Quantifying and Modeling Financial Market Fluctuations

DISSERTATION

zur Erlangung des Grades

„Doktor der Naturwissenschaften“

am Fachbereich Physik, Mathematik und Informatik

der Johannes Gutenberg-Universität Mainz

vorgelegt von

Tobias Preis

geboren am 23. Juli 1981 in Limburg an der Lahn

Mainz, im Juni 2010

Tag der mündlichen Prüfung: 28. Oktober 2010

Tobias Preis

Quantifizierung und Modellierung von Finanzmarktfluktuationen

— Zusammenfassung —

Die Entstehung eines Marktpreises für einen Vermögenswert kann als Superposition der einzelnen Aktionen der Marktteilnehmer aufgefasst werden, die damit kumulativ Angebot und Nachfrage erzeugen. Dies ist in der statistischen Physik mit der Entstehung makroskopischer Eigenschaften vergleichbar, die von mikroskopischen Wechselwirkungen zwischen den beteiligten Systemkomponenten hervorgerufen werden. Die Verteilung der Preisänderungen an Finanzmärkten unterscheidet sich deutlich von einer Gaußverteilung. Dies führt zu empirischen Besonderheiten des Preisprozesses, zu denen neben dem Skalierungsverhalten nicht-triviale Korrelationsfunktionen und zeitlich gehäufte Volatilität zählen. In der vorliegenden Arbeit liegt der Fokus auf der Analyse von Finanzmarktzeitreihen und den darin enthaltenen Korrelationen. Es wird ein neues Verfahren zur Quantifizierung von Muster-basierten komplexen Korrelationen einer Zeitreihe entwickelt. Mit dieser Methodik werden signifikante Anzeichen dafür gefunden, dass sich typische Verhaltensmuster von Finanzmarktteilnehmern auf kurzen Zeitskalen manifestieren, dass also die Reaktion auf einen gegebenen Preisverlauf nicht rein zufällig ist, sondern vielmehr ähnliche Preisverläufe auch ähnliche Reaktionen hervorrufen. Ausgehend von der Untersuchung der komplexen Korrelationen in Finanzmarktzeitreihen wird die Frage behandelt, welche Eigenschaften sich beim Wechsel von einem positiven Trend zu einem negativen Trend verändern. Eine empirische Quantifizierung mittels Reskalierung liefert das Resultat, dass unabhängig von der betrachteten Zeitskala neue Preisextrema mit einem Anstieg des Transaktionsvolumens und einer Reduktion der Zeitintervalle zwischen Transaktionen einhergehen. Diese Abhängigkeiten weisen Charakteristika auf, die man auch in anderen komplexen Systemen in der Natur und speziell in physikalischen Systemen vorfindet. Über 9 Größenordnungen in der Zeit sind diese Eigenschaften auch unabhängig vom analysierten Markt — Trends, die nur für Sekunden bestehen, zeigen die gleiche Charakteristik wie Trends auf Zeitskalen von Monaten. Dies eröffnet die Möglichkeit, mehr über Finanzmarktblasen und deren Zusammenbrüche zu lernen, da Trends auf kleinen Zeitskalen viel häufiger auftreten. Zusätzlich wird eine Monte Carlo-basierte Simulation des Finanzmarktes analysiert und erweitert, um die empirischen Eigenschaften zu reproduzieren und Einblicke in deren Ursachen zu erhalten, die zum einen in der Finanzmarktmikrostruktur und andererseits in der Risikoaversion der Handelsteilnehmer zu suchen sind. Für die rechenzeitintensiven Verfahren kann mittels Parallelisierung auf einer Graphikkartenarchitektur eine deutliche Rechenzeitreduktion erreicht werden. Um das weite Spektrum an Einsatzbereichen von Graphikkarten zu aufzuzeigen, wird auch ein Standardmodell der statistischen Physik — das Ising-Modell — auf die Graphikkarte mit signifikanten Laufzeitvorteilen portiert. Teilresultate der Arbeit sind publiziert in [PGPS07, PPS08, Pre11, PVPS09b, PVPS09a, PS09, PS10a, SBF$^+$10, BVP10, Pre10, PS10b, PSS10, SBF$^+$11, PB10].

Tobias Preis

Quantifying and Modeling Financial Market Fluctuations

— Abstract —

The formation of market prices for financial assets can be understood as superposition of individual actions of market participants, in which they provide cumulative supply and demand. This concept of macroscopic properties emerging from microscopic interactions among the various subcomponents of the overall system is also well-known in statistical physics. The distribution of price changes in financial markets is clearly non-Gaussian leading to distinct features of the price process, such as scaling behavior, non-trivial correlation functions and clustered volatility.

This thesis focuses on the analysis of financial time series and their correlations. A new method is introduced for quantifying pattern based correlations of a time series. With this methodology, evidence is found that typical behavioral patterns of financial market participants manifest over short time scales, i.e., that reactions to given price patterns are not entirely random, but that similar price patterns also cause similar reactions.

Based on the investigation of the complex correlations in financial time series, the question arises, which properties change when switching from a positive trend to a negative trend. An empirical quantification by rescaling provides the result that new price extrema coincide with a significant increase in transaction volume and a significant decrease in the length of corresponding time intervals between transactions. These findings are independent of the time scale over 9 orders of magnitude, and they exhibit characteristics which one can also find in other complex systems in nature (and in physical systems in particular). These properties are independent of the markets analyzed. Trends that exist only for a few seconds show the same characteristics as trends on time scales of several months. Thus, it is possible to study financial bubbles and their collapses in more detail, because trend switching processes occur with higher frequency on small time scales.

In addition, a Monte Carlo based simulation of financial markets is analyzed and extended in order to reproduce empirical features and to gain insight into their causes. These causes include both financial market microstructure and the risk aversion of market participants. The computationally expensive analysis procedures employed in these investigations are coded on a graphics card architecture which leads to a significant reduction of computing time. In order to demonstrate the wide range of possible applications, a standard model in statistical physics—the Ising model—is ported to graphics card architectures as well, resulting in large speedup values. Some of the results presented in this thesis have already been published in [PGPS07, PPS08, Pre11, PVPS09b, PVPS09a, PS09, PS10a, SBF$^+$10, BVP10, Pre10, PS10b, PSS10, SBF$^+$11, PB10].

Contents

List of Figures

List of Tables

Chapter 1

General introduction

This doctoral thesis focuses on the analysis of financial market data and the modeling of these markets from a physical point of view. Thus, the main issues addressed lie within the interdisciplinary research field *econophysics*. The first section of this chapter provides insight into this field. Section 1.2 motivates the questions which will be covered by this doctoral thesis. Included in these sections, selected models and concepts of econophysics are introduced in a nutshell. A structural overview of the remaining chapters concludes this general introduction.

1.1 Econophysics — Economy and Physics

As a result of the current financial crisis, which is still persisting, researchers have been further motivated to achieve an understanding of the formation of economic bubbles. The consequences of such a global crisis, which exceeds in intensity all recessions of recent decades, are not limited to privately held companies and individuals – members of the international community of nations have also struggled for their survival. For example, in the first quarter of 2010, it was revealed that Greece is one prominent example of the group of financially tarnished nations.

However, the real economy is also still suffering the consequences of this international financial crisis, which started as a real estate and mortgage market crisis in the United States of America in 2007. Originally focused on credit defaults in the U.S. housing market—the so-called *sub-prime* market— the crisis spread out, influencing the liquidity of interbank lending and thus affecting the money market. The high degree of propagation was caused by the fact that credit derivatives—introduced by banking institutions in order to manage and distribute credit risk—created a non-transparent world-wide inter-connected credit network. All phases of the crisis coincided with sometimes dramatic reactions of stock markets.

Thus, it is not very difficult to motivate the need for realistic financial market models. In the current crisis, the climax was reached with the falling out of five major U.S. investment banks in the third week of September 2008. Facing such structural changes,

anyone can perceive intuitively that coinciding movements of stocks in financial markets were definitely not solely the result of random fluctuations. For example, it is obvious that the strength of price fluctuations are clustered in time which is known as volatility clustering: Trading days with large price fluctuations are followed by days displaying again increased price changes. This dramatic effect could be observed in the impressive movements of international stock indices at the end of 2008.

Nevertheless, it has been traditionally assumed in economics that price dynamics in financial markets provide random walk statistics, which means that prices evolve completely randomly. Economists relegate impact events such as the current financial market crisis to the dustbin category of *outliers*.

Physicists do not like to do things this way. Physicists do not take Newton's law seriously part of the time, and then—if we suddenly see an example of what appears to be levitation—simply call it an *outlier*. They like to find laws that describe all examples of a phenomenon. Economists themselves, in a journal called *The Economist*, have admitted failure. This is a strong motivation for physicists to step in and to meet this scientific challenge. Also, practically speaking, catastrophic economic events have extreme social impacts. Widespread suffering is the usual outcome, especially among the poor. The ability to understand economic crashes and other large-scale risks would be of obvious utility to policy and research.

An additional reason why the economy is of interest to statistical physicists is that—like an Ising model, a model of ferromagnetism—it is a system made up of many subunits. The subunits in an Ising model are the interacting spins, and the subunits in the economy are market participants—buyers and sellers. During any time interval these subunits of the economy may be either positive or negative with respect to perceived market opportunities. People interact with each other, and this fact often produces what economists call the *herd effect*. The orientation of whether they buy or sell is influenced not only by neighbors but also by news. If we hear bad news, we may be tempted to sell. So the state of any subunit is a function of the states of all the other subunits and of a field parameter [PS10a].

One very illustrative example of the *herd effect* is shown in Fig. 1.1. The search engine Google offers the possibility of accessing information on how popular specific search terms are[1]. Thus, one can compare the interest in financial crisis related keywords such as "Subprime", "Lehman Brothers", and "Financial Crisis" with the fluctuations of the S&P 500 index which is an international benchmark index. It is easy to see that peaks in the search volume for the term "Subprime" coincide with dips in the S&P 500 time series. At the climax of the crisis, the collapse of "Lehman Brothers" caused the sell-out of stocks and the public was talking about the "Financial Crisis" afterwards. Figure 1.1 documents this over the course of time and shows that these activities steadily increased. The search volume profiles track the levels of escalation which can be seen as a prominent example of the *herd effect*.

These introductory words show that it is justified to address *econophysics* in the

[1]More details can be found on `http://www.google.com/trends`.

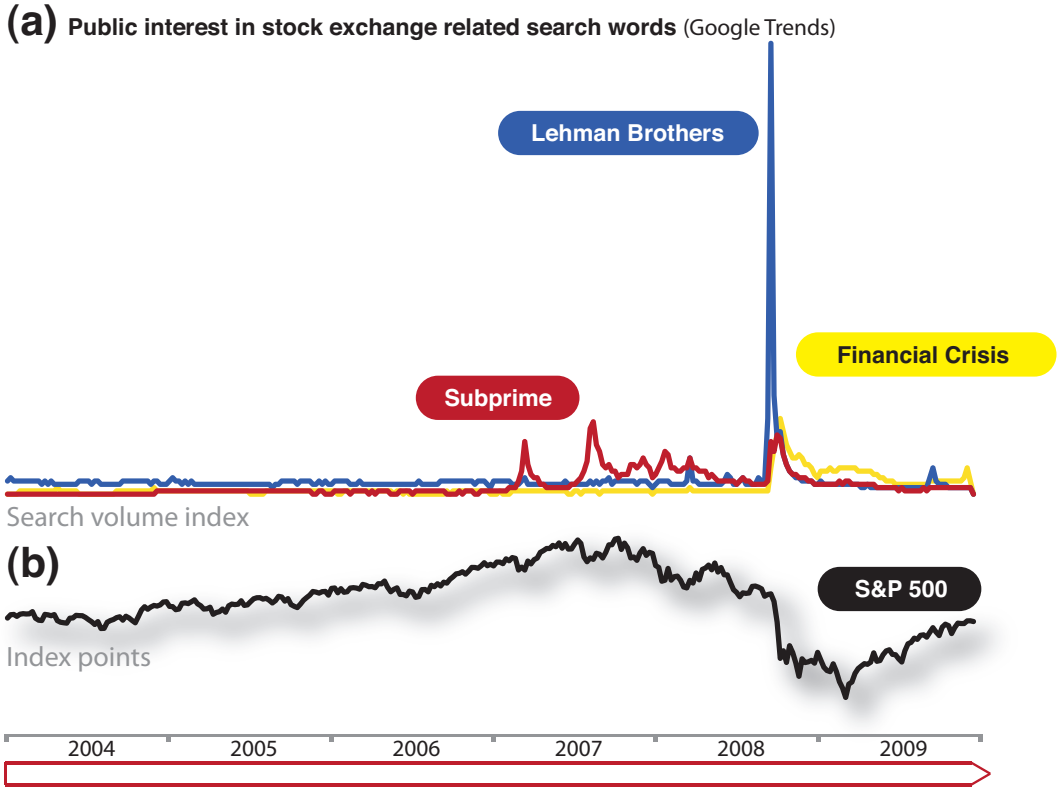

Figure 1.1: Keyword related search volume illustrates a behavior similar to a *Herd effect*: (a) Google Trends analyzes a portion of web searches to compute how many searches have been done for specified keyword terms [Goo10]. The relative number of searches can be plotted over time—here, the keyword terms "Subprime", "Lehman Brothers", and "Financial Crisis" are plotted for the time period from 2004 to 2009. The peak of "Lehman Brothers" coincide with the bankruptcy of this institution when the investment bank Lehman Brothers filed for Chapter 11 bankruptcy protection. (b) The U.S. stock index S&P 500 is shown for the same period of time.

framework of a doctoral thesis in physics. *Econophysics* or "physics of financial markets" forms the interdisciplinary interface between the two disciplines economics[2] and physics[3] both of which stem from Ancient Greek. The term *econophysics* was coined by H. E. Stanley[4] in the 1990s.

In order to provide further justify a doctoral thesis in the field of *econophysics*—which should be more elaborate than G. M. Obermair's[5] statement that physics is defined by that what physicists do—we sketch an overview of the history of *econophysics* and present overlaps between physics and economics.

The "experimental basis" of this interdisciplinary science *econophysics* is given by time series which can be used in their raw form or from which one can derive observables. Time series can be considered to be a link between economic and financial processes on the one hand and empirical analysis and physics based modeling on the other hand. Such historical price curves can be understood as a macroscopic variable for underlying microscopic processes. The price fluctuations are produced by the superposition of individual actions of market participants, thereby generating cumulative supply and demand for a traded asset—e.g. a stock. The analog in statistical physics is the emergence of macroscopic properties, which are caused by microscopic interactions among involved subunits as described earlier in this section.

A few decades ago, M. F. M. Osborne[6] [Osb95] and B. B. Mandelbrot[7] [Man63] analyzed time series related to financial markets realizing the non-Gaussian shape of the price change distributions [KSY06]—a Gaussian distribution is still an often used assumption for economic models. With these pioneering findings, they became early fathers of *econophysics*. Mandelbrot used for his analysis historical cotton times and sales records dating back to the beginning of the 20$^{\text{th}}$ century. His findings indicated that financial market time series obey a complex statistical behavior which has similarities with the non-Gaussian properties of critical fluctuations in physics [BP09, MS00, PB00]. The studies of Mandelbrot were based on data sets of very limited length. Alongside technological progress in information technology, trading processes in international financial markets adapted and fully electronic trading platforms have been established. Thus, a gargantuan amount of historical financial time series are available for

[2] *Ancient Greek*: οἰκονομία — housekeeping, management.

[3] *Ancient Greek*: φυσική τέχνη — art of handling the nature.

[4] **Harry Eugene Stanley** (*28 March 1941 in Oklahoma City, USA): an American physicist and University Professor at Boston University. He has made seminal contributions to statistical physics and was one of the pioneers of interdisciplinary science [Fou10].

[5] **Gustav M. Obermair**: German physicist and founding rector of the University of Regensburg who mainly investigated the fundamentals of quantum mechanics, quantum chaos, and non-linear dynamics but also worked on interdisciplinary problems [BBO$^+$01].

[6] **Maury F. M. Osborne**: Physicist. His publication "Brownian Motion in the Stock Market" [Osb59] marks the starting point in literature about the geometric Brownian motion in a financial mathematics context.

[7] **Benoît B. Mandelbrot** (*20 November 1924 in Warsaw, Poland): French and American mathematician who is known as the father of fractal geometry. He applies fractal geometry to various scientific fields [Fou10].

researchers and practitioners (see chapter 6) with an extremely high time resolution[8]. A century ago, only daily data were accessible, whereby the situation today is such that transaction records are available on time scales of milliseconds. The continuously increasing number of transaction records confirm the results obtained by Mandelbrot. Furthermore, the use of Gaussian approximations in finance has become more and more questionable.

However, the first work in this area which can be related to current *econophysics* was performed much earlier as reported for example in [PB00]. In 1900, a young PhD student of H. Poincaré[9] finished his thesis. His name was L. Bachelier[10]. On the one hand, Poincaré lauded the brilliant way in which the Gaussian distribution was derived. On the other hand, he pointed out that the subject of his thesis entitled "Théorie de la spéculation" [Bac00] was completely different from subjects of other theses. Bachelier proposed the *random walk* as model for price fluctuations in his PhD thesis. Thus, he developed the mathematics of Brownian motion with his description of financial market processes—five years before the famous publication of A. Einstein[11] on Brownian motion appeared in 1905 [Ein05]. Until the early 1940s Bachelier's work tended to be sidelined. But living in the shadows is not exceptional for revolutionary scientific contributions as exemplified by the famous publication by E. Ising[12] [Isi25] about a model which has become a standard model in statistical physics today—the Ising model. W. Heisenberg[13] cited this work in 1928. The breakthrough of the Ising model coincided with the exact solution for the Ising model on a two-dimensional square lattice achieved by L. Onsager[14] in 1944 [Ons44]. The Ising model can not only be used to describe ferromagnetism as originally intended; it can be extended to also describe more complex magnetic systems, such as spin glasses, in which competing interac-

[8] In August 2009, the Deutsche Börse Group reduced the network latency between Frankfurt and London to below 5 milliseconds, setting new standards in the industry [Deu10].

[9] **Jules Henri Poincaré** (*29 April 1854 in Nancy, France; †17 July 1912 in Paris, France): French mathematician, theoretical physicist, and a philosopher of science [Fou10].

[10] **Louis Jean-Baptiste Alphonse Bachelier** (*11 March 1870 in Le Havre, France; †26 April 1946 in St-Servan-sur-Mer, France): French mathematician who is credited with being the first person to model the stochastic process which is now called Brownian motion [Fou10].

[11] **Albert Einstein** (*14 March 1879 in Ulm, Germany; †18 April 1955 in Princeton, USA): German-born Swiss-American theoretical physicist, philosopher and author who is widely regarded as one of the most influential and best known scientists and intellectuals of all time. He is often regarded as the father of modern physics [Fou10].

[12] **Ernst Ising** (*10 May 1900 in Cologne, Germany; †11 May 1998 in Peoria, USA): German physicist and mathematician who is best remembered for the development of the Ising model [Fou10]. The Ising model describes the transition of an unordered paramagnetic material at high temperatures to a ferromagnetic material below a critical temperature—the Curie temperature T_C.

[13] **Werner Heisenberg** (*5 December 1901 in Würzburg, Germany; †1 February 1976 in Munich, Germany): Norwegian-born American physical chemist and theoretical physicist. He obtained the 1968 Nobel Prize in Chemistry [Fou10].

[14] **Lars Onsager** (*27 November 1903 in Kristiania, Norway; †5 October 1976 in Coral Gables, USA): Norwegian-born American physical chemist and theoretical physicist. He obtained the 1968 Nobel Prize in Chemistry [Fou10].

tions between various spins lead to frustration effects and slow ("glassy") dynamics. The energy landscapes of such systems exhibit a vast amount of local minima. The task is to find those spin configurations for which the energy is minimal. The inverse problem, namely the calculation of interactions for the proposed spin configurations which should form local minima in the energy landscape, is known as the Hopfield model for neural networks[15]. Today, the Ising model is used for various—ordered and disordered—complex systems in many areas of science. The various applications also include the interdisciplinary field of sociophysics [MdOdOS99] in which social interactions are treated by a modification of the Ising model—the Sznajd model [SWS00]. In this way, it is possible to investigate, for example, the stability of democracy in the German Federal State of Bavaria [SH05a] based on membership levels of the dominating political parties. Many further interdisciplinary streams of research in which physics' concepts and methods are used have appeared. One example is the physics of transport and pedestrians. An often asked question is how spontaneous traffic jams occur [NS92, Hel97]. Beside D. Helbing[16], who is one of the leading researchers in this field, one prominent scientist in this field of research is D. Brockmann[17]. His research on complex systems spans from transportation networks to infectious diseases. He has developed computational models, new analytic and numerical techniques, and large-scale quantitative and predictive computer simulations for studying various aspects of the dynamics of epidemics. For example, he has used data from `www.wheresgeorge.com`—a website where users enter the serial numbers of their dollar bills—in order to track their travels [BHG06]. In this way, patterns and laws of human mobility could be identified. From that information, Brockmann was able to reconstruct a multi-scale human mobility network of the U.S. including small scale daily commuter traffic, intermediate traffic, and long distance air travel. Based on this mobility network, Brockmann has modeled how diseases spread throughout the country. He and his research group have also created a map of large scale community boundaries in the United States, which sometimes differs from administrative boundaries. These effective maps show that some states, like Missouri or Pennsylvania, are essentially cut into halves. Other boundaries coincide with geographic features, such as the Appalachian Mountains. Brockmann also develops models for disease spreading via transportation networks, in order to quantify the susceptibility of various regions to epidemic threats and to develop more efficient containment strategies.

A field which is in some sense related to *econophysics* is the research area optimiza-

[15] In a spin glass, the interactions between individual spins are fixed. It is the aim—and sometimes a challenge—to find the ground state in the energy landscape. In a Hopfield network the configurations, which correspond to the ground states, are given and one has to find the interactions [SK06].

[16] **Dirk Helbing**: professor of sociology, in particular of modeling and simulation, at ETH Zurich since 2007 and external professor of the Santa Fe Institute since 2009.

[17] **Dirk Brockmann**: associate professor of engineering sciences and applied mathematics at the McCormick school of engineering and applied science at Northwestern University. He received the Young Scientist Award 2010 for socio- and econophysics from the physics of socio-economic systems division of the German Physical Society.

tion, especially stochastic optimization [SK06]. Recently, the German researcher Johannes J. Schneider located at the Johannes Gutenberg University Mainz was able to beat world records for a packing problem [MSS09]. They considered the problem of finding the densest closed packing of hard disks with proposed different radii in a circular environment, such that the radius of the circumcircle is minimal. With their approach, they were able to find denser packings for various problem instances than found in an international contest, in which 155 groups from 32 countries competed against each other. The work was recognized by *Time Magazine* [Tim09] to be one of "The 50 Best Inventions of 2009".

Bachelier did not gain such publicity in the first years after his work was published. Applying his ideas to describe French bonds took significant time. At that time, his interdisciplinary work was not popular in the scientific community at all—today, there actually exists a "Bachelier Finance Society"[18]. However, in the meantime it turned out that the work of A. Kolmogorov[19] was inspired by Bachelier's results as can be retraced in [CKB+00]. In 1944—almost half a century later—Itô[20] used Bachelier's PhD thesis as motivation for his famous calculus, now called Itô calculus. Later, Osborne introduced geometric Brownian motion which is an extension of Brownian motion. It is a continuous-time stochastic process in which the logarithm of the randomly varying quantity follows a Brownian motion. This is advantageous in financial market modeling in that it can only take positive values—a meaningful assumption for the prices of stocks. This compatibility with the axioms of economics was shown by P. A. Samuelson[21] for which he obtained the 1970 Nobel Prize in Economics. Asset prices could be assumed to be log-normally distributed. This implied normally distributed returns.

Based on that, F. Black[22] and M. Scholes [BS73]—as well as R. Merton[23] [Mer73] in-

[18] More information can be found on `www.bachelierfinance.org`.

[19] **Andrey Nikolaevich Kolmogorov** (*25 April 1903 in Tambow, Russia; †20 October 1987 in Moscow, Russia): one of the most famous mathematicians of the 20th century [Fou10].

[20] **Itô Kiyoshi** (*7 September 1915 in Hokusei-chô, Japan; †10 November 2008 in Kyôto, Japan): a Japanese mathematician whose work is now called Itô calculus. It facilitates mathematical understanding of random events. His theory is widely applied, for instance in financial mathematics [Fou10].

[21] **Paul Anthony Samuelson** (*15 May 1915 in Gary, USA; †13 December 2009 in Belmont, USA): an American economist, and the first American to win the Nobel Prize in Economics. The Swedish Royal Academies stated, when awarding the prize, that he "has done more than any other contemporary economist to raise the level of scientific analysis in economic theory" [Fou10].

[22] **Fischer Sheffey Black** (*11 January 1938 in Washington, USA; †30 August 1995 in New York, USA): an American economist, best known as one of the authors of the famous *Black-Scholes equation* [Fou10].

[23] **Myron Samuel Scholes** (*1 July 1941 in Timmins, Canada) and **Robert Cox Merton** (*31 July 1944 in New York, USA) were among the board of directors of Long-Term Capital Management (LTCM), a hedge fund that failed spectacularly in 1998 after losing USD 4.6 billion in less than four months. The Federal Reserve was so concerned about the potential impact of LTCM's failure on the financial system that Alan Greenspan arranged for a group of 19 banks and other firms to provide sufficient liquidity for the banking system to survive. Although these investors were eventually paid off, the reputations of Merton and Scholes were tarnished. LTCM used trading strategies such as *fixed income arbitrage*, *statistical arbitrage*, and *pairs trading*, combined with high leverage. *Arbitrage* is the practice of taking advantage of a price difference between two or more markets[Fou10].

dependently of both of them—developed a theory for evaluating stock options using geometric Brownian motion. This theory had a huge impact on financial markets. Up to that time, it had not been possible to obtain an "objective" valuation of options. Thus, the transaction numbers and transaction volumes of option contracts increased rapidly after the introduction of this theory. Today, modern *risk management* is inconceivable without options. In 1997, the Nobel Prize in Economics was awarded to Scholes and Merton for their work. However, it must also be mentioned that this theory, now known as the Black Scholes model, is accompanied by problematic aspects. Continuous time trading is a first assumption which is matched most closely by foreign exchange markets[24] but which is not fulfilled by other exchange platforms. The second requirement is a continuous price path which is unrealistic—this is already contradicted by the existence of opening gaps. An opening gap is a direct consequence of non-continuous trading. At the beginning of an exchange trading day a stock does not necessarily start being traded at the same price level as the day before. The difference between opening price and closing price one day before is known as the opening gap. If there is for example an important news release in the period of time without trading, larger opening gaps occur. In addition to these two assumptions, it is critical to note that the important input variable, the volatility (annualized standard deviation) of future stock returns, is not known in advance. Furthermore, the Black Scholes model is based on normally distributed asset returns. This aspect contradicts earlier findings mentioned in this section: thus, economics uses simplifying assumptions which were already falsified by Mandelbrot in 1963. Several papers confirm his findings and, most importantly, Stanley et al. were able to quantify the tails of the return distributions. They found that the tails are consistent with cubic power laws [MS00].

In this context, physicists used a wide range of methods in trying to examine the discrepancy between theory and the reality of economic and financial markets. They applied physical methods to the "many particle system" made up by financial markets. Non-trivial features of financial market time series are called *empirical stylized facts*. The focus lies on these facts, and they are shortly mentioned in chapter 3. Furthermore, the extraction of additional *empirical stylized facts* is one main aspect of this thesis. It is on the basis of commonly accepted facts that physicists started to model financial markets. Simple models were proposed as a result, for example at the end of the last century [BPS97, MS00]. A simple agent-based model[25] [BPS97] uses imitation and feedback. It

[24]The foreign exchange market (*FOREX*) is a worldwide decentralized over-the-counter financial market for the trading of currency pairs. There is a continuous trading established—an exception marks the weekend. The average daily trading volume of this non-regulated, global market was, e.g., roughly USD 1.9 trillion in April 2004 [BB04].

[25]This early contribution of physicists of modeling financial markets using agent based models is based on a reaction diffusion process [BPS97], $A + B \rightarrow 0$. It is able to reproduce some *empirical stylized facts*—e.g., non-trivial Hurst exponents which will be introduced in chapter 3 and non-Gaussian price change distributions. Large price changes occur more frequently than predicted by the Gaussian distribution. A few years later, a model was published [Mas00, MM01] incorporating two different order types—limit orders and market orders (see chapter 2). A model with an additional rule set was developed in 2003 [CS03]. There, a Poisson process is applied for the cancellation of submitted limit orders. In recent years,

is able to reproduce simple examples of the group of *empirical stylized facts*. In recent years, physicists started to investigate and understand the price formation process in detail on a microscopic level as simple models were not able to completely reproduce the behavior of financial markets. In this context, the statistical model of the continuous double auction [DFG+03, SFGK03] stand out. The continuous double auction is used for organizing trading processes at electronic financial exchanges and will be described in detail in chapter 2.

1.2 Motivation for this Doctoral Thesis

In physics and in other natural sciences, it is often a successful strategy to analyze the behavior of a system by studying the smallest components of that system. For example, the molecule is composed of atoms, the atom consists of a nucleus and electrons, the nucleus consists of protons and neutrons, and so on. The fascinating point about analyses on steadily decreasing time and length scales is that one often finds that the composite system exhibits properties which cannot solely be explained by the properties of its components alone. Instead, a complex behavior can emerge due to the interactions among these components [And72]. In financial markets, these components are comprised of the market participants who buy and sell assets—as emphasized in the previous section—in order to realize their trading and investment decisions. The superimposed flow of all individual orders submitted to the exchange trading system by market participants, and, of course, its change in time generate a complex system with fascinating properties, similar to physical systems.

One of the key conceptual elements in modern statistical physics is the concept of scale invariance, codified in the scaling hypothesis that functions obey specific functional equations whose solutions are power-laws [Sta71]. The scaling hypothesis has two categories of predictions, both of which have been remarkably well verified by a wealth of experimental data on diverse systems. The first category is a set of relations, called *scaling laws*, that relate the various critical-point exponents characterizing the singular behavior of functions such as thermodynamic functions. The second category is a sort of *data collapse*, where under appropriate axis normalization, diverse data "collapse" onto a single curve called a scaling function.

Econophysics research has been addressing a key question of interest: quantifying and understanding large stock market fluctuations. Previous work has focused on the challenge of quantifying the behavior of the probability distributions of large fluctuations of relevant variables such as returns, volumes, and the number of transactions. Sampling the far tails of such distributions require a large amount of data. A very large amount of precise historical financial market data has already been collected, many orders of magnitude more than for other complex systems. Accordingly, increasing

further agent based market models [SS99, LM99, KHH02, CS03, CMZ97] have been proposed in order to reproduce further *empirical stylized facts*.

numbers of scientists have been analyzing financial market data [Axt01, Tak06, KSY06, WTT07, GGPS03, PPS08, LFM03, PGGS02, CB00, KHH02, O'H97, VA97, EK06].

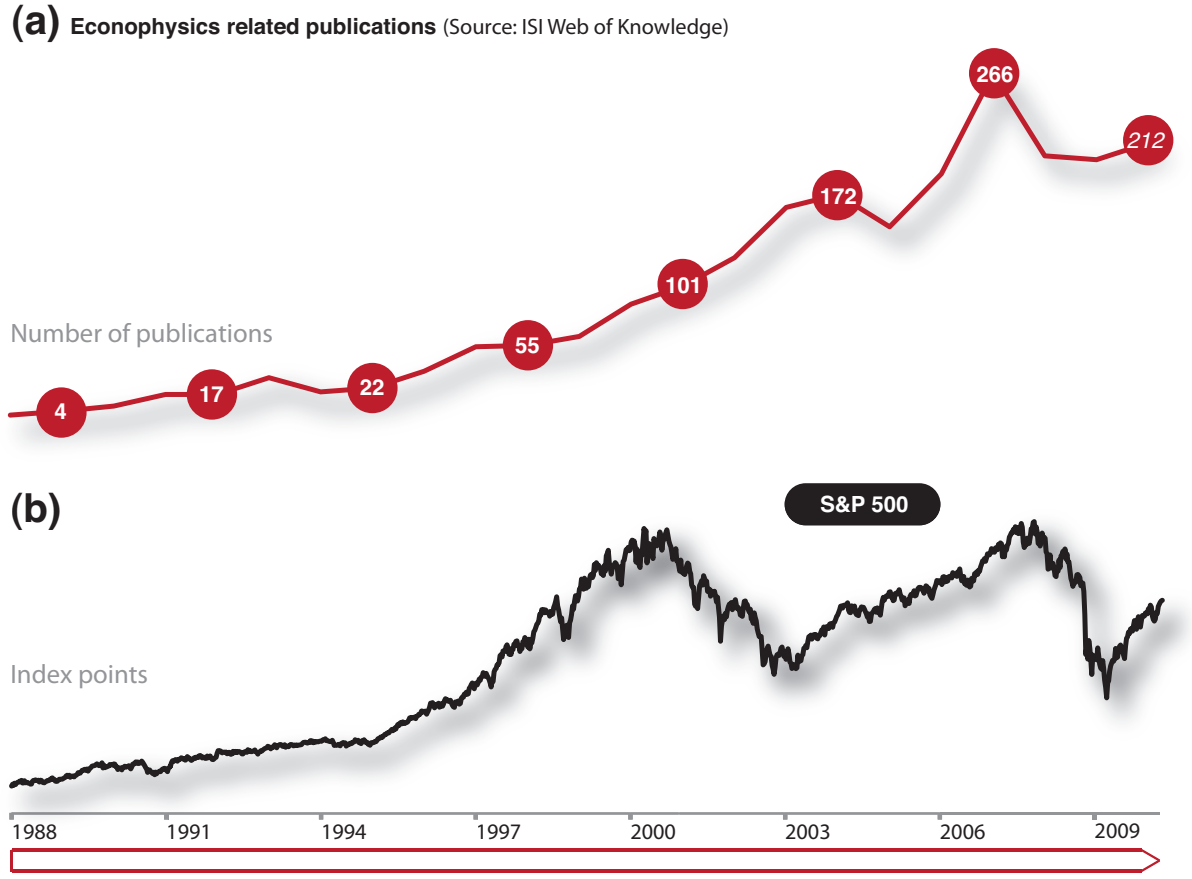

Figure 1.2: Number of publications related to *econophysics*: (a) Number of publications which fit the topics *econophysics*, *financial markets*, or *economy* in scientific journals aggregated yearly. (b) U.S. benchmark index S&P 500 plotted for the same period of time.

The increasing interest of the physics community in topics related to financial markets caused an increasing number of publications in physics journals which are related to *econophysics* as shown in Fig. 1.2. The results were limited to journals belonging to the areas of *physics*, *computer science*, and *mathematics*. The data were extracted from the *ISI Web of Knowledge* database. The number of publications for 2010—which is italicized—is estimated based on the number of publications for the first quarter of 2010. The increasing number of econophysics publication in the last two decades is linked to the increase in value of international benchmark indices—here, the S&P 500 is shown. The downfall of the *New Economy* after the year 2000 and the recent international financial market crisis had no dramatic impact on the number of publications.

Empirical analyses have focused on quantifying and testing the robustness of power-

law distributions that characterize large movements in stock market activity. Using estimators that are designed for serially and cross-sectionally independent data, findings thus far support the hypothesis that the power-law exponents that characterize fluctuations in stock price, trading volume, and the number of trades [Fam63, Lux96, GDD+97, PGR+99] are seemingly "universal" in the sense that they do not change their values significantly for different markets, different time periods, or different market conditions.

In contrast to these analyses of global financial market distributions, we focus on the temporal sequence of fluctuations in transaction volume and inter-trade times before and after a trend switching point. Our analysis can provide insight into switching processes in complex systems in general and financial systems in particular. The study of dramatic crash events is limited by the fortunately small number of such events. Currently, one may seek to better understand the current financial crisis by comparing it with the depression of the 1930's. Here we ask if the smaller financial crises—trend switching processes on all time scales—also provide information of relevance for large crises. If this is the case, then the large abundance of data on smaller crises should provide quantifiable statistical laws for *bubbles on all scales*. This fact alone provides the greatest motivation for a PhD thesis in the interdisciplinary field of econophysics in seeking to contribute to a better understanding of financial markets and to financial market crises in particular.

This PhD thesis continues a work which was started as a diploma thesis [Pre06] in 2005. There, the starting point was the microscopic statistical model [DFG+03, SFGK03] of the continuous double auction. In this model, transaction orders—limit orders and market orders (see chapter 2 for a detailed explanation of the market structure)—enter the system with constant rates uniformly distributed over an infinite price range. The simplified assumptions of the model are not realistic but necessary for solving the system in special cases. Based on this mechanism of order submission, a simple multi-agent based system was defined in the aforementioned diploma thesis [Pre06] which led to a publication in *Europhysics Letters* in 2006 [PGPS06]. Here, two different types of agents interact with each other. These agents can be identified as *liquidity providers* and *liquidity takers*. The aim of the agent-based Monte Carlo simulation[26] of financial markets is to analyze the properties of such a system in order to obtain a deeper and better understanding of the pricing processes in financial markets.

Certain aspects should not be neglected, however, concerning the general motivation for developing models for financial markets and for analyzing the corresponding data at this point. Therefore, it is necessary to introduce the fundamental concept of an *efficient market* which is an idealization. This means that the price process of financial assets can be modeled by a Markov chain stochastic process [PB00]. A Markov process is characterized by the property that the probability of reaching an arbitrary state at time $t+1$ depends only on the state of the system at time t. The state or states in which the system was in before time t are not relevant [Kre03]. A market is called *effective* if various criteria are fulfilled. Another important aspect is information. Market participants

[26] A Monte Carlo method is an algorithm that uses random numbers.

have to have instantaneous access to related information. The market has to provide a certain amount of liquidity, so that trading positions can be opened or closed at any time. Furthermore, the assumption of a liquid market includes the assumption that no price impact can be induced by a single order. A further criterion for an efficient market is low *market friction*. *Market friction* is a collective term for the effective trading costs arising from exchange rates, transaction costs, taxes, and the *bid ask spread*[27]. If the sum of all costs is negligibly small compared to the transaction volume then the market provides low friction. In the case that a market exhibits these properties then the *efficient market hypothesis* states that new information is efficiently incorporated into the market. This means that all relevant information is included in the current market price at any time. Assuming that this hypothesis is valid, it makes no sense to analyze historical time series in order to extract probabilities for future market movements.

This aspect can also be address though a brief excursion to the topic of *statistical arbitrage*. Let us assume that there is a significant correlation between the current market behavior and historical time series associated with the market [28]. This can lead to a trading strategy with positive expected value. If a market participant is aware of this pattern then it is possible to exploit the statistical pattern by statistical arbitrage. The term *arbitrage* in its strict sense can be described as follows: "Pure arbitrage is defined as generating risk-less profit today by statically or dynamically matching current and future obligations to exactly offset each, inclusive of incurring known financing cost" [Dub04]. A simple and often used example is a stock which is traded at two exchanges located in different places. If the bid price p_1 for selling at one exchange is higher than the ask price p_2 for buying at another exchange then one can realize a risk-less profit buying at p_2 and selling at p_1. However, *statistical arbitrage* is focused on profitable exploitation of correlations, i.e., there is a statistical mispricing of one or more assets based on the expected value of these assets. This kind of correlation disappears when market participants start to use this information and as this information spreads towards an increasing number of market participants. In an almost efficient market, such correlations do not occur, or they disappear instantaneously through an adaption of demand and offer price. In less efficient markets, in which the adaptation time to new information is finite and different from zero, correlations can be reduced or eliminated as they are exploited. The more liquid a market is, the less correlation can be observed. However, as there is a certain amount of market friction, correlations do not disappear completely. If the advantage is comparable to the level of market friction, then no positive expected value can be realized. This fact explains the autocorrelation coefficients of financial market time series on very small time scales which are significantly different from zero [Con01].

This brief discussion of efficient markets—on which financial valuation models rely—

[27] The bid ask spread is the non-zero gap between the lowest offering price and the highest demanding price in an order book, which stores all orders submitted by market participants. If a *liquidity taker* would like to sell using a market order, then an execution price at the highest demanding price is realized. Thus, the effective transaction fees are increased by the *bid ask spread*.

[28] This kind of pattern based correlations will be analyzed in detail in chapter 4.

makes clear the motivation (and fascination) in dealing with financial market topics from a physicist's point of view. One must be careful, however, not to succumb to the misconception that econophysics research can predict future market prices. Even a reasonable quantification of future risk is a major challenge as one can see in the current international financial market crisis.

1.3 Structure of the Thesis

The chapters of this thesis are organized as follows. Chapter 2 covers general topics concerning financial markets. Historical development is described along with an overview of how financial markets work and how they are structured. In chapter 3, theoretical background and empirical properties of financial markets are reported. As the time series analysis of a huge amount of data can require immense computational resources, the alternative concept of GPU computing[29] is applied. This technique is also used in chapter 4 in which pattern based correlations in financial markets are analyzed. However, GPU computing, which only recently became broadly available, can also accelerate computation for a standard model of statistical physics—the Ising model, which is also used in interpreting financial markets. The porting of a checkerboard update algorithm for the Ising model to a GPU architecture is demonstrated in chapter 5. Chapter 6 covers the definition and analysis of trend switching processes in financial markets. There, we ask how financial market quantities—such as volume or inter-trade waiting times—are quantifiable on the path from a local price minimum to a local price maximum and vice versa. In chapter 7, the multi-agent based order book model is presented which is able to reproduce a set of *empirical stylized facts*. Finally, chapter 8 summarizes and concludes the results of this doctoral thesis.

[29] GPU is an abbreviation for *Graphics Processing Unit*.

Chapter 2

Financial Markets

This chapter covers general topics related to financial markets. Their historical development is shown along with an overview of how financial markets work, including how they are structured, how the various market participants can interact by this means in order to exchange assets and realize a very diverse variety of trading strategies. These aspects are covered in general with specific emphasis on the context provided by the remaining chapters in which financial market fluctuations are analyzed and modeled in detail.

2.1 Historical Development of Financial Markets

Looking at the historical development of public financial markets, for which a term was only recently created as a collective term for capital and money markets, it is necessary to distinguish between spot markets (or cash markets) and futures markets (or derivatives markets).

Futures contracts which are traded on derivatives markets differ from contracts exchanged on cash markets. The major difference is the time of fulfillment. Spot market contracts—one example for spot market contracts are stocks—have to be fulfilled immediately after a contract between two parties is established[1]. On the other hand, futures contracts are fulfilled at a future date. One can additionally distinguish whether an instrument is a conditional or unconditional derivative instrument. This differentiation is made in order to present a complete overview of possible transaction types which should result in an overall deeper understanding of the inter-connectivity of financial markets. As mentioned in the general introduction, credit derivatives caused the high degree of propagation during the recent financial market crisis which started in 2007.

[1] In the Federal Republic of Germany, a spot market contract has to be fulfilled within a maximum time period of two exchange trading days. Fulfillment means that the underlying asset will be delivered, accepted, and paid for (*Clearing*). The same delivery time is valid for spot transactions of currencies on foreign exchange markets, originally due to the time it would take to transfer cash positions from one bank to another bank.

Banking institutions created and distributed such credit derivatives in order to manage credit based risks. As a consequence the real estate market, for example in the U.S., which can be seen as a spot market was tradable and became in some sense liquid. The risk of credit defaults within the real estate market was distributed globally.

We return to the two types of derivative instruments: An option as a conditional derivative instrument is a contract between a buying and a selling exchange member that gives the buyer the right, but not the obligation, to buy or to sell a specified asset—which is called the *underlying* asset—on (*European option*) or before (*American option*) the option's expiration time, at an agreed price. This price is called the strike price. In the opposite case, the seller of an option collects a payment—which is called the *premium*—from the buyer. Granting the option is also referred to as *selling* or *writing* the option. A call option gives the buyer of the option the right to buy the underlying at the strike price. A put option gives the buyer of the option the right to sell the underlying at the strike price. In the case that the buyer chooses to exercise this right, the seller is obliged to sell or buy the asset at the agreed price level[Dub04, Deu03]. The buyer may choose not to exercise the right and let it expire. The underlying asset can be a piece of property, a security (e.g., a stock or a bond), or a derivative instrument, such as a futures contract.

Unconditional contracts such as futures contracts secure for the buyer and for the seller the right and the obligation (in contrast to the previous paragraph) to buy or to sell a specified *underlying* on the future's expiration date, at an agreed upon price.

Based on this overview which contrasts the various types of exchange based transactions, for which the main differences can be found in the settlement date, we present in the main part of this historical outline a collection of important historical milestones of exchange based trading, covering also the Tulip Mania in the United Provinces and the Gold Fixing procedure in London. Additionally, we refer to commodity markets which can also be split into spot and derivative markets and which were already mentioned in chapter 1.1.

The historical development is also important for covering another aspect. From a physicist's point of view, financial markets exhibit behavior very much like a complex system in physics with non-Gaussian price increments. The price process of an exchange traded asset can be seen as the result of superimposed trading decisions of market participants. When modeling a financial market as a "complex particle system" with agent-based models using Monte Carlo techniques, market microstructure plays a huge role. "Market micro structure is the study of the process and outcomes of exchanging assets under explicit trading rules", as can be found in [O'H97]. However, these trading rules have changed throughout history. Thus, a historical outline is necessary in order to highlight the dramatic transformation from the beginning of floor trading to full electronic marketplaces in recent years.

2.1.1 Tulip Mania in the United Provinces

A classic example for exchange based trading can be found in the famous tulip mania [Gar89, Gar00, Gol07], which took place in the first part of the 17th century, in which tulips became a subject of speculation. This tulip mania had a serious impact on social life in the area now known as the Netherlands and is considered to be the first speculation bubble in history. Tulips, which were originally only located in Asian countries, were introduced in Europe in the middle of the 16th century. In 1560, first tulips were transported from Istanbul, the court of the Ottoman Empire, to Vienna. There, botanist Carolus Clusius was responsible for the imperial botanical garden of Emperor Maximilian II, where he cultivated these plants in 1573 after importing 1,500 seeds. The tulip cultivation in the United Provinces, which are today known as the Netherlands, started then in 1593 after Carolus Clusius had come from Vienna and when he was able to breed tulips under the particular, not ideal climate conditions in the Netherlands. This exotic flower began to enjoy great popularity and tulips subsequently came to be increasingly seen as a status symbol.

This triggered an avalanche and started a competition for possession of the rarest tulips. In the beginning the largest proportion of market participants in this "tulip market" was the upper class of society. Tulips with lines and flames on the petals were particularly popular. However, this effect was often caused by a special virus, which infects only tulips and is known as the Tulip Breaking potyvirus.

This development can be seen as the starting point of a speculation bubble. In the following time, the competition for possession of tulips escalated and very high price levels were reached when the demands for tulip bulbs exceeded the offers. In 1623, a bulb of a very rare and preferred tulip could change hands for a thousand Dutch florins. This has to be seen in relation to the general financial setting, in which the average income per year was only about 150 Dutch florins. In the beginning, bulbs were only traded during the planting season. Later with increasing demand, bulbs could change hands during the whole year with the consequence that bulbs were sold without the knowledge by the buyer of the exact tulip appearance. So, tulip trading became a speculation business. This aspect operated as a catalyst for paintings in order to give impressions of the probable future appearance of a tulip to interested clients.

In the 1630s, the situation had reached a turning point. One reason was, that cash settlement transactions were no longer the only type being recorded. Tulips were also traded via option contracts. In this case the specified underlyings were tulip bulbs. These financial instruments made it feasible to hedge risks due to price changes and allowed tulip producers to sell future crops. On the other hand, consumers were able to lock-in cheap buying prices for the future. As an instrument of speculation, option contracts, which are tradable in very different forms on today's derivatives markets, can exhibit a enormously high risk profile, as a large leverage effect is attainable with comparatively small trading accounts. These Dutch option contracts were the first derivative instruments in history, and the increased use of option contracts in Dutch tulip bulb trading can be regarded as the first evidence for the beginning of a crisis. Addi-

tionally, a high degree of debt for some market participants could be observed, which is one important ingredient for the formation of a speculation bubble. The prices of bulbs escalated by a factor of 50 in the years from 1634 to 1637 [Gol07]. In Amsterdam, for example, a mansion was sold for three tulip bulbs. And a transaction of 40 bulbs for 100,000 Dutch florins was observed and documented in 1635. Also a sale of a very rare tulip bulb—the Semper Augustus—was recorded with a price of 6,000 Dutch florins. These transactions have to be seen in comparison with the prevailing consumer price structure in the United Provinces in this time. A ton of butter cost roughly 100 Dutch florins. This example clearly illustrates the speculative nature of tulip bulb trading at the time.

In modern risk management, financial derivatives can be very efficient instruments for controlling the risk measures of a portfolio, which is a collection of various assets. In the Dutch tulip bulb crisis, however, option contracts were used in order to clear bottlenecks. This can also be seen in the context of the recent financial market crisis and its starting point which goes back to the U.S. housing market. In order to extend credit to future home owners, and to clear the associated bottleneck, existing credits were securitized and sold to international banks and institutions. It is possible to claim that the flourishing business of securitizations which were organized by investment banks caused more and more people to become borrowers in order to realize their plans for owning their own home. There was, however, as a result a larger portion of people who were not able to service their debts. Later the rising defaults on these sub-prime mortgages in the U.S. triggered the global crisis in the money markets. Of course, it must be added that an important ingredient in this was the policy of the Federal Reserve: increasing interest rates had a fatal effect on those debts, most of which had variable interest rates. Nevertheless, the securitization of debts before the beginning of the U.S. sub-prime crisis can also be seen as a clearing of bottlenecks similar, though not identical, to the Dutch tulip bulb crisis.

In the United Provinces, the speculation bubble boosted trading in various parts of Dutch society. In order to speculate in these rapidly developing markets, Dutch citizens of all social standings invested large parts of their wealth in tulips. As a result, some speculators made very large profits during this period of time while others lost their entire wealth or even more. The speculative bubble burst in 1637. The inflated prices for tulip bulbs could no longer be obtained by the tulip traders. Demand began falling and this caused a panic. Everybody wished to sell, and nobody bought tulip bulbs. The prices had rebounded in January, and in February a maximum was reached. As in every bubble, there were also here winners and losers. Some people were holding open option contracts for selling tulips at prices exorbitantly high in relation to current market values. The counterpart, who had the opposite position, was obliged to buy these bulbs for unrealistic prices. Thousands of Dutch people were financially ruined as a result.

2.1.2 History of Public Exchange Based Trading

Considering this short outline of history's first speculative bubble, it must be noted that trading in the Dutch tulip mania was unorganized and a kind of an *over-the-counter* market (OTC)[2]. Contracts were concluded individually by the parties. We now consider the historical development of public exchange based trading which itself goes back to the tulip bubble.

First evidence for exchange based trading structures can indeed be found in the tulip bulb bubble. At this time, options became important for economic processes for the first time. It was not only options on tulip bulbs, however, that were traded. Contracts of the East India Company, which was founded in 1602, also changed hands [Tra06]. In Germany, options became tradable for the first time in 1970. However, in the United Kingdom and in the United States of America (USA), options were already being traded in the 18th century. Despite the fact that the following outline is not exhaustive, the beginning of standardized futures trading should be mentioned, which occurred on the Chicago Board of Trade (CBOT) in 1865. In addition to the availability of an established OTC market, the trading of exchange based stock options in the USA was possible on the Chicago Board Options Exchange (CBOE) beginning in 1973. In Europe, the European Options Exchange (EOE) was founded in Amsterdam in the year 1978, and after difficulties getting started, the EOE was established as the primary market for options in Europe in the early 1980s. This remained so until the Deutsche Terminbörse (DTB) was created in 1990 in Germany. In 1998, then the merger of DTB and the Swiss Options and Futures Exchange (SOFFEX) led to a leader among international derivatives markets [Deu06], the European Exchange (EUREX), which was a purely electronic trading platform in contrast to established exchanges in the USA, which were based on floor trading. There, traders physically exchange assets on the floor. EUREX additionally bought the International Securities Exchange (ISE) in 2007, in order to play a pioneering role in the upcoming exchange consolidation process. However, this step also has to be seen in the context of power shifts among large exchange companies all over the world. In addition, the well-organized activities of hedge funds influenced this evolution, as they held large stock positions in the Deutsche Börse AG[3] and the London Stock Exchange (LSE). However, the initial intention of forcing a merger between both exchanges misfired. Instead, a merger between two spot markets, the New York Stock Exchange (NYSE) and Euronext was realized creating NYSE Euronext—a large competitor of Deutsche Börse AG and LSE[4].

International stock markets have a longer history. The NYSE is a large spot market

[2] *Over-the-counter* markets (OTC) differ from public exchange based trading, as they are not regulated. Furthermore, the fulfilling of OTC market contracts is not guaranteed by a central clearing house and so the additional risk that the contract counterpart is not able to fulfill an arranged contract exists.

[3] EUREX is a joint venture of Deutsche Börse AG and Swiss Exchange (SWX).

[4] Founded in 1801, the London Stock Exchange (LSE) is one of the largest stock exchanges in the world. The NASDAQ was interested in a takeover of LSE, but this offer was roundly rejected by LSE shareholders. The acronym NASDAQ originally stood for *National Association of Securities Dealers Automated Quotations*.

player with a long tradition having been founded in 1792[5]. Up to March 2006, almost all trading on NYSE was handled on the floor. Then, with the acquisition of the electronic trading platform Archipelago, the electronic era was ushered in at NYSE. This process was enforced after the merger with Euronext. Today, a clear tendency to electronic market places with the highest possible transparency can be observed. As could be read in September 2007 [Kul07], the NYSE intended to close two more trading rooms on the floor. In addition, roughly one-third of the trading volume at NYSE is now generated by algorithmic trading[6]. Algorithmic trading or program trading, which means the execution of automatic trading decisions generated by computer algorithms, can only be realized in a meaningful way on full-electronic trading platforms due to transparent rule sets and execution times. In addition to the NYSE, the XETRA system (the electronic spot market platform of Deutsche Börse AG, which replaced the old IBIS system in 1997) should be mentioned as well, and started in the same way as a great success story, and is now being exported to exchange institutions in Asia.

A very impressive example of the change from floor trading to electronic trading can be found in trading of the Euro-Bund future[7] (FGBL), which is a fixed income derivative. In the beginning of the 1990s, the FGBL contract was traded at the newly established DTB and also at the London International Financial Futures Exchange (LIFFE), which was founded in 1982 and which is today part of NYSE Euronext. At that time, LIFFE was purely a floor trading exchange, and the FGBL was favorably traded at LIFFE. However, an empirical comparison between these two exchanges reveals the advantage of the DTB, even though the transaction volume there was lower than that observed at LIFFE. One reason for the success of DTB was the higher level of transparency. The DTB's order book, which stores offers and demands of all market participants and matches orders against each other, was provided in parts to the market members. Thus, each market participant was able to use and observe market depth. In addition, the change to more anonymity boosted electronic trading. With an increasing liquidity, which is a function of market depth and the difference between best bid and best ask, the DTB was able to consolidate the entire trading volume of the Euro-Bund future (FGBL) into their own trading system. This was one milestone for the triumph of full-electronic exchange trading over floor based trading.

2.1.3 The Gold Fixing in London

Even if electronic trading has mostly displaced floor trading and manual trading processes, however, some relics can still be observed today. Naturally, some exchanges indulge in floor trading for public relation reasons. The Deutsche Börse AG has also

[5]The Philadelphia Stock Exchange is the oldest stock exchange in the United States of America which was founded in 1790 and has been owned by NASDAQ since 2007.

[6]Detailed information can be found in the news releases of NYSE Euronext on the website `http://www.nyse.com`.

[7]The underlying of the Euro-Bund future contract is a certain debt security issued by the Federal Republic of Germany with a remaining term of 8.5 to 10.5 years.

hung onto floor trading. Recently, the furniture on the floor was even modernized, in order to present to the general public a physical manifestation of financial markets.

A completely different trading process, which can be classified as a traditional institution, is the Gold Fixing in London [O'H97] which is a reference price for gold. The first Gold Fixing took place on September 12th, 1919 at 11.00 a.m. in London in physical presence of the original five founding members N. M. Rothschild & Sons, Mocatta & Goldschmid, Samuel Montagu & Co., Pixley & Abell, and Sharps & Wilkins[8]. Gold fixing is executed twice a day. At this time, the five members come together. Since 2004, a dedicated conference call has been used[9]. At the beginning of each fixing, the chairman of the meeting, who is one of the members, announces an opening price. The other four members then have the possibility of contacting their customers and including their orders into their decision. Then, each member has to declare if the member institution is a buyer or a seller at the announced price level. In the case, that there are buyers and sellers, also the quantity offered and demanded at this price is asked in units of bars. If only buyers or only seller exist at the announced price, or if the difference between offered and demanded gold bars at a price level is larger than 50 bars, then the procedure starts again and the price is moved. If a price finally fulfills the requirements, the chairman announces that the price is fixed. In order to be able to contact clients, each member has a verbal flag. As long as a flag is raised, a price cannot be declared fixed. However, this verbal flag has physical roots. Before the dedicated conference call was established, each member had a small Union Flag on their desk for this purpose.

Since 1919, gold price fixing has provided market participants with the opportunity to buy and sell gold at a specific price. Today the London Gold Fixing is still a benchmark for trading with gold. However, as seen above, former procedures were adapted to the progress achieved in information technology. Additionally, there exists a full electronic based marketplace for trading gold today. On the New York Mercantile Exchange (NYMEX), founded in 1882, and on the New York Commodities Exchange (COMEX), which is a sub division of NYMEX today, gold can be traded continuously. Recently, NYMEX became part of the Chicago Mercantile Exchange (CME).

2.2 Continuous Double Auction

The historical milestones and facts covered in the previous sections document the significant movement away from well-established floor trading toward fully electronic trading systems in the last two decades. There were various reasons for this development. On the one hand, there was the tendency toward more efficiency and transparency in trading processes. On the other hand, the potential of the technology for cost cutting acted as a catalyst.

[8] At the moment, the following institutions are members: The Bank of Nova Scotia - Scotia Mocatta, HSBC, Deutsche Bank AG London, Societe Generale Corporate & Investment Banking, and Barclays Capital.

[9] More information can be found on the website `http://www.goldfixing.com`.

In this section, we explain the underlying structures of electronic trading systems which are important for the analysis of high frequency financial market data sets. In addition, these structures are crucial ingredients for realistic financial market models[10] as they incorporate financial market micro structures.

On financial markets, a central order book stores offer and demand of all market participants. This process is called continuous double auction. A double auction is a process of buying and selling assets in which potential buyers submit their bids and potential sellers simultaneously submit their ask prices to a central auctioneer (in this case the central order book). Then the exchange system chooses some price p that clears the market. The word continuous indicates that during this period all submitted orders will be executed immediately against offer and demand stored in the order book if they fit the price constraints of an exiting order. If they do not fit the given price constraint then they can be also stored in the order book for a later execution. Continuous trading is the main part of trading. Beside continuous trading, one can also find opening and closing auctions as well as underlying auctions during the trading day. However, this depends on whether the traded product is supervised by a market maker. If a market maker agrees to continuously quote bid and ask prices for an underlying then the market maker will obtain reduced trading fees for this product.

In this context, the market determined observable "price", which represents the time dependent value of the underlying financial instrument in monetary units, is not given as a continuous variable. Instead, it is given as a discreet multiple of a product-specific minimum price change, which is called *tick size*, p_{tick}.

Economic reasoning tells us that the price of an asset cannot become negative if it secures a right and not an obligation (e.g., a stock). Thus, the price is given at any time t by

$$p(t) = n(t) \cdot p_{\text{tick}} \tag{2.1}$$

with $n(t) \in \mathbb{N}_0$. The highest price level, at which market participants are demanding (i.e., a demanding order exists at this price level) is called the *best bid*, p_b. Analogously, the *best ask* p_a defines the lowest price, at which a sell offer is available. The non-zero gap between *best bid* and *best ask* is called the *spread* which is given by

$$s = p_a - p_b \tag{2.2}$$

as shown in Fig. 2.1. During market phases in which no trading activity is detectable, the spread is always larger than zero. In the case $s = 0$ (i.e., $p_b = p_a$) executable buy and sell orders are inserted in the order book. This results in an "immediate" matching procedure and a spread of $s = 0$ is thus not possible in practice. Nevertheless, time is also discretized. In the case of EUREX, the clock cycle of the order book was recently reduced to 10 milliseconds.

[10] The author of this doctoral thesis passed the trader exam ("Complete Exam") of the European Exchange EUREX. The overview of financial market structures in this section is based on this knowledge. More information can be found on `http://www.eurexchange.com`.

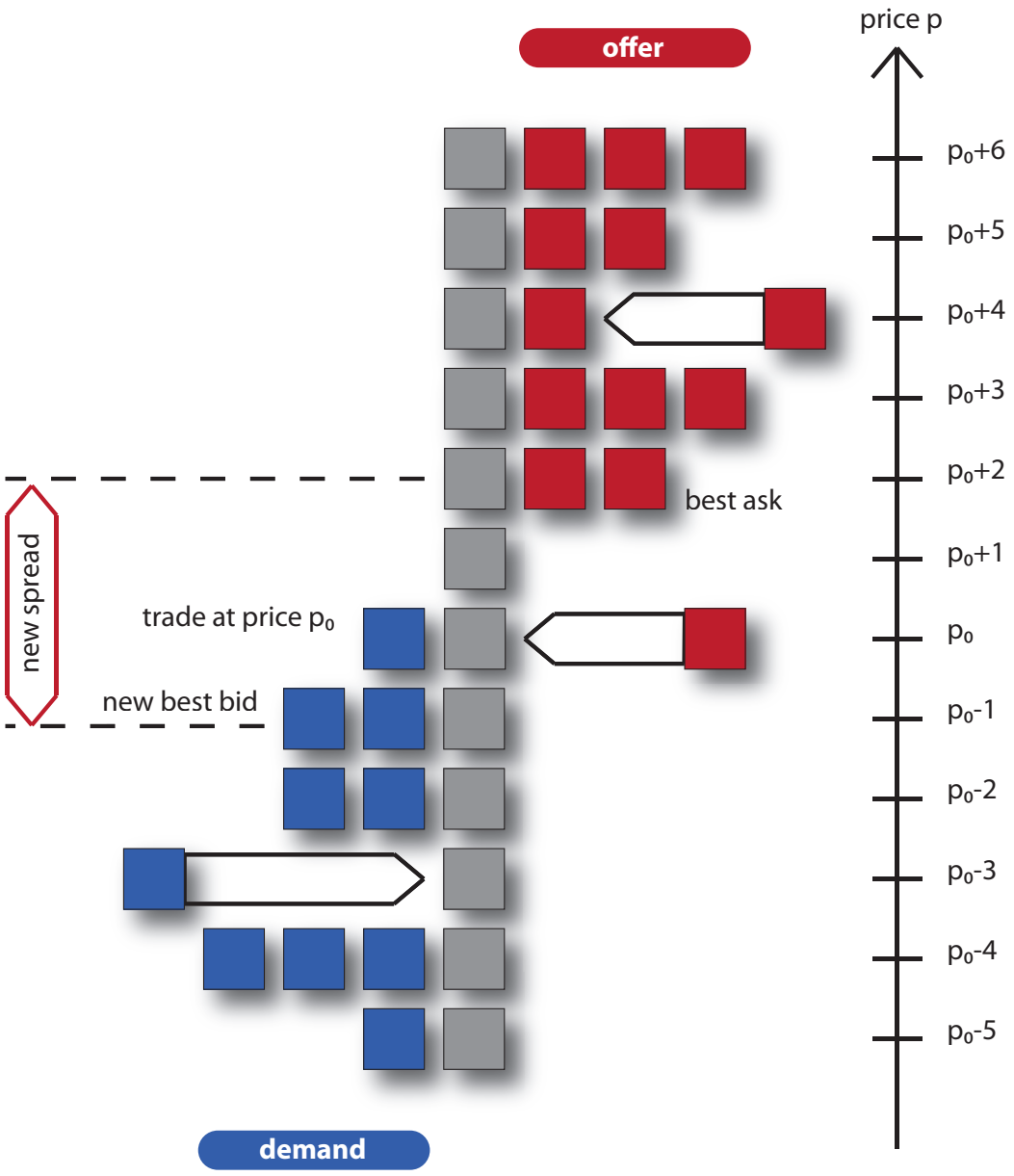

Figure 2.1: Structure of an order book: The buy (blue) and sell orders (red) are added chronologically at the corresponding discrete price level p. In this way, an allocation algorithm realized fulfilling price-time priority. The two orders at price level for the price p_0 are on the demand and offer side of the order book. Thus, they can be matched against each other. A transaction at price p_0 is thus established (p_0 is also known as last traded price or last price). This transaction causes an increase in the spread $s = p_a - p_b$. Before the arrival of the sell order at price level p_0, the spread was two *ticks*. After the execution, the spread is three *ticks*. In accordance with technical exchange settings, executable orders will be matched by the order book "immediately"—thus, a spread $s = 0$ is not possible in practice. Nevertheless, time is also discrete in this setting. In the case of EUREX, the clock cycle of the order book is currently 10 milliseconds.

Let us assume—without loss of generality—that a buyer would like to buy one unit of the underlying asset and that a seller would like to sell one unit. According to the given matching algorithm (which will be described in detail in section 2.4) the buyer obtains the asset traded in the order book from the seller. The seller receives the execution price $p^e = p_b = p_a$ in monetary units from the buyer.

2.3 Types of Orders

The variety of order types available is huge. The term order is used as a synonym for a transaction order. Trivially, one can distinguish between buy orders and sell orders. Additionally, this distinction can be combined with the order volume—the number of assets to be exchanged with a given order. Therefore, the volume k_i of an order i is positive if i is a buy order. Analogously, the volume k_i is negative if i is a sell order.

Furthermore, the group of order types can be executed in various ways which have to be taken into account. In the remainder of this section, we present limit orders in subsection 2.3.1 and then market orders in subsection 2.3.2 which are the two most important order types in real order books. Finally, subsection 2.3.3 documents a selection of important examples covering special order types.

2.3.1 Limit Orders

In addition to the volume $k_i^l \in \mathbb{Z}\backslash\{0\}$, a limit order i is also specified by a limit price $p_i^l = n \cdot p_{\text{tick}}$ with $n \in \mathbb{N}_0$. Whether an order is a buy or a sell order can be identified using the expression

$$\operatorname{sgn}\left(k_i^l\right) = \left\{ \begin{array}{rl} 1, & i \text{ is a buy order} \\ -1, & i \text{ is a sell order} \end{array} \right\}, \tag{2.3}$$

as already mentioned. Limit orders are executed at the limit price p_i^l or at a better price level:

$$p_i^e \leq p_i^l \tag{2.4}$$

can be applied for the execution price of a limit buy order i. Analogously, a limit sell order j leads to the expression

$$p_j^e \geq p_j^l. \tag{2.5}$$

If $|k_i^l| > 1$ it is possible that order i is executed in many partial executions. Here, a maximum of $|k_i^l|$ partial executions are possible. This can also lead to a variety of partial execution prices. Each of them has to fulfill Eq. (2.4) and (2.5), respectively.

There is also the possibility that a limit order includes restrictions. Here it is necessary to mention the possible options *immediate or cancel* (IOC) and *fill or kill* (FOK). IOC

orders are executed instantaneously, either completely or as completely as possible. The volume fraction of order i which cannot be executed at price p_i^l or better is deleted immediately. A FOK order also has the property that it is executed immediately. This is, however, only the case if the entire volume k_i^l is executable. In EUREX order books, the FOK restriction is only used for option contracts.

Limit orders without restrictions persist if there is no matchable limit order from a counterpart directly after submission. This persistence can be tuned by the submitting market participants. They have access to limit order properties. Unless specified otherwise, a limit order has a validity of one day, *good for day* (GFD). An unexecuted GFD order is removed from the order book at the close of the exchange.

The temporal settings *good till canceled* (GTC) and *good till date* (GTD) are also available. In the case of a GTC order, the time scale of order validity is not defined from the beginning. A GTC order is valid until execution or until the market participant decides to cancel the order. In the case of a derivatives exchange such as EUREX, there is an additional scenario in which a GTC order could be removed from the market: If the underlying contract passes the expiration date the limit order is also canceled. Also, limit orders stored for one day will be canceled at the close of the exchange. A GTD order is canceled if a specified date is reached.

2.3.2 Market Orders

A market order (for which volume can analogously be defined by k_i^m) has, in contrast to limit orders, no limit price. A market order will immediately be executed at the best available market price. This scenario guarantees to the market participant that market orders are executed (neglecting from the special case in which there are no limit orders stored in the order book). However, the market participant has no guarantee with respect to the realized execution price p_i^e.

The cumulative order volume at a given price level p is defined by

$$S(p) = \sum_{\{j \in \mathbb{N}_0 : p_j^l = p\}} k_j^l. \tag{2.6}$$

This expression calculates the summation over all limit orders j whose limit price p_j^l is equal to price level p. In the case that $k_i^m > 0$ and $|k_i^m| \leq |S(p_a)|$ are fulfilled at the same time, then the execution price of the market buy order is $p_i^e = p_a$. Analogously, the execution price p_i^e of a market sell order i is p_b if $k_i^m < 0$ and $|k_i^m| \leq |S(p_b)|$.

Given an empty order book, a market order which cannot be executed against limit orders will be stored in the order book until a corresponding order becomes available. Such a procedure is accompanied by the inherent risk that it might lead to an execution far from the fair value level of the underlying asset. Thus, it is not recommendable to use market orders in more illiquid order books.

It should be pointed out here, however, that it is also possible to simulate a market order using a limit order. Such an approach can also be practical, as some financial

market trading systems do not natively support market orders. If the limit price p_i^l of a limit buy order i is larger than the current *best ask* then the market participant obtains an execution at *best ask*. Additionally, $|k_i^l| \leq |S(p_a)|$ should be fulfilled. In order to reduce the probability that order i does not result in execution, one can adapt the limit price using a shift $\Delta p_i^l(k_i^l, \sigma_{\text{imp}}) = p_i^l - p_a$ depending on k_i^l and taking the prevailing implied volatility σ_{imp} into account. For a simulated market sell order, it is necessary to choose a limit price for the limit order p_i^l which must be below p_b.

2.3.3 Special Types of Orders

Besides limit orders and market orders, which are the most important order types, other order types in real exchange trading systems exists. Although not a comprehensive list, three other types of orders are mentioned here. Well-known is the stop order, which is supported only for futures contracts on the EUREX platform. The stop order can be understood as an emergency brake for a trading position, whose loss has grown. A stop order has a limit price, but this is not comparable with the limit price of limit orders. A stop order is triggered when the market price reaches its limit price. Then, the stop order becomes a market order, and this results in an execution at the best available market price. Thus, one obtains guaranteed execution and the level of loss can be defined more or less exactly. In more volatile markets, the guarantee of an execution comes with the risk that execution is performed farther from the triggering limit price. There are two trivial variants of stop orders, namely the stop buy order and the stop sell order. For a stop buy order, the limit price is located above the current market price (e.g., in order to close a short position). It will be triggered if the market moves up. A stop buy order then becomes a market buy order. Analogously, the limit price of a stop sell order is below the current *best bid* (e.g, in order to close a long position).

Far less known is the *market to limit* order, which is used in the XETRA trading system[11]. The *market to limit* order combines the positive properties of a market order (high probability of execution) with the advantage of a limit order with respect to the execution price. A *market to limit* is executed against the best limit price on the opposite side of the order book—either *best bid* or *best ask*. In the case that the order cannot be executed completely, the remaining volume of this order becomes a new limit order with a limit price which is identical to the already executed part. At first sight, the benefits of this type of order are not obvious compared to a limit order whose limit price is set to the *best bid* or the *best ask*. In volatile phases of market movement, however, orders are submitted in beats of milliseconds to the central order book. Thus, the current *best bid* and *best ask* at the time of order submission do not necessarily have to correspond to the situation when the order arrives at the order book after multiple milliseconds. The limit order thus remains unexecuted in the order book, while a *limit to market* order ad-

[11] More information can be found, e.g., at `http://www.wienerboerse.at`. The Vienna stock exchange uses the XETRA system (e**X**change **E**lectronic **TRA**ding) of the Deutsche Börse for the trading of Austrian shares, other equities and bonds since 1999.

justs to the market situation. The market participant obtains the immediately executable portion of the total volume.

Another type of order, the *iceberg* order, is characterized by a certain degree of discretion and is particularly suitable for the placement of large order volumes. It offers the possibility of putting a large order volume in the order book, and it's primary advantage is that the market is not able to obtain knowledge of the total volume. Such a mechanism is realized by specifying a limit price, the total volume and the so-called *peak* volume. The *peak* volume is the visible part of an *iceberg* order in the order book. During continuous trading, completely executed peak volume is replaced by a new peak volume order if there is still hidden volume in the *iceberg* order available. This procedure is carried out until the entire iceberg order is processed. Due to the nature of its construction, an *iceberg* order can be used for the discrete placement of large stock positions. If one were to instead use a limit order, which reveals to the market the total volume, then this limit order would have a lower execution probability due to legal versions of illegal *front running*[12].

2.4 Matching Algorithms

In electronic trading systems several processes of order allocation (*matching*) can be distinguished. The following subsections describe the two allocation algorithms that are most frequently encountered. However, combinations of these *matching* algorithms are also possible as will be mentioned in section 2.4.2. The allocation or *matching* algorithm answers the question of which owner of a limit order stored in the order book at a given price level (e.g., the *best bid* or *best ask*) will become the counterpart of a submitted market order or a submitted limit order which is immediately executable. Of course, this question is only non-trivial if there is more than one order at that price level.

2.4.1 Price Time Priority Allocation

The price time matching algorithm provides an order allocation which is firstly based on a price and secondly based on a time priority rule. If a new order is submitted to the electronic order book of the exchange, all stored limit orders will be examined in the order book. Limit orders with a better limit price are executed before limit orders with

[12] *Front running* is banned in most countries. It describes a trading strategy which is based on insider knowledge about upcoming large orders initialized by customers. The trader takes a position that corresponds to the customer's order. When the customer's order is executed the trader can benefit from the change of the market price. In this context, one has to mention the controversial practice of using *flash orders* which is still legal in U.S. Nevertheless, they are the subject of significant debate: Some customers are allowed to see the incoming order flow slightly earlier than general market participants—typically 30 milliseconds. For this advantage, they have to pay a fee. However, market participants with access to extremely powerful computing resources can conduct rapid statistical analysis of the changing market state using this information. They then carry out *high-frequency trading* a few milliseconds before "normal" market participants can react [And09]. Several U.S. exchanges banned this practice in August 2009.

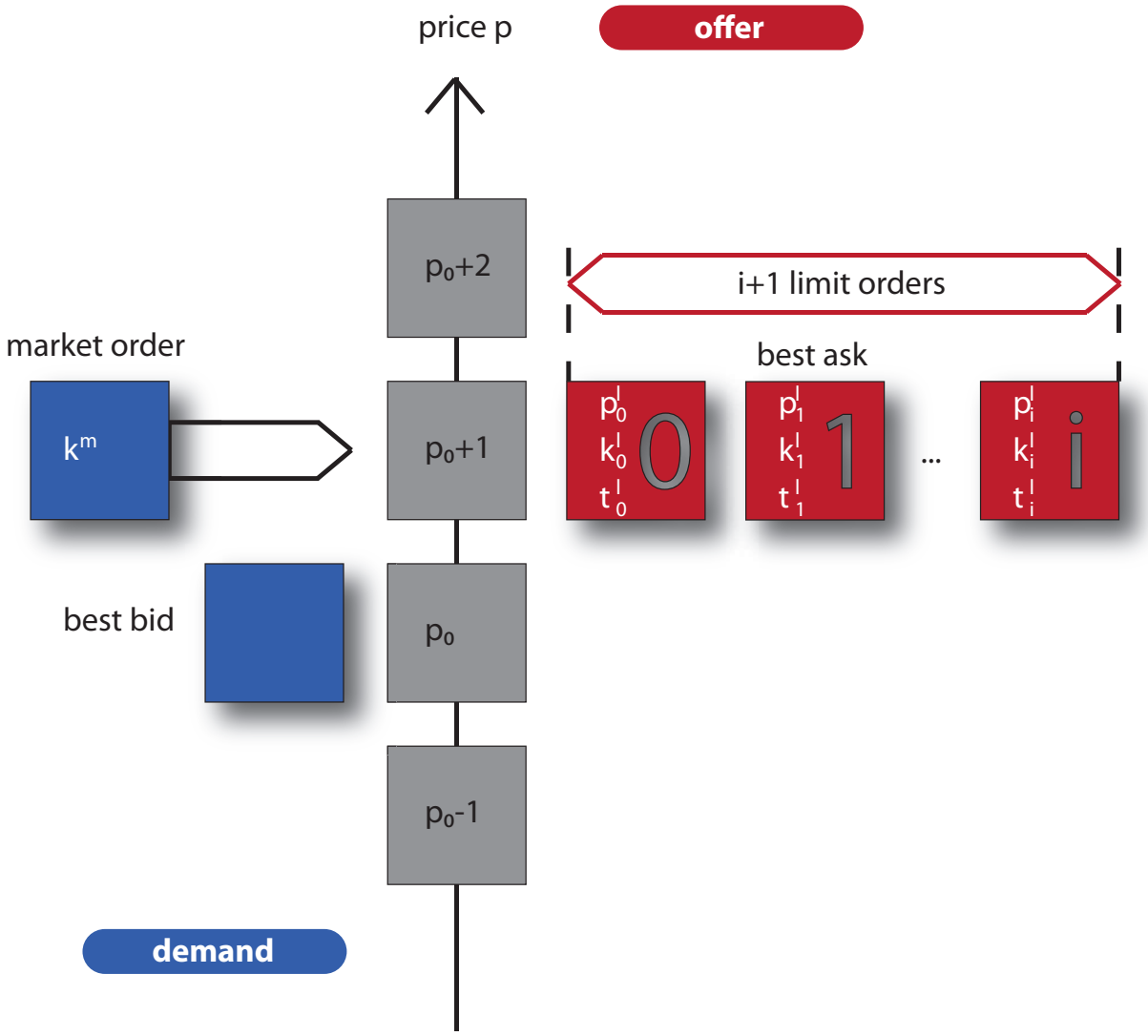

Figure 2.2: Order allocation: A market order of volume $k^m > 0$ meets $i+1$ limit orders with $k_j^l < 0 \;\; \forall j \in \{0,1,\ldots,i-1\}$. If $|k^m| < \sum_{j=0}^{i-1} |k_j^l|$ is valid, the *best ask* price p_a remains unchanged after the transaction. As all $i+1$ limit orders were inserted into the order book at price p_a, it is trivially fulfilled that $p_j^l = p_a \;\; \forall j \in \{0,1,\ldots,i-1\}$. The chronology of order submission implies $t_j^l < t_{j+1}^l \;\; \forall j \in \{0,1,\ldots,i-2\}$.

a worse limit price (i.e., price priority). It is worthwhile to mention here that the highest allocation priority is given for market orders and that a market order is matched before an executable limit order. In order to track the chronological order—for the purposes of achieving chronologically prioritized allocation—every order receives a time stamp when it is inserted into the central order book. If a market order of volume k^m meets $i+1$ limit orders at price level p_0+1, all of which have identical limit prices, price priority is not longer effective. If $|k^m| \leq |S(p_0+1)|$ is satisfied then the $i+1$ limit orders will be matched chronologically with the market order (see Fig. 2.2).

Table 2.1: Price time priority: In this first example, a market buy order with volume 1 is submitted to the central order book. As there are two limit sell orders sharing the price level p_0+1 with volumes 1 and 3, chronology is of critical importance. Since the limit sell order with volume 1 has a higher priority in time, this order is the corresponding order for the allocation of a market buy order. The limit sell order with volume 3 enters the market and, after the transaction, has the highest time priority at this price level.

Bids				Price p	*Asks*			
				p_0+3	1	3	3	1
				p_0+2	2	4		
				p_0+1	1	3		
			2	p_0				
	1	1	1	p_0-1				
2	1	4	1	p_0-2				
		2	3	p_0-3				

Bids				Price p	*Asks*			
				p_0+3	1	3	3	1
				p_0+2	2	4		
				p_0+1	3			
			2	p_0				
	1	1	1	p_0-1				
2	1	4	1	p_0-2				
		2	3	p_0-3				

This means that firstly, the order with the highest time priority is allocated to the market order. This is the limit order with time stamp t_0^l. If $|k^m| = |k_0^l|$ is valid then the market order and limit order 0 annihilate each other, and limit order 1 obtains the highest time priority for the ongoing trading process. In the case, however, that $|k^m| < |k_0^l|$ is fulfilled, the unexecuted part of order 0 remains in the order book with highest time priority. If the market order has a larger volume than the volume of the first limit order, that is $|k^m| > |k_0^l|$, then the chronologically next order in the order book is used for allocation to the extent necessary. This process continues until the entire volume of the market order is processed. If $|k^m| \leq |S(p+1)|$ holds, the market order can be carried out with a uniform execution price $p^e = p_0+1$. If this is not the case, one has to use for the matching procedure all $i+1$ limit orders at price level p_0+1 and in the end, unexecuted volume remains. Then, the matching principles are iteratively applied to higher price levels due to the price priority rule, first at price level p_0+2.

We sketch the principle of order allocation based on price time priority in two simple examples. The left sides of tables 2.1 and 2.2 each show the relevant part of the order book before the incoming order; on the right sides the situation in the order book after the matching has occurred is shown. The initial situation is in both cases the same—only the volume of the incoming market order varies. The red numbers indicate volume

Table 2.2: Price time priority: In this second example, a market buy order with volume 5 is submitted for execution in the central order book. Analogous to the example in Table 2.1, the market participant's order is partially executed at $p_0 + 1$ with volume 1. Furthermore, as there are still 4 units which have to be executed on the *market*, this leads to a second partial execution of 3 volume units at the same price level. After the second partial execution, the *best ask* is changed to $p_a = p_0 + 2$. The remaining volume unit of the market buy order is executed against the limit sell order providing 2 volume units at $p_0 + 2$. The extra volume unit of the limit sell order remains in the order book with the highest time priority at this price level.

Bids				Price p	*Asks*			
				$p_0 + 3$	1	3	3	1
				$p_0 + 2$	2	4		
				$p_0 + 1$	1	3		
			2	p_0				
	1	1	1	$p_0 - 1$				
2	1	4	1	$p_0 - 2$				
		2	3	$p_0 - 3$				

Bids				Price p	*Asks*			
				$p_0 + 3$	1	3	3	1
				$p_0 + 2$	1	4		
				$p_0 + 1$				
			2	p_0				
	1	1	1	$p_0 - 1$				
2	1	4	1	$p_0 - 2$				
		2	3	$p_0 - 3$				

of limit sell orders, while the blue marked numbers indicate the volume of limit buy orders. We use the convention that an order which is closer to the price column p has a higher time priority than an order which is further away. The field of limit orders which are used for matching are shaded.

2.4.2 Pro Rata Allocation

If an exchange traded asset is driven by only relatively small price fluctuations within a trading day—*intraday*, then the situation is such that limit orders stored behind the limit order with the highest time priority at the *best bid* or *best ask* suffer from a low execution probability. Those orders are waiting in the order book until the first order or first orders with higher priority are processed and removed from the order book. Sometimes, such priority orders can be very huge. Thus, it can be possible in some contracts that all other limit orders on the same side of the order book are blocked by a dominant order for the whole day. In order to address this problem many exchanges instead use the pro rata[13] matching algorithm for products with small price changes. Pro rata matching is wide-spread in U.S. derivatives markets. All *STIR*[14] *futures* are also handled with pro rata matching as they show a very low levels of daily fluctuation.

If a pro rata based allocation is used, then the corresponding time stamps of limit orders lose their meaning. Limit orders are then no longer distinguishable in terms of

[13] Latin: *pro rata* — proportional, relative

[14] *STIR* (*Short-Term Interest Rates*) denotes short-term money market interest rates. In Europe the *EURIBOR* (*European Interbank Offered Rate*) also belongs to this group.

Table 2.3: Pro rata allocation: Identical to the situation in Table 2.2, a market buy order with a volume of 5 units is submitted to the central order book. As in the previous example, the corresponding market participant who submitted the market order obtains a partial execution of 4 volume units at price level $p_0 + 1$. The remaining part of the market order is executed at price level $p_0 + 2$. It is necessary, however, to consider which of the two limit orders will be used for matching the remaining volume unit. If we calculate for each limit order its relative share in the total volume at this price level and multiply these relative shares with the volume which should be executed, then the execution volumes for both limit orders have to be rounded to zero. Thus, in contrast to the example presented in Table 2.2, the market order's remaining volume of 1 unit is matched with the larger limit order.

Bids				Price p	*Asks*			
				$p_0 + 3$	1	3	3	1
				$p_0 + 2$	2	4		
				$p_0 + 1$	1	3		
			2	p_0				
	1	1	1	$p_0 - 1$				
2	1	4	1	$p_0 - 2$				
		2	3	$p_0 - 3$				

Bids				Price p	*Asks*			
				$p_0 + 3$	1	3	3	1
				$p_0 + 2$	2	3		
				$p_0 + 1$				
			2	p_0				
	1	1	1	$p_0 - 1$				
2	1	4	1	$p_0 - 2$				
		2	3	$p_0 - 3$				

time priority. Only price priority is applied in this case. A limit order with a better price limit is carried out before an order with a worse price limit.

In order to illustrate the differences to a price time priority we refer to Fig. 2.2 again. One can no longer expect a temporal priority based allocation if a market buy order of volume k^m reaches the order book and $i + 1$ limit orders are logged at the *best ask*. For each limit order j, the relative share in the aggregated volume of all limit orders at this price level is calculated. This is given by

$$r_j = \frac{k_j^l}{S\left(p_j^l\right)} \tag{2.7}$$

with $S\left(p_j^l\right)$ as defined in Eq. (2.6). The share r_j in the volume of the market order will be allocated to limit order j. Because assets can only be exchanged in discrete units, the executed volume k_j^e of limit order j is given by

$$k_j^e = \lfloor -k^m \cdot r_j \rfloor \tag{2.8}$$

at the arrival of a market order with volume k^m. The residual volume of the market order is matched against the limit order with the largest volume.

In Table 2.3, a detailed example of pro rata allocation is provided. In order to highlight the differences in matching algorithms, it uses the same initial configuration as in the example shown in Table 2.2 for the price time matching algorithm.

In practice, it is not possible to find pure forms of the pro rata allocation in most order books. In general, stock exchanges link the rules set out with the possibility to submit an order with a *priority flag*. Such an order is executed before other orders at the corresponding price level. A limit order qualifies for a *priority flag* if the submission of this order reduces the spread $s = p_a - p_b$.

In this way, the stock exchange offers an incentive for market participants to improve offer and demand prices. Usually, *priority flags* are coupled with a specific minimum volume or a volume ceiling of price improving limit orders. The other limit orders remaining in the order book are unaffected and are treated according to pro rata allocation.

2.5 Order Book Depth

If a market participant is able to track the *best bid* and *best ask* price and additionally, the cumulative volumes at both price levels, then the market participant has access to the so-called *level 1 data*. Furthermore, it is possible that the cumulative volumes of further price levels are available, i.e., price levels higher than the *best ask* and lower than the *best bid*. This is known as *order book depth* or *market depth*. It is common in retail markets for users to be able to see the best five or best ten bid and ask levels. Full exchange members have, in general, no access limitations. Thus, they are able to access the whole order book. The depth of the order book $n(p)$ can be described by

$$n(p) = |S(p)| \tag{2.9}$$

with $S(p)$ as defined in Eq. (2.6). The cumulative order book depth $N(p)$ for the offering side of the order book is given by

$$N_{\text{Ask}}(p) = \sum_{\hat{p}=p_a}^{p} n(\hat{p}) \tag{2.10}$$

for $p \geq p_a$ and analogously, for the demand side of the order book given by

$$N_{\text{Bid}}(p) = \sum_{\hat{p}=p}^{p_b} n(\hat{p}) \tag{2.11}$$

for $p \leq p_b$. The depth of the order book directly affects the price impact function, which is a measure of the liquidity provided by the execution of a market order. The price impact function gives the instantaneous price change relative to the *best ask* or *best bid* depending on the market order volume. It is straightforward to see that the price impact function is given by the inverse function of the cumulative order book depth [SFGK03].

2.6 Market Participants

The term financial market participant, stock market participant or trader characterizes individuals or groups who are active in financial markets. Banks, investment and hedge funds[15] as well as intermediates and individuals who act on their own account are all examples of traders.

Exchange members which can be identified as major banks and proprietary trading firms usually have access to the central order book via a fast direct link. They combine this speed advantage with high-end technologies—also supported by Graphics Processing Units (GPU)[16]—in order to run high-frequency trading strategies. Outside of this group, it is difficult to classify other market participants in terms of extent of their access to exchange systems in the internet age and fully electronic order book—especially as contrasted with historic floor trading and exchange quotes on a more or less daily basis during that time.

Due to the internet revolution, a private investor can now place a limit or market order within few seconds today just as the manager of an investment fund can. However, it is possible to distinguish among market participants on the basis of volume and strategies they apply. Of course, a huge variety of trading strategies exists. We would like to, therefore, limit ourselves to two main categories of trading strategies, *market making* strategies on the one hand and *trend following* strategies on the other hand. The main difference is that *market making* strategies provide liquidity and that *trend following* strategies consume liquidity.

This distinction aims for making a rough allocation of preferred order types. This is important in order to motivate the simplified assumptions used for the order book model in chapter 7. There, *liquidity providers* can submit *market orders*, and *liquidity takers* are able to submit *market orders*.

2.6.1 Liquidity Providers

A *market maker* provides liquidity for the market. The *market maker* agrees to buy and sell assets at any time, but not at identical prices. The selling price—the price of a limit sell order—is higher than the buying price—the price of a limit buy order. The continuous willingness to buy and sell makes continuous trading possible, especially in more illiquid products. The profit realized by a *market maker* is the difference between buying and selling prices which can be quantified on average by orders of the *spread*.

[15] Hedge funds are less-regulated investment vehicles for institutional market participants and high net worth individuals. Their investment strategies strongly vary and range from classical arbitrage—the risk-less exploiting of market inefficiencies—to huge takeovers covered intensely by media companies. Hedge funds try to realize a large *leverage effect* on equity capital. This can be achieved by additional credits.

[16] Bloomberg uses GPUs to speed up bond pricing. The two-factor model for calculating hard-to-price asset-backed securities runs on graphics processing units paired with Linux servers. More information can be found on `http://www.bloomberg.com/`.

However, it is necessary to distinguish between a "pure" *market maker* who has taken the responsibility for quoting products in exchange trading systems and market participants who use strategies related to *market making*. A pure *market maker* obtains reduced transaction fees or has to pay no transaction fees at all. Such a concept is used for low liquidity products as one can find, for example, for options or for stocks of companies with low shareholder capital. For such small companies, the exchange has an interest in creating liquidity in order to attract investors and to earn transaction fees. *Market maker* adapted strategies are, however, also used by market participants. This means that these traders try to earn the spread (also in very liquid products such as futures contracts). This behavior does not, however, need to be subsidized by the exchange.

All participants have to accept risk—even if the risk is smaller in liquid products. It sounds tempting to be able to earn the spread. But it should also be clear that such an approach can only work perfectly, when the market price is mean reverting or hardly moving. If a trend is established, all *market maker* strategies suffer losses. Given a trend, a *limit order* by the *market maker* is executed which has the "wrong" sign. The order profits from counter-movements. Thus, a persisting trend causes more and more losses for such a *market making* strategy. Therefore, a *market maker* has to place limit orders with a larger distance from the *best bid* and *best ask*, respectively. The alternative is to close positions with loss or to hedge risks using other derivative products.

Nevertheless, these risks are not unmanageable. Therefore, one aspect should be repeated which was already discussed in the motivation section in chapter 1. There we described one motivation—and also fascination—in financial market related science is to implement findings in order to perform profit-maximizing strategies. It is tempting to think that one can apply a strategy which was declared to be successful more than once, and in the best case permanently. Then, it would be possible to reuse profits in order to increase leverage from time to time. Why does this strategy not work out? Let us consider *market making* which is evidentially a successful strategy—the profit is based on the *spread*. This is due to the fact that a *market maker* requires market participants that submit *market orders* and are willing to spend the *spread*. In this context, *market making* strategies are passive strategies, and the profit is linked to transaction volumes. Thus, it is not possible to realize unlimited gains from a *market making* strategy, nor is it possible for other strategies. This could be described as a lack of scale freeness of a *market making* strategy.

2.6.2 Liquidity Takers

The opposite of a *market making* strategy is *trend following*. It is impossible to provide a list of all possible strategies which consume *liquidity*. One can, however, distinguish between trading approaches which are triggered by systematic or discretionary decisions. Trading decisions are thus based on a defined set of rules. Such rules can be executed by algorithmic trading systems—or based merely upon intuition. A huge variety of *trend following* approaches, ranging from approaches that are focused on dividend

yields to strategies based on simple or complex market indicators (e.g., using moving averages which smooth historic price movements can be used). Strategies based on the current weather or the phases of the moon are examples of more exotic, questionable approaches.

Even though trend following strategies can be very multifarious, all of these strategies have one aspect in common. They are consumers of the liquidity provided by *market makers* because they operate via market orders. But one has to keep in mind that *liquidity takers* are "paying" for this service. They have to deal with a systematic disadvantage. Each *market order* transaction initiated by a *liquidity taker* has to face *market frictions* which mainly consist of the *spread* which is paid to *market makers*. Thus, a profitable trend following strategy has to overcome this *market friction*[17] on average.

Finally, it has to be mentioned that this coarse point of view is true on an average level. Trading strategies can also be mixtures of both *liquidity taking* and *liquidity providing* strategies. This is also a result of the drastic reduction of entry barriers in electronic trading systems. Financial markets are very easily accessible on a retail level. Thus, all market participants are able to perform all possible types of orders. Only the *market making* in terms of exchanges' rule sets is associated with institutions.

[17] Other important components of *market friction* are exchange fees for transactions and clearing.

Chapter 3

GPU accelerated Empirical Analysis

This chapter[1] covers general empirical features of financial markets which are called *empirical stylized facts*. After an introductory section which is focused on *Brownian motion* and the *random walk* we will present examples for the empirically[2] found evidence for deviations from a random walk behavior accompanied by an accelerated computation approach of GPU computing. GPU computing provides an incredible amount of computing power which has recently become accessible also for non-graphical use. A brief introduction for GPU computing will be provided in this chapter as well. In the last part of this chapter, financial return distributions will be investigated.

3.1 Brownian Motion and Random Walk

Brownian motion in one dimension is also called a Wiener[3] process and is a stochastic process which provides time consistency and stationarity with normally distributed positive and negative jumps. A stochastic process is stationary if and only if all distribution functions or probability densities are independent of shifts of the time axis [Voi10].

In order to simulate such a process in the context of financial markets, one can use a one dimensional random walk with discrete values in discrete time. This results in only an approximation of a one dimensional random walk. Let p_+ be the probability of increasing a discrete variable d by a constant increment Δd. Vice versa, p_- denotes the probability for decreasing d by Δd, as indicated in Fig. 3.1 with $p_+ + p_- = 1$. In addition, let n_+ be the number of increasing steps and n_- the number of decreasing steps. Then, the total number of steps N is given by

$$N = n_+ + n_-. \tag{3.1}$$

[1] Results of this chapter are published in [PVPS09a].

[2] Ancient Greek: ἐμπειρία — experience, knowledge

[3] **Norbert Wiener** (*26 November 1894 in Columbia, USA; †18 March 1964 in Stockholm, Sweden): an American pure and applied mathematician. Wiener was a pioneer in the study of stochastic and noise processes [Fou10].

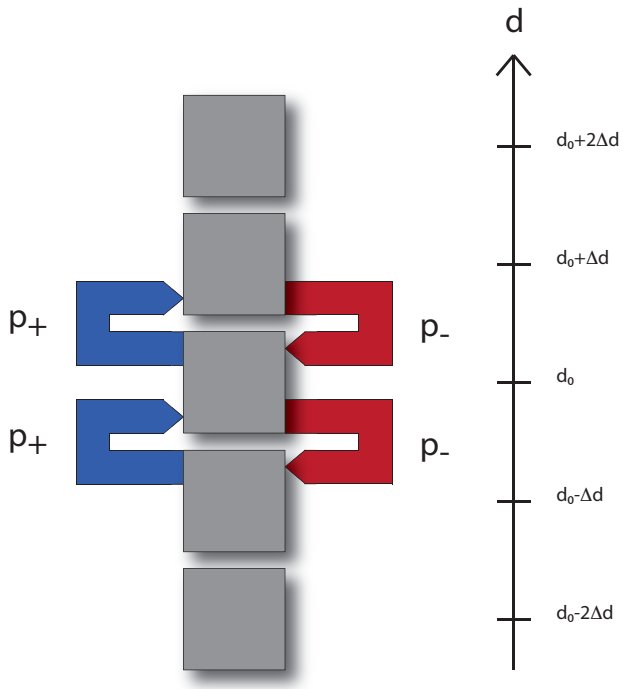

Figure 3.1: Sketch of a random walk in one dimension.

