Master’s Thesis

Multifractal dynamics of stock markets

submitted in fulfilment of requirements for M.Sc. in computer physics

Łukasz Czarnecki
Matricule number: 176884
Specialty: Computer Physics
Supervisor: dr D. Grech

Wrocław, 2009
Contents

1 Introduction 5

2 Fractals 7
   2.1 Monofractals 8
       2.1.1 Hausdorff dimension 9
       2.1.2 Generating data with a specific fractal dimension 10
   2.2 Multifractals 12
   2.3 Selected methods for calculating the fractal dimension 13
       2.3.1 Rescaled Range Analysis 13
       2.3.2 Power Spectrum Analysis 14
       2.3.3 Detrended Fluctuation Analysis 15
       2.3.4 Multifractal Detrended Fluctuation Analysis 16

3 Program for calculating the MFDFA 19
   3.1 Program structure 20
   3.2 Usage and possibilities of the application 24
   3.3 Further development 25

4 Analyzed data and initial problems 27
   4.1 Data sets 27
   4.2 Analysis of scaling range 29

5 Results 33
   5.1 Random midpoint displacement data 33
   5.2 Long-term data analysis 36
       5.2.1 Original data analysis 36
       5.2.2 Analysis of positive & negative returns 38
       5.2.3 Analysis of weekly returns 41
   5.3 Analysis of bearish and bullish periods 43
   5.4 Analysis of automatically divided data 44
CONTENTS

5.5 Improvement of scaling properties by division and “cutting out”
   special events ............................................................ 46
   5.5.1 Short term example - WIG and S&P500 .................... 46
   5.5.2 Long term example - S&P500 ............................... 48

6 Conclusions 53

A Abstract 59
Chapter 1

Introduction

Stock markets have been analyzed with many methods in the past. None seemed to fully describe this extremely complex system. To create a working model of this system, one should consider the influence of politics, current economic situation, foreign influences and even something as unpredictable as human decisions. Therefore most efforts tend to find a way of description of the stock markets since exact modeling is practically impossible. There are many quantitative ways of describing the market, all of which concentrate only on some part of the issue. Since the potential reward of understanding the stock markets mechanics is great, efforts to achieve that goal continue to this day. Some recent research showed that stock markets can be described in terms of fractals. Evolution of fractal analysis led to the concept of multifractals, one of the most complex structures in science. It led also to several ways of their description.

In the first chapter we will try to understand what a fractal is and later, how to describe it. After introducing the concept of a monofractal and how to create one, we shall investigate the idea of a multifractal. Finally we will introduce some methods of describing multi- and monofractals.

After analyzing the structure of the computer program used to calculate the MFDFA, and describing the possibilities and usage of that application, we will try to find ways to improve it. This will be followed by a brief description of the data that will be analyzed, and a discussion of the initial problems of the analysis.

Chapter five begins with an analysis of the multifractal noise created by short artificial data sets, and is followed by the multifractal analysis of two indices: S&P500 and WIG. The problems of that analysis lead to the next sections where we will gradually find that a more local approach gives better results. We
will show that dividing the series into parts greatly improves the effectiveness of the MF DFA method for long financial data.

We will propose a method of dividing the series and cutting out extreme events to improve the scaling range. Usefulness of that approach shall be proven in the form of several examples. This will be followed by a summary of the results.
Chapter 2

Fractals

The term *fractal* is a relatively new one, it was coined by Benoît Mandelbrot, during the 70’s of the XX century, mainly by the book “Fractals: Form, Chance and Dimension” [1] (which is a translation of an earlier French edition from 1975). The easiest way to understand what a fractal is, is to look at its main feature, i.e. self-similarity. Self-similarity is a property of an object to be an exact (or almost exact) copy of its magnified part. Although the idea of such objects dates back to the late XIX century, there are many definitions of a fractal in its current, very wide, meaning. Currently, a fractal can be understood as a set that fulfills most of the enumerated conditions [2]:

- has a infinitely complex, and irregular structure at all possible scales,
- can’t be described by euclidean geometry nor calculus,
- shows some sort of self-similarity or self-affinity,
- a “fractal dimension” (defined for that particular fractal) is strictly greater then its topological dimension,
- there is a very simple rule for generating it,
- has a “natural” appearance.

The definition of a fractal should include ones with exact, quasi-exact and statistical self-similarity, but since this work concentrates on time series, only statistical self-similarity is demanded in further chapters. One of the most accurate ways to describe a fractal, is to state that it’s an object with a Hausdorff dimension (described in Subsection 2.1.1) larger then its topological dimension. It is worth noting, that in complex systems, fractal characteristics are observed
2.1. MONOFRACTALS

when the system has no specific scale, which happens when its properties depend on the size of the system as a power law. In this thesis, our main point of interest lies in time series, most examples and equations (unless stated otherwise) will be made for time series that show statistical self-similarity.

The end of the XX century was a time of much research in the field of fractals. New methods of analyzing fractals were developed in many areas of science, from avalanche simulation [3], heart-rate fluctuations [4], DNA analysis [5], finances and econophysics [6, 7, 8], meteorology [9, 10], astronomy [11], etc.

2.1 Monofractals

There are many ways that a fractal may be modified when observed at different scales. But when there is a single rule, that object can be called a monofractal. A simple example of a deterministic monofractal is the Koch snowflake where at each step (beginning with an equilateral triangle) each line segment is divided into 3 equal parts, and the middle one is substituted by an equilateral wedge.

\[
F(s) \sim s^{\alpha_H},
\]

(2.1)

Figure 2.1: First four steps of creating the Koch snowflake [12].

When this fractal is observed at different resolutions, the number of elements that become visible through greater resolution, depends only on the increase in resolution, and not on the initial resolution. For fractal time series, that show statistical self-similarity, it can be written in a simple form:
where \( s \) denotes the scale selected for a particular problem, \( F(s) \) is a statistically defined function that measures geometrical or topological properties of time series (e.g. fluctuations) and \( \alpha_H \) is the scaling exponent. Due to the significantly different ways fractals can be created, their comparison is not straightforward. Therefore, one has to define a quantitative way of description, that would allow to find differences between them.

### 2.1.1 Hausdorff dimension

One of the most useful ways of describing a fractal is to determine its Hausdorff dimension [13]. Let us denote the fractal set as \( A \), and cover it with a finite or countable number of open sets of diameter not larger then \( \epsilon \). Then one can define the function \( m_H(A, \theta) \) as

\[
m_H(A, \theta) = \lim_{\epsilon \to 0} \inf_{G \in \mathcal{G}} \left( \sum_{U \in G} (\text{diam } U)^\theta \right),
\]

where \( G \) is a finite or countable coverage of \( A \) with open sets \( U \) of diameter less or equal then \( \epsilon \), and \( \theta \) is a positive real number. The Hausdorff dimension of the set \( A \) \( d_H(A) \) is defined as

\[
d_H(A) = \inf \{ \theta : m_h(A, \theta) = 0 \},
\]

which gives a good quantitative description of a fractal. This method however was invented for fractal sets. Therefore its application to many fractals, and especially for time series, proves to be considerably difficult. In many cases, an alternative way of defining the fractal dimension, (i.e. the Hurst-Hausdorff exponent), is much more practical. Usually, it is defined in one of two ways:

