Available online at www.sciencedirect.com
ScienceDirect
Physica A 387 (2008) 1218–1224
PHYSICA A
www.elsevier.com/locate/physa

Statistical properties of short term price trends in high frequency stock market data

Paweł Sieczka, Janusz A. Hołyst*

Faculty of Physics and Center of Excellence for Complex Systems Research, Warsaw University of Technology, Koszykowa 75, PL-00-662 Warsaw, Poland

Received 2 April 2007; received in revised form 20 September 2007
Available online 25 October 2007

Abstract

We investigated distributions of short term price trends for high frequency stock market data. A number of trends as a function of their lengths were measured. We found that such a distribution does not fit to the results following from an uncorrelated stochastic process. We proposed a simple model with a memory that gives a qualitative agreement with the real data.

PACS: 89.65.Gh

Keywords: Econophysics; Financial markets; Price trends

1. Introduction

Statistical analysis of stock prices is a rich source of information about the nature of financial markets. It was Louis Bachelier who used a stochastic approach to model financial time series for the first time [1]. Since that time the statistical analysis of stock prices has become a widely investigated area of interdisciplinary researches [2–5].

In 1973, Fischer Black and Myron Scholes published their famous work [6] where they presented a model for pricing European options. They assumed that a price of an asset can be described by a geometric Brownian motion. However, the behaviour of real markets differs from the Brownian property [7,8], since the price returns form a truncated Lévy distribution [9–11]. As a result of this observation many non-Gaussian models were introduced [2–4].

Another divergence from the Gaussian behaviour is an autocorrelation in financial systems. Empirical studies show that the autocorrelation function of the stock market time series decays exponentially with a characteristic time of a few minutes, while the absolute values of the autocorrelation of prices decay more slowly, as a power law function, which leads to a volatility clustering [12–15].

The issue of market memory was also considered by many authors (see references in Refs. [2,3]). It was observed [16], that for certain time scales, a sequence of two positive price changes lead more frequently to a subsequent positive change than a sequence of mixed changes, i.e. the conditional probability $P(+ \setminus ++)$ is larger than $P(+ \setminus +-)$. In this paper we investigated this effect for high frequency stock market data.

* Corresponding author. Tel.: +48 22 234 71 33; fax: +48 22 234 58 08.
E-mail address: jholyst@if.pw.edu.pl (J.A. Hołyst).

0378-4371/$ - see front matter
doi:10.1016/j.physa.2007.10.048

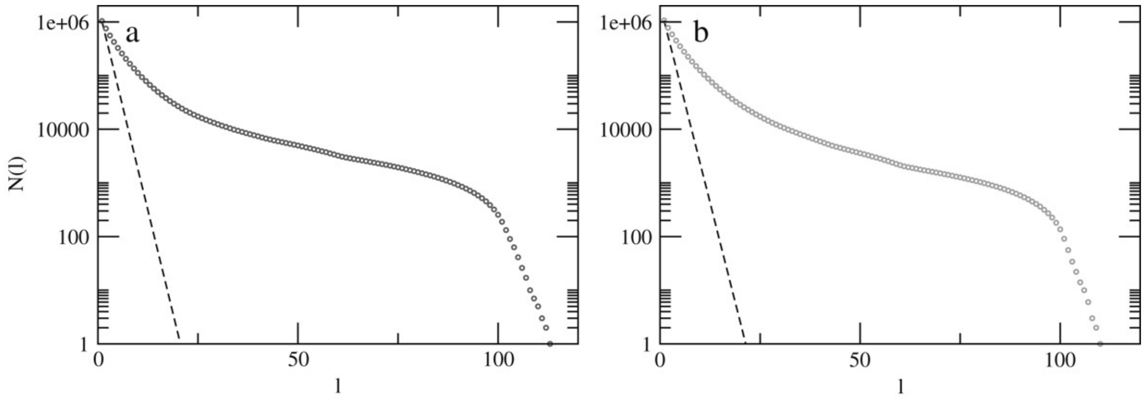

Fig. 1. Distribution $N(l)$ for uptrends (a) and downtrends (b) for the WIG20 index between the 13th June 2003 and the 3rd November 2006. Data are sampled every 15 s (circles), and were compared to uncorrelated process (line) Eq. (1).

2. Empirical data

Let us consider short term price trends for high frequency stock market data. By *short term uptrend/downtrend* we mean such a sequence of prices that a price is larger/smaller than the preceding one (see below for a more precise definition).

First, having a time series Y_t, which is in our case a history of a stock price or a market index, we build a series of variables s_t in the following way:

- $s_t = 1$ if $Y_t > Y_{t-1}$,
- $s_t = -1$ if $Y_t < Y_{t-1}$,
- $s_t = s_{t-1}$ if $Y_t = Y_{t-1}$.

A positive value of the variable s_t means that at time t the price Y_t did not decrease, and similarly a negative value means that the price did not increase.

In a series s_t we can distinguish subseries of identical values. For $a < b$ and $s_a = s_b = s$, $S(a, b, s)$ is such a subseries if and only if $\forall_{c \in (a,b)}\, s_c = s$. Subseries $S(a, b, s)$ can be identified with an uptrend lasting from $t = a$ till $t = b$ for $s = 1$, and with a downtrend for $s = -1$. The length l of such an uptrend/downtrend is equal to $b - a + 1$. Let us mention that a subseries of a length l includes two subseries of length $l - 1$, three subseries of length $l - 2$ etc.

Let $N(l)$ be a number of subseries of a length l with a fixed s in a series $s_1, \ldots, s_M$. If s_t were generated by an uncorrelated discrete stochastic process with a probability $P(s_t = 1) = p$, then the expected value of $N(l)$ would be equal to:

$$N(l) = (M - l + 1)p^l, \tag{1}$$

where M is a number of all elements in the basic series. Similarly the expected value of downtrend series of length l is $N(l) = (M - l + 1)(1 - p)^l$.

We have measured the distribution $N(l)$ for real market data and the same distribution for the corresponding uncorrelated process. Fig. 1 presents this distribution for the WIG20 index of Warsaw Stock Exchange (WSE) between the 13th June 2003 and the 3rd November 2006, and the distribution for the corresponding uncorrelated process. The results show a significant difference between the real data and the uncorrelated model. If variables s_t were uncorrelated, there would not be subseries longer than 25 ticks. In fact, subseries even longer than 100 ticks are present. The trends last for about 30 minutes. There are far more such trends than it would be if the process were uncorrelated. The distribution $N(l)$ was also calculated for particular stocks from WSE, NYSE and NASDAQ (Fig. 2). The stocks from WSE were: Bioton between the 31st March 2005 and the 3rd November 2006, and TPSA between the 17th November 2000 and the 3rd November 2006. The stock from NYSE was Apple, and the stock from NASDAQ was Intel, both between the 4th January 1999 and the 29th December 2000. For the index WIG20 trend periods were measured in real time, but for the stocks they were measured in a transaction time (see Section 4).

The observed difference between the uncorrelated model and the real markets is due to the strong autocorrelations in the process s_t. It is only seen in high frequency data. Choosing every nth element of the series s_t weakens