The probability of a unique sequence of n_+ increments and n_- decrements can be described with the N factors $p_+^{n_+} \cdot p_-^{N-n_+}$. If the probability is found to be d after N steps—where the starting value is d_0—then this results in

$$P(n_+, N) = \frac{N!}{(N-n_+)!\, n_+!}\, p_+^{n_+}\, p_-^{N-n_+} \tag{3.2}$$

according to the Binomial distribution as $d - d_0$ can be derived from n_+ and N due to

$$d - d_0 = n_+ - n_- = 2n_+ - N. \tag{3.3}$$

In the limit of large N, the Binomial distribution converges to the Gaussian distribution [Kre03] based on the law of large numbers. Thus, it is justified to approximate Brownian motion using such a random walk. An important modification of Brownian motion is geometric Brownian motion, which is the underlying stochastic process for the Black-Scholes model covered in chapter 1. Due to the continuous compounding of capital, a constant drift is inherent in stock prices. The drift parameter μ should be higher than the risk-free interest rate r as entrepreneurial activities are more risky from the shareholder's point of view. A continuously compounded investment with price S_I yields interest according to

$$\frac{\mathrm{d}S_I}{S_I(t)} = r\, \mathrm{d}t. \tag{3.4}$$

The drift component of the stock price S can be described by

$$\left(\frac{\mathrm{d}S}{S(t)}\right)_D = \mu\,\mathrm{d}t. \tag{3.5}$$

In addition, the stochastic component of the stock price can be modeled by

$$\left(\frac{\mathrm{d}S}{S(t)}\right)_S = \sigma\,\mathrm{d}W(t) \tag{3.6}$$

where $\mathrm{d}W(t)$ is the differential of the Wiener process and σ denotes the volatility, a quantity measuring the strength of statistical fluctuations (historical volatility and future volatility can be distinguished). In finance, the historical volatility most frequently refers to the standard deviation of the continuously compounded returns of an asset.

Combining the aforementioned components, we have the stochastic differential equation

$$\frac{\mathrm{d}S}{S(t)} = \mu\,\mathrm{d}t + \sigma\,\mathrm{d}W(t) \tag{3.7}$$

with parameters μ and σ known as geometric Brownian motion. Normal Brownian motion is given by $\mathrm{d}S(t) = \mu\,\mathrm{d}t + \sigma\,\mathrm{d}W(t)$. Geometric Brownian motion features log-normally distributed asset prices and normally distributed asset returns [PB00]. Furthermore, this model does not allow for negative asset prices which is a meaningful and valuable property. In addition, the absoluteness of price changes loses its importance.

In contrast to this geometric Brownian motion approach, financial price changes do not follow a Gaussian distribution as already mentioned in the introductory chapter 1. This aspect will be covered in detail in section 3.6 by analyzing the return distribution of a financial asset.

3.2 An Introduction to GPU Computing

Before focusing on the empirical differences between random walk behavior and financial market features, we introduce the concept of GPU computing.

In computer science applications and diverse interdisciplinary science fields such as computational physics or quantitative finance, the computational power requirements have continuously increased with time. The history of time series analysis has been especially driven by needs for computational power and simultaneously the opportunities arising from the use of it. Up to the present day, often a simplified assumption is used that price dynamics in financial time series obey random walk statistics—the discrete approximation of Brownian motion in one dimension—in order to simplify analytic calculations in financial applications.

However, such approximations, used for example in the option pricing model introduced by Black and Scholes [BS73] in 1973 as mentioned already in chapter 1, neglect

the real nature of financial market observables, and a large number of empirical deviations between financial market time series and models presuming only a random walk behavior have been observed and were calculated in the last decades [CB00, GPA+99, MS00, BP09, PB00]. However, in contrast to such calculations, which can be carried out with conventional computing facilities, a large computational power demand is driven by the quantitative hedge fund industry and also by modern market making, which requires primarily real time analytics. A market maker usually provides quotes for buying or selling a given asset. In the competitive environment of electronic financial markets, this cannot be done by a human market maker alone, especially if a large number of assets are quoted concurrently. The rise of the hedge fund industry in recent years and their interest in taking advantage of short time correlations boosted the real-time analysis of market fluctuations and the market micro-structure analysis in general, which is the study of the process of exchanging assets under explicit trading rules [O'H97], and which is studied intensively by the financial community [BHS95, BP05, BL98, BB04].

Such computing requirements, which can also be found in various interdisciplinary computer sciences such as computational physics (e.g., Monte Carlo- and molecular dynamics simulations [LB05, vMAF+08, KSRS08] or stochastic optimization [SK06]) make the use of high performance computing resources. This includes recent multicore computing solutions based on a shared memory architecture, which are accessible by OpenMP [DM98] or MPI [GFB+04] and can be found in recent personal computers. Furthermore, distributed computing clusters with homogeneous or heterogeneous node structures are available in order to parallelize a given algorithm by separating it into various sub-algorithms.

However, a recent trend in computer science and related fields is general purpose computing on graphics processing units (GPUs), which can yield impressive performance. Some applications have already been realized in computational physics[4]. With multiple cores connected by high memory bandwidth, today's GPUs offer resources for non-graphics processing. In the beginning, GPU programs used C-like programming environments for kernel execution such as OpenGL shading language [Ros04] or C for graphics (Cg) [FK03]. The common unified device architecture (CUDA) [NVI08a] is a conventional programming approach making use of the unified shader design of recent GPUs from NVIDIA corporation. The programming interface allows for implementing an algorithm using the standard C programming language without any knowledge of the native programming environment. A comparable concept "Close To the Metal" (CTM) [ATI06] was introduced by Advanced Micro Devices Inc. for ATI graphics cards. The computational power of consumer graphics cards roughly exceeds that of a central processing unit (CPU) by 1–2 orders of magnitude. A conventional CPU nowadays provides a peak performance of roughly 20×10^9 floating point operations per second (FLOPS) [vMAF+08]. The consumer graphics card NVIDIA GeForce GTX 280 reaches a theoretical peak performance of 933×10^9 FLOPS. If one were to try to realize the computational power of one GPU with a cluster of several CPUs, a much larger amount

[4] [TMB+05, vMAF+08, SPF+07, SEE+03, PZBG07, BBPZ08, LWK03, ALT08, YWC07]

of electrical power would be required. A GTX 280 graphics card exhibits a maximum power consumption of 236 W [NVI08b], while a recent Intel CPU consumes roughly 100 W.

We apply this general-purpose graphics processing unit (GPGPU) technology to methods of time series analysis, which includes determination of the Hurst exponent and equilibrium autocorrelation function—well-known empirical stylized facts of financial markets. Furthermore, we compare the recent GPU generation with the previous one. All methods are applied to a high frequency data set of the Euro-Bund futures contract and German DAX future (FDAX) contract traded at the electronic derivatives exchange Eurex.

The remaining parts of this chapter are organized as follows. In section 3.3, a brief overview of key facts and properties of the GPU architecture is provided in order to clarify implementation constraint details for the following sections. A GPU accelerated Hurst exponent estimation can be found in section 3.4. In section 3.5, the equilibrium autocorrelation function is implemented on a GPU. In each of these sections, the performance of the GPU code as a function of parameters is first evaluated for a synthetic time series and compared to the performance on a CPU. Then the time series methods are applied to a financial market time series followed by a discussion of numerical errors.

3.3 GPU Device Architecture

Table 3.1: Key facts and properties of the applied consumer graphics cards. The theoretical acceleration factor between GeForce 8800 GT and GeForce GTX 280 is given by the difference in number of cores $\times$ clock rate.

	GeForce 8800 GT	GeForce GTX 280
Global memory	512 MB	1024 MB
Number of multiprocessors	14	30
Number of cores	112	240
Constant memory	64 kB	64 kB
Shared memory per block	16 kB	16 kB
Registers available per block	8192	16384
Warp size	32	32
Clock rate	1.51 GHz	1.30 GHz

In order to provide and discuss information concerning implementation details on a GPU for time series analysis methods, key aspects of the GPU device architecture are briefly summarized in this section. As mentioned in the introduction, we use the

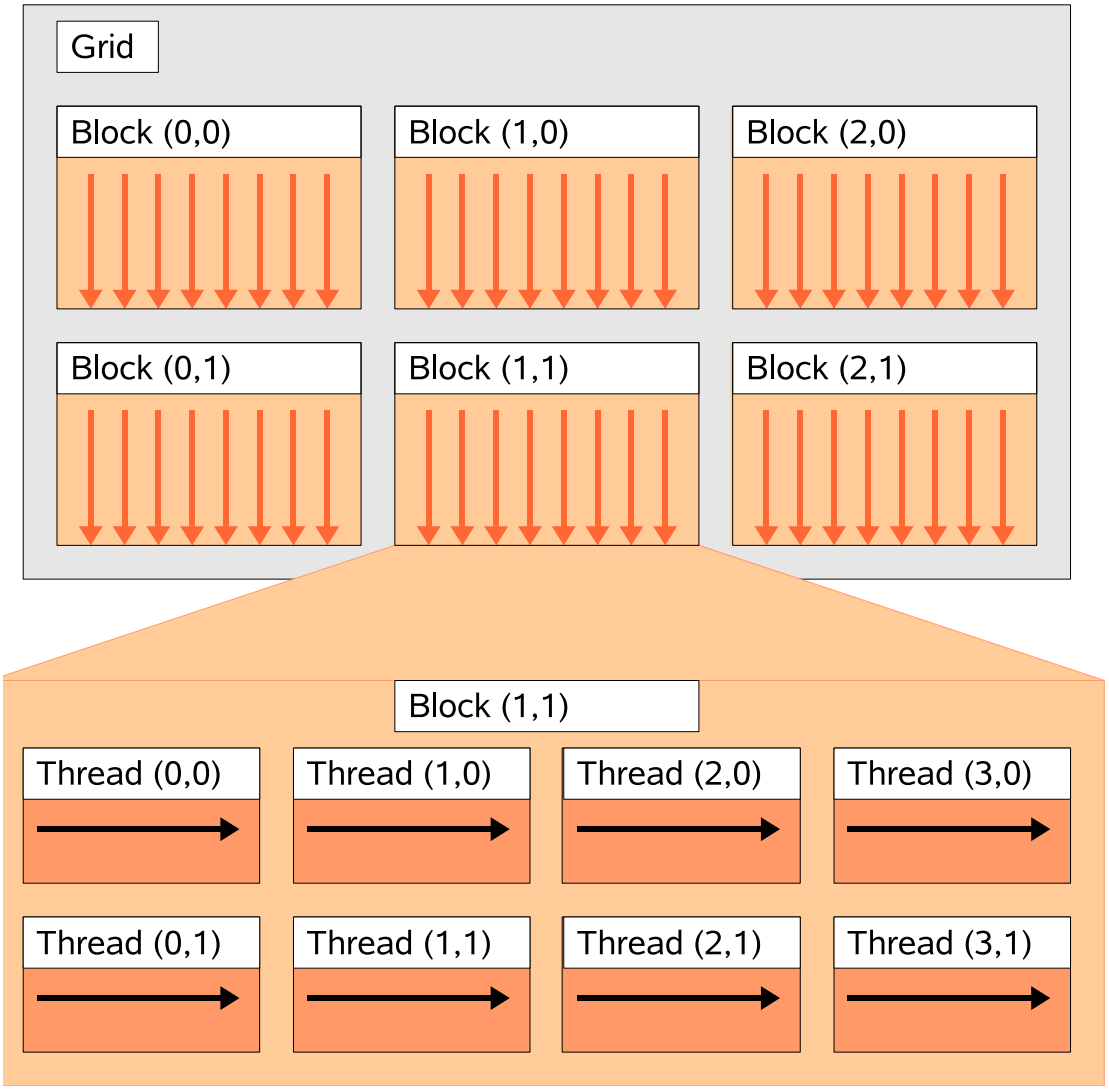

Figure 3.2: Schematic visualization of the grid of thread blocks for a two dimensional thread and block structure.

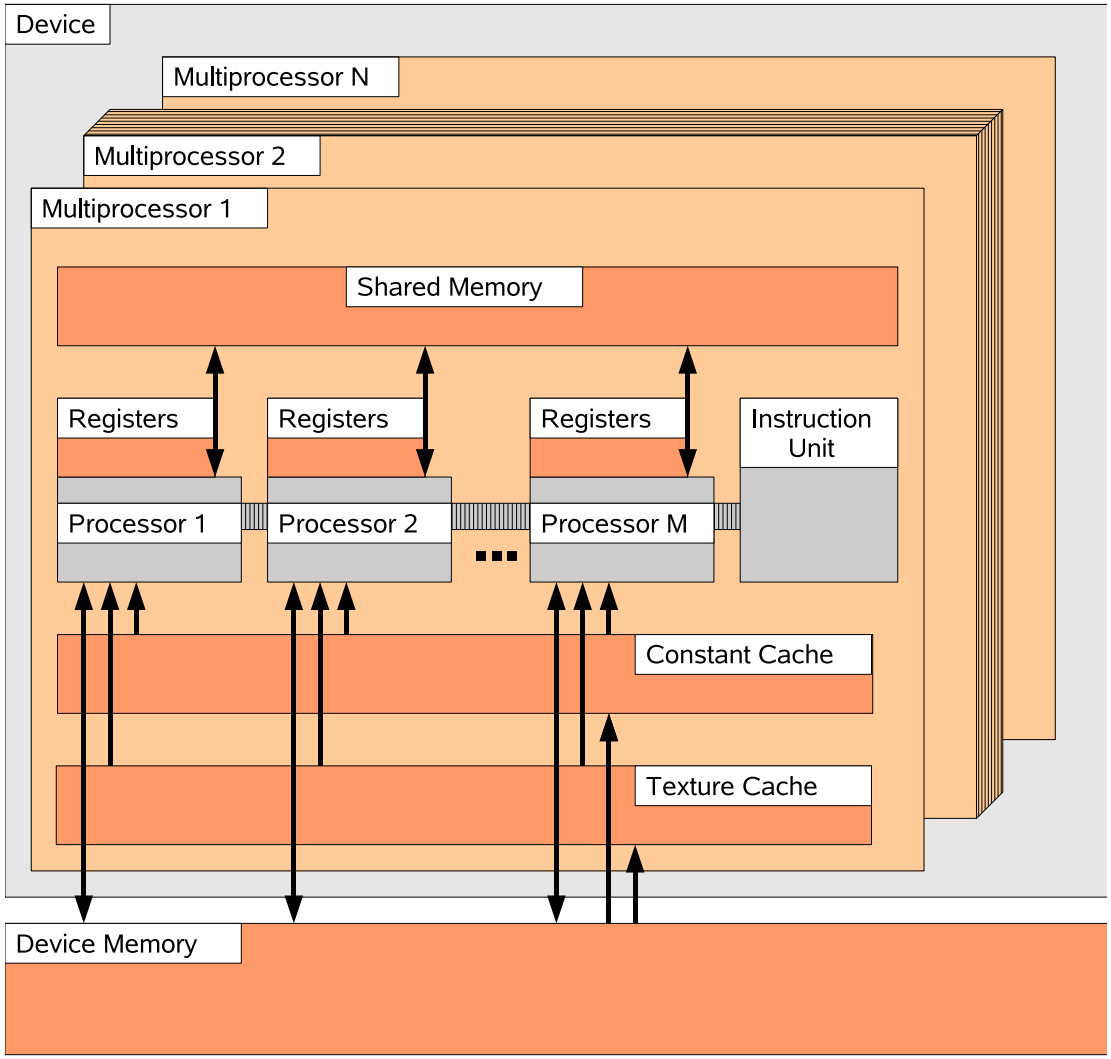

Figure 3.3: Schematic visualization of a GPU multiprocessor with on-chip shared memory.

compute unified device architecture (CUDA), which allows for implementation of algorithms using standard C with CUDA specific extensions. Thus, CUDA issues and manages computations on a GPU as a data-parallel computing device.

The graphics card architecture used in recent GPU generations is built around a scalable array of streaming multiprocessors (MPs) [NVI08a]. One such multiprocessor contains, amongst others, eight scalar processor cores, a multi-threaded instruction unit, and shared memory, which is located on-chip. When a C program using CUDA extensions and running on the CPU invokes a GPU kernel, which is a synonym for a GPU function, many copies of this kernel—known as threads—are enumerated and distributed to the available MPs, where their execution starts. For such an enumeration and distribution, a kernel grid is subdivided into blocks and each block is subdivided into various threads as illustrated in Fig. 3.2 for a two-dimensional thread and block structure. The threads of a thread block are executed concurrently in the vacant MPs. In order to manage a large number of threads, a single-instruction multiple-thread (SIMT) unit is used. An MP maps each thread to one scalar processor core and each scalar thread works independently of all the others. Threads are created, managed, scheduled, and executed by this SIMT unit in groups of 32 threads. Such a group of 32 threads forms a warp, which is executed on the same MP. If the threads of a given warp diverge via a data-induced conditional branch, each branch of the warp is executed serially and the processing time of this warp consists of the sum of the branches' processing times.

As shown in Fig. 3.3, each MP of the GPU device contains several local 32-bit registers per processor, memory which is shared by all scalar processor cores in an MP. Furthermore, constant and texture cache are available, which is also shared on the MP. In order to allow for reducing the number of involved MPs, the slower global memory can be used, which is shared among all MPs and is also accessible by the C function running in the CPU. Please note, that the GPU's global memory is still roughly 10 times faster than current main memory of personal computers. Detailed specifications of the consumer graphics cards 8800 GT and GTX 280 used in this study0 can be found in Table 3.1. Furthermore note that a GPU device only supports single-precision floating-point operations, with the exception of the most modern graphic cards starting with the GTX 200 series. However, the IEEE-754 standard for single precision numbers is not completely realized. In contrast, the GTX 200 series also supports double-precision floating-point numbers. However, each MP features only one double-precision processing core and so, the theoretical peak performance is dramatically reduced for double-precision operations. Further informations about the GPU device properties and CUDA can be found in [NVI08a].

3.4 Hurst Exponent

The Hurst exponent H [MH04] provides detailed information on the relative tendency of a stochastic process. A Hurst exponent of $H < 0.5$ indicates an anti-persistent behavior of the analyzed process, which means that the process is dominated by a mean

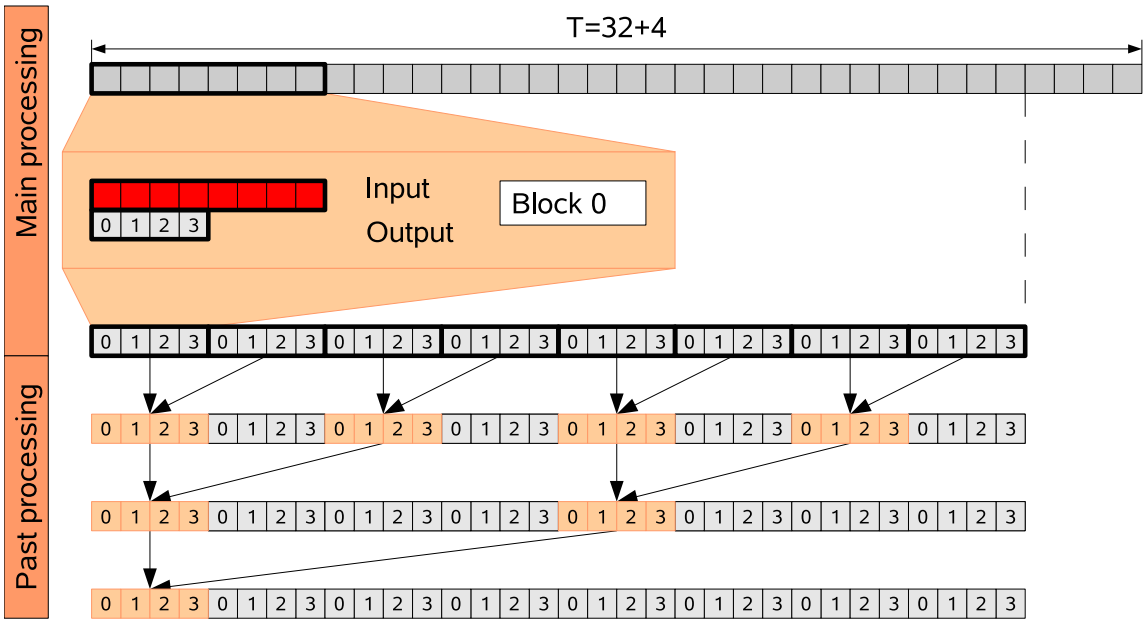

Figure 3.4: Schematic visualization of the determination of the Hurst exponent on a GPU architecture for $T = 32 + 4$ and $\Delta t_{\max} = 4$. The calculation is split into various processing steps to ensure that their predecessors are completed.

reversion tendency. $H > 0.5$ mirrors a super-diffusive behavior (persistent behavior) of the underlying process. Large values tend to be followed by large values, small values by small values. If the deviations of the current values of the time series from their mean value are independent, which corresponds to a random walk behavior, a Hurst exponent of $H = 0.5$ is obtained.

The Hurst exponent H was originally introduced by Harold Edwin Hurst [Hur51], a British government administrator. He studied records of the Nile river's volatile rain and drought conditions and noticed interesting coherences for flood periods. Harold Edwin Hurst observed in the eight centuries of records that there was a tendency for a year with good flood condition to be followed by another year with good flood conditions and vice versa. Nowadays, the Hurst exponent is well studied in context of financial markets [PPS08, DW00, Aus00, CCS04, GZ09]. Typically, an anti-persistent behavior can be found on short time scales due to the non-zero gap between offer and demand. On medium time scales, a super-diffusive behavior can be detected [Aus00]. On long time scales, a diffusive regime is reached, due to the law of large numbers.

For a time series $p(t)$ with $t \in \{1, 2, \ldots, T\}$, the time lag dependent Hurst exponent $H_q(\Delta t)$ can be determined by the general relationship

$$\langle |p(t + \Delta t) - p(t)|^q \rangle^{1/q} \propto \Delta t^{H_q(\Delta t)} \tag{3.8}$$

with the time lag $\Delta t \ll T$ and $\Delta t \in \mathbb{N}$. The brackets $\langle \ldots \rangle$ denote the expectation value. Apart from Eq. (3.8), there are also other calculation methods, e.g., rescaled range analysis [MH04]. We present the Hurst exponent determination implementation on a GPU for $q = 1$ and use $H(\Delta t) \equiv H_{q=1}(\Delta t)$. The process to be analyzed is a synthetic anti-correlated random walk, which was introduced in [PPS08]. This process emerges from the superposition of two random walk processes with different time scale characteristics. Thus, a parameter dependent negative correlation at time lag one can be observed. As a first step, one has to allocate memory on the GPU device's global memory for the time series, intermediate results, and final results. In a first approach, the time lag dependent Hurst exponent is calculated up to $\Delta t_{\max} = 256$. In order to simplify the reduction process of the partial results, the overall number of time steps T has to satisfy the condition

$$T = (2^{\alpha} + 1) \cdot \Delta t_{\max}, \tag{3.9}$$

with α being an integer number called the length parameter of the time series. The number of threads per block – known as block size – is equivalent to $\Delta t_{\max}$. The array for intermediate results has length T as well, whereas the array for the final results contains $\Delta t_{\max}$ entries. After allocation, the time series data have to be transferred from main memory to the GPU's global memory. When this step is completed, the main calculation part can start. As illustrated in Fig. 3.4 for block 0, each block, which contains $\Delta t_{\max}$ threads each, loads $\Delta t_{\max}$ data points of the time series from global memory to shared memory. In order to realize such a high-performance loading process, each thread[5] loads one value and stores this value in the array located in shared memory, which can be accessed by all the threads of a block. Analogously, each block also loads the next $\Delta t_{\max}$ entries. In the main processing step, each thread is in charge of one specific time lag. Thus, each thread is responsible for a specific value of Δt and summarizes the terms $|p(t + \Delta t) - p(t)|$ in the block subsegment of the time series. As the maximum time lag is equivalent to the maximum number of threads and as the maximum time lag is also equivalent to half of the data points loaded per block, all threads have to sum the same number of addends resulting in a uniform workload in the graphics card. However, as it is only possible to synchronize threads within a block, and native block synchronization does not exist, partial results of each block have to be stored in block-dependent areas of the array for intermediate results, as shown in Fig. 3.4. The termination of the GPU kernel function ensures that all blocks are executed. In a post processing step, the partial arrays have to be reduced. This is realized by a binary tree structure, as indicated in Fig. 3.4. After this reduction, the resulting values can be found in the first $\Delta t_{\max}$ entries of the intermediate array and a final processing kernel is responsible for normalization and gradient calculation. The source code of these GPU kernel functions can be found in the appendix of [PVPS09a].

For the comparison between CPU and GPU implementation, we use an Intel Core 2 Quad CPU (Q6700) with 2.66 GHz and 4096 kB cache size, of which only one core

[5] Thread and block IDs are accessible in standard C language via built-in variables.

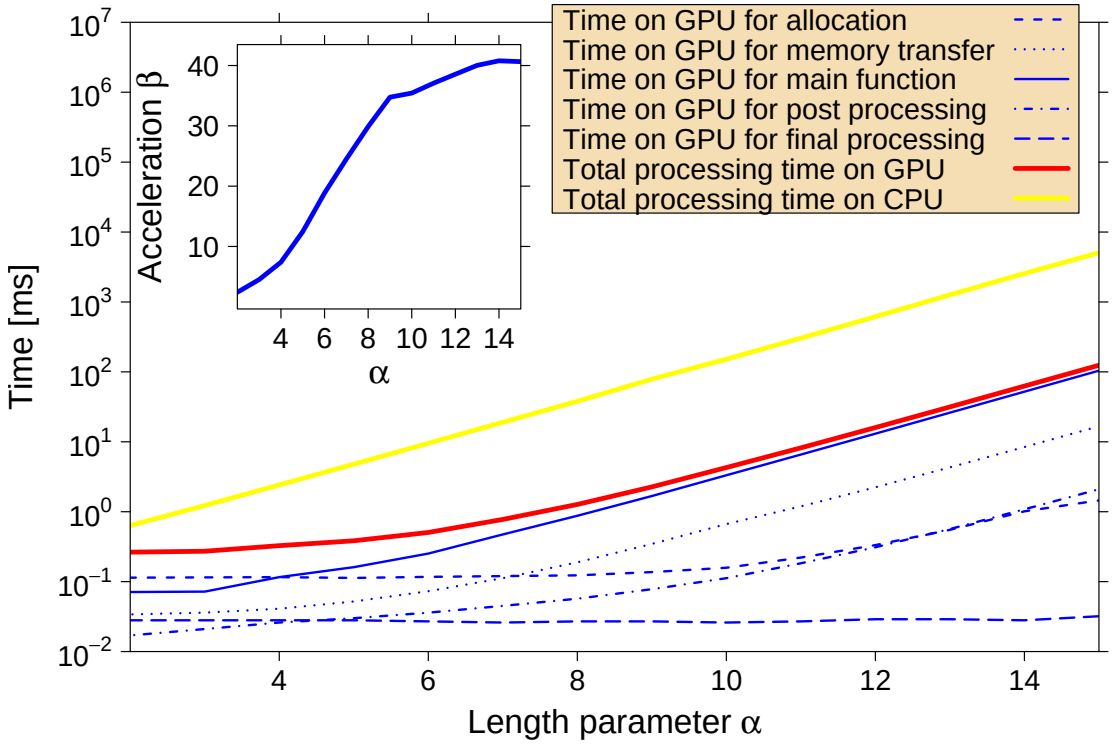

Figure 3.5: Processing times vs. length parameter α for the calculation of the Hurst exponent $H(\Delta t)$ on GPU and CPU for $\Delta t_{\max} = 256$. The time series contains $T = (2^{\alpha} + 1) \cdot \Delta t_{\max}$ data points. The consumer graphics card 8800 GT is used as GPU device. The total processing time on the GPU can be broken into allocation time, time for memory transfer, time for main processing, time for post processing, and time for final processing. The acceleration factor β is shown in the inset. A maximum acceleration factor of roughly 40 can be obtained. Furthermore, at $\alpha = 9$ there is a break of the slope of the acceleration which is influenced by cache size effects.

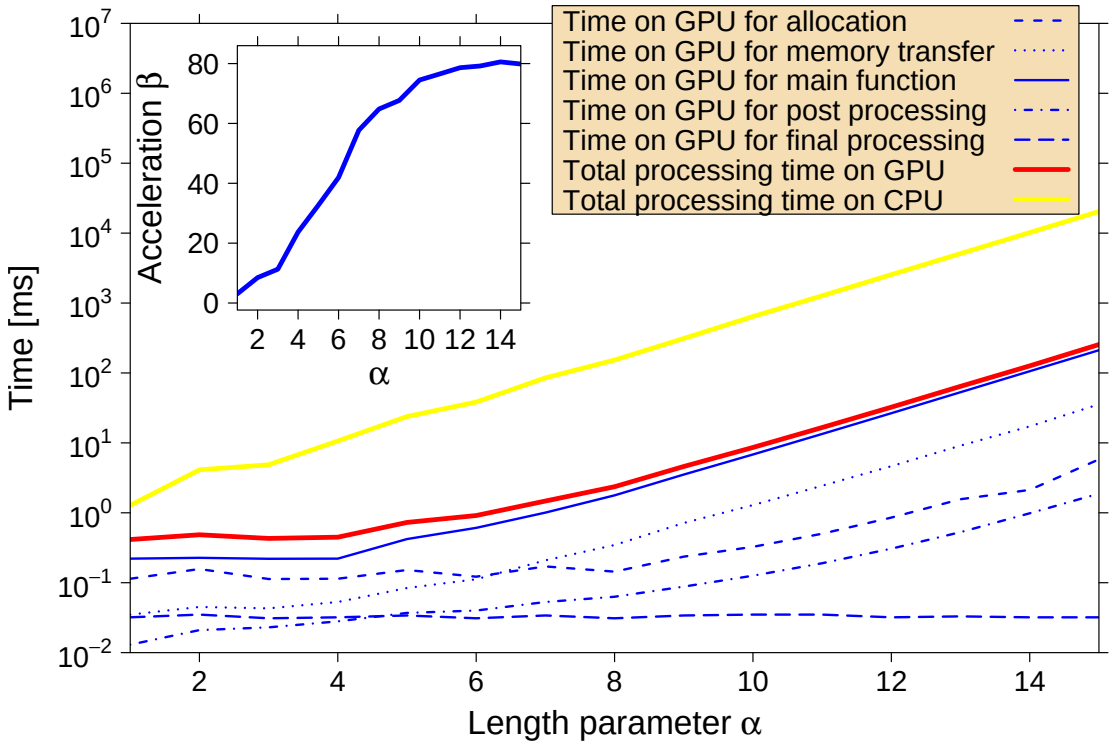

Figure 3.6: Processing times for the calculation of the Hurst exponent $H(\Delta t)$ on GPU and CPU for $\Delta t_{\max} = 512$. These results are obtained on the GTX 280. The total processing time on GPU can also be slit into allocation time, time for memory transfer, time for main processing, time for post processing, and time for final processing. A maximum acceleration factor of roughly 80 can be reached, which is shown in the inset.

is used. The standard C source code executed on the host is compiled with the gcc compiler (version 4.2.1). The results for $\Delta t_{\max} = 256$ and the consumer graphics card 8800 GT can be found in Fig. 3.5. The acceleration factor β, which is shown in the inset, reaches a maximum value of roughly 40, and is determined by the relationship

$$\beta = \frac{\text{Total processing time on CPU}}{\text{Total processing time on GPU}}. \tag{3.10}$$

A smaller speed-up factor can be measured for small values of α, as the relative fraction of allocation time and time for memory transfer is larger than the time needed for the calculation steps. The corresponding analysis for the GTX 280 yields a larger acceleration factor β of roughly 70. If we increase the maximum time lag $\Delta t_{\max}$ to 512, which is only possible for the GTX 280, a maximum speed-up factor of roughly 80 can be achieved, as shown in Fig. 3.6. This indicates that $\Delta t_{\max} = 512$ leads to a higher efficiency on the GTX 280.

At this point, we can also compare the ratio between the performances of 8800 GT and GTX 280 for our application to the ratio of theoretical peak performances. The latter is given as the number of cores multiplied by the clock rate, which amounts to roughly 1.84. If we compare the total processing times on these GPUs for $\alpha = 15$ and $\Delta t_{\max} = 256$, we obtain an empirical performance ratio of 1.7. If we use the acceleration factors for $\Delta t_{\max} = 256$ on the 8800 GT and for $\Delta t_{\max} = 512$ on the GTX 280 for comparison, we get a value of 2.

Following this performance analysis, we apply the GPU implementation to real financial market data in order to determine the Hurst exponent of the Euro-Bund futures contract traded on the European exchange (Eurex). We also validate the accuracy of the GPU calculations by quantifying deviations from the calculation on a CPU. The Euro-Bund futures contract (FGBL) is a financial derivative. As noted earlier, a futures contract is a standardized contract to buy or sell a specific underlying instrument at a proposed date in the future (called the expiration time of the futures contract) at a specified price. The underlying instruments of the FGBL contract are long-term debt instruments issued by the Federal Republic of Germany with remaining terms of 8.5 to 10.5 years and a coupon of 6 percent. We use the Euro-Bund futures contract with expiration time June 2007. The time series shown in Fig. 3.7 contains 1,051,982 trades, recorded from 8 March 2007 to 7 June 2007. In all presented calculations of the FGBL time series on the GPU, α is fixed to 11. Thus, the data set is limited to the first $T = 1,049,088$ trades in order to fit the data set length to the constraints of the specific GPU implementation. In Fig. 3.8, the time lag dependent Hurst exponent $H(\Delta t)$ is presented. On short time scales, the well-documented anti-persistent behavior is detected. On medium time scales, small evidence is observed, that the price process reaches a super-diffusive regime. For long time scales the price dynamics tend to random walk behavior ($H = 0.5$), which is also shown for comparison. The relative error

$$\epsilon = \left| \frac{H_{\text{GPU}}(\Delta t) - H_{\text{CPU}}(\Delta t)}{H_{\text{CPU}}(\Delta t)} \right| \tag{3.11}$$

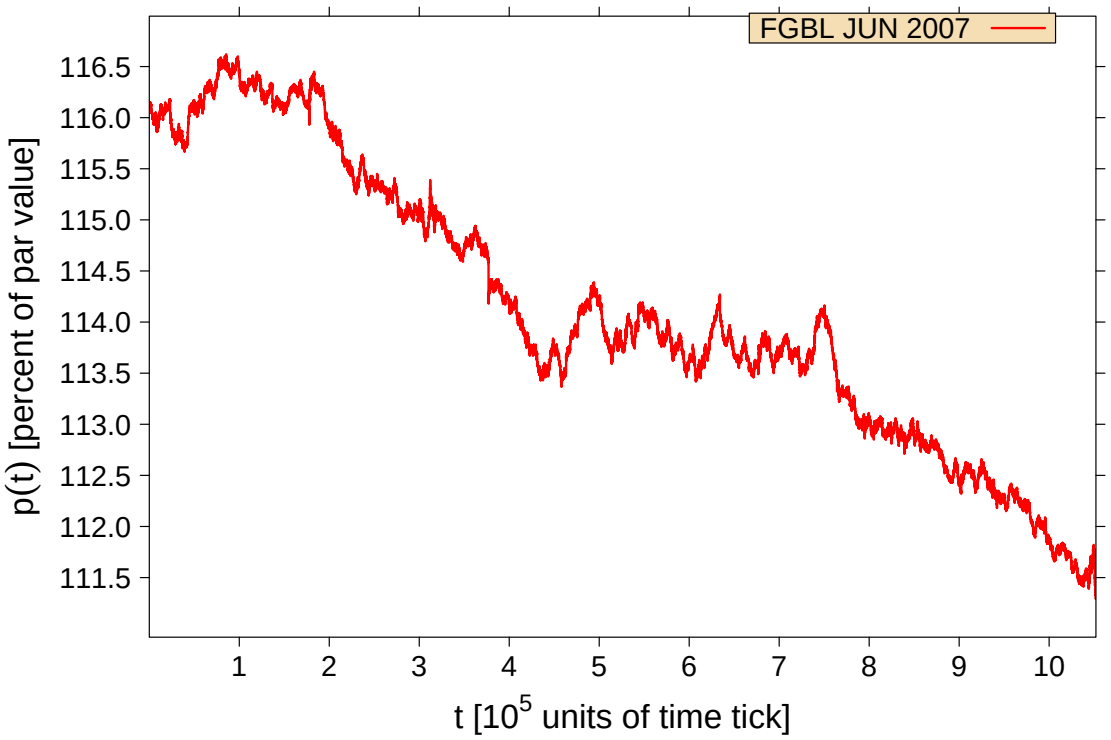

Figure 3.7: High frequency financial time series of the Euro-Bund futures contract (FGBL) with expiration time June 2007 traded at the European Exchange (Eurex). The time series contains 1,051,982 trades recorded from 8 March 2007 to 7 June 2007. The price is shown in units of percent of par value. In bond markets, the par value (as stated on the face of the bond) is the amount that the issuing firm is to pay to the bond holder at the maturity date. In calculations of the FGBL time series on a GPU presented here, α is fixed to 11. Thus, the data set is limited to the first $T = 1,049,088$ trades in order to fit the data set length to the constraints of the specific GPU implementation.

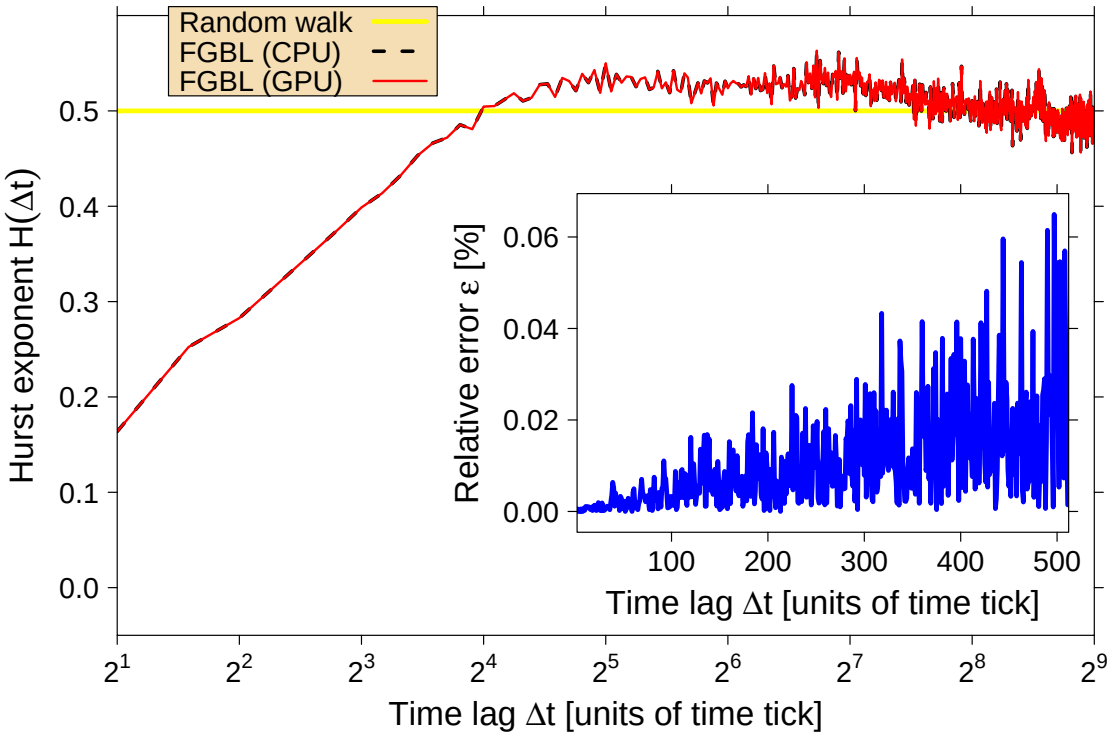

Figure 3.8: Hurst exponent $H(\Delta t)$ in dependence of time lag Δt calculated on CPU and GPU. Additionally, the theoretical Hurst exponent of a random walk process ($H = 0.5$) is included for comparison. One can clearly see the well-known anti-persistent behavior of the FGBL time series on short time scales ($\Delta t < 2^4$ time ticks). Furthermore, evidence is given that the process reaches a slightly super-diffusive region ($H \approx 0.525$) on medium time scales (2^4 time ticks $< \Delta t < 2^7$ time ticks). On long time scales, an asymptotic random walk behavior can be found. In order to quantify deviations from calculations on a CPU, the relative error ϵ (see main text) is presented for each time lag Δt in the inset. It is typically smaller than 10^{-3}.

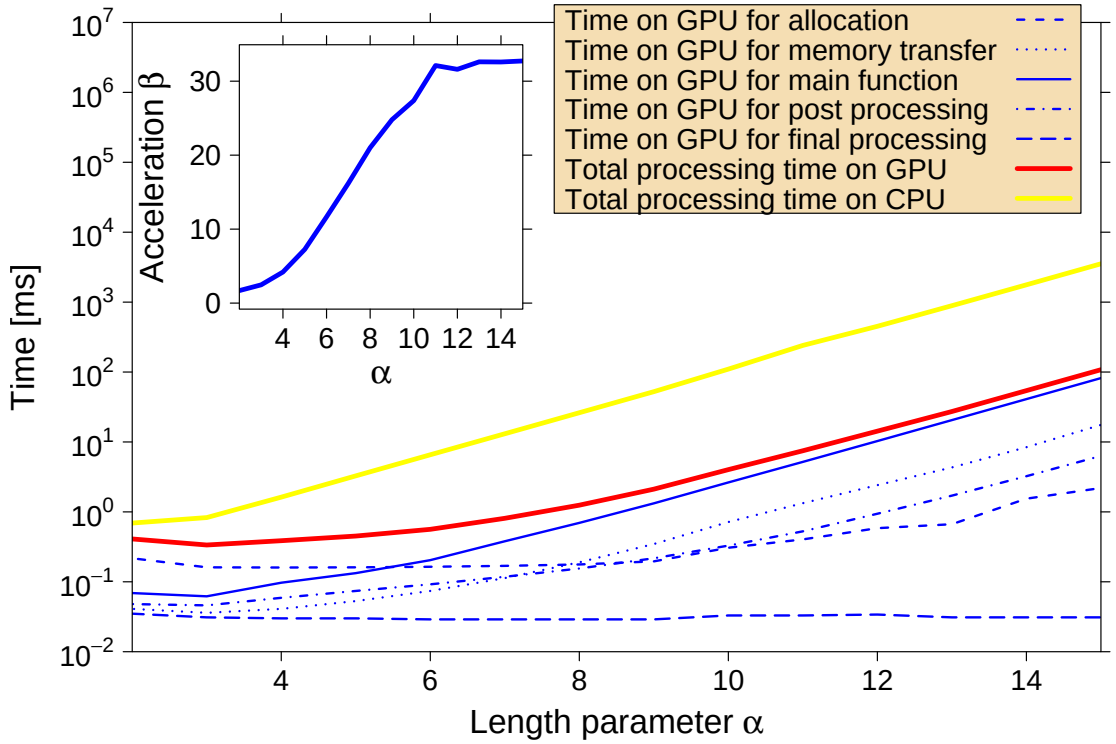

Figure 3.9: Processing times for the calculation of the equilibrium autocorrelation function $\rho(\Delta t)$ on GPU and CPU for $\Delta t_{\max} = 256$. The graphics card 8800 GT is used as GPU device. The total processing time on GPU is broken into allocation time, time for memory transfer, time for main processing, time for post processing, and time for final processing. The acceleration factor β is shown as inset. A maximum acceleration factor of roughly 33 can be obtained.

shown in the inset of figure 3.8 is smaller than one-tenth of a percent.

3.5 Equilibrium Autocorrelation

The autocorrelation function is a widely used concept for determining dependencies within a time series. The autocorrelation function is given by the correlation between the time series and the time series shifted by the time lag Δt through

$$\rho(\Delta t) = \frac{\langle p(t) \cdot p(t+\Delta t)\rangle - \langle p(t)\rangle\langle p(t+\Delta t)\rangle}{\sqrt{\langle p(t)^2\rangle - \langle p(t)\rangle^2}\sqrt{\langle p(t+\Delta t)^2\rangle - \langle p(t+\Delta t)\rangle^2}}. \tag{3.12}$$

For a stationary time series, Eq. (3.12) reduces to

$$\rho(\Delta t) = \frac{\langle p(t) \cdot p(t+\Delta t)\rangle - \langle p(t)\rangle^2}{\langle p(t)^2\rangle - \langle p(t)\rangle^2}, \tag{3.13}$$

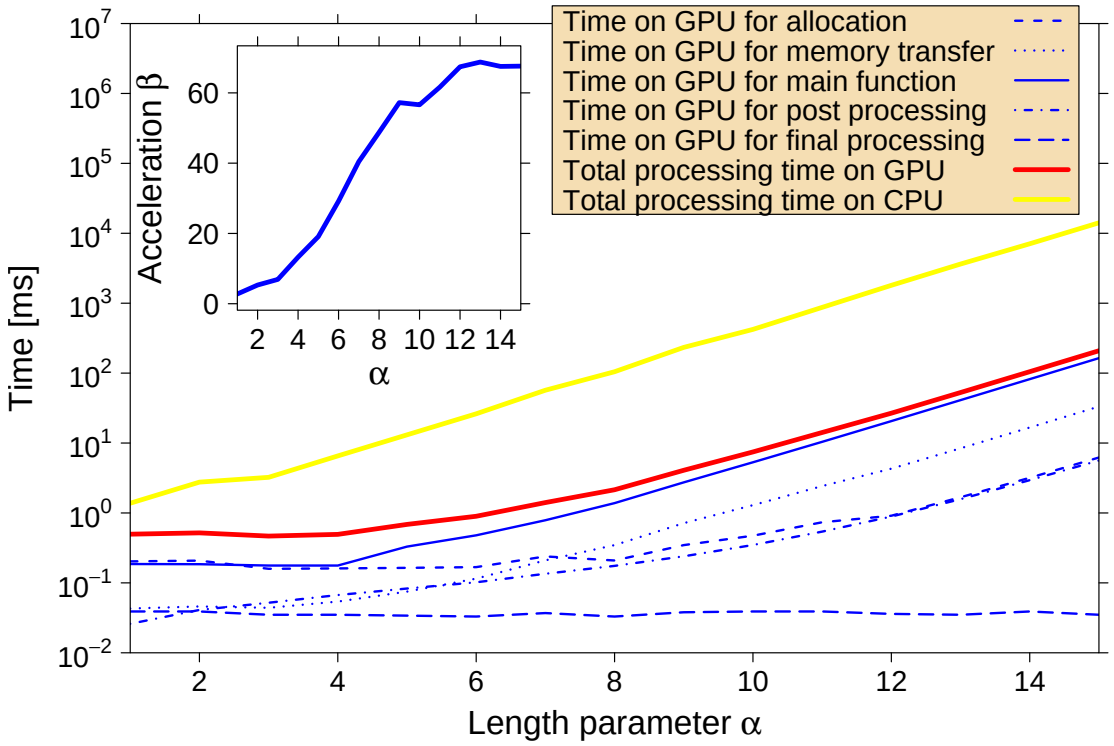

Figure 3.10: Processing times for the calculation of the equilibrium autocorrelation function $\rho(\Delta t)$ on GPU and CPU for $\Delta t_{\max} = 512$. The GTX 280 is used as GPU device. A maximum acceleration factor of roughly 68 can be obtained.

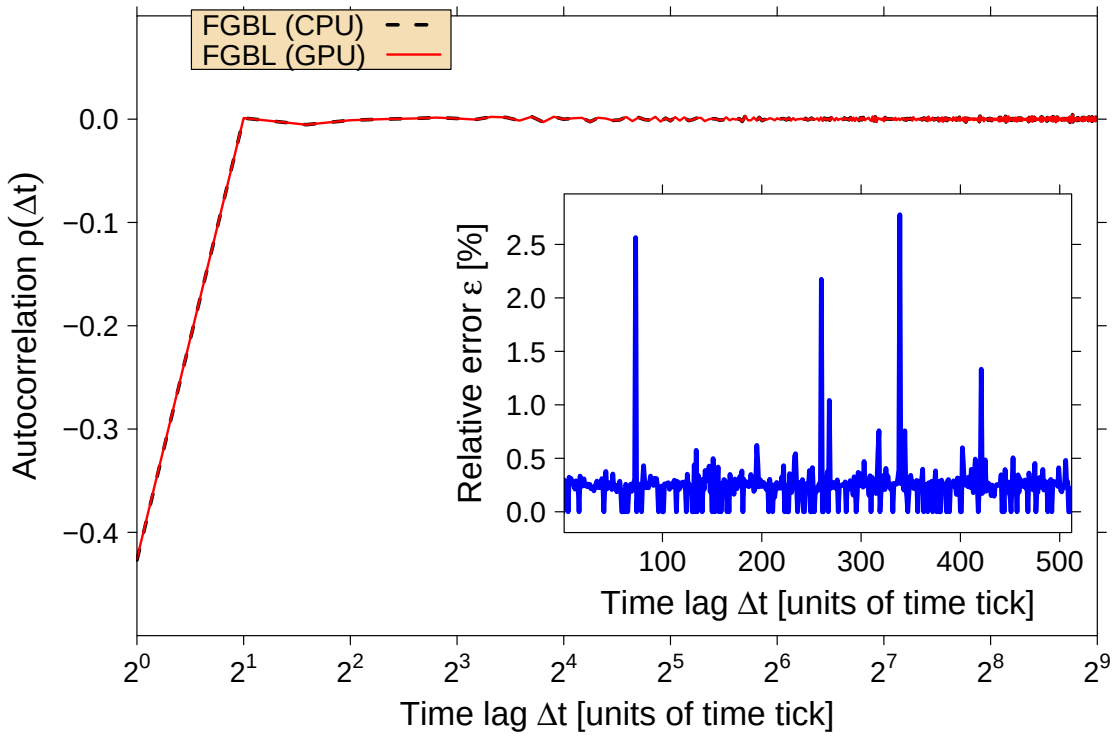

Figure 3.11: Equilibrium autocorrelation function $\rho(\Delta t)$ in dependence of time lag Δt calculated on CPU and GPU. One can clearly see the well-known negative autocorrelation of financial time series at time lag $\Delta t = 1$ also in this figure. In order to quantify deviations from calculations on a CPU, the relative error ϵ is presented for each time lag Δt in the inset. The relative error is always smaller than 3×10^{-2}.

as the mean value and the variance stay constant for a stationary time series, i.e., $\langle p(t) \rangle = \langle p(t + \Delta t) \rangle$ and $\langle p(t)^2 \rangle = \langle p(t + \Delta t)^2 \rangle$.

It can be observed that the autocorrelation function of price changes for a financial time series exhibits a significant negative value for a time lag of one tick, whereas it vanishes for time lags $\Delta t > 1$. Furthermore, the autocorrelation of absolute price changes or squared price changes, which is related to the volatility of the price process, decays slowly [PGPS07]. In order to implement Eq. (3.13) on a GPU architecture, steps similar to those covered in section 3.4 are necessary. The calculation of the time lag dependent part $\langle p(t) \cdot p(t + \Delta t) \rangle$ is handled analogously to the determination of the Hurst exponent on the GPU. The input time series, which is transferred to the GPU's main memory, does not contain prices but price changes. However, in addition one needs the results for $\langle p(t) \rangle$ and $\langle p(t)^2 \rangle$. For this purpose, an additional array of length T is allocated, in which a GPU kernel function stores the squared values of the time series. Then, time series and squared time series are decomposed with the same binary tree reduction process as in section 3.4. However, as this procedure produces arrays of length $\Delta t_{\max}$, one has to sum these values in order to obtain $\langle p(t) \rangle$ and $\langle p(t)^2 \rangle$.

The processing times for determining the autocorrelation function for $\Delta t_{\max} = 256$ on a CPU and a 8800 GT can be found in Fig. 3.9. Here we find that allocation and memory transfer dominate the total processing time on the GPU for small values of α and thus, only a fraction of the maximum acceleration factor $\beta \approx 33$ (shown as an inset) can be reached. Using the consumer graphics card GTX 280, we obtain a maximum speed-up factor of roughly 55 for $\Delta t_{\max} = 256$ and 68 for $\Delta t_{\max} = 512$ as shown in Fig. 3.10. In Fig. 3.11, the autocorrelation function of the FGBL time series is shown. For a time lag of one, the time series exhibits a large negative autocorrelation, $\rho(\Delta t = 1) = -0.43$. In order to quantify deviations between GPU and CPU calculations, the relative error ϵ is presented in the inset of Fig. 3.11. Note that small absolute errors can cause relative errors of up to three percentage point because the values $\rho(\Delta t > 1)$ are close to zero.

For some applications, it is interesting to study larger maximum time lags of the autocorrelation function. To do this on the basis of our GPU implementation, the program code must be modified in the following way. So far, each thread was responsible for a specific time lag Δt. In a modified implementation, each thread would be responsible for more than one time lag in order to realize a maximum time lag, which is a multiple of the maximum number of 512 threads per block. This way, one obtains a maximum speed-up factor of roughly 84 for $\Delta t_{\max} = 1024$ using the GTX 280.

3.6 Price Change Distributions

Last but not least we investigate return distributions in this section. The distribution of financial returns is of great interest in econophysics as the far tails of a return distribution differ significantly from the shape indicated by a Gaussian distribution [LFM03, GGPS03, PGGS02] and thus from that of a random walk. However, as the

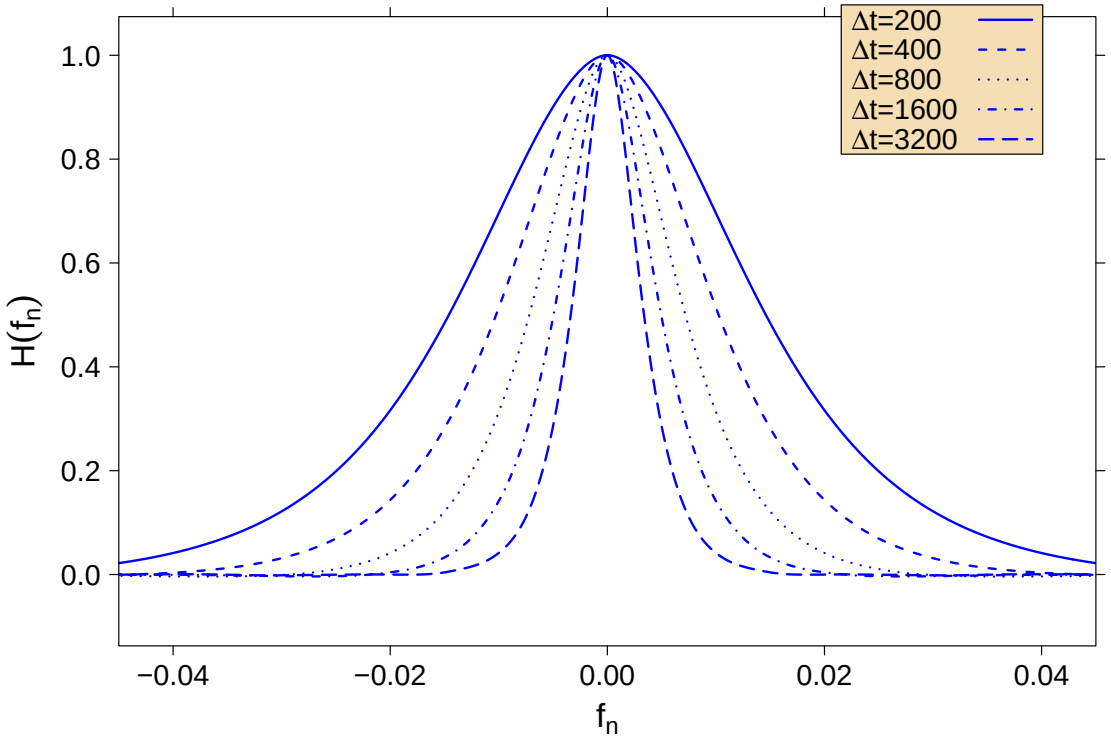

Figure 3.12: (a) Fourier transformed price change distributions of the FDAX time series comprising 37,933,953 transactions for various values of Δt.

computational requirements are not overwhelming, a conventional approach is used without a specific GPU implementation.

In general, we study returns

$$\Delta \ln p\,(t, \Delta\tau) = \ln p\,(t + \Delta\tau) - \ln p\,(t) \tag{3.14}$$

derived from a financial time series where Δp is the absolute price change and Δt is the time lag between two transaction prices—the time lag Δt is measured in units of *ticks* which means the temporal indices of successive individual transactions. Equivalent to that description, one can use

$$\Delta p\,(t, \Delta\tau) = p\,(t + \Delta\tau) - p\,(t) \tag{3.15}$$

if there are no large price movements and if the analysis focuses only on short time scales [PB00]. For the random walk, the probability distribution functions of price changes are Gaussian distributions. It is possible to scale the various Gaussian functions to each other which belong to various time lags Δt. For this rescaling, one can use the time lag dependent Hurst exponent. The rescaled distribution curves $P\,(\Delta p) \cdot \Delta\tau^H$, dependent on $\Delta p / \Delta\tau^H$, are congruent for all time lags Δt, due to $H \equiv 0.5$. For financial market returns, this is not possible due to their fat-tailed distributions.

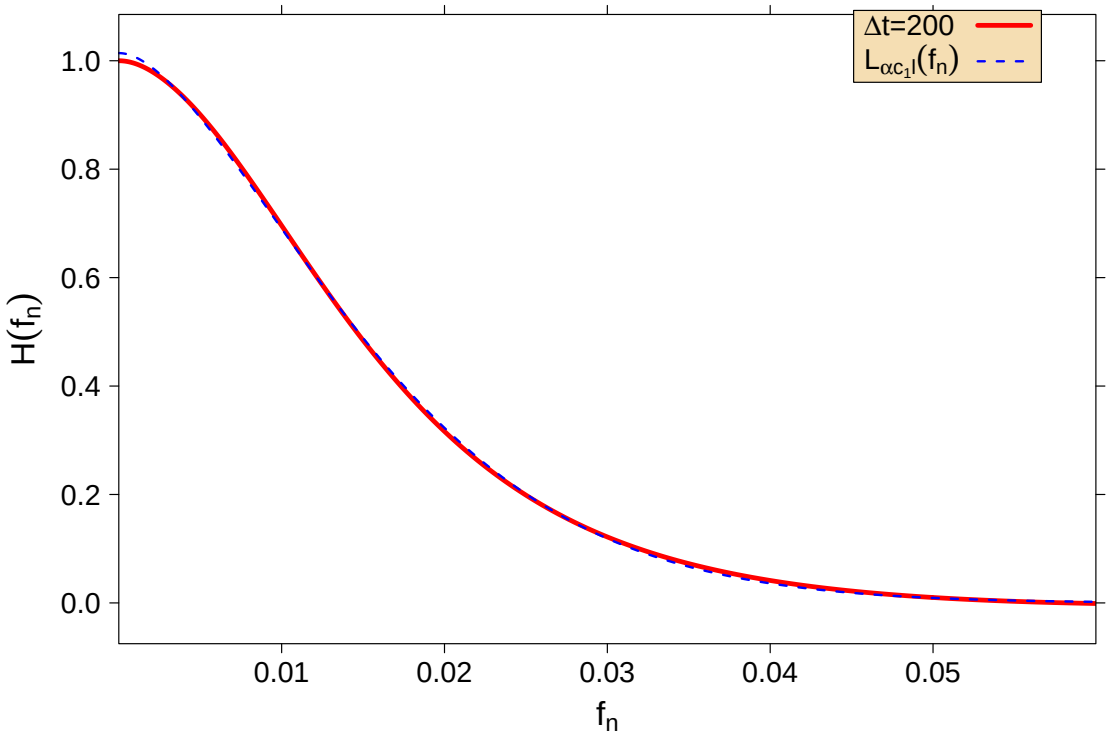

Figure 3.13: Fit of $\Lambda_{\alpha,c_1,l}(f_n)$ for $\Delta\tau = 200$.

We consider a German DAX future (FDAX) data set recorded from 1991 to 2003 for a detailed empirical analysis of financial return distributions. The data set contains 37,933,953 transactions. For the modeling of empirical return distributions, the *truncated Lévy flight* has been proposed [MS00, PB00, Voi10]. A Lévy stable non-Gaussian distribution satisfies scaling relations but has an infinite variance [MS00]. A truncated Lévy flight (TLF) is a stochastic process with finite variance and scaling relations in a large but finite interval. This process was proposed the R. N. Mantegna and H. E. Stanley [MS94]. Here, we use a modification which was introduced by I. Koponen [Kop95] and provides an exponential cut off. Its characteristic function—the characteristic function is the Fourier transform of the original function—is given by

$$\Lambda_{\alpha,c_1,l}(f_n) = \exp\left(c_0 - c_1 \frac{(f_n^2 + 1/l^2)^{\alpha/2}}{\cos(\pi\alpha/2)} \cos\left(\alpha \arctan(l|f_n|)\right)\right) \tag{3.16}$$

with the scaling factors

$$c_0 = \frac{l^{-\alpha}}{\cos(\pi\alpha/2)} \tag{3.17}$$

and c_1. Only the characteristic function of the TLF is available in analytical form. Thus, it is necessary to perform a discrete Fourier transform of the price change distributions in order to fit the parameters of $\Lambda_{\alpha,c_1,l}(f_n)$.

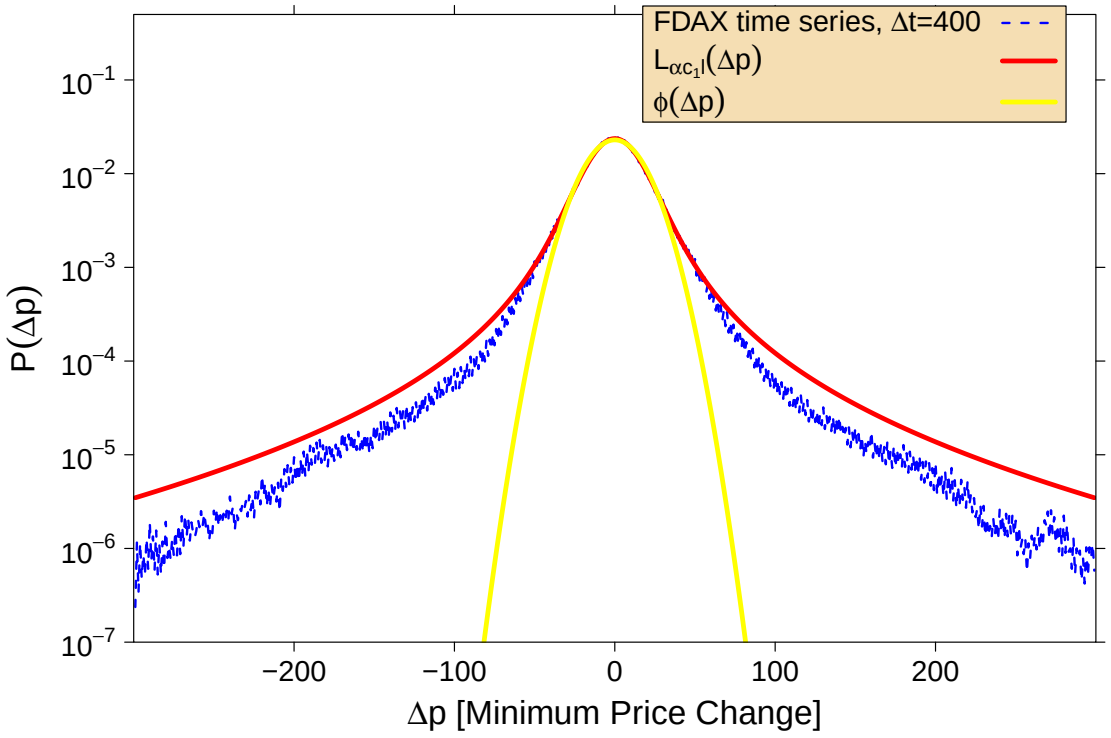

Figure 3.14: FDAX price change distribution with Gaussian fit and TLF distribution fit. The parameters of the Gaussian fit $\phi(\Delta p) = a \cdot \exp\left(-b \cdot \Delta p^2\right)$ are $a = 0.0229 \pm 0.0001$ and $b = 0.00185 \pm 0.00002$. The parameters of the TLF distributions $L_{\alpha,c_1,l}(\Delta p)$ can be found in Table 3.2. The Gaussian distribution underestimated dramatically the probability for large price changes.

A physical process can be described in the time domain and in the frequency domain. An observable therefore has a representation $h(t)$ (where t stands for time) and a representation $H(f)$ with f representing a (usually complex) frequency with $-\infty < f < \infty$. Switching between both representations of the same process is possible through the Fourier transform equations

$$H(f) = \int_{-\infty}^{\infty} h(t)\, e^{2\pi i f t} dt \tag{3.18}$$

and

$$h(t) = \int_{-\infty}^{\infty} H(f)\, e^{-2\pi i f t} df. \tag{3.19}$$

which is well-documented in the literature, e.g., in [PFTV92]. The price change distributions are approximately symmetric, which affects in turn an important property of the

Fourier transform. The imaginary part of coefficients $H(f)$ disappears for even functions $h(t)$. Here we will apply the Fourier transform to a sequence of data points which requires the discrete Fourier transform to be used. There are N equally spaced data points with a constant distance Δ. Furthermore, these return data points are localized symmetrically around 0. Thus, the original coefficients are given by $h_k \equiv h(t_k)$ with $k = 0, 1, 2, \ldots, N-1$. Note, that one has to use $t_k \equiv -N/2 + k \cdot \Delta$ due to the symmetry of the original function. In the frequency domain, the frequencies take on discrete values $f_n \equiv n/(N\Delta)$ between the positive and negative Nyquist frequency f_c ($f_c = -\frac{1}{2}$ and $f_c = \frac{1}{2}$) with $n = -N/2, \ldots, N/2$. Furthermore, Eq. (3.18) has to be approximated by a discrete sum:

$$\begin{aligned} H(f_n) = \int_{-\infty}^{\infty} h(t)\, e^{2\pi i f_n t} dt &\approx \sum_{k=0}^{N-1} h_k \exp(2\pi i f_n t_k)\, \Delta \\ &= \Delta \sum_{k=0}^{N-1} h_k \exp\left(2\pi i \frac{n}{N\Delta}\left(-\frac{N}{2} + k \cdot \Delta\right)\right) \\ &= \Delta \sum_{k=0}^{N-1} h_k \exp\left(-\frac{\pi i n}{\Delta}\right) \cdot \exp\left(\frac{2\pi i n k}{N}\right) \\ &\overset{\Delta=1}{=} \sum_{k=0}^{N-1} h_k \cos(-\pi n) \cdot \exp\left(\frac{2\pi i n k}{N}\right) \end{aligned} \tag{3.20}$$

Table 3.2: Fit parameters of the TLF distribution: The fit parameters α, c_1, and l are shown for various values of Δt. Additionally, errors of the least mean square fits are provided.

Δt	α	c_1	l
200	$1,508 \pm 0,002$	441 ± 4	1195 ± 16
400	$1,492 \pm 0,002$	695 ± 5	1818 ± 22
800	$1,478 \pm 0,002$	1097 ± 10	2772 ± 37
1600	$1,426 \pm 0,002$	1465 ± 15	3948 ± 45
3200	$1,395 \pm 0,003$	2145 ± 27	5498 ± 58
6400	$1.399 \pm 0,003$	3736 ± 59	7815 ± 92

Now we can apply Eq. (3.20) in order to transform the price change distributions of the FDAX time series. The resulting transformed price change distribution can be found in Fig. 3.12 for various time lags Δt. In Fig. 3.13, the numerical fit of $\Lambda_{\alpha,c_1,l}(q)$ is shown for $\Delta t = 200$. The fit parameters are shown in Table 3.2 together with the results for the other values of Δt. In agreement with previous results for stock market data [MS94, MS95], Lévy exponents α are obtained in the range from 1.4 to 1.5. In order to

visualize the results, one has to perform an inverse Fourier transformation of a numerically generated data set of the fitted TLF distribution. The result is shown in Fig.3.14 for the price change distribution with $\Delta t = 400$. A Gaussian fit $\phi(\Delta p)$ and the TLF distribution fit $L_{\alpha,c_1,l}(\Delta p)$—the Fourier transform of $\Lambda_{\alpha,c_1,l}(q)$—are shown as well. Thus, financial price change distributions are not normally distributed as assumed in many economic approaches. The TLF distribution is a better approximation for the empirical data even if it overestimates the wider nature of the tails of the distribution. The extent of this overestimation is visually increased by the semi-logarithmic representation of the data.

3.7 Chapter Summary

In this chapter, we applied the compute unified device architecture—a programming approach for issuing and managing computations on a graphics processing unit (GPU) as a data-parallel computing device—to methods of fluctuation analysis. Firstly, the calculation of the Hurst exponent on a GPU was presented and the differences from a random walk were documented. These results of the scaling behavior of a stochastic process can be obtained up to 80 times faster on a GPU than on a modern central processing unit (CPU) core and the relative absolute error of the results obtained from CPU and GPU is smaller than 10^{-3}. The calculation of the equilibrium autocorrelation function was also migrated to a GPU device successfully and applied to a financial market time series. In this case, acceleration factors of up to roughly 84 were realized.

Furthermore, we could verify that the current GPU generation[6] is roughly two times faster than the previous one. The presented methods were applied to a FGBL time series of the European Exchange (Eurex), which exhibits an anti-persistent structure on short time scales. Evidence was found that a super-diffusive regime is reached on medium time scales. On long time scales, the characteristics of the FGBL time series closely correspond to those of a random walk. Furthermore, the negative correlation for a time lag of one—an empirical stylized fact of financial market time series—was verified. Finally, it could be verified that financial market price change distributions differ from random walk statistics.

As already mentioned in the GPU introduction part, the main advantage of general-purpose computations on graphics processing units is that one does not need special-purpose computers. Although GPU computing opens a large variety of possibilities, the recent development of using graphic cards for scientific computing will perhaps also revive special-purpose computing as GPU implementations are not appropriate for all problems and contexts.

[6]Current GPU generation refers to the GTX 200 series. In the meantime, the Fermi generation was released by NVIDIA in April 2010.

Chapter 4

Pattern Formation in Financial Markets

As seen already, often the assumption is made that the price dynamics of financial markets obey random walk statistics. However, real financial time series show deviations from this assumption [GPA+99, MS00, BP09, PB00, CB00], such as fat-tailed price change distributions [KSY06, Man63]. Scaling behavior, short-time negatively correlated price changes and volatility clustering [BP01, KHH02] are also well known and can be reproduced, e.g., by a statistical model of a continuous double auction [DFG+03, SFGK03] or by various agent-based models [Mas00, MM01, SP98, SS99, LM99, PGPS06, PGPS07, BPS97, CS03].

Both price formation processes and cross correlations [LCBP99, PGR+99] between different stocks and indices have been studied with the intention to optimizing asset allocation and portfolios. The analysis of market reactions to specific price movements can be regarded as a first step in investigating their conditional probability distribution functions. Such studies have already been undertaken in the field of stock markets, which display a reversion tendency after large price movements [ZKA04, ZAK06]. The rise of the hedge fund industry in recent years and their interest in taking advantage of short time correlations has also boosted interest in the analysis of market microstructure, the study of the process of exchanging assets under explicit trading rules [O'H97], which is studied and modeled intensely by the financial community [AFS09, BHS95, BP05, BL98, BB04] in order to minimize order execution costs. Market movements and complex correlations on short time scales are highly relevant to the financial industry. Each significant, exploitable correlation can be a focus of leveraged trading strategies.

In this chapter[1], we study autocorrelations of financial market time series in the anti-persistent short-time regime. For this purpose, we analyze the randomness of financial markets employing specific conditional probability distribution functions, which reflect the primary market response to given price impacts. It is commonly accepted that the anti-persistence on short time scales is due to the bid ask bounce. In order to account for this effect, we introduce a simple stochastic model in which the price is the sum of a ran-

[1] Results of this chapter are published in [Pre11, PVPS09a].

dom walk part and a second part specifically describing the bid ask bounce. We show that beyond the correlations which are due to the bid ask bounce there are correlations in the fluctuation patterns, which we call "complex correlations". In order to identify such complex correlations, we introduce a new method for quantifying pattern-based correlations of a time series on short time scales. Finally, in section 4.4, a GPU accelerated version of the fluctuation pattern formation determination is designed which leads to a significant reduction in computing time.

4.1 Reproduction of the Scaling Behavior on Short Time Scales

Scientific market modeling can only be based on price time series, which are the outcome of the trading decisions of market participants comprising the "many particle system" of a financial market. The following analysis is based on historic price time series of the German DAX future contract (FDAX) traded at the European Exchange (EUREX), which is one of the world's largest derivatives exchanges. The time series, which is displayed in the inset of Fig. 7.5, contains 2,709,952 trades recorded from 2 January 2007 to 16 March 2007.

A futures contract is a contract to buy or sell a proposed underlying asset—in this case the German DAX index—at a specific date in the future at a specified price. The analysis of futures time series has the advantage that the prices are created by trading decisions alone. Stock index data, on the contrary, are derived from a weighted summation of stock prices. With a large liquidity and inter-trade waiting times as short as 10^{-2} seconds, an impressive data base is available which contains the transaction prices, the volumes, and the appropriate time stamps. Let $p(t)$ be the transaction price at time t, which is a discrete variable $t = 1, 2, \ldots, T$. As shown in Fig. 4.1, the time-lag–dependent Hurst exponent $H(\Delta t)$ indicates an anti-persistent behavior of financial data sets on short time scales. As in chapter 3, the Hurst exponent is calculated by a local derivative of the mean-square displacement, i.e., the relationship

$$\langle |p\left(t+\Delta t\right) - p\left(t\right)|^{q} \rangle^{1/q} \propto \Delta t^{H_q(\Delta t)} \tag{4.1}$$

is used with $q = 2$. The anti-persistent behavior on short time scales is is a trivial consequence of the negative autocorrelation of price time series at time lag 1, caused by the non-zero bid-ask spread—the gap between the best available offer and the best available demand in an order book, which stores offers and demands during the trading process [PGPS07]. These jumps around the spread can be added synthetically to a random walk. Let $p^*_\phi(t)$ be the time series of the synthetic negatively correlated random walk created in a Monte-Carlo simulation through

$$p^*_\phi(t) = a_\phi(t) + b(t). \tag{4.2}$$

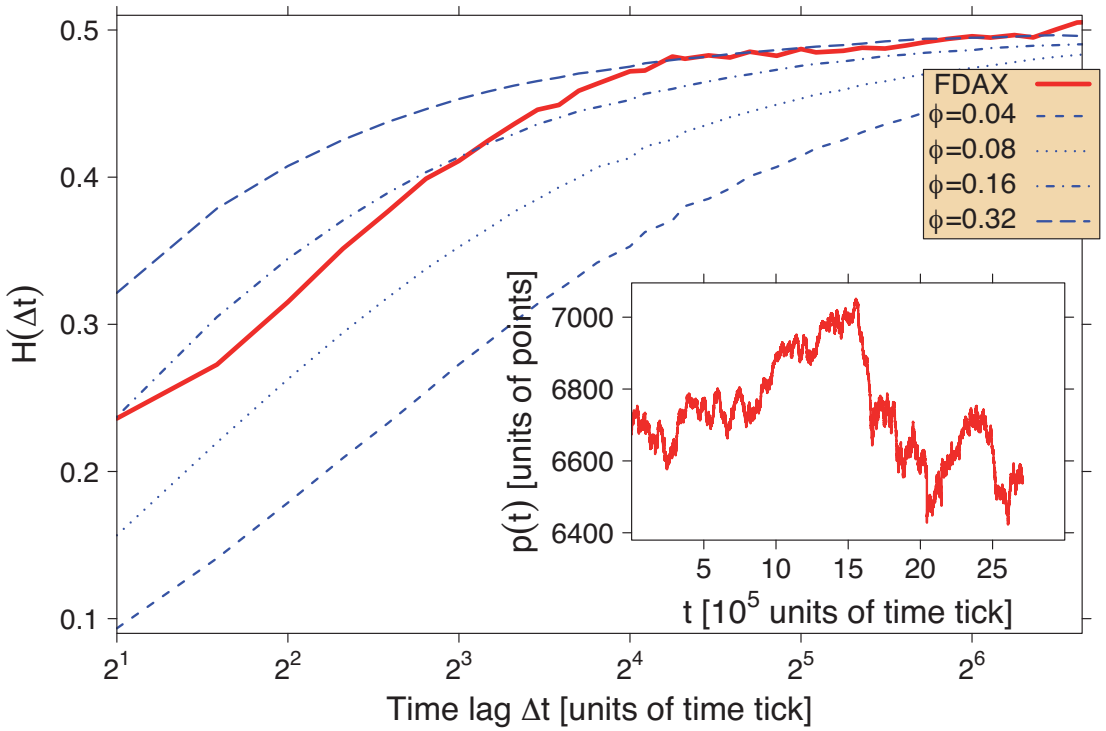

Figure 4.1: Hurst exponent $H(\Delta t)$ [Hur51, DW00, Aus00, CCS04] in dependence of time lag Δt calculated by the relationship $\langle |p(t+\Delta t) - p(t)|^q \rangle^{1/q} \propto \Delta t^{H_q(\Delta t)}$ with $q = 2$ for the FDAX time series. Also shown is the Hurst exponent for a synthetic anti-correlated random walk $p^*_\phi(t)$ (ACRW) for various values of the random walk control parameter ϕ. The optimal value $\phi = 0.16$ is found by fitting the $\Delta t = 1$ anti-correlation of the price time series to the one of the ACRW. Then also the time lag dependence of the Hurst exponent is reasonably approximated. The FDAX time series is shown in the inset.

With probability $\phi \in [0; 1/2]$ the expression

$$a_\phi(t+1) - a_\phi(t) = +1 \tag{4.3}$$

is applied and with probability ϕ a decrement

$$a_\phi(t+1) - a_\phi(t) = -1 \tag{4.4}$$

occurs. With probability $1 - 2\phi$ the expression

$$a_\phi(t+1) = a_\phi(t) \tag{4.5}$$

is used. The stochastic variable $b(t)$ models the bid-ask-spread and can take the value 0 or 1 in each time step, each with probability 1/2. Thus, by changing ϕ the characteristic time scale of the process a_ϕ can be modified as compared to process b. As shown

in Fig. 4.1, the strength of the anti-persistence is controllable. For $\phi = 0.16$, the anti-persistence of the FDAX data is reproduced also showing reasonable agreement with the observed time dependence of the Hurst exponent.

4.2 Probability Distribution Function

A fat-tailed overall probability distribution function (PDF) of price changes is shown as an example for the time interval $\Delta t^- = 10$ in the upper part of Fig. 4.2, for the time interval $\Delta t^- = 45$ in the upper part of Fig. 4.3, and for the time interval $\Delta t^- = 80$ in the upper part of Fig. 4.4. In order to examine the randomness of future price movements in the time interval Δt^+ dependent on previous price changes Δp in the time interval Δt^-, one can examine the conditional probability distribution functions (CPDF).

In the lower part of Fig. 4.2, this conditional expectation value

$$\langle p(t + \Delta t^+) - p(t) | p(t) - p(t - \Delta t^-) \rangle_t \tag{4.6}$$

is presented as a function of the time interval Δt^+ and the price jumps $\Delta p(\Delta t^-)$ for $\Delta t^- = 10$. A tendency to counterbalance jumps can be clearly identified. On average, a price reduction of for example 10 price ticks is counteracted by about 5 price ticks within 10 transactions. These results can only be reproduced qualitatively by the trivial random walk model with $\phi = 0.16$ which was introduced before. Trivially, process b can counteract maximally 1 tick. Qualitatively, the counteracting tendency is the same and is due to the anti-correlation of the time series for lag 1. However, the modified random walk has no fat tails by construction, reducing the counteracting effect. Also for other values $\Delta t^- \in [1; 100]$ the most significant counter-movements are present in the non-Gaussian tails. However, as the number of these events is very small despite the huge length of the considered time series, we introduce a limiting value λ, such that we only consider events occurring with reasonable probabilities.

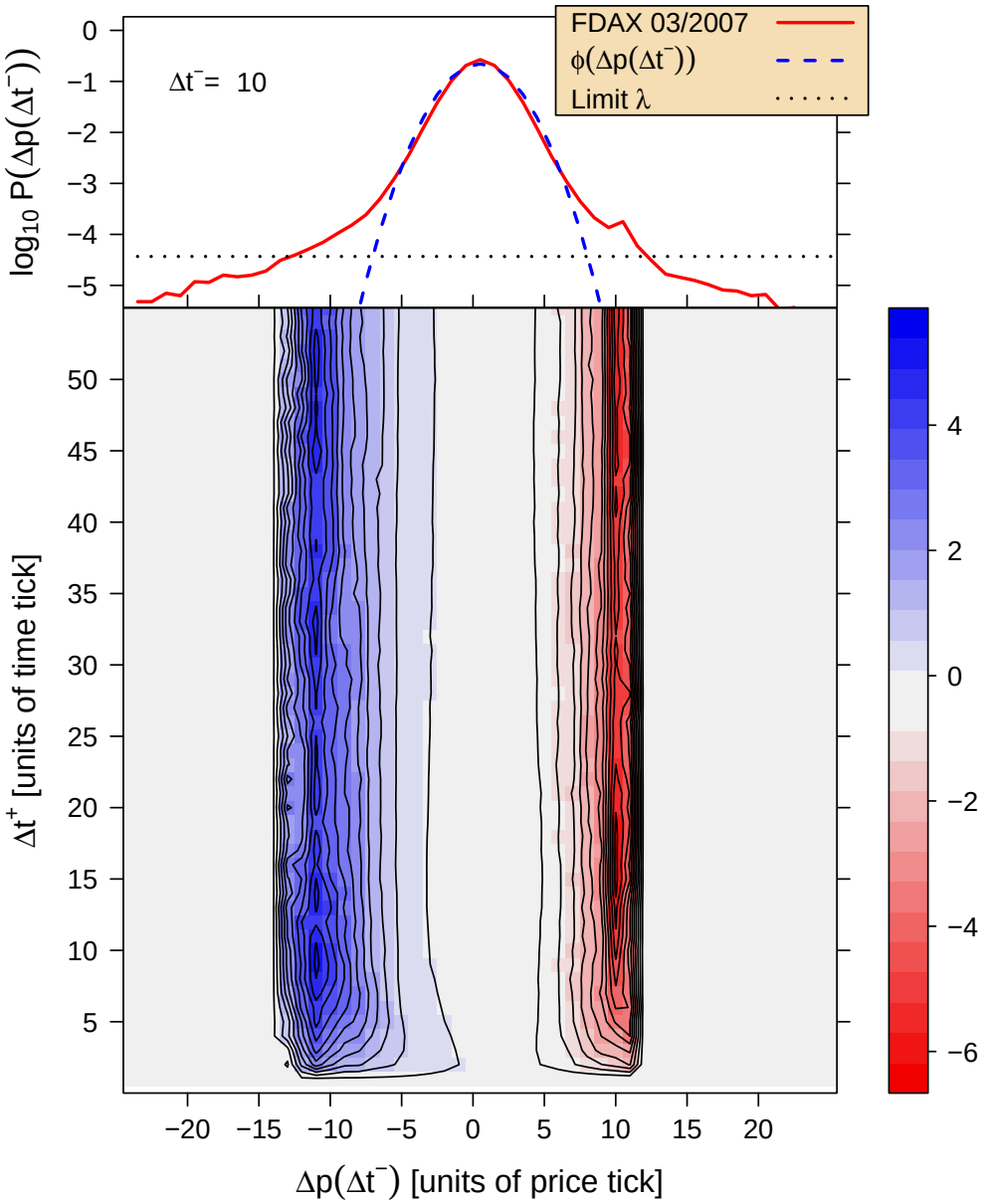

Figure 4.2: Probability distribution function (PDF) and conditional probability distribution function (CPDF) profile for FDAX time series: In the upper part, the PDF $P(\Delta p(\Delta t^-))$ is shown semi-logarithmically for $\Delta t^- = 10$. Additionally, a Gaussian least mean square fit $\phi(\Delta p(\Delta t^-)) = u \exp(-v\Delta p^2)$ is provided in order to exhibit the fat-tailed nature of the price change distributions. In the bottom part, the price changes are analyzed conditionally. The CPDF is only presented for price movements Δp, whose occurrence in the underlying data set is larger than a threshold value λ. The color code gives the conditional expectation value $\langle p(t+\Delta t^+) - p(t) | \Delta p(\Delta t^-) \rangle_t$ in dependence of Δt^+ and $\Delta p(\Delta t^-)$.

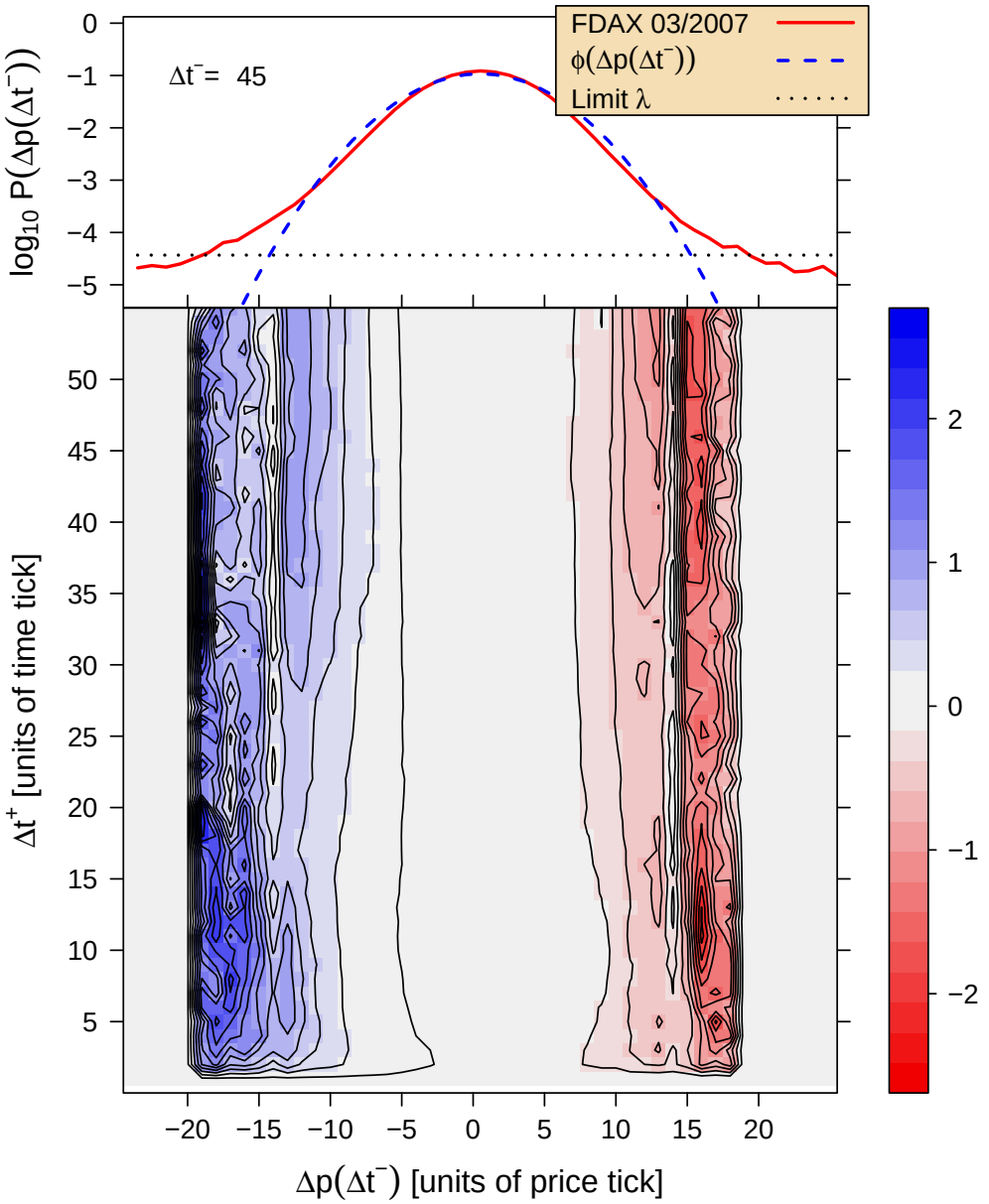

Figure 4.3: Probability distribution function (PDF) and conditional probability distribution function (CPDF) profile for FDAX time series for $\Delta t^- = 45$. In contrast to Fig. 4.2 with $\Delta t^- = 10$, the PDF is broader and also the corresponding CPDF in the lower part shows more details. A smaller counteracting tendency can be observed for this time interval.

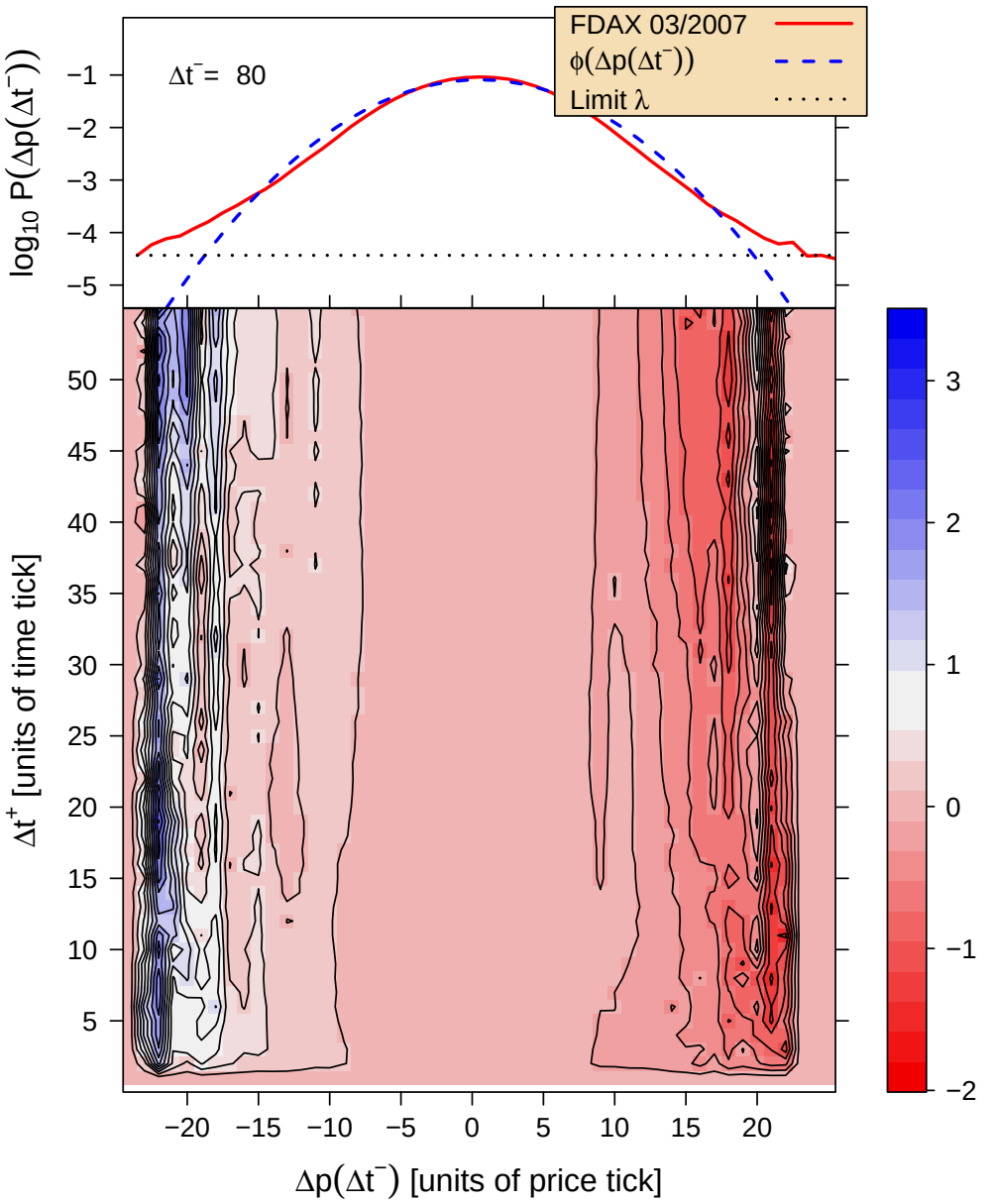

Figure 4.4: Probability distribution function (PDF) and conditional probability distribution function (CPDF) profile for FDAX time series for $\Delta t^- = 80$. In contrast to Fig. 4.3 with $\Delta t^- = 45$, the PDF is again broader and also the corresponding CPDF in the lower part shows more details.

4.3 Pattern Conformity

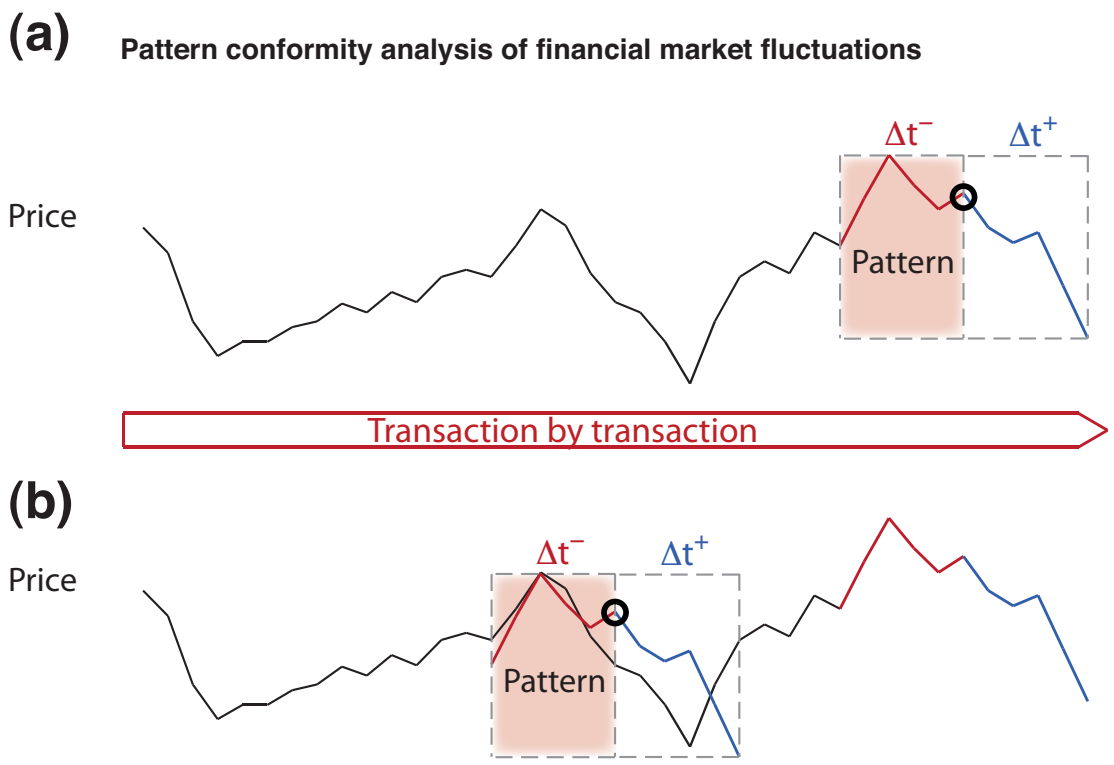

Figure 4.5: The aim is to compare a current pattern of a specific time interval length Δt^- (a) with all possible previous patterns of the time series (b).

This investigation supports the assumption that the CPDF profile of financial market data on short time scales is influenced, but not completely determined, by the negative correlation at time lag one. If complex correlations exist on these time scales, one has to find a sophisticated observable to quantify them. The existence of such correlations implies that market participants—human traders and most notably automated trading algorithms—react to a given time series pattern comparable to patterns in the past (see Fig. 4.5). On medium and long time scales, this is the basic assumption of the controversially discussed technical analysis. However, on tick by tick basis, the effect of algorithmic trading is larger. To quantify the additional correlations, we will define a general pattern conformity observable, which is not limited to the application to financial market time series.

