Statistical analysis of 5 s index data of the Budapest Stock Exchange

Imre M. Jánosi\textsuperscript{a,b,}*; Balázs Janecskó\textsuperscript{a}; Imre Kondor\textsuperscript{a}

\textsuperscript{a}Department of Physics of Complex Systems, Eötvös University, Pázmány Péter sétány 1, H-1117 Budapest, Hungary

\textsuperscript{b}John von Neumann - Institut für Computing (NIC), Forschungszentrum Jülich, D-52425 Jülich, Germany

Received 29 October 1998; received in revised form 11 January 1999

Abstract

A statistical analysis of the Budapest Stock Index (BUX) is presented. The high time resolution (5 s sampling) makes it possible to extract information on market functioning which does not emerge from daily data. The main results are as follows: from a statistical point of view the large drop in October 1997 was a “normal” event. Strong autocorrelation has been detected in the volatility and market activity data. Detrended fluctuation analysis reveals “superdiffusive” scaling without persistence. Finally, we report on a simple method for mapping local trends to represent sequences in order to obtain pattern statistics. © 1999 Elsevier Science B.V. All rights reserved.

\textit{PACS:} 05.40.+j; 05.90.+m

\textit{Keywords:} Financial markets; Stock index; Fluctuations

1. Introduction

Financial markets are being automated in the same way as every other industry, with huge sums being spent and wasted on computer networks all around the world. Prices are recorded several times a minute in liquid financial markets. The large amount of available data and the complexity of market structures has attracted considerable interest from statistical physicists in recent years. Research has focused on detailed statistical analysis of price fluctuations [1–4], modeling markets as complex interactive systems [5–7] and finding analogies between economic and other phenomena such as turbulence [8–10] or biological adaptation [11].

\* Corresponding author.

\textit{E-mail address:} janosi@hercules.elte.hu (I.M. Jánosi)
Understanding price fluctuations in financial markets is important for both derivative pricing and risk analysis [12]. As in other fields, sampling with a high time resolution allows better statistics to be obtained. High-resolution price or index data makes it possible to extract information on market functioning which may remain hidden when using the usual “opening–closing-high–low” data.

The Budapest Stock Exchange (BSE) was opened in 1864, and around the beginning of this century it became the fourth largest stock exchange in Europe. The BSE was reestablished in June 1990. Securities trading is entirely electronic. Since 1 of April 1997 the index (BUX) has been calculated and recorded continuously with a time resolution of 5 s during the main trading hours.

Here we report on a statistical analysis of 5 s BUX data covering a period of approximately one calendar year. Scaling properties, correlation functions and Fourier spectra has already been evaluated by Rotyis and Vattay [13] for the first third of the same data set. Here we analyze other aspects, namely the “crash” in October 1997, market activity measurement, scaling properties of detrended fluctuations, and pattern statistics, by means of a simple symbolic dynamics representation.

2. Draw down in October 1997

In Fig. 1 we show the evolution of the BUX index as a function of trading time in the period from 1 April 1997 to 10 April 1998 excluding the intervals when the market was closed. The data set consists of 296 390 records.

Stock indices can exhibit very large motion. Public policy, interest rates, economic and social conditions influence market behavior globally, especially in small and emerging markets such as the BSE. The large drop in October 1997, apparent in Fig. 1, is commonly attributed to the so-called “South-East Asian financial crisis”. In a recent paper [14], Johansen and Sornette found that the five or six largest crashes observed in this century in the Dow Jones Industrial Average are probably outliers, i.e. they do not belong to the fat tail of a continuous distribution function of index variations. The threshold value they obtained is \( \approx 15\% \), a draw-down exceeding this value can be considered as an extreme event outside the “normal” market functioning [14]. This also implies that outliers are probably triggered off by additional amplifying factors, and specific signatures of their presence could exist, as proposed in e.g. [15,16]. The draw-down of 14.76% on 28 of October 1997 in Budapest is very close to this threshold, therefore we can ask whether it is an outlier or not.

