Marschinski, Robert; Matassini, Lorenzo

Working Paper

Financial markets as a complex system: A short time scale perspective

Research notes in economics & statistics, No. 01-4

Provided in cooperation with:
Deutsche Bank Research

Suggested citation: Marschinski, Robert; Matassini, Lorenzo (2001) : Financial markets as a complex system: A short time scale perspective, Research notes in economics & statistics, No. 01-4, http://hdl.handle.net/10419/40278

Economics

Quantitative Analysis

Deutsche Bank Research

November 09, 2001

Research Notes in Economics & Statistics

Financial Markets as a Complex System: A Short Time Scale Perspective

In this paper we want to discuss macroscopic and microscopic properties of financial markets. By analyzing quantitatively a database consisting of 13 minute per minute recorded financial time series, we identify some macroscopic statistical properties of the corresponding markets, with a special emphasize on temporal correlations. These analyses are performed by using both linear and nonlinear tools. Multivariate correlations are also tested for, which leads to the identification of a global coupling mechanism between the considered stock markets. The application of a new formalism, called transfer entropy, allows to measure the information flow between some financial time series. We then discuss some key aspects of recent attemps to model financial markets from a microscopic point of view. One model, that is based on the simulation of the order book, is described more in detail, and the results of its practical implementation are presented. We finally address some general aspects of forecasting and modeling, in particular the role of stochastic and nonlinear deterministic processes.

Robert Marschinski and Lorenzo Matassini

Editor:
Ann-Charlotte Eliasson
+49 69 910-31832
ann-charlotte.eliasson@db.com

Deutsche Bank Research
Frankfurt am Main
Germany
E-mail: marketing.dbr@db.com
Fax: +49 69 910-31877

Managing Directors
Axel Siedenberg
Norbert Walter

RESEARCH NOTE RN-01-4

Financial Markets as a Complex System:

A Short Time Scale Perspective

Robert Marschinski*,†,‡ and Lorenzo Matassini*

November 2001

Abstract

In this paper we want to discuss macroscopic and microscopic properties of financial markets. By analyzing quantitatively a database consisting of 13 minute per minute recorded financial time series, we identify some macroscopic statistical properties of the corresponding markets, with a special emphasize on temporal correlations. These analysis are performed by using both linear and nonlinear tools. Multivariate correlations are also tested for, which leads to the identification of a global coupling mechanism between the considered stock markets. The application of a new formalism, called *transfer entropy*, allows to measure the information flow between some financial time series. We then discuss some key aspects of recent attemps to model financial markets from a microscopic point of view. One model, that is based on the simulation of the order book, is described more in detail, and the results of its practical implementation are presented. We finally address some general aspects of forecasting and modeling, in particular the role of stochastic and nonlinear deterministic processes.

Keywords: High-frequency data; temporal correlations; simulated markets; econophysics; time series analysis.

JEL Codes: C00, C19, C53

* Max-Planck-Institute for the Physics of Complex Systems, Nöthnitzer Str. 38, D-01187 Dresden, Germany. Phone: +49 351 8711803, Fax: +49 351 8711999.

† Università di Bologna, Dipartimento di Fisica.

‡ Corresponding author. E-mail: `robert@mpipks-dresden.mpg.de`.

1 Introduction

The here presented work has been carried out within the research group *Nonlinear dynamics and time series analysis*[1] at the *Max-Planck-Institute for the Physics of Complex Systems* in Dresden, Germany. It was motivated by the interest of seeing in how far it would be possible to apply advanced methods from the theory of time series analysis and nonlinear dynamics, developed[2] and routinely used within the group, to a system definitely *complex*, but nevertheless outside the usual scope of a physicist's research.

But what arguments justify the classification of a financial market as complex system? In order to respond, we first need to discuss the involved terms.

1.1 Financial markets as complex systems

What is complexity?

Unfortunately, it must be admitted that it is beyond our ability to give a rigorous definition of complexity or complex system, because it actually does not exist. Typically these terms refer to systems that, although governed by relatively simple - often nonlinear - equations, exhibit a rich dynamical behavior on temporal and spacial scales that are not explicitly contained in its constituents or associated equations. Therin we find reflected the difference between complicated and complex: in a scientific context, a complex system is not necessarily complicated, because there would be nothing special about a complicated system showing somewhat complex behavior.

Let us consider as a typical and very simple example a system of coupled spins of the Ising model type: although the regularly spaced magnetic dipoles of such a system have only two degrees of freedom (up or down), and their interaction is limited to a field induced coupling with their neighbors, such a system can exhibit global phenomena like phase transitions between a macroscopically magnetic and non magnetic state.

One might think of the climate as a second example: in a first approximation one could formulate a description in terms of gas on a solid surface, heated from above, and therefore governed by the classical Navier-Stokes equation of fluid dynamics. Nevertheless, long time climate forecasts remain a major challenge and even outside the scientific community the so called "butterfly effect", i.e. the possibility that a perturbation as small as the one caused by a flying butterfly can possibly have an impact so large as to cause a tornado for example, has become a widely known symbol of the unpredictability of climate dynamics. One has been able to simulate these phenomena numerically by means of the Lorenz equations [33], which lead to the important concept of the so called *strange* attractor.

[1] `http://www.mpipks-dresden.mpg.de/mpi-doc/kantzgruppe.html`
[2] TISEAN-Software: `http://www.mpipks-dresden.mpg.de/~tisean`

Even though the two systems just chosen to illustrate the concept of complexity are typical representatives of deterministic processes, this property defines no prerequisite. For instance, let us consider a two dimensional classical random walk generated by the linear stochastic Wiener-process: seen in any space of dimension two or higher, the associated brownian path constitutes an object not of dimension one, but due to its extremely intricate structure at all length scales, of dimension two.[3]

In what sense are financial markets complex?

A financial market generally consists of the so called agents (=traders), furnished by varying amounts of capital, and the interaction rules (e.g. commercial laws) of the trading platform. Every single one of these agents conducts his activities with the aim of realizing the highest possible profit, which he tries to achieve by selling and buying financial assets of all types at different times. In proportion to such a simple microscopic setup, the macroscopic behavior of financial markets appears rather rich: the seemingly uncorrelated ups and downs of financial indices and the extreme event of a crash constitute typical phenomena of complex systems; in fact, the financial market dynamics has often been described in terms borrowed from turbulence, and the financial crash has been compared with a phase transition of a physical many-body system.

It is not difficult to identify indicators for nonlinearity in financial markets, such as speculative bubbles and also the extreme diffusion of panic in cases of larger losses; these effects represent examples of typical nonlinear processes called *autocatalytic* [3], which are characterized by conditions in which small stimuli can be strongly amplified by means of the internal dynamics of the system. In addition, this also implies the absence of any stable state of equilibrium: if the price of an asset rises, agents will generally tend to buy it, thereby giving it still further potential to rise; the same is true in case of falling prices, where again there is no force "pushing" back to a presumed equilibrium price.

As an objection against the concept of financial markets as complex systems one could claim that there is no such thing as *the* financial market, that instead there are various markets all over the world, with very different products and even distinct commercial laws, and that hence the suggested unified treatment cannot be justified. Although this is true in principle, there are known phenomena, like fat tails in the distributions of price variations and scaling invariances[4], that have been found in the most diverse markets, often with varying parameters characteristic of the particular market, but of the same general form. This strange *universality* and the existence of typical pattern-like structures can perhaps be traced back to the common speculative character of all such markets, by

[3] For a further discussion of complexity see, e.g., [3].

[4] These phenomena will be discussed later.

which we intend that unique interplay between stochastic external influences and the (deterministic?) human psychology, with its strong desire to make profits and its angst of loosing.

1.2 Econophysics: origin and current issues

The interest of the exact sciences in some aspects of economics dates back to the doctoral thesis of L. Bachelier [2], a student of the famous Henri Poincaré at the Paris Ècole Normale Supérieure. In his thesis, entitled “Thèorie de la speculation”, he analyzed for the first time a speculative market, in his case the Paris stock exchange, and proposed the classical random walk as a model for the price evolution. Despite the simpleness of that model, it proved to be very successful: in 1973, Black, Scholes, and Merton published their famous *Black-Scholes* equation [7, 41], aimed at determining the “correct” price of an option, for which they were awarded the nobel price in economics in 1997. It is interesting to note that their original professional background was that of a physicist, mathematician and chemical engineer.

In 1963 the french mathematician Benoit Mandelbrot published his milestone paper “The variation of certain speculative prices” [35] in which for the first time the assumption of normally (i.e. according to a Gauss distribution) distributed price variations was rejected; instead, Mandelbrot suggested that a so called Lévy distribution, which like the normal distribution satisfies the request of stability, but assigns a much higher probability to the extreme events (a property called “fat tails”), would represent a better model. Still today Lévy distributions are considered to be the best model for the central part of the distribution describing the price variations. Mandelbrot continued to intervene [36], and it is to a considerable extend his merit, that the proper attention has been brought to the importance of scaling invariances and power laws in price dynamics.

The late 80’s and 90’s not only witnessed an immense growth of the worldwide financial markets, but also led to a much increased automatization and, consequently, to the electronic registration of huge amounts of financial data - on some markets virtually every transaction is recorded nowadays. This more and more attracted the interest of statistical physicists, who viewed the financial markets as a well-suited laboratory for their methods of extracting information from data, or for the verification of models describing a large number of independent units with a nonlinear interaction. Along with that, physicists became increasingly aware of strong analogies between speculative markets and some well known physical phenomena, as for instance universality [46], spin systems [11], self-organized criticality [34, 50], complexity, [39] or turbulence [22], almost all of which can be associated with the statistical mechanics branch of physics. The resulting publication of several articles on prestigious journals like “Nature” [34, 37, 38, 22, 51] and others, together with the appearance of monographs written by physicists on the phenomenology of financial markets [39, 57, 10, 44] marked the establishment of a new branch of physics for which the term *econophysics* was coined.

Today, one can divide the research activities within the interdisciplinary field of econophysics roughly in two areas: the "microscopic" approach investigates the financial market dynamics from the point of view of the single agents, with the long-term target of being able to derive the complex "macroscopic" behavior of the financial markets from microscopic equations [34, 11]. To thoroughly analyze the statistical properties of that "macroscopic" behavior is exactly what constitutes the second branch of econophysics [5, 15, 23, 42, 56]. This last field of research naturally profits in a special way from the immense amount of electronically recorded financial data available.

Following the second approach mentioned above, the present work will begin by carrying out an investigation based on empirical data, first in chapter 2 by using only linear, then in chapter 3 also nonlinear tools. An important question that all the same will not be addressed here concerns the analytical form of the empirical distribution of the price variations [23]. Instead, we will concentrate on what usually is described [39, 32] as *correlations in financial time series*.

After this empirical survey we will discuss in chapter 4 microscopic mechanisms that have been proposed for the explanation of the encountered phenomena, illustrated by a practical simulation of an artificial stock market. Finally, in chapter 5, we will address the question of forecasting, and the interplay between determinism and stochasticity, both generally and referred to the observed situation in financial markets.

1.3 Presentation of empirical data

In what follows some data analysis of real financial data will be performed, based on a total of 13 empirical series, recorded simultaneously at a one-minute rate by Deutsche Bank Research in Frankfurt, Germany, between May and December 2000. A list of all the series together with the number of available data points[5] is reported in Table 1.

As can be noted, there are three different types of series present: stock indices, currency exchange rates, and interest rates. While stock indices and interest rates are usually reported in the form of one definite value, the foreign exchange rates consist of two data points for every minute, the highest "bid" and the lowest "ask". As do most authors, we also defined a working series by taking the arithmetic mean value for every minute.

Generally, invalid values due to transmission errors or computer failure were carefully filtered out, and periods without trading activity (weekends, nighttime, holidays) were

[5] After the described filtering.

Series	Description	# of data points
CAC 40	French stock exchange index	67129
DAX	German stock exchange index	84133
Dow Jones	US industrial stock index	44396
NASDAQ	US technology stock index	46075
S&P 500	Index of 500 major US-stocks	44541
CAD/USD	Exchange rate Canada $ / US $	79446
CHF/USD	Exchange rate Swiss F / US $	101230
GBP/USD	Exchange rate British S / US $	100269
USD/EUR	Exchange rate US $ / Euro	106216
DEM 10YT	German Mark 10 year treasury bond	74791
EUR 3M	3 months interest rate Euro	11008
EUR 10Y	10 year interest rate Euro	15585
USD 10YT	10 year treasury bond US $	21813

Tab. 1: List of analyzed financial series

excluded, reconnecting afterwards the remaining parts of the original time series. This procedure of defining a new time scale to be called *trading time* has the obvious drawback that records notedly separated in real time may become close neighbors in the newly defined *trading time* series, but the relatively small number of such "critical" points prevents a statistically significant impact. For concreteness, the overall run of two series after the filtering procedure is shown in Fig. 1.

2 Linear time series analysis

In this chapter, we will first introduce the relevant variables needed when investigating financial time series quantitatively. Afterwards the basic but still very important tool of the linear autocorrelation function will be briefly explained and applied. Additional insight into long-term linear autocorrelations will be gained by means of a scaling analysis. Linear cross-correlations will finaly be discussed in the last section.

2.1 Basic definitions

As is evident also from Fig. 1, the raw financial time series cannot reasonably be assumed as stationary, a property yet essential for the validity of the forthcoming analysis. The standard solution to this problem is to define some new variable, that can be considered sufficiently stationary, or at least asymptotically stationary [39]. The relevant variables

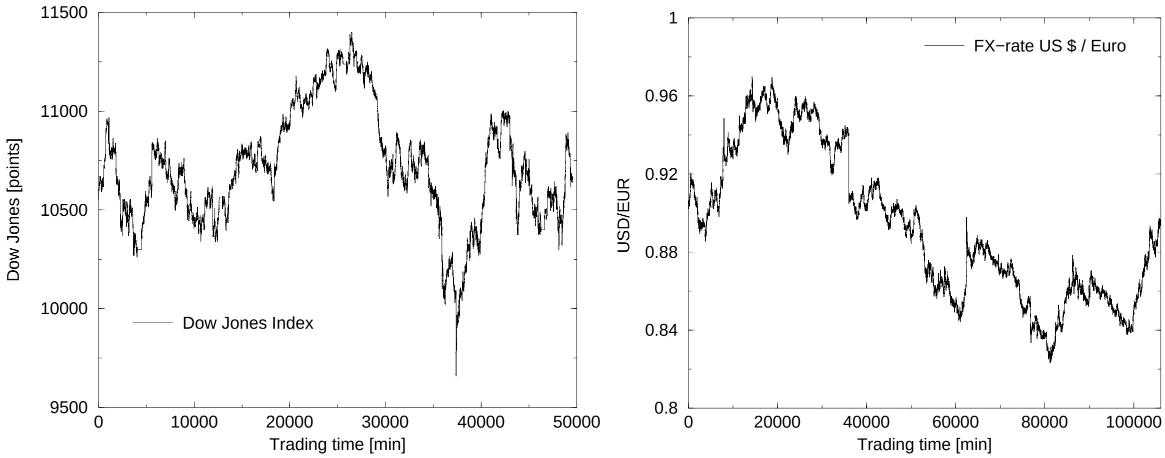

Fig. 1: Overall run of the Dow Jones Industrial Average index, and of the exchange rate US \$ / Euro, where the time axis is referring to the newly defined trading time.

chosen by most authors to describe a financial time series $x(t), t = 1 \dots N$ are:

price-change or increment

$$\delta x_\tau(t) := x(t+\tau) - x(t) \tag{1}$$

return

$$r_\tau(t) := \frac{x(t+\tau) - x(t)}{x(t)} = \frac{\delta x_\tau(t)}{x(t)} \tag{2}$$

log-return

$$s_\tau(t) := \ln\left[\frac{x(t+\tau)}{x(t)}\right] = \ln\left[x(t+\tau)\right] - \ln[x(t)]. \tag{3}$$

The choice of the variable does not affect the outcome of the present work; in fact, in the high-frequency regime they are approximately identical, or proportional to each other [39]. The usual quantity employed to characterize the fluctuations in financial data is the so called volatility, here[6] defined as

$$vol_\Delta(t) := \frac{1}{\Delta}\sum_{i=1}^{\Delta} |s_\tau(t+i)|, \tag{4}$$

where the parameter Δ refers to the chosen length of the time-window and τ (in our case always $\tau = 1$ min) denotes the basic time scale. To give some idea, the here considered time series have mean values of typically $\langle \hat{s}(t)\rangle_t \simeq \pm \cdot 10^{-6}$ for the log-returns, while the absolute log-returns, also interpretable as an estimate of the one-minute volatility, have mean values of the order of $\langle \hat{vol}_{1\ \mathrm{min}}(t)\rangle_t \simeq \cdot 10^{-4}$. However, as is generally known, the

[6] It can also be defined as mean square deviation.

degree of fluctuations in financial indices is subject to long-term correlated oscillations[7]. Still, in concordance with other authors [39], we assume a sufficiently long financial time series to be asymptotically stationary, i.e. leading to relevant results for the large time statistical properties of the analyzed data.

