days period. We can observe that the distribution shows excess kurtosis, disclosing a non-Gaussian behavior.

trading activity is not uniform during time. Clark assumed that trading volume is a plausible measure of the evolution of price dynamics. The model is based on the assumption that the time in which a transaction occurs are themselves a stochastic process and he used the cumulative distribution of trading volumes up to time t as the probability distribution to have a transaction at time t. In this picture the probability distribution of prices increments is subordinate to the one of trading volumes. In this way Clark was able to prove that the distribution of prices increments is a leptokurtic distribution with all the moments finite. This model has quantitatively pointed out the strict connection between trading volume fluctuations and price dynamics and hence the importance of a careful analysis of trading volume statistical properties to understand and model the dynamics of price fluctuations.

2.6 Bid and Ask

Financial market investors who wish to buy or sell assets, do so by contacting their broker. There are different kind of requests an investor can make. The first type of request is the so called *market order*, i.e. an order to buy or



(a) Volume fluctuations (b) Volume absolute fluctuations

Figure 2.10: The autocorrelation functions for the trading volume fluctuations, (a), and trading volume absolute fluctuations, (b), are shown. The time τ is measured in ticks (transactions). The volume fluctuations (a) seem to be anti-correlated for few ticks and then the correlation drops to zero. The autocorrelation function for the absolute trading fluctuations (b) behaves like a power law, disclosing a long range correlation.

sell a stock at the current market price. When a market order is placed, it is almost guaranteed that the order will be executed. Ultimately, however, this depends on whether or not there is a willing buyer or seller. The second type of request is the *limit order*. A limit order specifies the maximum (minimum) price at which an investor is willing to buy (sell) a certain number of shares. A market order is usually less expensive than a limit order, but one disadvantage of a market order is that the price is paid when the order is executed. The price may not always be the same as that presented by a real-time quote service. This often happens when the market is changing very quickly. Placing an order "at the market,"especially when it involves a large number of shares, offers a greater chance of getting different prices for different parts of the whole order.

The list of all buy and sell limit order, with their corresponding price and volume, at a given instant of time, is the *order book*. At a given instant of time, all limit buy orders are below the best buy order, called the *bid price*, while all sell orders are above the best sell order, called the *ask price*. Therefore at any instant two prices are quoted: the bid price and the ask price, which are called the *quotes*. When arrives a new buy order below the bid price or sell order above the ask price, it adds to the queue at the corresponding price. If someone requests for a market order to buy or with a limit price above the current ask price, he will then make a trade at the ask price with the person with the lowest ask. The corresponding price is printed as the *transaction price*. The activity of the market is therefore a succession of quotes and trades.

The person, or the firm, which manage the order book is the *market maker*. Market makers must be compensated for the risk they take: for example, if he buys your shares in IBM then IBM's stock price begins to fall before a willing buyer has purchased the shares, the market maker will lose money. To prevent this, the market maker maintains a spread (*bid-ask spread*) on each stock he covers. Using our previous example, the market maker may purchase your shares of IBM from you for $100 each (ask price) and then offer to sell them to a buyer at $100.05 (bid price). The difference between the ask and bid price is only $0.05, but by trading millions of shares a day, he's managed to pocket a significant chunk of change to offset his risk. The market maker must then update his calculation of the excess demand in the market, by adding up all the buy orders in the order book and subtracting all the sell orders. Due to the revised calculation of the excess demand, the market maker may then consequently wish to move the price of the asset.

The study of the order book is very interesting both for academic and practical reasons [29], in fact provides information on the processes of trading and price formation. In our dataset, see Sec. 2.1, complete data on the order book is not available, so we could not perform any statistical analysis on this kind of data. Systematic investigation of these datasets have been done [29, 21, 95, 96], and has motivated a number of interesting theoretical works [38, 16, 40, 80, 112, 47]. Some of the main results are [29]:

- the price at which new limit orders are placed is very broadly (power-law) distributed around the current bid/ask;

- the average order book has a maximum away from the current bid/ask, and a tail reflecting the statistics of the incoming orders, see Fig. 2.11;

- the distribution of volume at the bid (or ask) follows a Gamma distribution.

As already said, our dataset does not contain the information about the entire order book, but only gives, at any given time, the quotes, that are the bid price and the ask prices, together with the corresponding sizes (volumes). Therefore we was able to perform a statistical analysis of the dynamics of the midpoint between the bid price and the ask price. This gives an estimate of the "fair" value of the transaction price at the next step. The midpoint

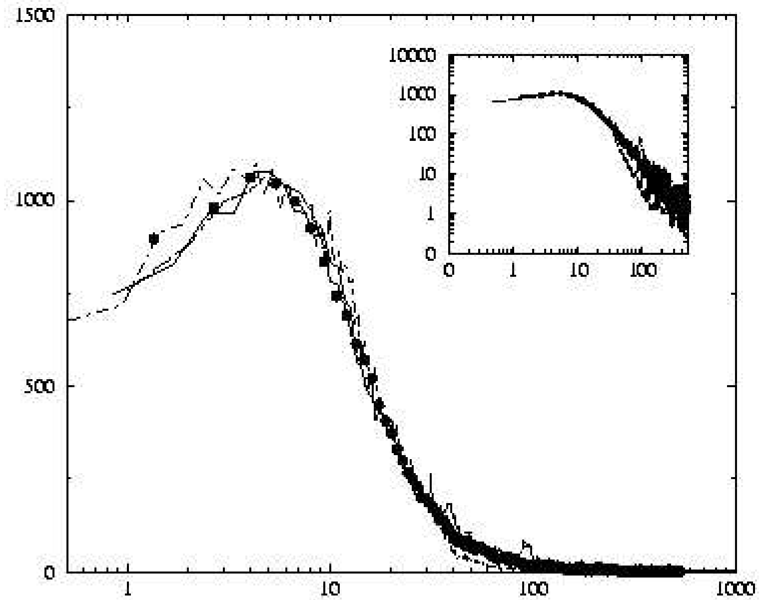


Figure 2.11: Average volume of the queue in the order book for three stocks, as a function of the distance from the current bid (or ask) in a log-linear scale. The axis have been rescaled in order to obtain a data collapse. In the inset are plotted the same data in log-log scale. We can see that the size of the queue reaches a maximum away from the best offered price. This results from the competition of two effects: the order flow is maximum around the current price but an order very near to the current price has a larger probability to be executed and disappear from the book. Figure from [29].

value is defined as:

$$M(t) = \frac{a(t) + b(t)}{2} \tag{2.12}$$

where $a(t)$ and $b(t)$ are the ask and bid prices. Similarly to the analysis performed for the stock price, we define a log-return for the midpoint price $M(t)$:

$$\mu(t) = \ln M(t + \Delta t) - \ln M(t). \tag{2.13}$$

We have studied the statistical properties of the autocorrelation function for the log-returns and the absolute log-returns of the midpoint price. In

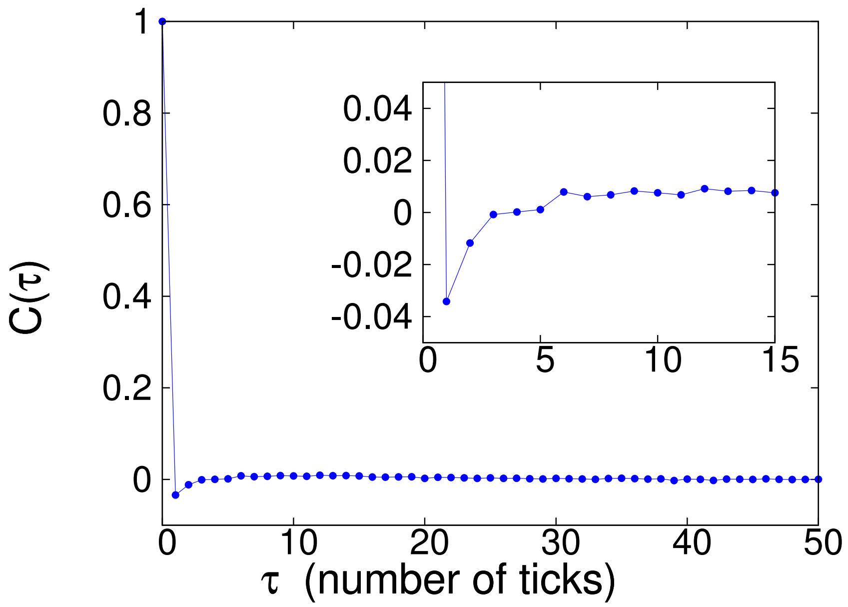


Figure 2.12: Autocorrelation of the log-return of the midpoint price fluctuations. We can observe an anti-correlation for few trade steps. In the inset are shown the same date in a magnified zone. Comparing this figure with Fig. 2.5 it is possible to observe that the level of correlation of the midpoint price fluctuation is an order of magnitude smaller than that one of the real price fluctuations.

Fig. 2.12 is plotted the autocorrelation function for the log-returns of the midpoint price $M(t)$ for the stock GE from NYSE calculated on an 80-trading days period. We can observe an anti-correlation of the midpoint price log-return, that persists for a few trades. Comparing this figure with Fig. 2.5 we can see that the level of anti-correlation is one order of magnitude larger for the price log-returns than for the midpoint log-returns. This implies that the prices log-returns tend more likely to invert sign from one step to another and the resulting signal for the price looks like a zigzag. The dynamic of midpoint price is instead smoother than that of the real price, this is probably due to the fact that the midpoint price is only the fair expectation for the price. The position of the price fluctuates in the bid-ask spread being sometimes nearer to the bid price and sometimes nearer to the ask price. In some rare case, when the price rapidly drop or rise, the price can have a value smaller than the bid price or larger than the ask price.

We have also studied the autocorrelation function of the absolute midpoint log-returns, as is shown in Fig. 2.13. In Fig. 2.13 we can observe the

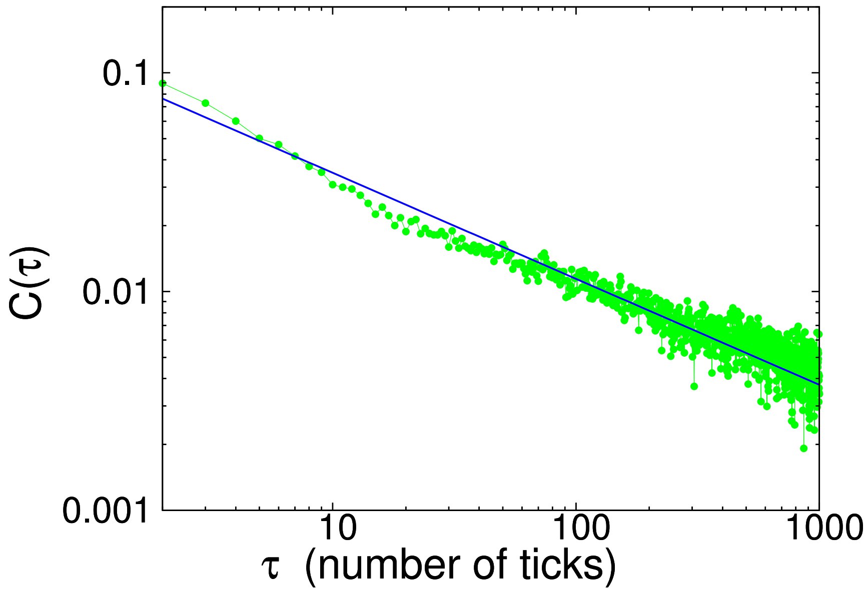


Figure 2.13: Autocorrelation function of the absolute log-return of the midpoint price fluctuations. Both axis are in a logarithmic scale. We can observed that the autocorrelation function shows a power law behavior disclosing long range correlations in the process.

power law behavior of the autocorrelation function of the absolute log-returns of the midpoint price. From a theoretical point of view this suggests a long-range correlation. The scaling exponent, estimated by a power law fit, is $\gamma = -0.485 \pm 0.002$. We have seen that those statistical properties of the prices fluctuations hold also for the midpoint price fluctuations, nevertheless this last has a smoother profile. In Fig. 2.14 is shown a profile of price and midpoint price for a nearly fifteens minutes time interval for a stock from NYSE. We can observe that the value of the price is almost never equal to the midpoint price and oscillates up and down the midpoint line. Instead the midpoint price is smoother and often remains constant for some time steps.

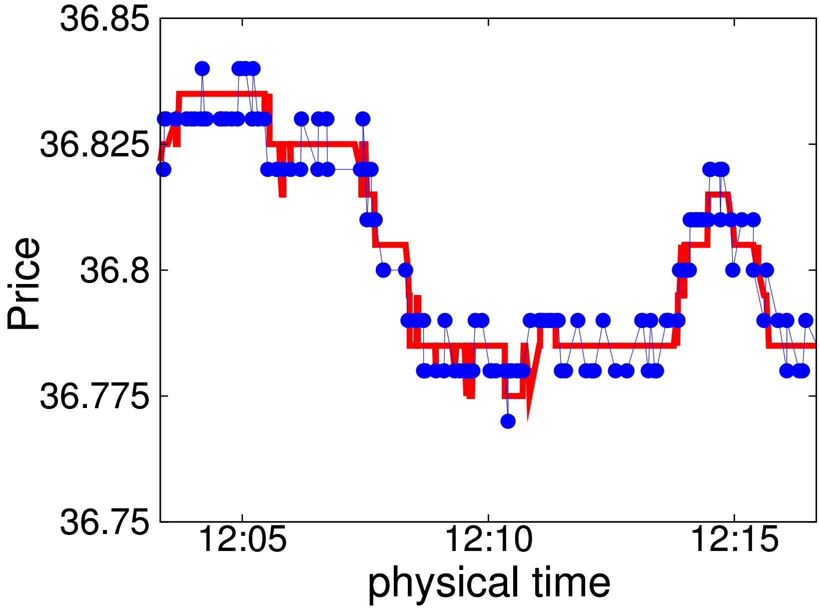


Figure 2.14: Time evolution for the real price (circle with blue line) and for midpoint price (red line). We can see that the real price dynamics is much oscillating while the midpoint price remains often constant.

Chapter 3

Theoretical Models

In Chapter 2 we have discussed the statistical properties of financial time series. This analysis has revealed a wealth of interesting stylized facts which seem to be common to a wide variety of markets, instruments and periods.

These properties has incited many researchers to develop models which can explain and reproduce these characteristics. Most of the work in modeling financial markets is addressed to describe the statistical properties of the time evolution of the price. In this regard, one of the main task is to find which is the stochastic model of stock price dynamics.

The standard theory of finance provides that the logarithm of a stock price executes a random walk. This means that price changes are uncorrelated in time and are drawn from a normal distribution.

While the first assumption is rather well satisfied, deviations from a normal distribution will lead to consider another class of stochastic processes. One of the most confirmed is the truncated stable Lévy process whose probability distribution function shows fat tails and which describes financial data much better than a normal distribution. This stable stochastic process was first introduced to model price dynamics in [87].

Both in random walk model and truncated stable Lévy process, volatility is assumed to be a constant. This is true over rather short time scales, at best. Empirical volatilities vary strongly with time and suggest considering volatility as a stochastic variable. This fact has led to the development of the ARCH and GARCH models [52, 24].

An alternative to model the price dynamic as a stochastic process is to develop highly simplified models of strategic interaction to capture the essence of the collective behavior in a financial market. In this case the prices dynamics is the result of the interaction of agents, whose behavior is modeled. One of the most famous agent based model is the "El Farol Bar" problem [10]. This model was simplified and abstracted by Challet e Zhang in a way that make it suitable to be studied with the methods of statistical mechanics. The new formulation of the problem is called “Minority Game” and is one of the main results of Physics applied to Economics [39].

Another very important theoretical result obtained by physicists studying Economics is the famous Black and Scholes equation for a rational option pricing [22]. We will see the main features of this model together with its limitations.

3.1 Random Walk Model

The first who modeled stock prices by a random walk was Bachelier [14] in 1900. Nowadays, the theoretical descriptions used in standard finance theory, are typically built around this assumption that asset prices follow some form of random walk. We therefore need to understand the details, and hence the limitations, of a random walk.

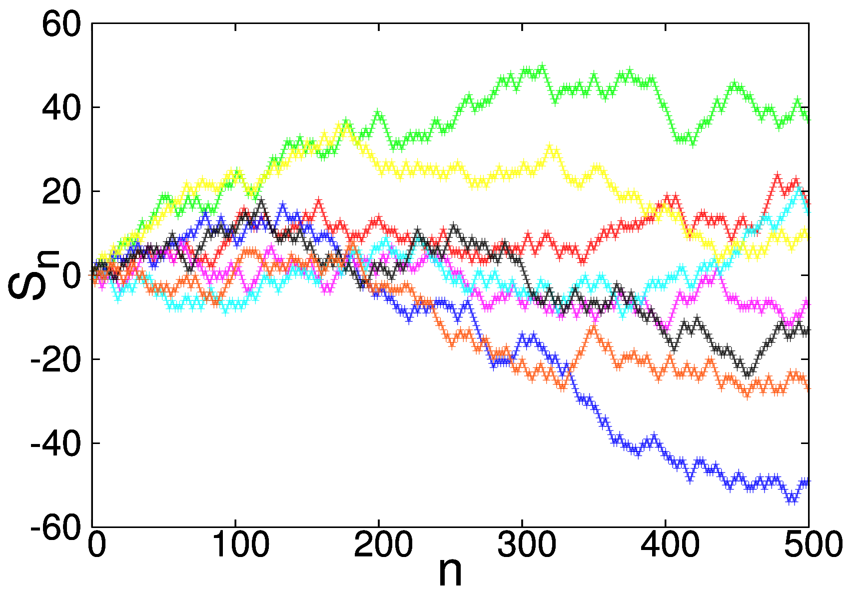


Figure 3.1: Eight different representation of a particle performing a random walk are plotted.

Suppose we are tossing a fair coin, hence heads and tails have equal probability of occurring. In this case p[heads] $= 1-q$[tails] $= 0.5$. In principle p could be different from q. These events can be thought to represent some increments X_i (price changes) in one time step Δt. For example we can denote $X_i = 1$ if at the $i-th$ time step the outcome of the game was head and $X_i = -1$ if it was tail. The successive partial sums $S_1, S_2, ..., S_n$ ($S_0 = 0$), where:

$$S_n = X_1 + X_2 + ... + X_n, \tag{3.1}$$

can be marked as point on a vertical axis and can represent the position of a particle performing a random walk [59]. In Fig. 3.1 are plotted some different representations of the time evolution of a particle performing a random walk. The probability of observing, in n events, α realizations of heads and $n - \alpha$ realizations of tails and hence to have $S_n = r = \alpha - (n - \alpha) = 2\alpha - n$, is given by the binomial distribution:

$$P(S_n = r) = \begin{pmatrix} n \\ \alpha \end{pmatrix} p^{\alpha}(1-p)^{n-\alpha}, \tag{3.2}$$

where is understood that the binomial coefficient is to be interpreted as zero unless α is an integer between 0 and n, inclusive. We can observe that in the case we are considering, $p = q = \frac{1}{2}$, the expected value of for the variable r is zero. In the limit $n \to \infty$ the binomial distribution tends to a Gaussian distribution:

$$p(r) = \frac{1}{\sqrt{2\pi npq}} \exp\Big(-\frac{r^2}{2npq}\Big). \tag{3.3}$$

For $p = q = \frac{1}{2}$ and setting $t = n\Delta t$, with Δt the unit time step, $r = x$,and $H = 4\Delta t$, we obtain:

$$p(x) = \frac{1}{\sqrt{2\pi\frac{t}{H}}} \exp\Big(-\frac{x^2}{2\frac{t}{H}}\Big). \tag{3.4}$$

So in this limit of large n one has passed from discrete time and discrete positions to continuous variables. To write the Gaussian in the standard form we set $\sigma^2 = \frac{t}{H}$, and we obtain:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{x^2}{2\sigma^2}\Big), \tag{3.5}$$

where σ^2 is the variance of the distribution. We can observe that the variance is proportional to the time, $\sigma^2 \sim t$ so the shape of the Gaussian distribution evolves with time becoming broader and broader as we can observe in Fig. 3.2.

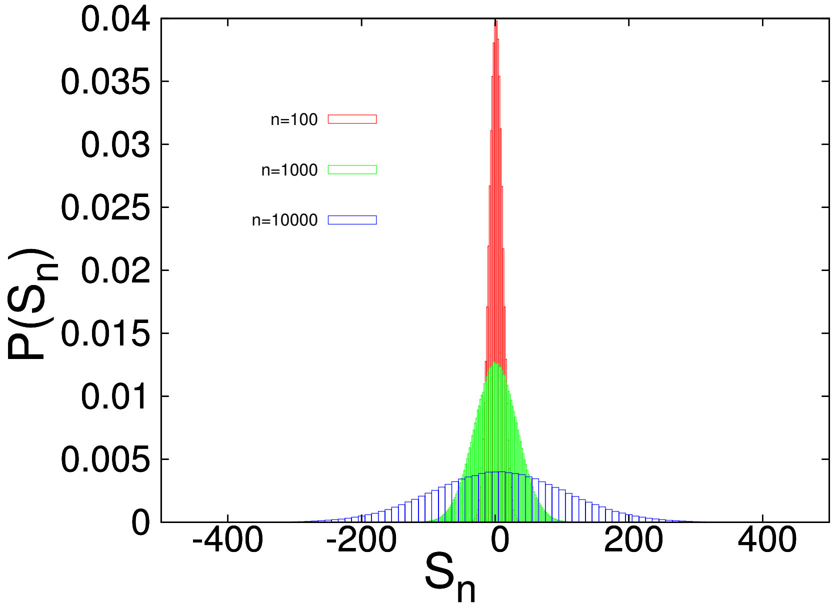


Figure 3.2: The probability density function for the variable S_n is plotted. The histogram is obtained performing 10^6 different realization of random walks of various size n. We can see that the pdf is a Gaussian that spreads its amplitude for increasing values of the time n.

Turning back to the time evolution of the walker we can write the following recursive equation:

$$S_n = S_{n-1} + X_n, \tag{3.6}$$

where X_i is an increment taking the value ± 1 with equal probability, therefore the increments X_i are independent and identically distributed random variables with zero mean and unit variance. This make sense for a coin-toss mechanism where successive coin-toss outcomes are indeed independent, and since we are using always the same coin, then the probability density function $\rho[X]$ is identical at every time step.

The Eq. 3.6 can be generalized substituting the increments X_i with a generic independent and identically distributed random variable with zero mean and unit variance drawn from a generic probability density function:

$$S_n = S_{n-1} + \xi_n. \tag{3.7}$$

In this perspective ξ_i can be Gaussian, exponential, uniform, etc. The random walk model for price asserts that the logarithm of prices performs a random walk. This is due to the fact that the quantity that is independent in time is not the price increments but the relative increments of price. Therefore the recursive equation for the logarithm of price is:

$$\ln P_t = \ln P_{t-1} + \xi_t \quad \text{and hence} \tag{3.8}$$

$$\ln \frac{P_t}{p_{t-1}} = \xi_t. \tag{3.9}$$

Substituting ξ with another random variable $\eta = \exp \xi$ we obtain the recursive equation for the prices:

$$P_t = P_{t-1}\eta_t. \tag{3.10}$$

Taking the limit of small price increments and small time steps in Eq. 3.8 we obtain a continuous stochastic differential equation for the logarithm of price $y = \ln P$:

$$dy = \xi, \tag{3.11}$$

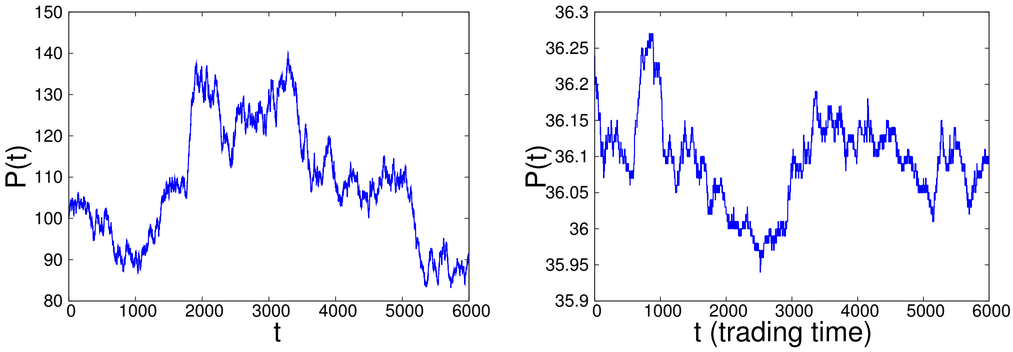
where ξ is a random variable taken for example from a Gaussian probability density function with zero mean and standard deviation proportional to $(dt)^{1/2}$. Physicists call equation 3.11 *Brownian motion* and use it to describe the random-walk dynamics of particles diffusing in a gas. Following the same steps we can obtain the stochastic equation for price taking the limit of small increments in Eq. 3.10:

$$\frac{dP}{P} = \xi. \tag{3.12}$$

Equation 3.12 is called *Geometric Brownian motion.*

In Fig. 3.3 we have plotted a price chart generated from a geometric Brownian motion. Also a comparison with the time evolution of the price of GE stock for a trading day time period is shown. We can observe that the similarity of the two behaviors is striking.

The random walk model for price is also consistent with the Efficient Market Hypothesis that we have discussed in subsection 1.3.1. In fact in this model is impossible to forecast the next outcome based on the past outcomes.



(a) Geometric Brownian motion (b) GE share index

Figure 3.3: Computer simulation of price chart as a geometric Brownian motion (a) and comparison with the evolution of the GE share index during a trading day period. At a glance we can observe that the similarity of the two behaviors is striking.

In other words, if we were betting in future outcomes, we would not be able to gain systematically over time. In Chapter 2 we have seen that real stock prices follow the random walk model only in first approximation. In fact, in the short time regime many deviations from the random-walk model appear, such as the tails in the distribution of returns. In the next section we will see a model for price dynamics which is able to describe the high frequency observations of price returns.

3.2 Lévy Distributions and Truncated Lévy Distributions

The most common stochastic model of stock price dynamics is the Gaussian behavior discussed above that assumes a geometric Brownian diffusion of the asset prices and a corresponding arithmetic Brownian motion of the logarithm of price differences. This model provides a first approximation of the behavior observed in empirical data.

However, the Gaussian probability distribution function is a universal consequence of the central limit theorem in the limit of long times on the condition that the financial market is in a stationary state. Indeed, the Gaussian law can possibly deviate considerably from the probability distribution function determined empirically for short timescales. In Chapter 2 we have observed serious systematic deviations from the Gaussian model predictions, which indicate that the empirically determined probability distributions exhibit a pronounced leptokurtic behavior. A highly leptokurtic function is characterized by a narrower and larger maximum and by fatter tails than in the Gaussian case. Obviously, the degree of leptokurtosis increases with decreasing time difference Δt between successive observations. Mandelbrot [84] in 1963 was perhaps the first person to challenge the paradigm that returns are normally distributed. He analyzed cotton prices on various exchanges in the United States and found evidences that their distribution of returns decays as a power law and hence much slower than a Gaussian. A lot of models had been developed in order to describe the short-time behavior of price fluctuations in terms of alternative probability functions. Among these models, a partial subset is the the Lévy stable model [84], the Student's t-distribution [23], the mixture of Gaussian distributions [42], the truncated Lévy flight [87], the jump diffusion model [98] and the hyperbolic-distributed stochastic process [50]. In this section we will expose the Lévy flight and truncated Lévy flight models because of their impact and acceptance in the scientific community.

Therefore, we will first introduce the so-called Lévy stable distributions and then the truncated Lévy flight distributions and their possible applications to financial data. To introduce the Lévy stable distributions it may be useful to recall the definition of characteristic function $\varphi(q)$ of a random variable x. Let $p(x)$ be the probability density function of the random variable x; hence the characteristic function of the stochastic process is the Fourier transform of $p(x)$:

$$\varphi(q) = \int_{-\infty}^{\infty} p(x) \exp(iqx) dx. \tag{3.13}$$

Now let x_1 and x_2 be two independent random variables with probability density functions $p_1(x_1)$ and $p_2(x_2)$. Since x_1 and x_2 are independent, the probability density function of the sum $x = x_1 + x_2$ is the convolution of the original probability density functions:

$$p(x) = \int_{-\infty}^{\infty} p_1(s) p_2(x-s) ds. \tag{3.14}$$

Let now $\varphi(q)$ denote the characteristic function of x. In view of the convolution theorem which states that the Fourier transform of the convolution is the product of the Fourier transforms, it then follows that

$$\varphi(q, 2) = \varphi_1(q) \varphi_2(q). \tag{3.15}$$

Suppose furthermore that x_1 and x_2 are identically distributed. In this case

$$\varphi_1(q) = \varphi_2(q) = \varphi(q). \tag{3.16}$$

From equations 3.15 and 3.16 follows that

$$\varphi(q, 2) = [\varphi(q)]^2. \tag{3.17}$$

In general, if $S_n = \sum_{i=1}^{n} x_i$, where the x_i's are independent and identically distributed (i.i.d) random variables, then:

$$\varphi(q, n) = [\varphi(q)]^n, \tag{3.18}$$

from which the probability density function of S_n can be obtained by calculating the inverse Fourier transform.