The aim is to compare the current reference pattern of time interval length Δt^- with all previous patterns in the time series $p(t)$. The current observation time shall be denoted by $\hat{t}$, and the reference interval is then given by $[\hat{t} - \Delta t^-; \hat{t})$. The forward evolution after this current reference interval—the distance to $\hat{t}$ is expressed by Δt^+—is compared

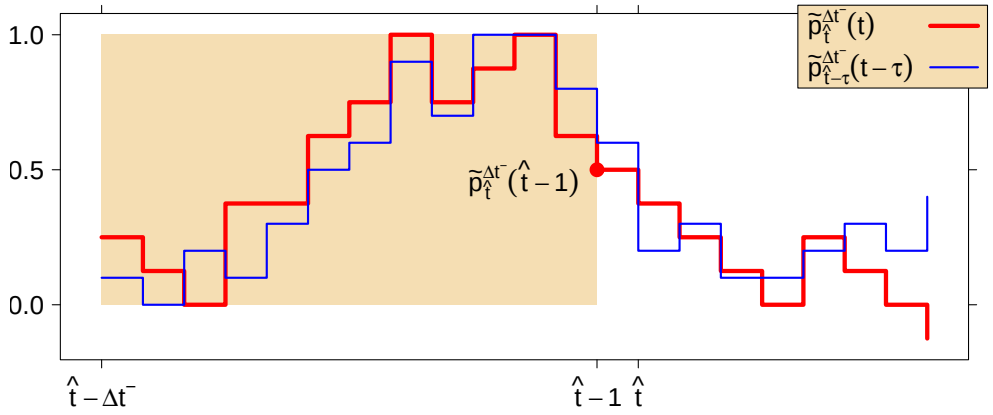

Figure 4.6: Schematic visualization of the pattern conformity calculation mechanism. The normalized reference pattern $\tilde{p}^{\Delta t^-}_{\hat{t}}(t)$ and the by τ shifted comparison pattern $\tilde{p}^{\Delta t^-}_{\hat{t}-\tau}(t-\tau)$ have the maximum value 1 and the minimum value 0 in $[\hat{t}-\Delta t^-;\hat{t})$, as illustrated by the filled rectangle. For the pattern conformity calculation, it will be checked for each time interval Δt^+ starting at $\hat{t}$ whether reference and comparison pattern are above or below the last value of the reference pattern $\tilde{p}^{\Delta t^-}_{\hat{t}}(\hat{t}-1)$. If both are above or below this level, then $+1$ is added to the non-normalized pattern conformity. If one is above and the other below, then -1 is added.

with the prediction derived from historical patterns. As the volatility is not constant, all comparison patterns have to be normalized with respect to the current reference pattern. For this reason, we use the true range—the difference between high and low. Let $p_h(\hat{t},\Delta t^-)$ be the maximum value of a pattern of length Δt^- at time $\hat{t}$ and analogously $p_l(\hat{t},\Delta t^-)$ be the minimum value. We construct a modified time series, which is true range adapted in the appropriate time interval, through

$$\tilde{p}^{\Delta t^-}_{\hat{t}}(t) = \frac{p(t)-p_l(\hat{t},\Delta t^-)}{p_h(\hat{t},\Delta t^-)-p_l(\hat{t},\Delta t^-)} \tag{4.7}$$

with $\tilde{p}^{\Delta t^-}_{\hat{t}}(t) \in [0;1] \;\forall\; t \in [\hat{t}-\Delta t^-;\hat{t})$, as illustrated in Fig. 4.6. At this point, the fit quality $Q^{\Delta t^-}_{\hat{t}}(\tau)$ between the current reference sequence $\tilde{p}^{\Delta t^-}_{\hat{t}}(t)$ and a comparison sequence $\tilde{p}^{\Delta t^-}_{\hat{t}-\tau}(t-\tau)$ for $t \in [\hat{t}-\Delta t^-;\hat{t})$ has to be determined by a least mean square fit through

$$Q^{\Delta t^-}_{\hat{t}}(\tau) = \sum_{\theta=1}^{\Delta t^-} \frac{\left(\tilde{p}^{\Delta t^-}_{\hat{t}}(\hat{t}-\theta)-\tilde{p}^{\Delta t^-}_{\hat{t}-\tau}(\hat{t}-\tau-\theta)\right)^2}{\Delta t^-} \tag{4.8}$$

with $Q^{\Delta t^-}_{\hat{t}}(\tau) \in [0,1]$ as a result of the true range adaption. With these elements, one can define an observable for the pattern conformity (PC), which is not yet normalized

by

$$\xi_\chi\left(\Delta t^+, \Delta t^-\right) = \sum_{\hat{t}=\Delta t^-}^{T-\Delta t^+} \sum_{\tau=\tau^*}^{\hat{t}} \frac{\operatorname{sgn}\left(\omega_{\hat{t}}^{\Delta t^-}(\tau, \Delta t^+)\right)}{\exp\left(\chi Q_{\hat{t}}^{\Delta t^-}(\tau)\right)}, \tag{4.9}$$

as motivated in Fig. 4.6, with

$$\tau^* = \begin{cases} \hat{t} - \hat{\tau} & \text{if } \hat{t} - \hat{\tau} - \Delta t^- \geq 0 \\ \Delta t^- & \text{otherwise} \end{cases} \tag{4.10}$$

Thus, we limit the evaluation for each pattern to the maximum of $\hat{\tau}$ historical patterns in order to save computing time. Furthermore, we use the standard definition:

$$\operatorname{sgn}(x) = \begin{cases} 1 & \text{for } x > 0 \\ 0 & \text{for } x = 0 \\ -1 & \text{for } x < 0 \end{cases} \tag{4.11}$$

The parameter χ weights terms according to their qualities. The larger χ, the stricter the pattern weighting in order to use only sequences with good agreement to the reference pattern. The expression $\omega_{\hat{t}}^{\Delta t^-}(\tau, \Delta t^+)$ in Eq. (4.9), which takes into account the value of reference and comparison pattern after $\hat{t}$ for a proposed Δt^+ relative to $\tilde{p}_{\hat{t}}^{\Delta t^-}(\hat{t}-1)$, is given by the expression

$$\begin{aligned} \omega_{\hat{t}}^{\Delta t^-}(\tau, \Delta t^+) = &\left(\tilde{p}_{\hat{t}}^{\Delta t^-}(\hat{t} - 1 + \Delta t^+) - \tilde{p}_{\hat{t}}^{\Delta t^-}(\hat{t}-1)\right) \times \\ &\left(\tilde{p}_{\hat{t}-\tau}^{\Delta t^-}(\hat{t} - \tau - 1 + \Delta t^+) - \tilde{p}_{\hat{t}}^{\Delta t^-}(\hat{t}-1)\right). \end{aligned} \tag{4.12}$$

We normalize the observable for pattern conformity by

$$\Xi_\chi\left(\Delta t^+, \Delta t^-\right) = \frac{\xi_\chi\left(\Delta t^+, \Delta t^-\right)}{\displaystyle\sum_{\hat{t}=\Delta t^-}^{T-\Delta t^+} \sum_{\tau=\tau^*}^{\hat{t}} \frac{\left|\operatorname{sgn}\left(\omega_{\hat{t}}^{\Delta t^-}(\tau, \Delta t^+)\right)\right|}{\exp\left(\chi Q_{\hat{t}}^{\Delta t^-}(\tau)\right)}}. \tag{4.13}$$

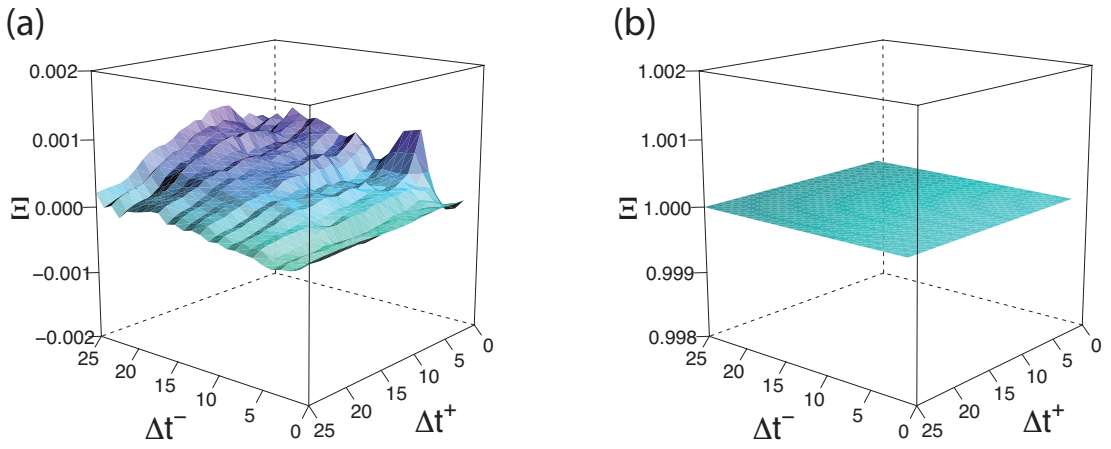

Figure 4.7: (a) Pattern conformity $\Xi_{\chi=10}(\Delta t^-, \Delta t^+)$ for a random walk time series with 3×10^6 time steps and $\hat{\tau} = 10^4$. It is close to 0 for all combinations of Δt^+ and Δt^-. (b) Pattern conformity with the same parameter settings for a time series of a straight line, which is exactly 1 for all parameter combinations. This reflects the perfect correlation of the underlying process.

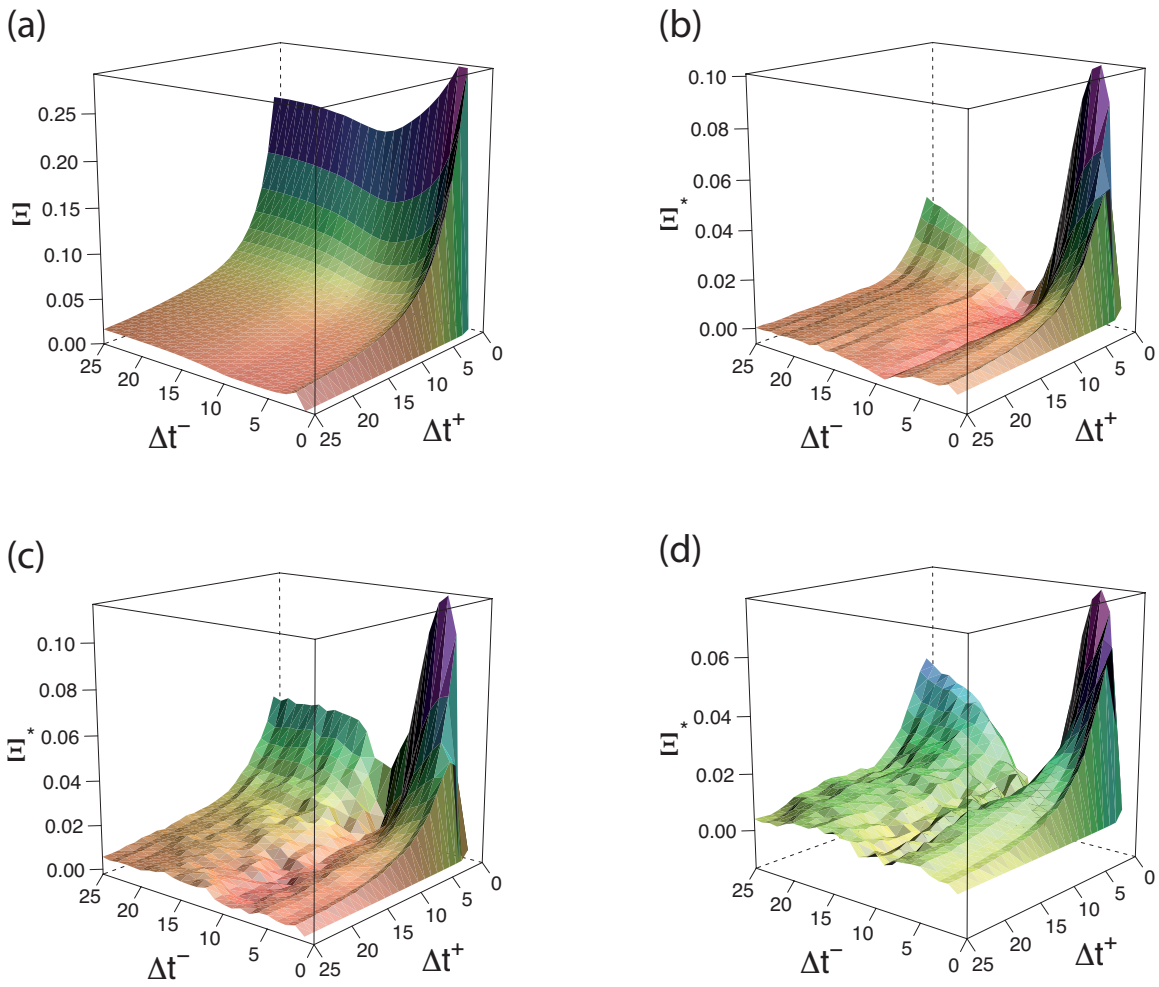

Figure 4.8: (a) Pattern conformity $\Xi^{\mathrm{FDAX}}_{\chi=100}(\Delta t^-, \Delta t^+)$ of FDAX time series with $\hat{\tau} = 10^4$ (b) FDAX pattern conformity corrected by the ACRW with $\phi = 0.16$ and with 3×10^6 time steps. Thus, $\Xi^* = \Xi^{\mathrm{FDAX}}_{\chi=100} - \Xi^{\mathrm{ACRW}}_{\chi=100}$ is shown. (c) Identical to (b), but the fit quality of a pattern is not only calculated based on prices. Appropriate transaction volumes are incorporated, too (see text). (d) Same as (b), but inter-trade waiting times are used in combination with prices to calculate the fit quality.

In Fig. 4.7, the pattern conformity for a standard random walk time series is shown (which exhibits no correlations by construction). The pattern conformity for a perfectly correlated time series—a straight line—is shown as well. With this method, it is possible to search for complex correlations in financial market data quantified through pattern conformity. In Fig. 4.8a, $\Xi_\chi(\Delta t^-, \Delta t^+)$ is shown for the FDAX time series, in which a significant pattern conformity can be detected. Parts of the correlations stem from the trivial negative autocorrelation for $\Delta t = 1$ caused by the jumps around the non-zero spread. In order to try to correct for this, in Fig. 4.8b, the pattern conformity of the ACRW with $\phi = 0.16$ is subtracted from the data of Fig. 4.8a. Obviously, the autocorrelation for $\Delta t = 1$ which is known from the order book structure is not the sole reason for the pattern conformity shown in Fig. 4.8a. Thus, impressive evidence is obtained that financial market data show pattern correlation on very short time scales beyond the simple negative correlation due to the non-zero gap between bid and ask prices in the order book.

So far, the comparison between reference and historic patterns was based on the price time series, $Q_{\hat{t}}^{\Delta t^-}(\tau) = Q_{\hat{t}}^{p,\Delta t^-}(\tau)$. Now, we also incorporate the time series of transaction volumes $v(t)$, i.e., $Q_{\hat{t}}^{\Delta t^-}(\tau) = Q_{\hat{t}}^{p,\Delta t^-}(\tau) + Q_{\hat{t}}^{v,\Delta t^-}(\tau)$, to improve the pattern selection. As a result, the pattern conformity is increased as shown in Fig. 4.8c. In contrast, using the inter-trade waiting time $\iota(t)$—$Q_{\hat{t}}^{\Delta t^-}(\tau) = Q_{\hat{t}}^{p,\Delta t^-}(\tau) + Q_{\hat{t}}^{\iota,\Delta t^-}(\tau)$—decreases the pattern conformity for small values of Δt^- as one can see in Fig. 4.8d. These results are qualitatively independent of the applied weighting method. If the exponential weighing of terms in Eq. (4.9) is replaced, for example by a cutoff rule for choosing terms, comparable results are achieved. It is only important that patterns with better agreement to historical patterns have a higher weight.

4.4 GPU Accelerated Pattern Conformity

As the determination of the fluctuation of patterns in a time series is very time consuming, relegating the calculation to a GPU is investigated in this section—analogous to the procedure in chapter 3. Again, pattern conformity is the most accurate measure for characterizing the short-term correlations of a general time series. It is essentially given by the comparison of subsequences of the time series. Subsequences of various lengths are compared with historical sequences in order to identify and extract similar reactions to similar patterns.

In order to implement a GPU implementation of the pattern conformity described in (Eq. 4.13), one has to allocate memory as in the Hurst exponent and for the autocorrelation function implementations in chapter 3. The allocation is needed for the array containing the time series, which has to be transferred to the global memory of the GPU, and for further processing arrays. The main processing GPU function is invoked with a proposed Δt^- and a given $\hat{t}$. In the kernel function, shared memory arrays for comparison and current pattern sequences are allocated and loaded from the global memory of

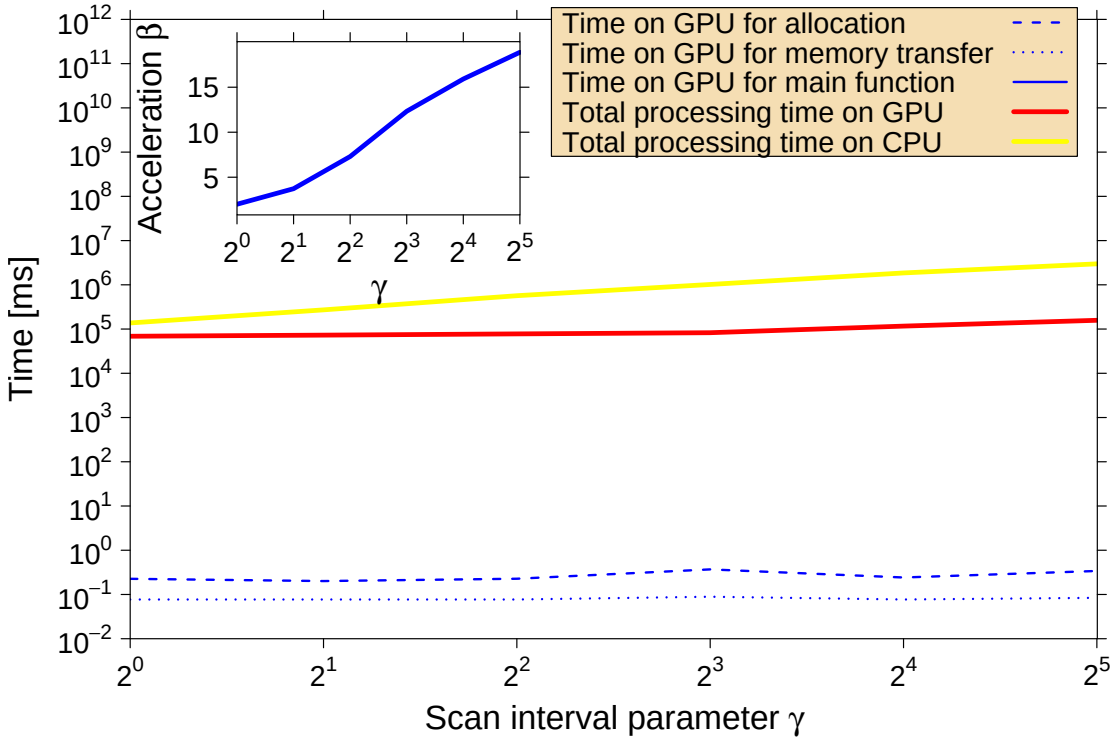

Figure 4.9: Processing times for the calculation of the pattern conformity on GPU and CPU for $\Delta t^{-}_{\max} = \Delta t^{+}_{\max} = 20$. The consumer graphics card NVIDIA GeForce GTX 280 is used as GPU device. The total processing time on GPU is split into allocation time, time for memory transfer, and time for main processing. The acceleration factor β is shown as inset. A maximum acceleration factor of roughly 19 can be obtained.

the GPU. In the main calculation part, each thread handles one specific comparison pattern, i.e., each thread is responsible for one value of τ and so, $\hat{\tau} = \gamma \times \sigma$ is applied with γ denoting the scan interval parameter and σ denoting the number of threads per block. Thus, γ corresponds to the number of blocks. The partial results of $\xi_\chi(\Delta t^+, \Delta t^-)$ are stored in a global memory based array of dimension $\hat{\tau} \times \Delta t^+$. These partial results have to be reduced in a further processing step, which uses the same binary tree structure as applied in chapter 3 for the determination of the Hurst exponent.

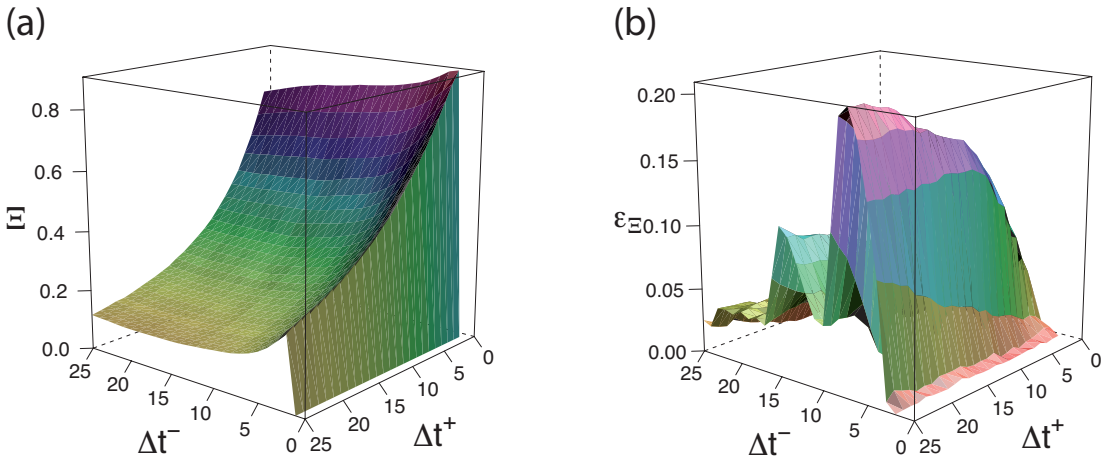

Figure 4.10: (a) Pattern conformity $\Xi^{\mathrm{GPU}}_{\chi=100}(\Delta t^-, \Delta t^+)$ of FGBL time series with $\hat{\tau} = 16384$ calculated on the consumer graphics card GTX 280. (b) Relative error ϵ_Ξ (in %) between calculation on GPU and CPU ($\Xi^{\mathrm{CPU}}_{\chi=100}(\Delta t^-, \Delta t^+)$, with the same parameter settings). The processing time on GPU was 5.8 hours; the results on CPU were obtained after 137.2 hours, which corresponds to roughly 5.7 days. Thus, for these parameters an acceleration factor of roughly 24 is obtained.

The pattern conformity for a random walk time series, which exhibits no correlation by construction, is 0. The pattern conformity for a perfectly correlated time series is 1 [PPS08]. A maximum speed-up factor of roughly 10 can be obtained for the calculation of the pattern conformity on the GPU and CPU for $\Delta t^-_{\mathrm{max}} = \Delta t^+_{\mathrm{max}} = 20$, $T = 25000$, $\chi = 100$, and $\sigma = 256$ using the 8800 GT. In figure 4.9, corresponding results for using the GTX 280 are shown as a function of the scan interval parameter γ. Here, a maximum acceleration factor of roughly 19 was realized.

With this method, which is able to detect complex correlations in a time series, it is also possible to search for pattern conformity based complex correlations in financial market data, as shown in Fig. 4.10 for the FGBL time series. In Fig. 4.10a, the results for the pattern conformity $\Xi^{\mathrm{GPU}}_{\chi=100}(\Delta t^-, \Delta t^+)$ are presented with $\hat{\tau} = 16384$ calculated on the GTX 280. For small values of Δt^- and Δt^+, large values of $\Xi^{\mathrm{GPU}}_{\chi=100}$ are obtained with a maximum value of roughly 0.8. For the results shown in Fig. 4.10b, the calculation of

the pattern conformity is executed on the central processing unit, and in Fig. 4.10c, the relative absolute error

$$\epsilon_{\Xi} = 10^2 \times \left| \frac{\Xi^{\mathrm{GPU}}_{\chi=100} - \Xi^{\mathrm{CPU}}_{\chi=100}}{\Xi^{\mathrm{CPU}}_{\chi=100}} \right| \tag{4.14}$$

is shown, which is smaller than two-tenths of a percent. This small error arises because the GPU device summarizes only a large number of the weighted values $+1$ and -1. Thus, the limitation to single precision has no significant negative effect on the result.

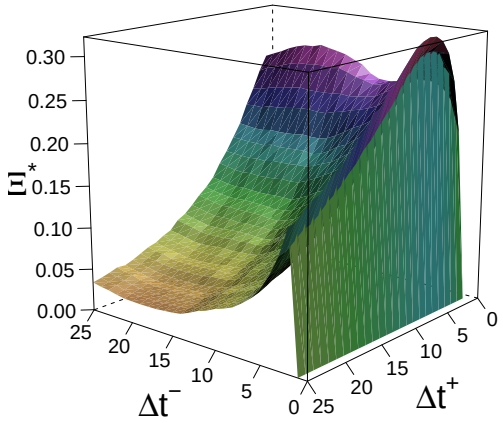

Figure 4.11: (a) FGBL pattern conformity corrected by the anti-correlated random walk (ACRW) with $\phi = 0.044$ and with 1.5×10^6 time steps. Thus, $\Xi^* = \Xi^{\mathrm{FGBL}}_{\chi=100} - \Xi^{\mathrm{ACRW}}_{\chi=100}$ is shown (see the text).

This raw pattern conformity is dominated by trivial pattern correlation caused by the jumps of the price process between best bid and best ask prices—the best bid price is given by the highest limit order price of all buy orders in an order book and analogously the best ask price is given by the lowest limit order price of all sell orders in an order book. As shown already, there are possibilities for reducing the influence of these trivial pattern conformity parts. For example, it is possible to add such jumps around the spread synthetically to a random walk. Let again p^*_ϕ be the time series of the synthetically anti-correlated random walk created in a Monte Carlo simulation through $p^*_\phi = a_\phi(t) + b(t)$. With probability $\phi \in [0; 0.5]$, the expression $a_\phi(t+1) - a_\phi(t) = +1$ is applied and a decrement $a_\phi(t+1) - a_\phi(t) = -1$ occurs with probability ϕ. With probability $1 - 2\phi$ the expression $a_\phi(t+1) = a_\phi(t)$ is used. The stochastic variable $b(t)$ models the bid-ask spread and can take the value 0 or 1 in each time step, each with probability 0.5. Thus, by changing ϕ, the characteristic time scale of the process a_ϕ can be modified as compared to the process b.

Parts of the pattern based correlations in Fig. 4.10 are caused by this trivial negative autocorrelation for $\Delta t = 1$. In order to try to correct for this, the pattern conformity

of the ACRW with $\phi = 0.044$, which reproduces the negative correlation of the FGBL time series at time lag $\Delta t = 1$, is subtracted from the data of Fig. 4.10a and shown in Fig. 4.11. Obviously, the autocorrelation for the time lag $\Delta t = 1$ is not the sole reason for the pattern formation conformity shown in Fig. 4.10a. Thus, evidence is obtained also in this example that financial market time series show pattern correlation on very short time scales beyond the simple anti-persistence due to the gap between bid and ask prices.

4.5 Chapters's Summary

Summarizing this section, we have analyzed high-frequency asset returns in detail. First of all, market impacts were investigated systematically by using conditional probability distribution functions (CPDF). The negative autocorrelation of the return time series for consecutive ticks results in a reversion tendency after a price shift on short time scales. Thus, the CPDF behavior can be reproduced qualitatively by a synthetically negatively correlated random walk, which in turn reflects the short-time anti-persistence of the Hurst exponent. Furthermore, we have introduced a method of measuring complex correlations within a time series by pattern conformity. This pattern conformity take a value of 0 for a random walk and 1 for a perfectly correlated time series. Checking the pattern conformity of financial data sets, we find that there is a small tendency to follow historic patterns on very short time scales. An increase of observed correlations occurs if the trading volume is included in the measure of agreement between the current part of the time series and the part to which it is compared.

Subsequently, the pattern formation conformity algorithm was ported to a GPU. For this application the GPU was up to 24 times faster than the CPU, and the values provided by the GPU and CPU implementations differ only by a maximum relative error of two-tenths of a percent.

The pattern conformity which is used is the most accurate measure for characterizing the short-term correlations of a general time series. It is essentially given by the comparison of subsequences of the time series. Subsequences of various lengths are compared with historical sequences in order to extract similar reactions to similar patterns. The pattern conformity of the FGBL contract exhibits large values of up to 0.8. However, these values also include the trivial auto-correlation property occurring at time lag one, which can be removed by the pattern conformity of a synthetic negatively correlated random walk. However, significant pattern based correlations are still exhibited after correction. Thus, evidence is obtained that financial market time series show pattern correlation on very short time scales beyond that of the simple anti-persistence which is due to the gap between bid and ask prices.

Chapter 5

GPU Accelerated 2D and 3D Ising Model

In the previous chapter, a GPU approach successfully applied in order to accelerate time series analyses. We now port a standard model of statistical physics—the Ising model—to the GPU architecture in this chapter[1]. As we will see in chapter 6, the Ising model can also be interpreted in terms of financial market fluctuations. A simple Ising model based simulation of an artificial stock market is presented for example in [ZS07], in which specific properties of financial markets such as the consequences of herding behavior can be reproduced.

The Ising model, which Ernst Ising studied in his PhD [Isi25], is a standard model of statistical physics and provides a simplified microscopic description of ferromagnetism. It was introduced to explain the ferromagnetic phase transition from the paramagnetic phase at high temperatures to the ferromagnetic phase below the Curie temperature T_C. A large variety of techniques and methods in statistical physics were originally formulated in terms of the Ising model and then generalized and adapted to related models and problems [BL01]. Supported by his results for a one dimensional spin chain, in which no phase transition occurs, Ising initially proposed in his PhD thesis that there is also no phase transition in higher dimensions which turned out to be a mistake. The Ising model on a two dimensional square lattice with no magnetic field was later analytically solved by Lars Onsager in 1944 [Ons44]. The critical temperature at which a second order phase transition between an ordered and a disordered phase occurs can be determined analytically for the two dimensional model ($T_C = 2.269185$ [Ons44]). Despite much effort, an analytic solution for the three dimensional Ising model still remains one of the great challenges in statistical physics. However, computer simulations in combination with finite-size scaling techniques [Bin81, FBV06, Heu93, Fis67] are able to determine $T_C \sim 4.5115$ [BL01] and the rest of the phase diagram with good accuracy. Starting in 1944, the Ising model not only became popular in main stream physics but also in various interdisciplinary fields, in particular in econophysics [MS00, PB00, PPS08]

[1] Results of this chapter are published in [PVPS09b, BVP10].

but also in sociophysics for explaining the temporal behavior of democratic societies [SWS00, SfSMdO00, Sch04, SH05c, SH05a, SH05b] and ghetto formation [Sch69].

Critical phenomena, scaling and universality properties of Ising models have been studied via Monte Carlo simulations for decades [BL01] with continuously improving accuracy, these studies having directly benefited from the increasing availability of computational resources. Such computing requirements, necessary not only for Monte Carlo simulations but also for various other tasks in computational physics including, e.g., molecular dynamics simulations [vMAF+08, KSRS08, LSVMW08, SPF+07] or stochastic optimization [SK06], need a large amount of high performance computing resources.

We employ the general purpose graphics processing unit technology for Monte Carlo simulations of a two dimensional square lattice and a three dimensional cubic lattice Ising models[2]. In [TMB+05], a GPU based version of the Ising model has already been proposed. This implementation was able to accelerate the Ising model computation by a factor of three by migration to a GPU. In this paper, we will demonstrate that a much better acceleration factor can be obtained. We demonstrate that our implementation works by calculating the critical temperatures of the phase transitions in the two and three dimensional Ising models.

This chapter is organized as follows. In a preparation step for Monte Carlo simulations, the generation of pseudo random numbers on a GPU device is described in section 5.1. In section 5.2, we provide the implementation details of the two dimensional ferromagnetic square lattice Ising model including acceleration factors and determination of the critical temperature by finite size scaling. In section 5.3, the Ising model implementation is expanded to three dimensions. Acceleration factors and critical temperatures are also provided. In the remaining sections of this chapter, the focus lies on a multi-spin coded version of the 2D Ising model which will be extended to a multi GPU approach. This extension was developed in collaboration with Benjamin Block. Our conclusions are summarized in section 5.7.

5.1 Random Number Generation

An efficient method for creating random numbers is essential for Monte Carlo simulations of the two dimensional ferromagnetic square lattice and the three dimensional ferromagnetic cubic lattice Ising model on a GPU in section 5.2 and section 5.3. For this purpose, we use an array of linear congruential random number generators (LCRNGs) applying one of the oldest and best-known algorithms for generation of pseudo random

[2]We use a GeForce GTX 280 as the GPU device if no other graphics card is specified. We are using CUDA (version 2.0) in combination with an NVIDIA graphics card driver (driver version 177.73). Early versions of the graphics card drivers for the NVIDIA Geforce GTX 200 series exhibit a bug if a large part of provided global memory is used. However, this problem has recently been fixed. For the comparison between CPU and GPU implementations, we use an Intel Core 2 Quad CPU (Q6700) with 2.66 GHz and 4096 kB cache size, of which only one core is used. The standard C source code executed on the host is compiled with the gcc compiler with option -O3 (version 4.2.1). Other compilers in combination with more sophisticated compiler options lead to even better processing times on the CPU.

numbers [SK06]. Starting at a seed value $x_{0,j}$, a sequence of random numbers $x_{i,j}$ with $i \in \mathbb{N}$ of the LCRNG j can be obtained by the recurrence relation

$$x_{i+1,j} = \left(a \cdot x_{i,j} + c\right) \mod m \tag{5.1}$$

where a, c, and m are integer coefficients. An appropriate choice of these coefficients in Eq. (5.1) is responsible for the quality of the produced random numbers. We use $a = 1664525$ and $c = 1013904223$ as suggested in [SK06]. In order to exploit the local 32-bit architecture provided by the GPU device, the parameter of the modulo operation m is set to 2^{32} as, by construction, results on a 32-bit architecture are truncated to the endmost 32 bits. Thus, if using signed integers, pseudo random numbers $x_{i,j} \in [-2^{31}; 2^{31} - 1]$ can be obtained, which have to be normalized according to $y_{i,j} = \text{abs}\left(x_{i,j}/2^{31}\right) \sim 4.656612 \cdot 10^{-10}\text{abs}\left(x_{i,j}\right)$ in order to get uniformly distributed pseudo random numbers $y_{i,j}$ in the interval $[0; 1]$. As we use an entire set of linear congruential random number generators in parallel, each LCRNG j of this array is initialized by a random number obtained from a further LCRNG through

$$x_{0,j+1} = \left(16807 \cdot x_{0,j}\right) \mod m \tag{5.2}$$

with $x_{0,0} = 1$.

In the GPU implementation, each thread of a thread block handles its own linear congruential random number generator. Denoting s as the number of involved thread blocks and denoting σ as the number of threads per block, in a GPU kernel, values of $s \cdot \sigma$ LCRNGs are determined in parallel. Each LCRNG calculates S pseudo random numbers. Thus, a total of $S \cdot s \cdot \sigma$ pseudo random numbers are created. In Fig. 5.1, a comparison of processing times of the random number generation between calculation on a GPU device and on a CPU core is presented for $S = 10^4$ and $\sigma = 512$ depending on the number of involved blocks s. On the CPU, the same $S \cdot s \cdot \sigma$ pseudo random numbers are used for the sake of comparison. The acceleration factor β, which is shown in the inset, is determined by the relationship

$$\beta = \frac{\text{Total processing time on CPU}}{\text{Total processing time on GPU}} \tag{5.3}$$

as defined in the previous chapter. A maximum acceleration factor $\beta \sim 130$ is obtained for $s = 30$, which corresponds to the number of multiprocessors of the NVIDIA GeForce GTX 280 consumer graphics card. Up to 30 groups of $\sigma = 512$ single linear congruential random number generators can be executed concurrently at the same time on this GPU device. If s is larger than 30, the 30 available multiprocessors are able to execute only the first 30 blocks in parallel, such that the remaining groups of $\sigma = 512$ single linear congruential random number generators have to be handled in a second step. Thus, processing time on the GPU device is doubled when $s = 31$ is used instead of $s = 30$. Note that the computation on the GPU becomes inefficient if s is not a multiple of the number of multiprocessors on the gpu device because some multiprocessors are idle at

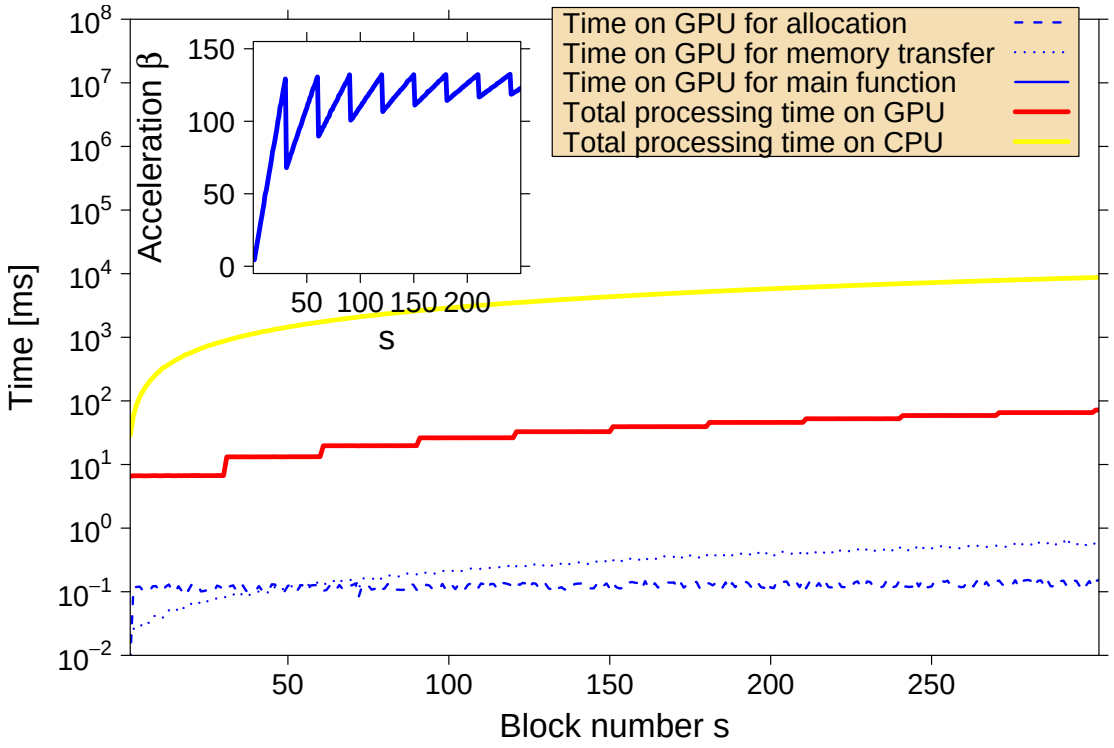

Figure 5.1: Processing times for the calculation of $S \cdot s \cdot \sigma$ pseudo random numbers on GPU and CPU for $S = 10^4$ and 512 threads per block in dependence of the number of involved thread blocks s. The total processing time on GPU is divided into allocation time, time for memory transfer, and for main processing. The acceleration factor β is shown as inset. A maximum acceleration factor of roughly 130 can be obtained for $s = 30$, which corresponds to the number of multiprocessors on our GPU device. Note that the computation on the GPU becomes inefficient if s is not a multiple of the number of multiprocessors on the gpu device because some multiprocessors are idle in the end of the calculation (see inset).

the end of the calculation (see inset of Fig. 5.1). The fraction of GPU processing time for allocation and memory transfer of the LCRNG seed values can be neglected. Please note that there are many other possibilities for implementing an efficient creation of pseudo random numbers on a GPU device. In [vMAF+08], e.g., an algorithm is presented for producing a set of pseudo random numbers in parallel on a GPU device with a single linear congruential random number generator which determines the set of numbers according to a serial rule. A CUDA based version of the Mersenne Twister random number generator [MN98] is able to generate pseudo random numbers in parallel as well.

5.2 Two Dimensional Ising Model

In this section, we present an implementation of the two dimensional ferromagnetic square lattice Ising model on a GPU. The simple Ising model, which is one of the simplest lattice models, consists of an arrangement of spins with are located on the sites of the lattice and which exhibit only the values $+1$ and -1 [LB05]. These spins interact with their nearest neighbors on the lattice with interaction constant $J > 0$. The Hamiltonian $\mathscr{H}$ for this model is given by

$$\mathscr{H} = -J \sum_{\langle i,j \rangle} S_i S_j - H \sum_i S_i \tag{5.4}$$

where $S_i = \pm 1$ represents a spin at site i and H denotes the coupling to an external magnetic field. We use a square lattice with n spins per row and column and periodic boundary conditions. The lattice contains $N = n^2$ spins. For the spin update, the Metropolis criterion [MRR+53] is applied which is an obvious choice for a transition probability satisfying the 'detailed balance' principle in the thermal equilibrium and leading to the fastest dynamics for single spin flip simulations. In the expression $P_a(t)W_{a\to b} = P_b(t)W_{b\to a}$ for 'detailed balance', $P_a(t)$ denotes the probability of the system being in state a at time t and $W_{a\to b}$ the transition rate for the move $a \to b$. Denoting $\Delta\mathscr{H}$ to be the energy difference between the two states a and b, which is given through $\Delta\mathscr{H} = \mathscr{H}_b - \mathscr{H}_a$, the probability for the move $a \to b$ is given by $W_{a\to b} = \exp(-\Delta\mathscr{H}/k_B T)$ if $\Delta\mathscr{H} > 0$ and by $W_{a\to b} = 1$ if $\Delta\mathscr{H} \leq 0$. As a single spin flip dynamics is applied, two successive states differ only by a single spin, such that $\Delta\mathscr{H}$ can be calculated as a local energy difference.

In the first step of the GPU implementation, one has to allocate memory on the GPU device's global memory for the two dimensional spin field and for the seed values of the LCRNGs. After a random initialization of the spin field on the CPU and initialization of seed values as described in section 5.1, spins and seeds are transferred to the GPU's global memory. The transition probability $W_{a\to b}$ depends on the temperature T which has to be passed to the GPU kernel function. The Boltzmann constant k_B is fixed to 1 in all simulations. We use a zero field ($H = 0$) and $J = 1$.

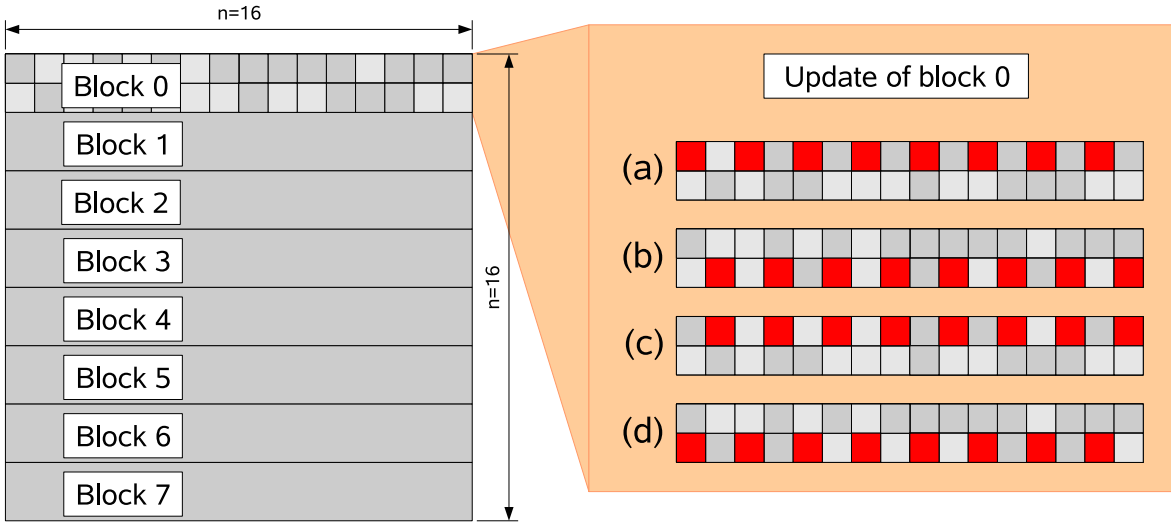

Figure 5.2: Schematic visualization of the implementation of a two dimensional Ising model on a GPU for $n = 16$ spins per row. The 2D spin field is split in strips of 16×2 spins each treated by a single block on the GPU. In each block, we use $n/2 = 8$ threads, which are used for the spin update corresponding to the scheme presented on the right side and described in the main text.

It must be taken into account that for Monte Carlo trial moves to be executed in parallel, the system has to be partitioned into non-interacting domains. Thus, for the update process of the spin lattice, a checkerboard algorithm is applied, that is there is a regular scheme for updating the lattice spin by spin. First all spins on the "white" squares of the checkerboard are updated and then all spins on the "black" squares. Please note that other methods for this updating process are also available (e.g., diverse cluster algorithms [SW87, Wol89], which exhibit faster convergence). However, the systematic scheme of the checkerboard algorithm is most suitable for demonstrating the migration to the GPU architecture realizing noninteracting domains where the Monte Carlo moves are performed in parallel. In fact, this partitioning for "real life" problems is a significant challenge.

On a GPU device, it is only possible to synchronize threads within a block. A native block synchronization does not exist. Therefore, only the termination of the GPU kernel function ensures that all blocks were executed. We divide the spin update process on the lattice into blocks on the GPU. In a single GPU kernel, only a semi-lattice can be changed without creating conflicts.

The spin updating process is subdivided into threads and blocks as illustrated in Fig. 5.2 for $n = 16$. The spin field is divided into strips of width 2. Each strip is handled by one block and each thread is responsible for a square sub-cell of 2×2 spins in order

to avoid idle threads. Thus, $n/2$ threads per block are necessary. A first GPU kernel handles the update process of the first semi-lattice, i.e., in each block $\sigma = n/2$ threads accomplish steps (a) and (b) of Fig. 5.2. The termination of this GPU kernel ensures that all blocks were executed and a second kernel can start in order to accomplish steps (c) and (d). Each sub-cell (i.e., each thread in each block) has access to its own LCRNG. As each thread in one GPU kernel function needs up to two random numbers and as this array is also used after two semi-lattice updates for the reduction process of partial energies or magnetizations of the Ising lattice, the seed values or current random numbers are transferred at the beginning of a GPU kernel to a shared memory array in order to take advantage of the increased memory speed. After finishing of the updating steps (steps (b) and (d)) the current value is transferred back to global memory. As CUDA does not provide native reduction functions for treating partial results, this must be managed manually. For this purpose, the shared memory array is used after step (d). A binary tree structure ensures a fast reduction of the partial values within a block. These partial results of each block are stored at block-dependent positions in global memory and finally transferred back to the host's main memory [PVPS09a]. The final summation of the results of $n/2$ blocks is carried out by the CPU. In Fig. 5.3, the processing times of the Ising model implementation on the GPU are compared with an Ising model implementation on one CPU core. For the largest system, $n = 1024$, an acceleration factor of roughly 60 could be achieved. For very small systems, the acceleration factor is smaller than 1 because the GPU device is not used efficiently for small thread numbers per block.

In order to verify the GPU implementation, it is insufficient to only compare values of energy and magnetization of a function of the temperature T between GPU and CPU versions. A very sensitive test is given by the determination of the critical temperature of the Ising model. For this purpose, we use finite size scaling and calculate the Binder cumulant [Bin81], which is given in a zero field by

$$U_4(T) = 1 - \frac{\langle M(T)^4 \rangle}{3\langle M(T)^2 \rangle^2}, \tag{5.5}$$

with M denoting the magnetization of a configuration at temperature T and $\langle \ldots \rangle$ being the thermal average. Near a critical point, finite size scaling theory predicts the free energy and derived quantities such as magnetization being functions of linear dimension L over correlation length $\xi \simeq (T - T_c)^{-\nu}$. Therefore, moment ratios of the magnetization like for example the Binder cumulant U_4, become independent of system size at the critical temperature. To test our implementation, we perform several simulations close to the critical point for different linear dimensions of the simulation box and determine U_4. As shown in Fig. 5.4, the curves of the Binder cumulants for various system sizes $N = n^2$ cross almost perfectly at the critical temperature derived by Onsager ($T_c \approx 2.269185$) [Ons44], indicated by a dashed line. Note that single spin flips and parallelization schemes based upon them are not particularly well-suited for the determination of critical properties because of critical slowing down. Nevertheless, we are able to determine T_c with reasonable accuracy ($T_c = 2.2692 \pm 0.0002$).

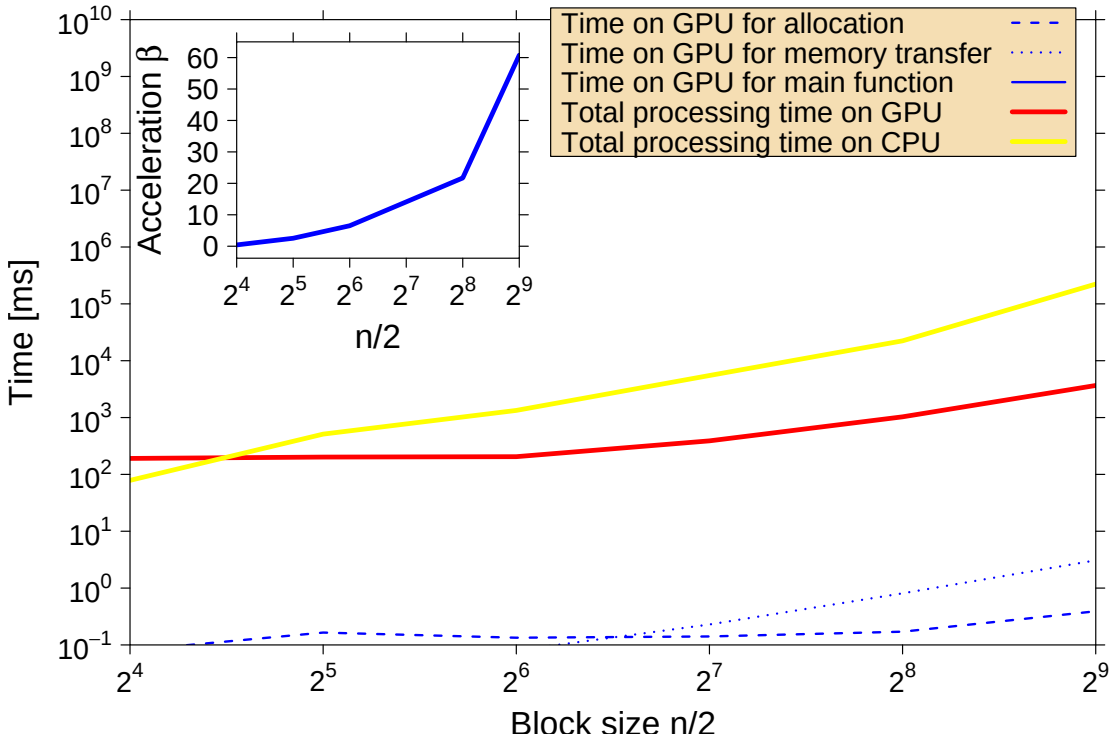

Figure 5.3: Processing times for a two dimensional ferromagnetic square lattice Ising model for a cooling down process. The temperature $T \in [2.0; 3.0]$ is stepwise reduced by the factor 0.99. In each temperature step, 100 sweeps are performed. The processing times are shown in dependence of the number of threads per block which is related to the system size by $\sigma = n/2$. The total processing time on GPU is divided into allocation time, time for memory transfer, and time for main processing. The acceleration factor β is shown as inset. A maximum acceleration factor of roughly 60 can be realized.

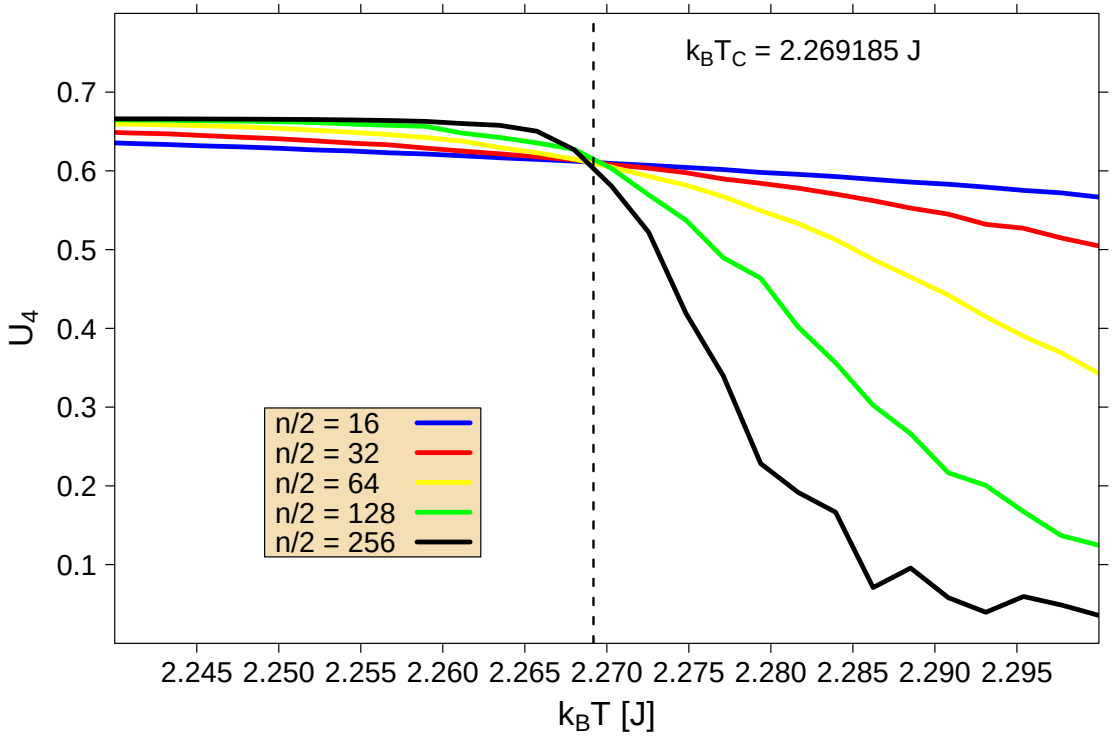

Figure 5.4: Binder cumulant U_4 in dependence of $k_B T$ for various numbers n of spins per row and column of the two dimensional square lattice Ising model. $n/2$ corresponds to the involved number of threads per block on the GPU implementation. The curves of the Binder cumulants for various system sizes $N = n^2$ cross almost perfectly at the critical temperature derived by Onsager [Ons44], which is shown additionally as a dashed line. In each temperature step, the average was taken over 10^7 measurements.

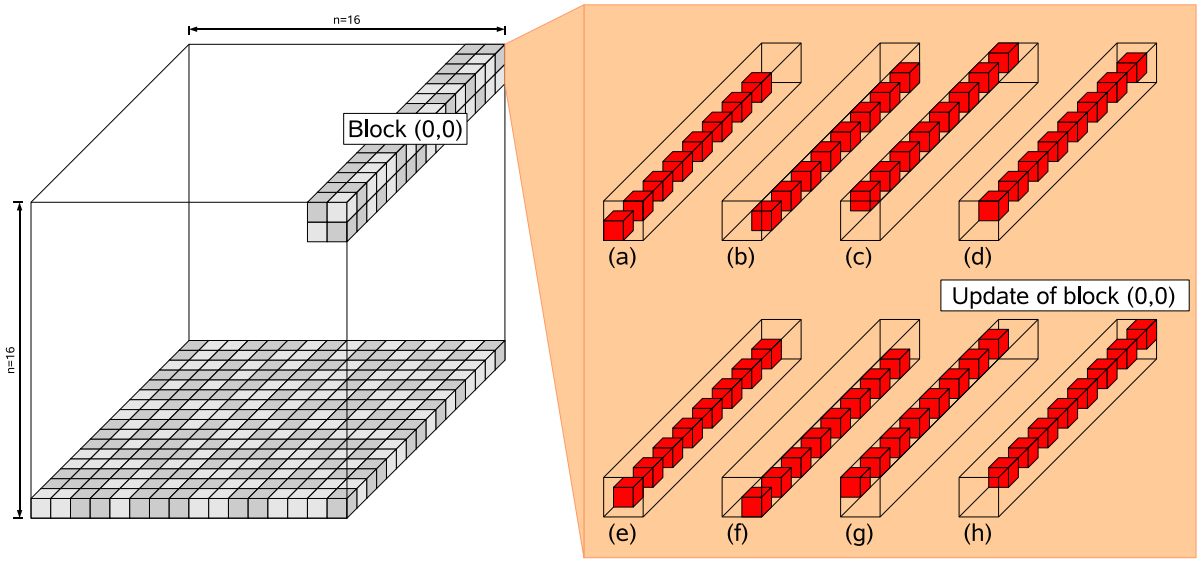

Figure 5.5: Schematic visualization of the three dimensional ferromagnetic cubic lattice Ising model implementation on a GPU for $n = 16$. The 3D spin field is split in cuboids (shown left) of $2 \times 2 \times 16$ spins each treated by a single block on the GPU. In each block, we use $n/2 = 8$ threads, which are used for the spin update corresponding to the scheme presented (shown right) and described in the main text.

5.3 Three Dimensional Ising Model

In this section, the GPU implementation of the two dimensional ferromagnetic square lattice Ising model is expanded to a three dimensional cubic lattice model version on the GPU. In a first step, we allocate memory analogously as in section 5.2. In addition, the three dimensional spin field with $N = n^3$ spins and the random number seeds have to be transferred to the GPU device. In the three dimensional case, the spin update process is also subdivided into threads and blocks. Now, the update scheme illustrated in Fig. 5.5 for $n = 16$ is applied. The 3D spin field is split in cuboids of $2 \times 2 \times 16$ spins each treated by a single block on the GPU. In each block, we use $n/2 = 8$ threads, which are used for the spin update corresponding to the scheme presented on the right side of Fig. 5.5. A first GPU kernel handles the update process of the first semi-lattice, i.e., in each block $\sigma = n/2$ threads accomplish steps (a), (b), (c), and (d) of Fig. 5.5. The termination of this GPU kernel ensures that all blocks were successfully executed and a second kernel can start in order to carry out steps (e), (f), (g), and (h). Each sub-cell (i.e., each thread in each block) which is responsible for 8 spins in the 3D case has access to its own LCRNG. In Fig. 5.6, the processing times of the three dimensional Ising model implementation on the GPU are compared with a corresponding Ising model

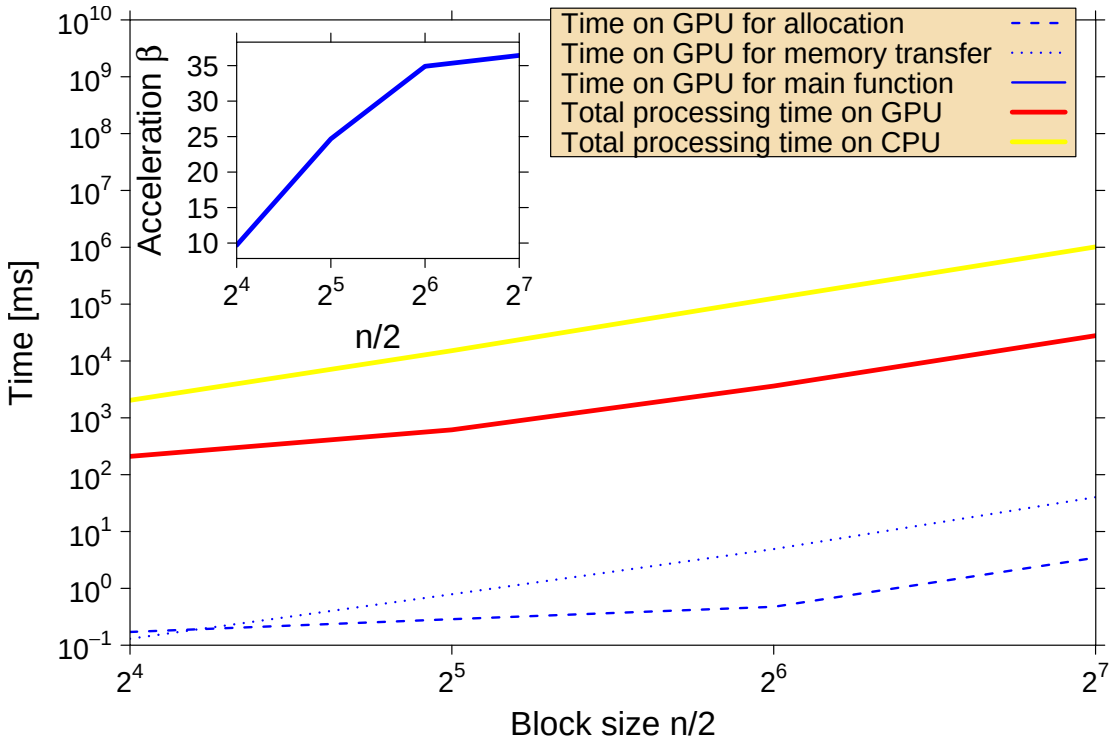

Figure 5.6: Processing times for a three dimensional ferromagnetic cubic lattice Ising model for a cooling process. The temperature $T \in [4.0; 5.0]$ is gradually reduced by the factor 0.99. In each temperature step, 100 sweeps are performed. The processing times are shown in dependence of the number of threads per block which is related to the system size by $\sigma = n/2$. The total processing time on the GPU is split into allocation time, time for memory transfer, and time for main processing. The acceleration factor β is shown as inset. A maximum acceleration factor of roughly 35 can be achieved.

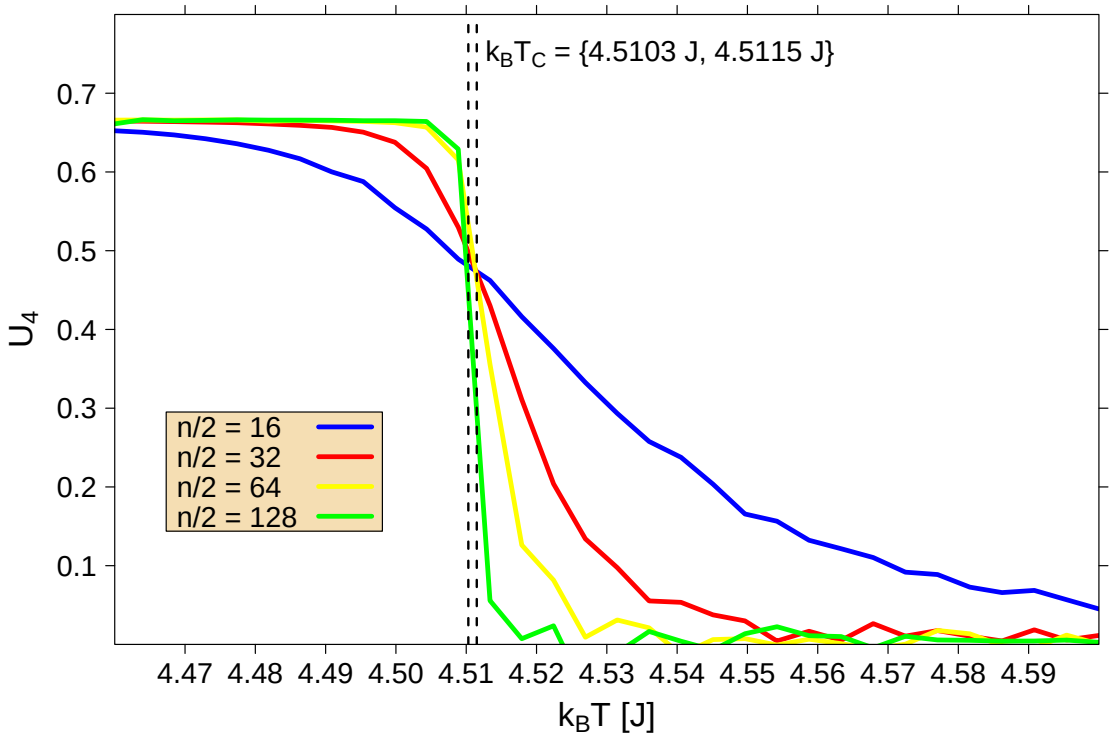

Figure 5.7: Binder cumulant U_4 in dependence of k_BT for various numbers n of spins per row and column of the three dimensional Ising model on the simple cubic lattice. $n/2$ corresponds to the involved number of threads per block on the GPU implementation. The curves of the Binder cumulant for various system sizes $N = n^2$ cross almost perfectly at the critical temperature $T_C \approx 4.51$, which is in agreement with a selection of previous simulation results which are presented as dashed lines. For each temperature step, the average was taken over 10^6 measurements.

implementation on a single CPU core. The largest system which can be realized on a GeForce GTX 280 in this way is $n = 256$. Global memory size limits the Ising system size to this value. Thus, a maximum acceleration factor of roughly 35 can be achieved.

In Fig. 5.7, the Binder cumulant for different system sizes $N = n^3$ is presented as a function of temperature T. The crossing point at $T_C \sim 4.51$ is in good agreement with previous simulation results [BL01], according to which the critical temperature is located at 4.5115 [Heu93] and 4.5103 [Fis67]. These values are shown as dashed lines in Fig. 5.7.

5.4 Multi Spin Coded 2D CPU Implementation

For our extended CPU reference implementation, we focus on a single spin-flip approach which performs well for large lattice sizes. Multi-spin coding refers to all techniques that store and process multiple spins in one unit of computer memory. In CPU implementations, update schemes have been developed that allow for processing more than one spin with a single operation [WZK84, ZHR81, IK90, IK88]. We use a scheme which encodes 32 spins into one 32-bit integer in a linear fashion. The 32-bit type is chosen because register operations of current hardware perform fastest on this data type. The key ingredient for an efficient update algorithm of these 32-bit patterns is to use pre-computed bit patterns that encode the evaluations of the flip condition expression

$$r < \exp(-\Delta\mathscr{H}/k_B T) \tag{5.6}$$

for every single spin bit—the variate r is an independent and identically distributed random number in $[0,1)$. Since there are only two possible energy differences $\Delta\mathscr{H}$ with $\Delta\mathscr{H} > 0$, two *boolean arrays* can encode the information of an evaluation of the flip condition. For reasonable results, N. Ito [IK88] suggested to use a pool of 2^{22} to 2^{24} Boltzmann patterns. We denote the two Boolean arrays "exp4" and "exp8". Our encoding is chosen such that a "1" is stored in exp4 if $\exp(-8J/k_B T) < r < \exp(-4J/k_B T)$ is satisfied and a "0" if not. A "1" is stored in exp8 if $r < \exp(-8J/k_B T)$ is satisfied and a "0" if not. For every spin update, the Monte Carlo simulation will choose a random pattern. To process a 32-bit spin pattern s_0, neighbor patterns s_1, s_2, s_3, s_4 have to be prepared which contain the neighbors of the i^{th} spin at their i^{th} digit.

Since the calculation is the same for every bit, it is convenient to look at just one bit. The first step is to transform the spin variables into "energy variables" to eliminate dependence on the initial state of s_0.

$$i_n = s_0 \hat{} s_n, \forall n \in \{1,2,3,4\} \tag{5.7}$$

where "^" denotes an XOR operation. Because of the special encoding of the Boltzmann patterns, the acceptance condition for each spin can be expressed in a simple way:

$$i_1 + i_2 + i_3 + i_4 + 2 \cdot \text{exp8}_s + \text{exp4}_s \geq 2 \tag{5.8}$$

where exp8_s and exp4_s denote the sth Boltzmann patterns that encode the spin flip condition, and s is a random position in the pool. It is possible to evaluate this expression for all 32 bits in parallel by applying a sequence of bitwise boolean operations [IK88].

As parallel updates are only allowed on non-interacting domains, an additional bit mask has to be applied to the update pattern that only allows to flip each second spin in parallel.

Table 5.1: Key facts and properties of the Tesla C1060 GPU[NVI09].

	Tesla C1060
Global memory	4096 MB
Streaming processor cores	240
Shared memory per block	16 KB
Clock rate	1.30 GHz
Memory clock	800 MHz
Maximal power consumption	187.8 W

5.5 Multi Spin Coded GPU Implementation

5.5.1 Problems Arising from a Straightforward Adaption

On GPUs, memory access is very costly compared to operations on registers. The great advantage of multi-spin coding is that only one memory access is needed to load several spins in parallel. The CPU implementation could be ported to a GPU with a kernel that uses less than 16 registers. This allows optimal usage of the GPU up to a maximum block size of 512 threads. Even though the update scheme presented in section 5.4 performs faster on the CPU compared to an implementation with integer representations of each spin, a straightforward GPU port of this scheme is not optimal.

The reason for the poor performance is that parallel threads in one warp have to access global memory in a random fashion which is very costly. The execution speed can be improved by choosing only one random position per block and letting all the threads in this block read the patterns linearly, starting from the chosen starting position. However, this approach reduces the quality of the flip patterns. In principle, this ansatz could be compensated for by using a significantly larger pool of random numbers. Another option is to calculate the spin flip patterns on the fly using a random number generator on the GPU instead of looking them up from global memory. It turns out, however, that the sophisticated update scheme does not longer benefit in this case. The performance of this implementation is compared in Table 5.2. It should be emphasized that the quality of random numbers differs between the implementations. In the next section, we present another update scheme that works well on the GPU, and which prevents pitfalls with the quality of the random numbers.

In this section, we use a NVIDIA Tesla C1060 as our CUDA enabled device, which offers 4 GB of global memory, see Table 5.1. This memory can store a multi-spin coded spin field of $100,000^2$ spins on one GPU. The reference CPU used in tests in sections 5.4, 5.5, and 5.6 is the Intel Xeon X5560 with a clock rate of 2.80 GHz and 8192 kB cache. The purpose of the CPU implementation is to have a fast and fair non-parallel reference implementation, not to benchmark a Core i7 CPU. Therefore, only one core of the CPU is used.

5.5.2 Extraction into Shared Memory

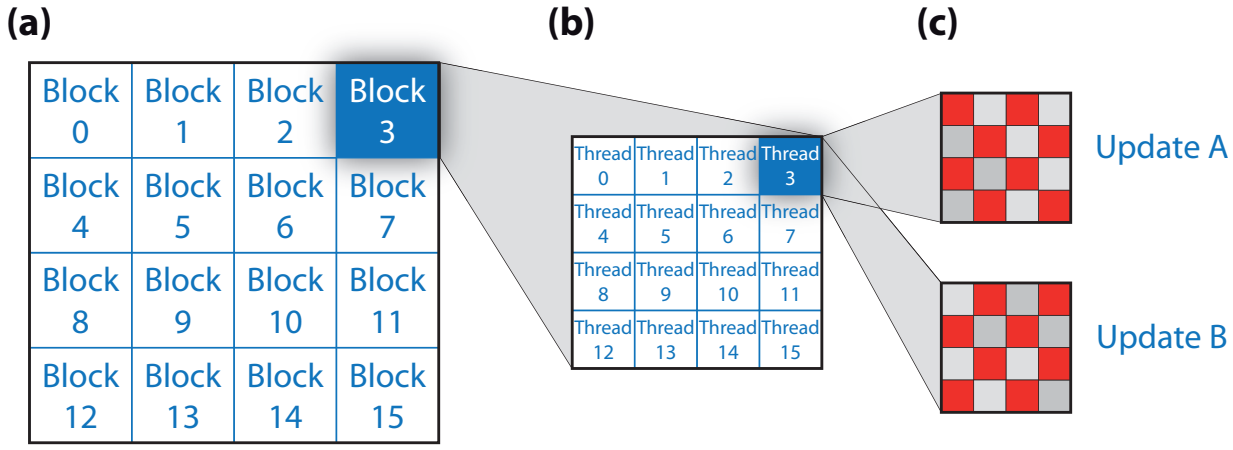

Figure 5.8: The spin lattice is processed by a variable number of blocks (a), where each block runs a variable number of threads (b). The threads update the spin lattice in two steps, *A* and *B*, using two kernel invocations (c).

The main goal of the implementation presented here is to reduce access to the global memory of the GPU, which is extremely costly. The best performance without pre-calculating flip patterns can be achieved by extracting the spins into shared memory and performing the calculations on integer registers. The spin field on the graphics card is encoded in quadratic blocks of 4×4 spins (hereafter referred to as "meta-spins") which can be stored as binary digits of one unsigned short integer (*2 bytes*), and thus can be accessed by a single memory lookup. Single spin values can be extracted from one meta-spin by using the expression

```
s[x,y] = (meta-spin & (1 << (y * 4 + x))) * 2 - 1
```

which returns a value of either -1 or 1. This is a slightly more complicated expression than for a linear layout, but it makes sense for a multi-GPU implementation in which border information has to be transferred between various GPUs (see section 5.6). This approach accounts for the fact that each spin uses exactly one bit of memory. The spin field is stored in global memory, which is expensive to access.

To process the spin field on the GPU, the spin field is subdivided into quadratic sub-fields which can be processed by threads grouped into one block (see Fig. 5.8). Each thread of this block processes a "meta-spin" of 4×4 spins. At the beginning of a kernel, it retrieves 5 meta-spins from the global memory, namely its own and its four neighboring "meta-spins" (Fig. 5.9a). This information is used to extract the information for the 4×4 spins. Each thread will store the spin field of 4×4 spins as well as the neighboring spins in a 6×6 integer array in shared memory, which allows for fast computation of the spin flips. The spin update is performed in two steps as described before. A first kernel is needed to update the "black" sites on a checkerboard pattern, and a second

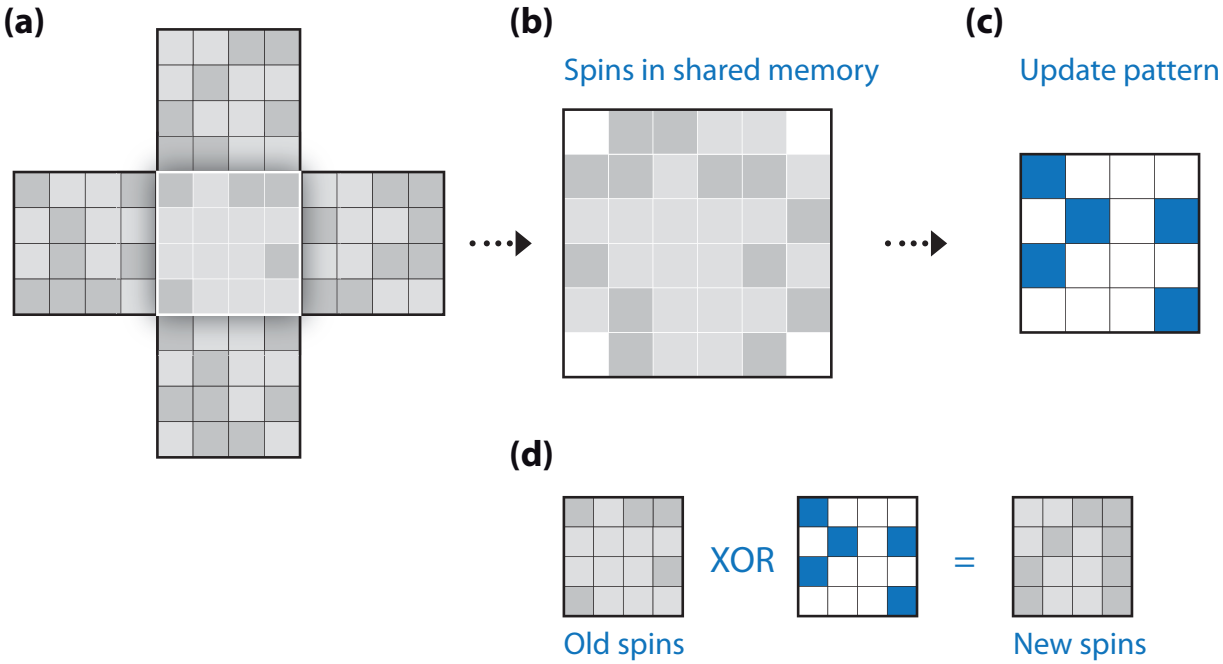

Figure 5.9: (a) The way a kernel processes a 4×4 meta-spin. (b) Spins are extracted into shared memory and an update pattern is created (c). (d) Afterwards, the new spins are obtained using the update pattern (Spins on blue sites will be flipped, spins on white sites will not be flipped), and written back to global memory.

processes the "white" sites. The update kernel for the "white" sites has to wait until all "black" sites have been updated. Thus, two separate kernels are needed. As discussed earlier, there is no other way to achieve global synchronization between the threads. Each kernel creates an update pattern, in which each binary digit indicates whether the associated spin has to be flipped or not. At the end of the kernel's execution, the 4×4 "meta-spins" are updated with one global memory write. In summary, each update thread executes the steps as follows: (a) Look up meta-spins from global memory. (b) Extract meta-spins into 6×6 integer array in shared memory which then contains the 4×4 "meta-spins" and the neighbors. (c) For all 8 "white"/"black" sites s_i in the 4×4 field, draw a random number and evaluate the Metropolis criterion. (d) Generate the update pattern (set the i^{th} bit to 1, if the flip of the i^{th} was accepted, otherwise to 0) (e) Update the "meta-spin" by an XOR operation with the update pattern to obtain the spins at the next time-step

Although this update scheme hardly sounds efficient, it dramatically reduces global memory access as compared to the previous implementation, which results in shorter computing times on GPU hardware. After the update is completed, the magnetization per spin $m(T)$ has to be extracted from the lattice. In a first step, the magnetization of each block can be aggregated using the shared memory of each block by employing a binary tree reduction and writing the total magnetization of the slice back to main

memory. The final summation of the magnetizations of the blocks can be either carried out on the CPU or on the GPU at about equal speeds.

5.5.3 Performance Comparison

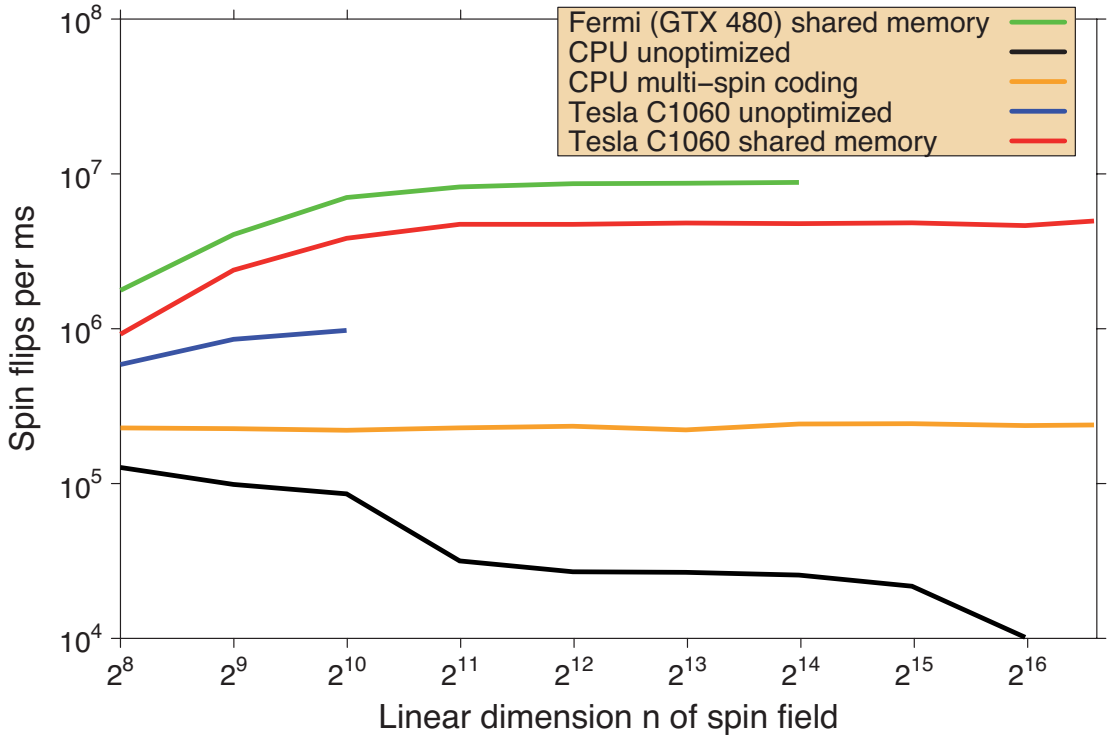

Figure 5.10: Benchmarking the implementations: The system is simulated at a constant temperature of $T = 0.99\ T_C$. The performance of the straightforward implementation varies strongly with lattice size because of the large block size of 512 threads, while the shared memory implementation offers stable performance over a wide range of sizes and offers better quality random numbers—comparable to the simple CPU implementation. For this benchmark, a GPU of NVIDIA's new Fermi generation could be used which became available in April 2010. A GeForce GTX 480 provides the following features: 1536 MB global memory, 480 streaming processor cores, 1.40 GHz processor clock, 1848 MHz memory clock, and a maximal power consumption of 250 W. The shared memory approach performs roughly two times faster than on a Tesla C1060.

The "multi-spin" implementations are compared to the "one spin per integer" implementations both on a CPU and a GPU. As a measurement for the performance of an implementation, we use the number of single spin flips per second, which also allows for comparing results given different lattice sizes. The temperature is set to $0.99\ T_C$. The GPUs perform most efficient for lattice sizes of a linear dimension beyond 4096×4096.

Table 5.2: Comparison at lattice size 4096 × 4096: *CPU simple* encodes one spin in one integer, *CPU multi-spin coding* uses the efficient "multi-spin" update scheme presented in section 5.4, *multi-spin unmodified* is a straightforward porting of this update to the GPU, *multi-spin coding on the fly* uses the same scheme but calculates the update patterns at each update step on the fly, and *multi-spin coding linear* determines one starting position in the random number pattern in the pool per block, and lets the threads read the random numbers linearly from that position on. The *shared memory* implementation (see section 5.5.2) provides random numbers with a better quality.

	Spinflips per μs	Relative speed
CPU simple	26.6	0.11
CPU multi-spin coding	226.7	1.00
shared memory	4415.8	19.50
multi-spin unmodified	3307.2	14.60
multi-spin coding on the fly	5175.8	22.80
multi-spin coding linear	7977.4	35.20

For this lattice size, a GPU is faster by a factor of about 15-35, depending on the implementation and the resulting quality of random numbers. For the implementation used in section 5.2, for ranges between 1024 × 1024 and 2048 × 2048 spins, the spin field size becomes comparable to the CPU L3 cache size, leading to a higher rate of costly L3 cache misses. This is the point at which the previous implementation becomes inefficient.

5.6 Multi GPU Approach

5.6.1 Implementation

The general idea is to extend the quadratic lattice by putting multiple quadratic "meta-spin" lattices next to each other in a super-lattice (see Fig. 5.11a for a 2×2 super-lattice) and letting each lattice be handled by one of the installed GPUs. On the border of each lattice, at least one of the neighboring sites is located in the memory of another GPU (see Fig. 5.11b). For this reason, the spins at the borders of each lattice have to be transferred from one GPU to the GPU handling the adjacent lattice. This can be realized by introducing four neighbor arrays containing the spins of the lattices' own borders, and four arrays for storing the spins of its adjacent neighbors (see Fig. 5.11c). At the beginning of execution, each MPI process initializes its own spin lattice, writes out its border spins into its own border arrays and sends them to its neighbors. In return it receives the adjacent borders from the according MPI processes. After this initialization phase, spins and random seeds are transferred to the GPU. Then, a single

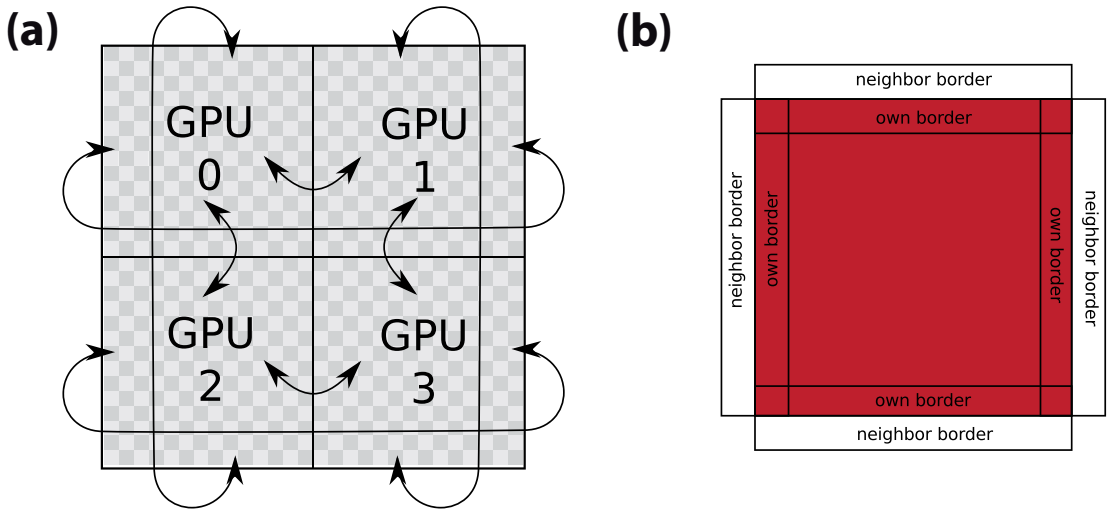

Figure 5.11: (a) Each GPU processes a "meta-spin" lattice of size $N = n^2$. The lattices are aligned on a super-lattice, and the outer borders are connected via periodic boundary conditions. In this example, 4 GPUs process a system of $2^2 \cdot N$ spins. (b) A meta-spin update needs the 4 nearest neighbor meta-spins. On the borders of a lattice, each GPU needs the spin information of the neighboring lattices. The border information has to be passed between the GPUs. In our implementation this is done by using 8 neighbor arrays.

lattice update is performed as follows: (a) Copy neighboring borders to GPU memory. (b) Call kernel to perform update A. (c) Call kernel to extract borders from the spin array to own borders array. (d) Copy own borders to host memory. (d) Exchange borders with the other MPI processes. (d) Copy neighbor borders to GPU memory again. (e) Call kernel to perform update B. (f) Call kernel to extract borders from spin array again. (g) Transfer own borders to host memory. (h) Exchange borders with other MPI processes. (i) Retrieve processed data from GPU.

It turns out that the transfer time was not the limiting factor for our purposes but rather the latency of the memory accessed.

5.6.2 GPU Cluster Performance

For performance measurements on the GPU cluster, the shared memory implementation (see section 5.5.2) was used, since it provided stable performance for various lattice sizes and because the memory layout is symmetric in the x and y directions, resulting in symmetric communication data. The tests were run on a GPU cluster with two Tesla C1060 GPUs in each node. Communication is established via Double Data Rate InfiniBand. The performance for various system sizes (see Fig. 5.12) provides evidence that for more than one GPU, spin flip performance scales nearly linearly with the number of GPUs. The drop from one GPU to four GPUs is due to the communication over-

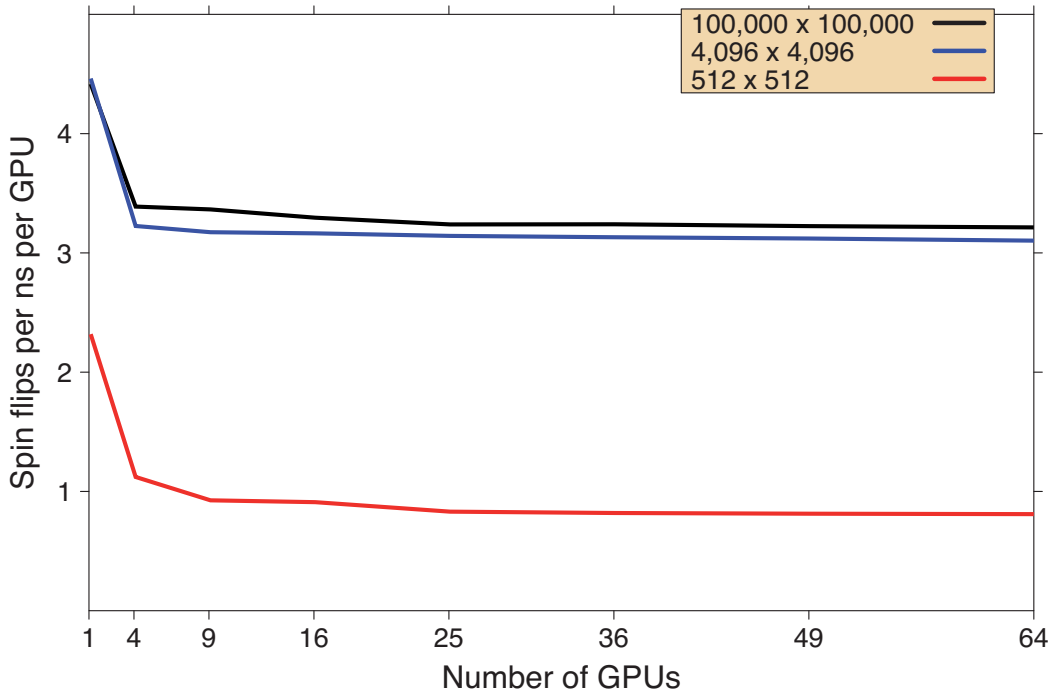

Figure 5.12: Cluster performance for various system sizes (per GPU). For more than one GPU, spin flip performance scales nearly linearly with the amount of GPUs. Again, optimal performance is reached at a lattice size of about 4096×4096 per GPU. Using 64 GPUs, a performance of 206 spin-flips per nanosecond can be achieved on a 800.000×800.000 lattice.

head resulting from the exchange of borders. For larger system sizes, the communication overhead per CPU/GPU remains constant. An optimal performance is reached for lattice sizes beyond 4096×4096 per GPU. For 64 GPUs—the NEC Nehalem Cluster maintained by the High Performance Computing Center Stuttgart (HLRS) provides 128 GPUs—a performance of 206 spin-flips per nanosecond can be achieved on a 800.000^2 2D Ising lattice. That is, the entire lattice can be updated in about three seconds.

5.7 Chapter's Summary

We presented a GPU accelerated version of the two dimensional ferromagnetic square lattice Ising model and the three dimensional ferromagnetic cubic lattice Ising model. For the GPU implementation, the compute unified device architecture was employed which is a new programming approach for utilizing a graphics processing unit (GPU) as a data-parallel computing device. The programming interface allows for the implementation of algorithms using extensions to the standard C programming language. With their continuously increasing number of cores in combination with a high memory bandwidth, a modern GPU offers viable, incredible resources for general purpose

computing. For our GPU based Monte Carlo simulations of the Ising model, we use a set of linear congruential random number generators on the GPU device. With the GPU implementation of a checkerboard algorithm of the two dimensional Ising model, results can be obtained up to 60 times faster than can be obtained from a CPU core. An implementation of a three dimensional Ising model on a GPU is able to generate results up to 35 times faster than an an implementation on a modern CPU core. As proof of our methodology for the GPU implementation, the critical temperatures of the 2D and 3D Ising models were determined successfully by finite size scaling. Both the theoretical result for the 2D Ising model and previous simulation results for the 3D Ising model can be reproduced.

Additionally, two major improvements were achieved. By using multi-spin coding techniques, we improved the computation to memory access ratio of our calculations dramatically, resulting in better overall performance. On a single GPU, up to 7.9 spin-flips per nanosecond are possible, 15 to 35 times faster than our multi-spin coded CPU version (see section 5.5), depending on the implementation and the quality of random numbers. The other improvement lies in the usage of GPU clusters, in which the 2D Ising lattice is distributed over many GPUs. We show that our implementation scales nearly linearly with the number of GPUs, allowing for the processing of huge Ising lattices on GPU clusters.

Chapter 6

Trend Switching Processes in Financial Markets

In physics and in other natural sciences, it is often a successful strategy to analyze the behavior of a system by studying its smallest components. For example, a molecule is composed of atoms, an atom consists of a nucleus and electrons, a nucleus consists of protons and neutrons, and so on. The fascinating point about analyses on steadily decreasing time and length scales is that the composite system often exhibits properties which cannot only be explained by the properties of its components alone. Instead, a complex behavior can emerge due to the interactions among these components [And72]. In financial markets, these components are comprised of market participants who buy and sell assets according to their trading and investment decisions. The superimposed flow of all individual orders submitted to the exchange trading system initiated by market participants and, of course, its change in time generate a complex system with fascinating properties, similar to physical systems.

One of the key conceptual elements in modern statistical physics is the concept of scale invariance, codified in the scaling hypothesis that specific functions are given as power-laws in the vicinity of a critical point [Sta71]. The scaling hypothesis has two categories of predictions, both of which have been remarkably well verified by a wealth of experimental data on diverse systems. The first category is a set of relations, called *scaling laws*, that serve to relate the various critical-point exponents characterizing the singular behavior of functions such as thermodynamic functions. The second category is known as *data collapse*, where under appropriate axis normalization, diverse data "collapse" onto a single curve called a scaling function.

Econophysics research has been addressing a key question of interest: quantifying and understanding large stock market fluctuations. Previous work was focused on the challenge of quantifying the behavior of the probability distributions of large fluctuations of relevant variables such as returns, volumes, and the number of transactions. Sampling the far tails of such distributions requires a large amount of data. A very large amount of pre-existing, precise financial market data has, however, already been collected; many orders of magnitude more so than for other complex systems. Accord-

ingly, financial markets are becoming a paradigm for complex systems, and increasing numbers of scientists are analyzing market data [Axt01, Tak06, KSY06, WTT07, GGPS03, PPS08, LFM03, PGGS02, CB00, KHH02, O'H97, VA97, EK06]. Empirical analyses have focused on quantifying and testing the robustness of power-law distributions that characterize large movements in stock market activity. Using estimators that are designed for serially and cross-sectionally independent data, findings thus far support the hypothesis that the power-law exponents that characterize fluctuations in stock price, trading volume, and the number of trades [Fam63, Lux96, GDD+97, PGR+99] seem to be "universal" in the sense that they do not change their values significantly for different markets, different time periods, or different market conditions.

In contrast to these analyses of "global" financial market distributions, in this chapter we focus on the temporal sequence of fluctuations in transaction volume and intertrade times before and after a trend switching point [1]. Our analysis provides insight into switching processes in complex systems in general and into financial systems in particular. The study of dramatic crash events is limited by the (fortunately) rare number of such events. Increasingly, understanding of the current financial crisis has been sought through comparison with the depression of the 1930's. Here we ask if the smaller financial crises—trend switching processes on all time scales—also provide information of relevance for large crises. If this is the case, then the large abundance of data on smaller crises should provide quantifiable statistical laws for *bubbles on all scales*.

6.1 Overview

In a variety of switching processes in nature, the underlying complex system abruptly changes from one state to another in a highly discontinuous fashion. Financial market fluctuations are characterized by many abrupt switchings creating upward trends and downward trends on time scales ranging from the macroscopic, persisting for hundreds of days, to microscopic trends lasting no more than a few minutes. An important question is whether or not these ubiquitous switching processes have quantifiable features independent of the time horizon studied. We find striking scale-independent behavior of the transaction volume after each switching event. Our findings can be interpreted as being consistent with the time-dependent collective behavior of financial market participants. We test the possible universality of our result by performing a parallel analysis of fluctuations in time intervals between transactions, and suggest that the well-known catastrophic bubbles that occur on large time scales—such as the most recent financial crisis—may not be outliers but rather single dramatic examples caused by the formation of increasing and decreasing trends on time scales varying over 9 orders of magnitude from very large down to very small.

As stated already, the study of dramatic crash events is limited by the fortunately rare number of such events. The large amount of available data on financial mar-

[1] Results of this chapter are published in [PS09, PS10a, SBF+10, PS10b].

kets, however, has allowed them to become a prime example for complex systems [Axt01], and increasing numbers of scientists are analyzing market data [Tak06, KSY06, WTT07, GGPS03, LFM03, PGGS02, CB00, KHH02] and modeling financial markets [Hom02, Sha06, LM99, O'H97, HLL08, HKK$^+$08, HH07, SFGK03, BP01]. The probability distribution function and the time autocorrelation function reveal interesting features, such as long-range power-law correlations in volatility and fat tails in the price change probability distribution function [VA97, EK06].

Increasingly, understanding of the current financial crisis has been pursued through comparisons with the depression of the 1930's. Here we pose the question of whether the smaller financial crises also provide information of relevance to large crises. If this is the case, then the larger abundance of data on smaller crises should provide quantifiable statistical laws for bubble formation and financial collapse on various scales. In answering this question, we perform parallel analyses of trend switching on two quite different time scales[2]: (i) from ≈ 10 ms to $\approx 10^6$ ms, and (ii) from $\approx 10^8$ ms to $\approx 10^{10}$ ms.

(i) German market: For the first analysis, we use a price time series of the German DAX Future (FDAX) traded on the Eurex. The time series is composed of $T_1 = 13,991,275$ trades in three disjoint three-month periods. The data base contains the transaction prices, the volumes, and the corresponding time stamps [JKL94, CF00, PS08, JCZ09], with a large liquidity and inter-trade times as low as 10 ms, which allows for the analysis of microtrends (see Fig. 6.1a).