1. When considering a time series \( x(t) \), one can make a rescaling of time:

\[
t \to t' = G^{-1}t, \quad G > 0
\]

\[
x(t) \to x(t') = x(G^{-1}t),
\]

and ask about time series magnitude rescaling \( S \), such that:

\[
Sx(G^{-1}t) \sim x(t),
\]

where \( \sim \) means selfsimilarity. It turns out that

\[
S = G^{\alpha_H}, \quad \alpha_H > 0
\]

where the \( \alpha_H \) denotes the Hurst-Hausdorff exponent.
2.1. MONOFRACTALS

When analyzing the so called second order fluctuation function of a detrended (which can be done in many ways, e.g. subtraction of a linear trend) time series, one can find:

\[ \langle (x(t + \tau) - x(t))^2 \rangle_{t \gg \tau} \sim \tau^{\alpha_H}, \quad \tau > 0 \quad (2.8) \]

where \( \alpha_H \) is again the Hurst-Hausdorff exponent. Several calculation methods for \( \alpha_H \) will be described in Section 2.3. It is worth noting at this point, that the Hurst-Hausdorff exponent is related to the Hausdorff dimension by:

\[ d_H = 2 - \alpha_H, \quad (2.9) \]

where \( \alpha_H \in (0, 1) \). Therefore a time series has a fractal dimension between 1 and 2. It is well known that the Hurst-Hausdorff exponent can measure persistence of the series. For persistent time series \( \alpha_H \in (0.5, 1) \) and anti-persistent when \( \alpha_H \in (0, 0.5) \). The case of \( \alpha_H = 0.5 \) corresponds to the so called Brownian motion.

2.1.2 Generating data with a specific fractal dimension

It is very useful to have a way to generate time series with a predetermined fractal dimension. One of the most important reasons is to test new methods. A simple method for generating monofractal time series is the Random Midpoint Displacement (RMD). The generating process starts with selecting two points as the first and last point of the series \( x_1 \) and \( x_N \), separated by distance (for time series this distance would be named “time”) \( L \). The second step is placing a random number generated from a normal distribution \( \mathcal{N} \left( \frac{x_1 + x_N}{2}, \sigma \right) \) halfway between them \( (\sigma \in \mathbb{R}_+ \setminus \{0\}) \). All following iterations consist of generating numbers in between those from the previous step. For each step \( n \) a number is generated from \( \mathcal{N} \left( \frac{x_i + x_j}{2}, \frac{\sigma}{2^n} \right) \) and placed in between \( x_i \) and \( x_j \), at a distance of \( \frac{L}{2^n} \) from both \( i \) and \( j \). \( \alpha_H \) has a visible impact on the generated series, it can be referred to as the “roughness” parameter. The plot of the artificial time series is very rough (see Figure 2.2) when \( \alpha_H \) is low (which corresponds to high \( d_H \)), for higher \( \alpha_H \) it gets smoother and eventually for \( \alpha_H = 0.9 \) it’s almost a straight line.

It is important to note, that the RMD algorithm achieves its assumed fractal dimension only when the process has an infinite number of steps, and hence the series is infinite. Therefore it is important to study “length effects” of various methods when working with real data. RMD generated time series will be used as base to see what properties should be expected in monofractal time series of specified length.
Figure 2.2: Examples of an RMD generated time series with Hurst-Hausdorff exponent $\alpha = 0.25, 0.5, 0.75$ respectively. All time series have the length of $N = 2^{16} = 65537$ points. The change in appearance of the series is clearly visible as the plot gets “smoother” with increasing $\alpha_H$. 11
2.2 Multifractals

The development of fractal analysis caused the need for creating a new, more complex, description of fractal phenomena. It seems that the monofractal approach is insufficient for explaining some of the encountered problems. Using techniques for monofractal time series analysis it was possible to find many interesting results (some mentioned in Chapter 2), but for many other problems a more exhaustive of the series is necessary. The two main problems of simple fractal analysis seem to be non-stationarity of the series and multifractality[14].

For many types of time series (including most, if not all, from the field of econophysics) the main concern was stationarity of the series. Most of the early methods were meant for stationary time series, where it is important to consider non-stationarity in series like stock market indices. Therefore new methods, like Wavelet Analysis [15], were created, and some older ones were modified. The most successful ones, were the modifications of Fluctuation Analysis [16]. The latter ones include DFA and DMA methods [5, 17]. Non-stationarity of time series also required for a less global approach. Since the evolution parameters of the series are not constant, a local approach seems to be more adequate. This idea led to development of the local-DFA [18, 19].

The second type of problem, as mentioned before, is multifractality. While the monofractal requires a single rule for every time scale, and for every area in the series, a multifractal is a structure with multiple scaling rules. Hence, it’s a generalisation of the monofractal, and it may be expressed by an equation analogous to Equation 2.1.

\[ F(s) \sim s^{\alpha_H(s)}, \]  
\[ (2.10) \]

where again, as in the monofractal case, \( s \) is denoting scale (defined for a problem at hand) and \( F \) is the statistical function depending on fluctuations. But, since \( \alpha_H \) is a function of scale for a multifractal object, it must be calculated in some other way. The most direct approach would be to try an approximation by expanding \( \alpha(s) \) in a Taylor series:

\[ \alpha(s) = \alpha(s_0) + \alpha'(s_0)\Delta s + \ldots. \]  
\[ (2.11) \]

One would want to take only a few first terms of this expansion, but this approach is only valid for small \( \Delta s \). However, since \( s \) denotes scale, that can be arbitrarily large. Because of that a different method should be used. Another approach is to notice that since \( F(s) \) is a statistical function, it’s possible that it will lose some crucial information required for a multifractal description. One way of reducing that effect is to artificially magnify small or large fluctuations.
and consider them separately. This approach is called the $q$-deformation, and can be easily put into an equation similar to Equation 2.8:

$$\langle |x(t + \tau) - x(t)|^q \rangle_{t \gg \tau} \sim \tau^{\alpha_H(q)}, \quad \tau > 0,$$

(2.12)

where $q \in \mathbb{R}$ is the deformation parameter. This approach is used in the Multi-fractal Detrended Fluctuation Analysis (MFDFA) [21].

At this point, there are two popular methods of description of multifractal time series: the Wavelet Transform Modulus Maxima method (WTMM) [20] and mentioned earlier MFDFA. The latter one will be described and used in this study due to its better performance for short sets of data [22].

With several scaling exponents, comes the obvious problem of the border between two regions that obey different power laws. Those crossover scales are subject of research [26, 27], as their proper localization and description is an important step in understanding of multifractal phenomena.

### 2.3 Selected methods for calculating the fractal dimension

Among many methods that are currently in use, only four of them will be described in this thesis. All of them have a significant meaning either for the fractal analysis in general, or for this study. The Rescaled Range Analysis [23] is the first method used for time series investigation. It is a great example of real life fractal analysis, although not working properly for non-stationary series. The Power Spectrum Analysis [24] is presented as an exceptionally easy method and one of the earliest. Detrended Fluctuation Analysis [5] is the base of many modifications and is showing great success in many fields. DFA is also the foundation of the last method – Multifractal Detrended Fluctuation Analysis [21] that will be used in this study, and is giving a way to describe multifractal phenomena.