Note that the probability density function of the sum of n i.i.d. random variables will in general be quite different from the probability density function of the individual variables. There is however a special class of distributions, the stable distributions, for which the probability density function of the sum has the same functional form of the individual probability density functions.

Definition 1. [117] *A probability distribution $p(x)$ is stable if for each $n \geq 2$ there exist numbers $a_n > 0$ and b_n, such that, if $x_1, x_2, ..., x_n$ are i.i.d. random variables with distribution p(x), then S_n and the variable $a_n x_i + b_n$ are identically distributed for every i.*

In other words, a distribution is stable if its form is invariant under addition, up to a rescaling of the variable by a translation and a dilation. The Gaussian distribution is stable. To see this, we recall that the Fourier transform of a Gaussian is a Gaussian and that the product of Gaussian is again a Gaussian, so from Eq. 3.18 it follows that the characteristic function $\varphi(q, n)$ will indeed be that of a Gaussian variable.

The class of stable distributions is rather small and was completely determined by the mathematicians P. Lévy [77] and A.Ya. Khintchine [76] in 1920's. They found that, restricting ourselves to the subclass of symmetric distributions, the most general form of a characteristic function of a stable process is:

$$\varphi_\alpha(q) = \exp(-\gamma |q|^\alpha), \tag{3.19}$$

where $0 < \alpha \leq 2$ is called the *stability exponent* and γ is a positive *scale factor*. Taking the inverse of $\varphi_\alpha(q)$ we obtain the corresponding probability density function $\varphi_\alpha(x)$:

$$p_\alpha(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \varphi_\alpha(q)\exp(-iqx)dq = \frac{1}{\pi}\int_0^{\infty} \exp(-\gamma q^\alpha)\cos(qx)dq. \quad (3.20)$$

The integral in equation 3.20 can be explicitly calculated only for two particular values of α for which the analytical form of the Lévy stable distribution is known:

- $\alpha = 1$ (Lorentzian or Cauchy distribution):

$$p(x) = \frac{2a}{\pi}\frac{1}{x^2 + 4a^2}; \quad (3.21)$$

- $\alpha = 2$ (Gaussian distribution):

$$p(x) = \frac{1}{\sqrt{2\pi a^2}}\exp(-\frac{x^2}{2a^2}). \quad (3.22)$$

Although, for arbitrary α the probability density function $p_\alpha(x)$ cannot be found in closed form, its asymptotic behavior for large x can be easily calculated from Eq. 3.20. Here one finds [111, 20] that:

$$p_\alpha(x) \sim \frac{\Gamma(1+\alpha)\sin(\pi\alpha/2)}{\pi|x|^{1+\alpha}} \sim |x|^{-(1+\alpha)}, \qquad |x| \to \infty. \quad (3.23)$$

We thus see that the Lévy distribution for $\alpha < 2$ has the interesting property that it shows scaling behavior for large x, i.e., $p(x)$ decays as a power-law. The power-law decay of Lévy distributions implies the absence of a characteristic scale. The downside of this is that all Lévy distributions with $\alpha < 2$ have *infinite* variance. In fact all moments of order higher than one are infinite, since $\mathbb{E}[|x|^k]$ diverges for $k \geq \alpha$, as can be shown from Eq. 3.23. Now we show that the Lévy stable distributions are self-similar [88]. This property can be obtained as follow. One can consider the probability of return to the origin $p(S_n = 0)$ starting from the the characteristic function for S_n:

$$\varphi(q, n) = \exp(-n\gamma|q|^n). \quad (3.24)$$

Performing the Fourier transform:

$$p(S_n) = \frac{1}{\pi}\int_0^\infty \exp\left(-n\gamma|q|^\alpha\right)\cos\left(qS_n\right)dq. \tag{3.25}$$

Hence

$$p(S_n = 0) = \frac{1}{\pi}\int_0^\infty \exp\left(-n\gamma|q|^\alpha\right)dq = \frac{\Gamma(1/\alpha)}{\pi\alpha(\gamma n)^{1/\alpha}}. \tag{3.26}$$

Then the $p(S_n)$ is properly rescaled by defining

$$\tilde{p}(\tilde{S}_n) = p(S_n)n^{1/\alpha}. \tag{3.27}$$

The normalization

$$\int_{-\infty}^{\infty} \tilde{p}(\tilde{S}_n)d\tilde{S}_n = 1, \tag{3.28}$$

is assured if:

$$\tilde{S}_n = \frac{S_n}{n^{1/\alpha}}. \tag{3.29}$$

When $\alpha = 2$ the scaling relations coincide with the standard normalization of a Gaussian variable by mean of $n^{1/2}$ (standard deviation).

In section 3.1 we have discussed the central limit theorem which asserts that the limit distribution of a variable that is the sum of independent and identically distributed random variables with finite variance, is the Gaussian stable distribution. This theorem is only a special case of a much more general theorem. The generalized central limit theorem [64, 65] affirms that if one consider a random variable that is the sum of i.i.d random variable with infinite variance then the asymptotic probability distribution is a Lévy stable distribution.

This property together with the self-similarity (stability) and the power-law asymptotic behavior have led to the development of the Lévy stable process as a model for price dynamics. This model was first proposed by Mandelbrot [84] in 1963 to describe the leptokurtotic behavior of the distribution of returns for a series of cotton prices. Also Fama [54] in 1965 used the Lévy stable process to model the price dynamics in his analysis of stock prices in the NYSE.

We have seen that the Lévy processes with $\alpha < 2$ have infinite variance. Processes with infinite variance are not physically plausible and a real Lévy distribution is not observed in financial data. To obtain a distribution with finite variance, several prescriptions to *truncated* Lévy distributions have been proposed. In particular we refer to the model proposed by Mantegna and Stanley in [87]. A truncated Lévy distribution is defined as:

$$p(x) = \begin{cases} 0 & x > l \\ c\, p_\alpha(x) & -l \leq x \leq l \\ 0 & x < -l, \end{cases} \tag{3.30}$$

where $p_\alpha(x)$ is the symmetrical Lévy stable distribution defined in equation 3.20.

Since a truncated Lévy distribution has finite variance, then by the central limit theorem the distribution of the sum of n i.i.d. variable with such a distribution will converge to a Gaussian distribution for large n. However this convergence is usually very slow and depends from α and l. In [87] is shown that the crossover, n_x form the Lévy to the Gaussian regime is given by:

$$n_x \sim A\, l^\alpha, \tag{3.31}$$

where A is an α-dependent factor.

Mantegna and Stanley performed a systematic investigation of the scaling behavior of the American S&P500 index [113]. They analyzed high-frequency data over the period from January 1984 to December 1989. Index changes $\Delta X(t, \Delta t) = X(t = \delta t) - X(t)$ have been determined over different timescales Δt ranging from 1 to 1000 minutes ($\sim$ 16 hours). If these data are drawn from a Lévy stable distribution, they should show a characteristic scaling behavior, i.e., one must be able, by a suitable change of scale, to collapse them onto a single master curve. So they rescale the variable and probability distribution according to equations 3.27 and 3.29:

$$\widetilde{\Delta X} = \frac{\Delta X}{(\Delta t)^{1/\alpha}} \quad \text{and} \quad \tilde{p}(\widetilde{\Delta X}) = \frac{p(\Delta X)}{(\Delta t)^{-1/\alpha}}. \tag{3.32}$$

They then computed the empirical probability density function $p(\Delta X, \Delta t)$ and analyzed the scaling of $p(0, \Delta t)$ with Δt. In log-log plot $p(0, \delta t)$ showed a linear behavior, as predicted by the Eq. 3.26, with a slope corresponding to $\alpha = 1.4$ for $30 < \Delta t < 10000$ minutes. For $\Delta t > 10^4$ the slope of $p(0, \Delta t)$

approaches -0.5 indicating convergence to a Gaussian behavior. The value $\alpha = 1.4$ was used to rescale the variables and distributions and indeed the data approximately collapse onto a single distribution.

In the Lévy regime (i.e. small Δt), however, the tail of their empirical probability density function decays slower than a Gaussian but faster than a pure Lévy distribution with the exponent α found from the scaling argument.

Analyzing the data, it is clear that Gaussian provides bad description of the data. The Lévy distribution is much better, especially in the central part of the distribution. For very large index fluctuations $\Delta t \geq 8\sigma$, the Lévy distribution seems to somewhat overestimate the frequency of such extremal events. These facts thus suggest that a truncated Lévy distribution would perhaps be more appropriate for modeling the actual distribution.

Indeed, Bouchaud and Potters [30] found that the probability of 15-minutes changes of the S&P500 index is well described by an exponentially truncated Lévy distribution with $\alpha = 1.5$.

Many other applications of Lévy processes in Finance have been discussed in the literature [109].

3.3 ARCH and GARCH processes

All the models described until now are stochastic processes with constant parameter and in particular with constant variance. In section 2.4 we have seen that this assumption is not true for empirical data where the variance (volatility) is a parameter that fluctuates as a function of time.

To overcome this problem Engle [53] proposed a stochastic process characterized by a time dependent variance. The phenomenon that indicates the fact the the variance is time dependent is called *heteroscedasticity*. In fact the model described by Engle is called ARCH that is the acronym of Auto-Regressive Conditional Heteroscedasticity, namely a stochastic process with non constant variance conditional on the past, but constant unconditional variance.

Before describing the ARCH process and its generalization GARCH process, let us explain a peculiarity of the processes with time dependent variance [30].

Let us consider a random process with a time dependent variance and construct the series of increments η_k. If the process is time dependent every increment η_i is drown from a different probability density function. So we have $p_1(\eta_1), p_2(\eta_2), ..., p_n(\eta_n)$ that are not all identical. Suppose that the distributions p_k varies sufficiently slowly that one can measure some of its moments (for example its mean and its variance) over a time scale which is long enough to allow a precise estimation of these parameters but short if compared to the time scale of the variation of p_k. Suppose for example that $p_k(\eta_k)$ is a Gaussian distribution with variance σ_k that is itself a random variable. So there are two kind of average: the average over the random variable σ_k ($\bar{\cdot}$) and the average over the probability distribution p_k ($< \cdot >$). In this case the empirical histogram of the variables $\eta_1, ..., \eta_k$ leads to an apparent distribution $\overline{p}$ that is non-Gaussian also if all the p_k are Gaussian. In fact, if we perform the average over the σ_k (in the continuous limit), we obtain:

$$\overline{p}(\eta) = \int p(\sigma) \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{\eta^2}{2\sigma^2}) d\sigma. \tag{3.33}$$

If now we calculate the kurtosis of $\overline{p}(\eta)$:

$$\overline{\kappa} = \frac{\overline{< \eta^4 >}}{(\overline{< \eta^2 >})^2} - 3 = 3\Big(\frac{\overline{\sigma^4}}{(\overline{\sigma^2})^2} - 1\Big). \tag{3.34}$$

For any random variable one has $\overline{\sigma^4} \geq (\overline{\sigma^2})^2$, where the equality is true only if σ is not time dependent. This means that the kurtosis is always positive, hence volatility fluctuations lead to fat tails.

Now we are going to describe the main features of the ARCH process [52]. Autoregressive conditional heteroscedastic (ARCH) processes are a form of stochastic process that are widely used in finance and economics for modeling conditional heteroscedasticity (time-varying volatility) and volatility clustering. First proposed by Engle [53], ARCH model considers the variance of the current step to be a function of the variances of the previous time steps. Here we consider the general ARCH(p) process where the variance at each time step depends on the variance of previous p steps. Specifically, let η_i be the increment and assume that:

$$\eta_i = \sigma_i \epsilon_i, \tag{3.35}$$

where ϵ is a Gaussian (but not necessarily Gaussian) random variable with zero mean and unit variance. Hence η is a random variable with zero mean and variance σ_i^2, characterized by a Gaussian conditional probability density function $f_i(\eta)$. The original ARCH(p) model has the variance equation:

$$\sigma_i^2 = \alpha_0 + \alpha_1 \eta_{i-1}^2 + \alpha_2 \eta_{i-2}^2 + ... + \alpha_p \eta_{i-p}^2 = \alpha_0 + \sum_{k=1}^{p} \alpha_k \eta_{i-k}^2, \tag{3.36}$$

where $\alpha_0, \alpha_1, ..., \alpha_p$ are positive random variables. By varying the number p of terms in Eq. 3.36, one can control the degree of memory of the variance σ_i^2.

The ARCH(p) process is particular case of weighted moving average process, over the variance fluctuations.

This model captures the conditional heteroscedasticity of stock prices fluctuations by using a moving average of the past squared errors: if a major market movement in either direction occurred at $m \leq p$ time steps ago, the error square will be large, and assuming its coefficient is nonzero, the effect will be to increase the actual variance. In other words, large fluctuations tend to follow large fluctuations of either sign. Mandelbrot [84] had first described this phenomenon, which today we call conditional heteroscedasticity.

The unconditional variance of an ARCH process is simply the average over the conditional variance using the probability distribution function of the corresponding ARCH(p) model.

Because ϵ is an i.i.d. random variables with zero mean and unit variance, we have:

$$< \eta_i \eta_j >= \delta_{ij}, \tag{3.37}$$

where δ_{ij} is the Kroeneker delta function ($\delta_{ij} = 1$ if $i = j$, 0 otherwise). Therefore the autocorrelation function of the increments η_k is given by:

$$\overline{< \eta_i \eta_j >} = \overline{\sigma_i \sigma_j} < \epsilon_i \epsilon_j >= \delta_{ij} \overline{\sigma^2}. \tag{3.38}$$

In particular we obtain:

$$\overline{\eta_k^2} = \overline{\sigma_k^2}. \tag{3.39}$$

Furthermore, the stationarity requires $\overline{\sigma_k^2} = \overline{\sigma^2}$ for all k. Hence, we obtain from equation 3.36:

$$\overline{\sigma^2} = \frac{\alpha_0}{1 - \sum_{k=1}^{p} \alpha_k}. \tag{3.40}$$

Therefore, to obtain a finite and positive unconditional variance the coefficient α_k must satisfy the constraint

$$\sum_{k=1}^{p} \alpha_k < 1. \tag{3.41}$$

For the sake of simplicity, we focus now on the ARCH(1) process:

$$\sigma_i^2 = \alpha_0 + \alpha_1 \eta_{i-1}^2. \tag{3.42}$$

If we consider the sum variable S_n, defined as:

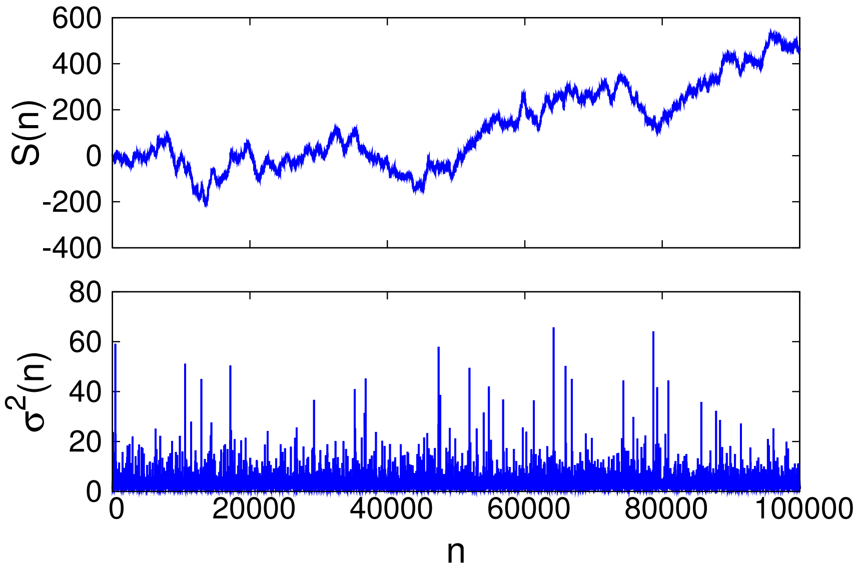


Figure 3.4: In this figure we have performed a numerical simulation of an ARCH(1) process. The parameters chosen are $\alpha_0 = 0.45$ and $\alpha_1 = .55$ while the random noise is Gaussian with zero mean and unit variance. In the top panel is plotted the time evolution of a the "price" S(n) and in the bottom panel is plotted the conditional variance $\sigma^2(n)$.

$$S_n = \sum_{i=1}^{i=n} \eta_i, \tag{3.43}$$

we can obtain the time evolution of the "price" of an ARCH(1) process. In Fig. 3.4 we have plotted the time evolution of an ARCH(1) process with parameters $\alpha_0 = 0.45$ and $\alpha_1 = 0.55$, together with the respective time evolution of the conditional variance σ_i^2.

In this case of ARCH(1) process, the unconditional variance, from equation 3.40, is given by:

$$\overline{\sigma^2} = \frac{\alpha_0}{1-\alpha_1}. \tag{3.44}$$

Furthermore, we get from the relation $<(\epsilon)^4>=3$ so that we find:

$$\overline{\sigma^4} = (\alpha_0)^2 + 2\alpha_0\alpha_1\overline{\sigma^2} + 3\alpha_1^2\overline{\sigma^4}. \tag{3.45}$$

Then the kurtosis of the ARCH(1) process is:

$$\kappa = \frac{\overline{<\eta^4>}}{\overline{<\eta^2>}^2} = \frac{3\overline{\sigma^4}}{(\overline{\sigma^2})^2} = 3 + \frac{6\alpha_1^2}{1-3\alpha_1^2}, \tag{3.46}$$

which is positive and finite for $0 \leq \alpha_1 \leq 1/\sqrt{3}$.

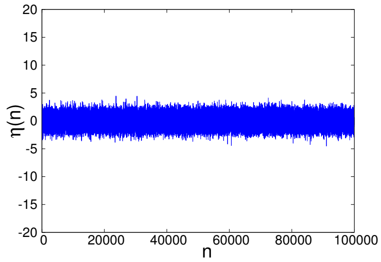
Hence, by varying α_0 and α_1, it is possible to obtain stochastic processes with the same unconditional variance but with different values of the kurtosis.

In Fig. 3.5 we have plotted the increments η_i as functions of the discrete time i, for a simple ARCH(1) process, for three different combinations of the values of the parameters α_0 and α_1, in a way to obtain the same unconditional variance ($\overline{\sigma^2} = 1$) but different values of the kurtosis ($\kappa = 3, 9, 23$). Also the Gaussian process ($\alpha_0 = 1, \alpha_1 = 0$, $\kappa = 3$) is plotted for comparison. We can observe that in the Gaussian case large jumps (jumps larger than 3 times the unconditional variance) are almost absent while are observed when the kurtosis $\kappa > 3$.

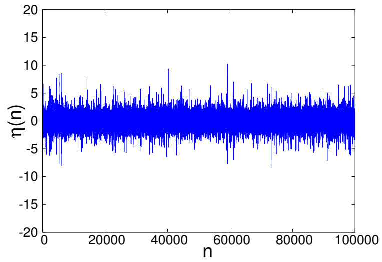
Also the probability distributions of the increments η will then be leptokurtotic. In Fig. 3.6 the probability density functions for the increments generated by three different ARCH(1) processes are shown. The three functions refer to the same values of Fig. 3.5. We can observe higher degree of leptokurtosis when $\kappa > 3$. When $\alpha_0 = 1$ and $\alpha_1 = 0$ the probability density function of the increments η is Gaussian. For $0 < \alpha_1 < 1$, the exact shape of the ARCH(1) probability density function is unknown.

Let us now discuss the correlation properties of an ARCH(1) process. From Eq. 3.38 we can observe that the increments η_k are uncorrelated random variables but, as we will see they are not independent since higher-order correlations are present and reveal a richer structure. For example consider the correlation of the square increments:

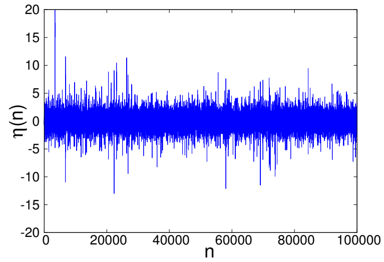
$$C(i,j) = \overline{<\eta_i^2\eta_j^2>} - \overline{<\eta_i^2><\eta_j^2>} = \overline{\sigma_i^2\sigma_j^2} - \overline{\sigma_i^2\sigma_j^2} \qquad i \neq j, \tag{3.47}$$



(a) $\kappa = 3$



(b) $\kappa = 9$



(c) $\kappa = 23$

Figure 3.5: Numerical simulation of increments generated by an ARCH(1) process for different values of the parameters α_0 and α_1. The couples of parameters are chosen in a way to obtain the same unconditional variance ($\sigma^2 = 1$) but different values of kurtosis κ. In (a) $\alpha_0 = 1$ and $\alpha_1 = 0$, hence from Eq. 3.46 we obtain $\kappa = 3$, that is the case of the Brownian motion. In (b) $\alpha_0 = \alpha_1 = 0.5$ and $\kappa = 9$. In (c) $\alpha_0 = 0.45$, $\alpha_1 = 0.55$ and hence the kurtosis is $\kappa = 23$.

which indeed has an interesting temporal behavior in financial time series as we have seen in Chapter 2. Notice that for $i \neq j$ this correlation can be zero either because σ is identically equal to a certain value σ_0, or because the correlations of σ are completely uncorrelated from one time to the next. Hence, let us now determine the volatility autocorrelation function of the ARCH(1) process. We start from

$$\overline{\sigma^2_{i+j+1}\sigma^2_i} = \overline{(\alpha_0 + \alpha_1\eta^2_{i+j})\sigma^2_i} = \alpha_0\overline{\sigma^2} + \alpha_1\overline{\sigma^2_{i+j}\sigma^2_n}, \tag{3.48}$$

for $j > 0$. Considering Eq. 3.44, one can observe that:

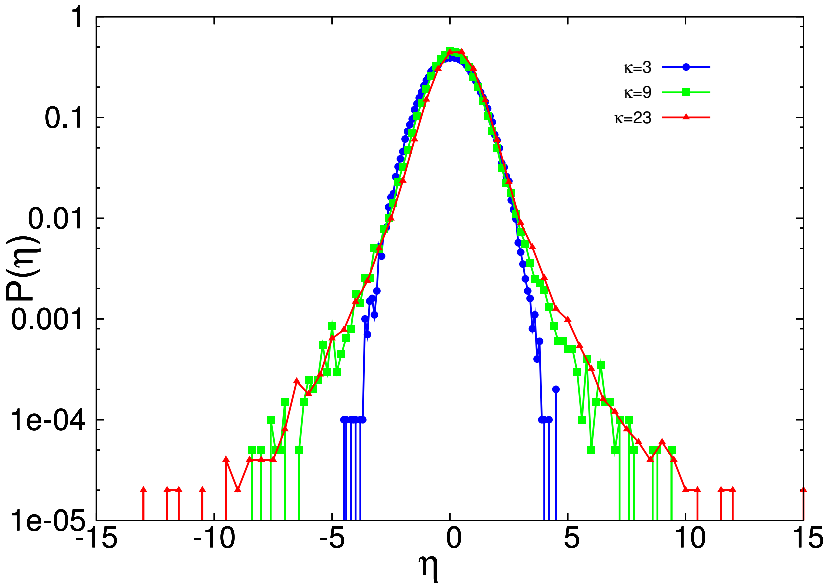


Figure 3.6: In this figure is plotted the probability density function for the increments η of an ARCH(1) process. In the figure are plotted three different histograms for ARCH(1) process generated with different values of the parameters α_0 and α_1. In particular the values used are that of Fig. 3.5. The blue line is the case of $\kappa = 3$ that is the simple RW. In fact in this case we obtain a Gaussian probability density function. In the other two curves the degree of κ increases and we can observe fat-tailed shapes.

$$\alpha_0\overline{\sigma^2} - \overline{\sigma^2}^2 = -\alpha_1\overline{\sigma^2}^2. \tag{3.49}$$

Then we arrive at

$$C(j+1) = \overline{\sigma^2_{i+j+1}\sigma^2_n} - \overline{\sigma^2}^2 = \alpha_1(\overline{\sigma^2_{i+j} - \overline{\sigma^2}^2}) = \alpha_1 C(j). \tag{3.50}$$

This recursion law is equivalent to

$$C(t) \sim C_0 \exp(t/\tau), \tag{3.51}$$

with $\tau = 1/\ln\alpha_1$. Thus, the autocorrelation of the ARCH(1) process shows an exponential decay in contradiction to various observations, suggesting a power law decay. This problem appears also for higher ARCH(p) processes [24], where the autocorrelation function is a weighted sum of various exponential decays.

Many different types of ARCH models have been proposed in the literature, such as ARMA-ARCH [118], CHARMA [116], threshold ARCH [120], or double threshold ARCH [78]. A model with quite attractive features is the generalized ARCH (GARCH) model [24].

In various applications using ARCH(p) model, a large value of p is required. This usually poses some problems in the determination of the parameters α_0 and α_n describing the evolution of a given time series. Overcoming this problem and some other inadequacies of the ARCH(p) model leads to the introduction of a generalized ARCH process, the so-called GARCH(p,q) process [24]. This generalized model adds q autoregressive terms to the moving averages of squared errors. The variance equation takes the form:

$$\sigma_i^2 = \alpha_0 + \alpha_1 \eta_{i-1}^2 + ... + \alpha_p \eta_{i-p}^2 + \beta_1 \sigma_{i-1}^2 + ... + \beta_q \sigma_{i-q}^2. \tag{3.52}$$

From here we obtain immediately that the unconditional variance of the GARCH(p,q) model is given by

$$\overline{\sigma^2} = \frac{\alpha_0}{1 - \sum_{k=1}^{p} \alpha_k + \sum_{k=1}^{q} \beta_k}. \tag{3.53}$$

Let us now focalize on the GARCH(1,1) process:

$$\sigma_i^2 = \alpha_0 + \alpha_1 \eta_{i-1}^2 + \beta_1 \sigma_{i-1}^2. \tag{3.54}$$

In this case the unconditional variance, from Eq. 3.53, is given by [15]:

$$\sigma^2 = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}, \tag{3.55}$$

and the kurtosis is given by the relation:

$$\kappa = 3 + \frac{6\alpha_1^2}{1 - 3\alpha_1^2 - 2\alpha_1\beta_1}. \tag{3.56}$$

Thus unconditional finite variance requires $\alpha_1 + \beta_1 < 1$ while a finite positive kurtosis appears for $2\alpha_1^2 + (\alpha_1 + \beta_1)^2 < 1$.

Recalling equation 3.35 and substituting it in equation 3.54, we obtain:

$$\sigma_i^2 = \alpha_0 + (\alpha_1 \epsilon_{i-1}^2 + \beta_1)\sigma_{i-1}^2. \tag{3.57}$$

Eq. 3.57 shows the multiplicative nature of the GARCH process.

The autocorrelation function of the random variable $\eta_k = \epsilon_k \sigma_k$, is proportional to a delta function, because of the nature of ϵ_k. Nevertheless, similarly to the ARCH process, also the GARCH process shows some higher-order correlations. Following [24], it can be shown that in a GARCH process, the variable η_k^2 is a Markovian random variable characterized by the time scale $\tau = |\ln(\alpha_1 + \beta_1)|^{-1}$. Hence, a GARCH(1,1) process provides an interesting example of a stochastic process η_k that is second-order uncorrelated, but is higher-order correlated.

An analogous computation as used for the derivation of Eq. 3.50, yields the recursion law for the volatility autocorrelation function,

$$C(j+1) = (\alpha_1 + \beta_1)C(j). \tag{3.58}$$

So we obtain again an exponential decay:

$$C(t) \sim C_0 \exp(t/\tau) \tag{3.59}$$

with the correlation time $\tau = 1/\ln(\alpha_1 + \beta_1)$.

However, an important difference between ARCH(1) and GARCH(1,1) is detected by comparing the characteristic timescale for these two processes. Let us consider only processes with finite unconditional variance and finite kurtosis. The finiteness of the kurtosis implies that the coefficient α_1 of ARCH(1) must be lower than $1/\sqrt{3}$. Hence, this process may be characterized by a relatively short maximal correlation time in the volatility fluctuations of $\tau = 1/\ln(1/\sqrt{3}) \sim 1.8$ time units whereas in the GARCH(1,1) process with finite variance and kurtosis we can observe a characteristic time scale longer than hundreds of time units, the only condition being that the β_1 parameter must be larger than 0.7.