2.2 Linear autocorrelation

Let us briefly recall this standard instrument's basic notations. If we express the estimate of the first moment, i.e. the mean value with respect to time[8], of a stationary time series $x(t), t = 1 \dots N$, by

$$\hat{\langle x \rangle}_t := \frac{1}{N} \sum_{i=1}^{N} x(t), \tag{5}$$

and the square of the standard deviation, or variance, by

$$\hat{\sigma}^2 := \frac{1}{N-1} \sum_{i=1}^{N} (x(t) - \langle x \rangle)^2 = \frac{1}{N-1} \left(\sum_{i=1}^{N} x(t)^2 - N \langle x \rangle^2 \right), \tag{6}$$

then the autocorrelation function is estimated with

$$C_{xx}(\nu) = \frac{1}{\sigma^2} \langle (x(t) - \langle x \rangle)(x(t-\nu) - \langle x \rangle) \rangle = \frac{\langle x(t)x(t-\nu) \rangle - \langle x \rangle^2}{\sigma^2}, \tag{7}$$

where ν represents the time lag.

For a stationary series, $C_{xx}(\nu)$ takes on values between +1 and -1, expressing thereby the linear dependency between the series $x(t)$ and its by ν time-units shifted copy $x(t-\nu)$. We can interpret the value $C_{xx}(\nu)$ as the cosine of the angle formed by the properly rescaled vectors $x(t)$ and $x(t-\nu)$: zero represents orthogonality and thus linear independence, +1 or -1 corresponds to parallel or antiparallel configurations, and therefore complete dependence. Trivially follows $C_{xx}(0) = 1$ and $C_{xx}(\nu) = C_{xx}(-\nu)$. We want to stress that $C_{xx} = 0$ does imply linearly uncorrelated data points, but not the absence of any other statistical dependency.

Differently from the mean value and the variance, the autocorrelation function already contains information about the temporal evolution of a system. Let us consider the autocorrelation function of some standard processes: a periodic process will be characterized by a periodic autocorrelation, a chaotic process by an exponential decay of C_{xx}, and stochastic processes show either an exponentially or a power-law decay of the autocorrelation function for growing ν, where the latter implies the presence of long-range correlations. It follows that in general it is not possible to distinguish between deterministic chaos and stochastic dynamics by means of the autocorrelation function.

[7] Known as correlated volatility.

[8] The index t will be omitted when superfluous.

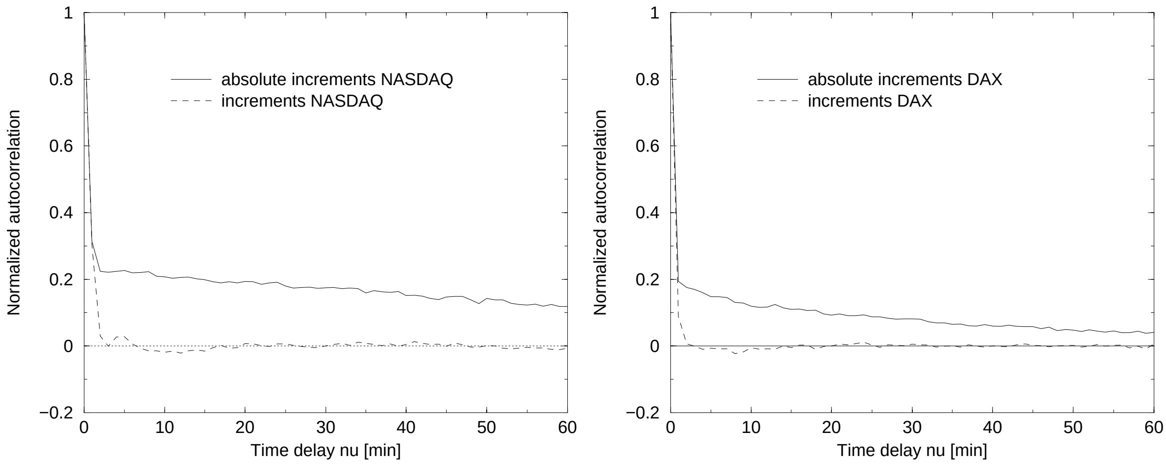

Fig. 2: Short-range autocorrelation for increments and absolute increments of the NASDAQ and DAX stock exchange index.

It could seem curious to speak of a linear autocorrelation function, when it actually, as can be seen from equation 7, contains a term of second order. The deeper reason is that any linear process, deterministic (trivially) or stochastic, can be completely characterized by its mean, variance, and the linear autocorrelation function. This includes e.g. the autoregressive processes of order m, AR(m), or its extended version with a moving average, ARMA(m,n), which play an important role in the simulation of price dynamics in finance. To see why this is true, one has to consider the associated power spectrum, which is uniquely determined by the process's parameters. The power spectrum, on the other hand, is ensured to be equivalent to the Fourier transform of the linear autocorrelation function by the theorem of Wiener-Khinchin [27].

In the following, some typical results from the empirical analysis will be reported - briefly, since this kind of approach is quite standard, and its outcome is widely known. In Fig. 2 we show for the NASDAQ and DAX the autocorrelation of the increments and also of the absolute increments, which, as was said before, can be interpreted as the one-minute volatility. For the increments we note an autocorrelation function falling to zero[9] within two or three minutes, as would be expected. The weak positive correlation in the first few minutes is too short to be exploited commercially, since possible profits would be consumed by the transaction fees. The case is different for the absolute increments, where we find a significant positive correlation between 0.2 and 0.1 for times of up to at least one hour. This known phenomenon has been termed *correlated volatility*, and cannot be used for making riskless profits either, and hence the findings are consistent with the efficient market hypothesis.

[9] Apart from very small statistical fluctuations without significance.

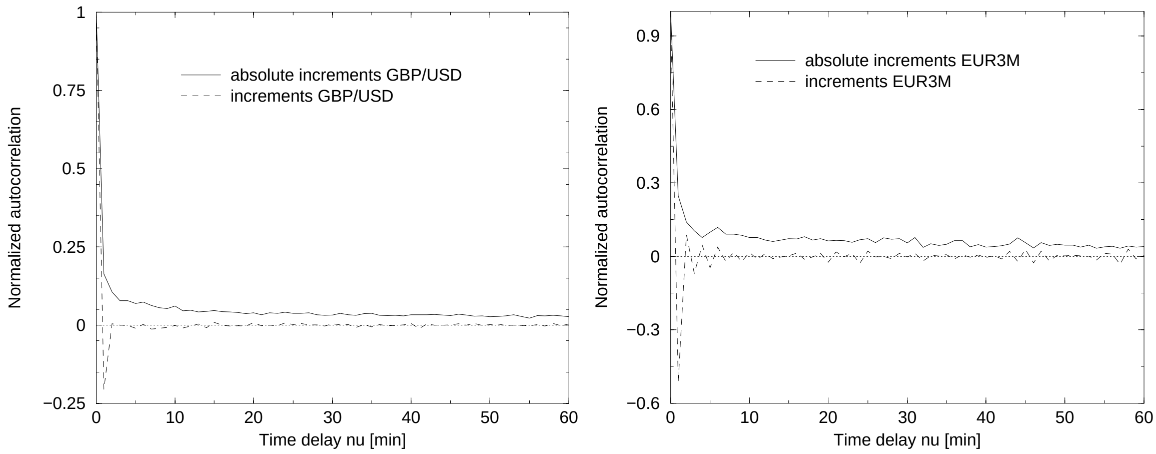

Fig. 3: Short-range autocorrelation for increments and absolute increments of the FX-rate GBP/US$, and the three months Euro interest rate.

The picture is similar for interest and foreign exchange rates. In Fig. 3 we can again observe the lasting, also called persistent, autocorrelation of the one-minute volatility, represented by the absolute increments, and an autocorrelation function that rapidly approaches zero for the raw increments. The only notable difference in comparison with the stock exchange indices consists in a strong anticorrelation at time lag one. This already noted [43] behavior is possibly induced by sudden price-jumps up or down, that after only one minute repeat themselves, thereby returning circa to the starting point. As numerical simulations have shown, for a dataset of 50000 points it is sufficient to insert ten such jumps with a relative change of 10% in order to reproduce a similar anticorrelation at lag one. The visible small periodic oscillations during the first ten minutes, observable for the interest rates, can be explained by noting that the interest rates often oscillate for minutes between two values separated by only one basis point.

As was shown, the fast decay of the autocorrelation function of the increments is compatible with the expected exponential decay of a stochastic process without temporal correlations, whereas the absolute increments show a decay possibly following a power law. In fact, some authors have confirmed the presence of a power law with characteristic exponents between 0.3 and 0.6 [25, 47, 1, 31, 12].

Normally, the time scale considered for the autocorrelation analysis of the absolute increments is limited to one or several hours, but what about its behavior on even longer time horizons? In Fig. 4 (left side) we show the long-time autocorrelation for the absolute and also raw increments, for a time lag of up to 10000 minutes trading time. While the increments itself produce only noise fluctuating around zero, the absolute increments show a surprisingly rich structure. We can confirm the existence of a truly long-range

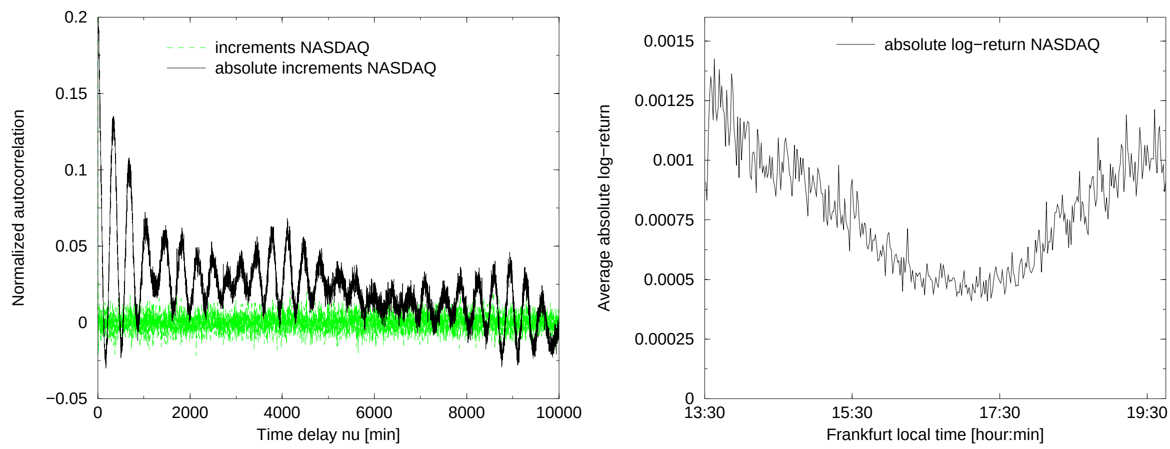

Fig. 4: Long-range autocorrelation for increments and absolute increments of the NASDAQ on the left, and the typical behavior of the NASDAQ one-minute volatility during one trading day on the right hand side.

correlated volatility, characterized by strong and quite regular modulations, corresponding to the daily volatility cycle. In fact, the distance measured in trading time between two neighboring peaks in the autocorrelation function corresponds to the length of one trading day, i.e. about six hours.

To illustrate better the origin of the observed pattern we report in Fig. 4 (right side) the typical behavior of the volatility during one trading day. The curve was generated by calculating for every minute of the trading day the mean value - with respect to all days available in the dataset - of the absolute log-return[10]. One notes a distinct pattern, with a high degree of fluctuation during the opening and closing phase, and a reduced volatility around the central hours of the trading day. This basic behavior can be reidentified 28 times on the left hand side of Fig. 4. However, we need to point out that this clearly visible influence of the daily seasonality does not explain by itself the phenomenon of long-range correlated volatility, as can already be understood by observing that the daily oscillations in Fig. 4 (left side) are not centered around zero. Numerical tests based on rescaled data, freed from the daily seasonality, have shown that the long-range autocorrelation between absolute increments persists.

2.3 Scaling laws

On graphs displaying the temporal evolution of stock prices one always needs to indicate the temporal horizon they are referring to; if not, it could turn out to be impossible to recognize whether what we see shows last week's or last year's price evolution. A graph

[10] Here we used the absolute log-returns because of their more intuitive meaning of percental change.

with such a property is called *self affine*, and the underlying process is said to be *scale invariant* [27]. With regard to financial markets, scale invariance means that the price evolution process can be described in terms of minutely, hourly, or daily recorded data, but the principle property of the process, i.e. the distribution of the price variations, will always be of the same general form, with only a scalar parameter that needs to be adjusted for a change of the time scale. The classical example for a scale invariant process is the random walk: however one choses the basic step-length of the random walker, the distribution describing the probability of finding the walker at a point x_0 after time t_0 will always be of the gaussian form. Formally expressed, this means that a scale invariant process satisfies an equation of the type

$$\langle \Delta x^2 \rangle \propto \Delta t, \tag{8}$$

where the constant of proportionality is called *diffusion constant*. In terms of the random walk, this equation expresses the fact, that the random walker in a time interval Δt is displaced by a quantity that is on the average proportional to the square root of Δt. There are, however, processes that require a generalization of equation 8, which is achieved by writing

$$\langle \Delta x^2 \rangle \propto \Delta t^{2\mathcal{H}}, \tag{9}$$

where $\mathcal{H}$ represents the so called *Hurst exponent*. Obviously, for $\mathcal{H} = \frac{1}{2}$ we recover the random walk and equation 8, which is classified as the so called standard diffusion. Standard diffusion is characterized by linearly independent increments and a finite variance. In order to obtain Hurst exponents different from one half, we have to renounce at at least one of these properties. Lévy flights are processes with an infinite variance, and are characterized by exponents of $\frac{1}{2} \leq \mathcal{H} \leq 1$. They are similar to random walks, except that they are not generated by a Gauss, but by a Lévy distribution, which has tails decaying with a power law. On the other hand, if we want to keep the variance finite, we have to admit long-range correlated increments, that hence do not satisfy the Markov property. Processes of that type are called *fractional brownian motion*, and are capable of reproducing Hurst exponents of $0 < \mathcal{H} < 1$. The case of $\frac{1}{2} < \mathcal{H} < 1$ corresponds to positively correlated increments, while anticorrelation is found for $0 < \mathcal{H} < \frac{1}{2}$.