#### 2.3.1 Rescaled Range Analysis

The Rescaled Range Analysis method was published in 1951 [23]. It was the first application of a fractal description to an engineering problem. It was used to determine the required capacity of a water reservoir in Africa to avoid floods. This method is based on previous observations of water levels. To calculate the Hurst-Hausdorff exponent (in the beginning it was named only after Hurst, but being later used as a fractal dimension equivalent Hausdorff’s name was added).
2.3. SELECTED METHODS FOR CALCULATING THE FRACTAL DIMENSION

The data is divided into non-overlapping segments of length of $s$. Therefore, one has $N_{\text{seg}} = \left\lfloor \frac{N}{s} \right\rfloor$ segments, where $N$ is the length of the entire series, and $\left\lfloor \kappa \right\rfloor$ denotes the largest integer not greater than $\kappa$. Mean from a segment is subtracted from every element of that segment and the result is integrated (integration in a discrete sense means a cumulative sum of the elements). It is worth noting that the data should describe changes of state, instead of the current state of the examined phenomena $x_i$ with $i = 1...N$. Thus

$$Y_k(j) = \sum_{i=1}^{j} (x_{ks+i} - \langle x \rangle_k), \quad (2.13)$$

where $\langle x \rangle_s$ denotes a mean over the segment $k$. This is followed by calculations of differences $R_k(s)$ between minimum and maximum values (ranges) and standard deviations for each segment $S_k(s)$. Respectively:

$$R_k(s) = \max_{j=1..s} Y_k(j) - \min_{j=1..s} Y_k(j), \quad S_k(s) = \left( \frac{1}{s} \sum_{j=1}^{s} Y_k^2(j) \right)^{\frac{1}{2}}. \quad (2.14)$$

Finally, the average over all segments is taken, and repeated for other $s$. One obtains the relation of the fluctuation function $F_{RS}(s)$ and the Hurst-Hausdorff exponent $\alpha_H$:

$$F_{RS}(s) = \frac{1}{N_{\text{seg}}} \sum_{k=1}^{N_{\text{seg}}} \frac{R_k(s)}{S_k(s)} \sim s^{\alpha_H} \quad \text{for} \quad s \gg 1. \quad (2.15)$$

2.3.2 Power Spectrum Analysis

This easy method was proposed by J. Geweke and S. Porter-Hudak in 1983 [24] for analyzing stationary time series. After calculating the power spectrum of the series, which is the squared magnitude of its Fourier transform. One can find that the spectrum obeys

$$\frac{1}{2\pi} \left| \hat{X}(\omega) \right|^2 = S(\omega) \sim \omega^{-\beta}, \quad (2.16)$$

where $\hat{X}$ is the Fourier transform of the time series. The $\beta$ exponent is a linear function of $\alpha_H$ [25, 28]:

$$\alpha_H = \frac{\beta - 1}{2} \quad \text{for} \quad \alpha_H \in (0.3, 0.9). \quad (2.17)$$

We have introduced three ways of describing a fractal: $\alpha_H, \beta, d_H$. We also presented some relations between the parameters that allow us to express one parameter in terms of the other two. It’s worth noting that there is a fourth,
commonly used, exponent for characterising fractals. It can be introduced by a
autocorrelation function $C(s)$ for a stationary time series $x(t)$ [29]:

$$
C(s) \sim \langle x(t)x(t+s) \rangle \sim s^{-\gamma},
$$

(2.18)

where $s$ is the time delay, and $\gamma$ the scaling exponent characterising the fractal.
Like for the previous ones, also the $\gamma$ exponent can be expressed by $\alpha_H$ [25, 30]:

$$
\alpha_H = 1 - \frac{\gamma}{2} \quad \text{for} \quad 0 < \gamma < 1.
$$

(2.19)

2.3.3 Detrended Fluctuation Analysis

Detrended Fluctuation Analysis (DFA) was first introduced to search for
correlations in DNA chains [5], since then this method has been in many fields
including physics [11], economics [6, 7, 8] and geophysics [31]. This method
consists of 5 steps:

1. Split the data of length $N$ into disjoint segments of length $s$, which gives
$N_{seg} = \lfloor \frac{N}{s} \rfloor$ segments. The data at the end of the series, that didn’t
fit into any segment, can be either discarded or included, for example by
copying the needed amount of data from the previous segment.

2. In each segment a polynomial function $y^r_k(j)$ of rank $r$ is fitted to the data,
and it is subtracted from the data (detrendisation)

$$
\tilde{X_k}(j) = X_k(j) - y^r_k(j).
$$

(2.20)
2.3. SELECTED METHODS FOR CALCULATING THE FRACTAL DIMENSION

![Graph](image)

Figure 2.4: Example log-log plot of the \( F(s) \) function for an RMD generated artificial time series with assumed \( \alpha_H = 0.7 \). Also shown is the fit estimating \( \alpha_H \) at 0.708 ± 0.03

3. Calculate mean square for each segment

\[
F^2(s, k) = \frac{1}{s} \sum_{j=1}^{s} \tilde{X}_k^2(j). \tag{2.21}
\]

4. Take square root from an average over all segments

\[
F(s) = \left( \frac{1}{N_{seg}} \sum_{k=1}^{N_{seg}} F^2(s, k) \right)^{1/2}. \tag{2.22}
\]

5. Calculate \( F(s) \) for other \( s \) to obtain the power law

\[
F(s) \sim s^{\alpha_{DFA}}. \tag{2.23}
\]

It can be proved that the \( \alpha_{DFA} \) coincides with \( \alpha_H \). This method to obtain the Hurst-Hausdorff exponent seems to be the most versatile and accurate of all currently available. Its usefulness has been noticed and many modifications of the detrending method have been proposed, for example: moving average detrending [17] and Fourier detrending [32].

2.3.4 Multifractal Detrended Fluctuation Analysis

Multifractal DFA is a more general approach to scaling exponents [21]. It is meant to find a set of exponents for different fluctuation scales. To achieve this a new parameter \( q \in \mathbb{R} \) is introduced. It’s purpose is to enlarge small or large
fluctuations for negative and positive $q$ respectively. MFDFA can be split into 6 steps:

1. Divide the data into disjoint segments of length $s$. To include the end of the data that might have been omitted (for some $s$) the division is made starting from the beginning and later from the end of the series. This procedure gives $2N_{seg} = 2\lfloor \frac{N}{2} \rfloor$ segments in total.

2. In each segment a polynomial function $y_k(j)$ of rank $r$ is fitted to the data, and it is subtracted from the data (detrendisation)

$$\hat{X}_k(j) = X_k(j) - y_k(j).$$

3. Calculate mean square for each segment

$$F^2(s,k) = \frac{1}{s} \sum_{j=1}^{s} \hat{X}_k^2(j).$$

4. Calculate the $q$-th order fluctuation function ($q \neq 0$)

$$F(s,q) = \left\{ \frac{1}{2N_{seg}} \sum_{k=1}^{2N_{seg}} \left[ F^2(s,k) \right]^{q/2} \right\}^{1/q}.$$ 

5. Find the $F(s,q)$ function for other $q$ and $s$ to obtain the relation

$$F(s,q) = s^{h(q)},$$

where $h(q)$ is the generalized Hurst-Hausdorff exponent, which behavior should reflect the scaling properties for various fluctuation magnitudes. It is important to notice the diverging exponent of the MFDFA method for $q \rightarrow 0$, to allow calculation at $q = 0$ a logarithmic averaging procedure is introduced

$$F(s,0) = \exp \left\{ \frac{1}{4N_{seg}} \sum_{k=1}^{2N_{seg}} \ln \left[ F^2(s,k) \right] \right\} \sim s^{h(0)}.$$ 

**Proof.** This can be easily shown with taking the limit $q \rightarrow 0$ and rewriting elements of the sum:

$$\lim_{q \rightarrow 0} \left\{ \frac{1}{2N_{seg}} \sum_{k=1}^{2N_{seg}} \left[ F^2(s,k) \right]^{q/2} - 1 + 1 \right\}^{1/q}.$$ 

Calculating the third element of the sum gives:

$$\lim_{q \rightarrow 0} \left\{ \frac{1}{2N_{seg}} \sum_{k=1}^{2N_{seg}} \left[ F^2(s,k) \right]^{q/2} - 1 + \frac{2N_{seg}}{2N_{seg}} \right\}^{1/q}. $$
2.3. SELECTED METHODS FOR CALCULATING THE FRACTAL DIMENSION

Using the identity:
\[ \lim_{n \to 0} a^n - 1 = \lim_{n \to 0} n \ln a \quad \text{for} \quad a > 0, \] (2.31)
one obtains:
\[ \lim_{q \to 0} \left\{ \frac{q}{4N_{seg}} \sum_{k=1}^{2N_{seg}} \ln \left[ F^2(s,k) \right] + 1 \right\}^{1/q}. \] (2.32)
Finally using:
\[ \lim_{n \to 0} (1 + an)^{1/n} = \exp a, \] (2.33)
one acquires Equation 2.28. 