Since the statistics for large index jumps are rather poor, we determined the probability of a percentage index variation \( \Delta x_t \) at a given instant \( t \) larger than \( \Delta x \). This relates to the probability density function \( P(\Delta x) \) as

\[
P(\Delta x_t > \Delta x) = \int_{\Delta x}^{\infty} P(\Delta x') \, d\Delta x' .
\]
Fig. 1. BUX values (a) and changes of the index (b) as a function of trading time between 1.4.1997–10.4.1998.

(For negative changes, the absolute values are considered.) As Fig. 2 shows, this probability has the scaling form

\[ P(\Delta x_t > \Delta x) \sim \Delta x^{-\alpha} \]

with \( \alpha \approx 1.55 \pm 0.08 \) for both positive and negative changes. The fit is not perfect (as indicated by the dotted lines), nevertheless, even the points of largest \( \Delta x \) values are not far from the curves. Extraordinarily large probabilities for large index changes would be reflected by an apparent flattening of the curves. The smooth shape of \( P(\Delta x_t > \Delta x) \) suggests that BUX variations close to 15% may belong to a continuous distribution function, similar to the DJIA [14]. Thus the October “crash” can be considered as a normal market event, at least from a statistical point of view.

Finally, it should be noted that the drop of 14.76% was not a 5-s event, it reflects the difference between the opening BUX value on 28 October and the closing value of the preceding day.
Fig. 2. The probability of an absolute index change larger than $\Delta x$ as a function of $\Delta x$ for the BUX data shown in Fig. 1. Circles and diamonds indicate positive and negative percentage variations, respectively, the latter is shifted downward for clarity. The bin size was 0.05%, altogether 100 057 cases are considered. The dotted lines are power law fits for different parts of the empirical curves. Note that the statistics is obtained over tick by tick changes, thus time (over-night gaps or intervals without index change) is not considered here.

3. Market activity

Definitions of market activity are connected with quantities such as traded volume, volatility, or frequency of transactions. Unfortunately volume data are scarce, therefore activity is usually characterized by means of the mean fluctuation of a market price over a certain time interval $\Delta t$, the volatility. To avoid the problem of changing scale (e.g. inflation, devaluation, etc.), we evaluate the logarithmic increment

$$S(t, \Delta t) = \ln[x(t + \Delta t)] - \ln[x(t)], \quad (3)$$

as usual [12]. The volatility is the average absolute value of $S(t, \Delta t)$ over a time window $T = n \Delta t$ with an integer $n$:

$$V_T(t) = \frac{1}{n} \sum_{k=0}^{n-1} |S(t_k, \Delta t)|. \quad (4)$$

Another simple measure of market activity is the frequency of index changes, which is simply

$$C_T(t) = \frac{\text{number of the index changed}}{\text{number of the index measured}}. \quad (5)$$
calculated in a time window $T$. Obviously, this measure is meaningful only if the
index is determined with a much higher time resolution than the average time between
transactions. In Fig. 3 the volatility $V_T(t)$ and the “activity” $C_T(t)$ are plotted as a
function of trading time, both are calculated with a $T = 5$ min window.

Both $V_T$ and $C_T$ are measures of market activity, but are of different kinds. While
volatility sometimes reflects huge price variations (the October crash is clearly seen
in Fig. 3a as well), the frequency of index changes does not show similarly erratic
fluctuations. On the other hand, the volatility is totally independent of the liquidity of
the market. The frequency of transactions is a better characteristic of liquidity, at least
in the absence of volume data. The white line in Fig. 3b illustrates a slow linear trend,
which correlates with the increasing liquidity of the BSE in the given period.

It is well known that, unlike price changes, the volatility has a strong autocorrelation
[12,17–19]. The normalized autocorrelation function of a given time series $y(t)$ is
Fig. 4. Autocorrelation function for (a) the volatility $V_T(t)$ and (b) the activity $C_T(t)$, calculated by Eq. (6).