(ii) US market: For the second analysis, which focuses on macrotrends, we use price time series of daily closing prices for all stocks included in the S&P500 index. The time series is composed of overall $T_2 = 2,592,531$ closing prices. The earliest closing prices in the dataset date back to 2 January 1962. The database contains the daily closing prices and the daily cumulative trading volumes.

To analyze the switching processes of financial fluctuations, we first lay out a framework within which such an analysis can be carried out. Let $p(t)$ be the transaction price of trade t, which is a discrete variable $t = 1, \ldots, T$. A transaction price $p(t)$ is defined to be a *local maximum* of order Δt if there is no higher transaction price in the interval $t - \Delta t \leq t \leq t + \Delta t$, and is defined to be a *local minimum* of order Δt if there is no lower transaction price in this interval (Fig. 6.1b).

Here, we perform an analysis of the volume fluctuations $v(t)$ from one price extremum to the next. The volume is the number of contracts traded in each individual transaction for microtrends in the German market and the number of traded stocks per day for macrotrends in the US market. For the analysis, we introduce a renormalized time scale ε between successive extrema. Thus, $\varepsilon = 0$ corresponds to the beginning of a trend and $\varepsilon = 1$ indicates the end of a trend (Fig. 6.1c). We analyze a range of ε for the interval $0 \leq \varepsilon \leq 2$, in order to consider analyze trend switching processes both before as well as after the critical value $\varepsilon = 1$ (Fig. 6.1). The renormalization is essential to assure that trends of various lengths can be aggregated and that all switching points have a

[2] In section 6.6, an additional data base is used covering 2,623,445,866 transactions.

common position in the renormalized time.

Figure 6.2a displays the volume $v^*(\varepsilon)$ averaged over all increasing and decreasing microtrends in the entire set of $T_1 = 13,991,275$ records and normalized by the average volume of all microtrends studied. In order to remove outliers (e.g., overnight gaps) only those microtrends are collected in which the time intervals between successive trades $\tau(t)$ [IYPL04] do not last longer than 1 minute, which is roughly 60 times longer than the average inter-trade time (≈ 0.94 s). Furthermore, transaction volumes do not have to be larger than 100 contracts (the average transaction volume is 2.55 contracts). As expected, new price extrema are linked with peaks in the volume time series but, surprisingly, we find that the usual cross-correlation between price changes and volumes is very small. In Fig. 6.2d, we show the averaged volume $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram. Surprisingly, average volume decreases on straight lines and thus indicates a power-law scaling behavior of the form

$$v^*(|\varepsilon - 1|) \sim |\varepsilon - 1|^{\beta_v} \tag{6.1}$$

with scaling parameters $\beta_v^- = -0.068 \pm 0.001$ (t-test, p-value $< 2 \times 10^{-16}$) before, and $\beta_v^+ = -0.155 \pm 0.004$ (t-test, p-value $= 9.2 \times 10^{-16}$) after price extremum. Such an extraction of slopes by performing least-squares linear regressions is not sufficient for the claim that the averaged volume follows a power-law distribution. However, additionally performed statistical tests enable us to conclude that our observations are indeed consistent with this hypothesis.

Next we test the possible universality of our result by performing a parallel analysis for trends on long time scales using the daily S&P500 closing price data. Note that for our parallel analysis on macroscopic time scales, the order of an extremum Δt is measured in units of days, and that $v^*(\varepsilon)$ is averaged over all trends and all closing price time series of all S&P500 components. In order to avoid biased contributions to the rescaled averaging caused by inflation based drifts over more than 47 years, the analyzed price time series $p(t)$ contains the logarithm of the daily closing prices. A log–log histogram of our parallel analysis for the US market on large time scales (Figs. 6.2b and 6.2e) provides evidence for similar behavior with scaling parameters $\beta_v^- = -0.052 \pm 0.001$ (t-test, p-value $= 1.7 \times 10^{-9}$) before, and $\beta_v^+ = -0.109 \pm 0.003$ (t-test, p-value $< 2 \times 10^{-16}$) after a price extremum. Statistical tests confirm the consistency with a power-law distribution.

In order to verify a possible universality, we analyze the behavior of the inter-trade times $\tau(t)$ of the German market during the short time interval from one price extremum to the next. The cross-correlation function between price changes and inter-trade times exhibits no reasonable correlation values as well. Thus, it is conceivable that the tendency to decreased inter-trade times at the end of positive microtrends is counteracted by the tendency for decreasing inter-trade times at the end of negative microtrends. The crucial issue is to distinguish between positive and negative microtrends realized by the renormalized time ε between successive extrema. In Fig. 6.2c, the averaged inter-trade times $\tau^*(\varepsilon)$ reflect the link between inter-trade times and price extrema.

Figure 6.2f shows $\tau^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram providing support for power-law behavior of the form

$$\tau^*(|\varepsilon - 1|) \sim |\varepsilon - 1|^{\beta_\tau} \tag{6.2}$$

with scaling parameters $\beta_\tau^- = 0.092 \pm 0.002$ (t-test, p-value $= 1.8 \times 10^{-15}$) before, and $\beta_\tau^+ = 0.118 \pm 0.002$ (t-test, p-value $< 2 \times 10^{-16}$) after a price extremum. Statistical tests confirm consistency with a power-law distribution in this case as well.

The straight lines in Figs. 6.2d, 6.2e, and 6.2f offer insight into financial market fluctuations: (i) a clear connection between volumes, inter-trade times, and price fluctuations on the path from one extremum to the next extremum seems to exist, and (ii) the underlying law, which describes the volumes and inter-trade times around extrema varying over 9 orders of magnitude starting from the smallest possible time scale, is a power-law with scaling parameters which quantitatively characterize the region around the trend switching point. As a direct consequence of the consistency with power-law distributions, the observed behavior does not depend on the time scale considered. Thus, we find identical behavior for other sub-intervals of $50 \leq \Delta t \leq 1000$.

In summary, we have seen that each type of trend—micro and macro—in a financial market starts and ends with a unique switching process, and each extremum shares properties of macroscopic cooperative behavior [Kru96, Shl00, HFV00, BKS02]. We have seen that the switching mechanism has no scale, for time scales varying over 9 orders of magnitude down to the smallest possible time scale (the scale of single transactions measured in units of 10 ms). Thus, the well-known catastrophic bubbles occurring on large time scales—such as the most recent financial crisis—may not be outliers but in fact single dramatic events caused by the inherent, scale-free behavior related to the formation of increasing and decreasing trends on time scales from the very large down to the very small.

The findings which were briefly covered in this overview are documented in detail in the following sections. The remainder of this chapter is organized as follows: Section 6.2 describes the underlying data sets used for the analyses of switching processes in financial markets. Section 6.3 documents the renormalization method in detail. Statistical tests confirming our results are found in section 6.4. As an additional test, we reshuffle the inter-trade time series and the volume time series in section 6.5 in order to destroy their link to the price development observable through the power-law behavior covered in this overview. Section 6.6 presents a dramatically extended analysis of switching processes in financial markets, in which we rely on a data base covering 2,623,445,866 transactions. Finally, section 6.7 summarizes the results of this chapter.

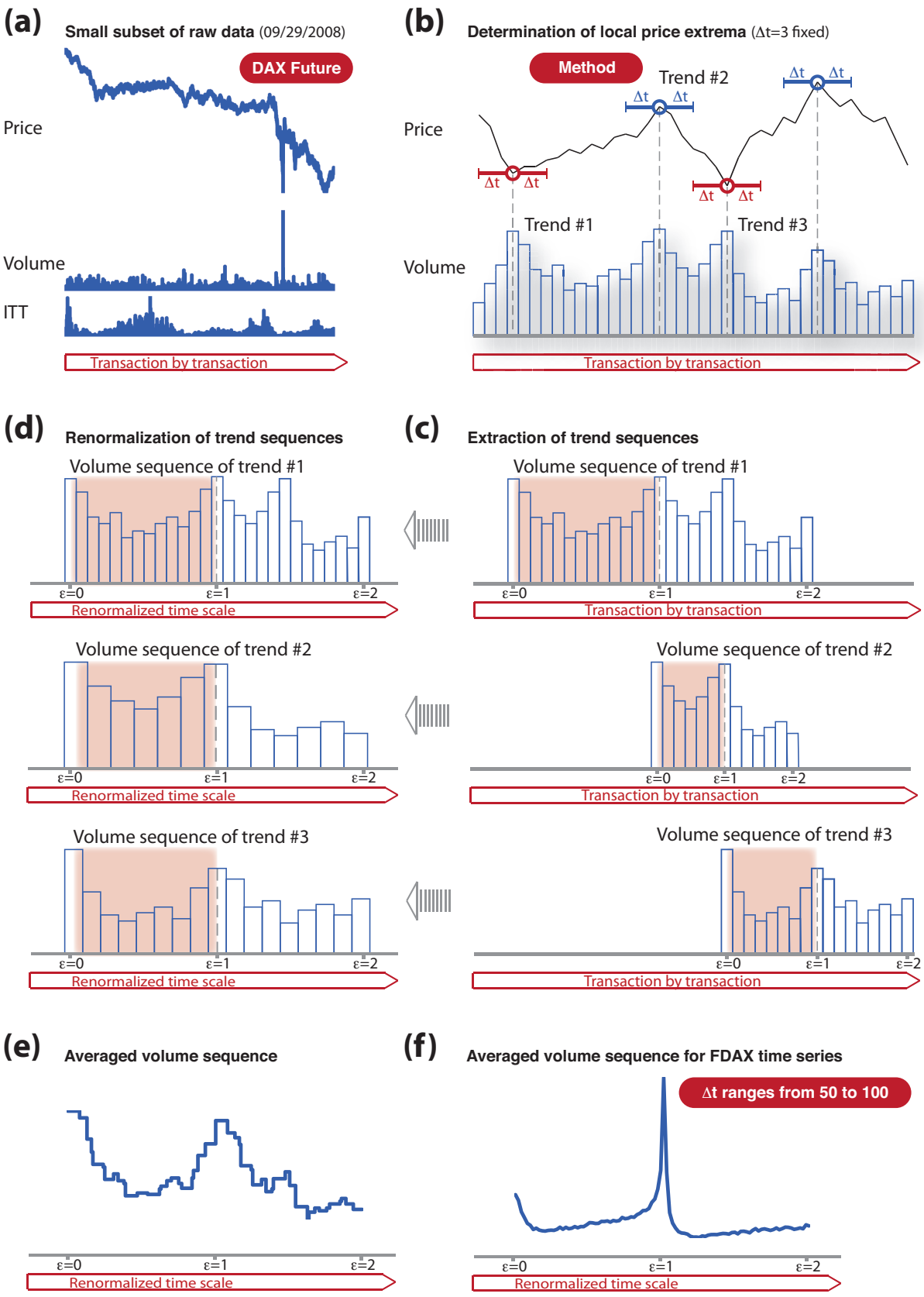
(a)
Small subset of raw data (09/29/2008)
DAX Future
Price
Volume
ITT
Transaction by transaction
(b)
Determination of local price extrema (Δt=3 fixed)
Method
Trend #2
Price
Trend #1
Trend #3
Volume
Transaction by transaction
(d)
Renormalization of trend sequences
Volume sequence of trend #1
ε=0
ε=1
ε=2
Renormalized time scale
Volume sequence of trend #2
Renormalized time scale
Volume sequence of trend #3
Renormalized time scale
(c)
Extraction of trend sequences
Volume sequence of trend #1
Transaction by transaction
Volume sequence of trend #2
Transaction by transaction
Volume sequence of trend #3
Transaction by transaction
(e)
Averaged volume sequence
Renormalized time scale
(f)
Averaged volume sequence for FDAX time series
Δt ranges from 50 to 100
Renormalized time scale

Figure 6.1: Segregation and rescaling of trend sequences in a multivariate time series in order to analyze financial market quantities on the path from one price extremum to the next. (a) Small subset comprising 121,400 transactions of the full data set (13,991,275 transactions) analyzed, extracted from the German DAX future (FDAX) time series which provides transaction prices, transaction volumes, and time intervals between transaction—inter-transaction times (ITT). This subset recorded on September 29, 2008 documents the volatile reaction of stock markets as the US government's $700 billion financial bailout plan was rejected by the House of Representatives on that day. (b) Schematic visualization of trend segregation for $\Delta t = 3$. Positive trends start at local price minima (red circles) and end at local maxima (blue circle)—and vice versa. A transaction price $p(t)$ is a local maximum if there is no higher transaction price in the interval $t - \Delta t \leq t \leq t + \Delta$. Analogously, $p(t)$ is a local minimum if there is no lower transaction price in the interval $t - \Delta t \leq t \leq t + \Delta$. (c) Segregated sequences of transaction volumes belonging to the three trends identified in (b). We assign $\varepsilon = 0$ to the start of each trend, and $\varepsilon = 1$ to the end of each trend. In order to study trend switching processes—both before as well as after the end of a trend—we consider additionally the subsequent volume sequences of identical length. (d) Visualization of the volume sequences in the renormalized time scale. The renormalization assures that trends of various lengths can be aggregated as all switching points have a common position in this renormalized scale. (e) Averaged volume sequence derived from the summation of the three trend sequences. (f) Average volume sequence $v^*(\varepsilon)$ for all trends in the full FDAX time series derived from summation over various values of Δt. Extreme values of the price coincide with peaks in the time series of the volumes.

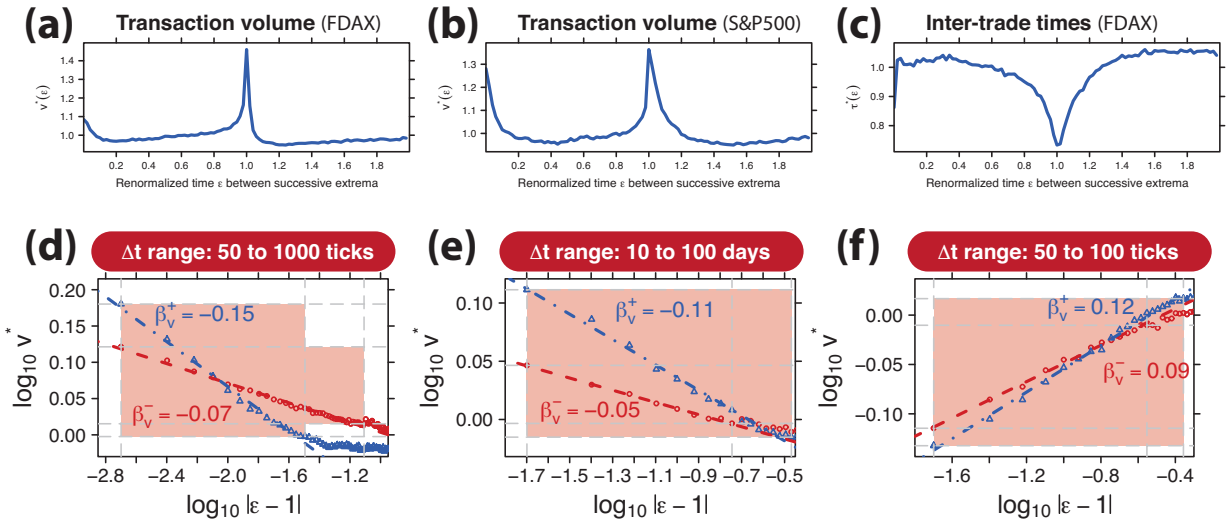

Figure 6.2: Renormalization time analysis and log–log plots of quantities with scale-free properties. (a) Averaged volume sequence $v^*(\varepsilon)$ of the German DAX Future time series. Δt ranges from 50 to 100 transactions (ticks). Extreme values of the price coincide with sharp peaks in the volume time series. (b) A very similar behavior is obtained for the averaged volume sequence $v^*(\varepsilon)$ of S&P500 stocks. Here, Δt ranges from 10 days to 100 days. (c) Averaged inter-trade time sequence $\tau^*(\varepsilon)$ of the German DAX Future time series. Extreme values of the price time series are reached with a significant decay of inter-trade times (50 ticks $\leq \Delta t \leq$ 100 ticks). (d) Log–log plot of the FDAX transaction volumes (50 ticks $\leq \Delta t \leq$ 1000 ticks) before reaching an extreme price value ($\varepsilon < 1$, circles) and after reaching an extreme price value ($\varepsilon > 1$, triangles). The straight lines correspond to power-law scaling with exponents $\beta_v^+ = -0.155 \pm 0.004$ (t-test, p-value $= 9.2 \times 10^{-16}$) and $\beta_v^- = -0.068 \pm 0.001$ (t-test, p-value $< 2 \times 10^{-16}$). The shaded intervals mark the region in which the empirical data are consistent with a power-law behavior. The left border of the shaded regions is given by the first measuring point closest to the switching point. The right borders stem from statistical tests of the power-law hypothesis (see section 6.4). (e) Log–log plot of the transaction volumes shown in (b) indicates a power-law behavior with exponents $\beta_v^+ = -0.109 \pm 0.003$ (t-test, p-value $< 2 \times 10^{-16}$) and $\beta_v^- = -0.052 \pm 0.001$ (t-test, p-value $= 1.7 \times 10^{-9}$) which are similar to our results on short time scales. (f) Log–log plot of the inter-trade times on short time scales (50 ticks $\leq \Delta t \leq$ 100 ticks) exhibits a power-law behavior with exponents $\beta_\tau^+ = 0.118 \pm 0.002$ (t-test, p-value $< 2 \times 10^{-16}$) and $\beta_\tau^- = 0.092 \pm 0.002$ (t-test, p-value $= 1.8 \times 10^{-15}$). An equivalent analysis on long time scales is not possible as daily closing prices are recorded with equidistant time steps.

6.2 Financial Market Data

To address the question of whether smaller financial crises also provide information of relevance to large crises, we perform parallel analyses of bubble formation and bursting using two different datasets on two quite different time scales: (i) from $\approx 10^1$ ms to $\approx 10^6$ ms, and (ii) from $\approx 10^8$ ms to $\approx 10^{10}$ ms.

6.2.1 German Market—DAX Future

For the first analysis, we use a multivariate time series of the German DAX Futures contract (FDAX) traded on the European Exchange (Eurex). A "futures" exchange or derivatives exchange is a central financial exchange where people can trade standardized "futures contracts". A "future" is a contract to buy or sell an underlying asset at a specified price at a specific future date—in this case the German DAX index, which measures the performance of the 30 largest German companies in terms of order book volume and market capitalization[3].

The time series contains $T_1 = 13,991,275$ transactions of three disjoint three-month periods (see Table 6.1). Each end of the three disjoint periods corresponds to a last trading day of the FDAX contract, which is fixed at the third Friday of the months March, June, September, and December, except for exceptions due to national holidays. The data set we analyze contains the transaction prices, the volumes, and the corresponding time stamps [JKL94, CF00, PS08, JCZ09], with large liquidity and inter-trade times as small as 10 ms, allowing for the analysis of microtrends.

Time series analysis of futures contracts has the advantage that prices are created by trading decisions alone. Stock index data, on the other hand, are derived from a weighted sum of a large number of stock prices. Furthermore, systematic drifts due to inflation are eliminated by construction. The theory of futures pricing based on arbitrage states that for an asset that can be stored at no cost and which does not yield any cash flows, the futures price F has to be equal to the spot price S plus the cost of financing the purchase of the underlying between the spot date and the expiry date [Dub04, Deu03]. This theoretical futures price can be referred to as *fair value*. In the case of the German DAX index, the underlying purchase can be financed until expiry at a standard loan rate. Using a continuously compounded rate r, the *fair value* equation can be written as

$$F(t) = S(t)e^{rt}, \tag{6.3}$$

where t denotes the remaining time until expiration. The expression for the theoretical futures price—see Eq. (6.3)—which simply reflects the *cost of carry*, compensates for interest rate related effects. At the time of expiry $t = 0$, the future's price and underlying price are identical.

[3]More detailed information about German DAX index constituents and calculation principles can be found on `http://www.deutsche-boerse.com`.

Table 6.1: Three disjoint three-month periods of the German DAX Future contract (FDAX) which we analyze. Additionally, the mean volume per transaction $\bar{v}$ and the mean inter-trade time $\bar{\tau}$ are shown.

Contract	Records	Time Period	$\bar{v}$	$\bar{\tau}$ [s]
JUN 2007	3,205,353	16 Mar 2007 – 15 Jun 2007	3.628[a]	2.485[b]
SEP 2008	4,357,876	20 Jun 2008 – 19 Sep 2008	2.558[a]	1.828[b]
DEC 2008	6,428,046	19 Sep 2008 – 19 Dec 2008	2.011[a]	1.253[b]

[a] measured in units of contract.
[b] including overnight gaps.

Cross Correlations

Cross-correlation is a measure of similarity of two time series as a function of a time-lag applied to one of them. For the analysis of trend switching points, we are interested in the question of to what extent the time series of price changes depends on the time series of inter-trade times, and vice versa. Figures 6.3a, 6.3b, and 6.3c provide evidence that only very small but at least a few significant correlation coefficients can be found for the DAX future time series. However, the small correlation coefficients for small time-lags differ strongly between the three disjoint data sets. Thus, one can conjecture that they are influenced by the randomly occurring over-night gaps which can be dramatically larger than price differences between two successive transactions during regular trading hours.

A parallel analysis of the cross-correlation functions (see Figs. 6.3d, 6.3e, and 6.3f) for price changes and transaction volumes uncovers absolute coefficients smaller than 0.017.

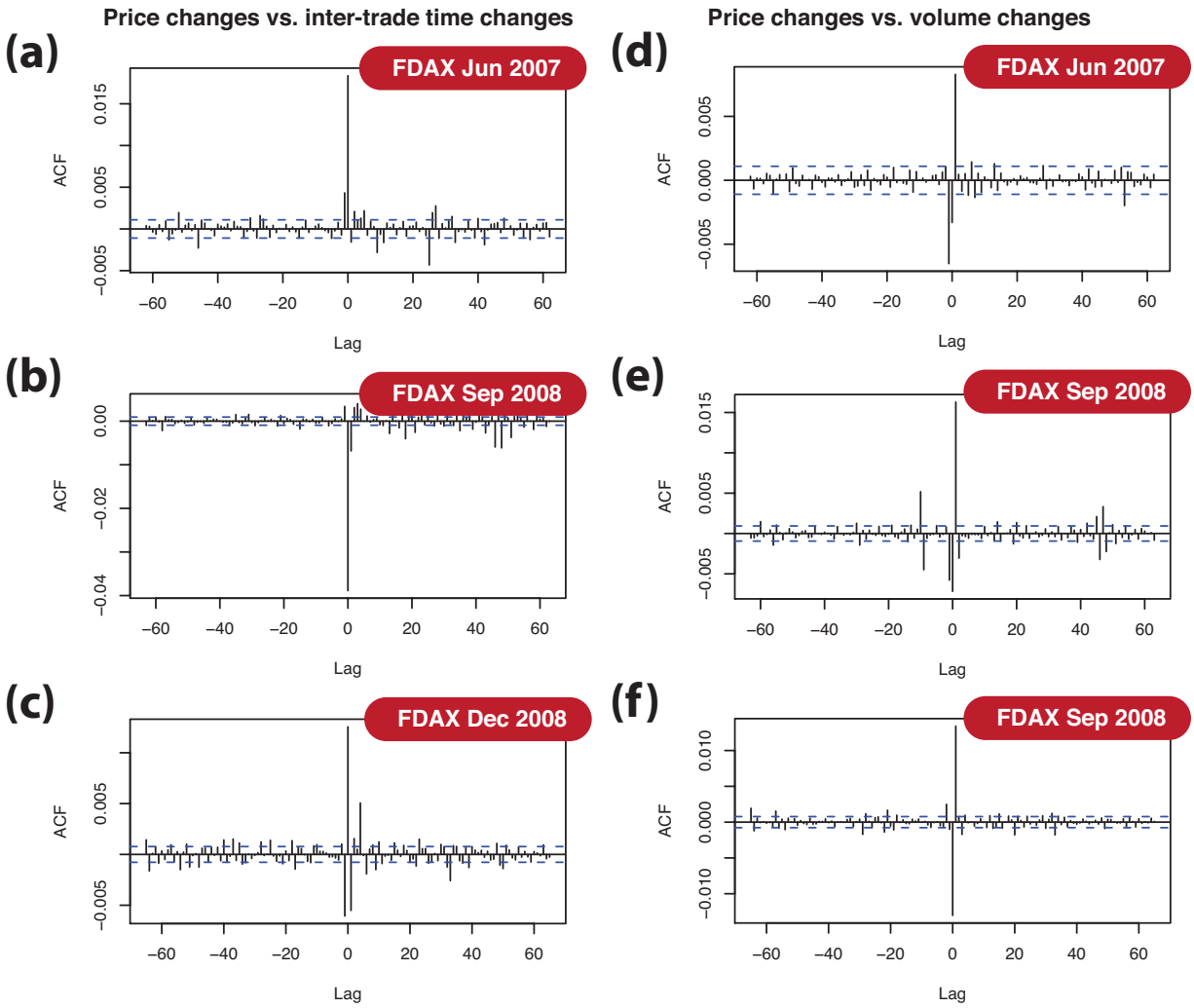

Figure 6.3: Cross correlation analysis of the quantities analyzed. (a) Time-lag dependent correlation between price changes and inter-trade times analyzed for the FDAX JUN 2007 contract. (b) Correlation between price changes and inter-trade times analyzed for the FDAX SEP 2008 contract. (c) Correlation between price changes and inter-trade times analyzed for the FDAX DEC 2008 contract. (d) Time-lag dependent correlation between price changes and volume changes analyzed for the FDAX JUN 2007 contract. (e) Parallel analysis for the FDAX SEP 2008 contract. (f) Parallel analysis for the FDAX DEC 2008 contract.

6.2.2 US Market—S&P500 Stocks

For the second analysis, which focuses on macrotrends, we use price time series of the daily closing prices of all stocks included in the S&P500 index. This index consists of 500 large-cap common stocks actively traded in the United States of America[4].

This time series includes $T_2 = 2,592,531$ closing prices of all US stocks which were included in the S&P500 on June 16, 2009. Our oldest closing prices date back to January 2, 1962. The data base of closing prices we analyzed contains the daily closing prices and the daily cumulative trading volume. The set of stocks which we analyze is shown in Tables 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9, and 6.10. Additionally, these tables provide the p-values of the Augmented Dickey-Fuller (ADF) test [SD84, BDGH93] and the p-values of the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test [KPSS92] for both considered quantities, namely prices and volumes. We perform these unit root tests to determine whether the time series of prices and volumes are stationary. The null hypothesis of the ADF tests is that a time series contains a unit root.

We find that the p-values for price time series of S&P 500 stocks support the rejection of the unit root hypothesis in 165 of 500 cases at the 5% significance level and in 54 of 500 cases at the 1% significance level. The p-values for the volume time series support the rejection of the unit root hypothesis in 489 of 500 cases at the 5% significance level and in 483 of 500 cases at the 1% significance level. This indicates that the volume time series are stationary. The number of lags used in the regression is $(T_i - 1)^{1/3}$ where T_i is the number of closing prices recorded for stock i.

As a second test we perform KPSS tests which are used for testing a null hypothesis that an observable time series is stationary around a deterministic trend. The time series is expressed as the sum of a deterministic trend, a random walk, and a stationary error. KPSS tests are intended to complement unit root tests such as the Dickey-Fuller tests. We find in 500 of 500 cases at a 5% significance level and in 498 of 500 cases for a 1% significance level. For the volume time series we obtain in 498 of 500 cases at 5% level of significance and in 492 of 500 cases at 1% level of significance.

As spot market prices are shifted by inflation over time periods of more than 40 years, we use the logarithm of stock prices for the analysis of trend switching processes instead of the closing prices. The results between the two different datasets on two quite different time scales can be better compared.

[4]More detailed information about S&P500 calculation principles can be found on `http://www.standardandpoors.com`.

Table 6.2: Trading symbols of S&P500 constituents with the corresponding date of the earliest closing price (YYYY/MM/DD). Additionally, the values of the ADF tests and KPSS tests for the price time series $p(t)$ and the volume time series $v(t)$ are shown with their corresponding p-values.

Symbol	First record	ADF_p	p-value	$KPSS_p$	p-value	ADF_v	p-value	$KPSS_v$	p-value
A	1999/11/18	−3.0055	0.1526	3.9842	0.01	−9.1849	0.01	0.5495	0.03052
AA	1962/01/02	−3.7079	0.02349	9.1188	0.01	−5.7815	0.01	15.6073	0.01
AAPL	1984/09/07	−2.4556	0.3854	7.3492	0.01	−9.1152	0.01	17.2867	0.01
ABC	1995/04/04	−2.5497	0.3456	3.1725	0.01	−7.2535	0.01	9.3419	0.01
ABT	1983/04/06	−3.9909	0.01	1.498	0.01	−7.8609	0.01	11.2965	0.01
ACS	1994/09/27	−2.8393	0.2229	7.1287	0.01	−8.8745	0.01	10.0268	0.01
ADBE	1986/08/14	−3.7252	0.02269	2.7708	0.01	−12.0336	0.01	10.7303	0.01
ADI	1984/07/19	−3.87	0.01544	11.4106	0.01	−12.092	0.01	26.6662	0.01
ADM	1983/04/06	−2.604	0.3225	6.9243	0.01	−7.07	0.01	14.2448	0.01
ADP	1983/04/06	−4.7252	0.01	1.9412	0.01	−9.2517	0.01	11.909	0.01
ADSK	1985/07/01	−4.8949	0.01	1.3272	0.01	−12.6106	0.01	17.4434	0.01
AEE	1998/01/02	−0.7001	0.9705	7.0075	0.01	−8.0109	0.01	13.2233	0.01
AEP	1970/01/02	−3.0133	0.1490	24.4386	0.01	−9.8982	0.01	18.5573	0.01
AES	1991/06/26	−2.2648	0.4662	6.3602	0.01	−8.343	0.01	14.7339	0.01
AET	1977/01/03	−3.1534	0.09584	8.3822	0.01	−11.4232	0.01	16.2625	0.01
AFL	1984/07/19	−3.4512	0.04694	18.5838	0.01	−6.3181	0.01	10.0441	0.01
AGN	1989/06/22	−2.3851	0.4153	20.1527	0.01	−9.88	0.01	16.8111	0.01
AIG	1984/09/07	−3.1493	0.09658	8.9116	0.01	−9.502	0.01	4.479	0.01
AIV	1994/07/22	−0.2938	0.99	5.6939	0.01	−4.3867	0.01	11.0744	0.01
AIZ	2004/02/05	−0.4565	0.9838	4.8726	0.01	−7.2676	0.01	4.5941	0.01
AKAM	1999/10/29	−3.1774	0.0918	2.8442	0.01	−8.0523	0.01	8.7294	0.01
AKS	1994/03/30	−2.15	0.5148	3.759	0.01	−5.066	0.01	13.6138	0.01
ALL	1993/06/03	−1.8231	0.6533	1.7758	0.01	−7.9799	0.01	10.9147	0.01
ALTR	1988/04/04	−3.1749	0.09217	4.6156	0.01	−7.2628	0.01	4.7353	0.01
AMAT	1984/09/07	−3.5872	0.03383	3.6274	0.01	−8.0035	0.01	14.6828	0.01
AMD	1983/03/21	−3.8622	0.01583	2.0913	0.01	−7.8068	0.01	23.5018	0.01
AMGN	1984/09/07	−4.4227	0.01	8.1773	0.01	−10.7281	0.01	1.2416	0.01
AMP	2005/10/03	−1.7914	0.6666	4.4397	0.01	−6.6324	0.01	6.3784	0.01
AMT	1998/06/05	−1.5579	0.7655	5.3705	0.01	−10.4172	0.01	11.8422	0.01
AMZN	1997/05/16	−3.3256	0.06623	2.7737	0.01	−6.6848	0.01	2.5631	0.01
AN	1992/03/03	−2.5511	0.3450	2.3972	0.01	−8.2073	0.01	6.914	0.01
ANF	1996/09/26	−1.6172	0.7404	6.2718	0.01	−9.5964	0.01	10.8549	0.01
AOC	1984/09/07	−2.6817	0.2896	3.1959	0.01	−9.5982	0.01	19.8387	0.01
APA	1981/12/31	−3.7196	0.02296	28.2564	0.01	−7.3313	0.01	27.6244	0.01
APC	1986/09/09	−3.6292	0.02979	15.2225	0.01	−8.1591	0.01	23.4856	0.01
APD	1983/04/06	−3.4789	0.04427	4.6475	0.01	−9.215	0.01	13.692	0.01
APH	1991/11/08	−3.4915	0.04307	11.5918	0.01	−10.2921	0.01	9.5402	0.01
APOL	1995/01/04	−3.0649	0.1274	15.3995	0.01	−10.135	0.01	15.7952	0.01
ATI	1999/11/30	−1.005	0.938	10.3369	0.01	−6.671	0.01	14.9206	0.01
AVB	1998/06/05	−0.5207	0.9809	14.5285	0.01	−4.2084	0.01	9.1641	0.01
AVP	1981/12/31	−2.9037	0.1956	15.217	0.01	−13.7795	0.01	0.7298	0.01083
AVY	1984/09/07	−2.2067	0.4908	24.4865	0.01	−9.2972	0.01	16.9445	0.01
AXP	1977/04/01	−2.5654	0.3389	4.5938	0.01	−6.4409	0.01	11.0398	0.01
AYE	1984/07/19	−2.422	0.3996	2.809	0.01	−9.2322	0.01	15.7826	0.01
AZO	1991/04/02	−2.0269	0.567	21.6767	0.01	−10.303	0.01	11.0425	0.01
BA	1962/01/02	−3.7278	0.02250	3.3708	0.01	−11.4676	0.01	26.0655	0.01
BAC	1986/05/29	−2.1601	0.5105	6.4955	0.01	−0.8188	0.96	4.8574	0.01
BAX	1981/10/27	−2.4497	0.3879	12.3025	0.01	−13.4449	0.01	14.3412	0.01
BBBY	1992/06/05	−5.0385	0.01	3.4572	0.01	−9.0439	0.01	13.2333	0.01
BBT	1990/03/26	−2.0561	0.5546	13.5389	0.01	−4.7333	0.01	10.2066	0.01
BBY	1985/04/19	−4.0601	0.01	15.2267	0.01	−10.4538	0.01	14.4203	0.01
BCR	1983/04/06	−3.5489	0.03752	27.146	0.01	−12.3609	0.01	3.2107	0.01
BDK	1981/12/31	−1.5909	0.7516	26.3152	0.01	−11.7101	0.01	20.7958	0.01
BDX	1983/04/06	−3.0137	0.1490	4.8623	0.01	−11.9256	0.01	2.5124	0.01
BEN	1984/09/07	−2.6164	0.3173	17.2831	0.01	−8.5573	0.01	16.0611	0.01
BF-B	1984/09/07	−3.2717	0.07546	4.3622	0.01	−10.4463	0.01	7.2164	0.01
BHI	1987/04/07	−2.6078	0.3210	19.8587	0.01	−9.1727	0.01	24.0566	0.01
BIG	1987/11/05	−2.4299	0.3963	5.2867	0.01	−6.8418	0.01	15.5507	0.01
BIIB	1991/09/17	−3.2099	0.08616	8.2143	0.01	−10.9615	0.01	17.0348	0.01
BJS	1990/07/20	−2.4796	0.3752	4.048	0.01	−10.4897	0.01	24.132	0.01

Table 6.3: Continuation of Table 6.2—Trading symbols of S&P500 constituents with the corresponding date of the earliest closing price (YYYY/MM/DD). Additionally, the values of the ADF tests and KPSS tests for the price time series $p(t)$ and the volume time series $v(t)$ are shown with their corresponding p-values.

Symbol	First record	ADF_p	p-value	$KPSS_p$	p-value	ADF_v	p-value	$KPSS_v$	p-value
BK	1984/09/07	−3.5723	0.03526	5.7965	0.01	−5.6703	0.01	11.2102	0.01
BLL	1984/09/07	−4.3542	0.01	9.7218	0.01	−11.3869	0.01	21.2702	0.01
BMC	1990/03/26	−3.7503	0.02144	12.7477	0.01	−9.8577	0.01	3.1146	0.01
BMS	1984/09/07	−3.1001	0.1124	10.2092	0.01	−7.4516	0.01	19.8755	0.01
BMY	1977/01/03	−2.8775	0.2066	6.1285	0.01	−9.3779	0.01	21.8408	0.01
BNI	1980/01/02	−2.9136	0.1914	1.4801	0.01	−9.1394	0.01	18.8315	0.01
BRCM	1998/04/17	−2.6647	0.2969	8.2622	0.01	−5.1249	0.01	3.8603	0.01
BSX	1992/05/19	−2.1186	0.5281	3.9129	0.01	−9.3846	0.01	17.5302	0.01
BTU	2001/05/22	−2.5916	0.3279	4.4802	0.01	−7.0815	0.01	12.1884	0.01
BXP	1997/06/18	−0.5072	0.9815	16.4786	0.01	−3.9226	0.01284	10.0569	0.01
C	1977/01/03	−3.0198	0.1463	1.6727	0.01	−6.2083	0.01	5.8179	0.01
CA	1984/09/07	−2.8212	0.2305	4.6479	0.01	−11.7315	0.01	7.5642	0.01
CAG	1984/09/07	−3.7407	0.02191	10.8228	0.01	−8.5647	0.01	18.8173	0.01
CAH	1988/01/04	−1.7575	0.681	15.1677	0.01	−11.4826	0.01	21.1124	0.01
CAM	1995/07/05	−3.4163	0.05058	0.6505	0.01805	−10.0746	0.01	14.3512	0.01
CAT	1962/01/02	−4.9207	0.01	7.2646	0.01	−9.417	0.01	30.3897	0.01
CB	1984/09/07	−3.4396	0.04806	5.8679	0.01	−9.9283	0.01	16.3532	0.01
CBE	1982/01/04	−3.1726	0.09255	16.8246	0.01	−11.2335	0.01	10.771	0.01
CBG	2004/06/10	−2.4384	0.3927	5.7298	0.01	−5.749	0.01	8.0863	0.01
CBS	2006/01/03	−1.6224	0.7381	8.9081	0.01	−5.6555	0.01	7.0106	0.01
CCE	1986/11/24	−3.3652	0.05935	4.172	0.01	−8.7011	0.01	14.6832	0.01
CCL	1989/01/05	−2.8078	0.2362	6.5008	0.01	−7.8246	0.01	18.873	0.01
CEG	1984/09/07	−1.9608	0.595	10.9726	0.01	−10.9002	0.01	10.7329	0.01
CELG	1990/03/26	−6.1079	0.01	16.8338	0.01	−8.4752	0.01	17.3251	0.01
CEPH	1991/04/25	−3.6515	0.02767	22.7988	0.01	−10.483	0.01	16.7804	0.01
CF	2005/08/11	−1.3056	0.8723	7.4823	0.01	−4.3318	0.01	7.1503	0.01
CHK	1993/02/16	−2.5585	0.3418	3.8457	0.01	−5.7468	0.01	13.8348	0.01
CHRW	1997/10/16	−3.6061	0.03207	12.3278	0.01	−10.1356	0.01	15.369	0.01
CI	1982/03/31	−2.9328	0.1833	6.7302	0.01	−13.1122	0.01	8.4373	0.01
CIEN	1997/02/07	−2.7404	0.2648	6.173	0.01	−8.1693	0.01	9.3086	0.01
CINF	1990/03/26	−3.8587	0.01602	8.9185	0.01	−7.893	0.01	11.7549	0.01
CIT	2002/07/05	−0.9264	0.9503	4.217	0.01	−5.7846	0.01	8.9123	0.01
CL	1977/01/03	−3.6328	0.02940	29.7954	0.01	−11.7431	0.01	9.8288	0.01
CLX	1983/03/21	−2.9634	0.1703	6.8708	0.01	−11.1321	0.01	14.4636	0.01
CMA	1990/03/26	−1.941	0.6033	6.1984	0.01	−5.6604	0.01	11.6741	0.01
CMCSA	1988/07/07	−2.523	0.3569	8.4123	0.01	−5.8823	0.01	19.4914	0.01
CME	2002/12/24	−0.6119	0.9767	9.9649	0.01	−8.0427	0.01	8.6248	0.01
CMI	1984/12/18	−2.9198	0.1888	2.4558	0.01	−9.6581	0.01	22.6685	0.01
CMS	1984/12/31	−2.0292	0.566	5.1848	0.01	−7.5705	0.01	20.0985	0.01
CNP	1970/01/02	−1.9241	0.6106	21.4908	0.01	−15.2832	0.01	18.9635	0.01
CNX	1999/04/30	−2.7945	0.2420	10.9317	0.01	−6.613	0.01	14.9819	0.01
COF	1994/11/16	−2.6439	0.3057	2.6006	0.01	−6.3123	0.01	9.3072	0.01
COG	1990/02/08	−3.7686	0.02053	18.7423	0.01	−6.3361	0.01	16.8178	0.01
COH	2000/10/06	−3.7656	0.02070	1.9785	0.01	−6.7372	0.01	6.1439	0.01
COL	2001/07/03	−1.0839	0.9252	12.773	0.01	−7.0815	0.01	7.7929	0.01
COP	1982/01/04	−3.286	0.07298	24.9679	0.01	−5.3222	0.01	15.3492	0.01
COST	1986/07/09	−2.6586	0.2994	5.5312	0.01	−11.0011	0.01	18.5104	0.01
CPB	1985/07/01	−3.5418	0.03821	4.0603	0.01	−8.0781	0.01	12.0084	0.01
CPWR	1992/12/16	−3.1948	0.08877	16.1519	0.01	−10.5801	0.01	1.272	0.01
CRM	2004/06/23	−1.5926	0.7508	8.3617	0.01	−8.9589	0.01	4.994	0.01
CSC	1981/12/31	−3.2401	0.08091	8.2218	0.01	−12.6106	0.01	20.6563	0.01
CSCO	1990/03/26	−3.7711	0.0204	8.9465	0.01	−9.9227	0.01	4.2235	0.01
CSX	1980/11/03	−2.9271	0.1857	2.3918	0.01	−5.5032	0.01	15.7262	0.01
CTAS	1990/03/26	−3.1886	0.08982	3.3018	0.01	−10.4156	0.01	17.1743	0.01
CTL	1987/11/05	−3.8785	0.01503	7.0824	0.01	−8.7379	0.01	19.0937	0.01
CTSH	1998/06/19	−3.147	0.09703	1.2374	0.01	−7.264	0.01	11.197	0.01
CTX	1985/07/11	−2.4937	0.3692	5.8428	0.01	−5.0167	0.01	17.4069	0.01
CTXS	1995/12/08	−3.6585	0.02702	6.4047	0.01	−10.4816	0.01	2.0917	0.01
CVG	1998/08/13	−2.6752	0.2925	7.1798	0.01	−8.8003	0.01	3.6675	0.01
CVH	1991/04/17	−1.8535	0.6404	11.7775	0.01	−11.4121	0.01	5.8828	0.01

Table 6.4: Continuation of Table 6.3—Trading symbols of S&P500 constituents with the corresponding date of the earliest closing price (YYYY/MM/DD). Additionally, the values of the ADF tests and KPSS tests for the price time series $p(t)$ and the volume time series $v(t)$ are shown with their corresponding p-values.

Symbol	First record	ADF_p	p-value	$KPSS_p$	p-value	ADF_v	p-value	$KPSS_v$	p-value
CVS	1984/12/17	−3.6866	0.02461	1.1429	0.01	−7.3937	0.01	19.0126	0.01
CVX	1970/01/02	−3.7038	0.02372	17.5702	0.01	−5.6901	0.01	24.2023	0.01
D	1984/10/03	−1.5086	0.7865	25.5608	0.01	−10.1913	0.01	20.1654	0.01
DD	1962/01/02	−3.06	0.1292	30.8953	0.01	−8.9736	0.01	30.1054	0.01
DE	1982/01/04	−3.0263	0.1436	16.4771	0.01	−10.4908	0.01	19.6658	0.01
DELL	1988/08/17	−3.1551	0.0956	4.3244	0.01	−6.6666	0.01	8.8295	0.01
DF	2002/01/09	−3.6592	0.02698	7.534	0.01	−8.5014	0.01	8.8123	0.01
DFS	2007/06/14	−3.4085	0.05197	7.1282	0.01	−5.963	0.01	2.8034	0.01
DGX	1996/12/26	−2.047	0.5584	5.5873	0.01	−7.6873	0.01	4.136	0.01
DHI	1992/06/05	−2.2386	0.4773	9.3171	0.01	−6.2941	0.01	17.1853	0.01
DHR	1987/11/05	−4.3769	0.01	24.7644	0.01	−8.9777	0.01	19.4931	0.01
DIS	1962/01/02	−4.1232	0.01	6.6899	0.01	−12.3307	0.01	14.2851	0.01
DNB	1983/06/10	−1.8041	0.6613	25.5467	0.01	−10.0798	0.01	1.9334	0.01
DNR	1996/01/03	−2.1485	0.5155	15.7566	0.01	−8.667	0.01	17.6784	0.01
DO	1995/10/11	−1.7995	0.6632	10.0657	0.01	−7.4373	0.01	14.0761	0.01
DOV	1985/07/01	−3.9211	0.01289	2.5087	0.01	−8.7332	0.01	17.2147	0.01
DOW	1977/01/03	−2.0688	0.5492	5.9632	0.01	−5.1601	0.01	11.4543	0.01
DPS	2008/05/07	−1.885	0.6246	2.9785	0.01	−5.2289	0.01	0.827	0.01
DRI	1995/05/17	−3.1373	0.09868	17.8669	0.01	−7.3442	0.01	15.5552	0.01
DTE	1970/01/02	−2.1033	0.5346	38.6776	0.01	−15.9487	0.01	22.7195	0.01
DTV	2003/12/23	−2.3609	0.4255	11.457	0.01	−8.229	0.01	9.4469	0.01
DUK	1983/04/06	−2.5559	0.3429	6.4489	0.01	−7.1182	0.01	21.376	0.01
DV	1991/06/28	−2.8789	0.2062	4.0846	0.01	−10.3215	0.01	13.2355	0.01
DVA	1995/10/31	−1.9289	0.6084	13.1169	0.01	−10.0123	0.01	1.745	0.01
DVN	1992/03/17	−3.6544	0.02739	21.155	0.01	−7.8089	0.01	22.5147	0.01
DYN	1993/11/10	−1.6892	0.71	4.6479	0.01	−8.193	0.01	13.8888	0.01
EBAY	1998/09/24	−4.0266	0.01	10.2083	0.01	−9.3634	0.01	6.3924	0.01
ECL	1988/01/05	−4.0727	0.01	15.5771	0.01	−7.2538	0.01	14.5338	0.01
ED	1970/01/02	−3.4199	0.04992	22.4928	0.01	−12.9184	0.01	17.223	0.01
EFX	1987/11/05	−3.3709	0.05838	15.5032	0.01	−9.6053	0.01	14.5147	0.01
EIX	1980/01/02	−2.1825	0.501	2.7198	0.01	−15.614	0.01	7.7212	0.01
EK	1962/01/02	−4.1943	0.01	29.6269	0.01	−12.0573	0.01	29.86	0.01
EL	1995/11/17	−3.1511	0.0963	4.4226	0.01	−11.739	0.01	12.8198	0.01
EMC	1988/12/16	−2.2347	0.479	3.7076	0.01	−8.1678	0.01	14.5027	0.01
EMN	1993/12/14	−2.6819	0.2896	2.4547	0.01	−10.1868	0.01	16.3375	0.01
EMR	1982/01/04	−3.3924	0.05465	9.3106	0.01	−7.8927	0.01	16.6504	0.01
EOG	1989/10/04	−2.8157	0.2329	18.346	0.01	−10.0976	0.01	22.9826	0.01
EP	1992/03/13	−2.026	0.5673	8.8826	0.01	−8.5295	0.01	16.916	0.01
EQ	2006/05/18	−1.9574	0.5963	5.175	0.01	−6.2531	0.01	1.0156	0.01
EQR	1993/08/12	−1.8825	0.6281	4.9283	0.01	−4.3365	0.01	12.8789	0.01
EQT	1987/11/05	−3.7714	0.02038	15.0002	0.01	−7.8475	0.01	18.4262	0.01
ERTS	1990/03/26	−3.5254	0.03981	11.665	0.01	−7.6922	0.01	12.2741	0.01
ESRX	1992/06/09	−4.5456	0.01	11.9304	0.01	−7.9609	0.01	11.4699	0.01
ESV	1991/01/02	−2.5076	0.3634	16.8336	0.01	−10.8406	0.01	23.9642	0.01
ETFC	1996/08/16	−5.0775	0.01	4.3471	0.01	−5.6394	0.01	4.3906	0.01
ETN	1985/07/01	−4.211	0.01	2.1801	0.01	−8.2274	0.01	17.9285	0.01
ETR	1981/12/31	−2.0578	0.5539	27.3713	0.01	−11.3261	0.01	20.1229	0.01
EXC	1980/01/02	−3.3153	0.06793	29.5115	0.01	−14.5337	0.01	20.8342	0.01
EXPD	1990/03/26	−3.9443	0.01174	12.6631	0.01	−11.233	0.01	19.0323	0.01
EXPE	2005/07/21	−1.6317	0.7342	2.589	0.01	−9.2536	0.01	4.3561	0.01
F	1977/01/03	−2.668	0.2954	14.0529	0.01	−7.1946	0.01	12.6768	0.01
FAST	1990/03/26	−3.8804	0.01493	10.8063	0.01	−11.5156	0.01	8.9469	0.01
FCX	1995/07/10	−2.1593	0.5109	13.109	0.01	−4.1307	0.01	14.1092	0.01
FDO	1987/11/05	−3.8257	0.01767	18.6506	0.01	−7.5079	0.01	16.5063	0.01
FDX	1980/01/02	−3.5956	0.033	10.1274	0.01	−8.8877	0.01	13.8763	0.01
FE	1997/11/10	−1.524	0.7799	17.3031	0.01	−8.5841	0.01	11.2424	0.01
FHN	1990/03/26	−2.2755	0.4617	2.14	0.01	−5.3992	0.01	9.0436	0.01
FII	1998/05/14	−1.9839	0.5851	9.239	0.01	−7.9858	0.01	12.4012	0.01
FIS	2001/09/04	−1.6011	0.7472	1.9402	0.01	−6.9185	0.01	10.6131	0.01
FISV	1990/03/26	−3.8035	0.01878	11.9561	0.01	−11.3877	0.01	20.0457	0.01

Table 6.5: Continuation of Table 6.4—Trading symbols of S&P500 constituents with the corresponding date of the earliest closing price (YYYY/MM/DD). Additionally, the values of the ADF tests and KPSS tests for the price time series $p(t)$ and the volume time series $v(t)$ are shown with their corresponding p-values.

Symbol	First record	ADF_p	p-value	$KPSS_p$	p-value	ADF_v	p-value	$KPSS_v$	p-value
FITB	1990/03/26	−2.4037	0.4074	6.6045	0.01	−3.1391	0.09836	6.4778	0.01
FLIR	1993/06/22	−2.6519	0.3023	12.9573	0.01	−10.1648	0.01	15.5896	0.01
FLR	2000/12/22	−1.9464	0.601	9.3768	0.01	−7.3045	0.01	11.9164	0.01
FLS	1990/03/26	−2.3422	0.4334	12.578	0.01	−5.8094	0.01	14.1579	0.01
FO	1977/01/03	−1.8485	0.6425	26.655	0.01	−9.3322	0.01	15.4046	0.01
FPL	1983/06/10	−3.8188	0.018	23.4448	0.01	−12.2743	0.01	21.6648	0.01
FRX	1988/04/19	−3.0212	0.1459	5.3859	0.01	−11.6135	0.01	14.4796	0.01
FTI	2001/06/18	−1.5293	0.7776	11.7515	0.01	−9.1686	0.01	10.0788	0.01
FTR	1990/03/26	−2.5344	0.3520	12.0491	0.01	−9.4531	0.01	20.755	0.01
GAS	1984/10/22	−4.2268	0.01	23.5768	0.01	−8.112	0.01	14.7493	0.01
GCI	1985/07/01	−0.6119	0.9768	9.4286	0.01	−3.6529	0.02751	10.4806	0.01
GD	1977/01/03	−3.5577	0.03664	30.0855	0.01	−8.1814	0.01	8.5472	0.01
GE	1962/01/02	−4.0484	0.01	8.1871	0.01	−8.6344	0.01	14.451	0.01
GENZ	1990/03/26	−3.4819	0.04399	11.7835	0.01	−9.0338	0.01	15.7234	0.01
GILD	1992/01/22	−3.5776	0.03479	14.1921	0.01	−6.6862	0.01	16.5193	0.01
GIS	1983/06/10	−3.6449	0.02827	7.2301	0.01	−9.371	0.01	14.6628	0.01
GLW	1981/12/31	−4.0637	0.01	4.1013	0.01	−9.5412	0.01	23.5589	0.01
GME	2002/02/14	−1.3462	0.8551	11.2625	0.01	−8.7376	0.01	12.5751	0.01
GNW	2004/05/25	−1.65	0.7265	6.9698	0.01	−6.129	0.01	6.6902	0.01
GOOG	2004/08/19	−1.7818	0.6707	6.5972	0.01	−6.4788	0.01	4.644	0.01
GPC	1983/04/06	−3.4405	0.04797	17.4406	0.01	−6.6762	0.01	13.2229	0.01
GPS	1987/07/23	−4.0168	0.01	13.7598	0.01	−10.0016	0.01	7.4173	0.01
GR	1985/07/01	−3.6194	0.03073	6.1093	0.01	−8.5304	0.01	19.3925	0.01
GS	1999/05/04	−1.9747	0.589	11.7586	0.01	−6.2115	0.01	10.0937	0.01
GT	1970/01/02	−2.1683	0.5071	5.3317	0.01	−8.6689	0.01	21.5688	0.01
GWW	1984/12/17	−3.4283	0.04915	7.5765	0.01	−8.5804	0.01	17.3474	0.01
HAL	1981/12/31	−3.1514	0.0962	7.7457	0.01	−7.8471	0.01	23.1429	0.01
HAR	1988/10/05	−1.9734	0.5896	17.5597	0.01	−12.9448	0.01	19.3785	0.01
HAS	1984/12/18	−4.2103	0.01	3.0633	0.01	−11.079	0.01	12.2336	0.01
HBAN	1990/03/26	−0.979	0.9422	3.2322	0.01	1.8705	0.99	6.8347	0.01
HCBK	1999/07/13	−2.2418	0.4759	3.8053	0.01	−7.4433	0.01	10.7184	0.01
HCN	1992/03/17	−2.6496	0.3033	18.9682	0.01	−3.9348	0.01222	9.6634	0.01
HCP	1987/11/05	−3.8056	0.01867	0.8512	0.01	−3.8096	0.01847	8.6866	0.01
HD	1984/08/20	−3.2611	0.0773	5.8882	0.01	−7.9329	0.01	12.8937	0.01
HES	1983/04/06	−4.1954	0.01	21.8442	0.01	−6.8309	0.01	21.9455	0.01
HIG	1995/12/18	−1.4974	0.7912	1.1683	0.01	−4.1728	0.01	5.1514	0.01
HNZ	1984/12/17	−3.5831	0.03423	13.5621	0.01	−8.2663	0.01	17.2014	0.01
HOG	1987/11/05	−3.4006	0.05325	8.8786	0.01	−7.9137	0.01	13.9195	0.01
HON	1970/01/02	−4.4417	0.01	3.8386	0.01	−13.4509	0.01	23.5794	0.01
HOT	1987/11/05	−1.5287	0.778	20.441	0.01	−7.0637	0.01	19.8114	0.01
HPQ	1962/01/02	−3.8706	0.01536	3.1366	0.01	−12.3925	0.01	34.2061	0.01
HRB	1986/11/12	−3.8084	0.01853	3.2735	0.01	−11.5333	0.01	13.8633	0.01
HRL	1990/01/02	−3.8644	0.01574	15.1089	0.01	−12.0685	0.01	14.4387	0.01
HRS	1981/12/31	−3.7221	0.02283	19.0976	0.01	−12.8725	0.01	15.3025	0.01
HSP	2004/05/03	−2.2207	0.4849	2.2041	0.01	−8.6852	0.01	1.2025	0.01
HST	1983/04/06	−2.6685	0.2952	12.0263	0.01	−4.662	0.01	10.3295	0.01
HSY	1985/07/01	−3.5147	0.04082	9.4283	0.01	−10.2163	0.01	11.8352	0.01
HUM	1981/12/31	−2.4451	0.3898	3.8248	0.01	−11.9698	0.01	14.6865	0.01
IBM	1962/01/02	−3.5555	0.03680	31.6581	0.01	−9.6225	0.01	33.0781	0.01
ICE	2005/11/16	−1.73	0.6926	3.3346	0.01	−4.8048	0.01	0.8679	0.01
IFF	1981/12/31	−2.3334	0.4371	4.5552	0.01	−10.3394	0.01	16.2625	0.01
IGT	1990/03/26	−2.9742	0.1658	3.2999	0.01	−10.6173	0.01	0.927	0.01
INTC	1986/07/09	−2.8629	0.2129	6.715	0.01	−11.8965	0.01	1.0996	0.01
INTU	1993/03/22	−4.5193	0.01	3.7203	0.01	−8.7407	0.01	2.9796	0.01
IP	1970/01/02	−2.7013	0.2812	16.6901	0.01	−5.7043	0.01	23.5906	0.01
IPG	1987/11/05	−2.7304	0.269	13.566	0.01	−8.6428	0.01	20.6742	0.01
IR	1985/07/01	−3.4751	0.04464	4.8672	0.01	−6.9318	0.01	19.4436	0.01
IRM	1996/02/01	−3.8953	0.01421	2.0012	0.01	−10.7992	0.01	7.5105	0.01
ISRG	2000/06/23	−1.9037	0.6191	14.8385	0.01	−9.2311	0.01	12.3955	0.01
ITT	1995/12/18	−1.5568	0.766	13.3514	0.01	−10.6671	0.01	9.6982	0.01

Table 6.6: Continuation of Table 6.5—Trading symbols of S&P500 constituents with the corresponding date of the earliest closing price (YYYY/MM/DD). Additionally, the values of the ADF tests and KPSS tests for the price time series $p(t)$ and the volume time series $v(t)$ are shown with their corresponding p-values.

Symbol	First record	ADF_p	p-value	$KPSS_p$	p-value	ADF_v	p-value	$KPSS_v$	p-value
ITW	1987/11/05	−2.8334	0.2254	8.8891	0.01	−9.1156	0.01	22.3562	0.01
IVZ	1995/08/25	−3.2819	0.07376	11.3688	0.01	−5.518	0.01	8.8862	0.01
JAVA	1987/03/11	−2.9871	0.1603	6.0917	0.01	−11.6903	0.01	1.1285	0.01
JBL	1993/05/03	−3.1006	0.1123	3.2337	0.01	−8.7652	0.01	10.6325	0.01
JCI	1985/03/27	−2.5256	0.3557	14.722	0.01	−8.2237	0.01	16.5115	0.01
JCP	1982/01/04	−2.7575	0.2575	2.6247	0.01	−5.363	0.01	18.9171	0.01
JDSU	1993/11/17	−2.5412	0.3492	5.7402	0.01	−10.5045	0.01	19.2696	0.01
JEC	1990/01/12	−3.7286	0.02252	18.4192	0.01	−6.6099	0.01	16.0963	0.01
JNJ	1970/01/02	−5.094	0.01	6.0484	0.01	−10.3656	0.01	19.016	0.01
JNPR	1999/06/25	−2.3792	0.4178	8.6518	0.01	−6.1404	0.01	1.247	0.01
JNS	2000/06/26	−2.7692	0.2527	3.7939	0.01	−7.9925	0.01	9.3913	0.01
JPM	1983/12/30	−2.3402	0.4342	4.0545	0.01	−4.3722	0.01	11.9063	0.01
JWN	1986/07/09	−3.658	0.02702	1.1786	0.01	−8.4734	0.01	6.7696	0.01
K	1984/12/17	−3.311	0.0687	11.1839	0.01	−8.1239	0.01	15.152	0.01
KBH	1987/11/05	−1.9251	0.61	14.4917	0.01	−6.7402	0.01	20.7943	0.01
KEY	1987/11/05	−1.572	0.7596	3.4431	0.01	1.3742	0.99	7.6231	0.01
KFT	2001/06/14	−3.3865	0.05576	4.6179	0.01	−6.1119	0.01	8.5616	0.01
KG	1998/06/25	−2.9095	0.1932	11.1767	0.01	−8.8607	0.01	10.3463	0.01
KIM	1991/11/22	−1.8866	0.6264	10.5548	0.01	−4.2409	0.01	7.5838	0.01
KLAC	1990/03/26	−2.9267	0.1859	7.8489	0.01	−4.7388	0.01	12.0209	0.01
KMB	1984/12/17	−4.3849	0.01	5.3457	0.01	−9.8674	0.01	17.9709	0.01
KO	1962/01/02	−3.9797	0.01	18.0126	0.01	−8.3735	0.01	24.5632	0.01
KR	1977/01/03	−3.7922	0.01931	2.8825	0.01	−8.63	0.01	24.1289	0.01
KSS	1992/05/19	−4.2622	0.01	3.9295	0.01	−7.125	0.01	20.3655	0.01
L	1987/07/10	−3.7195	0.02298	13.3838	0.01	−8.8311	0.01	13.1909	0.01
LEG	1987/11/05	−4.0338	0.01	10.3021	0.01	−6.3673	0.01	13.8576	0.01
LEN	1987/11/05	−2.4183	0.4012	8.9729	0.01	−6.6282	0.01	18.0083	0.01
LH	1990/03/29	−2.296	0.453	11.556	0.01	−10.0168	0.01	11.8647	0.01
LIFE	1999/02/26	−2.8397	0.2228	2.3256	0.01	−9.2339	0.01	1.9573	0.01
LLL	1998/05/19	−2.8438	0.2211	8.9914	0.01	−6.5572	0.01	6.2795	0.01
LLTC	1990/03/26	−3.6047	0.03217	4.2802	0.01	−9.1508	0.01	19.279	0.01
LLY	1982/01/04	−5.0234	0.01	5.1521	0.01	−11.6578	0.01	15.2323	0.01
LM	1987/11/05	−1.3978	0.8334	19.2096	0.01	−5.8196	0.01	15.9655	0.01
LMT	1977/01/03	−3.308	0.06917	10.6966	0.01	−10.2475	0.01	16.7925	0.01
LNC	1984/10/05	−2.4403	0.3919	7.0652	0.01	−2.1523	0.5139	7.6477	0.01
LO	2008/06/10	−1.9103	0.6137	2.1431	0.01	−5.0074	0.01	0.6149	0.02128
LOW	1985/07/01	−3.8017	0.01886	8.3406	0.01	−9.6982	0.01	17.8009	0.01
LSI	1990/01/12	−2.3571	0.4271	5.6043	0.01	−8.1512	0.01	14.8025	0.01
LTD	1985/07/01	−3.8066	0.01861	1.0344	0.01	−8.3599	0.01	16.9497	0.01
LUK	1987/11/05	−3.1756	0.09206	16.0281	0.01	−6.8184	0.01	13.121	0.01
LUV	1980/01/02	−4.661	0.01	14.1496	0.01	−9.2247	0.01	9.8029	0.01
LXK	1995/11/15	−2.4753	0.3771	2.8826	0.01	−12.0755	0.01	10.7524	0.01
M	1992/02/05	−1.5838	0.7546	6.2261	0.01	−6.9529	0.01	16.263	0.01
MA	2006/05/25	−1.2633	0.8902	6.0353	0.01	−7.436	0.01	0.6001	0.02262
MAR	1993/10/13	−1.5949	0.7499	11.9713	0.01	−8.7058	0.01	15.7571	0.01
MAS	1983/06/10	−3.2817	0.07374	2.671	0.01	−6.5568	0.01	20.6405	0.01
MAT	1982/01/04	−2.8327	0.2256	8.9577	0.01	−9.455	0.01	19.2545	0.01
MBI	1987/07/02	−1.8317	0.6496	5.5977	0.01	−5.1926	0.01	8.7663	0.01
MCD	1970/01/02	−4.4043	0.01	9.3399	0.01	−12.0879	0.01	13.2897	0.01
MCHP	1993/03/19	−4.1585	0.01	1.7114	0.01	−8.8596	0.01	10.9318	0.01
MCK	1994/11/15	−2.5393	0.35	2.6417	0.01	−9.5888	0.01	10.6985	0.01
MCO	2000/10/04	−1.6862	0.7112	3.8881	0.01	−7.3757	0.01	10.7021	0.01
MDP	1985/07/01	−3.4529	0.04678	2.3061	0.01	−6.8724	0.01	11.1821	0.01
MDT	1981/12/31	−3.9523	0.01132	4.8226	0.01	−11.1683	0.01	7.7483	0.01
MEE	1981/12/31	−2.6083	0.3207	5.1434	0.01	−3.8459	0.01664	12.6882	0.01
MET	2000/06/23	−0.9138	0.9515	11.4083	0.01	−5.1916	0.01	7.0052	0.01
MFE	1992/10/07	−2.4079	0.4056	2.3509	0.01	−8.6549	0.01	4.1464	0.01
MHP	1985/07/01	−3.734	0.02224	1.8703	0.01	−7.9415	0.01	6.9852	0.01
MHS	2003/08/21	−1.9369	0.605	5.0876	0.01	−7.6084	0.01	0.7094	0.01269
MI	1990/03/26	−2.0443	0.5596	1.7809	0.01	−1.7802	0.6714	9.8333	0.01

Table 6.7: Continuation of Table 6.6—Trading symbols of S&P500 constituents with the corresponding date of the earliest closing price (YYYY/MM/DD). Additionally, the values of the ADF tests and KPSS tests for the price time series $p(t)$ and the volume time series $v(t)$ are shown with their corresponding p-values.

Symbol	First record	ADF_p	p-value	$KPSS_p$	p-value	ADF_v	p-value	$KPSS_v$	p-value
MIL	1988/10/05	−3.6362	0.02912	17.5843	0.01	−11.0998	0.01	17.0494	0.01
MKC	1990/03/26	−3.5188	0.04044	7.8963	0.01	−11.1928	0.01	2.1619	0.01
MMC	1987/12/30	−2.7703	0.2521	12.8808	0.01	−12.2755	0.01	14.2464	0.01
MMM	1970/01/02	−3.6885	0.02448	4.0137	0.01	−10.1923	0.01	25.8067	0.01
MO	1970/01/02	−4.2583	0.01	2.8926	0.01	−10.8454	0.01	16.3543	0.01
MOLX	1990/03/26	−3.6244	0.03027	4.5292	0.01	−9.9566	0.01	14.4927	0.01
MON	2000/10/24	−1.9033	0.6192	14.7793	0.01	−7.6109	0.01	9.553	0.01
MOT	1977/01/03	−3.6781	0.02504	9.4877	0.01	−11.4699	0.01	24.9579	0.01
MRK	1970/01/02	−3.5379	0.03853	7.1548	0.01	−9.1393	0.01	20.2582	0.01
MRO	1970/01/02	−3.1674	0.09339	22.7585	0.01	−7.8058	0.01	24.5249	0.01
MS	1993/02/23	−3.0825	0.1200	2.5067	0.01	−4.8162	0.01	8.0655	0.01
MSFT	1986/03/13	−3.6812	0.02488	11.4289	0.01	−12.218	0.01	0.1171	0.1
MTB	1991/10/04	−2.0918	0.5395	6.7672	0.01	−7.0194	0.01	11.4305	0.01
MTW	1990/03/26	−2.1332	0.522	5.3028	0.01	−5.2232	0.01	13.4821	0.01
MU	1989/05/16	−3.6498	0.02782	5.6707	0.01	−8.1003	0.01	13.344	0.01
MUR	1983/04/06	−3.9211	0.01288	27.0771	0.01	−6.4108	0.01	23.4427	0.01
MWV	1985/07/01	−4.2631	0.01	11.0627	0.01	−6.7693	0.01	21.8876	0.01
MWW	1996/12/13	−2.8094	0.2356	3.0781	0.01	−8.0965	0.01	9.2631	0.01
MYL	1987/12/18	−3.1855	0.09035	2.7616	0.01	−8.6646	0.01	9.8043	0.01
NBL	1982/01/04	−4.3508	0.01	26.9951	0.01	−7.4338	0.01	22.9801	0.01
NBR	1991/02/28	−1.7394	0.6887	15.8759	0.01	−7.9528	0.01	22.9483	0.01
NDAQ	2002/07/02	−1.1132	0.9205	11.4891	0.01	−9.2207	0.01	14.1836	0.01
NEM	1983/04/06	−3.3973	0.05381	6.1803	0.01	−6.8308	0.01	25.3942	0.01
NI	1984/10/19	−1.5394	0.7734	12.9055	0.01	−6.5501	0.01	14.4894	0.01
NKE	1987/08/19	−3.9421	0.01185	6.0037	0.01	−9.3688	0.01	3.853	0.01
NOC	1981/12/31	−2.3251	0.4407	19.5598	0.01	−7.5025	0.01	19.0565	0.01
NOV	1996/10/29	−2.4611	0.3831	8.6134	0.01	−6.6222	0.01	16.7886	0.01
NOVL	1988/08/17	−3.3041	0.0699	19.5356	0.01	−10.625	0.01	4.0536	0.01
NSC	1982/06/02	−2.5912	0.3280	1.9276	0.01	−6.1272	0.01	14.6773	0.01
NSM	1982/01/04	−3.7918	0.01934	5.8291	0.01	−9.9791	0.01	25.3173	0.01
NTAP	1995/11/21	−4.3485	0.01	4.5046	0.01	−6.361	0.01	6.1612	0.01
NTRS	1990/03/26	−3.1732	0.09248	6.4877	0.01	−8.4428	0.01	11.9342	0.01
NU	1984/08/29	−2.8763	0.2072	20.3592	0.01	−11.0395	0.01	11.8512	0.01
NUE	1983/09/01	−4.6673	0.01	2.4915	0.01	−13.002	0.01	19.5985	0.01
NVDA	1999/01/22	−3.2922	0.072	5.0811	0.01	−6.1887	0.01	1.6759	0.01
NVLS	1990/03/26	−3.8194	0.01799	4.3331	0.01	−6.2237	0.01	7.7469	0.01
NWL	1984/07/19	−3.0145	0.1487	2.9287	0.01	−7.8501	0.01	17.1572	0.01
NWSA	1996/03/11	−2.3461	0.4318	4.7694	0.01	−8.2518	0.01	11.4681	0.01
NYT	1986/07/16	−2.5475	0.3465	4.6984	0.01	−9.1172	0.01	17.8239	0.01
NYX	2004/08/12	−1.122	0.919	3.456	0.01	−6.4175	0.01	8.6507	0.01
ODP	1990/03/26	−2.6825	0.2893	4.5467	0.01	−8.6555	0.01	10.4412	0.01
OI	1991/12/11	−2.1653	0.5084	4.4853	0.01	−8.9155	0.01	14.3712	0.01
OMC	1990/03/26	−2.0184	0.5706	12.8211	0.01	−8.2652	0.01	11.4931	0.01
ORCL	1988/03/02	−3.7655	0.02068	4.093	0.01	−13.1183	0.01	2.4497	0.01
ORLY	1993/04/23	−3.7252	0.02271	1.6054	0.01	−8.7455	0.01	9.3011	0.01
OXY	1981/12/31	−2.6137	0.3184	15.1109	0.01	−7.9006	0.01	20.9872	0.01
PAYX	1990/03/26	−3.7827	0.01982	3.0901	0.01	−11.2571	0.01	9.1867	0.01
PBCT	1990/03/26	−1.8901	0.6249	15.6506	0.01	−7.741	0.01	9.2388	0.01
PBG	1999/03/31	−2.9303	0.1845	1.3342	0.01	−9.0376	0.01	3.8308	0.01
PBI	1982/01/04	−4.0341	0.01	1.8247	0.01	−9.1291	0.01	5.3073	0.01
PCAR	1986/07/09	−3.7543	0.02123	2.5496	0.01	−9.229	0.01	19.2857	0.01
PCG	1980/01/02	−2.7822	0.247	28.5399	0.01	−13.4356	0.01	17.671	0.01
PCL	1989/06/02	−3.8561	0.01615	19.8454	0.01	−4.2469	0.01	11.1873	0.01
PCP	1990/01/12	−2.7079	0.2785	13.4936	0.01	−9.1018	0.01	20.623	0.01
PDCO	1992/10/28	−2.5865	0.3300	6.5572	0.01	−11.12	0.01	16.2254	0.01
PEG	1980/01/02	−3.0467	0.135	28.9051	0.01	−12.7593	0.01	18.8575	0.01
PEP	1977/01/03	−5.2267	0.01	9.0682	0.01	−10.0552	0.01	5.7096	0.01
PFE	1982/01/04	−3.2746	0.07495	10.1358	0.01	−6.9907	0.01	15.6321	0.01
PFG	2001/10/23	−0.6575	0.9743	5.4511	0.01	−4.3522	0.01	4.5404	0.01
PG	1970/01/02	−4.8445	0.01	1.9669	0.01	−10.1927	0.01	23.4255	0.01

Table 6.8: Continuation of Table 6.7—Trading symbols of S&P500 constituents with the corresponding date of the earliest closing price (YYYY/MM/DD). Additionally, the values of the ADF tests and KPSS tests for the price time series $p(t)$ and the volume time series $v(t)$ are shown with their corresponding p-values.

Symbol	First record	ADF_p	p-value	$KPSS_p$	p-value	ADF_v	p-value	$KPSS_v$	p-value
PGN	1984/10/01	−3.2784	0.07432	5.7957	0.01	−12.0833	0.01	12.7995	0.01
PGR	1986/07/09	−2.2977	0.4523	6.4949	0.01	−8.0819	0.01	20.4889	0.01
PH	1985/07/01	−3.3375	0.06412	23.8014	0.01	−9.8424	0.01	22.1474	0.01
PHM	1985/07/01	−2.6772	0.2916	12.6235	0.01	−5.4372	0.01	18.2744	0.01
PKI	1983/04/06	−2.542	0.3488	0.8199	0.01	−9.7955	0.01	21.341	0.01
PLD	1995/08/18	−0.2129	0.99	12.2362	0.01	−3.0763	0.1226	7.2168	0.01
PLL	1991/08/21	−3.7486	0.02153	13.1201	0.01	−11.8549	0.01	15.3469	0.01
PM	2008/03/17	−1.6848	0.7094	4.7875	0.01	−4.739	0.01	0.7466	0.01
PNC	1988/09/07	−2.2558	0.47	22.4559	0.01	−4.5977	0.01	9.4542	0.01
PNW	1984/07/19	−1.9407	0.6035	26.4087	0.01	−12.5607	0.01	11.5071	0.01
POM	1987/01/02	−6.323	0.01	1.2501	0.01	−6.3044	0.01	13.4343	0.01
PPG	1983/04/06	−3.7535	0.02127	20.7787	0.01	−6.9107	0.01	14.9081	0.01
PPL	1985/04/08	−2.8344	0.2250	4.2529	0.01	−11.6645	0.01	19.1432	0.01
PRU	2001/12/14	−0.7953	0.962	7.618	0.01	−4.5386	0.01	5.7395	0.01
PSA	1987/12/30	−2.4064	0.4062	27.0127	0.01	−4.63	0.01	11.744	0.01
PTV	1999/10/27	−2.1865	0.4993	12.6465	0.01	−7.38	0.01	10.4949	0.01
PX	1992/06/17	−3.1741	0.09233	18.7913	0.01	−9.2383	0.01	8.364	0.01
PXD	1997/08/08	−2.6471	0.3043	13.1582	0.01	−7.1598	0.01	15.2744	0.01
Q	1997/06/24	−3.0958	0.1144	15.6232	0.01	−8.182	0.01	11.4704	0.01
QCOM	1991/12/16	−4.3481	0.01	1.676	0.01	−7.725	0.01	3.7066	0.01
QLGC	1994/02/28	−3.16	0.09477	3.988	0.01	−3.7723	0.02035	4.4095	0.01
R	1980/01/02	−2.7538	0.2591	16.7461	0.01	−8.1982	0.01	19.9735	0.01
RAI	1999/06/15	−1.6413	0.7302	6.7413	0.01	−8.9577	0.01	1.392	0.01
RDC	1983/04/06	−3.1201	0.1039	23.1225	0.01	−6.984	0.01	25.3611	0.01
RF	1990/03/26	−1.8159	0.6563	3.6125	0.01	−6.4834	0.01	5.185	0.01
RHI	1992/03/10	−3.8206	0.01793	2.0192	0.01	−10.7677	0.01	18.9132	0.01
RL	1997/06/12	−2.2347	0.479	16.2726	0.01	−7.3776	0.01	12.2742	0.01
ROK	1981/12/31	−2.6362	0.3089	20.3257	0.01	−9.0044	0.01	19.5361	0.01
RRC	1992/12/28	−1.7353	0.6904	14.927	0.01	−4.8333	0.01	15.8571	0.01
RRD	1985/07/01	−3.2019	0.08751	5.4517	0.01	−6.4401	0.01	15.5925	0.01
RSG	1998/07/01	−2.3412	0.4339	13.5746	0.01	−8.4108	0.01	1.875	0.01
RSH	1982/01/04	−3.8872	0.01457	6.5469	0.01	−8.5709	0.01	7.2832	0.01
RTN	1981/12/31	−3.4321	0.04877	7.974	0.01	−7.9006	0.01	22.8144	0.01
RX	1998/06/23	−3.4468	0.0474	3.2621	0.01	−8.6291	0.01	0.1695	0.1
S	1984/11/08	−2.45	0.3878	3.5355	0.01	−6.323	0.01	15.4374	0.01
SBUX	1992/06/26	−3.1558	0.09548	1.5325	0.01	−11.1399	0.01	1.6778	0.01
SCG	1987/12/30	−2.3552	0.4279	3.1797	0.01	−8.454	0.01	14.8021	0.01
SCHW	1989/06/30	−3.1318	0.09962	4.739	0.01	−8.1395	0.01	11.3952	0.01
SE	2007/01/03	−1.7647	0.6779	6.7287	0.01	−5.7302	0.01	4.144	0.01
SEE	1987/12/30	−3.2866	0.07292	2.9818	0.01	−12.2552	0.01	6.8667	0.01
SGP	1983/04/06	−3.3552	0.06107	15.5055	0.01	−10.2432	0.01	7.6246	0.01
SHLD	2003/05/02	−1.0806	0.9257	4.39	0.01	−6.3615	0.01	1.4493	0.01
SHW	1985/07/01	−3.073	0.1239	14.9475	0.01	−9.3385	0.01	14.8674	0.01
SIAL	1990/03/26	−3.5152	0.04079	4.585	0.01	−9.7355	0.01	5.8902	0.01
SII	1983/04/06	−3.0964	0.1140	21.694	0.01	−8.1843	0.01	25.1641	0.01
SJM	2000/08/29	−2.7416	0.2644	12.6601	0.01	−7.3468	0.01	4.0286	0.01
SLB	1981/12/31	−3.4704	0.04508	11.6049	0.01	−9.466	0.01	25.1542	0.01
SLE	1986/05/05	−3.4566	0.04642	11.5157	0.01	−7.3044	0.01	14.2683	0.01
SLM	1988/01/05	−3.0107	0.1503	3.0155	0.01	−8.5275	0.01	2.9642	0.01
SNA	1985/07/01	−3.9148	0.01320	11.4394	0.01	−9.9165	0.01	17.3611	0.01
SNDK	1995/11/08	−2.7009	0.2816	1.3578	0.01	−9.4496	0.01	16.4103	0.01
SNI	2008/06/12	−0.7642	0.964	4.2635	0.01	−4.9411	0.01	1.2068	0.01
SO	1981/12/31	−3.675	0.02536	30.2985	0.01	−12.6641	0.01	12.2024	0.01
SPG	1993/12/14	−1.3817	0.8402	18.5412	0.01	−3.2787	0.0743	8.7543	0.01
SPLS	1990/03/26	−4.1217	0.01	3.1902	0.01	−10.0391	0.01	3.1609	0.01
SRCL	1996/08/23	−2.9488	0.1766	18.0012	0.01	−8.9002	0.01	7.9477	0.01
SRE	1998/06/29	−2.4614	0.383	18.6866	0.01	−7.8565	0.01	9.8722	0.01
STI	1987/12/30	−0.4552	0.9839	18.5333	0.01	−3.3994	0.05347	8.1247	0.01
STJ	1989/12/07	−3.2962	0.07127	3.68	0.01	−12.0547	0.01	1.2813	0.01
STR	1987/12/30	−2.4232	0.3991	15.0206	0.01	−8.1331	0.01	18.9746	0.01

Table 6.9: Continuation of Table 6.8—Trading symbols of S&P500 constituents with the corresponding date of the earliest closing price (YYYY/MM/DD). Additionally, the values of the ADF tests and KPSS tests for the price time series $p(t)$ and the volume time series $v(t)$ are shown with their corresponding p-values.

Symbol	First record	ADF_p	p-value	$KPSS_p$	p-value	ADF_v	p-value	$KPSS_v$	p-value
STT	1986/07/09	−2.7804	0.2478	11.3983	0.01	−7.5294	0.01	10.1079	0.01
STZ	1992/03/17	−2.5938	0.3269	4.789	0.01	−9.6805	0.01	18.3019	0.01
SUN	1983/04/06	−2.5286	0.3545	6.0589	0.01	−6.2477	0.01	22.7427	0.01
SVU	1985/07/01	−3.2025	0.0874	7.578	0.01	−5.7025	0.01	16.6217	0.01
SWK	1985/07/01	−3.69	0.02445	17.2579	0.01	−10.8076	0.01	21.5227	0.01
SWN	1987/12/30	−2.8218	0.2303	8.3461	0.01	−6.2074	0.01	18.602	0.01
SWY	1990/04/26	−2.2548	0.4704	4.8796	0.01	−11.6405	0.01	19.7997	0.01
SYK	1988/02/01	−3.8522	0.01634	18.66	0.01	−10.2525	0.01	12.285	0.01
SYMC	1990/03/26	−3.5629	0.03619	2.538	0.01	−10.7929	0.01	16.7217	0.01
SYY	1987/07/23	−4.8496	0.01	1.3812	0.01	−10.0418	0.01	13.8436	0.01
T	1984/07/19	−3.4352	0.04848	16.1246	0.01	−4.4268	0.01	19.8989	0.01
TAP	1984/09/07	−2.4757	0.3769	24.8704	0.01	−8.6879	0.01	19.0499	0.01
TDC	2007/10/01	−1.5835	0.7537	5.6542	0.01	−4.9632	0.01	1.279	0.01
TE	1984/10/29	−3.3513	0.06175	17.2217	0.01	−6.8371	0.01	15.9094	0.01
TEG	1988/01/05	−1.253	0.8947	23.7114	0.01	−6.6978	0.01	14.281	0.01
TEL	2007/06/14	−1.4947	0.7922	6.6738	0.01	−4.635	0.01	4.2238	0.01
TER	1987/03/11	−2.8397	0.2227	3.8685	0.01	−11.8058	0.01	23.7241	0.01
TGT	1983/04/06	−4.0326	0.01	3.2715	0.01	−7.7915	0.01	5.7367	0.01
THC	1982/01/04	−2.367	0.4229	2.4587	0.01	−10.1076	0.01	14.9357	0.01
TIE	1996/06/05	−2.8304	0.2267	3.3316	0.01	−9.6219	0.01	7.7101	0.01
TIF	1987/12/30	−4.1833	0.01	1.1601	0.01	−8.991	0.01	20.7456	0.01
TJX	1988/01/05	−3.3642	0.05954	10.7126	0.01	−7.5396	0.01	17.5468	0.01
TLAB	1990/03/26	−3.078	0.1218	9.5543	0.01	−9.2833	0.01	6.9861	0.01
TMK	1987/12/30	−3.1928	0.08908	14.391	0.01	−5.9644	0.01	9.9487	0.01
TMO	1987/09/01	−2.2971	0.4525	2.6941	0.01	−5.9319	0.01	14.8628	0.01
TROW	1989/09/13	−4.2591	0.01	8.0583	0.01	−8.2765	0.01	11.6288	0.01
TRV	1986/07/09	−2.9936	0.1575	2.1414	0.01	−5.9976	0.01	15.786	0.01
TSN	1986/07/09	−3.9419	0.01185	10.5755	0.01	−7.0329	0.01	15.549	0.01
TSO	1983/04/06	−2.7498	0.2608	11.4	0.01	−4.6484	0.01	18.4274	0.01
TSS	1989/06/30	−4.4062	0.01	3.4532	0.01	−5.4317	0.01	10.2643	0.01
TWC	2007/01/10	−1.4697	0.8028	6.5446	0.01	−2.5195	0.3584	3.671	0.01
TWX	1992/03/19	−3.5075	0.04153	9.285	0.01	−6.8664	0.01	2.468	0.01
TXN	1981/12/31	−3.9277	0.01255	9.2026	0.01	−10.3291	0.01	25.2201	0.01
TXT	1984/10/24	−1.682	0.713	12.1514	0.01	−6.958	0.01	7.9288	0.01
UNH	1990/03/26	−3.6902	0.02445	2.3073	0.01	−9.9408	0.01	8.3355	0.01
UNM	1986/11/06	−2.7365	0.2665	9.7854	0.01	−8.9269	0.01	16.0197	0.01
UNP	1980/01/02	−3.9321	0.01233	5.3897	0.01	−6.281	0.01	21.1066	0.01
UPS	1999/11/10	−1.8384	0.6467	5.1418	0.01	−9.8493	0.01	7.0036	0.01
USB	1987/11/05	−2.4707	0.379	2.9654	0.01	−5.2899	0.01	11.47	0.01
UTX	1970/01/02	−4.8647	0.01	20.5892	0.01	−12.8623	0.01	26.9016	0.01
VAR	1988/01/05	−3.6885	0.02453	8.2004	0.01	−10.205	0.01	7.6905	0.01
VFC	1985/07/01	−2.9202	0.1886	16.2467	0.01	−7.5235	0.01	11.956	0.01
VIA-B	2005/12/05	−1.5809	0.7557	7.9834	0.01	−6.5211	0.01	3.881	0.01
VLO	1982/01/04	−2.6679	0.2955	23.9914	0.01	−5.8158	0.01	21.9905	0.01
VMC	1988/01/05	−3.3009	0.07044	1.269	0.01	−4.9235	0.01	16.1354	0.01
VNO	1988/01/05	−2.1316	0.5226	4.7882	0.01	−4.153	0.01	10.1834	0.01
VRSN	1998/01/30	−2.4822	0.3742	7.3946	0.01	−7.1287	0.01	1.2546	0.01
VTR	1997/05/05	−2.4915	0.3702	18.717	0.01	−5.2722	0.01	7.1218	0.01
VZ	1983/11/21	−4.231	0.01	18.3962	0.01	−5.7313	0.01	23.4511	0.01
WAG	1985/07/01	−5.0315	0.01	0.6426	0.01876	−7.9333	0.01	13.7793	0.01
WAT	1995/11/17	−2.9255	0.1865	1.453	0.01	−9.2428	0.01	7.5397	0.01
WEC	1984/10/26	−2.2051	0.4915	22.471	0.01	−9.3906	0.01	15.0932	0.01
WFC	1984/11/01	−2.9267	0.1859	8.5998	0.01	−4.0377	0.01	7.4342	0.01
WFMI	1992/01/23	−1.9609	0.5949	8.3661	0.01	−12.2542	0.01	12.618	0.01
WFR	1995/07/13	−1.7694	0.676	6.093	0.01	−5.828	0.01	17.4503	0.01
WHR	1983/06/10	−3.6234	0.03034	15.9436	0.01	−8.21	0.01	20.3086	0.01
WIN	2005/02/09	−1.7967	0.6644	6.6177	0.01	−5.5047	0.01	9.1021	0.01
WLP	2001/10/30	−2.6064	0.3216	3.2619	0.01	−9.0571	0.01	1.717	0.01
WMB	1981/12/31	−2.6313	0.3110	5.393	0.01	−9.177	0.01	17.7866	0.01
WMI	1991/09/30	−2.3805	0.4172	7.6952	0.01	−8.2027	0.01	13.5552	0.01

Table 6.10: Continuation of Table 6.9—Trading symbols of S&P500 constituents with the corresponding date of the earliest closing price (YYYY/MM/DD). Additionally, the values of the ADF tests and KPSS tests for the price time series $p(t)$ and the volume time series $v(t)$ are shown with their corresponding p-values.

Symbol	First record	ADF_p	p-value	$KPSS_p$	p-value	ADF_v	p-value	$KPSS_v$	p-value
WMT	1972/08/25	−4.8676	0.01	15.64	0.01	−8.2751	0.01	21.1582	0.01
WPI	1993/02/17	−3.4465	0.04741	3.5348	0.01	−11.238	0.01	9.1622	0.01
WPO	1990/01/24	0.437	0.99	23.4133	0.01	−9.6355	0.01	10.4187	0.01
WU	2006/10/02	−2.0205	0.5696	3.1053	0.01	−6.2486	0.01	3.1863	0.01
WY	1977/01/03	−2.9061	0.1945	33.2266	0.01	−10.544	0.01	20.6952	0.01
WYE	1982/01/04	−3.8951	0.01418	9.1345	0.01	−9.8538	0.01	17.2622	0.01
WYN	2006/07/19	−1.8769	0.6304	9.0214	0.01	−4.3627	0.01	6.3129	0.01
WYNN	2002/10/25	−0.8897	0.9536	8.4118	0.01	−7.5548	0.01	9.5582	0.01
X	1991/04/12	−2.7773	0.2492	8.2568	0.01	−3.0501	0.1336	14.1958	0.01
XEL	1985/09/24	−2.6392	0.3077	17.5381	0.01	−8.1787	0.01	18.1643	0.01
XL	1991/07/19	−1.0123	0.9368	5.3703	0.01	−6.5734	0.01	8.7546	0.01
XLNX	1990/06/18	−3.7544	0.02124	4.6573	0.01	−9.6466	0.01	6.2797	0.01
XOM	1970/01/02	−3.402	0.05294	2.1779	0.01	−5.8817	0.01	21.2867	0.01
XRAY	1991/04/25	−3.4425	0.04779	2.8132	0.01	−11.3189	0.01	7.0003	0.01
XRX	1977/01/03	−2.4366	0.3934	9.6837	0.01	−12.9665	0.01	20.3898	0.01
XTO	1993/05/12	−2.9509	0.1757	18.3783	0.01	−6.3202	0.01	17.9536	0.01
YHOO	1996/04/12	−3.4411	0.04794	5.7468	0.01	−8.1257	0.01	0.6964	0.01387
YUM	1997/09/17	−2.9815	0.1628	1.1658	0.01	−9.6112	0.01	1.5115	0.01
ZION	1990/03/26	−2.4743	0.3775	4.0995	0.01	−5.0047	0.01	9.1927	0.01
ZMH	2001/08/09	−0.9836	0.9414	6.0544	0.01	−9.5533	0.01	4.328	0.01

6.3 Renormalization Method

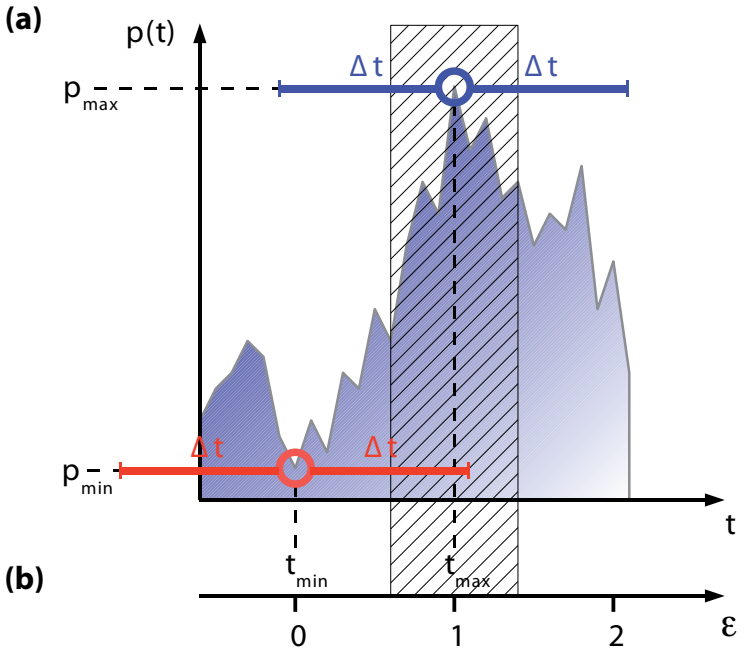

Figure 6.4: Visualization of a *microtrend* in the price movement $p(t)$. (a) Positive microtrend starting at a local price minimum $p_{\min}$ of order Δt and ending at a local price maximum $p_{\max}$ of order Δt. The hatched region around $p_{\max}$ indicates the interval in which we find scale-free behavior. This behavior is consistent with "self-organized" [Kru96] macroscopic interactions among many traders [Shl00], not unlike "tension" in a pedestrian crowd [HFV00, BKS02]. The reason for tension among financial traders may be found in the risk aversions and profit targets of financial market participants. (b) Renormalized time scale ε between successive extrema, where $\varepsilon = 0$ corresponds to the start of a microtrend, and $\varepsilon = 1$ corresponds to the end. The hatched region is surprisingly large, starting at $\varepsilon = 0.6$ and ending at $\varepsilon = 1.4$.

The various jagged functions of time characterizing complex financial fluctuations on time scales as short as a few milliseconds are much less studied than the large fluctuations of major national stock indices such as the S&P500 and their constituents. These functions also do not yield themselves to mathematical analysis at first sight, because they are characterized by sudden reversals between up and down microtrends (see Fig. 6.4 and Fig. 6.5a), which can also be referred to as microscopic *bubbles* on small

time scales. On these small time scales, evidence can be found [PPS08] that the three major financial market quantities of interest—price, volume, and inter-trade times—are connected in a complex way, and thus overwhelming the standard tools of time series analysis such as linear cross-correlation analyses. More sophisticated methods are thus necessary for analyzing these complex financial fluctuations responsible for such complex financial market patterns.

In order to study the switching processes in price movements on microscopic time scales, we first lay out a framework in which such a switching process can be quantitatively analyzed. Let $p(t)$ be the transaction price of trade t, which will be treated as a discrete variable $t = 1, \ldots, T$. Each transaction price $p(t)$ is defined to be a *local maximum* $p_{\max}(\Delta t)$ of order Δt, if there is no higher transaction price in the interval $t - \Delta t \leq t \leq t + \Delta t$. Thus, if $p(t) = p_{\max}(t, \Delta t)$, $p(t)$ is a *local maximum*, where

$$p_{\max}(t, \Delta t) = \max\{p(t) | t - \Delta t \leq t \leq t + \Delta t\}. \tag{6.4}$$

Analogously, a transaction price $p(t)$ is defined to be a *local minimum* of order Δt, if there is no lower transaction price in this interval. With

$$p_{\min}(t, \Delta t) = \min\{p(t) | t - \Delta t \leq t \leq t + \Delta t\}, \tag{6.5}$$

it follows that $p(t)$ is a *local minimum* if $p(t) = p_{\min}(t, \Delta t)$. In this sense, the two points in the time series in Fig. 6.4 marked by circles are a local minimum and a local maximum, respectively.

For the analysis of financial market quantities and their dependence on trends, we introduce a renormalized time scale ε between successive extrema as follows: Let $t_{\min}$ and $t_{\max}$ be the time (measured in units of ticks) at which the corresponding transactions take place for a successive pair of *local minimum* and *local maximum* (see Fig. 6.4). For a positive microtrend, the renormalized time scale is given by

$$\varepsilon(t) \equiv \frac{t - t_{\min}}{t_{\max} - t_{\min}}, \tag{6.6}$$

with $t \geq t_{\min}$, and for a negative microtrend by

$$\varepsilon(t) \equiv \frac{t - t_{\max}}{t_{\min} - t_{\max}}, \tag{6.7}$$

with $t \geq t_{\max}$. Thus, $\varepsilon = 0$ corresponds to the beginning of the microtrend and $\varepsilon = 1$ indicates the end of the microtrend. We analyze a range of ε within the interval $0 \leq \varepsilon \leq 2$, so we can analyze trend switching processes both before and after the critical value $\varepsilon = 1$ (Fig. 6.4). The renormalization is essential to assure that microtrends of various lengths can be aggregated and that all switching points have a common position on the renormalized time scale.

