

Fluctuation Phenomena on the Stock Market

Ph.D. thesis

Zoltán Eisler

Budapest University of Technology and Economics
Department of Theoretical Physics

Supervisor:

Prof. János Kertész

Budapest

September 14, 2007

Contents

1	Preface	5
2	Introduction	7
2.1	Financial markets	8
2.2	Stylized facts	10
3	Introduction to fluctuation scaling	13
3.1	Fluctuation scaling	14
3.2	Empirical results	15
3.3	Universal values of α	18
3.4	Summary of observations	20
4	Scaling theory and size dependent...	21
4.1	New stylized facts of trading activity	22
4.2	Scaling theory	28
4.3	Conclusions	33
4.4	Appendix: The components of the fluctuation σ^2	35
5	Stylized facts revisited	37
5.1	Capitalization and basic measures of trading activity	38
5.2	Traded value distributions revisited	40
5.3	Correlations in traded value time series	42
5.4	Multiscaling distribution of intertrade times	48
5.5	Conclusions	51
5.6	Appendix: The estimation of tail exponents	53
6	Introduction to order books	57
7	Time scales in order books	61
7.1	Monthly and longer time scales	62
7.2	Daily time scale	64
7.3	Intraday and tick-by-tick time scales	68
7.4	Conclusions	72

8	Diffusive approximation...	73
8.1	The first passage time	74
8.2	Time to fill, time to cancel	76
8.3	A simple model	83
8.4	The predictions of the model	86
8.5	The validity of the lifetime process	90
8.6	A generalization to $\Delta \leq 0$	93
8.7	Conclusions	95
9	Summary	97
9.1	Background	97
9.2	Goals	98
9.3	Methods	99
9.4	New scientific results	99
10	Összefoglaló (Summary in Hungarian)	101
10.1	A kutatások előzménye	101
10.2	Célkitűzések	102
10.3	Vizsgálati módszerek	102
10.4	Új tudományos eredmények	103

Köszönetnyilvánítás

A dolgozatot szüleimnek ajánlom.

Ez a munka nem jöhetett volna létre Kertész János belém vetett bizalma és állandó támogatása nélkül. Köszönöm Mód Melindának a sok türelmet és megértést az utazásaim és a számítógép előtt töltött éjszakák alatt. Köszönettel tartozom társszerzőimnek Rosario Mantegnának, Jaume Masolivernek, Fabrizio Lillonak, Josep Perellónak, Barabási Albert-Lászlónak és Bartos Imrének az ötleteikért és tanácsaikért. Továbbá köszönöm Jánosi Imrének a bátorítást és kritikai megjegyzéseit. Hálás vagyok a l'Ecole de Physique des Houches-nak, ahol az ajánlati könyvről szóló fejezet jelentős részét készítettem és Jean-Philippe Bouchaudnak a meghívásért. Köszönet Maria, Claudia és Angelo Coronnellonak, valamint Michele Tumminellonak a Palermóban töltött hetekért és Antonios Garasnak a megannyi ebédért. A droidszoba lakóinak: Kitartás, egyszer kiszabadultok.

A kutatást támogatta: OTKA K60456, OTKA T049238, COST-STSM-P10-917, MIUR "Dinamica di altissima frequenza nei mercati finanziari" és NEST-DYSONET 12911.

Acknowledgments

This thesis is dedicated to my parents.

This work would not have been possible without the confidence and unending support of János Kertész. I thank Melinda Mód for all the patience and understanding during my travels and nights spent in front of the computer. I owe thanks to my coauthors Rosario Mantegna, Jaume Masoliver, Fabrizio Lillo, Josep Perelló, Albert László Barabási and Imre Bartos for their ideas and advice. Furthermore I thank Imre Jánosi for his encouragement and critical remarks. I am grateful to the Ecole de Physique des Houches, where I wrote a significant part of the chapter on order books, and to Jean-Philippe Bouchaudnak for inviting me there. I thank Maria, Claudia and Angelo Coronnellonak and Michele Tumminellonak for the weeks spent in Palermo, and Antonios Garasnak for countless lunches ebédért. To my fellow PhD students: Hang in there, eventually you'll earn your freedom.

Support was provided by: OTKA K60456, OTKA T049238, COST-STSM-P10-917, MIUR "Dinamica di altissima frequenza nei mercati finanziari" and NEST-DYSONET 12911.

Chapter 1

Preface

In the last decade tens of books and thousands of research papers have been published by physicists in the field of finance. Such publication records reflect the conviction that these and similar ideas and techniques will be helpful to understand the mechanisms of the economy [BP00, MS99, Man97]. This new trend is only one of many fueled by the breakthrough of the early 70's in statistical physics. The advances of this period brought up several concepts and models like (fractal and multifractal) scaling, frustrated disordered systems, or far from equilibrium phenomena and we have obtained very efficient tools to treat them. But how reliable can insights be if they are based on principles that apply to particles and we would like to extend them to social or economic systems?

A generalization of ideas from physics to finance is, the least to say, counterintuitive. Physics is the science of clearly defined natural forces that are unchanged in time, can be isolated with sufficient care, and can be experimented with. Finance is often very pragmatic and less of a science than an art. Moreover, "experiments" on the stock market can be extremely costly, if at all possible, so one has to be satisfied with passive observation instead. Finally, the economy is never in a steady state, its characteristics change without end, and it is difficult to understand why and for how long any non-trivial observation remains valid. The current level of understanding is also very different between the two fields. While physics has some generally accepted models that have very good predictive power, no such models exist for human behavior and the financial markets that it controls.

This is not to say that economics lacks abundant theoretical background, but when contrasted with each other, the neoclassical economic theory and the way physicists look at markets (or physics itself for that matter) are worlds apart. Physicists are extremely critical of economics and they feel that the theory does not put enough emphasis on consistency with the observations. From the other side, according to economists,

"physicists suffer from the belief that there must be universal rules" [Bal06] combined with a "Tarzan complex" to solve any problem that arises in any field outside their own [Far99].

Despite all these conflicts it appears that physics, or as this area is sometimes called "econophysics", indeed has a contribution to make. At this point such a contribution is not as much to solving key questions that thrill economists as to providing innovative techniques to look at financial data. These techniques appear to be welcome by practitioners in banking and investment, but it is clear, that there is much room for improvement. And naturally, physicists already have their eyes on a more ambitious goal: a "microscopic" theory of markets, as much as quantum mechanics is a microscopic theory of matter. Even though no one can tell whether there is any hope for one, there is a lot to gain from still trying.

This thesis has been the product of more than three years of research which hoped to contribute to these early efforts. Beyond statistical finance, it was inspired by several other areas in complex systems, such as complex networks, random processes and population dynamics. While this manuscript attempts to give a broader overview of financial fluctuations, it has a clear focus on three subjects: financial data and non-universality, the scaling properties of fluctuations, and the temporal dynamics of the limit order book. This is why I decided to include three separate chapters discussing the respective background information, these are called "Introduction" which is more general, "Introduction to fluctuation scaling" and "Introduction to order books". The remaining chapters contain my own results.

Chapter 2

Introduction

There exist predominantly two approaches to the analysis of investment decisions: *fundamental analysis* and *technical analysis*. The aim of "fundamentalists" is to evaluate public information such as the financial statements of the company in order to estimate its "true" value per share. If, e.g., this is higher than the current stock price, then the stocks are undervalued in the market and this might be a good opportunity to buy them. The catch is of course that even if the company's assets are worth more than what we pay for the stocks, one cannot just liquidate the company and obtain this money. Technical analysis is quite different: Technical traders (also called "chartists") are trying to infer market trends instead, which reflect the market's valuation of the company's shares, and which can be to a large extent unrelated to the actual value of the company's assets. One extreme example is WorldCom, whose asset value was gradually inflated by 12 billion USD during the years 2000–2002 by misinformation and accounting fraud. The eventual fall and bankruptcy of the company in 2002 is one of the greatest scandals in US stock market history. The other key difference is that fundamental analysis takes a relatively long-term approach, as the relevant changes and decisions that affect a company's outlook happen on the time scale of months. Shifts in the investors' expectations occur much more rapidly, and technical indicators are often used to analyze day-to-day or even minute-to-minute changes.

The approach of technical analysis is akin to how physics relates to experimental results. Both involve extensive statistical analysis of an observed set of data. In finance data are generated by market activity, such as past prices and trading volume. These are compared with some benchmark models, based on which conclusions can be drawn. The actual models in technical analysis are extremely varied and often phenomenological. They range from trend lines and charts of past prices to quantitative indicators and time series models.

In the broadest sense there are two strategies for obtaining profit by

speculation: market making and directional trading. These have a very clear difference. Anyone with a market making strategy is in the market for both buying and selling the same share at the same time, of course at different prices to ensure some profit. Meanwhile a directional trader is in the market either to buy or to sell, but never both simultaneously for the same share.

One of the highest esteemed principles of finance and possibly all economics is the Efficient Market Hypothesis. There are several ways to postulate it, but all forms represent the common notion that markets digest all information infinitely fast, so it becomes immediately and perfectly represented in prices. Thus it should be impossible to make any profit from directional trading, because the market price is always the "fair" price. Practice certainly shows otherwise.

There are significant profits to be made. To provide an idea about their order of magnitude, let us consider a simple example set forth in Ref. [FL99]. One-month U.S Treasury Bills are considered one of the safest investments in the world. If in January, 1926, an individual invested 1 USD in these, and continued reinvesting the money until December, 1996, the original investment would have grown to 14 USD. If, on the other hand, the individual invested the same 1 USD in the Standard and Poor's 500 stock market index, an arguably much riskier asset, for the same 71-year period, the sum would have grown to 1,370 USD. Now let us suppose that the individual had a perfect forecasting ability, and each month it was able to choose whichever of the two investments would provide a higher profit. A simple calculation shows that after the same 71-year period this strategy would have provided a whopping 2,296,183,456 USD of final wealth.

In the light of these numbers it is easy to see that even some very limited forecasting ability could result in large gains. In practice these are difficult to realize, and the opportunities are short-lived, but they exist. Obviously, excess profits are not available to all market participants. This is very easy to see: because technical trading is to a large extent unrelated to any growth of the underlying worth of companies there is no value produced: i.e., if someone wins money, that comes from someone else losing it. This pressure on trading firms to continuously innovate and renew their strategies in order to remain profitable is probably the single most important reason while physicists and their fresh ideas are welcome in this field.

2.1 Financial markets

Even though their ultimate functions are the same, different stock markets can have very different structures. The textbook example is the comparison between the New York Stock Exchange and NASDAQ. NYSE trades

the shares of the best known companies such as Citigroup or General Electric, while NASDAQ trades mostly technological shares like those of Intel or Microsoft. An even more important difference is how they are organized. The NYSE is an auction market that uses floor brokers for the majority of the transactions. Each stock has a so called "specialist", who is an actual person responsible for all the trades for that particular stock. To them transactions are forwarded through a floor broker or the DOT system, which is a computer system that relays orders from trading firms directly to the specialist. A transaction happens when someone is willing to pay the buy or sell at the sell quote of the specialist. The specialist does not only mediate between buyers with sellers, but himself holds an inventory of the stock to balance demand and supply. He compensates his expenses by the difference between his buy and sell quotes.

In contrast, the NASDAQ is not a physical entity. It is a purely electronic over-the-counter (OTC) market, where anyone is allowed to trade. Liquidity is provided by market makers. Their role is similar to a specialist's, but they do not have to be appointed by the exchange, and important stocks can have tens of them. NASDAQ functions as a communication network of computers that broadcast the market makers' buy and sell quotes. Transactions can be initiated via online execution systems, and market makers compete for the orders of customers by their prices published on these systems.

If an order (i.e., an offer to buy/sell) cannot be traded immediately because its price is too low/high, it is stored in a database called the limit order book, until prices change to a level where it can be executed. When there are several channels (e.g., different computer networks) for order submission it is possible to have several independent order books for the same share. Order books will be discussed in more detail in Chapter 6.

For the purposes of technical analysis it is essential to have access to as much information as possible. Currently there is a range possibilities, typically the following:

- Low-frequency data: Daily or hourly opening/closing/lowest/highest transaction prices and trading volumes.
- High-frequency, transaction-level (so called tick-by-tick or Level I) data: The price and volume of every transaction, plus the best bid/ask at every time.
- Partial book data: Typically the best five bid/ask levels and the volume available (also called Level II data).
- Complete book data: The full history of all transactions and orders.

The costs of obtaining these data grow very rapidly with the level of detail.

Daily data are usually freely available online, while a subscription to good quality order book data can cost tens of thousands of dollars per month.

The empirical work in this thesis is based on two datasets. For Chapters 4 and 5 we used the TAQ database [taq] which records all transactions of the New York Stock Exchange and the NASDAQ during the period 1993 – 2003. For Chapters 7 and 8 we used the complete order book data of London Stock Exchange for the year 2002.

Time will always be measured by continuous trading time, i.e., the time periods between market closure and open are ignored for the calculation of durations. Finally, note that throughout the thesis we use 10-base logarithms.

2.2 Stylized facts

Financial time series display certain general properties. These are considered highly robust to the variation of many parameters, such as which market and stock we are looking at, and even the time period under investigation as long as there are no major alterations in trading rules. Although a significant portion of this thesis will be devoted to a critical inquiry into them, it is inevitable to outline here some of these supposedly universal properties, or as they are usually called, the *stylized facts*.

The majority of both finance and "econophysics" studies concentrates on the behavior of price. Let us consider $i = 1 \dots I$ stocks traded simultaneously at the same market. For stock i the price at time t will be denoted by $p_i(t)$. The most investigated quantities are without doubt the return r and the logarithmic return x . These are defined for some fixed time scale Δt as

$$r_i^{\Delta t}(t) = \frac{p_i(t) - p_i(t - \Delta t)}{p_i(t - \Delta t)} \quad (2.1)$$

and

$$x_i^{\Delta t}(t) = \log p_i(t) - \log p_i(t - \Delta t). \quad (2.2)$$

A first order Taylor expansion of Eq. (2.2) easily shows that $r_i^{\Delta t}(t) \approx x_i^{\Delta t}(t)$ as long as the changes are small. Compared to the price change $p_i(t) - p_i(t - \Delta t)$, x and r have the advantage that they are dimensionless and the value of the stock is scaled out.

The logarithmic return time series $x^{\Delta t}(t)$ for $\Delta t < 1$ year is known to have a fat tailed, non-normal distribution which means that large price changes are much more common than one would expect if the logarithmic price followed a Brownian motion with Gaussian increments. There have been attempts to describe the decay of the density function by various analytic forms, such as a truncated Lévy distribution [MS94], or as an asymptotic power law [GMAS98, PGA⁺99]. As one increases Δt the fat

tails are diminished, and the distribution becomes approximately Gaussian for yearly or longer time horizons [BP00, MS99, KTK⁺99].

There is overwhelming evidence that returns are not serially correlated beyond the time scale of a few minutes. If they were, that would result in an (at least statistical) predictability of future price changes, and it would be possible *for everyone* to earn a higher profit than the mean return $\langle r^{\Delta t} \rangle$ without any additional risk. As we pointed out earlier, this is not possible because because the total wealth of all investors can only grow with the mean return.

This is not to say, that returns are independent. For the disappearance of serial autocorrelations it is enough, if the *signs* of the returns are independent. Other quantities, such as absolute or squared returns indeed have very long memory. Qualitatively this means that small returns are usually followed by small returns, while large returns also tend to group together. The typical size of returns at a given time is called *volatility* and this long memory property is *volatility clustering*. There exists a very large number of definitions and empirical regularities for volatility. The explanation here is intentionally limited, because volatility will not be investigated quantitatively in this thesis.

Instead, let us turn to the properties of trading activity. For a given time period $[t, t + \Delta t)$ the total traded value (activity) of the stock i can be calculated as

$$f_i^{\Delta t}(t) = \sum_{n=1}^{N_i^{\Delta t}(t)} V_{i,n}^{\Delta t}(t), \quad (2.3)$$

where $N_i^{\Delta t}(t)$ is the number of transactions with stock i during $[t, t + \Delta t)$, and $V_{i,n}^{\Delta t}(t)$ is the value of the n 'th of those. The value of a transaction is calculated as the product of the transaction price, and the number of exchanged stocks $[\tilde{V}_{i,n}^{\Delta t}(t)]$.

The price usually changes only a little from trade to trade, while the number of stocks traded in consecutive deals varies heavily. Thus, the fluctuations and the statistical properties of the transaction value V are basically governed by those of \tilde{V} . Price only serves as a conversion factor to US dollars, that makes the comparison of the shares of various companies possible. This way, one also automatically corrects the data for stock splits¹. The statistical properties (normalized distribution, correlations, etc.) are otherwise practically indistinguishable between transaction volume and value.

¹It is called a stock split, when a company's existing shares are divided into multiple shares. The number of shares thus increases by a factor, while stock price decreases by the same factor. Splits usually happen when a company's share price is so high that to many investors the shares are too expensive to buy in the standard lots of 100. Similarly, there are "inverse splits" when several shares are combine into one, and the price is increased accordingly.

The 1-minute time average of f will be denoted by $\langle f \rangle$. Although there exists a large number of other definitions, the quantity $\langle f \rangle$ can be understood as a simple measure of the liquidity of the stock. We will use this quantity as the proxy of liquidity.

Both the traded value $[f^{\Delta t}(t)]$, and the number of transactions $[N^{\Delta t}(t)]$ are non-negative, and they both have broad distributions. Refs. [GPGS00] and [PGA⁺00] present a detailed statistical analysis of these two, and conclude that their (non-cumulative) distributions decay asymptotically as $\mathbb{P}(f) \propto f^{-(1+\lambda)}$ with $\lambda = 1.53 \pm 0.07$ up to $\Delta t = 1$ day, and $\mathbb{P}(N) \propto N^{-(1+\lambda_N)}$ with $\lambda_N = 3.40 \pm 0.05$. It is especially important to emphasize, that according to these studies the distribution of the traded value f has infinite variance and obeys the generalized central limit theorems related to Lévy stable distributions. Both quantities are reported to have long range power-law autocorrelations.

As with prices and volatility, there exist a number of additional stylized facts for trading activity. However, this short introduction focused only on those needed in this thesis.

Chapter 3

Introduction to fluctuation scaling

Interacting systems of many units with emergent collective behavior are often termed "complex". Such complex systems are ubiquitous in many fields of research ranging from engineering sciences through physics and biology to finance. An advantage of the related multi-disciplinary approach is that the universal appearance of several phenomena can be revealed more easily. Such generally observed characteristics include (multi-) fractality or scale invariance [Vic92, BTW87], the related Pareto or Zipf laws [Par96, Zip29], self-organized and critical behavior.

In this chapter we introduce such a general feature related to the scaling properties of the fluctuations in complex systems. This type of scaling relationship is called Taylor's law by ecologists after L.R. Taylor and his influential paper on natural populations in 1961 [Tay61]. The law states that for any fixed species the fluctuations in the size of a population (characterized by the standard deviation) can be approximately written as a constant times the average size of that population to a power α :

$$\text{fluctuations} \approx \text{const.} \times \text{average}^\alpha$$

for a wide range of the average.

The phenomenon was – to our knowledge – first discovered by H. Fairfield Smith in 1938 [Smi38], who wrote an equivalent formula for the yield of crop fields though his paper has, surprisingly, received much less attention than Taylor's work. The same relationship was explored recently by Menezes and Barabási [dMB04a] for dynamics on complex networks, and later termed "fluctuation scaling" [EK06a] in the physics literature. There the temporal fluctuations and the averages of the network's traffic were measured at the different nodes.

3.1 Fluctuation scaling

Let us consider some additive, positive quantity f , and the dependence between its mean and standard deviation. (While the discussion here will be quite general, for the purposes of this thesis one can think of f as the traded value of some stock, and in this case the notations will be consistent with the other chapters.) By dependence we mean the behavior of f over a multitude of observations. Say, if we can observe the same dynamical variable in several settings where it has different means, how does the standard deviation change with the value of the mean?

In order to determine this dependence one needs many realizations. These can be simultaneous temporal observations for different elements (nodes, subsystems) of a large complex system. The measured means and standard deviations are then calculated in time¹, and the subsystems are compared: for subsystems with a larger mean f are the fluctuations larger as well?

In a general context we will call the points of measurement as *nodes*. The additive quantity under study, be it activity, population, traffic, or whatever else, will be denoted by f_i , where i indicates the node of measurement. The time average of f over time windows of size Δt can be calculated as

$$\langle f_i^{\Delta t} \rangle = \frac{1}{Q} \sum_{q=0}^{Q-1} f_i^{\Delta t}(q\Delta t) = \frac{1}{Q} \sum_{q=0}^{Q-1} \sum_{n=1}^{N_i^{\Delta t}(q\Delta t)} V_{i,n}^{\Delta t}(q\Delta t), \quad (3.1)$$

where $Q = T/\Delta t$, and T is the total time of measurement. From the definitions it is trivial that $\langle f_i^{\Delta t} \rangle = \Delta t \langle f_i^{\Delta t=1} \rangle$. We will use $\langle f_i \rangle$ without the upper index to denote this latter quantity.

On any time scale the variance can be obtained as a time average:

$$\sigma_i^2(\Delta t) = \langle [f_i^{\Delta t}]^2 \rangle - \langle f_i^{\Delta t} \rangle^2,$$

this quantity characterizes the fluctuations of the activity of a fixed node i from interval to interval.

It is often observed that the relationship between the standard deviation and the mean of f is given by a power law:

$$\sigma_i(\Delta t) \propto \langle f_i^{\Delta t} \rangle^\alpha,$$

where one varies the node i , and Δt is fixed. The dependence of the right hand side on Δt is trivial, since $\langle f_i^{\Delta t} \rangle \equiv \Delta t \langle f_i \rangle$. Thus throughout this

¹In other cases f is *not* considered as time dependent, only as a fixed value for every subsystem. Then the averages are taken over an ensemble of subsystems of equal size, and the standard deviation characterizes the variation of f between subsystems of the same size. We will not discuss this possibility here.

work we will use $\langle f_i \rangle$ as the scaling variable:

$$\sigma_i(\Delta t) \propto \langle f_i \rangle^\alpha. \quad (3.2)$$

We will call this relationship *fluctuation scaling* (FS in short). The exponent α is usually in the range $[1/2, 1]$.

3.2 Empirical results

For the collection of data appropriate for FS it is necessary to have multi-channel measurements, simultaneously monitoring the behavior of a range of elements. With the unbroken growth of computing infrastructure, many technological networks now offer appropriate datasets, several ones publicly available. A summary of such datasets is shown in Table 3.1.

Complex networks

Menezes and Barabási [dMB04a, dMB04b], in part inspired by Taylor’s original paper, found FS for several complex networks. Menezes and Barabási analyzed web page visitations, river flow, microchip logical gates and highway traffic. They proposed that the datasets should fall into two “universality classes” with $\alpha = 1/2$ and 1. There also exists a growing body of literature on transport processes on networks, and the scaling of fluctuations in such systems [EK05, DA06, YdM05, vT06].

Another good example is the analysis of Internet traffic [dMB04a], which was later revisited by Duch and Arenas [DA06]. In their study they analyzed the traffic of the Abilene backbone network. The nodes i correspond to routers, and the mean and variance of their data flow was calculated. In Fig. 3.1 we show their results for weekly data traffic, the best fit is achieved with $\alpha \approx 0.75$.

Ecology

Ecologists have made many discoveries regarding FS, but the literature is far from unequivocal. The basic concept is to monitor many populations of a given species for an extended period of time. Then for each population i one calculates the temporal mean $\langle f_i \rangle$ and standard deviation σ_i of abundance. These are typically power law related according to FS, examples are shown in Fig. 3.2.

Classical population dynamics offers several benchmark models [Lot25, Vol25, May74], but simple deterministic and Markovian models cannot explain the observed α values between $1/2$ and 1. After a range of small populations where they show realistic behavior, they cross over to either $\alpha = 1/2$ or 1 [Kee00]. The model of Kilpatrick and Ives [KI03] suggested that the interaction between species and feedback mechanisms between

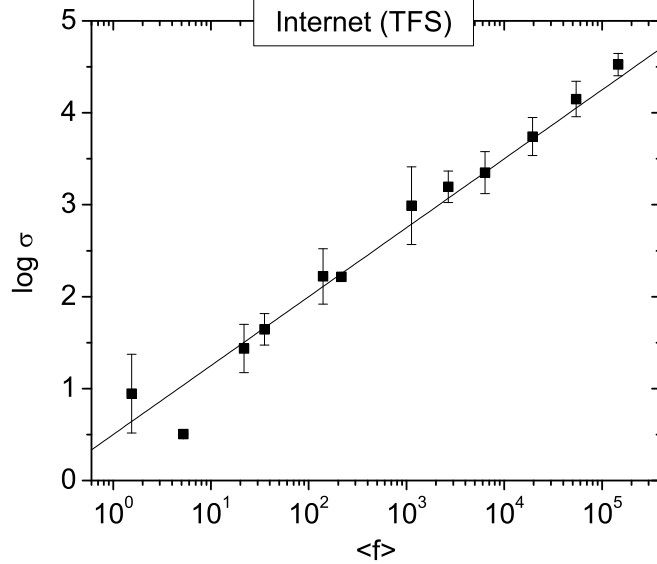


Figure 3.1: Fluctuation scaling for the traffic of Internet routers. For every router the temporal average of weekly data traffic $\langle f_i \rangle$ and its variance σ_i^2 was calculated separately. Then these points were binned logarithmically for better visibility. The fitted exponent is $\alpha = 0.75$. Data courtesy of Jordi Duch and Alex Arenas [DA06].

their fluctuations can give rise to any value of α . Perry proposed an even simpler chaotic model [Per94]. Both of these models can yield various exponents, but still only when populations are small enough.

There have been several findings for plant species. In a series of papers Ballantyne and Kerkhoff showed that the reproductive (yearly seed count) variability of trees follows FS with $\alpha \approx 1$. Here we consider three subsets of the dataset [upo], those collected by Tallqvist [Tal78], Franklin [Fra68] and Weaver and Forcella [WF86], including 4 – 17 years of observations for 44, 148 and 28 sites, respectively. The fits for FS are given in Fig. 3.3. The exponents for the three subsets were found to be $\alpha = 0.97$, 0.93 and 0.90. Given the quality of the fits it is not possible to outrule that for all three datasets $\alpha = 1$ (as suggested by Ref. [Kp03]), although there have been attempts for a theory that predicts this deviation [EBK07].

The same value $\alpha = 1$ is found for the Satake-Iwasa [SI00] forest model. There the trees are modeled by interacting oscillators which synchronize above a critical value of the coupling [pK05, pK07]. The synchronization transition² coincides with a transition from $\alpha = 1/2$ to $\alpha = 1$.

²Similar synchronization mechanism has also been observed in the reproduction of

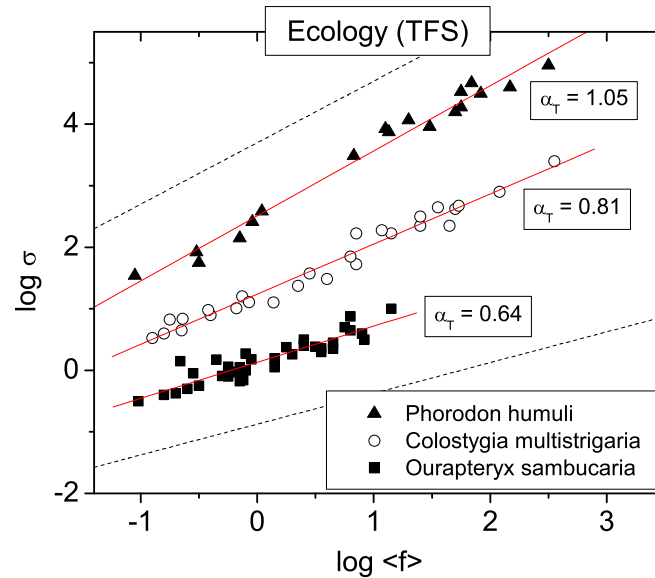


Figure 3.2: Fluctuation scaling of the population of three species. A point represents the temporal mean $\langle f_i \rangle$ and variance σ_i^2 of a population. The bottom dashed line corresponds to $\alpha_T = 1/2$, the top one to $\alpha_T = 1$. Points were shifted both vertically for better visibility. Data courtesy of Marm Kilpatrick [KI03, Tay86].

Life sciences

Keeling and Grenfell [KG99] suggested FS for the size of epidemics, and found both empirically and by a simple Markov chain model of population dynamics that vaccination in general decreases not only the size of epidemics but also the value of α . FS was later found by Woolhouse *et al.* [WTH01] to also hold between different pathogens.

FS has been found in the cell-to-cell variation of protein transcription by Bar-Even *et al.* [BEPM⁺06], albeit with a crossover and a rather narrow range. In particular, in the bacterium *Saccharomyces cerevisiae* the proteins with higher mean abundance also tend to exhibit greater fluctuations of their level, and the dependence can be fitted by a power law.

animals [GWF⁺98].

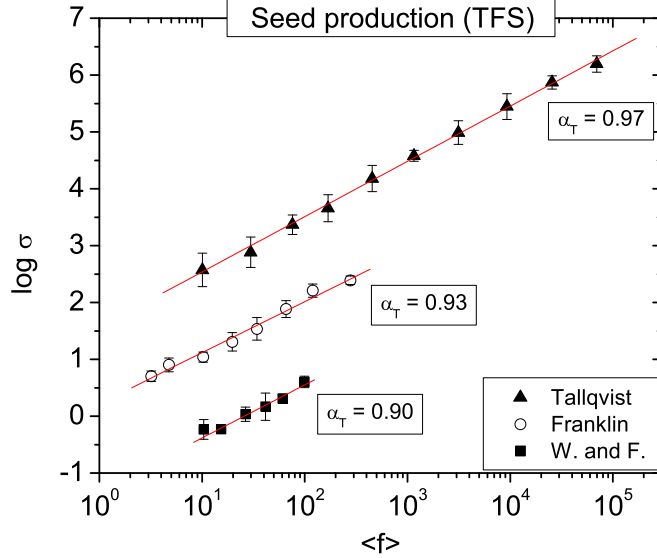


Figure 3.3: Fluctuation scaling for the yearly seed count (reproductive activity) of trees from three studies. The fitted exponents are $\alpha = 0.90, 0.93, 0.97$. Points were logarithmically binned and $\log \sigma$ was averaged for better visibility, the error bars represent the standard deviations inside the bins. The estimates are close to, but below 1.

Other areas

Fluctuation scaling was found in various electronic databases of human activity. Ref. [EBK07] considers the number of email sent and also the number of documents printed by individuals, and finds that for very short time windows $\alpha = 1/2$, while longer scale fluctuations display higher, non-trivial values.

The daily precipitation measured at weather stations also displays approximate fluctuation scaling [EBK07] with $\alpha \approx 0.73$, but the scaling plots are very broad. Substantial corrections to scaling are due to geographical position and height above sea level.

3.3 Universal values of α

There is evidence, that the exponent α carries information about the collective dynamical properties of the whole system. Based on this knowledge, a classification scheme was outlined in Refs. [dMB04a, EK05, KE05b]. Those studies assume, that the activities of all nodes are temporally un-

Subj.	System	Refs.
Networks	Random walk	[dMB04a, dMB04b, EK05]
	Network models	[vT06, YdM05]
	Highway network	[dMB04a, dMB04b]
	World Wide Web	[dMB04a, dMB04b]
	Internet	[dMB04a, dMB04b, DA06]
S./e.	Stock market	[EKYB05, EK06a, EK06b, ZQJZ07]
	Email traffic	[EBK07]
	Printing activity	[EBK07]
Cl.	River flow	[JG99, DJ05]
	Precipitation	[EBJ, EBK07]
Ecology	Forest reproductive rates	[Kp03, pK07]
	Satake-Iwasa forest model	[pK05]
	Animal populations	[Tay61, AGCH82, RH04, MT02]
	Population growth	[KS98, KABS02]
	Interacting population model	[KI03]
Life sc.	Protein expression	[BEPM ⁺ 06]
	Gene expression	[NOA05, ZTWT06]
	Epidemiology	[KG99, WTH01]

Table 3.1: A list of some studies where fluctuation scaling was directly applied or implied by a similar formalism. Groups were assigned by subject areas, S./e. = Sociology/economics, Cl. = Climatology.

correlated³. In this case, there are two known universality classes with respect to the value of α .

In certain systems, the activity comes from nearly equivalent, independent events. The difference between nodes with smaller and greater mean activity comes basically from the different mean *number* of events. Then, the *central limit theorem* can be applied to these events and this yields $\alpha = 1/2$ automatically. Examples include simple surface growth models and the data traffic of Internet routers [dMB04a].

Other systems' dynamics is under a *dominant external driving force*: Activity fluctuations are mainly caused by the variations of this external force, and this leads to proportionality between the average and the standard deviations at the nodes: $\alpha = 1$, regardless of the internal structure or the laws governing the time evolution. This is observed for the statistics of web page visitations and highway traffic [dMB04a].

Several processes are known to give rise to intermediate measured values $1/2 < \alpha < 1$: For example, some finite systems display a crossover be-

³Note that in general, instead of uncorrelated dynamics, it is enough if the activity of every node displays the same Hurst exponent, $H_i = H$. This is the direct consequence of arguments in Section 4.2.

tween $\alpha = 1/2$ and $\alpha = 1$ at a certain node strength $\langle f \rangle^*$, due to the competition of external driving and internal dynamics [dMB04a, dMB04b]. There is an *effective* value of α , but in fact, scaling breaks down. Other possible scenarios will be discussed in Chapter 4.