This allows to obtain the behavior of \( F(s,q) \) for any place in its two dimensional domain. A useful representation of this data is the Legendre transform of the \( F \) function. Let us define:
\[ \tau(q) = qh(q) - 1. \] (2.34)
One defines also the Hölder exponent also called the singularity strength:
\[ \alpha = \frac{d\tau}{dq}, \] (2.35)
then the Legendre transform will take form:
\[ f(\alpha) = q\alpha - \tau, \] (2.36)
where \( f(\alpha) \) is the dimension of the parts that can be described by \( \alpha \) [33]. As proof of versatility of the method, recent research presented a generalization of this process into higher dimensions [34].
Chapter 3

Program for calculating the MF DFA

Determination of the fractal dimension can be an analytical problem. This is the case when one deals with deterministic fractals, or when analyzing rules for a probabilistic fractal. One can calculate the Hausdorff dimension as in 2.1.1, or approach it in a different manner, but still analyzing only the rules. When the rules are not known, only numerical methods of fractal dimension derivation are available. For the attempted multifractal description of financial markets, the MF DFA approach was selected, and implemented in C/C++ languages. C is mentioned because of the GNU Scientific Library [35], which was used for finding linear fits for the power laws in log-log scale, as well as two libraries from C Standard Library: math and stdio. From the C++ Standard Template library the vector structure was used extensively due to its dynamic properties. Because of the difficult task of finding the scaling range for each time series for all values of the parameter $q$, the procedure was implemented in a separate class for maximum control over this process. The MF DFA class is responsible for the Multifractal DFA procedure, to the point of finding the fluctuations for each pair $(s, q)$. It is important to note that the application is meant for a Linux environment, but simple changes should allow to run in under Windows® [36] (these changes are described in 3.3). The source code and all other needed files are present on the attached DVD.
3.1 Program structure

The application source consists of 5 files, accompanied by a Makefile and a mdfda.pro file (both have been created by qmake application, a part of the QT4 developer suite [37], which was used due to the plans for a graphical user interface to be created in the future). To allow fast correction of errors, two bash [38] scripts are present. They are used for generation of figures needed to control the scaling range as well those presenting the final result. Depending on the option selected when running the program, either all figures are created, or only the latter. All parts of the application were written and tested under Ubuntu 8.04.2 (2.6.24) [39], and compiled using g++ version 4.2 [40]. Figure 3.1 shows a diagram describing the general structure of the two main classes of the application.

MFDFA class allows calculation of the Multifractal DFA up to the moment of finding $F(s,q)$ for a specified range of $s$ and $q$. Since the scaling range of the power law is not a constant, obtaining the behavior of $h(q)$ is left for the fiter class.

Attributes of the MFDFA class:

- **out** - contains the final result of the calculation, each vector holds the value of $q$ and values of $F(s,q)$ for consecutive $s$ values,
- **pudelka** - created for each $s$ to hold all the segments of that length, each segment is held in a separate vector,
- **myInput** - holds the original data for creation of segments,
- **tabq** - vector containing all values of $q$ for which MFDFA should be calculated,
- **tabstd** - contains values of mean square (Equation 2.25) for each segment,
- **rozmiar** - holds the size of the original data,
- **poczatek** - value of smallest $s$ to be used for calculation,
- **koniec** - value of largest $s$ to be used for calculation,
- **coIle** - incrementation step for $s$.

Usage and intent of MFDFA methods:

- **MFDFA(&inputData, poczatek, coIle, koniec, &qtabelka)** - constructor of the class, requires 4 arguments:
CHAPTER 3. PROGRAM FOR CALCULATING THE MFDFA

**Fiter (fiter.h, fiter.cpp)**

- `igreki`: vector<double>
- `iksy`: vector<double>
- `wx`: vector<double>
- `wy`: vector<double>
- `rozmiar`: integer

```cpp
fiter(dane1:vector<double>, nowe2:vector<double>): constructor
fituj(dolny:double, gorny:double, q:double): double
```

**MFDA (MFDA.h, MFDA.cpp)**

- `out`: vector<vector<double>>
- `pudelka`: vector<vector<double>>
- `myInput`: vector<double>
- `tabelka`: vector<double>
- `rozmiar`: integer
- `poczatek`: integer
- `colle`: integer
- `koniec`: integer

```cpp
MFDA(inputData:&vector<double>, poczatek1:integer, colle1:integer, koniec1:integer, tabelka1:&vector<double>): constructor
Tworca(dlugosc:integer): integer
Detrend(dlugosc:integer, il_pud:integer): void
Calculate(): void
qstd(il_pud:integer, q:double): double
nstd(dlugosc:integer, il_pud:integer): void
```

---

Figure 3.1: Pseudo-UML diagram describing the framework of `fiter` and `MFDA` classes and file structure of the application.
3.1. PROGRAM STRUCTURE

- `std::vector<double> &inputData` - pointer to the original data in a vector
- `int poczatek1, int coIle1, int koniec1` - first value, incrementation and last value of `s`
- `std::vector<double> &qtabelka` - pointer to a vector with all values of `q` for which to perform calculation,

- `int Tworca(dlugosc)` - fills `pudelka` with segments of specified length (creates all `2N_{seg}`), returns number of segments, requires 1 argument:
  - `int dlugosc` - length of the segments

- `void Detrend(dlugosc, il_pud)` - calculates and subtracts the trend for each segment in `pudelka`, requires 2 arguments:
  - `int dlugosc` - length of the segments
  - `int il_pud` - number of segments

- `void Calculate()` - calculates the Multifractal DFA, requires no arguments,

- `double qstd(il_pud, q)` - calculates the `q`-th order fluctuation function (Equation 2.26) based on the `tabstd`, returns the `F(s,q)` value, requires 2 arguments:
  - `int il_pud` - number of segments
  - `double q` - value of `q` parameter

- `void nstd(dlugosc, il_pud)` - calculates the mean square for each segment (Equation 2.25), requires 2 arguments:
  - `int dlugosc` - length of segment
  - `int il_pud` - number of segments

To obtain the relation `F(s,q) = s^{h(q)}` and then to determine the behavior of `h(q)`, one needs to find the scaling range for all `q` in the analyzed region. Since there is trouble with defining an algorithmic way of selecting that range, it has to be set manually. Hence it is necessary to have a clear view of the `F(s,q)` function. This, and the fitting process itself is handled by the fiter class. In the following class description the `F` function refers to an abstract function for which the linear fit is calculated. Attributes of the fiter class include:
CHAPTER 3. PROGRAM FOR CALCULATING THE MF DFA

- std::vector<double> igreki - holds $F$ function values for consecutive arguments from iksy

- std::vector<double> iksy - holds arguments of the $F$ function (all available points)

- std::vector<double> wx - contains of the arguments of the $F$ function within a specified range

- std::vector<double> wy - values of the $F$ function for the arguments from wx

- int rozmiar - length of the full data (length of iksy)