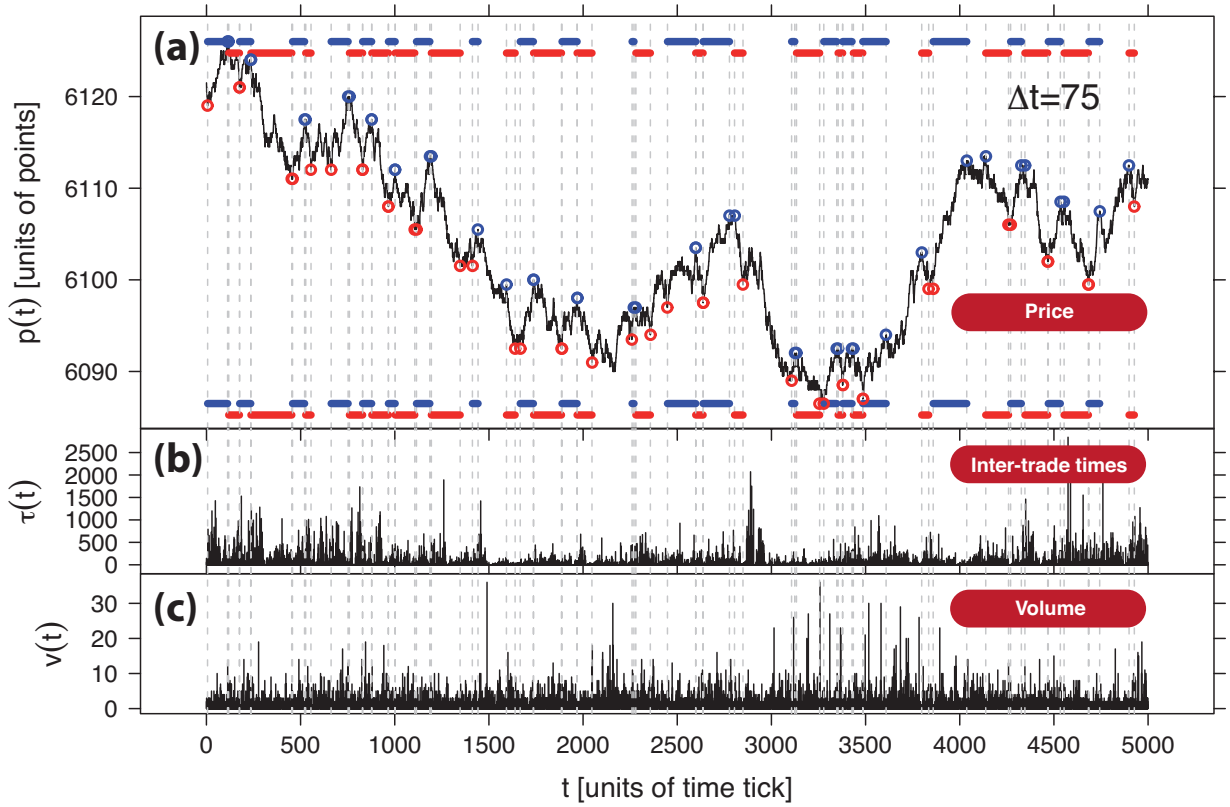

Figure 6.5: Visualization of the quantities analyzed. (a) A small subset comprising 5000 trades (0.04%) of the full $T_1 = 13,991,275$ trade data set analyzed, extracted from the German DAX future time series during part of one day. The extrema of order Δt are shown as circles, defined to be the extremum in the interval $t - \Delta t \leq t \leq t + \Delta t$. We performed our analysis for $\Delta t = 1, 2, \ldots, 1000$ ticks; in this example, $\Delta t = 75$ ticks. Positive microtrends are indicated by blue bars, which start at a Δt-minimum and end at the next Δt-maximum. A negative microtrend (red bars) starts at a Δt-maximum and ends at the consecutive Δt-minimum. (b) Time series of the corresponding inter-trade times $\tau(t)$ reflecting the *natural* time between consecutive trades in units of 10 ms, where $t = 1, 2, \ldots, 5000$ is the transactions index. (c) The volume $v(t)$ of each trade t in units of contracts.

6.3.1 Volume Analysis

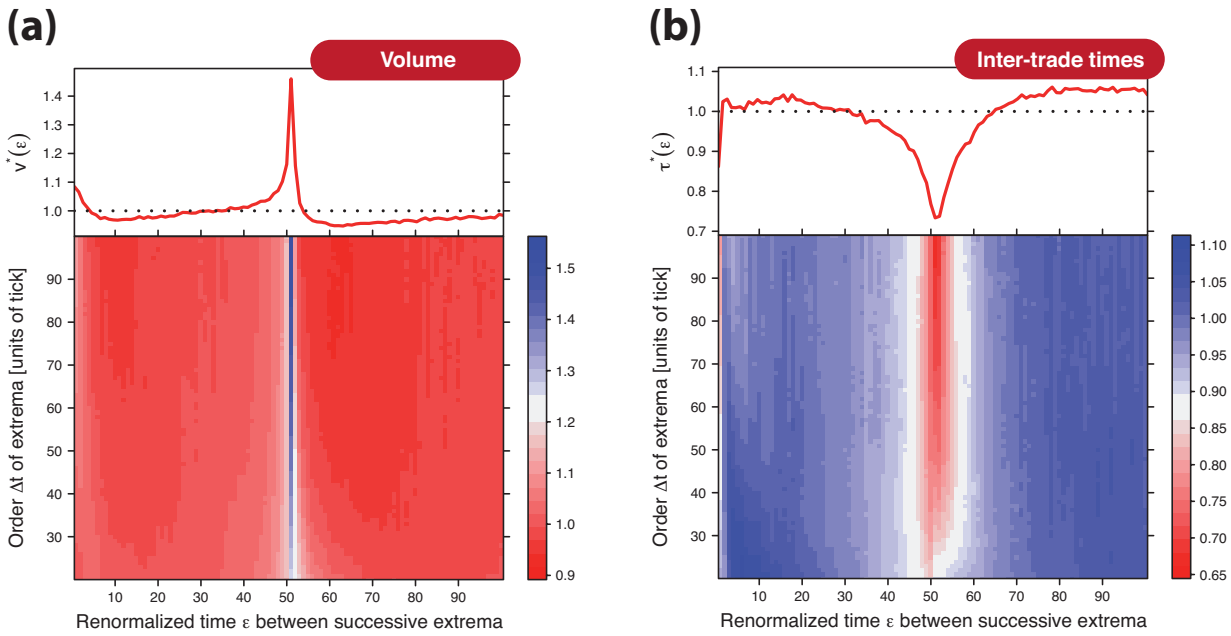

Figure 6.6: Renormalization time analysis of transaction volumes v, and inter-trade times τ for all microtrends—increasing and decreasing microtrends. (a) Volume profile, averaged over all microtrends in the German FDAX time series and normalized by the average volume of all microtrends studied. We analyze both positive and negative microtrends. The color code represents the normalized mean volume $\langle v \rangle(\varepsilon, \Delta t)/\bar{v}$. The color profile reflects the link between mean volume and price evolution. The volume is connected to the price evolution: new extreme values of the price coincide with peaks in the volume time series, as indicated by the vertical blue regions close to $\varepsilon = 1$. The top panel shows the volume aggregation $v^*(\varepsilon)$, where $v^*(\varepsilon)$ is the average over layers with $50 \leq \Delta t \leq 100$ ($\Delta t_{\text{cut}} = 50$). The sharp maximum in $v^*(\varepsilon)$ is shown in the top panel. (b) The colored inter-trade time profile—averaged over all microtrends in the time series and normalized by the average inter-trade times of all microtrends studied—is performed analogously to our study of volume. New extreme values of the price time series are reached with a significant decay of the inter-trade times.

First the volume fluctuations $v(t)$ during the time interval of increasing microtrends from one price minimum to the next price maximum and decreasing microtrends from one price maximum to the next price minimum are analyzed. The quantity studied is the number of contracts traded in each individual transaction (see Fig. 6.5c) in the case of microtrends for the German market. For the US market, it is the cumulative number of traded stocks per day. For the analysis of $v(t)$ and its dependence on trend

fraction, we use the renormalization time scale ε. In Fig. 6.6, the color key represents the mean volume $\langle v \rangle(\varepsilon, \Delta t)$ depending on ε and Δt, normalized by average volume $\bar{v}$, where the brackets $\langle \ldots \rangle$ denote the average over all increasing microtrends and all decreasing microtrends in the entire time series of $T_1 = 13,991,275$ records. If there are $N_{\text{pos}}(\Delta t)$ positive microtrends and $N_{\text{neg}}(\Delta t)$ negative microtrends, each of order Δt in the time series, letting $v_i(\varepsilon)$ denote the local volume at position ε in the i-th positive or i-th negative microtrend, then the mean volume is given by

$$\langle v \rangle(\varepsilon, \Delta t) = \frac{1}{N_{\text{pos}}(\Delta t) + N_{\text{neg}}(\Delta t)} \sum_{i=1}^{N_{\text{pos}}(\Delta t)+N_{\text{neg}}(\Delta t)} v_i(\varepsilon) \tag{6.8}$$

for positive and negative microtrends. This mean volume has to be normalized by the average volume $\bar{v}$, which is determined by

$$\bar{v} = \frac{\varepsilon_{\text{bin}}}{\varepsilon_{\max} \Delta t_{\max}} \sum_{\varepsilon=0}^{\varepsilon_{\max}/\varepsilon_{\text{bin}}} \left(\sum_{\Delta t=0}^{\Delta t_{\max}} \langle v \rangle(\varepsilon, \Delta t) \right), \tag{6.9}$$

where $\varepsilon_{\max}$ is the maximum value of the renormalization time scale ε studied (fixed to $\varepsilon_{\max} = 2$) and ε_{bin} denotes the bin size of the renormalized time scale. The maximum value of Δt analyzed is given by $\Delta t_{\max}$. For convenience, we relate the bin size to $\Delta t_{\max}$ through

$$\varepsilon_{\text{bin}} = \frac{2}{\Delta t_{\max}}. \tag{6.10}$$

As the absence of significant changes of the colored volume profiles in Fig. 6.6a is consistent with a *data collapse* for Δt values larger than a specific cut-off value Δt_{cut}, we calculate a volume aggregation $v^*(\varepsilon)$. This volume aggregation $v^*(\varepsilon)$ is the average of the mean volume $\langle v \rangle(\varepsilon, \Delta t)$, averaged only for Δt-layers with $\Delta t_{\text{cut}} \leq \Delta t \leq \Delta t_{\max}$. It is given by

$$v^*(\varepsilon) = \frac{1}{\Delta t_{\max} - \Delta t_{\text{cut}}} \sum_{\Delta t=\Delta t_{\text{cut}}}^{\Delta t_{\max}} \frac{\langle v \rangle(\varepsilon, \Delta t)}{\bar{v}}. \tag{6.11}$$

The colored volume profile (see Fig. 6.6a) shows the mean volume $\langle v \rangle(\varepsilon, \Delta t)$ averaged over all increasing and all decreasing microtrends. The color profiles exhibit a link between volume and price evolution. In Fig. 6.2d, we show the averaged volume $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ in a log–log histogram. Surprisingly, the average volume decreases on straight lines and indicating a power-law scaling behavior of the form

$$v^*(|\varepsilon - 1|) \sim |\varepsilon - 1|^{\beta_v} \tag{6.12}$$

with scaling parameters $\beta_v^- = -0.068 \pm 0.001$ (t-test, p-value $< 2 \times 10^{-16}$) before, and $\beta_v^+ = -0.155 \pm 0.004$ (t-test, p-value $= 9.2 \times 10^{-16}$) after a price extremum. Such an

extraction of slopes by performing least-squares linear regressions is not sufficient for supporting the claim that the averaged volume consistent with a power-law behavior. However, additional performed statistical tests (see section 6.4) enable us to conclude that our observations are indeed consistent with this hypothesis.

We test the possible universality of our result by performing a parallel analysis for trends on long time scales using the daily closing price dataset of S&P500 stocks. Note that for our parallel analysis on macroscopic time scales, Δt is measured in units of days, and that $v^*(\varepsilon)$ is averaged over all trends and all the closing price time series of all S&P500 individual stocks. In order to avoid biased contributions to the rescaled averaging caused by inflation based drifts over more than 47 years, the analyzed price time series $p(t)$ contains the logarithm of the daily closing prices. A log–log histogram of our parallel analysis for the US market on large time scales (Figs. 6.2b and 6.2e) provides evidence for similar behavior with scaling parameters $\beta_v^- = -0.052 \pm 0.001$ (t-test, p-value $= 1.7 \times 10^{-9}$) before, and $\beta_v^+ = -0.109 \pm 0.003$ (t-test, p-value $< 2 \times 10^{-16}$) after a price extremum. Statistical tests confirm, also in this case, consistency with a power-law distribution.

6.3.2 Inter-Trade Time Analysis

In order to verify a possible universality, we further analyze the behavior of the inter-trade times $\tau(t)$ on the German market during the short time interval from one price extremum to the next (see Fig. 6.5b). In Fig. 6.6b, the mean inter-trade time $\langle\tau\rangle(\varepsilon, \Delta t)/\bar{\tau}$ is shown for positive and negative microtrends reflecting the link between inter-trade times and price extrema. The mean inter-trade time begins decreasing far from the switching point $\varepsilon = 1$. After the formation of a new local price maximum, the mean inter-trade times increase and return to the average value in a very symmetric way.

In the top panel of Fig. 6.6b, the aggregation of the inter-trade time profile $\tau^*(\varepsilon)$ is calculated and shown for all values of Δt between $\Delta t_{\text{cut}} = 50$ and $\Delta t_{\text{max}} = 100$. Figure 6.2f shows $\tau^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram further supporting a power-law behavior of the form

$$\tau^*(|\varepsilon - 1|) \sim |\varepsilon - 1|^{\beta_\tau} \tag{6.13}$$

with scaling parameters $\beta_\tau^- = 0.092 \pm 0.002$ (t-test, p-value $= 1.8 \times 10^{-15}$) before, and $\beta_\tau^+ = 0.118 \pm 0.002$ (t-test, p-value $< 2 \times 10^{-16}$) after a price extremum. Statistical tests confirm consistency with a power-law distribution here as well. A log–log histogram of a parallel analysis for the US market on large time scales is not obtainable as the inter-trade times between successive closing prices are given by the constant value of one day (exceptions are weekends and holidays).

6.4 Test of the Power-Law Hypothesis

Using linear regressions, we are able to fit power-law curves to empirical data sets, and obtain thus an estimate of the scaling parameter β. However, this procedure does not prove that there is indeed a power-law underlying. Irrespective of the true functional law, underlying the process generating the empirical data, it is often possible to fit them in seemingly good agreement to a power-law graph. We need to find out whether the fit is a good match to the data. To achieve this, we test the power-law hypothesis quantitatively [CSN09]. The fact that distributions seem to be roughly straight on the log-log plot is a necessary but not sufficient condition for power-law behavior of the data. Unfortunately, there is no straightforward way to prove whether the underlying law is indeed a power-law. Even if data are sampled from a power-law distribution, their observed distribution is unlikely to exactly follow the power-law form. There will always be some small deviations caused by the random nature of the sampling process. As described in [CSN09], the challenge is to distinguish between deviations of this type and those that arise because the data are drawn from a non-power-law distribution.

For the test of the power-law hypothesis, we use the following approach: a large number of synthetic data sets are sampled from a true power-law distribution. The extent to which these synthetic data sets deviate from the power-law form is then measured, and these distances are compared to the results of the same measurement for the empirical price time series. If the deviation of the empirical data set from the power-law is larger than corresponding deviations of a specific fraction of synthetically created data sets, then the power-law is not a plausible fit to the data set. As a measure for the distance between distributions, we use the Kolmogorov-Smirnov (KS) statistic [PFTV92], which is simply the maximum distance between the cumulative distribution functions (CDFs) of the empirical data $v^*(x)$ (or $\tau^*(x)$) and the fitted model $q(x) = Cx^\beta$:

$$D = \max_{x_0 \leq x \leq x_{\text{cut}}} |V^*(x) - Q(x)|. \tag{6.14}$$

Here, C is a constant, $V^*(x)$ is the cumulative distribution function (CDF) of the empirical data $v^*(x)$ for the observations with values on the interval $[x_0, x_{\text{cut}}]$, and $Q(x)$ is the CDF of the power-law model $q(x)$ that provides the best fit to the data in this region. The variable x represents $|\varepsilon - 1|$.

First, we fit our empirical data $v^*(x)$ to the power-law model using linear regression, determining the value of the scaling parameter β. For this fit, we calculate the KS statistic. Next, we generate a large number of power-law distributed synthetic data sets with scaling parameter β as follows: a power-law distribution $q(x)$ is generated from a uniform distribution $q(y)$ with $q(x) = Cx^\beta$ for $x \in [x_0, x_{\text{cut}}]$. Then, normalization requires

$$\int_{x_0}^{x_{\text{cut}}} q(x)dx = C\frac{\left[x^{\beta+1}\right]_{x_0}^{x_{\text{cut}}}}{\beta+1} = 1. \tag{6.15}$$

Thus, the constant C is given by

$$C = \frac{\beta + 1}{x_{\text{cut}}^{\beta+1} - x_0^{\beta+1}}. \tag{6.16}$$

Let Y be a uniformly distributed variate on the interval $[0, 1]$. Then,

$$\int_{x_0}^{x} q(x')dx' = C \int_{x_0}^{x} x'^{\beta} dx' = \frac{C}{\beta + 1} \left(x^{\beta+1} - x_0^{\beta+1} \right) \equiv y, \tag{6.17}$$

and the variate given by

$$x = \left(\frac{\beta + 1}{C} y + x_0^{\beta+1} \right)^{1/(\beta+1)} = \left[\left(x_{\text{cut}}^{\beta+1} - x_0^{\beta+1} \right) y + x_0^{\beta+1} \right]^{1/(\beta+1)} \tag{6.18}$$

is distributed as $q(x)$ [PFTV92, Ben98]. Using Eq. (6.18), we generate 1000 synthetic data sets and fit each data set individually to its own power-law with its own values for C and β. Then, the KS statistic is calculated for each one. The parameter x_0 is the first measuring point closest to the switching point at $\varepsilon = 1$. The right border of the fitting region is $x_{\text{cut}} \equiv |\varepsilon - 1|_{\text{cut}}$. As a conservative approach, one synthetic data set contains $T_i(x_{\text{cut}} - x_0)$ power-law distributed data points, where T_i denotes the number of empirical observations. The number of effectively used empirical transactions is marginally smaller, as microtrends have to fulfill the requirements that time intervals between successive trades not be longer than one minute and transaction volumes not be larger than 100 contracts.

After the generation of 1000 synthetic data sets, we simply determine the fraction of time for which the resulting statistic is larger than the same value for the empirical data. This fraction is our p-value. It is of crucial importance that for each synthetic data set we compute the KS statistic relative to the best power-law fit for that data set. Thus, we ensure that we are performing for each synthetic data set the same calculation that we performed for the empirical data set. Once the p-value is calculated, a decision must be made concerning whether it is small enough to rule out the power-law hypothesis or whether the hypothesis is a plausible one for the data in question. If the p-value is close to 1, then the difference between the empirical data and the model can be attributed to statistical fluctuations alone. If the p-value is small, the model is not a plausible fit to the data.

In our calculations, the relatively conservative choice was made that the power-law is ruled out if the p-value is smaller than 0.1. Tables 6.11, 6.12, and 6.13 provide the p-values for the empirical data shown in Figs. 6.2d, 6.2e, and 6.2f as a function of $|\varepsilon - 1|_{\text{cut}}$. The second largest value of $|\varepsilon - 1|_{\text{cut}}$, for which the power-law hypothesis cannot be ruled out, is used as a limit for the fitting intervals in Fig. 6.2.

The results of the statistical test shown in Tables 6.11, 6.12, and 6.13 confirm consistency with power-law distributions.

Table 6.11: Statistical test of power-law hypothesis for the FDAX volume time series: Scaling parameters of the hypothesized power-law model are shown for both $v^*(\varepsilon)$ before (β_v^-) and $v^*(\varepsilon)$ after (β_v^+) the trend switching point $\varepsilon = 1$ in dependence of $|\varepsilon - 1|_{\text{cut}}$. Additionally, the corresponding values of the KS statistic, D_v^- and D_v^+, are given. The power-law hypothesis is supported if the p-value is larger than 0.1.

$\lvert\varepsilon - 1\rvert_{\text{cut}}$	β_v^+	D_v^+	p-value	β_v^-	D_v^-	p-value
0.004	−0.161	0	**0.568**	−0.062	0	**1**
0.006	−0.162	0.0001	**0.98**	−0.07	0.0014	**0.609**
0.008	−0.165	0.0007	**0.934**	−0.071	0.0011	**0.837**
0.01	−0.169	0.0017	**0.722**	−0.074	0.0014	**0.773**
0.012	−0.172	0.0021	**0.637**	−0.075	0.0016	**0.707**
0.014	−0.173	0.002	**0.679**	−0.074	0.0012	**0.873**
0.016	−0.17	0.0015	**0.873**	−0.074	0.001	**0.934**
0.018	−0.168	0.0012	**0.942**	−0.073	0.0008	**0.965**
0.02	−0.167	0.0014	**0.893**	−0.073	0.0008	**0.981**
0.022	−0.165	0.0018	**0.822**	−0.073	0.0007	**0.977**
0.024	−0.162	0.0021	**0.696**	−0.073	0.0006	**0.992**
0.026	−0.16	0.0027	**0.465**	−0.073	0.0006	**0.992**
0.028	−0.159	0.0031	**0.28**	−0.074	0.0006	**0.998**
0.03	−0.157	0.0034	**0.181**	−0.073	0.0005	**1**
0.032	−0.155	0.0038	0.065	−0.073	0.0005	**0.998**
0.034	−0.152	0.0042	0.019	−0.072	0.0007	**0.98**
0.036	−0.15	0.0046	0.004	−0.072	0.0006	**0.996**
0.038	−0.148	0.0051	0.002	−0.072	0.0006	**0.99**
0.04	−0.146	0.0055	0	−0.072	0.0007	**0.974**
0.042	−0.145	0.0059	0	−0.071	0.0008	**0.928**
0.044	−0.143	0.0061	0	−0.071	0.0008	**0.936**
0.046	−0.142	0.0062	0	−0.071	0.0009	**0.921**
0.048	−0.14	0.0064	0	−0.071	0.0008	**0.911**
0.05	−0.138	0.0066	0	−0.071	0.0008	**0.925**
0.052	−0.136	0.0068	0	−0.071	0.0008	**0.935**
0.054	−0.134	0.007	0	−0.07	0.0008	**0.911**
0.056	−0.132	0.0074	0	−0.07	0.0008	**0.902**
0.058	−0.131	0.0076	0	−0.07	0.0008	**0.922**
0.06	−0.129	0.0079	0	−0.07	0.0008	**0.936**
0.062	−0.127	0.0084	0	−0.07	0.0009	**0.881**
0.064	−0.125	0.0088	0	−0.07	0.0008	**0.876**
0.066	−0.124	0.009	0	−0.069	0.0009	**0.847**
0.068	−0.122	0.0092	0	−0.069	0.0009	**0.827**
0.07	−0.121	0.0095	0	−0.069	0.001	**0.734**
0.072	−0.119	0.0098	0	−0.068	0.0011	**0.624**
0.074	−0.118	0.01	0	−0.068	0.0012	**0.562**
0.076	−0.116	0.0101	0	−0.068	0.0013	**0.518**

Table 6.12: Statistical test of power-law hypothesis for the volume time series of S&P500 stocks: Scaling parameters of the hypothesized power-law model are shown for both $v^*(\varepsilon)$ before (β_v^-) and $v^*(\varepsilon)$ after (β_v^+) the trend switching point $\varepsilon = 1$ in dependence of $|\varepsilon - 1|_{\text{cut}}$. Additionally, the corresponding values of the KS statistic, D_v^- and D_v^+, are given. The power-law hypothesis is supported if the p-value is larger than 0.1.

$\lvert\varepsilon - 1\rvert_{\text{cut}}$	β_v^+	D_v^+	p-value	β_v^-	D_v^-	p-value
0.04	−0.084	0	**1**	−0.055	0	**1**
0.06	−0.098	0.0023	**0.272**	−0.052	0.0005	**0.792**
0.08	−0.111	0.0042	0.032	−0.053	0.0003	**0.981**
0.1	−0.113	0.0038	0.046	−0.053	0.0002	**0.997**
0.12	−0.115	0.0034	0.061	−0.054	0.0003	**0.999**
0.14	−0.116	0.003	**0.134**	−0.054	0.0003	**0.999**
0.16	−0.114	0.0026	**0.208**	−0.053	0.0007	**0.917**
0.18	−0.113	0.0023	**0.288**	−0.052	0.0008	**0.927**
0.2	−0.115	0.0021	**0.384**	−0.05	0.0015	**0.413**
0.22	−0.115	0.002	**0.435**	−0.048	0.0022	0.087
0.24	−0.115	0.0018	**0.516**	−0.047	0.0024	0.047
0.26	−0.114	0.0017	**0.618**	−0.047	0.0023	0.047
0.28	−0.113	0.0016	**0.625**	−0.046	0.0023	0.038
0.3	−0.112	0.0016	**0.651**	−0.045	0.0023	0.027
0.32	−0.11	0.0016	**0.66**	−0.044	0.0024	0.024
0.34	−0.109	0.0018	**0.463**	−0.044	0.0023	0.033
0.36	−0.107	0.0023	**0.151**	−0.043	0.0022	0.019
0.38	−0.106	0.0026	0.062	−0.042	0.0023	0.012
0.4	−0.105	0.003	0.01	−0.041	0.0025	0.001
0.42	−0.103	0.0035	0	−0.04	0.0027	0.001
0.44	−0.101	0.0038	0	−0.04	0.0027	0
0.46	−0.1	0.004	0	−0.04	0.0026	0.001

Table 6.13: Statistical test of power-law hypothesis for the FDAX inter-trade time series: Scaling parameters of the hypothesized power-law model are shown for both $\tau^*(\varepsilon)$ before (β_τ^-) and $\tau^*(\varepsilon)$ after (β_τ^+) the trend switching point $\varepsilon = 1$ in dependence of $|\varepsilon - 1|_{\text{cut}}$. Additionally, the corresponding values of the KS statistic, D_v^- and D_v^+, are given. The power-law hypothesis is supported if the p-value is larger than 0.1.

$\|\varepsilon - 1\|_{\text{cut}}$	β_τ^+	D_τ^+	p-value	β_τ^-	D_τ^-	p-value
0.04	0.089	0	**1**	0.097	0	**1**
0.06	0.096	0.0013	**0.21**	0.089	0.0013	**0.195**
0.08	0.105	0.0028	0.002	0.095	0.0013	**0.24**
0.1	0.112	0.0034	0	0.098	0.0016	**0.166**
0.12	0.115	0.0034	0	0.095	0.001	**0.538**
0.14	0.116	0.0031	0	0.096	0.0009	**0.708**
0.16	0.114	0.0025	0.01	0.096	0.0008	**0.837**
0.18	0.116	0.0023	0.023	0.095	0.0007	**0.928**
0.2	0.118	0.0025	0.006	0.094	0.0008	**0.885**
0.22	0.119	0.0025	0.008	0.094	0.001	**0.762**
0.24	0.12	0.0027	0.001	0.093	0.0011	**0.632**
0.26	0.121	0.0025	0	0.093	0.001	**0.749**
0.28	0.121	0.0025	0.002	0.092	0.0011	**0.714**
0.3	0.121	0.0023	0.008	0.091	0.0013	**0.435**
0.32	0.121	0.0021	0.042	0.089	0.0022	0
0.34	0.121	0.0019	**0.114**	0.089	0.0021	0.003
0.36	0.12	0.0017	**0.192**	0.088	0.0023	0
0.38	0.12	0.0016	**0.331**	0.087	0.0025	0
0.4	0.12	0.0015	**0.447**	0.087	0.0024	0
0.42	0.119	0.0014	**0.585**	0.087	0.0024	0
0.44	0.118	0.0013	**0.67**	0.086	0.0025	0
0.46	0.117	0.0014	**0.526**	0.085	0.0026	0

6.5 Random Reshuffling

To confirm that our results are a consequence of the exact time series sequence and thus sensitive to the temporal ordering of the original time series of volumes and inter-trade times, we randomly reshuffle γT pairs of data points in both the volume time series and the inter-trade time series in order to weaken their connection with the price evolution. We find that the clear link between volume fluctuations and price evolution (see Fig. 6.7a) and between inter-trade times and price evolution (see Fig. 6.7b) disappears with increasing γ and entirely vanishes for microtrends for $\gamma \geq 1$. The dip of the inter-trade times at $\varepsilon = 1$ becomes less pronounced with increasing γ and, correspondingly, the peak of the volume maximum decreases. For the S&P500 data set (Fig. 6.7c), the volume peak disappears with increasing γ, following the same pattern. These shuffling induced processes may be characterized by power-law relationships as well, which support our finding that a fluctuating price time series passes through a sequence of switching points with scale-free properties. This disappearance phenomenon is consistent with a power-law behavior. The maximum value of $v^*(\varepsilon)_\gamma$ at $\varepsilon = 1$ scales with the exponent $\beta_v^s = -0.12$ for microtrends (Fig. 6.7d). The minimum value of $\tau^*(\varepsilon)_\gamma$ at $\varepsilon = 1$ scales with the exponent $\beta_v^s = 0.09$ as shown in Fig. 6.7e. In the case of the maximum of $v^*(\varepsilon)_\gamma$ at $\varepsilon = 1$ on large time scales, the log-log plot can be fit with a straight line with a power-law exponent $\beta_v^s = -0.10$ for the S&P500 stocks (Fig. 6.7f). However, some deviations can be observed for macrotrends, which are caused by the limited number of closing prices in the S&P500 data set ($T_2 \ll T_1$).

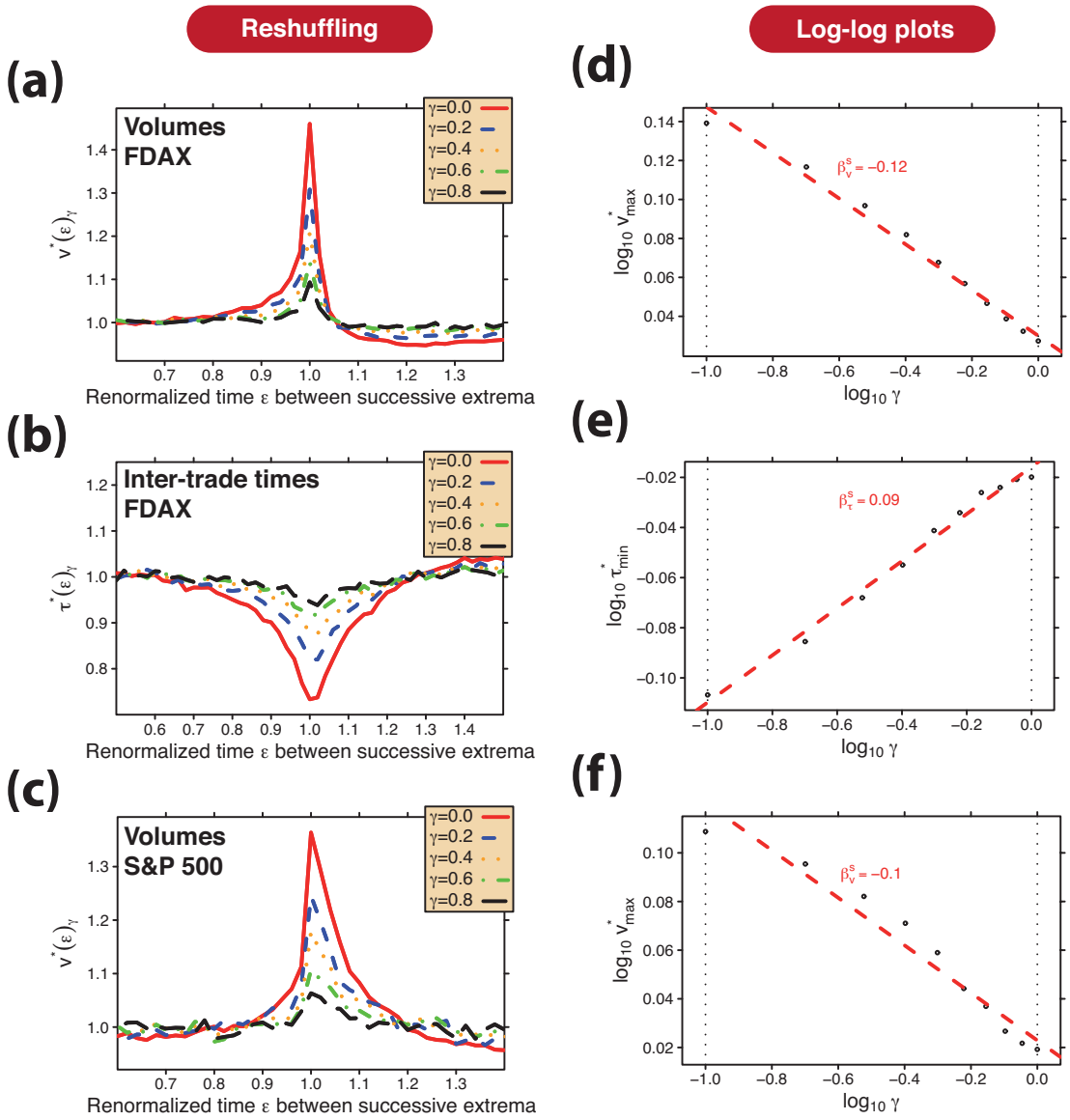

Figure 6.7: Stability test of the power-law dependence. (a) If randomly reshuffling γT pairs of volume entries in the multivariate time series, the significant link between volume and price evolution starts to disappear as γ increases. (b) If γT pairs of inter-trade time entries are randomly reshuffled, the inter-trade time dip starts to disappear. (c) We find an identical behavior for the volume peak on long time scales, using daily closing prices of S&P500 stocks. (d) The disappearance phenomenon also appears to follow a power-law behavior. The maximum value of $v^*(\varepsilon)_\gamma$ at $\varepsilon = 1$ scales with exponent $\beta_v^s = -0.115 \pm 0.005$. (e) The minimum value of $\tau^*(\varepsilon)_\gamma$ at $\varepsilon = 1$ scales with exponent $\beta_v^s = 0.094 \pm 0.004$. (f) In the case of the maximum of $v^*(\varepsilon)_\gamma$ at $\varepsilon = 1$ for the S&P500 stocks, the plot may be fitted by a power-law with exponent $\beta_v^s = -0.095 \pm 0.008$.

6.6 Analysis of Dow Jones Industrial Average Index Components

In this section, we perform a parallel analysis of the 30 assets contributing to the *Dow Jones Industrial Average (DJIA)*[5]. The time series stem from the *Trade and Quote* (TAQ)[6] database and cover 2,623,445,866 transactions, roughly 200 times the length of the *FDAX* time series analyzed in section 6.1.

The TAQ database makes it possible to study the switching phenomenon of transaction volume on an intraday time scale for all 30 stocks individually[7]. For the sake of clarity and presentation, we only present the results for one stock in the next subsection. Results for all the other stocks considered can be found in the appendix A. For each stock, we document the development of the stock price, the aggregated volume as a function of ε, and $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram with two power-law fits. The fitting exponents are reported in the upper right corner. Each fitting range is based on statistical tests identical to those applied in the previous analyses: The shaded intervals mark the region in which the empirical data are consistent with a power-law behavior. The left border of the shaded regions is given by the first measuring point closest to the switching point. The right borders stem from statistical tests of the power-law hypothesis.

As is evident in the presented results, the *DJIA* components confirm our previous findings and provide evidence for power-law exponents between $\beta_v^- = -0.06$ and $\beta_v^- = -0.01$ before the switching point $\varepsilon = 1$ and between $\beta_v^+ = -0.46$ and $\beta_v^+ = -0.12$ after the switching point $\varepsilon = 1$.

[5] We analyze all US stocks which were included in the DJIA in March 2010.

[6] The New York Stock Exchange's TAQ database contains data on all trades and quotes for stocks from The New York Stock Exchange (NYSE), The American Stock Exchange (AMEX) and The National Association of Securities Dealers Automated Quotation (Nasdaq).

[7] The quality of the time stamp is not comparable to that of the FDAX time series. Thus, only the volume is studied.

6.6.1 Alcoa Inc. Common Stock (AA)

The AA price time series contains 63,548,627 transactions.

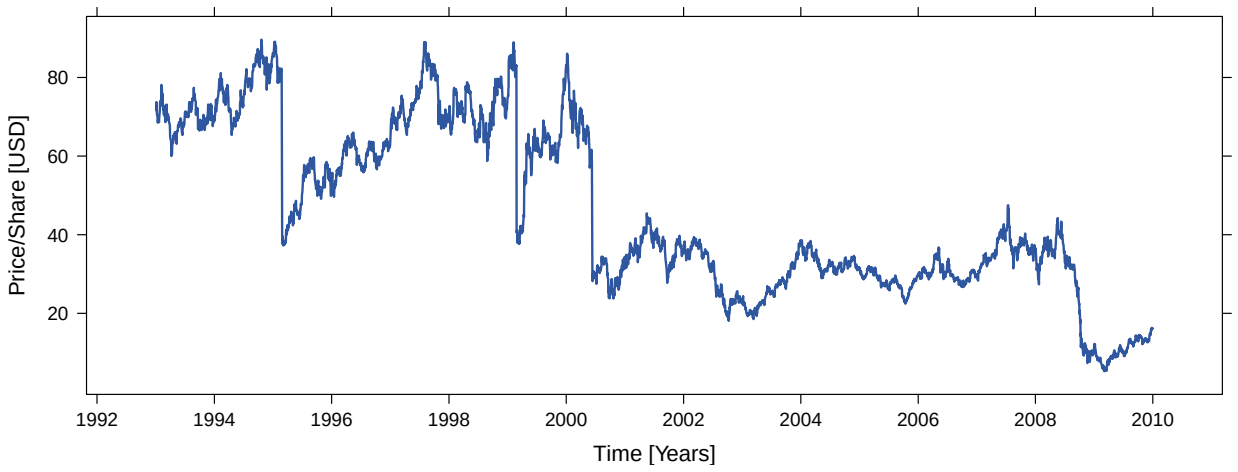

Figure 6.8: Stock splits occurred on 27 February 1995 [split ratio 2:1], 26 February 1999 [split ratio 2:1], and 12 June 2000 [split ratio 2:1].

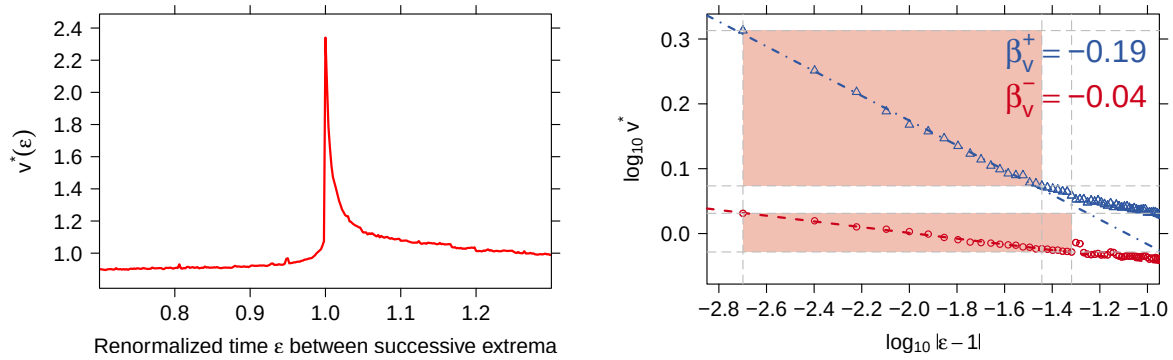

Figure 6.9: *left*: aggregated volume $v^*(\varepsilon)$. *right*: $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram.

The results for all other stocks in the dataset are included in appendix A.

Table 6.14: Statistical test of power-law hypothesis for the AA volume time series: Scaling parameters of the hypothesized power-law model are shown for both $v^*(\varepsilon)$ before (β_v^-) and $v^*(\varepsilon)$ after (β_v^+) the trend switching point $\varepsilon = 1$ in dependence of $|\varepsilon - 1|_{\text{cut}}$. Additionally, the corresponding values of the KS statistic, D_v^- and D_v^+, are given. The power-law hypothesis is supported if the p-value is larger than 0.1.

$\lvert\varepsilon - 1\rvert_{\text{cut}}$	β_v^+	D_v^+	p-value	β_v^-	D_v^-	p-value
0.004	−0.204	0	**1**	−0.038	0	**1**
0.006	−0.199	0.0009	**0.58**	−0.043	0.0008	**0.557**
0.008	−0.204	0.0011	**0.707**	−0.041	0.0005	**0.843**
0.01	−0.206	0.0013	**0.78**	−0.041	0.0003	**0.956**
0.012	−0.203	0.0009	**0.938**	−0.041	0.0003	**0.986**
0.014	−0.199	0.0018	**0.726**	−0.042	0.0007	**0.779**
0.016	−0.198	0.002	**0.7**	−0.043	0.001	**0.355**
0.018	−0.197	0.0019	**0.843**	−0.045	0.0013	**0.204**
0.02	−0.197	0.0017	**0.903**	−0.045	0.0013	**0.213**
0.022	−0.197	0.0016	**0.962**	−0.045	0.0012	**0.269**
0.024	−0.197	0.0015	**0.982**	−0.045	0.0011	**0.34**
0.026	−0.197	0.0014	**0.995**	−0.045	0.001	**0.381**
0.028	−0.195	0.0014	**0.997**	−0.045	0.0009	**0.498**
0.03	−0.193	0.0022	**0.795**	−0.045	0.0008	**0.546**
0.032	−0.193	0.0023	**0.717**	−0.044	0.0007	**0.582**
0.034	−0.191	0.0027	**0.373**	−0.045	0.0007	**0.614**
0.036	−0.19	0.0029	**0.146**	−0.044	0.0007	**0.653**
0.038	−0.189	0.0032	0.037	−0.044	0.0006	**0.708**
0.04	−0.188	0.0035	0.003	−0.044	0.0006	**0.714**
0.042	−0.186	0.0038	0	−0.044	0.0006	**0.744**
0.044	−0.185	0.004	0	−0.044	0.0006	**0.766**
0.046	−0.183	0.0044	0	−0.044	0.0005	**0.774**
0.048	−0.182	0.0046	0	−0.044	0.0005	**0.798**
0.05	−0.181	0.0046	0	−0.042	0.0013	0.027
0.052	−0.18	0.0048	0	−0.04	0.0024	0
0.054	−0.179	0.005	0	−0.04	0.0022	0
0.056	−0.178	0.005	0	−0.04	0.002	0
0.058	−0.177	0.0053	0	−0.041	0.0018	0
0.06	−0.176	0.0054	0	−0.041	0.0017	0
0.062	−0.174	0.0057	0	−0.041	0.0017	0
0.064	−0.173	0.0058	0	−0.041	0.0016	0
0.066	−0.172	0.0059	0	−0.041	0.0015	0.003
0.068	−0.171	0.006	0	−0.041	0.0015	0.002
0.07	−0.169	0.0062	0	−0.04	0.0016	0
0.072	−0.168	0.0064	0	−0.04	0.0017	0
0.074	−0.167	0.0067	0	−0.04	0.0017	0
0.076	−0.166	0.007	0	−0.04	0.0016	0

6.7 Chapter's Summary

The straight lines in Fig. 6.2 offer insight into financial market fluctuations: (i) a clear connection exists between volumes, inter-trade times, and price fluctuations on the path from one extremum to the next extremum, and (ii) the underlying distribution, which describes the tails of volumes and inter-trade times around extrema varying over 9 orders of magnitude starting from the smallest possible time scale, is a power-law with unique exponents which quantitatively characterize the region around the trend switching point. As a direct consequence of the existence of power-law tails, this behavior does not depend on the scale. Thus, we find identical behavior for other sub-intervals of $50 \leq \Delta t \leq 1000$. If the value of Δt is decreased, the number of local minima and maxima increases (see Fig. 6.4), around which we find scale-free behavior for exactly the same ε interval $0.6 \leq \varepsilon \leq 1.4$. The characterization of volume and inter-trade times by power-law relationships in the time domain supports the hypothesis that a fluctuating price time series passes through a sequence of "phase transitions".

Before concluding, one might ask "what kind of phase transition" could the end of a microtrend or macrotrend correspond to, or is the end of a trend an altogether different kind of phase transition which resembles all other phase transitions by displaying a regime of scale free behavior characterized by a critical exponent. It may be premature to speculate on possible analogies, so we limit ourselves here to describing what seems to be the leading candidate-explanation. Consider a simple Ising magnet characterized by one-dimensional spins that can point north or south. Each spin interacts with some (or even with all) of its neighbors with positive interaction strength J, such that when J is positive, neighboring spins lower their energy by lining up in parallel. The entire system is bathed in a magnetic field that interacts with all spins equally, and with a strength parametrized by H, such that when H is positive the field points north and when H is negative the field points south. Thus when H is positive, the system lowers its energy by pointing each spin north. There are thus two competing control parameters J and H. If, for example, the system is set to a state with the majority of spins pointing north, while the field H is pointing south, the competition will be between the relative effects of J and H: the J interaction pulls the spins to point to the north but the H interaction pulls the spin to point south. Such a system is termed *metastable* since if each north-pointing spin suddenly flips its state to point south, the system can achieve a lower total energy. This flipping will occur in time in a fashion not unlike the trading frequency near the end of a trend: first one or two spins will randomly switch their state, then more, and suddenly in an "avalanche" the majority of spins will point South. This phase transition is known as a spinodal singularity, characterized by its own set of exponents. Why should the end of microtrends or macrotrends exhibit similarities with a metastable physical system? Presumably because near the end of a positive trend, all market participants watching the market begin to sense that the market is metastable and that if they do not sell soon, it could be too late to make any profit because the price will drop. First a few traders sell, pushing the market imperceptibly lower. Then additional traders, sensing this microscopic downturn, may decide that now is the time

to sell and they too sell. Then an "avalanche" of selling begins, with traders all hoping to protect their profits by selling before the market drops [Sch04]. Thus, the set of N market participants "holding their position" are in this sense analogous to the set of N mostly North-pointing spins, bathed in a South-pointing magnetic field.

The above analogy may not be the best and it will be the challenge of future research to find a coherently convincing explanation for why the end of a microtrend or macrotrend displays such striking parallels to phase transitions. In any case, the set of interacting spins surely is analogous to the set of interacting traders.

The end of a negative microtrend or macrotrend is the same mechanism but with everything reversed. The N Ising spins point mostly south, the magnetic field is oriented north, and the spins flip from south to north one by one and they result in an avalanche corresponding to the spinodal singularity. Analogously, the N traders begin to suspect that the market is becoming metastable, so they one by one start to buy and as all the traders observe increasing prices, they jump in to buy before the price becomes too high [SH05b].

In summary we have seen that each trend—microtrend and macrotrend—in a financial market starts and ends with a unique switching process, and each extremum shares properties of macroscopic cooperative behavior. We have seen that the mechanism of bubble formation and bubble bursting has no scale for time scales varying over 9 orders of magnitude down to the smallest possible time scale (the scale of single transactions measured in units of 10 ms). On large time scales, histograms of price returns provide the same scale-free behavior. Thus, the formation of positive and negative trends on all scales is a fundamental principle of trading, starting on the smallest possible time scale, which leads to the non-stationary nature of financial markets as well as to crash events on large time scales. Thus, the well-known catastrophic bubbles which have occurred on large time scales—such as the most recent financial crisis—may not be outliers but in fact single dramatic representatives caused by the scale-free behavior of the forming of increasing and decreasing trends on time scales from the very large down to the very small.

Chapter 7

Statistical Analysis of the Order Book Model

As seen in the previous chapters, financial market time series exhibit a very large number of complex statistical properties. In this chapter[1], we analyze and extend an agent-based Order Book Model—the basic version was introduced in the author's diploma thesis [Pre06] and in [PGPS06]—which can be used to obtain a "mechanistic understanding" of the price formation process and which leads to the so-called stylized facts observed in financial market data. This chapter provides a detailed analysis of this model and covers the components which lead to important properties such as non-Gaussian price change distributions and the fact that persistent price dynamics on intermediate time scales (Hurst exponent $H > 1/2$) can be identified and realized individually.

Section 7.1 presents the definition of the basic model. In section 7.2 the parameter space of this model is analyzed to identify meaningful regions of this space. In section 7.3 we then discuss several augmentations of the model leading to a non-trivial Hurst exponent and a non-Gaussian return distribution. Section 7.4 studies switching processes in the Order Book Model in the context of the previous chapter. Finally, section 7.5 provides conclusions.

7.1 Definition of the Order Book Model

In this section, the Order Book Model is defined in its basic form. Its development was inspired by the model for the continuous double auction introduced in [DFG$^+$03, SFGK03]. Our aim is to accurately reproduce the structure and the mechanisms of an order book for real financial markets, as shown in Fig. 7.1 (which is similar to the one already provided in chapter 2). In our simulations, we limit ourselves to only one order book in which one individual asset is traded. This asset can be, e.g., a share, a loan,

[1] Results of this chapter are published in [PGPS07].

or a derivative product. Of the various types of orders which can be found on real financial markets we use only the two most important types, namely limit orders and market orders. The Order Book Model contains two different types of agents, liquidity providers and liquidity takers, which differ in the types of orders they are permitted to submit.

On the one hand, N_A liquidity providers only submit limit orders. In the case of a limit sell order, an agent offers an asset for sale at a given limit price (or any better price). Analogously, a limit buy order indicates demand for buying a traded asset and will be executed at a given limit price or a lower price. Let p_a be the so-called best ask, which is the lowest price level at which at least one limit sell order in the order book exists, and analogously p_b the so-called best bid, being the highest price level for which at least one limit buy order is stored in the order book. In our model, limit orders are placed around the midpoint

$$p_m = \frac{p_a + p_b}{2} \tag{7.1}$$

with a rate α. That is, $\alpha \cdot N_A$ new limit orders are inserted into the order book per time step. We denote q_{provider} to be the probability with which a limit order to be placed is a limit buy order. Thus, with probability $1 - q_{\text{provider}}$, the limit order to be placed is a limit sell order. The liquidity provider (*market maker*) supplies the order book with liquidity in this way. The aim of these market participants is to use the non-zero spread $s = p_a - p_b$ to earn money: they intend to sell an asset at price p_a or higher and then to buy it back at price p_b or lower, thus having earned at least the spread s (if it has remained constant between the sale and the purchase of the asset). Of course, they can analogously try to make money the other way around by first buying at price p_b or lower and then selling at price p_a or higher. Like in real financial markets, we allow agents to sell assets even if they do not possess them and thus to perform *short sales*.

On the other hand, N_A liquidity takers only submit market orders at a rate μ. That is, $\mu \cdot N_A$ market orders are inserted per time step. (There are, thus, a total of $2N_A$ agents in the system.) A market order is immediately executed after arriving in the order book: a market sell order is executed at price p_b, a market buy order at price p_a. The market order of a liquidity taker is a market buy order with probability q_{taker} and a market sell order with probability $1 - q_{\text{taker}}$. In the first version of our Order Book Model, we simply use

$$q_{\text{provider}} = q_{\text{taker}} = \frac{1}{2}. \tag{7.2}$$

Thus, limit orders and market orders are produced symmetrically around the midpoint. Limit orders stored in the order book can expire or can be deleted. The removal of each limit order occurs with probability δ per time unit.

As there are overall $2N_A$ agents in the system, each Monte Carlo step (MCS) consists of $2N_A$ moves, in which one agent is randomly selected and can perform one action. If the agent is a liquidity provider, then this agent submits a limit order with probability α.

Subsequently, independent of whether it came to an order placement or not, each limit order of this liquidity provider is deleted with probability δ. On the other hand, if the randomly selected agent is a liquidity taker, then this agent places a market order with probability μ which is immediately executed.

In the investigations of the Order Book Model provided here, we only consider order volumes of 1 "i.e., only one asset can be offered or demanded with a single order). The matching algorithm works according to price-time priority (as described in chapter 2). That is, first those limit orders at the best ask and the best bid are executed. If there is more than one order at a given price level, then the orders are executed in the same chronology as they were inserted into the order book.

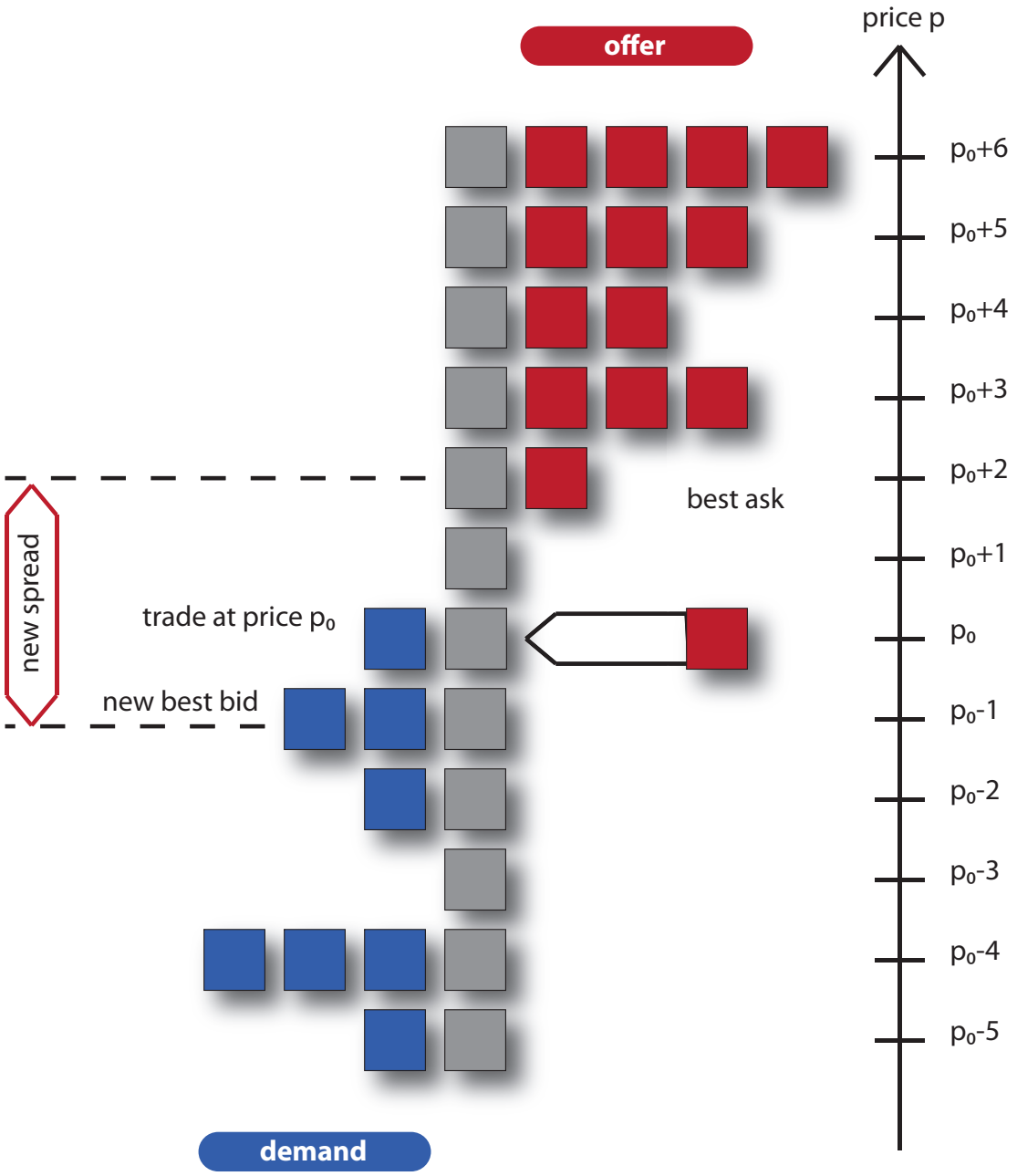

Figure 7.1: Structure of the order book: The limit buy (blue, $p \leq p_0$) and limit sell orders (red, $p > p_0$) are added chronologically to the appropriate discrete price level p. Here a matching algorithm with price time priority is implemented. In this example, a new sell order is placed at price p_0, at which there is already a buy order stored in the order book. The two orders at price p_0, which are on the demand and offer side of the order book, are executed against each other. Thus, a trade is performed at price p_0, which is then called last traded price. The spread s, which is the difference between the best ask and the best bid price, was two ticks before the arrival of the executable sell order at price level p_0 and increases to three ticks after this trade.

7.1.1 Liquidity Providers and Liquidity Takers

Based on this model definition, first an unrealistic independent, identically distributed (iid) order placement is assumed within two intervals (for the buy and for the sell orders), both of which having width p_{int}. Accordingly, every liquidity provider enters his/her limit buy orders to each price level within the interval $[p_a - 1 - p_{\text{int}}; p_a - 1]$ with the same probability. Also, limit sell orders are submitted as uniformly distributed on the interval $[p_b + 1; p_b + 1 + p_{\text{int}}]$. With these assumptions, we were able to reproduce the results of [DFG+03, SFGK03]. In [DFG+03, SFGK03], an appropriate microscopic-dynamic-statistical model for the continuous double auction is examined with analytic approximations under the assumption of an iid order flow and the limit $p_{\text{int}} \to \infty$.

Even with this comparatively simple realization of the Order Book Model, the profit loss distribution of the agents can be analyzed. In the Order Book Model, the main distinction between the agents is that they are either liquidity providers or liquidity takers, which is reflected in the types of orders they are permitted to submit. We first thus intend to investigate how the type of a trader influences the temporal development of his/her account balance.

Let $\kappa_i(t)$ be the account balance of agent i at time t. Each agent i possesses no money at the beginning of the simulation at $t = t_0$, such that $\kappa_i(t_0) = 0$ applies to $i = 1, 2, \ldots, 2N_A$. In order not to restrict their trades, we provide an unlimited credit line to each agent, free of charge. Furthermore, each agent buys and sells assets over time. The number of assets which agent i possesses at time t is given by $\pi_i(t)$. Here we also set $\pi_i(t_0) = 0$ for all $1 \leq i \leq 2N_A$, such that each agent possesses zero assets at the beginning of the simulation. Note that we do allow negative values of $\pi_i(t)$, as an agent can sell an asset he/she does not possess.

Of course, when buying an asset, agent i has to pay the price of this asset, such that $\pi_i(t)$ is incremented by 1 with this trade at time t, but $\kappa_i(t)$ is decreased by $p(t)$, with $p(t)$ being the transaction price of this trade, which then becomes the last traded price. Analogously, when selling an asset, $\pi_i(t)$ is decremented by 1 and $\kappa_i(t)$ is increased by $p(t)$.

The overall wealth $\gamma_i(t)$ of agent i at time t thus consists both of the account balance $\kappa_i(t)$ and the number of assets $\pi_i(t)$:

$$\gamma_i(t) = \kappa_i(t) + \pi_i(t)p(t) \tag{7.3}$$

The change of the wealth of agent i between time t_0 and t is given by

$$\Delta\gamma_i = \gamma_i(t) - \gamma_i(t_0) = \gamma_i(t) \tag{7.4}$$

as $\gamma_i(t_0) = 0$.

When simulating this Order Book Model with Monte Carlo techniques, we always observe significant difference between the wealth distributions of the liquidity providers and of the liquidity takers. The liquidity takers are systematically disadvantaged. Although it is possible that some liquidity takers obtain a positive trading result, a separation of the two groups nonetheless arises, because liquidity takers have to pay the

spread $s = p_a - p_b$ to liquidity providers if opening or closing a position in the traded asset (i.e., if either first buying and then reselling or if first selling and then re-buying an asset).

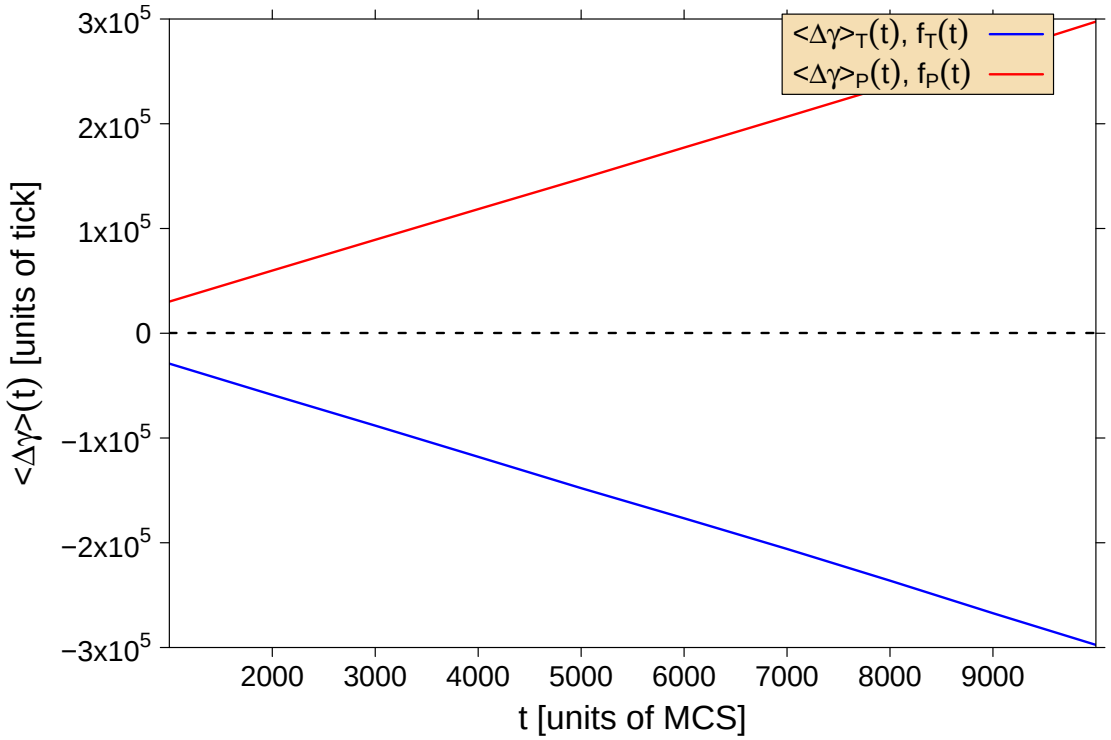

Figure 7.2: Liquidity providers vs. liquidity takers: Averaged wealth shift of the liquidity providers and the liquidity takers, rsp., as a function of t for $\alpha = 0.15$, $\mu = 0.1$, $\delta = 0.025$, and $N_A = 250$. These dependencies can be fitted with the functions $f_{\mathrm{T}}(t) = a_{\mathrm{T}} \cdot t$ and $f_{\mathrm{P}}(t) = a_{\mathrm{P}} \cdot t$ with the parameters $a_{\mathrm{T}} \approx -29.45$ and $a_{\mathrm{P}} \approx 29.75$. The difference between the absolute values of these two gradients is due to statistical inaccuracies occurring during the determination of the averages of the distributions.

The wealth values for liquidity takers and liquidity providers drift apart linearly in time as shown in Fig. 7.2. The identical absolute values of the gradients a_{P} and a_{T} are related to the spread s: an individual liquidity taker loses on average $|a_{\mathrm{T}}|$ ticks per MCS. This is the average gain achieved by a liquidity provider per MCS, whose wealth is increased on average by $|a_{\mathrm{P}}|$ ticks per MCS. Thus, on average $N_A \cdot |a_{\mathrm{T}}|$ ticks are transferred from liquidity takers to liquidity providers per MCS. Since $N_A \cdot \mu$ market orders are submitted to the order book and therefore $N_A \cdot \mu$ trades take place, the average wealth transfer per transaction is given by

$$\Gamma = \frac{N_A \cdot \langle |a_i| \rangle}{N_A \cdot \mu} = \frac{\langle |a_i| \rangle}{\mu} \tag{7.5}$$

with $\langle |a_i| \rangle = (|a_{\mathrm{P}}| + |a_{\mathrm{T}}|)/2$ being the averaged absolute value of the gradients. For the

parameters $\alpha = 0.15$, $\mu = 0.1$, $\delta = 0.025$, $p_{\text{int}} = 2000$, and $N_A = 250$ used in the simulations for the results shown in Fig. 7.2, the averaged wealth transfer per transaction can be determined to $\Gamma = 296$ ticks. This is approximately half of the averaged spread s, which was determined in the simulation to be $\langle s \rangle = 606$ ticks:

$$\Gamma = \frac{\langle s \rangle}{2} \tag{7.6}$$

The factor $1/2$ is to be attributed to the fact that a liquidity provider needs two transactions to earn the complete spread s—the liquidity provider has to buy once and sell once. Therefore the agent earns only $\langle s \rangle / 2$ per transaction on average.

If comparing these results with the situation in real markets, it must be mentioned that liquidity takers are in fact disadvantaged financially as compared to the group of liquidity providers. Our distinction in the order book model between liquidity providers and liquidity takers reflects the two different types of orders which are used—limit orders and market orders. In real financial markets, there is no strict distinction between these two groups of traders. In general, each market participant can submit both types of orders to the electronic order book, so that the idealized situation of Fig. 7.2 would be difficult to verify in real markets by looking at the wealth evolution of individuals.

7.1.2 Exponential Order Placement Depth

An iid order placement depth, such as that considered in the previous section, is not in agreement with the conditions found in real financial markets. In contrast to the uniform cumulative order volume generated by the iid order flow of [DFG+03, SFGK03], the order book depth of real markets can be described by a log-normal distribution [MM01].

To take this finding into account, we replace the iid limit order placement on the fixed interval p_{int} around the midpoint p_m with an exponentially distributed order placement depth. For placing a limit order i, the limit price p_i^l is determined for a limit buy order by

$$p_i^l = p_a - 1 - \eta \tag{7.7}$$

and for a limit sell order according to

$$p_i^l = p_b + 1 + \eta \tag{7.8}$$

whereby η is an exponentially distributed integer random number created by

$$\eta = \lfloor -\lambda_0 \cdot \ln(x) \rfloor \tag{7.9}$$

with x a uniformly distributed random number on the interval $[0;1)$ and $\lfloor z \rfloor$ denote the integer part of z. Thus, the submission of limit orders minimizes a potentially existing

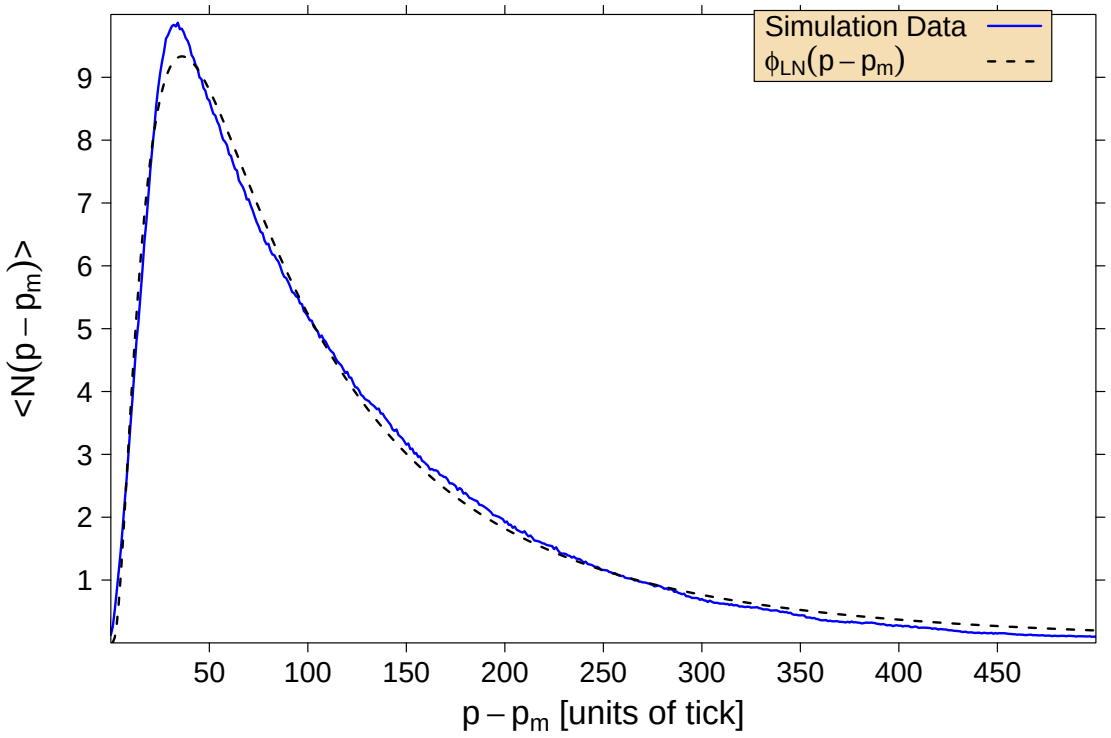

Figure 7.3: Exponential order placement depth: With the parameter selection $\alpha = 0.15$, $\mu = 0.025$, $\delta = 0.025$, $\lambda_0 = 100$, $N_A = 250$, and $q_{\text{provider}} = q_{\text{taker}} = 0.5$, one gets the order book depth shown here. The results were averaged over 10^4 MCS. A fit to the log-normal distribution function $\phi_{\text{LN}}(p - p_m) = a/x \cdot \exp\{-[\ln(p - p_m) - b]^2/c\}$ leads to the parameters $a = 527 \pm 1$, $b = 4.483 \pm 0.004$, and $c = 1.79 \pm 0.01$.

spread. Also the situation in which a limit order becomes instantaneously executable at p_a or p_b is avoided, because such a limit order degenerates into market order.

Using this exponentially distributed order placement depth, a log-normally distributed order book depth is achieved, an example of which is shown in Fig. 7.3. This implementation of the Order Book Model is from now on referred to as the basic version of the Order Book Model, for which we will show the most important time series characteristics. In Fig. 7.4, first exemplary price subsequences are shown. The autocorrelation of the corresponding time series $\delta p(t) = p(t+1) - p(t)$ of the price changes shows the same behavior as can be observed in real financial market data. A significant negative autocorrelation exists for a time lag $\Delta t = 1$. Thus, a positive price change is followed by a negative price change with large probability and vice versa. This autocorrelation vanishes for $\Delta t > 1$ [Pre06].

Figure 7.5 shows the Hurst exponent $H(\Delta t)$ for various agent numbers N_A. In general, the Hurst exponent $H(q)$ is calculated via the relationship

$$\langle |p(t + \Delta t) - p(t)|^q \rangle^{1/q} \propto \Delta t^{H(q)}, \tag{7.10}$$

as defined already in chapter 3 and for example in [DW00]. Unless mentioned oth-

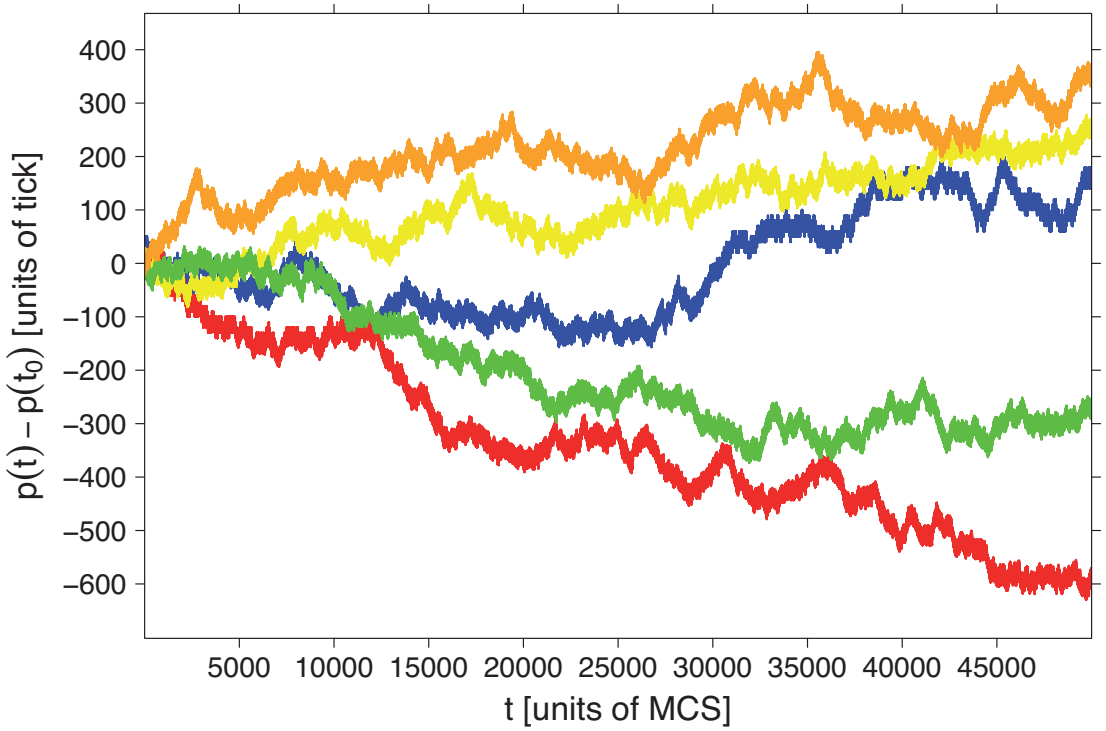

Figure 7.4: Exponential order placement depth: For $\alpha = 0.15$, $\mu = 0.025$, $\delta = 0.025$, $\lambda_0 = 100$, $N_A = 250$, and $q_{\text{provider}} = q_{\text{taker}} = 0.5$, exemplary price subsequences of 50,000 MCS are shown.

erwise, we use $H(\Delta t) \equiv H(\Delta t, q = 2)$ in this chapter. For comparison, the constant Hurst exponent $H = 1/2$ of the random walk is given in Fig. 7.5. On short time scales, the price process indicates an anti-persistent behavior, which is due to the order book structure. On long time scales, the process converges towards a diffusive regime.

This anti-persistent price behavior on short time scales can be found in actual financial time series [PGPS07] and is a consequence of the "mechanics" of the order book. Given a constant order influx, an executed market order of any kind automatically increases the probability that the next transaction price will be negatively correlated with the previous one.

The price change distributions shown in Fig. 7.6 exhibit no fat tails, but can be rather well approximated by a Gaussian distribution. Deviations from the Gaussian distribution are found for large price changes, where the Gaussian distribution overestimates the probability for these price changes. Note that the introduction of the exponentially distributed order placement depth influences only the order book depth, but the Hurst exponent and the price change distributions are not qualitatively influenced in comparison with the results for the iid order placement approach.

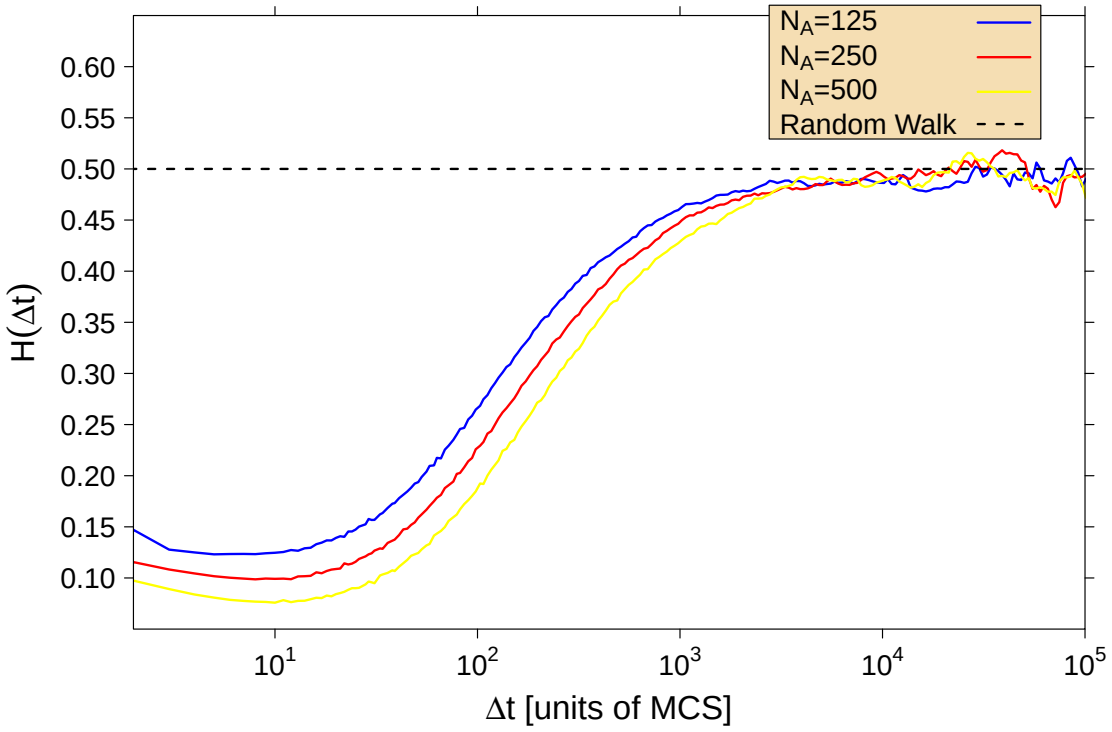

Figure 7.5: Exponential order placement depth: For the same parameters as in Fig. 7.4, the Hurst exponent is shown here for different numbers of agents N_A, averaged over 50 simulation runs each.

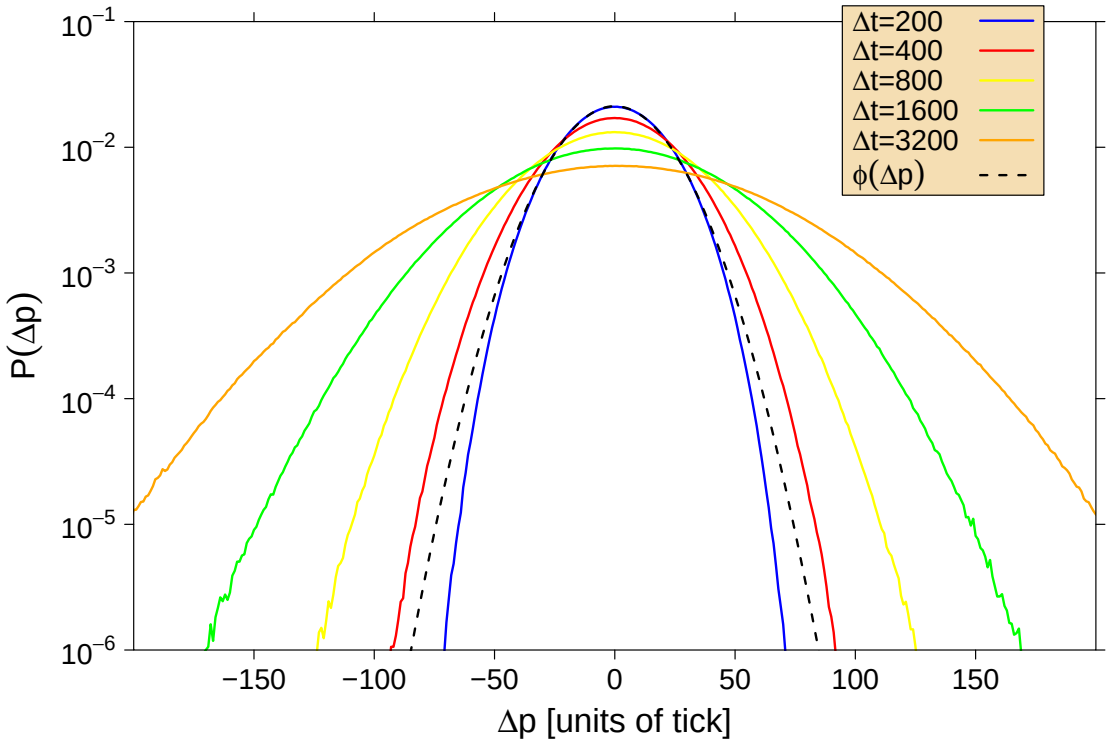

Figure 7.6: Exponential order placement depth: distributions of price changes for various Δt for the same parameter values as in Fig. 7.4. Again the results were averaged over 50 simulation runs each. For $\Delta t = 200$, the fit parameters for the Gaussian distribution function $\phi(\Delta p) = a \cdot \exp(-b \cdot \Delta p^2)$ are given by $a = 2.122 \times 10^{-2} \pm 4 \times 10^{-5}$ and $b = 1.388 \times 10^{-3} \pm 10^{-6}$.

7.2 Parameter Space

In order to obtain reasonable information concerning the parameter selection for future extensions of this Order Book Model, the parameter space of this basic version of the Order Book Model has to be analyzed in detail. As the order book depth has to be considered a vital criterion for the stability of the order book in the Monte Carlo simulations, it is quantitatively examined as a function of the two most important parameters λ_0 and μ. For the other parameters, we use the values $N_A = 500$, $\alpha = 0.15$, and $\delta = 0.025$ within this section. After a transient time of a few thousand MCS, the order book depth is stationary and log-normally distributed according to

$$P_{\mathrm{LN}}(x) = A \frac{1}{Sx\sqrt{2\pi}} \exp\left(-\frac{(\ln x - M)^2}{2S^2}\right) \tag{7.11}$$

with the parameters A, S^2, and M. In our simulation runs, we wait 22,500 MCS, as the order book reaches equilibrium within a few thousand MCS, and then take the average over the subsequent 2,500 MCS.

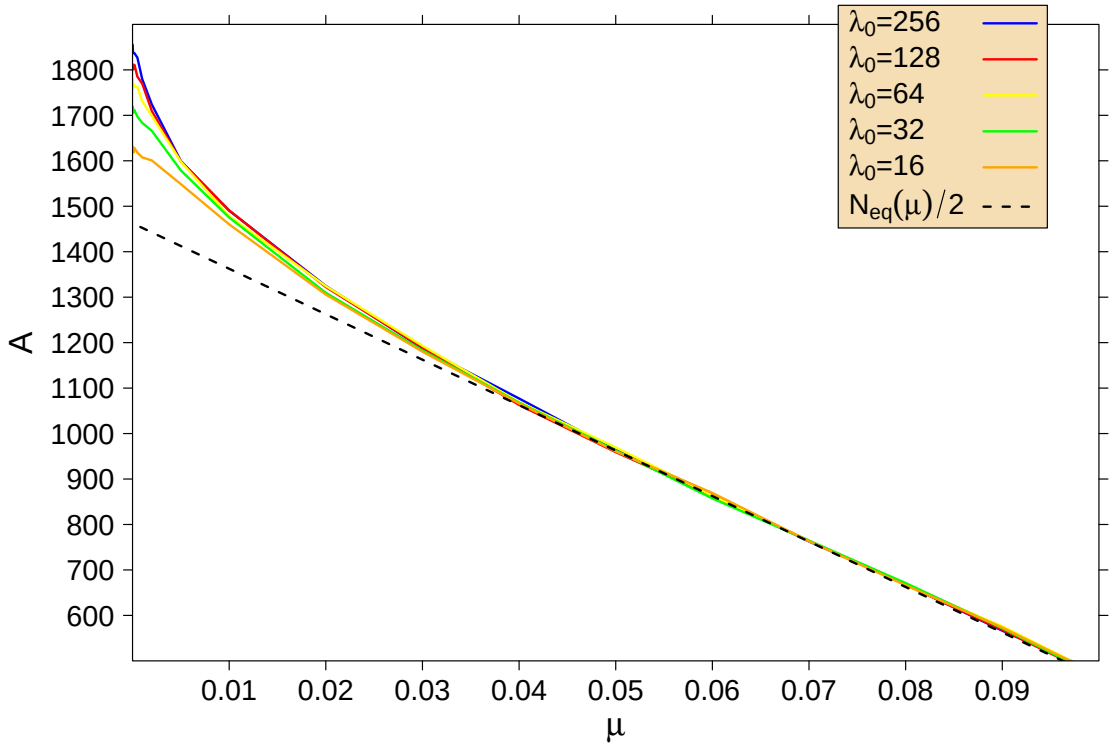

Figure 7.7: Parameter space: Fit parameter A as a function of the market order rate μ for different values of the order placement depth λ_0.

In Fig. 7.7, the influence of the parameter μ on the scaling factor A is illustrated for various values of λ_0. A is the factor in Eq. (7.11) which determines the size of the area

under the curve of the log-normal distribution, as the integral over Eq. (7.11) gives the value A. We find that A decreases linearly for medium and large values of μ independent of the value of λ_0. Only for small values of μ can deviations from this linear behavior be found. This result can be analytically explained if we consider the order book having already reached its equilibrium: in its stationary state, the order book depths on the bid and on the ask side are identical. Thus, the scaling factor A, as it is the area under the curve of the log-normal distribution, corresponds to half the total number of orders stored in the order book. The total number of limit orders at time $t+1$ can be described recursively by the order rates α, δ, μ, the number of agents N_A, and the number $N(t)$ of limit orders at time t by

$$N(t+1) = N(t) + \alpha N_A - (N(t) + \alpha N_A)\,\delta - \mu N_A. \tag{7.12}$$

In equilibrium we have

$$\frac{N_{\text{eq}}}{N_A} = \alpha\left(\frac{1}{\delta} - 1\right) - \frac{\mu}{\delta} \tag{7.13}$$

for the number of limit orders per liquidity provider. Defining an effective limit order rate $\alpha^* = \alpha\,(1-\delta)$, this relation results in

$$\frac{N_{\text{eq}}}{N_A}\,\delta = \alpha^* - \mu. \tag{7.14}$$

A stable order book is therefore achieved on average if the conditions $\alpha^* > \mu$ and $\delta > 0$ are fulfilled.

Inserting the values we used in the simulations in this equation, we obtain $N_{\text{eq}}/2 = 1462.5 - 10^4\mu$. This relation is drawn as a dashed line in Fig. 7.7. We find that the theoretical conclusions and the simulation results coincide for medium and large values of μ. The deviations for small μ result from the fact that the order book depth no longer approaches a log-normal distribution in this case, such that a fit to the log-normal distribution function is no longer valid.

As shown in Fig. 7.8, the center point M of the log-normal distribution is also affected by a modification of the parameters λ_0 and μ. The larger the market order rate μ, the more limit orders are removed from the bid and the ask side of the order book by transactions. The more limit orders are removed from the inside of the distributions at the bid and the ask side, however, the more M deviates from the midpoint p_m. As can be clearly seen in Fig. 7.8, this behavior only occurs for market order rates $\mu \gtrsim 10^{-2}$. An increase in the exponentially distributed order placement depth leads to a larger value of M.

For completeness, the squared variance S^2 of the log-normal distribution is shown in Fig. 7.9 as a function of μ for different values of λ_0. The parameter S^2 decreases with increasing μ, this decrease is rather independent of the value of λ_0 for large values of μ.

Significant deviations from the log-normal behavior discussed so far are found for parameter combinations with small values for either λ_0 or μ: for very small μ, the distribution of the order book depth resembles the exponentially distributed order placement

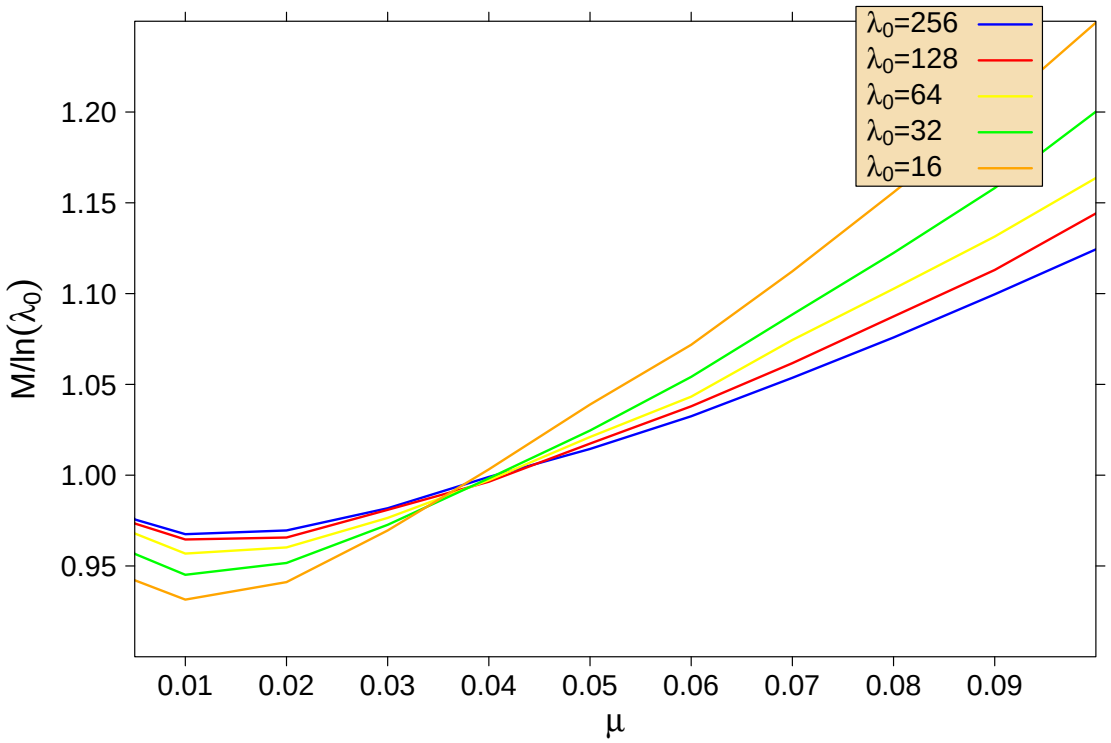

Figure 7.8: Parameter space: Fit parameter M as a function of the market order rate μ for different values of the order placement depth λ_0.

depth, as the market orders are not sufficient to remove the limit orders at the bounded random walk and again keep $q_{\text{provider}} = 1/2$ constant best ask and best bid. For small λ_0, limit orders are placed closely around the midpoint, and there are therefore such a large number of limit orders at the best bid and best ask that there is a limiting value for μ (depending on λ_0) under which the market orders are not able to shift p_b and p_a anymore. In both cases, when considering the extreme situations of vanishing μ and λ_0, the order book freezes in a way such that p_a and p_b become constant and $s = p_a - p_b = 1$. The price then jumps between p_a and p_b.

Another limiting situation is given by large market order rates μ. The more market orders are submitted, the larger becomes the probability that one side of the order book is completely cleared and trading stops. The total number of limit orders $N(t)$, which are stored in the order book at time t, is crucial for the stability of the simulation. Independent of the value chosen for μ, fluctuations in the selection of agents can empty the bid or ask side of the order book, if the order book depth is too small. This can result in a price crash or a price explosion. Especially, at the beginning of the simulation, limit orders have to be placed in a pre-opening phase which we choose to last 10 MCS.

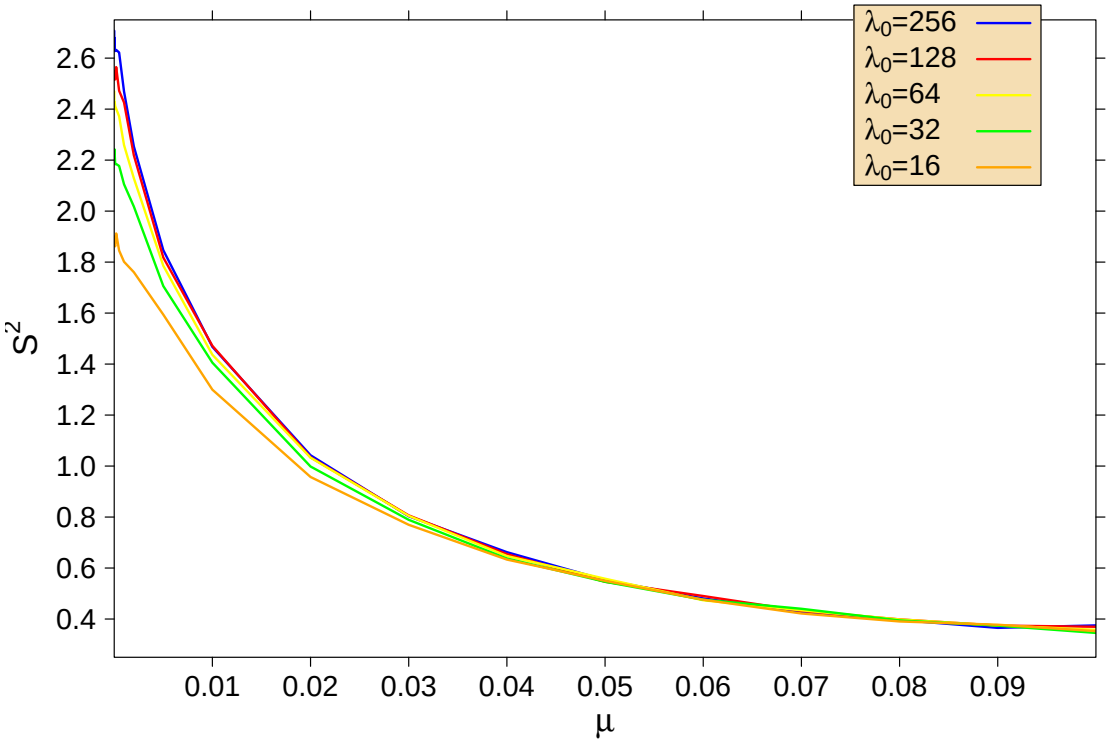

Figure 7.9: Parameter space: Fit parameter S^2 as a function of the market order rate μ for different values of the order placement depth λ_0.

7.3 Augmentation of the Model

7.3.1 Deterministic Perturbation

Following investigation of the parameter space of the basic version of the Order Book Model, extensions of this model are now considered. The basic model is only able to reproduce the anti-persistent price behavior on short time scales as well as the diffusive price behavior on long time scales, both of which are also found on real financial markets. However, the persistent behavior on medium time scales is not reproduced in this basic version which models a stationary market. Moreover, there are no fat tailed price change distributions found in the basic variant. So far, we have always considered symmetric buy and the sell probabilities, both for the group of liquidity providers and for the group of liquidity takers, as $q_{\text{provider}} = q_{\text{taker}} \equiv 1/2$. This is an appropriate assumption for presuming a stationary behavior of a financial market. Such stationary behavior is, however, not compatible with financial-economic conditions. Real order rates exhibit asymmetries. In a bull market, an increased buy probability can be found and one can measure an increased sell probability in a bear market. This applies not only for long time movements, but also for short time trends on intraday time scales. Therefore, the symmetry $q_{\text{provider}} = 1 - q_{\text{provider}} = 1/2$ and $q_{\text{taker}} = 1 - q_{\text{taker}} = 1/2$ used so far will now be broken such that q_{provider} stays at its constant value of $1/2$, whereas q_{taker} now changes with time while still having an average value of $1/2$. One could say that the market is oscillates around its stationary state through the modulation of q_{taker}.

For the practical realization of such a market oscillation, a deterministic symmetry disturbance is first investigated. As a simple ansatz, q_{taker} shall be varied via a saw tooth modulation

$$q_{\text{taker}} = \begin{cases} \frac{1}{2} + t\Delta S & \text{for } 0 \leq t \leq S/\Delta S \\ \frac{1}{2} + 2S - t\Delta S & \text{for } S/\Delta S \leq t \leq 3S/\Delta S \\ \frac{1}{2} - 4S + t\Delta S & \text{for } 3S/\Delta S \leq t \leq 4S/\Delta S \end{cases}, \tag{7.15}$$

which is periodically repeated (with amplitude value $S = 1/20$ and the step size $\Delta S = 1/1000$)and by which the variable q_{taker} is changed after each MCS time step denoted by t. This saw tooth modulation has a period of $4S/\Delta S$ and thus q_{taker} returns to the value of $1/2$ after every $t_r = 2S/\Delta S = 100$ MCS. We choose a saw tooth modulation because it exhibits a constant residence distribution.