3.4 Summary of observations

In sum, fluctuation scaling appears to be a surprisingly general concept that can be recognized in virtually any discipline where the proper data are available. The fluctuations of positive additive quantities appear to have the structure

$$\text{fluctuations} \approx \text{const.} \times \text{average}^\alpha.$$

There exist two known "universality" classes with respect to the value of α . The central limit theorem leads to $\alpha = 1/2$, while strongly driven systems display $\alpha = 1$. However, many real systems fall between these two values, and for non-trivial reasons.

Chapter 4

Scaling theory and size dependent fluctuations in stock market data

There is immense literature on the origins of fluctuations in various systems, ranging from gene networks [Pau04] through complexity [dMB04b] to animal populations [pG01, EBI98, pTE⁺00, GWF⁺98]. The common point of all these works is that fluctuations originate from two factors: internal and external. Naturally, the dynamics, the structure, and the interaction of the nodes vary from case to case. We expect that, e.g., Internet router traffic and the reproduction of trees is very different. The discovery of FS as a common pattern can be a good start to point out further analogies and to build a broader picture.

In several studies, the value of α is used as a proxy of the dominant factors of internal dynamics [dMB04a, dMB04b, EKYB05]. While the general idea works well in a number of settings, most cases when $1/2 < \alpha < 1$ are poorly understood. This chapter contains a parallel, step-by-step analysis of data from both NYSE and NASDAQ from this point of view.

Section 4.1 presents some new stylized facts regarding stock market trading activity, and then Section 4.2 connects all those observations with fluctuation scaling and its previously identified two universality classes. We deal with a mechanism that explains how stock markets can display a non-universal value of $\alpha \approx 0.68$. Finally, we describe how dynamical correlations are reflected in the time scale dependence of the exponent α .

4.1 New stylized facts of trading activity

4.1.1 Size-dependent correlations

The presence of long-range autocorrelations in various measures of trading is a well-known fact [EK06b, IYPL04, YI05]. For example, stock market volatility [BP00, MS99, Con01] and trading volumes [GPGS00, EK06b] show strong persistence. Correlations in stock i can be characterized by the Hurst exponent H_i [Vic92, KZKB⁺02] defined as

$$\sigma_i(\Delta t) = \left\langle [f_i^{\Delta t}(t) - \langle f_i^{\Delta t}(t) \rangle]^2 \right\rangle \propto \Delta t^{H_i}, \quad (4.1)$$

where $\langle \cdot \rangle$ denotes time averaging. There is a range of methods [KZKB⁺02, PBH⁺94, MBA91] to estimate the Hurst exponent, and the understanding of the results is well established [Vic92]. The signal is said to be correlated (persistent) when $H_i > 1/2$, uncorrelated when $H_i = 1/2$, and anticorrelated (antipersistent) for $H_i < 1/2$.¹

It is intriguing, that stock market activity has a much richer behavior, than simply all stocks having Hurst exponents statistically distributed around an average value, as assumed in Ref. [GPGS00]. Instead, there is a crossover [EK06b, IYPL04, YI05] between two types of behavior. We located this threshold by a technique that will be discussed in Section 4.2.2. An essentially uncorrelated regime was found when $\Delta t < 60$ min for NYSE and $\Delta t < 2$ min for NASDAQ, while the time series of larger companies become strongly correlated when $\Delta t > 390$ min for NYSE and $\Delta t > 60$ min for NASDAQ. As a reference, we also calculated the Hurst exponents H_i^{shuff} of the shuffled time series. Results are given in Fig. 4.1.²

One can see, that for shorter time windows, correlations are practically absent in both markets, $H_i \approx 0.51 - 0.53$. For windows longer than the crossover time, however, while small $\langle f \rangle$ stocks again display only very weak correlations, larger ones show up to $H \approx 0.9$. Furthermore, there is a distinct logarithmic trend in the data:

$$H_i = H_*^\pm + \gamma_t^\pm \log \langle f_i \rangle. \quad (4.2)$$

¹Note that $H_i \neq 0.5$ is possible even for memoryless processes, given that they do not have stationary increments [MRGS00]. To underline that our results indeed originate from correlations, we applied the Detrended Fluctuation Analysis technique [KZKB⁺02] to calculate $\sigma_i(\Delta t)$. This method uses piecewise polynomial fits to remove non-stationarities from the data, and often produces good estimates for H . We varied the order of the detrending polynomials in the range $d = 1 \dots 5$. Higher order polynomials should have been able to filter out more of the non-stationarity, whereas our findings did not depend strongly on d .

²We also investigated the effect of randomly shuffling the Fourier-phases of the data, which destroys the possible non-linearities in the time series. One finds, that the crossover behavior persists after such a transformation.

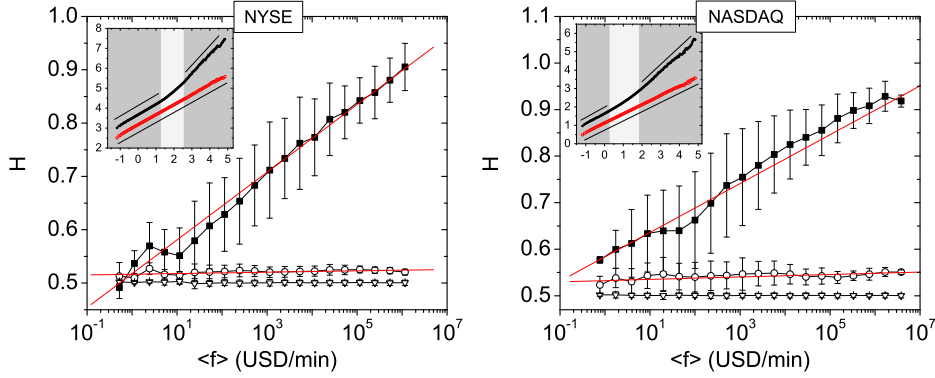


Figure 4.1: **(left)** Behavior of the Hurst exponents H_i in the period 2000 – 2002. For short time windows (\circ), all signals are nearly uncorrelated, $H_i \approx 0.51 - 0.52$. For larger time windows (\blacksquare), the strength of correlations depends logarithmically on the mean trading activity of the stock. Shuffled data (∇) display no correlations, thus $H_i^{\text{shuff}} = 1/2$, which also implies $\gamma_t = 0$. **(left)** Results for NYSE. The fitted slopes are $\gamma_t^-(\Delta t < 60 \text{ min}) = 0.001 \pm 0.002$ and $\gamma_t^+(\Delta t > 390 \text{ min}) = 0.06 \pm 0.01$. The inset shows the $\log \sigma - \log \Delta t$ scaling plot for General Electric (GE). The slopes corresponding to Hurst exponents are 0.53 and 0.93; the slope for shuffled data is 0.51. Shuffled points were shifted vertically for better visibility. **(right)** Results for NASDAQ. The fitted slopes are $\gamma_t^-(\Delta t < 2 \text{ min}) = 0.003 \pm 0.002$ and $\gamma_t^+(\Delta t > 60 \text{ min}) = 0.05 \pm 0.01$. The inset shows the $\log \sigma - \log \Delta t$ scaling plot for Dell (DELL). The slopes corresponding to Hurst exponents are 0.54 and 0.90; the slope for shuffled data is 0.50. Shuffled points were shifted vertically for better visibility.

The indices $-/+$ correspond to $\Delta t < 60 \text{ min}$ and $\Delta t > 390 \text{ min}$ with $\gamma_t^- \approx 0$ and $\gamma_t^+ = 0.06 \pm 0.01$ for NYSE. Similarly, $\Delta t < 2 \text{ min}$ and $\Delta t > 60 \text{ min}$ with $\gamma_t^- \approx 0$ and $\gamma_t^+ = 0.05 \pm 0.01$ for NASDAQ. Shuffled data, as expected, show $H_i^{\text{shuff}} \approx 1/2$ on all time scales and without significant dependence on $\langle f_i \rangle$.

It is to be emphasized, that the crossover is not simply between uncorrelated and correlated regimes. It is instead between homogeneous (all stocks show $H_i \approx H_*$, $\gamma_t = 0$) and inhomogeneous ($\gamma_t > 0$) behavior. One finds $H_*^\pm \approx 1/2$, but very minor $\langle f \rangle$ stocks do not depart much from this value even for large time windows, while major ones do. This is a clear relation to liquidity, which (as in Section 5.1 we will show) is a monotonically growing function of company capitalization [EK06b].

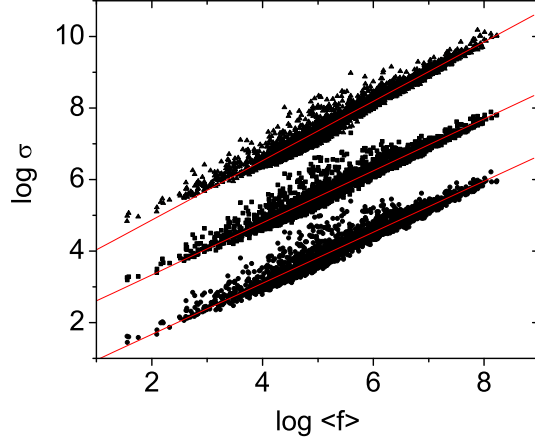


Figure 4.2: Examples of $\sigma(\langle f \rangle)$ scaling plots for NYSE, years 2000 – 2002. The window sizes from bottom to top: $\Delta t = 10$ min, 1/2 day, 2 weeks. The slopes are $\alpha = 0.70, 0.71, 0.80$, respectively. Points were shifted vertically for better visibility.

4.1.2 Fluctuation scaling of f

Due to the long time period investigated here (3 years), the data are highly instationary. Thus, unlike our earlier studies [EKYB05], here we applied the DFA procedure [KZKB⁺02, PBH⁺94] to estimate $\sigma_i(\Delta t)$. We determined the values of α for traded value fluctuations by fits to Eq. (3.2), examples are shown in Fig. 4.2.

The exponent α strongly depends on the size Δt of the time windows. Recently Refs. [ZAK04, ZAK06, YI05] pointed out that the trading activity of NYSE and NASDAQ display very different temporal correlations, possibly due to their different trading mechanisms. Still, fluctuation scaling does hold regardless of market and Δt . Furthermore, the functions $\alpha(\Delta t)$ agree qualitatively. The exponents are shown for NYSE and NASDAQ in Fig. 4.3. One can see, that α is a non-decreasing function of Δt , and in large regimes it is, to a good approximation, either constant or logarithmically increasing.

4.1.3 Multiscaling

Fluctuation scaling can be generalized to higher central moments. In this case one assumes that

$$\sigma_i^q = \langle |f_i - \langle f_i \rangle|^q \rangle \propto \langle f_i \rangle^{q\alpha(q)}. \quad (4.3)$$

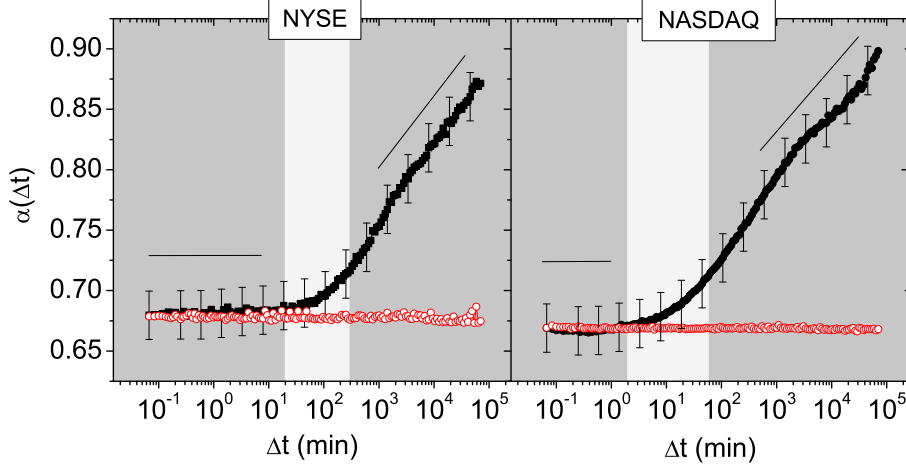


Figure 4.3: The dependence of the scaling exponent α on the window size Δt . The darker shaded intervals have well-defined Hurst exponents and values of γ_t , the crossover is indicated with a lighter background. **(left)** NYSE: without shuffling (\blacksquare) the slopes of the linear regimes are $\gamma_f^-(\Delta t < 60 \text{ min}) = 0.00 \pm 0.01$ and $\gamma_f^+(\Delta t > 390 \text{ min}) = 0.06 \pm 0.01$. For shuffled data (\circ) the exponent is independent of window size, $\alpha(\Delta t) = 0.68 \pm 0.02$. **(right)** NASDAQ: without shuffling (\blacksquare) the slopes of the linear regimes are $\gamma_f^-(\Delta t < 2 \text{ min}) = 0.00 \pm 0.01$ and $\gamma_f^+(\Delta t > 60 \text{ min}) = 0.06 \pm 0.01$. For shuffled data (\circ) the exponent is independent of window size, $\alpha(\Delta t) = 0.67 \pm 0.02$. *Note:* There is a deviation from linearity around $\Delta t \approx 1$ trading week. It is larger for NASDAQ, but it is still between the error bars. A possible cause is the weekly periodic pattern of trading which was not removed manually.

This means that all q 'th order central moments of the activity, which characterize fluctuations around the mean behavior, scale as power-laws with the mean. In this notation $\alpha(2)$ corresponds to α in Eq. (3.2). Figure 4.4 shows some examples that such scaling is actually observed. The $\alpha(q; \Delta t)$ dependence is non-trivial, but qualitatively similar to the case of $q = 2$. This is shown in Fig. 4.5.

4.1.4 Fluctuation scaling of N and V

One can carry out a fluctuation scaling analysis of other quantities, here we limit ourselves to two of those. The first one, the number of trades of stock i in size Δt time windows, which was denoted by $N_i^{\Delta t}(t)$, and its variance by $\sigma_{N_i}^2(\Delta t)$. The second one was also introduced before, $V_{i,n}^{\Delta t}(t)$

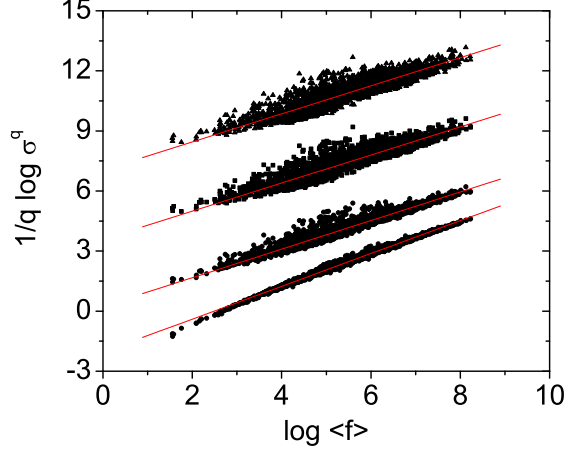


Figure 4.4: Examples of $1/q \times \sigma^q$ versus $\langle f \rangle$ scaling plots for NYSE, years 2000–2002. The window size is always $\Delta t = 10$ min. From bottom to top: $q = 1, 2, 3$ and 8 . The slopes are $\alpha = 0.83, 0.72, 0.67, 0.68$, respectively. Points were shifted vertically for better visibility.

is the value exchanged in the n 'th trade of stock i . The corresponding variance will be $\sigma_{V_i}^2$.

Dimensional analysis predicts

$$\sigma_{V_i}^2 \propto \langle V_i \rangle^2, \quad (4.4)$$

which is remarkably close to the observed behavior, shown in Fig. 4.6(left). Also, when the size of the time windows is chosen sufficiently small ($\Delta t \ll 1$ min), the probability that two trades of the same stock happen in the same period is negligible. In this limit $N^{\Delta t} = 0$ or 1 and $\langle N^{\Delta t} \rangle \ll 1$. Then $\langle N^{\Delta t} \rangle \approx \langle [N^{\Delta t}]^2 \rangle$, and thus

$$\sigma_N^2(\Delta t) = \langle [N^{\Delta t}]^2 \rangle - \langle N^{\Delta t} \rangle^2 \approx \langle N^{\Delta t} \rangle, \quad (4.5)$$

which again agrees very well with empirical data shown for $\Delta t = 1$ sec in Fig. 4.6(right).

4.1.5 Dependence of typical trade size on trading frequency

The final observation to be discussed here is that for a large group of stocks, the average rate of trades $\langle N \rangle$ and their mean value $\langle V \rangle$ are connected by a power law:

$$\langle V_i \rangle \propto \langle N_i \rangle^\beta. \quad (4.6)$$

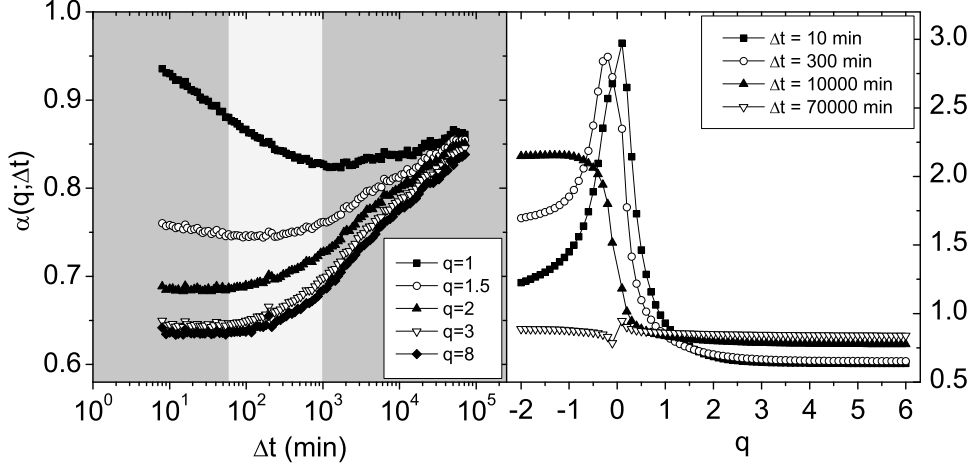


Figure 4.5: **(left)** The dependence of the scaling exponents $\alpha(q; \Delta t)$ on the window size Δt . In the the darker intervals the dependence is approximately logarithmic, an intermediate crossover regime is indicated with a lighter background. **(right)** The dependence of the scaling exponents $\alpha(q; \Delta t)$ on the power q .

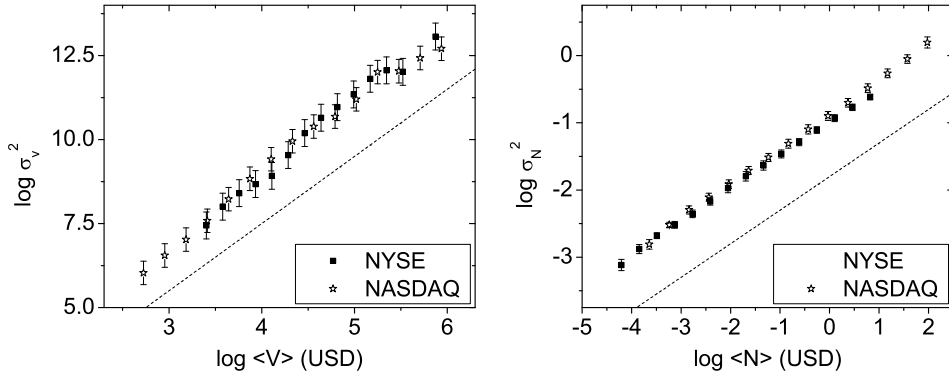


Figure 4.6: **(left)** Plot verifying the validity of (4.4) for stock market data, typical error bars are given. The straight line would correspond to $\sigma_{V_i}^2 \propto \langle V_i \rangle^2$. **(right)** Plot verifying the validity of (4.5) for stock market data, typical error bars are given. The straight line would correspond to $\sigma_{N_i}^2 \propto \langle N_i \rangle$. The size of the time windows is $\Delta t = 1$ sec.

Such relationships are shown in Fig. 4.7 for both NYSE and NASDAQ. The measured exponents are $\beta_{\text{NYSE}} = 0.59 \pm 0.09$ and $\beta_{\text{NASDAQ}} = 0.22 \pm 0.04$, although they are restricted to large enough stocks. The estimate based on Ref. [Zum04] for the stocks in London's FTSE-100, is $\beta \approx 1$.

The values of β_{NYSE} and β_{NASDAQ} , and especially the marked difference between them appears to be very robust for various time periods [EK06c]. One major contribution to this is probably the difference in trading mechanisms between the two markets [upo, YI05] (cf. Section 2.1).

One very crude interpretation of the existence of $\beta > 0$ in general is the following. Smaller stocks are exchanged rarely, but transaction costs must limit from below the value that is still profitable to be exchanged at once. This minimal unit is around the order of 10^4 USD for both markets. Once the speed of trading and liquidity grow, it becomes possible to exchange larger packages. Trades start to "stick together", their average value starts to grow. Although this tendency reduces transaction costs, the price impact [GGPS03, PGG04, FL04, FGL⁺04] of the trade also increases, which in practice often limits package sizes from above. These two mechanisms may have a role in the formation of (4.6). Also, as they vary strongly from market to market, such very different values of β might be justified.

4.2 Scaling theory

In this section we present a framework that unifies the – seemingly unrelated – observations of Section 4.1. This is centered around fluctuation scaling:

$$\sigma_i(\Delta t) \propto \langle f_i \rangle^{\alpha(\Delta t)}.$$

As this phenomenon is not at all specific to stock market data, the following arguments can be applied to a wide range of complex systems [EBK07]. We will only use the language of finance in order to make the arguments easier to follow.

4.2.1 Non-universal values of α

The activities $f_i(t)$ originate from individual trades that take place for each stock. For a given size Δt of time windows, the observed time series is given by

$$f_i^{\Delta t}(t) = \sum_{n=1}^{N_i^{\Delta t}(t)} V_{i,n}^{\Delta t}(t). \quad (2.3)$$

If the random process that represents the size of a trade is independent from the one that determines when the trade occurs, one can find a simple

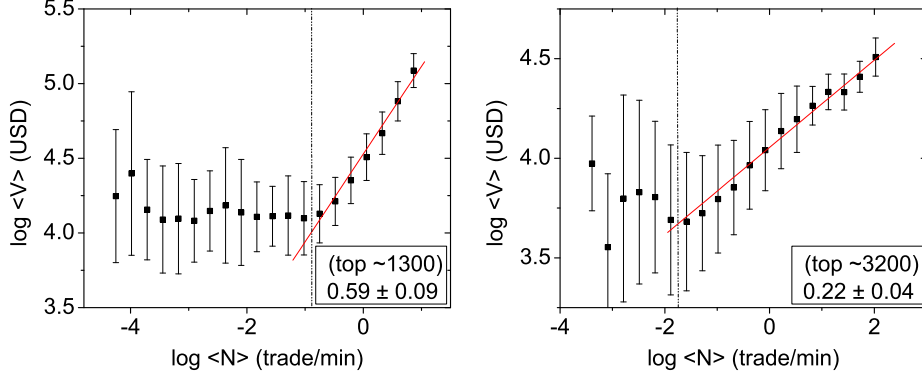


Figure 4.7: The dependence of the mean value per trade $\langle V_i \rangle$ on the average rate of trades $\langle N_i \rangle$. Calculations were done for the period 2000–2002, **(left)** shows NYSE and **(right)** shows NASDAQ. Points were binned and their logarithm was averaged for better visibility, error bars show the standard deviations in the bins. For the smallest stocks there is no clear trend at either exchange. However, larger stocks at NYSE and all except the minor ones at NASDAQ, show scaling between the two quantities, equivalent to that given in (4.6). The slopes are $\beta_{\text{NYSE}} = 0.59 \pm 0.04$ and $\beta_{\text{NASDAQ}} = 0.22 \pm 0.04$.

formula (see Appendix 4.4) that shows how fluctuations of f are composed:

$$\sigma_i^2(\Delta t) = \sigma_{N_i}^2(\Delta t) \langle V_i \rangle^2 + \sigma_{V_i}^2 \langle (N_i^{\Delta t})^{2H_{V_i}} \rangle, \quad (4.7)$$

where $\langle V_i \rangle$ and $\sigma_{V_i}^2$ are the mean and the standard deviation of the transaction size distribution. $\langle N_i^{\Delta t} \rangle$ and $\sigma_{N_i}^2(\Delta t)$ are similar, only for the number of events in time windows of length Δt . Under these conditions, it is also trivial, that $\langle f_i \rangle = \langle N_i \rangle \langle V_i \rangle$.

All the above can be expected from simple principles. Two more relationships are necessary and are often realized, they are basically the same as (4.4) and (4.5). The only strong assumption to account for non-universal values of α is the following. Recall that stocks with higher average activity do not only experience more trades, but those are also larger. Moreover, there scaling between the two quantities:

$$\langle V_i \rangle \propto \langle N_i \rangle^\beta. \quad (4.6)$$

Then, α can be expressed [EK05], by combining all the formulas, as

$$\alpha = \frac{1}{2} \left(1 + \frac{\beta}{\beta + 1} \right). \quad (4.8)$$

This expression is true as long as $H_{V_i} = 1/2$ or the second term dominates in Eq. (4.7). Alternatively, in stock market data for $\Delta t = 1$ sec $N_i = 0$ or 1 and thus $N_i^{2H_{V_i}} = N_i$, which has the same effect as if $H_{V_i} = 1/2$.

The intermediate values $1/2 < \alpha < 1$ interpolate between the square root type of $\langle N \rangle \propto \sigma_N^{1/2}$ and the linear $\langle V \rangle \propto \sigma_V$, while the conditions ensure that scaling is preserved (compare Figs. 4.2 and 4.6).

Similar conditions are satisfied exactly in a random walker model on complex networks [EK05]. Consequently, its behavior is well described by (4.8). However, such arguments can also be applied to stock market trading dynamics when $\Delta t \ll 1$ min to ensure the validity of (4.5). By substituting the observed values of β , one finds the estimates $\alpha_{\text{NYSE}}^* = 0.69 \pm 0.03$ and $\alpha_{\text{NASDAQ}}^* = 0.59 \pm 0.02$. The actual values are $\alpha_{\text{NYSE}}(\Delta t \rightarrow 0) = 0.68 \pm 0.02$ and $\alpha_{\text{NASDAQ}}(\Delta t \rightarrow 0) = 0.67 \pm 0.02$. The agreement for the NYSE data is good, for NASDAQ it is only approximate. Moreover, Eq. (4.6) only fits the data for large enough stocks, while FS gives an excellent fit over the whole range available. Therefore, this explanation is only partial, however, it indicates that $\alpha > 1/2$ is to be expected.

4.2.2 Time scale dependence of α

Section 4.1 revealed, that the exponent α of stock market activity fluctuations shows a strong dependence on the time window Δt . In an early study this was attributed to the effect of external factors [EKYB05]. On the time scale of minutes news, policy changes, etc. have no time to diffuse in the system. Thus, temporal fluctuations are dominated by internal dynamics, $\alpha < 1$. By increasing Δt to days or weeks, the importance of this external influence grows and α approaches 1, which is characteristic in the presence of strong external driving. However, the effect just described is a crossover, while observations show the persistence of scaling, only the exponent α changes. This section offers an alternative description that has no such shortcoming.

The key is to extend the analysis to $H_i \neq 1/2$ systems. We start from the relations (3.2) and (4.1). These define the Hurst exponent H_i and the FS exponent α in terms of $\sigma_i(\Delta t)$. The roles of the two variables Δt and $\langle f_i \rangle$ are analogous in the two equations. When they hold simultaneously, from the equality of their left hand sides, one can write the third proportionality

$$\Delta t^{H_i} \propto \langle f_i \rangle^{\alpha(\Delta t)}.$$

After taking the logarithm of both sides, differentiation as $\partial/\partial(\log \Delta t)$ and $\partial/\partial(\log \langle f_i \rangle)$ yields the asymptotic equality

$$\gamma_t \sim \frac{dH_i}{d(\log \langle f_i \rangle)} \sim \frac{d\alpha(\Delta t)}{d(\log \Delta t)} \sim \gamma_f. \quad (4.9)$$

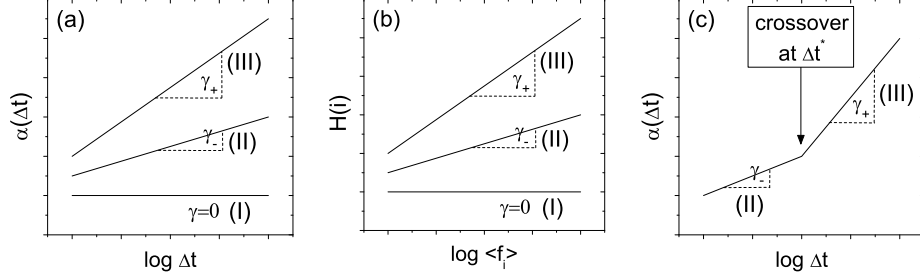


Figure 4.8: **(a-b)** Possible scenarios where both $\sigma_i(\Delta t) \propto \Delta t^{H_i}$ and $\sigma_i(\Delta t) \propto \langle f_i \rangle^{\alpha(\Delta t)}$ can be satisfied simultaneously. (I) In systems, where $\gamma = 0$, α is independent of window size and H is independent of node. (II) When $\gamma = \gamma^- > 0$, $\alpha(\Delta t)$ and H_i depend logarithmically on Δt and on $\langle f_i \rangle$, respectively, with the common slope γ_1 . (III) For a larger value, $\gamma = \gamma^+ > \gamma^-$, the dependence is stronger. **(c)** Example of a crossover between different values of γ . There, α still depends on Δt in a logarithmic way, but the slope is different in two regimes. In this case, for every node there are two Hurst exponents, H_i^- and H_i^+ , that are valid asymptotically, for $\Delta t \ll \Delta t^*$ and $\Delta t \gg \Delta t^*$, respectively. Then, both of these must independently follow the logarithmic law shown in (b): $H_i^- = H_*^- + \gamma^- \log \langle f_i \rangle$ and $H_i^+ = H_*^+ + \gamma^+ \log \langle f_i \rangle$.

This means that both partial derivatives are constant and they have the same value, which we will denote by $\gamma = \gamma_t = \gamma_f$.

The possibilities how this can be realized are sketched in Figures 4.8(a)-(b):

- (I) In systems, where $\gamma = 0$, the exponent $\alpha(\Delta t) = \alpha^*$, it is independent of window size. At the same time all nodes must exhibit the same degree of correlations, $H_i = H$.
- (II) In the case, when $\gamma = \gamma^- > 0$, $\alpha(\Delta t)$ actually depends on Δt . This dependence must be logarithmic: $\alpha(\Delta t) = \alpha_*^- + \gamma^- \log \Delta t$. At the same time, the Hurst exponent of the nodes depends on the mean flux in a similar way: $H_i = H_* + \gamma^- \log \langle f_i \rangle$. Moreover, the slope of the logarithmic dependence is the same.
- (III) When the constant γ is larger, for example $\gamma^+ > \gamma^-$ in Figures 4.8(a)-(b), α changes faster with Δt , while also H_i changes faster with $\langle f_i \rangle$.

Finally, the combination of these options is also possible. Systems may display a crossover between different values of γ at a certain time scale Δt^* , an example is given in Figures 4.8(b)-(c). There, α depends on Δt in

a logarithmic way, but the slope of the trend is different in two regimes. In this case, there is no unique Hurst exponent of $f_i(t)$. Instead, for every node there are two values, H_i^- and H_i^+ , that are valid asymptotically for $\Delta t \ll \Delta t^*$ and $\Delta t \gg \Delta t^*$, respectively. Then, both of these must independently follow the logarithmic law: $H_i^- = H_*^- + \gamma^- \log \langle f_i \rangle$ and $H_i^+ = H_*^+ + \gamma^+ \log \langle f_i \rangle$.

Stock markets belong to this last group. For $\Delta t \leq 60$ min for NYSE and $\Delta t \leq 2$ min for NASDAQ, $\alpha(\Delta t) \approx \alpha^*$. Correspondingly, H must be independent of $\langle f \rangle$, as it was found in Section 4.1. On the other hand, for $\Delta t > 390$ min for NYSE and $\Delta t > 60$ min for NASDAQ, $\alpha(\Delta t)$ is approximately logarithmic with the common coefficient $\gamma = 0.06 \pm 0.01$. This, again, must equal the slope of H_i plotted versus $\log \langle f_i \rangle$. There is agreement between error bars with the results of Section 4.1.

The fact that the local derivative $\frac{d\alpha(\Delta t)}{d(\log \Delta t)}$ also shows the degree of logarithmic trend in the Hurst exponents, gives a visual method to detect the change in this collective behavior of the market. Those regimes in Δt , where $\alpha(\Delta t)$ is constant, correspond to time scales where all stocks have the same level (Hurst exponent) of activity correlations. Where $\alpha(\Delta t)$ is logarithmically changing, the slope γ gives the degree of inhomogeneity in H_i . Finally, the function is curved near crossovers, where the degree of the mean flux dependence in correlation strengths is changing.

In order to underline, that the $\alpha(\Delta t)$ dependence comes from temporal correlations, we carried out the same measurement, but with all time series shuffled randomly. It is trivial, that if Δt equals the $\delta = 1$ sec resolution of the dataset, shuffling does not affect the estimates of $\sigma_i(\Delta t = \delta)$, it merely rearranges the terms used in averaging³. Hence, the fitted slope cannot change either, $\alpha_{\text{shuff}}(\delta) = \alpha(\delta)$. On the other hand, shuffling gives uncorrelated time series, $H_i^{\text{shuff}} \equiv 1/2$ (see Section 4.1). Correspondingly, $\gamma_{\text{shuff}} = \frac{dH_{\text{shuff}}}{d\log \langle f \rangle} = 0$. Hence, according to (4.9), $\alpha_{\text{shuff}}(\Delta t) = \alpha^*$, regardless of window size. The measurement results – in excellent agreement with the above reasoning – are shown by empty circles in Figs. 4.3.