Usage and intent of fiter methods:

- fiter(dane1,dane2) - constructor of the class, requires 2 arguments:
  - std::vector<double> dane1 - arguments of the $F$ function
  - std::vector<double> dane2 - values of the $F$ function

- double fituj(dolny,gorny,q) - calculates the fit for the specified range of data and saves it in the current directory. The data is saved in files fity* and fityc* for the data in the specified range and full data respectively. The * in the filename is substituted by the current $q$ parameter. Returns only the slope of the fitted linear function, requires 3 arguments:
  - double dolny - lower limit of the fitting range $dolny \in [0,1]$
  - double gorny - upper limit of the fitting range $gorny \in [0,1]$
  - double q - $q$ parameter (required for proper file naming)

There are two additional purposes of main.cpp (apart from only creating instances of the described classes). One is to calculate the earlier mentioned representation of the MF DFA output (Equation 2.34). The second is to manage files with the relation $F(s,q) = s^{h(q)}$ for each $q$, and to create plots of that data (if the user selects this option). Creation of the plots is handled by gnuplot [41]. The needed directory structure, for saving data and plots, can be changed by editing skr.true.sh or skr.sh. The first file is used when figures of the scaling ranges are needed. The second file is used when only plots of the final result are required.
3.2 Usage and possibilities of the application

The application must be run with 9 parameters:

- Filename of the input data - it has to be given in a single column as integers or floats (in scientific notation).

- Filename of the output file - the output consists of 5 columns, that contain $q$, $h(q)$, $\tau(q)$ (introduced in Equation 2.34), $\alpha$ (from Equation 2.35) and $f(\alpha)$ (from Equation 2.36).

- Lower limit of the scaling range for positive $q$ - value denotes how long is the shortest segment in the relation $F(s,q) \sim s^{h(q)}$.

- Upper limit of the scaling range for positive $q$.

- Lower limit of the scaling range for negative $q$.

- Upper limit of the scaling range for negative $q$.

- Whether to create plots of scaling ranges - true when value of the parameter equals 1, false when 0.

- Lower limit of the $q$ parameter.

- Upper limit of the $q$ parameter.

It is important to note how the limits of the scaling range should be provided. Because box sizes may vary, it is important to describe box size by a dimensionless value. Therefore it is expressed by percentage of the entire set of the segments. Hence, setting limits for positive $q$ as: $0.01 \ tense 0.25$, means that the first percent of the segments (counting from the shortest), and the last 75 percent won’t be included in calculation of $F(s,q) \sim s^{h(q)}$.

The lower boundary of segment sizes is automatically selected as 0.01 of the length of the series, or 5 depending on which is larger. The upper limit is set to 0.25 of the length of the series. The $F(s,q)$ function is calculated with the $s$ changing by one, and $q$ with a step of 0.1.

As mentioned earlier an appropriate directory structure is needed for proper saving of the generated figures. Files with data for the figures, are saved in directory `fits`, and the figures themselves are saved in subfolder `wykresy`.

Four plots are created after all of the calculations. They are all named appropriately to the content, all files are preceded with `mf_wynik_` and followed by names of the axes (first $x$, then $y$). Plotted functions represent:

1. $h \rightarrow q$
CHAPTER 3. PROGRAM FOR CALCULATING THE MF DFA

2. $q \rightarrow \tau$
3. $q \rightarrow \alpha$
4. $\alpha \rightarrow f(\alpha)$

3.3 Further development

One of the most important modifications that will be made, is to allow use of the application under Windows®. For a quick way to compile, run the program, and get the needed results, one has to delete lines from 106 to 121 in main.cpp (this will remove all figure creating and file moving). To have full usability, one must rewrite the bash scripts for Windows® command line, as well as to modify main.cpp in the lines mentioned earlier, substituting Linux console commands with proper Windows® commands. Since gnuplot is available for Windows®, it is possible to have required figures (given that appropriate scripts have been modified). One other simple modification to the bash scripts (or Windows® command line scripts) is to alter the directory structure needed for the application. To achieve that, it is only required to change directories in lines 119 and 120 of the main.cpp.

Future research is likely to involve longer sets of data. Therefore it’s very important to allow fast analysis of long series. This can be achieved in two ways (both of which should be included in later version of the application):

1. Multi-threading - although the class structure allows for easy multi-threading, it is still necessary to rewrite some parts of the code to allow parallel calculations.

2. Box size scaling - since the scaling range is sought after in a log-log scale. The calculation of the fluctuation function for all large segment sizes, can be unnecessary. Since only the linear approximation of the function is calculated, some of the $F(s,q)$ function values can be discarded (for large segments). An obvious approach is to choose segment size in an exponential progression, to minimize calculation of excess data points.

The final modification is an addition of a graphical user interface. Its main purpose would be easier control of the scaling range, and elimination of the need for gnuplot. Therefore, apart from a mechanism for managing parameters of the calculation, a plotting method, with a temporary linear fitting procedure, should be implemented for a fully functional GUI.
3.3. FURTHER DEVELOPMENT
Chapter 4

Analyzed data and initial problems

4.1 Data sets

The main goal of this research is to analyze the multifractal structure of two stock market indices, and to find possible differences between developed and developing markets. The Standard & Poor’s 500 index represents a developed market, composed from companies from NASDAQ and New York Stock Exchange (NYSE). The developing market is represented by the Warszawski Indeks Giedowy (WIG) from the Warsaw Stock Exchange (WSE). However analysis of those indices must be backed up by a study of an artificial monofractal signal. The artificial time series study is important to see how length effects manifest themselves in the MF DFA analysis. Since there are no commonly used methods of generating artificial multifractal series, a monofractal approach has to be sufficient.
4.1. DATA SETS

![S&P 500 Value Chart](image1.png)

Figure 4.1: Plot of the available data for the S&P500 (10.12.1927 - 3.09.2008)

![WIG Value Chart](image2.png)

Figure 4.2: Plot of the available data for the WIG (16.04.1991 - 10.10.2008)

The artificial data for the analysis has the length of $2^{14} = 16384$ points, which is comparable to the real data. For the S&P 500 index a data set of 17856 points was selected. This covers all of the available closing values from 30th of December 1927 to 3rd of September 2008. For the WIG the date span is from 16th of April 1991 to 10th of October 2008, which gives 3925 points that also note the closing value of the index. Both data sets were obtained from the Internet. The S&P 500 data was downloaded from the financial subpage of Yahoo.com [43] and WIG from a stock market subpage of Wirtualna Polska [44].
4.2 Analysis of scaling range

At this point it is necessary to describe the early problems with the MFDFA procedure for real data. This is needed to properly understand the reasons why, in the next chapter, the analysis is conducted in specific areas, and to comprehend the underlying cause for this kind of approach. One of the key issues in the MFDFA scheme is the selection of a proper scaling range for a given set of data. As there are no widely used algorithms for that task, therefore this has to be done manually. It can be easily done for an artificial series, even for short series the power law is preserved for the entire range of segment sizes. Unfortunately the real data proves to be very troublesome. Both, short and long real series, show quick deterioration of the power law. Thus obtaining the values \( h(q) \) is problematic and has a wide margin of error.