How can one estimate the Hurst exponent $\mathcal{H}$ from financial time series? A method recently introduced [45, 53], and particularly suited for data possibly nonstationary due to trends, is the *detrended fluctuation analysis* (DFA). It is implemented by dividing a time series $y(n)$, $n = 1 \dots N$, into N/t non-overlapping sub-sequences of length t. In each of them we eliminate the local linear trend by subtracting a least square fit $z(n) = an + b$

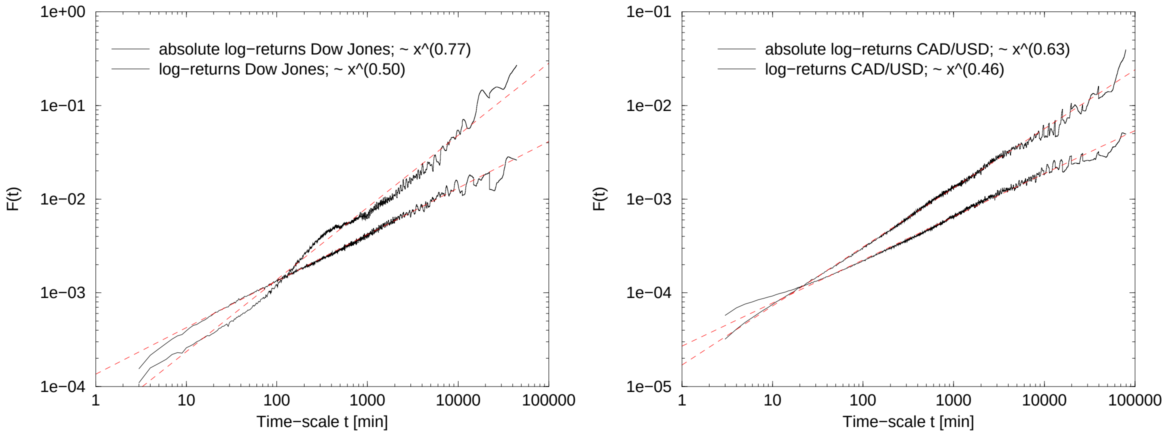

Fig. 5: Double logarithmic representation of the results of the detrended fluctuation analysis (DFA), for log-returns and absolute log-returns of the Dow Jones (left), and exchange rate CAD/USD (right). Shown are also best-fits by a power law.

from the data. Then the mean variance of the new series $y(n) - z(n)$ is calculated by averaging over all N/t subsequences. Expressed formally:

$$F(t)^2 = \frac{t}{N} \sum_{k=0}^{\frac{N}{t}-1} \frac{1}{t} \sum_{n=kt}^{(k+1)t-1} (y(n) - z(n))^2 \, . \tag{10}$$

For a scale invariant process one expects to find $F(t) \propto t^{\mathcal{H}}$, with a constant exponent $\mathcal{H}$, which can be retrieved by graphing $F(t)$ in a double logarithmic plot and measuring the slope of the obtained curve - if it is straight.

In Fig. 5 some typical empirical results are reported. Generally it can be said that while the analyses using log-returns show scaling with a Hurst exponent compatible with $\mathcal{H} = \frac{1}{2}$, this is different for the absolute log-returns, where we observe exponents significantly higher than one half, and hence, after double checking against the hypothesis of a Lévy flight by shuffling the data, persistence is implied. This can clearly be seen in Fig. 5 on the right hand side for the FX-rate CAD/USD. The Dow Jones, on the left side of the same figure, shows the same general behavior, but for time scales between 100 and 1000 minutes one finds an irregularity. The reason for that deviation from the straight line can be found again in the daily seasonality of the volatility: for time scales that roughly coincide with the length of one trading day, a higher correlation is found, and thus the slope of the DFA curve is higher also. This argument is supported by observing the results shown in Fig. 6 (right hand side) of another DFA of the Dow Jones, but this time with absolute log-returns that have been freed from the influence of the daily cycle. As a consequence, the irregularity has disappeared, and the curve has become straighter, but along with that also all other temporal correlations on short time scales ($t < 100$ min) have

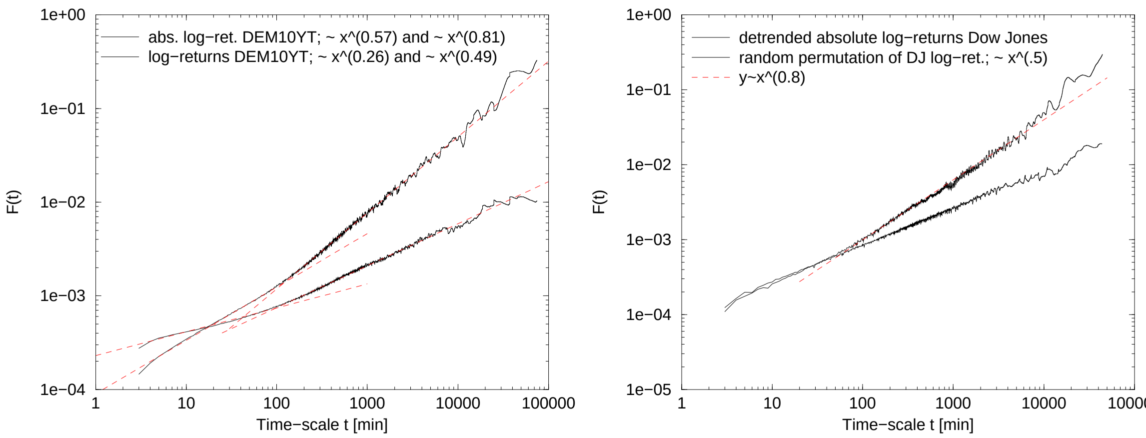

Fig. 6: Double logarithmic representation of the results of the DFA for log-returns and absolute log-returns of the DEM 10 year interest rate (left), and for the detrended absolute log-returns of the Dow Jones (right). Shown are also best-fits by a power law.

been canceled out, leading there to an exponent of one half. A slightly different picture is found for the interest rates, as can be seen in Fig. 6, left hand side. There are two different temporal regimes, less and more than 100 minutes, which show both scaling, but with a different Hurst exponents. For $t < 100$ min we find anticorrelated log-returns, and roughly uncorrelated absolute log-returns, whereas for $t > 100$ min the log-returns become uncorrelated, and the absolute log-returns show the same persistency that has also been observed for currency exchange rates and stock indices. An explanation could be, that the interest rates are actually oversampled when observed in the short-time scale regime, where they basically fluctuate around a constant value. Only for time scales longer than 100 minutes we recover the same behavior we found before for the other financial time series. In Table 2 we report measured Hurst exponents $\mathcal{H}$ for all the available financial time series.

To summarize this section, when analyzing log-returns we found most financial times series compatible with the random walk hypothesis, i.e. scaling invariance on almost four time decades with a Hurst exponent close to one half, implying uncorrelated returns. These findings are supported by the findings of section 2.2. For the absolute values of the log-returns we also found scaling invariance, but with Hurst exponents between 0.7 and 0.85. Since this effect was not seen anymore after shuffling the data[11], an enhanced diffusion generated by a Lévy flight could be excluded, and the existence of long-term correlations between the absolute log-returns can be confirmed[12]. For the interest rate

[11] Not explicitly shown here.
[12] For Hurst exponents obtained by other authors, see, e.g. [56].

Series	$\mathcal{H}$ for log-returns	$\mathcal{H}$ for absolute log-returns
CAC 40	0.495±0.01	0.73±0.025
DAX	0.48±0.01	0.77±0.02
Dow Jones	0.497±0.008	0.77±0.02
NASDAQ	0.495±0.01	0.86±0.02
S&P 500	0.497±0.008	0.83±0.02
CAD/USD	0.46±0.01	0.63±0.015
CHF/USD	0.49±0.02	0.70±0.02
GBP/USD	0.49±0.02	0.70±0.015
USD/EUR	0.47±0.01	0.70±0.02
DEM 10YT	0.255±0.01($t < 100$)	0.57±0.01($t < 100$)
	0.448±0.015($t > 100$)	0.81±0.015($t > 100$)
EUR 3M	0.12±0.01($t < 100$)	0.62±0.01($t < 100$)
	0.32±0.025($t > 100$)	0.77±0.025($t > 100$)
EUR 10Y	0.16±0.01($t < 90$)	0.47±0.01($t > 90$)
	0.42±0.02($t > 90$)	0.65±0.025($t > 90$)
USD 10YT	0.48±0.01	0.62±0.02

Tab. 2: Hurst exponent for various financial time series, result of the DFA and best fits according to a power law.

time series the picture was slightly modified on time scales below 100 minutes. Uncorrelated absolute log-returns and long-range anticorrelated returns were observed on these time scales, which was interpreted as being an artefact of the - for this particular financial market - inappropriately high sample rate of one minute.

The performed analysis for scaling invariances might seem somewhat academic at a first glance, but potentially it has a considerable practical value: if one is interested in the exact distribution of price variations on a daily time-scale, it is possible to first reconstruct a distribution on the minute time-scale, using the huge amount of data available on that time-scale, and then rescale the distribution to the daily time-scale. Distributions of price variations find important applications e.g. in option pricing.

2.4 Linear cross-correlation

The generalization of the autocorrelation from equation 7 to the linear cross-correlation function $C_{xy}(\nu)$ is straightforward and therefore needs not to be stated explicitly. The obtained values give information about the linear dependency between two distinct series,

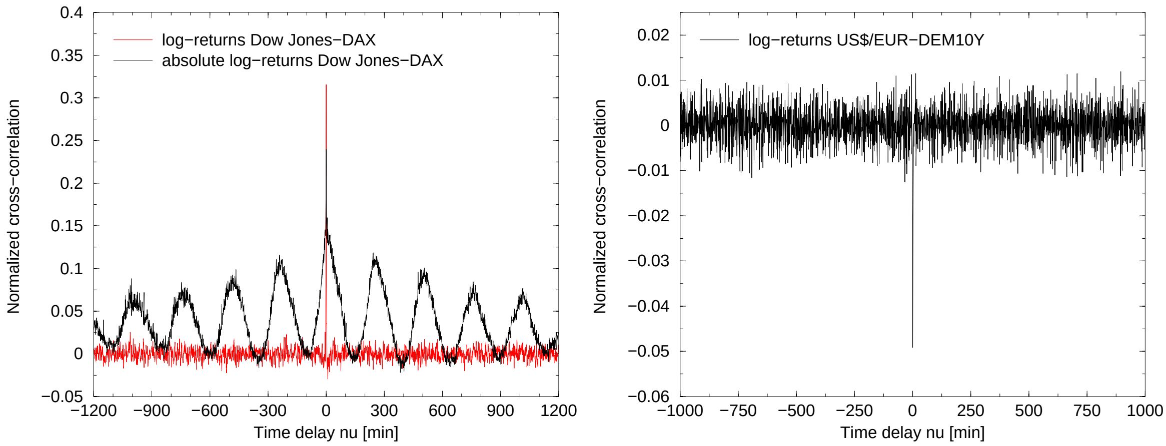

Fig. 7: Linear cross-correlation between log-returns and absolute log-returns of the Dow Jones and DAX stock index (left), and FX-rate USD/EUR and DEM 10 year interest rate (right).

$x(t)$ and $y(t)$; naturally symmetry is lost and therefore $C_{xy}(\nu) \neq C_{xy}(-\nu)$ and $C_{xy}(0) \neq 1$ in general.

In order to perform the data analysis, we first have to transform any pair of series to be taken in exam into one new synchronized series, that contains pairs of two values corresponding to the same minute. Since the various markets in general differ in their opening and closure time, this leads to a partial loss of data, which possibly worsens the statistical accuracy of the calculations presented in the following. On the other hand, such a simultaneous analysis can potentially give interesting insights into the coupling between these different markets.

Two representative graphical results are shown in Fig. 7. On the left hand side we can note a very regular pattern in the cross-correlation between the absolute returns of Dow Jones and DAX, while their returns result uncorrelated, except at time lag zero, where we find a relatively high cross-correlation of slightly more than 0.3. On the right hand side of the same figure we observe that the returns of the exchange rate US$/Euro and the DEM ten year interest rate are anticorrelated at time lag 1, while there was no significant result for the absolute log-returns (not shown). Further results for the "instantaneous" coupling of the series, in form of cross-correlations in the vicinity of time lag zero, are reported in Table 3. Notably high values are found for the cross-correlation between stock indices; series belonging to different types of financial markets show no or very little instantaneous cross-correlation.

Maybe the most interesting result presented in this section is the regular pattern found in the cross-correlation between the absolute log-returns of Dow Jones and DAX, as seen

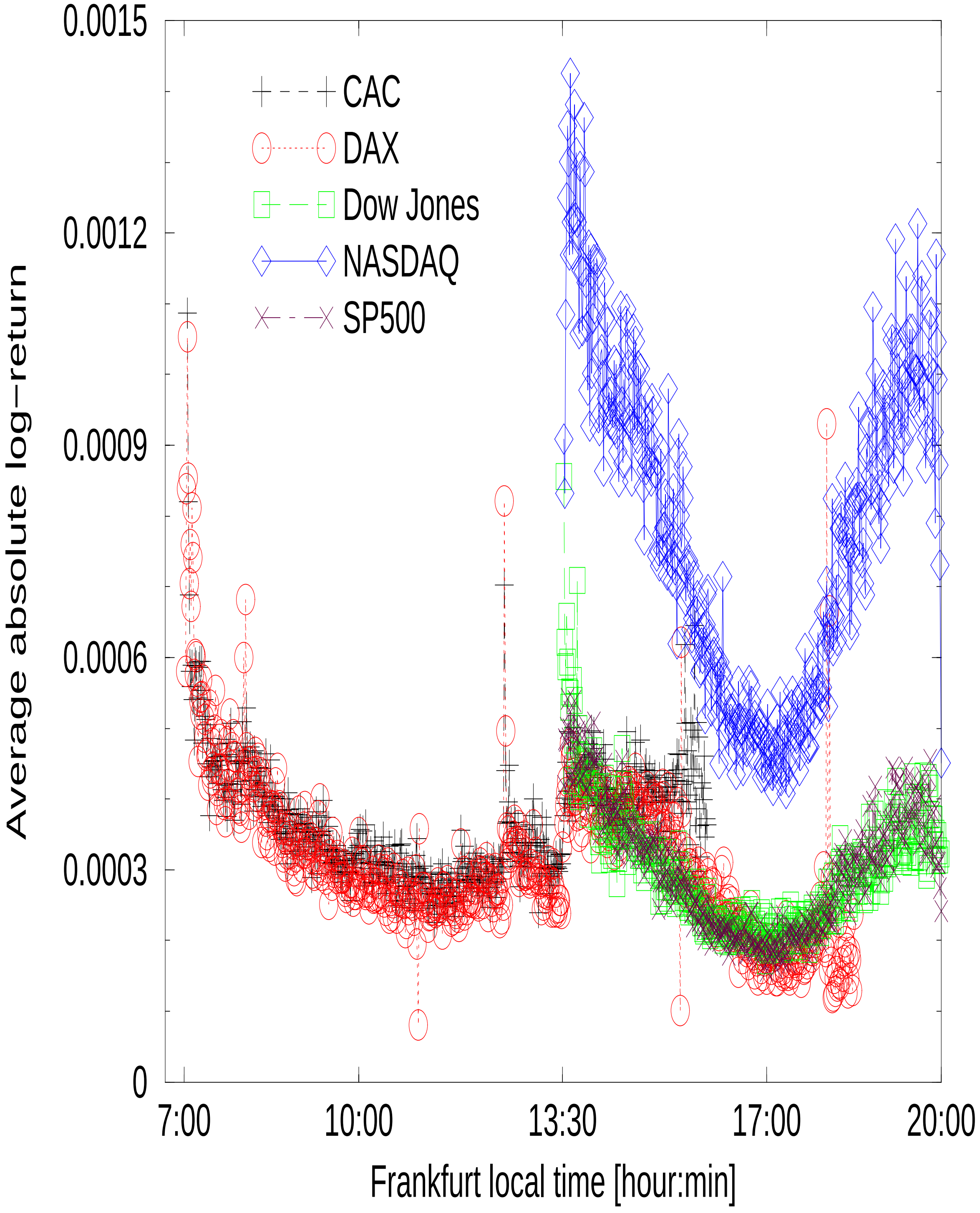

Fig. 8: Average values for the one-minute volatility during one trading day for various stock indices. Until 13:30 only the french CAC and the german DAX are traded, then also the US-american stock exchanges join in.

