UNIVERSITÀ DEGLI STUDI DI PADOVA

FACOLTÀ DI SCIENZE MM.FF.NN.

DIPARTIMENTO DI FISICA

Scuola di dottorato di ricerca in Fisica
Ciclo XXI - Tesi di dottorato:

Scaling and modelization of financial time series

Direttore e tutore: Ch.mo Prof. Attilio Stella

Co-tutore: Dott. Fulvio Baldovin

Dottorando: Dario Bovina

Contents

Introduction

In many fields of the physics of complex systems, like seismology [66, 67], fluid turbulence [61, 65] and finance [53, 60], repeated measurements of a given observable quantity provide us with time series of outcomes. The aim of a physicist who wants to reproduce the behaviour of a given time series, is to find a mathematical model able to display the same empirical statistical features characterizing the time series itself. As soon as new statistical features are discovered and classified, thus becoming "stylized facts", new models are formulated in order to take them into account.

In 1900 L. Bachelier first proposed to model the behaviour of financial prices by means of what is nowaday known as Brownian motion, thus opening the way for the interpretation of the price fluctuations as generated by a stochastic process [14]. Brownian motion accounts for the unpredictability of price returns, but fails to explain other important stylized facts, like the positive and slowly decaying of second-order correlations between the absolute values of two returns, and the returns distribution with fat power law tails [41].

Since the financial risk is much more sensitive to the large absolute returns rather than to the small ones and since the pricing of futures or options is essentially based on the volatility forecasting [11, 15], during the last century many models replaced the Brownian motion. In particular in 1982 R. Engle introduced the class of the ARCH models in order to reproduce the observed volatility clustering, which is a fingerprint of the second-order long range correlations of absolute returns [30]. A few years later T. Bollerslev generalized this model [29] and afterwards many other variants [31, 36] have been used to account for the long memory of price fluctuations [37, 40]. However, a stochastic model accounting for *all* the detected empirical features is not yet known.

Some properties of a time series can be of much help in finding the most suitable class of models in order to well reproduce its empirical behaviour. One such a property is the invariance under scale transformations. Like every invariance property, this symmetry is a clue on the nature of the physical system we are studying and it can help us to understand what kind of process can better describe the empirical data.

Scaling invariance means that the fluctuations of a given time series behave

similarly at every scale, so that if a little part of this time series is taken into account and the time axis and the price axis are properly rescaled, then the enlarged part statistically looks like the whole time series. In particular this means the presence of fluctuations at every scale and the scaling of the distributions for returns over different time windows [54].

The generalized Hurst exponent analysis is the main statistical tool to detect and characterize the scaling properties of a given time series [59, 64, 68]. In finance the invariance under rescaling means the scaling of the return distributions with time. This in turn implies the scaling of the q-order moment of these distributions according to a scaling exponent $D(q)$, which depends on the order q. The generalized Hurst exponent H is the slope of $D(q)$. If $D(q)$ is linear, and thus H is constant, the time series shows simple scaling, otherwise it manifests multiscaling.

Theoretically the scaling exponent $D(q)$ for random walks with independent increments and finite second order moment is a linear function with slope $H = \frac{1}{2}$, but in 1951 H. Hurst pointed out that many hydrological time series display an anomalous scaling exponent with a larger slope [70]. There are many possible causes for such a behaviour: long-term correlations [2], non-stationarity [72, 73] and finite sample effects [71] are all possible sources [69].

The scaling exponent $D(q)$ has been found to be non-linear in the order q for most of the time series in several other fields of complex systems physics, such as geophysics [66] and turbulence [61, 62]. Also in finance the Hurst exponent has been often used to detect multiscaling properties of the empirical time series [57, 60]. The exponent H has been helpful in classifing Stock Markets maturity too [4].

In finance scale invariance has started to play an increasingly important role in recent years. Especially after the discovery of multiscaling effects similar to those met in turbulence [26], many authors did put scale invariance properties at the basis of their modelizations of the dynamics of markets. The first attempt was made by B. Mandelbrot [21, 23]. Soon afterwards some models, based on a multiplicative cascade [19, 20] or a multi-agent structure [17, 18], also succeeded in reproducing the empirical observed multiscaling. More recently in [24, 25] a model based on a scale invariant non stationary process was able to reproduce the power law decay of the volatility correlation between two returns.

In conclusion the multiscaling character of most of the real time series, in finance and in many other fields, is a widespread feature. Like every statistical tool, the determination of the exponent H should be followed by an estimation of its uncertainty, or at least by a search of the factors that could somehow falsify the results.

A first problem concerns the accuracy of the Hurst exponent when it is evaluated on a finite time series: every statistical tool deals with a set of data whose size is necessarily finite and thus we cannot expect the outcome to be an infinitely precise

determination. To evaluate the uncertainty associated to a given outcome, in some cases it is possible to perform the calculation on several independent samples, all having the same size, thus obtaining a set of different results. The interval covered by these results provides us with an estimate of the uncertainty. This procedure implies the possibility to get many such samples, for instance by repeating the experiment. However this is not always possible.

In finance and in many other fields a single empirical time series is only available and we cannot restart the history to get another sample. One way to get around such difficulty would be to have a very long history and to extract many samples from the whole time series. Then one can regard these samples as shorter histories generated by the same physical mechanism. This is the main idea of the Bootstrap method [49, 52], which we are going to apply in order to get confidence intervals for the Hurst exponent determinations. We stress that recently, in seismology, some doubts about the actual stability of H, when it is calculated on a finite size series, have been raised [3]. However, to our knowledge, there has not been any previous attempt to give an estimation of the Hurst exponent uncertainy in finance.

A second problem concerns the very existence of the Hurst exponent. We already stressed that the calculation of the Hurst exponent $H(q)$ is based on the q-order moment of the return distribution, but this moment does not necessarily exist at all. It is worth noting that here we are dealing with the theoretical moment, namely with the moment we would expect for the density function of returns we asymptotically extrapolate from our data. Of course the empirical moments, being based on a finite sample of data, are always finite, but when the theoretical limit does not exist the consistency of the Hurst exponent should be carefully checked.

The possible presence of fat tails in the density functions of the returns of the financial time series [41] makes the theoretical q-order moment infinite if the order q is too large. Indeed, when the tails asymptotically decay as a power law with exponent γ, say, then the q-order moment exists for $q < q_0$, with $q_0 = \gamma - 1$. Hence, there is a threshold order q_0 beyond which the empirical moments cannot converge to anything as the number of data available grows.

To test the consistency of the Hurst exponent we can determine it for a time series generated by an exactly solvable model and check whether the outcoming scaling exponent agrees or not with the expected one. We know that for a Brownian motion the agreement is excellent [4], but the Gauss' distribution has tails which decay faster than exponentially, and thus all the theoretical moments exist, no matter how large is the order q. Therefore, we decided to determine the Hurst exponent for a time series based on a simulation of another exactly solvable model belonging to the class of the Levy flights: a random walk with independent increments distributed according to a Cauchy distribution. In such a case the density functions of the returns have tails which decay as a power law with exponent $\gamma = 2$, so the theoretical q-order moments exist for $q < 1$ only.

We stress that while this process is strictly simple scaling with $H = 1$, the Hurst exponent calculated on a simulated history is not constant. More precisely the scaling exponent $D(q)$ behaves proportional to q for $q < 1$, but fails to show the correct slope beyond the threshold order $q_0 = 1$. Indeed, for $q > 1$ one gets $D(q) = 1$, namely a constant scaling exponent. This spurious multiscaling is already known [5], but one of our main contributions is to describe a convincing probabilistic mechanism able to account for it. In this thesis we show on rigorous basis that such a mechanism works in general when one single time series is available and when the density functions of the returns have power law tails.

This mechanism we are going to describe has potential deep consequences on many multiscalings reported so far. Since in finance a single time series is available and since the large returns are power law distributed, this mechanism most likely affects the Hurst exponent analysis performed on the real data. Hence, the observed multiscaling could be spurious in character, not only for Levy flights, but also for the financial time series. Under this perspective, while the invariance under rescaling would be a robust fact, the multiscaling could be merely a finite sample effect.

To enforce this suspicion we took into account correlations. Indeed, a main difference between a Levy flight generated series and the financial time series is that the former has independent increments while the latter has long-range correlated returns. We thus repeated the calculation on our financial time series after a reshuffling of all the daily returns, namely after a removing of all correlations. The new scaling exponent behaves as a function of q in a way qualitatively very similar to that of a Levy flight model; which we know to be spurious in character. We propose a qualitative argument according to which a mechanism of spurious multiscaling could also apply to time series with dependent returns.

It's worth noting that some remarks about the reliability [55, 56] and the sample dependence [6] of the Hurst exponent H were already made. However, the financial literature, to our knowledge, still lacks a systematic study of the effects due to the finiteness of the sample, to Pareto tails and to the volatility clustering. This work aims at filling such a gap.

This thesis is organized as follows. In the first chapter we recall some of the most important stylized facts concerning the financial time series, illustrating them for some indexes. They will be the empirical background for all the subsequent work.

In the second one we will develop the theory behind the concepts of scaling invariance, simple scaling and multiscaling. We will define the Hurst exponent too; and estimate it for some simple scaling processes whose returns are distributed according to a law with exponentially decaying tails. Then we estimate the Hurst exponent on our financial time series and evaluate its uncertainty by means of a bootstrap method.

In the last chapter we present our main result: the Hurst exponent shows a spurious multiscaling in the case of a Levy random walk; moreover we prove that

such spurious multiscaling results for simple scaling processes if a single simulation is considered and if large increments are power law distributed. Then we give evidences for the multiscaling of the financial indexes to be spurious too by means of an argument which takes into account the effect due to the volatility correlations. We conclude by claiming that the Hurst exponent is reliable only for low orders, when the theoretical moments exist.

Chapter 1

Some financial stylized facts

Reproducing as best as possible the statistical features of a given empirical time series is the main task of every theretical model which aims at simulating the underlying process. But the importance of a careful statistical analysis of the series under investigation goes further: it is not only a tool for checking the goodness of a model, it often provides some insights into the unknown process which help the construction of the model itself. Generally symmetries, like invariance properties, are very useful to this end. This is why in this first chapter we start by studying the main statistical properties of the financial series we will deal with. In particular we are going to focus our attention on the invariance under scale trasformations.

1.1 Stock Market indexes

Since the accuracy of every statistical analysis is very sensitive of the size of the available data set, we consider some famous Stock Market indexes whose life is among the longest in finance:

Dow Jones Industrial (DJI): surely one of the most famous index of the New York Stock Exchange. It comprises the 30 biggest U.S. industrial companies among the *blue chips*. The origin of the DJI index dates back to 1900.

Dow Jones Transportation (DJT): the most long-lived index of the New York Stock Exchange: its origin dates back to 1897. At present it comprises 20 transportation-related companies.

Standard & Poor's 500 (S&P): stock market index which contains the stocks of the 500 U.S. companies with the largest capitalization. It was created in 1957, but it has been extrapolated back in time.

Standard & Poor's 90 (SPC): this S&P Composite index, on a daily basis only, contains the stocks of 90 U.S. companies from all economic areas. It was created in 1928.

For each index we collected the prices daily, at closure, from its origin until now, neglecting the days without trading activity like week-ends and holidays; see table (1.1).

We show in fig. (2.6), as an example, the graph of the price $S(t)$ of the DJI index versus time. Can we guess something about the process? There are some outstanding features, common to the graphs of all indexes:

- the prices are always positive: $S > 0$;
- one can easily see a global average growth with many fluctuations;
- fluctuations become larger as the prices become higher.

The last point is quite reasonable as only *rates* between two prices really matter. Hence we can guess an exponential growth with both a deterministic and a stochastic component:

$$S(t) = \exp[f(t) + x(t)] \tag{1.1}$$

where $f(t)$ is a trend, namely some function of the time, and $x(t)$ is the real stochastic process. This expression (eq. 1.1) allows the price S to be always positive regardless of the stochastic component x.

Since the stochastic component is only responsible for the statistical behaviour of our time series, we are interested in modelling the process $x(t)$ alone. To get $x(t)$ we simply extract the logarithm of $S(t)$ and assume a linear trend:

$$f(t) = At + B$$

where a linear best-fit of the logarithm of the whole time series provides easily with the parameters on the right hand side of eq. (1.1). Fig. (1.2) shows the detrended logarithm of the price $S(t)$ for the DJI index, while table (1.1) lists the estimations of the parameters A and B for every index.

At first sight the linearity of the trend $f(t)$ seems a rather unjustified hypothesis, but one should compare the result in fig. (1.2) with the stochastic fluctuations. Clearly any non-linear correction to the function $f(t)$ would be completely overwhelmed by the stochastic noise; thus every higher order term would not yield any sensible contribution. On the other hand we are going to see that any error on the slope A affects our statistical analysis very little, while the actual value of B does not matter at all.

Index	from	until	data set size N	$A \times 10^{-4}$	B
Dow Jones Ind.	02/01/1900	14/11/2005	26589	1.90	3.52
Stand.&Poor's 90	03/01/1928	29/12/2006	19837	2.59	1.82
Dow Jones Trans.	02/02/1897	31/08/2005	27260	1.26	3.47
Stand.&Poor's 500	03/01/1950	18/12/2006	14333	2.82	3.06

Table 1.1: The Stock Market indexes taken into account. The time interval inside which data have been collected and the number of trading days are shown. In the last column the parameters of the linear best-fit of $\ln S(t)$ are also listed.

Since in finance we are interested in the rate of change between two prices separated by a certain time gap, in what follows we will consider the returns $r_t(\tau)$ over a *time window* τ:

$$r_t(\tau) = x(t+\tau) - x(t) \tag{1.2}$$

All the statistical quantities, which we are going to study, deal with returns[1]; see fig. (1.3) for that of DJI index. Here we come to the central issue: the behaviour of our time series under a scale transformation. For a fixed window τ we calculate the corresponding return at every instant t and regard this set of returns as a realization of the process on a scale τ. For $\tau = 1$ we simply get the daily returns of our time series; for $\tau = 2$ we are looking the process as we would collect returns every two days; for $\tau = 3$ every three days and so on. To study the behaviour of the process $x(t)$ under rescaling means to study the statistical properties of the stochastic variables $r(\tau)$.

Before going on two very important remark about returns are in order:

Stationarity. To regard the empirical return $r_t(\tau)$ at each instant t as a realization of the same stochastic variable $r(\tau)$ we need to assume stationarity, at least when the process is sampled on a daily basis. That is we make the hypothesis that the outcoming returns $r_t(\tau)$ have all the same density function $p_\tau(u)$; otherwise we could not obtain any statistical information from the set of these returns. Indeed to do this we often need to calculate the average over the time t of some statistical quantity depending on $r_t(\tau)$, like momenta.

Sliding window. This is a more subtle point which concerns the actual calculation of returns. Consider a time series $x(t)$ of size $N+1$. Obviously there are N daily returns. If we would collect prices $r(\tau)$ every τ days, then our set of returns would have a size $\left[\frac{N}{\tau}\right]$: a great decrease in the available data. Since we cannot get a series of size $\sim \tau N$ for each window τ because in finance a

[1]Hence it is clear that the parameter B in the detrending procedure does not matter. It is only a translation along the vertical axis which cannot affect the right side of eq. (1.2)

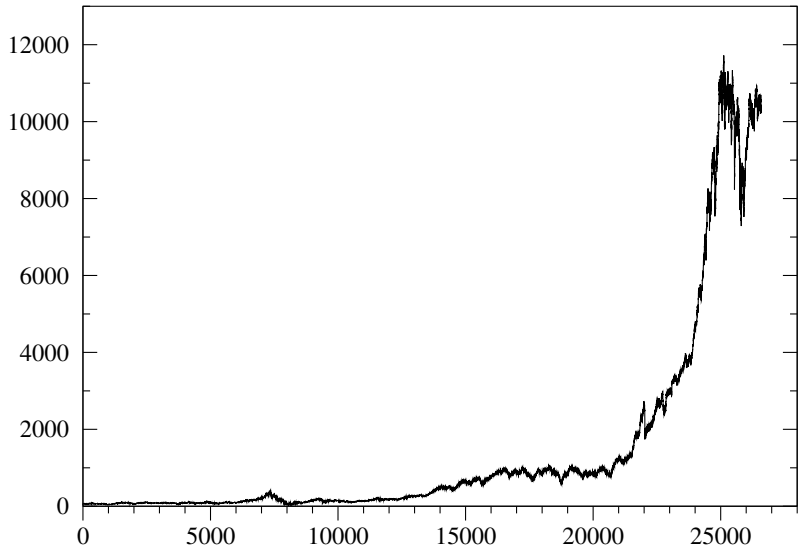

Figure 1.1: Graph of the daily price $S(t)$ of the Dow Jones Industrial index from the origin until 2005.

single history is only available, we can avoid the problem simply by calculating $r(\tau)$ every day, i.e. at each instant t as in eq. (1.2). This method is known as *sliding window* sampling. It is the only way to get around such difficulty, but as we will see in chapter (3) the set of returns thus obtained does not behave as a statistical ensemble in some cases.

In the following sections we are going to study some of the most important statistical features of the time processes $r_t(\tau)$. Often this features are common to a wide class of empirical time series and thus referred to as 'stylized facts'. This study will provide us with the empirical basis for all the ensuing work.

1.2 Volatility clustering

Perhaps the very first question about returns one should answer is whether they are independent or not. For processes with independent increments we have a number of powerful probabilistic theorems and a well developed stochastic theory. On the contrary, the treatment of processes with long-range memory is rather hard and little is known on the subject. Looking at fig. (1.3) one can easily see many changes in the width of the process; there are time intervals where r_t has wild fluctuations and others where the size of the returns is smaller. This is a strong evidence for dependence between two returns at different instants; indeed, if returns were strictly

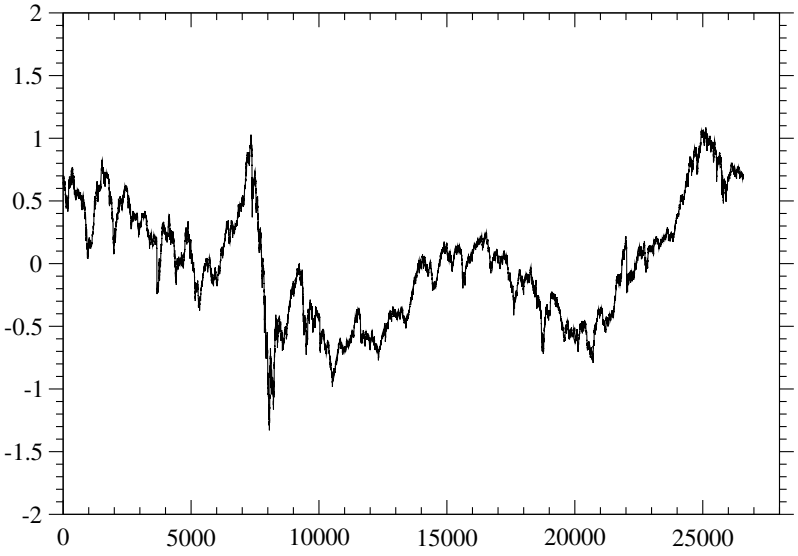

Figure 1.2: Graph of the linearly detrended logarithm of the price: $x(t) = \ln S(t) - At - B$

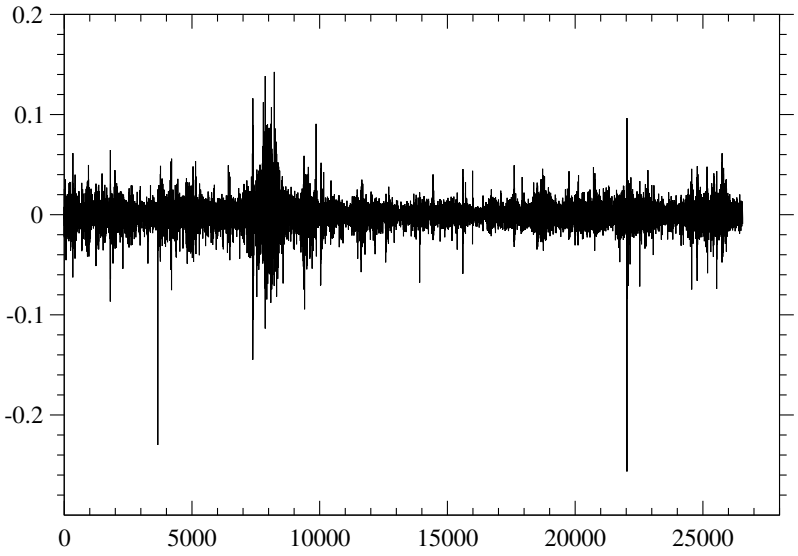

Figure 1.3: Graph of the daily returns $r_t(1) = x(t+1) - x(t)$

independent one could not observe these changes. The alternation of periods with large and small returns is known as *volatility clustering*.

Mathematically we can detect the correlations in a given time series by means of the empirical autocorrelation (AC) functions. Consider a time series with N returns, then the *linear* normalized autocorrelation coefficient reads[2]:

$$A_{lin}(s) = \frac{\langle r_{t+s}\, r_t\rangle - \langle r_t\rangle^2}{\langle r_t^2\rangle - \langle r_t\rangle^2}$$

Here the average $\langle\cdot\rangle$ is taken over all times t:

$$\langle f(r_t)\rangle = \frac{1}{N}\sum_{t=1}^{N} f(r_t) \tag{1.3}$$

where $f(\cdot)$ is any real function. It is easy[3] to show that:

- $|A_{lin}(s)| \leq 1$
- $A_{lin}(0) = 1$
- if r_{t+s} and r_t are independent then $A_{lin}(s) = 0$.

The AC function measures how much the sign of the present return allows us to forecast that of another return after a time gap s: the closer is $|A_{lin}(s)|$ to 1, the more reliable our expectation. If the AC function is close to 1 then the two signs are likely both positive or both negative, otherwise if $A_{lin}(s) \simeq -1$ then they are very likely opposite. In fig. (1.4) we can see the linear AC function calculated on the time series of the DJI index: it vanishes at once[4]. However, this does not mean that the daily returns are independent; it only means that any forecasting about the sign of a future return is impossible, even for the very next one. On the other hand, this condition seems a very reasonable one, at least for well developed markets: riskless gains are not allowed.

This stylized fact is usually reffered to as 'market efficiency' and in order to reproduce it one usually uses a martingale to model the process $x(t)$. For a martingale the expectation of the conditional distribution of a return, given all the previuos ones, is zero. In other words, at each step we cannot forecast the sign of the next return and thus the linear AC function vanishes, but this does not imply independence. Indeed, other characteristics, like variance, could depend on the past history.

[2] to simplify in this section we write r_t instead of $r_t(1)$ for the daily return at time t

[3] if our data set is large enough, i.e. if $N \gg 1$, than taking the average over t or $t+s$ does not affect the result; for example $\langle r_t\rangle = \langle r_{t+s}\rangle$

[4] it falls inside the stochastic noise for $s > 0$

A direct proof that returns are strongly dependent comes from the non-linear autocorrelation functions. Here we take into account the normalized volatility AC function:

$$A_{vol}(s) = \frac{\langle |r_{t+s}\, r_t| \rangle - \langle |r_t| \rangle^2}{\langle r_t^2 \rangle - \langle |r_t| \rangle^2}$$

This function has the same properties as the former, in particular it vanishes for $s > 0$ if the returns are independent, but its meaning is rather different. While the linear AC provides informations on the sign of returns, this AC function deals with the absolute value of returns and allows us to forecast the size of a subsequent return starting from the size of the present one.

In fig. (1.4) again we can see the volatility AC function in the case of the DJI index. Now the graph shows a slow decay until it becomes negligible around $s_0 \approx 10^3$. The time s_0 is called 'decorrelation time', or process memory and roughly indicates the time beyond which two returns are expected to be independent. The large value of s_0 is due to a long-range correlation among returns. In fig. (1.5) we can see, in a log-log plot, that $A_{vol}(s)$ behaves according to a power law for small time gaps:

$$A_{vol}(s) \propto \frac{1}{s^\ell}$$

where the exponent $\ell = 0.21$ has been estimated using a power law best-fit. The next one, fig. (1.6), shows an exponential decay after $s \approx 100$, but such decay is still quite slow due to a very large mean lifetime.

The slow pwer law decay of A_{vol} is common to many financial time series and indicates that the returns are actually strongly correlated. The volatility clustering seen at the beginning of this section just arises from such a slow decay: for a long period, large returns are likely to be followed by other large returns. This is because A_{vol} stays positive for a long time. The curve A_{vol} has also very deep practical consequences because modelling as well as possible the volatility clustering is fundamental in every strategy of risk control. Therefore the interest in a model that correctly reproduces the volatility AC curve goes beyond pure theory.

Summing up the central issue in this section we saw that the market prevent us from forecasting the sign of the next return, but the past history gives us many informations about the size of such return. This suggests a kind of model where the return is made of two factors, a process w_t with zero mean and independent increments and a function σ which depends on the past returns:

$$r_t = \sigma(r_{t-1}, r_{t-2} \dots) w_t$$

Usually the process w_t is a Brownian random walk and $\sigma(\cdot)$ fixes its conditioned variance, namely the width of the Gaussian distribution according to the past history. A process like this, which is easily proved to be a martingale, has been named ARCH, or *AutoRegressiv Conditional Heteroskedastic*.

The width $\sigma(r_{t-1} \dots r_{t-n})$ is usually a linear combination of the squares of the last n returns:

$$\sigma^2 = a_0 + \sum_{i=1}^{n} a_i r_{t-i}^2$$

where the weights a_i must be estimated by means of some maximum likelihood procedure. Since σ^2 must be positive always, no matter what happened in the past, then the weights are all positive too: $a_i \geq 0 \ \forall i = 1 \dots n$. Furthermore it can be shown that the unconditioned variance $\sigma_U = \langle r_t^2 \rangle$ is asymptotically constant:

$$\sigma_U = \frac{a_0}{1 - \sum_{i=1}^{n} a_i}$$

This is just the global average width of the whole process. Since it must be positive and finite, in order the process to be stationar, we find another constraint on our weights:

$$\sum_{i=1}^{n} a_i < 1$$

These constraints are the most important ones, but there are others. If $n = 1$, for instance, it can be shown that the unconditioned kurtosis is $K_U = \frac{6a_1^2}{1-3a_1^2}$; hence we need to impose $a_1 \leq \frac{1}{\sqrt{3}}$ in order to make K_U finite.