In Chapter 2 we have shown the evidence that the variance (or the absolute value) of returns has a power-law behavior. The correlations of the square increments of a GARCH process is exponential, so one should reject the GARCH process as a model for price dynamics. In spite of this limitation, GARCH(1,1) model is widely used to describe financial time series. The limitation is overcome by using values of the sum $(\alpha_1 + \beta_1)$, that defines the time scale, close to 1 for the empirical analysis [1, 3]. In fact, a value of the sum $(\alpha_1 + \beta_1)$, implies a time memory that could be of the order of months. For example, the empirically determined coefficients of the US dollar rate GARCH(1,1) volatility for sterling and the Japanese yen, obtained in the time period 1983 to 1991 by analyzing the daily data [3], are $\alpha_1 = 0.052$ and $\beta_1 = 0.931$ (sterling) and $\alpha_1 = 0.094$ and $\beta_1 = 0.839$ (yen). Thus we obtain a correlation time of $\tau = 58$ trading days in the sterling case and $\tau = 15$ trading days for the yen. Such a long time memory mimics in an approximate way the power law correlation of the volatility over a finite window.

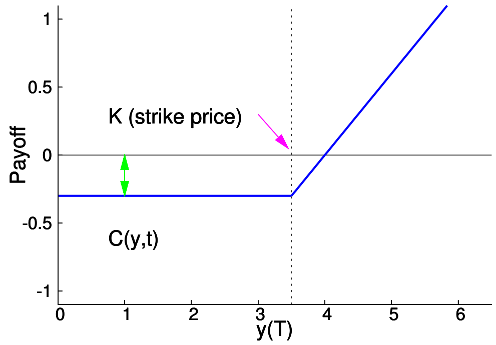
In [88] Mantegna and Stanley have performed an empirical test to investigate if the GARCH(1,1) model could describe a set of S&P500 high-frequency data. They compared the probability density function and the scaling properties of the empirical data with the same for a GARCH(1,1) process with the same variance and kurtosis measured in the time series of the S&P500 data. The chose the set of parameter $\alpha_0, \alpha_1, \beta_1$ which better described the data with the imposition that the kurtosis an the variance of the GARCH(1,1) process were the same of the empirical ones. From the empirical analysis of the $\Delta t = 1$ minute data, they found that $\sigma^2 = 0.00257$ and $\kappa = 43$, so they obtain $\alpha_0 = 2.30 \cdot 10^{-5}$ and $\alpha_1 = 0.09105$. The variable β_1 is fixed to the value 0.9 as is used to be in literature [1]. They found that the process describes well the $\Delta t = 1$ minute probability density function. In reality this agreement does not assure that the same process describes well the empirical data for any time horizon Δt. In fact, from a scaling analysis, they found that the GARCH(1,1) process fails to to describe the scaling properties of the probability density functions for all time horizon using the same parameters.

3.4 Option Pricing

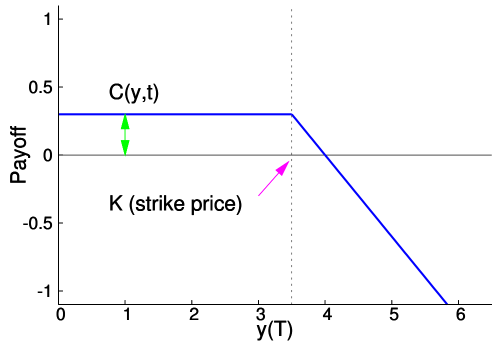
In this section we are going to describe an important theoretical tool, i.e. the famous Black and Scholes equation [22], to estimate the price of financial contracts, in particular of options. Financial contracts [72] are financial derivative products whose price depend upon the price of another financial product that for example can be the price of a stock. Before introducing the Black and Scholes equation, let us spend some words to describe some financial contracts. The simplest financial contract is a *forward contract*. This contract is an agreement between two parties to buy or sell an asset at a pre-agreed future point in time (the delivery date T) at a given forward price (the delivery price K). Therefore, the trade date and delivery date are separated. The buyer of the contract is said to have a long position, the seller of the contract is in a short position. The price $y(T)$ of the asset at the delivery date T usually differs from the pre-assigned delivery price K.

This contract is a symmetrical contract for the two parties (long and short), when one has a positive payoff the other has a negative one and vice versa. In fact the short position will have a positive payoff if $y(T) < K$. In this case the trader who has this short position can sell his stocks at a price K and immediately buy again stocks at price $y(T) > K$. Following the opposite reasoning, the long position has a positive payoff if $y(T > K)$.

Another important financial contracts are the *options*. An option is a contract between two people that conveys the right to buy or sell a given stock (or another financial product) at a specified price K (called the strike price) at time T (called the maturity date). We want to underline that in an option contract one has the right but not the obligation to execute the contract. There are two forms of option contracts: calls and puts. A call option is a contract that permits the holder to buy an asset at a specified maturity date for a designed strike price and requires the seller to sell. At the time the contract is written, the two parties agree to both the strike price and the maturity date and the buyer of the option contract has to pay an amount of money $C(y, t)$ to the seller of the option. A put option allows the holder to sell rather than purchase the stock. When the exercise of the contract can be done only at maturity date the option is called European. If the contract can be exercised at any time up to the maturity date, the option is called American. Here we discuss only European Options. Let us see some example of how works an option. Suppose that the stocks ABC are quoted in the market for 3,50 USD. How much one is disposed to pay to have the right to buy ABC stocks at 3,50 USD in a month? If one thinks that the price of ABC will rise up to 4,00 USD the maximum price one will be disposed to pay to have the right to buy ABC stocks is 0,50 USD. If instead one supposes that the price of ABC will go down, this right is not worth a bean. If one has the right to buy the only interesting thing is that price will rise. Suppose that one thinks that in a month the price of ABC will rise up to 4,00 USD and he buy a call option paying 0,30 USD ($C(y, t) = 0.30$ USD) per stock to have the right to buy ABC stocks at 3,50 USD in a month. After a month the price of ABC can be larger, smaller or equal to the actual price 3,50 USD. If the price will be 4,00 USD the option will be exercised and the holder of the contract will buy ABC stocks at 3,50 USD with the possibility of immediately sell them at 4,00 USD earning 0,50 USD per stock. If the price will go down up to 3,00 USD the holder of the option will not exercised his right to buy ABC stocks. One can observe that the holder of an option has a limited risk. If the price will go down he will loose only the price of the option but, the hypothetical gain if the price will rise can be unlimited. The seller of the option will surely gain the cost of the option but can meet an unlimited loss. In Fig. 3.7 this concept is graphically represented.



(a) buyer of a call option



(b) seller of a call option

Figure 3.7: In this figures we show the payoff for a buyer of a call option (a) and a seller of a call option (b), as function of the price of the asset y at maturity date T. The cost of the option is $C(y,t)$ and the strike price is K. We can observer that the holder of the option can obtain an unlimited gain.

In the case of a put option we have the opposite situation.

At this point we derive the most important result of option pricing theory: the Black and Scholes equation [22].

The Black and Scholes equation is valid under the following assumptions:

1 The market is assumed to be in a steady state at least over the time of the option contract.

2 There are no costs associated with exercising the option.

3 There are no riskless arbitrage opportunities.

4 The holder will exercise the option if it is profitable to do so.

5 There is no possibility of default on the contract.

6 Selling of securities is possible at any time.

7 The trading is continuous.

8 The stock pays no dividends during the option's life.

Black and Scholes assume that the stock price follows a geometric Brownian motion therefore the stochastic differential equation for the price y is:

$$dy = \mu y dt + \sigma y dW, \tag{3.60}$$

where μ is the drift, σ the standard deviation per unit time and W the Wiener process. With this assumption, it can be shown that any function of the price, and hence also the price of the option $C(y,t)$, must be a solution of the following differential equation:

$$dC = \left[\frac{\partial C}{\partial y}\mu y + \frac{\partial C}{\partial t} + \frac{1}{2}\frac{\partial^2 C}{\partial y^2}\sigma^2 y^2\right]dt + \frac{\partial C}{\partial y}\sigma y dW. \tag{3.61}$$

Now we consider a simplified strategy with a portfolio of only one type of stocks. The generalization to a multicomponent portfolio is always possible. We suppose further that the portfolio is constructed by buying n shares of price $y(t)$ and selling an option of price $C(y, T-t, K)$. Thus, the value of the portfolio is:

$$p(t) = ny(t) - C(y, T-t, K). \tag{3.62}$$

The fluctuation of the share price y leads to fluctuations of the portfolio:

$$\Delta p \approx n\Delta y - \left[\frac{\partial C}{\partial y}\right]\Delta y. \tag{3.63}$$

A riskless investment requires

$$n = \left[\frac{\partial C}{\partial y}\right]. \tag{3.64}$$

The value of the portfolio is therefore

$$p = y\left[\frac{\partial C}{\partial y}\right] - C \tag{3.65}$$

From here, we get the total differential

$$dp = \frac{\partial C}{\partial y}dy - dC. \tag{3.66}$$

Substituting equations 3.60 and 3.61 in this last equation, we obtain:

$$dp = \Big[-\frac{\partial C}{\partial y}\mu y - \frac{\partial C}{\partial t} - \frac{1}{2}\frac{\partial^2 C}{\partial y^2}\sigma^2 y^2 + \frac{\partial C}{\partial y \mu y}\Big] dt - \frac{\partial C}{\partial y}\sigma y dW + \frac{\partial C}{\partial y}\sigma y dW. \quad (3.67)$$

and therefore, simplifying,

$$dp = \Big[-\frac{\partial C}{\partial t} - \frac{1}{2}\frac{partial^2 C}{\partial y^2}\sigma^2 y^2 \Big] dt \quad (3.68)$$

Now we use the hypothesis of absence of arbitrage. This means that the change in the value of the portfolio must equal the gain obtained by investing the same amount of money in the corresponding security providing an average return r per unit of time. If r is constant:

$$\frac{dp}{dt} = rp. \quad (3.69)$$

By equating equations 3.68 and 3.69, we obtain the famous Black and Scholes differential equation.

$$rC = \frac{\partial C}{\partial t} + ry\frac{\partial C}{\partial y} + \frac{1}{2}\frac{\partial^2 C}{\partial y^2}\sigma^2 y^2. \quad (3.70)$$

To obtain the appropriate option price function C, we have to fix the boundary conditions. In the case of a call option, one executes the contract if $y(T) > K$ and the buyer gets assets of the value $y(T)$ at price K. Thus the option has the cost $y(T) - K$. If $y(T) < K$ the contract is not executed, and the value of the option is 0. Hence

$$C = \max\Big\{ y(T) - K, 0 \Big\} \quad (3.71)$$

To solve the differential equation 3.70, Black and Scholes make the substitution:

$$C(y,t) = \exp(r(t-T))y(x,\tau), \quad (3.72)$$

where

$$x \equiv \frac{2}{\sigma^2}\left(r\frac{\sigma^2}{2}\right)\left[\ln\left(\frac{y}{K}\right) - \left(r - \frac{\sigma^2}{2}\right)\left(t - T\right)\right], \tag{3.73}$$

and

$$\tau \equiv -\frac{2}{\sigma^2}\left(r - \frac{\sigma^2}{2}\right)\left(t - T\right). \tag{3.74}$$

With this substitution, the Black and Scholes equation becomes equivalent to the heat transfer equation in Physics, which is the standard form of a parabolic partial differential equation

$$\frac{\partial y(x,\tau)}{\partial \tau} = \frac{\partial^2 y(x,\tau)}{\partial x^2} \tag{3.75}$$

which can be solved exactly for the boundary condition in Eq. 3.71. The final solution is:

$$C(y,t) = y\Phi(\eta_1) - K\exp(r(t-T))\Phi(\eta_2) \tag{3.76}$$

where

$$\eta_1 = \frac{\ln y - \ln K + (2r - \sigma^2)(T-t)}{2\sigma\sqrt{T-t}} \tag{3.77}$$

and

$$\eta_2 = \eta_1 - \sigma\sqrt{T-t} \tag{3.78}$$

and Φ is the cumulative density function for a Gaussian variable with zero mean and unit standard deviation

Equation 3.76 allows the determination of the value of a European option for all times before the maturity date in order to provide a fair price only from the knowledge of the strike price, the actual price of the underlying stock, and the time remaining to maturity. While in practice more advanced models are often used, many of the key insights provided by the Black and Scholes formula have become an integral part of market conventions.

3.5 Agent Based Models and the Minority Game

In the preceding sections, we described the price fluctuations of financial assets in statistical terms. We did not ask question about their origin, and how they are related to individual investments decisions. In the language of Physics, the precedent approach was macroscopic and phenomenological. We considered the macro-variables (prices, returns, volatilities) and checked the internal consistency of the phenomena observed. In this section, we wish to discuss how these macroscopic observable are possibly related to the microscopic structure and rules governing capital markets.

In this perspective, a typical approach is to study the economic system starting from the behavior of a set of interacting agents. This is the field of game theory which deals with decision making and strategy selection under constraints. Game theory as applied by economists is built on one standard assumption of Economics that is agents behave in a rational manner. The non-trivial problem comes from constraints, and conflicting through similar behavior of other agents. This assumption of rationality eliminates randomness from the games, and make them essentially deterministic. In this perspective, games involves an optimization problem. The benefits of an agent are often described by a utility function which, of course, depends on the strategies of all players. Under the assumption of complete information sharing between players, the solution of the game is a Nash equilibrium. A Nash equilibrium is a state which is locally optimal simultaneously for each player (a local maximum of all utility functions).

In a Physics perspective, such Nash equilibria in deterministic games, might be viewed as a zero-temperature solution, where all possible fluctuations are frozen [93]. Introducing fluctuations, or randomness, then would correspond to finite-temperature properties.

One may wonder to what extent game theory can improve our understanding of financial markets. Financial Markets certainly provide the basic ingredients of game theory: a common goal and the necessity of strategy selection and decision making under constraints. However, uncertainty is an essential feature of capital markets, and while some information is available at high frequency and quality, the information an the strategies of other players is very limited.

In this section, we will explore a very simple game where agents have to make decisions using strategies chosen from a finite given set. In this model agents are far from being hyper-rational and their behavior often changes over time. Thus the dynamical approach towards the steady state is considered through evolution, adaptation and learning. This game is called Minority Game and was introduced by Challet and Zhang in 1997 [39].

The Minority Game is the mathematical formalization of "El Farol Bar" problem considered By Brian Arthur [10, 11]. The "El Farol Bar" problem was proposed as an example of inductive reasoning [99] as an alternative to the deductive one, which is usually assumed in Economics and that is extremely useful to generate theoretical solutions [60]. Inductive reasoning assumes that by feeding back the information about the game outcome, agents could eventually reach perfect knowledge about the game and arrive to a steady state. On the contrary, deductive reasoning assumes that the precedents contain full information about the game and then there is not any dynamical approach to the steady state, which is attained in a single step.

The economist Brian Arthur proposed the "El Farol Bar" model to embody why the economy is so complex. In particular it shows the importance of out-of-equilibrium behavior in a financial system, where agents behave inductively since they only have limited information available. Let us discuss it in some detail. There is a bar, the "El Farol" bar in Santa Fe, and N agents have to decide whether to go or not on a given night each week. The goal of each agent is to attend provided he can get a seat. The problem is that space is limited and only aN with $a < 1$ of the possible N agents can get a seat. Thus the agents have two actions: go if they expect the attendance to be less than aN people or stay at home if they expect the bar will overcrowded. There is no interaction among agents, the only information available is the numbers who came in the past weeks. Based on this, each player tries to predict whether the bar is likely to be overcrowded the following week (implying he stay at home) or under-crowded (implying the agent should make the effort to go to the bar). Hence, there is no correct expectation model; if everybody made the same decision, it would be automatically the wrong decision since everyone would either stay at home or go to the bar. A so-called mean-field theory describing a "typical" agent will not work.

This situation can be simulated on a computer choosing a pool of possible prediction methods given a particular set of recent outcomes, and randomly assign a few such prediction methods to each agent. Computer simulations of this model [10, 11] show that the attendance fluctuates around aN. Thus, predictors self-organize so that this structure emerges in the complex dynamics of the system.

The El Farol bar problem can be considered as a kind of very simple "market toy model"[91, 89, 121]: at each time step agents can buy (go to the bar) or sell or sell an asset and after each time step the price of the asset is determined by a simple supply-demand rule.

Since the actual attendance numbers range from 0 to N, the number of possible patterns of past attendance is enormous and hence so is the number of possible strategies. So in [39] Challet and Zhang have abstracted the basic features of this model using a binary approach in the so-called Minority Game (MG).

In order to describe the MG, let us consider a population of N (odd) players, each with a finite number of strategies S. At every time step t, every player i must choose one of two alternatives, ± 1, so his action will be $a_i(t) = \pm 1$. Those players who are in the minority side win. After a round, the total action is:

$$A(t) = \sum_{i=1}^{N} a_i(t). \tag{3.79}$$

At each time step, if a player is in the minority side, he collects a payoff $-a_i(t)g[A(t)]$, with g and odd function of $A(t)$. Challet and Zhang's initial choice was $g(x) = \text{sign}(x)$, but other analytical functions can be more suitable. The information about the winning group is released to the agents only in terms of the sign of $A(t)$, without the actual attendance number. So, at time step $t+1$, the information given to the agents regarding the game at time t is $W(t+1) = \text{sign}A(t)$. Let us assume that the players are limited in their analyzing power, then they can only retain last M winning sides, i.e. the last M numbers W. Moreover they base thir decisions $a_i(t)$ on these last M bits only. To this end they have a set of $s \geq 2$ strategies. A strategy is just a mapping from the sequence of the last M winning sides to the action of agent i. So different strategies are distinguished by the different predictions from the same signal. An example of strategy is illustrate in table 3.1 in the case $M = 3$.

There are 2^M possible signals of M bits, and two predictions for each signals. The space of strategies with memory M therefore is of size 2^{2^M}. At the beginning of the game each agent is given a set of S strategies randomly drawn from the total 2^{2^M} possible strategies. The heterogeneity among agents is in the fact that the set of S strategies could be very different for different agents. Adaptation comes in the way agents choose at each time step one of their S strategies. They take the strategy within their own set of strategies whose performance over time to predict the next winning group is biggest. In order to do that, each agent i assigns a virtual point to his strategy after each time step when it predicted correctly the winning group. However this points are only virtual points as they record only the agents' strategies performance and serve only to rank the strategies within each agent set. After time t each agent choose the first strategy in his personal ranking which tells him what to do in the future.

Signal	Prediction
-1 -1 -1	1
-1 -1 1	-1
-1 1 -1	-1
-1 1 1	1
1 -1 -1	1
1 -1 1	-1
1 1 -1	1
1 1 1	-1

Table 3.1: An example of strategy for $M = 3$. In the first column there are all the possible outcomes of the game in the case $M = 3$, i.e. 8 possible different outcomes. In the second column there is the action to make in accordance with the given strategy for any given outcome of the game.

When the game is simulated, in the case $g[A(t)] = \text{sign}(A(t))$ the times series $A(t)$ oscillates rather randomly around 0. Instead, the fluctuations of $A(t)$ around zero, given by the variance $\sigma^2 =< [A(t)- < A(t) >]^2 >$, have a more interesting behavior [71, 70]. First, note that large fluctuations imply large waste since still more players could have taken the winning side without harm done to the others. On the other hand, smaller fluctuations imply more efficient usage of available resources, in general this would require coordination and cooperation, which are not built-in explictly. In [39] is shown that, for fixed values of N and S, the fluctuations around $A(t) = 0$ are smaller for increasing values of M, so populations having larger memory cope with each other better.

The behavior of σ^2 as a function of the parameters of the model M, S and N shows quite remarkably behavior. Numerical [108] and analytical studies [74, 35, 36] show that in this game exists an order parameter $\alpha = 2^M/N$, which value controls the phase in which is the system.

In fact, it was found by extensive simulations that σ^2/N is only a function of α for each value of S, as is shown in Fig. 3.8. For large values of α, σ^2/N approaches the value for the random choice game $\sigma^2/N = 1$, i.e. the game in which each agent randomly chooses $a_i(t) = 1$ or $a_i(t) = -1$ independently and with equal probability at each time step. At low values of α, the average values of σ^2 is vary large, it scales like $\sigma^2/N \sim \alpha^{-1}$ which means that $\sigma \sim N$ and thus the size of the losing group is much larger than $N/2$. At intermediate values of α, the volatility σ is less than in the random case. In this region the size of the losing group is close to $N/2$.

The fact that σ gets below the random case for a given interval of values

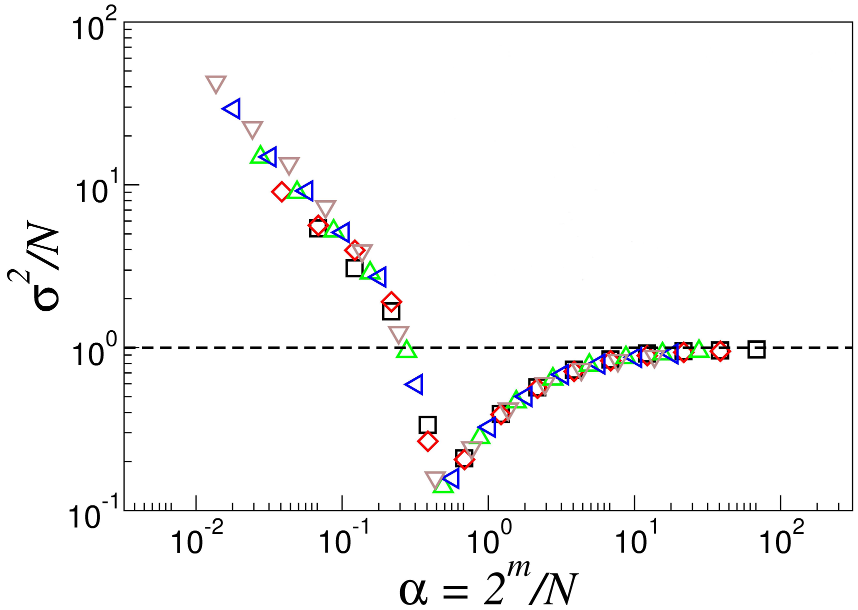


Figure 3.8: Volatility as a function of the control parameter $\alpha = 2M/N$ for $s = 2$ and different number of agents $N = 101, 201, 301, 501, 701$. The different symbols in the graph correspond to different values of N. We can see that, for small α the volatility is greater than the random case, while in the region of large α is smaller. Figure adapted form [99].

of α suggests the possibility that agents coordinate in order to reach a state in which less resource are globally wasted [108]. In the market formalism, this means that agents can exploit information available and predict future market movements. This fact led some authors to study the information contained in the time series of $A(t)$. Specifically, it was found that $W(t+1) = \text{sign}A(t)$ is independent of the sequence of the last attendances in the high volatility region (α small), while there is a strong dependence for α large. To quantify this behavior, it was proposed in [36] to measure the information as:

$$H = \frac{1}{2^M} \sum_{\nu=1}^{2^M} < W(t+1)|\mu(t) = \nu >^2, \tag{3.80}$$

where $\mu(t)$ is the binary representation of the string of M bits giving the last M outcomes of the game. If there is no significant dependence between $W(t+1)$ and $\mu(t)$ then $< W(t+1) >=< \text{sign}A(t) >= 0$ and hence $H = 0$. In the market context, $H \neq 0$ indicates the presence of information or arbitrage in the signal $A(t)$. Simulations in [36] shows that $H = 0$ for $\alpha < \alpha_c$ and $H \neq 0$ for $\alpha > \alpha_c$ with $\alpha_c \sim 0.3$. This fact suggests the possibility that there is a phase transition at $\alpha = \alpha_c$ which separates those two efficient and inefficient phases.

Since information is fed back into the system, many studies was concentrated in the possibility that the system of N agents could exploit this information to achieve better coordination. However, A. Cavagna in [34] showed by means of simulations that if the information $\mu(t)$ on the M past outcomes of the game that is given to the agents at any time steps is chosen randomly and independently of time from its possible values, the behavior of the MG remains the same regarding time averaged extensive quantities like the volatility, the information, etc. This means that in this game coordination do not come through exploitation of available information in the sequence of winning groups.

We have seen that in the phase $\alpha > \alpha_c$ the total waste is smaller than in the random solution. We can interpret this fact saying that the agents' dynamics tends to maximize the global efficiency. The concept that the system is trying to minimize a given quantity is very appealing to statistical mechanics. If this quantity exists, the system can be studied by considering its minima and perturbation around them. However, due to the rules of MG, the game never settles down because some of the agents keep on changing their strategies forever. However, since the actions of agents $a_i(t)$ depend on the points of the strategies that depend on the past history of actions, one could wonder whether there is any time pattern in the long run that agents follow. It might be that $a_i(t)$ never come to a rest, but $m_i(t) = \sum_{\tau=0}^{t} a_i(\tau)$ can converge when $t \to \infty$ to a given quantity. This is the key point in the first attempt at a solution of the MG in [37, 92]. In [37, 92] is shown that a slightly modified MG can be solved exactly using methods from spin-glass Physics in the limit $N \to \infty$. Agents do not simply choose the strategy with the highest virtual score, but proceed in a probabilistic manner: a strategy is chosen with a probability which depends exponentially on its virtual score in the game. To see the essentials, we limit our ourselves to $S = 2$ strategies which would correspond to spin 1/2. Therefore each agent $i = 1, ..., N$ has $S = 2$ strategies, denoted by $a_{\pm,i}$, which are randomly drawn from the set of 2^{2^M} possible strategies. For each history μ, i.e. the string of the last M action taken by the minority, a strategy a specifies a fixed action a^μ. We can define:

$$\omega_i^\mu = \frac{a_{+,i}^\mu + a_{-,i}^\mu}{2}, \qquad \xi_i^\mu = \frac{a_{+,i}^\mu - a_{-,i}^\mu}{2} \tag{3.81}$$

so that the strategies of agents i can be written as $a_{s_i,i}^\mu = \omega_i^\mu + s_i \xi_i^\mu$ with

$s_i = \pm$. The current best strategy of agent i, which he shall adopt at time t, is that which has the highest cumulated payoff. Let us define $\Delta_{i,t} = U_{i,t}^{(+)} - U_{i,t}^{(-)}$ as the difference between the cumulated payoff of strategies + and - for agent i at time i. Therefore the choice is:

$$s_i = \mathrm{sign}\Delta_{i,t}, \tag{3.82}$$

The difference in population of agents choosing the + and - sign, at time t, is then

$$A_t = \sum_{i=1}^{N} a_{s_i,i}^{\mu_t} = \Omega^{\mu_t} + \sum_{i=1}^{N} \xi_i^{\mu_t} s_i, \tag{3.83}$$

where $\Omega^{\mu} = \sum_i \omega_i^{\mu}$.

An important quantity is the variance $\sigma^2 =< A^2 >$, where $< \cdot >$ stands for a time average. Then, the variance is given by:

$$\sigma^2 = \overline{< A^2 >} = \overline{\Omega^2} + \sum_i [\overline{\xi_i^2} + 2\overline{\Omega\xi_i} < s_i >] + \sum_{i \neq j} \overline{\xi_i \xi_j} < s_i >< s_j > . \tag{3.84}$$

Here, $< x >$ denotes the time average of the quantity x and $\overline{x}$ is the average over histories. This allows us to decompose a temporal average into one conditioned on history $< x^h >$, followed by one over histories, i.e. $< x >= \overline{< x^h >}$. By symmetry, $< A >= 0$. However, for particular histories, there may be a finite expectation value $< A^h > \neq 0$. One may then calculate the average over the histories of the history dependent expectation values of A:

$$\overline{< A^h >^2} = \overline{\Omega^2} + 2\sum_i \overline{\Omega\xi_i} < s_i > + \sum_{i \neq j} \overline{\xi_i \xi_j} < s_i >< s_j >= \mathcal{H}. \tag{3.85}$$

When the strategies scores are updated using:

$$U_{s,i}(t+1) = U_{s,i}(t) - \frac{a_{s,i}(t)}{2^M} A(t) \tag{3.86}$$

and a probabilistic strategy selection rule $P[s_i(t) = s] \sim \exp[\Gamma U_{s,i}(t)]$ is adopted, the evolution of $< s_i >$ can be cast in the form:

$$\frac{d < s_i >}{dt} = -\Gamma(1 - < s_i >^2)\left(\frac{\partial \mathcal{H}}{\partial < s_i >}\right). \tag{3.87}$$

Formally, these are the equations of motion for magnetic moments $m_i =< s_i >$ in local magnetic field $\overline{\Omega\xi_i}$ interacting with each other through exchange integrals $\overline{\xi_i\xi_j}$.