Series	Absolute maximum of cross-correlation C_{xy}
DAX - CAC40	0.27 for $\tau = 0$
DAX - US$/EUR	-0.018 for $\tau = 0$
DOW - CAC40	0.28 for $\tau = 0$
DOW - DAX	0.32 for $\tau = 0$
USD10YT - CA$/US$	no peak observable
USD10YT - CHF/US$	no peak observable
USD10YT - DEM10Y/US$	0.07 for $\tau = 1$
USD10YT - GBP/US$	no peak observable
USD10YT - NASDAQ/US$	0.12 for $\tau = 0$
US$/EUR - DEM10Y	-0.05 for $\tau = 1$
US$/EUR - EUR10Y	-0.03 for $\tau = 2$
US$/EUR - EUR3M	no peak observable
US$/EUR - USD10YT	0.03 for $\tau = 0$

Tab. 3: List of highest absolute linear cross-correlations between the log-returns of some series.

in Fig. 7. This type of curve, reminiscent of the periodically oscillating autocorrelation functions seen in section 2.2, was also confirmed for other pairs of stock exchange series (not shown). As we could explain the periodic modulation in the autocorrelation function by looking at the typical daily cycle of the series' volatility, we will now try to understand the periodicity of the cross-correlation by looking at how the various series' cycles of volatility relate to each other. This is reported for the five considered stock indices in Fig. 8. The similarity in the evolution of the curves corresponding to the different stock indices is quite striking, and demonstrates how strongly the world financial markets are coupled; it seems that indeed one can speak of *the* financial market. Except for the NASDAQ, which shows the same general form, but on a higher basis level, all indices follow show the same evolution. Until 13:30 only the french CAC and the german DAX, which are almost indistinguishable, are traded, then the volatility rises when the US-american stock exchanges are opened. Some time afterwards the french index stops to be recorded due to closure, but the german DAX still follows exactly the behavior of the US-american indices. A broader discussion about the role of seasonality in financial markets can be found, e.g., in [14].

3 Nonlinear time series analysis

We will now turn to methods that allow to quantify statistical dependencies in a more general way than the linear instruments presented in the last chapter. These methods were inspired by works of Shannon [49] and Kolmogorov [28] on the theory of information, and belong to a field called *symbolic dynamics*. The great advantage of the formalism to be described in the following lies in its model-free approach, i.e. it makes no assumption about the underlying dynamics of the considered system, except for stationarity. On the other hand, a certain disadvantage is represented by the necessity to encode continuous or unproportionally high resolved data by a discrete set of symbols.

3.1 Detecting redundancies with entropy

We will begin by recalling the elementary notions. Let us consider a discrete and stationary signal $I(t)$, with $p(i)$ being the probability[13] to observe symbol $i, i \in \{1, 2, \ldots, S\}$, and S denoting the number of symbols in the alphabet. According to Shannon, the average number of bits needed to optimally encode the signal I is given by

$$H_I := -\sum_{i=1}^{S} p(i)\log_2 p(i), \qquad 0 \le H_I \le \log_2 S, \tag{11}$$

called Shannon entropy. It expresses the average amount of information contained in every realization of a variable that is drawn according to the probability distribution $p(i)$, and becomes maximal in the case of equalprobability $p(i) = \frac{1}{S}$. By writing $p(i_1, i_2, \ldots, i_m)$ for the probability of observing the subsequence $(i_1, i_2, \ldots, i_m)$, one can generalize the Shannon entropy and define the block-entropy of order m:

$$H_I(m) := -\sum_{i_1,i_2\ldots,i_m=1}^{S} p(i_1, i_2, \ldots, i_m)\log_2 p(i_1, i_2, \ldots, i_m). \tag{12}$$

The differences of block-entropies of neighboring order constitute the conditional entropies:

$$h_I(m) := H_I(m+1) - H_I(m), \qquad 0 \le h_I(m) \le H_I. \tag{13}$$

$h_I(m)$ expresses the average amount of information (in bits) still transmitted by the latest observation $I(m+1)$, when the last m observations of I are known and their information has been completely exploited; or, equivalently, the missing information for a correct forecast of $I(m+1)$ with the help of the m preceding historical observations. By using equation (12) and some elementary algebra, one can rewrite equation (13) as

$$h_I(m) = -\sum p(i_1, i_2, \ldots, i_m, i_{m+1})\log_2 p(i_{m+1}|i_1, i_2, \ldots, i_m), \tag{14}$$

[13] Time independent, since we assumed stationarity.

namely as Shannon entropy of the conditional probabilities, here denoting by

$$p(i_{m+1}|i_1, i_2, \ldots, i_m) = p(i_1, i_2, \ldots, i_m, i_{m+1})/p(i_1, i_2, \ldots, i_m) \tag{15}$$

the probability to observe symbol (i_{m+1}) immediately after the sequence $(i_1, i_2, \ldots, i_m)$. This also explains the name *conditional* entropy. From how $h_I(m)$ behaves for different values of m, one can draw conclusions about the deterministic or stochastic character of the underlying process. If, in the first case, $h_I(m)$ remains constant at its maximum value H_I for all m, it means that the analyzed time series is completely random, and that no information about future values can be gained from observing the past. If, in a second case, the values first decrease but then, from some value $m > M$ on, remain constant and non-zero, we can describe the corresponding process as markovian of order M, meaning that there is exploitable memory in the M past observations. If, in the last case, $h_I(m)$ drops to zero after some $m > M$, the observed process is periodic, and hence completely deterministic, with period M. In other words, any time we find $h_I(m) < H_I$ systematically, we can confirm the existence of temporal correlations, or redundancies, in the analyzed time series, and hence the knowledge of past values can potentially contribute to the prediction of future values.

In practice, however, the estimation of Shannon entropies is complicated by the finite size of any data set, see, e.g., [24, 26]. Especially when S, the number of the employed symbols and m, the considered block lengths, are relatively high, the conditional entropy $h_I(m)$ tends to be systematically underestimated. One thus has to evaluate carefully whether any observed fall off in the conditional entropy really corresponds to a statistical dependency in the time series, or is just an artefact of the finite sample size. This is achieved by the use of shuffled datasets, by which one can benchmark the entropy estimation. Since in the shuffled data all possible temporal correlations have been destroyed, any observed fall off for $h_I(m)$ can be traced back to the finite sample effect.

The first step in the practical analysis of real data with tools based on symbolic dynamics like conditional entropies is to discretize the data by some coarse graining. Although the financial data is actually already in a discrete form, its resolution is by far too high with respect to the amount of records available. For more robust statistics and especially in the case of multifractal phenomena it is often recommendable to work with coverings and use generalized Renyi entropies instead of partitions and Shannon entropies [27].

In the present case, however, a straightforward implementation defining a partition with marginal equalprobability for every symbol will lead to sensible results. Such a partition is generated by dividing the range of the given dataset into S (size of the alphabet) disjoint intervals, such that the number of data points in every interval is constant and

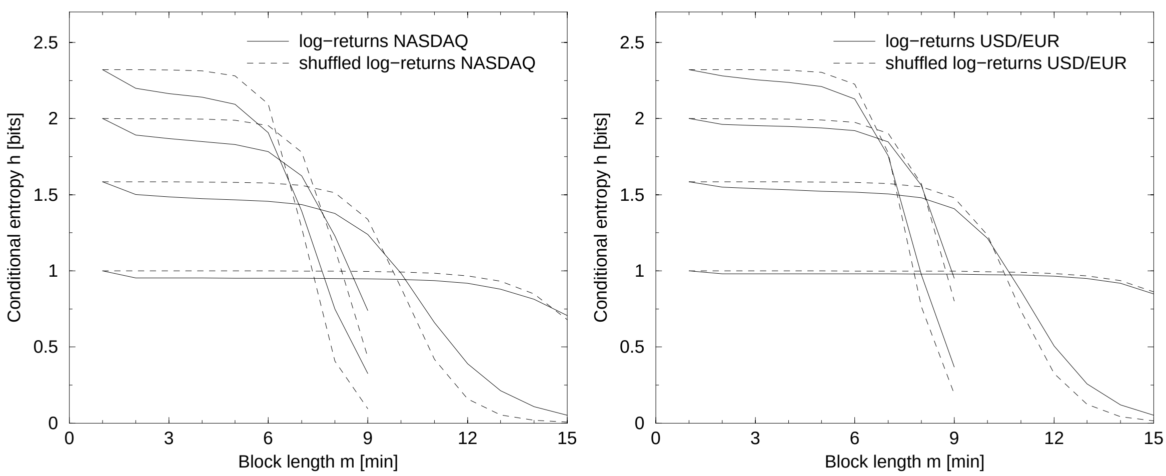

Fig. 9: Conditional entropy for the NASDAQ (left), and for the USD/EUR exchange rate (right). Calculations were done for four different partitions of S=2,3,4,5 symbols (bottom to top). Shown also the curves resulting from the same calculation done with shuffled data sets.

therefore $p(i) = 1/S$, and consequently $H_I = -\sum_{i=1}^{S} p(i)\log_2 p(i) = -S\frac{1}{S}\log_2\frac{1}{S} = \log_2 S$ automatically holds for every empirical time series I, where every data point has now been replaced uniquely by the label of its proper interval. Apart from its simpleness, this approach has the advantage of neutralizing undesirable effects due to very inhomogeneous histograms, and it also ignores the trivial information gain obtained by just observing marginal distributions. Furthermore, for data with an approximately symmetric distribution, the concrete meaning of partitions consisting of few symbols is quite intuitive: two symbols (S=2) only take the sign of the increments into account, three correspond to the three possible moves (i) larger gain, (ii) roughly neutral, (iii) larger loss etc. Numerical outcomes for entropy related quantities will of course depend on the specific partition chosen; however, by varying the partitions one tries to find approximately invariant results.

In Fig. 9 we report for four different partitions (S=2,3,4,5) the conditional entropy $h_I(m)$ for two series, the NASDAQ and the FX-rate US$/Euro. The results found when shuffling the series prior to the calculation are also shown as benchmark. As was said before, for longer block lengths all curves drop to zero due to finite sample effects, and the more symbols are used for the encoding, the faster will this effects be seen. Nevertheless, redundancies in the time series clearly show up for short block lengths, since the corresponding empirical curves are located below the ones corresponding to the shuffled data. However, it is difficult to interpret the results seen in Fig. 9, and therefore we will introduce other quantities with a clearer meaning. Intuitively, the difference in $h_I(m)$ between the uncorrelated shuffled and the empirical series corresponds to the amount of detected redundancy. We therefore define *effective redundancy* (ER) as the difference of the conditional entropy calculated for the shuffled series and the usual conditional entropy, calculated for the empirical series:

$$R_I^*(m) := h_{I_{\text{shuffled}}}(m) - h_I(m). \tag{16}$$

$R_I^*(m)$ expresses the quantity of information about future observations that one can extract from the last m historical observations of I. In order to get a better idea of how much the identified quantity of redundant information can explain, we put it in relation to the amount of information contained in an observation of I when ignoring past observations, which is nothing but the Shannon entropy from equation (11). *Relative explanation* (RE) is therefore defined as:

$$RE_I(m) := \frac{R_I^*(m)}{H_I}, \tag{17}$$

which in our case of partitions with marginal equalprobability $p(i) = \frac{1}{S}$ simplifies to

$$RE_I(m) = \frac{R_I^*(m)}{\log_2 S}. \tag{18}$$

With $RE_I(m)$ we now dispose of a quantity with a very intuitive meaning: what percentage of the information contained in a future observation of I can be explained by the last m historical observations? In Fig. 10 we report effective redundancy and relative explanation for the NASDAQ, and relative explanation only also for the exchange rate US$/Euro and the 10 year US treasury bond interest rate.

The quantitative results can now be easily interpreted: the redundancy detected in the NASDAQ amounts to almost 5% explanatory power in case of a binary encoding, but to about 8% when partitioning the data with three, four or five symbols. This clearly indicates a combination of linear and nonlinear correlations, of which the bivariate partition can only "feel" the linear part. Also, the linear memory of the process extends to only one past observation, while the nonlinear memory extends to time horizons of at least eight minutes; in fact, the amount of data available does not permit to estimate the temporal extension of the nonlinear memory. For the government bond interest rate USD10YT and the FX-rate USD/EUR, shown in the lower part of Fig. 10, the same conclusions can be drawn: while the results for the bivariate partition support a Markov process hypothesis, the higher order partitions possibly imply a long-range correlation. Interestingly, the partition employing S=4 symbols leads to lower relative explanation values than the three or five symbol partitions. This implies that the symmetry induced by uneven partitions is better suited to represent the temporal correlations; the long rang-correlated volatility discussed earlier fits in well as possible explanation for these observed phenomena. To give an overview of results for the other financial series, we report in

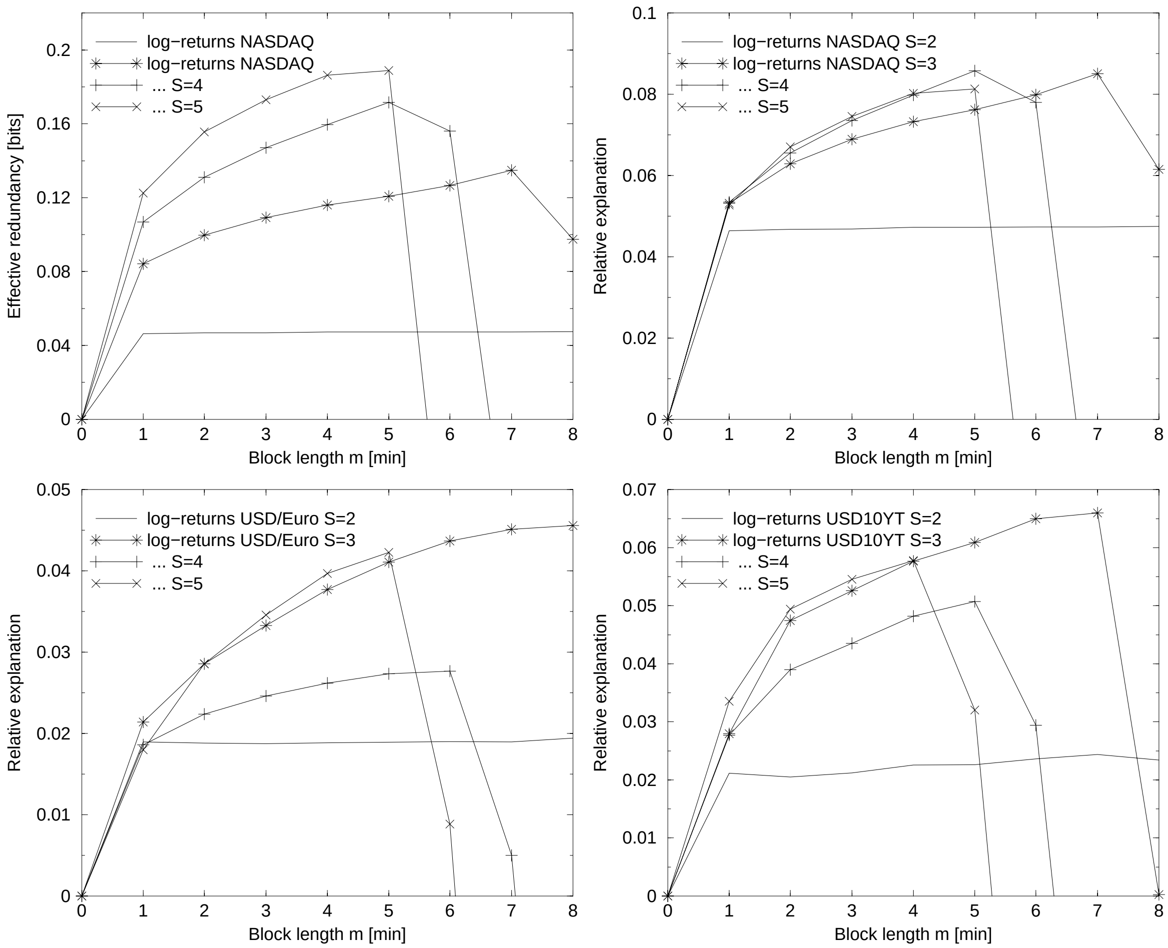

Fig. 10: Above: for NASDAQ log-returns, effective redundancy (left) and relative explanation (right). Below: relative explanation for USD/Euro FX-rate (left), and the interest rates on the 10 year treasury bond US$ (right).