Finally, we must emphasize that the value of γ is not *a priori* known for real systems. Consequently, α does not reflect the type of internal dynamics in the straightforward fashion suggested by Ref. [dMB04a]. Instead, a careful analysis, including the dependence on Δt , must be undertaken, in order to interpret the results correctly. The only exception is when we can assume homogeneous correlations, i.e., $\gamma = 0$ and so $H_i = H$.

4.2.3 Corrections to fluctuation scaling

As one can see from the broadening of the scaling plot of Fig. 4.3, fluctuation scaling is not perfect for stock markets. While obviously there are

³In fact, the DFA procedure can only be applied for $\Delta t \geq 4\delta$, but the effect of this difference is negligible.

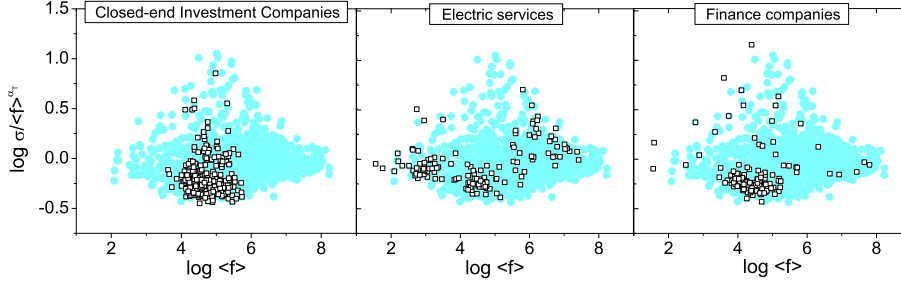


Figure 4.9: $\log[\sigma/\langle f \rangle^\alpha]$ plotted versus $\log \langle f \rangle$ for 10-second resolution traded value data of stocks. The data points were not binned or altered in any way, which makes visible the deviations from the original scaling law, which would correspond to a horizontal line. The lighter points indicate all stocks, while boxes (\square) highlight the distribution of points for the three indicated economic sectors. The great degree of clustering both horizontally and vertically. Clustering along the $\log \langle f \rangle$ axis only suggests that the sector has some typical trading activity. Systematic corrections to fluctuation scaling are indicated by clustering along the $\log \sigma/\langle f \rangle^\alpha$ axis. If a sector is clustered in the lower/higher half of the dataset, it means that its trading activity has typically lower/higher fluctuations than the market average. The presence of such sector dependent clustering suggests that the corrections to fluctuation scaling are not purely random.

some statistical errors as well, it is also expected that the trading activity of different market sectors will fluctuate to a different extent. In Fig. 4.9 we plot $\sigma_i/\langle f_i \rangle^\alpha$ versus $\langle f_i \rangle$, which represents the corrections to fluctuation scaling. By highlighting the alignment of three industrial sectors one can see that they form clusters, so the deviations are – to some degree – systematic. The fact that scaling is well preserved suggests that markets have a robust dynamics characterized by a value of α . The role of the corrections is not substantial in the formation of fluctuations.

4.3 Conclusions

In the above, we generalized the fluctuation scaling relation to the case when temporal correlations are present in the individual time series. In such an analysis, one measures the time-scale dependent scaling exponent $\alpha(\Delta t)$. In addition to previous studies, we found that even in the presence of strong temporal correlations, α still remains very characteristic to the internal dynamics. Indeed, its time scale dependence reveals additional

information. For the persistence of fluctuation scaling at all time scales, it is inevitable that the strength of correlations in *all* the individual time series be connected by a logarithmic law. Such a relationship is a peculiar feature of collective dynamics, which is not explained by the number or the size distribution of the events.

The framework was applied to reveal the connections between stylized facts of stock market trading activity. Empirical data for both NYSE and NASDAQ show qualitatively similar behavior. For short times when there are no correlations between the trades of an individual company, the non-trivial value of α comes from the highly inhomogeneous trade sizes of the different companies. For increasing time windows, we observe a logarithmic law in correlation strengths as the function of liquidity $\langle f \rangle$.

Our results imply that the trading of assets of companies with very different size and liquidity cannot be described in a universal manner⁴. There have been studies pointing out such asset-to-asset variations, and the key role of liquidity [PS03, CRS01, FGL⁺04, FL04], however, they have been consistently overlooked by some econophysics groups.

It is misleading to calculate averages for stocks with a wide range of liquidities as done in, e.g., Refs. [GPGS00, KOD05, MAS03, MdSQ06]. A "typical" $\tau(q)$ or multifractal spectrum of assets is not meaningful in the presence of this clear, systematic dependence.

Note that according to Eq. (4.3) fluctuation scaling (at least in stock markets) is valid for other moments of the fluctuations, not only the variance. Size dependent correlations are present there as well, we will discuss those in Section 5.3.2. In principle it would be possible to give a similar, consistent scaling theory for these other moments too. The relationship between correlations in other moments and the time-dependence of the related $\alpha(q)$'s can be straightforward. One expects that the full $\alpha(q; \Delta t)$ spectrum contains much more information about the system than $\alpha(\Delta t)$. However, the variance is very convenient for analytical calculations [cf. Eq. (4.7)], models and empirical results are widely available. At present it is completely unknown how the additional information from other moments can be extracted, and there are very few studies on this subject.

⁴A recent preprint [YI05] shows similar effects with respect to the market where the stocks are traded. More indications of similar behavior can be found in Refs. [EK06b, BLM00]

4.4 Appendix: The components of the fluctuation σ^2

A large part of this and the following chapter is concerned with the standard deviation of the sums of random variables. This is defined as

$$\sigma^2 = \left\langle \left(\sum_{n=1}^N V_n \right)^2 \right\rangle - \left\langle \sum_{n=1}^N V_n \right\rangle^2,$$

where V_n are the individual (not necessarily independent) random variables, and N is the number of these variables which itself can be random.

Let $\mathbb{P}(N)$ be the probability that the number of variables is N . The sum of N variables can be written as $V_N = \sum_{n=1}^N V_n$. Let $\mathbb{P}(V_N)$ denote the density function of this sum when N is fixed. Then the standard deviation of the sum when N itself is a random variable is

$$\begin{aligned} \sigma^2 &= \sum_N \mathbb{P}(N) \int dV_N \mathbb{P}(V_N) V_N^2 - \\ &\quad \left(\sum_N \mathbb{P}(N) \int dV_N \mathbb{P}(V_N) V_N \right)^2 = \\ \sum_N \mathbb{P}(N) &\left[\underbrace{\int dV_N \mathbb{P}(V_N) V_N^2}_{\langle V_N^2 \rangle} - \underbrace{\left(\int dV_N \mathbb{P}(V_N) V_N \right)^2}_{\langle V_N \rangle^2} \right] + \\ &\quad \underbrace{\sigma_{V_N}^2 = N\sigma_V^2} \\ &\quad \sum_N \mathbb{P}(N) \underbrace{\left(\int dV_N \mathbb{P}(V_N) V_N \right)^2}_{\langle V_N \rangle^2 = N^2 \langle V \rangle^2} - \\ &\quad \left(\sum_N \mathbb{P}(N) \underbrace{\int dV_N \mathbb{P}(V_N) V_N}_{\langle V_N \rangle = N \langle V \rangle} \right)^2 = \underbrace{\sigma_V^2}_{\langle N \rangle} \sum_N \mathbb{P}(N) N + \\ &\quad \underbrace{\langle V \rangle^2}_{\sigma_N^2} \left[\underbrace{\sum_N \mathbb{P}(N) N^2}_{\langle N^2 \rangle} - \underbrace{\left(\sum_N \mathbb{P}(N) N \right)^2}_{\langle N \rangle^2} \right] \end{aligned}$$

Thus finally

$$\sigma^2 = \sigma_V^2 \langle N \rangle + \langle V \rangle^2 \sigma_N^2.$$

In the case when the V_n 's are strongly (i.e., power law) correlated $\sigma_{V_N}^2 = \sigma_V^2 N^{2H_V}$ where H_V is the Hurst exponent as defined in Eq. (5.3), and so

$$\sigma^2 = \sigma_V^2 \langle N^{2H_V} \rangle + \langle V \rangle^2 \sigma_N^2. \quad (4.7)$$

The correlations in N are not reflected directly in this expression. Instead, they affect how σ_N changes with the time window size Δt as pointed out in Section 4.2.

Chapter 5

Stylized facts of trading activity revisited

Studies in econophysics concentrate on the possible analogies, even though there are important differences between physical and financial systems. This is, of course, a trivial statement – it is enough to refer to the possibility of influencing the system by its characterization or to the intrinsic non-stationarity of economic processes. The most crucial difference is, however, the discrepancy in the levels of description. In the case of a physical system undergoing a second order phase transition, it is natural to assume scaling on profound theoretical grounds and the (experimental or theoretical) determination of, e.g., the critical exponents is a fully justified undertaking. There is no similar theoretical basis for the financial market whatsoever, therefore in this case the assumption of power laws should be considered only as one possible way of fitting fat tailed distributions. Also, no reference to universality should be plausible as the robustness of *qualitative* features – like the fat tail of the distributions – is a much weaker property. Therefore, e.g., averaging distributions over different companies can sometimes be questionable. While we fully acknowledge the process of understanding based on analogies as an important method of scientific progress, we emphasize that special care has to be taken in cases where the theoretical support is sparse.

The stylized facts listed in Section 2.2 represent the widespread opinion in econophysics. Some of their limitations are obvious: For example, in real markets even if traded value were asymptotically power law distributed, that would have an upper cutoff simply because there is a finite number of shares of any company available. Similar upper cutoffs (or related finite size effects) are well-known in physics. However, aside from these, there is still a common belief that many concepts from statistical physics such as power laws, scaling and universality are intrinsic in financial data. For motivation it is (often very vaguely) argued that markets are

like physical systems due to their high complexity and the many, strongly interacting elements. In the last few years all sorts of physical principles have been applied to finance without moderation.

Very much contrary to this picture, the previous chapter discussed how fluctuation scaling and the systematic dependence of the Hurst exponents on the typical traded value (liquidity) of the stock are related. This is in clear opposition with the assumption of universality. Liquidity appears to act as a continuous parameter that determines the quantitative character of trading activity. Another such example from the previous chapter is that more frequently traded stocks tend to be traded in larger transactions on average. This means that the transaction size distribution will also depend on liquidity [EK07a].

Motivated by all these points, the aim of the present chapter is to present an in-depth analysis of the high resolution data of the New York Stock Exchange with special emphasis on the effects caused by the size of the companies or their liquidity. The particular database was chosen to be consistent with most recent studies.

This chapter is organized as follows. Section 5.1 presents the results on the capitalization dependence of various measures of trading activity. In Section 5.2 we show that the distribution of the traded values is not Lévy stable as suggested previously [GPGS00]. Consequently the variance and the Hurst exponent of f , as used in the previous chapter, are meaningful. Section 5.3 presents an extended inquiry into the non-universality of correlations in trading activity, including its origins and multiscaling properties. Section 5.4 deals with the time intervals between trades and we give indications, that their distribution is better described by a multiscaling ansatz than by gap scaling proposed earlier [IYPL04]. Finally, Section 5.5 concludes.

5.1 Capitalization and basic measures of trading activity

Many previous studies of trading focus on the stocks of large companies. These certainly have the appealing property that price and returns are well defined even on short time scales due to the high frequency of trading. Nevertheless, other quantities regarding the activity of trading, such as traded value/volume or the number of trades can be defined even for those stocks where they are zero for most of the time. In this section we extend the study of Zumbach [Zum04] which concerned companies of the top two orders of magnitude in capitalization at the London Stock Exchange. Instead, we analyze the 3347 stocks¹ that were traded continuously at

¹Note that many minor stocks do not represent actual companies, they are only, e.g., preferred class stocks of a larger enterprise.

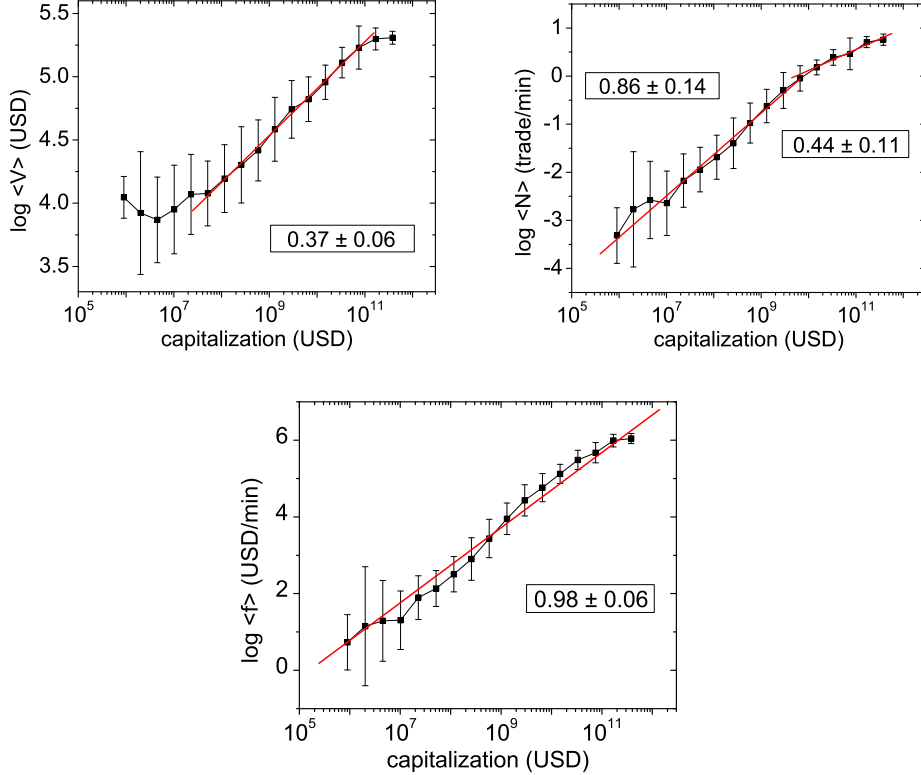


Figure 5.1: Capitalization dependence of certain measures of trading activity in the year 2000. The graphs are monotonically increasing and are (piecewise) well approximated by power laws as indicated. All three tendencies curve downward for large capitalizations. **(top left)** Mean value per trade $\langle V \rangle$ in USD. The fitted slope corresponds to the regime $5 \times 10^7 < C < 7.5 \times 10^{10}$ in USD. **(top right)** Mean number of trades per minute $\langle N \rangle$. The slope on the left is from a fit to $C < 4.5 \times 10^9$ USD, while the one on the right is for $C > 4.5 \times 10^9$ USD. **(bottom)** Mean trading activity (exchanged value per minute) $\langle f \rangle$ in USD. The plots include 3347 stocks that were continuously available at NYSE during 2000.

NYSE for the year 2000. This gives us a range of approximately $10^6 \dots 6 \times 10^{11}$ USD in capitalization.

Following Ref. [Zum04], we quantify the dependence of trading activity on company capitalization C_i . Mean value per trade $\langle V_i \rangle$, mean number of trades per minute $\langle N_i \rangle$ and liquidity (traded value per minute) $\langle f_i \rangle$ are plotted versus capitalization in Fig. 5.1. Ref. [Zum04] found that all three quantities have power law dependence on C_i , however, this simple ansatz does not seem to work for our extended range of stocks. While mean

trading activity can be approximated as $\langle f_i \rangle \propto C_i^{0.98 \pm 0.06}$ to an acceptable quality, neither $\langle V \rangle$ nor $\langle N \rangle$ can be fitted by a single power law in the whole range of capitalization. Nevertheless, there is – not surprisingly – a monotonic dependence: higher capitalized stocks are traded more intensively.

5.2 Traded value distributions revisited

The statistical properties of the trading volume of stocks have previously been investigated in Ref. [GPGS00]. That work finds that the cumulative distribution of traded volume in $\Delta t = 15$ min windows has a power-law tail with a tail exponent $\lambda = 1.7 \pm 0.1$. This is the so called *inverse half cubic law*. Formally, this corresponds to

$$\mathbb{P}_{\Delta t}(f) \propto f^{-(\lambda+1)}, \quad (5.1)$$

where $\mathbb{P}_{\Delta t}$ is the probability density function of traded volume (value) on a time scale Δt .

Ever since great effort was devoted to explain this exponent in terms of the *inverse cubic law* of stock returns [GMAS98, GGPS03, PGGS04]. However, the exact distribution and the possible exponents are still much debated [FL04, Que05].

The estimation of the tail exponent is a delicate matter. Following the methodology of Ref. [GPGS00] – and for the same 1994 – 1995 period of data – we repeated these measurements. Our results for the $\Delta t = 15$ min distribution are shown in Fig. 5.2 for three major stocks. The tails of these distributions can be fitted by a power law over an order of magnitude, for the top 5 – 10% of the events. The exponent λ we find is significantly higher than 1.7, it is around 2.2 for these examples.

For systematic calculations of λ , there is a range of mathematical tools available. We used three variants of Hill’s method [Hil75, Alv01] to estimate the tail exponent, details can be found in Appendix 5.6. All three have a common parameter: the number k of largest events that belong to the tail. The statistical weight associated with the tail events is $p = k/L$, where L is the total length of our time series. From Fig. 5.2 one can see, that $p \approx 5 - 10\%$ is the proper choice as a threshold for the asymptotic regime.

For the two-year period 1994 – 1995 and separately for the single year 2000, we took the 1000 stocks with the highest total traded value in the TAQ database. We detrended their trading activity by the well known *U*-shaped intraday pattern (see, e.g., Ref. [EKYB05]). Then we calculated the distribution of λ over these stocks. The median and the width of this distribution (characterized by the half distance of the 25% and 75% quantiles) is shown in Table 5.1 for various time windows Δt .

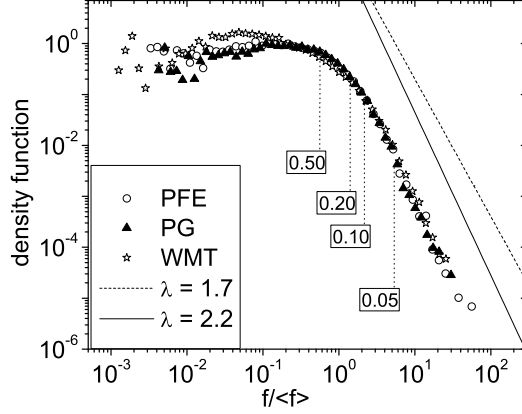


Figure 5.2: Distributions of traded value in $\Delta t = 15$ min time windows, divided by the mean. The plot displays three example stocks for the period 1994–1995. The numbers show some upper quantiles of the distribution (probability of values higher than indicated by the corresponding dashed line). The dashed and solid diagonal lines represent power-laws with exponents corresponding to $\lambda = 1.7$ and 2.2 , respectively.

The choice $p = 0.06$ in Hill’s method provides results in line with Ref. [GPGS00]. For $\Delta t = 15$ min time windows, one finds $\lambda = 1.71 \pm 0.20$ for the period 1994–1995. However, other estimates are significantly higher, $\lambda > 2$. Moreover, two estimators show a strong tendency of increasing λ with increasing time windows. Monte Carlo simulations on surrogate datasets show, that this is beyond what could be explained by decreasing sample size. It is well known, that for $\lambda < 2$ the distribution would have to converge to the corresponding Lévy distribution when $\Delta t \rightarrow \infty$. The measured λ ’s should also be independent of Δt . On the other hand, for $\lambda > 2$, the $\Delta t \rightarrow \infty$ limit distribution is a Gaussian. Accordingly, for finite samples, the measured effective value of λ increases with Δt .

One must keep in mind, that all three methods *assume* that the variable is asymptotically distributed as (5.1) and none of them *proves* it. If this does not hold, then the estimates of exponents are only a parametric characterization of the unknown functional form, nevertheless, they do suggest that the second moments exist. If the distribution is indeed of the limiting form (5.1), then although for short time windows ($\Delta t < 60$ min) there is a fraction of stocks whose estimate gives $\lambda < 2$, even those display $\lambda > 2$ for larger Δt .

Based on these results we conclude that the second moments of the distribution exists for large enough Δt , therefore for the related time series the calculation of the Hurst exponent – and possibly also higher order correlation exponents – is meaningful.

Δt	Hill's ($p = 0.06$)	Shifted Hill's λ	Shifted Hill's φ	Fraga Alves ($p = 0.1$)
1 min	1.43 ± 0.09	2.15 ± 0.15	3.0	1.98 ± 0.25
5 min	1.56 ± 0.13	2.29 ± 0.25	2.8	2.04 ± 0.25
15 min	1.71 ± 0.20	2.55 ± 0.35	2.8	2.1 ± 0.3
60 min	2.06 ± 0.30	2.85 ± 0.45	1.8	2.1 ± 0.4
120 min	2.3 ± 0.4	3.15 ± 0.70	1.6	2.1 ± 0.4
390 min	2.7 ± 0.6	3.7 ± 0.9	1.2	no estimate

Δt	Hill's ($p = 0.06$)	Shifted Hill's λ	Shifted Hill's φ	Fraga Alves ($p = 0.1$)
1 min	1.63 ± 0.13	2.40 ± 0.23	2.6	2.16 ± 0.25
5 min	1.91 ± 0.25	2.8 ± 0.5	2.4	2.30 ± 0.35
15 min	2.15 ± 0.40	3.1 ± 0.6	2.0	2.35 ± 0.40
60 min	2.6 ± 0.5	3.45 ± 0.8	1.2	2.2 ± 0.4
120 min	2.8 ± 0.6	3.8 ± 1.1	1.2	no estimate
390 min	3.2 ± 1.0	5.1 ± 0.8	1.6	no estimate

Table 5.1: Median of the tail exponents of traded value calculated by three methods. The width of the distributions is given with the half distance of the 25% and 75% quantiles. **(top)** Results for 1994 – 1995. **(bottom)** Results for 2000.

5.3 Correlations in traded value time series

5.3.1 The components of correlations

Section 5.1 showed that liquidity is a monotonically increasing function of company capitalization, and thus it is not incorrect to use the two terms as synonyms. Section 4.2.2 extensively studied the liquidity (as measured by $\langle f \rangle$) dependence of the Hurst exponents of traded value. But what aspect of the trading dynamics is the origin of this dependence and hence non-universality?

As Eq. (2.3) suggests, the source of fluctuations in f is the fluctuation of N and V (see also Ref. [EK05]). Thus, it is very instructive to define the Hurst exponents of these two processes in analogy with Eq. (4.1). One can introduce the H_{N_i} Hurst exponent of the time series $N_i^{\Delta t}(t)$ as

$$\sigma_{N_i}^2(\Delta t) = \left\langle [N_i^{\Delta t}(t) - \langle N_i^{\Delta t} \rangle]^2 \right\rangle \propto \Delta t^{2H_{N_i}}. \quad (5.2)$$

This H_N describes the temporal correlations of the number of trades.

In order to reduce noise in the following measurements, we selected five groups of stocks with respect to $\langle f_i \rangle$: those with $10^2 \text{ USD/min} \leq \langle f_i \rangle < 10^3 \text{ USD/min}$, those with $10^3 \text{ USD/min} \leq \langle f_i \rangle < 10^4 \text{ USD/min}$, \dots , and finally $10^6 \text{ USD/min} \leq \langle f_i \rangle$. Then, we averaged the $\sigma_{N_i}^2(\Delta t)$ variances within each group².

²This averaging procedure decreases the noise present in the data, without affecting our main conclusions. Also note that data were first corrected by the well-known U-shape pattern of daily trading activity (see, e.g., Ref. [EKYB05]), calculated independently for each group.

The results for the group averages of $\sigma_{N_i}^2$, and the asymptotically valid exponents $H_{N_i}^\pm$ are shown in Fig. 5.3(left). A comparison with σ_i^2 [Fig. 5.4(left)] shows that both quantities behave similarly³: Fluctuations in the number of trades N display the crossover and the liquidity dependence in the strength of correlations, just like f .

The H_{V_i} Hurst exponent of the so-called tick-by-tick data V can be defined as

$$\sigma_{V_i}^2(N/\langle N_i \rangle) = \left\langle \left[\sum_{n=1}^N V_{i,n}^{\Delta t}(t) - \left\langle \sum_{n=1}^N V_{i,n}^{\Delta t}(t) \right\rangle \right]^2 \right\rangle \propto (N/\langle N_i \rangle)^{2H_{V_i}}, \quad (5.3)$$

which has to hold for any Δt .

In order to make group averages as before, it is an important point that here the scaling variable is the N number of consecutive trades. This is divided by the $\langle N_i \rangle$ mean number of trades per minute. Such a normalization is crucial, because the trading frequency of stocks varies over many orders of magnitude. Thus N trades correspond to a different time span depending on trading frequency, i.e., on the stock. The scaling variable $N/\langle N_i \rangle$ has a dimension of minutes (just like Δt), and its fixed value always means the same time window size, regardless of the stock.

Moreover, when applying Eq. (5.3), there is a natural lower limit in window size: one cannot take less than one trade, and so $N \geq 1$. Consequently, a group average for $\sigma_{V_i}^2$ is not well defined, where the scaling variable would be $N/\langle N_i \rangle < 1/\langle N_i \rangle$ for any stock in the group⁴. For more liquid stocks, $\langle N_i \rangle$ is larger, thus the minimal effective window size is smaller.

The results are shown in Fig. 5.3(right). H_V^- is only defined for the two groups, whose stocks are traded at least every 10 minutes, and they indicate weak or no liquidity-dependence. H_V^+ exists for all groups and follows the same trend of increasing correlations for greater liquidity.

The number of transactions in a given time window $[t, t + \Delta t)$ is – to a good approximation – independent from the value of the single transactions⁵. As noted before, one can show that

$$\sigma_i^2(\Delta t) = \sigma_{N_i}^2(\Delta t) \langle V_i \rangle^2 + \sigma_{V_i}^2 \left\langle (N_i^{\Delta t})^{2H_{V_i}} \right\rangle, \quad (4.7)$$

where $\langle V_i \rangle$ is the mean, and $\sigma_{V_i}^2$ is the standard deviation of the value of

³Systematic dependence of the exponent of the power spectrum of the number of trades N on capitalization was previously reported in Ref. [BLM00], based on the study of 88 stocks. This quantity is closely related to the Hurst exponent for the time series of the number of trades (see Ref. [IYPL04]).

⁴We allowed up to 10% of such missing data.

⁵This means, that $N_i^{\Delta t}(t)$ is independent from $f_i^{\Delta t}(t)/N_i^{\Delta t}(t)$. The R^2 values of regressions between the logarithms of these two quantities are typically of the order 0.03 in the data.

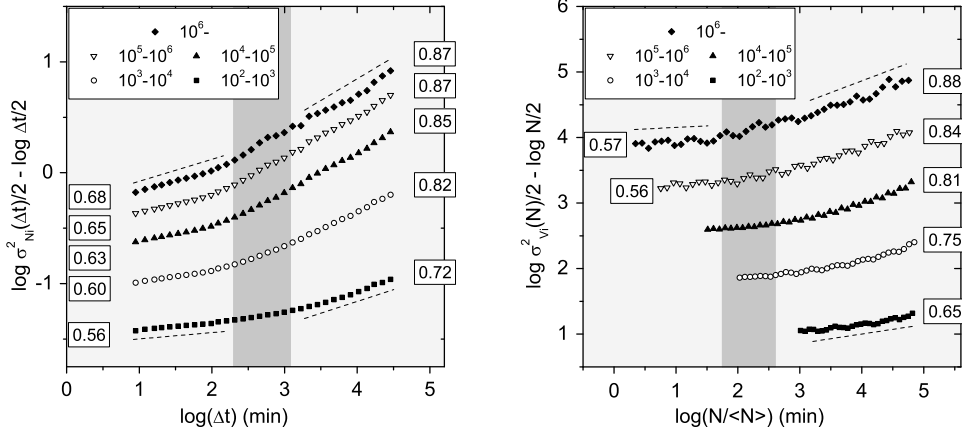


Figure 5.3: **(left)** The normalized partition function $\frac{1}{2} \log \sigma_{N_i}^2(\Delta t) - \frac{1}{2} \log \Delta t$ for the five groups of companies. A horizontal line would mean the absence of autocorrelations in the data. The crossover regime is for slightly longer times, $\Delta t \approx 160 - 1200$ min. Above the crossover the strength of correlations in N , and thus the slopes corresponding to $H_N^+ - 0.5$, increase with the liquidity of the stock. **(right)** Same as (left), but for $\frac{1}{2} \log \sigma_{V_i}^2(N/\langle N \rangle) - \frac{1}{2} \log N$. The darker shaded area corresponds to the crossover regime of f at $N/\langle N \rangle \approx 60 - 390$ mins. Small stocks are traded infrequently, therefore they have no data points below the crossover. *Note:* Stocks grouped by $\langle f \rangle$, increasing from bottom to top, ranges given in USD/min. In both plots, fits are for the regimes below 60 min and above 1700 min.

$\langle f \rangle$ (USD/min)	H_V^-	H^-	H_N^-	H_V^+	H^+	H_N^+
$10^6 -$	0.57	0.60	0.68	0.88	0.89	0.87
$10^5 - 10^6$	0.56	0.57	0.65	0.84	0.83	0.87
$10^4 - 10^5$	–	0.56	0.63	0.81	0.80	0.85
$10^3 - 10^4$	–	0.55	0.60	0.75	0.77	0.82
$10^2 - 10^3$	–	0.54	0.56	0.65	0.70	0.72

Table 5.2: Hurst exponents for f , N and V for the 5 groups of stocks. For all groups, every exponent is higher above the crossover than below it. Moreover, above the crossover there are no large differences between H_V^+ , H^+ , H_N^+ . From the fits, the errors are estimated to be ± 0.03 . *Note:* H_V^- is not defined for the 3 groups, whose stocks are not traded at least every 10 minutes.

individual transactions. The origins of the two terms in the formula are the following [EK05]:

1. The first term describes the effect of fluctuations in the number of transactions. Let us assume, that the size of the transactions is constant, so $V_{i,n} = \langle V_i \rangle$, and $\sigma_{V_i}^2 = 0$. Then, the second term is zero, and Eq. (4.7) simplifies to $\sigma_{N_i}^2(\Delta t) \langle V_i \rangle^2$.
2. The second term describes the effect of fluctuations in the value of individual transactions. If one assumes that the number of transactions is the same in every time window, $N_i^{\Delta t}(t) = \langle N_i^{\Delta t} \rangle$, then $\sigma_{N_i}^2 = 0$. The first term becomes zero, and Eq. (4.7) reduces to $\sigma_{V_i}^2 \langle N_i^{\Delta t} \rangle^{2H_{V_i}}$.

Thus the correlations in f originate from the correlations in N and V . By definition, the l.h.s. of Eq. (4.7) is proportional to Δt^{2H_i} . The first term on the r.h.s. is proportional to $\Delta t^{2H_{N_i}}$, while the second term can be estimated to scale as $\Delta t^{2H_{V_i}}$. For $\Delta t \rightarrow 0$ the behavior of σ^2 is dominated by the smaller of H_N and H_V , while for $\Delta t \rightarrow \infty$ by the larger one.

This is consistent with the results summarized in Table 5.2. The table also shows that above the crossover there are no major differences between H , H_N and H_V . This means that neither of the two processes dominates long range fluctuations in general.

5.3.2 Multiscaling

Although Section 5.2 [EK06b] only showed that the standard deviation of f exists, in principle it is possible to extend the analysis of correlations to higher moments. This is also motivated by Ref. [Que05], which finds finite moments of f up to a power of 4 or higher. Let us define the q -th order partition function of traded value in the following way:

$$\sigma_i^q(\Delta t) = \left\langle \left| f_i^{\Delta t}(t) - \langle f_i^{\Delta t}(t) \rangle \right|^q \right\rangle \propto \Delta t^{\tau_i(q)}. \quad (5.4)$$

For any fixed stock i , the formula defines a $\tau_i(q)$ set of exponents, indexed by q . These are often written in the form $\tau(q) = qH(q)$, and $H(2)$ is the Hurst exponent, while other $H(q)$'s are called the generalized Hurst exponents. If $H(q) \equiv H$ is independent of q , the signal is self-affine, while any other q -dependence is called multiscaling or multi-affinity.

For the calculation of $\sigma_i^q(\Delta t)$ we again used the same dataset, the method of Detrended Fluctuation Analysis, the test of robustness to the order of detrending, as in the previous section. As an example, the group averages of the partition function for $q = 2$ are shown in Fig. 5.4(left).

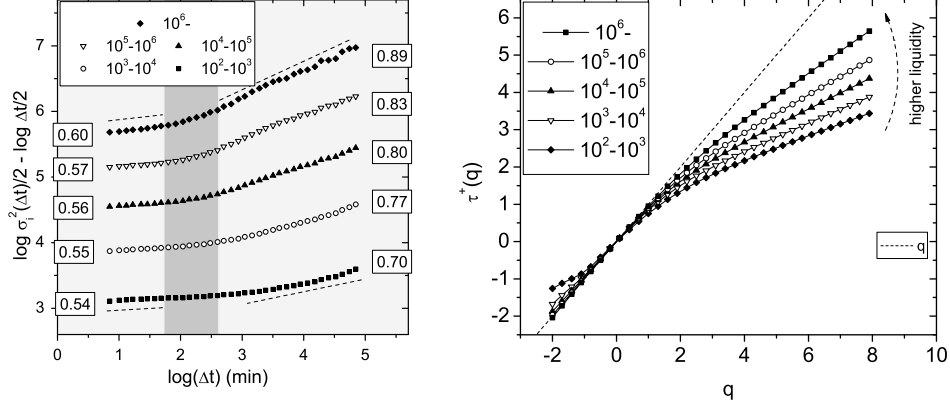


Figure 5.4: **(left)** The normalized partition function $\frac{1}{2} \log \sigma_i^2(\Delta t) - \frac{1}{2} \log \Delta t$ for the five groups of companies. A horizontal line would mean the absence of autocorrelations in the data. Instead, one observes a crossover phenomenon in the regime $\Delta t = 60 - 390$ mins. Below the crossover all groups show weakly correlated behavior. Above the crossover (fluctuations on daily and longer scales), the strength of correlations, and thus the slope corresponding to $H^+ - 0.5$, increases with the liquidity of the stock. *Note:* Fits are for the regimes below 60 min and above 1700 min. **(right)** The values of the scaling exponents $\tau^+(q)$, valid for time scales longer than a trading day, for the five groups of companies. Companies with higher average traded value exhibit stronger correlations, and weaker multiscaling than their smaller counterparts. Correspondingly, their $\tau^+(q)$ is greater, and the shape of the curve is closer to the linear relationship $\tau^+(q) = q$.

