On the Distribution of Stock Market Data

V.V. Ivanov and P.V. Zrelov
Laboratory of Information Technologies,
Joint Institute for Nuclear Research

1. Introduction. The time series originating from the stock market have a very complicated character due to a large number of factors influencing the underlying processes. The formulation of a realistic mathematical model capturing the salient features of the underlying process will serve as a basis for the development of effective mathematical tools for processing of stock market data and decision making. The review of the history of the stock prices data investigations can be found in [1, 2].

We analyze here the statistical relations between prices and corresponding traded volumes - for some stock market data the statistical distribution of closing prices normalized by the corresponding traded volumes, fits well the log-normal law. For some other stocks and for traded volumes the log-normal law is obtained after application of a de-trending procedure. Different schemes for the trend determination are considered.

2. Selection of data sets. We investigate the closing prices and traded volumes of the US stock markets: IBM, Xerox, Boeing, Lockheed Martin, Delta, Airlines, Exxon Mobil, Texaco, Ford, General Motors, General Electric, Reebok, Merk, Bank of America (New York Stock Exchange - NYSE) and Microsoft, Intel, Oracle, Dell, Sun (Nasdaq Stock Market). In addition, we analyze closing prices and traded volumes for some stocks in the EU markets: DaimlerChrysler (Exchange Electronic Trading - XETRA, Germany), Solvay and Arbed (Brussels Stock Exchange, Belgium). We also study different indices of the US stock market: S&P500, Dow Jones Industrial Average (DJIA), NYSE Composite, NASDAQ Composite, etc.

3. Analysis of original stock market data. It is very important to find a stochastic variable which is a function of the original, non-transformed stock market data, that forms a stable statistical distribution. Any transformation of raw data may destroy the main features of the assumed underlying dynamical system. In this connection, we propose a new indicator based on daily closing prices and corresponding traded volumes.

3.1 Daily closing prices, traded volumes and their ratio. The time series of daily closing prices and traded volumes for most companies demonstrate the increasing trends. Aiming to minimize the influence of the trend, we introduce a new stochastic variable Ξ, which is the closing price normalized by the corresponding traded volume ("Price/Volume"). The variable Ξ has a relatively stable mean value and dispersion, compared to the original time series of closing prices and traded volumes. The statistical distributions of Ξ are presented in the right plot of Fig. 1. They follow the log-normal law (in a generalized form):

$$f(x) = b + \frac{A}{\sqrt{2\pi}\sigma(x-c)} \exp^{-\frac{1}{2\sigma^2}(\ln(x-c)-\mu)^2}, \tag{1}$$

where σ is the dispersion, μ is the mean value, and A is a normalizing factor, b and c - additional parameters.

For the Boeing company: $\chi^2/\text{ndf} = 91.45/94 \simeq 0.973$ (significance level $\alpha \simeq 55\%$), $\mu = -10.64$, $\sigma = 0.526$, $c = -0.2708 \cdot 10^{-5}$, $A = 0.7984 \cdot 10^{-2}$, $b = 0$.

The Boeing example is typical for most companies under consideration. It is very surprising that such sufficiently stable results are obtained directly from the original data.