The importance of the ARCH models come from their capability of reproducing the volatility clustering observed in many financial time series. It can be shown, according to these models, that the volatility AC function decays very fast; if $n = 1$, for instance, it decays exponentially:

$$A_{vol}(s) = \exp(-\frac{s}{s_*})$$

where $s_* = |\ln a_1|^{-1}$ is the mean lifetime. However the model can fit quite well the volatility AC function if the weights are properly chosen and above all if n is large enough. Indeed the theoretical AC function $A_{vol}(s)$ decays rapidly for $s > n$ only[5], while for $s < n$ its decay is much slower. This is why n is a sort of memory range: a returns is strongly correlated with all the latter n ones[6].

At present, in finance, the ARCH models and their generalizations, like GARCH, EGARCH and many others, are the most used theoretical tools in order to simulate the behaviour of the real time series. Unfortunately the need of a large n means in turn the need of many unknown weights. So many free parameters are a problem from both a practical and a theoretical viewpoint; indeed evaluating them on data is the main difficulty in every simulation.

[5] hence in order to fit the long-range dependence of real data one must take n very large: tipical values ranges around $n \approx 100$.

[6] but the linear AC function still vanish; an ARCH is a martingale: $A_{lin}(s) = 0$.

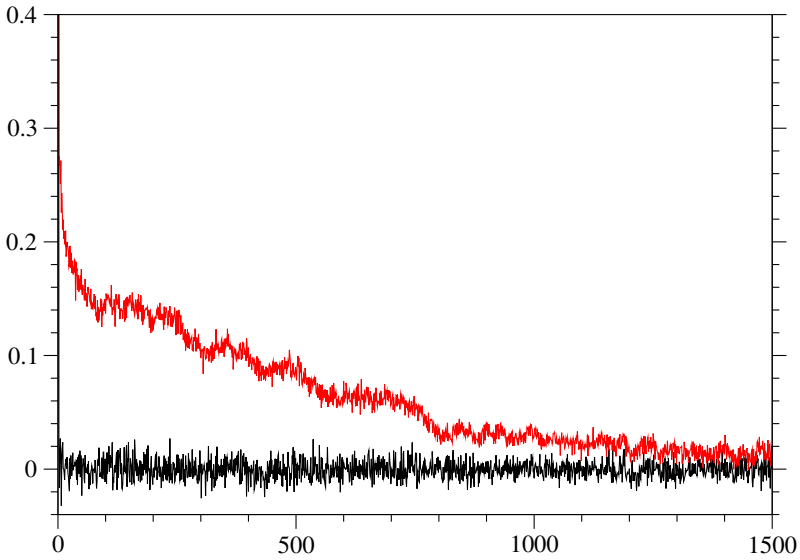

Figure 1.4: Graph of the autocorrelation function calculated on the data set of the DJI index. Both the linear (*black*) and the volatility (*red*) AC functions are shown.

The set of non-linear AC functions includes many others: for every positive real number q we define a generalization of $A_{vol}(s)$

$$A_q(s) = \frac{\langle |r_{t+s}|^q \, |r_t|^q \rangle - \langle |r_t|^q \rangle^2}{\langle |r_t|^{2q} \rangle - \langle |r_t|^q \rangle^2}$$

One gets A_{vol} setting $q = 1$, the quadratic AC setting $q = 2$ and so on. All these provides essentially the same informations about volatility clustering, however the curve A_{vol} is often less affected by the stochastic noise because it deals with the lowest powers of the returns[7].

However a different kind of AC functions may provides us with new informations about some different properties of our time series. The so called 'leverage effect' involves the sign of the present return and the size of a future one:

$$L(s) = \langle r_{t+s}^2 \, r_t \rangle - \langle r_t \rangle \langle r_t^2 \rangle$$

In finance $L(s)$ generally starts from a negative value at $s \gtrsim 0$ and slowly decays to zero, suggesting that a negative return is likely followed by a period of large volatility. However this is a rather weak effect and often neglected, so we will not consider this effect in the present work.

[7] we will see in the section (1.4) that this is a theoretical advantage too: for high q the average $\langle |r_t|^q \rangle$ may have no limit when $N \to \infty$.

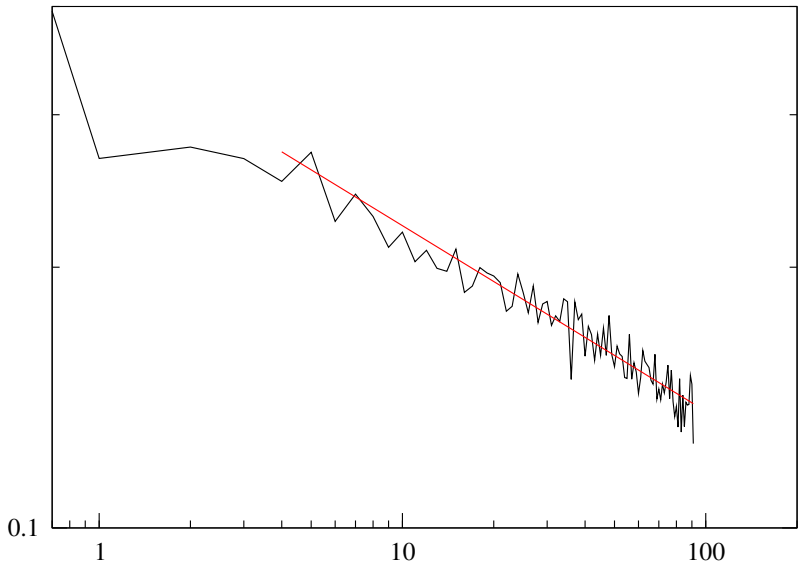

Figure 1.5: Log-log plot of the volatility AC function (*black*); the power law best-fit with decaying exponent $\ell = 0.21$ is shown in (*red*).

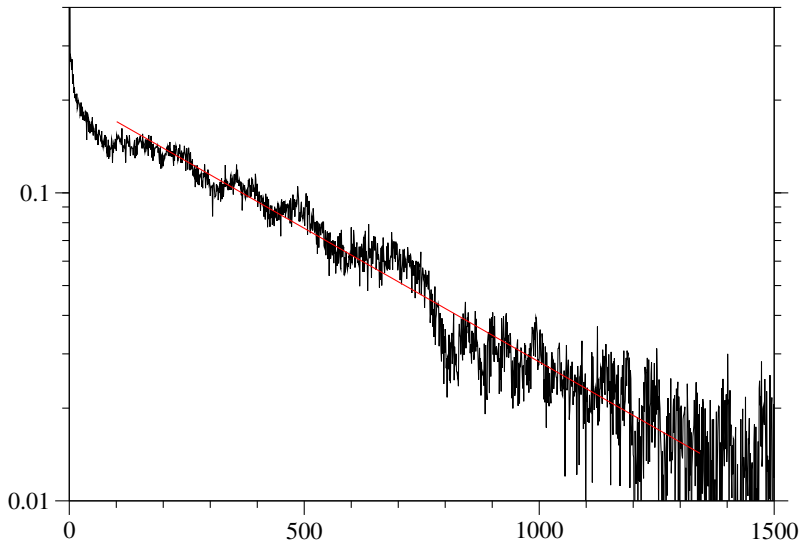

Figure 1.6: Plot of the volatility AC function (*black*); the exponential decay $\exp(-\frac{s}{s_*})$ with mean lifetime $s_* = 500$ is shown in (*red*).

1.3 Invariance under rescaling

In the last section we showed some very important results about the long-range correlations among returns and sketched the most common theoretical model able to display the volatility clustering. We will need these results again. Now we come to the central issue: the scaling invariance property of the graph of our empirical time series.

Looking at $x(t)$ in fig. (1.2) the feature which draws our attention at once is the high roughness of the graph in comparison with any usual deterministic function. At first sight one could think to make the graph smoother simply by enlarging it, i.e. by looking what happens between two consecutive oscillations. However one realizes soon that new wild fluctuations appear at every scale[8]; moreover if both axis are properly enlarged then the graph displays always the same pattern. This means that, for a correct choice of the scale factors on the time and x axis, a small part of our graph statistically looks like the whole. They have the same statistical behaviour and there is no way to distinguish between them.

Enlarging the time axis of a factor τ simply means to collect the points of our time series $x(t)$ every τ days instead of collecting them daily. In other words it means to consider the returns $r_t(\tau)$. A process $x(t)$ is invariant under rescaling if its returns over a window τ are equally distributed up to a scale factor depending on τ. Otherwise we could tell the scale of a graph since its statistical properties would be different from a scale to another. In particular the process is self-affine if all the stochastic variables $\tau^{-\alpha} r(\tau)$ have the same distribution:

$$r(\tau) \sim \tau^{\alpha} r(1) \tag{1.4}$$

where $\sim$ means 'equally distributed'. This in turn implies that the empirical probability density function (PDF) $p_\tau(u)$ of the returns $r(\tau)$ satisfies:

$$p_\tau(u) = \frac{1}{\tau^\alpha} g(\frac{u}{\tau^\alpha}) \tag{1.5}$$

The exponent α is called *Hölder exponent* and links the scale factor of the time and the x axis together in order for the scaling invariance to hold. It is both a sort of diffusive exponent and a measure of the intensity of the stochastic fluctuations, or roughness.

- Suppose a very small Hölder exponent $\alpha \simeq 0$, then eq. (1.4) tell us that the returns over a time window τ are nearly distributed like the daily returns. Hence we can have large fluctuations even during a small period, i.e. our

[8] obviously we cannot enlarge too much because our empirical time series is necessarily finite. In the examples considered, a period of one day is the finest scale available and it would be meaningless to go further.

graph has very wild oscillations. On the other hand, since as τ grows the PDF of $r(\tau)$ has approximately the same variance, thus the same width, then the process spreads rather slowly.

- On the contrary, when α is larger then the PDF of the returns over a large time window τ has a much higher variance than the PDF of the daily returns; therefore its spreading is fast. Furthermore the intensity of the fluctuations becomes sensible for large τ only and the graph appears more regular.

To check for the validity of eq. (1.5) we simply plot the PDFs p_τ together after rescaling both axis by a factor τ^α. However the actual value of the exponent α must be guessed and checked only a posteriori[9]. If, for a given α, all graphs collapse in a single density function, namely that of the daily returns, then we can state the scaling invariance. In fig. (1.7) we show the PDFs of the returns $r_t(\tau)$ of the DJI index rescaled according to $\alpha = 0.52$. The time window ranges from 1 to 32: $\tau = 2^k$ with $k = 0 \dots 5$.

The collapse is very good, but this procedure is not careful enough to determine α precisely, as it is not a direct evaluation, but a likelihood test; moreover it does not seem the collapse of the PDFs to be much sensitive to α. However, even if the statistical uncertainty prevents us from finding a more precise value, we can expect the scaling invariance to hold with

$$\alpha \approx \frac{1}{2} \tag{1.6}$$

To make a comparison with a widely known example, the Hölder exponent of a Brownian random walk is exactly $\alpha = \frac{1}{2}$. We will meet again the Hölder exponent in the next chapter.

It is worth to stress here that the scaling invariance cannot be exact, or better that there is a limit to the scale factor τ beyond which the former collapse does not hold any longer. Suppose that eq. (1.5) holds for $\tau \approx s_0$, where s_0 is the decorrelation time we defined in the previous section, then the correlation among the returns $r_t(\tau)$ is negligible and they can be considered as independent increments. Furthermore we will see in the next section that the scaling PDF g in the right hand side of eq. (1.5) is *not* gaussian, but has finite variance.

Under these condition the Central Limit Theorem (CLT) states that the sum of many such returns must follow the Gauss' law. Since a return over a window $n\tau$ is just a sum like that:

$$r_t(n\tau) = \sum_{i=0}^{n-1} r_{t+i\tau}(\tau)$$

[9] in the next chapter we will see how to get the Hölder exponent directly

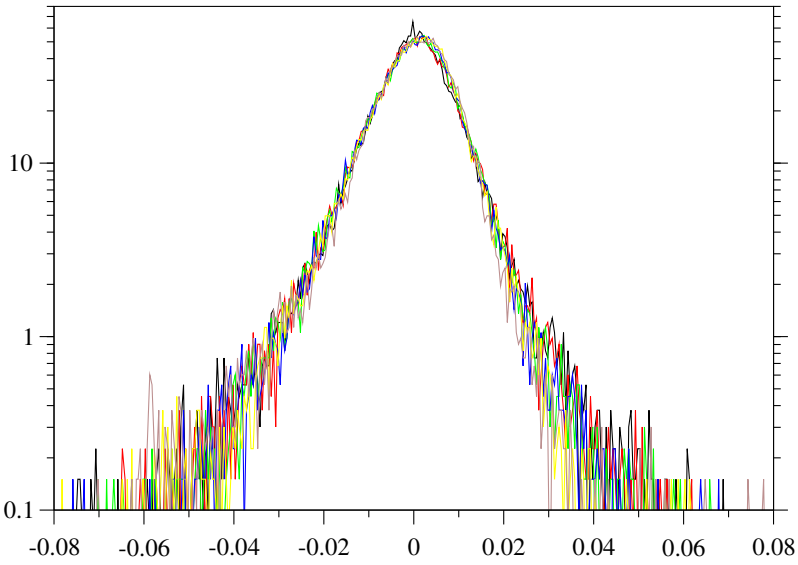

Figure 1.7: PDFs $p_\tau(u)$ of the returns of the DJI index for several windows $\tau = 2^k$ with $k = 0 \dots 5$. Each PDF has been rescaled by a factor τ^α with $\alpha = 0.52$.

then according to the CLT the density function of these returns, for n large enough, converges to a Gauss' law. Therefore at very large time window eq. (1.5) fails[10] and we observe the so called 'scaling breakdown', namely a crossover from the scaling PDFs towards a gaussian. For small windows τ the scaling invariance is a very good approximation; in particular inside the range considered scaling is almost perfect.

We conclude this section with a remark on the linear detrending procedure outlined in section (1.1). Suppose a small error $\varepsilon_A \ll A$ was made in the evaluation of the slope parameter A. In such case all returns $r_t(\tau)$ are shifted by an amount $\varepsilon_A \tau$, hence their PDF has mean $\varepsilon_A \tau$ and not zero. Since this PDF is then rescaled by a factor $\tau^\alpha \approx \sqrt{\tau}$ its graph in fig. (1.7) should be shifted by an amount $\varepsilon_A \sqrt{\tau}$. Using the data in table (1.1) we find that this quantity is much smaller than 1.6×10^{-3} for $\tau \leq 32$. Therefore any error in the evaluation of the slope A would be completely overwhelmed by the statistical uncertainty, see fig. (1.7), at least for small time windows.

[10] actually the collapse fails well before s_0. This is because we need a strong correlation to avoid the convergence to a Gauss' law

1.4 Power law tails

In the last section we have just seen that all the PDFs $p_\tau(u)$, if properly rescaled, collapse onto a single one $g(u)$, namely the *scaling density function*, due to the scaling invariance, eq. (1.5). Now a very careful study of the function g is in order as it is the starting point for much of the subsequent work. In the case of the DJI index, see fig. (1.7), we note at once a big central 'bell' like that of a gaussian law, but clearly the tails do not behave exponentially: they decay much slower. This is another very important stylized fact: the PDFs of returns have 'fat tails'.

The tails of g are made of large returns, namely of extreme events. Since in many financial applications, like the risk management, extreme events matter, it is fundamental for a model to produce large returns whose decay follows the same law as those of g. This is indeed one of the reason for giving up to the brownian motion as a theoretical model for the time series $x(t)$: it does not display large returns at all[11]. Now we are going to find a functional form which approximate as best as possible the main features of $g(u)$.

Fig. (1.8) and (1.9) show the scaling PDF $g(u)$, obtained by rescaling the PDFs $p_\tau(u)$, for each time window τ. We can see that the tails of $g(u)$ asymptotically follow a power law[12] of parameter γ for large arguments u:

$$g(u) \approx Au^{-\gamma} \tag{1.7}$$

where A is a multiplicative factor. Here the decay exponent γ was estimated using a maximum likelihood method equivalent to the Hill's estimator [46, 47]:

1. consider the right tail[13] and guess a starting point u_*, namely a point which clearly belongs to the tail and not to the central bell;

2. takes all the returns $r_1 \dots r_k$ larger than u_* and the density function:

$$h_\gamma(u) = \begin{cases} \dfrac{\gamma - 1}{u_*} \left(\dfrac{u}{u_*} \right)^{-\gamma} & \text{if } u \geq u_* \\ 0 & \text{otherwise} \end{cases}$$

3. for a given γ the probability to get the observed returns is[14]: $\prod_{i=1}^{k} h_\gamma(r_i)$;

[11] the tails of its scaling PDF are gaussian

[12] sometimes tails decaying as a power law are called 'Pareto tails' because of Vilfredo Pareto who first found that in many economical and social fields power laws are widespread relations

[13] the same procedure applies to the left tail is

[14] we suppose here that these observations are indipendent, which surely is not the case. However one can prove that the Hill's estimator is still consistent.

Index	Right tail γ	Left tail γ
DJI	4.2 ± 0.4	3.5 ± 0.2
DJT	3.8 ± 0.2	3.5 ± 0.1
S&P	5.8 ± 1.1	4.5 ± 0.5
SPC	3.6 ± 0.2	3.4 ± 0.2

Table 1.2: Decay exponents for our stock market indexes obtained by using $u_* = 0.02$ as a starting point. The difference between the left and the right exponent is due to the asymmetry of the scaling PDF $g(u)$ and gives rise to a non-zero skewness, see table (1.3).

4. therefore one get the best value of γ simply by demanding this probability[15] to be the highest:

$$\gamma = 1 + \frac{k}{\sum_{i=1}^{k} \ln \frac{r_i}{u_*}}$$

The outcome of the procedure above does not depend on u_* provided that the value chosen belongs to the tail[16]. A good starting point u_* of the tails can be evaluated by fig. (1.7): for $u_* \approx 0.02$ one get $\gamma = 4.2 \pm 0.4$. In table (1.2) we show the values of γ for both the left and right tail of our indexes[17].

The value of γ cannot be evaluated more precisely because the tails are not well sampled, due to the fact that most of returns fall into the central bulk. In order to better explore the tails, we would need a data set of larger size than the presently available one. Furthermore we will see in chapter (3) that collecting returns using a sliding window, see the remarks at the end of section (1.1), affects the behaviour of the largest returns. This means that when $\tau \neq 1$ the maximum likelihood extimation above may be slightly biased by a few extreme events. However looking at fig. (1.8) and (1.9) one can roughly guess a power law decay with exponent

$$\gamma \approx 4 \tag{1.8}$$

and this will be enough for our purposes.

The main question arising from eq. (1.8) is whether the moments of g exist or not. Indeed the theoretical q-order moment M_q of the scaling PDF $g(u)$ is

$$M_q = \int_{-\infty}^{\infty} du\, |u|^q g(u) \tag{1.9}$$

[15] actually its logarithm

[16] one cannot take u_* too large in order to make sure of this, otherwise the number k of returns available for the procedure would become so small to make the resulting exponent unreliable

[17] the starting point u_* is nearly the same for all indexes as their width is nearly the same too, see table (1.3).

and thus it exists provided the order q is not too high. If $g(u)$ asymptotically behaves according to a power law of exponent γ then the existence of the integral on the right hand side of eq. (1.9) is guaranteed for $q < q_0$ only; where q_0 is the *threshold* order beyond which the theoretical moments diverge:

$$q_0 = \gamma - 1$$

Using the estimate of γ in eq. (1.8) the threshold order takes the value $q_0 \approx 3$. In particular this means that the variance is theoretically finite; a very important fact which displays the rôle of the correlations as cause of the scaling invariance, see the end of the latter section.

In arguing that the moments of the scaling PDF do not exist for high orders $q \geq q_0$ we implicitely supposed to know the whole form of $g(u)$, i.e. its behaviour on the whole real axis, but actually it is not so. Even if we expect the central bulk of $g(u)$ to be well sampled obviously we cannot pretend the same for very large arguments u simply because large returns may not appear if our time series is too short. The larger the return we want to observe, the longer must be the time series in order the observation of this return to be likely. Therefore, stating that $g(u)$ has power law tails is correct only inside the range of arguments u explored until now. It may be that for longer time series the asymptotical behaviour of g does not follow a power law; for instance, outside the range of the present observations the PDF g may have an exponential cut-off, or even a sharp one[18]. In both cases the theoretical q-order moment M_q would be finite for every order q. However, even if this would be the case, every statistical quantity we can calculate on the *real* data will behave as if g would have power law tails. This is because at present our empirical data set cannot see any cut-off: from an empirical viewpoint $g(u)$ has fat tails on the whole real axis and the theoretical q-order moments M_q exist for low orders q only.

On the contrary the *empirical* moments $M_q^{(e)}$ are always finite as any real time series is finite too:

$$M_q^{(e)} = \frac{1}{N} \sum_{t=1}^{N} |r_t|^q \tag{1.10}$$

where N is the size of our data set[19]. Table (1.3) shows the main statistical properties of the scaling PDF of our indexes. Since in calculating the q-order moments we obviously use the above summation, and not the theoretical integral in eq. (1.9), one may think to avoid the existence problem, but another one arises.

Generally in performing an average like that in eq. (1.10) we are searching for a given statistical feature, namely the moment of g in this case. We know from the

[18] on the other hand in real Markets sometimes transactions are suspendend for excess of fall or rise in prices.

[19] here we obtained the returns r_t of g simply by rescaling the returns $r_t(\tau)$ according to τ^α, for each window τ; see eq. (1.4).

	DJI	DJT	S&P	SPC
mean	-1.2×10^{-7}	3.2×10^{-5}	2.7×10^{-5}	-3.8×10^{-5}
stand. dev.	1.136×10^{-2}	1.252×10^{-2}	8.959×10^{-3}	1.163×10^{-2}
skewness	-0.931	-0.1735	-1.322	-0.4865
kurtosis	27.47	15.52	35.14	20.65

Table 1.3: Some unconditioned empirical properties of the scaling density functions g which refer to our Stock Market indexes. The positive value of the kurtosis is due to the power law tails and stands for

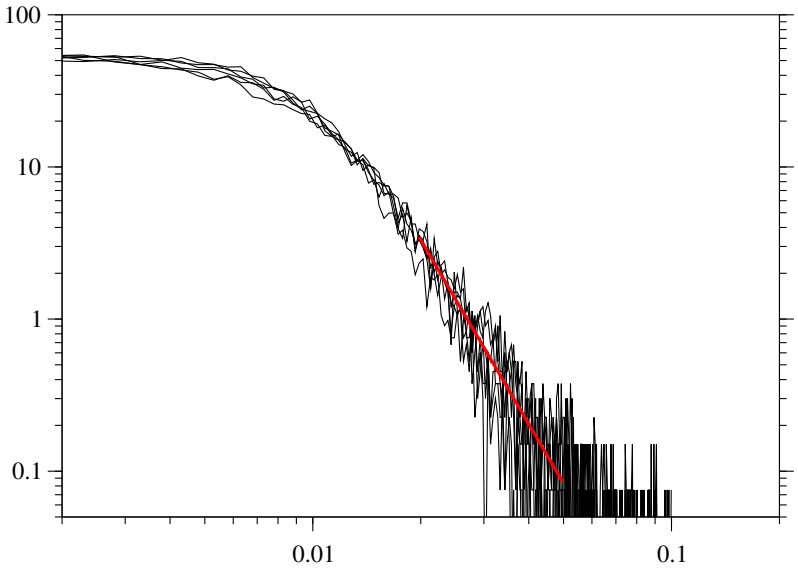

Figure 1.8: The right tail of every PDF shown in fig. (1.7) are drawn in *black*. The *red* thick line is the maximum likelihood power law of slope $\gamma = 4.2$ in the range $0.02 \leq u \leq 0.05$.

Law of Large Numbers that

$$M_q^{(e)} \to M_q$$

as $N \to \infty$, provided that the limit M_q exist. Therefore we are confident that our empirical average represents a well defined statistical property of g, at least for N large enough. If the summation in eq. (1.10) is not bounded, as the theoretical moment M_q diverge, one may ask whether it makes sense or not. When the empirical average does not converge, as the size N of our data set increases, one should pay much attention in handling and interpreting $M_q^{(e)}$: this will be the main issue of chapter (3).

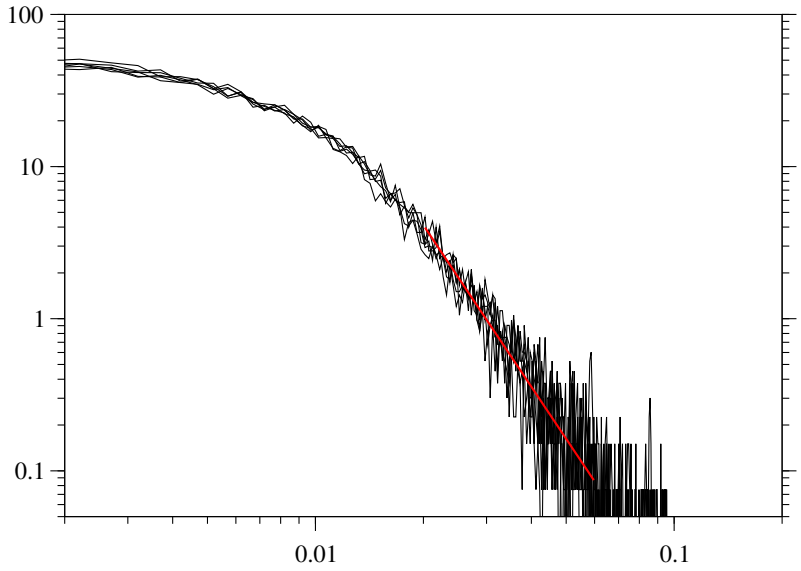

Figure 1.9: The left tail of every PDF shown in fig. (1.7) are drawn in *black*. The *red* thick line is the maximum likelihood power law of slope $\gamma = 3.5$ in the range $-0.02 \leq u \leq -0.06$.