Figure 4.3: An example plot of the \( F(s, q) \sim s^{h(q)} \) relation for a series with assumed \( \alpha_H \) equal 0.7 for \( q = 13 \) (top) and \( q = -13 \) (bottom)

Figure 4.4: An example plot of the \( F(s, q) \sim s^{h(q)} \) relation for a series with assumed \( \alpha_H \) equal 0.7 for \( q = -5 \) (top), \( q = 0 \) (middle) and \( q = 5 \) (bottom)
4.2. ANALYSIS OF SCALING RANGE

This deviation from the power law (or even lack of it) is worse for high $q$ where large fluctuations are dominant. For very small $q$ this effect can be observed but still allows proper calculation. Those effects were observed for both WIG and S&P 500. Again, the scaling in artificial series seems to be independent of $q$. Since there is a possibility of setting the scaling range separately for positive and negative $q$, one might consider setting them differently. This however leads very quickly, and for all kinds of analyzed data, to discontinuity of $h(q)$, even with a slight difference in the scaling ranges.

![Figure 4.5](image1.png)

Figure 4.5: Example plots of the $F(s, q) \sim s^{h(q)}$ relation for the entire S&P 500 series with $q = 13$

![Figure 4.6](image2.png)

Figure 4.6: Example plots of the $F(s, q) \sim s^{h(q)}$ relation for the entire S&P 500 series with $q = -5$ (top), $q = 0$ (middle) and $q = 5$ (bottom).
CHAPTER 4. ANALYZED DATA AND INITIAL PROBLEMS

Taking this problems into account, all calculations were made with the same scaling ranges for positive and negative values of $q$. The deformation parameter $q$ was always set for the range $(-15, 15)$ in all calculations.

**Note** More figures representing the $F(s, q)$ function for various $q$ have been provided on the attached DVD.
4.2. ANALYSIS OF SCALING RANGE
Results

Several attempts were made to improve the scaling range for real data. The first approach was to test the properties of returns instead of the original (integrated) data. Those however were worse, compared to the original signal. The scaling range was indeed wider, but very small changes in the selection of the scaling range, led to a dramatically different \( h(q) \) function. Knowing that the scaling properties of stock market time series are distorted during market crashes, the second approach was to cut out all extreme events from the time series. This was done by finding high absolute values of returns, and then removing all surrounding points that in 5 day windows gave a 10% or higher returns. This procedure did manage to decrease the fluctuations visible on Figure 4.5, but that still wasn’t enough to allow easy selection of scaling range for high \( q \). The last, at that stage, attempt to find the reason for such bad scaling properties was to remove the visible exponential trend. This however showed no significant impact on the results. Therefore the results described in the next two sections were obtained from unchanged data.

5.1 Random midpoint displacement data

As mentioned earlier, it is very important to have an idea of the “noise” that would come from a short monofractal signal. Three RMD generated time series of length 16384 were analyzed, and the results of MFDFA confirm their pre-set \( \alpha_H \) as the peak of the singularity spectrum is near the assumed Hurst (or in this representation the Hölder) exponent.

**Proof.** Let us rewrite the expression for \( \alpha \) by calculating it explicitly:

\[
\alpha = \frac{d\tau}{dq} = h(q) + q\frac{dh(q)}{dq}. \tag{5.1}
\]
5.1. RANDOM MIDPOINT DISPLACEMENT DATA

Knowing that $\frac{dh(q)}{dq}$ equals 0 for a monofractal ($h(q) = const$) it is clear that all values of $\alpha$ will be equal. Since calculating $h(2)$ is simply the DFA procedure, all values reflect the $\alpha_H$. This proof only applies to a “perfect” (in the case at hand – infinite) monofractal. However, for finite monofractal series the peak of the singularity spectrum still coincides with $\alpha_H$.

![Graph](image.png)

Figure 5.1: $h(q)$ function of 3 artificial RMD series with assumed $\alpha_H = 0.3$, 0.5 and 0.7 respectively. Their Legendre transformation is shown on Figure 5.2

These are only sample realizations of an RMD series, and since the method should be used in a statistical context (calculations performed on a large set of realizations of an RMD algorithm, and averaged), these results should be only treated as a sample “background noise” of a short monofractal. As described in Subsection 2.1.2, accuracy of setting the $\alpha_H$ in RMD series depends on length of the series, therefore it’s important to see if width of the spectrum is getting narrower with increasing length. This effect would indicate that a single scaling exponent is dominant in the entire series, and the “noise” is fading. The table below shows width of the singularity spectra for several lengths of series, and seems to confirm a slow narrowing of the spectra.

The narrowing effect was observed for series representing a random walk, and was analyzed for longer sets and from a statistical point of view [22]. Statistical analysis of this problem (for RMD and Fourier Transform Filtering method series) will be the topic of future studies.
CHAPTER 5. RESULTS

<table>
<thead>
<tr>
<th>Series length</th>
<th>Limits of spectrum</th>
<th>Width of spectrum</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2^{12}$</td>
<td>0.27 – 0.62</td>
<td>0.35</td>
</tr>
<tr>
<td>$2^{13}$</td>
<td>0.30 – 0.60</td>
<td>0.30</td>
</tr>
<tr>
<td>$2^{14}$</td>
<td>0.33 – 0.67</td>
<td>0.34</td>
</tr>
<tr>
<td>$2^{15}$</td>
<td>0.31 – 0.62</td>
<td>0.31</td>
</tr>
<tr>
<td>$2^{16}$</td>
<td>0.48 – 0.69</td>
<td>0.15</td>
</tr>
<tr>
<td>$2^{17}$</td>
<td>0.29 – 0.57</td>
<td>0.28</td>
</tr>
</tbody>
</table>

Table 5.1: Dependency between length and singularity spectrum width of an artificial RMD generated series.

This is a good moment to quote a fragment of the paper by J.W. Kantelhardt [21]. It explains why one should expect the $h(q)$ function to be monotonically decreasing. This fact will prove to be very important in the analysis of the MFDFA results.

Usually the large fluctuations are characterized by a smaller scaling exponent $h(q)$ for multifractal series than the small fluctuations. This can be understood from the following arguments: For the maximum scale $s = N$ the fluctuation function $F_q(s)$ is independent of $q$, since the sum in Eq. (4) (Equation 2.26 in this thesis) runs over only two identical segments ($N_s \equiv \lfloor N/s \rfloor = 1$). For smaller scales $s \ll N$ the averaging procedure runs over several segments, and the average value $F_q(s)$ will be dominated by the $F^2(v, s)$ from the segments with small (large) fluctuations if $q < 0$ ($q > 0$). Thus, for $s \ll N$, $F_q(s)$ with $q < 0$ will be smaller than $F_q(s)$ with $q > 0$, while both become equal for $s = N$. Hence, if we assume an homogeneous scaling behavior of $F_q(s)$ following Eq. (5) (Equation 2.27 in this thesis), the slope $h(q)$ in a loglog plot of $F_q(s)$ with $q < 0$ versus $s$ must be larger than the corresponding slope for $F_q(s)$ with $q > 0$. Thus, $h(q)$ for $q < 0$ will usually be larger than $h(q)$ for $q > 0$.

In Figure 5.2 the singularity spectrum of the signal with assumed $\alpha_H = 0.7$ has a peculiar shape. Its origin is the non-monotonic behaviour of the $h(q)$ function. All three $h(q)$ functions are shown in Figure 5.1. For RMD series the non-monotonic behaviour is stronger in series with higher assumed $\alpha_H$. The lack of monotonic behaviour is further amplified in real data, due to the problem of selecting the scaling range.

35
5.2. **LONG-TERM DATA ANALYSIS**

![Singularity spectra of 3 artificial RMD series with assumed $\alpha_H = 0.3, 0.5$ and $0.7$ respectively. The peaks of the spectra reflect the pre-set $\alpha_H$ in terms of the Hölder exponent.](image_url)

**Figure 5.2:** Singularity spectra of 3 artificial RMD series with assumed $\alpha_H = 0.3, 0.5$ and $0.7$ respectively. The peaks of the spectra reflect the pre-set $\alpha_H$ in terms of the Hölder exponent.

