

Scaling properties of financial time series

Diploma thesis

by

David Schreier

(181592)

submitted to

Prof. Dr. Wolfgang Härdle



in partial fulfilment of the requirements for the degree

Diplom Kaufmann

Berlin, 2007-12-06

Acknowledgment

I thank Professor Dr. Wolfgang Härdle for giving me the opportunity and motivation to write this thesis. Special thanks go to Enzo Giacomini who gave excellent guidance all the time.

After all I am indebted to my whole family, especially in remember of Prof. Haim Bakai and my uncle Rechtsanwalt Peter Schreier for giving me intellectual advice.

Declaration of Authorship

I hereby confirm that I have authored this diploma thesis and independently and without use of others than the indicated sources. All passages which are literally or in general matter taken out of publication or other sources are marked as such.

Berlin, November, 2007-12-06

David Schreier

Abstract

This thesis will first criticize standard financial theory. The focus will be on return distributions, efficient market hypothesis and the independence of returns. Part two gives the intuition to look at markets in a different view. Namely the one proposed by B. Mandelbrot who has shown that nature itself can often be described with fractals. Then the relationship between fractal power laws and scaling will be explained.

The main part focuses on the estimation of the tail index as a scaling parameter with the help of three different techniques: 1. OLS regression on a log-log plot, 2. Hill estimator and 3. the alpha exponent within the stable distribution.

In the last section a different power law exponent will be estimated to test for long memory effects (i.e. nonperiodical cycles) in return distributions. The last section gives a conclusion.

Keywords: efficient market hypothesis, fractal, self-similarity, scaling, power law, Hurst, fractional Brownian motion

Content

1. Introduction.....	11
1.1 Assumptions of Modern Finance.....	12
1.1.1 Return distributions in Finance.....	12
1.1.2 Dependence properties of returns.....	15
1.2 Models of financial markets.....	16
1.2.1 Random walk.....	16
1.2.2 The efficient market hypothesis.....	17
1.2.2.1 Theoretical challenges to the EMH.....	18
1.2.2.2 Criticism by behavioural finance literature.....	19
1.2.2.3 Empirical challenges to the EMH.....	21
1.3 Criticism of standard financial theory.....	26
2. The Fractal View.....	27
2.1 Dimension of fractals.....	27
2.2 Examples of fractals.....	28
2.2.1 Brownian motion as a fractal	28
2.2.2 Sierpinski triangle.....	29
2.3 Relation of scaling, self similarity and power laws.....	31
2.4 Fractal property of finance.....	31
3. Scaling.....	33
3.1 Scaling as a universal law.....	33
3.1.1 Pareto's power law of income distribution.....	33
3.1.2 Zipf's law.....	36
3.1.3 Scaling behaviour of company growth.....	37
3.2 The tail exponent and stable distributions.....	39
3.2.1 Numerical properties of stable distributions.....	41
3.2.2 Characteristic function representation.....	43
3.2.3 Stable distribution computation.....	45

3.2.4 Estimation of the tail and of stable parameters.....	45
3.2.4.1 Direct tail estimation.....	46
3.2.4.2 The Hill Estimator.....	49
3.2.4.3 Estimation by using the characteristic function.....	53
3.2.5 Scaling of Volume and other power laws in finance.....	57
4. Long Memory and R/S Analysis.....	59
4.1 Long memory of time series.....	59
4.1.2 Properties of long memory processes.....	60
4.2 R/S Analysis.....	64
4.2.1 Estimation of the Hurst exponent.....	67
4.3 Fractional Brownian Motion.....	68
4.4 Multiscaling of financial time series.....	70
5. Conclusion.....	71
Appendix A : Xplore Code of Sierpinski Triangle.....	72
Appendix B : Power law mathematics.....	73
Appendix C : Multifractal model.....	75
References.....	77

List of tables

1.1	Probability for large events in a Gaussian world.....	13
1.2	Value of Kurtosis for six Dax stocks.....	14
3.1	Parameters for stable distribution, DAX tick data.....	56
3.2	Parameters for Gaussian distribution, DAX tick data.....	56
4.1	Estimation of H and Lo's alternative for six Dax stocks.....	65

List of figures

1.1 Autocorrelation function for Telekom.....	14
1.2 Value function from Prospect Theory.....	19
1.3 Chart of S & P 500 compared to Dividends.....	21
1.4 Chart of P/E ratios compared to interest rates.....	22
1.5 Returns of buying winners and losers	24
2.2 Self similarity of Brownian motion.....	28
2.3 Sierpinski triangle.....	29
3.1 Income distribution according to Pareto.....	35
3.2 Zipf's law applied to web documents.....	36
3.3 Density of company growth rates.....	38
3.4 Scaling behaviour of company growth.....	39
3.5 Convergence of stable distribution to power law.....	40
3.6 Cauchy, Levy and Gaussian.....	41
3.7 Importance of sample size for log-log OLS.....	45
3.8 log-log OLS regression for Dax tick data.....	-
3.11 January 2006, April 2006, Mai 2006, January 2007.....	47

3.12 Overestimation of Hill estimator in a small sample.....	49
3.13 Hill estimator for Dax tick data.....	-
3.16 January 2006, April 2006, Mai 2006, January 2007.....	50
3.17 stable parameter estimation for Dax tick data.....	-
3.20 January 2006, April 2006, Mai 2006, January 2007.....	54
3.21 OLS of IBM volume on log-log plot.....	57
4.1 Hurts estimation of H 1951.....	64
4.2 ACF for simulated FGN with $H=0,35$	67
4.3 Simulated time series for FBM and FGN with $H=0,35$	67
4.4 ACF for simulated FGN with $H=0,75$	68
4.6 Simulated time series for FBM and FGN with $H=0,75$	68

Abbreviations

ACF	Autocorrelation Function
AR()	Autoregressive Process of order (...)
ARMA	Autoregressive Moving Average Process
B/M	Book to market ratio
ARCH	Autoregressive Conditional Heteroscedasticity
APV	Adjusted Present Value Method
CAPM	Capital Asset Pricing Model
DCF	Discounted Cash Flow Method
EMH	Efficient Market Hypothesis
EUREX	European Exchange
FIGARCH	Fractionally Integrated GARCH
FBM	Fractional Brownian Motion
FGN	Fractional Gaussian Noise
GARCH	General Autoregressive Conditional Heteroscedasticity
LTCM	Long Term Capital Managed Fund
NASDAQ	National Association of Securities Dealers Automated
NYSE	New York Stock Exchange
OLS	Ordinary Least Squares
R/S	Rescaled Range Analysis
S & P	Standard & Poors

1. Introduction

Modern financial theory is of very heavy use nowadays. It is thought in almost every business school. In practice people make decisions based on it. They use it to decide if to invest in a project or not.

Corporate finance models like DCF or APV rely on risk factors given by the CAPM. In option pricing, the famous Black and Scholes model is of heavy use. But what are the underlying assumptions and how correct are they?

Louis Bachelier introduced in (1900) the first „modern” model for stock prices. In his PhD thesis he proposed that prices follow a random walk. Which is interpreted as that investors cannot profit from a given strategy. Prices are unpredictable. One implication of this is that markets are supposed to be efficient.

But what is the explanation for crashes like that of wall street 1929 and 1987 or the Asian crises 1997 or the internet bubble in 2000. An another example which is often cited is the Russian default on government bonds. The LTCM hedge fund who had about 200 billion dollar under exposure and a capital base of only 4,8 billion dollar almost collapsed after the Russian crises. The interesting aspect here is that two of the leading financial theorists were managing this fund. Namely Myron Scholes and Robert C. Merton. Both received the Nobel price in economics in 1997 for there well known option pricing formula. Only with the help of the Federal Reserve the fund did not collapse (Sornette, 2003).

Also the stock market crash of 1987 with 20 % happened without any fundamental news. It has been the biggest decrease of the U.S. stock market in

a single day so far. After events happened analysts have reasonable explanations of why it happened but were not able to predict.

Is such a drop compatible with rational behaving agents? It seems that human behaviour is more irregular than standard financial models assume.

The behavioural finance literature has several examples where the market or individuals fail to behave as the standard fundamental analysis suggests.

The next two introductory sections look closer at statistical and economical reasoning's the standard theory has to offer and where the theory might fail.

1.1 Assumptions of modern finance

1.1.1 Return distributions in finance

In almost all popular models return distributions are assumed to be Gaussian. In Markowitz's portfolio theory risk is measured with respect to the mean and the variance. But if returns do not follow the normal, not only standard portfolio theory would be useless also the CAPM and hence all corporate finance models based on beta as a measure of risk.

To test if the mean and the variance do describe return distributions properly we can measure the skewness and kurtosis (i.e. the third and fourth moment). Skewness is defined as

$$S = E \left[\frac{(y - \mu_y)^3}{\sigma_y^3} \right] \quad (1.1)$$

Kurtosis as :

$$K = E \left[\frac{(y - \mu_y)^4}{\sigma_y^4} \right] \quad (1.2)$$

The estimators are defined as

$$\hat{S} = \frac{1}{T\hat{\sigma}_y^3} \sum_{t=1}^T (y_t - \hat{\mu}_y)^3 \quad (1.3)$$

for the skewness and

$$\hat{K} = \frac{1}{T\hat{\sigma}_y^4} \sum_{t=1}^T (y_t - \hat{\mu}_y)^4 \quad (1.4)$$

for kurtosis.

A joint test for normality is the well known Jarque-Bera procedure. It tests for skewness and excess kurtosis at the same time. The test is

$$\frac{T}{6} \hat{S}^2 + \frac{T}{24} (\hat{K} - 3)^2 \sim \chi^2(2) \quad (1.5)$$

Empirical tests have shown that most financial time series exhibit no skewness but they do have kurtosis greater than 3. Which means they have fat tails (also known as leptokurtosis). The probability for large values is higher than under the normal distribution. A stylized fact of financial time series is that the shorter the time horizons the more excess Kurtosis we have (Cont, 2001).

To give an impression for the probability of extreme events, Table 1.1 shows the probability for some values in the Gaussian bell shaped world.

X	Probability _{>}	One in N events	Calendar waiting time
1	0.317	3	3 days
2	0.045	22	1 month
3	0.0027	370	1.5 year
4	6.3×10^{-5}	15,787	63 years
5	5.7×10^{-7}	1.7×10^6	7 millenia
6	2.0×10^{-9}	5.1×10^8	2 million years
7	2.6×10^{-12}	3.9×10^{11}	1562 million years
8	1.2×10^{-15}	8.0×10^{14}	3 trillion years
9	2.3×10^{-19}	4.4×10^{18}	17,721 trillion years
10	1.5×10^{-23}	6.6×10^{22}	260 million trillion years

Table 1.1: X is the number times the standard deviation for some value in the Gaussian world. Next, the probability for the given value of X is shown. The last column translates the probability in calendar time (days in this example) one has to wait to experience such a return. Source.: Sornette (2003)

Here a market return of more than 4% should be observed only once in 63 years. In fact, the market drop October 19, 1987 should never happened.

The theoretical question is that: are large market drops outliers or do we face a distribution where “large” deviations are within the usual range? An alternative family of distribution is described in part three. These do exhibit fat tails in a natural way. In table 1.2 the Kurtosis is tested for six stocks and shows high values, an indication that the normal distribution is not a good candidate to describe the distribution of returns.

Stock	Kurtosis
Telekom	5,63
SAP	8,05
Allianz	7,15
Pfizer	5,50
INTEL	8,55
Microsoft	8,57

Table 1.2: Estimation of Kurtosis for 3 stocks of DAX and 3 of Dow Jones indicating that the tails are fatter than those of the normal distribution.(daily returns).

1.1.2 Dependence properties of returns

It is well established that return of financial returns do not exhibit significant autocorrelation (Cont, 2001). This means that the price increments are uncorrelated and one cannot predict prices (which will be more discussed in the next section). However, important is that the ACF of absolute and squared dies out in hyperbolic way which gives rise to some nonlinear dependence. Could it be that very distant observations have some influence of current ones? This property known as *long memory* is discussed and tested in section for.

Figure 1.1 shows the typical picture of the ACF for absolute or squared returns.

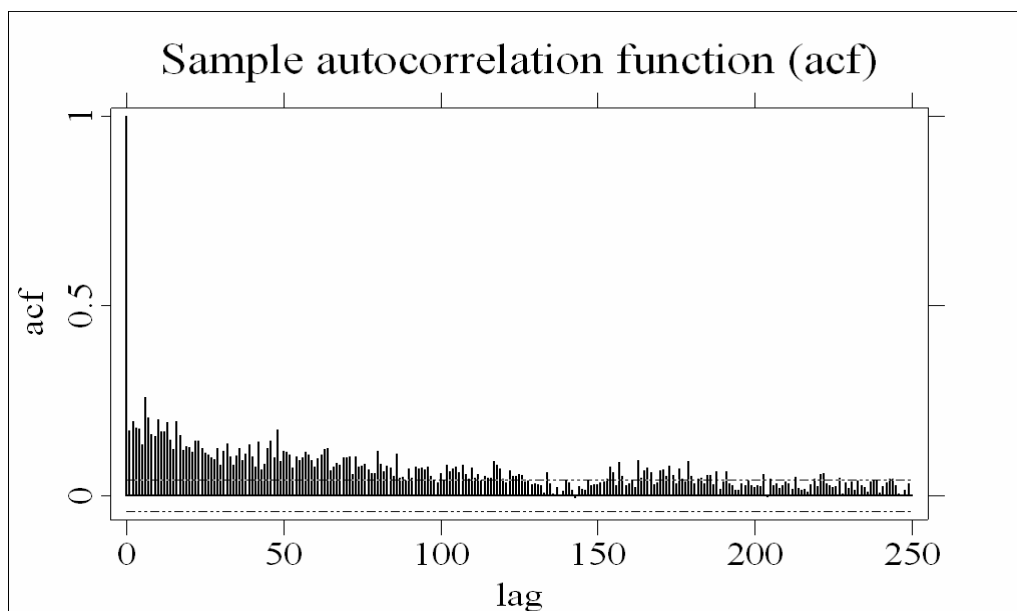


Figure 1.1: ACF for Deutsche Telekom, showing the typical hyperbolic decay in contrast to exponential decay inherent in most models

An another stylized fact of financial time series is the phenomenon of *volatility clustering* and is well studied and resulted in the invention of ARCH and GARCH models. Large price deviations are more likely to be followed by large price deviations.

Another empirical finding contradicting the independence assumption of returns is the existence of *drawdowns* (Sornette, 2003). These are defined as persistent decreases in prices over consecutive days. Hence, it is the cumulative loss from the last maximum to the next minimum. The persistence of such *drawdowns* is not measured by return distributions because they do not recognize the relative position of losses. Important here is to note that these dependence structures do only appear at special times when we have a few large losses in a row. The probability to have three consecutive losses of 10 % in an independent Gaussian world would be 10^{-9} . However, for example the Dow Jones experienced seven *drawdowns* with cumulative losses of more than 15 % and the highest with 30,7 %. To reject independence in returns completely cannot be justified by these facts!

1.2 Models of financial markets

This section shows the main theoretical concepts from an economist point of view regarding the behaviour of financial markets. It is better to begin with the random walk theory because the efficient market hypothesis is supposed to be a result of the former.

1.2.1 Random walk

A variable X_t follows a random walk with a drift δ , if

$$X_{t+1} = \delta + X_t + \varepsilon_{t+1} \quad (1.6)$$

with ε_{t+1} as an identically and independently distributed random variable.

The efficient market hypothesis is supposed to be a logical result of a random walk. It will be stressed in the next section that we have to distinguish between

predictability and the correctness of prices and hence between predictability and market efficiency.

1.2.2 The efficient market hypothesis

Market efficiency is still the hottest debated topic in finance. M.C. Jensen (1978) stated “there is no other proposition in economics which has more solid empirical evidences supporting it than the efficient market hypothesis”. It means that prices do reflect all relevant information. On average, it is not possible to make abnormal profits with any given strategy. Fama who defends the efficient market states in his review article (1991) states that there are only slight deviations. It has become a dogma since.

To show the EMH formally it is helpful to the concept of the martingale.

Mathematically speaking, prices follow a martingale if we assume that agents are risk neutral, they all have the same discount factor $\rho \left(\delta = \frac{1}{1+\rho} \right)$ and have access to the same information Ω_t , (Cuthbertson, 1996).