In the case (b), the time series is detrended by the line shown in Fig. 3.

The autocorrelation function is given by

$$A\{y(t)\}(\tau) = \frac{\langle y(t+\tau)y(t) \rangle}{\langle y^2(t) \rangle}$$

where $\tau$ is a time lag measured in units of sampling time $t$. Fig. 4 shows the autocorrelation for both the volatility $V_T$ and activity $C_T$.

Two features of the autocorrelations are apparent in Fig. 4. Firstly, the overall correlation decays slowly, both the volatility and the activity have a “memory” of several trading hours. Secondly, there is strong periodicity, suggesting a marked intraday pattern. The characteristic time $t_c \approx 100$ min covers the sampled interval of a day, 5 s BUX calculation running approximately from 11:35 to 13:15 each day. As for Fig. 4a, the sharp periodicity is a consequence of the short trading “day” of the BSE. The London Stock Exchange, where half of the stocks determining the BUX are also traded, opens earlier and closes later than the BSE. Since stock prices suffer changes in London (some of them in New York and Frankfurt too), the opening BUX value reflects the variations in large jumps compared to the last closing
value. The details of intraday and other patterns will be analysed in a subsequent paper [20].

It is clear from Fig. 3b that the sampling time of 5 s is small enough to obtain a reasonable value for the activity $C_T(t)$. Very often no deal is made within 5 s, thus the index remains unchanged. In Fig. 5 we show the histogram of the length of silent periods $\Delta l$ on a double logarithmic scale. The smooth part of the histogram can be well fitted with a parabola

$$\ln[P(\Delta l)] = p_0 - p_1 \ln(\Delta l) - p_2 [\ln(\Delta l)]^2,$$

the constants are $p_0 = 10.57$, $p_1 = 0.2367$, and $p_2 = 0.6327$. This histogram is characteristic of the so-called parabolic fractal distributions [21]. The most remarkable property of this parabolic fractal distribution is the existence of a finite compact support [22], which contrasts with other distributions that are defined for arbitrarily large arguments. If we assume that a smooth parabolic fractal distribution characterizes the silent periods, breaks without any deal longer than approximately 5 min could be ascribed to market disturbances. The large breaks (the largest being approx. 77 min on 18 November 1997), clearly out of the tail, are signatures of market regulations: trade is suspended for a while after “shocking” jumps of individual stock prices or index values.

The empirical parabolic fractal distribution has been introduced to account for the existence of curvature in a log–log plot [22]. The properties of a stochastic process obeying such a distribution is not analyzed in details yet. Nevertheless, the hypothesis of a pure Poisson process for trades gives a slightly worse fit, as shown with the
dashed line in Fig. 5. A possible explanation may be based on the analysis of subtle correlations in the trading processes.

4. Detrended fluctuation analysis

The time series of BUX values are apparently far from being stationary (see Fig. 1a). Clear trends are visible for shorter and longer periods, but these local trends change rather erratically. Detrended fluctuation analysis (DFA), developed originally to investigate correlations along DNA sequences [23,24], aims to overcome the difficulties of irregular local trends. The usefulness of the method is illustrated in analyzing USD/DEM time series, too [25].

The simplest realization of linear detrending is illustrated in Fig. 6 on a piece of the data set in Fig. 1. The key parameter of the method is the length $L$ of non-overlapping windows, $m$ of which cover the original data set of length $n = mL$. The local linear trend is obtained by least-square fits

$$x(t) = s_i t + r_i, \quad i = 1, \ldots, m$$

for each window. Then a detrended time series $d(t)$ is produced by subtracting the piecewise trend function from the original data set.