Note that $\mathcal{H}$ is a positive definite quadratic form, which has a unique minimum.

$$\mathcal{H} = \overline{< A >^2} = \sigma^2 - \sum_i \overline{\xi_i^2}(1- < s_i >^2). \tag{3.88}$$

This implies that the stationary states of the MG is described by the ground state properties of $\mathcal{H}$.

The calculations in [37, 92] showed that there is a phase transition at $\alpha_c \sim 0.33740$. In fact, for $\alpha < \alpha_c$ one has $\mathcal{H} = 0$, while for $\alpha > \alpha_c$ one has $\mathcal{H} \neq 0$.

A variety of extensions of the MG have been formulated. For example the agent population can be made heterogeneous in various dimensions such as memory size, strategy diversification, etc., and agents may choose to stay out of the market (grancanonical MG). One may further diversify the traders population in terms of wealth, investment size and investment strategy (trend following versus contrarians).

The MG is only the most famous agent based models studied by economists and physicists. For a review of these agent based models one can consult [115].

A model different to the MG is the mixed minority-majority where agents are divided into two groups: fundamentalists and chartists. This is a very simplified model to describe the traders behavior in a market. In spite of his simplicity, this model can describe many of the stylized facts of the real price fluctuations observed in the financial market [81].

Another class of models that can describe the stylized facts of real financial market are the so called *zero-intelligence* models [46, 57]. These models describe the agents action in a random manner, compatibly with the market constraints.

Chapter 4

Roughness and Finite Size Effect in the NYSE Stock-Price Fluctuations

The dynamics and fluctuations of stock-prices is represented, at the simplest level, by a random walk which guarantees for the basic property of an efficient market. In the past years it has become clear that one faces a rather subtle and complex form of random walk. Simple correlations of price change are indeed zero at the shortest time but many other features, often related to power law behavior have been discovered [85]. Among the most preeminent one may mention the power law distributions of returns ("fat tails") and the volatility clustering [88, 30] These properties, however, are far from exhaustive and other approaches have been introduced in the attempt of describing the subtle correlations of stock-price dynamics.

One of these methods is the attempt to characterize the "roughness" of the dynamics which can provide additional information with respect to the fat tails and volatility. The scaling properties of the roughness can be defined via the Hurst exponent H [73], through the so called R/S analysis. We consider the roughness problem for high frequency NYSE stock-prices. This means that we take into account all the transactions which occur (tick by tick).

First we discuss the statistical properties of the data set and show that finite size effects are unavoidable and very important [44]. Then we show that fat tails and correlations affect the value of the Hurst exponent in an important way [75]. Finally we analyze the real stock-price fluctuations and argue that the Hurst exponent alone cannot properly characterize their roughness [6]. To this purpose we use a new method to study the roughness which is able to automatically eliminate the trend problem. This is based on the deviation from a suitable moving average and it resolves various ambiguities of the Hurst's R/S analysis.

4.1 Database properties

For our roughness analysis we consider as database the price time series of all the transactions of a selection of 20 NYSE stocks. These have been selected to be representative and with intermediate volatility. This corresponds to volumes of $10^5 - 10^6$ stocks exchanged per day. We consider 80 days from October 2004 to February 2005. This data are a subset of the dataset described in Sec. 2.1

The time series we consider are by a sequential order tick by tick. This is not identical to the price value as a function of physical time but we have tested that the results are rather insensitive to this choice.

The number of transactions per day ranges from 500 to 5000 implying a typical time interval between transactions of a few seconds. The density of operations within a day is characterized by a concave shape which is rather universal as we have observed in section 2.3 and shown in figure 2.3. This means that, with respect to the physical time there are systematic density fluctuations up to a factor of two with a minimum around the center. This effect is obviously eliminated in our tick by tick time, in which physical time is not considered and we have tested that it is not relevant for the roughness properties.

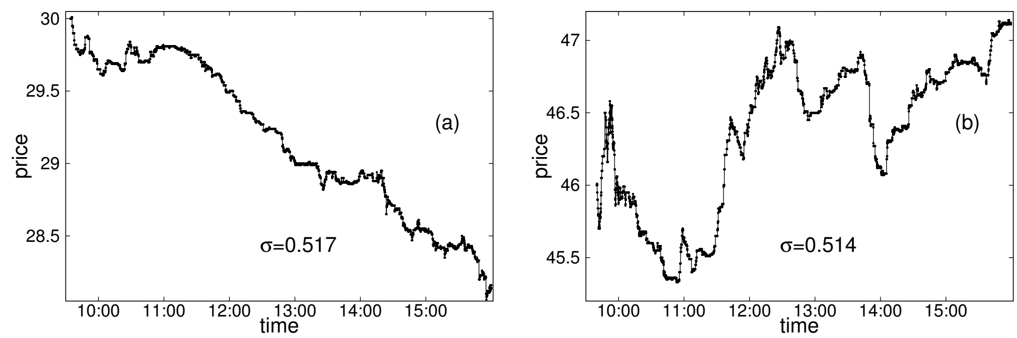
Another problem which is very important and that we have also discussed in section 2.2 is the fact that the closing price of a given day is usually very different than the opening price of the next day. A typical behavior is illustrated in figure 2.2. and it shows that these jumps are serious problem in linking the data of one day to those of the next day.

This means that the data are reasonably homogeneous from the time scale of a few seconds to a few hours but going to longer times can be rather arbitrary due to these large night jumps.

This leads to a very serious problem if one tries to extend these time series beyond the time scale of a single day. In fact, if one simply continues to the next day, one has anomalous jumps for the night which cannot be treated as a standard price change. An alternative possibility could be to artificially eliminate the night jumps and rescale the price correspondingly. This would produce a homogeneous data set which, however, does not correspond to the original data.

This discussion clarifies that there is a fundamental problem in extending the data beyond a single day. Since the transactions within each day range from 500 to 5000, this leads to an important problem of finite size effects in relation to the roughness exponent. In the next section we are going to discuss these finite size effects and show that they are strongly amplified by the fat tail phenomenon.

4.2 Roughness and Hurst exponent



(a) Smooth (b) Rough

Figure 4.1: Examples of the day price dynamics of two stocks whose behavior, with respect to the roughness, appears very different on a visual inspection. The two stocks have a very similar variance (σ) for the price distribution. Surprisingly, also the Hurst exponent will be similar for the two cases.

The importance of a characterization of the roughness properties is clearly illustrated in Fig. 4.1. Here we see the behavior of the price of two stocks which are clearly very different with respect to their roughness properties. The visual difference in roughness, however, does not influence the day volatility σ, which is almost identical. The idea is therefore to add new concepts to characterize their different behavior. We are going to see in the end that even the Hurst exponent is not really optimal to this purpose and the challenge of this new characterization should proceed along novel lines which we outline at the end of the paper.

We first consider the problem of the characterization of the roughness in the Hurst exponent including the finite size effects. The roughness exponent characterizes the scaling of the price fluctuation as a function of the size of the interval considered.

Originally this exponent was introduced for the time series of the levels of the floods of the Nile river. He invented a new statistical method, which he called the *rescaled range analysis* (R/S analysis). The basic idea was to construct a profile from these series and analyze its roughness. This implied some peculiar construction which we can avoid because we have the profile directly.

The characterization of the roughness is complicated by the fact that it corresponds to a problem of anisotropic scaling [18] and it can lead to confusing results in its practical applications [105]. An example of these difficulties is illustrated by the fact that for the growth of a rough profile the Renormalization Group procedure has to be implemented in a rather sophisticated and unusual way [33]. An illustration of this problem is also given by the fact that the value of the fractal dimension of a rough surface is crucially dependent on the type of procedure one considers [33]. The usual approach is to take the limit of small length scales for which the relation between the dimension of the profile, D, and the Hurst exponent is [18]:

$$D = 2 - H \quad . \tag{4.1}$$

However, if one consider the limit of large scales (not rigorous mathematically but often used in physics), one can get $D = 1$ for the Brownian profiles which does not correspond any more to equation 4.1.

In the data analysis one is forced to consider a finite interval and necessarily the two tendencies get mixed. Even considering equation 4.1 one can have various ambiguities. In fact a large Hurst exponent corresponds to small value of the fractal dimension which may appear strange.

Various problems contribute to this possible confusion. The first is how one looks at a scaling law for an anisotropic problem. The scaling for roughness links the vertical fluctuation Δh as a function of the interval considered:

$$\Delta h(\Delta L) \sim \Delta L^H \tag{4.2}$$

In a physical perspective one has typically a lower cutoff and looks at the behavior for large values of ΔL. Since for a random walk (Brownian profile) one has $H = 1/2$ one could say that if $H > 1/2$ this corresponds to a case which is more rough than the Brownian profile. However this is in apparent contradiction with equation 4.1 because the value of D, if $H > 1/2$, results smaller than the Brownian value ($D = 3/2$). This is because equation 4.1 is derived in the limit $\Delta L \to 0$ in the spirit of the coverage approach to derive the fractal dimension.

A similar confusion can be given by the existence of trends in the dynamics of the system. Consider for example a straight line behavior for which

$\Delta h \sim \Delta L$. In this case one would have $H = 1$ and $D = 1$, namely the system is not rough in the $\Delta L \to 0$ perspective but it is very rough in the $\Delta L \to \infty$ view. In such a situation one should realize that a trend is present and that the system is smooth. However, this distinction is not possible with the Hurst's R/S analysis.

Actually in the real data one has an upper and a lower limit for ΔL, due to the intrinsic statistical limitation of the sample. The exponent H is then obtained by a fit in a certain range of scales and all the above problems are difficult to sort out.

4.3 Roughness in a finite size Random Walk

In this section we discuss the role of finite size effects in the determination of the Hurst exponent. We start by deriving some analytical results for a finite size random walk. Consider the function:

$$\mathcal{R}(n) = \langle \max_{k=(ln+1),(ln+n)} (X_k) - \min_{k=(ln+1),(ln+n)} (X_k) \rangle_l \tag{4.3}$$

where $l = 1, 2, ..., \frac{N}{n}$ and $\{X_1, X_2, ..., X_N\}$ are N record in time of a variable X. The function $\mathcal{R}(n)$ describes the expectation value of the difference between maximum and minimum over an interval of size n. $\mathcal{R}(n)$, for many records in time is very well described by the following empirical relation:

$$\mathcal{R}(n) \propto n^H \tag{4.4}$$

where H is the Hurst exponent. Now we want to check which is the effect of the finite size in estimating the Hurst exponent. To perform this analysis we consider a random walk and try to make an analytical calculation of the function $R(n)$ [4].

Suppose that $\delta x_1, \delta x_2, ..., \delta x_n$ are independent random variables, each taking the value $+1$ with probability p, and -1 otherwise. Consider the sums:

$$X_n = \sum_{i=1}^{n} \delta x_i \tag{4.5}$$

then the sequence $X = \{X_i : i \geq 0\}$ is a simple random walk starting at the origin. In order to compute the expectation value of the maximum and the minimum of the walk after n steps, is useful to consider the following theorem.

Theorem 1 (Spitzer's identity). [69] *Assume that X is a right-continuous random walk, and let $M_n = \max\{X_i : 0 \geq i \geq n\}$ be the maximum of the* *walk up to time n. Then, introducing the auxiliary variables s and t, for $|s|,|t|<1$ one has* Eq.4.6*,*

$$\log\left(\sum_{n=0}^{\infty} t^n\,\mathbb{E}(s^{M_n})\right)=\sum_{n=1}^{\infty}\frac{1}{n}\,t^n\,\mathbb{E}\left(s^{X_n^+}\right) \tag{4.6}$$

where $X_n^+=\max\{0,X_n\}$ and $\mathbb{E}$ is the expectation value.

The term *right-continuous* refers essentially to the fact that the step size should be bounded by a characteristic maximum size. In this respect the random walk can be both continuous or discrete in terms of the step distribution. For a mathematically rigorous definition see Ref. [69]

Considering the exponential of both member of Eq.4.6 one has:

$$\sum_{n=0}^{\infty} t^n\,\mathbb{E}(s^{M_n})=\exp\left(\sum_{n=1}^{\infty}\frac{1}{n}\,t^n\,\mathbb{E}\left(s^{X_n^+}\right)\right) \tag{4.7}$$

The k-derivative with respect to t of the left hand side of Eq.4.7 for $t=0$, gives:

$$\left.\frac{\partial^k}{\partial t^k}\right|_0 = k!\,\mathbb{E}(s^{M_k}) \tag{4.8}$$

Defining the right side member of Eq.4.7 as $f(t)$ one can derive:

$$\begin{aligned}\left.\frac{\partial^k}{\partial t^k}f(t)\right|_0 &= \sum_{j=1} k\frac{(k-1)!}{(k-j)!}\mathbb{E}(s^{X_j^+})\left.\left(\frac{\partial^{(k-j)}}{\partial t^{(k-j)}}f(t)\right)\right|_0\\ f(0) &= 1\end{aligned} \tag{4.9}$$

By equating Eq.4.8 and Eq.4.9 one obtains:

$$\mathbb{E}(s^{M_k})=\frac{1}{k}\sum_{j=1}^{k}\frac{\mathbb{E}(s^{X_j^+})}{(k-j)!}\left.\left(\frac{\partial^{(k-j)}}{\partial t^{(k-j)}}f(t)\right)\right|_0 \tag{4.10}$$

In order to obtain $\mathbb{E}(M_k)$ from the function $\mathbb{E}(s^{M_k})$ it is useful to consider the following expansion which holds for a symmetrical probability density function:

$$
\begin{aligned}
\mathbb{E} \quad (s^{X_j^+}) &= \int_{-\infty}^{\infty} s^{X_j^+} P_j(X_j)\, dX_j \qquad (4.11)\\
&= \int_{-\infty}^{0} P_j(X_j)\, dX_j + \int_{0}^{\infty} s^{X_j} P_j(X_j)\, dX_j\\
&= \frac{1}{2} + \int_{0}^{\infty} s^{X_j} P_j(X_j)\, dX_j\\
&\simeq \frac{1}{2} + \int_{0}^{\infty} (1 + X_j \ln(s) + \frac{1}{2} X_j^2 (\ln(s))^2 + ...) P_j(X_j)\, dX_j\\
&\simeq 1 + \ln(s) \int_{0}^{\infty} X_j P_j(X_j)\, dX_j + \mathcal{O}(\ln(s))^2\\
&= 1 + \frac{1}{2}\, \mathbb{E}(|X_j|)\ \ln(s) + \mathcal{O}(\ln(s))^2
\end{aligned}
$$

We now insert this result into Eq. 4.10. Considering also the identity:

$$\mathbb{E}(M_k) = \lim_{s \to 1} \frac{\mathbb{E}(s^{M_k}) - 1}{\ln(s)} \qquad (4.12)$$

we finally obtain:

$$\mathbb{E}(M_k) = \sum_{i=1}^{k} \frac{\mathbb{E}(|X_i|)}{2i} \qquad (4.13)$$

Now we consider various possibilities for the specific nature of the random walk:

a) If the increments δx are independent and corresponding to a Gaussian distribution with $\mathbb{E}(\delta x) = 0$ and variance $\sigma^2 = 1$, one obtains:

$$\mathbb{E}(|X_i|) \quad = \quad \sqrt{\frac{2i}{\pi}} \qquad (4.14)$$

b) If δx have values ± 1 with equal probability one gets:

$$\mathbb{E}(|X_i|) \quad = \quad \sum_{X_i = -i}^{i} |X_i| P_i(X_i) \qquad (4.15)$$

where

$$P_i(X_i) = \binom{i}{\frac{X_i + i}{2}} \left(\frac{1}{2}\right)^i \qquad (4.16)$$

This leads to

$$\mathbb{E}(|X_i|) \quad = \quad \begin{cases} \frac{(i-1)!!}{(i-2)!!} & \text{if } i \text{ is even} \\ \frac{(i)!!}{(i-1)!!} & \text{if } i \text{ is odd} \end{cases}$$

These explicit results permit now to compute the exact expectation for finite size random walk properties [4]. Note that Eq.4.13 has a general value with the only assumption that the increments are not correlated and symmetrically distributed. This means, for example, that one could test the properties of stock-prices for finite size samples being able to separate the role of fat tail, included in Eq.4.13, from the role of correlations.

Eq.4.13 gives an exact relation between the expectation value of the maximum value M_k of a symmetric random walk of k steps and $\mathbb{E}(|X_i|)$, for a given probability distribution of the individual step. In terms of Monte Carlo simulations this would correspond to an infinite number of samples. Since the Monte Carlo method will be applied also to cases for which the analytical result is not available, we can use the present case as a test for the convergency of the Monte Carlo method.
This comparison is shown in Fig.4.2 (a and b) where the two inserts show precisely the degree of convergency as a function of the samples considered.

From this expression it is possible to derive explicitly the expected value of the maximum as a function of the number of steps of the walk. By considering that a similar expression holds also for the minimum, one can directly compute the effective Hurst exponent for random walk of any size. Replacing the results obtained for $\mathbb{E}(M_n)$ in the expression of $\mathcal{R}(n)$ we can plot the average span as a function of n and execute a fit to estimate the value of H. Executing a fit in the region $[10, 1000]$ we obtain a value of the slope that is grater then the asymptotic one. In Fig. 4.3 we shows the result for the effective Hurst exponent that we have obtained performing the fit in the region $[10, n]$ for the random walk with two identical steps (± 1). One can see that finite size effects are very important and a seriously affect the apparent value of H.

The random walk models considered until now have a distribution of individual steps corresponding to a Gaussian distribution or to two identical steps. Real price differences however, are characterized by a distribution of sizes which strongly deviates from these ("fat tails"). For example, if we consider the histogram of the quantity:

$$S(t) = \ln P(t+1) - \ln P(t), \tag{4.17}$$

we find a distributions with large tails, as shown in Fig. 4.4.

To analyze the effect of the fat tails in the evaluation of the Hurst exponent, we can consider a model of random walk with increments that take the values $\delta x = \pm\epsilon$ with probability 0.45 and $\delta x = \pm 10\epsilon$ with probability 0.05. The histogram in Fig. 4.4 represent such a model. We have performed a numerical analysis of the Hurst exponent for a random walk with fat tails to

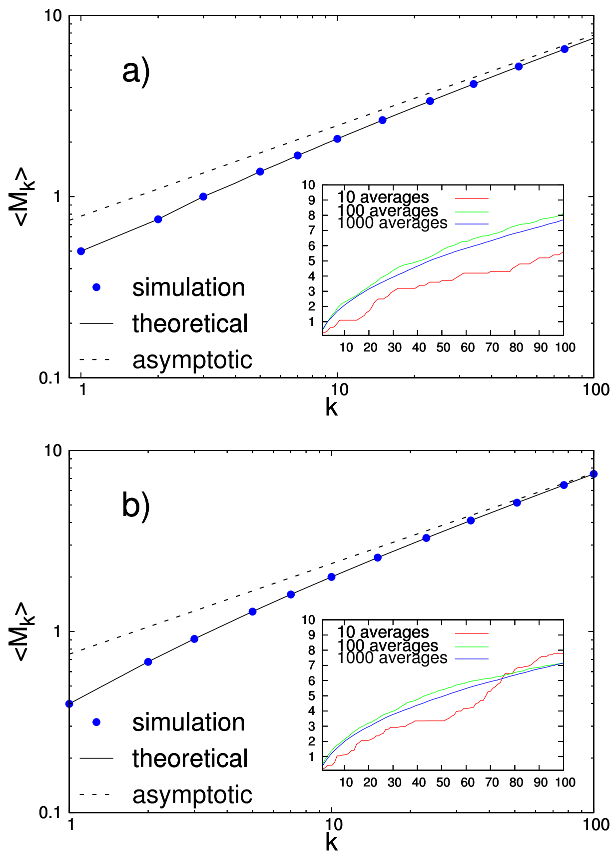


Figure 4.2: The log-log graph shows $\mathbb{E}(M_k)$ as a function of k in the case of a) random walk with two identical steps and b) Gaussian random walk. In the inserts we show the convergency of the simulations to expected value as a function of the number of realizations considered.

study their role on the finite size effects. To this purpose we have generated 1000 random walks of this kind of size n with $n = [100 : 5000]$ and we have calculated the function $R(n)$ for each sample. After calculating the average of $R(n)$, we have considered the plot $R(n)$ as a function of n and the evaluation of $H(n)$ has been performed in the region $[\frac{n}{100}; \frac{n}{10}]$. Figure 4.5 shows the result obtained, a comparison with a normal and a correlated random walk and real data is also shown.

The fact that fat tails and correlations enhance the finite size effects is easy to understand. In case of correlated random walks the effective number of independent steps is strongly reduced. In the case of fat tails instead only the tails give the main contribution to the profile.

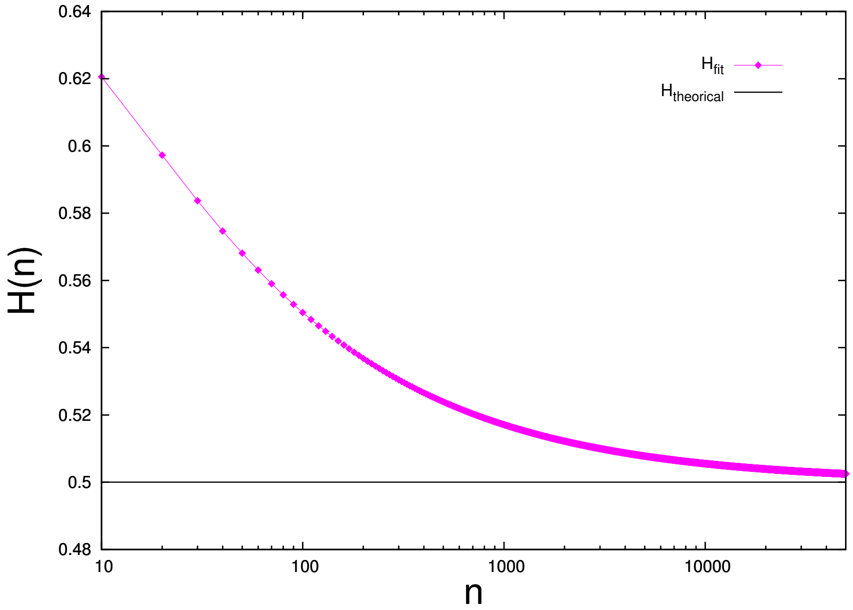


Figure 4.3: This plot shows the trend obtained fitting the curve $\mathcal{R}(n)$ for different values of the size n. The results shows a systematic overestimate of Hurst exponent for small size, due to finite size effects. This is a general result and it shows that finite size effects always enhance the apparent Hurst exponent. This enhancement can be understood by considering that, in some sense, a single step would correspond to $H = 1$, so the asymptotic value $H = 1/2$ is approached from above.

This findings could also have implications for very long times if combined with the non stationarity of the price dynamics. It should be considered the possibility that even the asymptotic regime is still altered by these effects. This could suggest a different interpretation of the deviation of H from the value 1/2, which is usually proposed in terms of long range correlations [97].

The Fig. 4.5 shows the inefficiency of the Hurst exponent's approach to the study of the roughness for systems with a small size. The results are clearly affected by the effect of a finite size and the interpretation of $H > 1/2$ as a long range correlation could be misleading.

4.4 Analysis of NYSE stocks

First we consider the Hurst's R/S analysis for the two stocks plotted in Fig. 4.1 and the relative results are shown in Fig. 4.6.

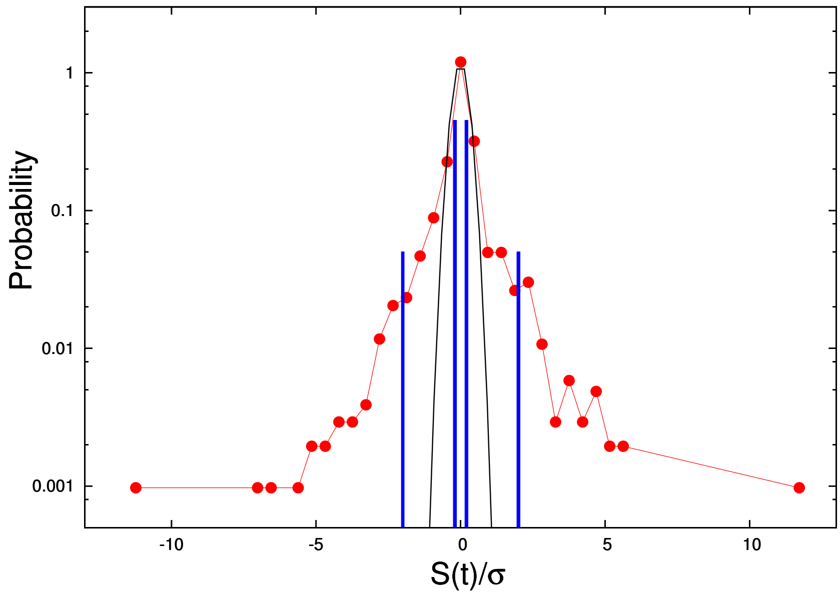


Figure 4.4: Probability function for high frequency price differences of the BRO stock during a day. The solid line is the Gaussian fit of the data. The boxes represent a model to estimate the effect of the tails for the random walk. The probability is estimate by an histogram given by a value $\pm\epsilon$ which has 0.45 of probability and a tail $\pm 10\epsilon$ with probability 0.05. In this plot $\epsilon = 0.2$.

The values of the two exponents H are very similar in spite of the large difference of the two stocks in their apparent roughness properties [6]. This shows that the exponent H is not suitable to characterize the different roughness properties of the two stocks.

We then consider the entire series of 20 stocks and the results are reported in Tab. 4.1. Here H represents the daily value averaged over 80 days. Then H_{max} and H_{min} are the maximum and the minimum values respectively, σ is the variance averaged over the 80 values and $< N >$ is the average number of transactions per day. In Fig. 4.7 we report the time behavior of $H(t)$ for the 80 days for the two stocks of Fig. 4.1. With respect to previous analysis of the time dependence of H(t) [32], we can observe that the daily variability of single stocks is much larger than that of global indices over long times. In addition also the average is appreciably larger.

A general result is that the value of H is systematically larger than 1/2. The usual interpretation would be to conclude that long range correlations are present [97]. However, in view of our previous discussion we would in-

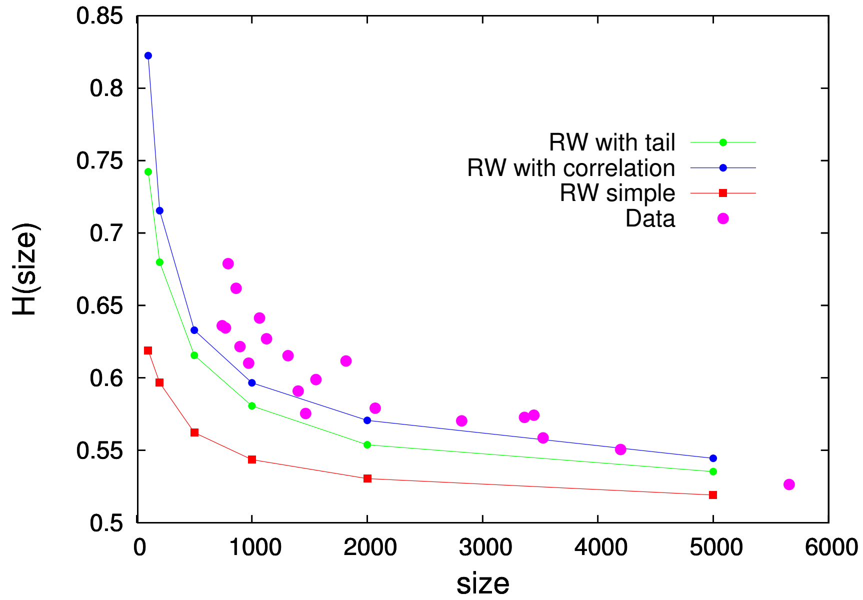


Figure 4.5: The value of $H(n)$ is shown for three different random walk models (normal, correlated and with fat tails) and for real data. Finite size effect already present in the normal random walk are amplified by the presence of fat tails and correlations. In the x-axis is plotted the effective size, that is $\frac{n}{10}$. The values are averaged over 1000 realizations.