An arbitrage free price would satisfy:

$$p_t = \frac{1}{1+\rho} E[p_{t+1} + d_{t+1} | \Omega_t] \quad (1.7)$$

which means that the current price is equal to the discounted expected value of the dividend, d_{t+1} , plus the expected resale price in the next period, conditional on information I_t . After transformation we have

$$E \frac{[p_{t+1} + d_{t+1} | \Omega_t] - p_t}{p_t} = \rho \quad (1.8)$$

or

$$E\left[\frac{p_{t+1} + d_{t+1} - p_t}{p_t} - \rho \mid \Omega_t\right] = \quad (1.9)$$

For high-frequency data we have that $d_{t+1} \approx 0$ and $\rho \approx 0$, which leads to

$$p_t = E[p_{t+1} \mid \Omega_t] \quad (1.10)$$

1.2.2.1 Theoretical challenges to the EMH

In the following, a few theoretical drawbacks of the efficient market hypothesis are shown. First of all the *no-trade-theorem* says that any offer that deviates from previous prices reveals private information. Other agents would be reluctant to accept the offer because they would lose from trading with a presumably better informed agent. The price would adjust and incorporate the information without trade taking place, (Cuthbertson 1996). So even if we have asymmetric information, it is not sufficient to stimulate trade. This argument stays in absolute contrast to the immense trading we see nowadays in financial markets.

The second theoretical drawback is known as the *Grossman-Stiglitz-Paradox* (Grossman Stiglitz, 1980). If the market is in equilibrium and hence all arbitrage opportunities are eliminated is it possible that the market can be always in equilibrium? No. Arbitrageurs then make no return from their costly information gathering activity.

This leads us to think about arbitrage. It is a very important argument in every analysis of markets. All modern finance models build on the assumption of the possibility of perfect arbitrage. As shown by Shleifer and Vishny (1997) arbitrage is limited. Not every asset has a perfect substitute. Arbitrage is risky. The S & P 500 Index Future does have a perfect substitute, namely the underlying stocks of the index. But what is the substitute for GM, Chrysler?

What if mispricing becomes even more severe? Noise traders could bring the price even more from its fundamental value. This risk is called *noise trader risk*. Shleifer (2000) shows a recent example. In 1997 the S & P 500 Index reached a very high value compared to its historical price earnings ratio. Federal Reserve chairman Alan Greenspan called it in 1997 a “irrational exuberance”. An arbitrageur should short the S & P 500 in this situation. The problem was that the situation became even more worse. The arbitrageur would have lost 33,4 percent and 28,6 percent the next year.

1.2.2.2 Criticism by behavioural finance literature

How do investors behave? Is it according to models like CAPM? The growing behavioural finance literature pointed out so many examples of people not acting in a rational way. As Fischer Black (1986) described it, many investors trade on noise rather than information. They follow the advice of financial Gurus, do not diversify correctly, sell winning stocks and hold on losing stocks with increasing tax liabilities (Odean, 1998). They hold expensively managed funds although these funds do not even generate there fees (Carhart, 1997). Nobel Psychologists Kahneman and Tversky have shown that investors deviate from maxims of economic rationality in a systematic way. People have difficulties with simple probability reasoning. *Prospect theory* states that people do not asses risky gambles as to achieve the highest possible outcome. They look at gains and losses in respect to some reference point which varies from situation to situation. This psychological bias can be shown by a loss function which is steeper than the gain function. Shleifer (2000) argues that this could be a possible explanation for the equity premium puzzle (Mehra and Prescott, 1985), the fact that stocks earned 7 percent on a yearly basis and bonds only 1 percent. Figure 1.2 shows the function from the original paper by Kahnemann and Tversky.

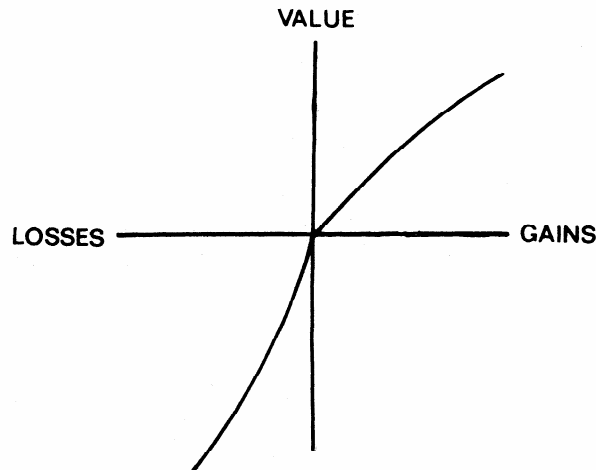


Figure 1.2: The value function of Kahneman and Tverskys „Prospect Theory” Source: Kahneman and Tversky (1979)

For explaining the puzzle Benartzi and Thaler (1993) add to loss aversion the fact that people exploit *mental accounting*. It will be explained by a simple example. Suppose an agent has the following utility function

$$U(x) = \begin{cases} x & x \geq 0 \\ 2,5x & x < 0, \end{cases}$$

If he is asked to do a bet, in case of winning he will receive 200 with probability 50 percent and in case of losing he must pay 100 with probability 50 percent. P. Samuelson asked this question his colleague who wouldn't accept this bet.

With two bets the colleague would have an positive expected value (400,0,25;100,0,50;-200,0,25). The crucial point is that this applies only if he would not evaluate his bet after the first session. For the stock market suppose we have a risky asset with an expected return of 7 percent a year and a standard deviation of 20 percent. On the other hand we have the riskless bond with a return of just 1 percent a year. This together with loss aversion is called *Myopic Loss Aversion* and is an attractive explanation why people invest in bonds in such an environment. But it fails to take into account that stocks are much more risky than seen by standard risk measures!

There are many more examples in which people do deviate from Bayesian rationality. The argument for the efficient market hypothesis lies in that these heuristics cancel each other out. But as shown by the theories from Kahneman and Tversky people tend to behave in same direction! We cannot assume that these mistakes are uncorrelated. Investors herd!

All these criticism of the efficient market hypothesis from the behavioural finance literature should make us sensible for questioning the underlying assumptions of modern finance and for finding new ways to model financial markets. As Richard Thaler has put it in his survey (2002) „Directly testing the validity of a model’s assumptions is not common practice in economics, perhaps because of Milton Friedman’s influential argument that one should evaluate theories based on the validity of their predictions rather than the validity of their assumptions. Whether or not this is sound scientific practice, we note that much of debate over the past twenty years has occurred precisely because the evidence has not been consistent with the theories, so it may be a good time to start worrying about the assumptions. If a theorist wants to claim that fact X can be explained by behaviour Y, it seems prudent to check whether people actually do Y”.

1.2.2.2 Empirical Challenges to the EMH

There are also a lot of empirical facts showing that the efficient market hypothesis is not true. One of the first important works was Shiller’s finding that stock prices are much more volatile than justified by dividends. With the efficient market hypothesis stock prices should only change when there are some news since the hypothesis states that prices are correct at all times! The volatility of stock prices shouldn’t be much higher than of the dividends. Figure 1.3 shows the big discrepancy between 1881 and 2003. For markets to be efficient the stock price should jump around the present value of dividends not in such dramatic way. Short recessions cannot be the explanation for this large volatility.

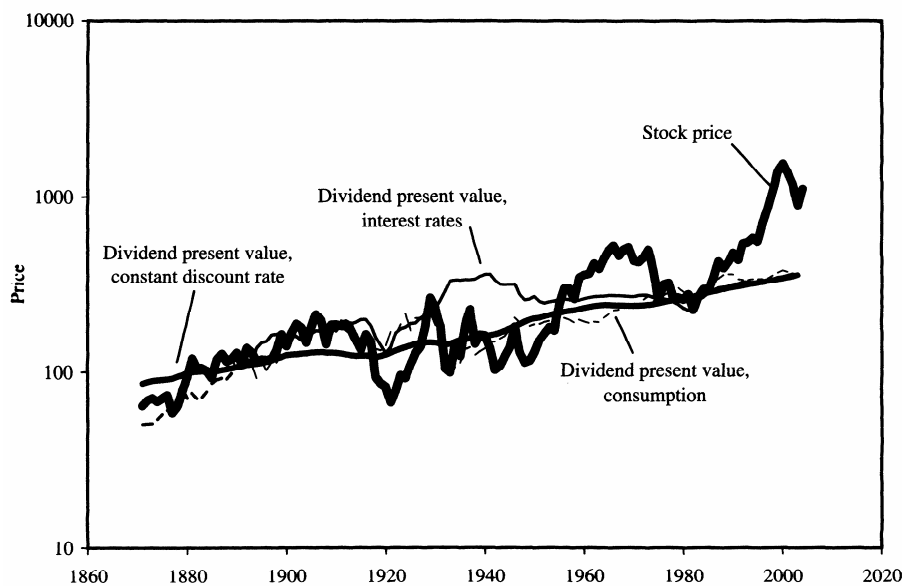


Figure 1.3: S & P 500 Index, 1881-2004 (heavy irregular curve), and present values, 1881-2003, of real dividends calculated by three different discount rates. Source: Shiller (2004)

The point is that it is true that forecasting stock prices is a very difficult task, but this does not mean that prices reflect the underlying fundamental value. A random walk must not imply correct prices. Shiller (2004) argues, the fact that we cannot predict day to day changes does not mean that we cannot predict any change. As described by the noise trader models, arbitrage is risky and smart investors are not able to bring prices down to fundamental value at every point in time. Shiller received hard critique, notably from nobel laureate Robert C. Merton who also managed the LTCM Hedge Fund described in the Introduction.

It is very important to emphasize that the efficient market hypothesis does not only mean that one cannot make excess returns given some Information set. It implies also that prices are right at every point in time. The Nasdaq high of 5000 in 2000 and also other indices around the world is a remarkable example of prices deviating from fundamentals in a extraordinary way. There were

companies with skyrocket P/E ratios. Figure 1.4 shows this for the whole US market. In peaks P/E ratios reached levels one cannot justify by the fundamental valuation formula where the price is just the sum of all discounted dividends.

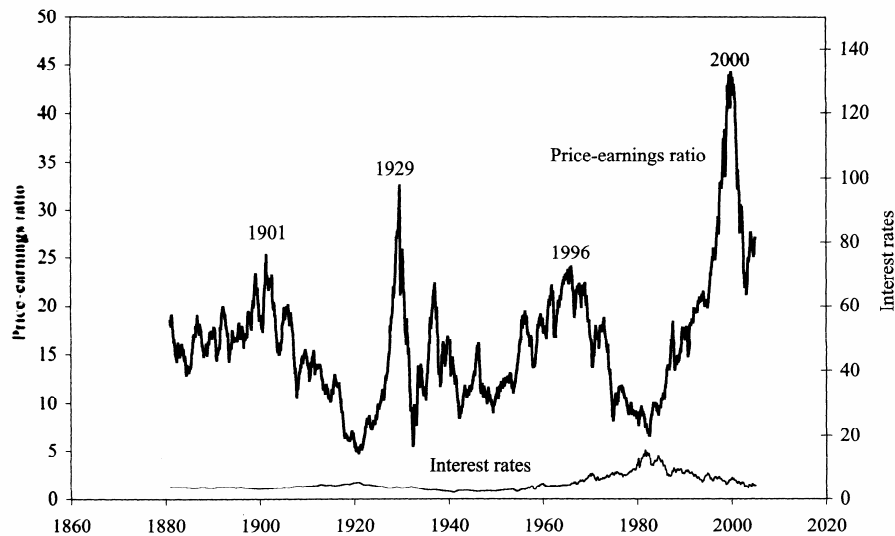


Figure 1.4: Price-earnings ratio, monthly, January 1881 to January 2005. Numerator: real (inflation-corrected) S &P Price Index, January. Denominator: moving average over preceding ten years of real S & P Composite earnings. Years of peaks are indicated. Interest rate is the yield of long term U.S. government bonds January 1881 to January 2005. Source: Shiller (2004)

Thaler (2002) finds that statements like „prices are right” and there is „no free lunch” are not equivalent. Both are true in an efficient market, but the „no free lunch” condition holds also in an inefficient market. That prices are away from fundamental value does not mean that we can predict anything or that there are any excess risk-adjusted average return available.

Thus,

„prices are right” \Rightarrow „no free lunch”

but

„no free lunch” \nRightarrow „prices are right”

Still many researcher point out the fact that fund managers do not generate excess returns is strong evidence for market efficiency and conclude that prices are right. The arbitrage argument with fundamental risk and noise trader risk does provide an explanation why this is not true. Also as shown in Figure 1.4 it seems very strange to argue that the market values assets always by its fundamental values.

Empirically there are also some other findings contradicting the efficient market hypothesis. In the literature they are called *anomalies*. For example the *size premium*, meaning that small cap stocks earn a higher return than justified by some equilibrium model (e.g. CAPM). Fama and French (1992) document that the average return on small cap stocks have a return 0,74 percent higher than the largest decile in terms of market capitalization.

Also well known is the study from De Bondt and Thaler (1985). They rank all stocks traded on NYSE by their prior three-year cumulative return and form „winner” and „loser” portfolios. Afterwards average returns are measured subsequent to there formation. The result is that loser portfolios outperform winner portfolios by 8 percent a year. Figure 1.5 shows this remarkable finding.

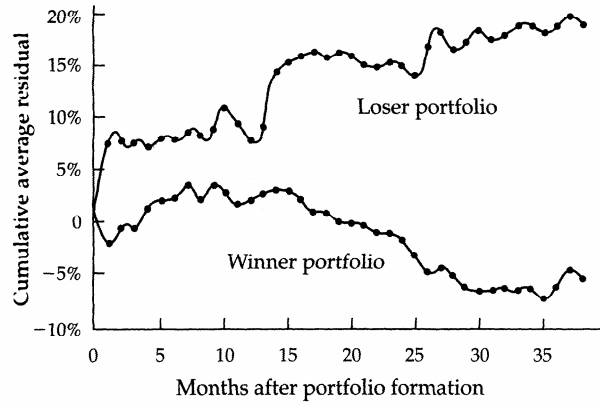


Figure 1.5: Cumulative average residuals for winner and loser portfolios of 35 stock (1-36 months into the test period) Source: De Bondt and Thaler (1985)

In relation to the finding of Thaler and De Bondt is that of Jagadeh and Titman (1993). They found that when looking on a shorter time periods the opposite is true. Winners continue to win and losers continue to perform poorly which is now known as *momentum*.

The last factor, which should not influence the risk premium, found statistically significant, is the *book-to-market ratio*. Fama and French (1992) grouped stocks of NYSE and NASDAQ into deciles based on their B/M ratios. The highest decile produced a monthly return 1,53 percent higher than the lowest decile. This suggests that value strategies do outperform growth strategies.

As a result the following regression is used to study abnormal returns, e.g. Carhart (1997) the performance of mutual fund managers.

$$(R_{it} - r_t) = \alpha_i + \beta_{1i}(R_{mt} - r_t) + \beta_{2i}SMB_t + \beta_{3i}HML_t + \beta_{4i}PRIYR_t + \varepsilon_{it} \quad (1.11)$$

It basically says that the return over the risk free interest rate r_t a mutual fund manager i earns can be explained by the CAPM beta (β_{1i}), a small cap (β_{2i}), market to book (β_{3i}) and a momentum factor (β_{4i}). Only the return corrected by these risk factors is what he has generated by his “strategy”. The excess return is measured by $\hat{\alpha}_i$.

1.3 Criticism of Standard financial theory

The efficient market hypothesis implies that one only has to hold the market portfolio and does not need to care about future upcoming events, since the investor always earns his appropriate return with respect to risk. All news are quickly incorporated in prices. If this were true, why do we see so much actively managed funds? Stock picking would be a useless activity. Why is there so much trading around the world without a comparable amount of upcoming relevant news?

In corporate finance, it is very important to have meaningful discount factors. The rate of return is a measure of the opportunity costs of funds corrected for risk. It is used to discount cash flows from physical investment projects. If stock prices are very low compared to its true value then the corporate treasurer will accept projects that he otherwise would have not.

In the case of new share issues there is no gain in delaying a project in hope that financing conditions will improve. Under the efficient market hypothesis the current price is the correct price and the project is correctly incorporated in price. If the stock price is irrationally low then projects are rejected that would be accepted under efficient markets.

To sum up, we have at least five assumptions which do not apply in the real world: 1.the assumption of rationality, 2.efficient markets, 3.the assumption of normality, 4. independence and 5. perfect arbitrage.

Thus, the discussion so far should lead us to view financial markets in a different way. One alternative is the one of *fractal*, *scaling* and *self similarity* explained in section two and then applied to finance in section three and for.

2. The fractal view

The discussion in the first part suggests that the description of financial markets by models of modern finance seems to be not very accurate. A view that is not based on any economic assumption comes from Benoit Mandelbrot. In his book „The fractal geometry of nature” he describes that Euclidian geometry describes spheres, cones and circles in a nice way. But nature is not smooth. Mountains are not like cones, Coastlines not like circles (Mandelbrot 1982). The world is much more complex. Hence, we need other tools to describe properties of nature.

The main concept which drives the fractal view is *self-similarity*. For example the mammalian lung consists of the main branch trachea and then divides into sub branches and these two divide further (Peters 1994). At each branch generation the average diameter decreases according to a *power law*. But within a generation we have randomness. This is the key concept of self similarity, global determinism (average branch size) and local randomness (individual branch size).