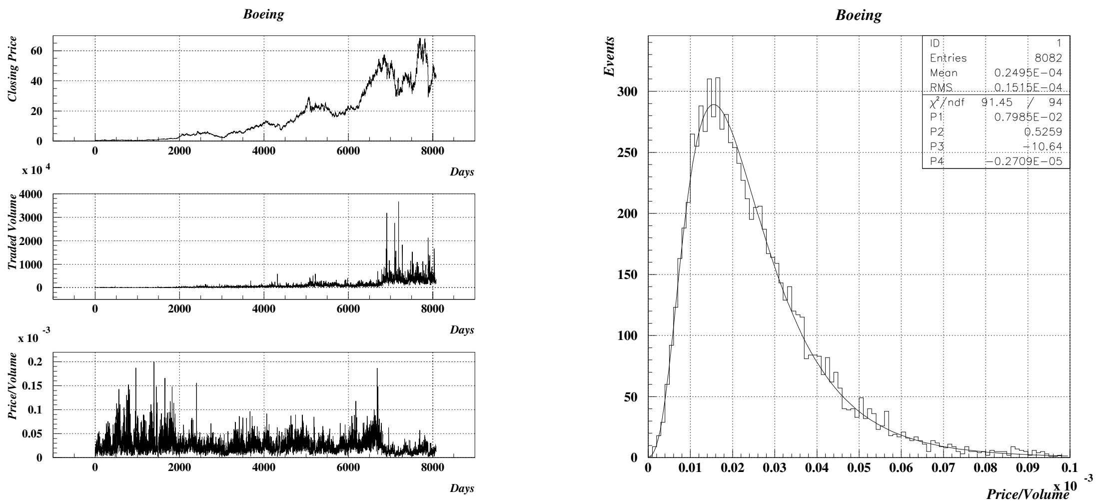

Figure 1: Left plots (from top to bottom): a) Boeing daily closing prices, b) daily traded volumes, c) daily closing prices normalized by corresponding traded volumes. Right plot: the distribution of daily close prices normalized by corresponding traded volumes for Boeing stocks and its approximation by function (1)

There are exceptions for some data sets. For example the distributions for Exxon Mobil and Bank of America stocks have heavy long tails which do not fit a single log-normal distribution. However, they are approximated quite well by a weighted sum of two log-normal distributions. This may mean either a difference in dynamics, as compared to other stocks, or some drawbacks of application of the "Price/Volume" indicator Ξ directly for raw data.

Taking into account all the above mentioned exceptions, we expect to get better results when standard methods of time series analysis are applied to original data.

3.2 Stock market indices. The statistical analysis of the index data shows that these series have a more complicated character, compared to stocks. As an example, in Fig. 2 we present the results for the NASDAQ index: left plots show the daily closing prices (top plot) and the logarithm of daily closing prices normalized by traded volumes (bottom plot). The later series clearly demonstrates the decreasing trend. The right plot shows the distribution of daily closing prices normalized by traded volumes superimposed by the fitting curve of the weighted sum of two log-normal functions. One can clearly see that the analyzed distribution qualitatively fits in the sum of two log-normal distributions. Similar results are obtained for the DJIA, NYSE Composite and S&P500 indices.

4. Analysis of de-trended data. The application of standard methods of time series analysis may improve the situation concerning the data sets which do not follow the log-normal form. A first step in this direction is the elimination of a trend component from the analyzed original data.

The form of the decomposition of the analyzed time series, described by the deterministic (d) and stochastic (ϵ) terms, can be different. We shall employ a multiplicative model $x_i = d_i \times \epsilon_i, i = 1, 2, ..., n$ in our analysis, because it ensures a more "homogeneous" character of the transformed series.

Our comparative analysis (see details in [2]) of the schemes based on the Principal Component Analysis ("Caterpillar" approach [3]) and the wavelet approximation for trend determination has shown that both of them lead to close results. However, the