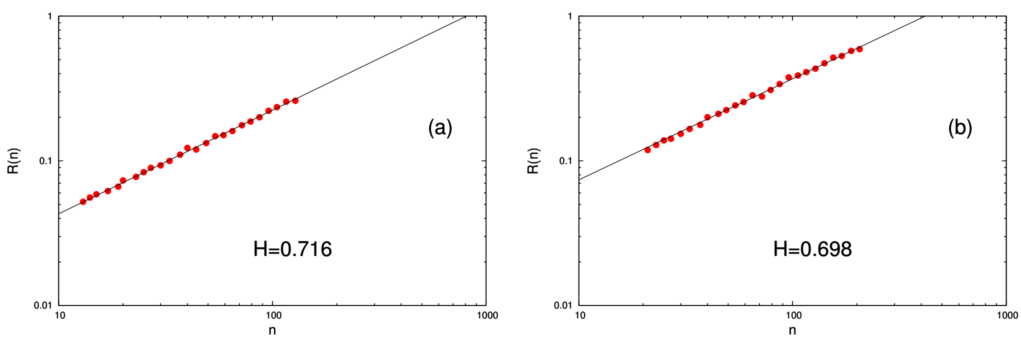


Figure 4.6: Analysis in terms of the Hurst exponent of the two stocks shown in Fig. 4.1. The case (a) refers to the stock which appears smooth, while (b) is the other one. One can see that the value of H is very similar despite the apparent differences between the two behavior (Fig. 4.1).

stock	H	H_{max}	H_{min}	σ	$< N >$
AH	0.599	0.732	0.489	0.0215	1535.77
AVO	0.615	0.785	0.501	0.0170	1296.71
BA	0.573	0.694	0.478	0.0106	3323.37
BRO	0.662	0.792	0.557	0.0161	853.91
CAI	0.641	0.751	0.478	0.0232	1052.58
DRI	0.575	0.699	0.445	0.0106	1446.65
GE	0.526	0.653	0.406	0.0065	5598.83
GLK	0.627	0.780	0.484	0.0105	1114.01
GM	0.574	0.677	0.462	0.0083	3405.84
JWN	0.579	0.738	0.457	0.0125	2025.67
KSS	0.570	0.686	0.438	0.0135	2789.09
MCD	0.559	0.691	0.417	0.0076	3480.63
MHS	0.612	0.750	0.460	0.0113	1792.51
MIK	0.591	0.752	0.456	0.0132	1377.84
MLS	0.635	0.914	0.496	0.0204	759.27
PG	0.551	0.662	0.456	0.0091	4135.80
TXI	0.636	0.776	0.473	0.0296	733.68
UDI	0.679	0.781	0.524	0.0147	774.25
VNO	0.622	0.777	0.506	0.0244	883.78

Table 4.1: Hurst exponent for 20 NYSE stocks. H is the average daily value over the 80 days. H_{max} and H_{min} are the maximum and the minimum and σ the variance. $< N >$ is the average number of transactions per day.

stead propose that this deviation from 1/2 is precisely due to finite size effects, combined with the fat tail phenomenon. A further support to this interpretation is that if we built a long time series by eliminating the night jumps, one observes a convergency towards the value 1/2. Also one may note that stocks with a relatively large number of transactions per day ($< N >$), like for example GE stock, are much closer to the random walk value $H = 1/2$.

The fact that apparently different profiles with respect to the roughness lead to value of H which are very similar is due to a variety of reasons. The overall enhancement with respect to the standard value 1/2 is, in our opinion, mostly due to the finite size effects phenomenon. However, this does not explain why two profiles which appear very different, like those in Fig. 4.1, finally, lead to very similar values of H. This is probably due to the fact that the Hurst's R/S analysis tends to mix the role of trends with fluctuations and in the next section we are going to propose a different method to resolve this problem.

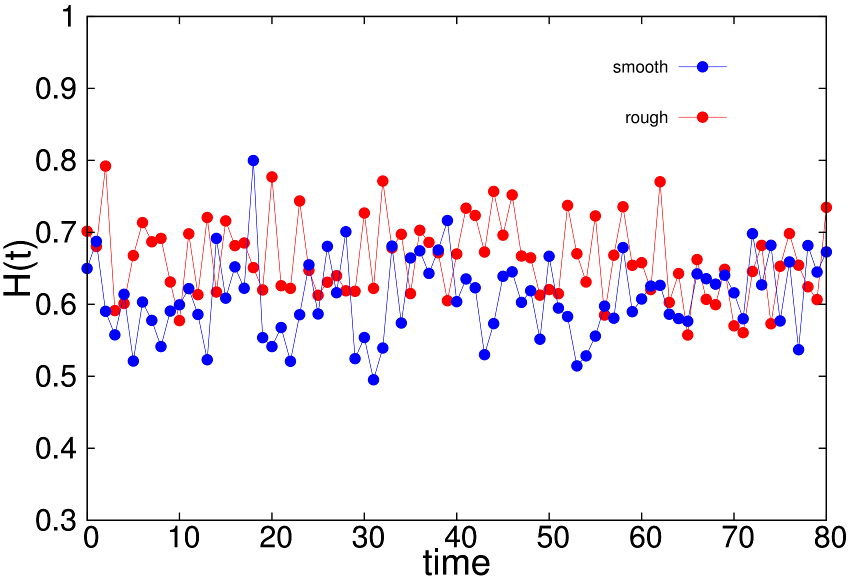


Figure 4.7: Time dependence of the Hurst exponent $H(t)$ for the two stocks shown in Fig. 4.1.

To complete our analysis, we consider the generalized Hurst exponent in the spirit of Ref. [12]. To this purpose we analyze a q-th order price difference correlation function defined by:

$$G_q(\tau) =< |P(t) - P(t+\tau)|^q >^{\frac{1}{q}} \tag{4.18}$$

The generalized Hurst exponent H_q can be defined from the scaling behavior of $G_q(\tau)$:

$$G_q(\tau) \sim \tau^{H_q} \tag{4.19}$$

For a simple random walk $H_q = H = 1/2$ independently of q. We have calculate the function $G_q(\tau)$ for the two test-stocks.

The results are shown in Fig. 4.8 and show that H_q is not a constant but strongly depends on q. This result provides an evidence that the characteristics of the profile are dominated by the large jumps due to the fat tail properties that are present in the plotted data.

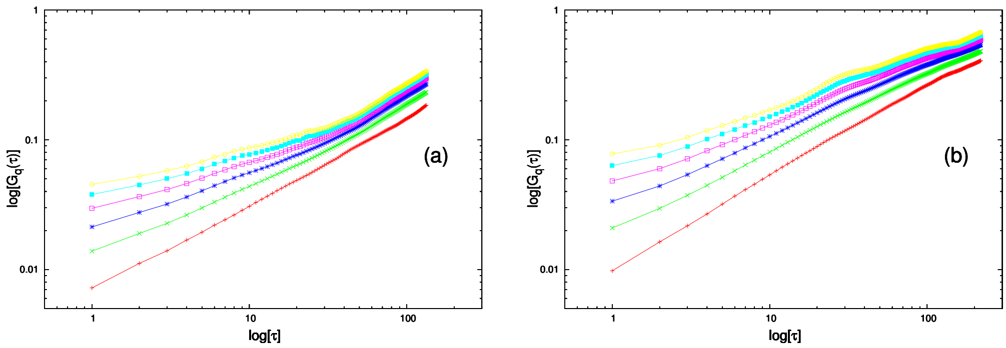


Figure 4.8: $G_q(\tau)$ as a function of τ in a log-log scale for the two test-stock ((a) is the smooth and (b) the rough one). In both (a) and (b), from bottom to top $q = 1, 2, 3, 4, 5, 6$.

4.5 New approach to roughness as fluctuation from Moving Average

In this section we consider a new method to characterize the roughness. The basic idea is to be able to perform an automatic detrendization of the price signal. This can be achieved by the difference between the price variable and its moving average defined in an optimal way. At each transaction point t_i we define the moving average of the price $P(t_i)$, with a characteristic time τ, as:

$$P_\tau(t_i) = \frac{1}{N_\tau}\sum_j P(t_j) \tag{4.20}$$

where N_τ are the number of transactions in the time interval $[-\tau/2 : \tau/2]$. This function corresponds to the symmetric average over an interval of size N_τ around t_i.

One can then consider the maximum deviation of $P(t_i)$ from $P_\tau(t_i)$ over an interval of a certain size, in our case we consider a single day:

$$R_\tau = \max_i |P(t_i) - P_\tau(t_i)| \tag{4.21}$$

This may appear similar to the standard definition of roughness which gives the absolute fluctuation in a time interval τ. Instead the use of R_τ corresponds to an automatic detrendization which appears more appropriate

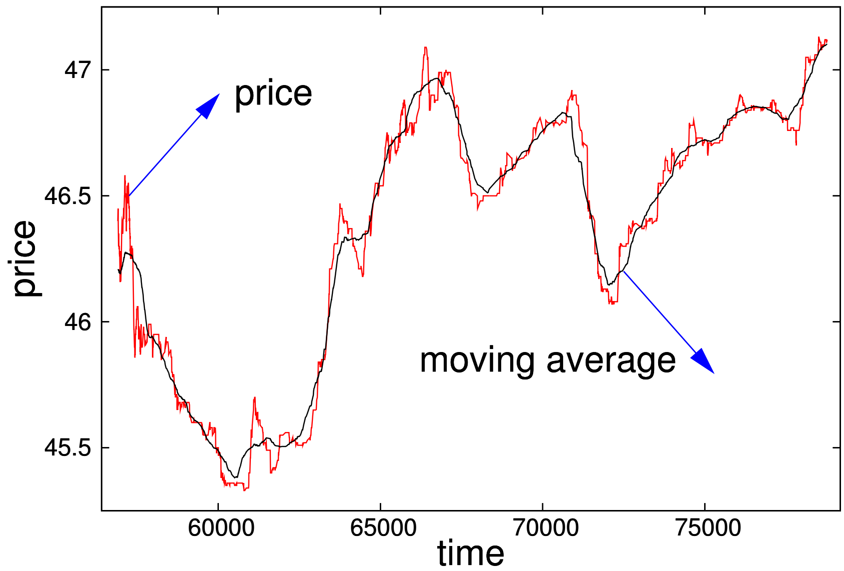


Figure 4.9: Example of price fluctuations and the corresponding moving average. In our case we consider a symmetrized moving average defined as the average of the price over a symmetric interval of total size τ.

to study the roughness. Our approach is similar to the one of Ref. [2], but with the difference that we use a symmetrized definition of the moving average while Ref. [2] defines the moving average only with respect to a previous time interval.

In Fig. 4.10 we show the values of R_τ for the two stocks shown in Fig. 4.1 and, for comparison, the same stocks analyzed with the R/S analysis. One can see that the fluctuations from the moving average are more appropriate to describe the difference between these stocks which cannot be detected with the standard Hurst's R/S analysis.

4.6 A finite range test for correlation in time series

Detecting complex correlations in finite size time series is not an easy task. Linear correlations are easy to identified by means the usual statistical tools, such as the autocorrelation function. If one has to identify some high order correlations, for example three point correlations, the simple correlation function become hard to calculate expecially in the case of short time series.

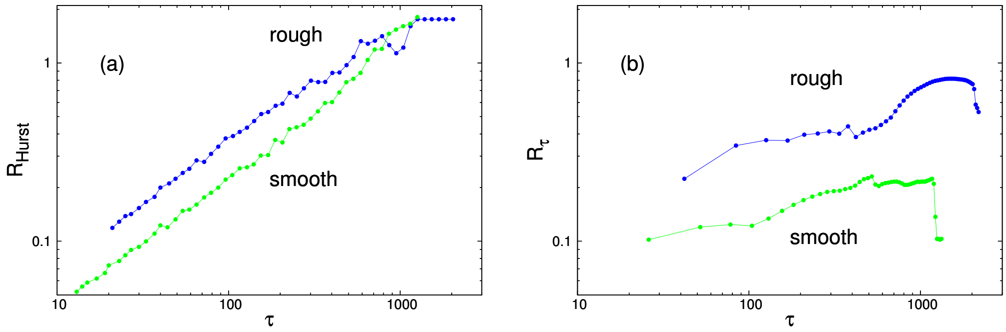


Figure 4.10: (a) Fluctuations over intervals of different size (τ) defined by the differences between maximum and minimum values over intervals of size τ. These curves were used in the previous sections to compute the Hurst exponent in the standard way. The two curves refer to the two stock of Fig. 4.1. One can observe that the slope is similar for the two cases and, at relatively large scale (τ), even the amplitudes become rather similar. (b) In this case the amplitudes are defined by the fluctuations from the moving averages as given by Eq. (4.21). In this case there is a marked difference in slope and even more in amplitude. This example clarifies that this new definition of roughness can be more useful to classify the stock dynamics.

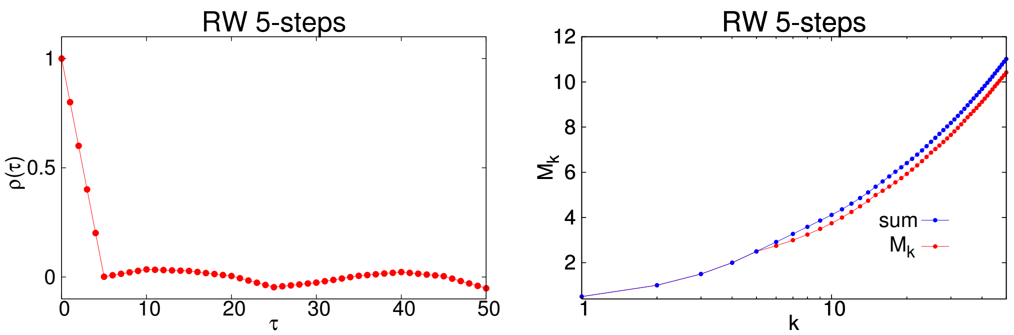
In this perspective, one of main challenge is to find new statistical tools able to identify complex correlations from experimental data. In this section we present a new tool to analyze real time series which seems to work in this direction [9]. To introduce this method to to extract information about correlations in time series, let's start recalling the Eq. 4.13 valid for of a finite RW. We write again Eq. 4.13 for a clearness discussion.

$$\mathbb{E}(M_k) = \sum_{i=1}^{k} \frac{\mathbb{E}(|X_i|)}{2i}$$

This result is valid for any uncorrelated process X which is the sum of limited and symmetrically distributed increments. In section 4.3 we have verified that this equation is valid for a simple Random Walk. One expects that this equation is not still valid considering correlated processes. Our strategy to detect correlations is to separately evaluate both the sides of Eq. 4.13 and after check if they match or not. Our aim is to verify if this method is able to detect complex correlations in data. We have chosen three different processes to test our method. The first is simple k-steps correlated Random Walk which has simple linear correlations. The second process we have considered is the ARCH(1) process that we have described in section 3.3. This process has no linear correlations but displays some higher order correlations. At last we analyze the two dataset described Fig. 4.1. As we have shown is sections 4.2 and 4.4 these two stocks look very different but have the same roughness properties and no linear correlations. We will see that this method is suitable to distinguish these two datasets in terms of their complex correlations properties.

4.6.1 Simple k-steps correlated RW

The simplest process for illustrating the method proposed here is a RW with a simple correlation in the increments. So we consider a RW with increments η taking the values $+1$ with probability $p = 1/2$ and -1 otherwise with the further rule that if $\eta_i = 1$ then $\eta_{i+1} = \eta_{i+2} = ... = \eta_{i+k-1} = 1$ and the same if $\eta_i = -1$. We have generated a correlated RW with this rules and we have calculated both the members of Eq. 4.13 performing the averages over an ensemble of 1000 realizations. In Fig. 4.11(a) is plotted the autocorrelation function $\rho(\tau)$ of the increments $\delta x_i = X_{i+1} - X_i$ for a 5-steps correlated RW. In Fig. 4.11(b) are plotted both the side of Eq. 4.13 as function of the RW size k.



(a) (b)

Figure 4.11: In this figure are shown the results for a 5-steps correlated RW. In (a) is plotted the autocorrelation function $\rho(\tau)$ for the increments δx. This function is zero when $\tau > 5$. In (b) are plotted the two side of Eq. 4.13 as a function of the size of the RW. We can see that the correlations in the increments δx lead to a mismatch of the two side.

As we can see from Fig. 4.11 the presence of positive correlations in the increments η is reflected in a mismatch in Eq. 4.13. We can see that the right-side value of equation Eq. 4.13 is greater then the left-side ($\mathbb{E}(M_k)$).

4.6.2 ARCH process

To test our method we use the simple ARCH(1) process we have described in section 3.3. We recall the ARCH(1) model equation for the variance σ_i^2 and the increments η_i:

$$\sigma_i^2 = \alpha_0 + \alpha_1 \eta_{i-1}^2 \qquad \eta_i = \sigma_i \epsilon_i, \tag{4.22}$$

where ϵ is a Gaussian random variable with zero mean and zero variance.

It is easy to see that the autocorrelation function for the increments η is a delta function but the ARCH(1) process shows some higher-order correlation. In particular the correlation of square increments have an exponentially decaying correlation function.

$$\rho(\tau) =< \eta_i^2 \eta^2 x_{i+\tau} >= A e^{-\tau/\lambda}, \tag{4.23}$$

where λ and A are functions of α_0 and α_1. In Fig. 4.12 are shown the autocorrelation functions for the increments and for the square-increments for an ARCH(1) process. with parameters $\alpha_0 = 0.45$ and $\alpha_1 = 0.55$.

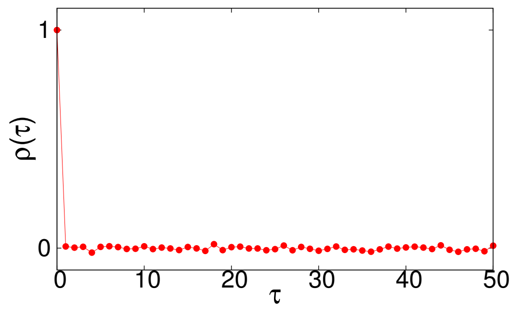
To see if our test is able to detect also these higher-order correlation, we have performed our analysis. In Fig. 4.13 we have plotted the two side of Eq. 4.13 as usual. As we can see from Fig. 4.13, although the autocorrelation function of the increments δx is a delta function, the two side of Eq. 4.13 do not match, showing that this method is able to identify and visualize the complex correlations of the process.

4.6.3 Test on Real Data from NYSE Stock-Market

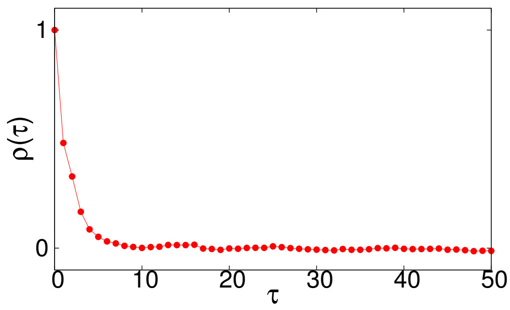
In this section we perform our correlation test on some data from the NYSE Stock-Price. In particular we want to check is our test is able to distinguish the two stock plotted in Fig. 4.1

To make our correlation test we calculate the value of $\mathbb{E}(M_k)$ and of the sums of the right side of Eq. 4.13. The expectation values are estimated splitting our data set in windows of size k and than averaging over all the values for each window.

We have choose the two time series in Fig. 4.1 to have the same statistical characteristics. The total fluctuation of the price during the day and also the volatility calculated in the day have very similar values for the two stocks.



(a)



(b)

Figure 4.12: In (a) the autocorrelation function for the increments of an ARCH(1) process is shown. The parameters of the model are $\alpha_0 = 0.45$ and $\alpha_1 = 0.55$. The autocorrelation function of price increments is a delta function. The autocorrelation function for the square-increments is plotted in (b) and we can see that it decays exponentially to zero.

Despite these similar features, we can see that the stock plotted in Fig. 4.1(a) looks very smooth with a well defined trend, instead the one in Fig. 4.1(b) has larger fluctuations and looks very rough. In this perspective we suppose to observe some kind of correlations in the stock plotted in Fig. 4.1. First we look to the autocorrelation function of the price increments. As we can see from Fig. 4.14 both the stocks plotted in Fig. 4.1 have uncorrelated price increments.

Now we try to detect some higher-order correlations using our test based on Eq. 4.13. We have calculated separately the two side of Eq. 4.13 subdividing the size of both data sets in windows of length n. In each window we numerically calculate M_k and $|X_k|$ then we average all the values. Varying the windows length we can perform our correlation test. The results are shown in Fig. 4.15. We can see from Fig. 4.15 that for the stock plotted in Fig. 4.1(a), our test points out a correlation that does not appear studying the only autocorrelation function of the price increments. On the other hand, for the stock plotted in Fig. 4.1(b), our test do not shows any correlations confirming the absence of trend that we can see "at sight".

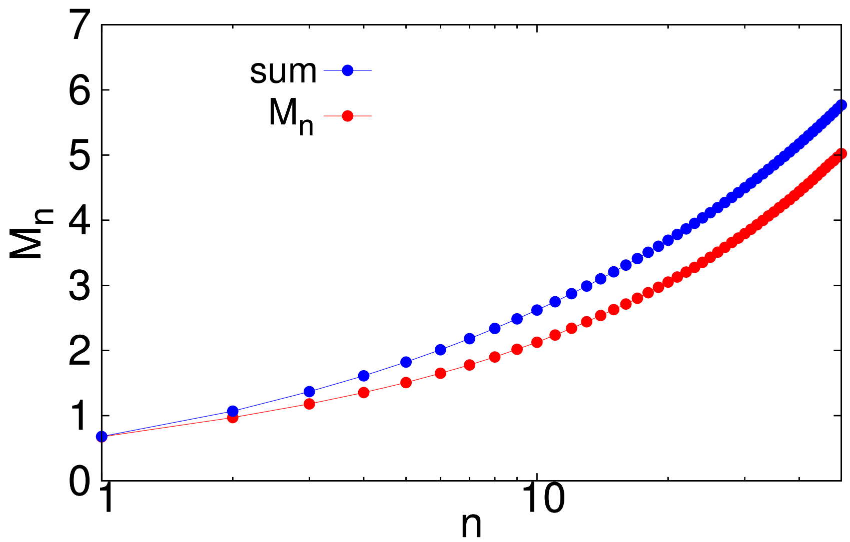
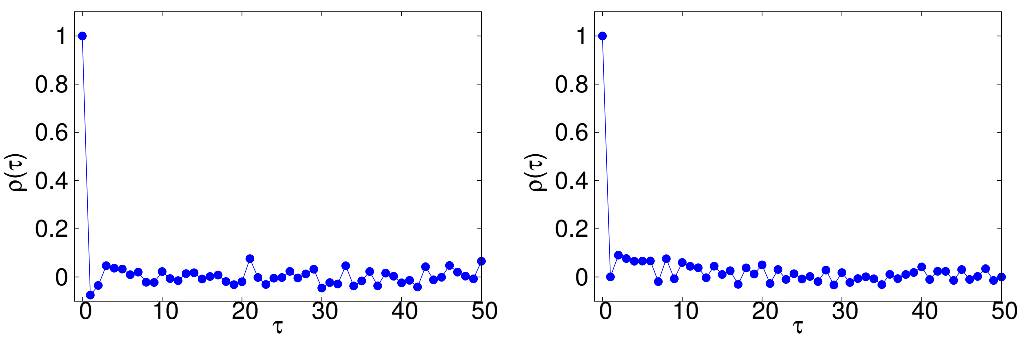


Figure 4.13: In this figure are plotted the two side of Eq. 4.13 for an ARCH(1) process with parameters $\alpha_0 = 0.45$ and $\alpha_1 = 0.55$. We can see that also in this case where the simple correlations are zero the two side of Eq. 4.13 do not match because of the presence of higher-order correlations.



(a) (b)

Figure 4.14: Autocorrelation function (ρ) for the price increments of the two stocks plotted in Fig. 4.1. One can see that there are no correlations between price differences for both the stocks.

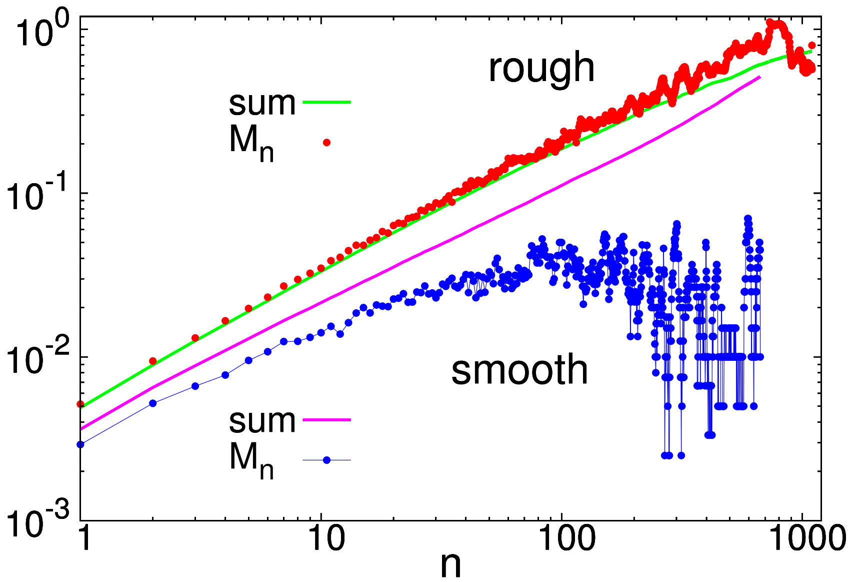


Figure 4.15: The two side of Eq. 4.13 are plotted for the two stocks in Fig. 4.1. We can observe that for the smooth stock with a trend, the one in Fig. 4.1(a), the two side do not match showing complex correlations not detectable with the usual statistical indicators. The stock plotted in Fig. 4.1(b) satisfies the Eq. 4.13. In this case our test does not reveal hidden correlations, as we expect.

4.7 Discussion and Conclusions

We have considered the roughness properties as a new element to characterize the high frequency stock-price fluctuations. The data considered include all transactions and show a large night jump between one day and the next. For these reasons the dataset are statistically homogeneous only within each day. This leads to a serious problem of finite size effects which we have analyzed by using various random walk models as examples. We have computed the effective Hurst exponent as a function of the size of the system. The basic result is that the finite size effects lead to a systematic enhancement of the effective Hurst exponent and this tendency is amplified by the inclusion of fat tails and eventual correlations.

An analysis of real stock-price behavior leads to the conclusion that most of the deviations from the random walk value ($H = 1/2$) are indeed due to finite size effects. Considering the importance of non-stationarity phenomenon one may conjecture that the finite size effects could be important even for long series of data.

Concerning the roughness analysis we conclude that the standard Hurst's R/S analysis is not very sensitive in order to characterize the various stock-price behaviors. We propose a different roughness analysis based on the fluctuations from a symmetrized moving average. This has the advantage of an automatic detrendization of the signal without any *ad hoc* modification of the original data. This new method appears much more useful than the standard one in order to characterize the fluctuations behavior of different stock as shown clearly by the analysis of the two cases in Fig. 4.1. An other improvement has be obtained considering a new tool for detecting complex correlations in finite size time series.

Chapter 5

Hidden Forces and Fluctuations from Moving Averages

The concept of moving average is very popular in empirical trading algorithms [101] but, up to now, it has received little attention from a scientific point of view [85, 88, 30]. In [6] we have proposed that a new definition of roughness can be introduced by considering fluctuations from moving averages with different time scales. This new definition seems to have various advantages with respect to the usual Hurst exponent in describing the fluctuations of high frequencies stock-prices.

A more specific analysis of these fluctuations can be found in two recent papers [19, 114, 5] which attempt to determine the tendency of the price to be attracted or repelled from its own moving average (Fig. 5.1). This is completely different from the use of moving averages in finance, in which empirical rules and predictions are defined in terms of a priori concepts [101]. The idea is instead to introduce a statistical framework which is able to extract these tendencies from the price dynamics.