A tree also has qualitatively self-similar branches, on the other hand every branch is unique. On different scales the branches look the same.

In physics complex systems such as turbulence are described by *power laws*.

2.1 Dimension of fractals

Another key to fractals is dimension. For a smooth curve the length $L(r)$ is given by the product of the number N of straight-line elements of length r needed to step along the curve from one end to the other and the length is r :

$L(r) = N \cdot r$. When the step size r goes to zero, $L(r)$ approaches a finite limit L (Schroeder, 1991).

With fractals the product $N \cdot r$ goes to infinity as r goes to zero. Asymptotically this divergence behaves like a *power law*. Thus, there is a exponent $D_H > 1$ so that the product $N \cdot r^{D_H}$ stays finite. If the exponent is smaller than D_H the product diverges to infinity and for larger values it tends to zero. The explained exponent is called the *Hausdorff dimension*. So we have

$$D_H = \lim_{r \rightarrow 0} \frac{\log N}{\log(1/r)} \quad (2.1)$$

With fractals the dimension lies somewhere between 1 and 2.

2.2 Examples of fractals

2.2.1 Brownian motion as a fractal

The often used Brownian motion as a model of the behaviour of stock market prices can also be seen as a statistically *self-similar* process (Schroeder, 1991). This is illustrated in figure 2.1. Here we have a typical Brownian motion of a particle between points A and B (A). Between point A and B the particle does not move in a straight line which is shown in (B). The particles motion is shown 100 times faster and the result is magnified 10 diameters. Now does the particle move in a straight line between C and D? No. If we look 100 times more often between C and D we get the same result as with points A and B. We can do the same iteration once again and would get the same result. In general if we increase the spatial resolution by a factor of $1/r$ we get $N(r) \sim 1/r^2$ more pieces to cover. Thus, the Hausdorff dimension is given by

$$D_H = \frac{\log N(r)}{\log(1/r)} = 2 \quad (2.2)$$

The scaling range of Brownian motion is between powers of 10^8 to 1 covered by a continuum of intermediate scales.

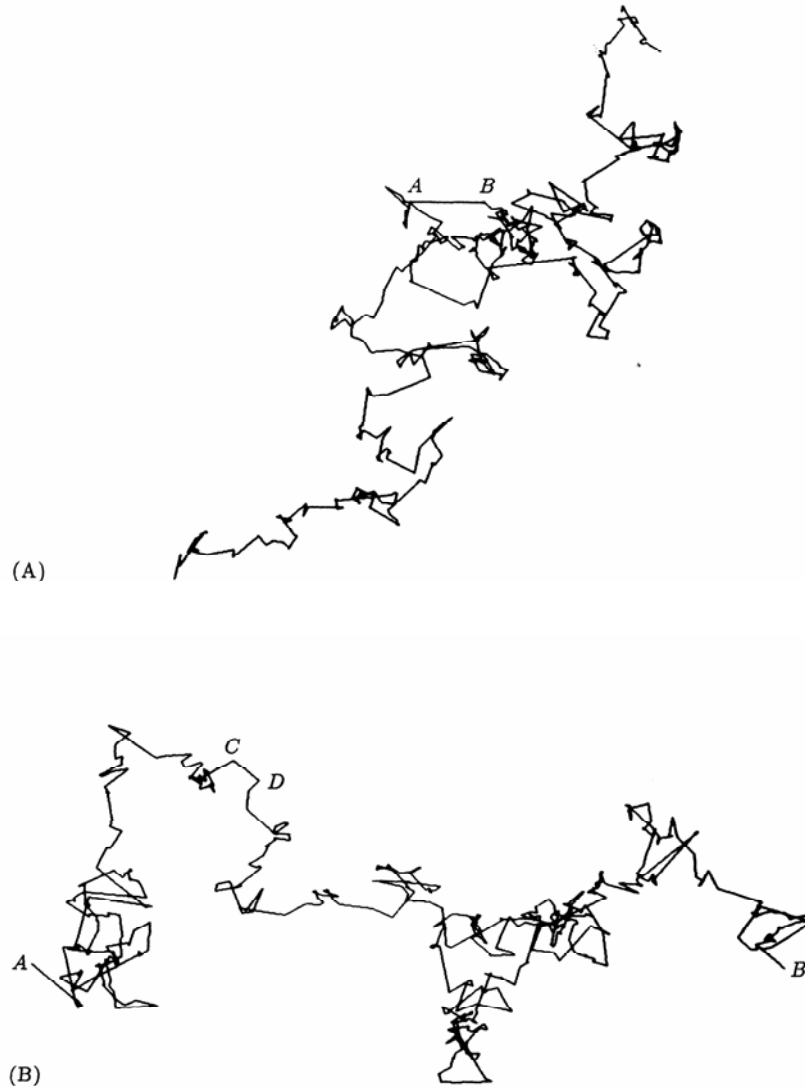


Figure 2.1: (A) Brownian motion. (B) The segment AB of the Brownian motion in part A sampled 100 times more frequently and magnified 10 times. Source.: Schroeder (1991).

2.2.2 The Sierpinski triangle

In this section a geometrical fractal will be shown. The algorithm behind is rather simple. We have a triangle. Then we choose one point within this triangle. Next, we choose one corner of the triangle, move half away and draw the next point. This is done 50000 times. What comes out is a *prima facie* example of self similarity. If one looks closer at the triangle you have the same structure. The same sort of triangle. But the process behind is random.

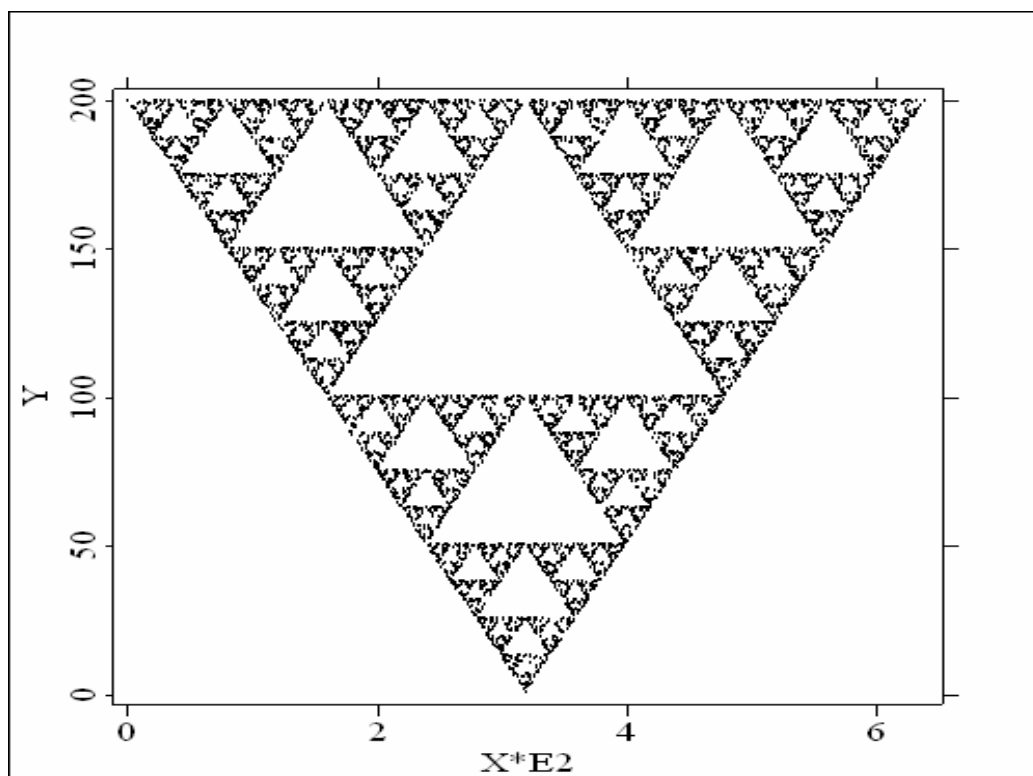


Figure 2.3: Sierpinski triangle, an example of a fractal which displays self similarity, local randomness and global determinism (also a property of chaos theory).

Figure 2.3 shows the triangle computed in Xplore. The code is provided in Appendix A.

2.3 Relation of scaling, self similarity and power laws

Formally, if we have an observable ψ which depends on a parameter x then under an arbitrary change $x \rightarrow \lambda x$, x is said to be *scale invariant* if there exists a number $\mu(\lambda)$ such that

$$\psi(x) = \mu\psi(\lambda x) \quad (2.3)$$

The solution of 2.3 is simply a power law $\psi(x) = x^\alpha$ (Sornette, 2003). The exponent α is given by

$$\alpha = -\frac{\ln \mu}{\ln \lambda} \quad (2.4)$$

This can be shown by inserting 2.4 in 2.3. The importance here lies in the fact that the ratio $\frac{\psi(\lambda x)}{\psi(x)} = \lambda^\alpha$ does not depend on x and that the relative values depend only on the ratio of two scales. A stochastic process encountering this self similar property is shown in section 4.3. In physics a branch called *renormalization group* uses the above relation to describe complex systems.

2.4 Fractal property of finance

The main object of this thesis is to show two statistical features of stock market time series where the fractal view applies. Namely, the power law behaviour of the tail of stock market returns and the detection of long term dependence.

Benoit Mandelbrot invented both ideas to finance and gave them biblical names according to two stories in the old testament. He called the power law behaviour of stock returns the *Noah effect* and the long memory property as the *Joseph effect*. In Noah's six hundredth year, God ordered a great flood to clean a wicked world. As with market crashes the flood came without warning.

The long memory property, meaning that the market moves in cycles, comes from the story of Joseph a Hebrew slave. He reported that the pharaoh of Egypt saw that after seven years of very good corn harvest seven bad will follow. Joseph advised the pharaoh to stock in good years. One could see this as a first arbitrage behaviour.

3. Scaling

In this section it will be shown that the Gaussian assumption is rejected in many instances and that tail of financial returns *scales*. The distribution is than well described by a *power law*. This approach is proposed by a new branch called *econophysics*. Before coming to the study of returns, I will show that this approach is applied in many fields. In fact, it came from the study of income distributions.

3.1 Scaling as a universal law

3.1.1 Paretos power law of income distribution

One of the earliest power laws in economics has been introduced by Vilfredo Pareto an Italian economist. He was on of the first economists who looked at data to study the distribution of income. His findings contradicted the classical economic school, because he found that income is far more distributed to the rich. He developed the following density function (Chipman, 1976) :

$$f(m) = \frac{\alpha u^\alpha}{m^{\alpha+1}} \quad (0 < u \leq m < \infty ; \alpha > 1) \quad (3.1)$$

Let $N(x)$ be the number of individuals with incomes exceeding x ; then the cumulative distribution function, defining the proportion p of the population earning incomes less than or equal to x , is given by

$$p = F(x) = \int_u^m f(\xi) d\xi = 1 - \frac{N(m)}{N(u)} = 1 - \left(\frac{u}{m}\right)^\alpha \quad (3.2)$$

hence

$$N(m) = N(u) \int_u^\infty f(\xi) d\xi - N(u) \left(\frac{u}{m}\right)^\alpha = \frac{A}{m^\alpha} \quad (3.3)$$

where

$$A = N(u)u^\alpha$$

The proportion $q = 1 - p$ of the total population with incomes exceeding x is then

$$q = G(m) = 1 - F(m) = \frac{N(m)}{N(u)} = \int_m^\infty f(\xi) d\xi = \left(\frac{u}{m}\right)^\alpha \quad (3.4)$$

Let $R(x)$ be the sum total of incomes exceeding x thus,

$$R(x) = \int_m^\infty \xi \left(-\frac{dN(\xi)}{d\xi} \right) d\xi = N(u) \int_m^\infty f(\xi) d\xi = \frac{A\alpha}{\alpha - 1} \frac{1}{m^{\alpha-1}} \quad (3.5)$$

The proportion r of total income which exceeds x , i.e. the relative share of those earning more than m is

$$r = \psi(m) = \frac{R(m)}{R(u)} = \frac{1}{\mu} \int_m^\infty \xi f(\xi) d\xi = \left(\frac{u}{m}\right)^{\alpha-1} \quad (3.6)$$

where

$$\mu = \int_u^\infty \xi f(\xi) d\xi = \frac{R(u)}{N(u)} = \frac{\alpha}{\alpha - 1} u \quad (3.7)$$

is the mean of the distribution, equal to per capita income (Chipman, 1976).

The expression $\left(\frac{u}{m}\right)^\alpha$ gives the probability that the proportion q exceeds the income level m , where x is often defined as the minimum income. Pareto found power laws in the region of $3/2$.

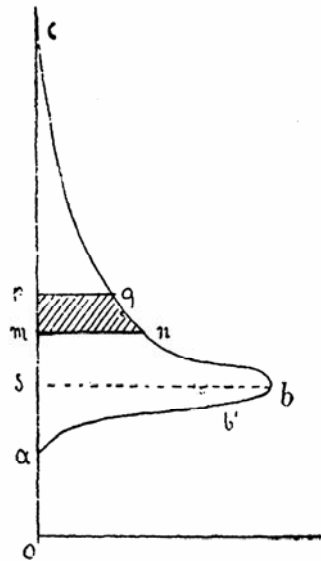


Figure 3.1: Pareto's original diagram from 1909 of how wealth is distributed through any human society. In any age or country. Rising income is on the vertical scale, population on the horizontal (which was switched later on). The number of people with income between levels m and p is represented by the shaded area. Source: Mandelbrot (2004).

Edgeworth criticized "that a close fit to a given statistics is not, *per se* and apart from a priori reasons, a proof that the curve in question is *the* form proper to the matter in hand".

Pareto replied a „rational” theory should be given more weight than to an empirical law. The point is that the later has to precede the theory (Chipman,

1976). Like the critics of the efficient market hypothesis, Pareto was accused of contradicting the well-established theory.

3.1.2 Zipf's law

An important example of scaling comes from linguistic studies. George Kingsley Zipf found that the frequency of any word is roughly inversely proportional to its rank, which means that the most frequent word will occur twice as much the second. In mathematical terms the power law exponent α is one:

$$Q(r) \sim \frac{A}{r^{-1/\alpha}} \quad (3.9)$$

Related to the above finding is the so called Zipf plot. Here the rank order axes and the frequency axes are both plotted in log form. Zipf's law has been applied in many fields such as in biology. An example of the application where Zipf's law applies is if we look at the accesses of documents in the world wide web. Crovella, Taquu and Bestravos (1998) found that for a data set of 46.830 unique files the slope of a log-log plot with the number of references to each file as a function of the files rank in reference count is -0.986 with $R^2 = 1.00$. The exponent is thus nearly 1.

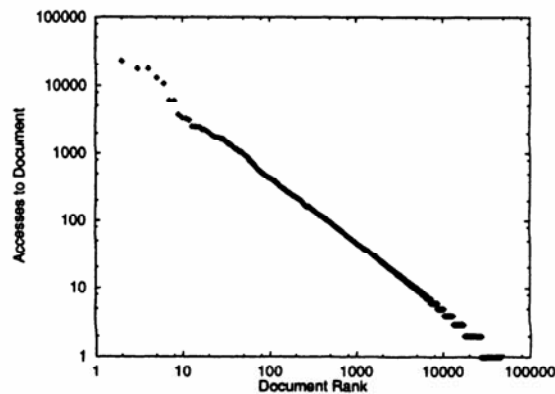


Figure 3.2: Zipf's law applied to web documents. Source: Crovella et al (1998)

3.1.3 Scaling behaviour of company growth

Stanley et al (1996) examined the statistical behaviour of corporate growth. It is just another instance where economic theory fails to take into account the empirics of the data. For example Gibrat's model assumes that growth is independent of the company size and uncorrelated in time. The process is therefore

$$S_{t+\Delta t} = S_t (1 + e_t) \quad (3.10)$$

with e_t uncorrelated and a mean close to zero and a standard deviation smaller than one. As a result $\log S_t$ follows a random walk and the firm size is log-normally distributed. This model is still used as a benchmark in current studies of firm size behaviour.

The data set used by Stanley et al (1996) consisted of 16 years, they defined the growth rate as $R = S_t / S_0$ with S as sales in the given year. In logarithmic scales $r = \log(S_t / S_0)$ and $s_0 = \log S_0$.

They calculated the conditional distribution $p(r|s_0)$ of the growth rates.

It was found that the data did not fit to a Gaussian but rather an exponential distribution with the form of

$$p(r|s_0) = \frac{1}{\sqrt{2}\sigma(s_0)} \exp\left(-\frac{\sqrt{2}|r - \bar{r}(s_0)|}{\sigma(s_0)}\right) \quad (3.11)$$

depending on the chosen initial value s_0 (Figure 3.3).