1.5 Fitting the scaling PDF

To conclude this chapter we show how to fit the density function $g(u)$ of the DJI index shown in fig. (1.7); or better how to obtain a stochastic variable distributed according to $g(u)$ starting from a uniformly distributed one. The key idea is to fit the central bulk with a gaussian bell and the two tails with a power law. It is worth to stress here that $g(u)$ is not symmetric. Its asymmetry is due to the leverage effect, see section (1.2), which makes the left tail fatter than the right one. This implies in turn a negative skewness, see table (1.3). Then, according to table (1.2), we take two different decay exponent $\gamma_+ = 4.2$ for the right tail and $\gamma_- = 3.5$ for the left one. Furthermore, since the two tails need not to leave the central bulk at the same point, we can choose two different starting points $-u_+$, for the right tail, and u_- for the left one[20].

Consider a power law PDF $g_+(u)$ to fit the right tail, another one $g_-(u)$ to fit

[20]note that $u_- \geq 0$: properly speaking it is the distance from zero of the left tail starting point.

the left tail and a gaussian bell $g_0(u)$ for the central bulk:

$$g_+(u) = \begin{cases} \mathcal{A}_+(u - c_+)^{-\gamma_+} & \text{if } u \geq u_+ \\ 0 & \text{otherwise} \end{cases}$$
$$g_-(u) = \begin{cases} \mathcal{A}_-(-u - c_-)^{-\gamma_-} & \text{if } u \leq u_- \\ 0 & \text{otherwise} \end{cases}$$
$$g_0(u) = \begin{cases} \frac{1}{\mathcal{A}_0}\exp(-\frac{1}{2}u^2) & \text{if } -u_- \leq u \leq u_+ \\ 0 & \text{otherwise} \end{cases}$$

where $\mathcal{A}_0 = \int_{-u_-}^{u_+} du\, \exp(-\frac{1}{2}u^2)$ and $\mathcal{A}_\pm = (\gamma_\pm - 1)\left(\frac{\gamma_\pm}{u_\pm}\right)^{\gamma_\pm - 1}$ are normalization constants. Then we put them together in a single PDF:

$$g_*(u) = a_- g_-(u) + a_0 g_0(u) + a_+ g_+(u) \tag{1.11}$$

There are five parameters to fix. First we must impose the normalization of the new PDF in (1.11):

$$a_- + a_0 + a_+ = 1 \quad \Rightarrow \quad a_0 = 1 - a_- - a_+$$

then we can fix the other four parameters by imposing the continuity of g_* and the continuity of its derivative too:

$$c_\pm = u_\pm - \frac{\gamma_\pm}{u_\pm}$$

$$a_\pm = \frac{\varphi_\pm}{1 + \varphi_- + \varphi_-}$$

where

$$\varphi_\pm = \frac{\gamma_\pm}{\mathcal{A}_0(\gamma_\pm - 1)u_\pm}\exp(-\frac{1}{2}u_\pm^2)$$

How can we get a stochastic variable with the same density function of g_* in eq. (1.11)? A transformation T_0 from two independent uniformly distributed variables to a single variable Y distributed according to a gauss' law is well known[21], see [10]. Two transformations from a uniformly distributed variable $0 < X \leq 1$ to a variable Y distributed according to $g_\pm(u)$ are available too:

$$Y = T_+(X) = c_+ + \frac{\gamma_+}{u_+}X^{-\mu_+}$$

and

$$Y = T_-(X) = -c_- - \frac{\gamma_-}{u_-}X^{-\mu_-}$$

[21] such procedure gives a Gauss' law on the whole real axis; to limit it between $-u_-$ and u_+ simply repeat the procedure when you get a Y outside this range

where $\mu_\pm = \dfrac{1}{\gamma_\pm - 1}$.

Therefore we can use these three transformations to get the PDF g_* from uniformly distributed variables: we choose one and generate the variable Y. If we choose the transformation T_0 we get Y in the central bulk, while if we choose T_- or T_+ we get Y on the left or right tail respectively. At each stage we must choose one among these trasformations. In order Y to be distributed according to $g_*(u)$ we must choose every transformation with the right frequency. Since the frequency for Y to stay in the central bulk or in the tails correspond to the weights a_0 and $a_\pm$, then at each stage we can extract again an uniformly distributed variable $0 < \mathcal{X} \leq 1$ to decide the right transformation:

$$\begin{array}{lcl} \text{if } \mathcal{X} \leq a_- & \Longrightarrow & \text{transformation } T_- \\ \text{if } 1 - \mathcal{X} \leq a_+ & \Longrightarrow & \text{transformation } T_+ \\ \text{otherwise} & \Longrightarrow & \text{transformation } T_0 \end{array}$$

In order to fit the scaling PDF g with g_*, in eq. (1.11), a further step is in order: due to asymmetry the mean of g_* does not vanish; moreover its standard deviation, namely its width, may be different from that of g, see table (1.3). The mean m_Y and the variance σ_Y^2 of $g_*(u)$ are:

$$m_Y = \frac{(\frac{u_+}{\gamma_+} + \frac{\gamma_+}{\gamma_+ - 2} u_+)\varphi_+ - (\frac{u_-}{\gamma_-} + \frac{\gamma_-}{\gamma_- - 2} u_-)\varphi_-}{1 + \varphi_+ + \varphi_-}$$

$$\sigma_Y^2 = \frac{1 + [\frac{2\gamma_+}{\gamma_+ - 2}(1 + \frac{\gamma_+}{\gamma_+ - 3} u_+^{-2}) + \frac{u_+^2}{\gamma_+}]\varphi_+ + [\frac{2\gamma_-}{\gamma_- - 2}(1 + \frac{\gamma_-}{\gamma_- - 3} u_-^{-2}) + \frac{u_-^2}{\gamma_-}]\varphi_-}{1 + \varphi_+ + \varphi_-} - m_Y^2$$

It worth noting that these quantities make sense as we set $\gamma_\pm > 3$. Therefore once we got our set of realizations of the variable Y we perform a final transformation:

$$Z = \frac{\sigma}{\sigma_Y}(Y - m_Y)$$

where σ is the standard deviation of g, see table (1.3). The density function of Z is like that of $g_*(u)$ in eq. (1.11), but it has a vanishing mean and the right variance. In fig. (1.10) we put this density function on the scaling PDF $g(u)$ of the DJI index. Taking $\gamma_\pm$ according to the table (1.2), and setting the starting points $u_- = 1.5$ and $u_+ = 1.9$, the overlap is almost perfect.

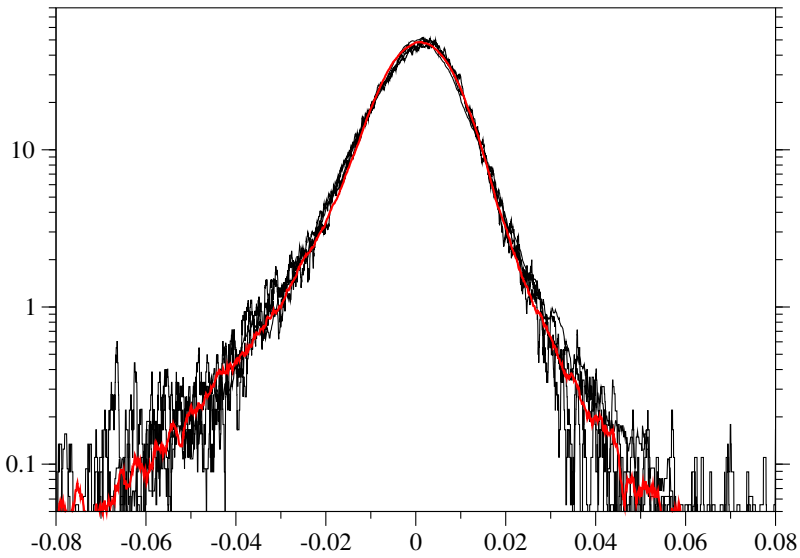

Figure 1.10: Comparison between the scaling PDF $g(u)$ as obtained by rescaling the PDFs $p_\tau(u)$ of the returns $r_t(\tau)$ (*black curves*) and the density function of the variable Z defined in the text (*red curve*).

Chapter 2

The generalized Hurst exponent analysis

We saw in section (1.3) that the time series of our financial indexes are scale invariant since all the density functions of the returns, if properly rescaled according to a certain Hölder exponent, collapse in a single PDF; see eq. (1.5). In this chapter we are going to deal with the generalized Hurst exponent analysis: the main statistical tool to detect and characterize the scaling invariance of any empirical time series.

We first determine the Hurst exponent of two well known, exactly soluble models: the brownian motion and the fractional Brownian motion; then we estimate its uncertainty by means of a bootstrap method. The comparison between the empirical results and the expected ones provides us with some interesting information about the Hurst exponent reliability. We conclude by calculating the Hurst exponent for the time series of the financial indexes seen in the previous chapter.

2.1 Some notions of probability theory

We start by reviewing more precisely some concepts already met in the previous chapter, like that of stochastic variable, moment and process. Consider a *probability space* $(\Omega, \mathcal{P})$, namely a space of events Ω equipped with a probability measure $\mathcal{P}$. A *stochastic variable* is any measurable[1] function $x : \Omega \mapsto \mathbb{R}$; and its *probability density function* (PDF) $p_x : \mathbb{R} \mapsto \mathbb{R}$ is defined as:

$$\int_B p_x(u)\, du = \mathcal{P}(x \in B)$$

[1] a measurable function between two spaces maps a measurable set of the first space into a measurable set of the second one. All functions we usually meet are measurable.

where

$$(x \in B) : \omega \in \Omega | x(\omega) \in B$$

is the measurable set containing all the events for which our stochastic variable takes values inside any given measurable[2] subset $B \in \mathbb{R}$. Starting from the very definition of the probability $\mathcal{P}$, see [8], it is easy to check that the following properties hold for any PDF[3]:

- $p_x(u) \geq 0$
- $\int_{\mathbb{R}} p_x(u)\, du = 1$

In particular the last property says that p_x is normalized. What is the physical meaning of the density function p_x? Suppose to perform an experiment on some physical system in order to measure a certain observable physical quantity x. Every time we get the realization this observable an event $\omega \in \Omega$ occurs and the outcome of x is a certain number $x(\omega)$. The measure $\mathcal{P}(A)$ gives us the probability for the occurred event to belong to a given subset $A \in \Omega$ of events, so $p_x(u_0)$ is just the probability for the outcome $x(\omega) = u_0$ to occur in our experiment[4].

We remark that the space Ω is often completely unknown and usually it does not matter either: it is a mathematical device only. On the contrary, the stochastic variables and their PDFs are really observable. Furthermore, since the PDF provides us with all the informations about the stochastic variable, whatever the probability space $(\Omega, \mathcal{P})$ may be, it is actually the only important quantity[5]. If this PDF is not known a priori, one has to get it from the experiment. We can build the empirical PDF $p_x^{(e)}$ of a physical observable x simply by repeating the experiment many times and calculating the frequency of each outcome $x(\omega)$. Suppose that N realizations of x have been performed, thus obtaining a set of N outcomes $u_i \quad i = 1 \dots N$, then

$$p_x^{(e)} = \frac{1}{N} \sum_{i=1}^{N} \delta(u - u_i)$$

where $\delta(u)$ is the usual Dirac pointwise distribution.

Even if the PDF p_x contains all the information about the stochastic variable x, however we need a tool to extract the particular information we are interested

[2] according to the usual Lebesgue measure

[3] moreover a given function which satisfies both these conditions is the density function of some stochastic variable

[4] or better, as the name *density* itself suggests, the integral $\int_{u_0-\varepsilon}^{u_0+\varepsilon} p_x(u)\, du$ gives the probability for $x(\omega)$ to lie between $u_0 - \varepsilon$ and $u_0 + \varepsilon$

[5] moreover it can be shown that each PDF comes from a single stochastic variable, so the association $x \mapsto p_x$ is an injection. This is why in many theoretical applications p_x is the very definition of x

in. In dealing with a stochastic quantity x, whose outcome $x(\omega)$ at each realization cannot be forecasted exactly, surely the most interesting information is its mean $\langle x \rangle$, or mathematical expectation. This is because it is in a sense the best forecasting at any realization, or better because $\langle x \rangle$ has the least distance from every possible outcome $x(\omega)$. Using the density function of the stochastic variable x it is possible to calculate the *mean* of any observable function $y = f(x)$. It is defined by an average over all the outcomes of the variable x:

$$\langle y \rangle = \int_{\mathbb{R}} f(u)\, p_x(u)\, du \tag{2.1}$$

From a statistical point of view, knowing the best expected value of x is not enough: we also need to know how much it is good. To this end one can calculate the average square distance between the mean $\langle x \rangle$ and any possible outcome of the stochastic variable x, namely the *variance*:

$$\begin{aligned} \sigma^2(x) &= \langle (x - \langle x \rangle)^2 \rangle \\ &= \langle x^2 \rangle - \langle x \rangle^2 \end{aligned}$$

The larger the variance the further are the realizations of x from the mean $\langle x \rangle$. Hence, from a geometrical viewpoint, the PDF p_x is centred in the mean $\langle x \rangle$ and has a characteristic width equal to the standard deviation[6] $\sigma(x)$.

Empirically the mean of a given function $f(x)$ can be achieved by using $p_x^{(e)}(u)$ in eq. (2.1), thus getting an expression like[7] eq. (1.3) in section (1.2):

$$\langle f(x) \rangle_e = \frac{1}{N} \sum_{i=1}^{N} f(u_i) \tag{2.2}$$

where $u_i \quad i = 1 \dots N$ are, as before, N realizations of the stochastic observable x coming from a given experiment. In particular, according to the previous interpretation the empirical mean $\langle x \rangle_e$ is regarded as the best value of the available data set and the empirical standard deviation as its uncertainty.

[6] there are other interesting quantities whose value help us in describing other characteristics of p_x, such as the skewness and the kurtosis; see table (1.3) in the previous chapter. The skewness

$$skew(x) = \frac{\langle (x - \langle x \rangle)^3 \rangle}{\sigma^3(x)}$$

vanishes only if p_x is symmetric and the kurtosis

$$kurt(x) = \frac{\langle (x - \langle x \rangle)^4 \rangle}{\sigma^4(x)} - 3$$

is positive or negative according to whether p_x is sharper or flater than a Gauss' law

[7] here we put the subscript e in order to tell the empirical mean from the theoretical one

It is worth noting, at this stage, that the empirical mean in eq. (2.2) depends on the actual outcomes of N realizations of the stochastic variable x while the theoretical mean in eq. (2.1) is a deterministic quantity depending only on x itself. Since the PDF of a physical observable is not known a priori and thus the theoretical means $\langle\cdot\rangle$ cannot be calculated, we can only get the empirical ones $\langle\cdot\rangle_e$ and hope they are good approximations to the formers, at least for N large enough. The following theorem [12] warrants that this is actually the case, provided that the theoretical limit exists.

Theorem 1 (Law of Large Numbers). *Given N independent realizations of a stochastic variable x and any function $f(x)$, if the mean $\langle f(x)\rangle$ exists then*

$$\langle f(x)\rangle_e \to \langle f(x)\rangle$$

almost surely for $N \to \infty$.

It is worth stressing again that the mean $\langle x\rangle$ is a number, while the empirical mean $\langle x\rangle_e$ is a stochastic variable depending on N realization of x. The almost sure convergence in the previous theorem means that the set of events for which the convergence fails has probability $\mathcal{P} = 0$.

There are some special functions of a stochastic variable, other than the variance, which play a key role in probability theory.

The characteristic function (CF) is the mean of e^{ikx}:

$$\varphi_x(k) = \int_{\mathbb{R}} e^{iku} p_x(u)\, du \tag{2.3}$$

This may seem a rather abstract quantity, but it turns out to be very useful especially when the sum of many independent stochastic variable is taken into account; as in the Central Limit Theorem. Their importance mainly come from the injective character of the relation $p_x \mapsto \varphi_x$, i.e. two PDFs cannot have the same CF[8].

The q-order absolute moments are the mean of $|x|^q$:

$$M_q(x) = \int_{\mathbb{R}} |u|^q p_x(u)\, du \tag{2.4}$$

[8]indeed one can get some information about the PDF directly from its CF. It is easy to prove that

- if φ_x is symmetric then so is p_x and viceversa;
- $\langle x\rangle = -i\varphi_x'(0)$ by derivation.

where $q \in \mathbb{R}^+$ is any positive real number. While the CF always exist, the theoretical q-order moment $M_q(x)$ may exist or not depending on the behaviour of the PDF $p_x(u)$ for large arguments $|u| \gg 1$, see discussion in section (1.4). In particular if p_x has tails which follow a power law of parameter γ then the moment $M_q(x)$ does not exist for $q \geq \gamma - 1$, since the integral in eq. (2.4) diverges.

From now on we will refer to $M_q(x)$ simply as moments[9]. In the coming section they will be the main tool to study the invariance properties of a given time series, so their existence is a very important point for us. In particular it is worth noting that the empirical moments $\langle |x|^q \rangle_e$ exist whatever the order q may be. Indeed suppose again that a set of realizations $u_i \quad i = 1 \dots N$ is available, then the calculation of the empirical q-order moment

$$M_q^{(e)}(x) = \frac{1}{N} \sum_{i=1}^{N} |u_i|^q \tag{2.5}$$

always gives a finite number, since the summation is necessarily finite. However if the theoretical limit $M_q(x)$ does not exist, the empirical moments (eq. 2.5) do not approximate anything, since theorem (1) does not hold anymore; and thus they become meaningless. During the whole chapter we will deal with stochastic variables x whose PDF decays rapidly enough to warrant the existence of the theoretical moments $M_q(x)$ for all orders q. The problems arising from PDFs which make the moments infinite will be the main issues of the next chapter.

2.2 Stochastic processes

The formalism above generalizes at once to many stochastic variables $\vec{x} : \{x(1) \dots x(n)\}$: the joint density function $p_{\vec{x}}$ is defined as

$$\int_B p_{\vec{x}}(\vec{u})\, du_1 \dots du_n = \mathcal{P}(\vec{x} \in B)$$

where

$$(\vec{x} \in B) : \omega \in \Omega | \vec{x}(\omega) \in B$$

and $B \in \mathbb{R}^n$ is any measurable subset. The joint PDF of many variables must satisfy conditions similar to the PDF of a single variable:

- $p_{\vec{x}}(\vec{u}) \geq 0$

[9] for the sake of precision the moments are the mean of x^m with the order m compelled to take only integer values

- $\int_{\mathbb{R}^n} p_{\vec{x}}(\vec{u})\, du_1 \dots du_n = 1$
- $p_{x_k}(u_k) = \int_{\mathbb{R}^{n-1}} p_{\vec{x}}(\vec{u})\, du_1 \dots du_{k-1} du_{k+1} \dots du_n$

Where the last condition is known as marginal probability formula.

The expected value of some function $y = f(\vec{x})$ reads

$$\langle y \rangle = \int_{\mathbb{R}^n} f(\vec{u}) p_{\vec{x}}(\vec{u})\, du_1 \dots du_n$$

In particular we are often concerned with the sum of many stochastic variables $y = \sum_{i=1}^{n} x_i$ since in most of applications an observable effect is often the result of a large number of microscopical contributions. It is easy to show from the definition that

$$\langle y \rangle = \sum_{i=1}^{n} \langle x_i \rangle$$

but in this general framework we cannot go further. In order to get more informations about the distribution of y we need to impose the independence of the variables x_i.

The variables x_i are said to be *independent* if their joint PDF is the product of all the PDFs p_{x_i}:

$$p_{\vec{x}}(\vec{u}) = \prod_{i=1}^{n} p_{x_i}(u_i) \tag{2.6}$$

This relation warrants that the knowledge of the realization of some variable does not affect the realization of the others. Indeed, for the sake of simplicity, take $n = 2$ and suppose to know that the first variable belong to a certain interval: $x_1 \in A$. In this case, using eq. (2.6), the density of x_2 writes

$$\int_{u_1 \in A} p_{x_1\, x_2}(u_1\, u_2)\, du_1 = \mathcal{P}(x_1 \in A)\, p_{x_2}(u_2)$$

However $\mathcal{P}(x_1 \in A) = 1$ whatever the interval A be, since we already know that this result has occurred; so the density function of x_2 is just p_{x_2}.

While very few theoretical result are known when the variables $\vec{x}$ are dependent, many powerful theorem are available when they are independent, especially for what concern their sum y. A first result is the summation rule for the variance:

$$\sigma^2(y) = \sum_{i=1}^{n} \sigma^2(x_i)$$

which does not hold in general for dependent variables. However the most important consequence of independence, coming at once from the substitution of eq. (2.6) in

the very definition of CF, eq. (2.3), is that the CF of y writes in a quite simple form:

$$\varphi_y(k) = \prod_{i=1}^{n} \varphi_{x_i}(k) \tag{2.7}$$

Remembering the injective character of the association between a PDF and its CF, we can thus build, at least in theory, the PDF of y starting from those of the variables x_i. This possibility is one of the key points which allow the Central Limit Theorem to work.

The generalization to many stochastic variables allows us to define a process:

Definition 1. *a stochastic process is a continuos set of random variables $x(t)$ at each instant t.*

A process behaves like a collection of uncountably many stochastic variables. Given an event $\omega \in \Omega$ we get a realization of our process as a *sample path* in $\mathbb{R}^2$:

$$\mathcal{G}(\omega) : (t; x(t, \omega)), t \in \mathbb{R}$$

which is simply the collection of all the realizations $x(t, \omega)$ along the time t.

In many physical experiments we get a set of results of a certain observable $x(t)$ which depend on time; namely a time series. From an empirical viewpoint the time t can take only discrete values, namely the instants at which we get the realization of $x(t)$. Hence a time series can be regarded as the discretization of a stochastic process. The financial indexes we studied in the first chapter provide some examples of a real time series while fig. (1.1) provides an example of sample path. From now on, in dealing with a mathematical process or with an empirical time series, we will denote both the discrete time and the continuos one with the same symbol t.

We could define the density function of $x(t)$ for each instant t, but for our purposes we will be concerned mainly with the increments of the process, rather than with the process itself. This is because, as explained in section (1.3), in considering the increments over different time windows we are actually looking at the process on different time scales. Hence, the increments are the most suitable variables to the study of the scaling invariance. The *increments* over a time window τ

$$r_t(\tau) = x(t + \tau) - x(t)$$

are also stochastic variables, like $x(t)$, but they depend on both the instant t and the window τ, see eq. (1.2) in section (1.1). Therefore their density function $p_{t,\tau}(u)$, and thus the respective q-order moment

$$M_q(t, \tau) = \int_{\mathbb{R}} |u|^q p_{t,\tau}(u)\, du$$

also depends on t and τ.

The moments $M_q(t,\tau)$ of the increments $r(t,\tau)$ are just the quantities with which we are going to deal in the next section. In order to empirically calculate $M_q(t,\tau)$ we need a set of realizations of the increment $r_t(\tau)$ for each instant t and each time window τ. In other words, for a fixed pair (t,τ) we should repeat the experiment many times, thus getting a set of time series, and draw out from each realization a single increment, namely that at the instant t of window τ. However, often this is not possible since a single history is only available, like in finance for instance. One can get around such a difficulty by means of the stationarity assumption and of the so called sliding window method, see the remarks at the end of section (1.1).

In a process with*stationary increments* the distribution of $r_t(\tau)$ does not depend on the instant t, but on the window τ only. Hence, for such a process, we can drop the subscript t and call p_τ the common distribution of all increments $r_t(\tau)$ along the time; the increments themselves will be denoted simply with $r(\tau)$. Moreover the q-order moment of these increments $r(\tau)$ becomes:

$$M_q(\tau) = \int_{\mathbb{R}} |u|^q p_\tau(u)\, du \tag{2.8}$$

We stress again that the importance of the stationary assumption obviously is not merely a matter of notations: often it is the only way to handle empirical data, as it allows us to collect all the returns $r(\tau)$ along the same time series.

Let the size of our stationary time series be N, for instance; then, due to the stationarity, we can collect N returns $r(1)$ for a time window $\tau = 1$. However for larger windows $\tau \geq 2$ a problem arises: if we collect the returns $r(\tau)$ along the same time series without overlapping we do not get N returns again, but $\approx \dfrac{N}{\tau}$ only. This is because the number of disjoint returns reduces rapidly as the size of the window τ increases. Since we need as many data as possible to perform accurate statistical analysis, we will collect the returns along our time series by means of a so called *sliding window* method.

The sliding window method consists in collecting a return $r(\tau)$ for each time instant, no matter if they overlap. Using this method we can collect $N - \tau \approx N$ returns $r(\tau)$ for each time window τ, thus avoiding any loss of data. Hence the empirical moments writes

$$M_q^{(e)}(\tau) = \frac{1}{N} \sum_{i=1}^{N} |r_i(\tau)|^q \tag{2.9}$$

In what follows we are going to consider processes with stationary increments only. We are going also to assume stationarity in all the empirical time series we study, like the financial ones we saw in the first chapter, and to collect the returns with a sliding window method.

2.3 Self-affine processes

Now we are ready to define self-affine processes. A stochastic process with stationary increments $x(t)$ is said *self-affine* if there exist $\alpha \geq 0$ such as, for any time window τ:

$$r(\tau) \sim \tau^{\alpha} r(1) \tag{2.10}$$

where $\sim$ means *equally distributed to*, and α is called the Hölder exponent. This equation, which involves returns, implies at once a scaling structure for their PDFs:

$$p_{\tau}(u) = \frac{1}{\tau^{\alpha}} g(\frac{u}{\tau^{\alpha}}) \tag{2.11}$$

where $g(u)$ is the scaling PDF, actually the PDF of the returns over a unit time window $\tau = 1$, see section (1.3). Eq. (2.11) has in turn a very interesting consequences on the q-order moments $M_q(\tau)$ defined in eq. (2.8):

$$M_q(\tau) = \tau^{\alpha q} M_q \tag{2.12}$$

where $M_q = \langle |r(1)|^q \rangle$ is the q-order moment of the scaling PDF g.

As we have already said, the theoretical moments may not exist, or better the integral in eq. (2.8) may diverge. In the case of a strictly self-affine process the existence of the moments depends on the behaviour of the scaling function g only: in order to make M_q finite the tails of g must decay faster than $u^{-(q+1)}$. Since the moments will be the main tools to study self-affinity and more generally the invariance under rescaling, in this chapter we will always consider scaling PDFs whose tails decay fast enough to ensure the existence of the theoretical moment M_q for every order q. This in turn means that $g(u)$ decays asymptotically faster than any power law. For instance, an exponential decay for large arguments $|u| \gg 1$, like that of the Gauss' law, guarantees the convergence of the integrals in eq. (2.8), no matter how large is q. The problems arising from scaling PDFs whose tails decay asymptotically as a power law, and thus too slow to ensure the existence of M_q for q beyond a certain threshold, will be the main issue of the next chapter.