### 5.2 Long-term data analysis

The second step of the calculations is to observe the multifractal structure of the entire real time series. As mentioned earlier this calculation is performed on unchanged data, and on the entire available time span. We will call this analysis as “long-term data analysis”.

#### 5.2.1 Original data analysis

The singularity spectra for the S&P 500 and its shuffle, are a very good example of effects of long and short scaling range. The “twist” on the top part of Figure 5.4 is the effect of a problematic scaling range selection for $q$ around 3. Where the shuffled data spectrum reflects the monotonic behaviour of its $h(q)$ function. This is produced by a long scaling range for $q$ within the range $(-15, 15)$. Positioning of the S&P500 spectrum suggests a complex structure, with high $\alpha$ values (which probably reflects the constant growth in the first 60% percent of the series).
Figure 5.3: $h(q)$ function of the entire S&P 500 time series and its shuffle

Figure 5.4: Singularity spectra of the entire S&P 500 time series and its shuffle

Figure 5.5: $h(q)$ function of the entire WIG time series and its shuffle
5.2. LONG-TERM DATA ANALYSIS

Figure 5.6: Singularity spectra of the entire WIG time series and its shuffle

When analyzing WIG, proper selection of the scaling range is problematic for both positive and negative $q$. Combined with the “quick” deterioration of the scaling range it produces the peculiar shape of the spectrum. The maximum of the spectrum is in the vicinity of 0.7, and with the shift of the entire $f(\alpha)$ function towards higher values of $\alpha$, one should expect some degree of persistence in the series (again the high $\alpha$ values might be explained by small fluctuations around the main trend in the first part of the series).

As expected, in both cases the maximum of the spectrum for shuffled data lies close to $\alpha = 0.5$.

5.2.2 Analysis of positive & negative returns

Once we know that the description of the whole index is proving somewhat difficult, it is time to divide the series into pieces the easiest way possible. It is a well known fact that the distribution of returns for stock markets is not symmetric [45, 46], but shifted toward the positive sign. This fact may indicate different scaling properties of positive and negative returns. The calculations were made for integrated series of positive (and negative) returns. It means that we took positive returns $x_k^+$ and created a series:

$$Y_k = \sum_{i=1}^{k} x_i^+.$$  \hspace{1cm} (5.2)

An analogous procedure was used to create the series of negative returns.
CHAPTER 5. RESULTS

Figure 5.7: $h(q)$ function of the positive and negative returns from the S&P500 for all available data.

Figure 5.8: Singularity spectra of the positive and negative returns from the S&P500 for all available data.

Figure 5.9: $h(q)$ function of the positive and negative returns from the WIG for all available data.
5.2. LONG-TERM DATA ANALYSIS

![Singularity spectra of the positive and negative returns from the WIG for all available data](image)

Figure 5.10: Singularity spectra of the positive and negative returns from the WIG for all available data

The maximum of the negative spectrum is to the left of the positive maximum for both markets. For the S&P 500 the whole spectrum is in the range of higher $\alpha$. The “twisted” top part of all spectra is the effect extremely quick deterioration of the scaling range for series with very high persistence. Even the DFA method must be used with caution for series with high persistence [33], thus the regions of very high (and low) $q$’s must be treated with utmost caution. This is also the reason for the very high values of $\alpha$ for S&P500.

In an attempt to improve the scaling properties of the real data (which are especially bad for returns) some modifications were made to the calculation procedure. The first modification was to remove an exponential trend from the data. The second one was to alter the MFDFA scheme by using a higher degree polynomial (second and third degree polynomials were tested). None of those methods managed to improve the scaling range. And for the third degree polynomial detrending, a decrease of the scaling range was observed.

![F(s,q) function for the positive returns of S&P 500 for q = 5 and a third degree polynomial as the detrending function](image)

Figure 5.11: $F(s, q)$ function for the positive returns of S&P 500 for $q = 5$ and a third degree polynomial as the detrending function (MFDFA3)
CHAPTER 5. RESULTS

5.2.3 Analysis of weekly returns

Let us now create a series from weekly returns for both indices. To avoid data shortening, we will use a special procedure. Let us differentiate the series and divide it into overlapping segments of length 5. The sum of all elements, of a given segment, denotes a weekly return. If that return is positive all elements of that segment are moved to the “positive series”. An analogous approach is used to create the “negative series”. Data at the end of the series that is too short to form a segment of length 5 is discarded. Both series are then integrated (see Equation 5.2). The integrated series are then used to calculate the MFDFA.

Figure 5.12: $h(q)$ functions of the positive and negative weekly trends from the S&P500. All available data was used to create the “returns series”.

Figure 5.13: Singularity spectra of the positive and negative weekly trends from the S&P500. All available data was used to create the “returns series”.

41
5.2. LONG-TERM DATA ANALYSIS

Figure 5.14: $h(q)$ functions of the positive and negative weekly trends from the WIG. All available data was used to create the “returns series”.

Figure 5.15: Singularity spectra of the positive and negative weekly trends from the WIG. All available data was used to create the “returns series”.

Again, as in the returns analysis, values of the Hölder exponents ($\alpha$) are higher for S&P 500, and the shift towards higher values of $\alpha$ of the negative spectrum is visible for both indices. This shows that a more general approach to the multifractality of positive and negative returns gives a similar result. One should also note that the method (in other words the window on which the return is calculated) used to create the returns series influences the the shift of the spectrum.

Results from this and the previous chapter confirm earlier research [42] concerning different multifractal structure of positive and negative returns. They also suggest an influence of the time scale of the multifractal structure of returns.
CHAPTER 5. RESULTS

However, one must keep in mind the scaling range problems. These problems introduce a high level of uncertainty. Therefore the results of this part of the analysis must be treated with caution.

5.3 Analysis of bearish and bullish periods

Recent multifractal research of the DAX index, showed that the spectrum for negative returns is shifted towards higher values of $\alpha$ when compared to positive returns spectrum [42]. However, those calculations were made for a shorter period of time and for intraday data. A similar result was obtained in Subsections 5.2.2 and 5.2.3. However, due to bad scaling properties it should not be considered as completely accurate. The mentioned paper also contains a study of bearish and bullish phases of the markets, and finds similar shifting of spectra. An analogous study was performed for the WIG and S&P500, with the difference of analyzing daily closing prices, instead of intraday data. It is necessary to note that that the bullish phase for the S&P500 was modified to improve the scaling range. These modifications will be described in detail in Section 5.5.

![Figure 5.16: Parts of WIG and S&P500 consisting of a bearish and bullish phase. The point of the trend change that splits the series into the two phases is marked as a vertical line (spectra of each is calculated separately).](image)

The spectra show the expected shifting of the bearish phase towards higher values of $\alpha$, and with much better scaling ranges, these result should be considered as more reliable. Comparison of the markets gives the opposite results then those from the previous sections. Both spectra of the S&P500 are now shifted to the left when compared to the WIG. This casts further doubt on the validity of the results when analyzing series that consist of data from long periods of
5.4. Analysis of automatically divided data

With a great improvement in the scaling range that was achieved in the previous section, came the obvious question: Is this improvement an effect of careful division into phases on the market, or is this effect present in every part of the series? To find the answer the S&P500 time series was divided into 5, 10 and 20 fragments of equal length, and the WIG in 5 and 10 fragments (division
CHAPTER 5. RESULTS

into 20 fragments would produce too short series).