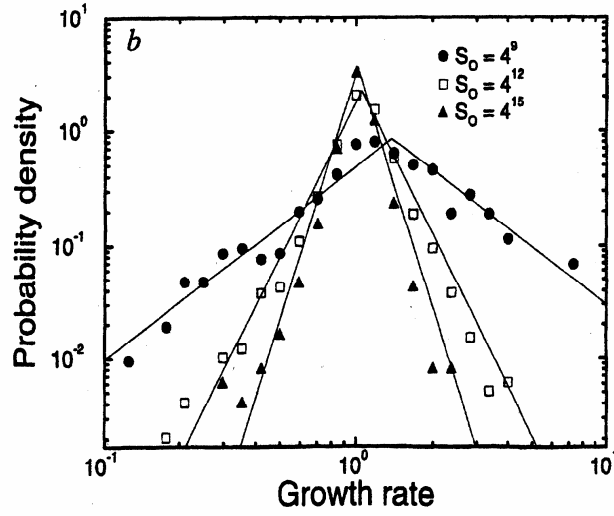


Figure 3.3: Probability density $p(r|s_0)$ of the annual growth rate, for three different bins of initial sales: $4^{8.5} < S_0 < 4^{9.5}$ (circles), $4^{11.5} < S_0 < 4^{12.5}$ (squares) and $4^{14.5} < S_0 < 4^{15.5}$ (triangles). The data were averaged over all 16 one-year periods between 1975 and 1991. The solid lines are fits to equation (1) using the mean $\bar{r}(s_0)$ and standard deviation $\sigma(s_0)$ Source: Stanley et. al (1996).

The plot in Figure 3.3 suggests that the data is well described by equation 3.11. The regression from sales of 10^4 up to 10^{11} dollar

$$\sigma(s_0) = a \exp(-\beta s_0) = a S_0^{-\beta} \quad (3.12)$$

yielded $a \approx 6.66$ and $\beta = 0.15 \pm 0.03$. To look for the scaling phenomena Stanley et al graphed the dependent variable against the scaled independent variable. If scaling holds the curves must collapse. Figure 3.4 shows this remarkable result.

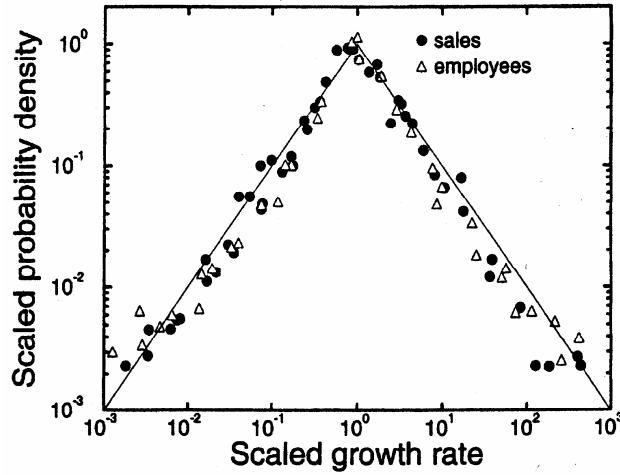


Figure 3.4: Scaled probability density $p_{scal} = 2^{1/2} \sigma(s_0) p(r|s_0)$ as a function of the scaled growth rate $r_{scal} = 2^{1/2} [r - \bar{r}(s_0)] / \sigma(s_0)$ of sales (circles). The values were rescaled using the measured values of $\bar{r}(s_0)$ and $\sigma(s_0)$. The same analysis has been done with the number of employees. Source: Stanley et. al. (1996)

It is notable that Stanley et al. get the same result with the number of employees (Figure 3.4). Here they found a slope of $\beta = 0.16 \pm 0.03$.

Also three indicators of growth are described by equation 3.11 and 3.12. Namely, cost of goods sold $\beta = 0.16 \pm 0.03$, assets $\beta = 0.17 \pm 0.04$ and property plant and equipment $\beta = 0.18 \pm 0.04$.

3.2 The Tail Exponent and Stable Distributions

Related to the topic of power laws is the proposition that financial returns follow a stable distribution. Other distributions one could use are the Student's t , hyperbolic, normal inverse Gaussian. One advantage of using stable distributions is that they are supported by the generalized *Central Limit Theorem*, which states that a stable law is the only possible limit distribution for properly normalized sums of independent identically distributed random

variables (Härdle et al, 2005). Stable laws were introduced by Paul Levy when he investigated the behaviour of sums of independent random variables. A sum of two independent random variables having an α -stable distribution with index α is again α -stable with the same index α . This invariance property does not hold for different α 's.

Another advantage of stable distributions is that they can resemble fat tails and asymmetry at the same time. If one wants to model extreme events like market crashes or natural catastrophes the stable distribution is a good candidate. The fractal property can be seen as that with a power law behaviour of the tail of the distribution the ration of $\frac{P_{>x}}{P_{>nx}}$ does not depend on x .

To describe the stable distribution one needs four parameters: the tail index $\alpha \in [0,2]$, a skewness parameter $\beta \in [-1,1]$, a scale parameter $\sigma > 0$ and a location parameter $\mu \in \mathbb{R}$. If $\alpha = 2$ then we have the Gaussian distribution. If $\alpha < 2$ the variance is infinite and the tails are asymptotically like a Pareto law and hence like a power law.

The current empirical research finds that log returns $r_t = \ln(p_t) - \ln(p_{t-1})$ exhibit an exponent of 3 so we have a *cubic law* (Stanley et. al. 2003)

$$\Pr(|r_t| > x) \cong x^{-3} \quad (3.13)$$

It is remarkable that this was found for 10 countries (Australia, Canada, France, Germany, Japan, Hong-Kong, Netherlands, South Korea, Spain, United Kingdom) and for foreign exchange markets as well. The literature interprets this as an exclusion of the stable distribution family. However, Weron (2001) shows that if $\alpha > 2$ we do not need to exclude stable Levy distributions. In fact, a much lower alpha is confirmed by some studies e.g. Kaizoji (2004), Mantegna (1996).

3.2.1 Numerical properties of stable distributions

By using the Central Limit Theorem it can be shown that

$$\lim_{x \rightarrow \infty} x^\alpha P(X > x) = C_\alpha (1 + \beta) \sigma^\alpha \quad (3.14)$$

$$\lim_{x \rightarrow \infty} x^\alpha P(X > -x) = C_\alpha (1 + \beta) \sigma^\alpha$$

where

$$C_\alpha = \left(2 \int_0^\infty x^{-\alpha} \sin(x) dx \right)^{-1} = \frac{1}{\pi} \Gamma(\alpha) \sin \frac{\pi\alpha}{2} \quad (3.15)$$

The convergence to power law tails depends on α . It is slower for „large” values (Härdle et al, 2005), which is shown in Figure 3.

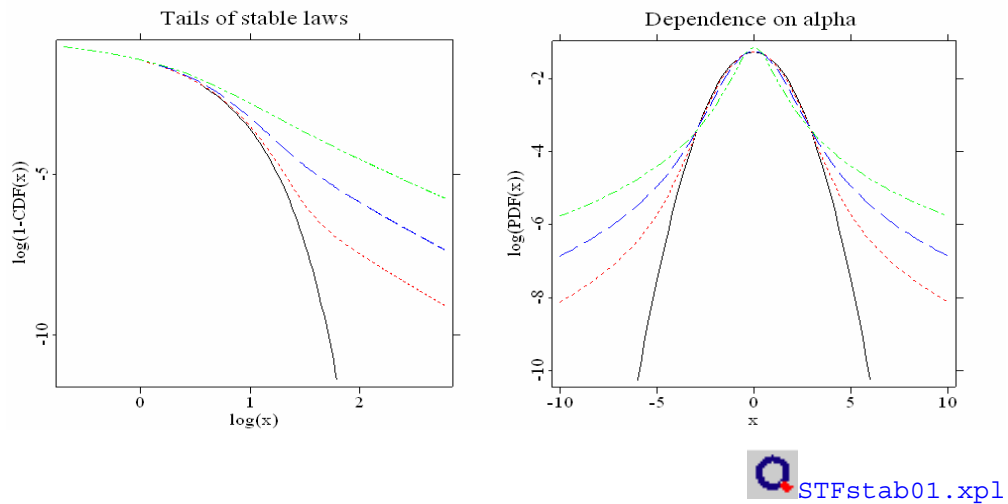


Figure 3.3.2.: Left: A semilog plot of symmetric ($\beta = \mu = 0$) α -stable probability density functions for $\alpha = 2$ (black solid line), $\alpha = 1.8$ (red line), $\alpha = 1.5$ and $\alpha = 1$ (green line). The Gaussian ($\alpha = 2$) density forms parabola and is the only distribution with exponential tails. Right: tails of symmetric α -stable cumulative distribution function (right tails) for $\alpha = 2$ (black line), 1.5 (green line), 1.8 (red line) on double logarithmic scale. For $\alpha < 2$ the tails have straight lines with slope $-\alpha$. Source : (Härdle, et al 2005)

If $\alpha > 1$, the mean of the distribution exists and is equal to μ . Generally, the p th moment of a stable random variable is finite if and only if $p < \alpha$. The distribution is skewed to the right if the skewness parameter β is positive and to the left if the parameter is negative. With α going to 2, β loses its effect and the distribution approaches the Gaussian regardless of β (Härdle et al, 2005). σ and μ are the scale and location parameters, σ determines the width and μ the peak of the density.

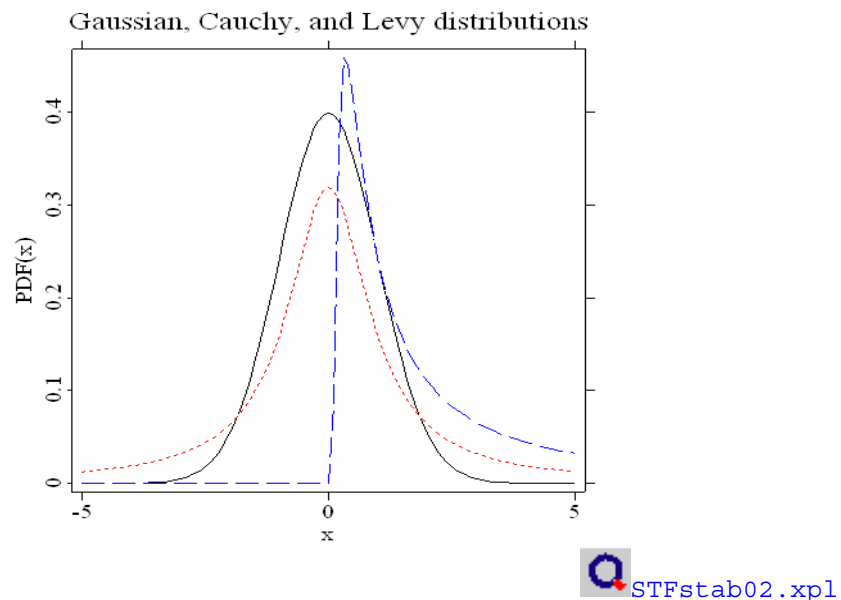


Figure 3.6.: Only three densities have closed form formulas, Gaussian (black line), Cauchy with $\alpha = 1$ (red line) and Levy $\alpha = 0.5$ $\beta = 1$ (blue line). The Levy is totally skewed in \mathbb{R}_+ . In general, for $\alpha < 1$ and $\beta = 1$ (-1) the distribution totally skewed. Source: (Härdle et al, 2005)

There are only three instances where the Levy family has a closed form formula which is shown in Figure 3.6. For our economic examples the Cauchy would mean that income is distributed more to the rich and for the growth of companies example that business is more concentrated with large firms.

3.2.2 Characteristic Function Representation

Because in most circumstances we have no closed form formula it is convenient to express the stable distributions by its characteristic functions (i.e. there inverse characteristic functions). A popular parameterization of the characteristic function of a stable random variable $X \sim S_\alpha(\sigma, \beta, \mu)$

$$\ln \phi(t) = \begin{cases} -\sigma|t|^\alpha \left\{ 1 - i\beta \text{sign}(t) \tan \frac{\pi\alpha}{2} \right\} + i\mu t, & \alpha \neq 1 \\ -\sigma|t| \left\{ 1 + i\beta \text{sign}(t) \frac{2}{\pi} \ln(|t|) \right\} + i\mu t, & \alpha = 1 \end{cases} \quad (3.16)$$

The location parameters of the representations are related by $\mu = \mu_0 - \beta\sigma \tan \frac{\pi\alpha}{2}$ for $\alpha \neq 1$ and $\mu = \mu_0 - \beta\sigma \frac{2}{\pi} \ln \sigma$ for $\alpha = 1$.

It is notable that the traditional scale parameter σ_G of the Gaussian:

$$f_G(x) = \frac{1}{\sqrt{2\pi\sigma_G}} \exp \left\{ -\frac{(x-\mu)^2}{2\sigma_G^2} \right\}, \quad (3.17)$$

is not the same σ in the above formulas. The relation is

$$\sigma_G = \sqrt{2}\sigma \quad (3.18)$$

3.2.3 Stable Distribution computation

Because of the lack of closed form formulas one has to use numerical procedures to for maximum likelihood estimation. However, these are computationally extensive. There are two ways of solving the problem. The first is fast Fourier transform (FFT) and the second direct numerical

integration. The advantage of the FFT approach is that it is faster for larger samples (Härdle et al, 2005).

If we set $\varsigma = -\beta \tan \frac{\pi\alpha}{2}$ then we can express the density $f(x; \alpha, \beta)$ of a α stable random variable in S^0 , i.e. $X \sim S^0_\alpha(1, \beta, 0)$ as

$$f(x; \alpha, \beta) = \frac{\alpha(x - \varsigma)^{\frac{1}{\alpha-1}}}{\pi|\alpha-1|} \int_{-\varsigma}^{\frac{\pi}{2}} V(\theta; \alpha, \beta) \exp\left\{-(x - \varsigma)^{\frac{\alpha}{\alpha-1}} V(\theta; \alpha, \beta)\right\} d\theta, \quad (3.19)$$

if $\alpha \neq 1$ and $x > \varsigma$

$$f(x; \alpha, \beta) = \frac{\Gamma\left(1 + \frac{1}{\alpha}\right) \cos(\xi)}{\pi(1 + \varsigma^3)^{\frac{1}{3\alpha}}} \quad (3.20)$$

if $\alpha \neq 1$ and $x = \varsigma$

$$f(x; \alpha, -\beta) = f(-x; \alpha, -\beta) \quad (3.21)$$

if $\alpha \neq 1$ and $x < \varsigma$

$$f(x; 1, \beta) = \begin{cases} \frac{1}{2|\beta|} e^{\frac{\pi x}{2\beta}} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} V(\theta; 1, \beta) \exp\left\{-e^{\frac{\pi x}{2\beta}} V(\theta; 1, \beta)\right\} d\theta, & \beta \neq 0 \\ \frac{1}{\pi(1+x^2)}, & \beta = 0 \end{cases} \quad (3.22)$$

where:

$$\xi = \begin{cases} \frac{1}{\alpha} \arctan(-\varsigma), \\ \frac{\pi}{2}, \end{cases} \quad (3.23)$$

and

$$V(\theta; \alpha, \beta) = \begin{cases} (\cos \alpha \xi)^{\frac{1}{\alpha-1}} \left(\frac{\cos \theta}{\sin \alpha(\xi + \theta)} \right)^{\frac{\alpha}{\alpha-1}} \frac{\cos\{\alpha \xi + (\alpha-1)\theta\}}{\cos \theta}, & \alpha \neq 1 \\ \frac{2}{\pi} \left(\frac{\frac{\pi}{2} + \beta \theta}{\cos \theta} \right) \exp\left\{ \frac{1}{\beta} \left(\frac{\pi}{2} + \beta \theta \right) \tan \theta \right\}, & \dots \alpha = 1, \beta \neq 0 \end{cases} \quad (3.24)$$

Formula 3.19 requires numerical integration of the function $g(\cdot) \exp\{-g(\cdot)\}$,

where $g(\theta; x, \alpha, \beta) = (x - \varsigma)^{\frac{\alpha}{\alpha-1}} V(\theta; \alpha, \beta)$.

3.2.4 Estimation of the tail exponent and stable of parameters

The main problem when we want to estimate stable parameters is, as noted above, the lack of closed form formulas in most circumstances. As shown in the previous section, numerical integration is very demanding in calculation.

If the data comes not from an α stable distribution we can apply direct tail estimation or the Hill estimator.

3.2.4.1 Direct tail estimation

The easiest way to estimate values of α empirically is to use log-log complementary distribution (CD) plots. These show the complementary distribution function $\bar{F}(x) = 1 - F(x) = P[X > x]$ on log-log axes (Taqqu et al, 2000). Now we have the property that

$$\frac{d \log \bar{F}(x)}{d \log x} \sim -\alpha \quad (3.25)$$

for large values of x . A simple possibility in practice is to select a minimal value of x_0 from which the log-log plot appears to be linear. However, the method is very sensitive to the size of the sample. Figure 3.7 illustrates that a regression in a small sample can lead to a false conclusion that we are outside the stable family (Weron, 2001).