Again, regardless of group, the $\log \sigma^q(\Delta t)$ versus $\log \Delta t$ plots are not straight lines. Instead, one observes a crossover phenomenon⁶ [EK06b, IYPL04]: There are two regimes of Δt for which different $\tau(q)$ -s can be identified. For $\Delta t < 60$ min, we are going to use the notation $\tau^-(q)$, while for $\Delta t > 390$ min, $\tau^+(q)$. One can define the related generalized Hurst exponents as $\tau^\pm(q) = qH^\pm(q)$. Systematically, $H^+(q) > H^-(q)$, which means that correlations become stronger when window sizes are greater than 390 min.

Similarly to the case of Hurst exponents, there is only a weak systematic difference between groups for $\tau^-(q)$, and a strong one for $\tau^+(q)$. More

⁶The fact, that the properties of stock market time series are different on time scales shorter than and longer than 1 trading day, was pointed out by many sources. The most common examples is are the distribution of returns and the autocorrelations of volatility [BP00, MS99].

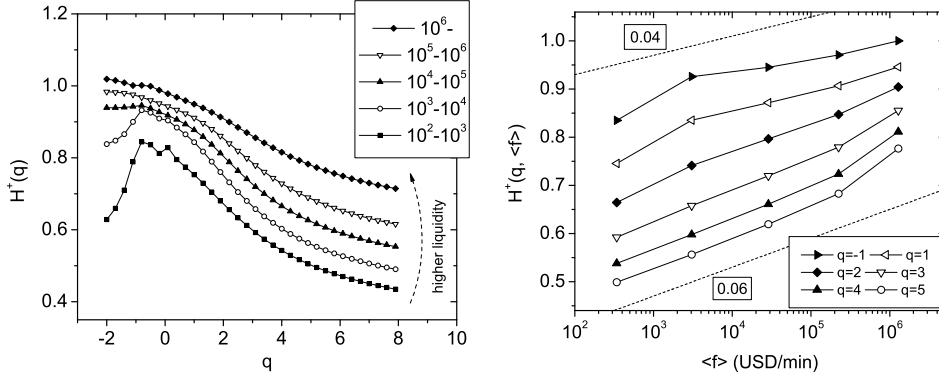


Figure 5.5: **(left)** The values of the generalized Hurst exponents $H^+(q)$, valid for time scales longer than a trading day, for the five groups of companies. The difference in the strength of correlations, and thus $H^+(q)$, is present for all powers q . This implies, that such a dependence on liquidity is present in both low and high trading activity periods. **(right)** The values of $H^+(q)$, from top to bottom $q = -1, 1, 2, 3, 4, 5$. The points represent the average value for one of the five groups of companies. One can see, that $H^+(q)$ changes in an approximately logarithmic fashion with $\langle f \rangle$. *Note:* Stocks grouped by $\langle f \rangle$, increasing from bottom to top, ranges given in USD/min.

of this dependence can be understood if one examines the scaling exponents for several powers of q . This was done by first evaluating the value of $\tau_i^+(q)$ for the independent stocks, and then averaging that for the elements within each group. As expected, more liquid stocks (greater $\langle f \rangle$) display stronger correlations than their less liquid (smaller $\langle f \rangle$) counterparts, for any order $q > 0$. This is realized in a way that the degree of multiscaling decreases, and the scaling exponents tend to the fully correlated self-affine behavior with the limiting exponents $\tau^+(q) = q$, $H^+(q) = 1$.

Fig. 5.5(left) shows the corresponding values of $H^+(q)$. The difference in the $H^+(q)$'s between the groups is present throughout the whole range of q 's, not only for large q 's which are sensitive to the high trading activity. This indicates that the higher level of correlations in more liquid stocks cannot be exclusively attributed to periods of high trading activity. Instead, it is a general phenomenon, that is present continuously⁷.

⁷One may notice, that there is a strong deviation in the case of stocks with low liquidity, and $q < -1$. The origin of this artifact is a finite size effect: The stocks are traded in lots of 100, and thus they cannot be traded in values less than price \times 100. This minimum acts as a cutoff in small fluctuations, to which $q < -1$ moments are very sensitive.

In Fig. 5.5(right), we plot vertical "cuts" of Fig. 5.5(left). These show, that for a fixed value of q , $\tau^+(q)$ increases with $\langle f \rangle$ in an approximately logarithmic way:

$$H^+(q; \langle f_i \rangle) = H_*^+ + \gamma^+(q) \log \langle f_i \rangle, \quad (5.5)$$

where $\gamma^+(q) \approx 0.04 - 0.06$. For $q = 2$ this is equivalent to what was found before in Section 4.2.2.

5.4 Multiscaling distribution of intertrade times

Finally, we analyzed the intertrade interval series $T_i (n = 1 \dots \nu_i - 1)$, defined as the time spacings between the n 'th and $n + 1$ 'th trade. ν_i is the total number of trades for stock i during the period under study.

Previously, Ref. [IYPL04] used 30 stocks from the TAQ database for the period 1993–1996 and proposed that the distribution of T_i scales with the mean $\langle T_i \rangle$ as

$$\mathbb{P}(T, \langle T \rangle) = \frac{1}{\langle T \rangle} F(T / \langle T \rangle), \quad (5.6)$$

and the universal scaling function F is well modeled by a Weibull distribution of the form

$$F(x) = \frac{\delta}{X} \left(\frac{x}{X} \right)^{\delta-1} \exp \left[- \left(\frac{x}{X} \right)^\delta \right], \quad (5.7)$$

where $X \approx 0.94$ and $\delta \approx 0.72$ for all the 30 stocks, with some statistical deviations.

We analyzed the data by including a large number of stocks with very different capitalizations. First it has to be noted that the mean intertrade interval has decreased drastically over the years. In this sense the stock market cannot be considered stationary for periods much longer than one year. We analyze the two year period 1994 – 1995 (part of that used in Ref. [IYPL04]) and separately the single year 2000. We use all stocks in the TAQ database with $\langle T \rangle < 10^5$ sec, a total of 3924 and 4044 stocks, respectively.

In order to check the validity of the formula (5.7), we divided the stocks into two groups⁸ with respect to $\langle T \rangle$. Then, we generated the distribution of $T / \langle T \rangle$ for the groups, a comparison for the year 2000 is shown in Figure 5.6. This already raises doubts about the generality of Eq. (5.7): The tails of the distribution seem to possess more weight for the group with small $\langle T \rangle$ (blue chips). The direct visual comparison of these distributions is, however, not always a reliable method to evaluate universality. Instead, we take a less arbitrary, indirect approach.

⁸The groups were constructed to have an approximately equal total number of trades. Small $\langle T \rangle$ (top 246 stocks): $6.48 \text{ sec} < \langle T \rangle < 47.8 \text{ sec}$ (other 3797 stocks), large $\langle T \rangle$: $47.8 \text{ sec} < \langle T \rangle < 10^5 \text{ sec}$.

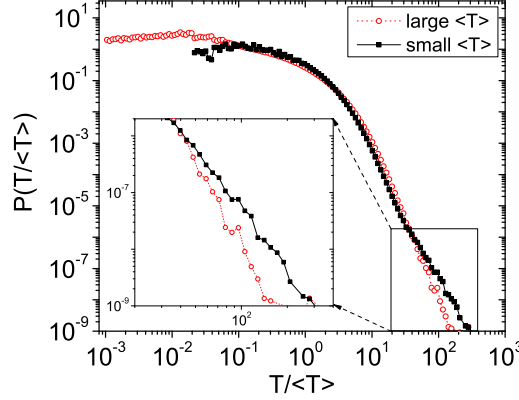


Figure 5.6: The distribution of $T/\langle T \rangle$ in the year 2000 for two groups of stocks with different mean intertrade times $\langle T \rangle$. The group with the most frequently traded stocks (blue chips) has a considerably greater weight for waiting times. This implies that the distribution $\mathbb{P}(T, \langle T \rangle)$ may not be universal.

The consequence of the universal distribution (5.6) would be that the moments of T should show gap scaling: The difference between the exponents of the q -th and $q + 1$ -th moments would be independent of q [HJK⁺86, MKG96]:

$$\langle T_i^q \rangle = C(q) \langle T_i \rangle^{-\tau(q)}, \quad (5.8)$$

with a scaling function⁹ $-\tau(q) \equiv q$.

Instead, we find a systematic dependence of $-\tau$ on q , see Fig. 5.7 for several examples of fitting and Fig. 5.8 for all results. There is good fit to a power law of type (5.8) over 4 orders of magnitude in $\langle T \rangle$ and with non-trivial exponents.

The intuitive meaning of $-\tau(q \gg 1) < q$ is simple: Intertrade times of larger (more frequently traded) stocks exhibit larger relative fluctuations. In line with our observation from Figure 5.6, this difference must come from the tail of the distribution, as the deviation becomes more pronounced for higher moments.

The absence of simple universal scaling raises the question of the liquidity dependence of the Hurst exponent for the time series T_i , defined

⁹We keep the negative sign to conform with usual conventions.

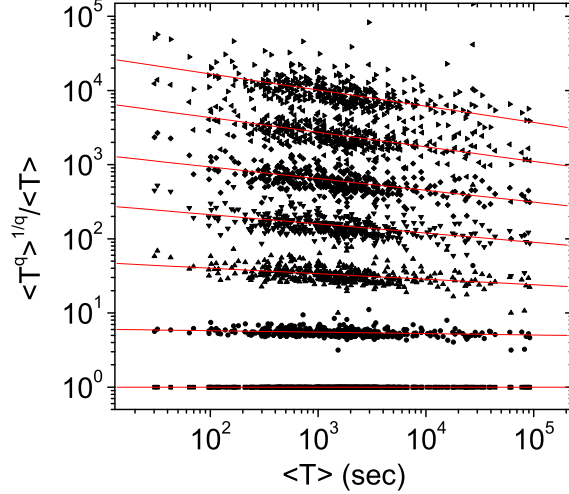


Figure 5.7: Scaling of integer moments of T , $q = 1, 2, 4, 6, 8, 12, 16$ (increasing from bottom to top). The plot shows $\langle T^q \rangle^{1/q} / \langle T \rangle$, the slopes correspond to $-\tau(q)/q - 1$. If the normalized distribution of T were universal, the points would align on horizontal lines. *Note:* The points were shifted vertically for better visibility. Only 400 points are shown per moment, the sample period was 1994 – 1995.

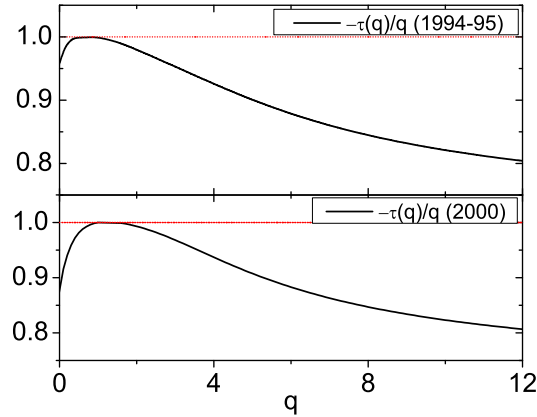


Figure 5.8: Scaling exponents for the moments of intertrade interval distributions defined in Eq. (5.8). The values $-\tau(q) \equiv q$ would imply a universal distribution that is independent of stock. The fact that $-\tau(q)/q < 1$, shows less frequently traded stocks display relatively lower variations in their trading dynamics. For large q , the effect increases monotonically with q . This suggests a difference between small and large stocks in the tail of the distribution, which corresponds to longer periods of inactivity.

analogously to Eq. (5.3) as

$$\sigma_{T_i}^2(N) = \left\langle \left[\sum_{n=1}^N T_i(n) - \left\langle \sum_{n=1}^N T_i(n) \right\rangle \right]^2 \right\rangle \propto N^{2H_{T_i}}. \quad (5.9)$$

The data show a crossover, similar to that for f , N and V , from a lower to a higher value of H_{T_i} when the window size is approximately the daily mean number of trades (for an example, see the inset of Fig. 5.9). For the restricted set studied in Ref. [IYPL04], the value $H_T \approx 0.94 \pm 0.05$ was suggested for window sizes above the crossover.

Much similarly to the case of other quantities analyzed earlier, the inclusion of more stocks¹⁰ reveals the underlying systematic non-universality. Again, less frequently traded stocks appear to have weaker autocorrelations as H_T decreases monotonically with growing $\langle T \rangle$. One can fit an approximate logarithmic law^{11,12} to characterize the trend:

$$H_{T_i} = H_{T^*} + \gamma_T \log \langle T_i \rangle, \quad (5.10)$$

where $\gamma_T = -0.10 \pm 0.02$ for the period 1994 – 1995 (see Fig. 5.9) and $\gamma_T = -0.08 \pm 0.02$ for the year 2000 [upo].

In a yet unpublished work, Yuen and Ivanov [YI05] have independently discovered the same tendency (5.10) for intertrade times of NYSE and NASDAQ in a different set of stocks.

5.5 Conclusions

In this chapter we revisited some “stylized facts” of stock market data and found in several ways alterations from earlier conclusions. The main difference in our approach was – besides the comparative application of extrapolation techniques – the extension of the range of capitalization of the studied firms. This enabled us to investigate the dependence of the trading characteristics on capitalization/liquidity. In fact, in many cases we found fundamental dependence on these parameters.

We have shown that liquidity $\langle f \rangle$, the number of trades per minute $\langle N \rangle$ and the mean size of transactions $\langle V \rangle$ display non-trivial, but monotonic dependence on company capitalization.

¹⁰For a reliable calculation of Hurst exponents, we had to discard those stocks that had less than $\langle N \rangle < 10^{-3}$ trades/min for 1994 – 1995 and $\langle N \rangle < 2 \times 10^{-3}$ trades/min for 2000. This filtering leaves 3519 and 3775 stocks, respectively.

¹¹As intertrade intervals are closely related to the number of trades per minute $N(t)$, it is not surprising to find the similar tendency for that quantity, cf. Section 4.2.2.

¹²Note that for window sizes smaller than the daily mean number of trades, intertrade times are only weakly correlated and the Hurst exponent is nearly independent of $\langle T \rangle$. This is analogous to what was seen for traded value records in Section 5.3.

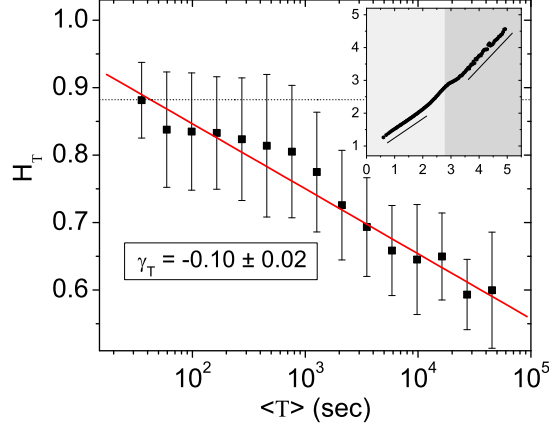


Figure 5.9: Hurst exponents of T_i for time windows greater than 1 day, plotted versus the mean intertrade time $\langle T_i \rangle$. Stocks that are traded less frequently, show markedly weaker persistence of T for time scales longer than 1 day. The dotted horizontal line serves as a reference. We used stocks with $\langle T \rangle < 10^5$ sec, the sample period was 1994 – 1995. The inset shows the two regimes of correlation strength for the single stock General Electric (GE) on a log-log plot of $\sigma(N)$ versus N . The slopes corresponding to Hurst exponents are 0.6 and 0.89.

We have given evidence that the distribution of traded value in fixed time windows is not Lévy stable. If a power law is fitted to the tail of the distribution, a careful analysis yields to an exponent λ , which is – even for short time windows – in most cases greater than 2, and then increases with increasing time window indicating the existence of the second moment of the distribution. This seems like a technical issue, but it has far-reaching consequences for a recent theory of price impact [GGPS03]. According to this theory price changes directly follow from trading volume. The return (also called price impact) induced by a transaction of size V is $r \propto V^\mu$ with $\mu \approx 1/2$. Consequently the distributions of the two quantities are not independent. In particular the tail exponent λ_r of the returns and the tail exponent λ of trading volume are related as $\lambda = \mu\lambda_r$. Previous analyses found the former value to be $\lambda_r \approx 3.1$ [GPGS00], while the latter $\lambda \approx 1.7$ [GMAS98, PGA⁺99]. This theory and the values of the exponents have been disputed extensively in the recent literature [FGL⁺04, EK07a, EK06b, FL04, Que05].

Another consequence of the revised distribution of f is that the Hurst exponent H for its variance can be defined. For long time horizons this, and the related exponents of other similar quantities show strong dependence on liquidity well described by some logarithmic laws. These can be

extended to the case of a multiscaling formalism for higher moments.

The distribution of the waiting times between trades is better described multiscaling than by gap scaling. It is characterized by an increase in both correlations and relative fluctuations with growing trading frequency.

Our findings indicate that special care must be taken when concepts like scaling and universality are applied to financial processes. The modeling of the market should be extended to the capitalization dependence of the characteristic quantities and this seems a real challenge at present.

5.6 Appendix: The estimation of tail exponents

In this appendix, as an addition to Section 5.2, we detail the techniques applied for the estimation of the tail exponent. For every measurement therein we will give the median estimates of λ for the 1000 stocks with highest traded value during the investigated period. The error bars correspond to the half distance between the 25% and 75% quantiles of the λ distribution.

5.6.1 Hill's estimator

Hill's estimator [Hil75] is a statistically consistent method to estimate the tail exponent λ from random samples taken from a distribution that asymptotically has the power-law form (5.1). The procedure first sorts the sample $f(t = 1 \dots L)$ in decreasing order. We are going to denote this series by $f[t]$, so that $f[1] > f[2] > f[3] > \dots$. Then, one defines the tail of the distribution by setting an arbitrary number k of points to be included in the estimation procedure. The estimate of the inverse tail exponent is

$$\hat{\lambda}^{-1}(k) = \left[\frac{1}{(k-1)} \sum_{t=1}^{k-1} \log f[t] \right] - \log f[k], \quad (5.11)$$

given that $k \rightarrow \infty$ and $p = k/L \rightarrow 0$. If the sampled distribution is of the form (5.1), then by increasing k , the estimator converges rapidly to the actual value of λ^{-1} . However, in the case of traded value data, this turns out not to be the case.

The inset of Fig. 5.10(top left) – a so called Hill plot – shows, that there is a systematic dependence of λ on p and no convergence is observed. With the inclusion of less tail events, the exponent increases sharply, beyond the $\lambda = 2$ threshold for Lévy stability. Further evidence for the lack of Lévy stability is that on increasing the time scale Δt , the estimated tail exponents also increase further as shown in Fig. 5.10(a).

This type of behavior is not new to mathematical statistics (see, e.g., Ref. [Alv01]). It is possible, that the distribution decays faster than a

power law and thus no finite λ exists. Alternatively, the power law may not be centered around zero, but instead it can be of the form

$$\mathbb{P}_{\Delta t}(f) \propto (f + f_0)^{-(\lambda+1)}. \quad (5.12)$$

In this latter case, there is a finite λ , but as the sample size is usually too small, the estimator displays the above bias. One can either try to approximate the value of f_0 and shift the data accordingly, so that Hill's estimator converges properly, or try to find another estimator that is insensitive to this shifting constant.

We have tried both approaches and they yielded qualitatively similar results.

5.6.2 Shifted Hill's estimator

One can apply Hill's estimator to the points $f[t = 1 \dots L] + \varphi \langle f \rangle$, where φ is a constant parameter and look for a value, where the estimator $\lambda(k)$ becomes independent¹³ of k , i.e., Hill's estimator truly finds a power-law decay that is now consistent with Eq. (5.12). This happens, when $\varphi \langle f \rangle = f_0$. How this shift by $\varphi \langle f \rangle$ affects the Hill plots is shown in Fig. 5.10(bottom left) for the case of $\Delta t = 15$ min. One finds, that in this case $\varphi \approx 2.8$ gives reasonable results, while $\lambda = 2.55 \pm 0.35$. One can repeat the procedure for various time scales Δt . The median Hill plots are shown in Fig. 5.10(bottom right), while $\lambda(\Delta t)$ and $\varphi(\Delta t)$ are given in Table 5.1(top). Again, one finds a significant increase of the tail exponent with growing Δt . This underlines our previous expectation that traded value distributions are not Lévy stable and thus have a finite variance.

5.6.3 Fraga Alves estimator

A more sophisticated approach to estimate tail exponents of distributions of the type (5.12), is a recent variant of Hill's method, proposed by Fraga Alves [Alv01]. The algorithm is described in detail in Appendix 5.6.4 and its estimates of λ are – in an exact mathematical sense – independent of the shift f_0 present in the density function, unlike those of the original Hill's estimator (5.11).

We applied the estimator to the same dataset, the Hill plots for $\Delta t = 1, 15, 60$ min are shown in Fig. 5.10(bottom right). What one finds is a very different behavior from the shifted Hill's estimator. The estimate of λ increases with growing p , i.e., the more points included. This is due to that the Fraga Alves estimator converges much slower than Hill's

¹³More precisely, we increased φ from 0 by increments of 0.2 and looked for $\lambda(k, \varphi) \approx \lambda(\varphi)$. The method is very sensitive to the proper choice of φ . For high values of Δt , there is a low number of data points, and the estimates of λ may be very noisy. In this case we chose φ , where the estimate of λ is lower.

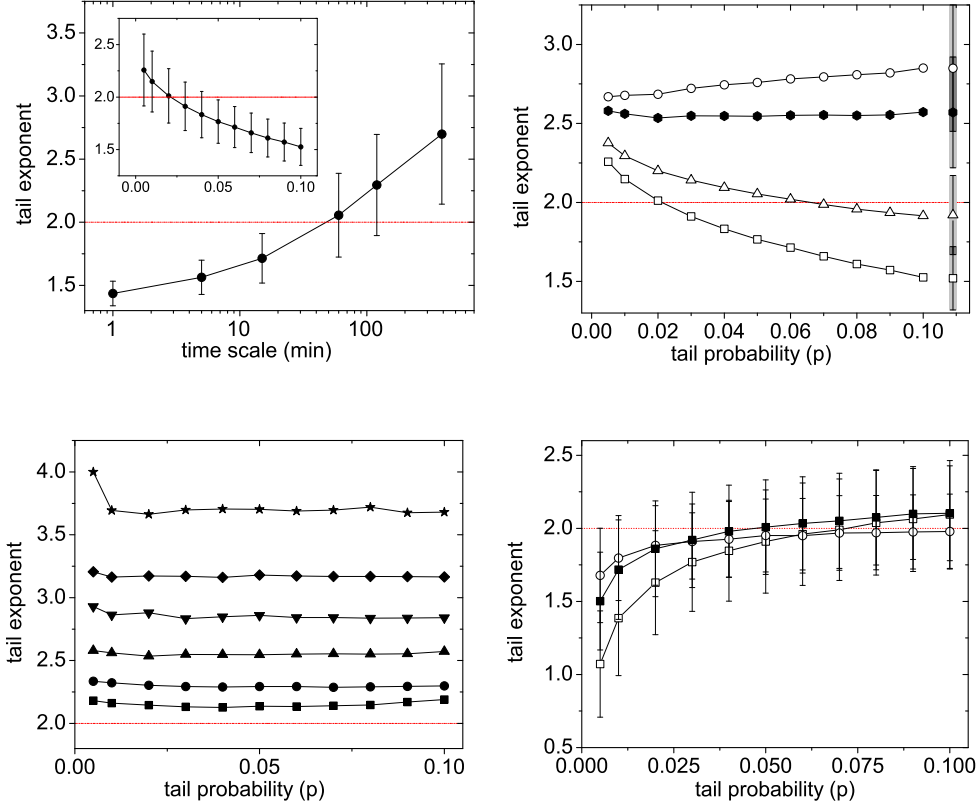


Figure 5.10: **(top left)** Hill's estimates of λ for different sizes of the time window with the tail probability set as $p = 0.06$. The monotonic trend indicates that the distribution is not Lévy stable. The inset shows, that for $\Delta t = 15$ min the effective tail exponent λ depends monotonically on the choice for tail probability p . Thus, Hill's estimates are unreliable, because they depend strongly on an arbitrary parameter. **(top right)** Dependence of the Hill plots for $\Delta t = 15$ min on the shifting constant φ . The values of φ from bottom to top: 0 (\square), 1 (\triangle), 2.8 (\bullet , optimal shift), 3.0 (\circ). Typical error bars are given on the right, darker gray indicates the regimes where they overlap. **(bottom left)** Hill plots of the optimally shifted Hill's estimators for various time windows. The values of Δt from bottom to top: 1 min (\blacksquare), 5 min (\bullet), 15 min (\blacktriangle), 60 min (\blacktriangledown), 120 min (\blacklozenge), 390 min (\blackstar). One finds $\lambda > 2$ and the strong increasing tendency in λ with Δt implies that the distribution is not Lévy stable. **(bottom right)** Hill plots of the Fraga Alves estimator for three time window sizes Δt : 1 min (\circ), 15 min (\blacksquare), 60 min (\square). The method gives a lower estimate of $\lambda \approx 2$.

estimator, and – as Fig. 5.10(bottom right) and Monte Carlo simulations on surrogate datasets indicate – it converges from below. On the other hand, setting the threshold as high as $p = 0.1$ may include events that no more belong to the power law regime, which also results in a reduced, effective exponent due to the shape of the distribution, shown in Fig. 5.2. Consequently, this method provides a lower estimate of λ . Still, the calculated values are mostly above 2. Finally, one must note that for $\Delta t \geq 120$ min, the number of points was inadequate to provide a proper estimate.

5.6.4 The algorithm of the Fraga Alves estimator

Ref. [Alv01] describes a method to approximate the parameter λ from a sample of a random variable that is asymptotically distributed as

$$\mathbb{P}_{\Delta t}(f) \propto (f + f_0)^{-(\lambda+1)}.$$

First, one sorts the sample $f(t = 1 \dots L)$ in decreasing order. We denote this series by $f[t]$, so that $f[1] > f[2] > f[3] > \dots$. Then, the procedure consists of the five steps formulated below:

1. $k_0^* = 2k^{2/3}$

- 2.

$$\hat{\lambda}^{-1}(k_0^*, k) = \frac{1}{k_0^* - 1} \sum_{t=1}^{k_0^*-1} \log \frac{f[t] - f[k]}{f[k_0] - f[k]}$$

- 3.

$$k_0 = C_0^{1/(2\hat{\lambda}^{-1}(k_0^*, k)+1)} k^\alpha,$$

where

$$C_0 = \frac{(1 + \hat{\lambda}^{-1}(k_0^*, k))^2}{2\hat{\lambda}^{-1}(k_0^*, k)},$$

and

$$\alpha = \frac{2\hat{\lambda}^{-1}(k_0^*, k)}{2\hat{\lambda}^{-1}(k_0^*, k) + 1}.$$

- 4.

$$\hat{\lambda}^{-1}(k_0, k) = \frac{1}{k_0 - 1} \sum_{t=1}^{k_0-1} \log \frac{f[t] - f[k]}{f[k_0] - f[k]}$$

5. Finally, the estimate of the inverse tail exponent is given by

$$\lambda^{-1}(k_0, k) = \hat{\lambda}^{-1}(k_0, k) - \sqrt{\frac{\hat{\lambda}^{-1}(k_0, k)}{2k_0}}.$$

$\lambda^{-1}(k_0, k)$ converges to the inverse tail exponent, if $L \rightarrow \infty$, $k/L \rightarrow 0$ and $k_0/k \rightarrow 0$.

Chapter 6

Introduction to order books

The previous chapters mainly focused on the phenomenological properties of trading activity. However, it is an at least equally interesting question how this activity is structured on the level of individual investors. The trading process, price formation and the consequences of market organization are classical areas of the analysis of market microstructure [BGS05]. A large number of studies has been focusing on modeling the dynamics of trading, with and without specialists, in financial markets. They comprise both theoretical [Glo94, CH95, Sep97, Par98, DFIS03, Luc03] and empirical studies [BHS95, HS96, HH96, Kav99, San01, MM01, CS01, LMZ02, PB03, HMS04, FGL⁺04, FPZ05, MF05, WR05, PLM06] devoted to the determinants of price formation, the trading process and market organization. The market mechanism, along with the complex interactions among market participants result in the emergence of a collective action of continuous price formation. This is an extremely difficult problem, and its studies can contribute greatly to the success of the modeling of financial markets.

The order book

A particular type of double¹ auction based on the limit order book is used to organize most developed markets. Some studies of the order book have used an agent based modeling approach. Examples are market models described in terms of agents interacting through the book by using simple rules [CP03] and models where the assumptions about the trading strategies are kept as minimal as possible (see for example Refs. [DFIS03, MF05]). One of the most striking findings was that even if trends and investor strategies are neglected, purely random trading may be adequate to describe certain basic properties of the order book [FPZ05].

Our approach will instead focus on the empirical analysis of order

¹i.e., both buyers and sellers compete among themselves.

books, based on the trading data of the electronic market (SETS) of London Stock Exchange (LSE) during the year 2002. The general concepts are the same in other markets, although there can be differences in regulations and the details of allowed transaction types².

The limit order book is basically a database of buy and sell limit orders, i.e., offers to buy or sell a given volume of a certain asset at a predetermined price. The limit orders with better prices are executed first, while those with the same price are executed in the order of their placement. The highest price to buy is usually called the best bid price $b(t)$, and the lowest price to sell the best ask price $a(t)$. The distance between the two is called the bid-ask spread s , $s(t) = a(t) - b(t)$.

Except for very special cases, there are already other limit orders waiting inside the book when one wants to place a new one. Let $b(t) - \Delta$ denote the price of a new buy limit order, and $a(t) + \Delta$ the price of a new sell limit order. The mechanism of limit order placement is illustrated in Fig. 6.1. Let us take buy orders for example. An incoming buy order can be placed in three regimes with respect to the distance Δ from the current best: (i) exactly at the existing best bid price, $\Delta = 0$, (ii) at a better (higher) price, $\Delta < 0$, this means that the order is placed inside the spread, (iii) or at a lower price, $\Delta > 0$ below the current best bid/ask, deeper "inside the book". For sell orders one can give analogous definitions. Any limit order which was not executed can be canceled at any time by the trader who placed it. The order can also have a predetermined validity after which it is automatically removed from the book, this is called expiration.

There exists another type of orders called market orders, these are characterized by a given volume and the traded asset, but they have no prefixed price. Instead, they are requests to buy/sell immediately at the best price available from counterparties with currently standing limit orders. It is also possible to have limit orders with such large negative values of Δ that they cross the spread, i.e., $\Delta \leq b(t) - a(t)$. These orders are called crossing orders, and they can be immediately paired with orders from the other side of the book. Given that enough volume of opposite limit orders is available, they are fully executed right away. Otherwise the unmatched part of the crossing limit order remains in the book as a new limit order. Since a trader would place a crossing limit order to execute part of it immediately, we will not consider them as limit orders in our analysis.

The opening times of LSE are divided into three periods. The intervals 7:50–8:00 and 16:30–16:35 are called the opening and closing auction, respectively. These follow different rules and thus also observe different statistical properties than the rest of the trading. Therefore one usually

²For example in the Chinese market – among other peculiarities – traders are not allowed to both buy and sell stocks on the same day [ZQJZ07].

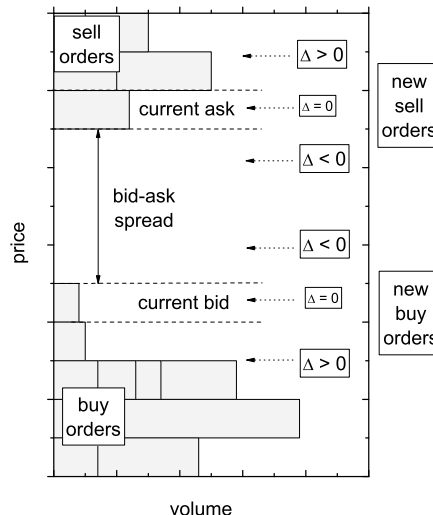


Figure 6.1: The schematic structure of the limit order book. Buy limit orders are placed at lower, sell limit orders at higher prices. The regime between the highest buy and lowest sell offer is the bid-ask spread. The placement of a new incoming limit order can be characterized by the Δ distance from the best offer on the same side of the book. $\Delta = 0$ for an order which is placed at the price of the current best offer. We define $\Delta < 0$ for orders that were placed inside the spread, and $\Delta > 0$ which are placed inside the bulk of previous orders.

neglects limit orders placed during these times, and focuses only on the periods of continuous (double) auction during 8:00–16:30. We also removed limit orders that were placed during 8:00–16:30 but were canceled (or expired) during the opening/closing auctions. Similarly to the case of TAQ data in Chapters 4 and 5 we measure time intervals as continuous trading time. In this case this starts from 8:00:00 on January 2, 2002.³

Whenever we refer to prices, unless stated otherwise, we exclude all transactions that were executed on the SEAQ market⁴ and not by means of the limit order book.