We already met the Hölder exponent α in section (1.3); moreover in that section we have tried to interpret it in a qualitative way. Now we are going to see more precisely what the meaning of α is[10].

Theorem 2. *A continuos self-affine process with Hölder exponent α has a graph $\mathcal{G}$ with Hausdorff dimension*

$$D_H(\mathcal{G}) = \max(1; 2 - \alpha)$$

[10]recall that a stochastic variable or a process satisfies *almost surely* a given condition if such condition holds for the whole space of events Ω, except for a subset of vanishing measure. In other words the probability of a realization of our stochastic variable or process to do not fulfill the condition vanishes

almost surely.

The Hausdorff dimension is a generalization of the usual concept of dimension to very irregular geometric objects, like a self-affine graph for instance. It is the main tool in fractal geometry in order to distinguish and classify a huge amount of fractal sets for which the tools of the classical geometry fail to apply. Among them there are some functions displaying wild fluctuations at every scale, just like a typical realization $\mathcal{G}$ of a self-affine stochastic process.

It can be shown that the Hausdorff dimension of a smooth function embedded in the plane $\mathbb{R}^2$ is equal to unity, as expected; moreover $D_H = 2$ for any open set of the plane. This is because the Hausdorff dimension of a regular set equal the usual dimension. However a more irregular function can have a fractional dimension $1 \leq D_H < 2$ due to its wild fluctuations and above all to their presence at every scale. Hence the dimension D_H is a measure of the roughness of the graph. The theorem (2) state that the lower is the Hölder exponent α the more the sample graph $\mathcal{G}$ of the process looks irregular, so we recover here the interpretation of α we gave in section (1.3) on the basis of an heuristic argument.

The empirical evaluation of the Hölder exponent α, by means of eq. (2.12), leads us to the definition of the *Hurst exponent* H. Suppose to have a time series of size N. Obviously we cannot get the moments $M_q(\tau)$ since they are theoretical quantities, but we can collect the returns on our time series and calculate the empirical moments $M_q^{(e)}(\tau)$ as defined in eq. (2.9). The Hurst exponent tell us how these empirical moments behave under a scaling transformation, i.e. when the window τ changes.

Definition 2. *A time series is simple scaling with Hurst exponent H if there exist a constant H such as:*

$$M_q^{(e)}(\tau) \propto \tau^{Hq} \tag{2.13}$$

where the proportionality constant may depend on the order q.

Assuming the existence of all theoretical moments $M_q(\tau)$ we know that:

$$\lim_{N\to\infty} M_q^{(e)}(\tau) = M_q(\tau)$$

according to the Law of Large Numbers. This means that for a very long history, namely for a data set of large size N, the empirical moments are equal to the theoretical ones for all practical purposes. Hence, if our time series is the realization of a self-affine process, we can use eq. (2.12) to get

$$H = \alpha \tag{2.14}$$

It is worth noting that the Hurst and Hölder exponents come from two quite different definitions, even if they turn out to be equal for self-affine processes. Indeed the

former is defined using the empirical moments while the latter concerns the PDFs of the increments. Their meaning should not be confused.

Since the calculation of the empirical moments performed on a single time series is a rather easy task, it provides us with a straightforward method to get the Hurst exponent; and in turn the Hölder exponent α in the case of a self-affine process. This procedures is called the *Hurst exponent analysis*:

1. for every order q and time window τ calculate $\ln M_q^{(e)}(\tau)$ using eq. (2.9);

2. plot $\ln M_q^{(e)}(\tau)$ against $\ln \tau$ for each order q and evaluate the slope $D(q)$ of the resulting line by means of a linear best-fit:

$$M_q^{(e)}(\tau) \propto \tau^{D(q)} \tag{2.15}$$

3. evaluate again the slope of the line $D(q)$.

Due to simple scaling, the path of a self-affine process has a linear *scaling exponent*:

$$D(q) = Hq$$

As we will see in the last section of this chapter, for processes that are not strictly self-affine, the scaling function $D(q)$ may have not a costant slope, and thus an Hurst exponent $H(q)$ may depend on the order q.

There are two main problems in performing the above procedure, both coming from the finiteness of our data set. The first concerns the very existence of the moments: as we already said in section (1.4), the finite size N of our time series makes the summation in eq. (1.10) always finite, but one should take the outcome very carefully if the theoretical limit $M_q(\tau)$ does not exist. The next chapter is entirely devoted to this problem. In this chapter we take into account process with converging moments of all orders q. However, even if every moment $M_q(\tau)$ exists a second problem still remains.

Once the process is given, the theoretical moments are completely determined; on the contrary the empirical ones depend on the particular realization, namely on the particular sample path. So they are stochastic variables themselves. This in turn means that the Hurst exponent H is a stochastic variable too. Hence, its calculation must be followed by an estimation of the statistical uncertainty.

Since the uncertainty associated to the Hurst exponent H, or to the scaling exponent $D(q)$, is due to its sample dependence, usually one can obtain a confidence interval simply by repeating the calculation on many different samples, each of size N. This set of outcomes for H is a statistical ensemble whose mean and standard deviations provides us with the best value and its uncertainty. Unfortunately in many applications, like finance, a single time series is only available. There is a

single history and we cannot repeat it. However, the *bootstrap* procedure allows us to get many samples from a single time series.

The bootstrap is a powerful statistical estimator which belongs to the wider class of the so called resampling methods. The key idea is to extract many sample of size $n < N$ from the whole time series and to perform the calculation on each subsample, thus obtaining a set of values. These samples are regarded as an ensemble and the set of values thus calculated as indipendent realization of H, or $D(q)$. Hence we can look at the dispersion of this set of values or at the spreading of this bundle of curves in order to evaluate a confidence interval. Essentially a small dispersion suggest a great robustness; on the contrary if our outcomes spread over a large range it means that H is more unreliable and sample dependent. In what follows we will always evaluate the uncertainty associated to the Hurst exponent H or to the scaling exponent $D(q)$ by means of a bootstrap method.

2.4 Some simple scaling processes

In this section we are going to apply the generalized Hurst exponent analysis to two well known processes: the Brownian motion (BM) and the fractional Brownian motion (FBM). They are two self-affine processes, so the simple scaling holds rigorously for them. Hence it is surely a good idea to check the Hurst exponent reliability for these processes. Our aim is to test the generalized Hurst exponent on a simulated time series for which a theoretical result is already known. Hence we apply the procedure of the previous section, namely the Hurst exponent analysis, on a single simulation and collect all the returns $r(\tau)$ along such history with a sliding window method. Lets start with the BM.

Definition 3 (Brownian Motion). *A process $x(t)$ is a Brownian motion if*

1. *it is continuous*
2. *its non overlapping increments are independent;*
3. *$r_t(\tau)$ is distributed according to a zero mean Gaussian with variance $\sigma^2 = \sigma_0^2 \tau$.*

where σ_0 is a width parameter.

It is obvious from the definition that the BM is a stationary process as the PDF of its increments $r_t(\tau)$ does not depend on the instant t. The most important property for what concern us is the scaling law of the BM. From the definition itself, the PDF of the increments $r(\tau)$ is

$$p_\tau(u) = \frac{1}{\sqrt{2\pi\tau}} \exp[-\frac{1}{2\tau}\left(\frac{u}{\sigma_0}\right)^2]$$

Hence the BM is a self-affine process with Hölder exponent $\alpha = \frac{1}{2}$:

$$p_\tau(u) = \frac{1}{\sqrt{\tau}} g(\frac{u}{\sqrt{\tau}})$$

Where the scaling PDF is just a Gauss' law with vanishing mean and variance σ_0^2:

$$g(u) = \frac{1}{\sqrt{2\pi}} \exp[-\frac{1}{2}\left(\frac{u}{\sigma_0}\right)^2] \tag{2.16}$$

Therefore

- the sample path of a BM has Hausdorff dimension $D_H = 1.5$, see theorem (2) in the previous section. This makes the graph of a BM rather irregular[11].
- we have simple scaling with Hurst exponent $H = 0.5$:

$$M_q^{(e)}(\tau) \propto \tau^{\frac{1}{2}q}$$

To check the above properties we must first simulate a Brownian motion. Following the definition we can take our increments $r(1)$ independently from the PDF g of eq. (2.16) and then construct the whole process by summation:

$$x(t) = \sum_{i=0}^{t-1} r_i(1) \qquad 0 \le t \le N$$

This is a general scheme for a so called Random Walk. It is easy to prove that non overlapping increments are independent too. Furthermore since the returns over a time window τ

$$r_t(\tau) = \sum_{i=t}^{t+\tau-1} r_i(1)$$

are the sum of many independent, identically distributed increments, according to eq. (2.7) their CF is the product of all the CFs of these increments:

$$\varphi_{r(\tau)}(k) = \prod_{i=1}^{\tau} \varphi_{r_i(1)}(k)$$

The CF of the Gauss' law in eq. (2.16) is $\exp(-\frac{1}{2}\sigma_0^2 k^2)$, thus

$$\varphi_{r(\tau)}(k) = \exp(-\frac{1}{2}\sigma_0^2 \tau k^2)$$

[11] indeed a Brownian motion has not derivative at any point almost surely, otherwise it would have Hausdorff dimension $D_H = 1$

which is again the CF of a Gauss' law[12], but with variance $\sigma_0^2\tau$. Hence the returns $r(\tau)$ have the right PDF and this random walk approximates a Brownian motion[13].

In fig. (2.1) we show the sample path of a BM with $N = 25000$ and $\sigma_0 = 0.01136$; the time series has about the same size of our financial indexes, see table (1.1), and the standard deviation σ_0 was chosen accordingly, see the unconditioned standard deviation of the daily returns of the DJI index in table (1.3). In fig. (2.2) we can see the scaling function $D(q)$ calculated on the whole time series and on several uniformly extracted samples of size $N = 5000$. Clearly the Hurst exponent analysis gives a simple scaling with a very low uncertainty. Our expectation $H = 0.5$ is fulfilled regardless of the sample extracted. However we stress that in this case all the theoretical moments exist.

So far we took a random walk with increments distributed according to a Gauss' law, but this demand is not necessary in order to have simple scaling with $H = 0.5$. Suppose the increments $r(1)$ are independent and equally distributed according to any density function with finite variance, then the Central Limit Theorem states that the returns $r(\tau)$ are asymptotically distributed according to a Gauss' law. Hence, at least asymptotically, the process would seem a BM. This why the particular value $H = 0.5$ is called *normal* scaling. In order to get *anomalous* scaling, namely a simple scaling process with $H \neq 0.5$, we need to relax one of the conditions which allow the CLT to work. In this chapter we will deal always with density functions whose moments exist for all orders q; we devote the next one to process for which high order moments fail to converge. However, we can consider dependent increments.

Definition 4 (Fractional Brownian Motion). *A process $x(t)$ is a fractional Brownian motion of index α if*

1. *its sample path is continuous;*
2. *$r_t(\tau)$ is distributed according to a zero mean Gaussian with variance $\sigma^2 = \sigma_0^2\tau^{2\alpha}$.*

where σ_0^2 is a width parameter, namely the variance of the increments $r(1)$.

If $0 < \alpha < 1$ a process satisfying the above definition exists [2]. As in the case of BM the self-affinity of a FBM is obvious from definition, but now the Hölder exponent is no longer equal to $\frac{1}{2}$:

$$r(\tau) \sim \tau^{\alpha} r(1)$$

There are two main consequences.

[12] in other words the Gauss' law is stable. We will meet again the class of the stable law in the next chapter

[13] one may think to the first condition as always satisfied since the continuity concerns a limit which does not make sense for a discrete random walk

- According to theorem (2) the sample path has dimension $D_H = 2 - \alpha$. Hence upon varying α we may have a very irregular path with large fluctuations, for α close to 0, or a more regular graph for an α close to 1.

- The scaling function $D(q)$ turns out to be linear with slope α.

In order to check this expectation we apply the generalized Hurst exponent analysis to a simulated FBM. However here a problem arises as we know that any random walk with independent increments cannot help us in simulating a FBM, so we must first describe a procedure to get such a process.

Note that the definition does not demand independence among increments; indeed they are not. Remembering that $x(t) = r_0(t)$ as $x(t) = 0$ one can calculate the linear correlation between the present value $x(t)$ of the process and the next return $r_t(\tau)$.

$$\langle x^2(t+\tau)\rangle = \langle x^2(t)\rangle + \langle r_t^2(\tau)\rangle + 2\langle x(t) r_t(\tau)\rangle$$

Hence, using that

$$\sigma^2 = \langle x^2(t)\rangle$$

due to the vanishing of the mean $\langle x(t)\rangle = 0$, one get:

$$\langle x(t) r_t(\tau)\rangle = \frac{1}{2}\sigma_0^2[(t+\tau)^{2\alpha} - (t)^{2\alpha} - (\tau)^{2\alpha}] \tag{2.17}$$

It is worth stressing that if $\alpha = \dfrac{1}{2}$, as in the case of a BM, then the last expression vanish. Otherwise the increments are correlated. For $\alpha < \frac{1}{2}$ the next increments is negatively correlated with the present value of the process: a positive past trend give rise to a negative coming return. On the contrary, for $\alpha > \frac{1}{2}$, positive values likely give rise to positive returns.

Using eq.(2.17) we easily find the correlation between the values of the process at two different instants $t_2 \geq t_1$. Since $\langle x(t_1)x(t_2)\rangle = \langle x^2(t_1)\rangle + \langle x(t_1) r_{t_1}(t_2 - t_1)\rangle$:

$$\langle x(t_1)x(t_2)\rangle = \frac{1}{2}\sigma_0^2[t_1^{2\alpha} + t_2^{2\alpha} - (t_2 - t_1)^{2\alpha}]$$

Then we use this result to define a random walk and show such random walk to be the discretization of a FBM.

1. Take the symmetric matrix $\mathbb{A}$ of size N with entries

$$\mathbb{A}_{j\,i} = \frac{1}{2}[j^{2\alpha} + i^{2\alpha} - |j - i|^{2\alpha}]$$

2. Let us consider the matrix $\mathbb{L}$ such as

$$\mathbb{L}^T\mathbb{L} = \mathbb{A}$$

The matrix $\mathbb{L}$ is just the Cholesky decomposition of $\mathbb{A}$ and, due to the symmetry of $\mathbb{A}$ itself, it is actually an upper-triangular matrix [16].

3. Let $\vec{B}$ be a vector of N independent realizations of a stochastic variable distributed according to a Gauss' law, as in eq. (2.16). The elements B_i of $\vec{B}$ are the increments of a BM.

4. Build the vector $\vec{F}$ by applying the matrix $\mathbb{L}$ to $\vec{B}$:

$$\vec{F} = \mathbb{L}\vec{B}$$

The elements F_i of $\vec{F}$ are the values of our FBM at time i.

To see that the random walk F_i for $j = 1 \dots N$ is a FBM of lenght N we start by proving that it has the right correlations. Indeed

$$\begin{aligned}
\langle F_j F_i \rangle &= \langle (\sum_{\mu=1}^{N} \mathbb{L}_{j\,\mu} B_\mu)(\sum_{\nu=1}^{N} \mathbb{L}_{i\,\nu} B_\nu) \rangle \\
&= \sum_{\mu\,\nu} \mathbb{L}_{j\,\mu}\mathbb{L}_{i\,\nu}\langle B_\mu B_\nu \rangle \\
&= \sigma_0 \sum_{\mu} \mathbb{L}_{j\,\mu}\mathbb{L}_{i\,\mu} = \sigma_0 \mathbb{A}_{j\,i}
\end{aligned}$$

where the independence and the variance σ_0 of the elements of $\vec{B}$ have been used. Our process has thus the expected correlations among their values. To show that the returns over a time window τ are distributed according to a Gauss' law with zero mean and standard deviation $\sigma^2 = \sigma_0\tau^\alpha$ we can use the characteristic function of F_j. Since the characteristic function of B_i is $\varphi_{B_i}(k) = e^{-\frac{1}{2}k^2}$ we get[14], using the

[14] here we need the scaling property of the characteristic functions. Let x be a stochastic variable and c any real constant, then cx is a stochastic variable too. From the definition of CF in eq. (2.3) it is easy to prove that

$$\varphi_{cx}(k) = \varphi_x(ck)$$

summation rule of the CFs (eq. 2.7):

$$\begin{aligned}\varphi_{F_i}(k) &= \prod_{\mu=1}^{N} \varphi_{\mathbb{L}_{i\,\mu} B_\mu}(k) \\ &= \prod_{\mu=1}^{N} \varphi_{B_\mu}(\mathbb{L}_{i\,\mu} k) \\ &= \exp(-\frac{1}{2}\sum_{\mu=1}^{N} \mathbb{L}_{i\,\mu}^2 k^2)\end{aligned}$$

Due to the symmetry $\mathbb{L}_{i\,\mu} = \mathbb{L}_{\mu\,i}$

$$\sum_{\mu=1}^{N} \mathbb{L}_{i\,\mu}^2 = \mathbb{A}_{i\,i}$$

and so

$$\varphi_{F_i}(k) = \exp[-\frac{1}{2}(i^\alpha k)^2]$$

This means that the process is self-affine with Hölder exponent α:

$$F_i \sim i^\alpha F_1$$

and, as expected, the variance grows accordingly.

In fig. (2.3) and (2.4) we show the sample path of a FBM with Hölder exponent $\alpha = 0.25$ and $\alpha = 0.75$ respectively. The size of both the time series is $N = 5000$ and the width parameter was chosen according to that of the DJI index, see table (1.3): $\sigma_0 = 0.01136$. It is worth noting that, according to theorem (2) the sample path of the first simulation looks much more irregular than that of the second one.

In fig. (2.5) we can see the scaling function $D(q)$ calculated on the whole time series and on several uniformly extracted samples of size $N = 1000$ for both the previous simulations. Clearly the Hurst exponent analysis gives a simple scaling with a quite low uncertainty. Our expectation $H = 0.25$ for the first simulation and $H = 0.75$ for the second one is fulfilled regardless of the sample extracted. However, we remark that, for every FBM, the q-order moment of the returns $r(\tau)$ always exists, no matter how large is q.

2.5 The concept of multiscaling

Until now we have regarded the Hölder exponent α as a constant. Indeed we defined it in eq. (2.10) for strictly self-affine processes only. In this section we are going

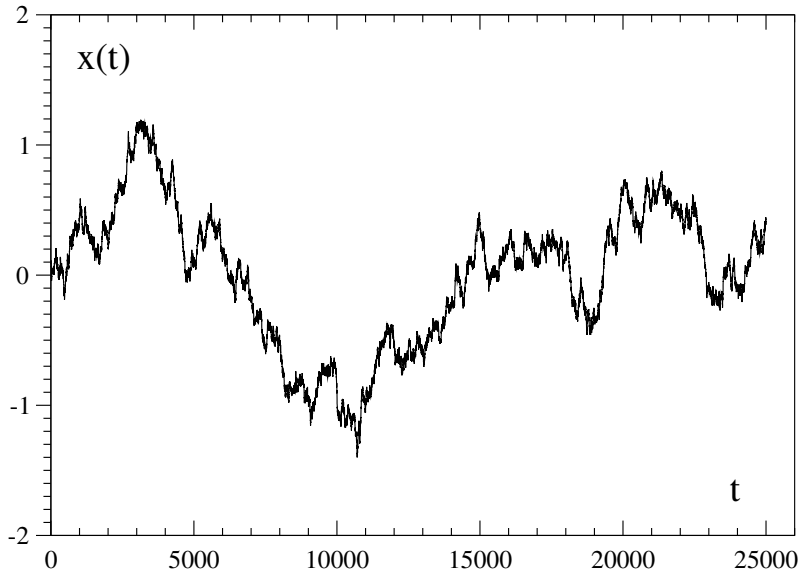

Figure 2.1: Random walk simulation of a Brownian motion of size $N = 25000$. The independent increments were drawn out from a zero mean Gauss' distribution with standard deviation $\sigma_0 = 0.001136$.

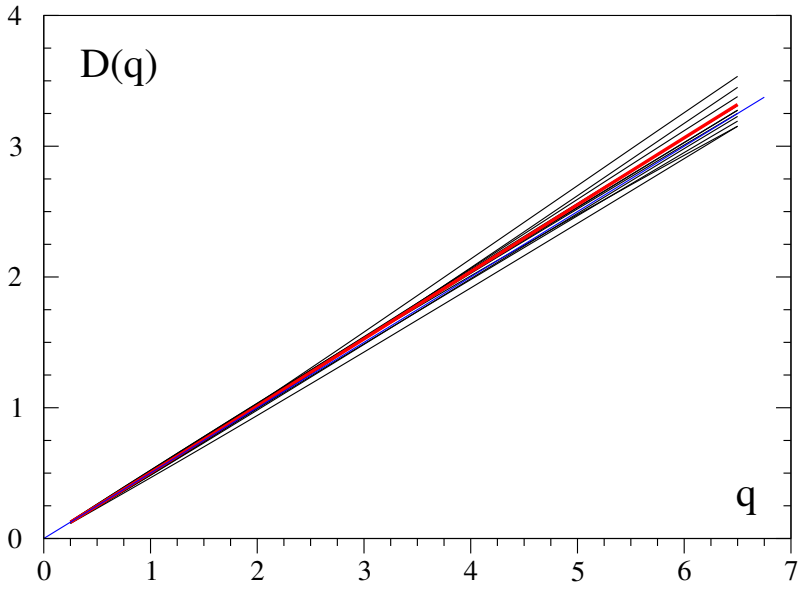

Figure 2.2: Scaling function $D(q)$ for several samples (N=5000,*black curves*) extracted from the whole time series (*red curve*) of fig. (2.1). The theoretical line $D(q) = \frac{1}{2}\,q$ is drawn in *blue*.

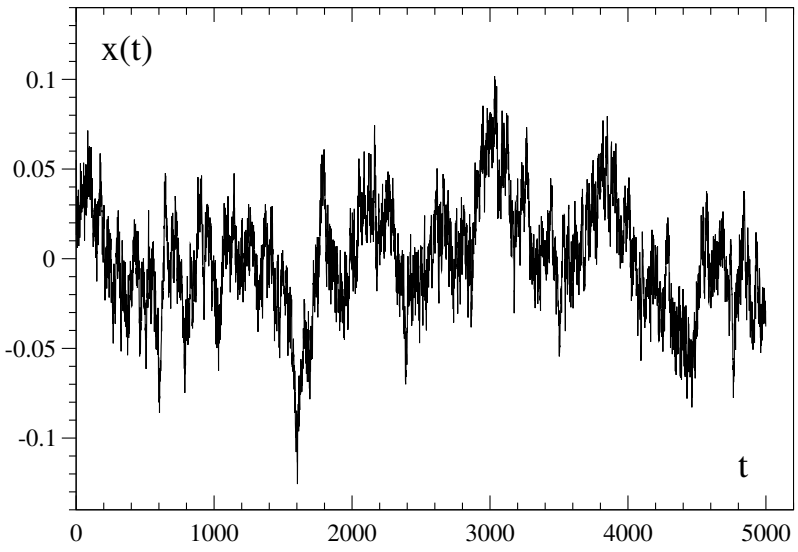

Figure 2.3: Random walk simulation of size $N = 5000$ of a Fractional Brownian motion with Hölder exponent $\alpha = 0.25$. The increments are distributed according to a zero mean Gauss' law with standard deviation $\sigma_0 = 0.001136$.

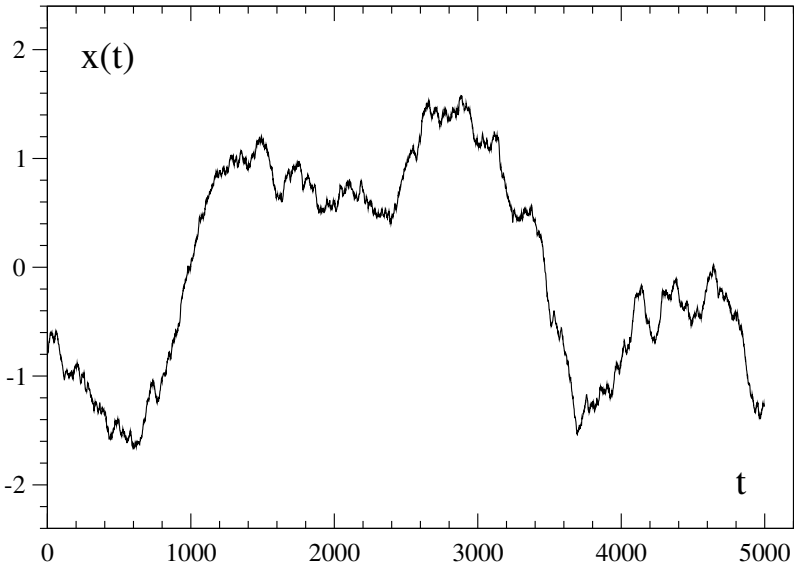

Figure 2.4: Random walk simulation of size $N = 5000$ of a Fractional Brownian motion with Hölder exponent $\alpha = 0.75$. The increments are distributed according to a zero mean Gauss' law with standard deviation $\sigma_0 = 0.001136$.