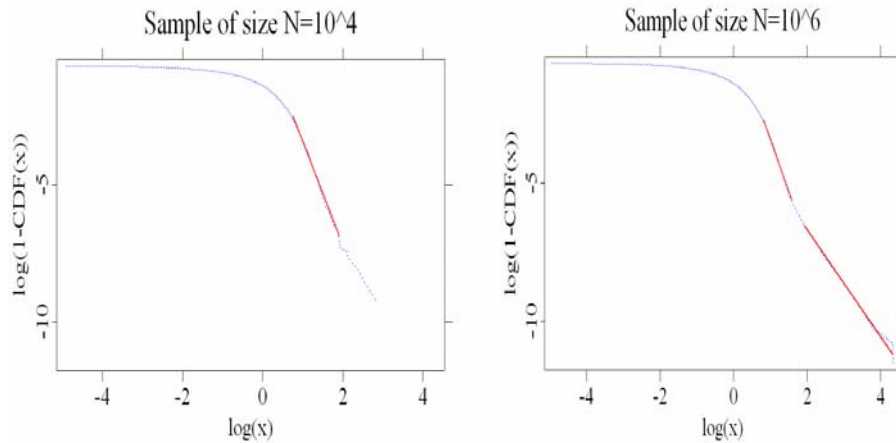
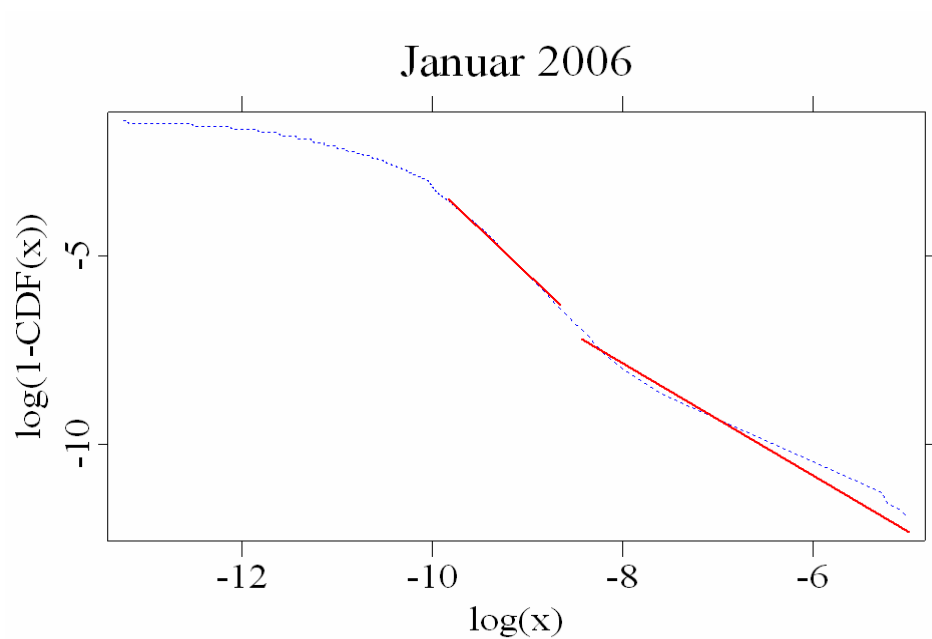
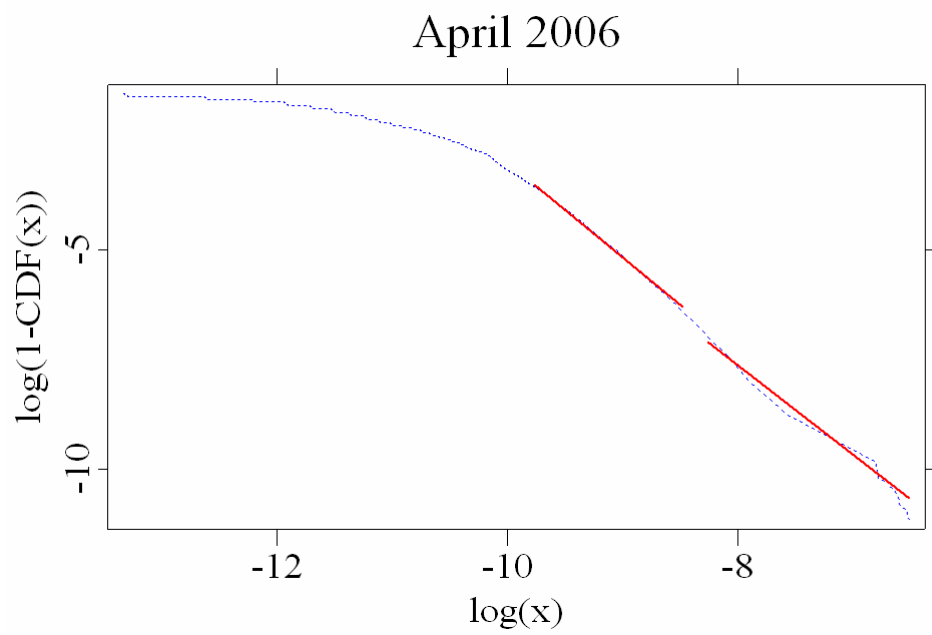


Figure 3.7: Double logarithmic plot of the right tail of an empirical symmetric 1.9 stable distribution function for a sample size of $N=10^4$ (left) and $N=10^6$ (right). Red lines present the regression fit. The tail index estimate $\hat{\alpha} = 3,7320$ with the smaller sample is close to initial power law like decay of the left plot. The far tail estimate of $\hat{\alpha} = 1,9309$ is close to the true value of α . Source: (Härdle et al, 2005)

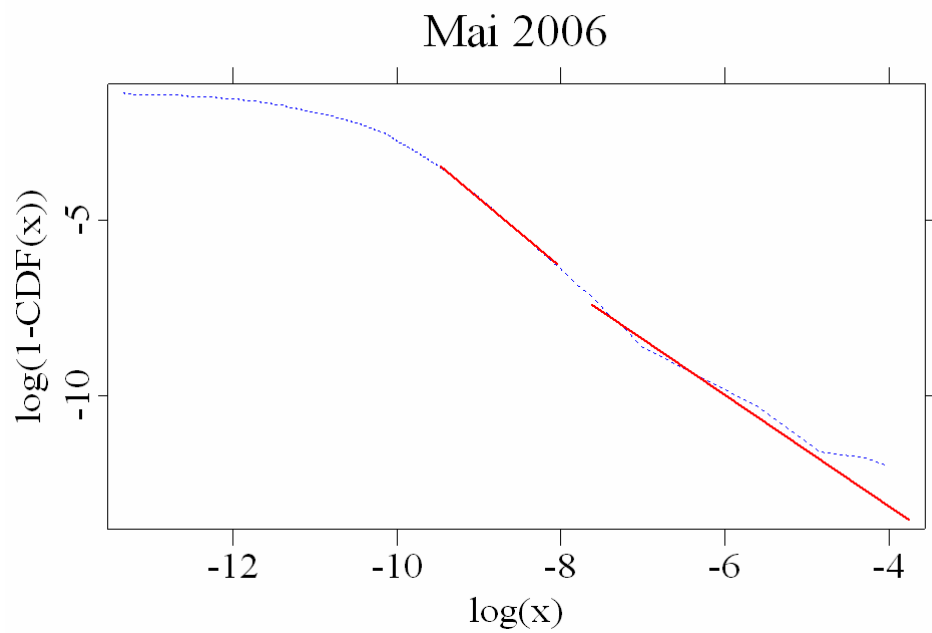
Next, the regression method described above will be applied to monthly tick data from the EUREX-Dax-Future (ca.600.000 observations per month). The estimated values indicate that we are not far from the stable regime. The tail index for January 2007 2006 with -1,60 and -1,48 fits to the theoretical analysis so far. However, the observed value mean that tick by returns are much more risky since these alphas are very low.



*Figure 3.8: Estimation of tail in by OLS regression for Eurex tick data Januar 2006.
The first value is -2,38 and for the tail -1,48.*



*Figure 3.9: Estimation of tail in by OLS regression for Eurex tick data April 2006.
The first value is -2,17 and for the tail -2,05.*



*Figure 3.10: Estimation of tail in by OLS regression for Eurex tick data Januar 2007.
The first value is -1,95 and for the tail -1,58.*

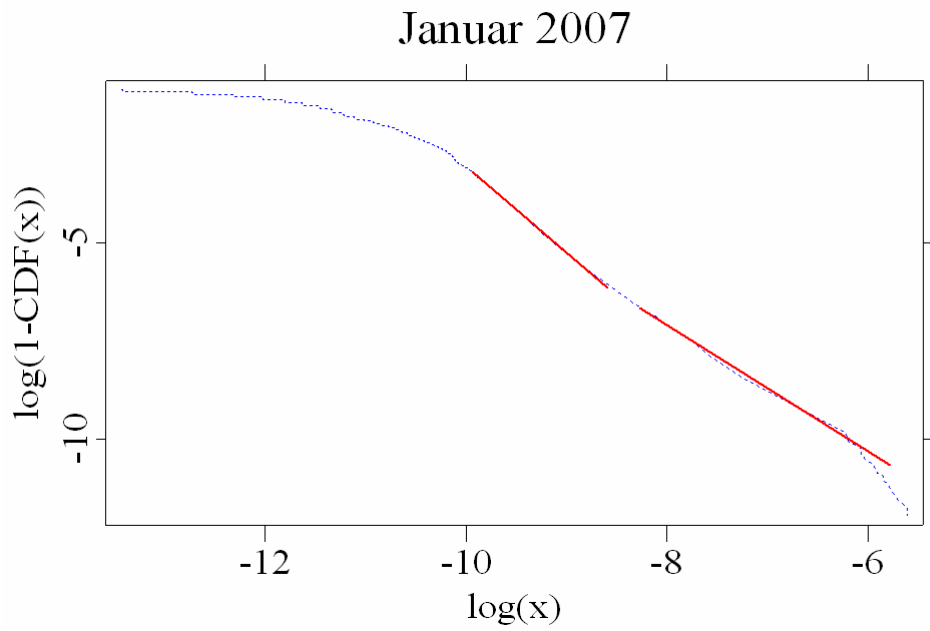


Figure 3.11: Estimation of tail in by OLS regression for Eurex tick data Januar 2007.
The first value is -2,19 and for the tail -1,60.

3.2.4.2 The Hill Estimator

The second method for estimating the tail index is the Hill estimator. The Hill estimator gives an estimate of α as a function of the k largest elements in the data set (Taqqu et al, 2000),

$$H_{k,n} = \frac{1}{k} \sum_{i=1}^k (\log X_{(i)} - \log X_{(k+1)}) \quad (3.26)$$

with $X_{(1)} \geq \dots \geq X_{(n)}$ denoting the dataset's order statistics. The procedure is nonparametric which means that it does not assume any distribution. The estimator tends to overestimate the tail index if α is close to two and the sample is small (Härdle et al, 2004).

The requirement of a large data set means that one should choose high-frequency data to analyse the tail of stock returns with this estimator. Weron (2001) shows by simulating the stable distribution that the Hill estimator can overestimate the tail exponent, see Figure 3.12.

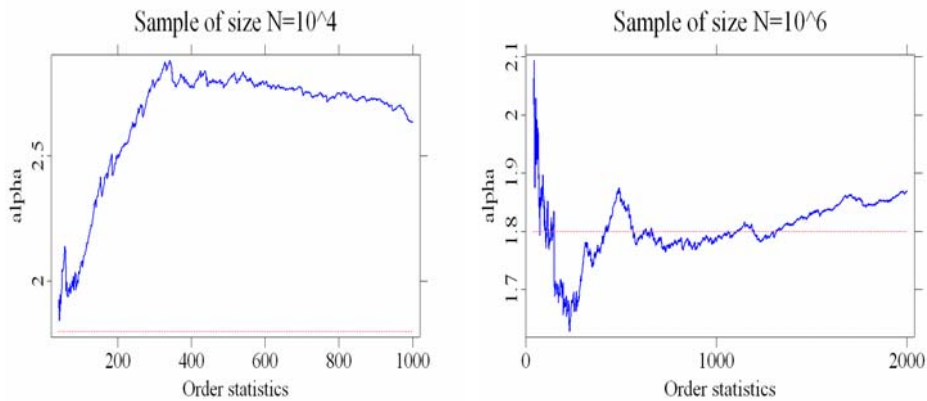


Figure 3.12: Plot of the Hill estimator versus the maximum order statistic k . The red line represents the true value of α . A good estimate is obtained only for values of $k=500, \dots, 1300$. Source (Härdle et al, 2005).

Now, the Hill estimator is applied to the EUREX tick data in first analysed in the preceding section. The exponents do not exactly equal those of the OLS regression but the indication that we cannot reject the hypothesis of being inside the stable distribution is given. All alphas are in the range 1,5-2,5 and hence we cannot reject Levy distributions. The overestimation noted by Weron (2001) does not seem to hold either by using OLS or the Hill estimator.

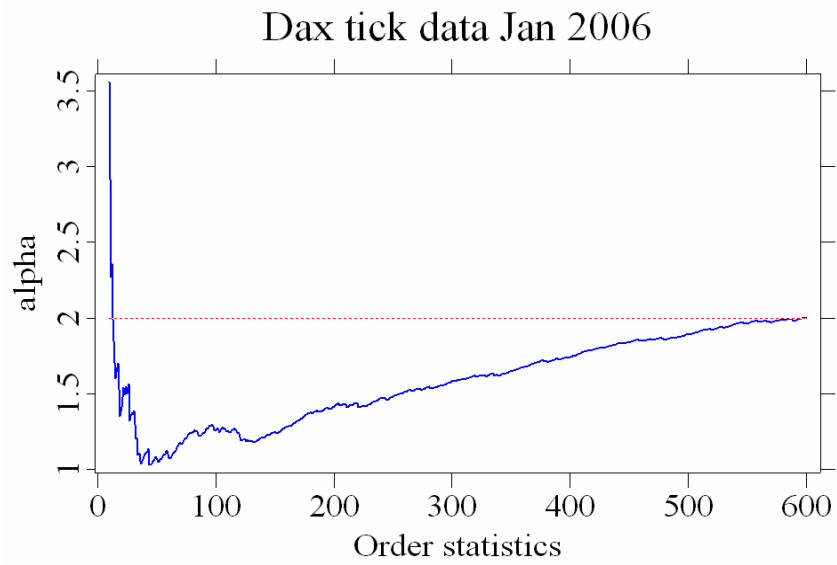


Figure 3.13: Estimation of Hill Estimator for January 2006, EUREX DAX tick data.

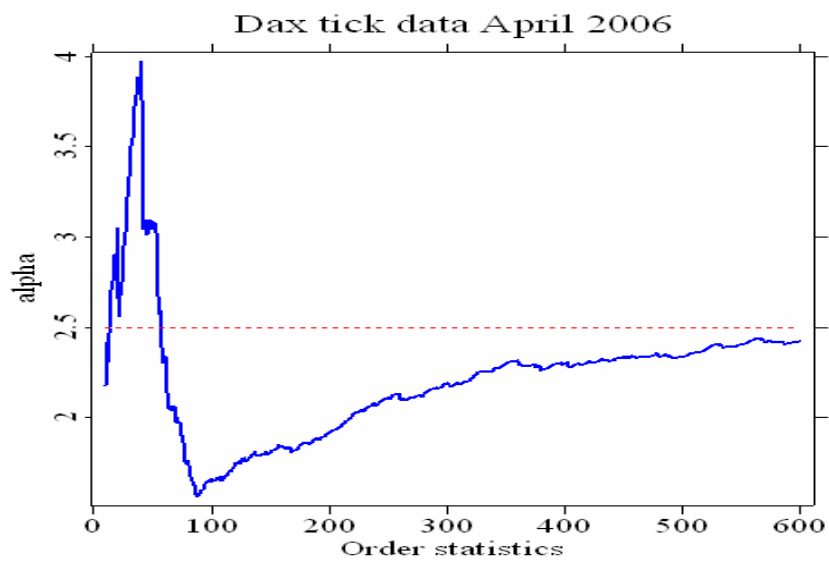


Figure 3.14: Estimation of Hill Estimator for April 2006, EUREX Dax tick data.