Finally, let us point out that in most of the literature the logarithm of the price is modeled, while we intentionally use price itself. Our study is concerned with very small price changes on the order of the spread, when there is little discrepancy between the two approaches. However it

³Every trading day lasts exactly 31501 seconds, except for December 24, 2002 and December 31, 2002, which are 17101 seconds long. In our analyses, we removed the data of trading on September 20, 2002. This is because on that day very unusual trading patterns were observed, including an anomalous behavior of the bid-ask spread.

⁴Many studies refer to this colloquially as the “upstairs” market.

is important to keep bare prices, as stocks have a finite tick size (minimal price change). Taking a log-transformation would mean that the minimal increment would depend on the price itself. Omitting the transformation will enable us to classify the orders into discrete categories by price difference. The size of ticks depends on the stock, the possible values are $1/4$, $1/2$ or 1 penny.

Chapter 7

Time scales and dynamics of order books

The most ambitious goal of econophysics would be to provide a solid microscopic theory of market participants, their interactions, supply and demand by integrating elements from economics and other areas. Based on this theory one could build from bottom up and see how the many rapidly fluctuating, interacting individuals create the long-term behavior, and finally contribute to macroeconomics. Such a unified theory seems currently out of reach, but there has been much progress on different levels. Behavioral finance and experimental economics [BT03] have clarified many aspects of how individuals make their decisions. At the other end of the spectrum, regarding the market as a whole, econometrics and more recently econophysics have uncovered several "stylized facts" of prices and trading activity [CLM96, BP00]. By analyzing the structure of limit order books recent studies hope to provide some of the missing links between these two levels [BGS05, MM01, FPZ05, SFGK03, BMP02, MF05].

At present every level of description has a characteristic time scale on which its models are valid. The typical price change on these durations acts as another important scale. In this chapter we will present a comparison of the phenomenology of stock market data as we go from monthly to tick-by-tick resolution. Limit order book data contains the maximal amount information about the state of the market that is available to traders. Our aim is to show how much of this microstructure is important at the different time scales. We present qualitative evidence that mainly supports recent theories, but also points out some of their limitations.

We investigated several liquid stocks, but here we will only present one: GlaxoSmithKline (GSK). The results are qualitatively similar for all stocks where the tick/price ratio is small enough. For the readers' convenience a comparison of the typical time and price scales is shown in Tables 7.1 and 7.2.

process/event	time scale (sec)
trading day	3.15×10^4
crossover to diffusivity	$3 \times 10^2 - 10^3$
relaxation of liquidity fluctuations	$10^2 - 10^3$
intertrade time	10

Table 7.1: Typical time scales for the stock GSK in the year 2002. Intertrade times were calculated as a mean for the dataset, over the 3000 – 28000'th seconds of every trading day. The time scale for the relaxation of liquidity fluctuations was taken from Ref. [PLM06]. The crossover time to diffusivity will be discussed in Chapter 8, cf. Ref. [EKLM07].

process/quantity	price scale (ticks)
midquote price	1.4×10^3
median-to-median width of book	68
1-month absolute return	1.4×10^2
1-day absolute return	23
distance of "stripes"	5
10-min absolute return	3
bid-ask spread	1.9
1-min absolute return	0.9

Table 7.2: Typical time scales for the stock GSK in the year 2002. The distance of "stripes" was determined from Fig. 7.2. All other quantities represent means for the dataset, over the 3000 – 28000'th seconds of every trading day.

7.1 Monthly and longer time scales

On monthly or longer time horizons stock prices are often modeled as an ordinary diffusion process with a drift [BP00, BS73]. Although this is an extreme simplification of the actual behavior, the autocorrelations of volatility and the fat tails of the return distribution are effectively diminished on such a long time scale, and so they need not be taken into account. The difference between bid and ask, opening and closing prices is also insignificant: the typical monthly absolute return is two orders of magnitude larger than the spread and about one order of magnitude larger than the difference between the opening and closing prices.

The time evolution of the order book during the whole year 2002 is shown in Fig. 7.1. From this plot one gains little additional information

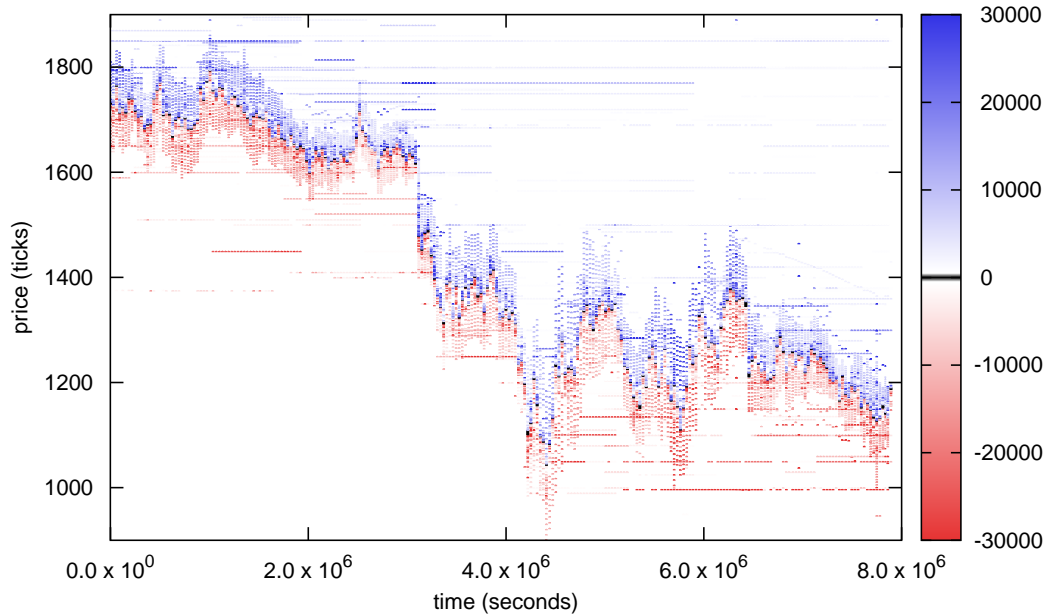


Figure 7.1: The order book of the stock GSK during the year 2002. For all 252 trading days during the year we made a snapshot of the order book at the 15000'th second of the day. The lack of orders is indicated by white color, buy/sell orders by red/blue, and the bid-ask spread by black. For every price level we indicated the total volume of limit orders by coloring as indicated on the right. The ends of the scale correspond to orders for 30000 shares or more.

as compared to a plot of only the price. The bid-ask spread, as noted before, is effectively negligible. One can also clearly see that the shape of the order book is – on average – symmetric between bids and asks. The book contains a significant number of orders very far, sometimes 100 ticks away from the best bid/ask. The width of the order book fluctuates in time, but one can see from Fig. 7.1 that it reverts to the typical value on the scale of weeks. This width can be calculated: on average, 50% of the total limit order volume (both bid and ask) is within a 68 tick range. This is comparable to the monthly absolute return, which is 1.4×10^2 ticks.

7.2 Daily time scale

From previous analyses of TAQ data (see Section 2.2) it is well known that daily returns look very different from monthly returns. In contrast with the ordinary diffusion picture of monthly data, daily returns are strongly fat tailed and day-to-day volatility autocorrelations are significant [BP00]. On this scale there is also a negative correlation between past returns and future volatility: a price drop induces a greater excess volatility than an equal price increase [BMP01]. The bid-ask spread is still small, its mean value is 1.9 ticks, while the daily absolute return is 23 ticks on average. Note that, for trading strategies with the time horizon of a few days the spread is no longer negligible, because although it is small compared to volatility, it is no longer small as compared to their expected profits.

The trading day is a characteristic time in many respects. Trading activity and volatility are known to follow intraday patterns, with typically higher values immediately after the opening and before the closing of the market. Even more importantly, for trading strategies a day is an important time horizon: day traders are required to close their positions at the end of the day, and even long-term strategies such as large investment portfolios are often adjusted daily. Consequently it is not surprising that most limit orders are canceled at the end of the trading day, and new limit orders are placed at the beginning of the next one. The first 32 trading days of the data period are shown in Fig. 7.2, where this daily cycle can be seen very clearly. The orders that remain in the book overnight are typically placed at technical levels [Osl03] constituted by round numbers. For example these are 1700 and 1650 ticks in the first week in Fig. 7.2.

Previous studies have reported a slow, continuous, symmetric decay of limit order volume as a function of the distance from the spread (see, e.g., Ref. [PB03]). While this true on average, it is very uncharacteristic of the state of the book at any given time. This difference has predominantly the following two origins:

(i) In Fig. 7.3(left) we plot the average shape of the order book over the whole year. Then in Fig. 7.4 the shape in the 15000'th second of the first 18 trading days is compared with the average pattern¹. The shape of the order book varies strongly from day to day, but it is always very different from the average. In other words, the *average* shape is not *typical*. For example, Fig. 7.3(right) shows the shape of the book as an average over the first 10 days of our data. Even such a time average looks very different from the yearly average.

(ii) From Fig. 7.2 one can see that the book consists of three types of orders. (a) There is a narrow range around the bid-ask spread where there

¹In fact the average shape of the book shows some variation during the trading day. Because Fig. 7.4 displays snapshots taken at the 15000'th second of trading days, the average was also calculated over such snapshots.

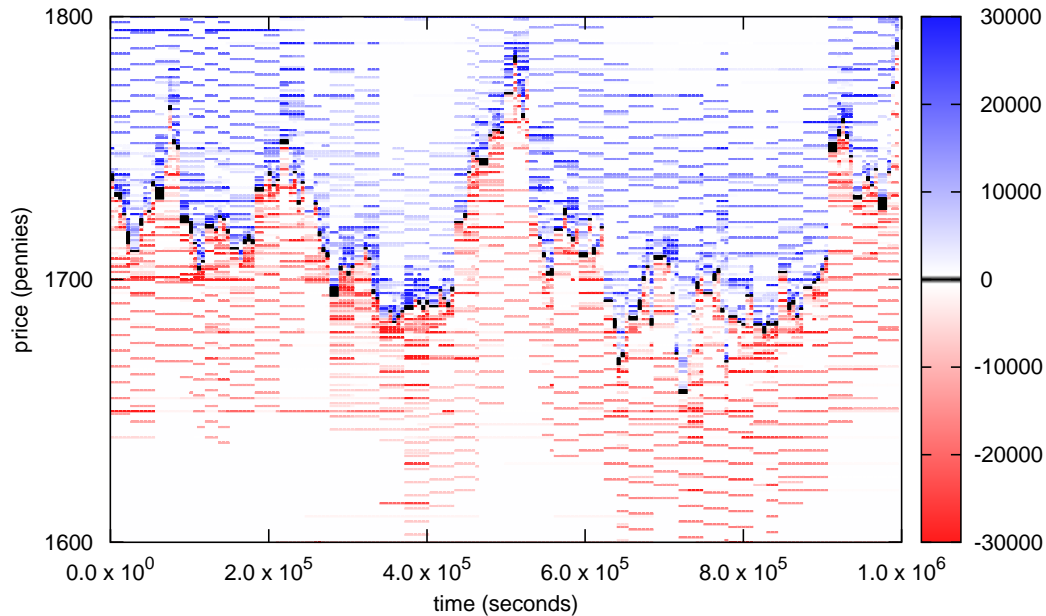


Figure 7.2: The order book of the stock GSK during the first 32 trading days of 2002. For all days during the period we made five snapshots of the order book at the 5000'th, 10000'th, \dots , 25000'th second of the day. The lack of orders is indicated by white color, buy/sell orders by red/blue, and the bid-ask spread by black. For every price level we indicated the total volume of limit orders by coloring as indicated on the right. The ends of the scale correspond to orders for 30000 shares or more.

is a high concentration of limit orders carrying a substantial fraction of the volume in the book. (b) Fig. 7.2 is also very "striped" along the price axis: there is a regular pattern of moderately sized orders every 5 ticks to a distance of up to 100 ticks from the spread. The same can be seen in most snapshots in Fig. 7.4. (c) In addition there appear to be much larger limit orders placed around round numbers, such as 1700 and 1650 ticks in this example (see Fig. 7.1).

The central part (a) will be addressed in the next section. The regularly striped pattern (b) contains relatively smaller volumes, and it does not show up at all in the yearly or even the 10-day average book shape in Fig. 7.3. The reason is that most of the orders constituting pattern (b) remain unchanged during the whole trading day, and hence they con-

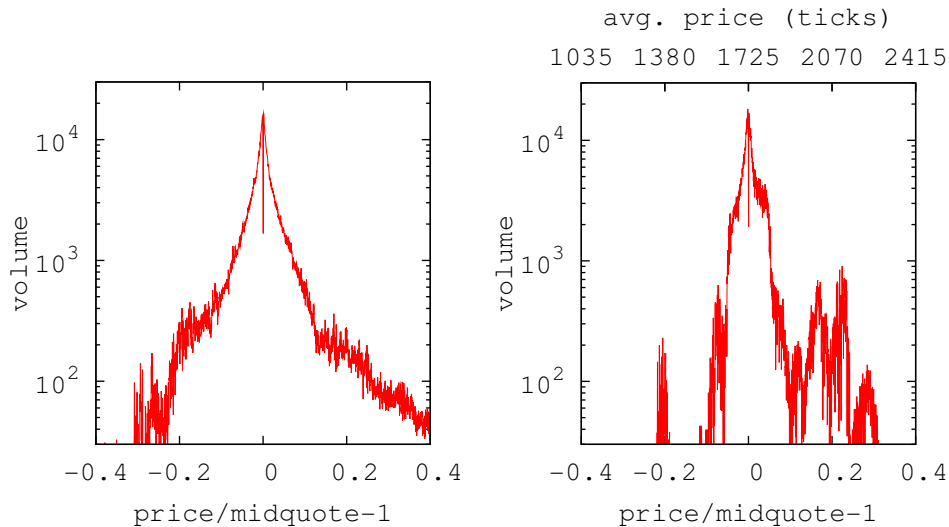


Figure 7.3: **(left)** The average shape of the order book of GSK calculated during the year 2002. The prices were then divided by the current midquote price, and we subtracted 1. **(right)** The average shape of the order book of GSK calculated during the first 10 trading days of the year 2002. The indicated average prices correspond to $1725 \times (\text{price}/\text{midquote} - 1)$ ticks. *Note:* Averages were calculated by taking snapshots of the book during every trading day of the period from its 3000'th to its 28000'th second every 1000 seconds. The prices were then divided by the current midquote price, and we subtracted 1.

tinuously change position relative to the midquote price. Thus a time average in which distance is calculated from the midquote will only indicate a continuous pattern. Also notice that the pattern is not fixed to round numbers or quotes ending to 5 or 0. If the price moves, the pattern is replaced at different levels on the next trading day.

Finally the orders at the round levels (c) are relatively large. They do not appear in the yearly average book shape, but some are large and stable enough to be seen in, e.g., the 10-day average in Fig. 7.3(right). The average was made in a period when the midquote price fluctuated around 1725 ticks, and so an approximate price axis could also be added to the plot. There are several clear peaks corresponding to limit orders that are very far from the spread.

From Fig. 7.4 one can see that not only the total volume inside the book fluctuates, but there is always an asymmetry present between the buy and sell sides, whose degree and direction varies in time. This imbalance can be quantified, for example, as follows. Let us denote the buy/sell

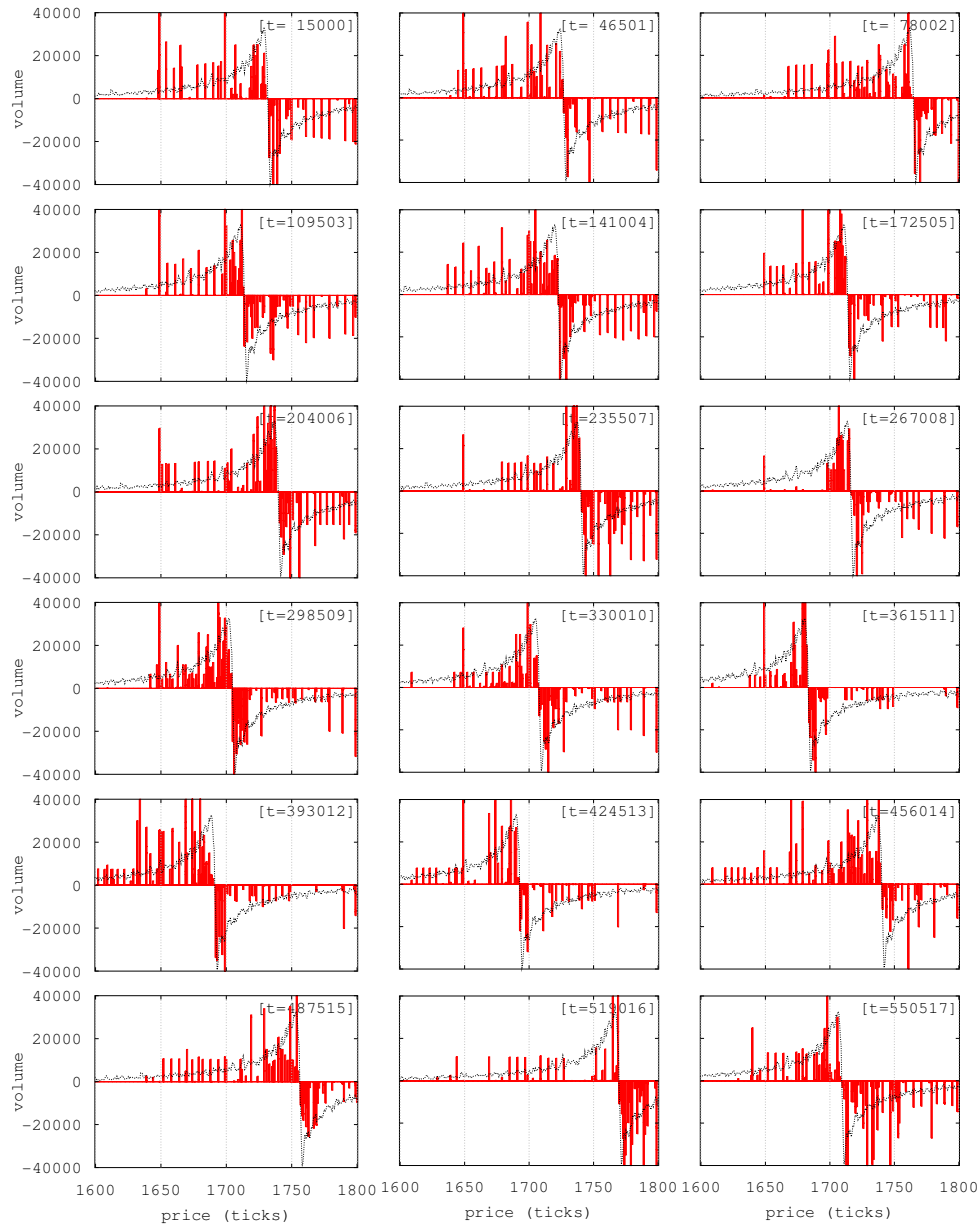


Figure 7.4: The order book of the stock GSK during the first 18 trading days of 2002. For all days we made snapshots of the order book at the 15000'th second of the day. The length of the bars indicates the total volume of the orders at that price level, the maximum length corresponds to volumes for 40000 shares or more. The dotted line corresponds to the time average of the book shape at the 15000'th second of trading days. The shape is similar, but the amplitude is somewhat larger as in Fig. 7.3(left).

volume at price p by $v_t^{\text{buy/sell}}(p)$. The total buy/sell volume is given by

$$V_t^{\text{buy}} = \sum_{p=0}^{b_t} v_t^{\text{buy}}(p)$$

and

$$V_t^{\text{sell}} = \sum_{p=a_t}^{\infty} v_t^{\text{sell}}(p).$$

We define the buy/sell imbalance as

$$I_t^{\text{buy}} = \frac{V_t^{\text{buy}}}{V_t^{\text{buy}} + V_t^{\text{sell}}}$$

and

$$I_t^{\text{sell}} = -\frac{V_t^{\text{sell}}}{V_t^{\text{buy}} + V_t^{\text{sell}}}.$$

Trivially $I_t^{\text{buy}} - I_t^{\text{sell}} \equiv 1$. The behavior of the imbalance is plotted in Fig. 7.5. The relationship between daily return and imbalance is less than straightforward. For example, there is a several day period around $t = 4 - 5 \times 10^5$, where there is a consistent upward price trend combined with a long lasting excess of buy limit orders. The pattern is then repeated downwards for $t = 5 - 7 \times 10^5$. On the other hand, in the period $t = 8 - 9 \times 10^5$ there is an upward trend and still there are excess sell limit orders. The lack of direct causality between limit order volume and returns is of course not very surprising. The efficient nature of the market would not allow such a simple predictor of returns to exist, but the relationship between imbalance, returns and predictability is a question still under debate [FGLM06, WR05, BKP06, FGL⁺04].

7.3 Intraday and tick-by-tick time scales

Finally, let us look at intraday and tick-by-tick data. The nature of the price formation process only becomes crucial on the very short time scales, comparable with the average time between consecutive trades. This was 10 sec for GSK in particular, but it is always of the order of seconds for any liquid stock or market [EKLM07, IYPL04].

The 8'th day of trading is shown in Fig. 7.6 from the order book perspective. For time horizons longer than 5 – 15 minutes any liquid stock is traded tens-hundreds of times, and there is evidence, that the qualitative picture of anomalous diffusion for prices describes limit order executions well [EKLM07]. The typical time scale when the mean absolute return exceeds the mean bid-ask spread is also similar: The mean 1 and

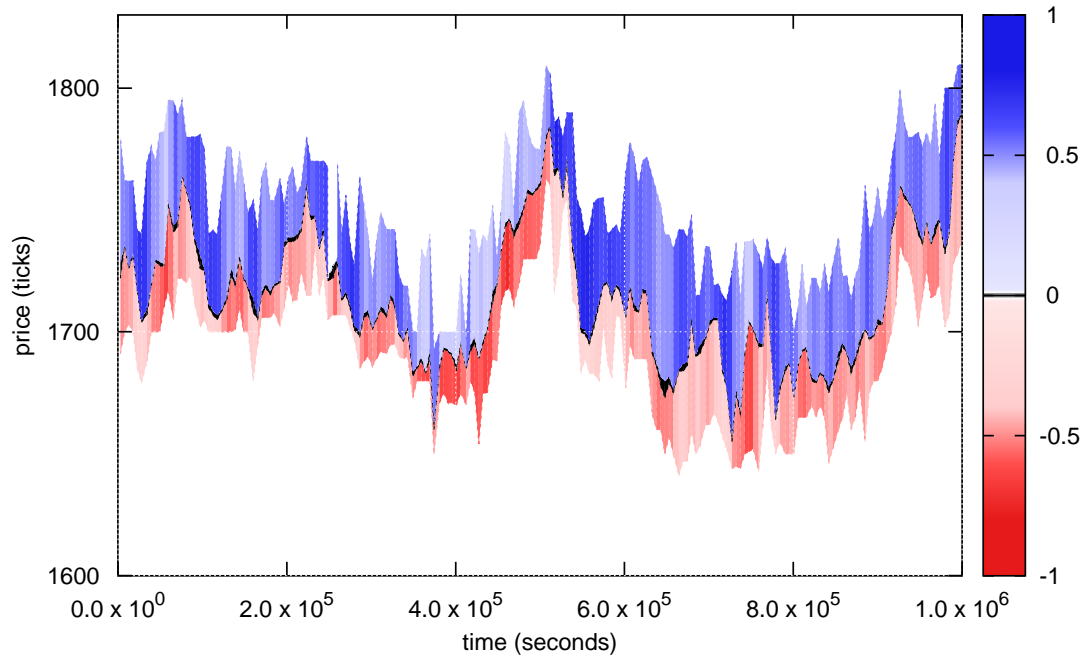


Figure 7.5: The order imbalance and the median-to-median width of the order book of GSK calculated during the first 10 trading days of the year 2002. We took six snapshots of the book during every trading day of the period from its 3000'th to its 28000'th second every 5000 seconds. Red/blue color corresponds to buy/sell orders, and the black line between them to the bid-ask spread. The coloring also indicates the value of I_t^{buy} and I_t^{sell} (see right for values), while the width of the graph shows the median of the book in terms of its total volume.

10-minute absolute returns are 0.9 and 3 ticks, respectively, while the mean bid-ask spread is 1.9 ticks.

As seen in Fig. 7.4, during usual trading there is a regime of up to 10 occupied levels immediately near the best prices. The regular striped pattern every 5 ticks can be clearly present and during the day the levels are constant.

In order to see more temporal structure one has to move down to the scale of individual events, where the microstructure of the order book truly begins to play a role. Fig. 7.7 shows a 4000 second period during the end of the same 8'th day. To visualize the dataset we used the technique intro-

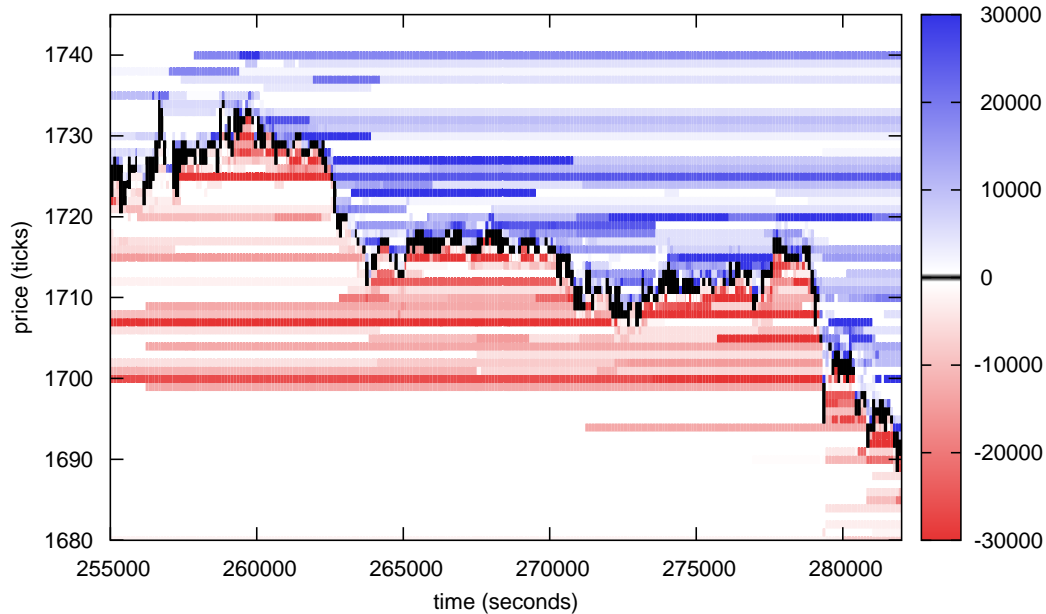


Figure 7.6: The order book of the stock GSK during the 8'th trading day of 2002. During the period a snapshots of the order book every 50 seconds. The lack of orders is indicated by white color, buy/sell orders by red/blue, and the bid-ask spread by black. For every price level we indicated the total volume of limit orders by coloring as indicated on the right. The ends of the scale correspond to orders for 30000 shares or more.

duced by Ponzi et al. [PLM06]. This is the finest level of price formation. Here we will very briefly sketch a qualitative picture as suggested by Refs. [PLM06, WR05, FGLM06, FGL⁺04].

During usual trading there is enough volume in a narrow range around the spread to satisfy market orders, and so trading is confined there. The bid/ask can basically change in three ways: (i) A typical market order has a volume which is less or equal to the best opposite limit order. If a market order arrives, it either leaves some volume at the bid/ask so the level does not move, or it removes the entire best offer but usually not more, and so the change in the quote equals the gap between the best and the second best opposite limit price [FGL⁺04]. (ii) A new limit order is placed inside the spread. (iii) All orders at the best price are canceled. The shift in the bid/ask price due to an event is called its *price impact*. Case (iii) is

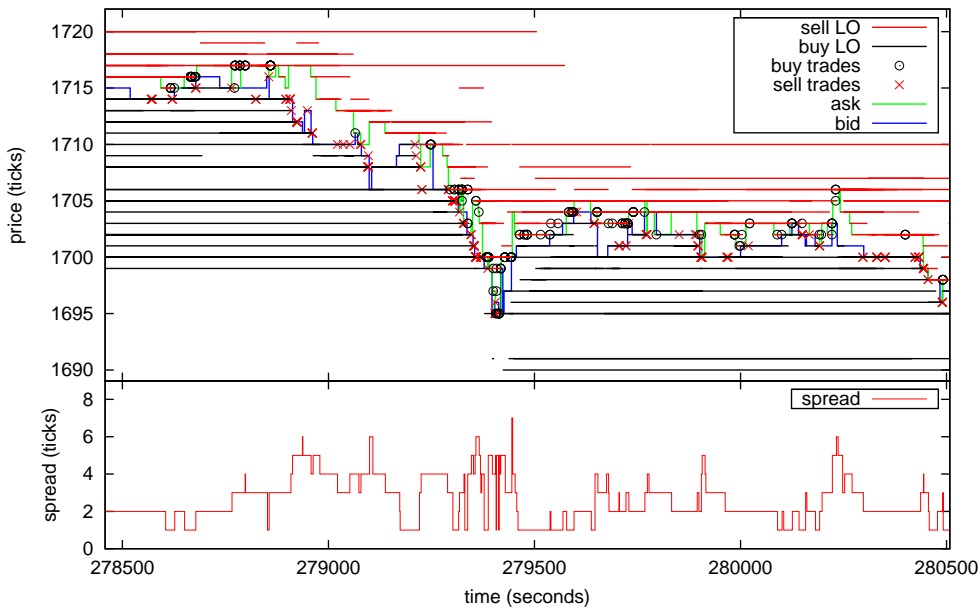


Figure 7.7: The trading activity of the stock GSK between the 26500 – 28500'th seconds of the 8'th trading day of 2002. In the top half of the plot price levels with buy/sell limit orders are indicated by black/red horizontal lines. The bid and ask are indicated by blue/green lines. During the period a snapshots of the order book every 50 seconds. The lack of orders is indicated by black color, buy/sell orders by red/blue, and the bid-ask spread by white. Buyer initiated transactions are marked by black circles, while seller initiated transactions by red crosses.

quite rare (see Section VI. of Ref. [PLM06]). The immediate impact of cases (i) and (ii) is the opposite, i.e., the spread increases or decreases. The information content of an order placement and an execution is also very different, so it is not surprising that the caused variation in the bid-ask spread will evolve further in different ways. The reaction is faster for an increase in the spread than for a decrease, and the mean reversion is stronger. Moreover, transactions have a stronger effect than other events. Because the type of market orders (buy or sell) are long time correlated, a buy market order for example is likely to induce further buy market orders or to be a reaction to those. Thus even if a buy market order *itself* does not change the ask price, the ask price is likely to increase on average due to subsequent events [PLM06].

Much more rarely one larger or many smaller limit orders arrive, which wipe out a large portion of the book and thus result in an abrupt increase of

the bid-ask spread which correspond to a decrease in liquidity. This is followed by intense fluctuations, see for example the period near $t = 279500$ sec in Fig. 7.7. After the shock the bid-ask spread closes. In our example this takes about 2 minutes. Complete relaxation of the spread/liquidity can be very slow, but the half-life of the decay is typically of the order $10^2 - 10^3$ seconds.

The price change corresponding to these large events is still much less than the width of the order book (with some very few exceptions) but they can represent a very substantial fraction of the total volume present at that time, because the orders very far from the best tend to be smaller. After the spread closes typically there is no complete reversion to the price levels before the event, and a new midquote price is formed [PLM06, ZAK06].

7.4 Conclusions

In this chapter, by using various representations of the order book, we have shown that the qualitative picture that describes the dynamics of trading strongly depends on the time scale. On a monthly scale price changes are well described qualitatively by an ordinary diffusion with a drift, and the fluctuations in supply and demand are not relevant. In contrast, for daily data price changes are very broadly distributed, and the volatility is clustered. This is mirrored by the large day-to-day fluctuations in the order book. The imbalance between the volume of buy and sell limit orders also varies strongly, but this variation is not reflected directly in the returns. The qualitative picture changes dramatically when one moves down to the resolution of individual transactions, where returns are governed by the price impact of individual events and liquidity fluctuations. Despite their potential importance in the execution of high-frequency trading strategies, at present no models can fully account for the rich structure of returns on very short time scales. There has been some recent success along these lines, but we believe that many questions still remain unanswered.

Chapter 8

Diffusive approximation for limit order executions

In contrast with the previous chapter, now we take a quantitative approach: We first demonstrate that some aspects of the order book dynamics are consistent with theoretical predictions obtained by assuming a random process with independent identically distributed and symmetrical increments. In particular, we focus on the difference observed between the time to fill a limit order, which is the time one has to wait before a limit order is executed, and the first passage time [Fel67], i.e., the time elapsed between an initial instant and the time when the transaction price crosses a given predefined threshold.

The chapter is organized as follows. In Section 8.1 we study the first passage time and in Section 8.2 the time to fill and the time to cancel. Section 8.3 describes a simple limit order model and in Section 8.4 we discuss the empirical testing of the model. Section 8.5 discusses the validity of one key assumption, namely random cancellation, whereas Section 8.6 extends the result to limit orders placed inside the spread. Finally Section 8.7 concludes.