Figure 5.19: The top figure represents an example of “good” data for which the calculations are performed. The bottom plot shows an example of the $F(s,q)$ function of this data. The bottom right figure shows the singularity spectrum.

Most of the fragments showed an improvement in the scaling range width. This enhancement was more visible for shorter series, therefore almost all fragments of the WIG show a good scaling range width.

Figure 5.20: The top figure represents an example of “bad” data for which the calculations are performed. On the bottom left an example $F(s,q)$ function of this data showing very poor scaling properties. The bottom right figure shows the “twisted” singularity spectrum, a result of a non-monotonic $h(q)$ function.

Some of the series presented no improvement, or even further deterioration,
5.5 IMPROVEMENT OF SCALING PROPERTIES BY DIVISION AND “CUTTING OUT” SPECIAL EVENTS

of the $F(s, q) \sim s^{h(q)}$ power law. When a long segment (one resulting from a division into 5 parts) showed this degeneration of scaling range, its sub-segments (a result of dividing into 10 parts), either all showed good scaling properties or at least one contained an extreme event. These sub-segments containing abrupt event also showed bad scaling properties. This led to a conclusion that the width (or even existence of) the scaling range depends on two factors:

- containing an abrupt event (a market crash or a very quick increase of the index value) - which means an occurrence of a very rare event,

- containing two distinctly different phases (regimes) of the series - for example contains the transition point between a bearish and a bullish phase.

5.5 Improvement of scaling properties by division and “cutting out” special events

Cutting out extreme events was used with no success for the entire series (see Chapter 4). However, results from the previous section seem to suggest that removing the extreme events might improve the scaling range for short series. This means that those events distort the multifractal image of the series. This is the reason why real long-time financial time series can’t be properly described by MFDFA. The second factor that determines the width of the scaling range, the “phase change” is the key to the failed attempt of only removing extreme events from the entire series. Since the series is composed of many phases, adding the “division approach” provides better results.

5.5.1 Short term example - WIG and S&P500

The method of removing special events should be tested on a fragment created in the previous chapter. Therefore one of the “bad” fragments has been selected and altered (the series is differentiated, the event with its close surrounding removed and the resulting series is integrated) to remove the abrupt event.
CHAPTER 5. RESULTS

Figure 5.21: On the left the analyzed signal, the crossed out region shows the extreme event that was removed. On the right the singularity spectra of the original part of the S&P500 and the modified one.

The visible improvement in the spectrum on Figure 5.21, is the result of the scaling range widening and therefore better estimation of the $h(q)$ function.

Figure 5.22: On the left, the analyzed signal, the crossed region shows the different regime. On the right, the singularity spectra of the original WIG fragment and the modified one.
5.5. IMPROVEMENT OF SCALING PROPERTIES BY DIVISION AND “CUTTING OUT” SPECIAL EVENTS

Two examples of the division method effects are visible on Figures 5.22 and 5.23. Again the improvement in the shape of the spectrum (due to the expected monotonic behaviour of the $h(q)$ function) reflects the widening of the scaling range. Only on Figure 5.23 a small “tail” is visible on the left end of the spectrum, which is the effect of bad scaling properties for very high $q$.

Further removing of extreme events should provide an even smoother $h(q)$ function, which removes the “edges” of the singularity spectrum.

5.5.2 Long term example - S&P500

It is time to try the “division and removal” approach for the entire series. In Chapter 4 we mentioned a failed attempt to improve scaling properties of long-time real series by removing extreme events. However, the S&P500 series is long enough for the division approach to give much better results, and provide a good description of the multifractal structure. One can easily divide the entire S&P500 series into two distinctly different regions. This division is shown on Figure 5.24.
CHAPTER 5. RESULTS

Figure 5.24: The entire S&P500 time series. The vertical line denotes the border between two regimes. One can see the two phases of low and high oscillations around the trend.

Analysis of the spectra and scaling range width show great improvement for both parts of the series. Widths of the spectra are very similar, but their positioning shows higher $\alpha$ values for the first part of the S&P500 index.

Figure 5.25: Singularity spectra of the first part of the S&P500 and for the entire signal
5.5. IMPROVEMENT OF SCALING PROPERTIES BY DIVISION AND “CUTTING OUT” SPECIAL EVENTS

Figure 5.26: Singularity spectra of the second part of the S&P500 and for the entire signal

One can notice the “edge” in the spectrum of the S&P500 second part for $\alpha = 0.58$ (Figure 5.26). This effect can be removed by finding an event that stands out from the ordinary behaviour of the series.

Figure 5.27: Second half of the S&P500 index with the highlighted region of an extreme event

Once we delete the highlighted region seen in Figure 5.27, a smooth monotonically decreasing $h(q)$ function is produced which in turn removes the “edge” from the singularity spectrum. The fact that the “edge” seems to be a representation of the deleted event is very interesting. This fact finds confirmation
in the position of the anomaly, since when the part with a very rapid fall (and therefore high persistence) is removed, the elevation of the spectrum in the high $\alpha$ region is flattened.

Figure 5.28: Singularity spectra of the original second part of the S&P500 and the altered version (with the removed event highlighted in Figure 5.27)
5.5. IMPROVEMENT OF SCALING PROPERTIES BY DIVISION AND “CUTTING OUT” SPECIAL EVENTS
Chapter 6

Conclusions

It is clear that more research is required to properly explore the significant multifractal noise generated by short RMD series. Most importantly a statistical approach to width of the singularity spectrum is needed. This should be done for a variety of lengths of the series, to estimate the rate of the spectrum narrowing, or in other words to assess the multifractal noise as a function of length.

At this point it seems that series that cover a long timespan ("real world" time) are very difficult to describe in the language of MFDFA. However, if one splits the series into several parts, a better description should be possible due to a wider scaling range. Which, in turn would give a better estimation of $h(q)$.

As shown in Section 5.5, not only the length of the series (their "real world" timespan) is a problem. Extreme events also seem to have a strong impact on the scaling range width. These events include market crashes and one-time fluctuations (e.g. large peaks “standing out” of the usual series behaviour).

High persistence in the series, especially in artificial series composed of returns, causes problems with the scaling range as well. Neither removing the dominant exponential trend, nor using higher degree polynomials gave any improvement. But despite that, it has been shown that WIG and S&P500 appear to have the same property of a shifted to the right spectrum for negative returns as noticed in [42]. For the bearish and bullish phases the shift is also present for both markets, which again confirms the results from [42].

The most important result is that division of time series into parts of distinctly different regimes (e.g. small and large fluctuations around the main trend), and removing any abrupt events improves the scaling range significantly. However, it requires further research to find if the division works because of difference of scale and regime, or is there a time factor involved (i.e. non-
stationarity).

It is highly probable that a time factor should be present when calculations are done for such long periods of time. One should expect different behaviour of the market in the beginning of the XX\textsuperscript{th} century in comparison with the present data. Therefore a more local approach might be more effective in such cases, and one should look for scaling exponents that include the time as a variable i.e. $h(q, t)$. 
BIBLIOGRAPHY

Bibliography


BIBLIOGRAPHY


[38] http://www.gnu.org/software/bash/
BIBLIOGRAPHY

[41] http://www.gnuplot.info/


Appendix A

Abstract

This thesis is an attempt to describe the multifractal structure of long-term financial time series. Two indices (WIG and S&P 500) have been analyzed with the MFDFA method. After a detailed description of the method we introduce its implementation in C/C++. We describe the problems encountered when analyzing long-term data and show the asymmetry in the multifractal structure of returns. Analysis of the series fragments suggests that non-stationarity plays a significant role in distorting the scaling range. We finally show that dividing a long-time series into fragments and removing extreme events improves the scaling range width.