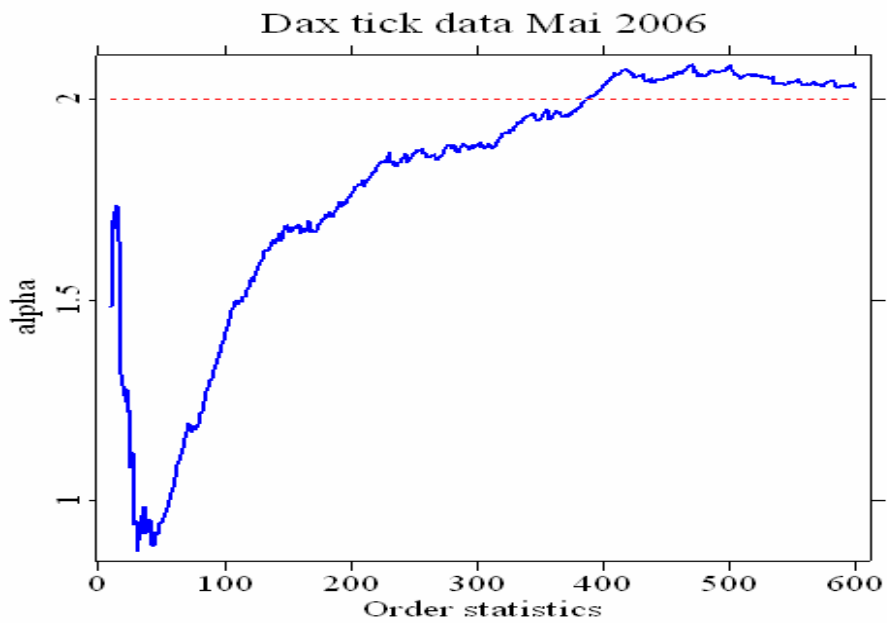


Figure 3.15: Estimation of Hill Estimator for Mai 2006, EUREX Dax tick data.

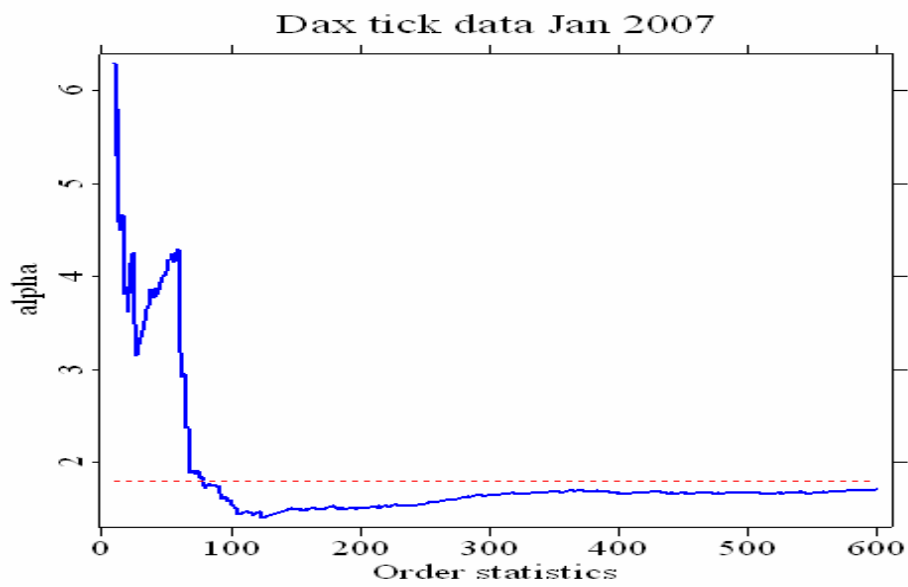


Figure 3.16: Estimation of Hill Estimator for January 2007, EUREX Dax tick data.

3.2.4.3 Estimation by using the characteristic function

If we have a sample x_1, \dots, x_n of independent and identically distributed (i.i.d.)

Random variables we define the characteristic function with

$$\hat{\phi}(t) = \frac{1}{n} \sum_{j=1}^n e^{itx_j}, \quad (3.27)$$

Because $|\hat{\phi}(t)|$ is bounded by unity all moments of $\hat{\phi}(t)$ are finite and, for any fixed t , it is the sample average of i.i.d. random variables $\exp(itx_j)$. It follows that $\hat{\phi}(t)$ is a consistent estimator of the characteristic function $\phi(t)$. A regression type method is developed by Koutrouvelis (1980). He starts with an initial estimate of the parameters and proceeds until some convergence criterion is satisfied (Härdle et al, 2005). One iteration is composed of two weighted regression runs. How many points we have to use depends on the sample size and the starting value of α . The convergence depends on the initial estimates.

We can derive equation from 3.16.

$$\ln(-\ln|\phi(t)|^2) = \ln(2\sigma^\alpha) + \alpha \ln|t| \quad (3.28)$$

Real and imaginary parts of $\phi(t)$ with $\alpha \neq 1$ are

$$R\{\phi(t)\} = \exp(-|\sigma|^\alpha) \cos\left[\mu t + |\sigma|^\alpha \beta \text{sign}(t) \tan \frac{\pi\alpha}{2}\right] \quad (3.29)$$

and

$$I\{\phi(t)\} = \exp(-|\sigma|^\alpha) \sin \left[\mu t + |\sigma|^\alpha \beta \text{sign}(t) \tan \frac{\pi\alpha}{2} \right] \quad (3.30)$$

these two equations lead to

$$\arctan \left(\frac{I\{\phi(t)\}}{R\{\phi(t)\}} \right) = \mu t + \beta \sigma^\alpha \tan \frac{\pi\alpha}{2} \text{sign}(t) |t|^\alpha, \quad (3.31)$$

Equation 3.31 depends only on α and σ we can therefore regress $y = \ln \left(-\ln |\phi_n(t)|^2 \right)$ on $w = \ln |t|$ in the model

$$y_k = m + \alpha w_k + \varepsilon_k, \quad k=1,2,\dots,K \quad (3.32)$$

with t_k as an appropriate set of real numbers, $m = \ln(2\sigma^\alpha)$, and ε_k as the error term. Koutrouvelis (1980) suggested to use $t_k = \frac{\pi k}{25}, k = 1, 2, \dots, K$; with K ranging between 9 and 134 for different estimates of α and sample sizes.

When α and σ have been fixed at some value, then we can also estimate β and μ by the use of 3.31. This procedure is then repeated until prespecified criteria are satisfied (Härdle et al, 2005). The above technique is now applied to the tick data set also used to estimate via OLS and Hill. As shown in Figures 3.17-3.21 the stable is much more accurate when describing tick by tick data. Moreover the very low alphas indicate that the estimates by OLS and Hill are not misleading. However the values of OLS fit more to the values in Table 3.1.

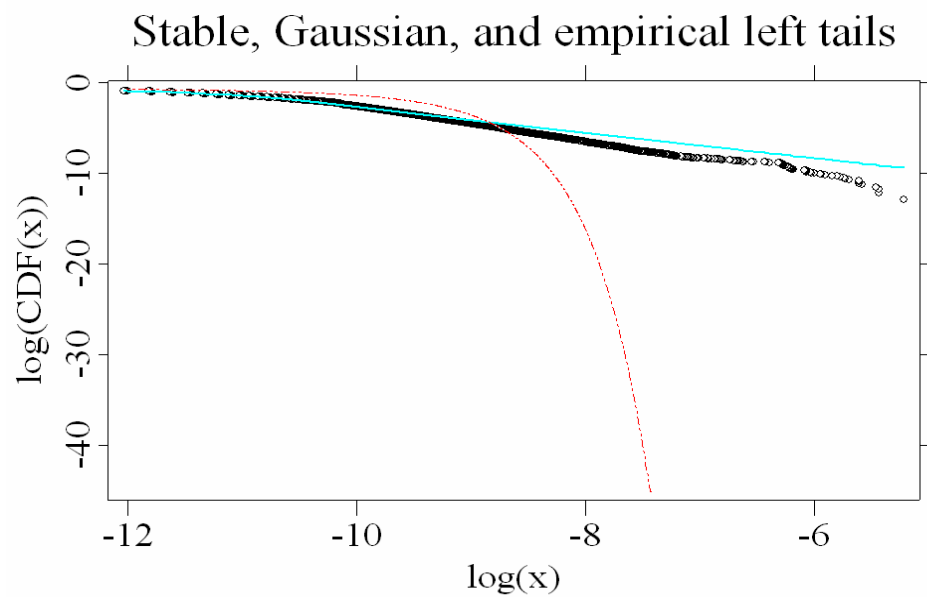


Figure 3.17: Estimation of the stable distribution and Gaussian (red line) for EUREX tick data January 2006. Also shown is the empirical cdf.

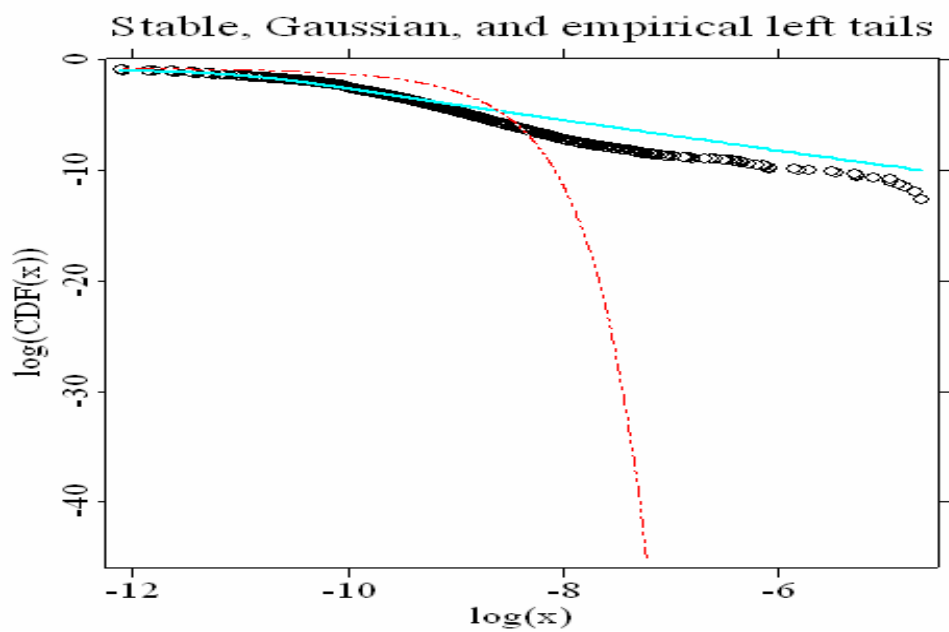


Figure 3.18: Estimation of the stable distribution and Gaussian (red line) for EUREX tick data April 2006. Also shown is the empirical cdf.

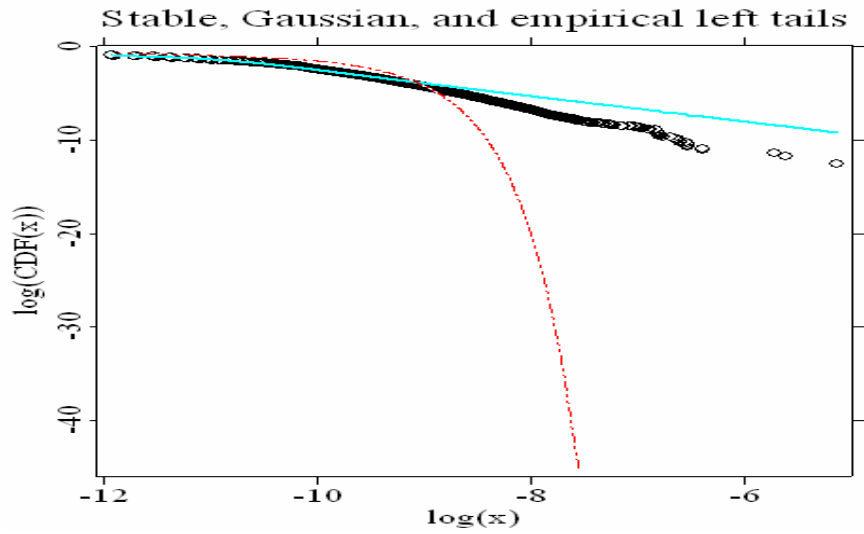


Figure 3.19: Estimation of the stable distribution and Gaussian (red line) for EUREX tick data April 2006. Also shown is the empirical cdf.

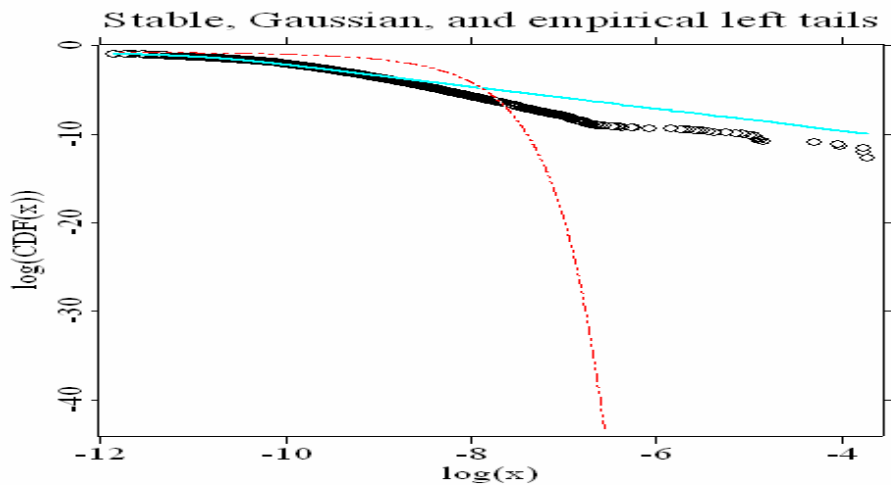


Figure 3.20: Estimation of the stable distribution and Gaussian (red line) for EUREX tick data April 2006. Also shown is the empirical cdf.

Month	alpha	sigma	beta	mu	A. D.	Kolm.
Jan 2006	1,2264	0,0000	-0,0053	0,000	252,38	13,91
Apr.2006	1,3334	0,0000	0,0059	0,000	231,45	14,57
Mai2006	1,3595	0,0000	-0,091	0,000	382,40	16,90
Jan 2007	1,3705	0,0000	-0,01	0,000	354,78	17,95

Table 3.1: Estimation of parameters and test values for the stable distribution for months of DAX tick data

Month	mean	std.	A.D.	Kolm.
Jan 2006	0,0000	0,0001	INF	120,39
Apr.2006	0,0000	0,0001	INF	75,29
Mai2006	0,0000	0,0001	INF	108,70
Jan 2007	0,0000	0,0002	INF	153,51

Table 3.2: Estimation of parameters and test values for the Gaussian for months of DAX tick data

3.3 Scaling of Volume and other power laws in finance

There are other power laws found in empirical analysis. For example Stanley et al (2003) report the scaling of price impact of trades, size of large investors and for the number of trades. Here, I will shortly examine the power law exponent of volume. This is maybe an important one, because one can imagine that more volume leads to more volatility. In Figure 3.21 a OLS regression on log-log plot is done with IBM tick data. Remarkably, as with returns, the line becomes straight suggesting a power law exponent is at work. The slope here is 1,8. Stanley et al (2003) report “half a cubic law”.

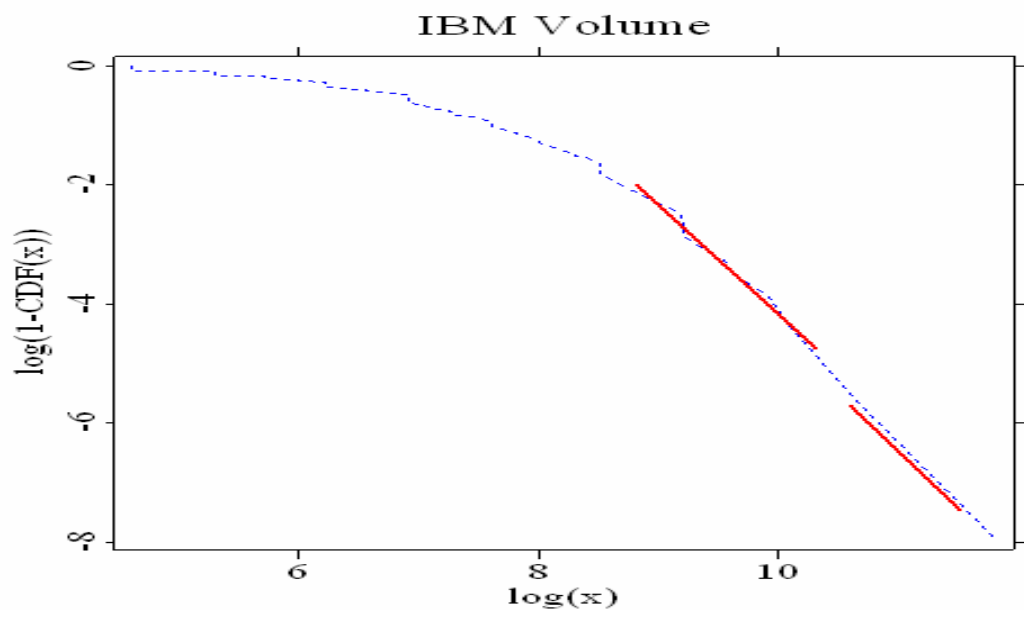


Figure 3.21: OLS regression on a log-log plot for IBM volume

4. Long Memory and R/S Analysis

A second property where the fractal view of financial markets comes in to play is the presence of long memory and hence fractional Brownian motion as an alternative for modelling stock price time series.