Table 4 the relative explanation potentially contained in the last four past observations of the various series.

Of course the values found for the Euro interest rates need to be commented. What we see in Table 4 is again an artefact of oversampling the series, since, as we said before in section 2.2, these interest rates tend to just oscillate between two values separated by one basis point. In fact, we have seen this in Fig. 3 in form of a strong anticorrelation at time lag $\nu = 1$. The otherwise most interesting results are certainly the high values obtained for the NASDAQ and SP500 (of course they are somewhat similar, since stocks of the NASDAQ are contained in the SP500), but also for the heavily traded exchange rate US$/Euro. In how far these results can be used for prediction remains an important question to investigate. Another interesting question is whether similar values could be obtained on other, commercially perhaps more relevant time scales. Similar approaches

Series	RE [%] $_{S=2}$	RE [%] $_{S=3}$	RE [%] $_{S=4}$	RE [%] $_{S=5}$
CAC	0.15	0.46	0.71	0.77
DAX	0.14	1.2	1.5	1.7
DJ	0.85	1.5	2.0	1.9
NASDAQ	4.7	7.3	8.0	8.0
S&P500	3.6	4.7	5.6	5.8
CAD/USD	3.4	4.2	2.3	6.8
CHF/USD	2.0	2.7	2.4	2.6
GBP/USD	2.1	2.2	2.7	3.3
USD/EUR	1.9	3.8	2.6	4.0
DEM10YT	4.8	6.4	5.7	6.4
EUR3M	10	13	12	13
EUR10Y	20	45	46	51
USD10Y	2.3	5.8	4.8	5.8

Tab. 4: Relative explanation: how much (here in percent) of $I(t+1)$ is explained by the information contained in $I(t), I(t-1), I(t-2)$ and $I(t-3)$?

in literature can be found, e.g., in [5, 42, 55].

3.2 Transfer entropy

Transfer entropy (TE) was recently introduced in [48], and is closely related to conditional entropy, but extends to two series, $I(t)$ and $J(t)$. The concept is the following:

Transfer Entropy =

+ information about future observation $I(t+1)$ gained from past observations of I and J

− information about future observation $I(t+1)$ gained from past observations of I only

= information flow from J to I.

This definition already reflects the key advantage of transfer entropy over other cross-correlation statistics: it is an asymmetric measure, that takes into account only statistical dependencies truly originating in the "source" series J, but not those deriving from a shared history, like in the case of a common external drive, as it would be the global daily cycle of volatility in our case, for instance. Expressing the above relationship with the conditional entropies h_m and using equation (14) leads to

$$\begin{aligned} T_{J\to I}(m,l) &:= h_I(m) - h_{IJ}(m,l) \\ &= \sum p(i_1,\ldots,i_{m+1},j_1,\ldots,j_l)\log_2\frac{p(i_{m+1}|i_1,\ldots,i_m,j_1,\ldots,j_l)}{p(i_{m+1}|i_1,\ldots,i_m)}, \end{aligned} \tag{19}$$

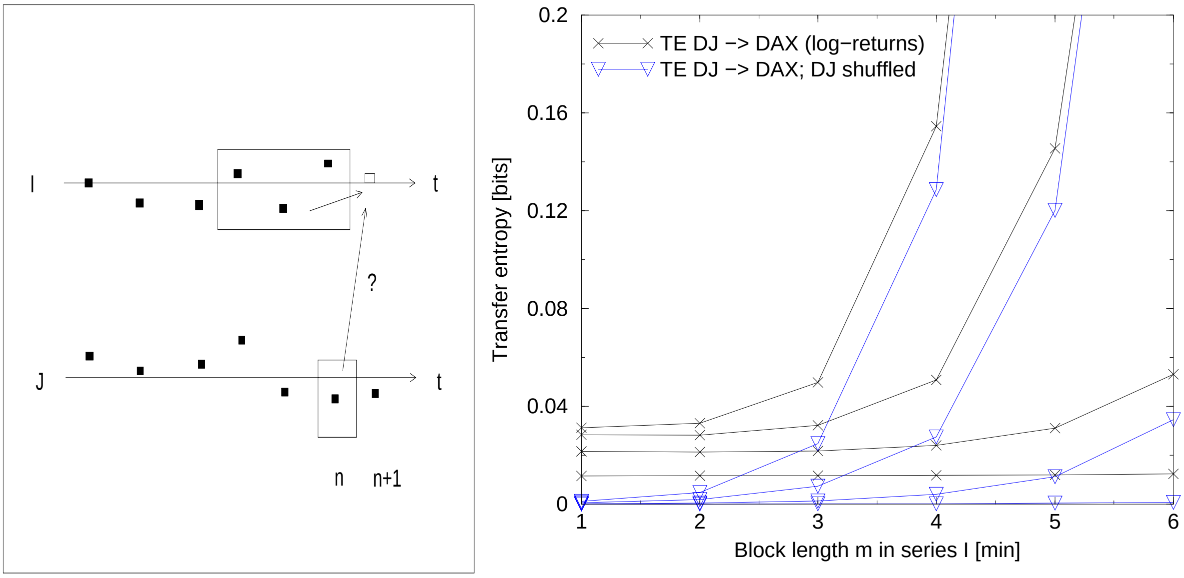

Fig. 11: Left: transfer entropy - basic concept (case m=3, l=1) from [48]. Right: transfer entropy measuring the information flow between Dow Jones and DAX series, using various partitions of S=2,3,4,5 symbols (bottom to top). Upper lines have been calculated on the log-returns of DJ and DAX, for the lower ones (triangles) the log-returns of the DJ series have previously been shuffled.

(20)

where the parameters m and l indicate the block lengths (=number of included past observations) in the I and J series, respectively. The sum must be taken over all possible states $i, j \in \{1, \dots, S\}$. The general concept is illustrated graphically in Fig. 11.

It would generally be desirable to choose the parameter m as large as possible, in order to avoid an erroneous misinterpretation of information present in the past of actually both series as information flow from J to I, but in practice the finite size of any real dataset imposes the need to find a reasonable compromise between unwanted finite sample effects (the amount of data required grows like $S^{(m+l)}$) and a higher accuracy. In a conservative approach it would thus be advisable to choose m as large as possible and set $l = 1$, which we will do in all forthcoming analyses. From equation (19) and (13) one deduces for the range of transfer entropy: $0 \leq T_{J \to I}(m, l) \leq H_I$.

In Fig. 11, right hand side, are displayed first results for the information transport from the DJ to the DAX series. The steady rise of the observed transfer entropy with increasing block length m is not compatible with the theoretical expectations, and therefore no information flow can be attributed to these "raw" findings. In order to investigate their significance we again use a correlation free, shuffled dataset, and confront the obtained results. As we said in section 3.1, the preprocessing of the "source" series J in form

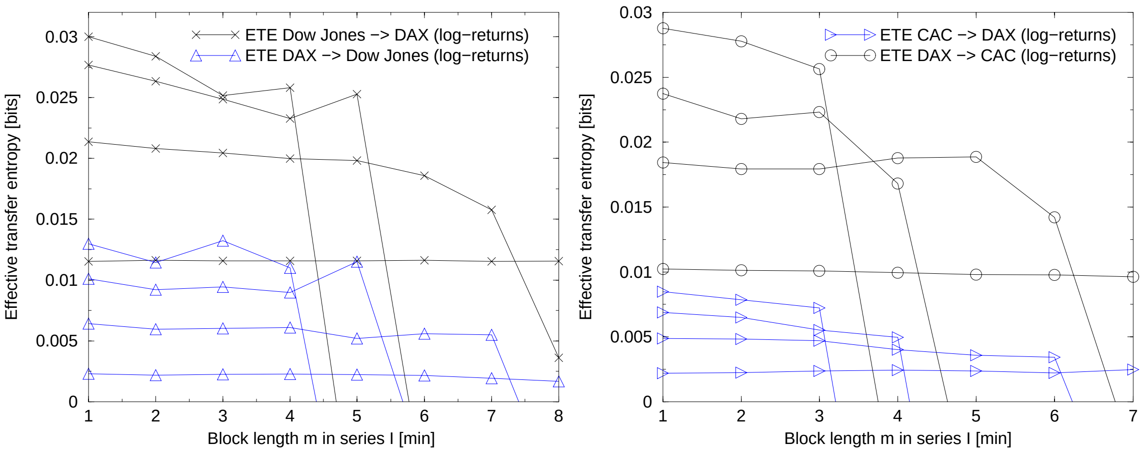

Fig. 12: Left: effective transfer entropy measuring the information flow between Dow Jones and DAX series, and vice versa, using four different partitions of S=2,3,4,5 symbols (bottom to top). Right: the same for the DAX and CAC series.

of shuffling destroys all possible correlations within that series, so that the afterwards observed information flow should be zero. However, as can be noted in Fig. 11 (right side), also the new curves calculated with the shuffled DJ log-returns rise monotonically and have similar values as their unshuffled counterparts. Since there cannot be any structure in the data, the observed non-zero values must be the artefact of the finite sample size, which also naturally accounts for the unexpected increase of the transfer entropy for growing block lengths m.

Intuitively, in order for transfer entropy to objectively confirm an information flow, the empirical curves need to be above the ones generated by the shuffled data, which can be interpreted as significance threshold. At this point, it is convenient to introduce a new variable, similar to what was done in section 3.1, that incorporates directly that intuitive point of view: we define *effective transfer entropy* (ETE) as the difference of the usual transfer entropy calculated for the empirical series and the transfer entropy calculated for the same series, but with the J series shuffled:

$$ET_{J\to I}(m,l) := T_{J\to I}(m,l) - T_{J_{\text{shuffled}}\to I}(m,l). \tag{21}$$

In Fig. 12 we show results of the effective transfer entropy for the cases Dow Jones vs DAX and DAX vs CAC[14], considering both possible directions of interaction. From the now much clearer overall picture the following conclusions can be deduced:

[14] Due to the reduced amount of data contained in the synchronized series, we found clear results only for few pairs of series, and will thus only discuss the two combinations for which the best results were obtained.

S	$\text{REA}_{DJ\to DAX}$[%]	$\text{REA}_{DAX\to DJ}$[%]	$\text{REA}_{DAX\to CAC}$[%]	$\text{REA}_{CAC\to DAX}$[%]
2	1.2	0.24	1.1	0.18
3	1.4	0.41	1.3	0.26
4	1.4	0.52	1.3	0.29
5	1.3	0.57	1.3	0.30

Tab. 5: Relative explanation added: how much (here in percent) of $I(t+1)$ can be explained only by $J(t)$?

- A flow of information from minute t of one series to the following minute of the other series is confirmed in both cases, and for both directions, thereby demonstrating that the interaction time of the global financial markets amounts to one minute or less.
- The series do not have the same relative "weight", i.e. more information is transferred from the DJ to the DAX, and from the DAX to the CAC, than vice versa, which in the case DJ/DAX may seem trivial as a purely economical fact, but it actually confirms in an independent way the validity of the transfer entropy formalism.

As was done in section 3.1, we will try to define some new variables, that allow a more straightforward interpretation of the numerical values obtained for the effective transfer entropy. Similar to the concept of *relative explanation*, we can relate the measured amount of information flow from J to I to the total flow of information in I. However, this does not correspond to the total explanatory power of the last observation of J with respect to a future observation of I, since any information contained in the past of J, but also in I, is not taken into account. Instead we are asking about how much of $I(t+1)$ is additionally explained, when we already know the past of series I, and then take into account the last observation of J, $J(t)$. Expressing this *relative explanation added* (REA) formally:

$$REA(m,l) := \frac{ET_{J\to I}(m,l)}{h_I(m)}, \tag{22}$$

for which we report quantitative results in Table 5.

The obtained values are smaller, but roughly of the same order as the ones reported in Table 4. Of course, the combined explanatory power of past observations of I must reach higher relative values as just the last observation of J, but its contribution is by no means negligible. For calculating the above figures in Table 5 we set the block length in I to $m=1$, which was justified by the fact that the relative explanation added varied very little when changing the block length (not shown here), and since $m=1$ gives the statistically most robust value, we ignored all others obtained for higher m.

It is interesting to note a certain clustering of the values in Table 5: especially for the cases Dow Jones → DAX and DAX → CAC, all values for partitions finer than the bipartition are rather close to each other. Also for the opposite directions of flow we observe a gap between the values found for the bipartition and all others. Since a bipartition has the special characteristic that it can only represent a linear statistical dependency, the observed jump in the information flow when going to higher resolutions possibly implies a nonlinear correlation between the series.

At the end of this chapter that illustrated some nonlinear approaches suitable for the analysis of financial time series, let us briefly discuss possible sources of errors. There are two aspects we retain the most important: the first one, concerning the stationarity of the data, constitutes a critical issue not only for this work, but for the whole data analysis branch of finance, econometrics, or econophysics. That financial data cannot be considered to be strictly stationary is widely accepted, but few attempts[15] have been made in order to develop statistical methods taking that into account appropriately. With reference to our case this means that we cannot assume total time independence for the single $p(i)$ and conditional $p(i|j)$ probabilities, and, in fact, in a moving window analysis fluctuations became apparent in the information flow between Dow Jones and DAX. This somewhat weakens the numerical results presented here, but the qualitative aspects, i.e. the existence of the information flow, should not be affected. Actually, the nonstationarity must not necessarily be disadvantageous, but instead could be used to identify periods of stronger and weaker coupling between the various indices - of course only for large enough datasets.

Since the measurement errors in the electronically elsewhere recorded data cannot be assessed here, the remaining cause of errors in our work is given by the statistical fluctuations in the performed calculations and estimates. For the sample length N analyzed here ($N > 10^5$) the error is rather small, and is judged to be negligible in comparison to the larger fluctuations induced by the weak stationarity of the data.

Apart from developing forecast algorithms that exploit the identified redundancies, a possible next step following the presented work could consist in measuring information flows between several financial time series, e.g. various FX-series, thereby deriving a currency taxonomy and a hierarchy of relative "weights".

