Available online at www.sciencedirect.com
ScienceDirect
Physica A 382 (2007) 193–198
PHYSICA A
www.elsevier.com/locate/physa

Medium and small-scale analysis of financial data

Andreas P. Nawroth, Joachim Peinke*

Institut für Physik, Carl-von-Ossietzky Universität Oldenburg, D-26111 Oldenburg, Germany

Available online 30 March 2007

Abstract

A stochastic analysis of financial data is presented. In particular we investigate how the statistics of log returns change with different time delays τ. The scale-dependent behaviour of financial data can be divided into two regions. The first time range, the small-timescale region (in the range of seconds) seems to be characterised by universal features. The second time range, the medium-timescale range from several minutes upwards can be characterised by a cascade process, which is given by a stochastic Markov process in the scale τ. A corresponding Fokker–Planck equation can be extracted from given data and provides a non-equilibrium thermodynamical description of the complexity of financial data.

Keywords: Econophysics; Financial markets; Stochastic processes; Fokker–Planck equation

1. Introduction

One of the outstanding features of the complexity of financial markets is that very often financial quantities display non-Gaussian statistics often denoted as heavy tailed or intermittent statistics [1–9]. To characterise the fluctuations of a financial time series $x(t)$, most commonly quantities like returns, log returns or price increments are used. Here, we consider the statistics of the log return $y(\tau)$ over a certain timescale τ, which is defined as

$$y(\tau) = \log x(t+\tau) - \log x(t), \tag{1}$$

where $x(t)$ denotes the price of the asset at time t. A common problem in the analysis of financial data is the question of stationarity for the discussed stochastic quantities. In particular we find in our analysis that the methods seem to be robust against nonstationarity effects. This may be due to the data selection. Note that the use of (conditional) returns of scale τ corresponds to a specific filtering of the data. Nevertheless the particular results change slightly for different data windows, indicating a possible influence of nonstationarity effects. In this paper we focus on the analysis and reconstruction of the processes for a given data window (time period). For further information concerning stationarity and the methods used here, refer to Refs. [10,11]. The analysis presented is mainly based on Bayer data for the time span of 1993–2003. The financial

*Corresponding author.
E-mail address: peinke@uni-oldenburg.de (J. Peinke).

0378-4371/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2007.03.041

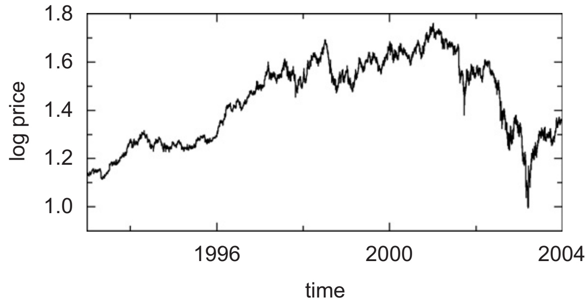

Fig. 1. Log price for Bayer for the years 1993–2003.

data sets were provided by the Karlsruher Kapitalmarkt Datenbank (KKMDB) [12]. In order to illustrate the underlying data, the graph of the logarithm of the price time series is shown in Fig. 1.

2. Small-scale analysis

The statistics of $y(\tau)$ are shown in Fig. 2. One remarkable feature of financial data is the fact that the probability density functions (pdfs) are not Gaussian, but exhibit heavy tailed shapes. Another remarkable feature is the change of the shape with the size of the scale variable τ. To analyse the changing statistics of the pdfs with the scale τ a non-parametric approach is chosen. The distance between the pdf $p(y(\tau))$ on a timescale τ and a pdf $p_T(y(T))$ on a reference timescale T is computed. As a reference timescale, $T = 1$ s is chosen, which is close to the smallest available timescale in our data sets and on which there are still sufficient events. The independence of the results with respect to the chosen timescale within a certain range is shown in Ref. [13]. In order to be able to compare the shape of the pdfs and to exclude effects due to variations of the mean and variance, all pdfs $p(y(\tau))$ have been normalised to a zero mean and a standard deviation of 1.

As a measure to quantify the distance between the two distributions $p(y(\tau))$ and $p_T(y(T))$, the Kullback–Leibler entropy [14]

$$d_K(\tau) := \int_{-\infty}^{+\infty} \mathrm{d}y\, p(y(\tau)) \ln\left(\frac{p(y(\tau))}{p_T(y(T))}\right) \tag{2}$$

is used. In Fig. 3 the evolution of d_K with increasing τ is illustrated. This quantifies the change of the shape of the pdfs. For different stocks we found that for timescales smaller than about 1 min a linear growth of the distance measure seems to be universally present, see Fig. 3a. If a normalised Gaussian distribution is taken as a reference distribution, the fast deviation from the Gaussian shape in the small-timescale regime becomes evident, as displayed in Fig. 3b. For larger timescales d_K remains approximately constant, indicating a very slow change of the shape of the pdfs, in accordance with Ref. [15]. The independence of this small-scale behaviour of the particular choice of the measure and on the choice of the stock is shown in Ref. [13].

3. Medium scale analysis

Next the behaviour for larger timescales ($\tau > 1$ min) is discussed. We proceed with the idea of a cascade. As it has been shown in Refs. [9,16,17] it is possible to grasp the complexity of financial data by cascade processes running in the variable τ. In particular it has been shown that it is possible to estimate directly from given data a stochastic cascade process in the form of a Fokker–Planck equation [16,17]. The underlying idea of this approach is to access statistics of all orders of the financial data by the general joint n-scale probability densities $p(y_1, \tau_1; y_2, \tau_2; \ldots; y_N, \tau_N)$. Here we use the shorthand notation $y_1 = y(\tau_1)$ and take without loss of generality $\tau_i < \tau_{i+1}$. The smaller log returns $y(\tau_i)$ are nested inside the larger log returns $y(\tau_{i+1})$ with common end point t.

The joint pdfs can be expressed as well by the multiple conditional probability densities $p(y_i, \tau_i | y_{i+1}, \tau_{i+1}; \ldots; y_N, \tau_N)$. This very general n-scale characterisation of a data set, which contains the