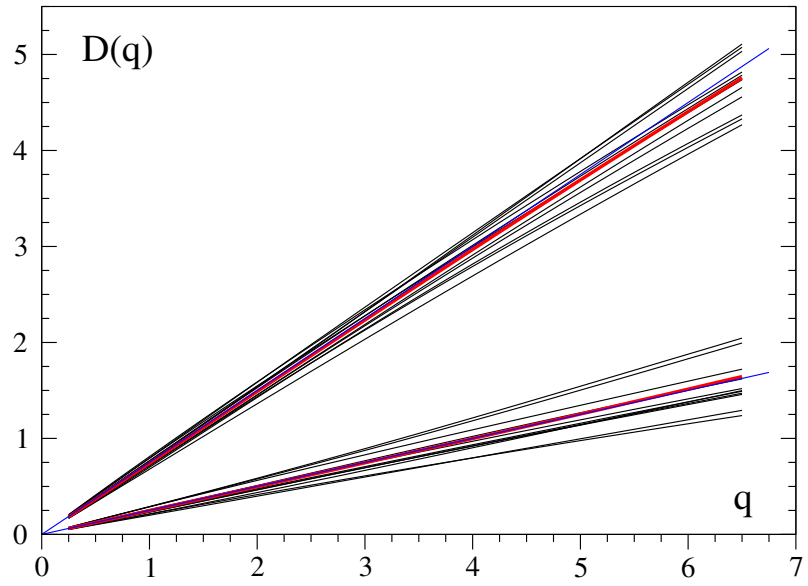

Figure 2.5: Scaling function $D(q)$ for several samples (N=1000,*black curves*) extracted from the whole time series (*red curve*) of fig. (2.3) and (2.4). The theoretical lines $D(q) = \frac{1}{4}\,q$ and $D(q) = \frac{3}{4}\,q$ are drawn in *blue*.

to generalize the concept of Hölder exponent to process for which self-affinity does not hold. Let $x(t)$ be a continuous stochastic process and $r_t(\tau) = x(t+\tau) - x(t)$ the returns over a time window τ, as usual; then the *local* Hölder exponent at the instant t is defined as follows[15] [9, 21]:

$$\alpha(t) = \lim_{\tau \to 0} \frac{\ln |r_t(\tau)|}{\ln \tau} \tag{2.18}$$

Some remarks are in order at this point.

- Empirically one cannot achieve the limit in eq. (2.18) since a discrete time series has necessarily a minimum time window: in dealing with our financial indexes, for instance, one cannot take a time window less than one day. However, since we are interested in the relationship between the Hölder exponent and the Hurst one, we will consider the local Hölder exponent from a theoretical viewpoint only and do not attempt to calculate it. In particular we will see that eq. (2.14) does not hold anymore if α is not a constant.

- We are considering processes with stationary increments, as in the previous sections. The local Hölder exponent is defined on a single realization, while

[15] sometime this limit is regarded as a logarithmic derivative [21]

stationarity concerns the density function $p_{t,\tau}$ of the increment $r_t(\tau)$: if such PDF does not depend on time then the process has stationary increments. In other words the limit in eq. (2.18) depend on the particular realization we are considering: if, for certain instant t, one gets a certain $\alpha(t)$ in the present realization, then for another realization, at the same instant, one may find a different result for $\alpha(t)$

It is worth noting that $\alpha(t)$ retains the meaning that the Hölder exponent has for self-affine processes. According to the interpretation we gave of α in section (2.3), if we rescale a small part of a self-affine time series we statistically get a graph equal to the whole time series, while the exponent α links together the rescaling factors on the time axis and on the vertical one. Looking at eq. (2.18) we recover the same interpretation if such rescaling takes place on a very small interval around a fixed instant t.

In order to compare eq. (2.18) with the previous one (eq. 2.10) it is worth showing their equivalence for a stationary self-affine process with Hölder exponent α_0. Using self-affinity we have

$$\frac{\ln|r_t(\tau)|}{\ln\tau} \sim \alpha_0 + \frac{\ln|r(1)|}{\ln\tau}$$

where $\sim$, as usual, means 'equally distributed'. Next, using the Chebyshev inequality [12] we find the probability for the rate in the righthand side to exceed any fixed positive and very small constant ε:

$$\begin{aligned} \forall\varepsilon \quad \mathcal{P}\left(\left|\frac{\ln|r(1)|}{\ln\tau}\right| \geq \varepsilon\right) &\leq \frac{1}{(\varepsilon\ln\tau)^2}\int_{-\infty}^{\infty} u^2 p_{\ln|r(1)|}(u)\,du \\ &= \frac{1}{(\varepsilon\ln\tau)^2}\int_{-\infty}^{\infty} (\ln|u|)^2 g(u)\,du \propto (\varepsilon\ln\tau)^{-2} \end{aligned}$$

where g is the PDF of the returns $r(1)$. Hence, when $\tau \to 0$, this probability vanishes for every ε implying that

$$\forall\varepsilon \quad \lim_{\tau\to 0}\mathcal{P}\left(\left|\frac{\ln|r_t(\tau)|}{\ln\tau} - \alpha_0\right| \geq \varepsilon\right) = 0$$

This in turn means that for every self-affine process:

$$\alpha(t) = \alpha_0$$

In section (2.3) we defined a simple scaling process and found that a self-affine process with Hölder exponent α_0 is simple scaling, i.e. it has a linear scaling exponent $D(q)$ with slope $H = \alpha_0$. Since we are interested just in the Hurst exponent H we

wonder the behaviour of $D(q)$ when a time series has a time varying Hölder exponent $\alpha(t)$.

Let $x(t) \quad 0 \leq t \leq T$ be the realization of a continuous process with fixed size T. Then divide the interval $[0;T]$ into m parts by means of m points $t_j \quad j = 1 \dots m$, thus obtaining a mesh of m subintervals $\{t_j; t_{j+1}\}$ of size $\tau = \frac{T}{m}$. Inside each subinterval we have a return $r_j(\tau) = x(t_{j+1}) - x(t_j)$ over a time window τ. Now let $m_\tau(\alpha, \varepsilon)$ be the number of instants t_j for which the return inside the corresponding subinterval satisfy:

$$\tau^{\alpha+\varepsilon} \leq |r_j(\tau)| \leq \tau^{\alpha-\varepsilon}$$

Provided the following double limit exists, we can define [9] the *multifractal spectrum* $S(\alpha)$ of our continuous process $x(t)$ as

$$S(\alpha) = \lim_{\tau \to 0\,,\, \varepsilon \to 0} -\frac{\ln m_\tau(\alpha, \varepsilon)}{\ln \tau} \tag{2.19}$$

We are going to show that this spectrum allows us to find the scaling exponent $D(q)$ and thus the Hurst exponent, but first some remarks are in order. The expression in eq. (2.19) may seem rather obscure, it has a straightforward physical interpretation instead [21]. Take the subset $E(\alpha) \in [0;T]$ containing the instant t whose local Hölder exponent is $\alpha(t) = \alpha$; then $S(\alpha)$ is just the Box-dimension of the set $E(\alpha)$.

In fractal theory the Box-dimension of a geometrical set is often used in place of the Hausdorff dimension D_H since it is much more easy to calculate, both from an empirical and a theoretical point of view, see [9]. Furthermore in many cases these two dimensions turn out to be equal[16]. Hence the spectrum $S(\alpha)$ provides an estimation of the size of the set $E(\alpha)$: the larger the spectrum the more numerous the instant t at which $\alpha(t) = \alpha$. In other words $S(\alpha)$ tells us how frequently $\alpha(t) = \alpha$ along our time series. In particular if $S(\alpha) < 1$ the set $E(\alpha)$ has a vanishing lenght and thus the instants at which $\alpha(t) = \alpha$ occur very rarely. However, as we are going to see, even such a rare α can affect the scaling exponent $D(q)$.

The following theorem [9] is a first step in order to relate the multifractal spectrum to the scaling exponent.

Theorem 3. *If $S(\alpha)$ exists, then*

$$\sup_\alpha [S(\alpha) - \alpha q] = 1 - \lim_{\tau \to 0} \frac{\ln M_q^{(e)}(\tau)}{\ln \tau} \tag{2.20}$$

Recall that $M_q^{(e)}(\tau)$, as defined in eq. (2.9), are the empirical moments of the PDF of the returns $r(\tau)$ collected along the time series. We remark that for the

[16] for the sake of precision it can be shown that the Hausdorff dimension $D_H(E)$ of any set E is never larger than the corresponding Box-dimension

validity of this theorem, as for the validity of all the results of this section, the existence of the moments is needed.

It is worth noting that for a self-affine process with Hölder exponent α_0 we get a pointwise spectrum. Since we have $M_q^{(e)}(\tau) \propto \tau^{\alpha_0 q}$, from eq. (2.12), then we get:

$$\forall q \in \mathbb{R} \quad \sup_{\alpha}[S(\alpha) - \alpha q] = 1 - \alpha_0 q \qquad \Longrightarrow \qquad S(\alpha) = \begin{cases} 1 & \text{if } \alpha = \alpha_0 \\ -\infty & \text{otherwise} \end{cases}$$

According to the previous interpretation of $S(\alpha)$ as the Box-dimension of the set $E(\alpha)$, this means that $\alpha(t) = \alpha_0$ always, at all instant. Therefore, as expected, we have a single Hölder exponent α_0 and the set of instants for which $\alpha(t) = \alpha_0$ has Box-dimension $S = 1$, namely the maximum possible[17]. If the spectrum has a support wider than a single point, then there are many exponents in our time series: the broader the spectrum the wider the range of exponents available. In this case we have *multifractality*.

The link between $S(\alpha)$ and the scaling exponent $D(q)$ comes at once from the definition of $D(q)$ in eq. (2.15):

$$M_q^{(e)}(\tau) \propto \tau^{D(q)}$$

Substituting the above expression in eq. (2.20) one finds that the scaling exponent and the multifractal spectrum are related by a Legendre transform:

$$\sup_{\alpha}[S(\alpha) - \alpha q] = 1 - D(q) \tag{2.21}$$

Then, once the spectrum $S(\alpha)$ is given, one can calculate $D(q)$.

For instance a pointwise spectrum with support on α_0, as that of a self-affine process, implies

$$D(q) = \alpha_0 q$$

namely a linear scaling function, which is distinctive of simple scaling. On the contrary a broader spectrum gives a non linear scaling function $D(q)$, and thus a non costant Hurst exponent $H(q)$. In this case we have multiscaling:

Definition 5. *A stationary time series is multiscaling if:*

$$M_q^{(e)}(\tau) \propto \tau^{D(q)} \tag{2.22}$$

where the scaling exponent $D(q)$ is not linear.

[17] this is because the interval $[0; T]$ containing all the sets $E(\alpha)$ has just this dimension. On the other hand it is easy to see from the definition (eq. 2.19) that $S(\alpha) \leq 1$

A comment is in order at this stage. We saw in section (2.3) that for a strictly self-affine process with Hölder exponent α_0 the invariance under rescaling holds, or better the PDFs p_τ of the returns $r(\tau)$ satisfy eq. (2.11). Now one may wonder about the validity of such invariance for multiscaling time series. Surely eq. (2.11) does not hold for a multifractal process as it implies simple scaling at once. However, according to the interpretation of the multifractal spectrum $S(\alpha)$ as the dimension of the set $E(\alpha)$, it turns out that the invariance under rescaling may still hold approximately.

Suppose $S(\alpha)$ to have a single maximum in α_*, namely: $S(\alpha_*) = 1$ and $S < 1$ elsewhere. Then such an exponent leads the whole time series as any other exponent $\alpha \neq \alpha_*$ has a set $E(\alpha)$ of vanishing lenght. In other words, even if the spectrum $S(\alpha)$ is broad, it may happen that a single exponent α_* is much more likely than all others, so that it leads the observable behaviour of our time series. In conclusion, even for a multiscaling time series, one may empirically find an approximate scaling invariance with Hölder exponent α_*:

$$p_\tau(u) \approx \frac{1}{\tau^{\alpha_*}} g(\frac{u}{\tau^{\alpha_*}})$$

While the estimation of an Hölder exponent may be quite hard to beat, due to the limit in eq. (2.18), the evaluation of $D(q)$ is much simpler. Hence, in order to get α_*, we use the scaling exponent $D(q)$. For each q, let $S(\alpha) - \alpha q$ to have a maximum in $\alpha(q)$, so that

$$\frac{\partial}{\partial \alpha} S[\alpha(q)] = q$$

Then, using eq. (2.21) we get $S[\alpha(q)] - q\alpha(q) = 1 - D(q)$ and, by derivation on both side:

$$\begin{aligned} \frac{\partial}{\partial q} D(q) &= q \frac{\partial}{\partial q} \alpha(q) + \alpha(q) - \frac{\partial}{\partial q} \alpha(q) \frac{\partial}{\partial \alpha} S[\alpha(q)] \\ &= \alpha(q) \end{aligned} \tag{2.23}$$

Now consider $q = 0$. Since[18] $D(0) = 0$, from eq. (2.21) we find that:

$$\sup_\alpha S(\alpha) = 1 \quad \Rightarrow \quad S[\alpha(0)] = 1$$

If the spectrum S has a single maximum, then $\alpha(0) = \alpha_*$ and using eq. (2.23):

$$\alpha_* \approx \frac{\partial}{\partial q} D(0) \tag{2.24}$$

[18] indeed, according to eq. (2.15) and (2.9), for $q = 0$ we get $M_0^{(e)}(\tau) = 1$ for each window τ, and thus $D(0) = 0$

2.6 Hurst exponent estimation for some financial indexes

Many real time series coming from several fields of the physics of the complex systems display a non linear scaling exponent [62, 66], and thus are multiscaling, according to eq. (2.22). In finance the non linearity of the scaling exponent is a common feature to a large number of time series [57, 58, 59]. In this section we are going to estimate the scaling exponent $D(q)$ for the time series of the financial indexes we analized in chapter (1) by performing the procedure outlined in section (2.3).

As in the case of the Brownian motion and of the fractional Brownian motion of section (2.4), in order to test the robustness of the outcoming scaling exponent $D(q)$ we use a bootstrap method: the generalized Hurst exponent analysis has been performed on the whole time series and on many uniformly extracted samples. This results in a bundle of curves $D(q)$ for each index. In fig. (2.6 - 2.9) the red curves are the scaling exponents estimated, for each index, on the available time series, while the black ones are the scaling exponent estimated on many samples extracted from the whole series itself.

Since the returns of our financial time series are strongly correlated, the extracted subsample should not overlap too much otherwise they cannot regarded as distincts: their separation should be greater than the decorrelation time, namely the time interval needed to regard two returns as independent, see section (1.2). However, the reliability of the bootstrap estimator is as greater as the extracted samples are longer and more numerous. Hence, the size of the samples we have drawn from the whole series has been set equal to $N = 5000$; moreover in the extraction procedure very overlapping subsamples have been avoided.

The first feature one can note is the common global pattern of all the graphs. Taking into account the scaling exponent related to the whole time series, namely the red thick curves, one can see that:

- for low orders the shape of $D(q)$ is nearly linear with slope $H \approx \frac{1}{2}$;
- as the order increases all the scaling exponents start to bend and their behaviour is no longer linear.

In section (1.3) we have seen that our financial time series are invariant under rescaling and so that the PDFs of the returns over different time windows τ, if properly rescaled according to τ^{α}, collapse in a singlescaling function. In that section we estimated empirically the value of α by attempting to obtain the best collapse possible. According to eq. (2.24), the slope of the scaling exponent $D(q)$ near the origin $q = 0$ accounts for the estimation $\alpha \approx 0.5$ we have given in eq. (1.6).

The global behaviour of $D(q)$ is not linear. It is worth noting that some indexes, like the DJI and specially the S&P, display a quite strong downward bending, while the others diverge slightly from the line $D(q) = \frac{1}{2}q$. This behaviour is generally accepted as multiscaling.

However, in taking into account also the bootstrap samples for each index, namely the black curves, one can easily distinguish two regions:

- for q lower than a certain order q_* the curves $D(q)$ related to the samples extracted from the whole time series are all close together, thus suggesting that the outcoming scaling exponent is not sample dependent;
- on the contrary, for $q > q_*$, the behaviour of $D(q)$ become very erratic and strongly sample dependent and accordingly its uncertainty becomes very high;
- for $q > q_*$ there is also a kind of clustering among the curves $D(q)$ estimated from different extracted samples. Their spreading does not span a continous area: the behaviour of $D(q)$ is nearly the same for some extracted samples and then changes abruptly.

One can approximately estimate

$$q_* \sim 3$$

for all the indexes considered.

In order to understand the erratic behaviour of the scaling exponents $D(q)$ it is worth going back to the case of the BM and of the FBM. In section (2.4) we performed the generalized Hurst exponent analysis on a single history and get the scalind exponent $D(q)$ for two kinds of exactly solvable models: the Brownin motion and the fractional Brownian motion. We saw that the outcoming $D(q)$ was a line with the expected slope H. Furthermore the uncertainty, estimated with a bootstrap method, was much lower than in the present case: the scaling exponent did not depend on the extracted sample.

One of the main differences between our financial indexes and the BM, or the FBM, is the asymptotic behaviour of the density function of their increments. While the BM and the FBM have tails that decay faster than exponentially, the return PDFs of our indexes are fat-tailed, see section (1.4). More precisely, for large absolute arguments, the scaling function $g(u)$ follows a power law:

$$g(u) \approx \frac{1}{u^\gamma} \quad \text{if} \quad |u| \gg 1$$

see eq. (1.7). This means that the theoretical q-order moments $M_q(\tau)$, as defined in eq. (2.8), do not exist for $q \geq q_0$; where $q_0 = \gamma - 1$. We estimated $\gamma \approx 4$, and thus

$$q_* \approx q_0$$

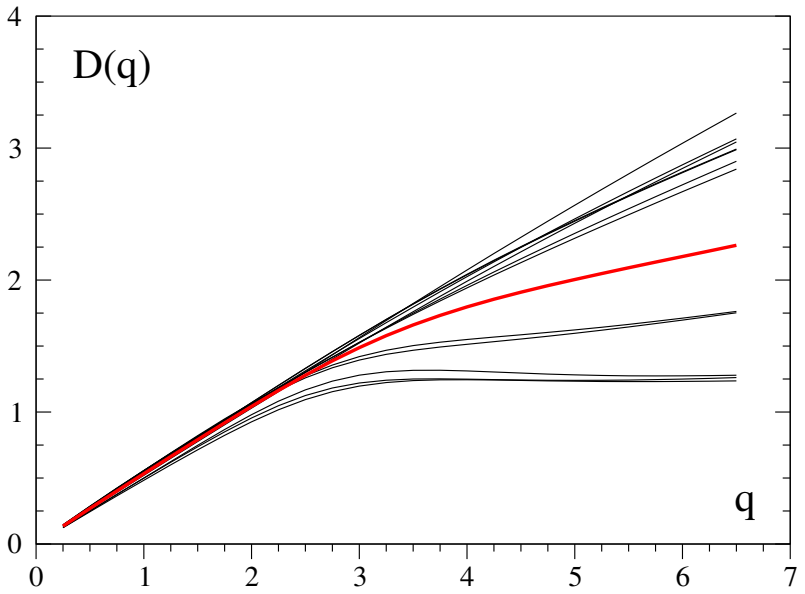

Figure 2.6: Scaling function $D(q)$ for several samples (N=5000,*black curves*) extracted from the whole time series (*red curve*) of the Dow Jones Industrial (**DJI**) index.

This means that the order at which the scaling exponent $D(q)$ becomes very sample dependent agrees with the threshold order q_0 beyond which the moments are expected to diverge. There is no any reason for such a coincidence. We believe it is a clue suggesting that the non-existence of a limit for the sum in eq. (2.9) makes the determination of the scaling exponent very sample dependent. This is the first evidence suggesting that the scaling exponent $D(q)$ resulting from the generalized Hurst exponent analysis becomes unreliable inside the range of the orders for which the corresponding moment does not exist. In the next chapter we will see another one, even more persuasive, with deep consequences on many multiscaling claimed until now in finance.

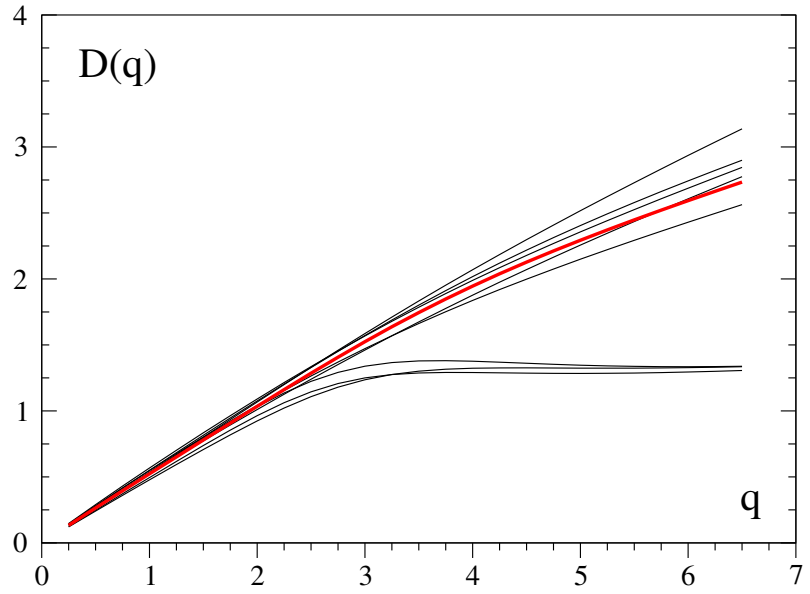

Figure 2.7: Scaling function $D(q)$ for several samples (N=5000,*black curves*) extracted from the whole time series (*red curve*) of the Standard & Poor's Composite (**SPC**) index.

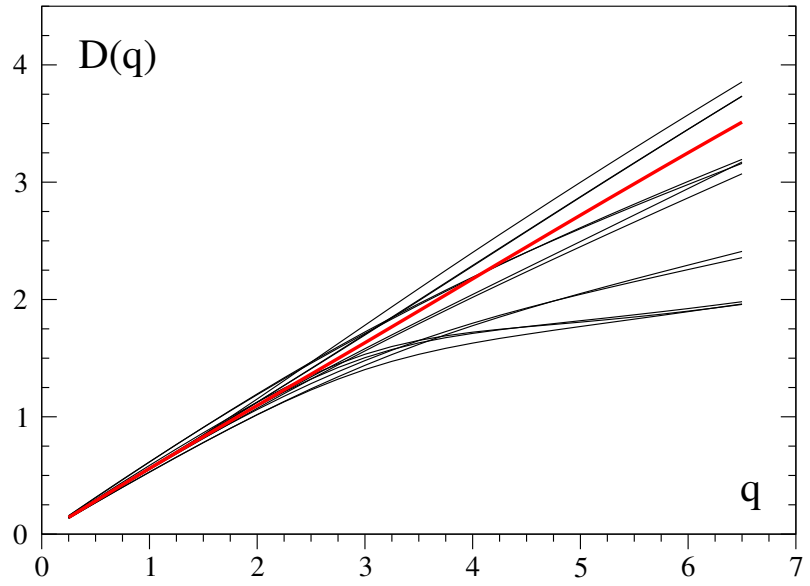

Figure 2.8: Scaling function $D(q)$ for several samples (N=5000,*black curves*) extracted from the whole time series (*red curve*) of the Dow Jones Transportation (**DJT**) index.

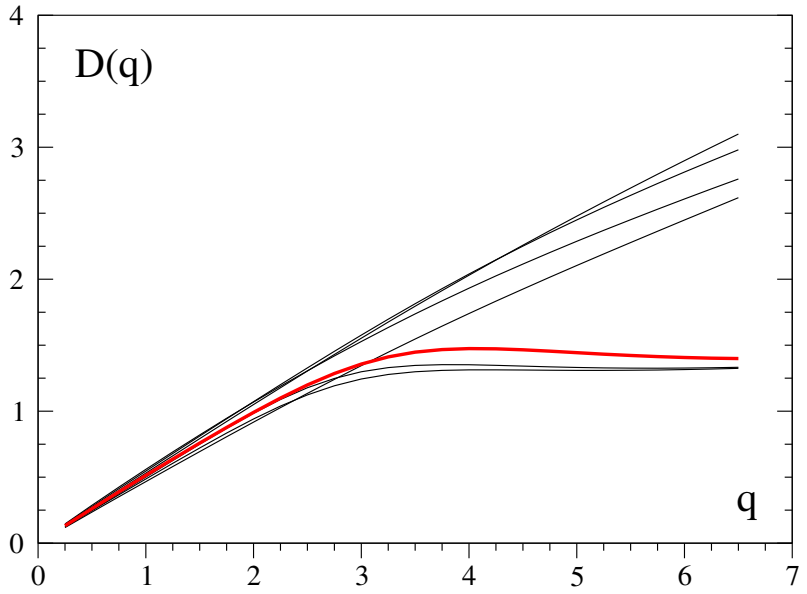

Figure 2.9: Scaling function $D(q)$ for several samples (N=5000,*black curves*) extracted from the whole time series (*red curve*) of the Standard & Poor's (**S&P**) index.

Chapter 3

Hurst exponent reliability

In the previous chapter the Hurst exponent analysis led correctly and with great accuracy to simple scaling when applied to simple scaling processes. On the contrary it led to multiscaling when performed on financial indexes. Can we claim such a multiscaling to be a reliable result as in the examples of section (2.3)? There are two main differences:

1. the statistical uncertainty associated to the scaling function for high orders is very important, much larger than in the case of BM and FBM, see section (2.6);

2. while the increments of a BM or a FBM are distributed according to a Gauss' law, those of our financial indexes follow a density function whose tails decay as a power law, see section (1.4).

In particular the second point implies that the theoretical moments no longer exist for large orders and therefore the scaling exponent becomes ill-defined, as it is based on moments' existence (eq. 2.13). We wonder whether the Hurst exponent still leads to a correct picture when large returns follow a Pareto tail.

3.1 Lorentzian Random Walks

To test the reliability and the meaning of the Hurst exponent when large increments are distributed according to a power law we need an exactly soluble model displaying this feature. In other words we need to calculate the theoretical scaling exponent $D(q)$ of our model and then to compare it with that coming from a simulation. From a mathematical viewpoint a random walk with independent identically distributed increments would make calculations easy to perform, but the density function g of these increments should be stable and fat-tailed.

A density function $g(u)$ is *stable* if the sum $y = \sum_{i=1}^{\tau} x_i$ of many stochastic variables x_i identically distributed according to $g(u)$ follows again the same rescaled PDF:

$$p_y(u) = \frac{1}{\tau^\alpha} g(\frac{u}{\tau^\alpha}) \tag{3.1}$$

If the variables x_i are the increments $r_{t+i}(1) \quad i = 1 \dots \tau$ then their sum y is the return $r_t(\tau)$ over a time window τ. Therefore the stability of g guarantees the self affinity of our random walk and the exponent α in eq. (3.1) is just the Hölder exponent.