We investigate 5 highly liquid stocks, AstraZeneca (AZN), Glaxo-SmithKline (GSK), Lloyds TSB Group (LLOY), Shell (SHEL), and Vodafone (VOD). We treat buy and sell orders together because we do not find a significant difference between their behavior, in contrast with Ref. [LMZ02] for US markets but in analogy with Ref. [HMS04] for the case of Ericsson stock traded at the Stockholm Stock Exchange. We therefore are not able to conclude whether the symmetric behavior we observe in the London Stock Exchange is common to most markets or specific to some of them or to certain time periods.

8.1 The first passage time

Let the price of an asset at time $t = 0$ be S_0 . The first passage time [Fel67] of price through a prescribed level $S_0 + \Delta$ for a fixed distance $\Delta > 0$ is defined as the first time t when $S(t) \geq S_0 + \Delta$. Then we repeat the same analysis for the level $S_0 - \Delta$. We will call the distribution of the quantity t the first passage time distribution to a distance Δ , and denote it by $\mathbb{P}_{\text{FPT};\Delta}(t)$.

Such first passage processes have been studied extensively [Red01]. For the following analysis of empirical data, it is useful to review the first passage time distribution for a Brownian motion. If the standard deviation for the price changes during a unit time (i.e., the volatility) is σ , then [Fel67]

$$\mathbb{P}_{\text{FPT};\Delta}(t) = \frac{\Delta}{\sqrt{2\pi\sigma^2}} t^{-3/2} \exp\left(-\frac{\Delta^2}{2\sigma^2 t}\right), \quad (8.1)$$

which is the fully asymmetric 1/2-stable distribution. For any fixed Δ the asymptotics for long times is

$$\mathbb{P}_{\text{FPT};\Delta}(t) \propto t^{-3/2}. \quad (8.2)$$

A recent study [CMG⁺03] has clarified that this asymptotic behavior is valid not only for Brownian motion but also for any Markov process with symmetric jump length distribution. This result is consistent with the Sparre-Andersen theorem [Red01]. Alternative descriptions obtained for the asymptotic time dependence of the FPT of Lévy flights which were hypothesizing a dependence of the distribution exponent from the index of the Lévy distribution have missed the fact that the method of the images, which is extremely powerful in Gaussian diffusion, fails for Lévy flight processes [CMG⁺03]. The behavior is of course more complex in the case of Lévy random processes described by using a subordination scheme. In these cases the first passage time asymptotic behavior depends on the complete properties of the subordination procedure [SM04].

Not only price changes are not described by continuous values, but transactions and order submissions are also separated by finite waiting times, which a continuous time random walk formalism could take into account [Sca05, MPM⁺05]. However, in this chapter we are mainly interested in relatively long time intervals as compared to these waiting times, so the discrete aspects of the dynamics are negligible. Thus, we will model prices as if they varied continuously in time.

Let us now investigate empirically the first passage time behavior. The first passage time distribution for the transaction price¹, bid and ask when

¹The reader is reminded that we only consider here transactions that take place through the order book, but not the SEAQ "upstairs" market.

$\Delta = 1$ tick is shown in Fig. 8.1 for the stock GSK. The distribution is obtained by sampling the first passage time at each second. One can see that there are no significant differences in the behavior of the three quantities. Qualitatively, the distribution is similar to Eq. (8.1), and the long time asymptotic of real data seems to decay approximately as $t^{-3/2}$. For times shorter than 1 minute the curves significantly deviate from the power law. We choose to fit the first passage time distribution with the function

$$\mathbb{P}_{\text{FPT};\Delta}(t) = \frac{Ct^{-\lambda_{\text{FPT}}}}{1 + [t/T_{\text{FPT}}(\Delta)]^{-\lambda_{\text{FPT}} + \lambda'_{\text{FPT}}}}. \quad (8.3)$$

This form, that we will use to fit also the other distributions introduced below, is characterized by two power law regimes. Normalization conditions of Eq. (8.3) imply that $\lambda_{\text{FPT}} > 1$ and $\lambda'_{\text{FPT}} < 1$. For $t \ll T_{\text{FPT}}(\Delta)$ it is $\mathbb{P}_{\text{FPT};\Delta}(t) \propto t^{-\lambda'_{\text{FPT}}}$, whereas for $t \gg T_{\text{FPT}}(\Delta)$ it is $\mathbb{P}_{\text{FPT};\Delta}(t) \propto t^{-\lambda_{\text{FPT}}}$.

Table 8.1 contains the fitted parameters λ_{FPT} , λ'_{FPT} , and $T_{\text{FPT}}(\Delta)$ for $\Delta = 1, \dots, 4$ ticks. The difference between the actual values of λ_{FPT} and $3/2$ known for Brownian motion or for any Markov process with symmetric jump length distribution is small. Other values would be possible due to clustered volatility but we find no evidence for a systematic deviation.

Given the level of noise in the sample, these arguments are not meant as conclusive evidence in favor of the form proposed in Eq. (8.3). For many of our arguments it will be enough that for large times \mathbb{P}_{FPT} is well approximated by a power law. According to Fig. 8.1, for $\Delta = 1$ tick we are in this regime when t is at least a few minutes. Finally, the inset of Fig. 8.1 shows the median first passage time as a function of Δ for the five investigated stocks. The behavior is not exactly quadratic (Δ^2) as one would expect from Eq. (8.1). If prices followed a Brownian motion, the q -th quantile (T_q) of the first passage time distribution would be

$$T_q = \frac{\Delta^2}{2\sigma^2[\text{erfc}^{-1}(q)]^2}, \quad (8.4)$$

where the median ($M[\text{FPT}]$) corresponds to $q = 0.5$. Instead, the power law behavior with Δ is not always evident in the inset of Fig. 8.1 and assuming a form $M[\text{FPT}] \propto \Delta^\eta$ would require an exponent varying between 1.5 and 1.8 depending on the specific stock and the precise Δ interval used for the estimation of η . A similar deviation from the prediction for Brownian motion was reported in Ref. [SJJ02] in the analysis of closure index values sampled at a daily time horizon.

We have also verified the role of the inclusion of SEAQ transactions in the determination of the η exponent. By considering time series including the SEAQ transactions, the exponent η assumes values within the range 1.7 – 1.9. This discrepancy might be due to deviation from the Gaussian

stock	$\Delta = 1$			$\Delta = 2$		
	λ	λ'	T	λ	λ'	T
AZN	1.50	0.14	58	1.50	0.22	140
GSK	1.52	0.16	62	1.52	0.18	230
LLOY	1.54	0.22	85	1.55	0.20	280
SHEL	1.52	0.20	83	1.53	0.27	160
VOD	1.57	0.43	150	1.54	-0.19	450

stock	$\Delta = 3$			$\Delta = 4$		
	λ	λ'	T	λ	λ'	T
AZN	1.50	0.18	240	1.49	0.11	350
GSK	1.50	-0.02	390	1.48	-0.21	520
LLOY	1.53	0.01	460	1.51	-0.12	630
SHEL	1.51	0.02	360	1.51	0.00	450
VOD	1.49	-0.69	720	1.51	-0.66	1500

Table 8.1: Parameters of the fitting function (8.3) for the distribution of first passage time for the five stocks. Δ is measured in ticks and all times are given in seconds. Typical standard errors for the quantities: ± 0.05 for λ_{FPT} , ± 0.05 for λ'_{FPT} , and $\pm 10\%$ for T_{FPT} .

assumption, to a superdiffusive behavior of price or to both. We have performed a series of shuffling experiments and preliminary results support the conclusion that the main role is played by the deviation from Gaussianity. This non-Gaussianity is well documented in the literature down to the scale of single transactions [BP00, FGL⁺04]. Moreover, the scaling exponent of the first passage time of Lévy flights with the variable Δ can be expected to be different from 2 and related to the index of the Lévy distribution [SW82]. At the present stage these observations and theoretical interpretations are preliminary and further investigations are needed to clarify this discrepancy.

8.2 Time to fill, time to cancel

As noted in Chapter 6, the life of every limit order ends with either execution or deletion. In this latter case deletion can be due to intentional cancellation or to expiration, but we will not distinguish between these mechanisms and we will call both of them cancellation. For an executed order the elapsed time between its placement and its complete execution is called *time to fill*. Orders are often not executed in a single transaction, thus one can independently define *time to first fill*, which is the time from order placement to the first transaction this order participates in. Finally, for canceled orders one can define the *time to cancel* which is the

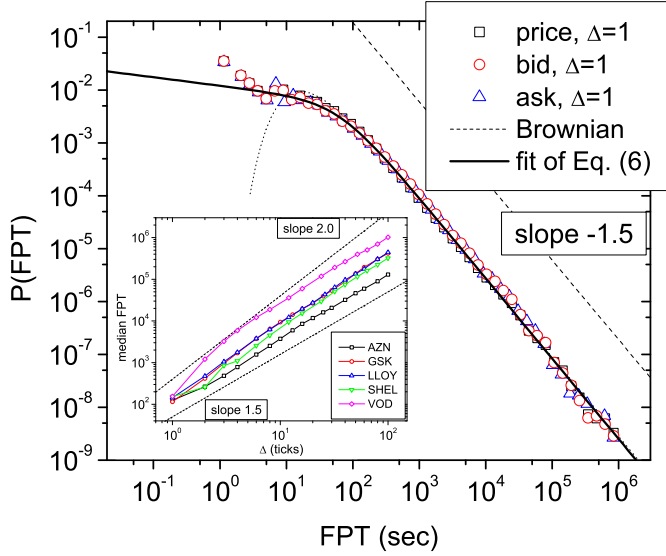


Figure 8.1: First passage time distributions for the price, bid and ask quotes of GlaxoSmithKline (GSK), distance $\Delta = 1$ tick. The dotted line is the first passage time distribution for Brownian motion with volatility $\sigma = 1/7$ penny \times sec $^{-1/2}$. The thick solid line is a fit with Eq. (8.3) as given in Table 8.1. The inset shows the median first passage time as a function of Δ .

time between order placement and cancellation. The distribution of these three quantities will be in the following denoted by $\mathbb{P}_{\text{TTF}}(t)$, $\mathbb{P}_{\text{TTF}}(t)$, and $\mathbb{P}_{\text{TTC}}(t)$, respectively.

By the nature of trading, limit orders with small Δ , especially ones with large negative values, are executed earlier. In most of the chapter we will only concern orders where $\Delta > 0$. Then, Section 8.6 will deal with the generalization to $\Delta \leq 0$.

8.2.1 Distributional properties

As a first characteristic of the order book, we investigate the distribution of time to fill and time to cancel for the stocks in our dataset. Fig. 8.2 shows these distributions for the stock GlaxoSmithKline (GSK) for different values of Δ . Similarly to first passage time, we fitted the empirical

density with the functional form

$$\mathbb{P}_{\text{TF};\Delta}(t) = \frac{C't^{-\lambda_{\text{TF}}}}{1 + [t/T_{\text{TF}}(\Delta)]^{-\lambda_{\text{TF}}+\lambda'_{\text{TF}}}}. \quad (8.5)$$

Table 8.2 shows the results for all five stocks. We found that λ_{TF} , which gives the asymptotic behavior of the distribution, ranges between 1.8 and 2.2 for up to $\Delta = 4$ ticks. This finding is in contrast with the value 1.5 obtained in Ref. [CS01] for the limit order data of a pool of stocks traded at NASDAQ. We have no explanation for this discrepancy. The exponent λ'_{TF} varies between -0.4 and 0.6 . Finally T_{TF} typically grows with larger Δ , orders placed deeper into the book are executed later. We will return to this observation in Section 8.2.3. For higher values of Δ the small number of limit orders in our sample does not allow us to make reliable histograms or give robust estimates for the shape of the distribution. Fig. 8.2 also gives a comparison of four further stocks (AZN, LLOY, SHEL and VOD) to show that our findings are quite general. The distribution of time to first fill is indistinguishable from time to fill [upo].

For time to cancel one finds a similarly robust behavior, also shown in Fig. 8.2. The respective distribution again decays as a power law and it is well fitted by the form

$$\mathbb{P}_{\text{TC};\Delta}(t) = \frac{C''t^{-\lambda_{\text{TC}}}}{1 + [t/T_{\text{TC}}(\Delta)]^{-\lambda_{\text{TC}}+\lambda'_{\text{TC}}}}, \quad (8.6)$$

where the long time asymptotic has an exponent λ_{TC} ranging between 1.9 and 2.4. Differently from the case of λ_{TF} previously discussed, the measured values of λ_{TC} are in agreement with those measured in Ref. [CS01] for the limit order data of a pool of stocks traded at NASDAQ. All results concerning the time to cancel are given in Table 8.3.

8.2.2 Comparison of characteristic times

Our emphasis is on the interplay between order execution, order cancellation and the first passage properties of price. To understand this relationship, consider the following argument proposed in Ref. [LMZ02]. Imagine that there are no cancellations. Let a buy order be placed at the price $b_0 - \Delta$, when the current best bid is at b_0 . How much time does it take until this order is executed? It is certain that the order cannot be executed before the best bid decreases to $b_0 - \Delta$, because until then there will always be more favorable offers in the book. On the other hand, once the price decreases to $b_0 - \Delta - \epsilon$ where ϵ is the tick size of the stock, it is certain, that all possible offers at the price $b_0 - \Delta$ have been exhausted, including ours. Therefore both time to fill and time to first fill for any order (a similar argument can be given for sell orders) placed at a distance Δ from the best offer is greater than the first passage time of price

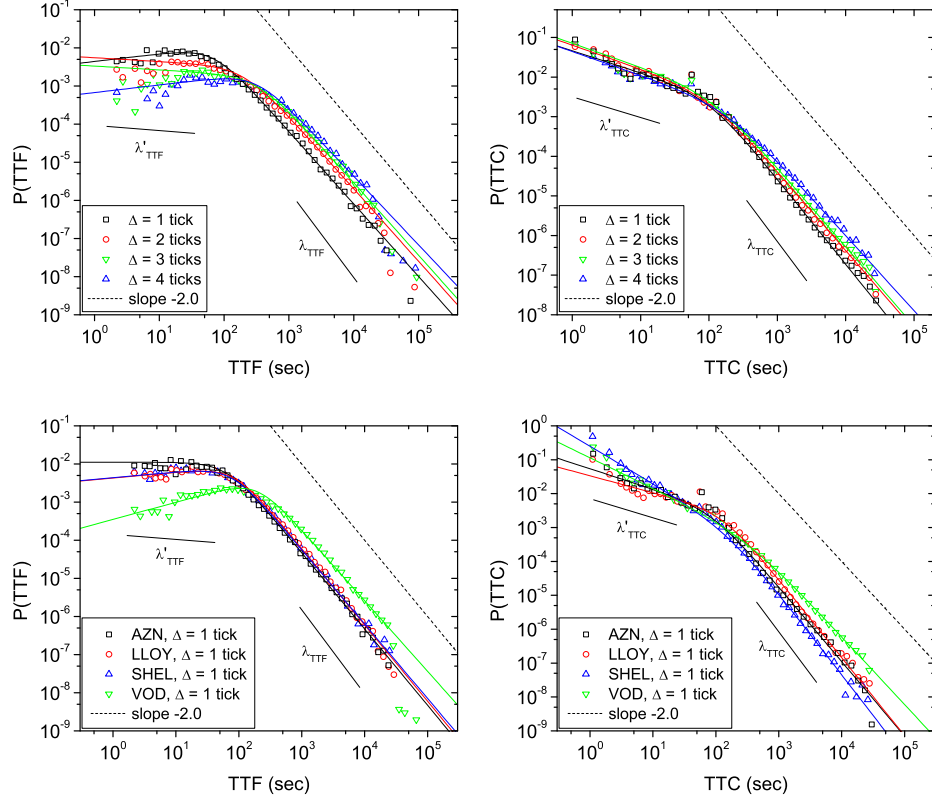


Figure 8.2: **(top left)** Distribution of time to fill of GSK for $\Delta = 1 \dots 4$ ticks, and fits with Eq. (8.5). The dashed line is a power law with exponent -2.0 . **(top right)** Distribution of time to cancel of GSK for $\Delta = 1 \dots 4$ ticks, and fits with Eq. (8.6). The dashed line is a power law with exponent -2.0 . **(bottom left)** Comparison of distributions of time to fill for three typical stocks, $\Delta = 1$ tick, and fits with Eq. (8.5). The dashed line is a power law with exponent -2.0 . **(bottom right)** Comparison of distributions of time to cancel for three typical stocks, $\Delta = 1$ tick, and fits with Eq. (8.6). The dashed line is a power law with exponent -2.0 .

to a distance Δ , and less than that to $\Delta + \epsilon$. Since this is true for every individual order, one expects the following inequality for the distribution functions of characteristic times:

$$\int_0^t \mathbb{P}_{\text{FPT};\Delta}(t') dt' \geq \int_0^t \mathbb{P}_{\text{TTF};\Delta}(t') dt' \geq \int_0^t \mathbb{P}_{\text{TTF};\Delta}(t') dt' \geq \int_0^t \mathbb{P}_{\text{FPT};\Delta+\epsilon}(t') dt'. \quad (8.7)$$

stock	$\Delta = 0$			$\Delta = 1$			$\Delta = 2$		
	λ	λ'	T	λ	λ'	T	λ	λ'	T
AZN	2.2	0.6	110	2.0	-0.0	65	1.9	0.0	100
GSK	2.2	0.5	110	1.9	-0.2	68	1.9	-0.2	150
LLOY	2.2	0.5	120	2.0	-0.1	85	1.9	-0.1	160
SHEL	2.2	0.4	110	1.9	-0.1	77	1.9	-0.2	110
VOD	2.1	0.5	160	1.8	-0.4	190	1.8	-0.5	490

stock	$\Delta = 3$			$\Delta = 4$		
	λ	λ'	T	λ	λ'	T
AZN	1.8	-0.0	120	1.9	0.0	200
GSK	1.8	-0.4	190	1.8	-0.3	320
LLOY	1.9	-0.2	240	1.9	-0.2	350
SHEL	1.9	0.0	270	1.8	-0.1	250
VOD	1.8	-0.4	980	-	-	-

Table 8.2: Parameters of the fitting function (8.5) for the distribution of time to fill for the five stocks and $\Delta \geq 0$ ticks. All times are given in seconds. Data are missing where the statistics was inadequate for fitting. Typical standard errors for the quantities: ± 0.1 for λ_{TTF} , ± 0.1 for λ'_{TTF} , and $\pm 10\%$ for T_{TTF} .

Using the empirical distributions above, a straightforward calculation yields

$$\lambda_{\text{FPT}} = \lambda_{\text{TTF}} = \lambda_{\text{TTF}} \quad (\text{wrong!}), \quad (8.8)$$

which is in clear disagreement with the data, where pronouncedly $\lambda_{\text{FPT}} < \lambda_{\text{TTF}} \approx \lambda_{\text{TTF}}$. This inequality for the tail exponents means that one finds less orders with very long time to (first) fill than expected. The solution of this apparent contradiction is the presence of cancellations: Orders which would have a too long time to wait until being executed are often canceled and thus removed from the statistic. The problem is that this argument is only qualitative, and it tells us nothing about the behavior of cancellation times. Actually, the measurement of the cancellation time distribution suffers from the same bias. The observed distribution of time to cancel does not characterize how traders actually cancel their orders, because from that statistic it is the executed orders that are missing. In fact, one can infer very little directly from \mathbb{P}_{TC} about how orders are canceled.

In Section 8.3 we will present a simple model that gives insight into the features pointed out so far. However, before doing so, we would like to present one further point concerning the empirical data.

stock	$\Delta = 0$			$\Delta = 1$			$\Delta = 2$		
	λ	λ'	T	λ	λ'	T	λ	λ'	T
AZN	2.3	0.6	130	2.2	0.6	87	2.2	0.6	90
GSK	2.1	0.6	130	2.2	0.5	110	2.0	0.5	90
LLOY	2.2	0.5	120	2.3	0.5	130	2.2	0.4	140
SHEL	2.3	1.0	220	2.4	1.1	150	2.3	1.1	140
VOD	2.0	0.7	200	2.0	0.9	300	2.1	0.8	1000

stock	$\Delta = 3$			$\Delta = 4$		
	λ	λ'	T	λ	λ'	T
AZN	2.2	0.6	85	2.2	0.7	100
GSK	1.9	0.5	94	1.9	0.6	170
LLOY	2.0	0.4	120	2.0	0.5	250
SHEL	2.3	1.1	68	2.2	1.0	56
VOD	2.2	0.6	1500	1.9	0.5	1000

Table 8.3: Parameters of the fitting function (8.6) for the distribution of time to cancel for the five stocks and $\Delta \geq 0$ ticks. All times are given in seconds. Data are missing when there were no orders at all, or the statistics was inadequate for fitting. Typical standard errors for the quantities: ± 0.1 for λ_{TTC} , ± 0.1 for λ'_{TTC} , and $\pm 25\%$ for T_{TTC} .

8.2.3 The role of entry depth

How do order execution times change as a function of the entry depth Δ ? Before answering the question, let us remark on a caveat: The distributions found above imply that the mean of time to fill/cancel might diverge, because the distributions decay very slowly. Therefore, in the following we will use the medians of all quantities as a measure of their typical value.

In Fig. 8.3 we show that the median of time to fill is empirically well described by

$$M[\text{TTF}] \propto \Delta^{1.4}, \quad (8.9)$$

a behavior which is quite different from $M[\text{TTF}] \propto \Delta^2$, which would be expected naively from Eq. (8.4) and a Brownian motion assumption. An explanation for this discrepancy can be broken down into steps based on Fig. 8.3(left), which summarizes the Δ -dependence of several quantities. In the following we will use the stock GSK as a representative example, but qualitatively the same arguments hold for other stocks as well.

Let us first make a surrogate experiment for GSK, which is similar to the calculation of time to fill. We select all filled orders, and from the time of their placement we calculate the first time when the transaction price becomes equal to or better than the price of the order. If one plots the median of this quantity versus the Δ of the orders, the re-

sulting curve is indistinguishable from the median of time to fill [curve labeled as "TTF/FPT filled ord."]. Because of this, we can also think of time to fill as a first passage time measurement for executed orders only. For these time to fill surrogates multiple sources can contribute to their non-diffusive behavior. There are several statistical biases: due to the exclusion of canceled orders (A), because Δ is correlated with volatility (B), and also because the number of incoming limit orders is correlated with volatility (C). The exclusion of the "upstairs" market transactions may also play a role (D).

First of all, for any Δ clearly $M[\text{TTF}]$ is smaller than $M[\text{FPT}]$ for the price, which supports our above statements concerning bias (A) due to cancellations [the two curves we refer to are labeled as "TTF/FPT filled ord." and "FPT, price (book only)"]. So how do we eliminate these biases?

We can make another surrogate in a similar way as before, but including all orders, not only the filled ones to eliminate bias (A). Moreover, to destroy correlations between volatility and order placement, we shuffle the Δ values between orders. We end up with the following dataset: Every time an order is placed, for buy orders we introduce a marker at the current best bid minus Δ (or at best ask plus Δ for sell orders), where due to the shuffling Δ is drawn randomly from its empirical distribution. Then we measure the time it takes for price to reach the marker (this time is labeled as "FPT shuff. all ord"). This new curve now agrees with the first passage time of price with upstairs trading excluded, but still only when $\Delta > 8$ ticks, which corresponds to a median time of about 1 – 2 hours [labeled as "FPT shuff. all ord." and to be compared with "FPT, price (book only)"]. The reason is simple: We are not sampling the first passage process uniformly in time, which refers to bias (C). We place a marker every time an order was placed, so periods with many incoming limit orders are oversampled. If we change to sampling every second, we recover the first passage time for order book transactions [line labeled as "FPT, price (book only)"].

We can also include upstairs trades (D) [line labeled as "FPT, price (upst. incl.)"]. The first passage time of this price now follows the approximate relation $M[\text{FPT}] \propto \Delta^{1.9}$ for GSK. The exponents for the other four stocks vary between 1.7 and 1.9, which are closer to, but consistently less than 2 expected for Brownian motion. As already noticed, this behavior is similar to the one detected in Ref. [SJJ02] in the investigation of the daily closures of a detrended version of the Dow Jones Industrial Average.

In summary, the most significant bias is (A), the presence of canceled orders. Once this is accounted for, the other curves have at least qualitatively similar behavior, and for $\Delta > 20$ they become indistinguishable. This price change corresponds to a median FPT of about one trading day.

The dependence of time to cancel on the entry depth Δ has a less clear functional form, as shown by Fig. 8.4. While $M[\text{TTC}]$ appears to be a

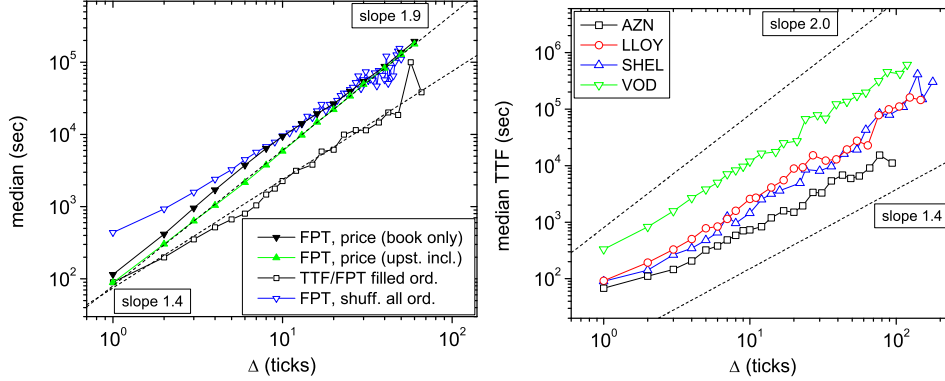


Figure 8.3: **(left)** The dependence of characteristic times on the entry depth Δ for GSK. Empty squares show time to fill, the curve for the surrogate first passage time assigned to filled orders is indistinguishable from this one. The dotted lines are $M[\cdot] \propto \Delta^{1.4}$ and $M[\cdot] \propto \Delta^{1.9}$. **(right)** The dependence of median time to fill on the entry depth Δ for four stocks. The curves show a strong similarity and they are significantly better fitted by $M[\text{TTF}] \propto \Delta^{1.4}$ than by the naive ansatz $M[\text{TTF}] \propto \Delta^2$ from Brownian motion.

monotonically increasing function of Δ , the curves for the different stocks show only a qualitative similarity. The reason is probably the non-trivial dependence of order lifetimes on Δ . Here we have no *a priori* expectation such as the quadratic form for the first passage times.

8.3 A simple model

In this section we present a simple model of limit order placement and cancellation.² We will see that the model gives falsifiable predictions that can be tested against real data. Moreover, it also gives indications on the statistical properties of a quantity that is directly unobservable: the lifetime an agent gives to a limit order. We do not suggest a detailed microstructure of any kind, but rather the following assumptions:

1. We consider one "agent". At time $t = 0$ the agent places a single buy³ limit order at a $\Delta > 0$ distance from the current best offer. We treat all the other agents on an aggregate level, which we will call

²Ref. [WBK⁺06] shows that similar arguments give a very good approximation for the average shape of the order book.

³Note that we use the language of buy orders, but analogous definitions can be given for sell orders. All measurements include both buy and sell orders.

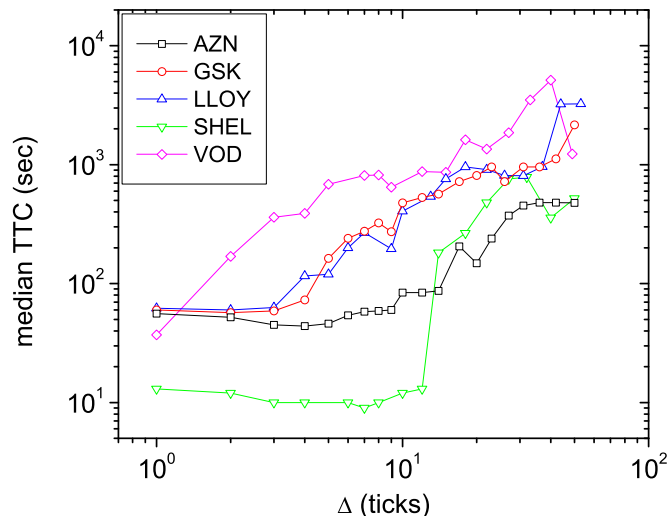


Figure 8.4: The dependence of median time to cancel on the entry depth Δ of the limit order. The curves have an increasing tendency and they are qualitatively similar across stocks. However, they do not follow any obvious functional form.

”the market”. The spirit of this assumption is similar to a ”mean field” approximation in physics.

2. The agent is not willing to wait indefinitely for the order to be executed. Instead, at the time of placement the agent also decides about a cancellation (or more appropriately expiration) time t' for the order. This is a value drawn randomly from the distribution $P_{LT;\Delta}(t')$. We will call this function the *lifetime distribution*. If the order is not executed until t' , then the order is canceled. The agent has no additional cancellation strategy. This assumption makes cancellations effectively independent from price changes.
3. The market is very liquid and tick sizes are small. As a consequence,
 - (a) before its execution, the effect of the agent’s limit order on the evolution of the market price is negligible.
 - (b) the time between when the best bid reaches the order price and when the agent’s order is executed is negligible.
 - (c) such immediate execution happens regardless of the volume of the order

These assumptions seem very restrictive, hence the next part of this section is devoted to motivating them (we keep the above scheme of numbering).

1. Our agent should be considered as a typical investor, or "representative agent" [Kir92]. In reality, the typical investor does not exist, as trading strategies are very heterogeneous. Mixing together a range of strategies often leads to the recognition that the final average no more resembles any individual market participant [HK95]. Thus when we analyze all limit orders regardless of who placed them (because such information is not available), our conclusions will also only apply on this aggregate level. As for the restriction $\Delta > 0$, for simplicity we only consider here such orders, and return to the $\Delta \leq 0$ case in Section 8.6.

2. The cancellation process is an essential part of the model. Ref. [MF05] finds that the per tick probability of an order being canceled is of the form

$$K_1 \left[1 - \exp \left(-\frac{\Delta(t)}{\Delta(0)} \right) \right], \quad (8.10)$$

where $\Delta(t)$ is the difference between the limit order price and the price of the opposite best at time t . $\Delta(0)$ is this difference when the limit order was placed, so that $\Delta(0)$ is equal to Δ plus the spread at time $t = 0$. It is a very crude assumption to say instead that traders decide about the expiry time of the limit orders when they place them, and do not cancel them otherwise. Nevertheless, this will not affect the quantities we are interested in, see Section 8.5 for more detail.

3. High liquidity and small relative tick sizes (minimal price change divided by stock price) do not hold for every stock. It has been pointed out by several studies that small liquidity [EK06b, EK07b] or too large tick sizes [MF05, WBK⁺06] can have a great impact on the microstructural properties of prices and limit order books. This is the reason why we included in our study SHEL and VOD which are known to have large tick/price ratios. Contrary to our expectations, we did not find any indication of anomalies, and the model proved valid for all five stocks.

(a) Before its execution, a limit order by definition has no direct impact on price formation. There is an indirect effect however. The appearance of a new, large limit order introduces significant new information about the state of the market that was previously hidden from traders. They may become more cautious, or try to take advantage of such disclosed demand/supply, even if it is not at the current best price. This leads to a hide-and-seek game of traders. Informed traders are trying to conceal their information (and thus their excess demand/supply) to decrease their losses due to the price impact of

their trades. Meanwhile, others are trying to infer this knowledge to increase their own profit from providing them liquidity. This is the main motivation behind strategic order splitting. While we fully acknowledge their existence, our model cannot take such effects into account [Bou, BDW93, Cha01].

- (b) When a transaction is set at a given price, it does not necessarily mean that all limit orders at that price are executed immediately. Moreover, even that limit order may possibly be only partially executed. A simple way to motivate that the volume present at a given price does not strongly affect execution times is to measure the typical ratio between time to fill and time to first fill as a function of the volume of the order. Fig. 8.5 shows that for at least 75% of the orders of any volume this is close to 1. The only exceptions can be very large orders with $\Delta = 1$. Here the price reaches the order quickly, but it takes about 20% longer to execute it completely. Our model neglects these effects on very large volumes (see also Ref. [LMZ02]).
- (c) The typical time to fill does not depend too strongly on the volume of the order, except for very large orders. Fig. 8.6 shows the median time to fill (left panel) and time to cancel as a function of the volume of the order. These median values are not correlated with volume except for very large orders which have a larger median time to fill. Our model neglects this behavior.

Under our assumptions one can write a joint density function that describes both the price diffusion process and cancellations. The probability $\mathbb{P}_\Delta(t, t')$ that the price reaches an order placed at a distance $\Delta > 0$ from the current best offer at time t (and it is then executed immediately), and that the agent decides to cancel the order a time t' can be written as a product of two independent distributions:

$$\mathbb{P}_\Delta(t, t') = \mathbb{P}_{\text{FPT};\Delta}(t)\mathbb{P}_{\text{LT};\Delta}(t'). \quad (8.11)$$

$\mathbb{P}_\Delta(t, t')$ is the joint probability that a limit order placed at a distance Δ with lifetime t' describes the following state for the limit order at time t : (i) the limit order is executed if $t < t'$ or (ii) it is canceled if $t > t'$. The two states are illustrated in detail in Fig. 8.7.