4 Microscopical perspective

The dynamics of the stock market is still object of great debate. The variation of stock prices are usually considered to be a random process, and various forms of statistical

[15] The DFA (detrended fluctuation analysis [45]) represents one of them.

distributions have been proposed in order to describe correctly the empirical return distribution. In any case, some universal features have been identified, as was shown in the preceding chapters. It would therefore be of great interest to develop a model that is able to reproduce these aspects by a proper tuning of its parameters. Provided these parameters have a definite physical meaning, one could then discuss their microscopic influence on the macroscopically observed properties.

4.1 General aspects of market models

The leap of faith required when modeling financial markets is the assumption that it is not necessary to fully understand the individual components of the systems, i.e. the human agents, but rather their way of interacting. In how far this assumption can be justified theoretically remains an open question, but activities dealing with the modelization of complex, socioeconomic systems like the stock market are constantly growing within the physics community.

To a physicist, the question of whether a financial market operates at a critical point close to a phase transition (that could correspond to a crash or a speculative bubble) is especially interesting. The traditional theory of critical phenomena states that a system will approach a critical point via deliberate tuning of a certain control parameter. This description does not seem to apply to markets, however. The rules governing market dynamics were not chosen in order to put the market in a critical state, but it appears to have arrived there spontaneously, without any external tuning. This phenomenon, originally proposed as a possible explanation for scaling in many natural phenomena, is known as self-organized criticality.

There are several approaches to modeling market mechanics. In one principle class of models. price fluctuations result from the trading activity of conscious agents, whose decisions to buy or sell are dictated by well defined strategies, evolving in time and giving rise to a slowly changing fluctuation pattern. There is little doubt that the evolution and dynamics of investors' strategies and beliefs influence the long term behavior of real market prices. For example, if some company does not manage to keep up with its competitors, investors will sooner or later become aware of that, and in the long-term the corresponding stock price will go down. However, the temporal evolution of investment strategies cannot explain the properties of stock price fluctuations at very short time scales, where time is not sufficient for traders to update their strategies, or for a company to change its profile.

Another problem with models explaining short time price fluctuations in terms of strategy evolution is that they inevitably lead their creators to shaky grounds of speculations about relevant and irrelevant psychological motivations of a typical trader in a highly heterogeneous trader population. The remarkable universality of general features of price fluctuations in markets of different types of assets, such as stocks, options, foreign currency, and commodities, indicates that in fact individual psychological factors play little role in determining their short time properties, and stimulates the research for simpler mechanisms giving rise to these features.

4.2 Recently proposed concepts

Several models are based on the assumption that two different kinds of economic agents are interacting in the market: some authors [6] call them *dealers* and *savers*, others [34] use the names *fundamentalists* and *noise traders* (further distinguishing between optimistic and pessimistic), still others [4] speak of *rationals* and *chartists*. *Fundamentalists* follow the premise of the efficient market hypothesis meaning that they expect the price to follow the fundamental value of the underlying asset. A fundamentalist's trading strategy consists of buying when the actual market price is believed to be below the fundamental value, and selling in the opposite case[16]. *Noise traders*, on the other hand, do not believe in an immediate tendency of the price to follow the underlying fundamental value: they try to identify price trends and consider the behavior of other traders as a source of information, giving rise to the tendency towards *herding*.

Since the details of the circumstances which govern the expectations and decisions of the various involved individuals are unknown to the modeler, the behavior of a large number of heterogeneous agents may best be formalized using a probabilistic setting. When thinking of the scaling laws and complex behavior exhibited by physical systems where large numbers of single units interact, there seems to be no necessity to introduce different classes of agents, although it is absolutely reasonable from a macroscopic point of view. But another reason for avoiding distinctions among classes of traders lies in the fact that they introduce some collateral problems: it may be necessary, for instance, to let people move from one group to another, and to introduce a mechanism for the estimation of the fundamental value, but these two requirements sound somehow artificial, and in any case create further ambiguity.

An important model which has received a remarkable resonance was proposed by Lux and Marchesi [34], in which they show that scaling in finance emerges from the interaction of a large ensemble of market participants, in contradiction to the prevalent efficient market hypothesis in economics, according to which scaling in price changes would simply reflect similar scaling in the incoming news about future earning prospects. It is their model which introduced the already mentioned two groups of traders, the fundamentalists and the noise traders. Switches between the two groups are possible, noise

[16] How to properly estimate the fundamental value is another non trivial question.

traders can alternate opinion between pessimistic and optimistic, and the fundamental value constitutes the external driving force acting on the market. Properties like fat tails in the return distribution and correlated volatility are absent in the input signal, but they appear in the output signal, being generated by the microscopic interactions of the agents. Also the empirically known alternation between tranquil and turbulent trading conditions emerges naturally when simulating the model. A main disadvantage of the model is that the output signal is almost equal to the input, except for the presence of fat tails and correlated volatility. Furthermore, it is not very realistic to assume the fundamental value of a stock to be a purely random sequence.

Most models in economics and finance assume that investors behave rationally. The model of Levy, Levy, and Solomon [30] is able to determine the effects on asset prices of the investors' deviation from rationality. Here, the traders possess an incomplete and varying knowledge of their complex environment, i.e. the market. As result, the known positive correlation between volume and absolute returns has been reproduced. This is a clear example of how microscopic diversity may influence a macroscopic observable.

According to Yukalov [58] a market develops self-similarly, and therefore a self-similar approach to the market should be appropriate. Following this model, the evolution of the price follows autonomously some internal laws of the market. The problem consists in discovering these hidden internal laws which define the system's character. An attempt is constituted by employing the so called self-similar approximation theory, which supplies the mathematical tools for identifying the rules of the self-similar evolution. For the success of the model it is therefore crucial to correctly identify the transformation functions that enable the passage from one time scale to another. For empirical applications the latter request is quite strong, since finite size effects begin to play an important role. In the proximity of crashes, however, the scaling law behavior of the market is more evident and the model could give new insights, provided one is able to tune it.

A good modelization of the herding mechanism was provided by Cont and Bouchaud [12] via an artificial stock market with a random communication structure between the agents. Their setup is able to reproduce the heavy tails in the distribution of stock price variations in form of an exponentially truncated power law, similar to what has been observed in empirical studies of high-frequency market data. This way they provide a direct link between a microscopic phenomenon (herding) and its empirical outcome (fat tails). In particular, the authors suggest a relation between the excess kurtosis observed in asset returns, and the tendency of market participants to imitate each other. Furthermore, identifying and transcripting mathematically the different processes influencing demand and supply of financial assets, they manage to derive a nonlinear Langevin equation for stock market fluctuations and crashes. As a result, they can formally conclude that the asymmetry of risk aversion constitutes the principle prerequisite for crashes and for the sudden collapse of speculative bubbles, since panic is much more self reinforcing than a rally condition.

Stauffer has proposed an Ising interpretation [52] of the model of Count and Bouchaud, where clusters of parallel spins in a square lattice are defined as groups of traders acting together (super-spins). To take into account their tendency to be influenced by the opinion of other groups of agents, interactions among super-spins are incorporated in the model. Stauffer then applies the so called percolation theory, in order to get further insights into the dynamics of his proposed setup. To get an idea what percolation theory is about, let us consider the following example: we describe a forest by a square lattice with regularly spaced trees. Every tree has the same global probability p to be set on fire when one of its neighbors is burning. The question being asked now, is how a fire started at one edge of the lattice will spread out, and, in particular, whether it will reach the opposite edge. Of course, if $p = 0$ nothing happens, and with $p = 1$ the whole forest will burn. Interestingly, there exists a critical value of p, called the percolation threshold, above which (on average) at least one path connecting the two edges will form. Similarly to the percolation problem, Stauffer's model shows a crossover from a power-law to gaussian behavior for the return distribution.

4.3 Modeling the order book

We now want to focus our attention on a simple model [21, 40] simulating the book which stores the "bids" and "asks" during the trading activity. It is characterized by only one type of investors, whose goal is to maximize the profit while minimizing the risk. Every trader has a limited amount of money, and a given inclination towards investment. Also time comes into account, since a given gain has a different meaning whether realized within few weeks or after several years. At the beginning of the simulation, one agent is supposed to play the role of the central bank responsible for the *Initial Public Offer* (IPO): all the shares consequently belong to one agent. We provide a mechanism to generate news and advertisement, as a way to introduce global coupling into our model. During the IPO, traders feel a strong pressure to buy and generally almost all of them will order some quotas. Since the bank is responsible for the IPO, and cannot buy back any shares at the moment, this transient has a limited length, and the simulation reaches the typical trading regime after few iterations. The main building blocks of the model are the price formation and the book, where all the pending orders are stored. Every trader, when willing to buy a share, has to find a - for him - reasonable price according to past market values, opinion of the media, and suggestions coming from acquaintances. This constitutes the price by which he would like to enter the market; starting from it, every

BUY ORDERS					SELL ORDERS			
time	trader	shares	price		price	shares	trader	time
21005	240	4	11122		11123	4	576	19802
25008	207	70	11121		11124	4	876	14706
24506	647	3	11118		11125	2	806	12150
19002	820	2	11108		11130	49	201	17203
20148	100	12	11106		11130	4	792	20101

Fig. 13: First five levels of a typical book. It is divided in two parts, one containing buy orders, the other related to sell orders. For every order the following information are stored: time of insertion, PIN of the trader, number of involved shares, and desired price. Both lists are sorted according to the price, starting from the highest bid for buy-, and from the lowest ask for sell-orders. In case of two orders of the same type with the same price, the arrival time is taken into account. A transaction takes place whenever the two prices in the first line of the book coincide.

agent "keeps in mind" a target price and a stop-loss price (respectively according to the desired gain and the maximum loss). They are of fundamental importance for deciding whether and when to sell some shares, together with a certain threshold in time. There is no need of a fundamental price and/or external input.

In our model, every trader is characterized by the following quantities: *(i)* initial amount of money, *(ii)* number of owned shares, *(iii)* invested money, to keep trace of the average buying price, *(iv)* desired gain, *(v)* maximum loss, *(vi)* threshold, i.e. the amount of time after which the trader may start to change ideas about the investment. The key ideas here are the interplay between time and money, and the risk aversion represented by the stop loss mechanism. Every order is stored in the corresponding list of the book, according to its type (buy or sell), together with the requested price and the time at which it was submitted (see Fig. 13).

A transaction occurs whenever the lowest price in the sell list matches with the highest offer in the buyers' list: this value is defined as the market price of the stock at that particular instant (tick). The difference between limit and market orders becomes clear when looking at Fig. 13. Suppose a trader wants to buy 15 shares. The effect of a limit order with a price of 11120 would be the insertion of a line in the left list at the third position. The effect of a market order would be the exchange of 15 shares in the following way: 4 shares from trader 576 at price 11123, 4 shares from trader 876 at 11124, 2 shares

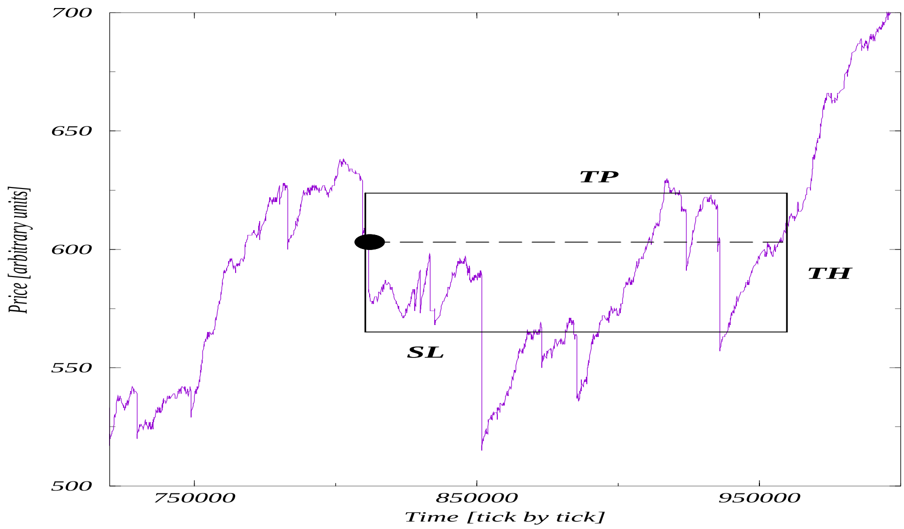

Fig. 14: **The trading rectangle**. Reported is the market price versus trading time. The filled circle indicates the moment in which the trader has bought shares. The dotted line, constant at the buying price, is plotted only for eye guide. The upper line is the target price (**TP**), the lower line refers to the stop loss price (**SL**), and the threshold in time (**TH**) defines the right end of the trading rectangle.

from 806 at 11125, and finally 5 shares from the trader with PIN = 201 at price 11130. A simple market order involving 15 shares would cause a price shift from the last transaction to 11130.

In accordance with our concept of avoiding any use of ambiguous fundamental rules, the model does not contain any rigorous mechanism which decides when the single agent enters the market. When randomly selected, a trader is willing to buy shares, if he neither possess any nor has a pending order. The empirical justification of such a behavior is that it is much more important to identify the right moment to sell than to buy, because it is only when you sell that you get the extra money you have won, or you realize your loss. Let us have a look at Fig. 14 for a better understanding of this concept. Suppose that a trader has bought shares at the price and the time marked by the filled circle. The basic strategy is represented by the *trading rectangle*, defined by the three following quantities: target price (upper horizontal line), stop-loss price (lower horizontal line), and threshold in time (rightmost vertical line).

As long as the market price is confined within the trading rectangle, the agent does not feel the need to trade, but once this condition has been violated, it is very likely for him to perform an operation. If the price goes beyond one of the two horizontal lines, a market order to sell the shares is very probable (either to cash the win or to limit the loss). If the price remains almost constant within the trading rectangle, and therefore the

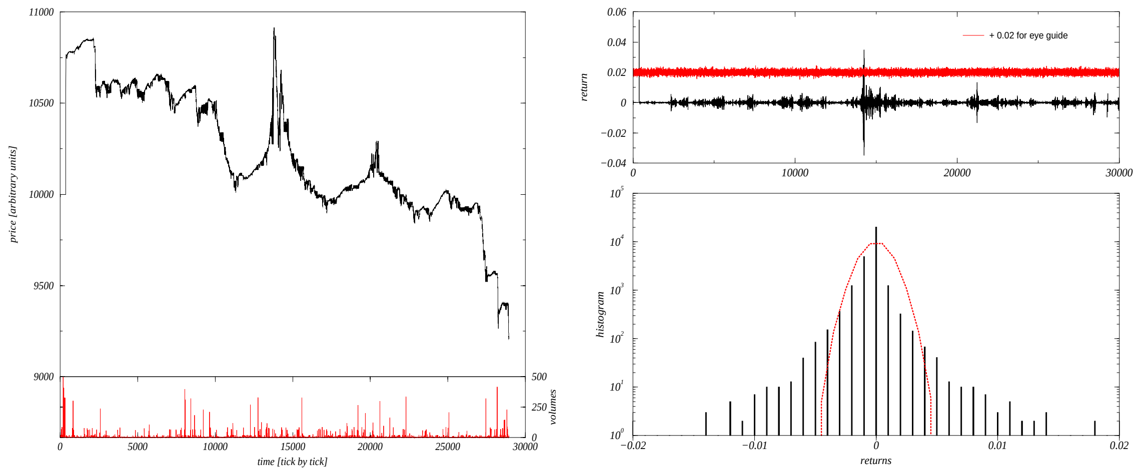

Fig. 15: Left side: typical time series segment from a simulation run (Upper panel: temporal evolution of market price. Lower panel: temporal evolution of the corresponding exchanged volume). Right side, upper panel: price returns, with the random series shifted upwards for eye guide. Right side, lower panel: return distribution. The comparison with a best-fitted normal distribution reveals the presence of *fat tails*.

time series ultimately crosses the rectangle at its rightmost vertical line, the decision of the trader depends on a global condition, which is given by the imbalance of the book, namely by the ratio between selling and buying orders. If too many people want to sell, this constitutes a good reason to leave the market as soon as possible (therefore with a market order). If a lot of agents are willing to buy, then it can be better to keep the shares, because their value could appreciate substantially in a near future.