Long memory is now studied in depth by econometricians (e.g. Granger 1993). ARIMA and FIGARCH models can be used to account for long memory in financial time series. Here we pursue the R/S analysis of the hydrologist Hurst who developed this powerful method to analyse the sequence of floods of the river Nil. His exponent can also be interpreted as a power law, but one that does account for memory in the data instead of the distribution.

4.1 long memory of time series

One of the earliest finding of a long memory effect has been in the 19th century with astronomy. Other examples include hydrology, geophysics, climatology, economics and agronomy.

We can divide the findings of long correlations in two categories (Beran 1994).

1. The long memory property is expected a priori due to the nature of the phenomenon
2. Observations are expected to be (more or less) independent

In economics Granger (1966) was the first who recognized the spectral shape of economic variables even after he removed trends in business cycles etc. For economic time series the shape of the spectral density is a function with the pole at the origin (Beran, 1994).

One of the first examples found for long memory was for spatial data in agronomy. The method used was called uniformity trials which is a method to determine the best size of a plot. It was found that the variance of the average yield of a plot converged slower to zero than if they were independent.

A third classic example comes from astronomy where the correlation was unexpected (the two previous examples refer to 1.). It was observed that errors affect whole groups of observations which drastically increases probable errors. Therefore the traditional error was too small. This sort of error is called *semi systematic error*.

4.1.2 Properties of long memory processes

Suppose we have a random sample X_1, X_2, \dots, X_n from a population and all variables have the same marginal distribution F . To show why it is important to study processes with long memory one has to recall what the assumption of classical statistics are.

1. that the population mean $\mu = E(X_i)$ exists and is finite
2. that the variance $\sigma^2 = \text{var}(X_i)$ exists and is finite
3. X_1, X_2, \dots, X_n are uncorrelated

The third assumption is the important assumption for the following analysis (the second has been the one for third part) and means that

$$\rho(i, j) = 0 \text{ for } i \neq j \quad (4.1)$$

where

$$\rho(i, j) = \frac{\gamma(i, j)}{\sigma^2} \quad (4.2)$$

is the autocorrelation between X_i and X_j , and

$$\gamma(i, j) = E[(X_i - \mu)(X_j - \mu)] \quad (4.3)$$

The first two assumptions depend only on the marginal distribution F .

The analysis now focuses on what happens if the third assumption is not valid.

If μ is constant the variance of $\bar{X} = n^{-1} \sum_{i=1}^n X_i$ is equal to

$$\text{var}(\bar{X}) = n^{-2} \sum_{i,j=1}^n \gamma(i, j) = n^{-2} \sigma^2 \sum_{i,j}^n \rho(i, j) \quad (4.4)$$

In the case that the correlations add up to zero we have

$$\sum_{i,j=1}^n \gamma(i, j) = \sum_{i=1}^n \gamma(i, i) = \sum_{i=1}^n \sigma^2 = n\sigma^2 \quad (4.5)$$

So if we have $\sum_{i \neq j}^n \rho(i, j) = 0$ the variance is just

$$\text{var}(\bar{X}) = \sigma^2 n^{-1} \quad (4.6)$$

as in elementary statistics. But the variance is

$$\sigma^2 [1 + \delta_n(\rho)] n^{-1} \quad (4.7)$$

if the observations are correlated. The correction term is

$$\delta_n(\rho) = n^{-1} \sum_{i \neq j} \rho(i, j) \quad (4.8)$$

In case that the correlation depends only on the first lag $|i - j|$ 4.8 gets

$$\delta_n(\rho) = 2 \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) \rho(k) \quad (4.9)$$

If we take an AR (1) process

$$X_i = aX_{i-1} + \varepsilon_i, \quad a \in (-1, 1) \quad (4.10)$$

the autocorrelation (ACF) is now

$$\rho(i, j) = a^{|i-j|} \quad (4.11)$$

As a result we get:

$$\text{var}(\bar{X}) = n^{-2} \sum_{i,j=1}^n \gamma(i, j) = \frac{\sigma^2}{n^2} \sum_{i,j=1}^n \rho(i, j) = \frac{\sigma^2}{n} \left[\sum_{i=1}^n 1 + \frac{1}{n} \sum_{i \neq j} a^{|i-j|} \right] \quad (4.12)$$

$$= \sigma^2 n^{-1} \left[1 + 2 \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) a^k \right] \quad (4.13)$$

which can be written as

$$\sigma^2 [1 + \delta_n(a)] n^{-1} = \sigma^2 c_n(a) n^{-1} \quad (4.14)$$

with

$$\delta_n = \frac{2a}{1-a} \left[1 - n^{-1} \frac{1}{1-a} + n^{-1} \frac{a^n}{1-a} \right] \quad (4.15)$$

Asymptotically:

$$\text{var}(\bar{X}) \approx \sigma^2 [1 + \delta(a)] n^{-1} = \sigma^2 c(a) n^{-1} \quad (4.16)$$

with

$$\delta(a) = \frac{2a}{1-a} \quad (4.17)$$

How much the correction factor $c(a)$ differs from the standard formula depends on a . If a tends to one then $c(a) \rightarrow \infty$ and if a tends to minus one $c(a) \rightarrow 0$. In these cases $\sigma^2 n^{-1}$ is a poor estimate of the variance and one needs a good estimate of $c(a)$.

The problem here is not only to find a appropriate value of $c(a)$ but also that the variance of \bar{X} decays more slowly to zero than n^{-1} . If the data generating process is such that the variance decays so slow than we cannot find a suitable value of c . Rather c is increasing with n . If one tries to fit an ARMA model in such an environment we would have many parameters. Statistical inference is then difficult to interpret. A natural simple way to model the slow decay would be $n^{-\alpha}$ with $\alpha \in (0,1)$,

$$\text{var}(\bar{X}) \approx \sigma^2 c(\rho) n^{-\alpha} \quad (4.18)$$

and

$$c(\rho) = \lim_{n \rightarrow \infty} n^{\alpha-2} \sum_{i \neq j} \rho(i, j) \quad (4.19)$$

If we consider specific lags and an increasing sample size it from 4.9 and 4.19 that

$$\sum_{k=-(n-1)}^{n-1} \rho(k) \approx \text{cons} \tan t \cdot n^{1-\alpha} \quad (4.20)$$

Because α is assumed to be less than one, we have

$$\sum_{k=-\infty}^{\infty} \rho(k) = \infty \quad (4.21)$$

which means that the correlations do not sum up.

As a result if 4.20 holds we have

$$\rho(k) \approx c_p |k|^{-\alpha} \quad (4.22)$$

when $|k|$ tends to infinity and c_p is a constant. It is now shown that 4.23 is a process with long memory.

4.2 R/S Analysis

A way to detect long memory processes (*Joseph effect*) is the so called R/S analysis. Hurst (1951) developed the analysis as follows if we want to find the ideal capacity of the Nil river under the assumption of no storage losses. The goal is that the outflow is uniform, that the reservoir is full and that the reservoir never outflows. X_i is the inflow at time i . The formula for the ideal capacity is:

$$R_n = \max_{0 \leq k \leq n} \sum_{i=1}^k [X_i - \bar{X}_n] - \min_{0 \leq k \leq n} \sum_{i=1}^k [X_i - \bar{X}_n] \quad (4.23)$$

R_n is the adjusted range. In a second step the above formula will be standardized by

$$S_n = \sqrt{n^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2} \quad (4.24)$$

As a result we have the R/S statistic.

$$R/S = \frac{R_n}{S_n} \quad (4.25)$$

The interesting thing is that as Hurst plotted the log of R/S with large values of k the scattering was around a straight line with a slope above $1/2$ which means (and is in a sense the fractal property here).

$$\log E[R/S] \approx a + H \log n, \quad \text{with } H > 1/2 \quad (4.26)$$

This finding was in sharp contradiction with standard Markov processes where the time series is assumed to be independent. With stationary processes and just some short range dependence we should have $H^{1/2}$. However, Hurst found long term dependence for many hydrological, geophysical and climatological cases. Figure 4.1 shows Hurst's original findings.

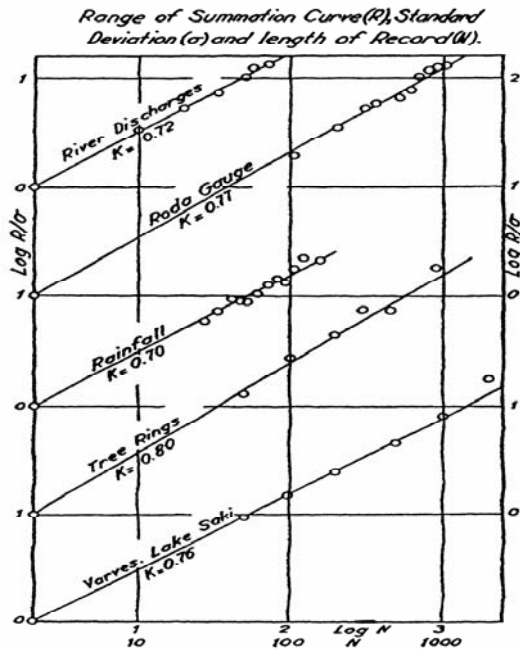


Figure 4.1: Hurst's original R/S analysis. In the derivation above H is equal to K

Source: Peters (1994)

4.2.1 Estimation of the Hurst exponent

Next, the R/S test is done along with a extension proposed by Lo (1991).

Instead of using S_n Lo uses $\hat{\sigma}_n(q)$

$$\hat{\sigma}_n^2(q) = \sigma^2 + 2 \sum_{j=1}^q 1 - \frac{j}{q+1} \hat{\gamma}_j \quad (4.27)$$

with σ^2 and $\hat{\gamma}_j$ as the sample variance and autocovariance estimators of X .

The truncation lag q can be chosen with respect to the data but should not be too small Lo (1991).

Stock	Hurst	Lo q 5	Lo q 10	Lo q 25
Telekom	0.36	5,16*	4,22*	3,08*
SAP	0.41	4,18*	3,62*	2,89*
Allianz	0,36	3,78*	3,09*	2,29*
Pfizer	0,41	3,29*	2,89*	2,37*
Intel	0,44	3,91*	3,49*	2,78*
Microsoft	0,39	3,74*	3,30*	2,64*

Table 4.1: Estimators for the Hurst exponent and Lo's alternative for six blue-chip stocks. Here daily data was used.

Table 4.1 shows the estimators for 6 “blue chip” stocks. Apparently the Hurst coefficient is lower than $\frac{1}{2}$ indicating rather negative persistence (i.e. a higher probability of positive returns to be followed by negative ones and vice versa). However, the Lo statistic does reject the null of no *long memory* in all cases. This finding shows that we still have to look for more powerful tests to detect

long range dependence. To improve the estimation technique one could also *bootstrap* H . This was done by Härdle et al (2005).

4.3 fractional Brownian motion

In this section a stochastic process will be introduced which resembles the property of self similarity explained in section two.

Definition 4.1: Beran (1994) defines a stochastic process as self similar if with any stretching factor c and the new time scale ct , $c^{-H}Y_{ct}$ has the same distribution as the original process.

This definition basically means that for any time point t_1, \dots, t_k and any positive constant c , $c^{-H}(Y_{ct_1}, Y_{ct_2}, \dots, Y_{ct_k})$ has the same distribution as $(Y_{t_1}, Y_{t_2}, \dots, Y_{t_k})$.

Definition 4.2: A stochastic process $B_H(t)$ with continuous sample path and the following properties is called fractal Brownian motion :

1. $B_H(t)$ is a Gaussian process (4.27)

2. $B_H(0) = 0$ (4.28)

3. $E[B_H(t) - B_H(s)] = 0$ (4.29)

4. $Cov[B_H(t), B_H(s)] = \frac{\sigma^2}{2}(|t|^{2H} - |t-s|^{2H} + |s|^{2H})$ (4.30)

for any $H \in (0,1)$ and σ^2 a variance scaling parameter. Only the covariance structure is different from the Gaussian case. A value of $H = 0,5$ means that we have the standard Brownian motion as a special case of the above process. The next charts show simulated series of fractional Brownian motion and fractional Gaussian noise which is the first difference of the former and stationary. The first simulation is for a short memory process ($H = 0,35$) and the second for a

long memory process ($H=0,75$). Strikingly, the ACF for the long memory example dies out in a hyperbolic way which is what we observe in reality and is shown in the Telekom example in section 1.1.2.

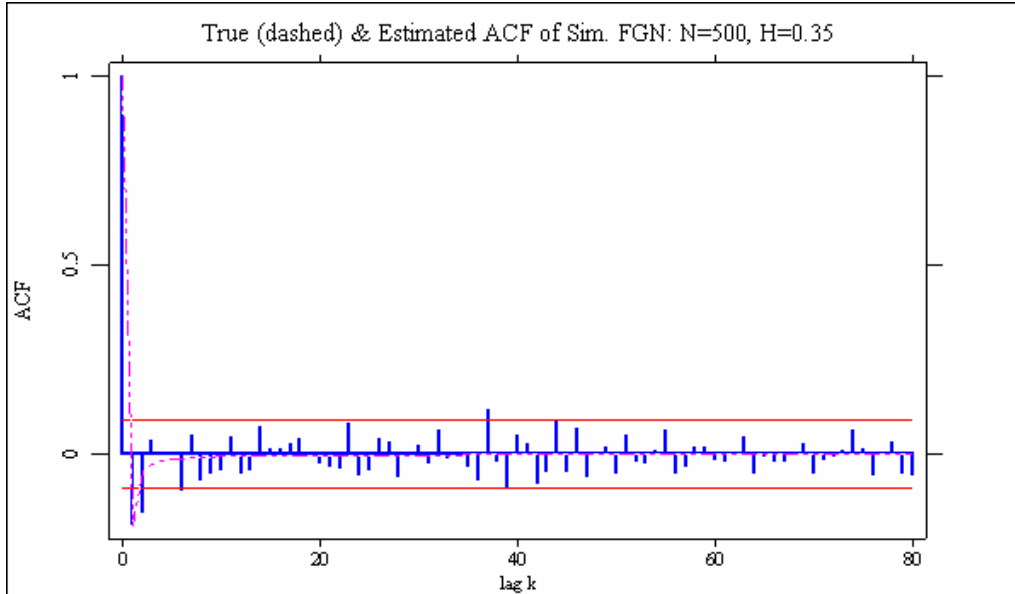


Figure 4.2: ACF for simulated fractional Gaussian noise with $H=0,35$. As opposed to a long memory process the ACF dies out exponentially. [XFGSimFBM.xpl](#)

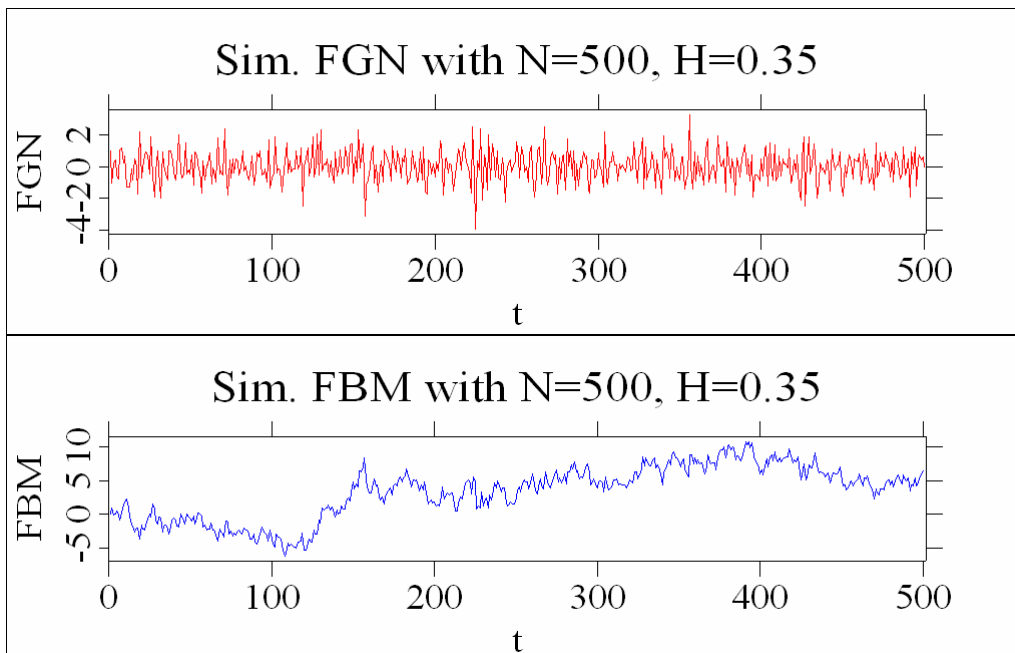


Figure 4.3: Simulated Fractional Gaussian Noise and Fractional Brownian Motion for a process with $H=0,35$ and hence short memory. [XFGSimFBM.xpl](#)