8.4 The predictions of the model

Let us denote distribution functions in the following way:

$$\mathbb{P}_{X;\Delta}(> t) = \int_t^\infty \mathbb{P}_{X;\Delta}(\tau) d\tau, \quad (8.12)$$

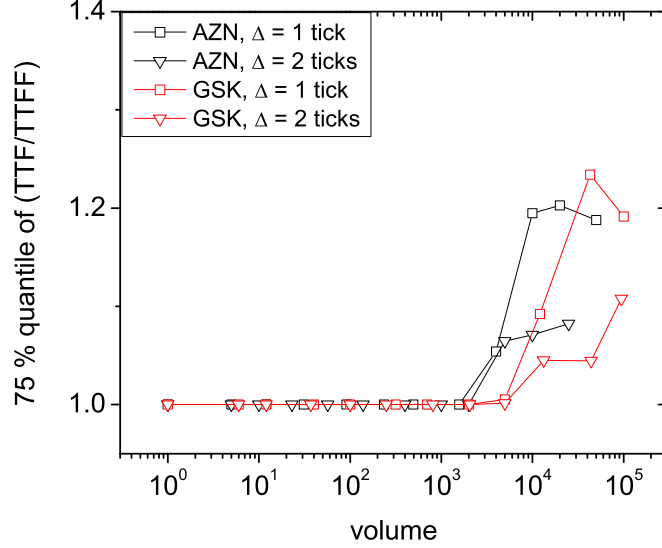


Figure 8.5: The 75% quantile of the ratio of time to fill and time to first fill for AZN and GSK. The ratio is around 1.1 – 1.2 for the largest orders and $\Delta = 1$ tick, but it decreases when $\Delta = 2$ ticks. These results show that except for very small Δ and very large orders, at least 75% of the orders time to fill and time to first fill is essentially equal.

where X can be any process introduced above (FPT, LT, TTF, TTF, TTC). We omit the lower index Δ for brevity. Let us first express the previously introduced quantities in terms of the joint probability $\mathbb{P}_\Delta(t, t')$ and via Eq. (8.11). For executed orders $t < t'$, thus the distribution of time to fill is given by

$$\mathbb{P}_{\text{TTF}}(t) = \frac{\mathbb{P}_{\text{FPT}}(t)\mathbb{P}_{\text{LT}}(> t)}{\int_0^\infty \mathbb{P}_{\text{FPT}}(\tau)\mathbb{P}_{\text{LT}}(> \tau)d\tau} = \mathcal{N}[\mathbb{P}_{\text{FPT}}(t)\mathbb{P}_{\text{LT}}(> t)]. \quad (8.13)$$

We introduced the operator $\mathcal{N}[\cdot]$, which normalizes a function to an integral of 1. Symmetrically for time to cancel $t < t'$:

$$\mathbb{P}_{\text{TTC}}(t) = \frac{\mathbb{P}_{\text{FPT}}(> t)\mathbb{P}_{\text{LT}}(t)}{\int_0^\infty \mathbb{P}_{\text{FPT}}(> \tau)\mathbb{P}_{\text{LT}}(\tau)d\tau} = \mathcal{N}[\mathbb{P}_{\text{FPT}}(> t)\mathbb{P}_{\text{LT}}(t)]. \quad (8.14)$$

As (8.13) and (8.14) are two equations with only one unknown function, namely the lifetime distribution $\mathbb{P}_{\text{LT}}(t)$, one can calculate that from, e.g., Eq. (8.13), and then see if the solution is consistent with Eq. (8.14). We can express from Eq. (8.13), that

$$\mathbb{P}_{\text{LT}}(> t) \propto \frac{\mathbb{P}_{\text{TTF}}(t)}{\mathbb{P}_{\text{FPT}}(t)} \quad (8.15)$$

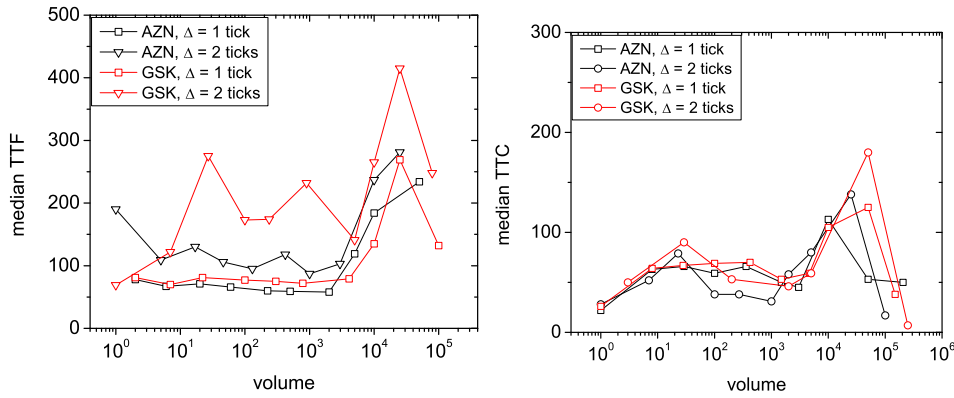


Figure 8.6: **(left)** Median time to fill for AZN and GSK, $\Delta = 1$ and 2 ticks, as a function of order volume. The value does not depend strongly on order size except for very large orders. **(right)** Median time to cancel for AZN and GSK, $\Delta = 1$ and 2 ticks, as a function of order volume. The value does not depend strongly on order size.

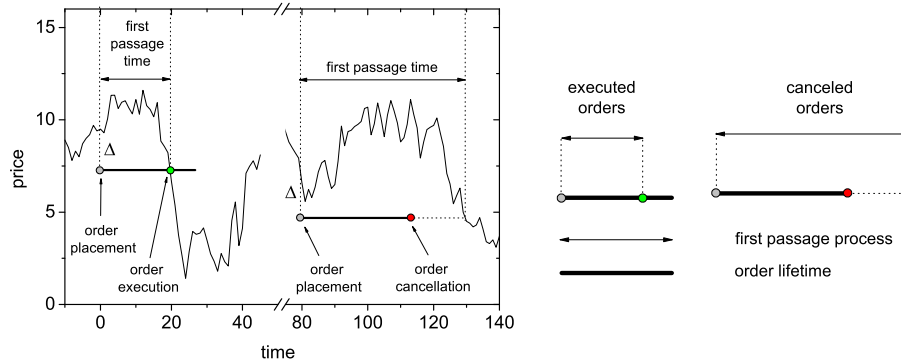


Figure 8.7: **(left)** The scheme of the diffusive model for execution times (example with buy orders). Orders are indicated by thick horizontal lines. The order is placed at a distance Δ below the current best bid. At the time of its placement the order is assigned a lifetime (the length of the thick line). If the bid crosses the line, then the order is executed at the time of crossing. The time between order placement and the crossing is the first passage time of the bid price to a distance Δ . If there is no crossing, the order is canceled at its cancellation time (the end of the thick line). **(right)** Orders are executed when the first passage time is less, and canceled when larger than the intended lifetime.

and thus

$$\mathbb{P}_{\text{LT}}(t) = -\frac{d}{dt}\mathbb{P}_{\text{LT}}(> t) = -\mathcal{N}\left[\frac{d\mathbb{P}_{\text{TTF}}(t)}{dt\mathbb{P}_{\text{FPT}}(t)}\right]. \quad (8.16)$$

It is also possible to estimate the same quantity directly from Eq. (8.14):

$$\mathbb{P}_{\text{LT}}(t) = \mathcal{N}\left[\frac{\mathbb{P}_{\text{TTC}}(t)}{\mathbb{P}_{\text{FPT}}(> t)}\right]. \quad (8.17)$$

Let us eliminate the lifetime distribution, and substitute the large t asymptotic power law behavior of all probabilities. After simple calculations one finds that

$$\lambda_{\text{TTF}} = \lambda_{\text{TTC}}. \quad (8.18)$$

Then we substitute this result back into Eq. (8.15) to find that the lifetime distribution also has to decay asymptotically as a power law:

$$\mathbb{P}_{\text{LT}}(t) \propto t^{-\lambda_{\text{LT}}}, \quad (8.19)$$

with

$$\lambda_{\text{LT}} = \lambda_{\text{TTF}} - \lambda_{\text{FPT}} + 1 = \lambda_{\text{TTC}} - \lambda_{\text{FPT}} + 1. \quad (8.20)$$

Eq. (8.18) is in good agreement with the results of Section 8.2, where $\lambda_{\text{TTF}} = 1.8 - 2.2$, and $\lambda_{\text{TTC}} = 1.9 - 2.4$. This is a clear improvement as compared to Eq. (8.8). The introduction of the simplest possible cancellation model gives a good prediction for the discrepancy between the asymptotic of first passage time and time to fill.

Moreover, one can now observe the hidden distribution of lifetimes. By substituting the typical values into Eq. (8.20), one gets $\lambda_{\text{LT}} \approx 1.6$. In comparison, a paper by Borland and Bouchaud [BB05] describes a GARCH-like model obtained by introducing a distribution of traders' investment horizons and the model reproduces empirical values of volatility correlations for $\lambda_{\text{LT}} = 1.15$, which is not far from our estimate. More recently it has been shown [Lil06] that the limit order price probability distribution is consistent with the solution of an utility maximization problem in which the limit order lifetime is power law distributed with an exponent $\lambda_{\text{LT}} \simeq 1.75$. The origin of the power law distribution of limit order lifetimes is not clear. Unfortunately the data do not allow us to separate individual traders. Therefore we do not know whether such a result arises from the broad distribution of the time horizons of each trader, or simply a distribution of traders with different investment strategies. Based on an empirical investigation at the broker level, in Ref. [Lil06] it is argued that heterogeneity of investors could be the determinant of the power law lifetime distribution. Notice, however, two points: (i) We are not speaking about how long the investors *hold* the stock. Instead, \mathbb{P}_{LT} is the distribution of *how long investors are willing to wait* for their limit orders to be executed and before they cancel or revise their offers. (ii)

None of the limit orders we are discussing here are truly long-term. Even the orders with relatively long lifetime spend at most a few days in the book.

8.5 The validity of the lifetime process

In Section 8.3 we motivated most assumptions of our model. However, one unrealistic assumption was the independence of the lifetime of the order from the evolution of price. This would mean that traders decide about an expiry time of their limit orders at the time of their placement, and then do not cancel them earlier, or for example that cancellation is simply random as in Ref. [FPZ05]. The natural procedure to validate this assumption would be to calculate the cross-correlation coefficient of first passage times and the lifetime process and show that this value is negligible. However, once again we are limited by the fact that the lifetime is hidden. It is not possible to calculate cross-correlations between time to fill and time to cancel either, because an order can only be either filled or canceled, thus the two variables cannot be observed simultaneously for the same order.

One solution to this puzzle is the following. One can observe the realized values of the lifetime for those orders that were actually canceled. Moreover, for the same orders it is possible to observe the value of the first passage time simply because price data is always available, even after the cancellation of the order. So under our model assumption, for all canceled orders we know the cancellation time and also the time when the order *would have been* executed. It is possible to quantify cross-correlations between these two quantities, but one has to keep in mind three points. (Note that we will consider orders with $\Delta = 1$ to have the largest possible sample.)

1. For very short times the price dynamics is dominated by bid-ask bounce, and other non-diffusive processes [Rol84]. Our model is not valid in this regime, because order executions are not governed by a first passage process. Hence we discard all orders which were canceled within $L = 4$ minutes of their placement.
2. At least for small Δ , both first passage times and the lifetime distributions may have diverging means. Therefore the cross-correlation coefficient is not always well-defined. Instead, we choose to evaluate Spearman's rank-correlation coefficient (ρ), which has favorable statistical properties, for example it is not very sensitive to the extreme events. Moreover, it can be defined for any two random variables regardless of the existence of any of their moments.
3. As we can only consider canceled orders, we know that $\text{FPT} > \text{LT}$. This constraint alone, and regardless of the choice of correlation mea-

sure, will cause strong positive correlations between the two quantities. The conditional joint distribution reads

$$\begin{aligned} \mathbb{P}(\text{FPT} = t, \text{LT} = t' | \text{FPT} > \text{LT}) = \\ \mathcal{N}[\Theta(t - t')\mathbb{P}_{\text{FPT}}(t)\mathbb{P}_{\text{LT}}(t')], \end{aligned} \quad (8.21)$$

where Θ is the Heaviside step function. Even if FPT and LT are independent, due to our restricted observations this is clearly not a product of two independent densities.

Instead, a proper null hypothesis is to measure the correlations between FPT/LT and LT. We know that for $\Delta = 1$ and $L = 4$ min the distribution of the first passage time is well described by the power law

$$\mathbb{P}_{\text{FPT}}(t | t > L) \sim \frac{\lambda_{\text{FPT}} - 1}{L^{\lambda_{\text{FPT}} - 1}} t^{-\lambda_{\text{FPT}}}. \quad (8.22)$$

If FPT and LT are independent, then

$$\begin{aligned} \mathbb{P}(\text{FPT}/\text{LT} = x, \text{LT} = t' | \text{FPT} > \text{LT}) = \\ \mathcal{N}[\Theta(x - 1)\mathbb{P}_{\text{FPT}}(xt')\mathbb{P}_{\text{LT}}(t')] = \\ \mathcal{N}[\Theta(x - 1)x^{-\lambda_{\text{FPT}}}] \times \mathcal{N}[\mathbb{P}_{\text{FPT}}(t')\mathbb{P}_{\text{LT}}(t')]. \end{aligned} \quad (8.23)$$

This is a product form in functions of x and of t' , which means that FPT/LT is independent from LT, given that we restrict ourselves to $\text{FPT} > \text{LT}$. Remember that the only assumption for this result is that first passage times are asymptotically power law distributed, which seems to hold very well both in our data down to $L \approx 4$ min, and any reference process described in Section 8.1.

As a next step, we calculated Spearman's rank correlations ρ_{res} between FPT/LT and LT in our restricted sample for various stocks. Results are summarized in Table 8.4. At all traditional significance levels one finds negative correlation between the two quantities.⁴ This means that those limit orders that would have been executed later were canceled earlier, i.e., that traders update their decision on when to cancel a limit order by tracking the price path. This is in line with the results of Ref. [MF05]. To prove that this value of ρ truly comes from correlations, we generated surrogate datasets by randomizing the pairs FPT/LT and LT while keeping the constraint $\text{FPT} > \text{LT}$. According to Table 8.4 this completely destroys the correlations between FPT/LT and LT, $\rho_{\text{surr}} = 0$.

It is a key point to remember that this value of ρ_{res} is not the actual correlation coefficient between the first passage time and the lifetime process.

⁴The error bars were estimated by the bootstrapping procedure suggested in Ref. [SS06] (for more details see refs. therein). For the application of bootstrapping to test the significance of correlations in financial data, see e.g. Ref. [TCL⁺07].

To quantify the true value of cross-correlations, we introduce ρ_{true} which is Spearman's rank-correlation coefficient between LT and FPT. While this cannot be measured directly, there is a procedure to estimate it from a known value of ρ_{res} based on Monte Carlo simulation. Let us assume that FPT and LT are adequately described by power law distributions with the known tail exponents. Moreover, let us model the cross-correlation between the two processes by Morgenstern's copula:

$$\mathbb{P}(> t, > t') = \mathbb{P}_{\text{FPT}}(> t)\mathbb{P}_{\text{LT}}(> t') \times \{1 + 3\rho_{\text{true}}[1 - \mathbb{P}_{\text{FPT}}(> t)][1 - \mathbb{P}_{\text{LT}}(> t')]\}, \quad (8.24)$$

with some $-1/3 < \rho_{\text{true}} < 1/3$, or by Frank's copula:

$$\mathbb{P}(> t, > t') = \frac{1}{\alpha} \ln \left[1 + \frac{(e^{\alpha \mathbb{P}_{\text{FPT}}(> t)} - 1)(e^{\alpha \mathbb{P}_{\text{LT}}(> t')} - 1)}{e^\alpha - 1} \right], \quad (8.25)$$

with some $-\infty < \alpha < \infty$. Here $\mathbb{P}(> t, > t') = \int_t^\infty d\tau \int_{t'}^\infty d\tau' \mathbb{P}(\tau, \tau')$ which is the joint distribution function. Copulas are widely used in the finance and insurance literature (for an authoritative introduction, see Ref. [FV98]).

Monte Carlo measurements based on random pairs from these copulas suggest a nearly linear relationship between the true and the restricted correlation coefficients. With the substitution of the typical values of λ_{FPT} and λ_{LT} one finds that the values of the true rank correlation coefficients ρ_{true} are given by

$$\rho_{\text{true}} = r \times \rho_{\text{res}}, \quad (8.26)$$

where $r \approx 1.66$ for Morgenstern's and $r \approx 1.55$ for Frank's copula. The resulting estimates are given in Table 8.4. Naturally, the shuffled surrogate datasets yield $\rho_{\text{true}} = \rho_{\text{res}} = 0$.

In conclusion we observe that there is a strong negative correlation between the first passage time and the lifetime of an order in agreement with Ref. [MF05] but contrary to our model assumption 2 and Eq. (8.11). So the key question is: How much does the presence of this correlation affect the predictions of our model? We performed a series of Monte Carlo simulation of the submission and cancellation processes by using the empirically observed value of tail exponents and cross correlations (Table 8.4). We found that for a fixed value of λ_{FPT} and λ_{LT} the introduction of such correlations increases the values of λ_{TTF} and λ_{TC} by about 0.1, which is comparable to the error bars of our estimates, and the power law behavior is well preserved. Moreover, the central part of our arguments, Eq. (8.18), remains valid [upo]. Thus the presence of a dynamic cancellation strategy does not significantly affect the validity of our model.

stock	ρ_{res}	ρ_{true} (Morg.)	ρ_{true} (Frank)	no. points
AZN	-0.12 ± 0.02	-0.19 ± 0.03	-0.18 ± 0.03	3277
GSK	-0.13 ± 0.01	-0.21 ± 0.02	-0.20 ± 0.02	8573
LLOY	-0.10 ± 0.01	-0.16 ± 0.02	-0.15 ± 0.02	8201
SHEL	-0.13 ± 0.02	-0.21 ± 0.03	-0.19 ± 0.03	2791
VOD	-0.134 ± 0.008	-0.22 ± 0.01	-0.21 ± 0.01	16392

Table 8.4: Estimates of the correlation between first passage times and the lifetimes for orders with entry depth $\Delta = 1$ tick. We discarded orders executed or canceled within $L = 4$ minutes of their placement. $\rho_{\text{surr}} = 0.000 \pm 0.002$ for all stock. The errors represent the standard deviation estimated from 300 bootstrap samples.

stock	$\Delta = -1$			$\Delta = -2$		
	λ	λ'	T	λ	λ'	T
AZN	2.2	1.0	230	2.1	1.2	320
GSK	2.1	1.1	180	2.2	1.3	410
LLOY	2.1	1.0	150	2.0	1.2	220
SHEL	2.1	1.0	120	2.1	1.0	140
VOD	2.0	1.1	130	–	–	–

Table 8.5: Parameters of the fitting function (8.5) for the distribution of time to fill for the five stocks and $\Delta < 0$ ticks. All times are given in seconds. Data are missing when there were the statistics was inadequate for fitting. Typical standard errors for the quantities: ± 0.1 for λ_{TF} , ± 0.1 for λ'_{TF} , and $\pm 25\%$ for T_{TF} .

8.6 A generalization to $\Delta \leq 0$

So far we only considered $\Delta > 0$, i.e., orders with prices which were worse than the best offer at the time of their placement. However, this group only accounts for less than half of the actual limit orders. Measurements for $\Delta \leq 0$ orders give the surprising result that execution times are described by statistics very similar to that for $\Delta > 0$. One example stock (GSK) is shown in Fig. 8.8(left), and fit results with formula (8.5) for all five stocks are given in Table 8.5.

According to our model, these orders should have been executed within a negligible time of placement. While this is true for a number of them, certainly not for all. Let us assume that we are placing a new buy limit order. If our order has $\Delta = 0$, then it will be among the best offers at the time of its placement. If our order has $\Delta < 0$, then it becomes the single best offer in the book, and hence it will trade with certainty if the next event is a sell market order. Why can our order still take a long time before being executed? The answer is naturally that before our order is

stock	$\Delta = -1$			$\Delta = -2$		
	λ	λ'	T	λ	λ'	T
AZN	2.0	0.7	90	1.9	0.8	72
GSK	1.9	0.7	90	1.8	0.8	50
LLOY	1.9	0.7	70	1.8	0.9	85
SHEL	2.2	1.0	130	2.0	1.0	160
VOD	1.8	0.8	120	-	-	-

Table 8.6: Parameters of the fitting function (8.6) for the distribution of time to cancel for the five stocks and $\Delta < 0$ ticks. All times are given in seconds. Data are missing when there were no orders at all, or the statistics was inadequate for fitting. Typical standard errors for the quantities: ± 0.1 for λ_{TTC} , ± 0.1 for λ'_{TTC} , and $\pm 25\%$ for T_{TTC} .

executed, a new buy limit order may enter the book. If this new order has $\Delta < 0$, it means that it has an even better price than our order and it will gain priority of execution. On the other hand, our order now effectively has $\Delta > 0$, and the original model can be applied.

To test such a hypothesis, we carried out the following calculation. For the sake of simplicity, we will consider the time to first fill instead of time to fill. Section 8.3 argued that for the majority of orders the difference between the two is negligible. From the time of its placement, we tracked every single at least partially filled $\Delta \leq 0$ order until the time it was first filled. We defined the reduced entry depth (Δ') and the reduced time to first fill (TTFF') for these orders as follows:

1. For orders, where from their placement to their first fill there were no even more favorable orders both placed and then at least partially filled, $\Delta' = 0$ and $\text{TTFF}' = \text{TTFF}$.
2. For orders where after their placement but before their first fill there was at least one new, more favorable order introduced with $\Delta_{\text{new}} < 0$ and then this new order was at least partially filled, we selected the first of such new orders placed after the original one and set $\Delta' = -\Delta_{\text{new}}$. Thus, Δ' is the new position of the original order, after the new one was placed. TTFF' is defined as the time to first fill of our order measured from the placement of this new order.

The typical distribution of TTFF' for different groups in Δ' is shown in Fig. 8.8(right). For orders with $\Delta' = 0$ this is – except for very short times – well described by a stretched exponential distribution $\mathbb{P}_{\text{TTFF}'}(t) = \frac{1}{25} \exp \left[- \left(\frac{t}{6} \right)^{1/2} \right]$. These are the orders, where there was no better offer made, and hence their execution times were purely determined by the incoming market orders. The distribution is very close to the distribution

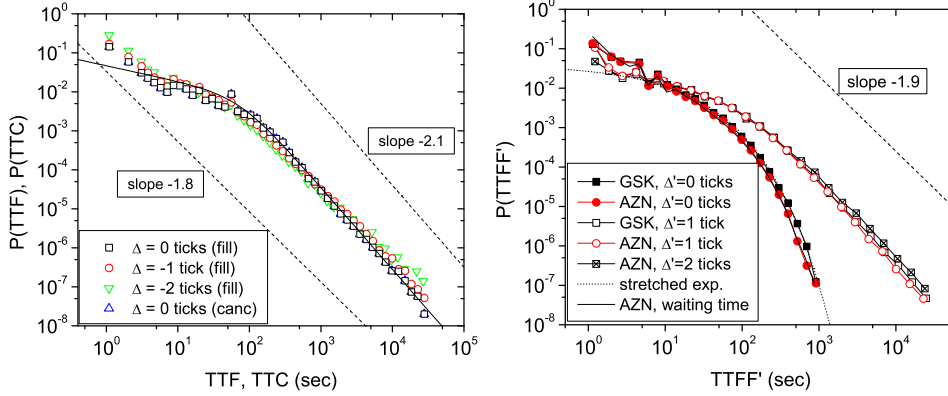


Figure 8.8: **(left)** Examples of the distribution of time to fill/cancel for $\Delta \leq 0$ buy limit orders of GSK. **(right)** The distribution of reduced time to first fill (TTFF') as a function of the reduced entry depth Δ' . For orders where $\Delta' = 0$, the tail of the distribution is well fitted by the stretched exponential $\frac{1}{25} \exp \left[- \left(\frac{t}{6} \right)^{1/2} \right]$. Where $\Delta' > 0$, the distribution decays asymptotically as a power law with an exponent close to -2.0 . The solid line is the distribution of waiting times between two consecutive trades of GSK.

of the times between two consecutive transactions of the stock (see, for an example the right panel of Fig. 8.8).

For orders with $\Delta' > 0$, one recovers the results of the previous sections, and the distribution of reduced time to first fill asymptotically decays as a power law with a power close to 2.0. Eqs. (8.18) and (8.20) are expected to be valid for orders with $\Delta < 0$ and $\Delta' > 0$ as well, given that we use them in terms of Δ' and TTFF' .

As a summary, time to first fill for orders with $\Delta \leq 0$ is a two-component process. If there is no better order submitted before the first fill, then time to first fill is identical to the waiting time distribution between opposite market orders. If there is a better offer submitted, then the order effectively becomes $\Delta > 0$, and the diffusion approximation applies. As this latter process has a much fatter tail than the former one, long waiting times and the tail exponent are again dominated by a first passage process.

8.7 Conclusions

The price formation of stock markets is organized by complex procedures which often involve a limit order book. There have been many attempts to

map some aspects of this auction process to less complicated probabilistic models. In this chapter we showed that the waiting time of an order until its execution presents statistical regularities which are quite robust and can be seen as "stylized facts" of the order book dynamics. These results offer some insight into how the random diffusion process of price can be characterized more precisely.

The difference between statistical properties of time to fill, first passage time and time to cancel are informative of the interdependence of these phenomena. With data of five highly capitalized stocks traded at the London Stock Exchange, we empirically observed that for the transaction time (and best bid and ask) the first passage time, the time to (first) fill and time to cancel are asymptotically power law distributed in time. Based on a simple model we showed that the characteristic exponents of the asymptotic power law behavior of the first passage time, the time to (first) fill and time to cancel are related to each other by simple rules which are in agreement with our empirical observations. These predictions are in contrast with another study (the NASDAQ data investigated in Ref. [CS01]). Therefore further investigation of our results is needed to clarify whether or not they are market dependent.

Although the proposed model makes many simplifying assumptions – for example that order volume is irrelevant for the execution time – we have verified that most of them are sufficiently supported by empirical evidence. Other important stylized facts have been neglected in the model, for example, the role of volatility fluctuations. In fact, the temporal fluctuations of volatility and liquidity result in execution rates that change from time to time. In short we have, at the moment, ignored the conditional aspects of the dynamics of the order book. Some of these aspects have been already investigated in the literature [LMZ02]. An extension of our study along these lines can be feasible.

In spite of its obvious limitations, our simple model is able to make one additional prediction: The placement/cancellation strategies followed by each single trade are strongly heterogeneous. This might come from a heterogeneity either among different traders, or already in the actions of each single trader in different time periods. It remains an open question to discriminate between these two kinds of heterogeneity. Nevertheless, the observation itself is still relevant, because as long as orders cannot be traced back to whoever placed them, this aggregate is what any investor sees as a market environment. Proprietary data firms providing financial services to institutional investors may be used to verify or falsify the results obtained from aggregated data by means of our simple model.

Chapter 9

Summary

9.1 Background

During the last decade physicists have published dozens of books and thousands of scientific papers in the field of finance. This new trend – like many others – stems from the huge development of statistical physics that started in the 70's. In this period many new concepts and models were born, such as fractal and multifractal scaling, frustrated disordered systems, or strongly non-equilibrium phenomena and the tools necessary for their description.

Inspired by these, studies in "econophysics" concentrate on possible analogies between financial and physical systems, but there is a large discrepancy in the level of understanding between the two fields. For example, in the case of a physical system undergoing a second order phase transition the assumption of scaling and the determination of universal exponents is a well motivated undertaking. In the behavior of financial markets there is no basis whatsoever for a similar treatment, and the assumption of power laws is just one of many possibilities for the description of empirical data. Universality, as present in physical systems, is also a much stronger property, than the existence of "stylized facts", the qualitative similarity of stock market data throughout various stocks and markets.

The general observations in stock markets include the broad, non-normal distribution of returns (change of the logarithm of price), the absence of linear autocorrelations in returns, and the long memory of volatility (time dependent standard deviation of returns). The trading activity of stocks (expressed in value or number of shares) is also characterized by a broad distribution and strong temporal autocorrelations.

Although these facts are well known phenomenologically, their accurate description and origin remains an important field of research even today. The most plausible reason for such interest is of course that by de-

vising the appropriate pricing models and trading strategies, it is possible to obtain substantial speculative profits in financial markets. Investment firms often employ physicists in the related areas.

There are, however, further reasons too. The market can be understood as a self-adaptive complex system whose operation is determined both by external influences and a complicated internal structure. The former is comprised by news and macroeconomic effects, while the latter by countless interconnected, interacting firms, banks, brokers. The understanding of the dynamics of market investment decisions could lead the way to significant improvements: The structure and legislation of markets could be changed in the hope of a more efficient operation. For example, it might become possible to dampen the fluctuations and the effects of bubbles that cause market crashes and potentially significant economic damage.

9.2 Goals

The most important goal of my research was to critically revisit the stylized facts regarding the dynamics of trading, and then describe these phenomena by scaling theories and random models. Besides this, I have put great emphasis on facilitating a stronger connection between the theory of complex systems and the description of markets. This was achieved through the phenomenon of fluctuation scaling and the introduction of a unified scaling theory. This theory and my empirical results also made it inevitable to revise several statements found in the recent literature.

The notion of stock market fluctuations can be approached empirically in two ways. One is based on the transaction data characterizing the trading activity of the whole market. For this study I used the TAQ database containing all transactions on the New York Stock Exchange and NASDAQ in the period 1993-2003. The other approach is the analysis of the so called limit order book that contains all individual buy and sell offers. The order book is the most complete source of information available to the participants of trading. It makes possible a very detailed inquiry into the nature of price formation, the trading process and market microstructure. My aim was to identify the most important features of the dynamics of the order book on different time scales. Based on these it became possible to analyze, understand and model the statistics of individual order executions. For this research I used the complete order book data of the London Stock Exchange for the year 2002.

9.3 Methods

My work included filtering and processing the above mentioned databases, their statistical analysis and comparison with Monte Carlo simulations. For the interpretation of the results it was necessary to create, analyze and simulate benchmark models. For the solution of the great majority of these tasks I used the programming language C. My theoretical calculations were based on tools of probability theory, the theory of random walks, and a phenomenological scaling theory.

9.4 New scientific results

1. Many complex systems are characterized by the fluctuation scaling relationship between its i elements. This means that for some positive, additive, time-dependent quantity $f_i(t)$ the standard deviation σ_i and the mean $\langle f_i \rangle$ is related by the law

$$\sigma_i \propto \langle f_i \rangle^\alpha .$$

I have shown that if one chooses $f_i(t)$ as the traded value of the i 'th share on the stock market during the period $[t, t + \Delta t)$ the above scaling law is valid, and α strongly depends on the window size Δt . [EKYB05]

2. I introduced a combined scaling theory of long-range temporal autocorrelations and fluctuation scaling. In this framework I showed that the behavior of the $\alpha(\Delta t \rightarrow 0)$ limit is explained by the relationship of the frequency and the size of the transactions. Furthermore, I showed that the time window dependence of α is equivalent to that for short times trading activity is uncorrelated, and for longer times it is characterized by the $H_i = H_* + \gamma \log \langle f_i \rangle$ Hurst exponents. Since $\langle f \rangle$ essentially characterizes liquidity, this logarithmic relationship is proof, that the strength of correlations displays systematic non-universality. [EK05, KE05b, EK06a, KE05a, EK06c]
3. I extended fluctuation scaling and the related scaling theory to the analysis of higher moments. In the generalized, multiscaling form

$$\langle |f_i - \langle f_i \rangle|^q \rangle \propto \langle f_i \rangle^{\alpha(q)} .$$

I showed that this expression can be applied to the distribution of stock market traded value. I also showed that the correlation properties of f also display multiscaling, and the $\tau(q)$ scaling function and the multifractal spectrum are both liquidity dependent. [EKYB05, EK07b]

4. By the comparative application of various fitting methods I showed, that the distribution of the f traded value, in contrast with earlier statements in the literature, has a finite second moment, and thus the related Hurst exponent is well defined. I determined that the correlations of traded value have two origins: correlations between the V size of consecutive transactions, and in the N number of trades in the given time interval. The time dependence of V and N is characterized by non-universal Hurst exponents, and both give a significant contribution to the correlations of f . [EK06b, EK07b, EK07a]
5. I showed, that order book level data display different qualitative behavior on different time scales. The distributions of time to fill and time to cancel decay asymptotically as power laws with the same exponent $\lambda_{\text{TF}} \approx \lambda_{\text{TC}} \approx 2$. This value is greater than the $\lambda_{\text{FPT}} \approx 1.5$ value characterizing the first passage time of price. Based on a random walker model I showed that the order lifetimes are also asymptotically power law distributed, and the related exponent can be expressed as $\lambda_{\text{LT}} = \lambda_{\text{TF}} - \lambda_{\text{FPT}} + 1$. This relationship can be used to calculate λ_{LT} in the case of real data where it cannot be observed directly. I investigated the dependence of the above processes as the function of a Δ distance from the best price. Based on Monte Carlo simulations I showed that the results of the model do not change significantly if one relaxes the simplifying assumptions. [EKLM07, EKL07]

Chapter 10

Összefoglaló (Summary in Hungarian)

10.1 A kutatások előzménye

Az utóbbi évtizedben könyvek tucatjait és tudományos cikkek ezreit publikálták a pénzügyek területén fizikusok. Ez az új trend – mint sok másik is – a statisztikus fizika a 70-es években kezdődő látványos fejlődéséből fakad. Ebben az időszakban számos új koncepció és modell született, mint például a fraktál és multifraktál skálázás, a frusztrált rendezetlen rendszerek, vagy az erősen nem-egyensúlyi jelenségek és a leírásukhoz szükséges eszközök.