In order to characterize the fluctuations of the detrended index, we determine three quantities as a function of window length $L$. Firstly, the average spread $R(L)$ is defined by the difference between the maximum and minimum at a given window:

$$R(L) = \frac{1}{m} \sum_{i=1}^{m} \max\{d(t)\}_i - \min\{d(t)\}_i.$$
Fig. 7. Average spread \( R(L) \) [Eq. (9)], average standard deviation \( \sigma(L) \) [Eq. (10)], and average detrended volatility \( f(L) \) [Eq. (11)] as a function of window size \( L \) for the BUX data of Fig. 1. The vertical dotted line indicates the border of scaling regimes I and II (see text); heavy dashed lines show power law fits.

Secondly, the average variance \( \sigma^2(L) \) is simply

\[
\sigma^2(L) = \frac{1}{m} \sum_{i=1}^{m} \left( \frac{1}{L} \sum_{j=1}^{L} d_i^2(t) \right). \tag{10}
\]

(Note that the \( i \)th piece of the detrended time series \( d_i(t) \) has a zero local average by definition.) Thirdly, the average absolute fluctuation or “detrended volatility” \( f(L) \) is given by

\[
f(L) = \frac{1}{m} \sum_{i=1}^{m} \left( \frac{1}{L} \sum_{j=1}^{L} |d_i(t)| \right). \tag{11}
\]

These quantities are calculated for the BUX and plotted in Fig. 7 on a double logarithmic scale.

All of these quantities show scaling of the form \( \sim L^x \) over a wide range. A closer inspection reveals that the curves have a breakpoint at around \( L_b \approx 100 \) min denoted in Fig. 7 by a vertical dotted line. Recall that this is the length of the record for one trading day. The scaling exponents are different for the two regimes:

\[
x_I = 0.690 \pm 0.015, \quad x_{II} = 0.566 \pm 0.031, \tag{12}
\]
indicating again the presence of an intraday pattern observed in volatility and market activity (see Fig. 4). Random processes with independent increments typically have an exponent $\alpha = \frac{1}{2}$ [26]. An exponent value $\alpha > \frac{1}{2}$ is usually associated with persistence, i.e. strong correlations between consecutive increments [26]. However, one should be careful with this interpretation for price changes. The quantities $R(L)$, $\sigma(L)$, or $f(L)$ characterize average fluctuations, the larger their value, the wider the range “visited” by the time series. There are two basically different ways of covering a given range with a fluctuating signal. Firstly, if the size of individual steps is limited, a wide range can be visited only with a strong correlation between sequential steps. A typical example of such a process is water level fluctuations in rivers [27], where the elementary steps of increase or decrease are limited for obvious reasons. Correlation between consecutive steps is not necessary for visiting a wide range, if the size of elementary steps is not limited, as in the case of e.g. Lévy flight [28]. Since the distribution of short time price and index variations is known to be non-Gaussian with fat tails [12], we think that the second interpretation explains the observed scaling better in Fig. 7.

We show the autocorrelation function Eq. (6) for the BUX variations in Fig. 8. It is clear that persistence is not present here, rather a slight anticorrelation for very short times is visible. As further illustration, we computed the histogram for step length of

![Autocorrelation function](image)
monotonous index changes, as shown in Fig. 8-Inset. The histograms can be well fitted with the exponential distributions

\[ P(l^-) \sim e^{-1.68 l^-}, \quad P(l^+) \sim e^{-1.48 l^+}. \]  

(13)

In order to quantify the speed of this exponential decay, let us assume that a monotonous sub-series is detected in which the index increased in 5 consecutive steps, an event of probability \( P(l^+ = 5) \approx 0.0016 \) itself. (Note that time is not considered here, it can be different between the steps.) The probability of the next step continuing the increasing trend and our getting a monotonous series of length \( l^+ = 6 \) can be estimated as \( P(5|6) = \exp(-1.48) \approx 0.23 \). This contrasts with a random walk of independent increments, where the same probabilities are \( P(5) = (1/2)^5 \approx 0.031 \) and \( P(5|6) = 1/2 \), by definition. Thus we can conclude that persistence does not hold for the BUX data.