5.1 The Effective Potential Model

The basic idea is to describe price dynamics in terms of an active random walk (RW) which is influenced by its own moving average. This induces complex long range correlations which cannot be determined by the usual correlation functions and that can be explored by this new approach [19, 114]. The basic ansatz is that price dynamics $P(t)$ can be described in terms of a stochastic equation of type:

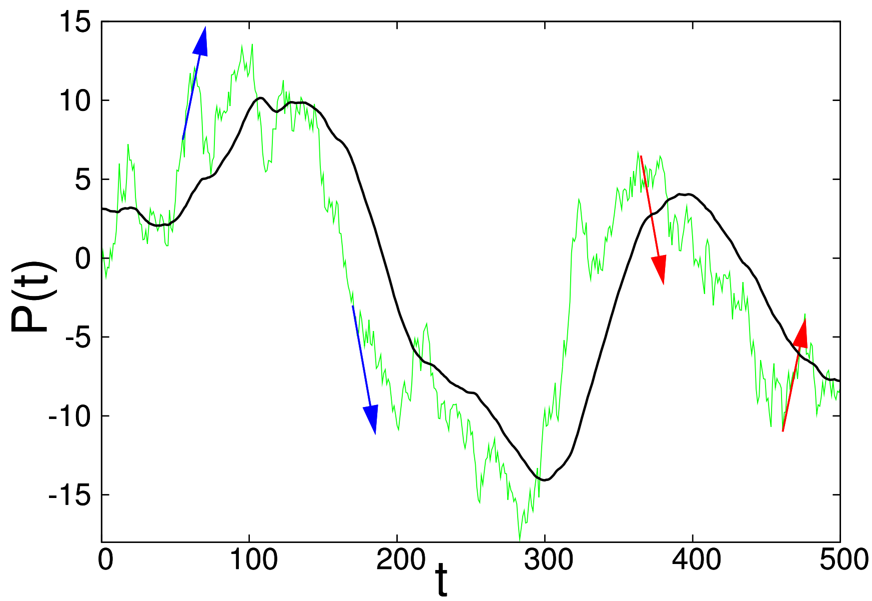


Figure 5.1: Example of a model of price dynamics (in this case a simple random walk) together with its moving average defined as the average over the previous 50 points. The idea is that the distance of the price from its moving average can lead to repulsive (blue arrows) or attractive (red arrows) effective forces.

$$
\begin{aligned}
P & (t+1) - P(t) = \\
= & -b(t) \frac{d}{d\Big(P(t) - P_M(t)\Big)} \Phi\Big(P(t) - P_M(t)\Big) + \\
+ & \sigma(t)\omega(t)
\end{aligned}
\tag{5.1}
$$

where $\omega(t)$ corresponds to a random noise with unitary variance and

$$
P_M(t) \equiv \frac{1}{M} \sum_{k=1}^{M} P(t-k) \tag{5.2}
$$

is the moving average over the previous M steps.

The potential Φ together with the pre-factor $b(t)$ describe the interaction between the price and the moving average. In both approaches [19, 114] it is assumed to be quadratic:

$$\phi\Big(P(t) - P_M(t)\Big) = \Big(P(t) - P_M(t)\Big)^2. \tag{5.3}$$

The time evolution of the "price" of such a process, is shown in Fig. 5.2,

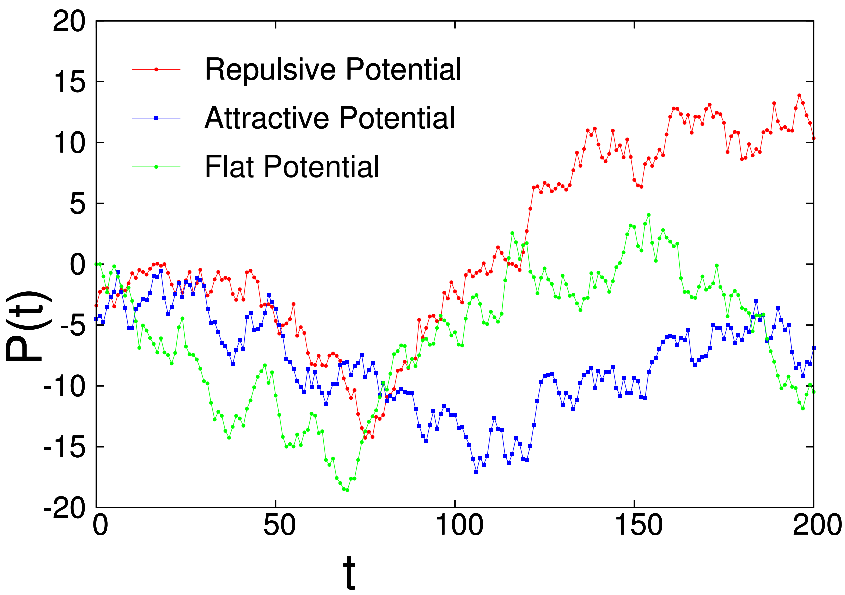


Figure 5.2: Time evolution of the price described by Eq. 5.1 in the case of a quadratic potential as Eq. 5.3. Three different behavior are plotted. The red lines represents the time evolution of a RW in a repulsive quadratic potential while the blue line is in an attractive quadratic potential. The green line is the case of flat potential (simple RW). The parameters are fixed to $M = 20$ and $b = \pm 1$. We can observe an over diffusion (under diffusion) in the case of repulsive (attractive) potential.

where we can observe three cases in which the potential is attractive, repulsive and constant (simple RW).

Despite this similar starting point the two studies proceed along rather different perspectives. In Ref. [19] the three essential parameters of the model $(b; M; \sigma)$ are considered as constants with respect to t. Then, by analyzing the price fluctuations over a suitable time interval and for a long time series, the values of the three parameters are identified.

In Ref. [114] instead the analysis is performed by looking directly at the relation between $P(t+1) - P(t)$ and $P(t) - P_M(t)$. In this way one can reconstruct the force of the process plotting $P(t+1) - P(t)$ as a function of $P(t) - P_M(t)$. Then, integrating from the center, one can obtain the potential. This permits to derive the form of the potential and to identify the parameter $b(t)$ and its time variation. In Fig. 5.3 are shown the potential obtained from a simulation of the process described in Eqs. 5.1 and 5.3 in the case of attractive force for various values of M.

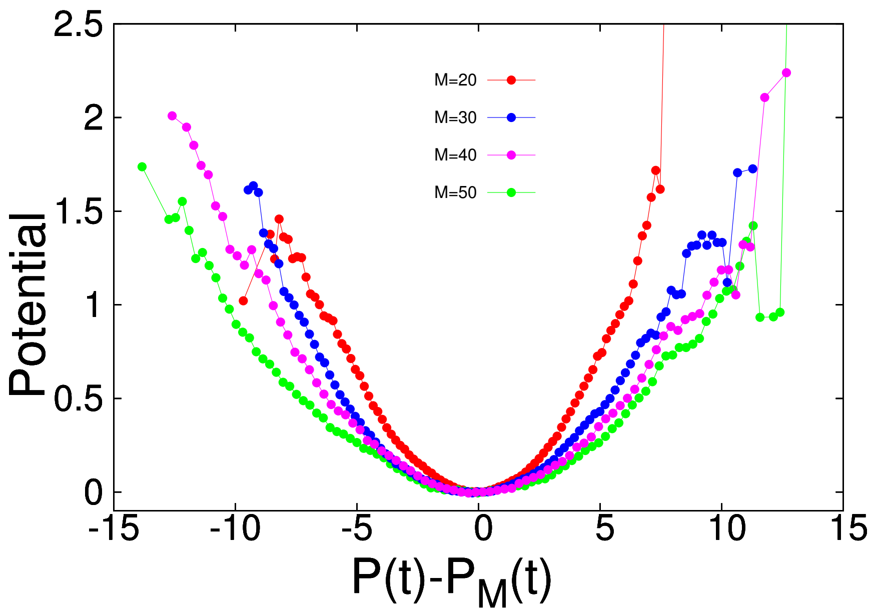


Figure 5.3: The plot shows the shapes of the quadratic attractive potentials defined by Eqs. 5.1 and 5.3. We can see that the amplitude of the potentials depend on the choice of the parameter M.

We can observe that the potentials have an amplitude (that is the slope of the linear force) which depends on M. In [114] is shown that such a dependence can be eliminated rescaling the potential by a factor $(M-1)$. This would imply that it is not necessary to specify the time scale of the moving average.

In Fig. 5.4 are shown the potentials plotted in Fig. 5.3, rescaled by the factor $(M-1)$. Indeed we can observe a good data collapse.

This idea of assuming a linear force in Eq. 5.1 has been tested on real data. In [114] a series of data from the Yen-Dollar exchange rates have been analyzed. The potential analysis for the case of the Yen-Dollar exchange rates indeed leads to the observation of rather quadratic potentials.

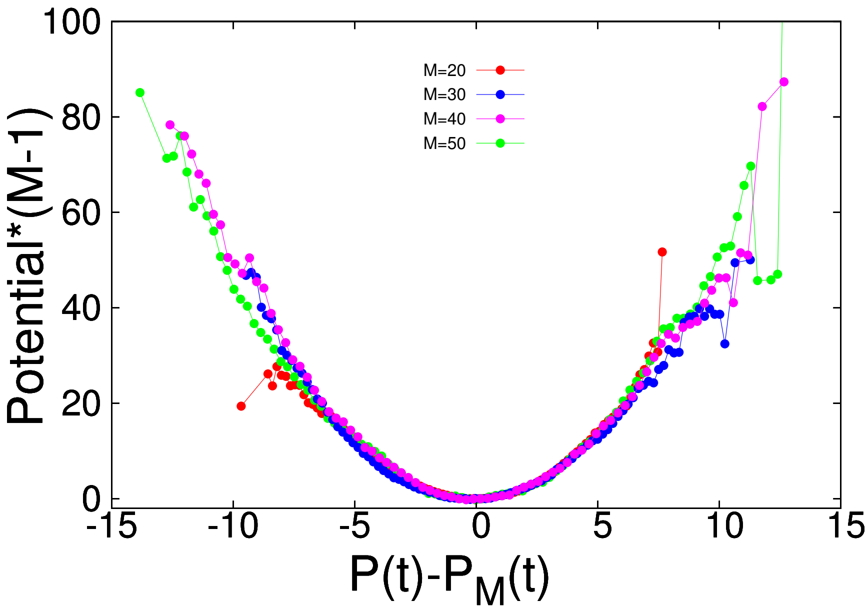


Figure 5.4: The different potential plotted in Fig. 5.3 are re-plotted scaling the potential with the factor $(M-1)$. We can see that in this way we obtain a good data collapse.

5.2 Test Studies

Given these different perspectives, which arise from the same basic model, we decided to perform a series of tests of this approach which we present in this thesis [5]. We believe that these tests can elucidate various properties and limitations of the new approach and represent a useful information for its future developments and applications.

In Fig. 5.1 we show a simple RW and a moving average which represents its own smoothed profile. The analysis is performed by plotting the values of $P(t+1) - P(t)$ as a function of $P(t) - P_M(t)$ and deriving the potential by integrating from the center [114]. The simple RW leads to a flat potential (no force) as expected (Fig. 5.5). Than we can take the smoothed profile (previous moving average) as a dataset by itself and repeat the analysis by comparing it to a new, smoother moving average (not shown). As one can see in Fig. 5.5 this leads to an apparent repulsive potential which should be considered as spurious. This is due to the fact that the smoothed curve implies some positive correlations as shown in Fig. 5.6. Therefore in this framework positive correlations lead to a destabilizing potential with respect to the moving average.

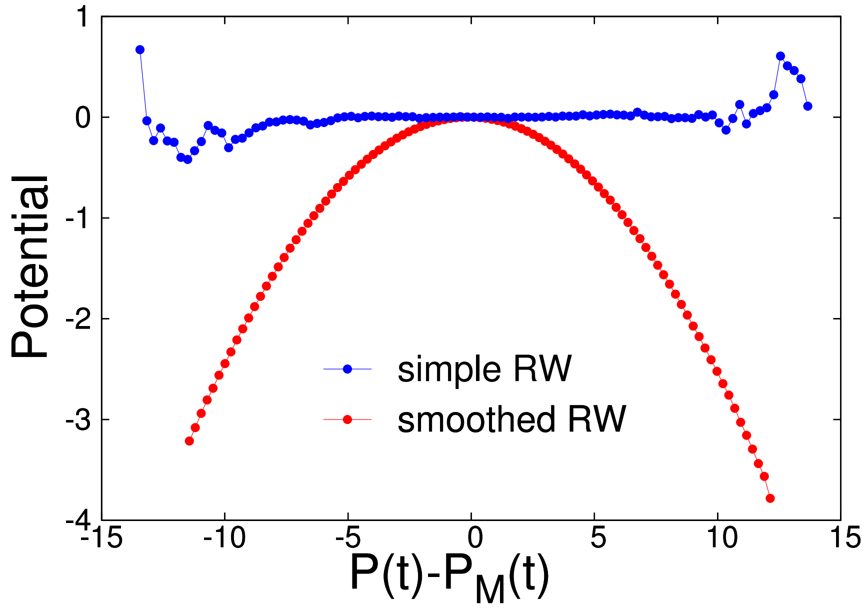


Figure 5.5: Effective potential for a random walk (flat line) and a smoothed random walk (convex parabola). The apparent repulsive potential corresponding to the smoothed RW is spurious and due to the correlations corresponding to the smoothing procedure. The units of the potential are defined by Eq.(5.1).

The opposite would happen for negative correlations (zig-zag behavior).

The interesting question is however if one can identify a non trivial situation in terms of the effective potential but in absence of simple correlations. This would be the new, interesting situation and the corresponding forces can be considered as hidden, in the sense that they do not have any effect in the usual correlation functions. Real stock-prices data clearly do not show any appreciable correlation, otherwise they would violate the simple arbitrage hypothesis. In the exchange rates instead there is a zig-zag behavior (negative correlation) at very short times which should be filtered with suitable methods in order to perform the potential analysis [114].

We now consider the model of the quadratic potential as in Refs. [19, 114]. The effective potential is easily reconstructed as shown in Fig. 5.7. We also show in Fig. 5.8 the behavior of the absolute price variations for different time steps. The correlation function for the price and volatility are shown in Fig. 5.9 which clarifies that, in this case, no simple correlation is present, nor is there any volatility clustering effect. This is an interesting result because it shows that the new method is able to detect hidden forces which have no

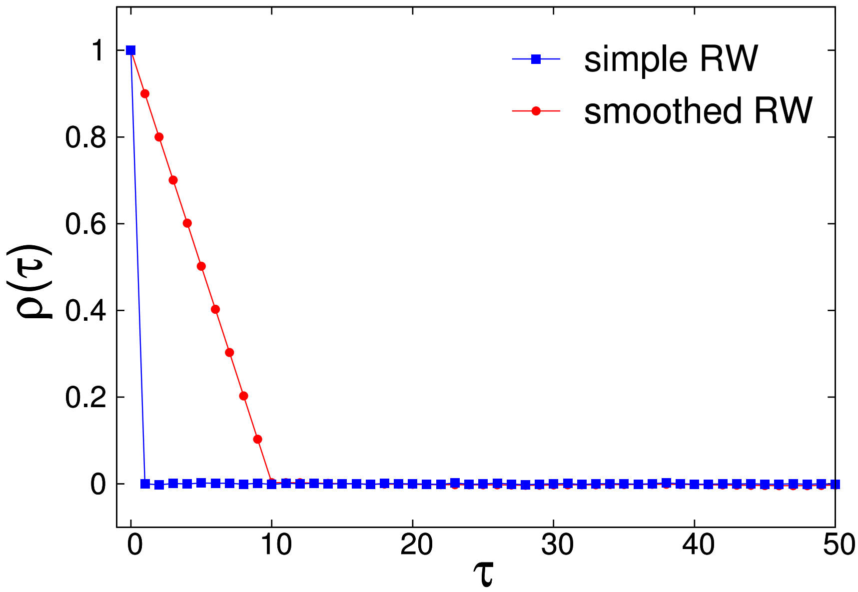


Figure 5.6: Autocorrelation (ρ) of the price increments for the simple RW and its smoothed profile. One can see that the smoothing procedure induces positive correlations up to the smoothing length (in this case 10 steps).

effect in the usual correlations of prices or volatility.

5.3 Probabilistic Models

We now consider some variations to the RW which depend on $P(t) - P_M(t)$. We modify the probability of a certain step rather than the size of the step as in Eq.(5.1). The simplest model is to add a constant drift, independent on the value of $P_M(t)$. The effective potential corresponding to this case is simply linear as shown in Fig. 5.10. One can see that in this case the point where $P(t) - P_M(t) = 0$ is not a special point and this model appears to be oversimplified with respect to the dataset analyzed up to now [19, 114].

A more interesting model is represented by the following dynamics for a RW with only up and down steps:

$$\begin{cases} p(\uparrow) = 1/2 + \epsilon_1 & \text{for} \quad P(t) - P_M(t) > 0 \\ p(\downarrow) = 1/2 - \epsilon_1 & \end{cases} \tag{5.4}$$
$$\begin{cases} p(\uparrow) = 1/2 - \epsilon_2 & \text{for} \quad P(t) - P_M(t) < 0 \\ p(\downarrow) = 1/2 + \epsilon_2 \;. & \end{cases}$$

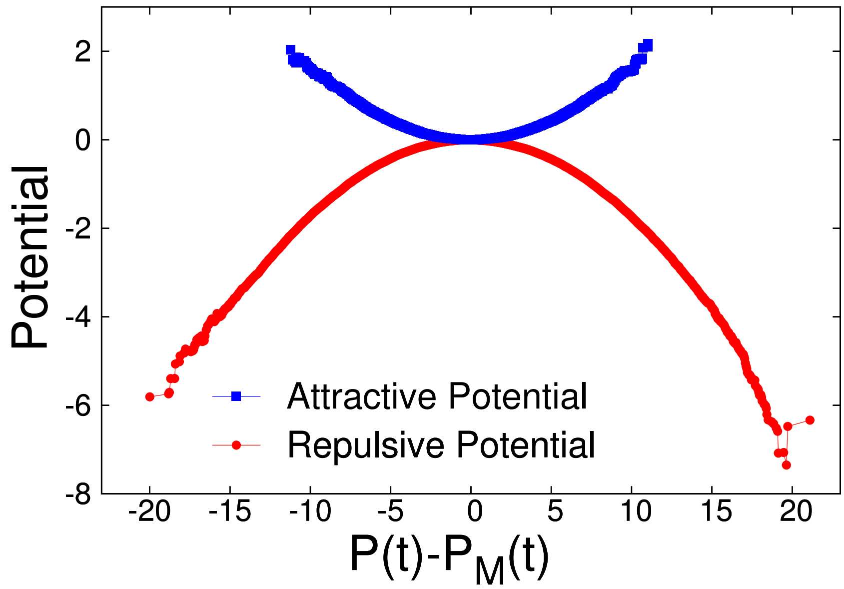


Figure 5.7: Effective potential reconstructed from a series of data obtained by a dynamics corresponding to Eqs.(5.1, 5.2) for the two cases of attractive and repulsive potentials. In this case $M = 30$ and $b = \pm 1$. The units of the potential are defined by Eq.(5.1).

This implies a tendency of destabilization (repulsion from $P_M(t)$) whose strength is only dependent on the sign of $P(t) - P_M(t)$. In principle the situation can be asymmetric with $\epsilon_1 \neq \epsilon_2$.

In Fig. 5.11 is shown the time evolutions of a price whose dynamical equations is given by Eq. 5.4 in the case of asymmetric repulsive potential. Also the comparison with a simple RW is shown.

The potential analysis for this case leads to a piecewise linear potential in which the slopes are related to ϵ_1 and ϵ_2 (Fig. 5.12).

One can also see that one line extends more than the other indicating an asymmetric distribution. Also in this case the correlation of the price variations and volatilities show no detectable effect as shown in Fig. 5.13. Clearly in this case the effective potential is just a representation of the correlations between $P(t+1) - P(t)$ and $P(t) - P_M(t)$ whose microscopic origin is instead in the modification of the probability for unitary steps.

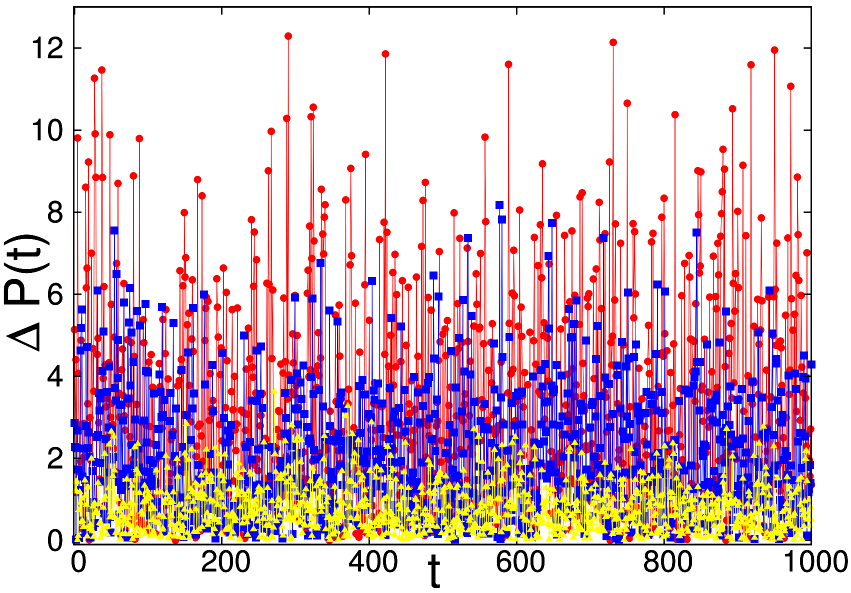


Figure 5.8: Absolute price variations for different time steps($\tau = 1$ (yellow); $\tau = 5$ (blue); $\tau = 10$ (red)) corresponding to the dynamics of Eqs.(5.1, 5.2).

5.4 Fractal Model

It may be interesting to consider also the case of a fractal model constructed by an iterative procedure [85], Fig. 5.14.

The fractal model does not have a specific dynamics but, since it is often considered as to capture some properties of real prices, we consider of some interest to study if this model would correspond to some type of effective potential. In Fig. 5.15 we can see that the effective potential is slightly attractive. Given the symmetry of the model construction, the asymmetry observed in the effective potential is probably due to the backward construction of the corresponding moving average.

5.5 Analysis of the Fluctuations

We now consider the nature of fluctuations from the moving average by analyzing the probability distribution $W\Big(P(t) - P_M(t)\Big)$ for the various models. In Fig. 5.16 we show the distributions corresponding to the quadratic potential as compared to that of a reference RW ($b = 0$). The first observation is that the repulsive potential makes the distribution broader (super diffusion)

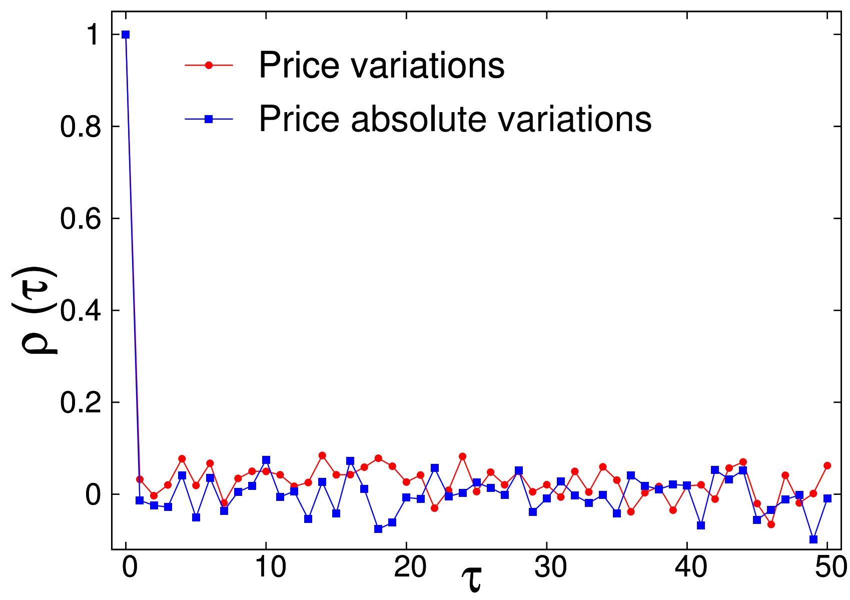


Figure 5.9: The correlation analysis of price variations shows no correlations between price differences and no volatility clustering effect. This implies that the presence of attractive or repulsive forces with respect to the moving averages is not detectable with the usual statistical indicators.

while the attractive potential makes it narrower (sub diffusion).

This behavior was already observed in Refs. [19, 114]. Less trivial is the fact that the distributions are well represented by Gaussian curves.

In Fig. 5.17 we show the same distributions corresponding to the probabilistic model of Eq. (5.4) for the case of asymmetric attractive and repulsive effects. In this case there is a marked deviation from the Gaussian behavior and the case of repulsive trend develops two separate peaks. It will be interesting to check the corresponding distribution on real stock-prices which we intend to perform in the future.

5.6 Application and test on an agent based model

It is instructive to analyze the effective potential scenario in agent-based models, where the price process is not defined explicitly but only through the aggregate choices of a group of traders. The simplest and most studied framework from a Statistical Physics perspective is that of Minority Games

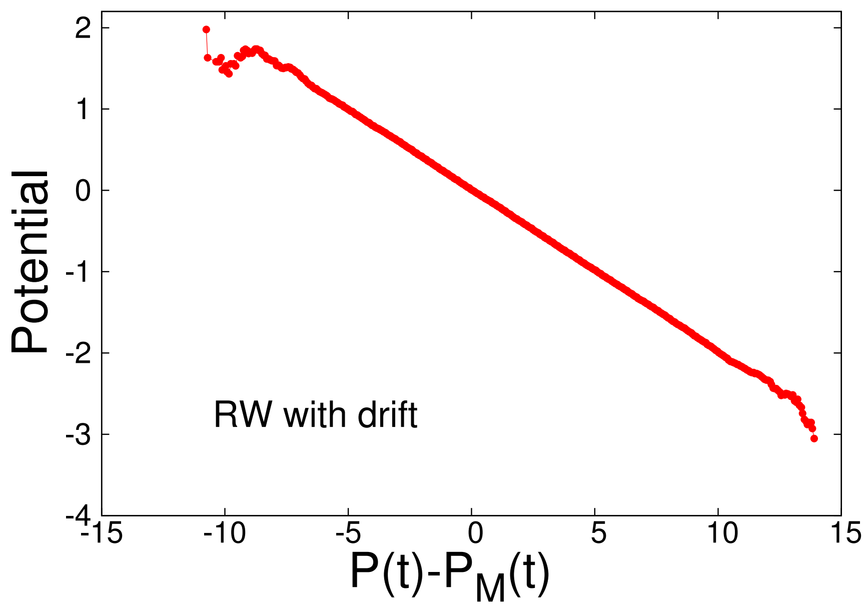


Figure 5.10: Effective potential corresponding to a RW with a constant drift which alters the probability for a step up or down. The units of the potential are defined by Eq.(5.1).

[34, 39], in which each of N agents must decide at every (discrete) time step whether to buy ($a_i(t) = 1$) or sell ($a_i(t) = -1$) an asset. The resulting price process is determined by the decisions of all agents through the "excess demand" $A(t) = \sum_{i=1}^{N} a_i(t)$. In particular, neglecting liquidity effects for the sake of simplicity, one can write that

$$P(t+1) - P(t) = A(t), \tag{5.5}$$

which amounts to defining the (log-)price as $P(t) = \sum_{t'<t} A(t')$.

It is clear that an agent's trading behavior will depend on his expectations about the future price increment $A(t)$, denoted by $\mathbb{E}_i[A(t)]$. For example, it has been argued [90] that if

$$\mathbb{E}_i[A(t)] = \psi_i A(t-1) + (1-\psi_i)A(t-2), \tag{5.6}$$

agent i behaves as a trend-follower for $\psi_i > 1$ (correspondingly he perceives the market as a Majority Game with payoff $\pi_i(t) = a_i(t)A(t)$), while he behaves as a fundamentalist for $0 < \psi_i < 1$ and plays a Minority Game with payoff $\pi_i(t) = -a_i(t)A(t)$.

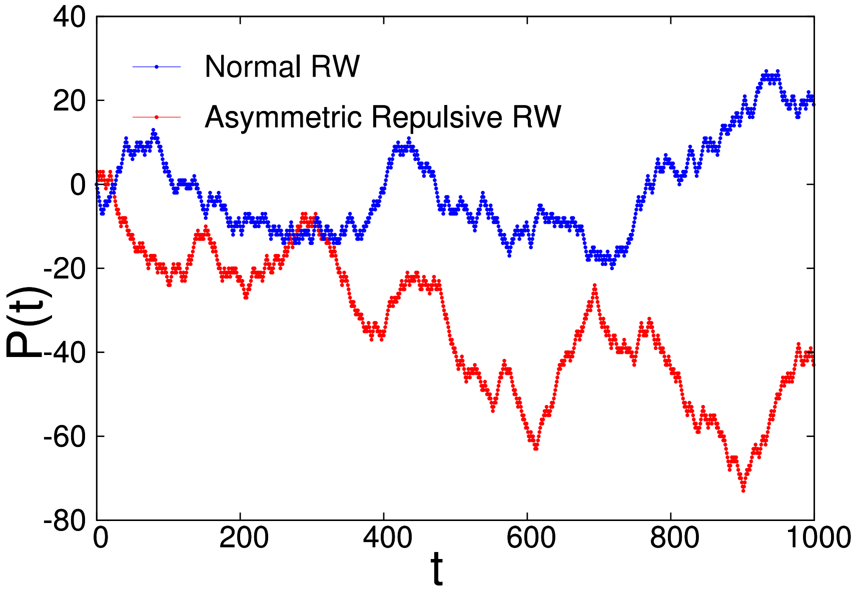


Figure 5.11: In figure (a) is plotted the time evolution the price whose dynamics is described by Eq. 5.4 , compared with the time evolution of a simple RW.