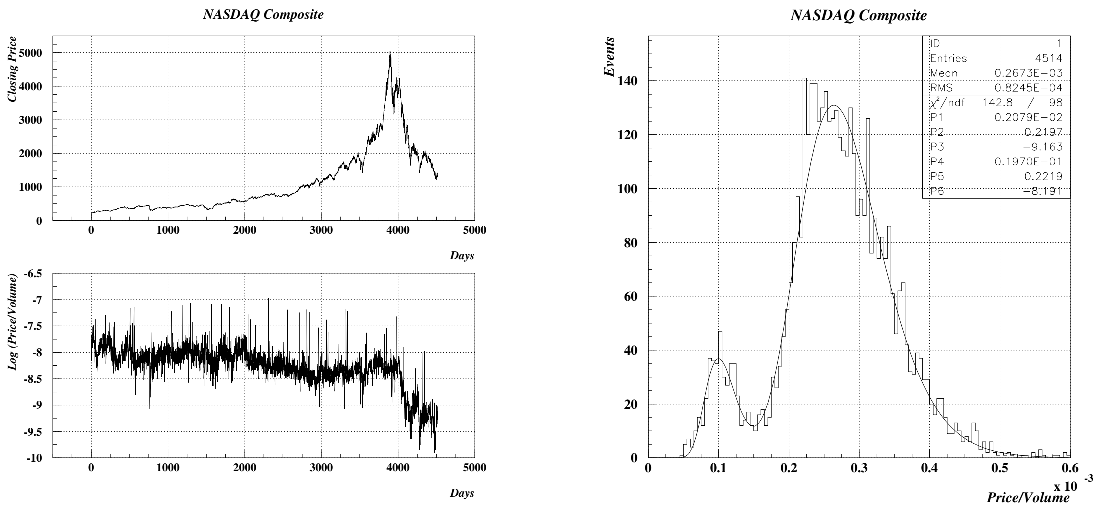

Figure 2: Left plot (from top to bottom): a) daily closing prices for NASDAQ Composite index, b) the logarithm of daily closing prices normalized by traded volumes. Right plot: the distribution of daily closing prices normalized by traded volumes superimposed by the weighted sum of two log-normal functions

"Caterpillar" algorithm provides additional and very useful qualitative information - the value of the contribution of the chosen leading components into the analyzed series.

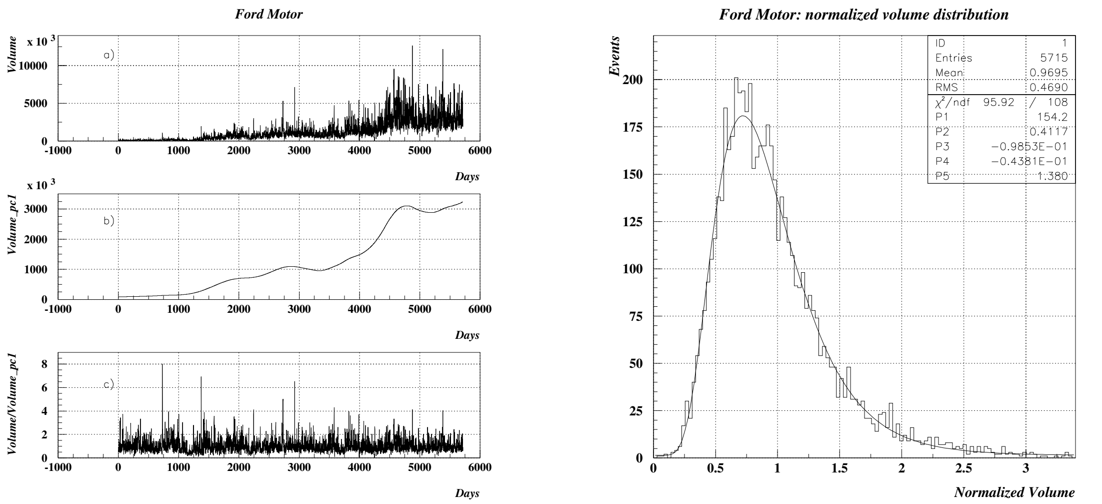

Figure 3: Left plots (from top to bottom): a) Ford Motor daily volumes, b) the trend term determined by the "Caterpillar" method with $L = 480$, c) Ford daily volumes normalized by trend. Right plot: the distribution of Ford Motor daily volumes normalized by trend and its approximation by function (1)

4.1 Statistical analysis of de-trended volumes. Fig. 3 shows the results of de-trending the time series of traded volumes for Ford stocks (L is the so-called "caterpillar length" - parameter of the method [3]). The statistical distributions of the de-trended volumes (the volumes normalized by trend), superimposed by the fitting curves corresponding to the log-normal functions, are shown in the right plots. These plots demonstrate a very high level of correspondence of the analyzed statistical distributions to the log-normal law.