Using such a saw tooth modulation for the implementation of an asymmetry in the order flow, price time series with new properties are produced. An averaged Hurst exponent is shown for example in Fig. 7.10. Each simulation lasted 10^6 MCS, and the average was taken of 50 simulation runs. The deterministic modulation of q_{taker} with a period of 200 MCS is reflected in the mean square displacement, leading to quasi-periodic oscillations of the Hurst exponent $H(\Delta t)$ at medium time scales. The Hurst exponent oscillates a few times with a period of approximately 200 MCS, before showing a diffusive behavior for the price development on long time scales. The more agents

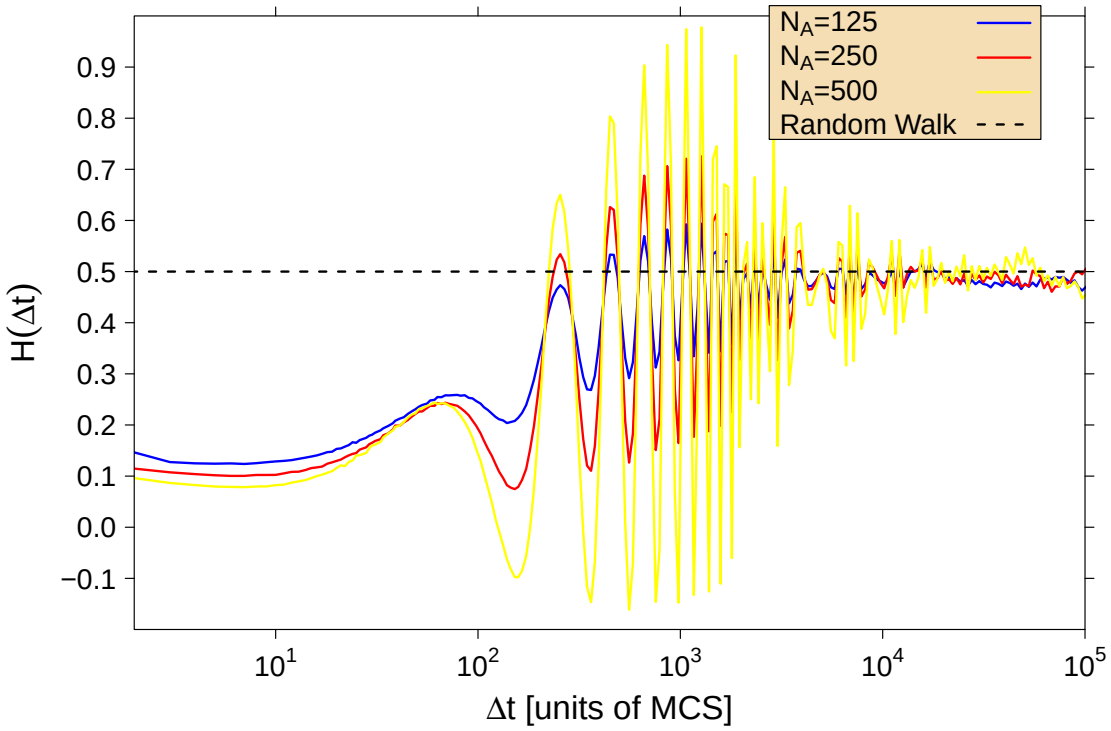

Figure 7.10: Deterministic perturbation realized by a sawtooth modulation of q_{taker}: Hurst exponent $H(\Delta t)$ for various agent numbers N_A as a function of Δt. The parameter values used are $\alpha = 0.15$, $\mu = 0.025$, $\delta = 0.025$, and $\lambda_0 = 100$.

are trading in the order book, the larger are the amplitudes of the oscillations. Such periodic oscillations of the Hurst exponent have nothing in common with the behavior the Hurst exponent displays for data from real financial markets.

Figure 7.11 shows the corresponding distributions of the price changes. As in the basic version of the Order Book Model with symmetric order flow, no fat tailed price change distributions are found here. A fit to a Gaussian distribution is appropriate here as well, except that it again overestimates the probability of large price changes. In conclusion, we can state that using a deterministic perturbation was not successful in producing a more realistic price behavior, but we were able to change the behavior of the Hurst exponent, especially on medium time scales with this periodic modulation.

7.3.2 Stochastic Perturbations

As any deterministic perturbation with a discrete return time spectrum will be reflected in oscillations of the Hurst exponent, we focus now on stochastic perturbations with continuous return time distributions. As a straightforward approach, we make the value of q_{taker} follow a bounded random walk with average value $1/2$ and again keep $q_{\text{provider}} = 1/2$ constant. This approach was already introduced in [PGPS06]. In this case, one obtains an anti-persistent price behavior on short time scales, a persistent price

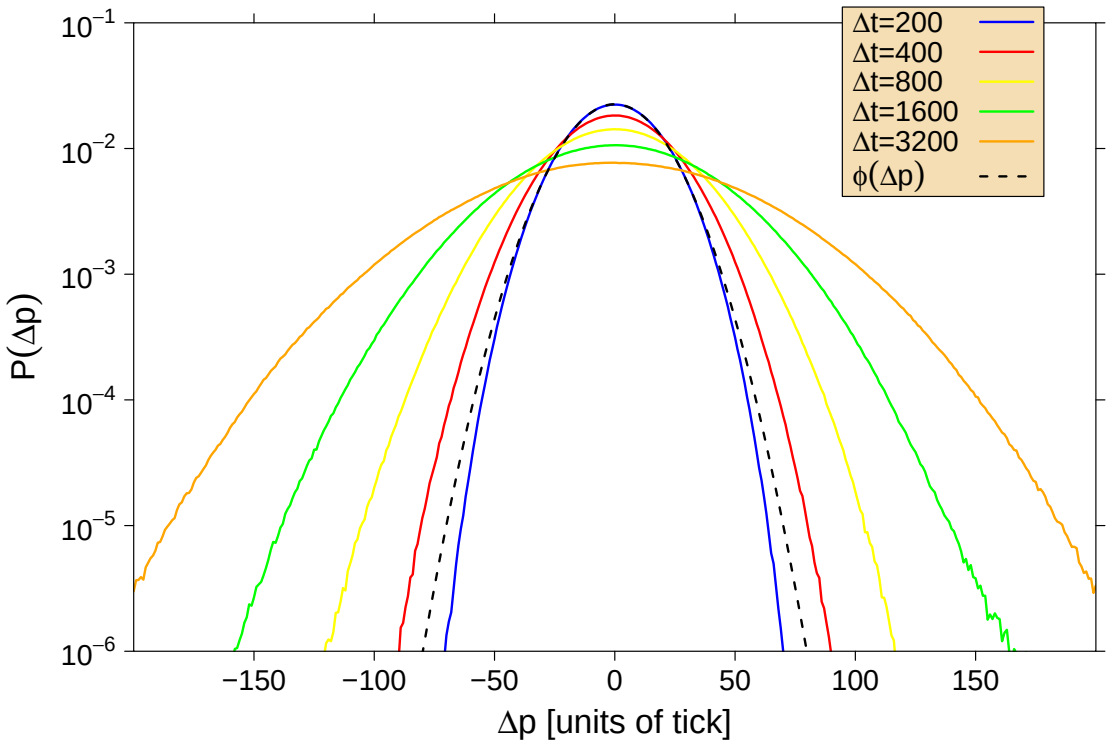

Figure 7.11: Deterministic perturbation realized by a sawtooth modulation of q_{taker}: Price change distributions for various values of Δt and for $N_A = 500$. For $\Delta t = 200$, the function $\phi(\Delta p) = a \cdot \exp(-b \cdot \Delta p^2)$ can be numerically fitted with $a = 2.251 \times 10^{-2} \pm 3 \times 10^{-5}$ and $b = 1.573 \times 10^{-3} \pm 5 \times 10^{-6}$.

behavior on medium time scales, and a diffusive behavior on long time scales. However, the maximum values of $H(\Delta t)$ found for the medium time scales are too large (up to 0.9) but according to [Man63], only a maximum value of ≈ 0.6 may be found. This is also the case for other financial time series, see for example, the results in [Aus00] for foreign exchange time series. The price change distributions of the bounded random walk approach exhibit a bimodal shape, which is in contrast to the behavior of price changes in time series of real financial markets. This wrong behavior is caused by the constant residence distribution of a bounded random walk. As q_{taker} returns slowly from the extreme areas of the modulation to the mean value of $1/2$, strong trend phases are created, which lead in turn to the bimodal shape.

Based on these insights, applying a feedback random walk (i.e., a random walk with increased probability for returning to the mean value) is a straightforward next step. This approach is able to produce a nontrivial Hurst exponent, which is comparable to those of time series found in real financial markets. This approach generates an almost Gaussian price change distribution.

The feedback random walk, which is again only applied to q_{taker}, whereas q_{provider} stays constant at $1/2$, functions as follows: at the beginning of the Monte Carlo simula-

tion, q_{taker} starts at the mean value of $1/2$. The variable q_{taker} is incremented and decremented by a value of ΔS after each MCS. But in contrast to a standard random walk, the probability for returning to the average value of $1/2$ is given by $1/2 + |q_{\text{taker}}(t) - 1/2|$ and thus the probability for departing from the mean value is given by $1/2 - |q_{\text{taker}}(t) - 1/2|$. This feedback random walk tends to return to its average value $\langle q_{\text{taker}} \rangle = 1/2$ more often compared to the bounded random walk: the expectation value of $q_{\text{taker}}(t+1)$ is given by

$$\langle q_{\text{taker}}(t+1) \rangle = (q_{\text{taker}}(t) - \Delta S) q_{\text{taker}}(t) + (q_{\text{taker}}(t) + \Delta S)(1 - q_{\text{taker}}(t)) \tag{7.16}$$

both for $q_{\text{taker}}(t) \geq 1/2$ and for $q_{\text{taker}}(t) \leq 1/2$. Thus, we generally have

$$\langle q_{\text{taker}}(t+1) \rangle - q_{\text{taker}}(t) = \Delta S(1 - 2q_{\text{taker}}(t)) \begin{cases} > 0 & \text{if } q_{\text{taker}}(t) < 1/2 \\ < 0 & \text{if } q_{\text{taker}}(t) > 1/2 \end{cases} \tag{7.17}$$

such that this feedback random walk on average approaches its expectation value of $1/2$. The stochastic process of the feedback random walk is characterized by a continuous return time spectrum, qualitatively comparable to that of the bounded random walk. However, the residence distribution of the probability q_{taker} exhibits an almost Gaussian shape in contrast to the bounded random walk, for which it is uniform. For the simulation results shown here, we again use $\Delta S = 1/1000$.

Figure 7.12 shows the behavior of the Hurst exponent $H(\Delta t)$, which was averaged over 50 simulation runs lasting 10^6 MCS each. Again one finds an anti-persistent behavior on short time scales, a persistent behavior on medium time scales, and a diffusive regime on long time scales. The maximum of the Hurst exponent increases with increasing agent number N_A. When comparing the results shown here with measurements of the Hurst exponent of financial time series achieved in real financial markets, we find that agent numbers in the range $150 \leq N_A \leq 500$ are best able to reproduce a realistic maximum value of the Hurst exponent.

In Fig. 7.13, price change distributions for this approach are shown. In contrast to the bounded random walk approach, here more reasonable, almost Gaussian shaped price change distributions are obtained.

This extension to the Order Book Model also fails, however, to produce fat tailed price change distributions. Note that identical results can be achieved for a temporal modulation of q_{provider} and a constant $q_{\text{taker}} = 1/2$ and if q_{provider} and q_{taker} are changed in time independently of each other through feedback random walks. After the next section, in which the relationship between the Hurst exponent and the autocorrelation is analyzed, a further extension will be introduced that results in fat tails.

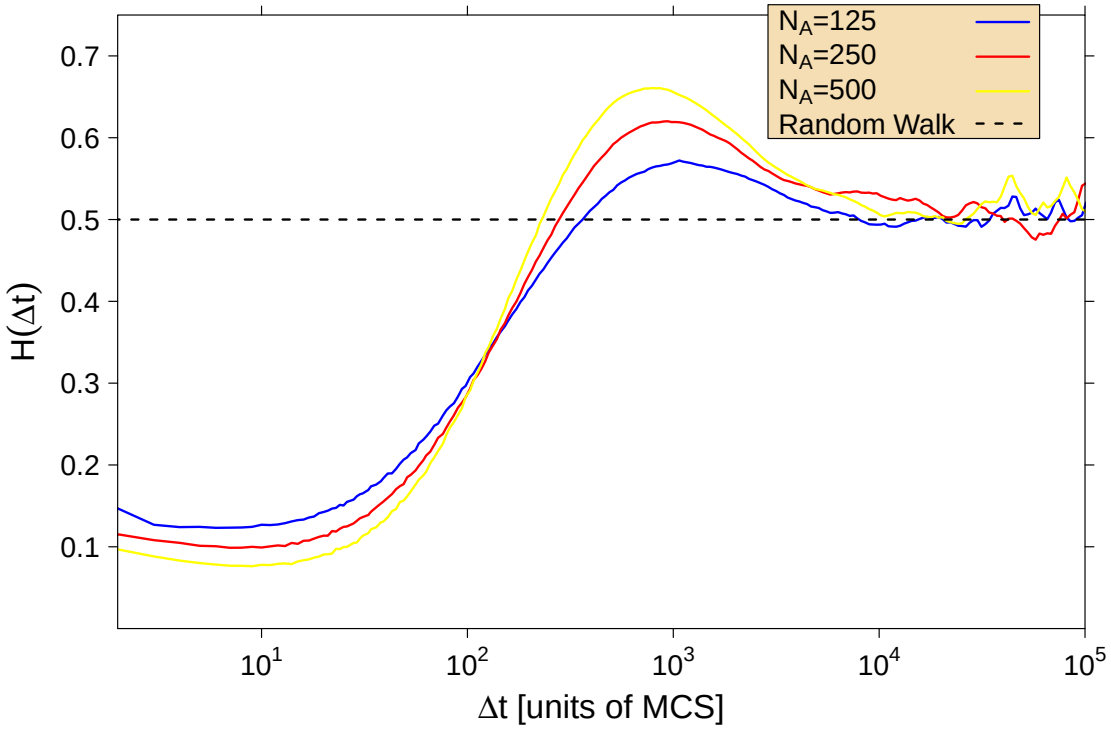

Figure 7.12: Stochastic perturbation realized by the feedback random walk modulation of q_{taker}: Hurst exponent $H(\Delta t)$ for various agent numbers N_A as a function of Δt. The parameters used are $\alpha = 0.15$, $\mu = 0.025$, $\delta = 0.025$, and $\lambda_0 = 100$.

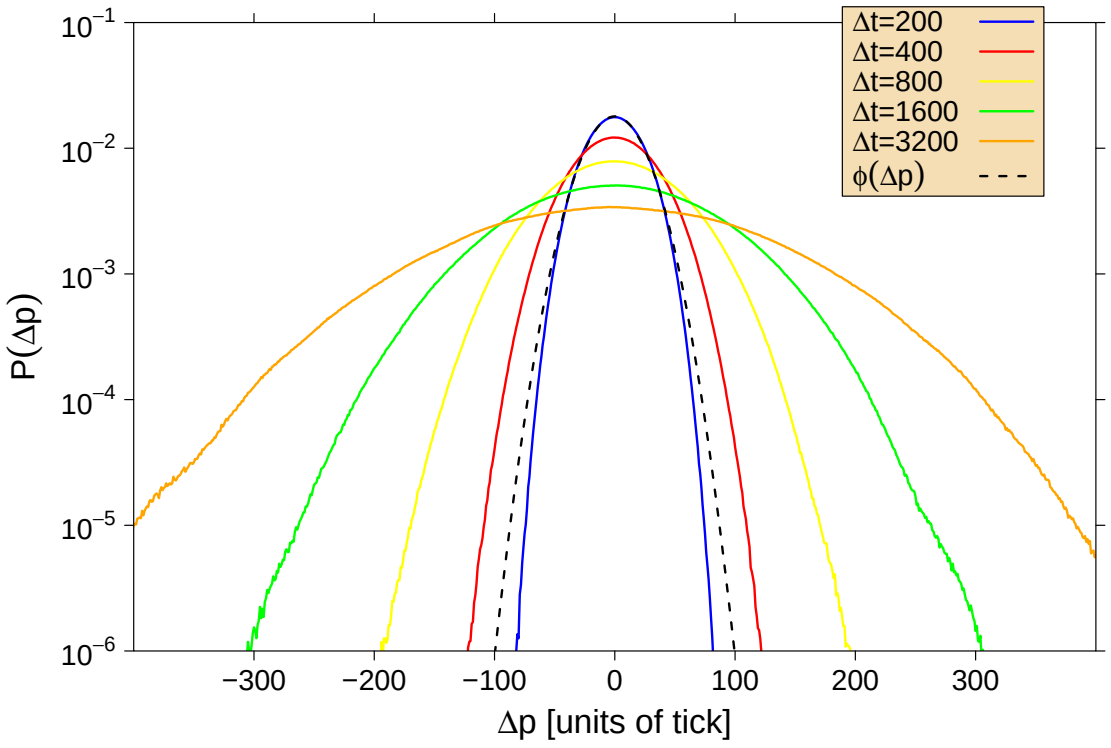

Figure 7.13: Stochastic perturbation realized by the feedback random walk modulation of q_{taker}: price change distributions for various values of Δt and for $N_A = 500$. The other parameters are the same as in Fig. 7.12. Fitting the function $\phi(\Delta p) = a \cdot \exp(-b \cdot \Delta p^2)$ to the data for $\Delta t = 200$, one gets $a = 1.79 \times 10^{-2} \pm 10^{-4}$ and $b = 9.9 \times 10^{-4} \pm 10^{-5}$.

7.3.3 Hurst Exponent and Autocorrelation

The Hurst exponent is often used for the characterization of stochastic processes. Often a connection to autocorrelations is drawn, which describe memory effects within stochastic processes. In the literature, it is widely assumed that a Hurst exponent of $H \neq 1/2$ implies long-time correlations, but recent theoretical work [BGM06, MGB07] shows that this is not necessarily true. A persistent behavior with $H = 1/2$ also occurs for Markov processes (i.e., processes without memory) in the case of non-stationary increments. How to reproduce this behavior was shown in the previous section. This result also affects the interpretation of the Hurst exponent of financial market time series. In this context, the Hurst exponent is used in order to measure the efficiency of a market. The Hurst exponent $H = 1/2$ of the random walk corresponds to an efficient market. However, this criterion alone is not sufficient for determination of efficiency according to the results of [BGM06, MGB07]. From a measurement of the Hurst exponent only, the existence of a long memory cannot be derived nor can the existence of an efficient financial market be deduced. Instead, an additional investigation of autocorrelations is necessary.

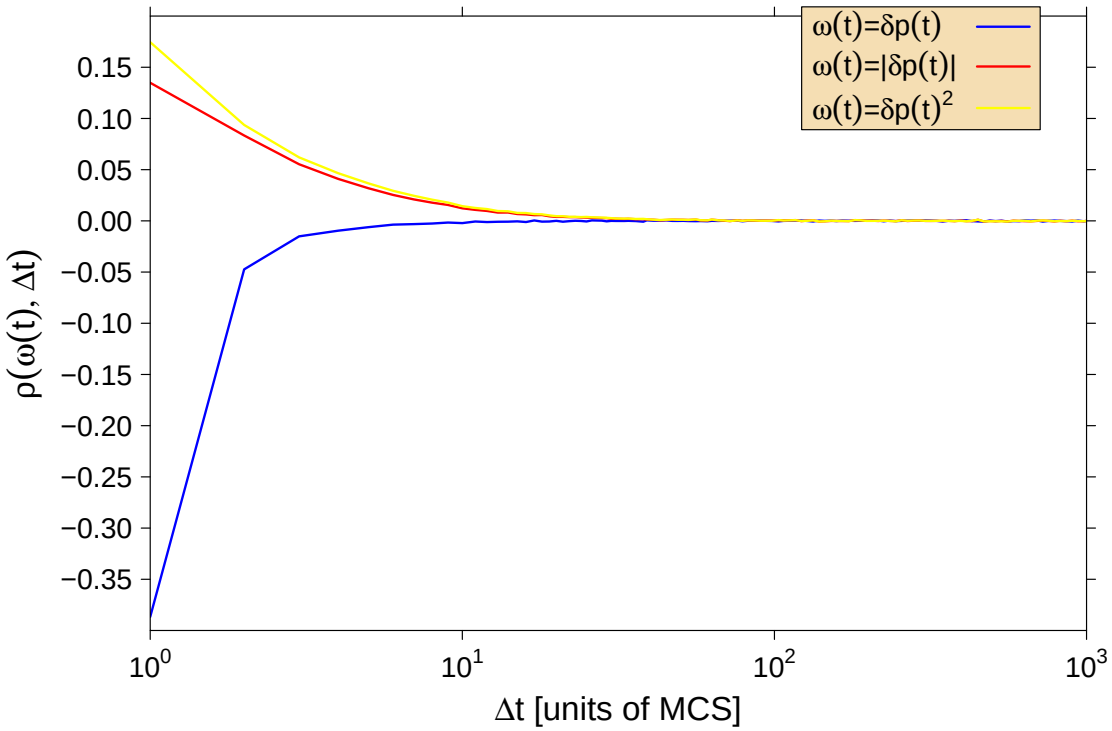

Figure 7.14: Autocorrelation $\varrho(\delta p(t), \Delta t)$ of price changes, autocorrelation $\varrho(|\delta p(t)|, \Delta t)$ of absolute price changes, and autocorrelation $\varrho(\delta p(t)^2, \Delta t)$ of squared price changes with $\delta p(t) = p(t+1) - p(t)$. The parameters are the same as for the results in Fig. 7.12 with $N_A = 125$.

Figure 7.14 shows the autocorrelation of the price change, of the absolute price

change, and of the quadratic price change time series created by the same parameter set as was used for the corresponding Hurst exponent shown in Fig. 7.12 for $N_A = 125$, but instead of averaging over 50 simulations, here only one simulation was performed whose calculation time was increased to 10^7 MCS in order to improve the statistics of autocorrelation coefficients. The autocorrelation $\varrho(\omega(t), \Delta t)$ of a time-dependent function $\omega(t)$ is given by

$$\varrho(\omega(t), \Delta t) = \frac{\langle \omega(t + \Delta t)\omega(t)\rangle - \langle \omega(t)\rangle^2}{\langle \omega(t)^2\rangle - \langle \omega(t)\rangle^2} \tag{7.18}$$

in the stationary case. We find that a nontrivial Hurst exponent does not coincide with long-time correlations, as the autocorrelation functions quickly converge to zero. Although the price time series show anti-persistence on short time scales and persistence on medium time scales, no non-vanishing autocorrelation $\varrho(\delta p(t), \Delta t)$ with $\delta p(t) = p(t + 1) - p(t)$ can be observed for $\Delta t > 10$. The autocorrelation functions for the quadratic price change and for the absolute price change, also shown in Fig. 7.14, are positive and converge roughly exponentially to zero. This result can be interpreted as volatility clustering on short time scales, which is also a stylized empirical fact of financial data and have been analyzed in detail in [Con01, PS99, SFG$^+$06]. The analysis in [PS99], for example, shows that for data sets from the New York Stock Exchange the volatility correlations are given by power-laws on time scales from one day to one year and that the exponent is not unique.

These results of the Order Book Model are in good agreement with the results in [BGM06, MGB07], showing that a Hurst exponent $H > 1/2$ does not necessarily imply long-time correlations.

7.3.4 Dynamic Order Placement Depth

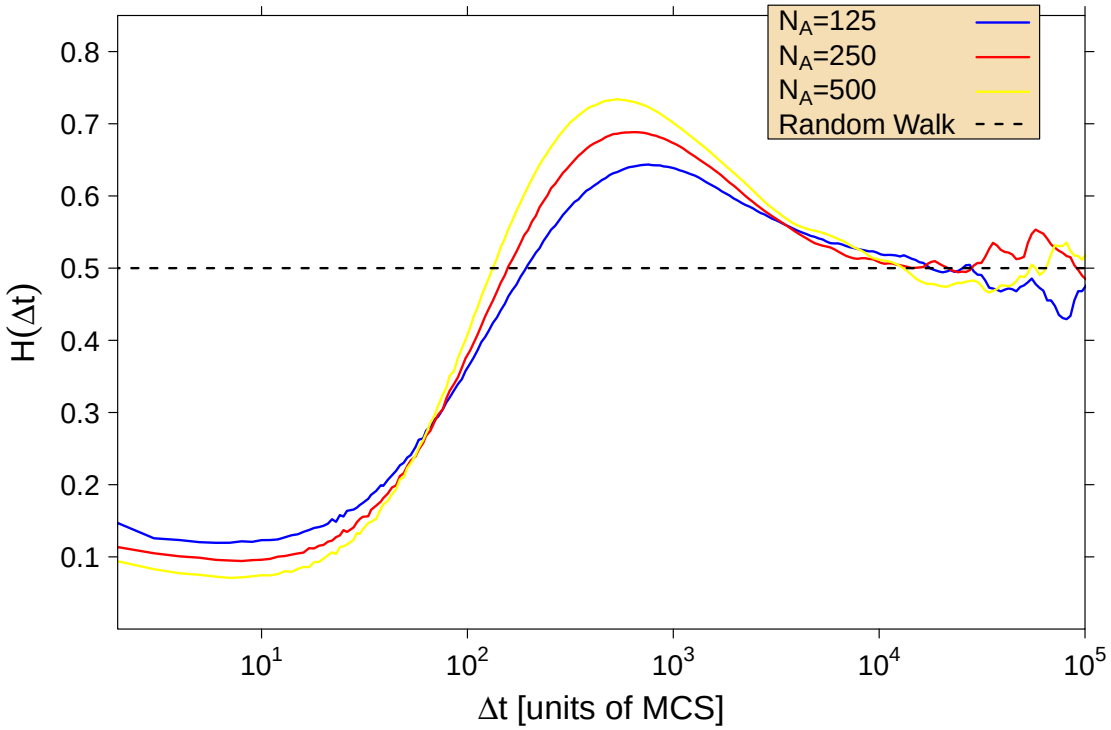

Figure 7.15: Dynamic order placement depth: Hurst exponent $H(\Delta t)$ for various values of N_A with $\lambda_0 = 100$, $C_\lambda = 10$, $\alpha = 0.15$, $\mu = 0.025$, $\delta = 0.025$, and $\Delta S = 0.001$.

In the previous approaches, a constant order placement depth λ_0 was used. According to the remarks in [WBK+08], one can expect an equilibrium on real markets between the effective costs of a market order and of a limit order. If the spread is large, it is advantageous to submit a limit order. In this case, the execution of a limit sell order at best ask or the execution of a limit buy order at best bid is connected with a smaller risk than in the case of a small spread. The risk consists of the establishment of a market trend, which is directed against the position entered by the limit order, leading to a loss. However, if a small risk exists, other liquidity providers are also ready to place orders around a smaller spread. The spread thus decreases down to a level, at which the risk and thus the effective costs of a market order and of a limit order are comparable [WBK+08].

From the above discussion it follows that the liquidity providers can reduce their risk exposure by adapting their limit order placement depth to the prevailing market conditions. In trend-less market phases, in which no large price fluctuations are to be expected, liquidity providers place their limit orders close to the midpoint, in order to be able to participate in small price movements. But if the volatility increases, which can be recognized for example, in strong trend phases, the risk of the liquidity providers to hold positions orientated against the prevailing market trend increases. In these market phases, the probability of closing such a position on the opposite side of the order book

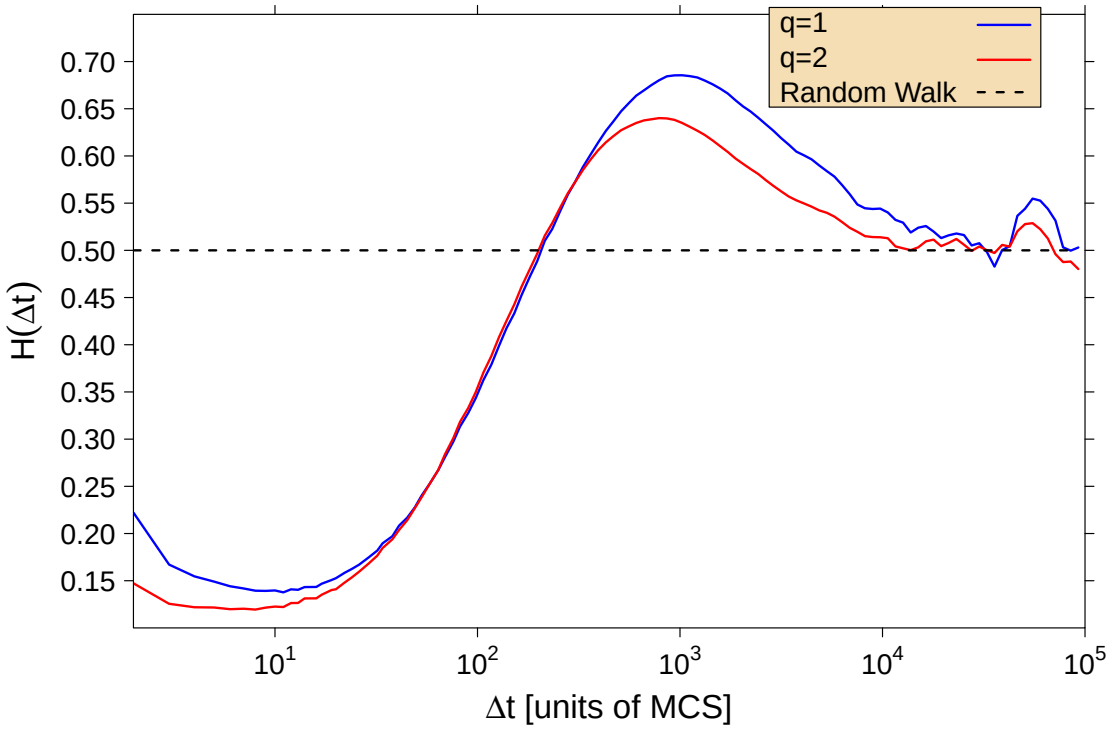

Figure 7.16: Dynamic order placement depth: Hurst exponent $H(\Delta t, q)$ for various values of q with $\lambda_0 = 100$, $C_\lambda = 10$, $\alpha = 0.15$, $\mu = 0.025$, $\delta = 0.025$, $\Delta S = 0.001$, and $N = 125$ for one simulation lasting 5×10^7 MCS.

by a limit order without loss also deceases. Therefore, it is an obvious consequence that liquidity providers adapt their characteristic order placement depth to changing conditions. The market risk is reduced by an enlargement of this characteristic order placement depth in times of a trend. In the Order Book Model, the strength of a trend is given by the deviation of the market order influx from the symmetric case $q_{\text{taker}} = 1/2$. We therefore replace the constant order placement depth λ_0 by:

$$\lambda(t) = \lambda_0 \left(1 + \frac{\left| q_{\text{taker}}(t) - \frac{1}{2} \right|}{\sqrt{\left\langle \left(q_{\text{taker}}(t) - \frac{1}{2} \right)^2 \right\rangle}} \cdot C_\lambda \right) \tag{7.19}$$

For $C_\lambda = 0$, this further extension of the Order Book model corresponds to the variant of the Order Book Model with static order placement parameter. The average value $\langle (q_{\text{taker}}(t) - \frac{1}{2})^2 \rangle$ is determined in a separate Monte Carlo simulation lasting 10^6 MCS before the main simulation starts.

The results of coupling the order placement depth λ to the prevailing trend are shown in Figs. 7.15 – 7.17. The average was taken of 50 simulation runs lasting 10^6 MCS each. With this additional extension of our Order Book Model, which already includes the feedback random walk approach from the first extension, it is now possible to

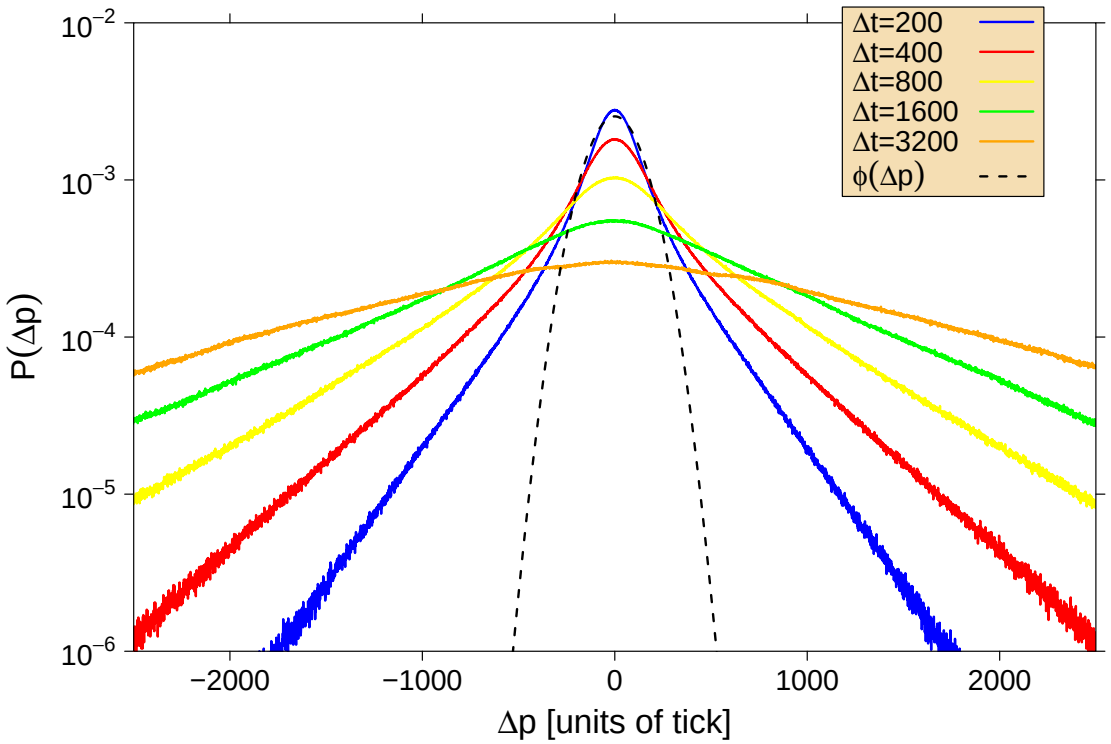

Figure 7.17: Dynamic order placement depth: Distributions of price changes for the same parameter values as in Fig. 7.15. One clearly finds fat tails. A Gaussian approximation $\phi(\Delta p) = a \cdot \exp(-b \cdot \Delta p^2)$ for $\Delta t = 200$ with the parameters $a = 2.54 \times 10^{-3} \pm 10^{-5}$ and $b = (2.80 \pm 0.02) \times 10^{-5}$ strongly underestimates the probability for large price changes.

produce not only a persistent Hurst exponent for medium time scales but also fat-tailed price change distributions.

In Fig. 7.16, the Hurst exponent $H(\Delta t) \equiv H(\Delta t, q = 2)$ is shown in comparison with $H(\Delta t, q = 1)$, as defined in Eq. (7.10). The scaling exponent for absolute price changes $H(\Delta t, q = 1)$ exhibits a larger persistent behavior on medium time scales and the anti-persistence on short time scales is smaller, consistent with earlier findings [PS99, SFG$^+$06].

A widely discussed problem in physics is the origin of the fat-tailed price distributions generated by complex system financial markets. Often the truncated Lévy distribution [MS00, PB00, Voi10] is considered as an approximation to fat-tailed price change distributions found on real financial markets. A Lévy stable distribution is scale invariant and exhibits an infinite variance [MS00]. The truncated Lévy distribution (TLD) [MS94, Kop95] has finite variance and shows scaling behavior in a large, but finite interval. However, the possibility of power law tails have also been discussed at great length in the physics community [MS00]. In [GPA$^+$99] it is shown for the S&P 500 index that the price change distributions for time lags of $\Delta t \leq 4$ days are consistent with a power-

law behavior with an exponent $\alpha \approx 3$, outside the stable Lévy regime ($0 < \alpha_L < 2$). For larger time lags a slow convergence to Gaussian behavior was found.

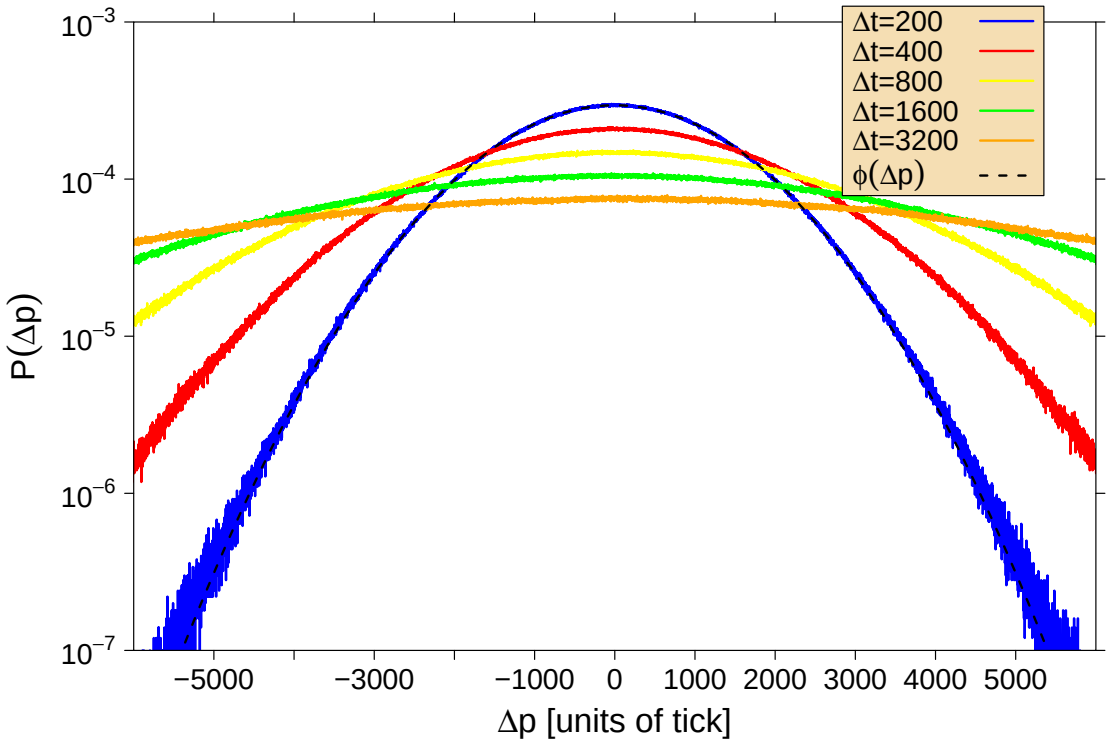

Figure 7.18: Dynamic order placement depth: Distributions of shuffled price changes for the same parameter values as in Fig. 7.17. After shuffling the returns on the time horizon of one MCS, the longer term aggregate returns appear to be Gaussian. For $\Delta t = 200$, the function $\phi(\Delta p) = a \cdot \exp(-b \cdot \Delta p^2)$ can be numerically fitted with $a = 2.9505 \times 10^{-4} \pm 4 \times 10^{-8}$ and $b = 2.7419 \times 10^{-7} \pm 9 \times 10^{-11}$.

One possibility for determining if a process generates true Lévy distributions or not, is to shuffle the short term returns of the time series [GPA+99, VFLS03]. If after shuffling, the longer term returns maintain the same power law exponent, then the stochastic process generates true Lévy distributions, as a stable Lévy process is invariant under folding. However, if after shuffling of the short term returns, the longer term returns appear to be Gaussian then the distributional tails are not as "fat" as Lévy fat-tails as a result of the central limit theorem. However, this argument assumes that the return distributions on short time scales are independent of each other. In our case, one finds the results shown in Fig. 7.17 after shuffling returns on the time horizon of one MCS in Fig. 7.18. It is obvious that the return distributions exhibit convergence to Gaussian behavior. This shows that the fat tails are indeed a result of the dynamic order entry depth, which is coupled with the strength of the prevailing market trend. So we find a qualitatively similar result for data shuffling as that found for independent returns, for different reasons however.

A TLD as described by Koponen [Kop95] shall be used to fit the fat-tailed price changing distributions of Fig. 7.17—this was already used for an analysis of real financial data in chapter 3. It features a smooth exponential cutoff and its characteristic function is given by

$$\Lambda_{\alpha_L,c_1,l}(f_n) = \exp\left(c_0 - c_1 \frac{(f_n^2 + 1/l^2)^{\alpha_L/2}}{\cos(\pi\alpha_L/2)} \cos\left(\alpha_L \arctan(l|f_n|)\right)\right) \tag{7.20}$$

with the scaling factors

$$c_0 = \frac{l^{-\alpha_L}}{\cos(\pi\alpha_L/2)} \tag{7.21}$$

and c_1. As only the characteristic function of the TLD is given in analytic form, a discrete Fourier transform of the price change distributions of our Order Book Model is necessary in order to fit $\Lambda_{\alpha_L,c_1,l}(f_n)$ to the simulation data.

Table 7.1: Fit parameters α_L, c_1, and l of the TLF distribution for $C_\lambda = 1$ in dependence on Δt: the errors given originate only from the process of fitting.

Δt	α_L	c_1	l
200	1.606 ± 0.002	4271 ± 40	3635 ± 52
400	1.477 ± 0.003	3799 ± 52	5433 ± 72
800	1.426 ± 0.004	5542 ± 113	8510 ± 124
1600	1.490 ± 0.005	16901 ± 460	12073 ± 210

Table 7.2: Corresponding to Table 7.1, the fit parameters are given for $C_\lambda = 5$.

Δt	α_L	c_1	l
200	1.232 ± 0.002	1872 ± 23	20903 ± 252
400	1.181 ± 0.002	2209 ± 32	39040 ± 497
800	1.148 ± 0.003	3215 ± 73	70650 ± 1293
1600	1.223 ± 0.005	11058 ± 398	75158 ± 1559

The values of the fit parameters are given in Tabs. 7.1, 7.2, and 7.3 for selected example values of C_λ. Compared to the Lévy exponent α_L measured for real financial data time series (which takes values in the range of $\approx 1.4 - 1.5$ [MS94, MS95]) $C_\lambda = 1$ seems to be the best approximation to real market behavior. It is also in perfect agreement with our findings in chapter 3. In Fig. 7.19, the Lévy exponent α_L is shown as a function of

Table 7.3: Corresponding to Table 7.1, the fit parameters are given for $C_\lambda = 10$.

Δt	α_L	c_1	l
200	1.101 ± 0.002	1427 ± 20	83954 ± 1376
400	1.046 ± 0.003	1583 ± 29	298864 ± 14000
800	1.039 ± 0.004	2773 ± 83	591232 ± 52480
1600	1.143 ± 0.006	12696 ± 642	265875 ± 10860

Δt for various values of C_λ. Again one clearly finds that α_L stays in the correct interval for $C_\lambda \approx 1$, whereas a larger value of C_λ leads to values too small and a smaller one to values that are too large.

Looking closely at Fig. 7.19, we furthermore find an interesting relation between the Lévy exponent α_L and the parameter C_λ for not too long time lags Δt: α_L depends on C_λ via a power law according to $\alpha_L = \tilde{\xi} \cdot C_\lambda^{-0.15}$, with the pre-factor $\tilde{\xi}$ only depending on the time lag.

The coupling of the order placement depth to the prevailing trend thus leads to fat-tailed price change distributions. This property is, however, independent of the persistence of the price time series on medium time scales, as we already obtained this persistence by imposing non-stationary increments in the price process. On the other hand, one can also produce fat tails without $H > 1/2$ for medium time scales. We achieved this scenario by determining $\lambda(t)$ according to a mean reverting random walk as in Eq. (7.19), but always assuming symmetric order placement behavior.

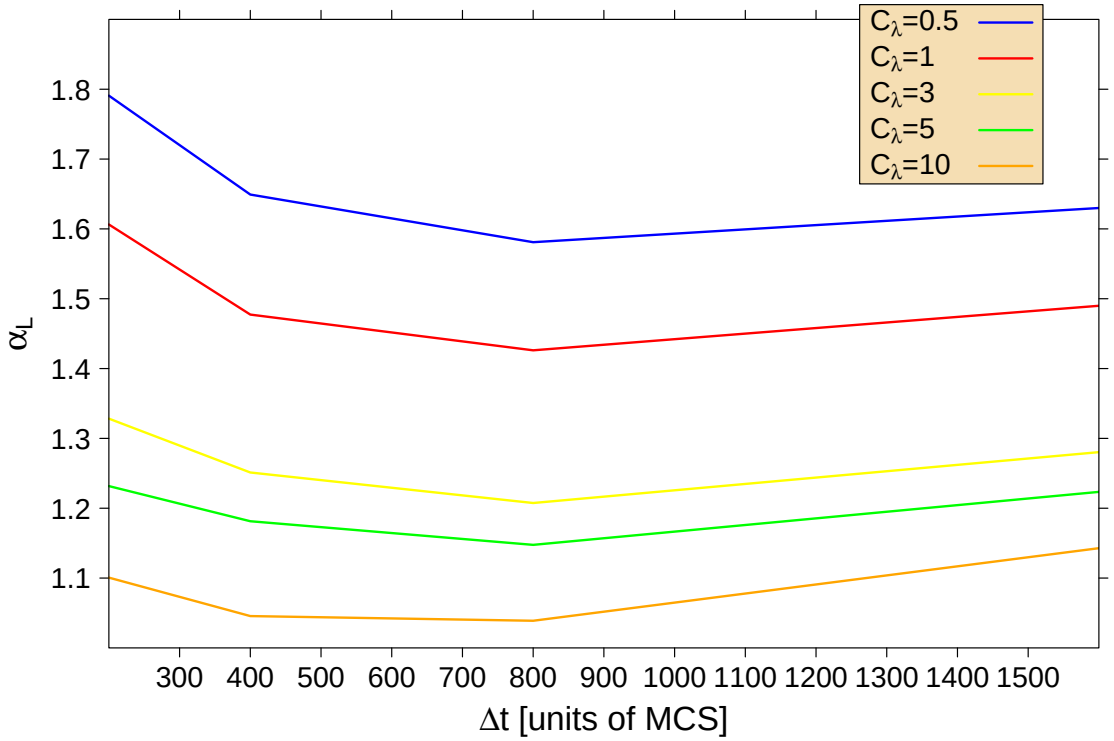

Figure 7.19: Dynamic order placement depth λ: Lévy exponent α_L as a function of $\Delta\tau$ for various C_λ.

7.4 Switching Phenomena in the Order Book Model

In this section, we analyze switching processes in times series generated by the Order Book Model. As seen in chapter 6, the distinct link between price changes and volume fluctuations, as well as between price changes and inter-trade times, vanishes when the sub time series of volumes or inter-trade times are randomly re-shuffled. Thus, a randomly constructed price development with a random volume time series—or random inter-trade time series—is not able to reproduce the switching phenomena shown in chapter 6. In the Order Book Model, it is possible to extract the number of traded contracts per MCS. In contrast, it is not trivial to determine the time intervals between single transactions in the Order Book Model as each agent has one chance to buy or to sell per MCS. Thus, a time resolution smaller than one MCS is not meaningful (i.e.,we limit ourselves to the analysis of volume fluctuations on the path from a locally extreme value to the next locally extreme value in the price). Here, we perform a parallel analysis of the volume fluctuations in the renormalized time ε.

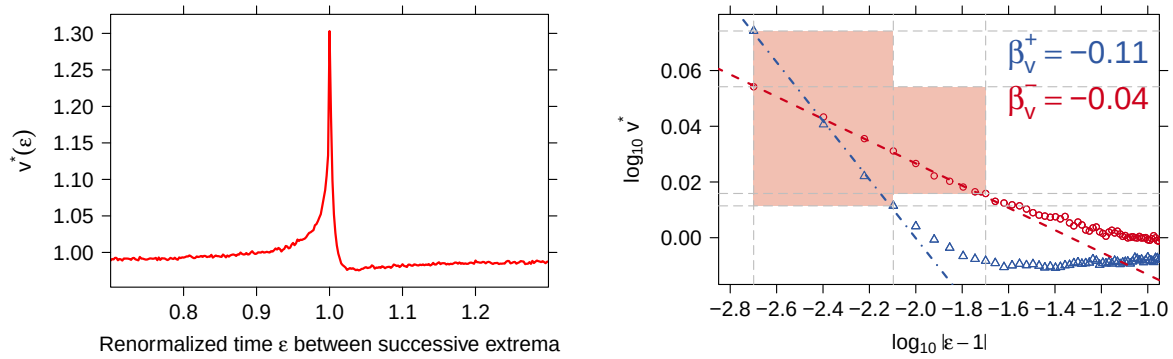

Figure 7.20: Switching analysis of the basic Order Book Model—*left*: aggregated volume $v^*(\varepsilon)$. *right*: $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram.

Figure 7.20 shows the aggregated volume $v^*(\varepsilon)$ (left) and $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ in a log–log histogram (right) for a price volume time series of the basic version of the Order Book Model with exponential order placement depth (50×10^6 MCS). The same parameters are used as in section 7.1.2. As shown in Fig. 7.20, the basic version of the Order Book Model is able to reproduce the link between price fluctuations and volume fluctuations with similar power-law exponents ($\beta_v^+ \approx -0.11$ and $\beta_v^- \approx -0.04$).

The advanced version of the Order Book Model with dynamic order placement depth and feedback random walk leads to almost the same exponents (see Fig. 7.21, 50×10^6 MCS). In this case, we also use the same parameters as in section 7.3.4. Power-law exponents $\beta_v^+ \approx -0.09$ and $\beta_v^- \approx -0.04$ are obtained.

Thus, we conclude that a completely random process is not suitable for modeling the *empirical stylized facts* of switching phenomena. However, realistic rule sets and a

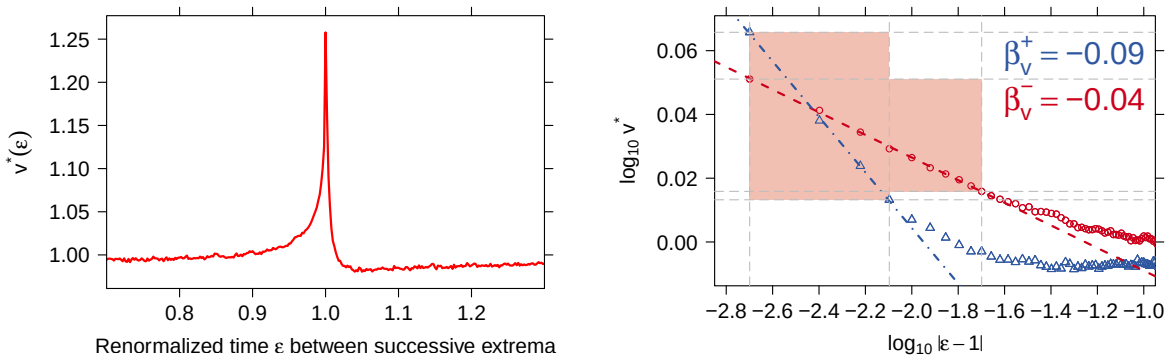

Figure 7.21: Switching analysis of the augmented Order Book Model—*left*: aggregated volume $v^*(\varepsilon)$. *right*: $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram.

realistic microstructure, which is implemented in the Order Book Model, are able to qualitatively reproduce the link between volume and price changes.

7.5 Chapter's Summary

In this chapter we examined the Order Book Model, introduced in [PGPS06] as a multi-agent system for the modeling of financial markets, in more detail. Following a bottom-up-approach, we started out with a simple variant of the model, whose key feature is the distinction between liquidity providers and liquidity takers. We showed that liquidity providers have a systematic advantage through their ability to submit limit orders. Moreover, this simple variant is able to reproduce the results of [DFG+03, SFGK03]. It does not, however, correspond to order books in real financial markets at all, because it exhibits a constant non-zero order book depth, even far away from the midpoint. The introduction of an exponentially distributed order placement depth of limit orders creates a log-normally distributed depth of the order book. This model variant is regarded as the basic version of the Order Book Model. Its price time series possesses an anti-persistent price behavior on short time scales which is due to the order book structure. On medium and long time scales the Hurst exponent converges to a diffusive regime, and the price change distributions exhibit an almost Gaussian shape. This basic version of the Order Book Model, which is characterized by a symmetry created by identical buy and sell probabilities, describes a stationary market. However, when asymmetry is introduced, the Order Book Model is displaced from its stationary state. This extension is implemented by a temporal modulation of the buy probability q_{taker} of the liquidity takers or the buy probability q_{provider} of the liquidity providers. Qualitatively identical results are achieved if both probabilities are modulated independently of each other. Applying a feedback random walk to introduce micro market trends into the market, a persistent price behavior on medium time scales is obtained. However, no fat tails can be reproduced with this asymmetric extension of the Order Book Model. When furthermore coupling the characteristic order placement depth to the prevailing market trend, widened price change distributions are achieved which exhibit so-called fat tails. A truncated Lévy distribution can be fitted to these price changes. Thus, with these extensions to our Order Book Model, we were able to demonstrate that the generation of a nontrivial Hurst exponent is independent of the generation of fat tails. This proves that the assumption which can often be found in the literature that a persistent price behavior corresponds to non-Gaussian price changes is wrong. Furthermore, we are able to support the statement in [BGM06, MGB07] that $H > 1/2$ does not necessarily imply long time correlations. In addition, the characteristics of switching phenomena can be reproduced by the underlying microstructure of the Order Book Model with scaling parameters close to those found for real financial markets.

Chapter 8

Summary and Final Remarks

This thesis introduces and analyzes concepts arising in the interdisciplinary research field between physics and economics known as *econophysics*. In addition to the quantification and modeling of financial market time series, new concepts are applied which contribute to reducing computing times for scientific studies in the data-driven field of econophysics and in statistical physics in general.

It was shown that a graphics card architecture—a graphics processing unit (GPU)—can be very successfully used for methods of time series analysis. An accelerated determination of scaling exponents can be performed on a GPU as well as the calculation of autocorrelation coefficients. Results of the scaling behavior of a stochastic process are obtained up to 80 times faster than on a modern central processing unit (CPU) core. In addition, the relative absolute error of the results in comparison to a CPU is smaller than 10^{-3}. These methods were applied to a German Bund future (FGBL) time series of the European Exchange (Eurex), which exhibits an anti-persistence on short time scales. Evidence is found that a super-diffusive regime is reached on medium time scales. In addition, a time series of the German DAX future (FDAX) is analyzed. Its price change distributions differ significantly from random walk statistics as they exhibit distributional tails which indicate higher probabilities for large price changes than those predicted by a Gaussian distribution. Furthermore, high-frequency asset returns are studied in detail. First, market impacts were investigated systematically by using conditional probability distribution functions (CPDF). The negative autocorrelation of the return time series for consecutive ticks results in a reversion tendency after a price shift on short time scales. Thus, the CPDF behavior can be reproduced qualitatively by a synthetically anti-correlated random walk, which reflects the short-time anti-persistence of the Hurst exponent. Furthermore, a method for measuring complex correlations within a time series by pattern conformity was introduced. Pattern conformity is used as an accurate measure for characterizing the short-term correlations of a general time series. It is essentially given by the comparison of subsequences of the time series. Subsequences of various lengths are compared with historical sequences in order to extract similar reactions to similar patterns. The value of this pattern conformity is 0 for a random walk and is 1 for a perfectly correlated time series. Checking the pattern conformity of finan-

cial data sets, we find that there is a tendency to match historic patterns on very short time scales. This method needs a huge amount of computing time. Thus, the pattern formation conformity algorithm was also implemented on a graphics card architecture. For the pattern conformity calculation, the GPU was up to 24 times faster than a recent CPU core. In addition, it is noteworthy that the pattern conformity values provided by the GPU and the CPU differ only in a relative error of up to two-tenths of a percent.

As data-driven methods were able to be significantly accelerated on a GPU, we used the concept of GPU computing for a prototype model in statistical physics in order to demonstrate the wide range of applicability. Thus, we presented a GPU accelerated version of the two dimensional ferromagnetic square lattice Ising model and the three dimensional ferromagnetic cubic lattice Ising model. This model is often adapted and interpreted in interdisciplinary contexts. For example, in the context of financial markets, spins correspond to market participants. The two possible states model the decision to buy or to sell. For our GPU based Monte Carlo simulations of the Ising model, we use a set of linear congruential random number generators on the GPU device. With the GPU implementation of a checkerboard algorithm of the two dimensional Ising model, results on the GPU can be obtained up to 60 times faster than on a recent CPU core. An implementation of a three dimensional Ising model on a GPU is able to generate results up to 35 times faster than on a modern CPU core. As proof of our conceptual method for the GPU implementation, the critical temperatures of the 2D and 3D Ising models were determined successfully by finite size scaling. Both the theoretical result for the 2D Ising model and previous simulation results for the 3D Ising model can be reproduced. The limitation of such an approach as well must also be mentioned. On a single graphics card, the simulation is limited to the amount of global memory. Algorithms simulating larger lattices have to communicate with the main memory which is the main bottleneck. Facing this problem, two major improvements could be realized as well. For the simple approach, one integer number was used for one single Ising spin. By using multi-spin coding techniques—one bit is used for one Ising spin—we improved the computation to memory access ratio of our calculations dramatically. On one GPU, up to 7.9 spin-flips per nanosecond are attainable this way. This update time is roughly 15 to 35 times faster than a multi-spin coded CPU version. The exact speed increase depends on the implementation and the quality of random numbers. A second major step was to overcome the memory limitation of the GPU global memory. It was possible to distribute a huge Ising lattice on many GPUs—each GPU updating a sub-lattice. This is very efficient for the Ising model as the spin configurations of sub-lattices can be stored on the GPUs. Only the boundaries of each sub-lattice have to be shared with neighboring GPUs hosting the neighboring lattices. This extension of the GPU based simulation of the Ising model is very attractive for GPU clusters. We have shown that the implementation scales nearly linearly with the number of GPUs.

Based on the GPU accelerated analysis of financial market time series—in particular high-frequency time series—the question arose as to whether or not the tendency to react in a similar way to given historic price paths can be uncovered in a time series on the way from a local maximum to a local minimum and vice versa. Furthermore, one has to

add that econophysics' research undertaken formerly was mainly focused on average correlations and distributions. This means, that global properties of the time series are analyzed with histogram methods. Thus, non-trivial correlations among components of a multivariate time series consisting of prices, transaction volumes, and inter-trade times cannot be detected. We were able to show that there is a clear connection between transaction volumes, inter-trade times, and price fluctuations on the path from one extremum to the next extremum. In addition, the underlying law describing the volumes and inter-trade times in the renormalized time ε is a power-law with unique exponents which quantitatively characterize the region around the trend switching point at $\varepsilon = 1$. We find identical behavior for all sub-intervals studied. With a decreasing value of Δt, the number of local minima and maxima increases, around which we find scale-free behavior for exactly the same ε interval $0.6 \leq \varepsilon \leq 1.4$. The characterization of volume and inter-trade times by power-law relationships in the time domain supports the hypothesis that a fluctuating price time series passes through a sequence of switching points.

We observed that each type of trend—micro and macro—in a financial market starts and ends with a unique switching process, and each extremum shares properties of macroscopic cooperative behavior. The mechanism of bubble formation and bubble bursting has no scale for time scales varying over 9 orders of magnitude down to the smallest possible time scale (the scale of individual transactions measured in units of 10 ms). For large time scales, histograms of price returns provide the same scale-free behavior. Thus, the formation of positive and negative trends on all scales is a fundamental principle of trading, starting on the smallest possible time scale, which leads to the non-stationary nature of financial markets as well as to crash events on large time scales. Thus, the well-known catastrophic bubbles occurring on large time scales—such as the most recent financial crisis—may not be outliers but in fact single dramatic representatives caused by the scale-free behavior of the forming of increasing and decreasing trends on time scales from the very large down to the very small.

In order to model the *empirical stylized facts* of financial markets we analyze and extend the Order Book Model, which we originally introduced in [PGPS06]. Following a bottom-up-approach, we began with a simple variant of the model, whose key feature is the distinction between liquidity providers and liquidity takers. We showed that liquidity providers have a systematic advantage through their ability to submit limit orders. The introduction of an exponentially distributed order placement depth of limit orders creates a log-normally distributed depth of the order book. This model variant is regarded as the basic version of the Order Book Model [Pre06]. Its price time series possesses an anti-persistent price behavior on short time scales which is due to the order book structure. On medium and long time scales, the Hurst exponent converges to a diffusive regime, and the price change distributions exhibit an almost Gaussian shape. This basic version of the Order Book Model, which is characterized by a symmetry created by identical buy and sell probabilities, describes a stationary market. However, when additionally introducing a symmetry disturbance, the Order Book Model is shifted from its stationary state. This extension is implemented using a temporal modu-

lation of the buy probability q_{taker} of the liquidity takers. Employing a feedback random walk to introduce micro market trends, one additionally obtains a persistent price behavior on medium time scales. However, no fat tails can be reproduced with such an asymmetric extension of the Order Book Model. When the characteristic order placement depth is further coupled to the prevailing market trend, widened price change distributions are achieved.

In addition, the link between price changes and volume fluctuations uncovered for real financial market datasets could also be reproduced by the Order Book Model. In contrast to random processes for price changes and transaction volumes, the underlying order book structure seems to cause the link between the two quantities through trading rules. Thus, no dynamic order placement depth is necessary in order to cause these switching properties. However, this structure is only able to qualitatively reproduce the volume profile in the renormalized time ε. Future efforts are necessary in order to model the link between price fluctuations and inter-trade times which is much more pronounced than the volume peak.

Appendix A

Dow Jones Industrial Average Index Components

A.1 American Express Company Common Stock (AXP)

The AXP price time series contains 57,393,316 transactions.

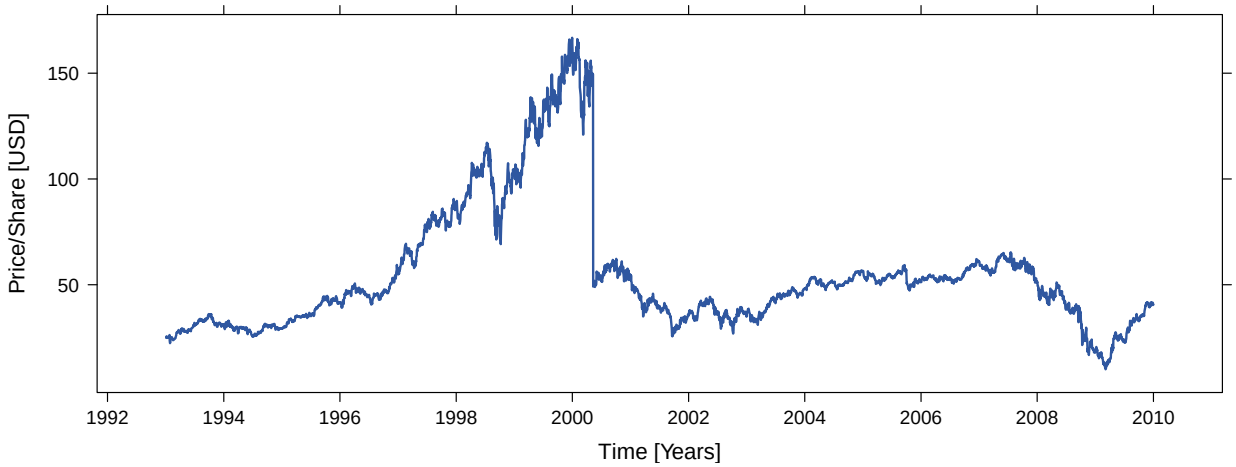

Figure A.1: A stock split occurred on 11 May 2000 [split ratio 3:1].

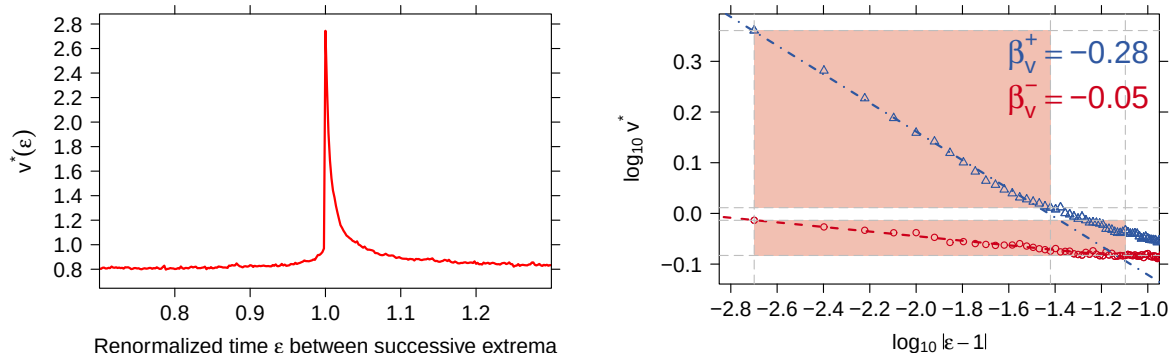

Figure A.2: *left*: aggregated volume $v^*(\varepsilon)$. *right*: $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram.

Table A.1: Statistical test of power-law hypothesis for the AXP volume time series: Scaling parameters of the hypothesized power-law model are shown for both $v^*(\varepsilon)$ before (β_v^-) and $v^*(\varepsilon)$ after (β_v^+) the trend switching point $\varepsilon = 1$ in dependence of $|\varepsilon - 1|_{\text{cut}}$. Additionally, the corresponding values of the KS statistic, D_v^- and D_v^+, are given. The power-law hypothesis is supported if the p-value is larger than 0.1.

$\lvert\varepsilon - 1\rvert_{\text{cut}}$	β_v^+	D_v^+	p-value	β_v^-	D_v^-	p-value
0.004	−0.262	0	**0.726**	−0.043	0	**1**
0.006	−0.278	0.0025	**0.157**	−0.041	0.0003	**0.821**
0.008	−0.286	0.003	**0.188**	−0.04	0.0002	**0.943**
0.01	−0.29	0.0031	**0.292**	−0.036	0.0015	**0.24**
0.012	−0.287	0.0024	**0.731**	−0.039	0.0007	**0.736**
0.014	−0.288	0.0021	**0.912**	−0.045	0.0025	0.009
0.016	−0.289	0.002	**0.977**	−0.046	0.0026	0.003
0.018	−0.291	0.0023	**0.976**	−0.047	0.0027	0.004
0.02	−0.295	0.0033	**0.713**	−0.048	0.0026	0.001
0.022	−0.295	0.0033	**0.813**	−0.049	0.0024	0.008
0.024	−0.295	0.0031	**0.903**	−0.048	0.0022	0.005
0.026	−0.295	0.0027	**0.985**	−0.047	0.0019	0.01
0.028	−0.294	0.0024	**0.999**	−0.046	0.0017	0.039
0.03	−0.292	0.002	**1**	−0.045	0.0015	0.055
0.032	−0.29	0.0018	**1**	−0.045	0.0015	0.063
0.034	−0.287	0.002	**1**	−0.045	0.0014	0.081
0.036	−0.285	0.0029	**0.988**	−0.045	0.0013	**0.118**
0.038	−0.282	0.0037	**0.623**	−0.046	0.0012	**0.136**
0.04	−0.279	0.0044	0.042	−0.045	0.0012	**0.153**
0.042	−0.276	0.0055	0	−0.046	0.0011	**0.191**
0.044	−0.273	0.0061	0	−0.046	0.001	**0.235**
0.046	−0.271	0.0067	0	−0.046	0.001	**0.27**
0.048	−0.269	0.007	0	−0.046	0.0009	**0.313**
0.05	−0.267	0.0074	0	−0.047	0.0009	**0.364**
0.052	−0.264	0.008	0	−0.047	0.001	**0.278**
0.054	−0.262	0.0083	0	−0.047	0.0008	**0.451**
0.056	−0.261	0.0085	0	−0.047	0.0008	**0.489**
0.058	−0.259	0.0088	0	−0.046	0.0008	**0.42**
0.06	−0.256	0.0091	0	−0.046	0.0008	**0.431**
0.062	−0.254	0.0094	0	−0.046	0.0007	**0.465**
0.064	−0.252	0.0096	0	−0.046	0.0007	**0.556**
0.066	−0.251	0.0096	0	−0.046	0.0007	**0.504**
0.068	−0.25	0.0097	0	−0.046	0.0007	**0.447**
0.07	−0.248	0.0097	0	−0.045	0.0007	**0.444**
0.072	−0.247	0.0097	0	−0.045	0.0007	**0.48**
0.074	−0.246	0.0096	0	−0.045	0.0007	**0.477**
0.076	−0.244	0.0098	0	−0.045	0.0007	**0.41**

A.2 Boeing Company (The) Common Stock (BA)

The BA price time series contains 36,717,524 transactions.

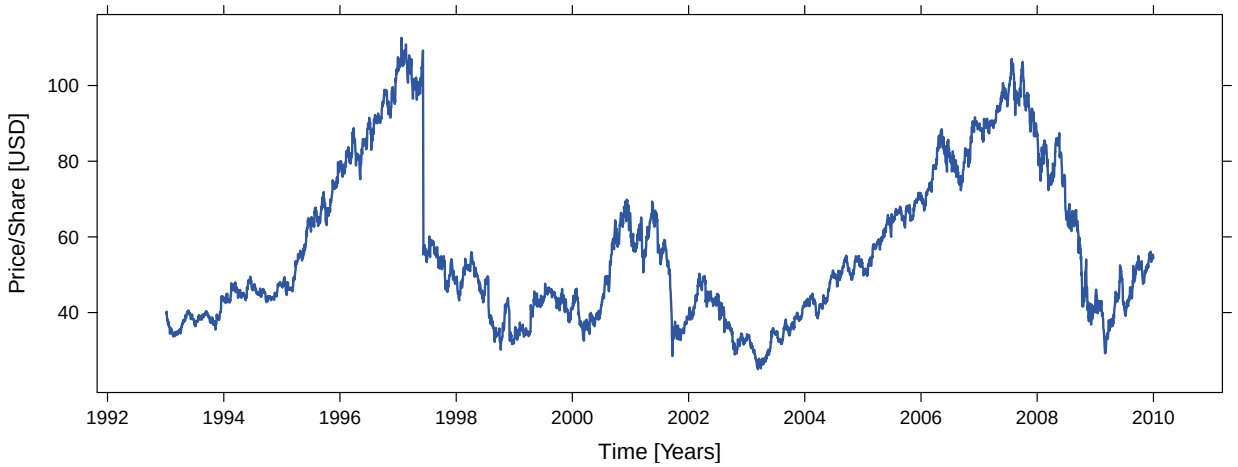

Figure A.3: A stock split occurred on 9 June 1997 [split ratio 2:1].

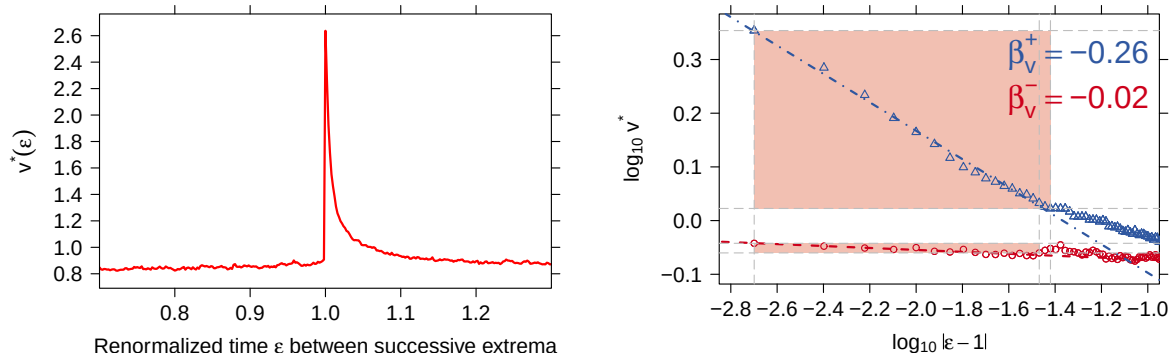

Figure A.4: *left*: aggregated volume $v^*(\varepsilon)$. *right*: $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram.

Table A.2: Statistical test of power-law hypothesis for the BA volume time series: Scaling parameters of the hypothesized power-law model are shown for both $v^*(\varepsilon)$ before (β_v^-) and $v^*(\varepsilon)$ after (β_v^+) the trend switching point $\varepsilon = 1$ in dependence of $|\varepsilon - 1|_{\text{cut}}$. Additionally, the corresponding values of the KS statistic, D_v^- and D_v^+, are given. The power-law hypothesis is supported if the p-value is larger than 0.1.

$\lvert\varepsilon - 1\rvert_{\text{cut}}$	β_v^+	D_v^+	p-value	β_v^-	D_v^-	p-value
0.004	−0.229	0	**1**	−0.017	0	**1**
0.006	−0.249	0.0032	**0.126**	−0.018	0	**0.956**
0.008	−0.267	0.0059	0.005	−0.018	0.0001	**0.987**
0.01	−0.273	0.0059	0.001	−0.014	0.0016	**0.351**
0.012	−0.276	0.0055	0.008	−0.016	0.0008	**0.81**
0.014	−0.282	0.0052	0.017	−0.017	0.0008	**0.764**
0.016	−0.286	0.0059	0.006	−0.015	0.0011	**0.552**
0.018	−0.286	0.0054	0.007	−0.015	0.0007	**0.862**
0.02	−0.285	0.0049	0.046	−0.017	0.0008	**0.712**
0.022	−0.283	0.0041	**0.288**	−0.018	0.0011	**0.407**
0.024	−0.28	0.0036	**0.569**	−0.018	0.0009	**0.59**
0.026	−0.277	0.0035	**0.647**	−0.018	0.001	**0.363**
0.028	−0.275	0.0034	**0.722**	−0.018	0.0008	**0.565**
0.03	−0.272	0.0034	**0.713**	−0.018	0.0009	**0.49**
0.032	−0.269	0.0033	**0.717**	−0.019	0.0008	**0.492**
0.034	−0.267	0.0035	**0.609**	−0.018	0.0007	**0.684**
0.036	−0.266	0.0037	**0.453**	−0.016	0.0014	0.062
0.038	−0.264	0.004	**0.261**	−0.014	0.0025	0
0.04	−0.262	0.0043	0.064	−0.012	0.003	0
0.042	−0.259	0.0049	0	−0.01	0.0041	0
0.044	−0.256	0.0055	0	−0.009	0.0042	0
0.046	−0.253	0.0058	0	−0.009	0.0042	0
0.048	−0.252	0.0062	0	−0.009	0.0041	0
0.05	−0.25	0.0067	0	−0.009	0.0039	0
0.052	−0.248	0.0073	0	−0.009	0.0036	0
0.054	−0.245	0.0078	0	−0.009	0.0036	0
0.056	−0.243	0.0082	0	−0.009	0.0035	0
0.058	−0.241	0.0085	0	−0.008	0.0035	0
0.06	−0.239	0.0088	0	−0.008	0.0034	0
0.062	−0.237	0.0091	0	−0.008	0.0033	0
0.064	−0.235	0.0094	0	−0.008	0.0031	0
0.066	−0.233	0.0096	0	−0.009	0.0029	0
0.068	−0.231	0.0097	0	−0.009	0.0028	0
0.07	−0.23	0.0098	0	−0.01	0.0026	0
0.072	−0.228	0.0099	0	−0.011	0.0025	0
0.074	−0.227	0.0101	0	−0.011	0.0024	0
0.076	−0.226	0.0103	0	−0.011	0.0023	0

A.3 Bank of America Corporation Common Stock (BAC)

The BAC price time series contains 223,277,158 transactions.

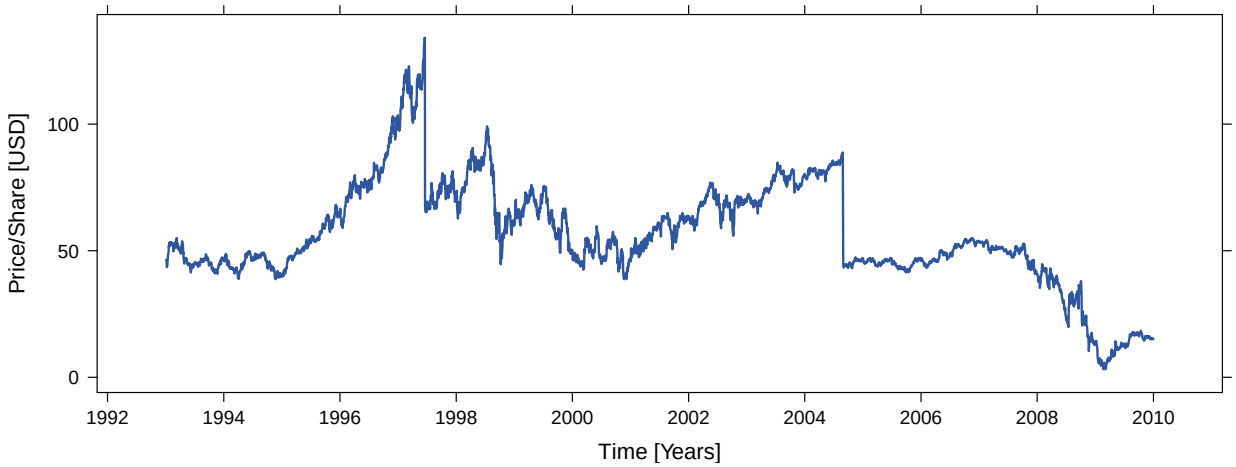

Figure A.5: Stock splits occurred on 28 February 1997 [split ratio 2:1] and 30 August 2004 [split ratio 2:1].

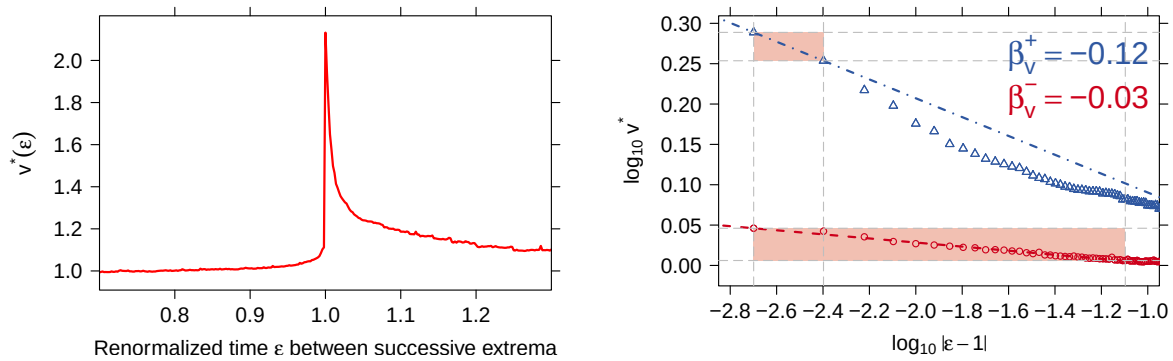

Figure A.6: *left*: aggregated volume $v^*(\varepsilon)$. *right*: $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram.

Table A.3: Statistical test of power-law hypothesis for the BAC volume time series: Scaling parameters of the hypothesized power-law model are shown for both $v^*(\varepsilon)$ before (β_v^-) and $v^*(\varepsilon)$ after (β_v^+) the trend switching point $\varepsilon = 1$ in dependence of $|\varepsilon - 1|_{\text{cut}}$. Additionally, the corresponding values of the KS statistic, D_v^- and D_v^+, are given. The power-law hypothesis is supported if the p-value is larger than 0.1.

$\lvert\varepsilon - 1\rvert_{\text{cut}}$	β_v^+	D_v^+	p-value	β_v^-	D_v^-	p-value
0.004	−0.116	0	**0.776**	−0.012	0	**0.743**
0.006	−0.146	0.0049	0	−0.021	0.0015	0.029
0.008	−0.152	0.0044	0	−0.026	0.0019	0.002
0.01	−0.161	0.0048	0	−0.028	0.0019	0.003
0.012	−0.163	0.0044	0	−0.028	0.0016	0.003
0.014	−0.166	0.0044	0	−0.028	0.0013	0.009
0.016	−0.166	0.0039	0	−0.028	0.0012	0.022
0.018	−0.165	0.0035	0	−0.028	0.001	0.036
0.02	−0.164	0.0031	0	−0.028	0.0009	0.052
0.022	−0.162	0.0028	0.003	−0.027	0.0008	0.068
0.024	−0.16	0.0026	0.014	−0.027	0.0008	**0.104**
0.026	−0.158	0.0024	0.071	−0.027	0.0007	**0.12**
0.028	−0.155	0.0023	0.076	−0.026	0.0007	**0.155**
0.03	−0.153	0.0027	0.001	−0.026	0.0006	**0.184**
0.032	−0.152	0.003	0	−0.026	0.0006	**0.235**
0.034	−0.15	0.0032	0	−0.026	0.0006	**0.252**
0.036	−0.149	0.0034	0	−0.026	0.0005	**0.31**
0.038	−0.148	0.0036	0	−0.026	0.0005	**0.341**
0.04	−0.146	0.0037	0	−0.026	0.0005	**0.386**
0.042	−0.145	0.0038	0	−0.026	0.0004	**0.458**
0.044	−0.144	0.0039	0	−0.026	0.0004	**0.514**
0.046	−0.142	0.004	0	−0.026	0.0004	**0.509**
0.048	−0.141	0.004	0	−0.026	0.0004	**0.547**
0.05	−0.14	0.0041	0	−0.026	0.0004	**0.555**
0.052	−0.139	0.0043	0	−0.026	0.0004	**0.559**
0.054	−0.137	0.0045	0	−0.026	0.0004	**0.6**
0.056	−0.136	0.0047	0	−0.026	0.0003	**0.619**
0.058	−0.135	0.0049	0	−0.026	0.0003	**0.584**
0.06	−0.133	0.0052	0	−0.026	0.0003	**0.626**
0.062	−0.132	0.0055	0	−0.026	0.0003	**0.645**
0.064	−0.131	0.0057	0	−0.026	0.0003	**0.662**
0.066	−0.129	0.006	0	−0.025	0.0003	**0.637**
0.068	−0.128	0.0062	0	−0.025	0.0003	**0.669**
0.07	−0.127	0.0065	0	−0.025	0.0003	**0.502**
0.072	−0.126	0.0067	0	−0.025	0.0003	**0.502**
0.074	−0.125	0.0069	0	−0.025	0.0004	**0.448**
0.076	−0.124	0.007	0	−0.025	0.0004	**0.337**

A.4 Caterpillar, Inc. Common Stock (CAT)

The CAT price time series contains 43,498,630 transactions.

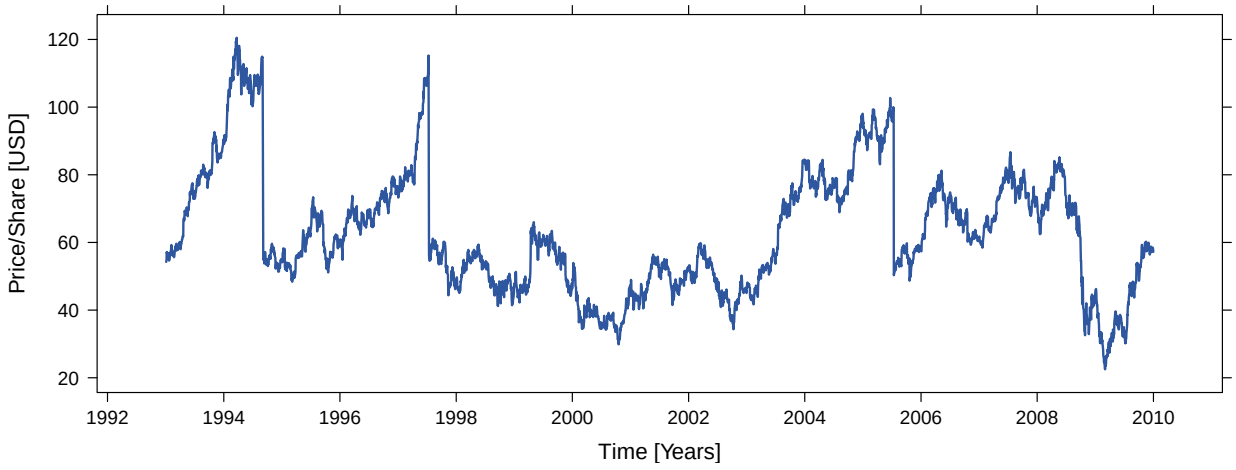

Figure A.7: Stock splits occurred on 6 September 1994 [split ratio 2:1], 14 July 1997 [split ratio 2:1], and 14 July 2005 [split ratio 2:1].

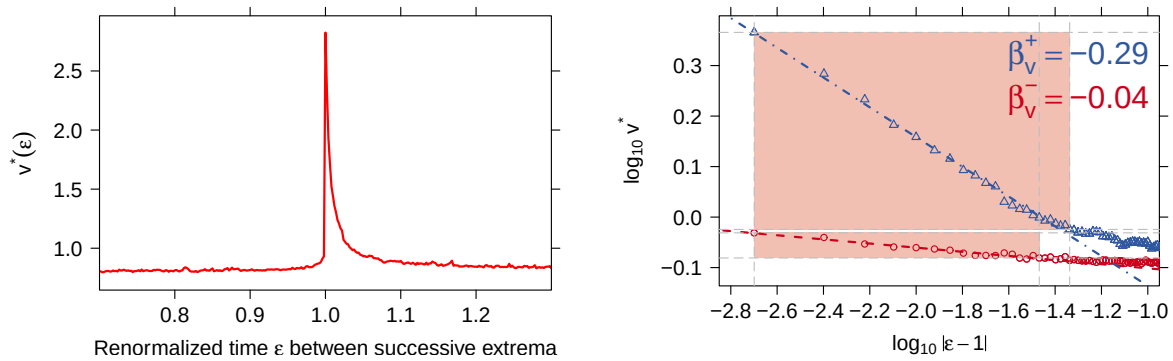

Figure A.8: *left*: aggregated volume $v^*(\varepsilon)$. *right*: $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram.

Table A.4: Statistical test of power-law hypothesis for the CAT volume time series: Scaling parameters of the hypothesized power-law model are shown for both $v^*(\varepsilon)$ before (β_v^-) and $v^*(\varepsilon)$ after (β_v^+) the trend switching point $\varepsilon = 1$ in dependence of $|\varepsilon - 1|_{\text{cut}}$. Additionally, the corresponding values of the KS statistic, D_v^- and D_v^+, are given. The power-law hypothesis is supported if the p-value is larger than 0.1.

$\lvert\varepsilon - 1\rvert_{\text{cut}}$	β_v^+	D_v^+	p-value	β_v^-	D_v^-	p-value
0.004	−0.272	0	**0.692**	−0.03	0	**1**
0.006	−0.277	0.0008	**0.717**	−0.044	0.0023	**0.13**
0.008	−0.298	0.0058	0	−0.047	0.002	**0.178**
0.01	−0.3	0.0052	0.009	−0.045	0.0015	**0.327**
0.012	−0.303	0.0048	0.052	−0.043	0.0012	**0.449**
0.014	−0.302	0.0042	**0.181**	−0.042	0.0011	**0.549**
0.016	−0.304	0.0039	**0.353**	−0.043	0.0009	**0.646**
0.018	−0.303	0.0035	**0.608**	−0.045	0.0012	**0.44**
0.02	−0.303	0.0032	**0.793**	−0.045	0.0011	**0.452**
0.022	−0.301	0.003	**0.918**	−0.045	0.0009	**0.561**
0.024	−0.305	0.0033	**0.866**	−0.042	0.0011	**0.443**
0.026	−0.307	0.0042	**0.467**	−0.041	0.0017	0.078
0.028	−0.308	0.0043	**0.393**	−0.041	0.0014	**0.186**
0.03	−0.307	0.0037	**0.807**	−0.042	0.0011	**0.364**
0.032	−0.307	0.0034	**0.924**	−0.041	0.0014	**0.114**
0.034	−0.306	0.003	**0.993**	−0.04	0.0014	**0.109**
0.036	−0.305	0.0027	**0.999**	−0.04	0.0014	0.095
0.038	−0.302	0.0022	**1**	−0.039	0.0015	0.049
0.04	−0.3	0.0022	**1**	−0.039	0.0014	0.071
0.042	−0.299	0.0023	**1**	−0.039	0.0014	0.049
0.044	−0.296	0.0032	**0.967**	−0.038	0.0016	0.016
0.046	−0.294	0.0039	**0.584**	−0.038	0.0016	0.017
0.048	−0.292	0.0045	0.057	−0.037	0.0016	0.014
0.05	−0.29	0.005	0.001	−0.037	0.0015	0.013
0.052	−0.288	0.0056	0	−0.037	0.0015	0.016
0.054	−0.286	0.0061	0	−0.037	0.0015	0.016
0.056	−0.283	0.0065	0	−0.037	0.0014	0.009
0.058	−0.28	0.0074	0	−0.037	0.0014	0.016
0.06	−0.277	0.0082	0	−0.037	0.0014	0.025
0.062	−0.274	0.0089	0	−0.036	0.0014	0.023
0.064	−0.272	0.0094	0	−0.036	0.0014	0.007
0.066	−0.269	0.01	0	−0.036	0.0014	0.015
0.068	−0.267	0.0106	0	−0.035	0.0014	0.003
0.07	−0.264	0.0112	0	−0.035	0.0014	0.004
0.072	−0.262	0.0116	0	−0.035	0.0015	0.002
0.074	−0.26	0.0119	0	−0.034	0.0017	0
0.076	−0.259	0.0122	0	−0.034	0.0018	0

A.5 Cisco Systems, Inc. (CSCO)

The CSCO price time series contains 232,622,629 transactions.

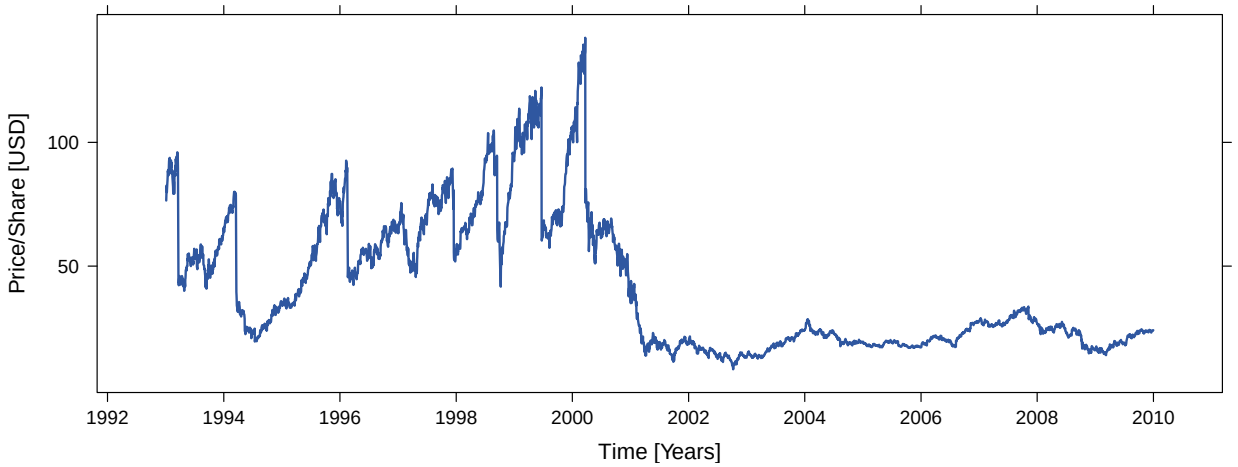

Figure A.9: Stock splits occurred on 22 Mar 1993 [split ratio 2:1], 21 March 1994 [split ratio 2:1], 20 February 1996 [split ratio 2:1], 17 December 1997 [split ratio 3:2], 16 September 1998 [split ratio 3:2], 22 June 1999 [split ratio 2:1], and 23 March 2000 [split ratio 2:1].

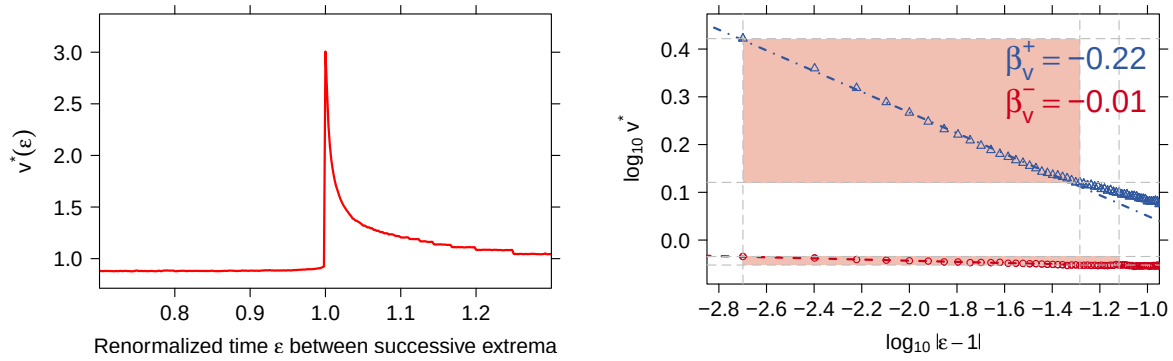

Figure A.10: *left*: aggregated volume $v^*(\varepsilon)$. *right*: $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram.

Table A.5: Statistical test of power-law hypothesis for the CSCO volume time series: Scaling parameters of the hypothesized power-law model are shown for both $v^*(\varepsilon)$ before (β_v^-) and $v^*(\varepsilon)$ after (β_v^+) the trend switching point $\varepsilon = 1$ in dependence of $|\varepsilon - 1|_{\text{cut}}$. Additionally, the corresponding values of the KS statistic, D_v^- and D_v^+, are given. The power-law hypothesis is supported if the p-value is larger than 0.1.

$\lvert\varepsilon - 1\rvert_{\text{cut}}$	β_v^+	D_v^+	p-value	β_v^-	D_v^-	p-value
0.004	−0.207	0	**1**	−0.01	0	**1**
0.006	−0.215	0.0014	**0.249**	−0.013	0.0004	**0.496**
0.008	−0.22	0.0017	**0.33**	−0.013	0.0003	**0.709**
0.01	−0.222	0.0018	**0.566**	−0.013	0.0002	**0.866**
0.012	−0.224	0.0017	**0.836**	−0.012	0.0002	**0.932**
0.014	−0.225	0.0018	**0.941**	−0.013	0.0002	**0.968**
0.016	−0.225	0.0015	**0.998**	−0.013	0.0001	**0.985**
0.018	−0.226	0.0014	**1**	−0.013	0.0001	**0.978**
0.02	−0.226	0.0013	**1**	−0.013	0.0001	**0.997**
0.022	−0.226	0.0013	**1**	−0.012	0.0001	**0.995**
0.024	−0.226	0.0012	**1**	−0.012	0.0003	**0.728**
0.026	−0.226	0.0011	**1**	−0.012	0.0002	**0.804**
0.028	−0.225	0.001	**1**	−0.012	0.0002	**0.778**
0.03	−0.225	0.0009	**1**	−0.012	0.0002	**0.75**
0.032	−0.224	0.0009	**1**	−0.012	0.0002	**0.858**
0.034	−0.224	0.0009	**1**	−0.012	0.0002	**0.816**
0.036	−0.223	0.0009	**1**	−0.012	0.0002	**0.863**
0.038	−0.223	0.0009	**1**	−0.012	0.0002	**0.892**
0.04	−0.222	0.0009	**1**	−0.012	0.0001	**0.916**
0.042	−0.221	0.0009	**1**	−0.012	0.0001	**0.944**
0.044	−0.22	0.0013	**1**	−0.012	0.0001	**0.949**
0.046	−0.219	0.0018	**1**	−0.012	0.0002	**0.823**
0.048	−0.218	0.0021	**1**	−0.012	0.0002	**0.782**
0.05	−0.217	0.0023	**0.987**	−0.012	0.0002	**0.83**
0.052	−0.216	0.0027	**0.508**	−0.012	0.0002	**0.745**
0.054	−0.215	0.003	0.025	−0.012	0.0002	**0.775**
0.056	−0.214	0.0032	0	−0.012	0.0002	**0.83**
0.058	−0.213	0.0035	0	−0.012	0.0002	**0.832**
0.06	−0.212	0.0037	0	−0.012	0.0001	**0.888**
0.062	−0.21	0.004	0	−0.012	0.0001	**0.888**
0.064	−0.21	0.0042	0	−0.012	0.0001	**0.949**
0.066	−0.208	0.0044	0	−0.012	0.0001	**0.957**
0.068	−0.208	0.0045	0	−0.012	0.0001	**0.964**
0.07	−0.206	0.0048	0	−0.012	0.0001	**0.959**
0.072	−0.205	0.005	0	−0.012	0.0002	**0.76**
0.074	−0.205	0.0052	0	−0.012	0.0003	**0.29**
0.076	−0.204	0.0054	0	−0.012	0.0003	**0.145**

A.6 Chevron Corporation Common Stock (CVX)

The CVX price time series contains 60,434,392 transactions.

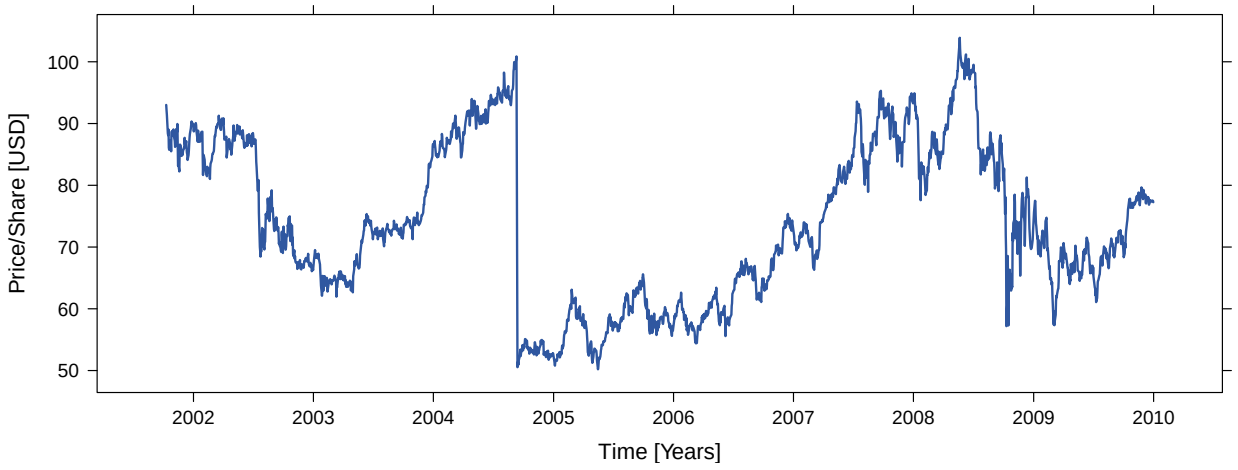

Figure A.11: Stock splits occurred on 13 June 1994 [split ratio 2:1] and 13 September 2004 [split ratio 2:1].