Let us now consider an agent who forms expectations with the simple assumption to adopt a linear dependence (quadratic potential):

$$P(t+1) - P(t) \propto P(t) - P_M(t) \tag{5.7}$$

It is easy to see that such an agent is described by a generalization of (5.6). Indeed, a direct calculation shows that (5.7) corresponds to

$$\mathbb{E}_i[A(t)] \propto \sum_{\tau=1}^{M-1} \frac{M-\tau}{M} A(t-\tau). \tag{5.8}$$

Agents thus tend to discount events further back in time and give larger weight to recent price changes when estimating the future returns. Clearly, such an agent has a more complicated reaction pattern than a pure Minority or Majority Game player and will be described by a payoff function that accounts for the possibility of behaving differently in different market regimes.

Models of this type have been introduced recently and appear to be an ideal testing ground to verify the emergence of the effective potential scenario

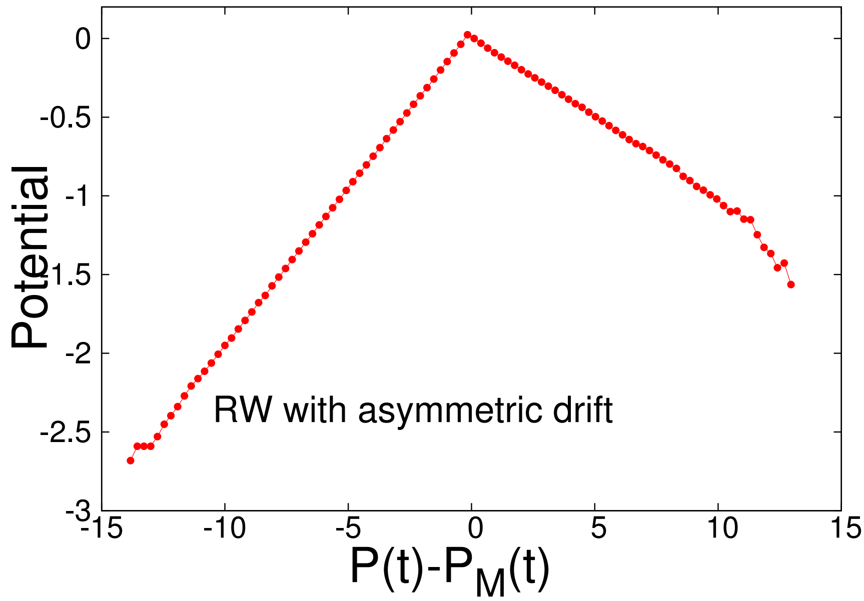


Figure 5.12: Effective potential corresponding to the dynamics of Eq.(5.4) with $\epsilon_1 = 0.05$ and $\epsilon_2 = 0.10$. One can see that in this case the distribution is asymmetric and it extends more in the direction for which the instability is stronger. In this model the effective force only depends on the sign of $P(t) - P_M(t)$ and not on its specific value. The units of the potential are defined by Eq.(5.1).

in a microscopic setting [7]. Specifically, we have tested it on a model in which agents may switch from a trend-following to a fundamentalist attitude (and vice-versa) depending on the market conditions they perceive, which was introduced in Ref. [94]. We refer the reader to the literature for a detailed account of the model's definition and properties. In a nutshell, it describes agents who strive to maximize the payoff

$$\pi_i(t) = a_i(t)[\mathcal{A}(t) - \epsilon\mathcal{A}(t)^3], \tag{5.9}$$

where $\mathcal{A}(t) = A(t)/\sqrt{N}$ is the normalized excess demand. The idea is that for small price movements ($A(t) \simeq 0$) agents perceive the game as a Majority Game as they try to identify profitable trends. However when price movements become too large, the game is perceived as a Minority Game, i.e. agents expect the price to revert to its fundamental value. As in most Minority Games, agents have fixed schemes ('strategies') to react to the receipt of one of P possible external information patterns and learn from experience

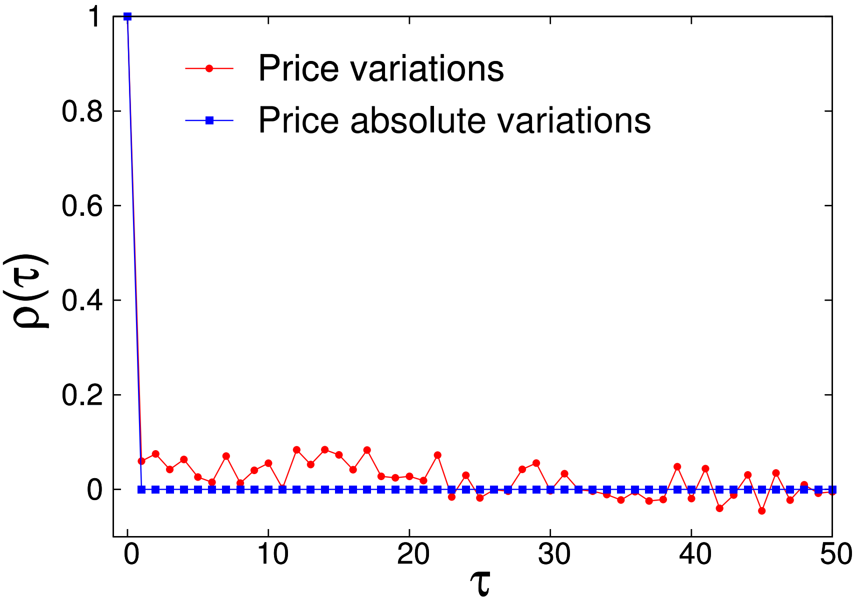


Figure 5.13: Correlation analysis of price variations and volatility for the model of Eq. (5.4). Also in this case no detectable correlations are present.

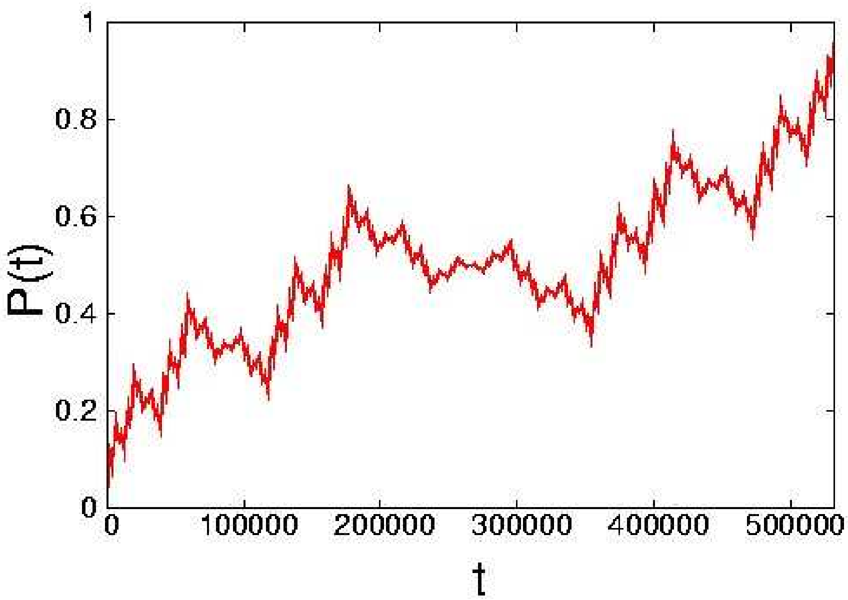


Figure 5.14: Example of a fractal model distribution of price

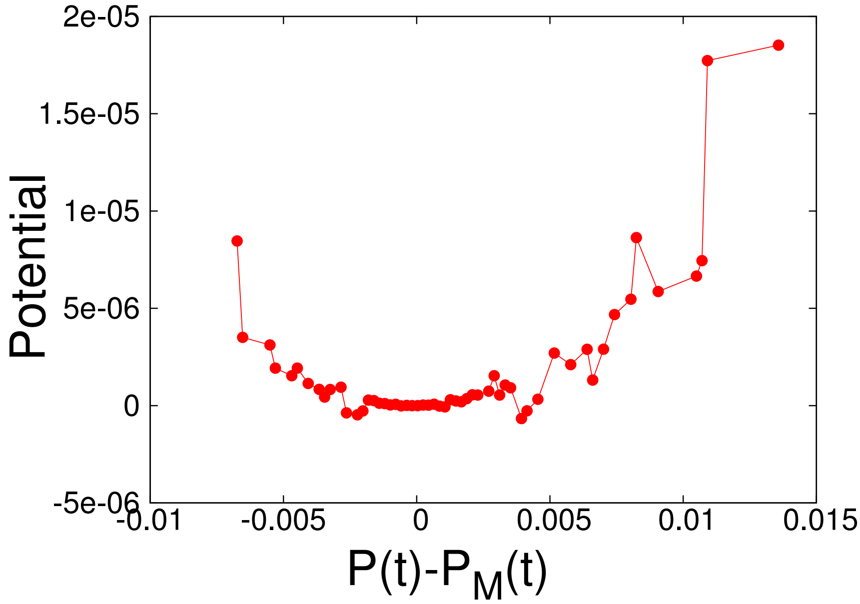


Figure 5.15: Effective potential corresponding to the fractal price model. The units of the potential are defined by Eq.(5.1).

to select the strategy and, in turn, the action $a_i(t)$ that is more likely to deliver a positive payoff. A realistic dynamical phenomenology is obtained in a whole range of values of the model parameter ϵ when the number N of players is large compared to the amount of information available to them P (this is measured by a parameter $\alpha = P/N$, see [94] for details).

In Fig. 5.18 is shown the time evolution of $P(t)$ for a game with parameters $\alpha = 0.05$ and $\epsilon = 1$. This choice of parameters corresponds to be in the range in which the competition between trend followers and contrarians is stronger. In fact, in Fig. 5.18, we can observe some "ordered" periods, where $A(t)$ is small and well defined trends in the price dynamics appear, but also "chaotic" periods where the dynamics of the price is dominated by the contrarians. In Fig. 5.18 we have identified two periods in which the different behaviors of the agents are well defined and we have used these periods as dataset for our potential analysis.

In Fig. 5.19 are plotted the potentials obtained performing the effective potential analysis with $M = 20$. We can observe that, when the market is dominated by contrarians, we obtain an attractive potential. This shape of the potential reproduces the agents' tendency to keep the price near its "fundamental" value. We can also note that this potential is not perfectly quadratic as in model described in Eqs. 5.1 and 5.3. In fact, plotting different

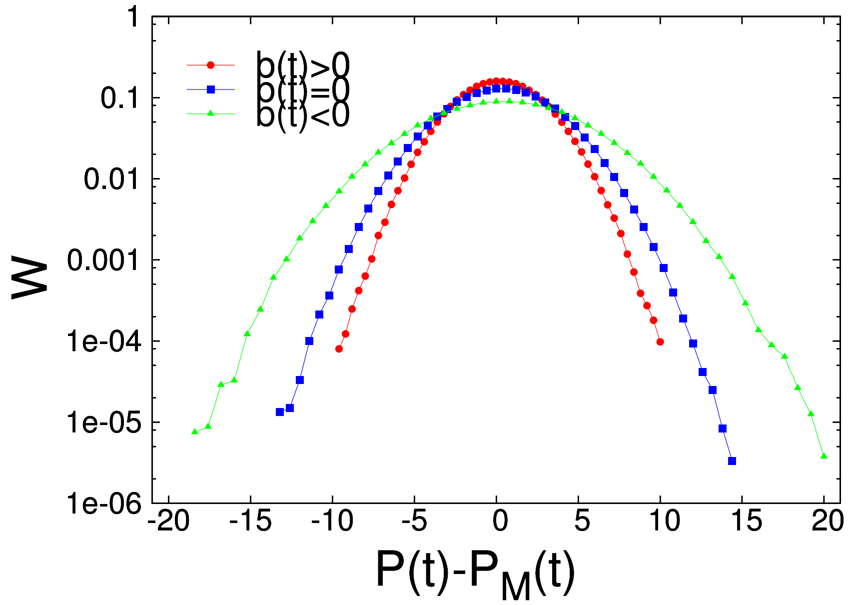


Figure 5.16: Distribution of the fluctuations, $W(P(t) - P_M(t))$, for the dynamics of Eq. (5.1-5.3) and different values of the parameter b.

potentials with various values of M we can not obtain a data collapse scaling the potentials with the factor $(M-1)$. In case of market dominated by trend followers, we can observe the presence of well defined trends (bubbles and crashes). In this case the agents try to follow the trends and the price tends to go away from its fundamental value. In this case we obtain a repulsive potential.

Therefore, the potential analysis is able to detect the agents' behavior based on microscopic rules only analyzing the data of a macroscopic variable, $P(t)$.

From the viewpoint of modeling real markets, it will be very interesting to introduce an agent based model in which agents perform their decision (buy/sell) by considering their expectations about the next price increment using Eq. 5.8, with different constant of proportionality and different values of the 'memory' M and not on the basis of a set of given strategies (as in the minority game framework). Work along these lines is currently in progress.

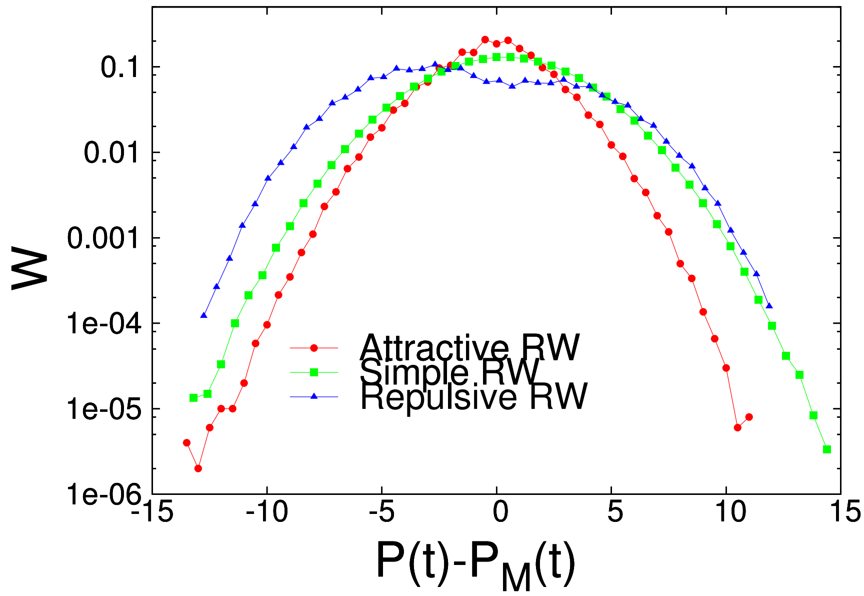


Figure 5.17: Distributions of the fluctuations, $W(P(t) - P_M(t))$, for the dynamics of Eq. (5.4). In this case the distributions became asymmetric due to different values of ϵ_1 and ϵ_2.

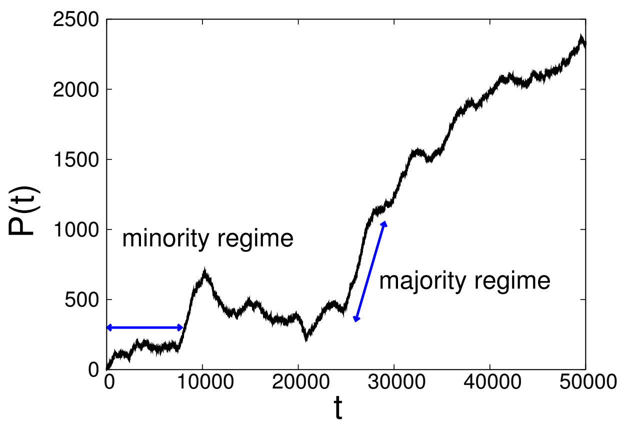
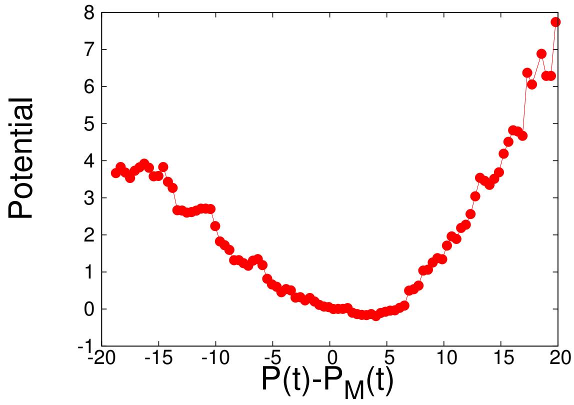
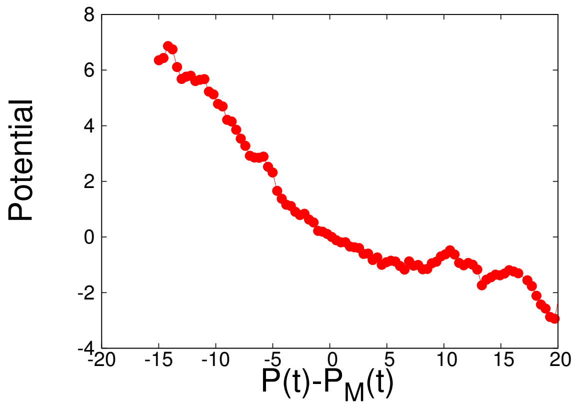


Figure 5.18: The time evolution of $P(t)$ for a minority-majority game with $\epsilon = 1$ and $\alpha = 0.05$ is shown. We can observe the alternation of different regimes. In the graph are indicated two periods by means of arrows. In the minority regime the price remains near to its fundamental value, while in the majority regime appears a well defined trend.



(a)

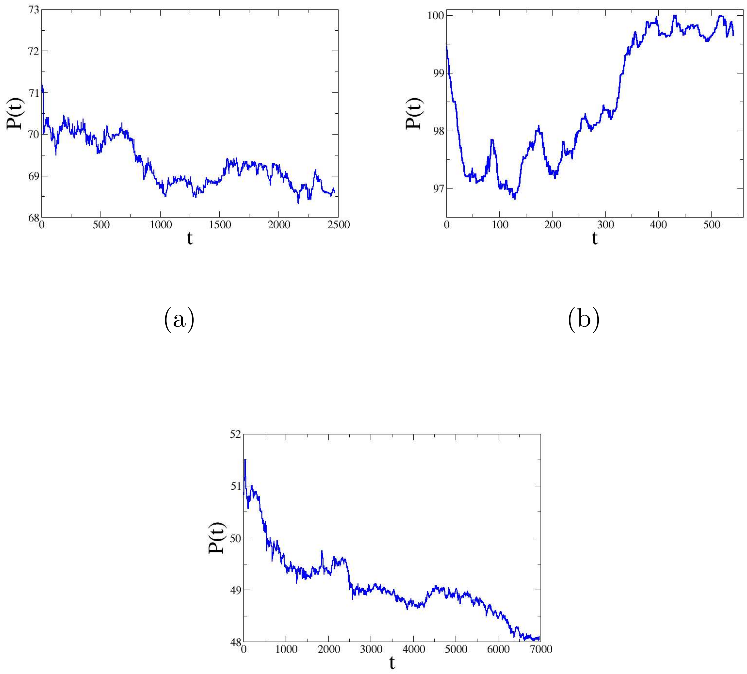


(b)

Figure 5.19: The potential analysis with $M = 20$ for the two periods indicated in Fig. 5.18 is shown. The analysis for the minority region leads to an attractive (even though not really quadratic) potential. Instead, analyzing the majority region we found a repulsive potential.

5.7 Results for Real Stock Prices from NYSE

For our potential analysis we consider as database the price time series of all the transactions of a selection of 20 NYSE stocks. These have been selected to be representative and with intermediate volatility. This corresponds to volumes of $10^5 - 10^6$ stocks exchanged per day. We consider 80 days from



(a) (b)

(c)

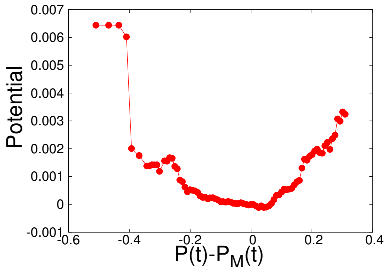
Figure 5.20: The time evolution of three stock indexes (CLF, ITU and TOL) is shown. The time is expressed in tick and correspond to one trading day.

October 2004 to February 2005.

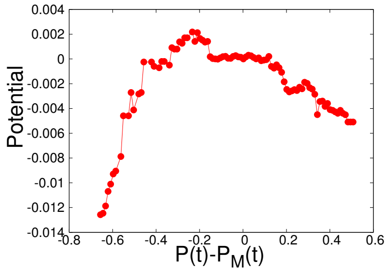
The time series we consider are by a sequential order tick by tick. This is not identical to the price value as a function of physical time but we have tested that the results are rather insensitive to this choice.

The statistical properties of these kind of data are relatively homogeneous within the time scale of a trading day but the large jumps of the prices between different days prevent the extension of the analysis to large times [6]. So we focus our potential analysis considering the stock-prices fluctuations within a trading day. In Fig. 5.20 are plotted the time evolutions of three stock indexes in a trading day.

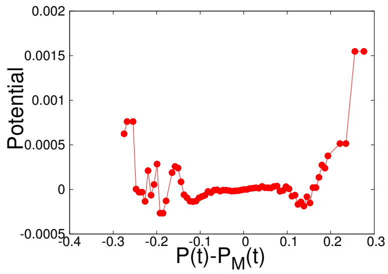
If we perform the effective potential method for a trading day of a given stock, we found shapes of the effective potential that are very irregular and often asymmetric. In Fig 5.21 are plotted the results obtained for the data plotted in Fig 5.20. We can see that the shapes of the potentials are not always quadratic. The potential in Fig. 5.21(a) it seems rather quadratic



(a)



(b)

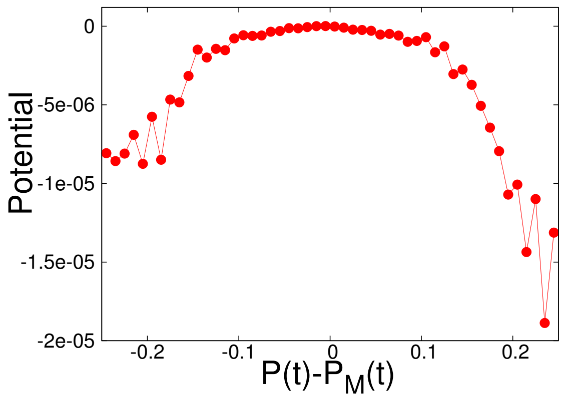


(c)

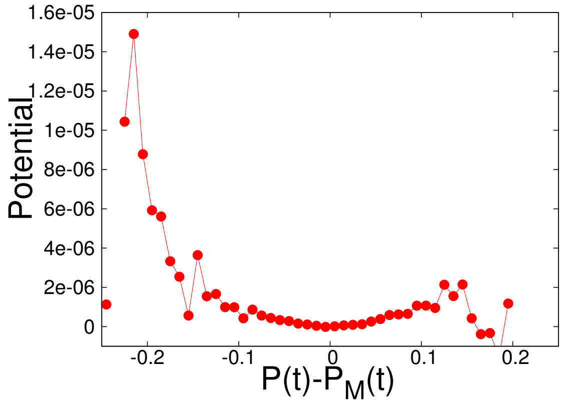
Figure 5.21: The effective potential method has been applied to the price evolution of the three stocks plotted in Fig. 5.20. We can see that the effective potential are rather asymmetric and not all quadratic. The potential in (a) seems quadratic and attractive in the central part but has asymmetric tails. The potential in (b) and (c) are not quadratic. In particular the potential in *b* is asymmetric and looks like piecewise linear as in Fig. 5.12. The potential in (c) is flat like a simple RW potential.

and attractive while in Fig. 5.21(b) has a piecewise linear shape similar to the potential plotted in Fig. 5.12. The potential in Fig 5.21(c) seems flat as one expects from a simple RW model.

Instead, if we consider some average over a long period (80 trading days) of the potentials obtained for a single day, the resulting potentials seems to be quadratic as in [114]. In Fig. 5.22 are shown the average shape of the potential for two stock indexes (BRO and PG). We can observe a rather quadratic shape for the potential. In Fig. 5.22(a) the potential is quadratic and repulsive while for the index PG the potential is attractive, as we can see in Fig. 5.22(b).



(a)



(b)

Figure 5.22: This plot shows the average shape of the potential over 80 trading days for two different stock indexes (BRO and PG). The shape is quite quadratic and symmetric, and in (a) is repulsive while in (b) attractive.

5.8 Conclusions and Perspectives

In summary the idea to consider price dynamics as influenced by an effective force dependent on the distance of price $P(t)$ from its own moving average $P_M(t)$ represents a new statistical tool to detect hidden forces in the market. The implementation of the analysis can be seriously affected by the eventual presence of positive or negative correlations. However, we have shown by a series of models and tests, that this new method is able to explore complex correlations which have no effect on the usual statistical tools like the correlations of price variation and the volatility clustering.

The method provides an analysis of the sentiment of the market: aggressive for the case of repulsive forces and conservative for attractive ones. In this respect it may represent a bridge between the financial technical analysis and the application of Statistical Physics to this field. In addition it may also be useful to analyze the results of the different strategies and behaviors which arise in agent based models.

Chapter 6

Agent Based Model in an Active Random Walk

Moving averages are one of the most popular and easy to use tools available to the technical analysts. They smooth a data series and make it easier to spot trends, something that is especially helpful in volatile markets. In this chapter we want to analyze an agent based model in which agents base their strategies on some suitable moving average. Our aim is to consider a model with evolutive and adaptive agents which can be either trend follower or trend adverse. The main idea comes from the potential model we have introduced in chapter 5. The macroscopic price emerging from this model will follow a Random Walk in which a term of force is present. The strength and the direction of this force depends on which type of agent dominates. This model is on a preliminary stage and, up to now, many of the main features have still to be considered. At the moment we have considered on oversimplified model with no interacting and evolutive agents, with the only purpose to check the basic running of the model [8].

6.1 Definition of the Model

We start from the hypothesis that the price fluctuation in a market can be described by the active Random Walk defined in Eq. 5.1. We recall this equation considering a quadratic potential:

$$P(t+1)-P(t)=B(t)\frac{d}{d\Big(P(t)-P_M(t)\Big)}\Big(P(t)-P_M(t)\Big)^2+\omega(t) \qquad (6.1)$$

where $\omega(t)$ is a Gaussian random noise with zero mean and unit variance. The idea is that $B(t)$ is an effective parameter that comes from the microscopic actions of each agents. In our model we consider N agents which at each time step t calculate an arithmetic moving average of the price. Agents differ from each other by the number of time steps back they use in performing their moving average. Every agent i calculates his moving average considering the previous M_i values of the price:

$$P_{M_i}(t) = \frac{1}{M_i} \sum_{j=1}^{M_i} P(t-j) \tag{6.2}$$

Each agents can also be either trend follower or adverse. In the first case he will always move in the direction which put the next price away from the moving average he has calculated. The opposite is true for a trend adverse agent. In this simplified stage we consider only two possible type of agent: a trend follower agent which has a microscopic positive force parameter $b_i = +1$, and a trend adverse agents which has $b_i = -1$. The contribution of each agent i in the price formation is

$$\rho_i = bi(t) \frac{d}{d[P(t) - P_{M_i}]} [P(t) - P_{M_i}]^2. \tag{6.3}$$

The macroscopic price will be give by:

$$P(t+1) - P(t) = \sum_{i=1}^{N} \rho_i + \omega(t) \tag{6.4}$$

The price dynamics will depend on the relative number, Φ of agents with positive b, and on the number, k of different steps M_i used in calculating the moving averages.

6.2 Simulations

Now we check if the simulations will reproduce what we expect. When trend followers dominate we expect to find an over diffusion of the price. On the contrary an over diffusion is expected when trend adverse agents dominate.