We know the Gauss' law to be stable, see section (2.3), but its tails decay exponentially and all its moments exist, so we must consider other stable density functions, namely the so called Levy stable law. Since we are dealing with the sum of independent increments, in order to find the class of the stable laws it is convenient to look at the CFs. Consider the characteristic function $\psi(k)$ of the stable law $g(u)$. Due to the main property of CFs (eq. 2.7), the CF φ_y of the sum y is written as

$$\varphi_y(k) = \psi^\tau(k)$$

Since the stability condition (eq. 3.1) is just a scaling relation among PDFs, then we can use the scaling rule for CFs

$$\varphi_y(k) = \psi(\tau^\alpha k)$$

to get

$$\psi^\tau(k) = \psi(\tau^\alpha k)$$

Taking logarithms, one obtains

$$\tau \ln \psi(k) = \ln \psi(\tau^\alpha k)$$

Here the time window τ is an integer as the summation can involve only an integer number of addends, but one can show that the above equation also holds for every real τ. Hence, the class of the CFs whose density is stable[1] can be written as:

$$\psi(k) = \exp(-|\sigma k|^{\frac{1}{\alpha}}) \tag{3.2}$$

where $\sigma > 0$ is a width parameter[2]. In order $\varphi(k)$ to be a CF, i.e. in order the corresponding PDF to be positive definite α must satisfy [8]: $\alpha \geq \frac{1}{2}$. Then for α

[1] and symmetric

[2] this parameter is positive because the modulus of a CF must be less than 1. Indeed recalling the very definition of CF in eq. (2.3) we find that

$$\forall k \in \mathbb{R} \quad |\varphi(k)| \leq 1$$

since $|e^{iku}| = 1$

ranging inside this interval $[\frac{1}{2};\infty]$, each CF in eq. (3.2) corresponds to a given Levy stable law, but finding its analytical form is a more difficult task. For $\alpha = \frac{1}{2}$ we recognize the CF of a Gauss' law, then the normal distribution belongs to the class of the Levy stable laws and the corresponding independent random walk is just the brownian motion of section (2.3).

For general values of $\alpha > \frac{1}{2}$ the exact form of the related stable PDF $L_\alpha(u)$ is not known, but one can show that asymptotically the tails of such PDF decay as a power law:

$$L_\alpha(u) \approx \frac{c}{u^{1+\frac{1}{\alpha}}}$$

This means that, if $\alpha \neq \frac{1}{2}$, the theoretical q-order moments of L_α exist for $q < \frac{1}{\alpha}$ only. In particular the Gauss' law is the only element of the class for which the variance exist[3], moreover it is the only element whose tails decay exponentially. Since we were looking for a stable PDF with power law tails to build our random walk, we can draw out the increments from any density $L_\alpha(u)$ save the Gauss' law.

For what follows we will need the analytical form of L_α. Fortunately for $\alpha = 1$ the CF $\varphi_1(k) = e^{-\sigma|k|}$ is simple enough to make the inverse Fourier transform exactly soluble. Hence

$$\begin{aligned}
L_1(u) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \varphi_1(k) e^{-iku}\, dk \\
&= \frac{1}{2\pi}\int_{0}^{\infty} \left[e^{-(\sigma+iu)k} + e^{-(\sigma-iu)k}\right]\, dk \\
&= -\frac{1}{2\pi}\left[\frac{e^{-(\sigma+iu)k}}{\sigma+iu} + \frac{e^{-(\sigma-iu)k}}{\sigma-iu}\right]_0^{\infty} \\
&= \frac{1}{2\pi}\left(\frac{1}{\sigma+iu} + \frac{1}{\sigma-iu}\right) \\
&= \frac{\sigma}{\pi}\frac{1}{\sigma^2+u^2}
\end{aligned}$$

This PDF is the Lorentz distribution and will be denoted simply with

$$L(u) = \frac{\sigma}{\pi}\frac{1}{\sigma^2+u^2} \tag{3.3}$$

thus dropping the subscript 1.

To conclude we perform a stationary random walk by extracting the independent increments $r(1)$ from a symmetric Lorentz distribution with zero mean and width parameter σ and summing them up to get a *Lorentzian Random Walk* (LRW). Since

[3]If the variance exist then the Central Limit Theorem would state that the sum of many independent variables distributed according to L_α follows a Gauss' law, therefore L_α could not be stable.

the process is stationary, as in section (2.2) we denote the increments $r_t(\tau)$ simply with $r(\tau)$, thus dropping the subscript t. The notation with the subscript will be used only when we will need to take into account these increments individually, like inside a summation. It is worth stressing some very important points.

- Since L does not depend on the instant t this random walk is stationary.
- Since this L is stable, the returns $r(\tau)$ over a time window τ are exactly self affine with Hölder exponent $\alpha = 1$:

  $$r(\tau) \sim \tau r(1)$$

  Hence

  $$p_\tau(u) = \frac{1}{\tau} g(\frac{u}{\tau}) \tag{3.4}$$

  where p_τ is the PDF of the returns $r_t(\tau)$, as usual, and $g = L$ is just the scaling PDF.
- Since the tails of a Lorentz distribution asymptotically follow a power law:

  $$L(u) \approx \frac{\sigma}{\pi u^2} \quad \text{when} \quad u \gg \sigma$$

  and thus the theoretical q-order moment M_q of L exists for $q < 1$ only.

Since the return PDFs p_τ are all equal up to a scale factor τ, we could forecast $D(q)$ to be a straight line with slope $H = 1$, as in the case of the Gaussian Random Walk, no matter how large the order q is, provided that the sample size N is large enough[4]:

$$D(q) = q \qquad \forall\, q \geq 0$$

Surely the above equation is true for $q < 1$, when theoretical moments $M_q(\tau)$ exist, see eq. (2.8). For $q \geq 1$ the theoretical moments diverge, but for a given sample of size N we empirically get a finite value $M_q^{(e)}(\tau)$, see eq. (2.9). However, even for $N \to \infty$, this value cannot converge to anything as the limit $M_q(\tau)$ does not exist, so we ask whether the scaling exponent $D(q)$ should be still linear.

In order to find the behaviour of $D(q)$, according to the procedure outlined in section (2.3), we are going to calculate the logarithm of the empirical moments $M_q^{(e)}(\tau)$. We can regard $\ln M_q^{(e)}(\tau)$ as a stochastic variable depending on a set of N returns $r(\tau)$. Then we prove the following theorem, which tells us that for a simple scaling the expected value of the scaling exponent $D(q)$ is linear no matter whether the theoretical moment exist or not.

[4]for an increasing sample size N the sum in eq. (2.9) approaches the theoretical value (eq. 2.8) better and better provided that the latter exists

Theorem 4. *For every self affine process with Hölder exponent α:*

$$\langle \ln M_q^{(e)}(\tau) \rangle \;=\; \langle \ln M_q^{(e)} \rangle \;+\; \alpha q \,\ln \tau \tag{3.5}$$

where $M_q^{(e)}$ is the empirical q-order moment of the scaling PDF g.

Proof:
Let us calculate the density function $P_q^{(\tau)}(u)$ of the logarithm of the empirical moments in term of the density p_τ of $r(\tau)$. The stochastic variables $|r_i(\tau)|^q$ are distributed according to[5]

$$|r_i(\tau)|^q \sim \frac{2}{q}\,\frac{p_\tau(u^{1/q})}{u^{1-1/q}}$$

hence their characteristic function is

$$\varphi_q^{(\tau)}(k) = \frac{2}{q}\int_0^\infty du\, \exp iku^q p_\tau(u)$$

Using the self-similarity of p_τ, see eq. (3.4), one easily gets[6]:

$$\varphi_q^{(\tau)}(k) = \varphi_1^{(q)}(\tau^{\alpha q}k)$$

Since $|r_\tau(i)|^q$ are independent, the characteristic function of their sum is the product of all characteristic functions, so the CF of $M_q^{(e)}(\tau)$ is:

$$\left[\varphi_q^{(\tau)}(\frac{k}{N})\right]^N = \left[\varphi_q^{(1)}(\tau^{\alpha q}\,\frac{k}{N})\right]^N$$

This means that the PDFs of $M_q^{(e)}(\tau)$ also are self-similar:

$$M_q^{(e)}(\tau) \sim \tau^{q\alpha} M_q^{(e)}$$

This in turn implies[7] that the PDFs $P_q^{(\tau)}$ of the logarithm of q-order moments are shifted by a quantity $\alpha q \,\ln\tau$:

$$\mathcal{P}_q^{(\tau)}(u) = \mathcal{P}_q^{(1)}(u - q\alpha \ln\tau)$$

so the result follows.

---

[5] if x is a stochastic variable with PDF p_x then the variable $y = x^q$ has PDF $p_y(u) = \frac{1}{q}\frac{p_x(u^{1/q})}{u^{1-1/q}}$. The factor 2 in front comes from the symmetry of $L_\lambda(u)$ for the transformation $u \to -u$

[6] for the sake of generality, since the theorem is valid for every α, in what follows we left the Hölder exponent unspecified, but it should be borne in mind that $\alpha = 1$ in the present case

[7] if the stochastic variable x has distribution p_x then $y = \ln x$ is distributed according to $p_y(u) = e^u\, p(e^u)$

□

In our case, a Lorentzian random walk, the theorem above states a linear average scaling exponent with slope $\alpha = 1$. Indeed, consider a large amount N of independent returns coming from a Lorentz law L. For $q < 1$ the theoretical moment exists and, due to the Law of Large Numbers, $M_q^{(e)}(\tau)$ are very close to $M_q(\tau)$, at least for N large enough. Therefore $P_q^{(\tau)}$ are very close to Dirac δ functions centered in $\ln M_q(\tau)$ and eq. (3.5) simply becomes:

$$\ln M_q(\tau) = \ln M_q(1) + D(q) \ln \tau \quad \text{with} \quad D(q) = q$$

as expected.

Suppose now $q \geq 1$. In this case $\lim_{N\to\infty} \langle \ln M_q^e(\tau) \rangle = \infty$ and $P_q^{(\tau)}$ cannot converge to a Dirac δ. Actually it cannot converge to any density function. This means that for $q \geq 1$ the empirical scaling exponent $D(q)$ may display important deviations from its average behaviour, even for a very large N. Anyway eq. (3.5) always holds, no matter how large is the order q. Therefore, once N is given, the dependence on τ of $P_q^{(\tau)}$ still implies that the empirical curve $D(q)$ should follow a straight line with slope $\alpha = 1$ if one consider statistical averages of $\ln M_q^{(e)}(\tau)$. In conclusion a self-affine process with independent increments has a linear scaling exponent $D(q)$ no matter whether the theoretical moments exist or not[8].

In order to verify empirically theorem (4) we

- fix the order q and generate a set of N returns $r(\tau)$ for each time window τ;
- calculate the corresponding moments $M_q^{(e)}(\tau)$ by means of eq. (2.9);
- find the scaling exponent $D(q)$ by means of eq. (2.15);
- repeat the whole procedure many times for each order q, thus getting a set of exponents $D(q)$.

Fig. (3.1) shows the theoretical line $D(q) = q$ and the scaling exponents calculated in this way together with their average at each q. For each order q we calculated 70 exponent $D(q)$, using the procedure just outlined above. When the theoretical moment exists, for $q < 1$, we can see that the scaling exponents $D(q)$ follow a straight line with slope $\alpha = 1$, while for $q \geq 1$ they start to behave very chaotically, although the size $N = 25000$ is quite large. However, their average follows the theoretical line almost perfectly, according to theorem (4). We stress that for every

[8] at least in a probabilistic sense. In deriving eq. (3.5) we made not any hypothesis on the existence of moments, but an exact equation is only possible when they are finite. However the key point here is that the scaling exponent $D(q)$ is expected to grow with the order q

order q and window τ the variables $\ln M_q^e(\tau)$ are calculated by means of different sets of returns, each of size N.

Since in many practical cases, like finance, one cannot restart the process to get returns from different histories, one may worry about the validity of the theorem (4) when a single time series is only available. Hence we generate a single time series of size N and calculate the scaling exponent by means of the returns collected on such history using a sliding window method.

A simulation of a lorentzian random walk is shown in fig. (3.2); as before in the case of the gaussian random walk $N = 25000$. When the generalized Hurst exponent calculation is performed on the simulated series and on many randomly extracted samples of size $N = 5000$ we get a somehow unexpected result: within a small error, all curves strongly bend downward at $q = 1$. It is very important to stress that now we collect returns from a single history, as we have to do for any financial index.

For smaller orders the slope is $H = 1$, but beyond $q = 1$ the Hurst exponent vanishes and the curves $D(q)$ become horizontal, see fig. (3.3). We stress that all the curves $D(q)$ become suddenly horizontal when $q = 1$, i.e. at the threshold order for diverging moments:

$$D(q) = \begin{cases} q & \text{if } q < 1 \\ 1 & \text{if } q \geq 1 \end{cases} \tag{3.6}$$

Furthermore such behaviour does not depend on the choice of the sample. However we just saw that the average of $D(q)$ should follow a straight line with slope $H = 1$, so our scaling exponent should statistically grow with q. Why does the Hurst exponent fail to show the correct simple scaling if estimated on the basis of a single history?

The bending of $D(q)$ is generally accepted as multiscaling, or bifractality, but in the next section we try to explain the behaviour of $D(q)$ otherwise. We claim that the process is really simple scaling, but when returns are collected along the same time series a very simple mechanism arising from power law tails bends the curve $D(q)$ and leads to a strong spurious multiscaling.

3.2 Pareto tails as generators of apparent multiscaling

Consider again a single Lorentzian random walk of size N and suppose the empirical sum $M_q^{(e)}(\tau)$ of eq. (2.9) can be well approximated by an integral like that in eq. (2.8), at least when N is large. For an order $q \geq 1$ this integral does not exist, but the sum does because the sample size is finite.

Hence, we stop the upper limit of integration at some finite value R. This could be a reasonable request: R can be seen as the largest return in our sample, see also

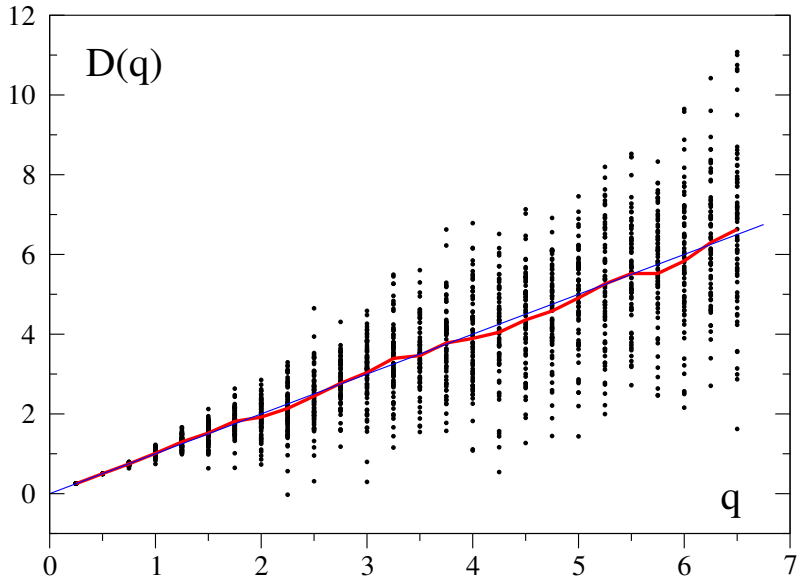

Figure 3.1: Scaling exponents of the empirical moments calculated on a set of returns of size $N = 25000$ (*black points*). For each order q we obtained 70 exponents $D(q)$. The *red* curve show the average scaling exponent, while the *blue* line is the theoretical expectation.

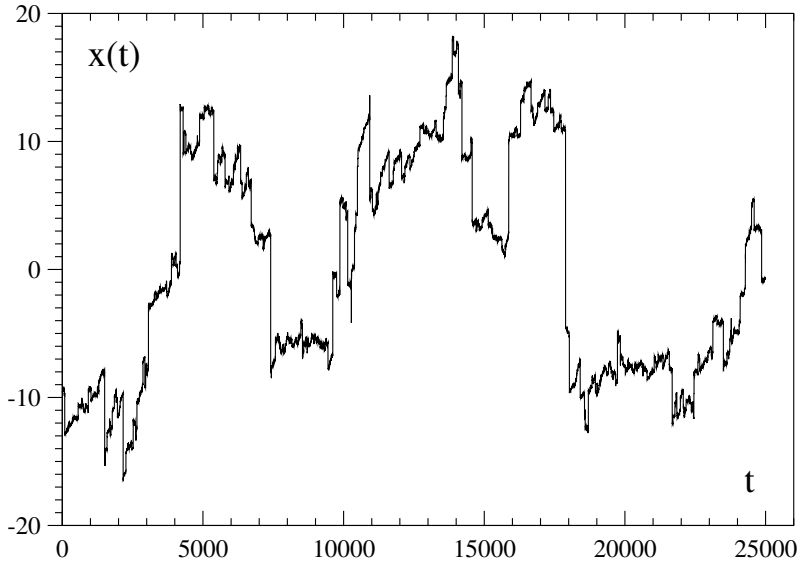

Figure 3.2: Simulation of a Lorentzian random walk of size $N = 25000$. The independent increments were drawn out from a zero mean Lorentz's distribution with width parameter $\gamma = 0.00614$. This parameter, see eq. (3.3) was set in order to fit the width of the PDF of the daily returns of the DJI index, see table (1.3).

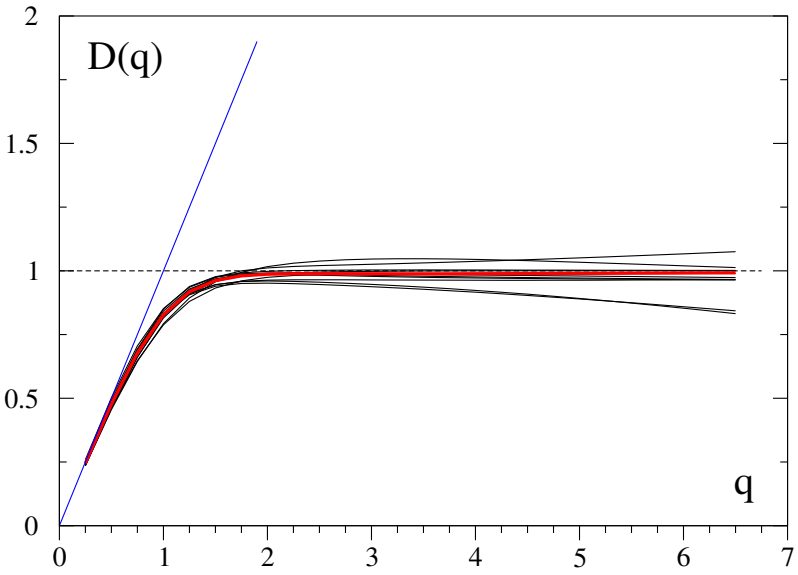

Figure 3.3: Scaling exponents $D(q)$ for several samples ($N = 5000$, *black curves*) extracted from the whole series (*red curve*) plotted in fig. (3.2). The theoretical line $D(q) = q$ is drawn in *blue*.

[5]. Actually R should be written as $R(\tau)$ because it depends on the time window[9] τ of the returns $r(\tau)$. Indeed we are collecting returns using a sliding window from a single history of size N; this means that for every τ we get a set of N returns; so, for every τ there is an upper bound to the corresponding set of returns.

In particular, one would expect $R(\tau)$ to be an incresing function of the time window τ. Consider for instance a Brownian motion: we know, from its definition in section (2.4), that the variance of $p(\tau)$ grows linearly with τ. Hence, for a given time series of fixed size N we expect the largest return among $r_t(\tau) \quad t = 1 \dots N$ to grow as $\sqrt{\tau}$. For a fractional Brownian motion of index α, see section (2.4) again, one expect $R(\tau) \propto \tau^\alpha$. Moreover, for a general self-affine process with a given Hölder exponent α, we would expect, on the basis of eq. (2.10), that

$$R(\tau) \propto \tau^\alpha$$

On the contrary our key theorem states that for a LRW $R(\tau)$ is costant.

Theorem 5. *Consider a single Lorentzian random walk, namely a single time series with independent increments distributed according to a Lorentz law L. If all the returns are collected along this history by means of a sliding window, then*

$$R(\tau) \approx R$$

[9] for example, in the case of a gaussian random walk one must take $R(\tau) = \sqrt{\tau} R(1)$ in order for the integral $\int_{|u|<R(\tau)} |u|^q p_\tau(u)\, du$ to behave like the empirical sum (eq. 2.9)

i.e. $R(\tau)$ does not depend on the window τ.

Surely R grows if the sample size N is increased, since with more data available one can explore tails better and better, but for a fixed N it is a constant. In the next section we are going to justify this somehow strange statement[10], but first we show that it can explain the observed behaviour of $D(q)$.

According to the theorem (5), for a large sample size N we assume as a starting point that[11]:

$$M_q^{(e)}(\tau) \simeq 2 \int_0^R u^q p_\tau(u)\, du \tag{3.7}$$

If this is true, take[12] $q = \dfrac{n}{m} \quad m \geq 1$.
Using simple scaling (eq. 3.4) and the substitution $u = w^m$

$$\begin{aligned} M_q^{(e)}(\tau) &= 2\tau^q \int_0^{R/\tau} du\, u^q L(u) \\ &= \frac{2m}{\pi}(\sigma\tau)^q \int_0^{\hat{R}} dw\, \frac{w^{n+m-1}}{1+w^{2m}} \\ &= \frac{2m}{\pi}(\sigma\tau)^q \int_0^{\hat{R}} dw \left(\sum_{i=0}^{n-m-1} a_i\, w^i \;-\; \sum_{i=0}^{2m-1} a_i\, \frac{w^i}{1+w^{2m}} \right) \end{aligned} \tag{3.8}$$

Where we put $\hat{R} = \left(\dfrac{R}{\sigma\tau}\right)^{1/m}$
The last equality (eq. 3.8) comes at once from this algebraic identity:

$$\begin{aligned} u^{n+m-1} &= (1+u^{2m}) \sum_{i=0}^{n-m-1} a_i\, u^i \;-\; \sum_{i=0}^{2m-1} a_i\, u^i \\ a_i &= \begin{cases} (-1)^{j+1} & \text{if } \exists\, j \mid i = n-1-(2j-1)m \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{3.9}$$

Actually, the last sum on the right hand side of eq. (3.8) contains a single term only. Indeed one can prove, using the definition of the coefficients a_i in eq. (3.9), that they all vanish except one in the range $0 \leq i \leq 2m-1$. Let $a_\ell = (-1)^{\ell+1}$ be this non zero coefficient; then, according to eq. (3.9) the integer ℓ must satisfy

$$0 \;\leq\; \phi \;\leq\; 2m-1 \quad \text{with} \quad \phi = n-1-(2\ell-1)m \tag{3.10}$$

[10] moreover we are going to generalize it to a random walk with independent increments asymptotically distributed as a power law

[11] The Lorentz law is symmetric

[12] It's easy to show that $M_q(\tau)$ is a continous function of q, so there isn't any loss of generality in restricting attention to rational orders $q = \frac{n}{m}$

Therefore eq. (3.8) is written as:

$$M_q^{(e)}(\tau) = \frac{2m}{\pi}(\sigma\tau)^q \left[\sum_{i=0}^{n-m-1} \frac{a_i}{i+1} \hat{R}^{i+1} + (-1)^\ell \int_0^{\hat{R}} du \, \frac{u^\phi}{1+u^{2m}} \right] \tag{3.11}$$

Suppose now the range of the time windows to have an upper bound. We stress that this is often the case in many practical cases: for instance, we already saw in section (1.3) that the invariance under rescaling of the time series of our financial indexes does not hold any more for very large windows τ. This is why until now we performed the generalized Hurst exponent analysis inside a bounded range of windows; even in dealing with models for which the invariance under rescaling holds exactly for all the windows τ, like BM and FBM in section (2.4) and the LRW in section (3.1).

Since the largest return in our time series R grows with the size N of our time series and since the windows τ are bounded, we have

$$\frac{R}{\sigma\tau} \gg 1 \tag{3.12}$$

for N very large. This allows us to perform the following approximations.

- If $q < 1$, then $n < m$. This means that only the integral appears into the right hand side of eq. (3.11); moreover $\ell = 0$ is the solution of eq. (3.10), and thus $\phi = n + m - 1$. Then, using eq. (3.12), after some changes of variable we find:

$$\begin{aligned} \frac{2m}{\pi}\sigma^q \int_0^{\hat{R}} \frac{u^{n+m-1}}{1+u^{2m}} du &= \frac{2}{\pi}\sigma^q \int_0^{\hat{R}^m} \frac{u^q}{1+u^2} du \\ &= 2\frac{\sigma}{\pi} \int_0^{\sigma\hat{R}^m} \frac{u^q}{1+u^2} du \\ &\approx 2\int_0^\infty u^q L(u)\, du = M_q \end{aligned}$$

  Therefore, since the q-order moment of L exists for $q < 1$, the integral is very close to a constant which does not depend on τ. Thus:

$$M_q^{(e)}(\tau) \approx M_q \tau^q \propto \tau^q$$

- If $q = 1$, as before, only the integral appears into the right hand side of eq. (3.11) and $\ell = 0$, but now $\phi = 2n - 1$. Using again eq. (3.12) we find

$$\begin{aligned} 2n \int_0^{\hat{R}} \frac{u^{2n-1}}{1+u^{2n}} du &= \ln(1+\hat{R}^{2n}) \\ &\approx 2\ln\frac{R}{\sigma\tau} \approx 2\ln R \end{aligned}$$

Thus:

$$M_q^{(e)}(\tau) \approx \frac{2}{\pi}\sigma \ln R\,\tau \propto \tau$$

- If $q > 1$, take into account the sum which appears into the right hand side of eq. (3.11). Due to eq. (3.12), the term of this sum which has the highest power of $\hat{R}$ overwhelms all other terms and the contribution of the integral as well. On the other hand, it is easy to prove, starting from eq. (3.9), that $a_{n-m-1} = 1$, and thus:

$$\begin{aligned} M_q^{(e)}(\tau) &\approx \frac{2m}{\pi}(\sigma\tau)^q \frac{\hat{R}^{n-m}}{n-m} \\ &= \frac{2\sigma}{\pi(q-1)}\tau^q \left(\frac{R}{\tau}\right)^{q-1} \\ &= \frac{2\sigma}{\pi}\frac{R^{q-1}}{q-1}\tau \propto \tau \end{aligned}$$

To summarize, if our starting theorem (5) is true, one gets

$$M_q^{(e)}(\tau) \propto \begin{cases} \tau^q & q \leq 1 \\ \tau & q > 1 \end{cases} \qquad \Rightarrow \qquad D(q) = \begin{cases} q & q < 1 \\ 1 & q \geq 1 \end{cases} \tag{3.13}$$

This is exactly the empirical behaviour observed for $D(q)$ in spite of self-affinity (eq. 3.4), which would imply $D(q) = q$ for every order q.