On the contrary, what we found is rather anti-persistence considering correlations between consecutive index increments. Nevertheless, a scaling exponent \( \alpha > 1/2 \) is possible, because the distribution of index variations is very wide [13].

5. Pattern statistics with DFA

Many speculators try to predict stock prices or index values by means of so-called “technical analysis” [29]. The basic assumption behind technical analysis is that patterns of human behavior (buyers and sellers) are repetitive in given situations, resulting in price patterns which are sufficiently characteristic to allow prediction of future price changes. Magic terms like “diagonal triangles”, “head & shoulder”, “saucers” and “inverse saucers”, “island reversals”, etc., are used to identify patterns in technical charts. Since most of these patterns are constructed from trendlines of different slopes and lengths, we can try to utilize DFA in pattern statistics.

As a first approximation, we neglect the slope of a local trend and evaluate only the sign of it. Recall that the key parameter of linear DFA is the length \( L \) of non-overlapping time windows covering the time series. A pattern of length \( Z \) can be formed from the signs of consecutive local slopes, e.g. ‘+ − + + − − +’. Note that a third obvious possibility is a zero local slope. However, if the length \( L \) is large enough, a local trend of exactly zero slope has a negligible probability. We can easily estimate the necessary window size by means of the distribution of silent periods shown in Fig. 5. A choice of \( L > 6 \) minutes for the BUX data satisfies this condition, then each of the patterns can easily be mapped to a binary number.

The parameter space \( \{L; Z\} \) of binary patterns is huge, here we show only representative examples for pattern statistics. In Fig. 9 relative probabilities are plotted for \( Z = 3, 4, 5 \) and \( L = 6, 12, 30 \) min. The probabilities of different patterns have very irregular distributions for each \( (Z, L) \) parameter pair. However, the statistics plotted in Fig. 9 represent a single realization of local linear fits starting from the first data point of our time series. Obviously, another realization from a shifted initial point can give different local slopes, and thus different pattern statistics. This
is a first test of statistical significance. We calculated pattern statistics for a given \((Z, L)\) parameter pair beginning the local fits from \(Z \cdot L\) initial points shifted gradually by one step. The average and standard deviation of pattern probabilities were obtained. As an example, we show the result for two parameter pairs in Fig. 10.

Obviously, the statistics are much better for patterns of short local trends. For \(L = 72\) (Fig. 10a), the two homogeneous patterns, namely \(\text{"---"}\) (code 0) and \(\text{"+++++"}\) (code 31) have maximal probabilities. This indicates the presence of quiet periods, when the index is sweeping more or less monotonously, for at least half an hour. The small local maximum at 15 (\(\text{"++++-"}\) and 16 (\(\text{"---++"}\)) reflects the same fact. As for longer period trends such as 30 min in Fig. 10b, the error bars overlap fully, indicating the lack of clear patterns.
Fig. 10. Average probabilities of different sign patterns of local trends for (a) $Z = 5$ and $L = 72$ (6 min), and (b) $Z = 5$ and $L = 360$ (30 min). The averages and standard deviations are obtained from $ZL$ initial points shifted step by step by 1 unit (5 s).

Note here that there is no contradiction between the results of Figs. 8 and 10a. Trend patterns allow fluctuations of the instantaneous index value around the trend lines, thus monotonous index changes (without any reverse steps) can be less probable than in the random walk case.

The situation is very similar for other $Z$ and $L$ parameters. Short time local fits give higher probabilities for patterns of the same signs. Preferred patterns cannot be extracted from local fits of wider time windows. Nevertheless, a direct comparison with the efficiency of technical analysis is not possible, because technical charts are based on daily data, and the pieces forming a given “magic” pattern usually have different lengths and slopes.

Acknowledgements

This work was partially supported by the Hungarian National Science Foundation (OTKA) under Grant No. F014967. I.M.J. thanks the Foundation for Research and
Higher Education for a Zoltán Magyary scholarship. Special thanks are due to the Budapest Stock Exchange for access to the BUX data.

References