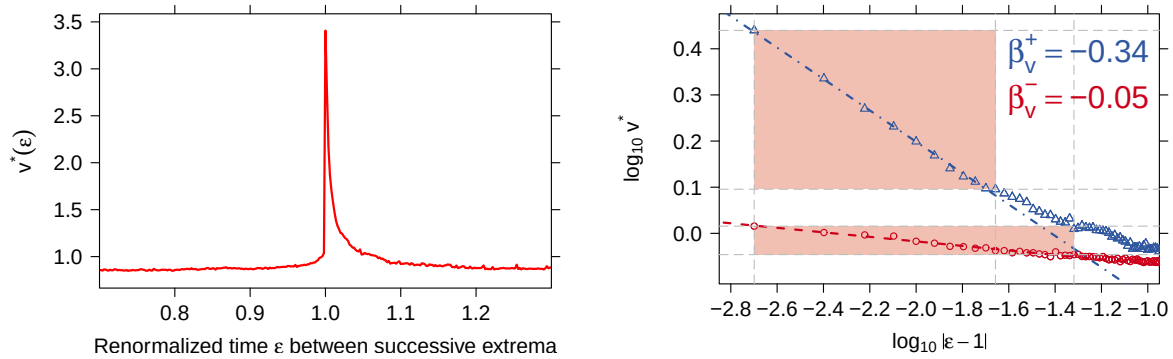

Figure A.12: *left*: aggregated volume $v^*(\varepsilon)$. *right*: $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram.

Table A.6: Statistical test of power-law hypothesis for the CVX volume time series: Scaling parameters of the hypothesized power-law model are shown for both $v^*(\varepsilon)$ before (β_v^-) and $v^*(\varepsilon)$ after (β_v^+) the trend switching point $\varepsilon = 1$ in dependence of $|\varepsilon - 1|_{\text{cut}}$. Additionally, the corresponding values of the KS statistic, D_v^- and D_v^+, are given. The power-law hypothesis is supported if the p-value is larger than 0.1.

$\lvert\varepsilon - 1\rvert_{\text{cut}}$	β_v^+	D_v^+	p-value	β_v^-	D_v^-	p-value
0.004	−0.346	0	**1**	−0.046	0	**1**
0.006	−0.354	0.0014	**0.553**	−0.04	0.0009	**0.494**
0.008	−0.349	0.0009	**0.94**	−0.036	0.0014	**0.309**
0.01	−0.346	0.0013	**0.969**	−0.042	0.0015	**0.232**
0.012	−0.346	0.001	**0.997**	−0.045	0.002	0.061
0.014	−0.35	0.001	**0.999**	−0.049	0.0024	0.011
0.016	−0.35	0.0011	**1**	−0.05	0.0022	0.022
0.018	−0.348	0.0009	**1**	−0.051	0.0022	0.011
0.02	−0.345	0.0018	**1**	−0.05	0.0018	0.052
0.022	−0.34	0.0035	**0.932**	−0.051	0.0018	0.044
0.024	−0.334	0.0053	0.038	−0.051	0.0017	0.078
0.026	−0.329	0.0068	0	−0.052	0.0016	0.059
0.028	−0.324	0.0081	0	−0.051	0.0015	**0.122**
0.03	−0.32	0.009	0	−0.051	0.0014	**0.133**
0.032	−0.317	0.0091	0	−0.051	0.0013	**0.183**
0.034	−0.315	0.0092	0	−0.051	0.0012	**0.194**
0.036	−0.312	0.0093	0	−0.051	0.0011	**0.245**
0.038	−0.31	0.0095	0	−0.051	0.0011	**0.298**
0.04	−0.308	0.0095	0	−0.05	0.001	**0.283**
0.042	−0.306	0.0095	0	−0.05	0.001	**0.285**
0.044	−0.303	0.0096	0	−0.05	0.001	**0.332**
0.046	−0.299	0.0099	0	−0.049	0.0009	**0.325**
0.048	−0.298	0.0098	0	−0.048	0.0012	**0.118**
0.05	−0.296	0.0097	0	−0.048	0.0014	0.018
0.052	−0.293	0.0101	0	−0.048	0.0015	0.008
0.054	−0.291	0.0106	0	−0.047	0.0015	0.006
0.056	−0.288	0.011	0	−0.047	0.0016	0.002
0.058	−0.285	0.0117	0	−0.047	0.0016	0.002
0.06	−0.283	0.0122	0	−0.046	0.0016	0.001
0.062	−0.28	0.0127	0	−0.046	0.0016	0
0.064	−0.278	0.0131	0	−0.046	0.0017	0
0.066	−0.275	0.0135	0	−0.046	0.0016	0.001
0.068	−0.273	0.014	0	−0.046	0.0015	0
0.07	−0.271	0.0142	0	−0.046	0.0015	0.002
0.072	−0.269	0.0145	0	−0.046	0.0015	0
0.074	−0.268	0.0146	0	−0.046	0.0015	0
0.076	−0.266	0.0148	0	−0.046	0.0014	0.001

A.7 E.I. du Pont de Nemours and Common Stock (DD)

The DD price time series contains 36,254,128 transactions.

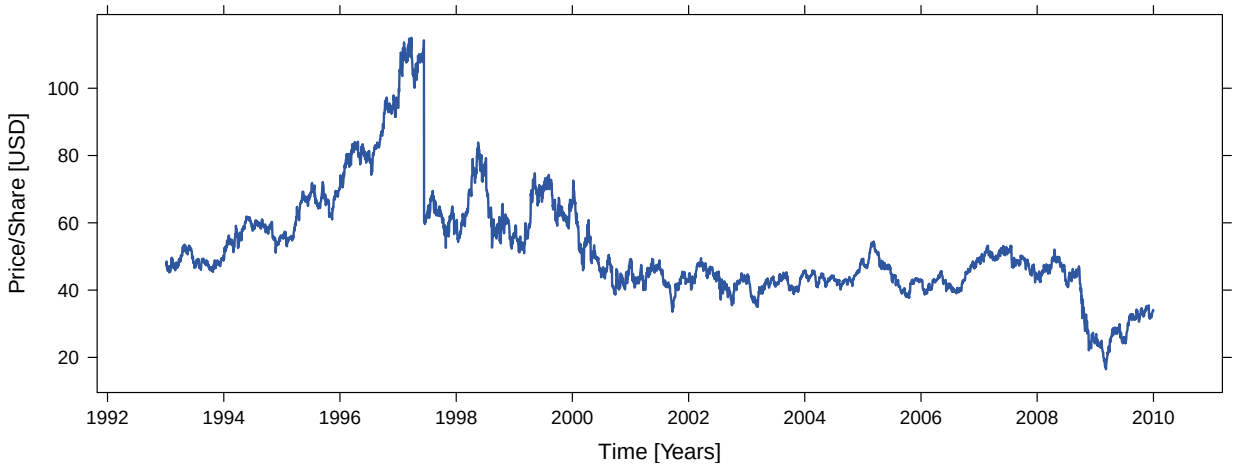

Figure A.13: A stock split occurred on 13 June 1997 [split ratio 2:1].

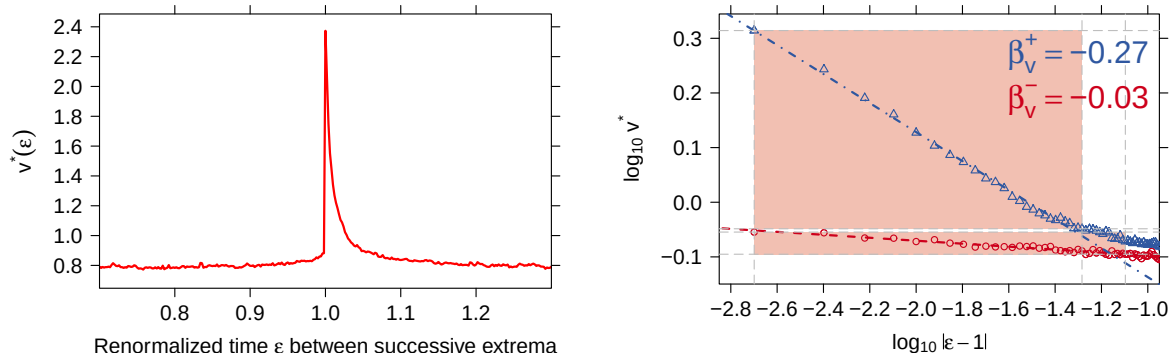

Figure A.14: *left*: aggregated volume $v^*(\varepsilon)$. *right*: $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram.

Table A.7: Statistical test of power-law hypothesis for the DD volume time series: Scaling parameters of the hypothesized power-law model are shown for both $v^*(\varepsilon)$ before (β_v^-) and $v^*(\varepsilon)$ after (β_v^+) the trend switching point $\varepsilon = 1$ in dependence of $|\varepsilon - 1|_{\text{cut}}$. Additionally, the corresponding values of the KS statistic, D_v^- and D_v^+, are given. The power-law hypothesis is supported if the p-value is larger than 0.1.

$\lvert\varepsilon - 1\rvert_{\text{cut}}$	β_v^+	D_v^+	p-value	β_v^-	D_v^-	p-value
0.004	−0.236	0	**1**	−0.004	0	**1**
0.006	−0.256	0.0033	**0.109**	−0.02	0.0027	**0.107**
0.008	−0.257	0.0027	**0.317**	−0.02	0.0021	**0.225**
0.01	−0.265	0.0038	**0.128**	−0.024	0.002	**0.203**
0.012	−0.271	0.0046	0.054	−0.022	0.0015	**0.358**
0.014	−0.273	0.0044	0.095	−0.024	0.0013	**0.413**
0.016	−0.272	0.0039	**0.261**	−0.025	0.0016	**0.26**
0.018	−0.273	0.0036	**0.403**	−0.027	0.002	0.061
0.02	−0.274	0.0033	**0.596**	−0.028	0.002	0.052
0.022	−0.273	0.0031	**0.762**	−0.028	0.0019	0.05
0.024	−0.273	0.0029	**0.876**	−0.028	0.0017	0.08
0.026	−0.274	0.0027	**0.943**	−0.028	0.0016	**0.111**
0.028	−0.275	0.0025	**0.977**	−0.027	0.0014	**0.162**
0.03	−0.276	0.0029	**0.933**	−0.026	0.0012	**0.188**
0.032	−0.277	0.0028	**0.945**	−0.026	0.0011	**0.264**
0.034	−0.277	0.0028	**0.97**	−0.026	0.0011	**0.262**
0.036	−0.276	0.0025	**0.993**	−0.025	0.001	**0.32**
0.038	−0.276	0.0023	**1**	−0.024	0.001	**0.327**
0.04	−0.275	0.002	**1**	−0.025	0.0009	**0.359**
0.042	−0.273	0.002	**1**	−0.025	0.0008	**0.398**
0.044	−0.271	0.0021	**0.997**	−0.026	0.0009	**0.325**
0.046	−0.269	0.0022	**0.999**	−0.026	0.0009	**0.332**
0.048	−0.268	0.0025	**0.997**	−0.026	0.0011	**0.14**
0.05	−0.267	0.003	**0.893**	−0.026	0.0009	**0.304**
0.052	−0.265	0.0036	**0.403**	−0.026	0.0008	**0.325**
0.054	−0.263	0.0041	0.068	−0.026	0.001	**0.127**
0.056	−0.261	0.0047	0.001	−0.027	0.0011	0.094
0.058	−0.259	0.0053	0	−0.026	0.0009	**0.2**
0.06	−0.257	0.0058	0	−0.026	0.0009	**0.194**
0.062	−0.255	0.0061	0	−0.026	0.0008	**0.263**
0.064	−0.253	0.0065	0	−0.026	0.0008	**0.293**
0.066	−0.251	0.0072	0	−0.026	0.0008	**0.251**
0.068	−0.248	0.0078	0	−0.027	0.0009	**0.211**
0.07	−0.247	0.0081	0	−0.026	0.0008	**0.268**
0.072	−0.245	0.0084	0	−0.027	0.0008	**0.28**
0.074	−0.244	0.0087	0	−0.027	0.0008	**0.294**
0.076	−0.242	0.0093	0	−0.027	0.0007	**0.284**

A.8 Walt Disney Company (The) Common Stock (DIS)

The DIS price time series contains 52,644,832 transactions.

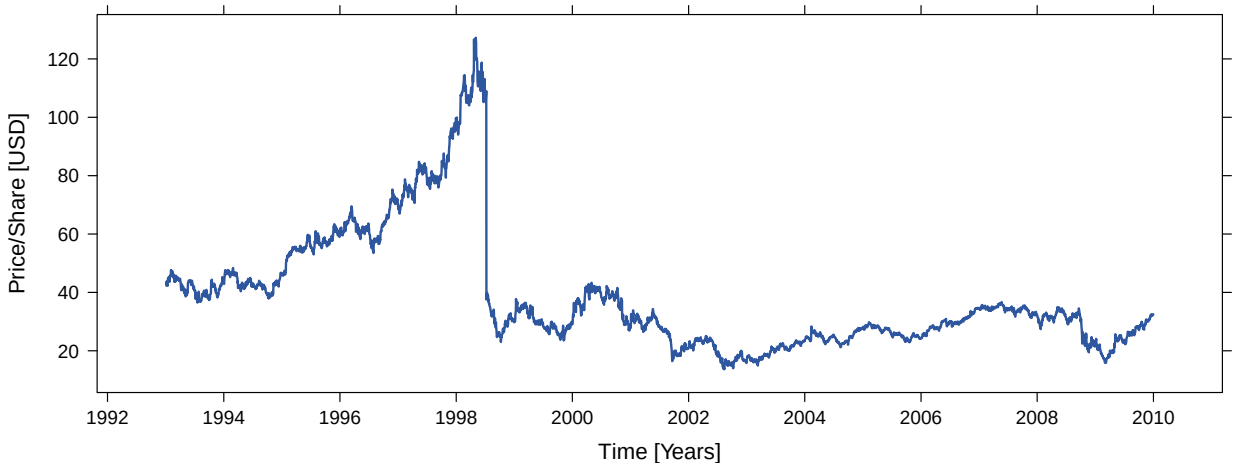

Figure A.15: A stock split occurred on 10 July 1998 [split ratio 3:1].

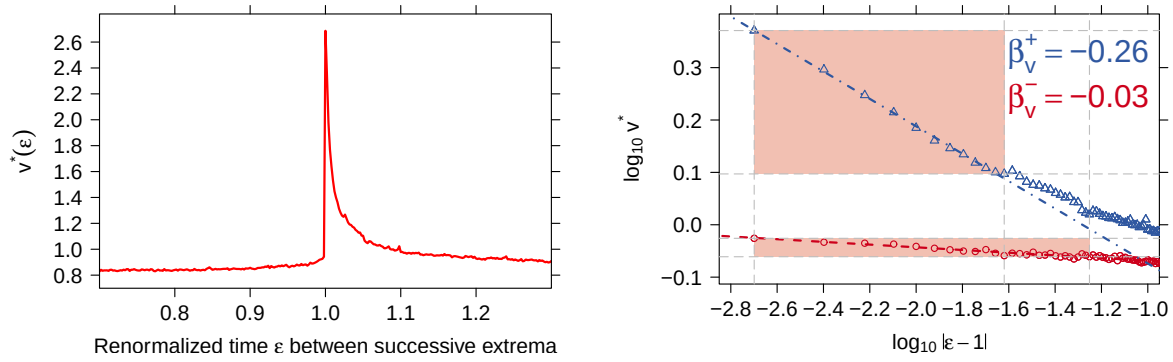

Figure A.16: *left*: aggregated volume $v^*(\varepsilon)$. *right*: $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram.

Table A.8: Statistical test of power-law hypothesis for the DIS volume time series: Scaling parameters of the hypothesized power-law model are shown for both $v^*(\varepsilon)$ before (β_v^-) and $v^*(\varepsilon)$ after (β_v^+) the trend switching point $\varepsilon = 1$ in dependence of $|\varepsilon - 1|_{\text{cut}}$. Additionally, the corresponding values of the KS statistic, D_v^- and D_v^+, are given. The power-law hypothesis is supported if the p-value is larger than 0.1.

$\lvert\varepsilon - 1\rvert_{\text{cut}}$	β_v^+	D_v^+	p-value	β_v^-	D_v^-	p-value
0.004	−0.247	0	**1**	−0.024	0	**1**
0.006	−0.257	0.0016	**0.383**	−0.02	0.0007	**0.589**
0.008	−0.259	0.0015	**0.643**	−0.018	0.0007	**0.721**
0.01	−0.264	0.0023	**0.516**	−0.02	0.0005	**0.889**
0.012	−0.269	0.0031	**0.362**	−0.022	0.0008	**0.643**
0.014	−0.269	0.0028	**0.637**	−0.024	0.0012	**0.299**
0.016	−0.267	0.0022	**0.888**	−0.025	0.0014	**0.152**
0.018	−0.267	0.002	**0.968**	−0.026	0.0014	**0.161**
0.02	−0.266	0.0018	**0.994**	−0.025	0.0011	**0.288**
0.022	−0.265	0.0016	**0.999**	−0.026	0.0011	**0.268**
0.024	−0.261	0.0015	**1**	−0.028	0.0016	0.038
0.026	−0.255	0.0041	0.08	−0.028	0.0014	0.091
0.028	−0.251	0.0056	0	−0.027	0.0013	**0.103**
0.03	−0.248	0.0062	0	−0.028	0.0012	**0.105**
0.032	−0.246	0.0066	0	−0.027	0.0009	**0.352**
0.034	−0.243	0.0072	0	−0.027	0.0009	**0.33**
0.036	−0.241	0.0076	0	−0.027	0.0008	**0.37**
0.038	−0.238	0.0079	0	−0.026	0.0007	**0.447**
0.04	−0.236	0.0079	0	−0.026	0.0007	**0.468**
0.042	−0.234	0.008	0	−0.026	0.0007	**0.508**
0.044	−0.233	0.0079	0	−0.026	0.0006	**0.512**
0.046	−0.231	0.0079	0	−0.026	0.0006	**0.562**
0.048	−0.23	0.0077	0	−0.027	0.0006	**0.633**
0.05	−0.228	0.0076	0	−0.027	0.0005	**0.605**
0.052	−0.228	0.0073	0	−0.026	0.0005	**0.572**
0.054	−0.229	0.0071	0	−0.025	0.0007	**0.374**
0.056	−0.229	0.0069	0	−0.025	0.0007	**0.267**
0.058	−0.228	0.0067	0	−0.025	0.001	0.054
0.06	−0.227	0.0065	0	−0.025	0.001	0.043
0.062	−0.226	0.0064	0	−0.024	0.0012	0.012
0.064	−0.225	0.0063	0	−0.024	0.0012	0.007
0.066	−0.225	0.0062	0	−0.024	0.0012	0.005
0.068	−0.224	0.0061	0	−0.024	0.001	0.025
0.07	−0.223	0.006	0	−0.024	0.001	0.028
0.072	−0.222	0.0059	0	−0.024	0.0011	0.015
0.074	−0.221	0.0058	0	−0.024	0.001	0.024
0.076	−0.22	0.0057	0	−0.024	0.0011	0.011

A.9 General Electric Company Common Stock (GE)

The GE price time series contains 160,362,970 transactions.

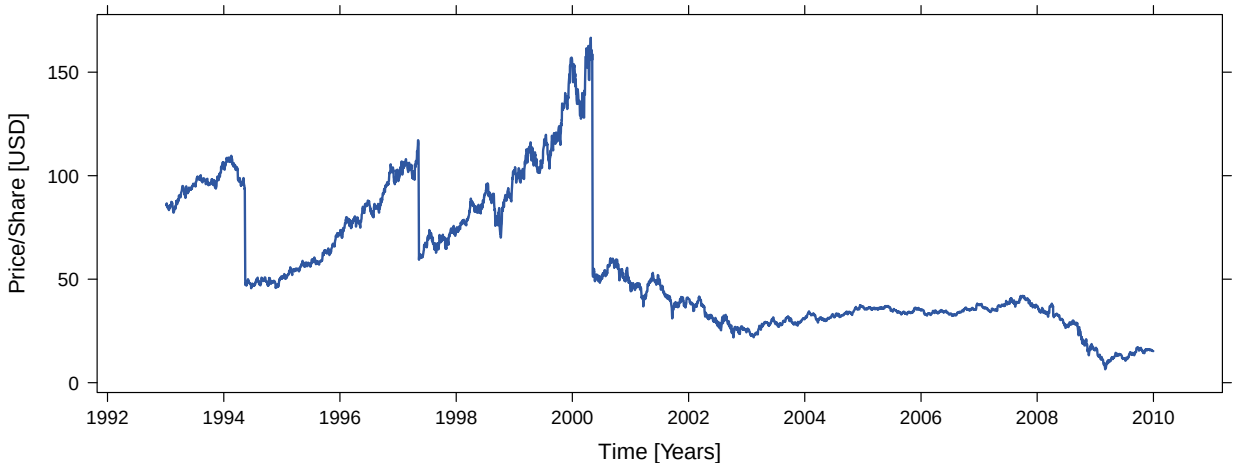

Figure A.17: Stock splits occurred on 16 May 1994 [split ratio 2:1], 12 May 1997 [split ratio 2:1], and 8 May 2000 [split ratio 3:1].

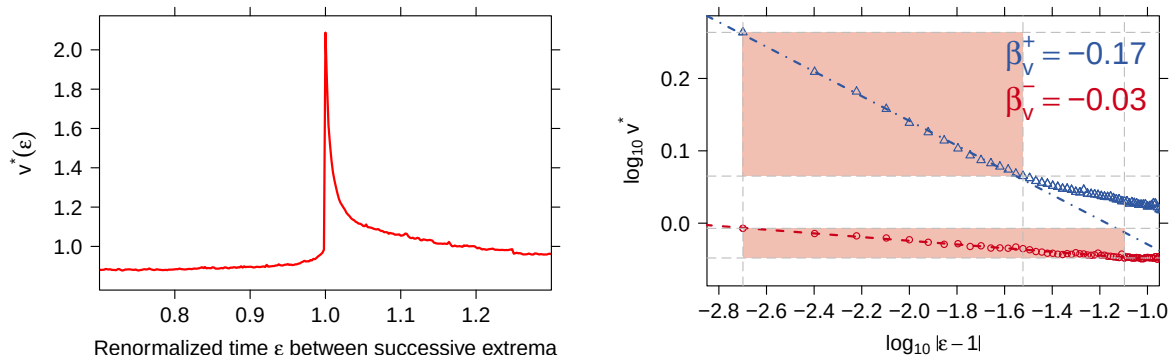

Figure A.18: *left*: aggregated volume $v^*(\varepsilon)$. *right*: $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram.

Table A.9: Statistical test of power-law hypothesis for the GE volume time series: Scaling parameters of the hypothesized power-law model are shown for both $v^*(\varepsilon)$ before (β_v^-) and $v^*(\varepsilon)$ after (β_v^+) the trend switching point $\varepsilon = 1$ in dependence of $|\varepsilon - 1|_{\text{cut}}$. Additionally, the corresponding values of the KS statistic, D_v^- and D_v^+, are given. The power-law hypothesis is supported if the p-value is larger than 0.1.

$\lvert\varepsilon - 1\rvert_{\text{cut}}$	β_v^+	D_v^+	p-value	β_v^-	D_v^-	p-value
0.004	−0.181	0	**1**	−0.023	0	**0.769**
0.006	−0.172	0.0014	**0.189**	−0.022	0.0003	**0.731**
0.008	−0.174	0.001	**0.681**	−0.021	0.0002	**0.919**
0.01	−0.176	0.0006	**0.943**	−0.022	0.0001	**0.97**
0.012	−0.177	0.0006	**0.995**	−0.023	0.0005	**0.61**
0.014	−0.177	0.0005	**1**	−0.024	0.0008	**0.209**
0.016	−0.177	0.0005	**1**	−0.024	0.0007	**0.351**
0.018	−0.177	0.0004	**1**	−0.025	0.0007	**0.279**
0.02	−0.177	0.0004	**1**	−0.025	0.0006	**0.372**
0.022	−0.176	0.0006	**1**	−0.025	0.0006	**0.326**
0.024	−0.174	0.0012	**0.998**	−0.025	0.0006	**0.42**
0.026	−0.173	0.0018	**0.928**	−0.025	0.0005	**0.442**
0.028	−0.171	0.002	**0.705**	−0.025	0.0005	**0.571**
0.03	−0.17	0.0023	**0.313**	−0.024	0.0004	**0.57**
0.032	−0.169	0.0027	0.031	−0.024	0.0004	**0.644**
0.034	−0.167	0.0029	0.009	−0.025	0.0004	**0.695**
0.036	−0.166	0.0031	0.001	−0.025	0.0005	**0.513**
0.038	−0.164	0.0033	0	−0.026	0.0007	**0.173**
0.04	−0.163	0.0036	0	−0.026	0.0008	0.087
0.042	−0.162	0.0039	0	−0.026	0.0009	0.031
0.044	−0.16	0.0042	0	−0.026	0.0009	0.01
0.046	−0.159	0.0044	0	−0.026	0.0009	0.014
0.048	−0.157	0.0048	0	−0.027	0.0009	0.012
0.05	−0.156	0.005	0	−0.026	0.0007	0.039
0.052	−0.154	0.0053	0	−0.026	0.0007	0.077
0.054	−0.153	0.0057	0	−0.026	0.0006	**0.108**
0.056	−0.151	0.0059	0	−0.026	0.0006	**0.141**
0.058	−0.15	0.0062	0	−0.026	0.0006	**0.141**
0.06	−0.148	0.0064	0	−0.026	0.0005	**0.221**
0.062	−0.147	0.0066	0	−0.025	0.0005	**0.317**
0.064	−0.146	0.0069	0	−0.025	0.0004	**0.37**
0.066	−0.144	0.0071	0	−0.025	0.0004	**0.399**
0.068	−0.143	0.0072	0	−0.025	0.0004	**0.409**
0.07	−0.142	0.0074	0	−0.025	0.0004	**0.453**
0.072	−0.141	0.0075	0	−0.025	0.0004	**0.449**
0.074	−0.14	0.0077	0	−0.025	0.0004	**0.464**
0.076	−0.138	0.0078	0	−0.025	0.0004	**0.516**

A.10 Home Depot, Inc. (The) Common Stock (HD)

The HD price time series contains 72,368,643 transactions.

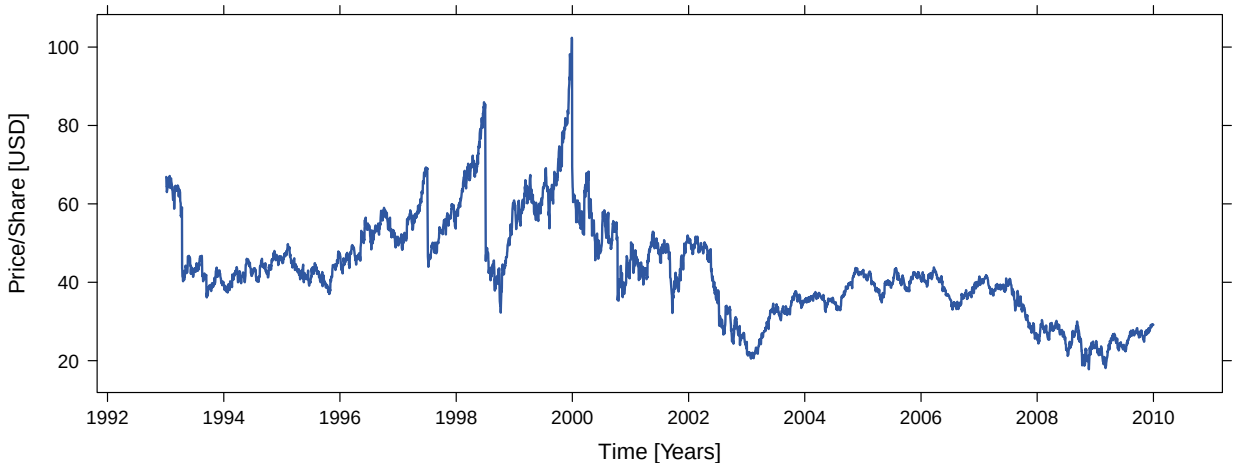

Figure A.19: Stock splits occurred on 14 April 1993 [split ratio 4:3], 7 July 1997 [split ratio 3:2], 6 July 1998 [split ratio 2:1], and 31 December 1999 [split ratio 3:2].

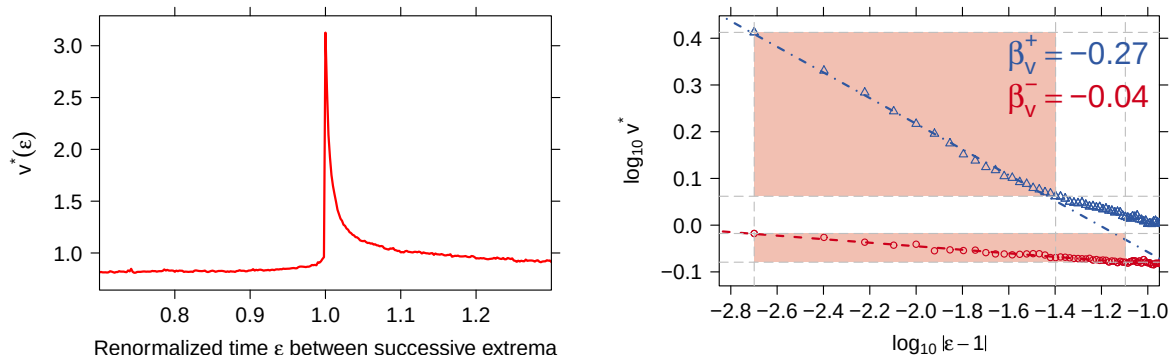

Figure A.20: *left*: aggregated volume $v^*(\varepsilon)$. *right*: $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram.

Table A.10: Statistical test of power-law hypothesis for the HD volume time series: Scaling parameters of the hypothesized power-law model are shown for both $v^*(\varepsilon)$ before (β_v^-) and $v^*(\varepsilon)$ after (β_v^+) the trend switching point $\varepsilon = 1$ in dependence of $|\varepsilon - 1|_{\text{cut}}$. Additionally, the corresponding values of the KS statistic, D_v^- and D_v^+, are given. The power-law hypothesis is supported if the p-value is larger than 0.1.

$\lvert\varepsilon - 1\rvert_{\text{cut}}$	β_v^+	D_v^+	p-value	β_v^-	D_v^-	p-value
0.004	−0.272	0	**1**	−0.027	0	**0.741**
0.006	−0.27	0.0003	**0.835**	−0.037	0.0017	**0.183**
0.008	−0.277	0.0019	**0.543**	−0.041	0.0016	**0.183**
0.01	−0.28	0.002	**0.74**	−0.037	0.0011	**0.383**
0.012	−0.28	0.0017	**0.919**	−0.043	0.0027	0.001
0.014	−0.281	0.0016	**0.986**	−0.043	0.0024	0.004
0.016	−0.286	0.0026	**0.846**	−0.043	0.002	0.021
0.018	−0.288	0.0031	**0.716**	−0.041	0.0015	0.084
0.02	−0.289	0.0033	**0.702**	−0.041	0.0014	**0.116**
0.022	−0.288	0.0029	**0.924**	−0.042	0.0013	**0.147**
0.024	−0.289	0.0028	**0.978**	−0.042	0.0012	**0.167**
0.026	−0.287	0.0023	**1**	−0.042	0.0011	**0.221**
0.028	−0.286	0.002	**1**	−0.041	0.001	**0.281**
0.03	−0.283	0.0017	**1**	−0.04	0.001	**0.3**
0.032	−0.282	0.0016	**1**	−0.039	0.001	**0.217**
0.034	−0.28	0.0022	**1**	−0.038	0.0014	0.036
0.036	−0.278	0.0026	**0.999**	−0.038	0.0015	0.01
0.038	−0.276	0.0033	**0.846**	−0.038	0.0013	0.04
0.04	−0.274	0.0037	**0.398**	−0.038	0.0012	0.077
0.042	−0.272	0.0042	0.031	−0.038	0.0012	0.063
0.044	−0.27	0.0048	0.001	−0.038	0.0012	0.05
0.046	−0.268	0.0051	0	−0.038	0.0011	0.042
0.048	−0.266	0.0054	0	−0.038	0.0011	0.071
0.05	−0.264	0.0058	0	−0.038	0.001	0.078
0.052	−0.262	0.0064	0	−0.038	0.001	0.069
0.054	−0.26	0.0068	0	−0.038	0.001	0.092
0.056	−0.258	0.0071	0	−0.037	0.001	0.078
0.058	−0.256	0.0075	0	−0.037	0.001	0.086
0.06	−0.254	0.0079	0	−0.037	0.0009	0.096
0.062	−0.252	0.0083	0	−0.037	0.0009	**0.126**
0.064	−0.25	0.0086	0	−0.037	0.0009	**0.132**
0.066	−0.248	0.0091	0	−0.037	0.0008	**0.129**
0.068	−0.246	0.0095	0	−0.037	0.0008	**0.144**
0.07	−0.244	0.0098	0	−0.037	0.0008	**0.18**
0.072	−0.242	0.0101	0	−0.037	0.0008	**0.203**
0.074	−0.241	0.0104	0	−0.037	0.0007	**0.203**
0.076	−0.239	0.0106	0	−0.037	0.0007	**0.226**

A.11 Hewlett-Packard Company Common Stock (HPQ)

The HPQ price time series contains 62,996,894 transactions.

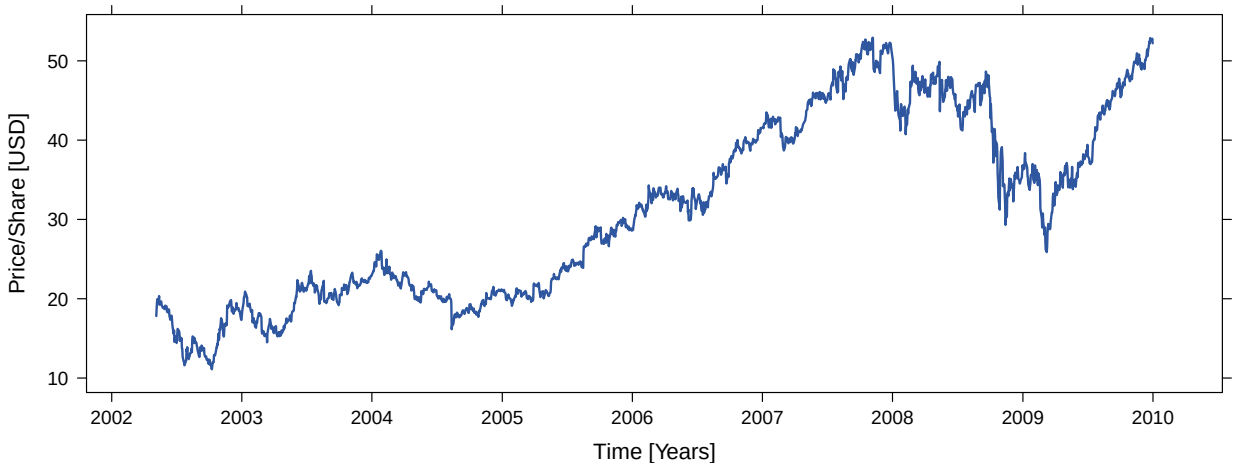

Figure A.21: Stock splits occurred on 17 April 1995 [split ratio 2:1], 16 July 1996 [split ratio 2:1], and 30 October 2000 [split ratio 2:1].

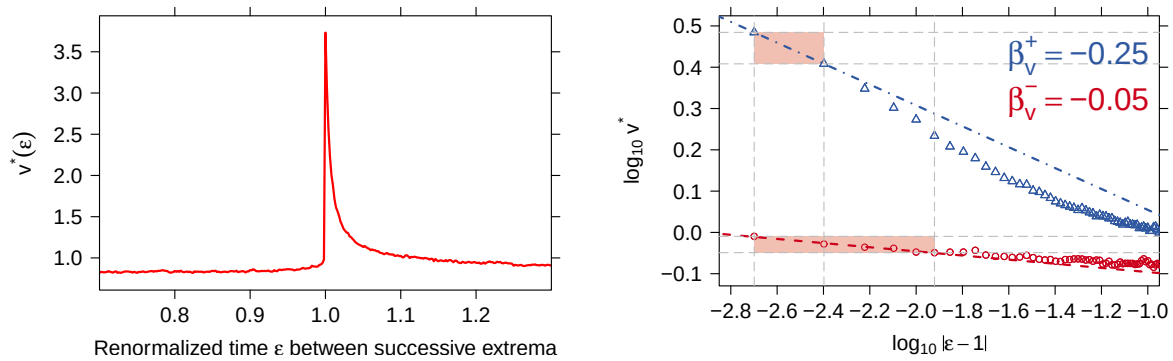

Figure A.22: *left*: aggregated volume $v^*(\varepsilon)$. *right*: $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram.

Table A.11: Statistical test of power-law hypothesis for the HPQ volume time series: Scaling parameters of the hypothesized power-law model are shown for both $v^*(\varepsilon)$ before (β_v^-) and $v^*(\varepsilon)$ after (β_v^+) the trend switching point $\varepsilon = 1$ in dependence of $|\varepsilon - 1|_{\text{cut}}$. Additionally, the corresponding values of the KS statistic, D_v^- and D_v^+, are given. The power-law hypothesis is supported if the p-value is larger than 0.1.

$\lvert\varepsilon - 1\rvert_{\text{cut}}$	β_v^+	D_v^+	p-value	β_v^-	D_v^-	p-value
0.004	−0.252	0	**1**	−0.061	0	**0.72**
0.006	−0.282	0.0048	0.006	−0.055	0.0009	**0.473**
0.008	−0.301	0.0067	0	−0.049	0.0021	0.09
0.01	−0.306	0.0065	0	−0.051	0.0013	**0.341**
0.012	−0.319	0.0083	0	−0.05	0.0012	**0.369**
0.014	−0.329	0.01	0	−0.046	0.0019	0.064
0.016	−0.331	0.0095	0	−0.043	0.0028	0
0.018	−0.331	0.0088	0	−0.038	0.0039	0
0.02	−0.334	0.0084	0	−0.039	0.0033	0
0.022	−0.335	0.008	0	−0.04	0.0029	0
0.024	−0.337	0.0076	0	−0.04	0.0026	0
0.026	−0.337	0.0071	0	−0.04	0.0024	0
0.028	−0.336	0.0068	0	−0.04	0.0022	0
0.03	−0.333	0.0064	0	−0.039	0.0021	0
0.032	−0.332	0.0061	0.001	−0.039	0.002	0.002
0.034	−0.33	0.0059	0	−0.039	0.0019	0.002
0.036	−0.329	0.0057	0.001	−0.04	0.0018	0.002
0.038	−0.327	0.0056	0.002	−0.04	0.0017	0.007
0.04	−0.326	0.0054	0.003	−0.04	0.0016	0.008
0.042	−0.325	0.0053	0.009	−0.041	0.0016	0.008
0.044	−0.323	0.0052	0.01	−0.04	0.0015	0.011
0.046	−0.322	0.0052	0.013	−0.04	0.0014	0.018
0.048	−0.32	0.0051	0.006	−0.039	0.0013	0.02
0.05	−0.319	0.0051	0.004	−0.038	0.0017	0
0.052	−0.316	0.0052	0	−0.038	0.0019	0
0.054	−0.314	0.0052	0.001	−0.037	0.002	0
0.056	−0.312	0.0053	0.001	−0.037	0.0021	0
0.058	−0.31	0.0053	0	−0.037	0.0021	0
0.06	−0.309	0.0053	0	−0.036	0.0022	0
0.062	−0.306	0.0055	0	−0.036	0.0022	0
0.064	−0.304	0.006	0	−0.036	0.002	0
0.066	−0.302	0.0065	0	−0.037	0.0019	0
0.068	−0.301	0.007	0	−0.036	0.0019	0
0.07	−0.299	0.0075	0	−0.037	0.0018	0
0.072	−0.297	0.0079	0	−0.037	0.0016	0
0.074	−0.296	0.0081	0	−0.037	0.0016	0
0.076	−0.294	0.0083	0	−0.037	0.0016	0

A.12 International Business Machines (IBM)

The IBM price time series contains 49,386,841 transactions.

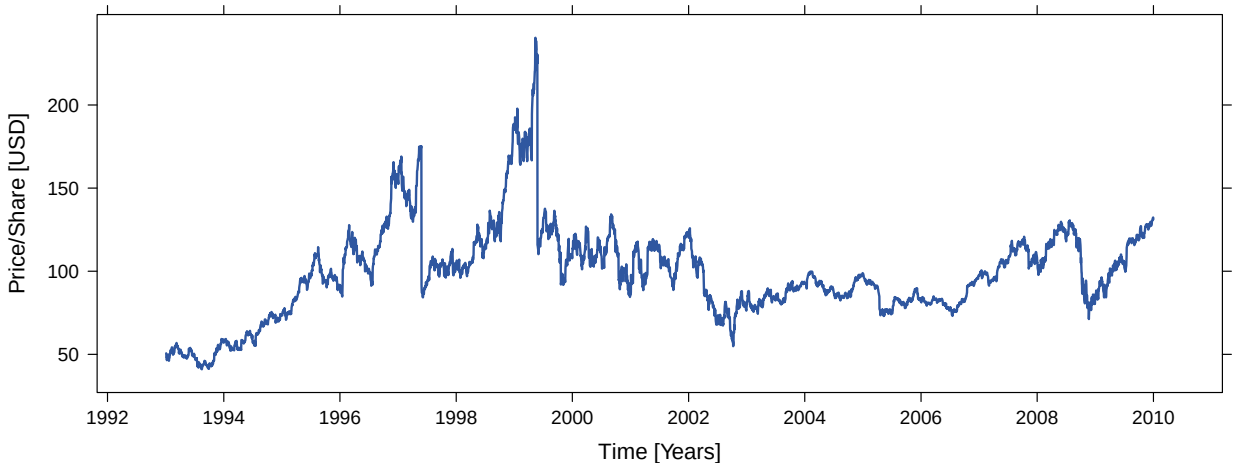

Figure A.23: Stock splits occurred on 28 May 1997 [split ratio 2:1] and 27 May 1999 [split ratio 2:1].

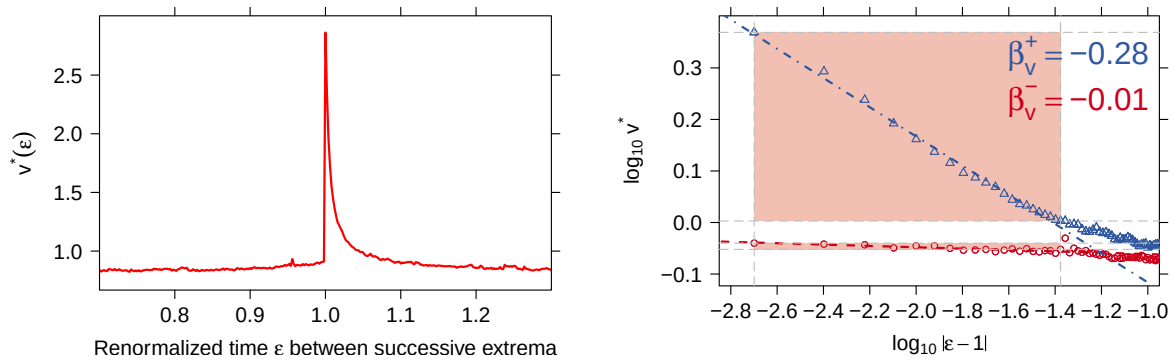

Figure A.24: *left*: aggregated volume $v^*(\varepsilon)$. *right*: $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram.

Table A.12: Statistical test of power-law hypothesis for the IBM volume time series: Scaling parameters of the hypothesized power-law model are shown for both $v^*(\varepsilon)$ before (β_v^-) and $v^*(\varepsilon)$ after (β_v^+) the trend switching point $\varepsilon = 1$ in dependence of $|\varepsilon - 1|_{\text{cut}}$. Additionally, the corresponding values of the KS statistic, D_v^- and D_v^+, are given. The power-law hypothesis is supported if the p-value is larger than 0.1.

$\lvert\varepsilon - 1\rvert_{\text{cut}}$	β_v^+	D_v^+	p-value	β_v^-	D_v^-	p-value
0.004	−0.251	0	**0.776**	−0.006	0	**1**
0.006	−0.271	0.0033	0.068	−0.006	0	**1**
0.008	−0.29	0.006	0.001	−0.013	0.0018	**0.215**
0.01	−0.298	0.0065	0	−0.01	0.0009	**0.634**
0.012	−0.302	0.006	0	−0.008	0.0008	**0.677**
0.014	−0.305	0.0056	0.003	−0.01	0.0007	**0.811**
0.016	−0.307	0.0051	0.032	−0.012	0.0014	**0.17**
0.018	−0.306	0.0046	0.096	−0.013	0.0015	**0.111**
0.02	−0.303	0.0043	**0.27**	−0.013	0.0014	**0.133**
0.022	−0.3	0.004	**0.416**	−0.014	0.0015	0.066
0.024	−0.299	0.0038	**0.577**	−0.014	0.0014	**0.103**
0.026	−0.298	0.0036	**0.735**	−0.013	0.0011	**0.196**
0.028	−0.297	0.0034	**0.846**	−0.014	0.0011	**0.195**
0.03	−0.295	0.0034	**0.869**	−0.013	0.0009	**0.267**
0.032	−0.293	0.0033	**0.934**	−0.013	0.0009	**0.311**
0.034	−0.291	0.0032	**0.95**	−0.013	0.0008	**0.329**
0.036	−0.289	0.0032	**0.928**	−0.013	0.0008	**0.343**
0.038	−0.287	0.0032	**0.919**	−0.013	0.0007	**0.368**
0.04	−0.285	0.0035	**0.774**	−0.013	0.0007	**0.431**
0.042	−0.283	0.004	**0.314**	−0.013	0.0007	**0.428**
0.044	−0.28	0.0046	0.018	−0.009	0.0027	0
0.046	−0.278	0.0051	0	−0.01	0.0023	0
0.048	−0.276	0.0057	0	−0.009	0.0025	0
0.05	−0.273	0.0064	0	−0.009	0.0024	0
0.052	−0.271	0.0067	0	−0.009	0.0021	0
0.054	−0.27	0.007	0	−0.009	0.002	0
0.056	−0.268	0.0074	0	−0.01	0.0018	0
0.058	−0.265	0.0079	0	−0.01	0.0016	0
0.06	−0.262	0.0086	0	−0.011	0.0014	0.002
0.062	−0.26	0.009	0	−0.011	0.0012	0.007
0.064	−0.258	0.0094	0	−0.012	0.0014	0
0.066	−0.256	0.0097	0	−0.012	0.0015	0
0.068	−0.254	0.0101	0	−0.013	0.0018	0
0.07	−0.252	0.0104	0	−0.014	0.0022	0
0.072	−0.25	0.0105	0	−0.014	0.0025	0
0.074	−0.249	0.0106	0	−0.015	0.0026	0
0.076	−0.247	0.0109	0	−0.016	0.0028	0

A.13 Intel Corporation (INTC)

The INTC price time series contains 253,972,099 transactions.

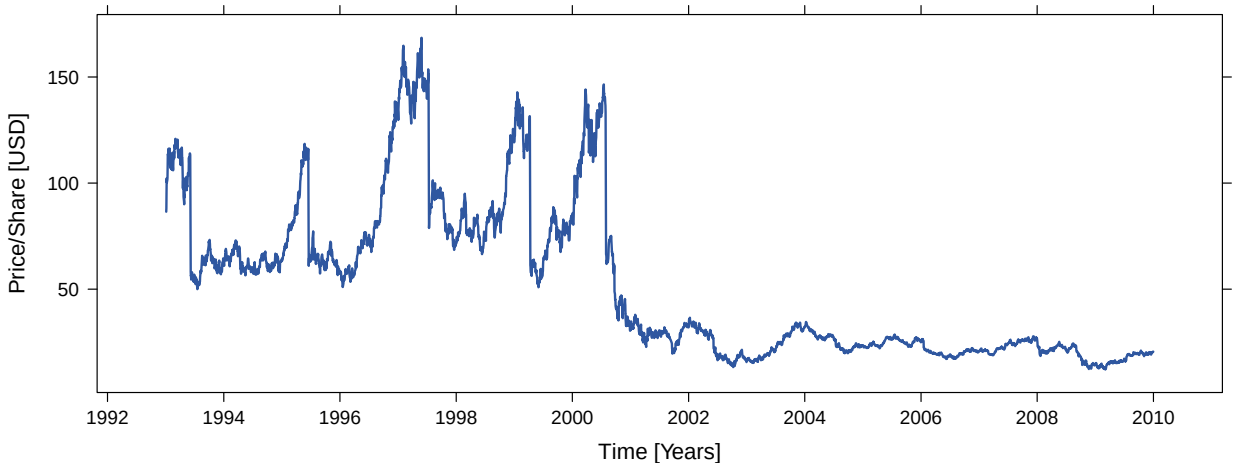

Figure A.25: Stock splits occurred on 7 June 1993 [split ratio 2:1], 19 June 1995 [split ratio 2:1], 14 July 1997 [split ratio 2:1], 12 April 1999 [split ratio 2:1], and 31 July 2000 [split ratio 2:1].

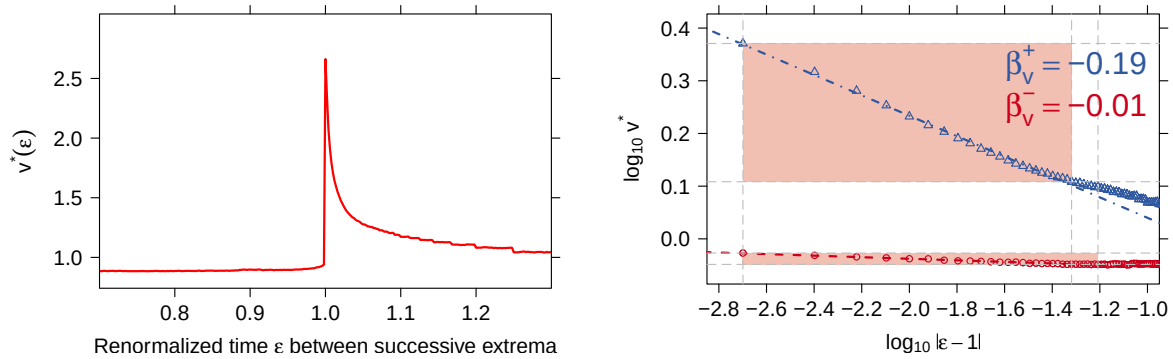

Figure A.26: *left*: aggregated volume $v^*(\varepsilon)$. *right*: $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram.

Table A.13: Statistical test of power-law hypothesis for the INTC volume time series: Scaling parameters of the hypothesized power-law model are shown for both $v^*(\varepsilon)$ before (β_v^-) and $v^*(\varepsilon)$ after (β_v^+) the trend switching point $\varepsilon = 1$ in dependence of $|\varepsilon - 1|_{\text{cut}}$. Additionally, the corresponding values of the KS statistic, D_v^- and D_v^+, are given. The power-law hypothesis is supported if the p-value is larger than 0.1.

$\lvert\varepsilon - 1\rvert_{\text{cut}}$	β_v^+	D_v^+	p-value	β_v^-	D_v^-	p-value
0.004	−0.179	0	**1**	−0.014	0	**0.793**
0.006	−0.187	0.0012	**0.275**	−0.015	0	**0.951**
0.008	−0.194	0.0022	0.064	−0.015	0	**0.998**
0.01	−0.198	0.0025	0.032	−0.015	0	**0.995**
0.012	−0.2	0.0025	0.075	−0.014	0.0002	**0.904**
0.014	−0.201	0.0023	**0.306**	−0.015	0.0002	**0.908**
0.016	−0.202	0.0021	**0.626**	−0.015	0.0003	**0.626**
0.018	−0.202	0.0019	**0.896**	−0.015	0.0003	**0.615**
0.02	−0.202	0.0018	**0.986**	−0.015	0.0003	**0.671**
0.022	−0.202	0.0017	**0.999**	−0.015	0.0003	**0.695**
0.024	−0.202	0.0015	**1**	−0.016	0.0003	**0.739**
0.026	−0.202	0.0014	**1**	−0.016	0.0002	**0.774**
0.028	−0.202	0.0013	**1**	−0.015	0.0002	**0.843**
0.03	−0.201	0.0013	**1**	−0.015	0.0002	**0.888**
0.032	−0.201	0.0012	**1**	−0.015	0.0002	**0.91**
0.034	−0.2	0.0012	**1**	−0.015	0.0002	**0.93**
0.036	−0.2	0.0012	**1**	−0.015	0.0001	**0.934**
0.038	−0.198	0.0012	**1**	−0.015	0.0001	**0.944**
0.04	−0.198	0.0012	**1**	−0.015	0.0001	**0.957**
0.042	−0.197	0.0014	**1**	−0.015	0.0001	**0.962**
0.044	−0.195	0.0018	**1**	−0.015	0.0001	**0.967**
0.046	−0.194	0.0022	**0.954**	−0.015	0.0001	**0.979**
0.048	−0.193	0.0025	**0.385**	−0.015	0.0001	**0.98**
0.05	−0.192	0.0027	0.025	−0.015	0.0001	**0.975**
0.052	−0.191	0.003	0	−0.015	0.0001	**0.946**
0.054	−0.19	0.0032	0	−0.015	0.0002	**0.819**
0.056	−0.189	0.0035	0	−0.015	0.0002	**0.569**
0.058	−0.188	0.0037	0	−0.015	0.0003	**0.488**
0.06	−0.187	0.004	0	−0.014	0.0003	**0.267**
0.062	−0.185	0.0043	0	−0.014	0.0004	**0.173**
0.064	−0.184	0.0045	0	−0.014	0.0004	0.097
0.066	−0.183	0.0047	0	−0.014	0.0004	0.045
0.068	−0.182	0.0049	0	−0.014	0.0005	0.031
0.07	−0.181	0.0051	0	−0.014	0.0005	0.011
0.072	−0.18	0.0053	0	−0.014	0.0006	0.002
0.074	−0.179	0.0055	0	−0.013	0.0006	0.001
0.076	−0.178	0.0057	0	−0.013	0.0007	0.001

A.14 Johnson & Johnson Common Stock (JNJ)

The JNJ price time series contains 56,780,934 transactions.

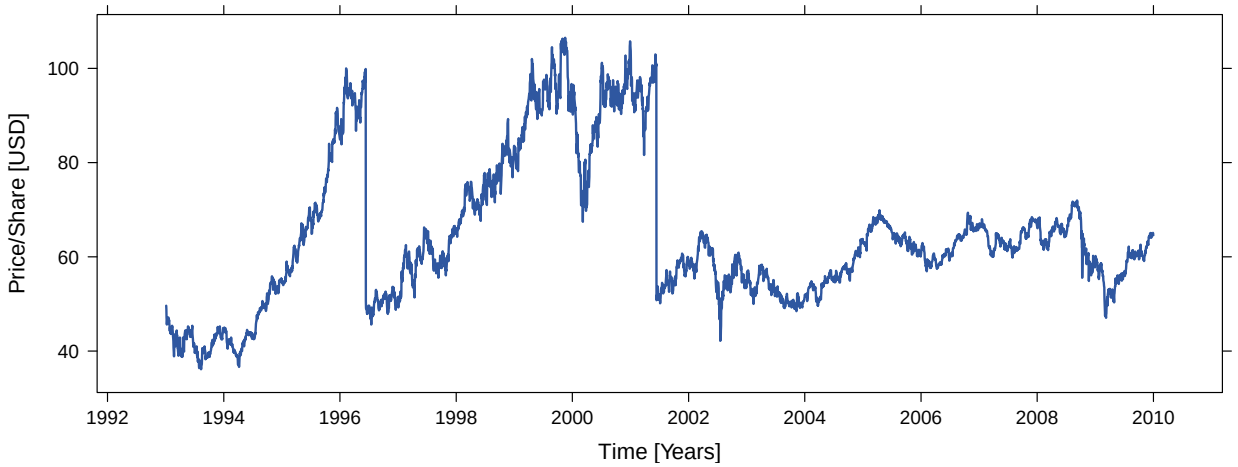

Figure A.27: Stock splits occurred on 12 June 1996 [split ratio 2:1] and 13 June 2001 [split ratio 2:1].

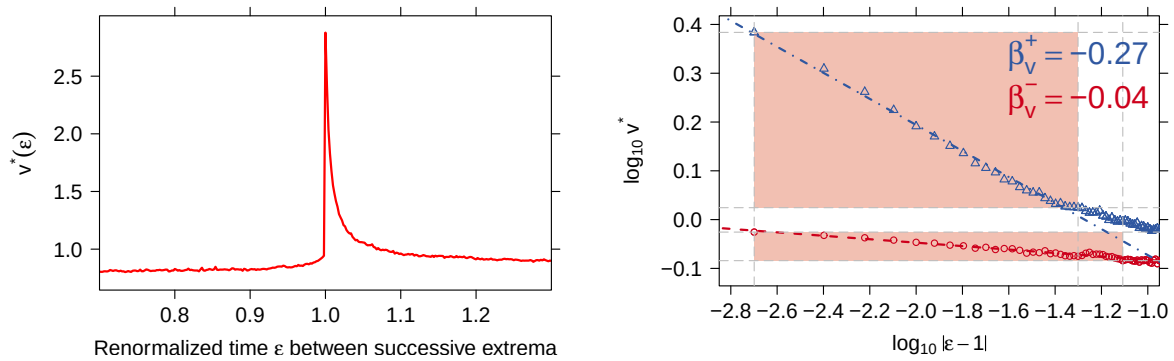

Figure A.28: *left*: aggregated volume $v^*(\varepsilon)$. *right*: $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram.

Table A.14: Statistical test of power-law hypothesis for the JNJ volume time series: Scaling parameters of the hypothesized power-law model are shown for both $v^*(\varepsilon)$ before (β_v^-) and $v^*(\varepsilon)$ after (β_v^+) the trend switching point $\varepsilon = 1$ in dependence of $|\varepsilon - 1|_{\text{cut}}$. Additionally, the corresponding values of the KS statistic, D_v^- and D_v^+, are given. The power-law hypothesis is supported if the p-value is larger than 0.1.

$\lvert\varepsilon - 1\rvert_{\text{cut}}$	β_v^+	D_v^+	p-value	β_v^-	D_v^-	p-value
0.004	−0.247	0	**0.758**	−0.022	0	**1**
0.006	−0.254	0.0012	**0.508**	−0.024	0.0003	**0.826**
0.008	−0.262	0.0025	**0.262**	−0.026	0.0007	**0.702**
0.01	−0.272	0.0043	0.025	−0.029	0.0011	**0.478**
0.012	−0.276	0.0046	0.018	−0.029	0.0011	**0.428**
0.014	−0.278	0.0044	0.042	−0.031	0.0012	**0.31**
0.016	−0.279	0.0041	**0.143**	−0.032	0.0012	**0.297**
0.018	−0.282	0.004	**0.21**	−0.033	0.0015	**0.113**
0.02	−0.283	0.0037	**0.411**	−0.033	0.0013	**0.152**
0.022	−0.283	0.0034	**0.646**	−0.034	0.0014	**0.117**
0.024	−0.283	0.0032	**0.799**	−0.034	0.0012	**0.175**
0.026	−0.282	0.003	**0.928**	−0.034	0.0011	**0.26**
0.028	−0.282	0.0028	**0.961**	−0.034	0.001	**0.245**
0.03	−0.282	0.0027	**0.986**	−0.035	0.0012	**0.168**
0.032	−0.281	0.0026	**0.996**	−0.035	0.0013	**0.105**
0.034	−0.279	0.0026	**0.997**	−0.036	0.0013	0.056
0.036	−0.277	0.0025	**0.994**	−0.035	0.001	**0.226**
0.038	−0.277	0.0025	**0.996**	−0.035	0.0011	**0.144**
0.04	−0.276	0.0025	**0.999**	−0.035	0.001	**0.212**
0.042	−0.274	0.0025	**0.999**	−0.036	0.001	**0.146**
0.044	−0.272	0.0025	**0.997**	−0.036	0.0011	0.099
0.046	−0.271	0.0028	**0.988**	−0.037	0.0012	0.076
0.048	−0.269	0.0033	**0.753**	−0.037	0.0012	0.064
0.05	−0.266	0.0039	**0.117**	−0.037	0.0012	0.058
0.052	−0.264	0.0048	0	−0.037	0.0011	0.067
0.054	−0.262	0.0053	0	−0.036	0.001	**0.145**
0.056	−0.26	0.0058	0	−0.035	0.0008	**0.264**
0.058	−0.258	0.0063	0	−0.035	0.0008	**0.277**
0.06	−0.256	0.0069	0	−0.035	0.001	0.075
0.062	−0.253	0.0076	0	−0.034	0.0012	0.011
0.064	−0.252	0.0081	0	−0.034	0.0014	0.004
0.066	−0.25	0.0084	0	−0.034	0.0013	0.004
0.068	−0.249	0.0086	0	−0.034	0.0013	0.003
0.07	−0.247	0.0088	0	−0.034	0.0012	0.006
0.072	−0.245	0.009	0	−0.034	0.0012	0.01
0.074	−0.244	0.0092	0	−0.034	0.0011	0.026
0.076	−0.243	0.0093	0	−0.034	0.001	0.052

A.15 JP Morgan Chase & Co. Common Stock (JPM)

The JPM price time series contains 150,161,709 transactions.

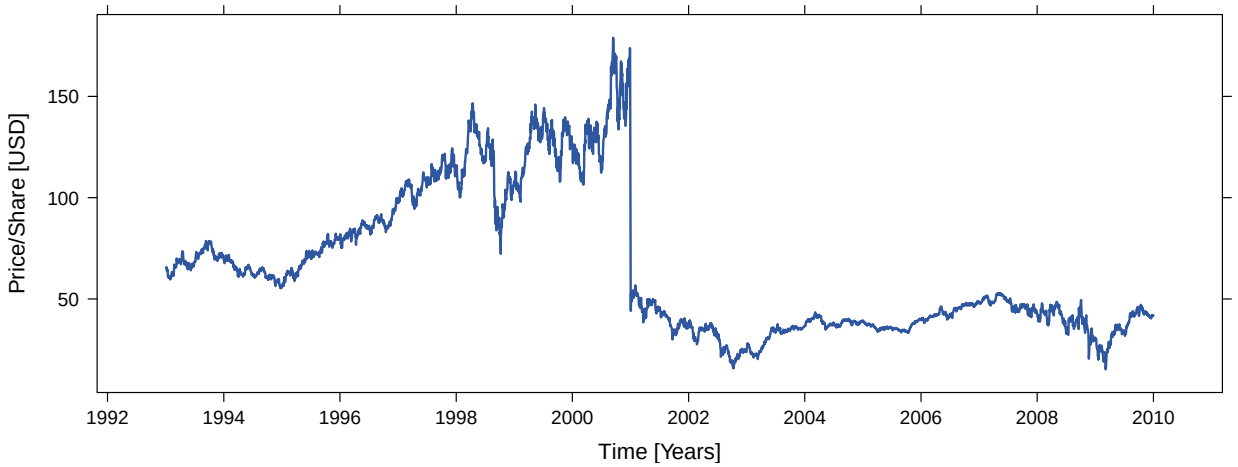

Figure A.29: Stock splits occurred on 15 June 1998 [split ratio 2:1] and 12 June 2000 [split ratio 3:2].

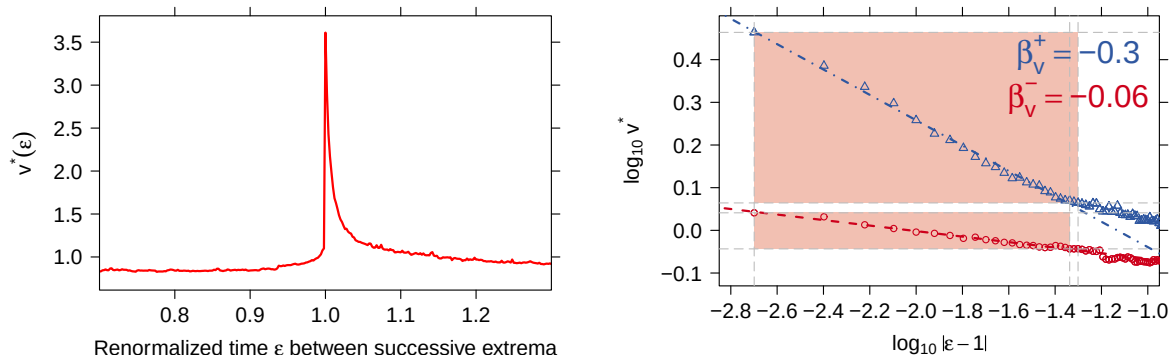

Figure A.30: *left*: aggregated volume $v^*(\varepsilon)$. *right*: $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram.

Table A.15: Statistical test of power-law hypothesis for the JPM volume time series: Scaling parameters of the hypothesized power-law model are shown for both $v^*(\varepsilon)$ before (β_v^-) and $v^*(\varepsilon)$ after (β_v^+) the trend switching point $\varepsilon = 1$ in dependence of $|\varepsilon - 1|_{\text{cut}}$. Additionally, the corresponding values of the KS statistic, D_v^- and D_v^+, are given. The power-law hypothesis is supported if the p-value is larger than 0.1.

$\lvert\varepsilon - 1\rvert_{\text{cut}}$	β_v^+	D_v^+	p-value	β_v^-	D_v^-	p-value
0.004	−0.26	0	**0.779**	−0.031	0	**0.747**
0.006	−0.267	0.0011	**0.508**	−0.055	0.0041	0
0.008	−0.274	0.0023	**0.259**	−0.062	0.0037	0
0.01	−0.288	0.0058	0	−0.066	0.0032	0
0.012	−0.301	0.008	0	−0.066	0.0027	0.001
0.014	−0.304	0.0076	0	−0.067	0.0023	0.001
0.016	−0.306	0.0071	0	−0.068	0.0021	0.006
0.018	−0.31	0.007	0	−0.066	0.0018	0.01
0.02	−0.313	0.0068	0	−0.066	0.0016	0.028
0.022	−0.313	0.0063	0	−0.066	0.0015	0.051
0.024	−0.314	0.006	0	−0.067	0.0014	0.083
0.026	−0.315	0.0057	0	−0.066	0.0013	**0.134**
0.028	−0.313	0.0053	0	−0.067	0.0012	**0.2**
0.03	−0.312	0.005	0.001	−0.067	0.0011	**0.297**
0.032	−0.31	0.0048	0.009	−0.067	0.001	**0.365**
0.034	−0.308	0.0047	0.013	−0.067	0.001	**0.459**
0.036	−0.306	0.0045	0.052	−0.067	0.0009	**0.544**
0.038	−0.305	0.0044	**0.104**	−0.067	0.0009	**0.572**
0.04	−0.304	0.0043	**0.182**	−0.066	0.0009	**0.548**
0.042	−0.303	0.0042	**0.29**	−0.065	0.0009	**0.483**
0.044	−0.302	0.0041	**0.337**	−0.064	0.001	**0.29**
0.046	−0.3	0.0041	**0.287**	−0.064	0.001	**0.189**
0.048	−0.299	0.0041	**0.211**	−0.064	0.0011	0.067
0.05	−0.297	0.0041	**0.128**	−0.063	0.0013	0.018
0.052	−0.295	0.0042	0.05	−0.063	0.0014	0.001
0.054	−0.293	0.0042	0.014	−0.063	0.0013	0.011
0.056	−0.291	0.0043	0.001	−0.063	0.0014	0.002
0.058	−0.288	0.0054	0	−0.063	0.0014	0.002
0.06	−0.285	0.0063	0	−0.062	0.0014	0
0.062	−0.283	0.0067	0	−0.062	0.0015	0
0.064	−0.282	0.007	0	−0.063	0.0012	0.003
0.066	−0.28	0.0074	0	−0.064	0.0016	0
0.068	−0.277	0.0081	0	−0.065	0.0023	0
0.07	−0.276	0.0085	0	−0.066	0.0027	0
0.072	−0.274	0.0089	0	−0.067	0.0031	0
0.074	−0.271	0.0097	0	−0.067	0.0032	0
0.076	−0.269	0.01	0	−0.067	0.0032	0

A.16 Kraft Foods Inc. Common Stock (KFT)

The KFT price time series contains 34,206,284 transactions.

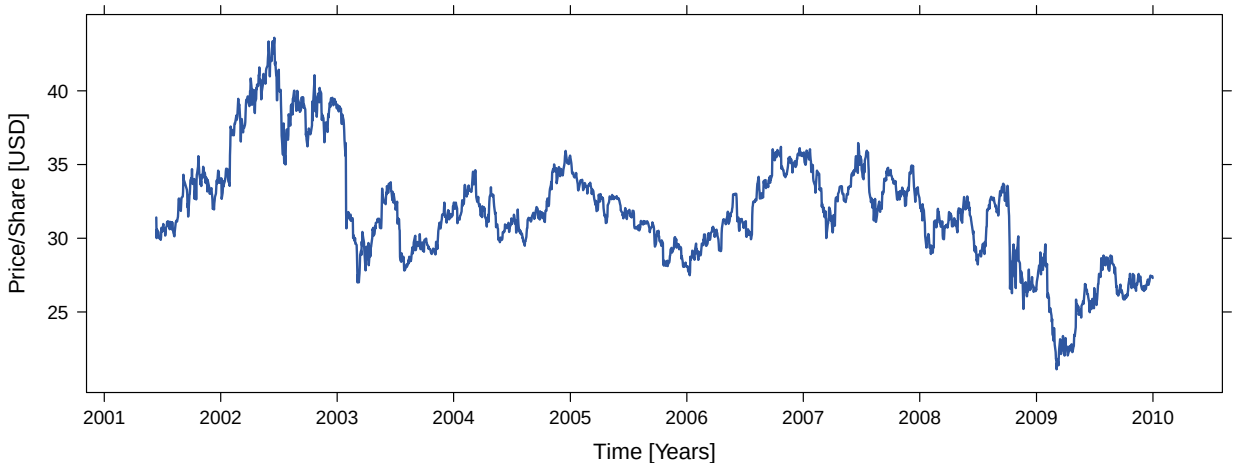

Figure A.31: There was no stock split in this period of time.

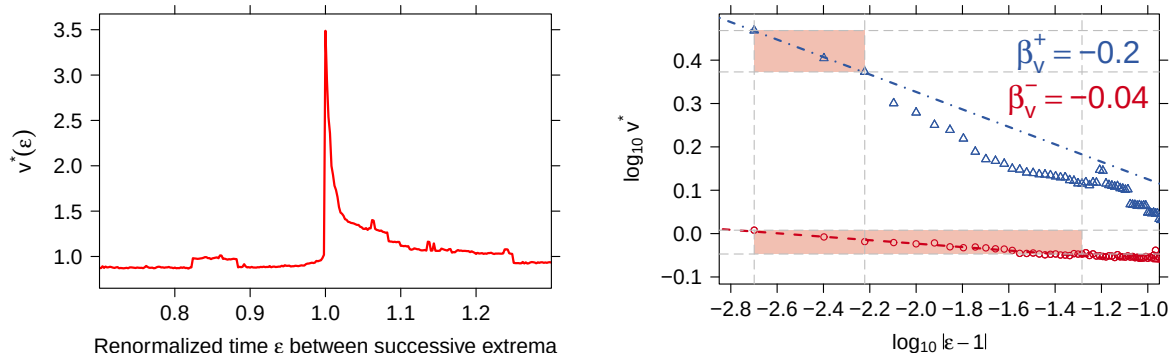

Figure A.32: *left*: aggregated volume $v^*(\varepsilon)$. *right*: $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram.

Table A.16: Statistical test of power-law hypothesis for the KFT volume time series: Scaling parameters of the hypothesized power-law model are shown for both $v^*(\varepsilon)$ before (β_v^-) and $v^*(\varepsilon)$ after (β_v^+) the trend switching point $\varepsilon = 1$ in dependence of $|\varepsilon - 1|_{\text{cut}}$. Additionally, the corresponding values of the KS statistic, D_v^- and D_v^+, are given. The power-law hypothesis is supported if the p-value is larger than 0.1.

$\lvert\varepsilon - 1\rvert_{\text{cut}}$	β_v^+	D_v^+	p-value	β_v^-	D_v^-	p-value
0.004	−0.212	0	**1**	−0.051	0	**1**
0.006	−0.201	0.0017	**0.393**	−0.054	0.0005	**0.783**
0.008	−0.258	0.0147	0	−0.049	0.0013	**0.526**
0.01	−0.273	0.0149	0	−0.046	0.0016	**0.366**
0.012	−0.284	0.0142	0	−0.04	0.0028	0.045
0.014	−0.285	0.0126	0	−0.041	0.0021	**0.128**
0.016	−0.288	0.0115	0	−0.041	0.0018	**0.191**
0.018	−0.296	0.0106	0	−0.04	0.0018	**0.151**
0.02	−0.303	0.0111	0	−0.039	0.0017	**0.151**
0.022	−0.304	0.0109	0	−0.038	0.0016	**0.186**
0.024	−0.304	0.0101	0	−0.038	0.0016	**0.164**
0.026	−0.304	0.0095	0	−0.038	0.0014	**0.214**
0.028	−0.302	0.0086	0	−0.039	0.0013	**0.317**
0.03	−0.3	0.0079	0	−0.04	0.0012	**0.352**
0.032	−0.297	0.0072	0	−0.041	0.0011	**0.398**
0.034	−0.294	0.0071	0	−0.041	0.001	**0.424**
0.036	−0.29	0.007	0	−0.041	0.001	**0.424**
0.038	−0.285	0.007	0	−0.041	0.001	**0.476**
0.04	−0.281	0.007	0	−0.041	0.0009	**0.555**
0.042	−0.277	0.007	0	−0.041	0.0008	**0.579**
0.044	−0.273	0.007	0	−0.041	0.0008	**0.582**
0.046	−0.27	0.0074	0	−0.041	0.0008	**0.58**
0.048	−0.266	0.0082	0	−0.04	0.0007	**0.695**
0.05	−0.264	0.0088	0	−0.04	0.001	**0.315**
0.052	−0.261	0.0093	0	−0.039	0.0013	**0.103**
0.054	−0.257	0.0099	0	−0.038	0.0018	0.003
0.056	−0.255	0.0103	0	−0.038	0.0018	0.002
0.058	−0.251	0.0108	0	−0.038	0.0021	0
0.06	−0.248	0.0113	0	−0.037	0.002	0.001
0.062	−0.242	0.0127	0	−0.037	0.002	0
0.064	−0.236	0.014	0	−0.037	0.002	0.001
0.066	−0.233	0.0143	0	−0.037	0.002	0
0.068	−0.231	0.0148	0	−0.037	0.002	0
0.07	−0.229	0.0152	0	−0.037	0.002	0
0.072	−0.226	0.0156	0	−0.037	0.002	0
0.074	−0.224	0.016	0	−0.036	0.0021	0
0.076	−0.222	0.0162	0	−0.036	0.002	0

A.17 Coca-Cola Company (The) Common Stock (KO)

The KO price time series contains 47,554,230 transactions.

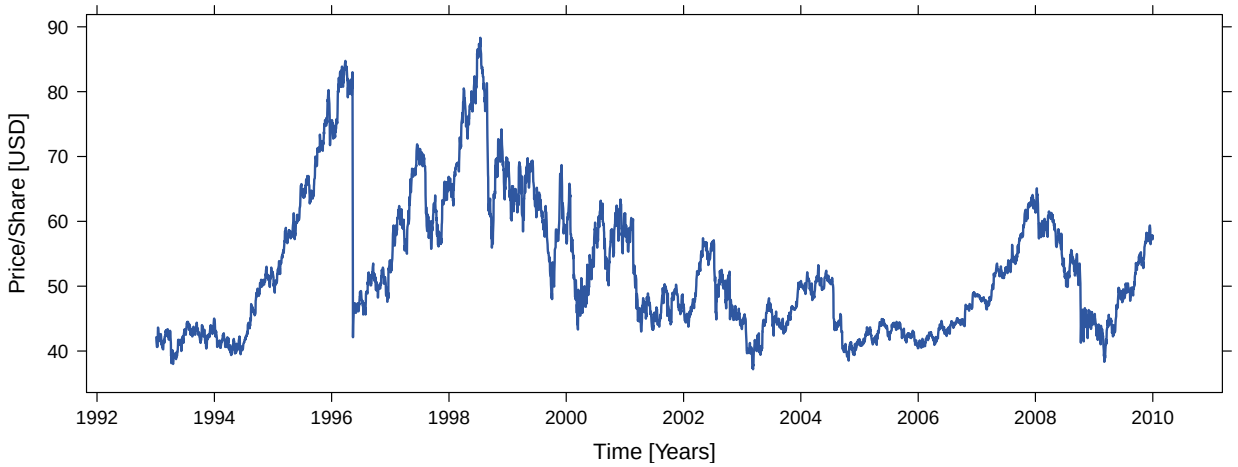

Figure A.33: A stock split occurred on 13 May 1996 [split ratio 2:1].

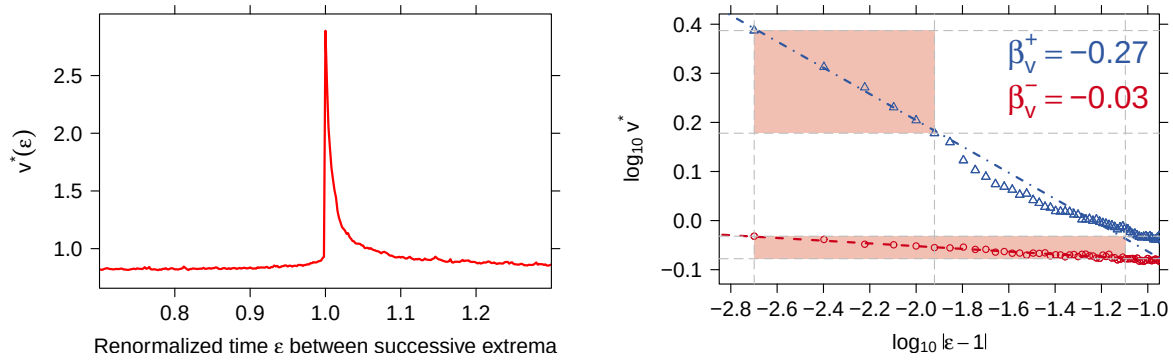

Figure A.34: *left*: aggregated volume $v^*(\varepsilon)$. *right*: $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram.

Table A.17: Statistical test of power-law hypothesis for the KO volume time series: Scaling parameters of the hypothesized power-law model are shown for both $v^*(\varepsilon)$ before (β_v^-) and $v^*(\varepsilon)$ after (β_v^+) the trend switching point $\varepsilon = 1$ in dependence of $|\varepsilon - 1|_{\text{cut}}$. Additionally, the corresponding values of the KS statistic, D_v^- and D_v^+, are given. The power-law hypothesis is supported if the p-value is larger than 0.1.

$\lvert\varepsilon - 1\rvert_{\text{cut}}$	β_v^+	D_v^+	p-value	β_v^-	D_v^-	p-value
0.004	−0.246	0	**1**	−0.022	0	**0.775**
0.006	−0.243	0.0005	**0.772**	−0.032	0.0018	**0.23**
0.008	−0.255	0.003	**0.166**	−0.031	0.0012	**0.455**
0.01	−0.26	0.0034	**0.142**	−0.03	0.0009	**0.657**
0.012	−0.267	0.0038	**0.17**	−0.03	0.0007	**0.785**
0.014	−0.27	0.0042	0.078	−0.029	0.0006	**0.847**
0.016	−0.283	0.0077	0	−0.027	0.0009	**0.57**
0.018	−0.293	0.0107	0	−0.027	0.0008	**0.69**
0.02	−0.299	0.0118	0	−0.027	0.0008	**0.591**
0.022	−0.305	0.0122	0	−0.028	0.0005	**0.875**
0.024	−0.307	0.0118	0	−0.028	0.0006	**0.818**
0.026	−0.307	0.0111	0	−0.029	0.0009	**0.48**
0.028	−0.307	0.0106	0	−0.029	0.0009	**0.423**
0.03	−0.305	0.0098	0	−0.03	0.0011	**0.212**
0.032	−0.304	0.0092	0	−0.03	0.001	**0.303**
0.034	−0.303	0.0088	0	−0.029	0.0009	**0.411**
0.036	−0.302	0.0084	0	−0.029	0.0007	**0.567**
0.038	−0.3	0.008	0	−0.029	0.0007	**0.475**
0.04	−0.299	0.0077	0	−0.029	0.0007	**0.542**
0.042	−0.297	0.0075	0	−0.029	0.0007	**0.549**
0.044	−0.294	0.0072	0	−0.03	0.0006	**0.55**
0.046	−0.291	0.0071	0	−0.029	0.0006	**0.613**
0.048	−0.289	0.007	0	−0.029	0.0006	**0.677**
0.05	−0.286	0.0069	0	−0.029	0.0005	**0.736**
0.052	−0.285	0.0068	0	−0.028	0.0005	**0.689**
0.054	−0.282	0.0067	0	−0.028	0.0008	**0.251**
0.056	−0.28	0.0067	0	−0.028	0.0009	**0.199**
0.058	−0.278	0.0067	0	−0.028	0.0009	**0.151**
0.06	−0.275	0.0067	0	−0.028	0.0008	**0.243**
0.062	−0.273	0.0071	0	−0.028	0.0007	**0.38**
0.064	−0.271	0.0076	0	−0.028	0.0008	**0.194**
0.066	−0.269	0.0079	0	−0.027	0.0008	**0.158**
0.068	−0.267	0.0083	0	−0.027	0.001	0.056
0.07	−0.265	0.0086	0	−0.027	0.0009	0.083
0.072	−0.263	0.0088	0	−0.027	0.0009	0.076
0.074	−0.261	0.009	0	−0.027	0.0008	**0.159**
0.076	−0.259	0.0093	0	−0.027	0.0007	**0.239**

A.18 McDonald's Corporation Common Stock (MCD)

The MCD price time series contains 42,285,966 transactions.

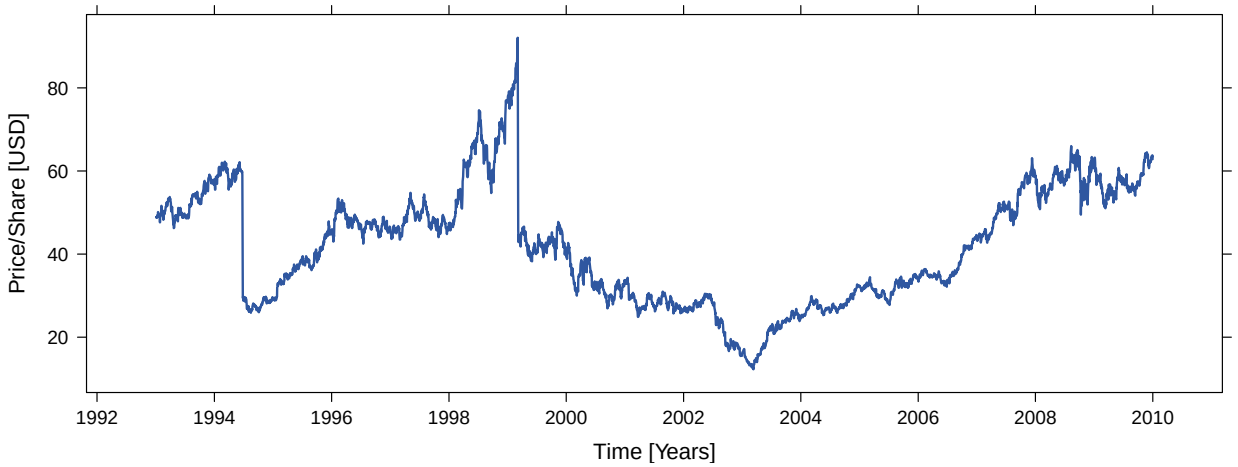

Figure A.35: Stock splits occurred on 27 June 1994 [split ratio 2:1] and 8 March 1999 [split ratio 2:1].

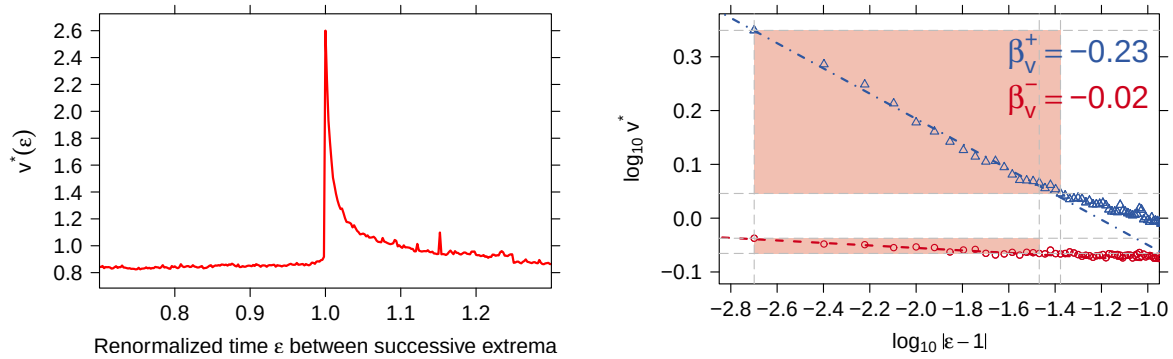

Figure A.36: *left*: aggregated volume $v^*(\varepsilon)$. *right*: $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram.

Table A.18: Statistical test of power-law hypothesis for the MCD volume time series: Scaling parameters of the hypothesized power-law model are shown for both $v^*(\varepsilon)$ before (β_v^-) and $v^*(\varepsilon)$ after (β_v^+) the trend switching point $\varepsilon = 1$ in dependence of $|\varepsilon - 1|_{\text{cut}}$. Additionally, the corresponding values of the KS statistic, D_v^- and D_v^+, are given. The power-law hypothesis is supported if the p-value is larger than 0.1.

$\lvert\varepsilon - 1\rvert_{\text{cut}}$	β_v^+	D_v^+	p-value	β_v^-	D_v^-	p-value
0.004	−0.209	0	**1**	−0.035	0	**1**
0.006	−0.21	0.0001	**0.934**	−0.026	0.0015	**0.323**
0.008	−0.221	0.003	**0.147**	−0.027	0.001	**0.609**
0.01	−0.238	0.0065	0	−0.023	0.0014	**0.376**
0.012	−0.243	0.0069	0	−0.022	0.0016	**0.235**
0.014	−0.248	0.0068	0	−0.025	0.0009	**0.673**
0.016	−0.251	0.0065	0	−0.024	0.0009	**0.617**
0.018	−0.252	0.006	0	−0.022	0.0012	**0.352**
0.02	−0.252	0.0055	0	−0.024	0.0007	**0.731**
0.022	−0.249	0.005	0.002	−0.025	0.001	**0.408**
0.024	−0.247	0.0046	0.015	−0.025	0.001	**0.395**
0.026	−0.246	0.0044	0.048	−0.025	0.0007	**0.752**
0.028	−0.247	0.0041	0.093	−0.025	0.0007	**0.603**
0.03	−0.246	0.0039	**0.161**	−0.024	0.0007	**0.66**
0.032	−0.244	0.0038	**0.18**	−0.023	0.0011	**0.206**
0.034	−0.241	0.0037	**0.206**	−0.023	0.0013	**0.103**
0.036	−0.24	0.0036	**0.236**	−0.022	0.0014	0.064
0.038	−0.237	0.0036	**0.182**	−0.021	0.0019	0.003
0.04	−0.235	0.0036	**0.161**	−0.02	0.0019	0.002
0.042	−0.234	0.0036	**0.144**	−0.02	0.0019	0
0.044	−0.233	0.0038	0.043	−0.02	0.0019	0
0.046	−0.232	0.0039	0.034	−0.019	0.002	0
0.048	−0.231	0.004	0.006	−0.019	0.002	0
0.05	−0.229	0.0043	0.001	−0.018	0.002	0
0.052	−0.227	0.0046	0	−0.019	0.0019	0
0.054	−0.226	0.0049	0	−0.019	0.0018	0
0.056	−0.224	0.0052	0	−0.019	0.0017	0
0.058	−0.223	0.0054	0	−0.019	0.0016	0
0.06	−0.221	0.0055	0	−0.019	0.0016	0
0.062	−0.22	0.0057	0	−0.019	0.0015	0
0.064	−0.218	0.0063	0	−0.019	0.0015	0
0.066	−0.216	0.0067	0	−0.019	0.0014	0
0.068	−0.215	0.0069	0	−0.018	0.0014	0.001
0.07	−0.214	0.007	0	−0.018	0.0014	0.002
0.072	−0.213	0.0071	0	−0.018	0.0014	0.002
0.074	−0.212	0.0072	0	−0.018	0.0014	0
0.076	−0.21	0.0076	0	−0.017	0.0014	0

A.19 3M Company Common Stock (MMM)

The MMM price time series contains 27,312,827 transactions.

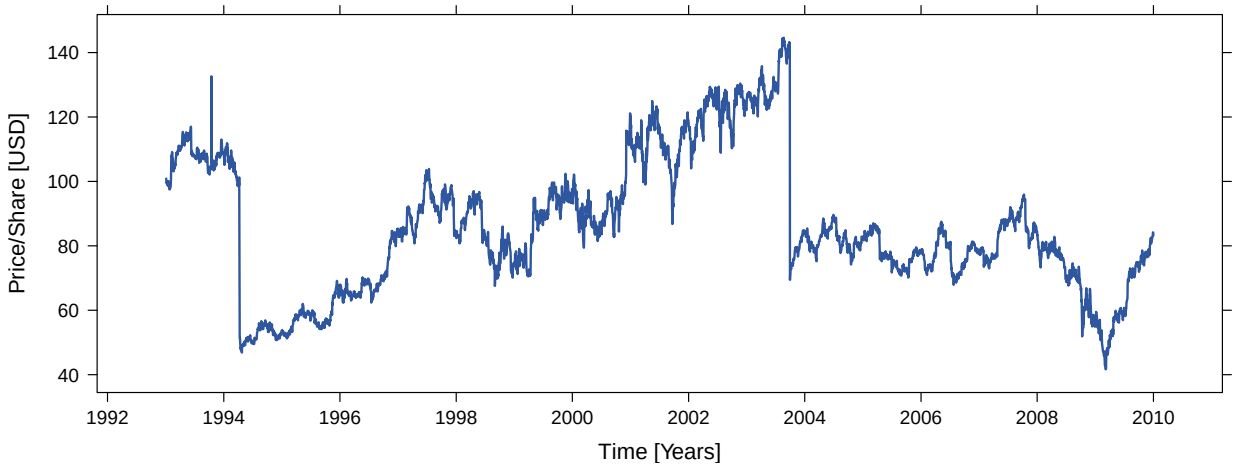

Figure A.37: Stock splits occurred on 11 April 1994 [split ratio 2:1] and 30 September 2003 [split ratio 2:1].

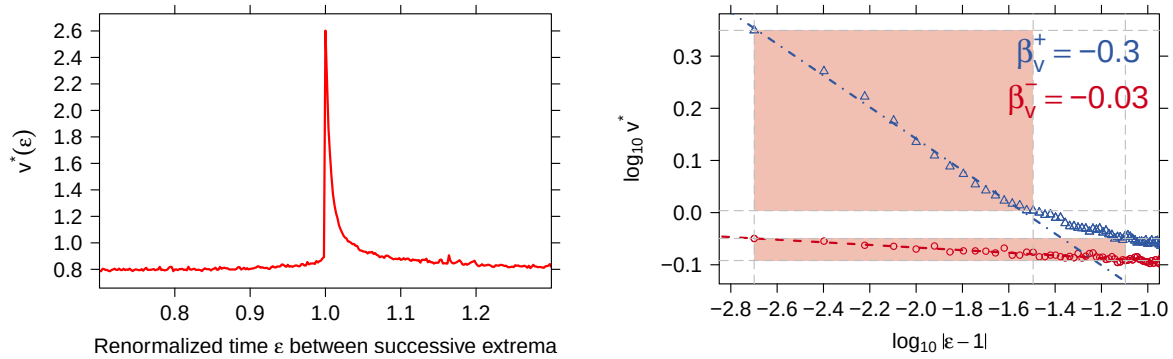

Figure A.38: *left*: aggregated volume $v^*(\varepsilon)$. *right*: $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram.

Table A.19: Statistical test of power-law hypothesis for the MMM volume time series: Scaling parameters of the hypothesized power-law model are shown for both $v^*(\varepsilon)$ before (β_v^-) and $v^*(\varepsilon)$ after (β_v^+) the trend switching point $\varepsilon = 1$ in dependence of $|\varepsilon - 1|_{\text{cut}}$. Additionally, the corresponding values of the KS statistic, D_v^- and D_v^+, are given. The power-law hypothesis is supported if the p-value is larger than 0.1.

$\lvert\varepsilon - 1\rvert_{\text{cut}}$	β_v^+	D_v^+	p-value	β_v^-	D_v^-	p-value
0.004	−0.259	0	**1**	−0.017	0	**1**
0.006	−0.265	0.0009	**0.701**	−0.026	0.0015	**0.444**
0.008	−0.282	0.0048	0.043	−0.026	0.0011	**0.696**
0.01	−0.3	0.0077	0	−0.028	0.0012	**0.639**
0.012	−0.309	0.0088	0	−0.023	0.0015	**0.49**
0.014	−0.314	0.0087	0	−0.027	0.0015	**0.404**
0.016	−0.315	0.0079	0	−0.027	0.0014	**0.444**
0.018	−0.317	0.0074	0	−0.026	0.0011	**0.622**
0.02	−0.317	0.0068	0.003	−0.026	0.0008	**0.771**
0.022	−0.316	0.0063	0.001	−0.026	0.0008	**0.783**
0.024	−0.315	0.0059	0.019	−0.023	0.0015	**0.236**
0.026	−0.313	0.0056	0.051	−0.025	0.0007	**0.831**
0.028	−0.309	0.0053	0.083	−0.025	0.0011	**0.449**
0.03	−0.306	0.0051	**0.101**	−0.024	0.0006	**0.89**
0.032	−0.303	0.005	**0.107**	−0.024	0.0008	**0.691**
0.034	−0.299	0.005	0.098	−0.025	0.0006	**0.87**
0.036	−0.295	0.005	0.088	−0.025	0.0009	**0.583**
0.038	−0.291	0.0058	0.002	−0.025	0.0008	**0.637**
0.04	−0.287	0.0068	0	−0.025	0.0007	**0.769**
0.042	−0.283	0.0079	0	−0.025	0.0007	**0.71**
0.044	−0.281	0.0084	0	−0.026	0.0009	**0.521**
0.046	−0.278	0.0087	0	−0.026	0.0008	**0.54**
0.048	−0.276	0.0091	0	−0.026	0.0008	**0.523**
0.05	−0.273	0.0094	0	−0.025	0.0005	**0.85**
0.052	−0.271	0.0097	0	−0.025	0.0006	**0.835**
0.054	−0.268	0.0101	0	−0.025	0.0005	**0.867**
0.056	−0.266	0.0103	0	−0.024	0.001	**0.238**
0.058	−0.263	0.0107	0	−0.023	0.0014	0.03
0.06	−0.26	0.011	0	−0.023	0.0014	0.026
0.062	−0.257	0.0113	0	−0.023	0.0014	0.034
0.064	−0.255	0.0115	0	−0.023	0.0013	0.034
0.066	−0.252	0.0118	0	−0.023	0.0012	0.086
0.068	−0.25	0.0122	0	−0.023	0.0011	**0.107**
0.07	−0.247	0.0125	0	−0.023	0.0012	0.067
0.072	−0.246	0.0127	0	−0.023	0.0011	**0.102**
0.074	−0.244	0.013	0	−0.024	0.0009	**0.229**
0.076	−0.242	0.0132	0	−0.024	0.0007	**0.398**

A.20 Merck & Company, Inc. Common Stock (MRK)

The MRK price time series contains 63,308,477 transactions.