Assuming eq. (3.7), the above argument accounts for the observed spurious multiscaling in the case of a LRW, namely a random walk whose independent increments are distributed according to a Lorentz law L. We wonder whether it still holds when such independent increments are drawn out from a general Levy stable PDF L_α. The main problem here is that the analytical form of L_α for a general $\alpha \neq 1$ is not known. Furthermore we wonder whether this argument still holds for a random walk whose returns, perhaps strongly dependent returns, are distributed according to any scaling PDF g with fat tails. We are going to show, at least qualitatively, that the previous result of eq. (3.13) is actually valid under these more general conditions too.

It is worth noting that, from an empirical viewpoint, this is a very important issue. Indeed in many practical applications, like finance, the PDF of the returns cannot be determined exactly: it can only be fitted by approximations. Take, for instance, the financial time series of the first chapter. In section (1.3) we have found that the PDFs of the returns are invariant under rescaling while in section (1.4) we have detected the power law behaviour of their tails and we have estimated the related decaying parameter. However, obviously the analytical form of the scaling PDF g is not known.

Suppose the scaling PDF g to have tails asymptotically distributed according to a power law of parameter γ:

$$\begin{aligned} p_\tau(u) &= \frac{1}{\tau^\alpha} g(\frac{u}{\tau^\alpha}) \\ g(u) &\simeq \frac{c}{u^\gamma} \quad \text{for} \quad |u| \gg 1 \end{aligned} \tag{3.14}$$

where c is a constant. Consider a very high order $q \gg 1$; then in the calculation of the empirical moments $M_q^{(e)}(\tau)$ by means of eq. (3.7) only very large returns $r(\tau) > R_*$ give a sensible contribution to the integral.

Then substituting eq. (3.14) in eq. (3.7) and using the scaling invariance:

$$\begin{aligned} M_q^{(e)}(\tau) &\approx 2 \int_{R_*}^{R} u^q p_\tau(u)\, du \\ &= 2\tau^{\alpha q} \int_{R_*/\tau^\alpha}^{R/\tau^\alpha} u^q g(u)\, du \approx 2c\tau^{\alpha q} \int_{R_*/\tau^\alpha}^{R/\tau^\alpha} u^{q-\gamma}\, du \\ &= \frac{2A}{q-\gamma+1} \tau^{\alpha q} \left[u^{q-\gamma+1}\right]_{R_*/\tau^\alpha}^{R/\tau^\alpha} = \frac{2A}{q-\gamma+1} \left[u^{q-\gamma+1}\right]_{R_*}^{R} \tau^{\alpha(\gamma-1)} \end{aligned}$$

Hence the empirical Hurst exponent $H(q)$ should vanish, at least for very high orders q. This suggests that the Hurst exponent analysis performed on a single time series whose returns are asymptotically power law distributed leads to a spurious multiscaling also for every self-affine process, and not only for a Lorentzian random walk.

Since the time series is invariant under rescaling we know that for low orders simple scaling holds. For the sake of precision, the scaling exponent $D(q)$ is linear with slope $H = \alpha$ for $q < q_0$, where

$$q_0 = \gamma - 1$$

is the threshold order beyond which the theoretical q-order moment M_q diverge, see section (2.3). In conclusion we can guess, on the basis of the above argument, that

$$D(q) = \begin{cases} \alpha q & \text{if } q < q_0 \\ \alpha q_0 & \text{if } q \geq q_0 \end{cases} \tag{3.15}$$

if theorem (5) is valid. In other words, for a time series with independent increments asymptotically distributed according to a power law, this theorem leads to a very strong form of multiscaling: the so called *biscaling*. Moreover, the empirical scaling exponent $D(q)$ displays such a multiscaling even for time series coming from a strictly self-affine process.

We remark that a key point for the argument outlined in this section to work is the divergence of the theoretical moments $M_q(\tau)$. Indeed, when the sum in eq.

(2.9) converges, we have not to care about the behaviour of $R(\tau)$ against τ because the large returns contribution is negligible. This means that in eq. (3.7) the upper bound of integration runs to infinity and, for a self-affine process, the expected simple scaling follows, see section (2.3). This is why in eq. (3.15) the scaling exponent is linear with the correct slope $H = \alpha$ until the threshold order q_0 and shows a ficticious biscaling beyond $q = q_0$.

3.3 The proof of our result: simple case

Let us now prove our theorem (5). Consider the simulation of a LRW, namely a random walk with N independent increments distributed according to a Lorentz law $L(u)$. For each time window τ collect the returns $r_t(\tau) \quad t = 1 \dots N$ along the time series using a sliding window. Then the constancy of R against τ means that the largest among the returns $r_t(1) \quad t = 1 \dots N$ is nearly equal to the largest among the returns $r_t(\tau) \quad t = 1 \dots N$. In the following we are going to show that this is actually the case, at least for $\tau = 1$ and $\tau = 2$:

$$R(1) \simeq R(2) \tag{3.16}$$

if the size N of the time series is large enough. The proof in the general case follows the same strategy with minor modifications and is given in the last section.

Consider two returns in succession: $r_i(1)$ and $r_{i+1}(1)$; their sum $r_i(2) = r_i(1) + r_{i+1}(1)$ is a return over a time window $\tau = 2$. Let the modulus of such return be much larger than σ:

$$|r_i(2)| = z \qquad z \gg \sigma$$

where σ is the width parameter of the scaling PDF L, see eq. (3.3). Since σ is only a constant scaling factor, it does not affect our calculations and thus we can put $\sigma = 1$ without any loss of generality. Now we have a large return in a time window $\tau = 2$ made of two single returns and we want to calculate the probability $\mathcal{P}$ that there is one between them whose magnitude is nearly equal to z. In other words $\mathcal{P}$ is the probability that $r_i(1) \simeq z$ or $r_{i+1}(1) \simeq z$.

Formally, setting $z \gg 1$, we are concerned with two events:

event A : $r_i(1) + r_{i+1}(1) = z$

event A′ : $|r_i(1)| \leq \sqrt{z} \quad \text{or} \quad |r_{i+1}(1)| \leq \sqrt{z}$

And we want to calculate

$$\mathcal{P} = Prob(A'|A)$$

Hence $\mathcal{P}$ is the probability that one of two returns r_1, which make the return $r_2 = z$, has an absolute value less than $\sqrt{z}$ and so it is negligible in comparison to z if $z \gg 1$. Let's first define the following two events on the basis of the formers:

event B : $B = A \cap A'$

event $\bar{\mathbf{B}}$: $\bar{B} = A - B$

By construction:

$$A = B \cup \bar{B} \quad \text{and} \quad B \cap \bar{B} = \emptyset$$
$$Prob(A) = Prob(B) + Prob(\bar{B})$$

Thus:

$$\mathcal{P} = \frac{Prob(A \cap A')}{Prob(A)} = \frac{1}{1 + \dfrac{Prob(\bar{B})}{Prob(B)}} \tag{3.17}$$

Since the increments are supposed to be independent with PDF $L(u)$ the probabilities defined above are written as:

$$Prob(B) \;=\; 2 \int_{-\sqrt{z}}^{\sqrt{z}} du\, L(u)\, L(z-u) \tag{3.18a}$$

$$Prob(\bar{B}) \;=\; 2 \int_{-\infty}^{-\sqrt{z}} du\, L(u)\, L(z-u) \;+\; 2 \int_{\sqrt{z}}^{z/2} du\, L(u)\, L(z-u) \tag{3.18b}$$

The factor 2 in front of the integrals comes from the restriction of the integration range to $z/2$. The probabilities $Prob(B)$ and $Prob(\bar{B})$ are invariant under returns exchange $r_i(1) \longleftrightarrow r_{i+1}(1)$ and such symmetry reflects on the range of integration[13] because

$$\int_{-\infty}^{z/2} du\, L(u)\, L(z-u) = \int_{z/2}^{\infty} du\, L(u)\, L(z-u)$$

Lorentz law has a single maximum in $u = 0$ and this allows us to find:

- In eq. (3.18a): $L(z-u) \geq L(z+\sqrt{z})$
- In eq. (3.18b): $L(z-u) \leq L(z+\sqrt{z})$ in the first integral, and $L(z-u) \leq L(z/2)$ in the second one.

Thus:

$$Prob(B) \;\geq\; 2\, L(z+\sqrt{z})\,(1-2\omega) \;\geq\; 2\, L(2z)\,(1-2\omega) \tag{3.19a}$$

$$Prob(\bar{B}) \;\leq\; 2\omega \left[L(z+\sqrt{z}) \;+\; L(\frac{z}{2}) \right] \;\leq\; 4\, L(\frac{z}{2})\, \omega \tag{3.19b}$$

Where we put

$$\omega = \int_{\sqrt{z}}^{\infty} du\, L(u)$$

[13] Moreover this does not depend on the actual shape of the density function

We now use the analytical form of $L(u)$ with $\sigma = 1$, see eq. (3.3), and thus we get:

$$Prob(B) \geq \frac{1}{2\pi} \frac{1-2\omega}{1+z^2} \tag{3.20a}$$

$$Prob(\bar{B}) \leq \frac{16}{\pi} \frac{\omega}{1+z^2} \tag{3.20b}$$

Substituting eq. (3.20) in eq. (3.17):

$$\mathcal{P} = Prob(A' \mid A) \geq \frac{1}{1+\frac{32\omega}{1-2\omega}} = \frac{1-2\omega}{1+30\omega} \tag{3.21}$$

We can find an upper bound on ω:

$$\omega \leq \frac{1}{\pi} \int_{\sqrt{z}}^{\infty} \frac{dx}{x^2} = \frac{1}{\pi\sqrt{z}}$$

If $z \gg 1$ then $\omega \to 0$ and this implies:

$$\mathcal{P} \to 1 \tag{3.22}$$

What does this mean? If z is very large then every time event A happens, event B happens too. A large two-period return $r_2(i)$ is very likely to be made of two one-period returns $r_1(i)$ and $r_1(i+1)$ one of which is very close to $r_2(i)$ and the other one is negligible.

It's worth noting that, as we demonstrated in section (3.2), this mechanism leads to multiscaling no matter how many events N there are in our time series, provided that their number is finite. The uncertainty in the determination of $D(q)$ may depend on N: for a larger N we get a confidence interval much smaller; but bifractality still holds, giving robustness to this spurious feature. The strong bending at $q = 1$, which is often regarded as a multiscaling effect, is due to Pareto tails and the finiteness itself of the size N.

The sliding window method of collecting returns and the availability of a single time series is a key point here. The very definition of event A at the beginning of our calculation says that the two-period return r_2 is made of two one-period returns r_1 and so it is not independently extracted: once the two returns r_1 are given then the return r_2 is also determined. If, for each window τ, we draw out the returns $r(\tau)$ from different simulations, then the theorem (5) would fail.

It's easy to check that eqs. (3.20) fail when the tails of return PDFs decay faster than any power law, like in the Gaussian case. Indeed for our reasoning to work[14]

[14]That is, to deduce eq. (3.20) from eq. (3.19). The latters are true for very general distribution, but they imply the formers for power law decay of tails only

it's fundamental that[15] $L(au) \simeq C(a)L(u)$ and, while this is true for any power law, it fails for a short tailed distribution like the Gaussian one.

In fact, consider a scaling PDF behaving as a power law for large returns[16]:

$$g(u) \simeq \frac{a}{u^{\gamma}}$$

then, going back to eq. (3.20), we get

$$\mathcal{P} \geq \frac{1-2\omega}{1+2\omega(2^{2\gamma}-1)}$$

with

$$\omega \leq \frac{a}{\gamma-1}\, z^{-\frac{\gamma-1}{2}}$$

Since $\lim_{z\to\infty} \omega = 0$ for $\gamma \geq 2$ then $\mathcal{P} \to 1$. This is very important since it allows us to generalize the result of the theorem (5) to any scaling PDF g with power law tails. Since in the case of our financial indexes the scaling PDF g actually has power law tails, see section (1.4), then the theorem (5) applies.

3.4 The effect due to the volatility clustering

Now we come back to section (2.6) where the generalized Hurst exponent analysis was performed on the financial indexes of chapter (1). Since the financial time series of our indexes are invariant under rescaling and since the scaling PDF g of their daily returns has power law tails, according to the previous two sections one may expect to find biscaling like that in eq. (3.15). However this not the case: in fig. (2.6) and in the subsequent ones we can clearly see a quite different global behaviour. In particular the observed scaling exponents $D(q)$ differ qualitatively from that of a LRW (eq. 3.13) for two reasons:

1. a very high uncertainty beyond the threshold order q_0 of diverging theoretical moments;

2. a weaker bending of $D(q)$ for $q \gtrsim q_0$.

A possible explanation of the great uncertainty and clustering exhibited by scaling functions $D(q)$ in the case of real financial time series in comparison to gaussian or lorentzian random walk (LRW) lies in the form of the scaling density functions $g(u)$ of the daily returns. May be there is a competition between the central bulk

[15] C(a) is a function of the multiplicative factor only

[16] for a Lorentzian Random Walk $\gamma = 2$

and tails. In contrast to a Lorentz distribution (eq. 3.3) we can see in fig. (1.10) that the tails of $g(u)$ start to be important for very large returns only. This means that, in order to well explore tails, we need the size N of the time series to be larger in comparison to that of a LRW.

If N is not large enough, then our extracted sample may not contain any large return at all and so cannot feel tails: it lies in the central bulk like for the Gaussian random walk. Therefore it behaves as if the theoretical moments would all exist and the argument of section (3.2) does not apply. In such a case we find a straight line with slope close to $H = \frac{1}{2}$. In alternative it may happen that a subsample contains a few large oscillations. Hence, due to the probabilistic mechanism outlined in the previous sections, it behaves according to eq. (3.15) and display a strong spurious multiscaling. This makes the Hurst exponent $H(q)$ very sample dependent, and thus unreliable, for $q \gtrsim q_0$.

The second difference is even more interesting. In section (3.2) we saw that, for independent increments decaying like a power law, the theorem (5) leads to a strong spurious multiscaling. However, the scaling function of our financial indexes does not agree with such a strong multiscaling, although their returns are collected along a single time series and are asymptotically distributed according to a power law. In what follows we are going to see, at least qualitatively, that this discrepancy comes from volatility clustering. Indeed, the sharp biscaling in eq.(3.15) strictly holds for independent returns as it is based on (3.16).

Our theorem (eq. 3.16) in the section (3.3) implies that, if a return $r(2)$ has size $z \gg 1$ then it is the sum of two returns $r(1)$, one of which has size very close to z, while the other is negligible. This is why, due to eq. (3.22), such configuration is more likely to happen than any other. We saw that this is true when dealing with a Lorentzian random walk (LRW) where returns are independent; moreover we showed that it is also true for every self-affine process with independent increments asymptotically distributed according to a power law.

Now we proceed in taking into account correlations among returns. We saw in section (1.2) that the autocorrelation function $A_{vol}(s)$ among the absolute daily returns $r(1)$ of our financial time series is a slowly decaying function of the time gap s between two such returns. Hence the increments of our financial time series are strongly correlated during large time intervals. What happens when the absolute value of two consecutive returns are positively correlated?

Roughly speaking, if returns are positively correlated it is more likely than in the independent case that two successive return are both large. Hence, the maximum absolute return $R(\tau)$ may be a function growing with τ and not a constant. So, our theorem (3.16) no longer holds. Indeed, since a large return is likely to be followed by another large return, we no longer need a return r_1 to be very close to z because the very next one will probably have rather the same size; so the first oscillation needs only to reach a fraction of the size z. Hence the configuration with a return

negligible in comparison to the other is no more the most probable one.

In order to forecast the consequences of this fact, we note that the case of Gaussian random walk is very similar: R grows with the time window, precisely

$$R(\tau) \approx \sigma_0 \sqrt{\tau}$$

where σ_0 is the standard deviation of the Gaussian scaling PDF. On the ground of the above qualitative argument we expect volatility clustering to attenuate the strong bending observed in $D(q)$ for an independent LRW. Hence:

- Below the threshold order q_0 of converging theoretical moments we still get a line with slope H.
- Beyond q_0, we expect a weaker bending in $D(q)$: perhaps we get a line with a slope smaller than H, but not necessarily an horizontal one.

We stress that this weaker bending in the function $D(q)$ would still be spurious as it would be generated by the same probabilistic mechanism outlined in the previous sections.

To empirically check the previous argument and test the influence of the volatility clustering we would need a correlated self-affine process displaying fat tails in return PDFs and which is analitically soluble. Then we could compare the scaling exponent coming from a simulation with the expected one. Otherwise we may consider our real financial indexes. Of course, we cannot solve analitically the problem, but we can destroy any correlation by reshuffling data randomly.

A permutation on the set of daily returns $r(1)$ available surely make the volatility clustering negligible. Then a comparison between the curve $D(q)$ obtained by real data and by the same reshuffled data should provided us with an understanding of the volatility clustering effect. So, we randomly reshuffled the whole series of the daily returns of our indexes and then performed the usual Hurst exponent analysis on the reshuffled series.

It's important to remember that for the PDF of the returns the variance σ^2 is very strongly suggested to exist, see section (1.4), thus the PDF of returns r_τ tends to a Gauss' law as $\tau \to \infty$, see the remark at the end of section (1.3). Correlations avoid such convergence until the time window τ become so large to make them negligible. If the correlations are destroyed one may worry about the existence of scaling properties themselves, i.e. the validity of eq. (1.5) which implies the same functional form for return PDFs regardless of the time window τ. Fortunately it seems that sums of variables distributed according to a law whith long tails converge to the Gaussian very slowly [7], so we can approximatively regard at scaling as still holding in the range $1 \leq \tau \leq 32$, see section (1.3).

In fig. (3.4 - 3.7) we show the result obtained. The new curves $D(q)$ compares with the original ones, see fig. (2.6 - 2.9). Here we find that the volatility clustering

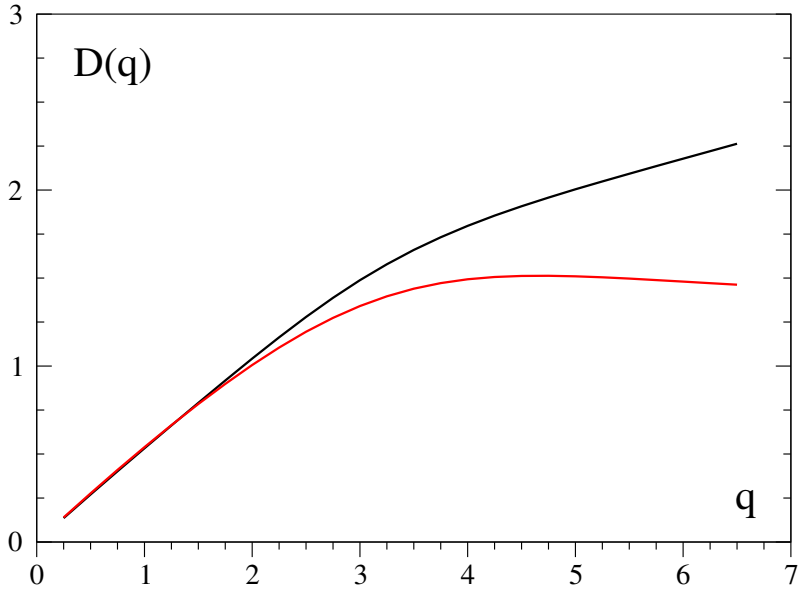

Figure 3.4: Scaling exponent of the **DJI** Index (*black curve*) and of the same time series with all the daily increments randomly reshuffled (*red curve*).

attenuates the multiscaling behaviour of our index; while correlation among returns should be one of the causes of multiscaling. When there is not dependence between returns, $D(q)$ behaves approximately according to eq. (3.15), i.e. as if it would be affected by the same spurious effect seen in section (3.2). In our opinion this is another fact supporting that the multiscaling observed in our financial indexes is ficticious: here the correlation does not attenuate a real multiscaling, but only the effects due to the mechanism described in section (3.3).

3.5 The proof of our result: general case

In the following we generalized the proof in section (3.3) when $\tau = n$:

$$R(n) \approx R(1)$$

namely when n returns $r(1) \quad i = 1 \ldots n$ are added up to get a return $r(\tau) = z$: this proves our theorem (5). As before we let $r(\tau)$ be very large:

$$z \gg 1$$

and try to prove that in such a case the probability that all returns $r(1)$ are negligible but one, which magnitude is then very close to z. Precisely, let $\mathcal{P}$ be the probability

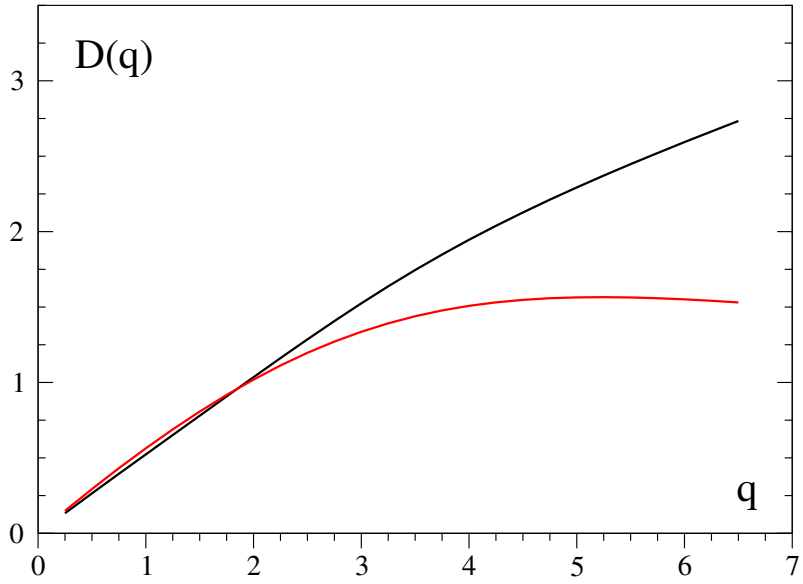

Figure 3.5: Scaling exponent of the **SPC** Index (*black curve*) and of the same time series with all the daily increments randomly reshuffled (*red curve*).

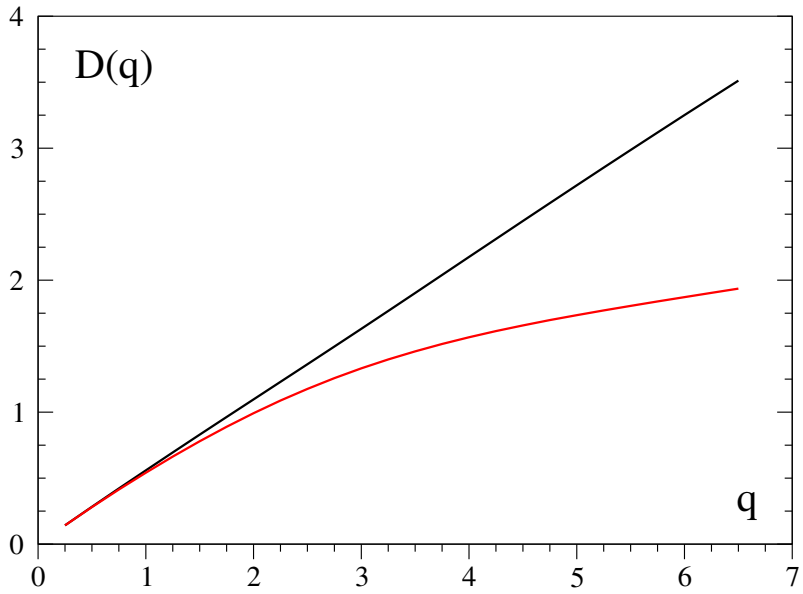

Figure 3.6: Scaling exponent of the **DJT** Index (*black curve*) and of the same time series with all the daily increments randomly reshuffled (*red curve*).

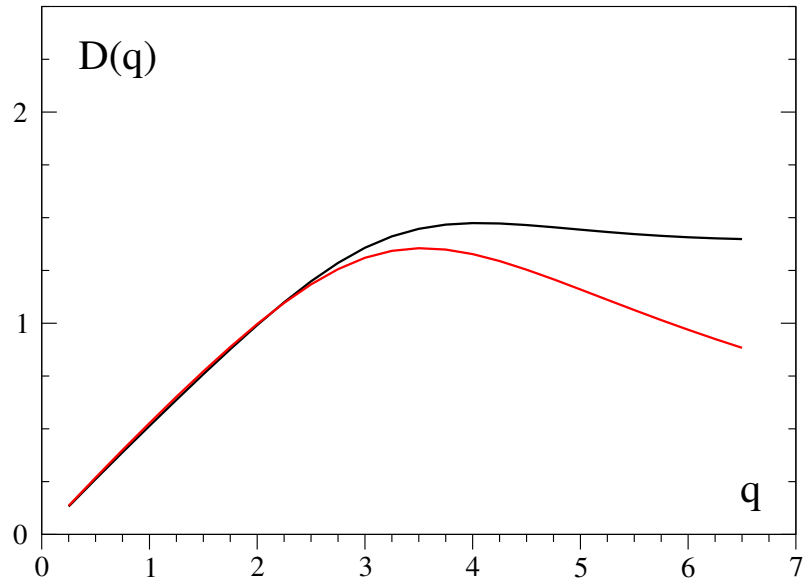

Figure 3.7: Scaling exponent of the **S&P** Index (*black curve*) and of the same time series with all the daily increments randomly reshuffled (*red curve*).

that $\forall i \neq j \ |r_i(1) \leq \sqrt{z}|$, then $\lim_{z\to\infty} \mathcal{P} = 1$. Hence the probability $\mathcal{P}$ that $\exists j \mid |r_1(j)| \geq z - (n-1)\sqrt{z}$ tends to 1 for $z \to \infty$. We recall that $L(u)$ is the PDF of the increments $r(1)$.