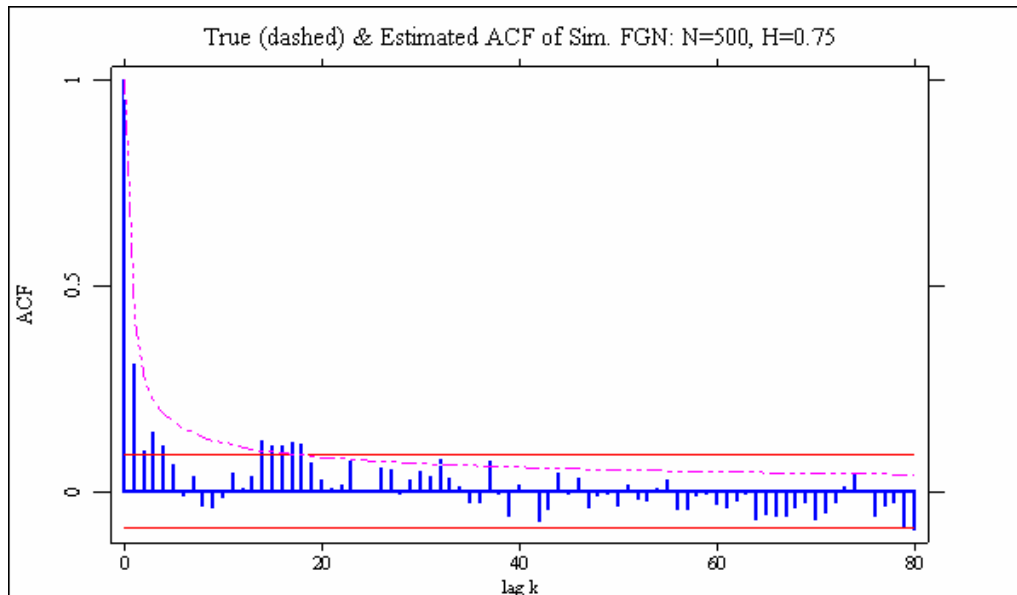


Figure 4.4: ACF for simulated fractional Gaussian noise with $H=0.75$. As empirical data show the ACF dies out exponentially.

[XFGSimFBM.xpl](#)

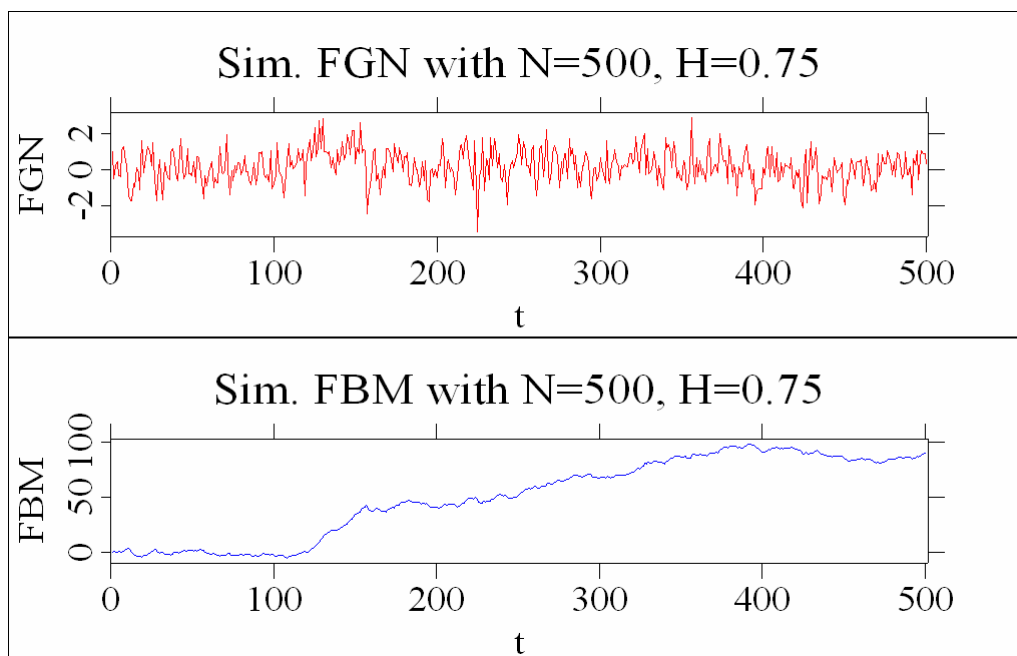


Figure 4.5: Simulated Fractional Gaussian Noise and Fractional Brownian Motion for a process with $H=0.75$ and hence short memory.

[XFGSimFBM.xpl](#)

4.4 Multiscaling of financial time series

Another power law is empirically observed for the autocovariance function (Lux, 2006).

$$E [|r_t - r_{t-\Delta t}|] \sim \Delta t^{-\gamma} \quad (4.31)$$

The value of γ estimated empirically is about 0,2-0,3 (Granger 1993). This power law is related to the analysed *Hurst exponent* from section 4.2.

Furthermore, the literature (Lux 2006) finds that returns can be described by *multi-fractality* meaning that markets exhibit *multi-scaling* properties. Different moments have different scaling laws. This can be shown such that:

$$E [|r_t^q - r_{t-\Delta t}^q|] \sim \Delta t^{-\gamma(q)} \quad (4.32)$$

A *multi-fractal brownian* process which accounts for such properties is shown in Appendix B. Lux (2006) has tested this model against GARCH, FIGARCH and stochastic volatility models and found a remarkable advantage when forecasting volatility.

5. Conclusion

In this thesis the scaling of returns and other financial time series was presented as an alternative to standard financial models. It was first shown that a few important assumptions these models make fail. Independence, normality of returns (also stationarity), rationality of agents and the market as a whole were among these. Most important, the efficient market hypothesis as a building block of “modern finance” also implies the correctness of prices at every time, which is hard state in such a volatile environment.

The notion of fractal and self-similarity was introduced and explained by some examples. Scaling and power laws provide a powerful environment to examine empirical time series. Physics use them to explain complex systems where fractal structures are relevant. In fact, Stanley et al develop a theory which is based on several power laws observed in finance. Some of them were studied in this thesis. One should build a theory on empirical facts and not the other way around.

It is very important for risk managers and traders in financial institutions to take these findings into account. The market seems much more riskier than assumed, even with a power law exponent of 3 which is not clear at least for the Dax-Future tick data set analysed in this thesis where a exponent of around 1,5 has been found.

The detection of long memory could not be confirmed with simple R/S but Lo’s alternative gave long memory in all cases. The topic remains important.

We still have to look for better techniques to model financial markets. One could be the proposed multifractal model of Lux (2006). However, a full explanation of markets is very hard to imagine, since we do not observe the generator behind it, only realizations, see Taleb (2007).

Appendix A

```
library("plot")
proc(ve)=p()

i = 1
xa0=330
ya0=150
ve = #(320,1)'#(1,200)'#(640,200)'#(xa0,ya0)'
x = xa0
y=ya0

while (i <10000)
  randomize2(i)
  r = uniform(1)

  if (r <0.34)

    x= x + (-(x-320)*(x<=320) - (x-320)*(x>320)).*0.5
    y= y - (y-1)/2

  else
    if (r>0.67)

      x= x-(x-1)/2
      y= y-(y-200)/2
    else

      x= x-(x-640)/2
      y= y-(y-200)/2

    endif
  endif
  ;x~y~r
  ve = ve|#(x,y)'

  i = i+1
endwhile
endp

obj = p()
pp = #(320,1)'#(1,200)'#(640,200)'|obj[100:rows(obj),]
disp = createdisplay(1,1)
show(disp,1,1,setmask(pp, "tiny"))
```


Appendix B

Here, some power law mathematics will be presented. The aggregation properties are especially interesting for theoretical and empirical work (Stanley et al, 2003).

A random variable has a power law behaviour if there is a $\varsigma_x > 0$ such that

$$P(X > x) \sim \frac{1}{x^{\varsigma_x}}$$

The probability density is then :

$$p(x) \sim \frac{1}{x^{\varsigma_x+1}}$$

One could also express the above definition as a slowly varying function $L(x)$ and a ς_x s.t. $p(x) \sim L(x)/x^{\varsigma_x+1}$, so that the tail follows a power law up to logarithmic corrections. $\varsigma_x < \varsigma_y$ is the expression for Y having fatter tails than X , thus large X 's are more frequent than large Y 's. A implication is that with $\alpha > 0$ and, $E\left[|X|^\alpha\right] = \infty$ for $\alpha > \varsigma_x$ and $E\left[|X|^\alpha\right] < \infty$ for $\alpha < \varsigma_x$. For example if returns $\varsigma_r = 3$, then $E\left[|r|^\alpha\right]$ for $\alpha > 3$. The Kurtosis of the returns is then infinite and the skewness almost infinite.

A technical advantage of power laws is that the property as a power law is valid after addition, multiplication, polynomial transformation and min, max. If we combine two power laws the smaller one (with the fatter tail) dominates.

$$\varsigma_{X+Y} = \varsigma_{X \times Y} = \varsigma_{\max(X,Y)} = \min(\varsigma_X, \varsigma_Y)$$

=power law of the fattest variable

$$\varsigma_{\min(X,Y)} = \varsigma_X + \varsigma_Y$$

If X is a power law ζ_x for $\zeta_x < \infty$, and Y is a power law variable with an exponent $\zeta_x \geq \zeta_y$ then $X + Y$, $X \times Y$, $\max(X, Y)$ are still power laws. This is also true if Y would be normal, lognormal or exponential (i.e. $\zeta_y = \infty$).

An advantage here is that estimating power law 's is still possible even if we have a lot of noise in the data (Stanley et al 2003). Small effects do not affect the power law exponent.

Appendix C

The model mentioned in part 4.4 takes the relation to turbulent flows also exhibiting multiscaling properties. To model the break-off of small eddies to bigger ones one starts with a uniform probability measure over the unit interval $[0,1]$. In the first step, this interval is split up into two subintervals (smaller eddies) which receive fractions p_1 and $1-p_1$ of their „mother intervals” (Lux 2006). This procedure is repeated infinitely. It generates a heterogeneous structure in which the final outcome after n steps of ever smaller eddies can take any value $p_1^m p_2^{n-m}$, $0 \leq m < n$. The process is highly autocorrelated, since on average several joint components. In the limit $n \rightarrow \infty$ multi-fractality can be shown to hold.

This Markov-Switching Multi-Fractal process (MSM) is a special case of Markov-Switching and stochastic volatility models. Returns over a unit time interval are modelled as:

$$r_t = \sigma_t u_t$$

with innovations u_t drawn from a standard Normal distribution $N(0,1)$ and instantaneous volatility σ_t being determined by the product of k volatility components or multipliers, $M_t^{(1)}, M_t^{(2)}, \dots, M_t^{(K)}$ and a constant scale factor δ :

$$\sigma_t^2 = \sigma^2 \prod_{i=0}^k M_t^{(i)}$$

Each volatility component is renewed at time t with probability

$$\gamma_i = 1 - (1 - \gamma_1)^{(b^k - 1)}$$

with parameters $\gamma_i \in [0,1]$ and $b \in (1,\infty)$. This specification is derived as a discrete approximation to a continuous-time multi-fractal process with Poisson arrival probabilities and geometric progression of frequencies. It can be shown when that when the grid size goes to zero, the above model converges to a continuous-time process.

References

- Bachelier, L. (1900). Theorie de la speculation, *Annales de l'Ecole Normale Supérieure*
- Beran, J. (1994). *Statistics for Long-Memory Processes*, New York Chapman & Hall
- Black, F. (1986). Noise, *Journal of Finance*, 41, 529-543
- Carhart, M. (1997). On Persistence in Mutual Fund Performance, *Journal of Finance* 52(1), 57-82
- Chipman, J. (1976). The Paretian Heritage, Working paper, University of Minnesota
- Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues, *Quantitative Finance* Volume one 223-236
- Crovella M.E., Taqu M.S., BestavosA. *Heavy-Tailed Distributions in the World Wide Web*, in Adler, R.J., Feldman, R.E. and M.S. Taqu (2000) *A Practical Guide to Heavy Tails*, Birkhäuser
- Cuthbertson, K. (1996). *Quantitative Financial Economics*, Wiley and Sons
- Cuthbertson, K. , Nitzsche, D. & O'Sullivan, N. (2005). Mutual Fund Performance : Skill or Luck?, *Working Paper Cass Business School, City University London*
- De Bondt, W. and Thaler R. H. (1985). Does the Stock Market Overreact, *Journal of Finance*, 793-805
- Fama, E. , French K. (1992). The Cross section of expected Stock Returns *Journal of Finance*, 47 (1992) 427-466
- Fama, E. (1991). Efficient capital markets. II, *Journal of Finance*, 46, 1575-617
- Granger, C. (1966). The typical spectral shape of an economic variable, *Econometrica*, 34 (1), 150-161
- Granger, C., Z.R. Ding, A Long Memory Property of Stock Market Returns and a new Model, *Journal of empirical Finance*, 1,1993, 83-106

- Grossman, S., Stiglitz, J. (1980). On the impossibility of informationally efficient markets. *American Economic Review*, 70, 39-408
- Härdle, W., Hall, P., Kleinow, T., Schmidt, P. (2000). Semiparametric bootstrap approach to hypothesis tests and confidence intervals for the Hurst coefficient, *Statistical Inference for stochastic Processes* 3
- Härdle, W., Borak, S., Weron, W. (2005). *Stable Distributions* in Härdle, W., Cizek, P., Weron, W. (2005) *Statistical Tools for Finance and Insurance*, Springer Berlin
- Hurst, H.E. (1951). Long-term storage capacity of reservoirs, Transactions of the Society of Civil Engineers, 116, 770-799, 800-808
- Jegadeesh, N., Sheridan, T. (1993). Returns to Buying Winners Selling Losers: Implications for Stock Market Efficiency, *Journal of Finance*, 48, 65-91
- Jensen, M. (1968). The Performance of Mutual in the Period 1945-1964, *Journal of Finance*, 23, 389-416
- Jensen, M. (1978). Some Anomalous Evidence Regarding market Efficiency, *Journal of Financial Economics*, 6, 95-102
- Kahnemann, D., Tversky, A. (1979). Prospect theory: An analysis of decision under risk, *Econometrica*, 47, 263-291
- Kaizoji, T., Kaizoji M. (2004). Power law for ensembles of stock prices, *Physica A*, 240-243
- Koutrouvelis, I.A. (1980) *Journal of American Statistical Association* 75, 918
- Lo, A.W. (1991). Long term memory in stock market prices, *Econometrica*, 59, 1279-1313
- Lux, T. (2006). Applications of Statistical Physics in Finance and Economics, forthcoming in the *Journal of Economic Dynamics and Control*
- Mandelbrot, B. (1982). *The Fractal Geometry of Nature*, Freeman San Francisco
- Mantegna, R., Stanley H. Scaling Behaviour in the Dynamics of an economic index, *Nature*, 46-49
- Odean, T. (1998). Are investors reluctant to realize their losses?, *Journal of Finance*, 53, 1775-98
- Peters, E. (1994). *Fractal Market Analysis*, Wiley & Sons

- Schroeder, M. (1991). *Fractals, Chaos and Power Laws*, Freeman
- Shleifer, A. (2000). *Inefficient Markets: An Introduction to Behavioral Finance*, Oxford University Press, New York
- Shleifer, A., Vishny, R. (1997). The limits of Arbitrage, *Journal of Finance*, 52, 35-55
- Shiller, R. (2004). *Irrational Exuberance*, second Edition, Princeton University Press
- Sornette, D. (2003). *Why Stock Markets Crash, Critical Events in Complex Financial Systems*, Princeton University Press
- Stanley, H.R. et. al. (1996) Scaling behavior in the growth of companies, *Nature*, 379, 804-806
- Stanley, H.R. et. al. (2003). A Theory of Large Fluctuations in Stock Market Activity, *MIT Working Paper No. 03-30*
- Taleb, N.N., (2007). *The Black Swan*, Random House
- Thaler, R. (2002). *A survey of behavioural finance*, in Advances in Behavioural Finance, Princeton University Press
- Weron, W. (2001). Levy-Stable Distributions Revisited: Tail Index $\alpha > 2$ does not exclude the Levy-Stable Regime, *International Journal of Modern Physics*, 12, 209-223