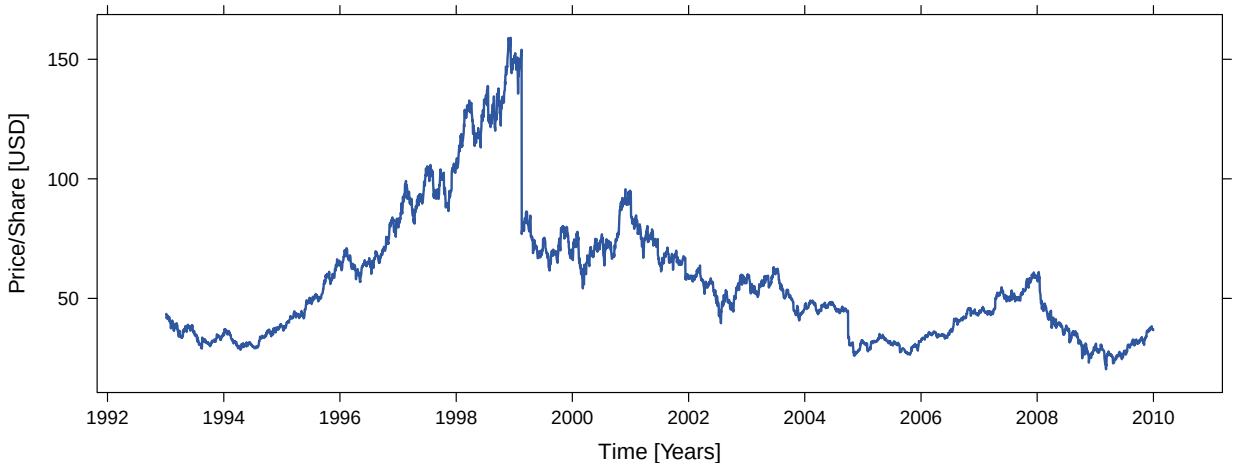

Figure A.39: A stock split occurred on 17 February 1999 [split ratio 2:1].

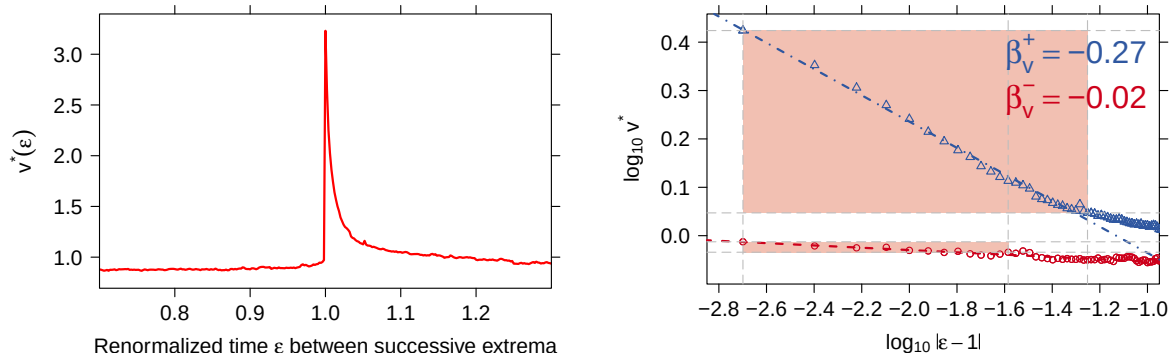

Figure A.40: *left*: aggregated volume $v^*(\varepsilon)$. *right*: $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram.

Table A.20: Statistical test of power-law hypothesis for the MRK volume time series: Scaling parameters of the hypothesized power-law model are shown for both $v^*(\varepsilon)$ before (β_v^-) and $v^*(\varepsilon)$ after (β_v^+) the trend switching point $\varepsilon = 1$ in dependence of $|\varepsilon - 1|_{\text{cut}}$. Additionally, the corresponding values of the KS statistic, D_v^- and D_v^+, are given. The power-law hypothesis is supported if the p-value is larger than 0.1.

$\lvert\varepsilon - 1\rvert_{\text{cut}}$	β_v^+	D_v^+	p-value	β_v^-	D_v^-	p-value
0.004	−0.237	0	**1**	−0.028	0	**1**
0.006	−0.246	0.0014	**0.414**	−0.026	0.0003	**0.829**
0.008	−0.254	0.0027	**0.204**	−0.02	0.0017	**0.173**
0.01	−0.26	0.0033	**0.153**	−0.022	0.001	**0.499**
0.012	−0.267	0.0044	0.017	−0.023	0.0007	**0.678**
0.014	−0.272	0.005	0.004	−0.024	0.0006	**0.822**
0.016	−0.276	0.0054	0	−0.023	0.0006	**0.746**
0.018	−0.278	0.0052	0.004	−0.024	0.0005	**0.894**
0.02	−0.282	0.0054	0	−0.025	0.0008	**0.464**
0.022	−0.284	0.0056	0.001	−0.026	0.0011	**0.216**
0.024	−0.286	0.0056	0	−0.025	0.0008	**0.425**
0.026	−0.287	0.0055	0	−0.023	0.0011	**0.13**
0.028	−0.286	0.005	0.002	−0.022	0.0014	0.036
0.03	−0.285	0.0046	0.041	−0.02	0.0022	0.001
0.032	−0.284	0.0043	**0.139**	−0.019	0.0025	0
0.034	−0.284	0.0041	**0.249**	−0.02	0.002	0
0.036	−0.285	0.0039	**0.39**	−0.021	0.0017	0.002
0.038	−0.284	0.0038	**0.519**	−0.022	0.0013	0.013
0.04	−0.284	0.0036	**0.695**	−0.023	0.0016	0.001
0.042	−0.283	0.0035	**0.77**	−0.024	0.0017	0
0.044	−0.282	0.0034	**0.852**	−0.024	0.0018	0
0.046	−0.281	0.0033	**0.862**	−0.024	0.0018	0
0.048	−0.279	0.0033	**0.879**	−0.025	0.0019	0
0.05	−0.278	0.0033	**0.891**	−0.025	0.0018	0
0.052	−0.274	0.0034	**0.72**	−0.025	0.0018	0
0.054	−0.273	0.0035	**0.603**	−0.025	0.0018	0
0.056	−0.271	0.0036	**0.34**	−0.025	0.0017	0
0.058	−0.269	0.0043	0.002	−0.025	0.0017	0
0.06	−0.267	0.005	0	−0.025	0.0016	0
0.062	−0.265	0.0054	0	−0.025	0.0015	0
0.064	−0.263	0.0061	0	−0.025	0.0015	0
0.066	−0.262	0.0066	0	−0.025	0.0014	0.001
0.068	−0.26	0.0071	0	−0.025	0.0014	0
0.07	−0.258	0.0075	0	−0.025	0.0013	0
0.072	−0.256	0.0079	0	−0.025	0.0013	0.001
0.074	−0.255	0.0082	0	−0.024	0.0012	0.001
0.076	−0.253	0.0086	0	−0.024	0.0012	0

A.21 Microsoft Corporation (MSFT)

The MSFT price time series contains 250,448,998 transactions.

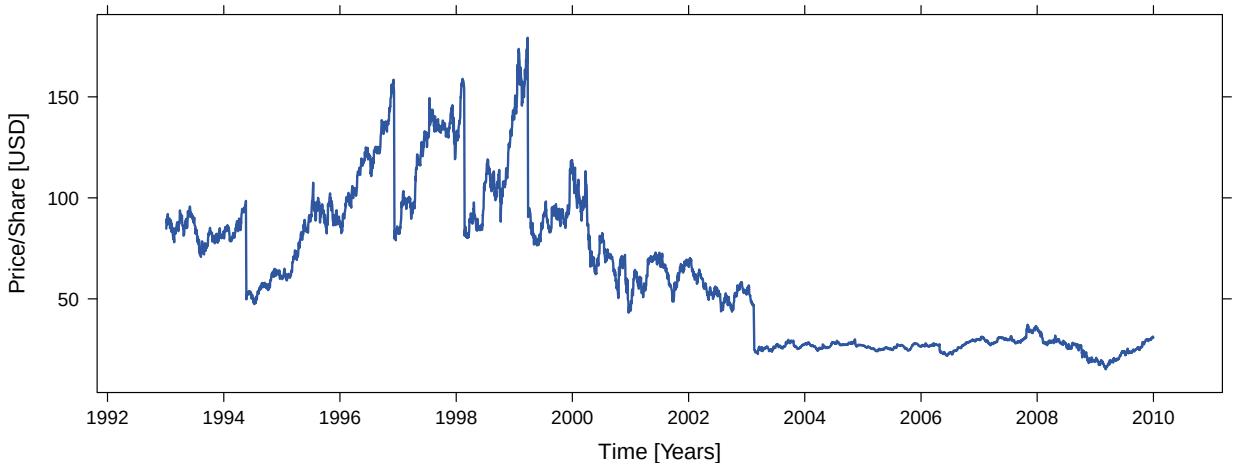

Figure A.41: Stock splits occurred on 23 May 1994 [split ratio 2:1], 9 December 1996 [split ratio 2:1], 23 February 1998 [split ratio 2:1], 29 March 1999 [split ratio 2:1], and 18 February 2003 [split ratio 2:1].

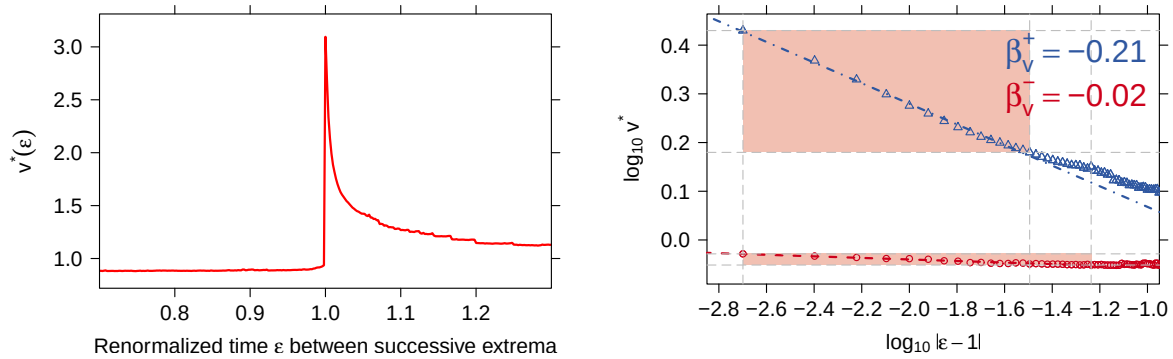

Figure A.42: *left*: aggregated volume $v^*(\varepsilon)$. *right*: $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram.

Table A.21: Statistical test of power-law hypothesis for the MSFT volume time series: Scaling parameters of the hypothesized power-law model are shown for both $v^*(\varepsilon)$ before (β_v^-) and $v^*(\varepsilon)$ after (β_v^+) the trend switching point $\varepsilon = 1$ in dependence of $|\varepsilon - 1|_{\text{cut}}$. Additionally, the corresponding values of the KS statistic, D_v^- and D_v^+, are given. The power-law hypothesis is supported if the p-value is larger than 0.1.

$\lvert\varepsilon - 1\rvert_{\text{cut}}$	β_v^+	D_v^+	p-value	β_v^-	D_v^-	p-value
0.004	−0.205	0	**0.771**	−0.015	0	**1**
0.006	−0.21	0.0008	**0.567**	−0.015	0	**0.992**
0.008	−0.216	0.0019	**0.214**	−0.015	0	**0.992**
0.01	−0.22	0.0023	**0.161**	−0.014	0.0003	**0.727**
0.012	−0.221	0.0021	**0.569**	−0.014	0.0004	**0.643**
0.014	−0.222	0.0019	**0.874**	−0.015	0.0003	**0.672**
0.016	−0.222	0.0017	**0.983**	−0.015	0.0005	**0.364**
0.018	−0.222	0.0015	**1**	−0.017	0.0008	0.068
0.02	−0.221	0.0014	**1**	−0.017	0.0008	0.069
0.022	−0.22	0.0013	**1**	−0.018	0.0009	0.03
0.024	−0.218	0.0012	**1**	−0.018	0.0009	0.02
0.026	−0.216	0.0013	**1**	−0.018	0.0008	0.014
0.028	−0.215	0.0018	**1**	−0.018	0.0007	0.058
0.03	−0.213	0.0024	**0.927**	−0.017	0.0006	**0.118**
0.032	−0.211	0.0028	**0.399**	−0.017	0.0006	**0.145**
0.034	−0.21	0.0031	0.036	−0.017	0.0005	**0.198**
0.036	−0.208	0.0033	0	−0.017	0.0005	**0.223**
0.038	−0.207	0.0036	0	−0.017	0.0004	**0.264**
0.04	−0.205	0.0038	0	−0.017	0.0004	**0.254**
0.042	−0.204	0.004	0	−0.016	0.0004	**0.294**
0.044	−0.203	0.0042	0	−0.016	0.0004	**0.293**
0.046	−0.201	0.0044	0	−0.016	0.0004	**0.309**
0.048	−0.2	0.0045	0	−0.016	0.0004	**0.338**
0.05	−0.198	0.0047	0	−0.016	0.0004	**0.285**
0.052	−0.196	0.0051	0	−0.016	0.0004	**0.284**
0.054	−0.195	0.0053	0	−0.016	0.0004	**0.259**
0.056	−0.194	0.0056	0	−0.016	0.0004	**0.112**
0.058	−0.192	0.006	0	−0.015	0.0004	**0.1**
0.06	−0.191	0.0062	0	−0.015	0.0005	0.052
0.062	−0.189	0.0064	0	−0.015	0.0005	0.034
0.064	−0.189	0.0065	0	−0.015	0.0005	0.017
0.066	−0.187	0.0067	0	−0.015	0.0006	0.005
0.068	−0.186	0.0068	0	−0.015	0.0006	0.002
0.07	−0.185	0.0069	0	−0.015	0.0006	0.005
0.072	−0.185	0.0068	0	−0.015	0.0006	0
0.074	−0.185	0.0067	0	−0.014	0.0007	0
0.076	−0.184	0.0066	0	−0.014	0.0007	0.001

A.22 Pfizer, Inc. Common Stock (PFE)

The PFE price time series contains 103,602,977 transactions.

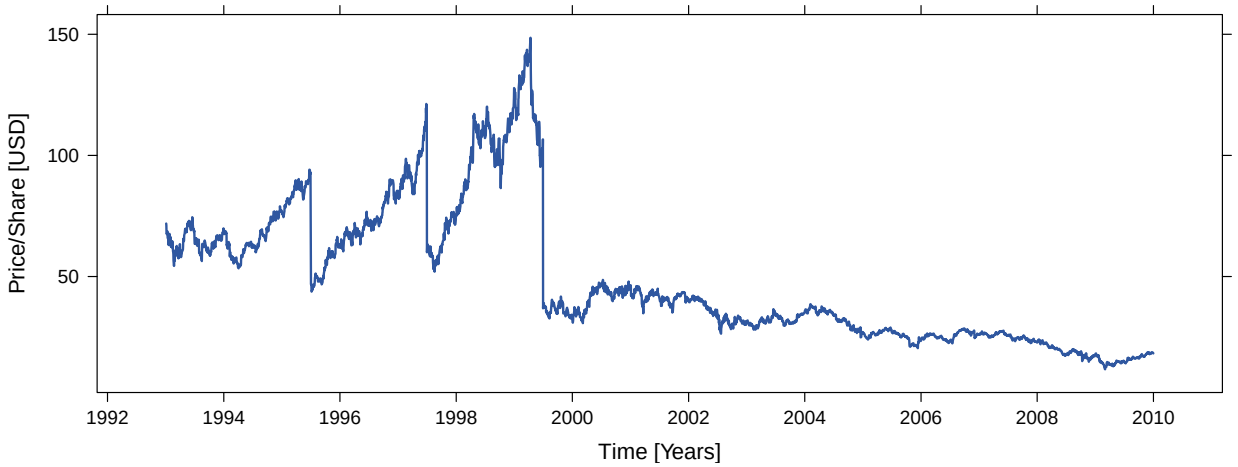

Figure A.43: Stock splits occurred on 3 July 1995 [split ratio 2:1], 1 July 1997 [split ratio 2:1], and 1 July 1999 [split ratio 3:1].

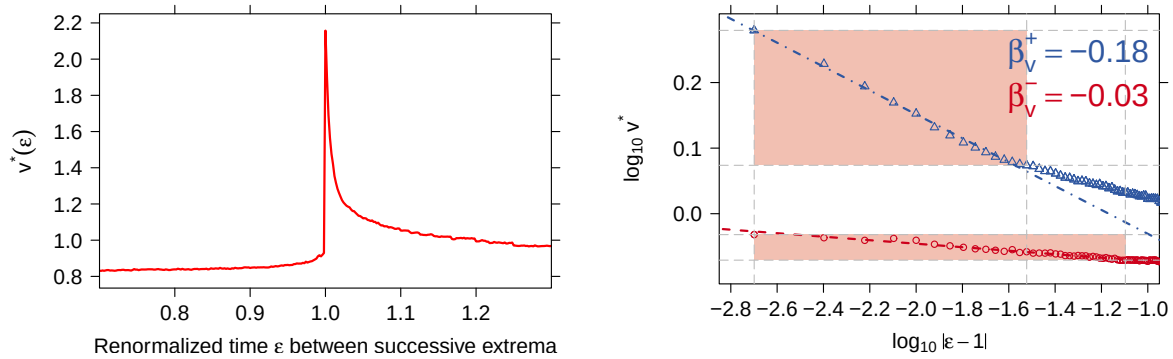

Figure A.44: *left*: aggregated volume $v^*(\varepsilon)$. *right*: $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram.

Table A.22: Statistical test of power-law hypothesis for the PFE volume time series: Scaling parameters of the hypothesized power-law model are shown for both $v^*(\varepsilon)$ before (β_v^-) and $v^*(\varepsilon)$ after (β_v^+) the trend switching point $\varepsilon = 1$ in dependence of $|\varepsilon - 1|_{\text{cut}}$. Additionally, the corresponding values of the KS statistic, D_v^- and D_v^+, are given. The power-law hypothesis is supported if the p-value is larger than 0.1.

$\lvert\varepsilon - 1\rvert_{\text{cut}}$	β_v^+	D_v^+	p-value	β_v^-	D_v^-	p-value
0.004	−0.17	0	**0.763**	−0.015	0	**1**
0.006	−0.178	0.0013	**0.301**	−0.018	0.0003	**0.724**
0.008	−0.182	0.0016	**0.379**	−0.011	0.0016	0.083
0.01	−0.183	0.0014	**0.61**	−0.011	0.0013	**0.133**
0.012	−0.188	0.0026	**0.125**	−0.015	0.0011	**0.164**
0.014	−0.19	0.0029	0.076	−0.017	0.0018	0.01
0.016	−0.191	0.0029	**0.108**	−0.019	0.002	0.002
0.018	−0.191	0.0026	**0.293**	−0.021	0.0021	0.002
0.02	−0.19	0.0023	**0.595**	−0.022	0.0021	0
0.022	−0.189	0.002	**0.821**	−0.023	0.0021	0
0.024	−0.188	0.0018	**0.953**	−0.024	0.002	0.001
0.026	−0.186	0.0016	**0.98**	−0.024	0.0018	0
0.028	−0.184	0.0016	**0.989**	−0.024	0.0018	0
0.03	−0.182	0.0024	**0.415**	−0.025	0.0017	0
0.032	−0.179	0.0033	0.003	−0.025	0.0016	0.001
0.034	−0.177	0.0039	0	−0.025	0.0015	0
0.036	−0.175	0.0044	0	−0.025	0.0014	0.001
0.038	−0.172	0.0049	0	−0.025	0.0013	0
0.04	−0.17	0.0054	0	−0.025	0.0013	0.003
0.042	−0.168	0.0057	0	−0.025	0.0012	0.004
0.044	−0.167	0.0058	0	−0.025	0.0012	0.004
0.046	−0.165	0.006	0	−0.025	0.0011	0.007
0.048	−0.164	0.0062	0	−0.026	0.001	0.007
0.05	−0.162	0.0062	0	−0.025	0.001	0.014
0.052	−0.161	0.0063	0	−0.025	0.001	0.01
0.054	−0.159	0.0065	0	−0.025	0.0009	0.016
0.056	−0.158	0.0066	0	−0.025	0.0009	0.023
0.058	−0.157	0.0067	0	−0.025	0.0009	0.027
0.06	−0.155	0.0068	0	−0.025	0.0008	0.03
0.062	−0.154	0.0069	0	−0.026	0.0008	0.025
0.064	−0.153	0.0069	0	−0.026	0.0008	0.073
0.066	−0.151	0.007	0	−0.026	0.0007	0.053
0.068	−0.15	0.0071	0	−0.026	0.0007	0.052
0.07	−0.149	0.0073	0	−0.026	0.0007	0.068
0.072	−0.148	0.0073	0	−0.026	0.0007	0.079
0.074	−0.147	0.0073	0	−0.026	0.0006	**0.1**
0.076	−0.146	0.0074	0	−0.026	0.0006	**0.124**

A.23 Procter & Gamble Company (The) (PG)

The PG price time series contains 56,194,379 transactions.

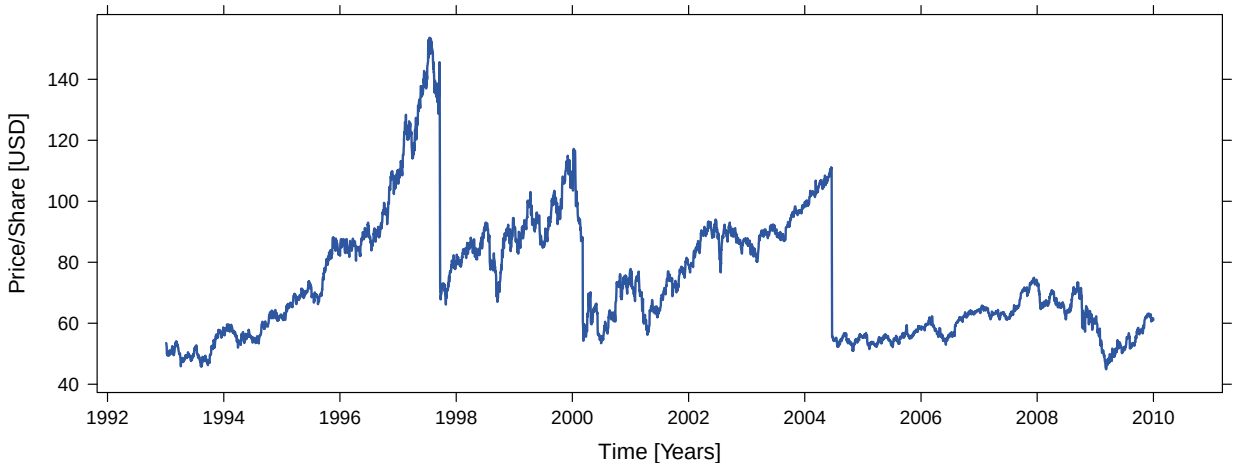

Figure A.45: Stock splits occurred on 22 September 1997 [split ratio 2:1] and 21 June 2004 [split ratio 2:1].

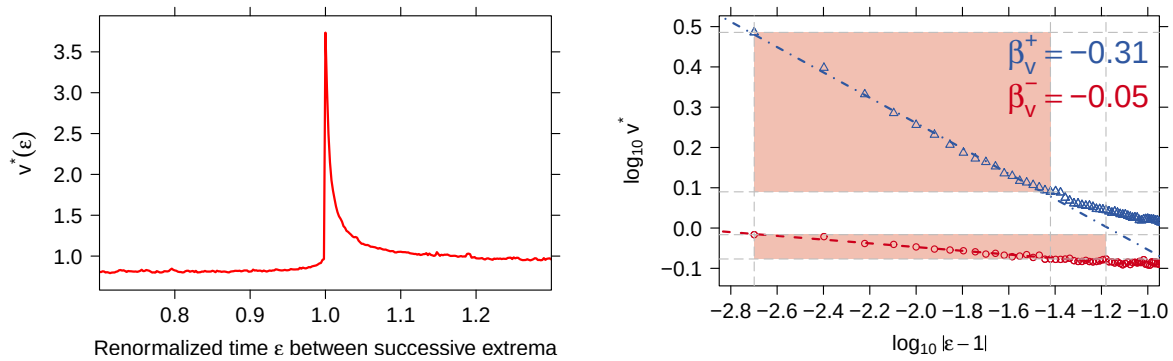

Figure A.46: *left*: aggregated volume $v^*(\varepsilon)$. *right*: $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram.

Table A.23: Statistical test of power-law hypothesis for the PG volume time series: Scaling parameters of the hypothesized power-law model are shown for both $v^*(\varepsilon)$ before (β_v^-) and $v^*(\varepsilon)$ after (β_v^+) the trend switching point $\varepsilon = 1$ in dependence of $|\varepsilon - 1|_{\text{cut}}$. Additionally, the corresponding values of the KS statistic, D_v^- and D_v^+, are given. The power-law hypothesis is supported if the p-value is larger than 0.1.

$\lvert\varepsilon-1\rvert_{\text{cut}}$	β_v^+	D_v^+	p-value	β_v^-	D_v^-	p-value
0.004	−0.292	0	**1**	−0.016	0	**1**
0.006	−0.318	0.0043	0.011	−0.042	0.0044	0.001
0.008	−0.332	0.005	0.005	−0.043	0.0034	0.005
0.01	−0.333	0.0045	0.05	−0.045	0.0028	0.02
0.012	−0.333	0.0038	**0.266**	−0.047	0.0024	0.025
0.014	−0.334	0.0035	**0.594**	−0.048	0.0021	0.044
0.016	−0.334	0.0032	**0.827**	−0.048	0.0018	0.071
0.018	−0.334	0.0029	**0.967**	−0.048	0.0016	**0.126**
0.02	−0.331	0.0027	**0.989**	−0.049	0.0015	**0.147**
0.022	−0.328	0.0025	**0.998**	−0.048	0.0013	**0.235**
0.024	−0.327	0.0024	**0.998**	−0.049	0.0012	**0.249**
0.026	−0.325	0.0023	**1**	−0.05	0.0014	**0.157**
0.028	−0.324	0.0022	**1**	−0.05	0.0013	**0.183**
0.03	−0.322	0.0022	**1**	−0.049	0.001	**0.408**
0.032	−0.32	0.0023	**1**	−0.049	0.0009	**0.466**
0.034	−0.318	0.0029	**1**	−0.048	0.0009	**0.453**
0.036	−0.316	0.0035	**0.956**	−0.048	0.0008	**0.537**
0.038	−0.314	0.0041	**0.615**	−0.048	0.0009	**0.41**
0.04	−0.311	0.005	0.024	−0.049	0.001	**0.368**
0.042	−0.307	0.0059	0	−0.049	0.0009	**0.382**
0.044	−0.306	0.0062	0	−0.048	0.0008	**0.543**
0.046	−0.304	0.0064	0	−0.048	0.0007	**0.726**
0.048	−0.303	0.0064	0	−0.048	0.0006	**0.73**
0.05	−0.302	0.0066	0	−0.048	0.0006	**0.752**
0.052	−0.3	0.0068	0	−0.048	0.0006	**0.761**
0.054	−0.299	0.007	0	−0.048	0.0006	**0.797**
0.056	−0.297	0.0073	0	−0.048	0.0005	**0.826**
0.058	−0.295	0.0076	0	−0.048	0.0005	**0.851**
0.06	−0.293	0.0078	0	−0.048	0.0005	**0.836**
0.062	−0.292	0.008	0	−0.047	0.0005	**0.811**
0.064	−0.29	0.0081	0	−0.047	0.0005	**0.783**
0.066	−0.288	0.0083	0	−0.046	0.001	**0.152**
0.068	−0.286	0.0084	0	−0.046	0.0011	0.081
0.07	−0.284	0.0088	0	−0.045	0.0012	0.046
0.072	−0.282	0.0093	0	−0.045	0.0011	0.044
0.074	−0.281	0.0097	0	−0.045	0.0012	0.014
0.076	−0.279	0.01	0	−0.045	0.0012	0.025

A.24 AT&T Inc. (T)

The T price time series contains 85,424,623 transactions.

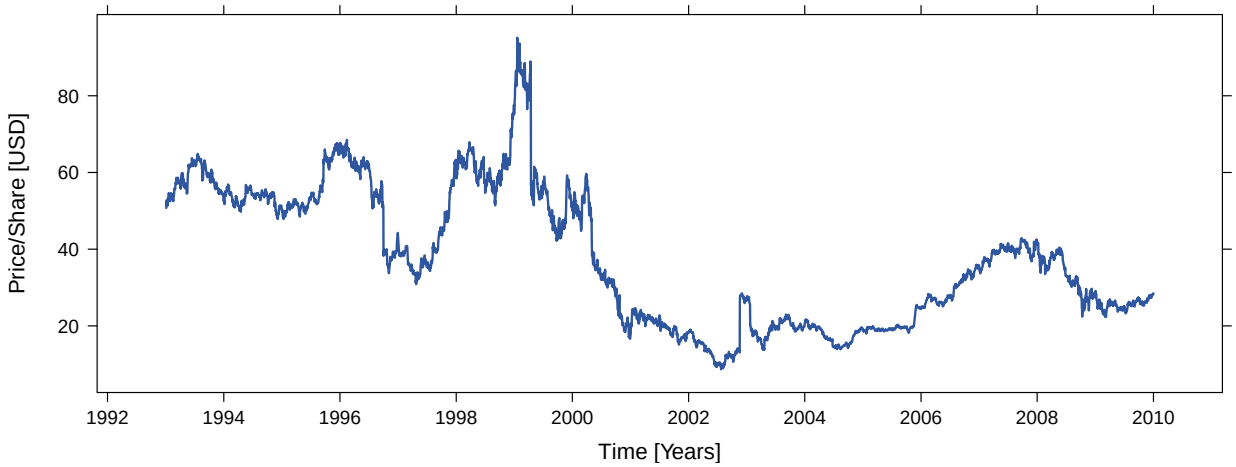

Figure A.47: Stock splits occurred on 26 May 1993 [split ratio 2:1] and 20 March 1998 [split ratio 2:1].

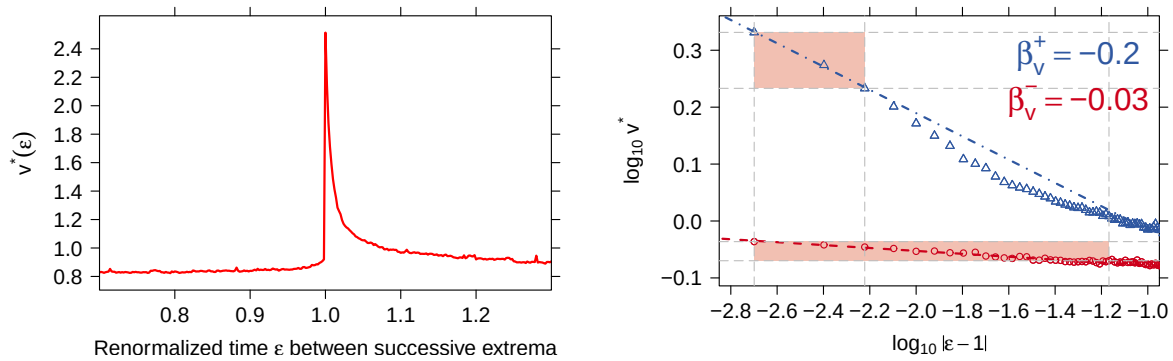

Figure A.48: *left*: aggregated volume $v^*(\varepsilon)$. *right*: $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram.

Table A.24: Statistical test of power-law hypothesis for the T volume time series: Scaling parameters of the hypothesized power-law model are shown for both $v^*(\varepsilon)$ before (β_v^-) and $v^*(\varepsilon)$ after (β_v^+) the trend switching point $\varepsilon = 1$ in dependence of $|\varepsilon - 1|_{\text{cut}}$. Additionally, the corresponding values of the KS statistic, D_v^- and D_v^+, are given. The power-law hypothesis is supported if the p-value is larger than 0.1.

$\lvert\varepsilon - 1\rvert_{\text{cut}}$	β_v^+	D_v^+	p-value	β_v^-	D_v^-	p-value
0.004	−0.19	0	**0.749**	−0.018	0	**1**
0.006	−0.204	0.0022	**0.123**	−0.019	0.0001	**0.887**
0.008	−0.215	0.0035	0.011	−0.02	0.0001	**0.984**
0.01	−0.226	0.0053	0	−0.023	0.001	**0.341**
0.012	−0.234	0.0064	0	−0.022	0.0007	**0.559**
0.014	−0.239	0.0065	0	−0.023	0.0007	**0.52**
0.016	−0.246	0.0071	0	−0.023	0.0006	**0.622**
0.018	−0.249	0.0071	0	−0.021	0.0005	**0.796**
0.02	−0.249	0.0066	0	−0.023	0.0008	**0.374**
0.022	−0.251	0.0065	0	−0.024	0.0013	0.06
0.024	−0.252	0.0063	0	−0.025	0.0015	0.016
0.026	−0.252	0.0059	0	−0.026	0.0017	0.006
0.028	−0.251	0.0055	0	−0.025	0.0014	0.008
0.03	−0.25	0.005	0	−0.025	0.0012	0.019
0.032	−0.248	0.0047	0	−0.026	0.0013	0.016
0.034	−0.246	0.0044	0.001	−0.026	0.0013	0.016
0.036	−0.244	0.0043	0.001	−0.026	0.0012	0.02
0.038	−0.241	0.0042	0.001	−0.026	0.0011	0.035
0.04	−0.24	0.0041	0.003	−0.026	0.0011	0.034
0.042	−0.238	0.004	0.002	−0.026	0.001	0.044
0.044	−0.236	0.004	0.003	−0.026	0.001	0.053
0.046	−0.234	0.004	0.001	−0.026	0.0009	0.069
0.048	−0.232	0.004	0	−0.027	0.0009	0.056
0.05	−0.231	0.0041	0	−0.027	0.0009	0.075
0.052	−0.228	0.0046	0	−0.027	0.0008	0.081
0.054	−0.226	0.005	0	−0.026	0.0008	**0.102**
0.056	−0.225	0.0053	0	−0.026	0.0008	0.096
0.058	−0.223	0.0057	0	−0.026	0.0008	**0.108**
0.06	−0.221	0.0061	0	−0.026	0.0007	**0.138**
0.062	−0.219	0.0066	0	−0.026	0.0007	**0.172**
0.064	−0.218	0.0068	0	−0.026	0.0007	**0.162**
0.066	−0.216	0.0072	0	−0.025	0.0007	**0.152**
0.068	−0.214	0.0075	0	−0.025	0.0007	**0.132**
0.07	−0.213	0.0077	0	−0.025	0.0008	0.049
0.072	−0.211	0.0079	0	−0.024	0.0009	0.029
0.074	−0.21	0.0081	0	−0.024	0.0009	0.009
0.076	−0.209	0.0083	0	−0.024	0.0009	0.004

A.25 The Travelers Companies, Inc. Common Stock (TRV)

The TRV price time series contains 16,994,915 transactions.

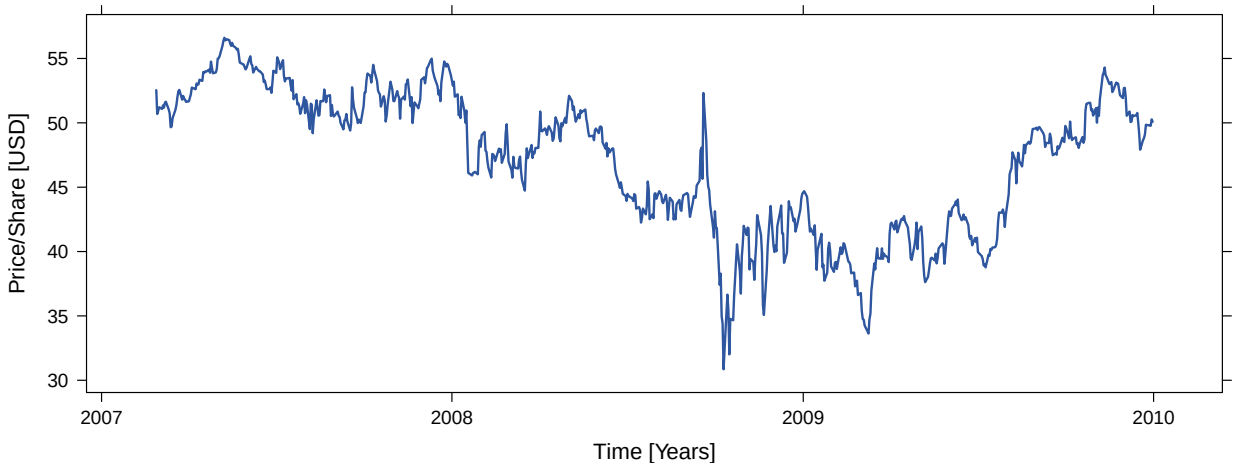

Figure A.49: Stock splits occurred on 7 June 1994 [split ratio 2:1] and 12 May 1998 [split ratio 2:1].

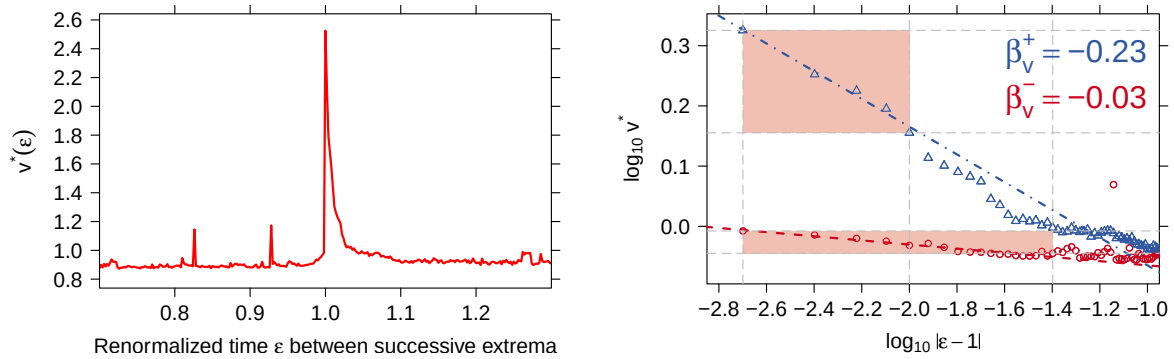

Figure A.50: *left*: aggregated volume $v^*(\varepsilon)$. *right*: $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram.

Table A.25: Statistical test of power-law hypothesis for the TRV volume time series: Scaling parameters of the hypothesized power-law model are shown for both $v^*(\varepsilon)$ before (β_v^-) and $v^*(\varepsilon)$ after (β_v^+) the trend switching point $\varepsilon = 1$ in dependence of $|\varepsilon - 1|_{\text{cut}}$. Additionally, the corresponding values of the KS statistic, D_v^- and D_v^+, are given. The power-law hypothesis is supported if the p-value is larger than 0.1.

$\lvert\varepsilon - 1\rvert_{\text{cut}}$	β_v^+	D_v^+	p-value	β_v^-	D_v^-	p-value
0.004	−0.242	0	**1**	−0.022	0	**1**
0.006	−0.213	0.0048	0.062	−0.025	0.0005	**0.825**
0.008	−0.212	0.0038	**0.174**	−0.027	0.0008	**0.868**
0.01	−0.23	0.0045	**0.12**	−0.032	0.0019	**0.479**
0.012	−0.254	0.0099	0	−0.03	0.0011	**0.848**
0.014	−0.264	0.0116	0	−0.031	0.0012	**0.809**
0.016	−0.267	0.0111	0	−0.035	0.0021	**0.329**
0.018	−0.267	0.01	0	−0.036	0.0025	**0.147**
0.02	−0.265	0.009	0	−0.037	0.0024	**0.133**
0.022	−0.269	0.0087	0	−0.037	0.0023	**0.141**
0.024	−0.272	0.0084	0	−0.038	0.0022	**0.116**
0.026	−0.277	0.008	0	−0.038	0.002	**0.182**
0.028	−0.28	0.0085	0	−0.038	0.0019	**0.193**
0.03	−0.28	0.008	0	−0.038	0.0018	**0.235**
0.032	−0.28	0.0075	0	−0.038	0.0016	**0.269**
0.034	−0.277	0.0066	0	−0.037	0.0015	**0.297**
0.036	−0.275	0.0064	0.001	−0.035	0.0014	**0.383**
0.038	−0.272	0.0063	0	−0.035	0.0013	**0.404**
0.04	−0.27	0.0061	0.001	−0.034	0.0015	**0.273**
0.042	−0.268	0.006	0.002	−0.032	0.0022	0.025
0.044	−0.265	0.006	0	−0.031	0.0026	0.005
0.046	−0.263	0.0059	0.001	−0.029	0.0033	0
0.048	−0.259	0.006	0	−0.027	0.0039	0
0.05	−0.256	0.0061	0	−0.026	0.0042	0
0.052	−0.253	0.0065	0	−0.026	0.0039	0
0.054	−0.249	0.0072	0	−0.027	0.0037	0
0.056	−0.246	0.0082	0	−0.026	0.0036	0
0.058	−0.243	0.0089	0	−0.026	0.0035	0
0.06	−0.24	0.0097	0	−0.026	0.0034	0
0.062	−0.238	0.0104	0	−0.026	0.0035	0
0.064	−0.234	0.0112	0	−0.025	0.0034	0
0.066	−0.231	0.012	0	−0.025	0.0035	0
0.068	−0.228	0.0127	0	−0.023	0.0036	0
0.07	−0.225	0.0133	0	−0.023	0.0037	0
0.072	−0.222	0.0137	0	−0.013	0.0092	0
0.074	−0.22	0.0139	0	−0.014	0.0084	0
0.076	−0.217	0.0142	0	−0.015	0.0077	0

A.26 United Technologies Corporation (UTX)

The UTX price time series contains 30,464,044 transactions.

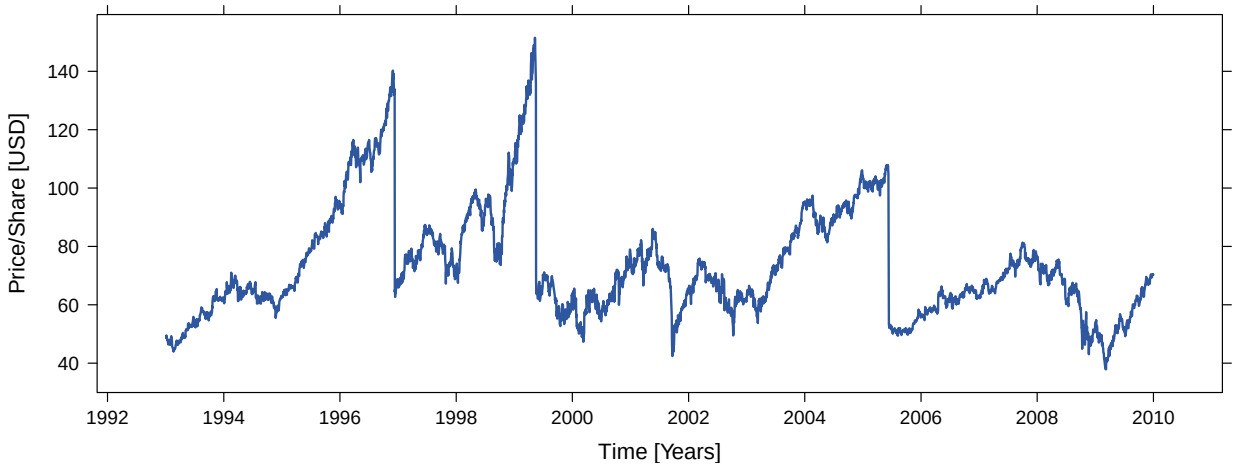

Figure A.51: Stock splits occurred on 11 December 1996 [split ratio 2:1], 18 May 1999 [split ratio 2:1], and 13 June 2005 [split ratio 2:1].

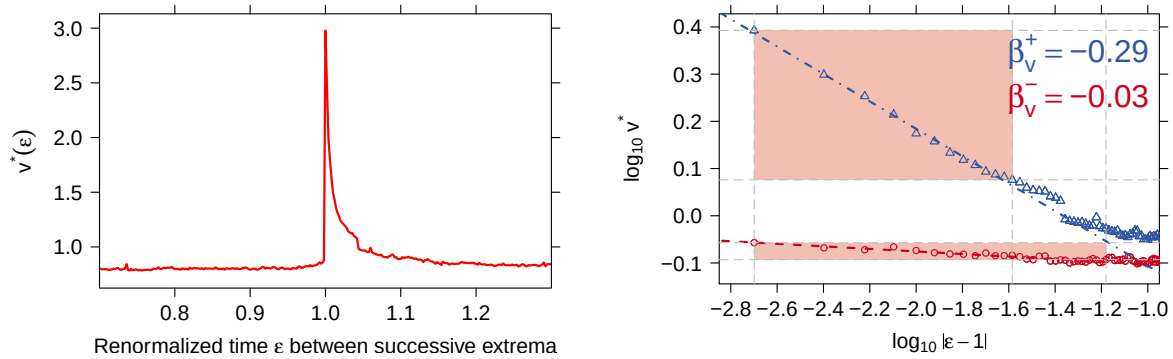

Figure A.52: *left*: aggregated volume $v^*(\varepsilon)$. *right*: $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram.

Table A.26: Statistical test of power-law hypothesis for the UTX volume time series: Scaling parameters of the hypothesized power-law model are shown for both $v^*(\varepsilon)$ before (β_v^-) and $v^*(\varepsilon)$ after (β_v^+) the trend switching point $\varepsilon = 1$ in dependence of $|\varepsilon - 1|_{\text{cut}}$. Additionally, the corresponding values of the KS statistic, D_v^- and D_v^+, are given. The power-law hypothesis is supported if the p-value is larger than 0.1.

$\lvert\varepsilon - 1\rvert_{\text{cut}}$	β_v^+	D_v^+	p-value	β_v^-	D_v^-	p-value
0.004	−0.312	0	**1**	−0.036	0	**1**
0.006	−0.294	0.0029	**0.216**	−0.031	0.0007	**0.701**
0.008	−0.294	0.0023	**0.478**	−0.018	0.0039	0.021
0.01	−0.304	0.0026	**0.582**	−0.019	0.0029	0.051
0.012	−0.303	0.002	**0.86**	−0.022	0.0021	**0.176**
0.014	−0.305	0.0021	**0.908**	−0.024	0.0017	**0.281**
0.016	−0.305	0.0018	**0.966**	−0.025	0.0015	**0.384**
0.018	−0.303	0.0015	**0.997**	−0.026	0.0013	**0.417**
0.02	−0.301	0.0013	**0.996**	−0.024	0.0011	**0.48**
0.022	−0.298	0.0021	**0.988**	−0.025	0.001	**0.541**
0.024	−0.294	0.0034	**0.801**	−0.025	0.001	**0.611**
0.026	−0.29	0.0047	**0.144**	−0.024	0.0009	**0.669**
0.028	−0.286	0.0058	0.001	−0.024	0.0008	**0.704**
0.03	−0.283	0.0063	0	−0.025	0.0008	**0.689**
0.032	−0.281	0.0067	0	−0.026	0.001	**0.418**
0.034	−0.278	0.0072	0	−0.025	0.0007	**0.729**
0.036	−0.274	0.0077	0	−0.025	0.0009	**0.459**
0.038	−0.272	0.0079	0	−0.026	0.0006	**0.784**
0.04	−0.269	0.0081	0	−0.026	0.001	**0.385**
0.042	−0.267	0.0082	0	−0.027	0.001	**0.3**
0.044	−0.271	0.0074	0	−0.027	0.0011	**0.243**
0.046	−0.273	0.0068	0	−0.028	0.0014	0.064
0.048	−0.275	0.0064	0	−0.028	0.0015	0.043
0.05	−0.276	0.0061	0	−0.028	0.0015	0.036
0.052	−0.276	0.0059	0	−0.028	0.0014	0.038
0.054	−0.277	0.0057	0	−0.028	0.0014	0.045
0.056	−0.277	0.0056	0	−0.028	0.0013	0.062
0.058	−0.276	0.0054	0	−0.027	0.0011	**0.133**
0.06	−0.274	0.0053	0	−0.027	0.0009	**0.23**
0.062	−0.273	0.0052	0	−0.027	0.001	**0.231**
0.064	−0.272	0.005	0.001	−0.027	0.0009	**0.235**
0.066	−0.271	0.0049	0.001	−0.027	0.0008	**0.311**
0.068	−0.271	0.0048	0.001	−0.026	0.0012	0.032
0.07	−0.27	0.005	0	−0.025	0.0016	0.001
0.072	−0.269	0.0052	0	−0.025	0.0016	0.003
0.074	−0.268	0.0057	0	−0.024	0.0017	0.001
0.076	−0.267	0.0059	0	−0.024	0.0018	0.001

A.27 Verizon Communications Inc. Common Stock (VZ)

The VZ price time series contains 55,684,512 transactions.

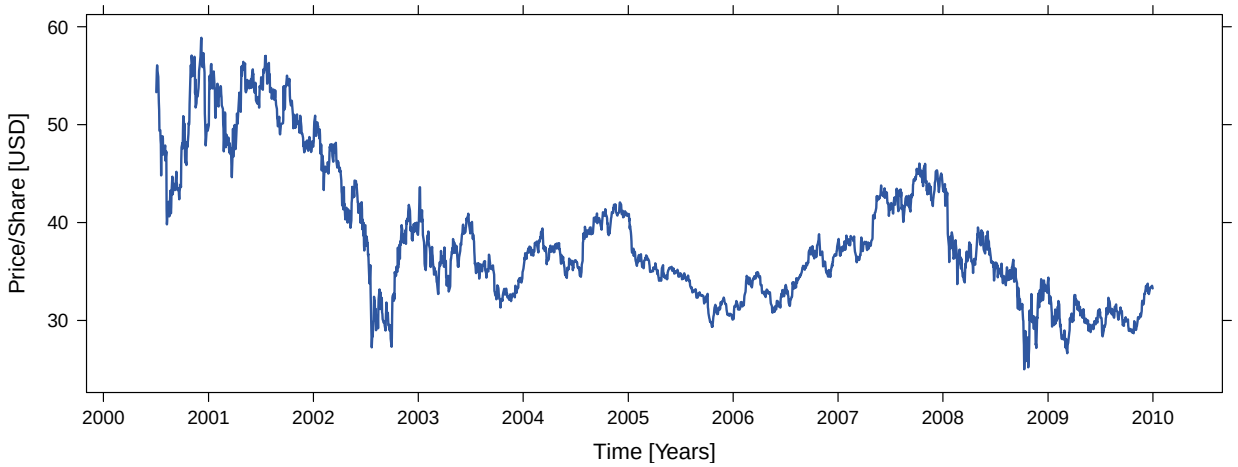

Figure A.53: A stock split occurred on 30 June 1998 [split ratio 2:1].

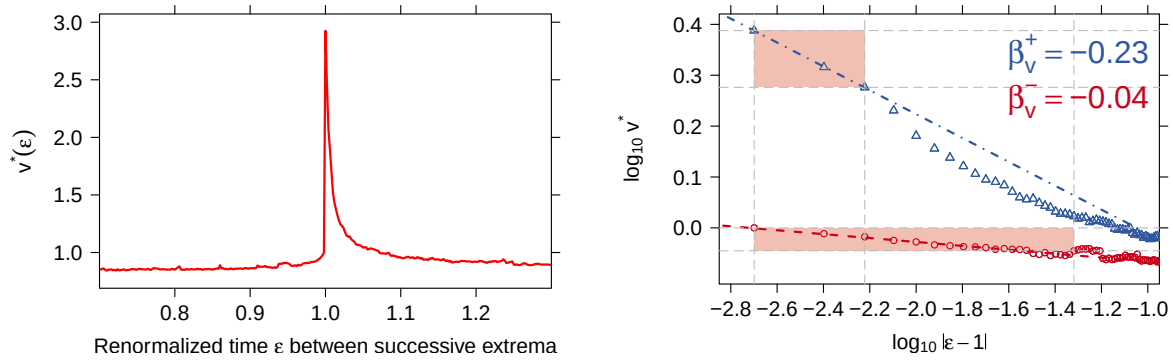

Figure A.54: *left*: aggregated volume $v^*(\varepsilon)$. *right*: $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram.

Table A.27: Statistical test of power-law hypothesis for the VZ volume time series: Scaling parameters of the hypothesized power-law model are shown for both $v^*(\varepsilon)$ before (β_v^-) and $v^*(\varepsilon)$ after (β_v^+) the trend switching point $\varepsilon = 1$ in dependence of $|\varepsilon - 1|_{\text{cut}}$. Additionally, the corresponding values of the KS statistic, D_v^- and D_v^+, are given. The power-law hypothesis is supported if the p-value is larger than 0.1.

$\lvert\varepsilon - 1\rvert_{\text{cut}}$	β_v^+	D_v^+	p-value	β_v^-	D_v^-	p-value
0.004	−0.239	0	**0.727**	−0.038	0	**1**
0.006	−0.234	0.0007	**0.688**	−0.036	0.0001	**0.916**
0.008	−0.254	0.0049	0.003	−0.04	0.0007	**0.697**
0.01	−0.283	0.0113	0	−0.039	0.0006	**0.839**
0.012	−0.298	0.0132	0	−0.041	0.0009	**0.612**
0.014	−0.304	0.0127	0	−0.042	0.0009	**0.568**
0.016	−0.306	0.0118	0	−0.042	0.0007	**0.708**
0.018	−0.308	0.0109	0	−0.041	0.0005	**0.896**
0.02	−0.308	0.0101	0	−0.04	0.0004	**0.947**
0.022	−0.305	0.0093	0	−0.04	0.0004	**0.948**
0.024	−0.302	0.0087	0	−0.039	0.0006	**0.791**
0.026	−0.3	0.0081	0	−0.039	0.0007	**0.692**
0.028	−0.299	0.0077	0	−0.038	0.0009	**0.425**
0.03	−0.296	0.0074	0	−0.038	0.001	**0.319**
0.032	−0.293	0.0072	0	−0.038	0.0007	**0.62**
0.034	−0.29	0.007	0	−0.039	0.0007	**0.578**
0.036	−0.288	0.0068	0	−0.039	0.0006	**0.74**
0.038	−0.285	0.0067	0	−0.039	0.0008	**0.44**
0.04	−0.283	0.0066	0	−0.039	0.0008	**0.504**
0.042	−0.281	0.0064	0	−0.039	0.0007	**0.511**
0.044	−0.279	0.0064	0	−0.04	0.0007	**0.507**
0.046	−0.276	0.0064	0	−0.039	0.0006	**0.708**
0.048	−0.274	0.0064	0	−0.038	0.001	**0.177**
0.05	−0.272	0.0063	0	−0.036	0.0018	0
0.052	−0.269	0.0064	0	−0.035	0.0027	0
0.054	−0.266	0.0068	0	−0.034	0.0034	0
0.056	−0.264	0.0072	0	−0.032	0.0039	0
0.058	−0.262	0.0078	0	−0.032	0.0042	0
0.06	−0.259	0.0085	0	−0.031	0.0044	0
0.062	−0.256	0.0091	0	−0.03	0.0046	0
0.064	−0.254	0.0096	0	−0.031	0.0041	0
0.066	−0.251	0.0101	0	−0.031	0.0037	0
0.068	−0.249	0.0104	0	−0.032	0.0035	0
0.07	−0.247	0.0109	0	−0.033	0.0032	0
0.072	−0.245	0.0111	0	−0.033	0.003	0
0.074	−0.244	0.0113	0	−0.033	0.0029	0
0.076	−0.242	0.0115	0	−0.033	0.0028	0

A.28 Wal-Mart Stores, Inc. Common Stock (WMT)

The WMT price time series contains 81,511,936 transactions.

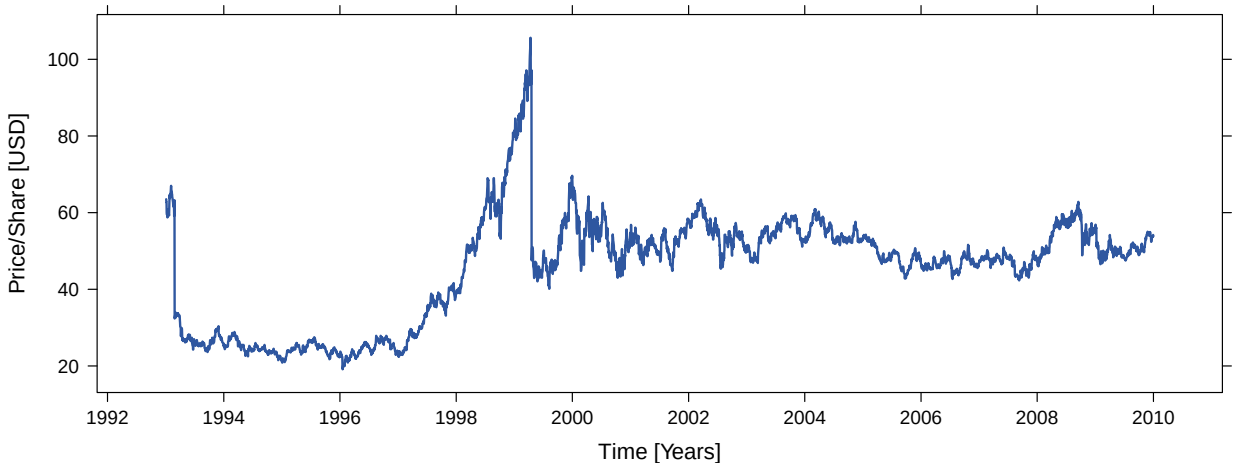

Figure A.55: Stock splits occurred on 26 February 1993 [split ratio 2:1] and 20 April 1999 [split ratio 2:1].

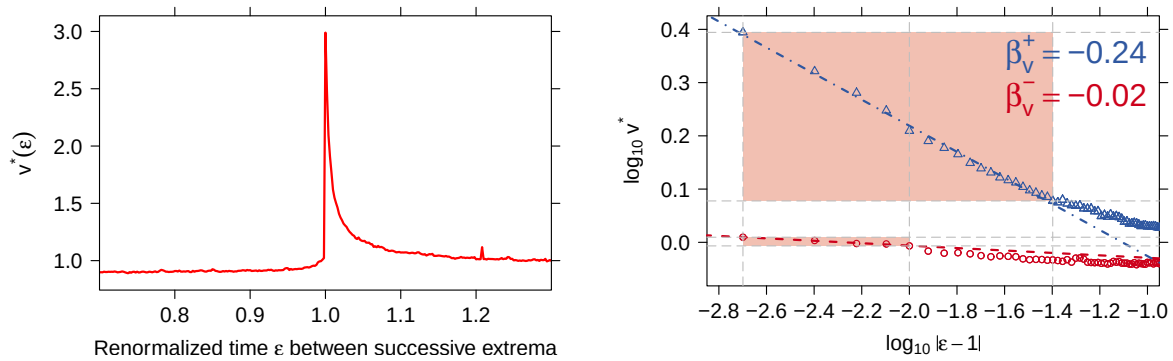

Figure A.56: *left*: aggregated volume $v^*(\varepsilon)$. *right*: $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram.

Table A.28: Statistical test of power-law hypothesis for the WMT volume time series: Scaling parameters of the hypothesized power-law model are shown for both $v^*(\varepsilon)$ before (β_v^-) and $v^*(\varepsilon)$ after (β_v^+) the trend switching point $\varepsilon = 1$ in dependence of $|\varepsilon - 1|_{\text{cut}}$. Additionally, the corresponding values of the KS statistic, D_v^- and D_v^+, are given. The power-law hypothesis is supported if the p-value is larger than 0.1.

$\lvert\varepsilon-1\rvert_{\text{cut}}$	β_v^+	D_v^+	p-value	β_v^-	D_v^-	p-value
0.004	−0.242	0	**1**	−0.02	0	**1**
0.006	−0.238	0.0006	**0.67**	−0.024	0.0006	**0.568**
0.008	−0.242	0.0007	**0.892**	−0.022	0.0004	**0.836**
0.01	−0.257	0.0051	0	−0.022	0.0002	**0.966**
0.012	−0.262	0.0057	0	−0.028	0.0025	0.003
0.014	−0.262	0.005	0	−0.033	0.0035	0
0.016	−0.26	0.0043	0.024	−0.034	0.0033	0
0.018	−0.26	0.0039	0.08	−0.034	0.003	0
0.02	−0.26	0.0036	**0.269**	−0.036	0.0029	0
0.022	−0.259	0.0033	**0.508**	−0.037	0.0027	0
0.024	−0.257	0.003	**0.71**	−0.036	0.0025	0
0.026	−0.256	0.0029	**0.853**	−0.036	0.0023	0
0.028	−0.253	0.0027	**0.924**	−0.036	0.0021	0
0.03	−0.252	0.0026	**0.958**	−0.036	0.002	0
0.032	−0.25	0.0026	**0.971**	−0.037	0.0018	0.003
0.034	−0.249	0.0028	**0.929**	−0.037	0.0017	0
0.036	−0.247	0.003	**0.755**	−0.037	0.0016	0
0.038	−0.245	0.0034	**0.365**	−0.037	0.0016	0.003
0.04	−0.244	0.0035	**0.203**	−0.036	0.0015	0.008
0.042	−0.243	0.0037	0.087	−0.036	0.0014	0.001
0.044	−0.24	0.0041	0.002	−0.036	0.0014	0.008
0.046	−0.239	0.0045	0	−0.036	0.0013	0.006
0.048	−0.237	0.0048	0	−0.036	0.0013	0.008
0.05	−0.235	0.0053	0	−0.035	0.0013	0.002
0.052	−0.233	0.0058	0	−0.034	0.0013	0.004
0.054	−0.231	0.0063	0	−0.033	0.0013	0.003
0.056	−0.229	0.0068	0	−0.033	0.0012	0.006
0.058	−0.227	0.0074	0	−0.033	0.0012	0.008
0.06	−0.226	0.0077	0	−0.033	0.0012	0.008
0.062	−0.224	0.008	0	−0.033	0.0011	0.005
0.064	−0.223	0.0081	0	−0.033	0.0011	0.011
0.066	−0.221	0.0082	0	−0.033	0.0011	0.011
0.068	−0.22	0.0083	0	−0.033	0.001	0.013
0.07	−0.219	0.0085	0	−0.033	0.001	0.012
0.072	−0.217	0.0086	0	−0.033	0.0011	0.001
0.074	−0.216	0.0088	0	−0.032	0.0012	0.001
0.076	−0.214	0.009	0	−0.032	0.0013	0.001

A.29 Exxon Mobil Corporation Common Stock (XOM)

The XOM price time series contains 116,030,372 transactions.

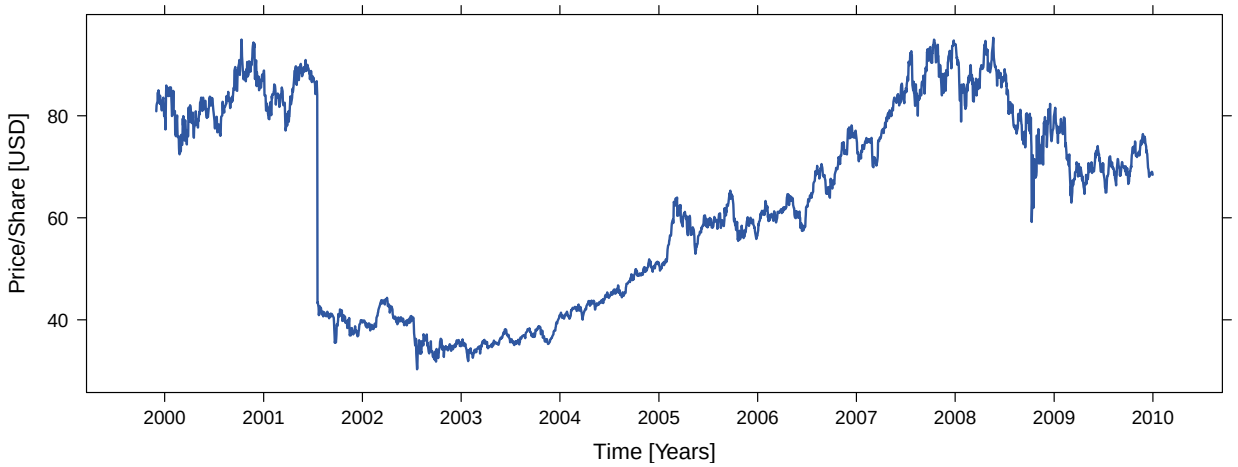

Figure A.57: Stock splits occurred on 14 April 1997 [split ratio 2:1] and 19 July 2001 [split ratio 2:1].

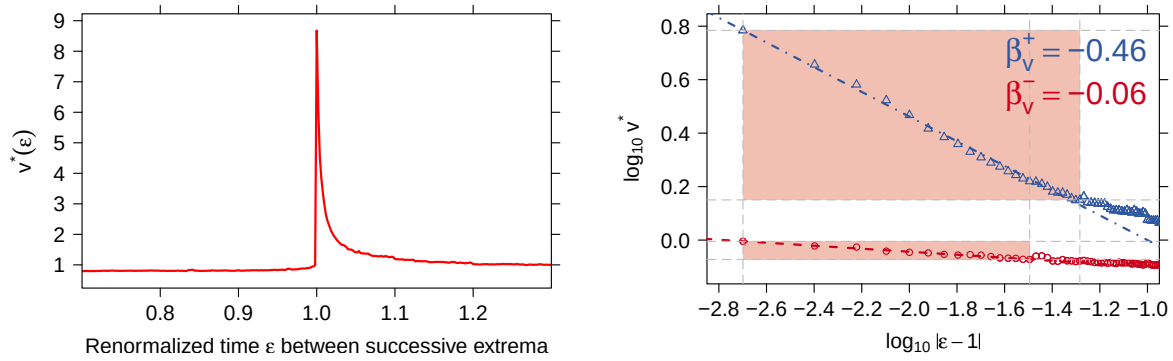

Figure A.58: *left*: aggregated volume $v^*(\varepsilon)$. *right*: $v^*(\varepsilon)$ versus $|\varepsilon - 1|$ as a log–log histogram.

Table A.29: Statistical test of power-law hypothesis for the XOM volume time series: Scaling parameters of the hypothesized power-law model are shown for both $v^*(\varepsilon)$ before (β_v^-) and $v^*(\varepsilon)$ after (β_v^+) the trend switching point $\varepsilon = 1$ in dependence of $|\varepsilon - 1|_{\text{cut}}$. Additionally, the corresponding values of the KS statistic, D_v^- and D_v^+, are given. The power-law hypothesis is supported if the p-value is larger than 0.1.

$\lvert\varepsilon-1\rvert_{\text{cut}}$	β_v^+	D_v^+	p-value	β_v^-	D_v^-	p-value
0.004	−0.425	0	**1**	−0.056	0	**1**
0.006	−0.425	0	**0.996**	−0.045	0.0017	0.068
0.008	−0.432	0.0017	**0.917**	−0.056	0.0023	0.013
0.01	−0.446	0.0054	0.005	−0.058	0.0022	0.008
0.012	−0.464	0.0088	0	−0.057	0.0018	0.036
0.014	−0.473	0.0103	0	−0.057	0.0016	0.078
0.016	−0.477	0.0104	0	−0.057	0.0014	**0.141**
0.018	−0.482	0.0104	0	−0.055	0.0012	**0.22**
0.02	−0.484	0.01	0	−0.054	0.0013	**0.115**
0.022	−0.486	0.0096	0	−0.054	0.0012	**0.143**
0.024	−0.486	0.0091	0	−0.055	0.0009	**0.429**
0.026	−0.486	0.0086	0	−0.055	0.0008	**0.557**
0.028	−0.486	0.0082	0.001	−0.055	0.0007	**0.592**
0.03	−0.486	0.0079	0.002	−0.055	0.0007	**0.695**
0.032	−0.484	0.0076	0.013	−0.056	0.0007	**0.735**
0.034	−0.481	0.0073	0.06	−0.053	0.0016	0.003
0.036	−0.478	0.0072	**0.12**	−0.05	0.003	0
0.038	−0.476	0.007	**0.188**	−0.049	0.0035	0
0.04	−0.475	0.0069	**0.382**	−0.05	0.0029	0
0.042	−0.473	0.0068	**0.459**	−0.051	0.0025	0
0.044	−0.47	0.0068	**0.414**	−0.05	0.0025	0
0.046	−0.467	0.0068	**0.344**	−0.051	0.0024	0
0.048	−0.466	0.0068	**0.356**	−0.051	0.0023	0
0.05	−0.464	0.0067	**0.294**	−0.051	0.0021	0
0.052	−0.461	0.0068	**0.156**	−0.051	0.0021	0
0.054	−0.457	0.007	0.009	−0.05	0.0021	0
0.056	−0.455	0.0071	0	−0.05	0.0021	0
0.058	−0.452	0.0072	0	−0.05	0.002	0
0.06	−0.449	0.0074	0	−0.05	0.0019	0
0.062	−0.446	0.0077	0	−0.05	0.0018	0
0.064	−0.443	0.0087	0	−0.05	0.0018	0
0.066	−0.439	0.0098	0	−0.05	0.0017	0
0.068	−0.437	0.0104	0	−0.05	0.0017	0
0.07	−0.435	0.0109	0	−0.05	0.0016	0
0.072	−0.433	0.0113	0	−0.05	0.0016	0
0.074	−0.431	0.0117	0	−0.05	0.0015	0
0.076	−0.428	0.0122	0	−0.05	0.0015	0

Bibliography

[AFS09] A. Alfonsi, A. Fruth, and A. Schied. Optimal execution strategies in limit order books with general shape functions. *Quantitative Finance*, 10(2):143–157, 2009.

[ALT08] J.A. Anderson, C.D. Lorenz, and A. Travesset. General purpose molecular dynamics simulations fully implemented on graphics processing units. *Journal of Computational Physics*, 227(10):5342–5359, 2008.

[And72] P.W. Andersen. More is different. *Science*, 177(4047):393–396, 1972.

[And09] J. Anderson. U.S. proposes ban on 'flash' trading on wall street. New York Times, 18 September 2009.

[ATI06] ATI Technologies Inc. *ATI CTM Guide, Technical Reference Manual Version 1.01*, 2006.

[Aus00] M. Ausloos. Statistical physics in foreign exchange currency and stock markets. *Physica A*, 285(1):48–65, 2000.

[Axt01] R.L. Axtell. Zipf distribution of U.S. firm sizes. *Science*, 293(5536):1818–1820, 2001.

[Bac00] L. Bachelier. *Théorie de la Spéculation*. PhD thesis, Faculté des Sciences de Paris, 1900.

[BB04] P. Bank and D. Baum. Hedging and portfolio optimization in financial markets with a large trader. *Mathematical Finance*, 14(1):1–18, 2004.

[BBO+01] J. Bentner, G. Bauer, G.M. Obermair, I. Morgenstern, and J.J. Schneider. Optimization of the time-dependent traveling salesman problem with Monte Carlo methods. *Physical Review E*, 64:036701, 2001.

[BBPZ08] R.G. Belleman, J. Bédorf, and S.F. Portegies Zwart. High performance direct gravitational n-body simulations on graphics processing units ii: An implementation in CUDA. *New Astronomy*, 13(2):103–112, 2008.

[BDGH93] A. Banerjee, J.J. Dolado, J.W. Galbraith, and D.F. Hendry. *Cointegration, Error Correction, and the Econometric Analysis of Non-Stationary Data*. Oxford University Press, 1993.

[Ben98] D.J. Bennett. *Randomness*. Harvard University Press, 1998.

[BGM06] K.E. Bassler, G.H. Gunaratne, and J.L. McCauley. Markov processes, hurst exponents, and nonlinear diffusion equations: With application to finance. *Physica A*, 369(2):343–353, 2006.

[BHG06] D. Brockmann, L. Hufnagel, and T. Geisel. The scaling laws of human travel. *Nature*, 439:462–465, 2006.

[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.

[Bin81] K. Binder. Finite size scaling analysis of Ising model block distribution functions. *Zeitschrift für Physik B Condensed Matter*, 43(2):119–140, 1981.

[BKS02] A. Bunde, J. Kropp, and H.J. Schellnhuber, editors. *The Science of Disaster: Climate Disruptions, Market Crashes, and Heart Attacks*. Springer, 2002.

[BL98] D. Bertsimas and A.W. Lo. Optimal control of execution costs. *Journal of Financial Markets*, 1(1):1–50, 1998.

[BL01] K. Binder and E. Luijten. Monte Carlo tests of renormalization-group predictions for critical phenomena in Ising models. *Phys. Rep.*, 344:179–253, 2001.

[BP01] J.-P. Bouchaud and M. Potters. More stylized facts of financial markets: Leverage effect and downside correlations. *Physica A*, 299(1-2):60–70, 2001.

[BP05] M.K. Brunnermeier and L.H. Pedersen. Predatory trading. *Journal of Finance*, 60(4):1825–1864, 2005.

[BP09] J.-P. Bouchaud and M. Potters. *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management*. Cambridge University Press, 2 edition, 2009.

[BPS97] P. Bak, M. Paczuski, and M. Shubik. Price variations in a stock market with many agents. *Physica A*, 246(3-4):430–453, 1997.

[BS73] F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.

[BVP10] B. Block, P. Virnau, and T. Preis. Multi-GPU accelereted multi-spin Monte Carlo simulation of the 2D Ising model. *Accepted for Publication in Computer Physics Communications*, tba:tba, 2010.

[CB00] R. Cont and J.-P. Bouchaud. Herd behavior and aggregate fluctuations in financial markets. *Macroeconomic Dynamics*, 4(2):170–196, 2000.

[CCS04] A. Carbone, G. Castelli, and H.E. Stanley. Time-dependent hurst exponent in financial time series. *Physica A*, 344(1-2):267–271, 2004.

[CF00] K. Chan and W.-M. Fong. Trade size, order imbalance, and the volatility-volume relation. *Journal of Financial Economics*, 57(2):247–273, 2000.

[CKB+00] J.-M. Courtault, Y. Kabanov, B. Bru, P. Crépel, I. Lebon, and A. Le Marchand. Louis bachelier on the centenary of théorie de la spéculation. *Mathematical Finance*, 10(3):341–353, 2000.

[CMZ97] G. Caldarelli, M. Marsili, and Y.-C. Zhang. A prototype model of stock exchange. *Europhysics Letters*, 40:479–484, 1997.

[Con01] R. Cont. Empirical properties of asset returns: Stylized facts and statistical issues. *Quant. Finance*, 1(2):223–236, 2001.

[CS03] D. Challet and R. Stinchcombe. Non-constant rates and over-diffusive prices in a simple model of limit order markets. *Quantitative Finance*, 3(3):155–162, 2003.

[CSN09] A. Clauset, C.R. Shalizi, and M.E.J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51:661–703, 2009.

[Deu03] H.-P. Deutsch. *Derivate und Interne Modelle. Modernes Risikomanagement.* Schäffer-Poeschel, 2003.

[Deu06] Annual report 2005. Deutsche Börse Group, 2006.

[Deu10] Deutsche Börse Systems expands ultra low latency network. Press Release, Deutsche Börse AG, January 2010.

[DFG+03] M.G. Daniels, J.D. Farmer, L. Gillemot, 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(10):108102, 2003.

[DM98] L. Dagum and R. Menon. Openmp: An industry-standard API for shared-memory programming. *IEEE Computational Science and Engineering*, 5(1):46–55, 1998.

[Dub04] R. Dubil. *An Arbitrage Guide to Financial Markets*. Wiley, 2004.

[DW00] G.A. Darbellay and D. Wuertz. Entropy as a tool for analyzing statistical dependences in financial time series. *Physica A*, 287(3-4):429–439, 2000.

[Ein05] A. Einstein. Über die von der molekularkinetischen Theorie der Wärme geförderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik*, 17:549–560, 1905.

[EK06] Z. Eisler and J. Kertész. Scaling theory of temporal correlations and size-dependent fluctuations in the traded value of stocks. *Physical Review E*, 73(4):046109, 2006.

[Fam63] E.F. Fama. Mandelbrot and the stable Paretian hypothesis. *Journal of Business*, 36(4):420–429, 1963.

[FBV06] B. Fierro, F. Bachmann, and E.E. Vogel. Phase transition in 2D and 3D Ising model by time-series analysis. *Physica B: Condensed Matter*, 384(1-2):215–217, 2006.

[Fis67] M.E. Fisher. The theory of equilibrium critical phenomena. *Reports on Progress in Physics*, 30(2):615–730, 1967.

[FK03] R. Fernando and M.J. Kilgard. *The Cg Tutorial: The Definitive Guide to Programmable Real-Time Graphics*, 2003.

[Fou10] Wikimedia Foundation. *http://www.wikipedia.org/*, April 2010.

[Gar89] P.M. Garber. Tulipmania. *Journal of Political Economy*, 97:535–560, 1989.

[Gar00] P.M. Garber. *Famous First Bubbles: The Fundamentals of Early Manias*. MIT Press, 2000.

[GDD+97] D.M. Guillaume, M.M. Dacorogna, R.R. Davé, U.A. Müller, R.B. Olsen, and O.V. Pictet. From the bird's eye to the microscope: A survey of new stylized facts of the intra-daily foreign exchange markets. *Finance and Stochastics*, 1(2):95–129, 1997.

[GFB+04] E. Gabriel, G.E. Fagg, G. Bosilca, T. Angskun, J.J. Dongarra, J.M. Squyres, V. Sahay, P. Kambadur, B. Barrett, A. Lumsdaine, R.H. Castain, D.J. Daniel, R.L. Graham, and T.S. Woodall. Open MPI: Goals, concept, and design of a next generation MPI implementation. *Proceedings, 11th European PVM/MPI Users' Group Meeting*, pages 97–104, 2004.

[GGPS03] X. Gabaix, P. Gopikrishnan, V. Plerou, and H.E. Stanley. A theory of power-law distributions in financial market fluctuations. *Nature*, 423(6937):267–270, 2003.

[Gol07] A. Goldgar. *Tulipmania: Money, Honor, and Knowledge in the Dutch Golden Age*. University of Chicago Press, 2007.

[Goo10] Google Corporation. *http://www.google.de/trends*, April 2010.

[GPA+99] P. Gopikrishnan, V. Plerou, L.A.N. Amaral, M. Meyer, and H.E. Stanley. Scaling of the distribution of fluctuations of financial market indices. *Physical Review E*, 60:5305–5316, 1999.

[GZ09] G.-F. Gu and W.-X. Zhou. On the probability distribution of stock returns in the Mike-Farmer model. *European Physical Journal B*, 67(4):585–592, 2009.

[Hel97] D. Helbing. Empirical traffic data and their implications for traffic modeling. *Physical Review E*, 55:R25, 1997.

[Heu93] H.-O. Heuer. Critical crossover phenomena in disordered Ising systems. *Journal of Physics A: General Physics*, 26(6):L333–L339, 1993.

[HFV00] D. Helbing, I. Farkas, and T. Vicsek. Simulating dynamical features of escape panic. *Nature*, 407(6803):487–490, 2000.

[HH07] A.D. Hall and N. Hautsch. Modelling the buy and sell intensity in a limit order book market. *Journal of Financial Markets*, 10(3):249–286, 2007.

[HKK+08] W. Haerdle, T. Kleinow, A. Korostelev, C. Logeay, and E. Platen. Semiparametric diffusion estimation and application to a stock market index. *Quantitative Finance*, 8(1):81–92, 2008.

[HLL08] P.R. Hansen, J. Large, and A. Lunde. Moving average estimators of integrated variance. *Econometric Reviews*, 27:79–111, 2008.

[Hom02] C.H. Hommes. Modeling the stylized facts in finance through simple nonlinear adaptive systems. *Proceedings of the National Academy of Sciences*, 99:7221–7228, 2002.

[Hur51] H.E. Hurst. Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 116:770–808, 1951.

[IK88] N. Ito and Y. Kanada. An effective algorithm for the Monte-Carlo simulation of the Ising-model on a vector processor. *Supercomputer*, 5:31, 1988.

[IK90] N. Ito and Y. Kanada. Random number generation on a vector processor. *Supercomputer*, 7(1):29–35, 1990.

[Isi25] E. Ising. Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift für Physik*, 31(1):253–258, 1925.

[IYPL04] P.C. Ivanov, A. Yuen, B. Podobnik, and Y. Lee. Common scaling patterns in intertrade times of U. S. stocks. *Phys. Rev. E*, 69(5):056107, 2004.

[JCZ09] Z.-Q. Jiang, W. Chen, and W.-X. Zhou. Detrended fluctuation analysis of intertrade durations. *Physica A*, 388(4):433–440, 2009.

[JKL94] C. Jones, G. Kaul, and M. Lipson. Transactions, volume, and volatility. *Review of Financial Studies*, 7(4):631–651, 1994.

[KHH02] A. Krawiecki, J.A. Holyst, and D. Helbing. Volatility clustering and scaling for financial time series due to attractor bubbling. *Physical Review Letters*, 89(15):158701, 2002.

[Kop95] I. Koponen. Analytic approach to the problem of convergence of truncated Lévy flights towards the gaussian stochastic process. *Physical Review E*, 52(1):1197–1199, 1995.

[KPSS92] D. Kwiatkowski, P.C.B. Phillips, P. Schmidt, and Y. Shin. Testing the null hypothesis of stationarity against the alternative of a unit root : How sure are we that economic time series have a unit root? *Journal of Econometrics*, 54:159–178, 1992.

[Kre03] U. Krengel. *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Vieweg, 2003.

[Kru96] P. Krugman. *The Self-Organizing Economy*. Wiley-Blackwell, 1996.

[KSRS08] H. Köstler, R. Schmid, U. Rüde, and C. Scheit. A parallel multigrid accelerated poisson solver for ab initio molecular dynamics applications. *Computing and Visualization in Science*, 11:115–122, 2008.

[KSY06] K. Kiyono, Z.R. Struzik, and Y. Yamamoto. Criticality and phase transition in stock-price fluctuations. *Physical Review Letters*, 96(6):068701, 2006.

[Kul07] N. Kuls. Das New Yorker Börsenparkett schrumpft auf die Hälfte. Frankfurter Allgemeine Zeitung, 27 September 2007.

[LB05] D.P. Landau and K. Binder. *A Guide to Monte Carlo Simulations in Statistical Physics*. Cambridge University Press, 2 edition, 2005.

[LCBP99] L. Laloux, P. Cizeau, J.-P. Bouchaud, and M. Potters. Noise dressing of financial correlation matrices. *Physical Review Letters*, 83(7):1467–1470, 1999.

[LFM03] F. Lillo, J.D. Farmer, and R.N. Mantegna. Econophysics: Master curve for price-impact function. *Nature*, 421(6919):129–130, 2003.

[LM99] T. Lux and M. Marchesi. Scaling and criticality in a stochastic multi-agent model of a financial market. *Nature*, 397(6719):498–500, 1999.

[LSVMW08] W. Liu, B. Schmidt, G. Voss, and W. Müller-Wittig. Accelerating molecular dynamics simulations using graphics processing units with CUDA. *Computer Physics Communications*, 179(9):634–641, 2008.

[Lux96] T. Lux. The stable paretian hypothesis and the frequency of large returns: An examination of major german stocks. *Applied Financial Economics*, 6(6):463–475, 1996.

[LWK03] W. Li, X. Wei, and A. Kaufman. Implementing lattice Boltzmann computation on graphics hardware. *Visual Computer*, 19(7-8):444–456, 2003.

[Man63] B. Mandelbrot. The variation of certain speculative prices. *Journal of Business*, 36:394–419, 1963.

[Mas00] S. Maslov. Simple model of a limit order-driven market. *Physica A*, 278(3):571–578, 2000.

[MdOdOS99] S. Moss de Oliveira, P.M.C. de Oliveira, and D. Stauffer. *Evolution, Money, War, and Computers*. Teubner, 1999.

[Mer73] R.C. Merton. Theory of rational option pricing. *Bell Journal of Economics and Management Science*, 4(1):141–183, 1973.

[MGB07] J.L. McCauley, G.H. Gunaratne, and K.E. Bassler. Hurst exponents, Markov processes, and fractal Brownian motion. *Physica A*, 379:1–9, 2007.

[MH04] B.B. Mandelbrot and R.L. Hudson. *The (Mis)behavior of Markets: A Fractal View of Risk, Ruin and Reward*. Basic Books, 2004.

[MM01] S. Maslov and M. Mills. Price fluctuations from the order book perspective—empirical facts and a simple model. *Physica A*, 299(1-2):234–246, 2001.

[MN98] M. Matsumoto and T. Nishimura. Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation*, 8(1):3–30, 1998.

[MRR+53] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[MS94] R.N. Mantegna and H.E. Stanley. Stochastic process with ultraslow convergence to a Gaussian: The truncated Lévy flight. *Physical Review Letters*, 73(22):2946–2949, 1994.

[MS95] R.N. Mantegna and H.E. Stanley. Scaling behaviour in the dynamics of an economic index. *Nature*, 376(6535):46–49, 1995.

[MS00] R.N. Mantegna and H.E. Stanley. *An Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge University Press, 2000.

[MSS09] A. Müller, J.J. Schneider, and E. Schömer. Packing a multidisperse system of hard disks in a circular environment. *Physical Review E*, 79(2):021102, 2009.

[NS92] K. Nagel and M. Schreckenberg. A cellular automaton model for freeway traffic. *Journal de Physique I France*, 2:2221–2229, 1992.

[NVI08a] NVIDIA Corporation. *NVIDIA CUDA Compute Unified Device Architecture, Programming Guide Version 2.0*, 2008.

[NVI08b] NVIDIA Corporation. *NVIDIA GeForce GTX 280 Specifications*, 2008.

[NVI09] NVIDIA Corporation. *NVIDIA Tesla C1060 Specifications*, 2009.

[O'H97] M. O'Hara. *Market Microstructure Theory*. Blackwell Publishing, 1997.

[Ons44] L. Onsager. Crystal statistics. I. A two-dimensional model with an order-disorder transition. *Physical Review*, 65(3-4):117–149, 1944.

[Osb59] M.F.M. Osborne. Brownian motion in the stock market. *Operations Research*, 7(2):145–173, 1959.

[Osb95] M.F.M. Osborne. *The Stock Market and Finance from a Physicist's Viewpoint*. Crossgar Press, 1995.

[PB00] W. Paul and J. Baschnagel. *Stochastic Processes: From Physics to Finance*. Springer, 2000.

[PB10] T. Preis and B. Block. A practical guide to massively parallel Monte Carlo simulations: The Ising model. In W. Hwu, editor, *Accepted in GPU Computing Gems*. Addison-Wesley, 2010.

[PFTV92] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 2 edition, 1992.

[PGGS02] V. Plerou, P. Gopikrishnan, X. Gabaix, and H.E. Stanley. Quantifying stock-price response to demand fluctuations. *Physical Review E*, 66(2):027104, 2002.

[PGPS06] T. Preis, S. Golke, W. Paul, and J.J. Schneider. Multi–agent-based order book model of financial markets. *Europhysics Letters*, 75:510–516, 2006.

[PGPS07] T. Preis, S. Golke, W. Paul, and J.J. Schneider. Statistical analysis of financial returns for a multiagent order book model of asset trading. *Physical Review E*, 76:016108, 2007.

[PGR+99] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral, and H.E. Stanley. Universal and nonuniversal properties of cross correlations in financial time series. *Physical Review Letters*, 83(7):1471–1474, 1999.

[PPS08] T. Preis, W. Paul, and J.J. Schneider. Fluctuation patterns in high-frequency financial asset returns. *Europhysics Letters*, 82:68005, 2008.

[Pre06] T. Preis. Investigation of a multi-agent based order book model of financial markets. Diploma thesis, 2006.

[Pre10] T. Preis. Simulating the microstructure of financial markets. *Journal of Physics: Conference Series*, 221:012019, 2010.

[Pre11] T. Preis. An order book model simulating modern financial markets. In J.J. Schneider, editor, *Accepted in Proceedings of the Summer School on Socio- and Econophysics*. Springer Berlin, 2011.

[PS99] M. Pasquini and M. Serva. Multiscale behaviour of volatility autocorrelations in a financial market. *Economics Letters*, 65(3):275–279, 1999.

[PS08] M. Politi and E. Scalas. Fitting the empirical distribution of intertrade durations. *Physica A*, 387(8-9):2025–2034, 2008.

[PS09] T. Preis and H.E. Stanley. How to characterize trend switching processes in financial markets. *Bulletin of the Asia Pacific Center for Theoretical Physics [Cover Image]*, 23:18–23, 2009.

[PS10a] T. Preis and H.E. Stanley. Switching phenomena in a system with no switches. *Journal of Statistical Physics*, 138:431–446, 2010.

[PS10b] T. Preis and H.E. Stanley. Trend switching processes in financial markets. In M. Takayasu, T. Watanabe, and H. Takayasu, editors, *Econophysics Approaches to Large-Scale Business Data and Financial Crisis*. Springer Tokyo, 2010.

[PSS10] T. Preis, J.J. Schneider, and H.E. Stanley. Switching processes in financial markets. *Submitted to Science*, 2010.

[PVPS09a] T. Preis, P. Virnau, W. Paul, and J.J. Schneider. Accelerated fluctuation analysis by graphic cards and complex pattern formation in financial markets. *New Journal of Physics*, 11:093024, 2009.

[PVPS09b] T. Preis, P. Virnau, W. Paul, and J.J. Schneider. GPU accelerated Monte Carlo simulation of the 2D and 3D Ising model. *Journal of Computational Physics*, 228:4468–4477, 2009.

[PZBG07] S.F. Portegies Zwart, R.G. Bellemana, and P.M. Geldof. High-performance direct gravitational n-body simulations on graphics processing units. *New Astronomy*, 12:641–650, 2007.

[Ros04] R.J. Rost. *OpenGL Shading Language*, 2004.

[SBF+10] H.E. Stanley, S.V. Buldyrev, G. Franzese, S. Havlin, F. Mallamace, P. Kumar, V. Plerou, and T. Preis. Correlated randomness and switching phenomena. *Physica A*, 389:2875–2888, 2010.

[SBF+11] H.E. Stanley, S.V. Buldyrev, G. Franzese, S. Havlin, F. Mallamace, M.G. Mazza, P. Kumar, V. Plerou, T. Preis, K. Stokely, and L. Xu. Switching phenomena. In C. Rangacharyulu and E. Haven, editors, *Accepted in Proceedings of the First Interdisciplinary CHESS Interactions Conference*. World Scientific Singpore, 2011.

[Sch69] T.C. Schelling. Models of segregation. *American Economic Review*, 59:488–493, 1969.

[Sch04] J.J. Schneider. The influence of contrarians and opportunists on the stability of a democracy in the Sznajd model. *International Journal of Modern Physics C*, 15:659–674, 2004.

[SD84] S.E. Said and D.A. Dickey. Testing for unit roots in autoregressive - moving average models with unknown order. *Biometrika*, 71:599–607, 1984.

[SEE+03] R. Susukita, T. Ebisuzaki, B.G. Elmegreen, H. Furusawa, K. Kato, A. Kawai, Y. Kobayashi, T. Koishi, G.D. McNiven, T. Narumi, and K. Yasuoka. Hardware accelerator for molecular dynamics: MDGRAPE-2. *Computer Physics Communications*, 155(2):115–131, 2003.

[SFG+06] M. Serva, U.L. Fulco, I.M. Gléria, M.L. Lyra, F. Petroni, and G.M. Viswanathan. A Markov model of financial returns. *Physica A*, 363(2):393–403, 2006.

[SFGK03] E. Smith, J.D. Farmer, L. Gillemot, and S. Krishnamurthy. Statistical theory of the continuous double auction. *Quantitative Finance*, 3(6):481–514, 2003.

[SfSMdO00] D. Stau ffer, A.O. Sousa, and S. Moss de Oliveira. Generalization to square lattice of Sznajd sociophysics model. *International Journal of Modern Physics C*, 11:1239–1245, 2000.

[SH05a] J.J. Schneider and C. Hirtreiter. The democracy-ochlocracy-dictatorship transition in the Sznajd model and in the Ising model. *Physica A*, 353:539–545, 2005.

[SH05b] J.J. Schneider and C. Hirtreiter. The impact of election results on the member numbers of the large parties in Bavaria and Germany. *International Journal of Modern Physics C*, 16:1165–1215, 2005.

[SH05c] J.J. Schneider and C. Hirtreiter. Scaling laws for the lifetimes of governments in the Sznajd democracy. *International Journal of Modern Physics C*, 16:157–165, 2005.

[Sha06] W.T. Shaw. Sampling Student's t distribution – use of the inverse cumulative distribution function. *Journal of Computational Finance*, 9(4):37–73, 2006.

[Shl00] A. Shleifer. *Inefficient Markets: An Introduction to Behavioral Finance*. Oxford University Press, 2000. cited By (since 1996) 252.

[SK06] J.J. Schneider and S. Kirkpatrick. *Stochastic Optimization*. Springer, 2006.

[SP98] D. Stauffer and T.J.P. Penna. Crossover in the Cont-Bouchaud percolation model for market fluctuations. *Physica A*, 256(1-2):284–290, 1998.

[SPF+07] J.E. Stone, J.C. Phillips, P.L. Freddolino, D.J. Hardy, L.G. Trabuco, and K. Schulten. Accelerating molecular modeling applications with graphics processors. *Journal of Computational Chemistry*, 28(16):2618–2640, 2007.

[SS99] D. Stauffer and D. Sornette. Self-organized percolation model for stock market fluctuations. *Physica A*, 271(3-4):496–506, 1999.

[Sta71] H.E. Stanley. *Introduction to Phase Transitions and Critical Phenomena*. Oxford University Press, 1971.

[SW87] R.H. Swendsen and J.-S. Wang. Nonuniversal critical dynamics in Monte Carlo simulations. *Physical Review Letters*, 58(2):86–88, 1987.

[SWS00] K. Sznajd-Weron and J. Sznajd. Opinion evolution in closed community. *International Journal of Modern Physics C*, 11(6):1157–1165, 2000.

[Tak06] H. Takayasu, editor. *Practical Fruits of Econophysics*. Springer, 2006.

[Tim09] The 50 best inventions of 2009. Time Magazine, November 23, 2009.

[TMB+05] S. Tomov, M. McGuigan, R. Bennett, G. Smith, and J. Spiletic. Benchmarking and implementation of probability-based simulations on programmable graphics cards. *Computers and Graphics*, 29(1):71–80, 2005.

[Tra06] S. Trautmann. *Investitionen. Bewertung, Auswahl und Risikomanagement*. Springer, 2006.

[VA97] N. Vandewalle and M. Ausloos. Coherent and random sequences in financial fluctuations. *Physica A*, 246(3-4):454–459, 1997.

[VFLS03] G.M. Viswanathan, U.L. Fulco, M.L. Lyra, and M. Serva. The origin of fat-tailed distributions in financial time series. *Physica A*, 329(1-2):273–280, 2003.

[vMAF+08] J.A. van Meel, A. Arnold, D. Frenkel, S.F. Portegies Zwart, and R.G. Belleman. Harvesting graphics power for MD simulations. *Molecular Simulation*, 34:259–266, 2008.

[Voi10] J. Voit. *The Statistical Mechanics of Financial Markets*. Springer, 2010.

[WBK+08] M. Wyart, J.-P. Bouchaud, J. Kockelkoren, M. Potters, and M. Vettorazzo. Relation between bid-ask spread, impact and volatility in double auction markets. *Quantitative Finance*, 8:41–57, 2008.

[Wol89] U. Wolff. Collective Monte Carlo updating for spin systems. *Physical Review Letters*, 62(4):361–364, 1989.

[WTT07] K. Watanabe, H. Takayasu, and M. Takayasu. A mathematical definition of the financial bubbles and crashes. *Physica A*, 383(1 SPEC. ISS.):120–124, 2007.

[WZK84] S. Wansleben, J.G. Zabolitzky, and C. Kalle. Monte Carlo simulation of Ising models by multispin coding on a vector computer. *Journal of Statistical Physics*, 37:271–282, 1984.

[YWC07] J. Yang, Y. Wang, and Y. Chen. GPU accelerated molecular dynamics simulation of thermal conductivities. *Journal of Computational Physics*, 221(2):799–804, 2007.

[ZAK06] A.G. Zawadowski, G. Andor, and J. Kertész. Short-term market reaction after extreme price changes of liquid stocks. *Quantitative Finance*, 6(4):283–295, 2006.

[ZHR81] R. Zorn, H.J. Herrmann, and C. Rebbi. Parallelization of the Ising simulation. *Computer Physics Communications*, 23:337–342, 1981.

[ZKA04] A.G. Zawadowski, J. Kertész, and G. Andor. Large price changes on small scales. *Physica A*, 344(1-2):221–226, 2004.

[ZS07] W.-X. Zhou and D. Sornette. Self-organizing Ising model of financial markets. *European Physical Journal B*, 55(2):175–181, 2007.