Consider the events:

$$\textbf{event A} : \sum_{i=1}^{n} r_i = z$$

$$\textbf{event } \mathbf{A'_j} : \forall i \neq j \quad |r_i| \leq \sqrt{z}$$

$$\textbf{event A}' = \bigcup_{j=1}^{n} A'_j$$

Then, as before:

$$\mathcal{P} = Prob(A'|A)$$

Now we define again the two events:

$$\mathbf{B} = A \cap A'$$

$$\bar{\mathbf{B}} = A - B$$

Hence, by construction:

$$\mathcal{P} = \frac{1}{1 + \dfrac{Prob(\bar{B})}{Prob(B)}} \tag{3.23}$$

We now obtain a lower bound for $Prob(B) = \bigcup_{j=1}^{n} A \cap A'_j$ using independence and symmetry among returns:

$$\begin{aligned} Prob(A \cap A'_j) &= \int du_j L(u_j) \int_{-\sqrt{z}}^{\sqrt{z}} du_1 \dots du_{j-1} du_{j+1} \dots du_n \\ &\quad \delta(z - \sum_{i=1}^{n} u_i)\, L(u_1) \dots L(u_{j-1}) L(u_{j+1}) \dots L(u_n) \\ &= \int_{-\sqrt{z}}^{\sqrt{z}} du_1 \dots du_{n-1}\, L(u_1) \dots L(u_{n-1})\, L\left(z - \sum_{i=1}^{n-1} u_i\right) \end{aligned}$$

Since $A'_j \cap A'_k =$ if $j \neq k$ we get:

$$Prob(B) = n \int_{-\sqrt{z}}^{\sqrt{z}} du_1 \dots du_{n-1}\, L(u_1) \dots L(u_{n-1})\, L\left(z - \sum_{i=1}^{n-1} u_i\right) \tag{3.24}$$

which corresponds to eq.(3.18a) in the case $n = 2$. Since Lorentz law[17] has a single maximum in $u = 0$ then[18], in eq. (3.24):

$$L\left(z - \sum_{i=1}^{n-1} u_i\right) \geq L[z + (n-1)\sqrt{z}] \geq L(nz) \tag{3.25}$$

which in turn implies:

$$Prob(B) \geq n(1 - 2\omega)^{n-1} L(nz) \tag{3.26}$$

which corresponds to eq.(3.19a) in the case $n = 2$. Here we put again

$$\omega = \int_{\sqrt{z}}^{\infty} du\, L(u)$$

As a second step we obtain an upper bound for $Prob(\bar{B}) = Prob(A) - Prob(B)$. The probability of the event A is:

$$Prob(A) = \int du_1 \dots du_{n-1}\, L(u_1) \dots L(u_{n-1})\, L\left(z - \sum_{i=1}^{n-1} u_i\right) \tag{3.27}$$

[17] which we denote with $p(u)$ here

[18] taking $z \gg 1$ we can obviously ask for $z \gg \sqrt{z}$ to hold

where the integration is performed on the whole $\mathbb{R}^{n-1}$. Hence, using eq.(3.24):

$$Prob(\bar{B}) = Prob(A) - n \int_{\mathcal{C}} du_1 \dots du_{n-1}\, L(u_1) \dots L(u_{n-1})\, L\left(z - \sum_{i=1}^{n-1} u_i\right) \quad (3.28)$$

where $\mathcal{C} : \vec{u} \in \mathbb{R}^{n-1} \mid |u_i| \leq \sqrt{z} \quad i = 1 \dots n-1$ is simply the cube of side $2\sqrt{z}$ centered in the origin $\mathcal{O}$. At this point we want to carry the last factor $L(z - \sum_{i=1}^{n-1} u_i)$ out of the integrals by replacing it with a constant upper bound, like in eq.(3.25) where we was able to find a constant lower bound and hence to greatly simplify the calculation.

However there are $n-1$ points $\vec{v}\ (j)$ which prevent us to find a good constant upper bound: indeed taking into account

$$\vec{v}\ (j) \mid v_i(j) = \begin{cases} 0 & \text{if} \quad i \neq j \\ z & \text{if} \quad i = j \end{cases}$$

the factor $L(z - \sum_{i=1}^{n-1} u_i)$ can be bounded by 1 only. To get around this difficulty we need a geometrical trick very similar to that used in writing eq.(20b)[19]. The idea is to restrict the integral in a region as far as possible from the points $\vec{v}\ (j)$ using the symmetry of the integrand.

Consider $n-1$ changes of coordinates $S_j : \vec{u} \Rightarrow \vec{u'}$ in $\mathbb{R}^{n-1}$:

$$S_j : \ u'_i = \begin{cases} u_i & \text{if} \quad i \neq j \\ z - \sum_{k=1}^{n-1} u_k & \text{if} \quad i = j \end{cases}$$

It is straightforward to show that the integral in eq.(3.27) is invariant for all these transformations:

$$\begin{aligned} \forall E \subseteq \mathbb{R}^{n-1} \qquad & \int_E du^{n-1}\, L(u_1) \dots L(u_{n-1})\, L\left(z - \sum_{i=1}^{n-1} u_i\right) = \\ & = \int_{S_j(E)} du^{n-1}\, L(u_1) \dots L(u_{n-1})\, L\left(z - \sum_{i=1}^{n-1} u_i\right) \end{aligned} \quad (3.29)$$

Furthermore, by construction:

$$\sum_{i=1}^{n-1} u'_i = z - u_j \quad (3.30)$$

[19] in that case we had to avoid the point z on the real line

for every transformation S_j.

Let

$$\Pi : \vec{u} \mid \sum_{i=1}^{n-1} u_i \leq (1 - \frac{1}{n})z$$

a subset of $\mathbb{R}^{n-1}$ and $S_j(\Pi)$ the transformed set under S_j. It's easy to prove that

$$\mathbb{R}^{n-1} \subseteq \Pi \cup \bigcup_{j=1}^{n-1} S_j(\Pi)$$

Indeed let $\vec{w} \in \mathbb{R}^{n-1}$, if $\sum_{i=1}^{n-1} w_i \geq (1 - \frac{1}{n})z$ then $\exists j \mid w_j \geq \frac{1}{n}$. By eq.(3.30) we get $\sum_{i=1}^{n-1} w'_i \leq (1 - \frac{1}{n})z$, i.e. it exist S_j such that $S_j \vec{w} \in \Pi$. We are now ready to write eq.(3.27) in a more suitable form:

$$\begin{aligned}
Prob(A) &= \int_{\mathbb{R}^{n-1}} du^{n-1} L(u_1) \dots L(u_{n-1}) L\left(z - \sum_{i=1}^{n-1} u_i\right) \\
&= \int_{\Pi \cup \bigcup_{j=1}^{n-1} S_j(\Pi)} du^{n-1} L(u_1) \dots L(u_{n-1}) L\left(z - \sum_{i=1}^{n-1} u_i\right) \\
&\leq \sum_{j=1}^{n-1} \int_{S_j(\Pi)} du^{n-1} L(u_1) \dots L(u_{n-1}) L\left(z - \sum_{i=1}^{n-1} u_i\right) \\
&+ \int_{\Pi} du^{n-1} L(u_1) \dots L(u_{n-1}) L\left(z - \sum_{i=1}^{n-1} u_i\right) \\
&= n \int_{\Pi} du^{n-1} L(u_1) \dots L(u_{n-1}) L\left(z - \sum_{i=1}^{n-1} u_i\right)
\end{aligned}$$

Where the last equality comes from eq.(3.29).

Hence, substituting in eq.(3.28), we find:

$$Prob\bar{B} = n \int_{\Pi - \mathcal{C}} du^{n-1} L(u_1) \dots L(u_{n-1}) L\left(z - \sum_{i=1}^{n-1} u_i\right) \tag{3.31}$$

which corresponds to eq.(3.18b) for $n = 2$. Since $L(u)$ has a single maximum in $u = 0$ we have

$$L\left(z - \sum_{i=1}^{n-1} u_i\right) \leq L(\frac{z}{n})$$

and by replacing in eq.(3.31):

$$\begin{aligned} Prob(\bar{B}) &\leq n\,L(\frac{z}{n}) \int_{\Pi-\mathcal{C}} du^{n-1}\, L(u_1)\ldots L(u_{n-1}) \\ &\leq n\,L(\frac{z}{n}) \int_{\mathcal{R}^{n-1}-\mathcal{C}} du^{n-1}\, L(u_1)\ldots L(u_{n-1}) \\ &= n\,L(\frac{z}{n})(1-\int_{\mathcal{C}} du^{n-1}\, L(u_1)\ldots L(u_{n-1})) \\ &= n\,L(\frac{z}{n})\,[1-(1-2\omega)^{n-1}] \end{aligned} \tag{3.32}$$

which corresponds to eq.(3.19b) for $n = 2$.

Finally we substitute eq.(3.26) and (3.32) in definition (3.23):

$$\mathcal{P} \geq \frac{1}{1+\frac{L(\frac{z}{n})[1-(1-2\omega)^{n-1}]}{L(nz)(1-2\omega)^{n-1}}}$$

As in the main text, making explicit the functional form of $L(u) = \frac{\sigma}{\pi}\frac{1}{\sigma^2+u^2}$ we find[20]

$$\frac{L(\frac{z}{n})}{L(nz)} = \frac{1+z^2n^2}{1+\frac{z^2}{n^2}} \leq n^4$$

Hence

$$\mathcal{P} \geq \frac{1}{1+\frac{1-w}{w}n^4} = \frac{w}{n^4-w(n^4-1)}$$

where

$$w = (1-2\omega)^{n-1}$$

This corresponds to eq.(3.21) for $n = 2$.

Since[21]

$$\lim_{z\to\infty} \omega = 0$$

it is straightforward to find that, for any given n

$$\lim_{z\to\infty} w = 1$$

[20] we again stress that the exact analytical form of $p(u)$ doesn't matter, see the end of section (3.3). The key point here is to find an upper bound to the rate $\frac{p(z/n)}{p(nz)}$ which does not depend on z. Such a constant bound exist for every $p(u)$ whose tails decay asymptotically according to some power law; while it does not exist for short tailed density functions, like the Gaussian one.

[21] in the main text we saw that $\omega \leq \frac{\lambda}{\pi\sqrt{z}}$

This means that, for a fixed n, the probability $\mathcal{P}$ is very close to one:

$$\mathcal{P} \to 1 \tag{3.33}$$

when z is very large; and this concludes our proof.

As stressed in section (3.3) $\mathcal{P}$ is the probability that, given a very large return $r_\tau = z$, the $\tau = n$ returns $r_1(i) \quad i = 1 \dots n$ are all negligible in comparison with z except one, whose magnitude is close to z. In other words it is the probability that there are not two or more large[22] returns in a sequence of n consecutive returns r_1. We remark that in order eq. (3.33) to hold we need to fix the maximum window $\tau = n$ and *then* to take $z \gg 1$, otherwise we it could not be true that a large return is surrounded by $n - 1$ smaller returns. The size of z, and thus the lenght of the time series, must be as larger as the maximum window n increases. In conclusion, fix a given range n within which the time window τ may vary; then $\mathcal{P} \to 1$ if the size of the time series is large enough.

[22] in comparison with z

Conclusions

This thesis has been devoted to the characterization of the invariance under rescaling of financial time series. Our main result is that the multiscaling observed until now in many financial time series could be a spurious effect. We proved that a probabilistic mechanism working in the empirical statistical analysis of a single time series and based on the power law tails of the density function of the returns, can affect the outcoming Hurst exponent and leads to a strong spurious multiscaling even for a strictly simple scaling underlying process. Since this effect is due only to the availability of a single empirical time series and to the presence of extreme events whose distribution follows a power law, we cannot exclude that our results could be relevant to fields of complex systems physics different from finance.

In chapter (3) we estimated the Hurst exponent for an exactly solvable model, the Lorentzian random walk, in order to compare it with the expected one. Based on the fact that this model is strictly self-similar, we first proved that the empirical scaling exponent $D(q)$ should follow a straight line with slope $H = 1$ for all orders q. However, we found $D(q)$ to display a strong form of multiscaling for orders larger than a certain threshold q_0 when the Hurst exponent analysis is performed on a single history. Remarkably, this threshold order equals the order at which the theoretical moments of the Lorentz's law no longer exist.

Next, we were able to account for this spurious multiscaling by proving that for every time series, generated by a simple scaling model, the empirical $D(q)$ leads to a strong ficticious multiscaling if the returns are independent and asymptotically power law distributed. More precisely, for orders less than the threshold order q_0 beyond which the moments diverge, the moment scaling exponent analysis gives the correct result one expects on the basis of the simple scaling. However, for $q \geq q_0$ it displays an horizontal line, thus leading to a spurious $H = 0$.

This spurious multiscaling has already been studied in [5], but the argument given does not hold when the increments are collected by means of a sliding window method, as it is often the case; moreover it cannot be generalized to time series with correlated increments. Our main contribution was to provide a convincing mechanism able to account for this multiscaling under very general conditions. In particular it retains its validity also in the case of strongly dependent increments.

Indeed, we generalized it to scaling invariant time series whose increments have a slowly decaying second order correlation, at least qualitatively. According to the argument given in section (3.4), a strong correlation among the absolute returns can attenuate the probabilistic effect previously outlined, and thus give rise to a weaker spurious multiscaling than in the independent case.

We tested this claim on our financial indexes directly. After a reshuffling of the increments of the time series, we calculated again the Hurst exponent and found that its behaviour agree with the ficticious multiscaling obtained in the case of the Lorentzian random walk. Hence, the observed empirical multiscaling observed in many financial time series could be a spurious effect due to the interplay between the volatility clustering and the mechanism which leads to a strong ficticious multiscaling in the case of independent returns. We stress also that this statement holds for any time series whose large absolute increments follow a Pareto tail and not for the financial ones only.

On the other hand, the Hurst exponent is defined on the basis of the existence of the moments. However, a well established stylized fact states that in finance the extreme events are power law distributed, and thus the summation used to calculate the empirical q-order moment does not converge for an order q larger than a threshold q_0.

In chapter (2) we estimated, for our financial indexes, the uncertainty associated to the scaling exponent $D(q)$ by means of a non-parametric bootstrap method. The same method has been used in [3] to estimate a confidence interval for the scaling exponent in the case of the spatial distribution of mining induced microearthquakes. The main empirical results are shown in fig. (2.6 - 2.9): the sample dependence of $D(q)$ becomes clearly important beyond the threshold order q_0 at which the moments no longer exist.

In conclusion we believe that the Hurst exponent analysis is reliable for low order q only, inside the range where the theoretical moments are expected to exist. This is because the existence of this moments prevents the mechanism outlined in chapter (3) to work and makes the sample dependence of the scaling exponent negligible.

It is worth mentioning that the first observation of multiscaling in finance was made on time series relative to exchange rates [26], and led to the suggestion of a close analogy between finance and turbulence. In spite of the fact that this analogy has been subsequently criticized [27, 28], the multiscaling of financial time series has been considered an important stylized fact by several authors [57, 60], and was put at the basis of some modelization of the market [18, 19, 22]. On the basis of the above evidencies, also theoretical models based on simple scaling could be good candidates for simulating the most important financial stylized facts, including the empirical observed multiscaling and the power law decay of the second order correlations. We believe that the elucidation of the probabilistic mechanism leading to this multiscaling is a useful contribution towards a better understanding of the

field.

Bibliography

[1] W. Feller (1951). 'The asymptotic distribution of the range of sums of independent random variables'. *Ann. Mat. Statist*, **22**, 427–432

[2] B.B. Mandelbrot and J.W. Van Ness (1968). 'Fractional brownian motions, fractional noises and applications'. *Soc. Ind. Appl. Math. Rev.*, **10**, 422–437

[3] M. Eneva (1994). 'Monofractal of multifractal: a case study of spatial distribution of mining induced seismic activity'. *Nonlinear Processes in Geophysics*, **1**, 182–190

[4] T. Aste, M.M. Dacorogna and T. Di Matteo (2005). 'Long-term memories of developed and emerging markets: using the scaling analysis to characterize their stage of development'. *Journal of Banking & Finance*, **29**, 827–851

[5] A.V. Chechkin and V.Y. Gonchar (2000). 'Self and spurious multi-affinity of ordinary Levy motion, and pseudo-Gaussian relations'. *Chaos Solitons & Fractals*, **11**, 2379–2390

[6] B. LeBaron (2001). 'Stochastic volatility as a simple generator of apparent financial power laws and long memory'. *Quantitative Finance*, **1**, 621–631

[7] R.N. Mantegna and H.E. Stanley (1994). 'Stochastic Process with Ultraslow Convergence to a Gaussian: The Truncated Levy Flight'. *Physical Review Letters*, **23**, 2946–2949

[8] B.V. Gnedenko and A.N. Kolmogorov (1954). 'Limit distributions for sums of independent random variables'. *Addison Wesley*

[9] K. Falconer (1990). 'Fractal geometry - mathematical foundations and applications'. *John Wiley & Sons*

[10] R. Toral and A. Chakrabarti (1993). 'Generation of Gaussian distributed random numbers by using a numerical inversion method'. *Computer Physics Communications*, **74**, 327–334

[11] J.P. Bouchaud and M. Potters (2003). 'Theory of financial risk and derivative pricing'. *Cambridge University Press*

[12] B.V. Gnedenko (1968). 'The theory of probability'. *Chelsea Pub. Co.*

[13] J.W. Kantelhardt (2008). 'Fractal and Multifractal Time Series'. *ArXiv: physics.data-an* 0804.0747v1

[14] L. Bachelier (1900). 'Theorie de la speculations'. *Annales de l'ecole Normale superieure*, **17**, 21–86

[15] D. Sornette (1998). 'Large deviations and Portfolio optimization'. *Physica A*, **256**, 251–283

[16] N.J. Higham (1990). 'Analysis of the Cholesky decomposition of a semi-definite matrix'. In 'Reliable numerical computation' *Oxford University Press*, 161–185

[17] T. Lux (2001). 'The multi-fractal model of asset returns: simple moment and GMM estimation'.

[18] T. Lux and M. Marchesi (1999). 'Scaling and criticality in a stochastic multi-agent model of a financial market'. *Nature*, **397**

[19] J.F. Muzy, D. Sornette, J. Delour and A. Arneodo (2001). 'Multifractal returns and Hierarchical Portfolio Theory'. *Quantitative Finance*, **1**, 131–148

[20] A. Arneodo, J.F. Muzy and D. Sornette (1998). 'Direct causal cascade in the stock market'. *The European Physical Journal B*, **2**, 277–282

[21] L. Calvet, A. Fisher and B. Mandelbrot (1997). 'Large Deviations and the Distribution of Price Changes'. *Cowles Foundation*, 1165

[22] B.B. Mandelbrot, A. Fisher and L. Calvet (1997). 'A Multifractal Model of Asset Returns'. Working Paper

[23] B.B. Mandelbrot (2001). 'Scaling in financial prices: IV. Multifractal concentration'. *Quantitative Finance*, **1**, 641–649

[24] F. Baldovin and A.L. Stella (2007). 'Central limit theorem for anomalous scaling due to correlations'. *Physical Review E*, **75**

[25] F. Baldovin and A.L. Stella (2007). 'Scaling and efficiency determine the irreversible evolution of a market'. *Proc. Nat. Acad. Sci.*, **104**, 19741–19744

[26] S. Ghashghaie, W. Breymann, J. Peinke, P. Talkner and Y. Dodge (1996). 'Turbulent cascades in foreign exchange markets'. *Nature*, **381**, 767–770

[27] R.N. Mantegna and H.E. Stanley (1997). 'Stock market dynamics and turbulence: parallel analysis of fluctuation phenomena'. *Physica A*, **239**, 255–266

[28] R.N. Mantegna and H.E. Stanley (1996). 'Turbulence and financial markets'. *Nature*, **383**, 587–588

[29] T. Bollerslev (1986). 'Generalized autoregressive conditional heteroscedasticity'. *Journal of Econometrics*, **31**, 307–327

[30] R.F. Engle (1982). 'Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation'. *Econometrica*, **50**, 987–1008

[31] B. Podobnik, P.C. Ivanov, I. Grosse, K. Matia and H.E. Stanley (2004). 'Arch-Garch approaches to modeling high-frequency financial data'. *Physica A*, **344**, 216–220

[32] R. Engle (2001). 'Garch 101: the use of Arch/Garch models in applied econometrics'. *Journal of Economic Perspectives*, **15**, 157–168

[33] D.B. Nelson (1991). 'Conditional heteroskedasticity in asset returns: a new approach'. *Econometrica*, **59**, 347–370

[34] R.F. Engle, D.M. Lilien and R.P. Robins (1987). 'Estimating time varying risk premia in the term structure: the Arch-M model'. *Econometrica*, **55**, 391–407

[35] R.T. Baillie, T. Bollerslev and H.O. Mikkelsen (1996). 'Fractionally integrated generalized autoregressive conditional heteroskedasticity'. *Journal of Econometrics*, **74**, 3–30

[36] R.A. Velasco Fuentes. 'Arch-type models'.

[37] R.F. Engle (1983). 'Estimates of the variance of U.S. inflation based upon the Arch model'. *Econometrica*, **15**, 286–301

[38] T. Bollerslev and H.O. Mikkelsen (1996). 'Modeling and pricing long memory in stock market volatility'. *Journal of Econometrics*, **73**, 151–184

[39] R.F. Engle and T. Bollerslev (1986). 'Modelling the persistence of conditional variances'. *Econometric Reviews*, **5**, 1–50

[40] J.M. Maheu (2002). 'Can Garch models capture the long-range dependence in financial market volatility?'.

[41] R. Cont (2001). 'Empirical properties of asset returns: stylized facts and statistical issues'. *Quantitative Finance*, **1**, 223–236

[42] A.W. Lo (1991). 'Long-term memory in stock market prices'. *Econometrica*, **59**, 1279–1313

[43] V. Teverovsky, M.S. Taqqu and W. Willinger (1999). 'A critical look at Lo's modified R/S statistic'. *Jour. of stat. planning and inference*, **80**, 211–227

[44] H.E. Stanley and V. Plerou (2001). 'Scaling and universality in economics: empirical results and theoretical interpretation'. *Quantitative Finance*, **1**, 563–567

[45] R. Cont (2005). 'Volatility clustering in financial markets: empirical facts and agent-based models'.

[46] B.M. Hill (1975). 'A simple general approach to inference about the tail of a distribution'. *The Annals of Statistics*, **3**, 1163–1174

[47] S. Resnick and C. Starica (1996). 'Tail index estimation for dependent data'. *Annals of Applied Probability*, **8**, 1156–1183

[48] H. Drees, L. de Hann and S. Resnick (2000). 'How to make a Hill plot'. *The Annals of Statistics*, **1**, 254–274

[49] E. Carlstein, K.A. Do, P. Hall, T. Hesterberg and H.R. Kunsch (1998). 'Matched-block bootstrap for dependent data'. *Bernoulli*, **4**, 305–328

[50] B. Efron (1979). 'Bootstrap Methods: Another Look at the Jackknife'. *The Annals of Statistics*, **7**, 1–26

[51] B. Efron and R. Tibshirani (1986). 'Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy'. *Statistical Science*, **1**, 54–75

[52] H.R. Kunsch (1989). 'The Jackknife and the Bootstrap for General Stationary Observations'. *The Annals of Statistics*, **17**, 1217–1241

[53] A. Carbone, G. Castelli and H.E. Stanley (2004). 'Time-dependent Hurst exponent in financial time series'. *Physica A*, **344**, 267–271

[54] C.J.G. Evertsz (1995). 'Fractal Geometry of Financial Time Series'. *Fractals*, **3**, 609–616

[55] J.P. Bouchaud, M. Potters and M. Meyer (2000). 'Apparent multifractality in financial time series'. *The European Physical Journal B*, **13**, 595–599

[56] Z.Q. Jiang and W.X. Zhou (2008). 'Multifractality in stock indexes: Fact or Fiction?'. *Physica A*, **387**, 3605–3614

[57] J. Fillol (2003). 'Multifractality: Theory and Evidence an Application to the French Stock Market'. *Economics Bulletin*, **31**, 1–12

[58] F. Schmitt, D. Schertzer and S. Lovejoy (2000). 'Multifractal fluctuations in finance'. *International Journal of Theoretical and Applied Finance*, **3**, 361–364

[59] K. Matia, Y. Ashkenazy and H.E. Stanley (2003). 'Multifractal properties of price fluctuations of stocks and commodities'. *Europhysics Letters*, **61**, 422–428

[60] L. Calvet, A. Fisher and B. Mandelbrot(). 'Multifractality of Deutschemark / US dollar exchange rates'.

[61] V.I. Belinicher, V.S. L'vov, A. Pomyalov and I. Procaccia (1998). 'Computing the scaling exponents in fluid turbulence from first principles: demonstration of multiscaling'. *Journal of Statistical Physics*, **93**, 797–832

[62] V.S. L'vov, E. Podivilov and I. Procaccia (1997). 'Temporal multiscaling in hydrodynamic turbulence'. *Physical Review E*, **55**, 7030–7036

[63] J. Herweijer and W. van der Water (1995). 'Universal shape of scaling functions in turbulence'. *Physical Review Letters*, **74**, 4651–4655

[64] M.H. Jensen (1999). 'Multiscaling and Structure functions in turbulence: an alternative approach'. *Physical Review Letters*, **83**, 76–80

[65] Z.S. She and E. Leveque (1994). 'Universal scaling laws in fully developed turbulence'. *Physical Review Letters*, **72**, 336–340

[66] E. Bonnet, O. Bour, N.E. Odling, P. Davy, I. Main, P. Cowie and B. Berkowitz (2001). 'Scaling of fracture systems in geological media'. *Reviews of Geophysics*, **39**, 347–383

[67] Y. Xu and P.W. Burton (2006). 'Time varying seismicity in Greece: Hurst's analysis and Monte Carlo simulation applied to a new earthquake catalogue for Greece'. *Tectonophysics*, **423**, 125–136

[68] L. Tennekoon, M.C. Boufadel, D. Lavallee and J. Weaver (2003). 'Multifractal anisotropic scaling of the hydraulic conductivity'. *Water Resources Research*, **39**, 1193–1205

[69] A. Montanari (2001). 'Dipendenza a lungo termine ed effetto di Hurst nelle serie temporali idrologiche'

[70] H.E. Hurst (1951). 'Long-term storage capacity of reservoirs'. *American Society of Civil Engineers*

[71] N.C. Matalas and C.S. Huzzen (1967). 'A property of the range of partial sums'. *Proc. Int. Hydrol. Symp.*, **1**, 252–257

[72] H.E. Hurst (1957). 'A suggested statistical model of some time series which occur in nature'. *Nature*, **180**

[73] R.N. Bhattacharya, V.K. Gupta and E. Waymire (1983). 'The Hurst effect under trends'. *J. Appl. Probab.*, **20**, 649–662