Az ezek által inspirált, "gazdaságfizikai" tanulmányok lehetséges analógiákra koncentrálnak, de a pénzügyi és fizikai rendszerek között a megértés szintjében jelenleg óriási a különbség. Például egy másodrendű fázisátalakuláson átmenő fizikai rendszerben a skálázás feltételezése és univerzális exponensek meghatározása jól motivált. Pénzügyi piacok viselkedésében erre nincsen hasonló alap és hatványfüggvény szerinti eloszlások feltételezése csak egy lehetőség a sok közül az empirikus adatok leírására. A fizikai rendszerekben meglévő univerzalitás szintén sokkal erősebb tulajdonság, mint az ú.n. stilizált tények léte, a tőzsdei adatok kvalitatív hasonlósága különböző részvényekre és piacokon.

Az általános tőzsdei megfigyelések közé tartozik a hozamok (az ár logaritmusának megváltozása) széles, nem-normális eloszlása, a lineáris autokorrelációk eltűnése a hozamokban és a volatilitás (a hozamok időfüggő szórása) hosszú memóriája. A részvények kereskedett mennyisége (értékben vagy a részvények számában mérve egyaránt) szintén széles eloszlást és erős időbeli korrelációkat mutat.

Bár ezen jelenségek fenomenologikusan jól ismertek, pontos leírásuk és eredetük továbbra is fontos kutatási terület. Érdekességük legkézenfekvőbb oka természetesen az, hogy alkalmas árazási modellek és kereskedési stratégiák kidolgozásával jelentős spekulatív profit érhető el a pénzü-

gyi piacokon. Befektetési cégek előszeretettel alkalmaznak fizikusokat az ehhez kapcsolódó területeken.

Vannak azonban ezen túlmutató okok is. A piac felfogható önadaptív komplex rendszerként, amely működését együttesen határozzák meg a külső hatások és egy összetett belső struktúra. Előbbit a hírek, gazdasági változások, utóbbit a számtalan, egymással kapcsolatban, "kölsönhatásban" lévő cég, bank, bróker alkotják. A piaci befektetési döntések dinamikájának megértésével egyúttal lehetőség nyílna arra is, hogy a tőzsde szerkezetét és a jogszabályokat a jobb működés érdekében megváltoztassuk. Így például csökkenthetők lennének a fluktuációk és a buborékok hatásai, amelyek súlyos esetben tőzsdei összeomlásokat és jelentős gazdasági károkat is okozhatnak.

10.2 Célkitűzések

Kutatásom legfontosabb célja a tőzsdei kereskedés dinamikájára vonatkozó stilizált tények felülvizsgálata, majd a jelenségek skálatörvényekkel és véletlen modellekkel történő leírása volt. Emellett kiemelt hangsúlyt kapott a komplex rendszerek elmélete és a tőzsdei jellemzők leírása közötti kapcsolat szorosabbá tétele. Ezt a fluktuáció skálázás jelenségén és egy egységes skálaelmélet kidolgozásán keresztül értem el. A kutatás során kidolgozott elmélet és az empirikus eredmények szükségessé tették az irodalomban fellelhető számos korábbi állítás felülvizsgálatát is.

A tőzsdei fluktuációk kérdéskörének empirikus szempontból kétféle megközelítése lehetséges. Az egyik a teljes tőzsde kereskedési aktivitását jellemző tranzakciós adatokon alapul. Ehhez a vizsgálathoz rendelkezésre állt a New York Stock Exchange-en és a NASDAQ-on az 1993-2003 közötti időszakban történt tranzakciókat tartalmazó TAQ adatbázis. A másik megközelítés a kereskedők egyes vételi és eladási ajánlatait tartalmazó ún. ajánlati könyv vizsgálata. Az ajánlati könyv a kereskedés résztvevői számára rendelkezésre álló lehető legteljesebb információforrás. Csak ennek segítségével lehetséges az árképzés, a kereskedési folyamat és a piaci mikrostruktúra részletesebb elemzése. Céлом az ajánlati könyv dinamikájának fő vonásainak azonosítása volt különböző időskálákon. Ezek alapján lehetővé vált az egyes tranzakciók végrehajtásának statisztikai elemzése és megértése, valamint az eredmények modell szintű leírása. Ehhez a kutatáshoz a London Stock Exchange-en a 2002-es év ajánlati könyvének adatai álltak rendelkezésre.

10.3 Vizsgálati módszerek

Munkámhoz hozzátartozott az említett adatbázisok feldolgozása és szűrése, az adatok statisztikai kiértékelése és összehasonlítása Monte Carlo szí-

mulációkkal. Emellett az eredmények értelmezéséhez elengedhetetlen volt modellek megalkotása, elemzése és szimulációja. A feladatok megoldásához döntő részben a C programozási nyelvet használtam. Elméleti számításaim valószínűségszámítási eszközökön, a véletlen bolyongások elméletén és egy fenomenologikus skálaelméleten alapultak.

10.4 Új tudományos eredmények

1. Számos komplex rendszer i elemeire jellemző a fluktuáció skálázási reláció, azaz hogy egy pozitív, additív, időfüggő $f_i(t)$ mennyiség szórása és várható értéke a

$$\sigma_i \propto \langle f_i \rangle^\alpha$$

relációval kapcsolható össze. Megmutattam, hogy $f_i(t)$ -t a tőzsdén az i . részvény $[t, t + \Delta t)$ időszakban kereskedett értékének választva igaz a fenti skálatörvény, és α erősen függ a Δt ablakmérettől. [EKYB05]

2. Kidolgoztam a hatványfüggvény szerinti időbeli autokorrelációk és a fluktuáció skálázás közös skálaelméletét. Ennek keretében megmutattam, hogy az $\alpha(\Delta t \rightarrow 0)$ határeset viselkedését a kereskedések számának és nagyságának összefüggése magyarázza. Megmutattam továbbá, hogy α időablak függése ekvivalens azzal, hogy részvények kereskedése rövid időkre korrelálatlan, hosszabb időkre pedig a $H_i = H_* + \gamma \log \langle f_i \rangle$ Hurst exponenssel jellemezhető. Mivel $\langle f \rangle$ lényegében a likviditást jellemzi, ez a logaritmikus összefüggés bizonyítéka annak, hogy a korrelációk erőssége szisztematikus nem-univerzalitást mutat. [EK05, KE05b, EK06a, KE05a, EK06c]
3. A fluktuáció skálázást és a hozzá tartozó skálaelméletet kiterjesztettem a magasabb momentumok vizsgálatára. Az általánosított, multiskálázást leíró alakban

$$\langle |f_i - \langle f_i \rangle|^q \rangle \propto \langle f_i \rangle^{\alpha(q)}.$$

Megmutattam, hogy ez a kifejezés alkalmazható a tőzsdei kereskedett érték eloszlására. Emellett megmutattam, hogy f korrelációs tulajdonságai szintén multiskálázást mutatnak, és hogy a $\tau(q)$ skálafüggvény valamint a multifraktál spektrum is likviditásfüggő. [EKYB05, EK07b]

4. Különböző illesztési módszerek alkalmazásával megmutattam, hogy az f kereskedett érték eloszlásnak a korábbi irodalmi állításokkal ellentétben létezik véges szórása, tehát a Hurst exponensek jól definiáltak. Megállapítottam, hogy a kereskedett érték korrelációi két forrásra vezethetők vissza: az egymást követő tranzakciók V értéke

közötti és az adott intervallumban kötött tranzakciók N számában jelentkező korrelációkra. V és N időfüggése egyaránt nem-univerzális Hurst exponensekkel jellemezhető, és mindkettő járuléka jelentős f korrelációihoz. [EK06b, EK07b, EK07a]

5. Megmutattam, hogy az ajánlati könyv szintű adatok különböző időskálákon eltérő kvalitatív viselkedést mutatnak. Az ajánlatok teljesítési (time to fill) és törlési idejének (time to cancel) statisztikája aszimptotikusan hatványfüggvény eloszlású, azonos, $\lambda_{\text{TTF}} \approx \lambda_{\text{TTC}} \approx 2$ körüli exponenssel. Ez az érték nagyobb az ár első áthaladási idejét (first passage time) jellemző $\lambda_{\text{FPT}} \approx 1.5$ exponensnél. Véletlen bolyongáson alapuló modell segítségével megmutattam, hogy az ajánlatok élettartamát (order lifetime) aszimptotikusan jellemző λ_{LT} exponens kifejezhető a $\lambda_{\text{LT}} = \lambda_{\text{TTF}} - \lambda_{\text{FPT}} + 1$ alakban. Ez a reláció felhasználható, hogy valódi adatokból is kiszámítsuk a közvetlen módon nem mérhető λ_{LT} exponenst. Megvizsgáltam a fenti folyamatok viselkedését a legjobb ártól mért Δ távolság függvényében. Szimulációk segítségével megmutattam, hogy a modell eredményei nem változnak az egyszerűsítő feltevések elhagyása esetén. [EKLM07, EKL07]

Bibliography

- [AGCH82] R.M. Anderson, D.M. Gordon, M. J. Crawley, and M. P. Hassell. Variability in the abundance of animal and plant species. *Nature*, 296:245–248, 1982.
- [Alv01] M. I. Fraga Alves. A location invariant hill-type estimator. *Extremes*, 4:199–217, 2001.
- [Bal06] P. Ball. Econophysics: Culture crash. *Nature*, 441:686–688, 2006.
- [BB05] L. Borland and J.-P. Bouchaud. On a multi-timescale statistical feedback model for volatility fluctuations. *physics/0507073*, 2005.
- [BDW93] M.J. Barclay, C.G. Dunbar, and J.B. Warner. Stealth and volatility: which trades move prices? *Journal of Financial Economics*, 34:281–306, 1993.
- [BEPM⁺06] A. Bar-Even, J. Paulsson, N. Maheshri, M. Carmi, E. O’Shea, Y. Pilpel, and N. Barkai. Noise in protein expression scales with natural protein abundance. *Nature Genetics*, 38:636–643, 2006.
- [BGS05] B. Biais, L. Glosten, and C. Spatt. Market microstructure: A survey of microfoundations, empirical results, and policy implications. *Journal of Financial Markets*, 8:217–264, 2005.
- [BHS95] B. Biais, P. Hillion, and C. Spatt. An empirical analysis of the limit order book and the order flow in the paris bourse. *Journal of Finance*, 50:1655–1689, 1995.
- [BKP06] J.-P. Bouchaud, J. Kockelkoren, and M. Potters. Random walks, liquidity molasses and critical response in financial markets. *Quantitative Finance*, 6:115–123, 2006.
- [BLM00] G. Bonanno, F. Lillo, and R.N. Mantegna. Dynamics of the number of trades of financial securities. *Physica A*, 280:136–141, 2000.

- [BMP01] J.-P. Bouchaud, A. Matacz, and M. Potters. Leverage effect in financial markets: The retarded volatility model. *Phys. Rev. Lett.*, 87:228701, 2001.
- [BMP02] J.-P. Bouchaud, M. Mézard, and M. Potters. Statistical properties of stock order books: empirical results and models. *Quantitative Finance*, 2:251–256, 2002.
- [Bou] J.-P. Bouchaud. Personal communication.
- [BP00] J.-P. Bouchaud and M. Potters. *Theory of Financial Risk*. Cambridge University Press, Cambridge, 2000.
- [BS73] F. Black and M. Scholes. The pricing of options and corporate liabilities. *The Journal of Political Economy*, 81:637–654, 1973.
- [BT03] N. Barberis and R. Thaler. *A survey of behavioral finance*, volume Handbook of the Economics of Finance, chapter 18, pages 1053–1128. Elsevier, 2003.
- [BTW87] P. Bak, C. Tang, and K. Wiesenfeld. Self-organized criticality: an explanation of $1/f$ noise. *Phys. Rev. Lett.*, 59:381–384, 1987.
- [CH95] S. Chakravarty and C. Holden. An integrated model of market and limit orders. *Journal of Financial Intermediation*, 4:213–241, 1995.
- [Cha01] S. Chakravarty. Stealth-trading: Which traders’ trades move stock prices? *Journal of Financial Economics*, 61:289–307, 2001.
- [CLM96] J. Y. Campbell, A. Lo, and A. C. MacKinlay. *The Econometrics of Financial Markets*. Princeton University Press, Princeton, 1996.
- [CMG⁺03] A. V. Chechkin, R. Metzler, V. Y. Gonchar, J. Klafter, and L. V. Tanatarov. First passage and arrival time density for lévy flights and the failure of the method of images. *J. Phys. A: Math. Gen.*, 36:L537–L544, 2003.
- [Con01] R. Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1:223–236, 2001.
- [CP03] M. Li Calzi and P. Pellizzari. Fundamentalist clashing over the book: a study of order-driven stock markets. *Quantitative Finance*, 3:470–480, 2003.

- [CRS01] T. Chordia, R. Roll, and A. Subrahmanyam. Market liquidity and trading activity. *Journal of Finance*, 56:501–530, 2001.
- [CS01] D. Challet and R. Stinchcombe. Analyzing and modeling 1+1d markets. *Physica A*, 300:285–299, 2001.
- [DA06] J. Duch and A. Arenas. Scaling of fluctuations in traffic on complex networks. *Phys. Rev. Lett.*, 96:218702, 2006.
- [DFIS03] M. G. Daniels, J. D. Farmer, G. Iori, and E. Smith. Quantitative model of price diffusion and market friction based on trading as a mechanistic random process. *Physical Review Letters*, 90:108102, 2003.
- [DJ05] K. Dahlstedt and H. J. Jensen. Fluctuations spectrum and size scaling of river flow and level. *Physica A*, 348:596–610, 2005.
- [dMB04a] M.A. de Menezes and A.-L. Barabási. Fluctuations in network dynamics. *Phys. Rev. Lett.*, 92:28701, 2004.
- [dMB04b] M.A. de Menezes and A.-L. Barabási. Separating internal and external dynamics of complex systems. *Phys. Rev. Lett.*, 93:68701, 2004.
- [EBI98] S. Engen, O. Bakke, and A. Islam. Demographic and environmental stochasticity-concepts and definitions. *Biometrics*, 54:840–846, 1998.
- [EBJ] Z. Eisler, I. Bartos, and I. M. Jánosi. *in preparation*.
- [EBK07] Z. Eisler, I. Bartos, and J. Kertész. Fluctuation scaling in complex systems: Taylor’s law and beyond. *submitted to Adv. Phys.*, *arXiv:0708.2053*, 2007.
- [EK05] Z. Eisler and J. Kertész. Random walks on complex networks with inhomogeneous impact. *Phys. Rev. E*, 71:057104, 2005.
- [EK06a] Z. Eisler and J. Kertész. Scaling theory of temporal correlations and size dependent fluctuations in the traded value of stocks. *Phys. Rev. E*, 73:046109, 2006.
- [EK06b] Z. Eisler and J. Kertész. Size matters: some stylized facts of the stock market revisited. *Eur. Phys. J. B*, 51:145–154, 2006. (also physics/0508156).
- [EK06c] Z. Eisler and J. Kertész. Why do Hurst exponents increase as the logarithm of company size? In B. K. Chakrabarti

- A. Chatterjee, editor, *Econophysics of Stock and Other Markets, Proceedings of Econophysics Kolkata II*, pages 49–58, 2006.
- [EK07a] Z. Eisler and J. Kertész. The dynamics of traded value revisited. *Physica A*, 382:66–72, 2007.
- [EK07b] Z. Eisler and J. Kertész. Liquidity and the multiscaling properties of the volume traded on the stock market. *Europhys. Lett.*, 77:28001, 2007.
- [EKL07] Z. Eisler, J. Kertész, and F. Lillo. The limit order book on different time scales. *Proc. SPIE*, 6601:66010G, 2007.
- [EKLM07] Z. Eisler, J. Kertész, F. Lillo, and R.N. Mantegna. Diffusive behavior and the modeling of characteristic times in limit order executions. *physics/0701335*, submitted to *Quantitative Finance*, 2007.
- [EKYB05] Z. Eisler, J. Kertész, S.-H. Yook, and A.-L. Barabási. Multiscaling and non-universality in fluctuations of driven complex systems. *Europhys. Lett.*, 69:664–670, 2005.
- [Far99] J. D. Farmer. Physicists attempt to scale the ivory towers of finance. *Comp. Sci. & Eng. (IEEE)*, 1:26–39, 1999.
- [Fel67] W. Feller. *An introduction to probability theory*, volume 1. Wiley & Sons, 3 edition, 1967.
- [FGL⁺04] J.D. Farmer, L. Gillemot, F. Lillo, S. Mike, and A. Sen. What really causes large price changes? *Quantitative Finance*, 4:383–397, 2004.
- [FGLM06] J. D. Farmer, A. Gerig, F. Lillo, and Sz. Mike. Market efficiency and the long-memory of supply and demand: is price impact variable and permanent or fixed and temporary? *Quantitative Finance*, 6:107–112, 2006.
- [FL99] J. D. Farmer and A. W. Lo. Frontiers of finance: Evolution and efficient markets. *Proc. Natl. Acad. Sci. USA*, 96:9991–9992, 1999.
- [FL04] J. D. Farmer and F. Lillo. On the origin of power law tails in price fluctuations. *Quantitative Finance*, 4:C7–C11, 2004.
- [FPZ05] J. D. Farmer, P. Patelli, and I. I. Zovko. The predictive power of zero intelligence in financial markets. *Proc. Natl. Acad. Sci. USA*, 102:2254–2259, 2005.

- [Fra68] J.F. Franklin. Cone production by upper slope conifers. Pacific NW Forest Range Experiment Station Research Paper No. PNW-60, 1968.
- [FV98] E. W. Frees and E. A. Valdez. Understanding relationships using copulas. *North American Actuarial Journal*, 2:1, 1998.
- [GGPS03] X. Gabaix, P. Gopikrishnan, V. Plerou, and H.E. Stanley. A theory of power-law distributions in financial market fluctuations. *Nature*, 423:267–270, 2003.
- [Glo94] L. Glosten. Is the electronic open limit order book inevitable? *Journal of Finance*, 49:1127–1161, 1994.
- [GMAS98] P. Gopikrishnan, M. Meyer, L.A.N. Amaral, and H.E. Stanley. Inverse cubic law for the distribution of stock price variations. *Eur. Phys. J. B*, 3:139–140, 1998.
- [GPGS00] P. Gopikrishnan, V. Plerou, X. Gabaix, and H.E. Stanley. Statistical properties of share volume traded in financial markets. *Phys. Rev. E*, 62:4493–4496, 2000.
- [GWF⁺98] B.T. Grenfell, K. Wilson, B.F. Finkenstädt, T.N. Coulson, S. Murray, S.D. Albonk, J.M. Pemberton, T.H. Clutton-Brock, and M.J. Crawley. Noise and determinism in synchronized sheep dynamics. *Nature*, 394:674–677, 1998.
- [HH96] L. Harris and J. Hasbrouck. Market vs. limit orders: The superdot evidence on order submission strategy. *Journal of Financial and Quantitative Analysis*, 31:213–231, 1996.
- [Hil75] B.M. Hill. A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3:1163–1174, 1975.
- [HJK⁺86] T.C. Halsey, M.H. Jensen, L.P. Kadanoff, I. Procaccia, and B.I. Shraiman. Fractal measures and the singularities: The characterization of strange sets. *Phys. Rev. A*, 33:1141–1151, 1986.
- [HK95] W. Härdle and A. Kirman. Nonclassical demand: A model-free examination of price-quantity relations in the marseille fish market. *Journal of Econometrics*, 92:227–257, 1995.
- [HMS04] B. Hollifield, R. A. Miller, and P. Sandas. Empirical analysis of limit order markets. *Review of Economic Studies*, 71:1027–1063, 2004.
- [HS96] P. Handa and R. Schwartz. Limit order trading. *Journal of Finance*, 51:1835–1861, 1996.

- [IYPL04] P.Ch. Ivanov, A. Yuen, B. Podobnik, and Y. Lee. Common scaling patterns in intertrade times of U.S. stocks. *Phys. Rev. E*, 69:56107, 2004.
- [JG99] I. M. Jánosi and J. A.C. Gallas. Growth of companies and water-level fluctuations of the river Danube. *Physica A*, 271:448–457, 1999.
- [KABS02] T. H. Keitt, L. A. N. Amaral, S. V. Buldyrev, and H. E. Stanley. Scaling in the growth of geographically subdivided populations: invariant patterns from a continent-wide biological survey. *Phyl. Trans. R. Soc. London B*, 357:627–633, 2002.
- [Kav99] K. Kavajecz. A specialist’s quoted depth and the limit order book. *Journal of Finance*, 54:747–771, 1999.
- [KE05a] J. Kertész and Z. Eisler. Limitations of scaling and universality in stock market data. in *Proc. International Conference for the Management of Risk in Economically Relevant Human Activities, physics/0512193*, 2005.
- [KE05b] J. Kertész and Z. Eisler. Non-trivial scaling of fluctuations in the trading activity of the nyse. *arXiv:physics/0503139, Proceedings of the III-rd Nikkei Econophysics Symposium*, pages 19–23, 2005.
- [Kee00] M. J. Keeling. Simple stochastic models and their power-law type behaviour. *Theoretical Population Biology*, 58:21–31, 2000.
- [KG99] M. J. Keeling and B. T. Grenfell. Stochastic dynamics and a power law for measles variability. *Phyl. Trans. R. Soc. London B*, 354:769–776, 1999.
- [KI03] A. M. Kilpatrick and A. R. Ives. Species interactions can explain Taylor’s power law for ecological time series. *Nature*, 422:65–68, 2003.
- [Kir92] A. P. Kirman. Whom or what does the representative individual represent? *Journal of Economic Perspectives*, 6:117–136, 1992.
- [KOD05] J. Kwapien, P. Oswiecimka, and S. Drozd. *Physica A*, 350:466–474, 2005.
- [Kp03] A. J. Kerkhoff and F. Ballantyne IV. The scaling of reproductive variability in trees. *Ecology Letters*, 6:850–856, 2003.

- [KS98] T. H. Keitt and H. E. Stanley. Dynamics of North American breeding bird populations. *Nature*, 393:257–260, 1998.
- [KTK⁺99] L. Kullmann, J. Töyli, J. Kertész, A. Kanto, and K. Kaski. Characteristic times in stock market indices. *Physica A*, 269:98–110, 1999.
- [KZKB⁺02] J.W. Kantelhardt, S.A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, and H. Eugene Stanley. Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A*, 316:87–114, 2002.
- [Lil06] F. Lillo. Limit order placement as an utility maximization problem and the origin of power law distribution of limit order prices. *physics/0612016*, 2006.
- [LMZ02] A.W. Lo, A.C. MacKinlay, and June Zhang. Econometric models of limit-order executions. *Journal of Financial Econometrics*, 65:31–71, 2002.
- [Lot25] A. J. Lotka. *Elements of Physical Biology*. Williams and Wilkins Company, Baltimore, 1925.
- [Luc03] H. Luckock. A steady-state model of the continuous double auction. *Quantitative Finance*, 3:385–404, 2003.
- [Man97] B.B. Mandelbrot. *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk*. Springer-Verlag, 1997.
- [MAS03] K. Matia, Y. Ashkenazy, and H.E. Stanley. Multifractal properties of price fluctuations of stocks and commodities. *Europhys. Lett*, 61:422–428, 2003.
- [May74] R. May. *Stability and Complexity in Model Ecosystems, 2nd edition*. Princeton University Press, 1974.
- [MBA91] J.F. Muzy, E. Bacry, and A. Arneodo. Wavelets and multifractal formalism for singular signals: Application to turbulence data. *Phys. Rev. Lett.*, 67:3515–3518, 1991.
- [MdSQ06] L.G. Moyano, J. de Souza, and S.M.D. Queirós. On the multi-fractal structure of traded volume in financial markets. *Physica A*, 371:118–121, 2006. *physics/0512240*.
- [MF05] Sz. Mike and J. Doyné Farmer. An empirical behavioral model of price formation. *physics/0509194*, 2005.

- [MKG96] K.P.N. Murthy, K.W. Kehr, and A. Giacometti. Multifractal scaling of moments of mean first-passage time in the presence of Sinai disorder. *Phys. Rev. E*, 53:444–449, 1996.
- [MM01] S. Maslov and M. Mills. Price fluctuations from the order book perspective – empirical facts and a simple model. *Physica A*, 299:234–246, 2001.
- [MPM⁺05] M. Montero, J. Perelló, J. Masoliver, F. Lillo, S. Micciché, and R. N. Mantegna. Scaling and data collapse for the mean exit time of asset prices. *Phys. Rev. E*, 72:056101, 2005.
- [MRGS00] F. Mainardi, M. Raberto, R. Gorenflo, and E. Scalas. Fractional calculus and continuous-time finance II: The waiting-time distribution. *Physica A*, 287:468–481, 2000.
- [MS94] R. N. Mantegna and H. E. Stanley. Stochastic process with ultraslow convergence to a Gaussian: the truncated Lévy flight. *Phys. Rev. Lett.*, 73:2946–2949, 1994.
- [MS99] R.N. Mantegna and H.E. Stanley. *Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge University Press, Cambridge, 1999.
- [MT02] B. A. Maurer and M. L. Taper. Connecting geographical distributions with population processes. *Ecology Letters*, 5:223–231, 2002.
- [NOA05] J.C. Nacher, T. Ochiai, and T. Akutsu. On the relation between fluctuations and scaling-law in gene expression time series from yeast to human. *Modern Physics Letters B*, 19:1169–1177, 2005.
- [Osl03] C. L. Osler. Currency orders and exchange rate dynamics: An explanation for the predictive success of technical analysis. *The Journal of Finance*, 58:1791–1820, 2003.
- [Par96] V. Pareto. *Cours d’économie politique*. Droz, Geneva Switzerland, Geneva, Switzerland, 1896.
- [Par98] C. Parlour. Price dynamics in a limit order market. *Review of Financial Studies*, 11:789–816, 1998.
- [Pau04] J. Paulsson. Summing up the noise in gene networks. *Nature*, 427:415–418, 2004.
- [PB03] M. Potters and J.-P. Bouchaud. More statistical properties of order books and price impact. *Physica A*, 324:133–140, 2003.

- [PBH⁺94] C.-K. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, and A.L. Goldberger. Mosaic organization of DNA nucleotides. *Phys. Rev. E*, 49:1685–1689, 1994.
- [Per94] J. N. Perry. Chaotic dynamics can generate Taylor’s power law. *Proc. R. Soc. Lond. B*, 257:221–226, 1994.
- [pG01] O. N. Bjørnstad and B. T. Grenfell. Noisy clockwork: Time series analysis of population fluctuations in animals. *Science*, 293:638–643, 2001.
- [PGA⁺99] V. Plerou, P. Gopikrishnan, L. A. N. Amaral, M. Meyer, and H. E. Stanley. Scaling of the distribution of price fluctuations of individual companies. *Phys. Rev. E*, 60(6519-6529), 1999.
- [PGA⁺00] V. Plerou, P. Gopikrishnan, Luís A. Nunes Amaral, Xavier Gabaix, and H. Eugene Stanley. Economic fluctuations and anomalous diffusion. *Phys. Rev. E*, 62:R3023–R3026, 2000.
- [PGGS04] V. Plerou, P. Gopikrishnan, X. Gabaix, and H.E. Stanley. On the origin of power-law fluctuations in stock prices. *Quantitative Finance*, 4:C11–C15, 2004.
- [pK05] F. Ballantyne IV and A. J. Kerkhoff. Reproductive correlation and mean-variance scaling of reproductive output for a forest model. *Journal of Theoretical Biology*, 235:373–380, 2005.
- [pK07] F. Ballantyne IV and A. J. Kerkhoff. The observed range for temporal mean-variance scaling exponents can be explained by reproductive correlation. *Oikos*, 116:174–180, 2007.
- [PLM06] A. Ponzi, F. Lillo, and R. N. Mantegna. Market reaction to temporary liquidity crises and the permanent market impact. *physics/0608032*, 2006.
- [PS03] L. Pástor and R.F. Stambaugh. Liquidity risk and expected stock returns. *Journal of Political Economy*, 111:642–685, 2003.
- [pTE⁺00] B.-E. Sæther, J. Tufto, S. Engen, K. Jerstad, O. W. Røstad, and J. E. Skåtan. Population dynamical consequences of climate change for a small temperate songbird. *Science*, 287:854–856, 2000.
- [Que05] S.M.D. Queirós. On the distribution of high-frequency stock market traded volume: a dynamical scenario. *Europhys. Lett.*, 71:339–345, 2005.

- [Red01] S. Redner. *A Guide to First-Passage Processes*. Cambridge University Press, 2001.
- [RH04] D. H. Reed and G. R. Hobbs. The relationship between population size and temporal variability in population size. *Animal Conservation*, 7:1–8, 2004.
- [Rol84] R. Roll. A simple implicit measure of the effective bid-ask spread in an efficient market. *Journal of Finance*, 39:1127–1140, 1984.
- [San01] P. Sandas. Adverse selection and competitive market making: Empirical evidence from a limit order market. *Review of Financial Studies*, 14:704–734, 2001.
- [Sca05] E. Scalas. The application of continuous-time random walks in finance and economics. *Physica A*, 362:225–239, 2005.
- [Sep97] D. Seppi. Liquidity provision with limit orders and a strategic specialist. *Review of Financial Studies*, 10:704–734, 1997.
- [SFGK03] E. Smith, J. D. Farmer, L. Gillemot, and S. Krishnamurthy. Statistical theory of the continuous double auction. *Quantitative Finance*, 3:481–514, 2003.
- [SI00] A. Satake and Y. Iwasa. Pollen coupling of forest trees: Forming synchronized and periodic reproduction out of chaos. *Journal of Theoretical Biology*, 203:63–84, 2000.
- [SJJ02] I. Simonsen, M.H. Jensen, and A. Johansen. Optimal investment horizons. *Eur. Phys. J. B*, 27:583–586, 2002.
- [SM04] I. M. Sokolov and R. Metzler. Non-uniqueness of the first passage time density of lévy random processes. *J. Phys. A: Math. Gen.*, 37:L609–L615, 2004.
- [Smi38] H. Fairfield Smith. An empirical law describing heterogeneity in the yields of agricultural crops. *J. Agric. Sci.*, 28:1–23, 1938.
- [SS06] F. Schmid and R. Schmidt. Bootstrapping Spearman’s multivariate rho. Proc. Compstat (in press), 2006.
- [SW82] V. Seshadri and Bruce J. West. Fractal dimensionality of lévy processes. *PNAS*, 79:4501–4505, 1982.
- [Tal78] K. Tallqvist. Results of long-time measurements of the quality of flowering and seed crop of trees. *Folia Forestali*, 364:1–60, 1978.

- [taq] Trades and Quotes Database for 1993-2003, New York Stock Exchange, New York.
- [Tay61] L.R. Taylor. Aggregation, variance and the mean. *Nature*, 189:732–735, 1961.
- [Tay86] L.R. Taylor. Synoptic dynamics, migration and the rothamsted insect survey. *Journal of Animal Ecology*, 55:1–38, 1986.
- [TCL⁺07] M. Tumminello, C. Coronello, F. Lillo, S. Miccichè, and R. N. Mantegna. A tool for filtering information in complex systems. *to appear in J. Bifurcation Chaos*, 17, 2007. physics/0605116.
- [upo] Data available upon request.
- [Vic92] T. Vicsek. *Fractal Growth Phenomena*. World Scientific Publishing, Singapore, 1992.
- [Vol25] V. Volterra. Variations and fluctuations of the number of individuals in animal species living together. In *Animal Ecology*, pages 409–448. McGraw-Hill, 1925.
- [vT06] M. Šuvakova and B. Tadić. Transport processes on homogeneous planar graphs with scale-free loops. *Physica A*, 372:354–361, 2006.
- [WBK⁺06] M. Wyart, J.-P. Bouchaud, J. Kockelkoren, M. Potters, and M. Vettorazzo. Relation between bid-ask spread, impact and volatility in double auction markets. cond-mat/0603084, 2006.
- [WF86] M. J. Weaver and F. Forcella. Cone production in pinus albicaulis forests. In Proceedings of a symposium on conifer tree seeds in the Inland Mountain west. USDA Forest Service General Technical Report INT-203, 1986.
- [WR05] P. Weber and B. Rosenow. Order book approach to price impact. *Quantitative Finance*, 5:357–364, 2005.
- [WTH01] M. E. J. Woolhouse, L. H. Taylor, and D. T. Haydon. Population biology of multihost pathogens. *Science*, 292:1109–1112, 2001.
- [YdM05] S.H. Yook and M.A. de Menezes. Fluctuation of incoming flux with multiplicative noise on a scale-free network. *Europhys. Lett.*, 72:541–547, 2005.

- [YI05] A. Yuen and P.Ch. Ivanov. Impact of stock market microstructure on intertrade time and price dynamics. physics/0508203, 2005.
- [ZAK04] A.G. Zawadowski, Gy. Andor, and J. Kertész. Large price changes on small scales. *Physica A*, 344:221–226, 2004.
- [ZAK06] Á. G. Zawadowski, Gy. Andor, and J. Kertész. Short-term market reaction after extreme price changes of liquid stocks. *Quantitative Finance*, 6(4):283–295, 2006.
- [Zip29] G. K. Zipf. Relative frequency as a determinant of phonetic change. *Harvard Studies in Classical Philology*, 15:1–95, 1929.
- [ZQJZ07] L. Guo Z.-Q. Jiang and W.-X. Zhou. Endogenous and exogenous dynamics in the fluctuations of capital fluxes: An empirical analysis of the chinese stock market. *Eur. Phys. J. B*, 57:347–355, 2007.
- [ZTWT06] J. Zivković, B. Tadić, N. Wick, and S. Thurner. Statistical indicators of collective behavior and functional clusters in gene networks of yeast. *Eur. Phys. J. B*, 50:255–258, 2006.
- [Zum04] G. Zumbach. How trading activity scales with company size in the FTSE 100. *Quantitative Finance*, 4:441–456, 2004.