4.4 Simulating the order book

Fig. 15 shows the representative result of a simulation of the market price evolution and the corresponding amount of exchanged shares. The model is able to reproduce all the typical features observed in empirical data. In the beginning, the price remains constant due to the ongoing IPO phase, meaning that the bank offers the shares to the traders at a fixed price, the IPO price. After that, one can see the typical pressure made by agents who did not get enough shares during the initial public offer: the volumes are high and the price tends to rise. Then, after a settlement, the price starts to oscillate, with very little volume being exchanged: traders with shares do not want to sell because they hope to get more money if they still wait; agents without shares do not buy because the price is too high, and there is no evidence of a trend. Then oscillations become stronger and stronger, and when the volumes are large as well, a small crash occurs and the price returns back to a more interesting value for potential buyers. As a consequence, volumes remain high, and the market experiences a so called rally period, followed again by a crash, maybe due to the fact that the bubble phase has been too optimistic.

As shown in Fig. 15, the probability density function of the returns of the simulated stock shows a strong leptokurtic nature. For comparison, the gaussian distribution with the same measured standard deviation is also reported. The time series given by the artificially generated returns exhibits a higher frequency of extreme events, and a clustering of volatility. The qualitative difference between the return time series of the model and gaussian noise can clearly be seen in the upper panel on the right side of Fig. 15. The estimation of the self-similarity parameter, the Hurst exponent $\mathcal{H}$, reveals a strong persistence in the volatility with $\mathcal{H} = 0.85$.

To summarize, in the last two sections we have discussed a model for the stock market, which is able to reproduce the two main characteristics of empirical data, namely correlated volatility and fat tails. We have performed this task avoiding the use of different classes of agents, and the artificial introduction of a fundamental price. We just made use of realistic assumptions about the behavior of traders, i.e. limited amount of money, limited time of liquidity, desired gain, and maximum acceptable loss.

5 Forecasting

5.1 Sources of unpredictability

It seems to be very intuitive to believe that once an accurate mathematical description of a physical system has been found, it automatically leads to a profound understanding of the system's properties, and, along with that, gives rise to the possibility of making significant predictions about its temporal evolution. In fact, these assertions have been proved and used for a wide variety of phenomena, ranging from the motion of planetary bodies to the fundamental constituents of matter. However, it is not difficult to show that these assumptions are not generally true when dealing with nonlinear phenomena and nonstationarity. This might seem surprising, since it is a common experience that although some details are missing, approximate versions of the "correct" laws may be used to make robust predictions about a system's behavior, predictions that are often confirmed experimentally with satisfying accuracy. However, if the future evolution of a system results unpredictable, this does not imply that the system is fundamentally random. The inverse is obviously true, but randomness is not the only source for the lack of forecasting power.

For illustration, let us consider the laws of planetary motion, as formulated by Newton. It is possible to predict the orbit of the moon around the earth with a very good accuracy, because the influence of other planets of the solar system can be ignored, and the so called two-body-problem can be solved analytically without particular effort. These predictions have been tested over centuries and were found to be robust. Now, if a smaller third planet is introduced into Newton's mathematical description of the gravitational interaction of massive bodies, we are led to an intractable three-body-problem. Newton solved various restricted versions of the complete problem, but he was unable to find a general solution to it. Two centuries later, Poincaré suggested that the motion of the third smaller planet orbiting in the gravitational field of two massive ones would generally be highly complicated. With today's computational power we can obtain very precise numerical solutions of the three-body-problem, and it can be shown that the orbit of the third planet is indeed unpredictable in practice: every small error in the setting of the initial conditions will drastically reduce the time horizon of the prediction.

The idea that almost nothing is really linear but can quite easily be linearized has become a widespread believe, naturally due to the considerable success achieved by approximate linear methods in a wide range of problems. The same is true for the concept of nonstationarity: the problem of the non-constancy of parameters can be overcome by dividing the time series into intervals and verifying their stationarity. However, many natural processes across the whole spectrum of science are inherently strongly nonlinear and nonstationary, and simple adaptation of known methods may not be sufficient to resolve important issues, such as prediction. Therefore, there is a need to develop new ways of dealing with complex processes, and, in fact, questions of nonstationarity and prediction constitute a very active field of current research.

5.2 Nonstationarity

In order to study an unknown system, one needs to extract information about it. The usual way of confronting this task consists in measuring some quantities related to the system, taking into account that a scientific measurement of any kind is useful only as far as it is reproducible, at least in principle. One has to be sure that the values obtained from the measurement device correspond to properties of the system, and not, e.g., of the measurement device. The concept of reproducibility, and therefore of meaningfulness, is strictly related to the notion of *stationarity*.

Stationarity means that all the parameters of the system remain constant during the measurement, but unfortunately, in most cases, one has no direct access to all the involved parameters (one might not even know how many relevant parameters there are), and therefore it is often difficult to affirm stationarity with a good degree of confidence in this rather abstract sense. Consequently, a practically utilizable definition of stationarity has to be related to the available time series, from which information about the system and any quantity of interest can be extracted. With regard to the information retrieved from a time series, we call a process stationary, if all transition probabilities from one state of the system to another are independent of time, at least within the observation period. This actually represents a stronger requirement than the constance of all parameters, since now the measurement additionally needs to be sufficiently long or precise to enable a statistically sound deduction of the system's transition probabilities. As further necessity in order to avoid erroneous results due to nonstationarity, one always needs to observe a system for a sufficiently long period of time, i.e. much longer than any characteristic time scale of the system itself. In field measurements, nonstationarity is ubiquitous.

One can try to divide different forms of nonstationarity into three basic types according to the following scheme:

- *Drift of parameters.* The control parameters of the dynamical system generating the time series are not constant. Different segments of the time series are related to different instantaneous dynamics.
- *Diffusive properties.* The transition probabilities are constant, but the marginal probabilities spread out, and therefore we get a lack of recurrence for the process. As an example, one can imagine a random walker on a line, moving to the right or to the left with the same probability. The mean position is the initial one, but the variance is increasing with time.
- *Trends and seasonality.* These are typical features of financial time series and make the estimation of several quantities not reliable. Sometimes trends can be overcome by using appropriate tools, such as the detrended fluctuation analysis (DFA).

A simple stationarity check consists of first dividing the dataset into several segments, then computing some quantity for each of them, and finally testing whether these quantities differ beyond their usual statistical fluctuation. If they do, the analyzed data might likely be nonstationary. Unfortunately it can also happen that a parameter drift does not produce any visible drift in the measurements. In such cases one needs special nonlinear dynamical relations; quantities to be compared for the different subsets of data can consist in the prediction error with respect to some nonlinear model for instance.

5.3 Linear models

The most popular class of stochastic models in time series analysis and modeling consists of linear filters acting on a series of independent noise inputs, and on past values of the signal itself. An obvious problem consists in choosing the right input series, which cannot be derived directly from the empirical data we want to reproduce. When building such a model, we have to estimate all the parameters from the output only; i.e. for every new set of parameters we also have to generate a new and independent noise input.

The moving average model (MA) is a filter on a series of gaussian white noise input η_n:

$$x_n = \sum_{j=0}^{M} b_j \eta_{n-j}, \tag{23}$$

where $\langle \eta_n \eta_m \rangle = \sigma^2 \delta_{nm}$ and $\langle \eta \rangle = 0$. The number M of adjustable parameters is called the order of the process. Note that x_n is also a gaussian random variable with zero mean. This model is also called *finite impulse response* filter, since the signal vanishes after M steps, if the input is given by a single pulse.

Alternatively, in a so called autoregressive model (AR), the output is given by a linear combination of the past signal, plus additive noise:

$$x_n = \sum_{j=1}^{N} a_j x_{n-j} + \eta_n. \tag{24}$$

This defines an AR model of order N, where η_n is white gaussian noise as in the previous model. Again, x_n is a gaussian random variable. AR models are able to create noisy harmonic motion, and are particularly appropriate if the spectrum of a time series is dominated by sharp peaks at distinct frequencies, in contrast to the MA model, which is preferable if the estimated spectrum is of the form of colored noise, i.e. without a prominent peak.

In principle, all gaussian linear stochastic processes can be modeled with arbitrary accuracy by either of the two approaches. The order of the process can become extremely large in particular circumstances. For example, to model an harmonic noisy oscillation with a MA process would require an infinite number of terms. A generalization of both is a combination of them, which is called autoregressive moving average process (ARMA). With such a process one is able to obtain a power spectrum with poles and a polynomial background. Since the noise input η_n is not known, it must be averaged over, which leaves the AR part of the model as the possible predictive part.

Real data are often not gaussian distributed. If one wants to model them using one of the discussed processes (MA, AR, ARMA), it is usually assumed that a nonlinear transformation has distorted the output of the originally gaussian random process, and thereby changed the distribution to the observed one. Such nonlinearities are called static, because they do not intervene directly in the dynamics of the system, and they also conserve the property of time reversal invariance. Before fitting a model to such data, one should render the distribution gaussian by inverting the nonlinear transformation. A typical problem consists in *overfitting*: since one can reproduce the data better by using more and more parameters, it is necessary to identify an appropriate maximum for the order of the process. We will return to this concept later on.

5.4 Nonlinear models

Linear methods interpret all regular structure in a dataset as linear correlations, implying that the intrinsic dynamics of the system are governed by the linear paradigm that small causes lead to small effects. Since linear differential equations of motion can only lead to exponentially growing or periodically oscillating solutions, all irregular behavior of the system has to be attributed to some random external input to the system. But random input is not the only possible source of irregularity in the output of a system. Nonlinear, chaotic systems can produce very irregular data with purely deterministic equation of motion.

Autoregressive models can be generalized by introducing nonlinearities. One important class [54] consists of threshold autoregressive models (TAR), consisting of a collection of standard AR models, where each single one is valid only within a certain domain. For the construction of the model one divides the reconstructed phase space into patches, and determines the coefficients of each single AR model as usual, using only data points of the corresponding patch. TAR models are therefore piecewise linear models and can be regarded as coarse-grained versions of local methods in phase space. Alternatively, AR models can be extended by nonlinear terms.

AR models are a special class of *Markov models*, which rely on the notion of a state space. A Markov model of order m is a model where the probability of finding the signal at time n in some state (e.g. a certain scalar interval) depends only on the values of the last m time steps of the signal, which define the state of the system. The concept of memory becomes very clear in the framework of Markov models. Even for deterministic systems, a stochastic description arises naturally if not all relevant variables are taken into account explicitly. Thus, if some coarse-graining procedure is applied to a deterministic system, the evolution of the coarse-grained variables might be stochastic, if the original system was chaotic.

If the time series is long enough and the noise level low, local methods can be very powerful. They derive neighborhood relations from the data and map them forward in time. They are conceptually simpler than global models but they can require a larger numerical effort. An AR model cannot cope with chaos, since chaos relies on nonlinearities. But one can construct AR models locally in a proper embedding space, by finding an approximation to the tangent plane. The neighborhood size is the result of the trade-off between a reliable determination of the coefficients of the local model (large size in order to include as many data points as possible), and the need to avoid an overlapping of different parts of the signal (small size). Usually, attractors can be embedded locally in fewer dimensions than are required for a global reconstruction.

5.5 Stochasticity vs. deterministic chaos

The problem of distinguishing between nonlinear determinism[17] and stochasticity has not yet been solved satisfyingly. Even knowing the nature of the underlying process does not lead automatically to the knowledge of the character of the signal, since it might also depend on the measurement setup, and in particular on the resolution. In fact, there are processes that, when observed on small length scales (= high resolution) appear completely deterministic, but when transiting to larger scales (= low resolution) they change their character and become more and more stochastic. Thus, one can define the notion of deterministic or stochastic behavior in dependence of the considered range of length-scales. Even the concept of the "real nature" of such a process becomes subtle in such a case, and therefore also the distinction, based only on data analysis, between a genuine deterministic system, and one with intrinsic randomness.

A detailed discussion of this topic is beyond the scope of the present paper. However, we can remark that typical methods [27] of approaching this problem consist in estimating the correlation dimension or Lyapunov exponents, and then interpreting any finite value found for these quantities as a sign for the deterministic nature of the signal. With the help of embedding techniques (i.e. methods to reconstruct a phase space starting from a scalar measurement), one can show that noise and other stochastic processes fill up all available dimensions in phase space; in the opposite case, a deterministic signal shows some sort of convergence once a dimensionality larger than the number of its active degrees of freedom is reached[18].

The type of model and its degree of "sophistication" is particularly important when trying to do prediction. Let us consider the following simple example. It is well known that two given distinct points define uniquely a line. Similarly, with three points one can identify a parabola. In general, a polynomial of degree n can be identified once $(n+1)$ of its points are given. Imagine now to have a time series of length n, and to have the intention of forecasting the $(n+1)$th point with the help of the previous n. Using a polynomial of degree $(n-1)$ would provide a perfect interpolation of the data, but the predictive power of such a model would probably be extremely poor. The principal problem of such an apparently perfect approach is that no attempt has been made to distinguish between signal and noise, and using the latter to forecast future values of the signal will naturally

[17] Chaos constitutes a special case of deterministic nonlinear dynamics. Nonlinearity is an essential ingredient of chaos, but by itself does not already imply chaos.

[18] This corresponds to the dimension of the so called attractor, which might be a non-integer number.

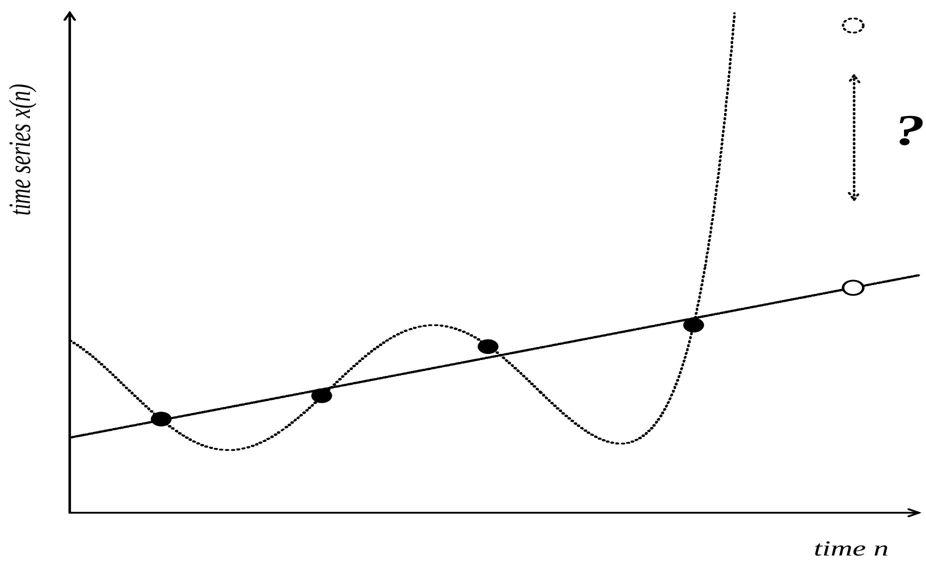

Fig. 16: *Overfitting.* Given four measurements (filled circles), there are several alternatives for a model that is supposed to predict the next value. It is generally not true that reducing the fitting error leads to a better forecast, if one just uses more parameters for the "improved" fit. The empty circle is the result of a linear forecast. A higher order model with a perfect interpolation would predict the dotted circle as next observation, which evidently represents no improvement of the forecast.

only produce nonsense.