We have performed a simulation with $N = 100$ agents, which are in a fraction Φ trend followers and $1 - \Phi$ trend adverse. The parameter of the possible different moving average is set to $K = 2$ (5, 25).

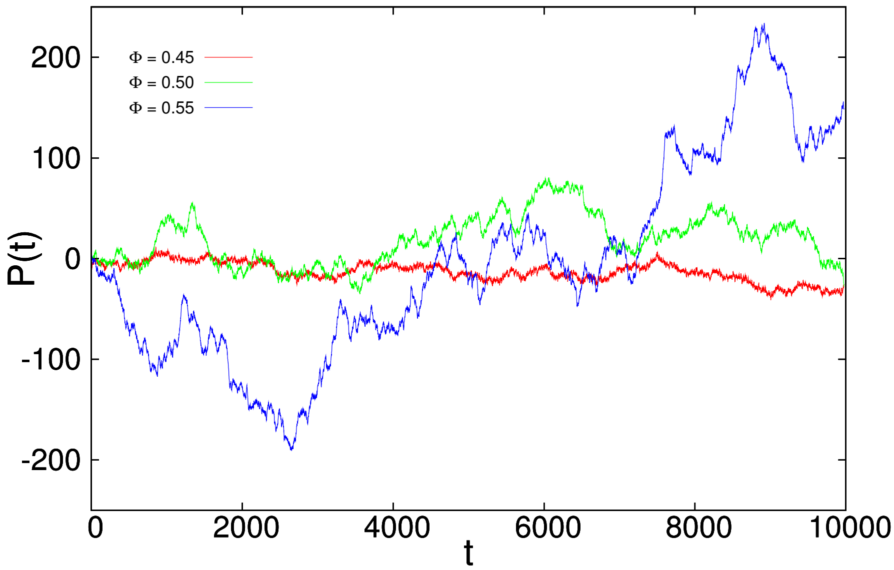


Figure 6.1: Three different output of the model are plotted. The red line represent a case in which trend adverse agents dominate. In this case we observe an under diffusion of the price. The contrary happens when trend followers dominate (blue line). Also an intermediate situation is plotted (green line) in which we observe a normal diffusion.

We have performed a simulation for three different values of Φ (0.45, 0.50, 0.55) to observe all the possible behaviors as one can see in Fig. 6.1.

In the future it will be interesting to see how the coefficient of diffusion of this active Random Walk is dependent on the fraction Φ of trend followers. Also a more detailed analysis of the collective behavior as a function of Φ is necessary. In fact we expect that for large number of Φ, the strength of the force due to the action of trend followers will overcome the Gaussian noise, and the price will grow up exponentially. Finally we will perform a statistical analysis of the distribution of price increments for the three regimes ($\Phi \lesseqgtr 0$).

6.3 Potential Analysis

In this preliminary stage, we conclude our test performing the potential analysis described in chapter 5 on the outputs of the simulations. As usual we expect to find attractive potential when the market is stable and under diffusive and a repulsive potential in the opposite case. A flat potential is expected when there is an equilibrium between agents ($\Phi = 0.5$).

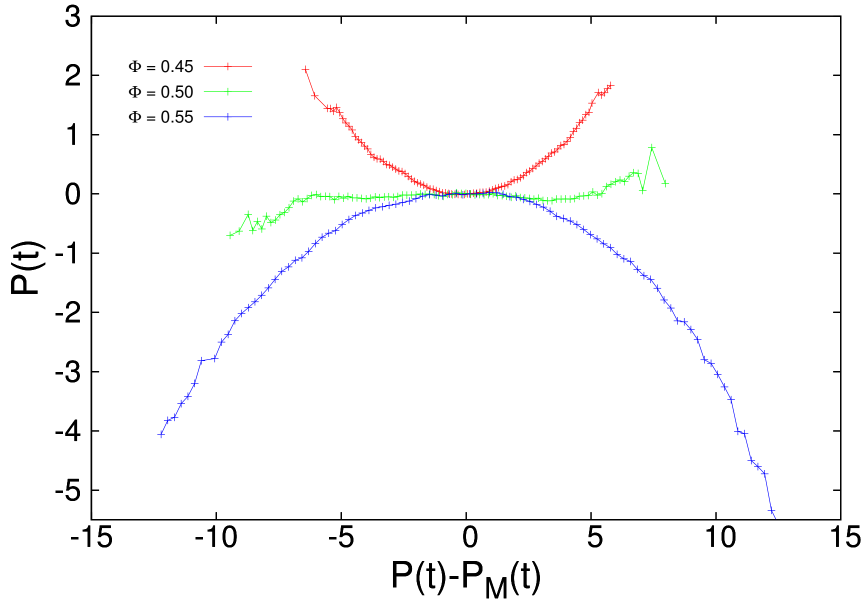


Figure 6.2: Three different output of the model are plotted. These different behaviors depends on the values of the parameter Φ. The red line represent a case in which trend adverse agents dominate. In this case we observe an stable potential. The contrary happens when trend followers dominate (blue line). In this case the potential is unstable. A flat potential is observed when $\Phi = 0.5$.

The results obtained from the potential analysis are plotted in Fig. 6.2. We can observe that the potential analysis is able to characterize the different phases of the game. A repulsive potential is observed when trend followers dominate. On the opposite the potential is stable when the market is dominated by trend adverse agents.

Conclusions

In this thesis we have collected a panoramic of analysis and models concerning the dynamics of stock price fluctuations. The most common stochastic model of stock price dynamics is the Gaussian behavior that assumes a geometric Brownian motion for the stock price. The Gaussian probability distribution is a universal consequence of the central limit theorem in the limit of long time series on the condition that the financial market is in a stationary states. This hypothesis is not always true, in fact empirical distributions of price increments disclose non Gaussian shapes. Further more the price fluctuations have complex correlations.

A way to characterize stock price dynamics is to consider the properties of the so called Roughness, that is to consider the price fluctuation as function of the length of the considered time interval and to estimate the relative scaling exponent. The usual interpretation of a non trivial roughness exponent is the presence of long range correlations in the underlying process. In this thesis we have shown that this interpretation is not always true because other elements can lead to non trivial roughness exponent without the presence of correlations and persistence in the analyzed process. One of this elements is the effect given by considering finite size samples. The need to consider finite size time series can be due to different reasons. In our case the problem concerns the nature itself of our financial data. We have analyzed data from the NYSE. The NYSE market is close during the night and the price undergoes a significant discontinuity. The overnight jump of price implies that our data are statistically homogeneous only within a single trading days. This limit our analysis in considering finite size time series.

Starting from this situation, we have analyzed the role of finite size effects in estimating the roughness exponent from a given time series. We have started our analysis considering a finite size Random Walk. By means of an analytical calculation, we have been able to show that considering finite size time series has the effect to systematically enhance the roughness exponent. In addition to the Random Walk behavior, we have also studied other cases in which the dynamics presents also fat tails and correlations. Simulations of these more complex models have led to a further bias in the estimation of the roughness exponent. This effect is intrinsic and unavoidable in studying the roughness properties of finite size datasets but it is often omitted in literature.

In this perspective, we have introduced new tools to estimate roughness. The first is by means the analysis of the fluctuations of a given variable from a suitable reference value. This new method appears much more useful than the standard one in order to characterize the absolute fluctuations of a given process. Another method we have proposed consists in a test of complex correlations. We have considered an exact analytical equation that is valid for uncorrelated variables and we have checked if it is still satisfied by other process. We have shown that this method is indeed suitable to detect complex, higher order correlations in finite size time series. In our view these methods provide an useful tool to detect and visualize the characteristics of absolute price fluctuations in finite size time series.

In addition to this information on the absolute value of the fluctuations, one may also consider directional ones. In fact the dynamics of directional fluctuations is another source of complexity. This fluctuations are often related to the traders' strategy. In facts two different strategies can react differently to a given deviation of the price from a given reference value. In this perspective, together with a group of Japanese researchers, we have introduced that idea that price dynamics can be influenced by its distance from a given reference value, i.e. a moving average of the price performed in the previous time steps. Starting from this idea, we have considered a model for the price dynamics that is a Random Walk in which is present a term of force. This force can be either attractive or repulsive reveling the stable or unstable nature of the market in a given time interval. In this way it has been possible to identify the prevailing strategy present in the market directly from the data.

To test this method we have considered a minority-majority game in which the emerging dynamics is an alternation of phases in which the market is stable and a "trend adverse" strategy is prevailing, and phases in which the market is unstable and "trend followers" dominate. Our method was indeed suitable to detect the agents' behavior analyzing the output of the game. When the market is unstable we have found repulsive forces and, on the contrary, attractive forces when the market is stable. An analysis of real data has shown that this method is also suitable to visualize the complex correlations of financial price fluctuations. We have analyzed our data from NYSE. If no correlations would have been in our data, nor force has to emerge from the analysis. Instead we have found that this method can detect the "hidden forces" present in the market. In fact, an extensive analysis has shown that the force is systematically different from zero and can be both attractive or repulsive. This is very interesting because, only looking to the data, we can obtain a direct measure of the "sentiment" of the market, i.e stable or unstable. It is important to highlight that this new method is able to explore complex correlations which have no effect on the usual statistical tools like the correlations of price variation and the volatility clustering. In a wide perspective, one can interpret the presence of non trivial forces in the stock price datasets as a universal features of financial data, such as the fat tail and volatility clustering phenomena. The importance to find new "stylized fact" is crucial. This is due to the fact that many different models for price dynamic can all reproduce the market features. Discover new universal characteristics can be a powerful tool to find more realistic and specific models. Hence, an essential point for the progress in this field is to identify further general properties which will permit a critical comparison between models.

In addiction to the analysis of real data, the above concepts can suggest a new framework for the modeling of the strategies in an Agent Based Models. In this perspective we have defined an agent based model in which the trading strategies are based on some different moving averages. The strategies differs between each other in the time scales in which the moving average are performed. The traders can be either trend followers or adverse. The resulting behavior depends on which kind of strategy is prevailing. If trend followers dominate the dynamic is unstable and well defined trends (bubbles or crashes) appear. On the other side, if trend adverse traders dominate the market is stable. This model is in a preliminary stage and other important features have to be considered. For example we want to introduce the possibility that agents switch their strategies to include role of evolution and adaptation.

Bibliography

[1] V. Akgiray. Conditional heteroskedasticity in time series of stock returns: Evidence and forecast. *J. Business*, 52:55–80, 1989.

[2] E. Alessio, A. Carbone, G. Castelli, and V. Frappietro. Second-order moving average and scaling of stochastic time series. *Eur. Phys. J. B*, 27:197–200, 2002.

[3] C. Alexander. Common volatility in the foreign exchange market. *Appl. Fin. Econ.*, 5:1–10, 1995.

[4] V. Alfi, F. Coccetti, M. Marotta, A. Petri, and L. Pietronero. Exact results for the roughness of a finite size random walk. *Physica A*, 370:127–131, 2006.

[5] V. Alfi, F. Coccetti, M. Marotta, L. Pietronero, and M. Takayasu. Hidden forces and fluctuations from moving averages: A test study. *Physica A*, 370:30–37, 2006.

[6] V. Alfi, F. Coccetti, A. Petri, and L. Pietronero. Roughness and finite size effect in the NYSE stock-price fluctuations. *European Physical Journal B*, Article in Press, 2006. available at http://www.edpsciences.org/articles/epjb/abs/first/b06064/b06064.html.

[7] V. Alfi, A. D. Martino, L. Pietronero, and A. Tedeschi. Detecting the traders' strategies in minority-majority games and real stock-prices. 2006. Submitted. E-print: physics/0609038.

[8] V. Alfi, A. D. Martino, L. Pietronero, and A. Tedeschi. A model of agents adaptive strategies in terms of an active random walk. 2006. Article in preparation.

[9] V. Alfi, A. Petri, and L. Pietronero. A finite range test for correlation in time series. 2006. Article to submit.

[10] B. W. Arthur. Inductive reasoning and bounded rationality: the el farol problem. *Am. Econ. Rev.*, 84:406–411, 1994.

[11] B. W. Arthur. Complexity and the economy. *Science*, 284:107–109, 1999.

[12] J. Asikainen, S. Majaniemi, M. Dubé, and T. Ala-Nissila. Interface dynamics and kinetic roughening in fractals. *Phys. Rev E*, 65:052104, 2002.

[13] R. Axelrod. *The Complexity of Cooperation: Agent-based Models of Competition and Collaboration*. Princeton University Press, Princeton, 1997.

[14] L. Bachelier. Théorie de la spéculation. *Annales Scientifiques de l'Ecole Normale Supérieure*, III-17:21–86, 1900. [Ph.D. thesis is mathematics].

[15] R. Baillie and T. Bollerslev. Conditional forecast densities from dynamics model with garch innovations. *J. Econometrics*, 52:91–113, 1992.

[16] P. Bak, M. Paczuski, and M. Shubik. Price variations in a stock market with many agents. *Physica A*, 246:430, 1997.

[17] P. Ball. Culture crash. *Nature*, 441:686–688, 2006.

[18] A. Barabasi and H. Stanley. *Fractal Concepts in Surface Growth*. Cambridge University Press, Cambridge, 1994.

[19] V. Baviera, M. Pasquini, J. Raboanary, and M.Serva. Moving averages and price dynamics. *International Journal of Theoretical and Applied Finance*, 5:575–583, 2002.

[20] H. Bergström. On some expansions of stable distributions. *Ark Mathematicae II*, 18:375–378, 1952.

[21] B. Biais, P. Hillion, and C. Spatt. An empirical analysis of the limit order book and the order flow in the paris bourse. *Journal of Finance*, 50(5):1655–89, 1995.

[22] F. Black and M. Scholes. The pricing of options and corporate liabilities. *J. Polit. Econ.*, 81:637–654, 1973.

[23] R. Blatteberg and N. Gonedes. A comparison of the stable and student distributions as a statistical model for stock prices. *J. Buisness*, 47:244–280, 1974.

[24] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *J. Econometrics*, 31:307–327, 1986.

[25] G. Bonanno, G. Caldarelli, F. Lillo, S. Miccichè, N. Vandewalle, and R. N. Mantegna. Networks of equities in financial markets. *Eur. Phys. J. B*, 38:363–371, 2004.

[26] G. Bonanno, F. Lillo, and R. Mantegna. High-frequency cross-correlation in a set of stocks. *Quantitative Finance*, 1:96–104, 2001.

[27] J. Bouchaud and M. Potters. *Théories des Risques Financiers.* Eyrolles, Aléa-Saclay, 1997.

[28] J. Bouchaud and D. Sornette. The black & scholes option pricing problem in mathematical finance: Generalization and extesions for a large class of stochastic process. *J. Phys. I France*, 4:863–881, 1994.

[29] J. P. Bouchaud, M. Mezard, and M. Potters. Statistical properties of stock order books: empirical results and models. Science & Finance (CFM) working paper archive 0203511, Science & Finance, Capital Fund Management, 2002.

[30] J. P. Bouchaud and M. Potters. *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management.* Cambridge University Press, 2003.

[31] J. Campbell, A. Lo, and C. MacKinlay. *The Econometrics of Financial Markets.* Princeton University Press, New Jersey, 1997.

[32] A. Carbone, G. Castelli, and H. Stanley. Time-dependent hurst exponent in financial time series. *Physica A*, 344:267–271, 2004.

[33] C. Castellano, M. Marsili, and L. Pietronero. Nonperturbative renormalization of the kpz growth dynamics. *Phys. Rev. Lett.*, 80:3527, 1998.

[34] A. Cavagna. Irrilevance of memory in the minority game. *Phys. Rev. E*, 59:R3783–R3786, 1999.

[35] A. Cavagna, J. Garrahan, I. Giardina, and D. Sherrington. Thermal model for adaptive competition in a market. *Phys. Rev. Lett.*, 83:4429, 1999.

[36] D. Challet and M. Marsili. Phase transition and symmetry breaking in the minority game. *Phys. Rev. E*, 60:R6271–R6274, 1999.

[37] D. Challet, M. Marsili, and R. Zecchina. Statistical mechanics of systems with heterogeneous agents: Minority games. *Phys. Rev. Lett.*, 84:1824, 2000.

[38] D. Challet and R. Stinchcombe. Analyzing and modelling 1+1d markets. *Physica A*, 300:285, 2001.

[39] D. Challet and Y. Zhang. Emergence of cooperation and organization in an evolutionary game. *Physica A*, 246:407–418, 1997.

[40] D. L. C. Chan, D. Eliezer, and I. I. Kogan. Numerical analysis of the minimal and two-liquid models of the market microstructure. Jan. 2001. cond-mat/0101474.

[41] P. Cizeau, Y. Liu, M. Meyer, C. Peng, and H. Stanley. Volatility distribution in the S&P 500 stock index. *Physica A*, 245:441–445, 1997.

[42] P. K. Clark. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, 41(1):135–55, 1973.

[43] R. Cont. Empirical properties of assets returns: stylized facts and statistical issues. *Quantitative Finance*, 1:223–236, 2001.

[44] M. Couillard and M. Davison. A comment on hurst processes in financial time series. *Physica A*, 384:404–418, 2005.

[45] G. Cuniberti, M. Raberto, and E. S. nad G. Susinno. L'econofisica, una scienza emergente. *Scienza & Buisness*, 1-2:27–32, 2001.

[46] M. Daniels, J. Farmer, L. Gillemot, G. Iori, and D. Smith. Quantitative model of price diffusion and marketfriction based on trading as a mechanistic random process. *Phys. Rev. Lett.*, 90:481, 2003.

[47] M. G. Daniels, J. D. Farmer, L. Gillemot, G. Iori, and E. Smith. A quantitative model of trading and price formation in financial markets, 2002. cond-mat/0112422.

[48] E. Dimson and M. Mussavian. A brief history of market efficiency. *European Financial Management*, 4:91–193, 1998.

[49] Z. Ding, C. Granger, and R. Engle. A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1:83–106, 1983.

[50] E. Eberlin and U. Keller. Hyperbolic distributions in finance. *Bernoulli*, 1:281–299, 1995.

[51] Editorial. Econophysicists matter. *Nature*, 441:667, 2006.

[52] R. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1008, 1982.

[53] R. Engle. Autoregressive conditional heteroskedasticity with estimates of the variance of u.k. inflation. *Econometrica*, 50:987–1008, 1982.

[54] E. Fama. The behavior of stock market prices. *J. Business*, 38:34–105, 1965.

[55] E. F. Fama. Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, 25(2):383–417, 1970.

[56] E. F. Fama. Efficient capital markets: II. *Journal of Finance*, 46(5):1575–617, 1991.

[57] J. Farmer, P. Patelli, and I. Zovko. The predictive power of zero intelligence in financial market. *preprint in PNAS*, 102:2254, 2005.

[58] J. D. Farmer, D. E. Smith, and M. Shubik. Is economics the next physical science? *Physics Today*, 58(9):37–42, 2005.

[59] W. Feller. *An Introduction to Probabbility and Its Applications, Vol. 1.* J. Wiley & Sons, New York, 1968.

[60] D. Fudenberg and J. Tirole. *Game Theory.* MIT Press, Cambridge, 1991.

[61] A. R. Gallant, P. E. Rossi, and G. Tauchen. Stock prices and volume. *Review of Financial Studies*, 5(2):199–242, 1992.

[62] M. Gallegati, S. Keen, T. Lux, and P. Ormerod. Worrying trends in econophysics. *Physica A*, Article in Press, 2006.

[63] D. Gershon. Taking your skill all the way to the bank. *Nature*, 393:496, 1998.

[64] B. Gnedenko. On the theory of domains of acctraction of stable law. *Uchenye Zapiski Moskow. Gos. Univ. Matematkia*, 45:61–72, 1940.

[65] B. Gnedenko and A. Kolmogorov. *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley, Cambridge MA, 1954.

[66] C. Goodhart and M. O'Hara. High frequency data in financial markets: Issues and applications. *Journal of Empirical Finance*, 4:72–114, 1997.

[67] P. Gopikrishnan, M. Meyer, L. Amaral, and H. Stanley. Inverse cubic law for the distribution of stock price variations. *Eur. Phys. J. B*, 3:139–140, 1998.

[68] P. Gopikrishnan, V. Plerou, X. Gabaix, and H. E. Stanley. Statistical properties of share volume traded in financial markets. *Phys. Rev. E*, 62(4):R4493–R4496, 2000.

[69] G. Grimmet and D. Stirzaker. *Probability and Random Processes*. Oxford University Press, Oxford, 2001.

[70] M. Hart, P. Jefferies, P. Hui, and N. Johnson. Crowd-anticrowd theory of multi-agent market games. *Eur. Phys. J. B*, 20:547, 2001.

[71] M. Hart, P. Jefferies, N. Johnson, and P. Hui. Crowd-anticrowd theory of the minority game. *Physica A*, 298:537, 2001.

[72] J. Hull. *Options, Features and Other Derivatives*. Prentice-Hall, Upper Saddle River NJ, 1997.

[73] H. Hurst. Long-term storage capacity reservoirs. *Transaction of the American Society of Civil Engineers*, 116:770–808, 1951.

[74] N. Johnson, M. Hart, and P. Hui. Crowd-anticrowd theory of multi-agent market games. *Physica A*, 269:1, 1999.

[75] S. Katsev and I. L'Heureux. Are hurst exponents estimated from short or irregular time series meaningful? *Computers & Geosciences*, 29:1085–1089, 2003.

[76] A. Y. Khintchine and P. Lévy. Sur le loi stables. *C. R. Acad. Sci. Paris*, 202:374–376, 1936.

[77] P. Lévy. *Calcul des probabilités*. Gauthier-Villars, Paris, 1925.

[78] C. Li and W. Li. On a double-threshold autoregressive heteroscedastic time series model. *J. Appl. Econ.*, 11:253, 1996.

[79] I. N. Lobato and C. Velasco. Long memory in stock-market trading volume. *Journal of Business & Economic Statistics*, 18(4):410–27, 2000.

[80] H. Luckock. A statistical model of a limit order market. 2001. Available at http://www.maths.usyd.edu.au/res/AppMaths/Luc/2001-9.html.

[81] T. Lux and M. Marchesi. Scaling and criticality in a stochastic multi-agent model of a financial market. *Nature*, 397:498–500, 1999.

[82] E. Majorana. Il valore delle leggi statistiche nella fisica e nelle scienze sociali. *Scientia*, 36:58–66, 1942.

[83] B. G. Malkiel. The efficient market hypothesis and its critics. *Journal of Economic Perspectives*, 17(1):59–82, 2003.

[84] B. Mandelbrot. The variation of certain speculative prices. *Journal of Business*, 36:394–419, 1963.

[85] B. Mandelbrot. *Fractals and Scaling in Finance*. Springer Verlag, New York, 1997.

[86] R. Mantegna. Lévy walks and enhanced diffusion in milan stock exchange. *Physica A*, 179:232–242, 1991.

[87] R. Mantegna and H. Stanley. Stochastic process with ultraslow convergence to a gaussian: The truncated lévy flight. *Phys. Rev. Lett.*, 73(22):2946–2949, 1994.

[88] R. N. Mantegna and H. Stanley. *An Introduction to Econophysics: Correlation and Complexity in Finance*. Cambridge University Press, New York, NY, USA, 2000.

[89] R. Manuca, Y. Li, R. Riolo, and R. Savit. The structure of adaptive competition in minority games. *Physica A*, 282:559, 2000.

[90] M. Marsili. Market mechanism and expectations in minority and majority games. *Physica A*, 299:93, 2001.

[91] M. Marsili. *Toy Models of markets with heterogenous interacting agents*, volume 503. Springer-Verlag, 2001.

[92] M. Marsili, D. Challet, and R. Zecchina. Exact solution of a modified al farol's bar problem: Efficiancy and the role of market impact. *Physica A*, 280:522, 2000.

[93] M. Marsili and Y.-C. Zhang. Fluctuations around nash equilibria in game theory. *Physica A*, 245:181, 1997.

[94] A. D. Martino, I. Giardina, A. Tedeschi, and M. Marsili. Generalized minority games with adaptive trend-followers and contrarians. *Phys. Rev. E*, 70:025104(R), 2004.

[95] S. Maslov. Simple model of a limit order-driven market. *Physica A*, 278:571, 2000.

[96] S. Maslov and M. Mills. Price fluctuations from the order book perspective - empirical facts and a simple model. *Physica A*, 299:234, 2001.

[97] T. D. Matteo, T. Aste, and M. Dacorogna. Long-term memories of developed and emerging markets: Using the scaling analysis to characterize their stage of development. *Journal of Banking and Finance*, 29:827–851, 2005.

[98] R. Merton. Option pricing when underlying stock returns are discontinuous. *J. Financial Econ.*, 3:125–144, 1976.

[99] E. Moro. *The Minority Game: an introductory guides*. Nuova Scientia Publisher, Inc., Advances in Condensed Matter and Statistical Physics, 2004.

[100] U. Müller, M. Dacorogna, R. Olsen, O. Pictet, M. Scwarz, and C. Morgenegg. Statistical study of foreign exchanges rates, empirical evidence of a price change scaling law and intraday analysis. *J. Banking and Finance*, 14:1189–1208, 1995.

[101] B. Murphy. *Technical Analysis of the Financial Markets*. Prentice Hall Press, 1999.

[102] R. B. Olsen, U. A. Müller, M. M. Dacorogna, O. V. Pictet, R. R. Davé, and D. Guillaume. From the bird's eye to the microscope: A survey of new stylized facts of the intra-daily foreign exchange markets. *Finance and Stochastics*, 1(2):95–129, 1997.

[103] A. Pagan. The econometrics of financial markets. *Journal of Empirical Finance*, 3(1):15–102, 1996.

[104] M. Pasquini and M. Serva. Multiscale behaviour of volatility autocorrelations in a financial market. *Economics Letters*, 65:275–279, 1999.

[105] L. Pietronero. *Order and Chaos in Nonlinear Physical System*. Plenum Publishing Corporation, New York, 1988.

[106] K. Popper. *Conjectures and Refutations: The Growth of Scientific Knowledge.* Harper & Row, New York, 1965.

[107] S. Reimann. Eco - no(?) - physics - comments and reflexions -. http://arxiv.org/abs/physics/0607033, 2006.

[108] R. Savit, R. Manuca, and R. Riolo. Adaptive competition, market efficiency, and phase transition. *Phys. Rev. Lett.*, 82:2203, 1999.

[109] W. Schoutens. *Lévy Processes in Finance: Pricing Financial Derivatives.* Wiley, 2003.

[110] M. Schulz. *Statistical physics and economics: concepts, tools, and applications.* Springer-Verlag, New York, 2003.

[111] A. Shiryaev. *Essentials for Stochastic Finance: Facts, Models, Theory.* World Scientific, Singapore, 1999.

[112] F. Slanina. Mean-field approximation for a limit order driven market model. *Physical Review E*, 64:056136, 2001.

[113] R. M. H. Stanley. Scaling behaviour in the dynamics of an economic index. *Nature*, 376:46–49, 1995.

[114] M. Takayasu, T. Mizuno, and H. Takayasu. Potentials of unbalanced complex kinetics observed in market time series. 2005. E-print http://arxiv.org/abs/physics/0509020.

[115] L. Tesfatsion. Introduction to the special issue on agent-based computational economics. *J. Econ. Dyn. Contr.*, 25:281, 2001.

[116] R. Tsay. Conditional heteroscedastic time series models. *J. Am. Stat. Assoc.*, 82:590, 1987.

[117] G. L. Vasconcelos. A guided walk down wall street: An introduction to econophysics. *Brazilian Journal of Physics*, 34:1039–1065, 2004.

[118] A. Weiss. Asymptotic theory for arch models: Estimation and testing. *Econ. Theory*, 2:107–131, 1986.

[119] R. Westfall. *Never at Rest: A Biography of Isaac Newton.* Cambridge University Press, Cambridge, 1980.

[120] J.-M. Zakoian. Threshold heteroskedastic models. *J. Econ. Dyn. Control*, 18:931–955, 1994.

[121] Y. Zhang. *Europhys. News*, 29:51, 2001.

Acknowledgment

First of all, I would like to thank Prof. Luciano Pietronero for giving me the opportunity to study the interesting subjects I have analyzed in this thesis. His precious teachings and encouragements have been essential for the achievement of this thesis.
I thank Prof. Mauro Rovere for being a kind supervisor and for the always interesting discussions.
I am also very grateful to Prof. Vittorio Loreto and Dr. Alberto Petri for the fruitful scientific collaboration from which I have learnt a lot.
Of course, I thank all the (many) people that have been in the wonderful room 206 during these three years; with them I have shared the office and a lot of nice moments.
At last, but not at least, I would like to thank all the persons that somehow have helped me in this work.