Fig. 16 provides a clear example of this concept, called *overfitting.* Imagine we have obtained four points from a measurement (filled circles), and we want to speculate about the next observation. The principle of parsimony[19] would suggest to use a linear fit, since it keeps the interpolation error already quite small. The linear approximation is reported in Fig. 16, together with the predicted next point (empty circle). Alternatively, one could reduce the interpolation error to zero using a higher order polynomial (dotted curve). The prediction obtained from that approach would even be outside the plotting range (dotted circle). Although we do not know the coming value of the time series, the second prediction does not seem to be reasonable at all. In any given model, the parsimony principle may help to avoid such artefacts in form of inconsistencies, ambiguities or redundancies. Last but not least, developing a simpler model will also be easier.

When constructing models, one usually aims at a *complete* description of the empirical system under exam. It is interesting to ask what such a complete description in the deterministic case signifies. In the mathematical sense, the system's equations together

[19] Ockham's razor: one should not increase, beyond what is necessary, the number of entities required to explain a given phenomenon. It is a principle attributed to the medieval philosopher William of Ockham.

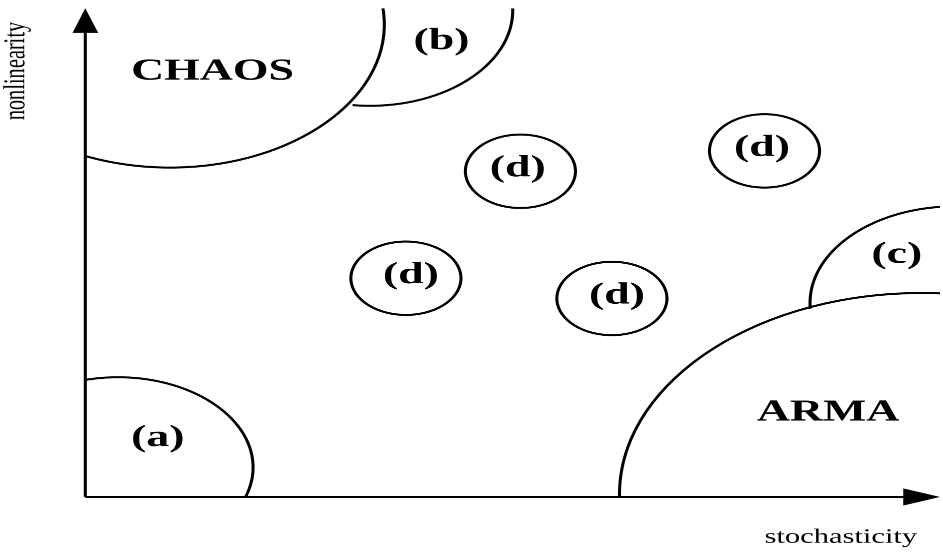

Fig. 17: Qualitative representation of the variety of processes involving stochasticity and nonlinearity. (a) Periodic oscillations. (b) Extension of chaos for small noise. (c) Extension of ARMA models for small nonlinearities. (d) Markov processes.

with the initial conditions are sufficient. For this to be true, the latter must be known with infinite precision, which is unphysical. If the system is chaotic, then even in the noise free case errors in the initial condition will grow exponentially with time. Of course, as soon as noise is present, the situation becomes worse. The two paradigms, nonlinear deterministic and linear stochastic behavior, are the extreme positions in the space spanned by the properties nonlinearity and stochasticity. They are singled out not only because they are particularly interesting for many real-world situations, but also because of their paradigmatic role and their well-known mathematical foundations.

There exists more than one way to switch from predictable to unpredictable as is illustrated in Fig. 17. Stochasticity and chaos both have the the property of severely limiting any forecast potential. In the qualitative representation of Fig. 17, several explored areas that correspond to some class of analytically well representable process are outlined. The simplest case (a) consists of periodic oscillations. When increasing the nonlinearity, but keeping a strictly deterministic setup, chaos may appear: sensitivity to initial conditions and exponential divergence of neighboring trajectories reduce drastically the predictability, at least to all practical effects (infinite precision, although unphysical, would help). Instead of considering nonlinear effects, one can introduce stochasticity to obtain what was previously discussed in the framework of ARMA models and their extensions. Of course, one can also take into account the effects of a small nonlinearity (c) in the model.

However, there are a lot of areas (d) in our plot where no real closed formalism is available to describe the corresponding process, situated between purely deterministic chaotic, and stochastic dynamics, like, e.g., the *Markov process*. Extending the concept of the state of a system, it is characterized by its order, which corresponds to the number of past states that contain information about the present one, or, in other words, the order defines the memory of the Markov process. According to current speculations, financial markets could be located in the vicinity of these islands.

6 Conclusion

Let us now briefly summarize the principle results of the analyses presented in this article. In the first part we investigated quantitatively a number of financial time series, that were recorded minute per minute for a time period of about one year. The following statistical properties were identified:

- With the help of the linear autocorrelation function we confirmed the existence of a long-range autocorrelation between the absolute price-changes, a phenomenon called correlated volatility.
- We observed scale invariance for all considered financial time series, even for the interest rates, which usually are not included in such tests. Hurst exponents for the price changes were found to be compatible with $\frac{1}{2}$, i.e. brownian motion. In case of the absolute price changes, Hurst exponents significantly larger than $\frac{1}{2}$ were found, confirming again the existence of long-range correlations.
- The role of the daily volatility cycle was recognized and elaborated; in particular it was shown how its presence induces a global coupling between all considered stock indexes.
- Significant linear cross-correlations were shown to exist between some series at time lag zero.
- By using information-theoretic nonlinear tools we identified general redundancies within all time series. An idea was given of how much "historical" values of a series can help to explain a future value.
- Applying the nonlinear tool of transfer entropy led to the detection and quantification of an information flow between two pairs of stock indexes, meaning that there is a causal interaction between those markets at a time scale of only one minute.

We then discussed some recent approaches regarding the microscopic modeling of financial markets, emphasizing on key concepts and principle problems. A particular model, which by simulating the order book is capable of reproducing the main statistical characteristics observed in financial markets, was discussed in more depth.

The last topic addressed was forecasting. After having introduced some standard models used frequently in finance, we showed that both stochasticity and nonlinearity can explain the absence of a significant forecast horizon, and how it is sometimes difficult to distinguish between these principle types of processes. We discussed the obstructive and yet fundamental role of nonstationarity and the problem of overfitting. Finally, it was tried to set up a general classification scheme of various processes in terms of nonlinearity and stochasticity, in which a possible localization of the financial markets as Markov process was indicated.

References

[1] A. Arneodo, J.F. Muzy, D. Sornette, *"Direct" causal cascade in the stock market*, European Physical Journal B 2, 277-282 (1998).

[2] L. Bachelier, *Théorie de la speculation*, Annales Scientifiques de l'Ècole Normal Supérieure, III (1900). English translation in [13].

[3] R. Badii and A. Politi, *Complexity: hierarchical structures and scaling in physics*, Cambridge University Press (1997).

[4] P. Bak, M. Paczuski, M. Shubik, *Price variations in a stock market with many agents*, Physica A 246, 430-453 (1997).

[5] R. Baviera, D. Vergni, A. Vulpiani, *Markovian approximation in foreign exchange markets*, Physica A 280, 566-581 (2000).

[6] G.I. Bischi, V. Valori, *Nonlinear effects in a discrete-time dynamic model of a stock market*, Chaos, Solitons and Fractals 11, 2103 (2000).

[7] F. Black and M. Scholes, *The pricing of options and corporate liabilities*, Journal of Political Economy 81, p.637 (1973).

[8] T. Bollerslev, *Generalized autoregressive conditional heteroskedasticity*, Journal of Econometrics 31, 307-327 (1986).

[9] T. Bollerslev, R.Y. Chou and K.F. Kroner, *ARCH modeling in finance: A review of the theory and empirical evidence*, Journal of Econometrics 52, 5-59 (1992).

[10] J.-P. Bouchaud and M. Potters, *Theory of financial risks*, Cambridge University Press (2000).

[11] D. Chowdhury and D. Stauffer, *A generalized spin model for financial markets*, European Physical Journal B, 8, 477-482 (1999).

[12] R. Cont, M. Potters, J.P. Bouchaud, *Scaling in stock market data: stable laws and beyond*, from: B. Dubrulle, F. Graner, D. Sornette (Eds.), *Scale invariance and beyond*, EDP Sciences, Les Vlis, France (1997).

[13] P.H. Cootner, *The random character of stock market prices*, The MIT Press, Cambridge, Massachusetts, USA (1964).

[14] M.M. Dacorogna, U.A. Müller, R.J. Nagler, R.B. Olsen, O.V. Pichet, *A geographical model for the daily and weekly seasonal volatility*, Journal of International Money and Finance 12, 413-438 (1993).

[15] G.A. Darbellay and D. Wuertz, *The entropy as a tool for analysing statistical dependences in financial time series*, Physica A 287, 429-439 (2000).

[16] R. D'Hulst and G.J. Rodgers, *Democracy versus dictatorship in self-organized models of financial markets*, Physica A 280, 554-565 (2000).

[17] C. Dunis and B. Zhou (edited by), *Nonlinear modeling of high frequency financial time series*, Wiley (1998).

[18] R.F. Engle, *Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation*, Econometrica 50, 987-1008 (1982).

[19] R.F. Engle and G. Gonzalez-Riviera, *Semiparametric ARCH models*, Journal of Business and Economic Statistics 9, 345-360 (1991).

[20] R.F. Engle and C. Mustafa, *Implied ARCH models from option prices*, Journal of Econometrics 52, 289-311 (1992).

[21] F. Franci, R. Marschinski, L. Matassini, *Learning the Optimal Trading Strategy*, Physica A 294, 213, (2001).

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

[23] P. Gopikrishnan, M. Meyer, L.A.N. Amaral and H.E. Stanley, *Inverse cubic law for the distribution of stock price variations*, The European Physical Journal B 3, 139 (1998).

[24] P. Grassberger, *Finite-sample corrections to entropy and dimension estimates*, Physics Letters A 128, 369-373 (1988).

[25] D.M. Guillaume et al., *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, 95, (1997).

[26] H. Kantz and T. Schürmann, *Enlarged scaling ranges in entropy and dimension estimates*, CHAOS 6, 167-171 (1996).

[27] H. Kantz and T. Schreiber, *Nonlinear time series analysis*, Cambridge University Press (1997).

[28] A.N. Kolmogorov, *Information theory and the theory of algorithms*, Selected works Vol.3, Kluwer, Dordrecht (1993).

[29] C.G. Lamoureux and W.D. Lastrapes, *Persistence in variance, structural change and the GARCH model*, Journal of Business and Economic Statistics 8, 225-234 (1990).

[30] M. Lévy, H. Lévy, S. Solomon, *Microscopic simulation of the stock market: the effect of microscopic diversity*, Journal de Physique I 5, 1087-1107 (1995).

[31] Y.H. Liu, P. Gopikrishnan, P. Cizeau, M. Meyer, C.K. Peng, H.E. Stanley, *Statistical properties of the volatility of price fluctuations*, Physical Review E 60, 1390-1400 (1999).

[32] Y.H. Liu, P. Cizeau, M. Meyer, C.K. Peng, H.E. Stanley, *Correlations in economic time series*, Physica A 245, 437-440 (1997).

[33] E.N. Lorenz, *Deterministic non-periodic flow*, J. Atmos. Sci. 20, 130-141 (1963).

[34] T. Lux and M. Marchesi, *Scaling and criticality in a stochastic multi-agent model of a financial market*, Nature 397, 498-500 (1999).

[35] B.B. Mandelbrot, *The variation of certain speculative prices*, The Journal of Business 36, 394-419 (1963).

[36] B.B. Mandelbrot, *Fractals and scaling in finance: Discontinuity, Concentration, Risk*, Springer, New York, USA (1997).

[37] R.N. Mantegna and H.E. Stanley, *Scaling behavior in the dynamics of an economic index*, Nature 376, 46 (1995).

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

[39] R.N. Mantegna and H.E. Stanley, *An introduction to econophysics. Correlations and complexity in finance*, Cambridge University Press (2000).

[40] L. Matassini and F. Franci, *On Financial Markets Trading*, Physica A 289, 526 (2001).

[41] R.C. Merton, *Theory of rational option pricing*, Bell Journal of Economics and Management 4, p.141 (1973).

[42] L. Molgedey and W. Ebeling, *Intraday patterns and local predictability of high-frequency financial time series*, Physica A 287, 420-428 (2000).

[43] J. Moody and W. Lizhong, *High frequency foreign exchange rates: Price behavior analysis and "true price" models*, in [17].

[44] W.Paul and J. Baschnagel, *Stochastic processes: from physics to finance*, Springer Verlag (1999).

[45] C.-K. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, A.L. Goldberger, *Mosaic organization of DNA nucleotides*, Physical Review E 49, 1685-1689 (1994).

[46] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral, H.E. Stanley, *Universal and nonuniversal properties of cross correlations in financial time series*, Physical Review Letter 83, 1471 (1999).

[47] M. Potters, R. Cont, J.-P. Bouchaud, *Financial markets as adaptive systems*, Europhysics Letters 41, 239-244 (1998).

[48] T. Schreiber, *Measuring information transfer*, Physical Review Letters 85, 461 (2000).

[49] C.E. Shannon and W. Weaver, *The mathematical theory of information*, University of Illinois Press, IL, USA (1949).

[50] D. Sornette, A. Johansen and J.-P. Bouchaud, *Stock market crashes, precursors and replicas*, Journal de Physique I 6, 167-175 (1996).

[51] M.H.R. Stanley et al., *Scaling behavior in the growth of companies*, Nature 379, 804 (1996).

[52] D. Stauffer and T.J.P. Penna, *Crossover in the Cont-Bouchaud percolation model for market fluctuations*, Physica A 256, 284-290 (1998).

[53] P. Talkner and R.O. Weber, *Power spectrum and detrended fluctuation analysis: Application to daily temperatures*, Physical Review E 62, 150-160 (2000).

[54] H. Tong, *Threshold models in non-linear time series analysis*, Notes in Statistics Vol. 21, Springer, New York (1983).

[55] N. Vanderwalle and M. Ausloos, *Coherent and random sequences in financial fluctuations*, Physica A 246, 454-459 (1997).

[56] N. Vanderwalle and M. Ausloos, *Sparseness and roughness of foreign exchange rates*, International Journal of Modern Physics C 9, 711-719 (1998).

[57] J. Voit, *The statistical mechanics of capital markets*, Springer (2001).

[58] V.I. Yukalov and S. Gluzman, *Self-similar crossover in statistical physics*, Physica A 273, 401-415 (1999).

Acknowledgement

The authors would like to thank Deutsche Bank Research, and in particular C. Kreuter, for having provided the financial time series.

All Deutsche Bank Research products are also available by e-mail. **Subscribers receive the electronic publication on average four days earlier than the printed version.** If you are interested in receiving this product by e-mail, please get in touch with your Deutsche Bank sales contact or with the DB Research Marketing Team: **marketing.dbr@db.com**

Printed by: HST Offsetdruck GmbH, Dieburg
Print: ISSN 1615-956X / Elektr.: 1615-9683 / Internet: ISSN 1616-0428