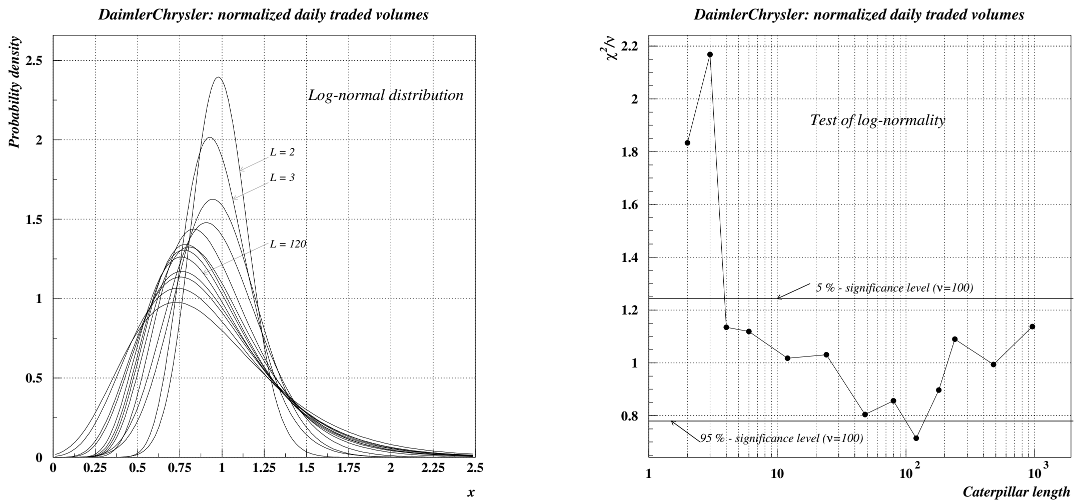

Figure 4: Left plot: log-normal densities approximating the distribution of DaimlerChrysler daily volumes normalized by trend determined using the "Caterpillar" method with $L = 2$, 3, 4, 6, 12, 24, 48, 80, 120, 180, 240, 480 and 960. Right plot: the dependence of χ^2/ν versus L, describing the quality of approximation of normalized DaimlerChrysler daily volumes by the log-normal ditribution

The presented results are typical for all analyzed stock market data. The choice of L value is controlled by the χ^2-test. It can be illustrated by the following example. In Fig. 4 the left plot demonstrates the density functions of the log-normal distributions that approximate the DaimlerChrysler daily volumes normalized by the trend calculated applying the PCA method with different caterpillar lengths: $L = 2$, 3, 4, 6, 12, 24, 48, 80, 120, 180, 240, 480 and 960. The right plot in Fig. 4 shows a quality of the approximation of this statistical distribution by the log-normal function applying the χ^2-test.

In this connection, we may conclude that the above described procedure permits one to establish another important property of the stock market data - the distributions of traded volumes normalized by their trends fit in the log-normal form with a reliable accuracy.

4.2 Statistical analysis of de-trended closing prices. A Similar de-trending procedure applying to the closing prices data leads to Gaussian like distributions. The interval of acceptable Caterpillar lengths L can be established by means of the χ^2-test together with ω_n^2 criterion [4] (test of symmetry and test of Gaussianity) (see detailed analysis in [2]).

4.3 Double normalization scheme. We analyze the statistical distributions of a new variable – the ratio of de-trended prices on de-trended volumes, which we denote $\tilde{\Xi}$. Fig. 5 shows the results of this analysis for the Exxon Mobil stock market data. The left plot presents the time series of the de-trended daily closing prices, the de-trended volumes and their ratio. The right plots demonstrate the statistical distributions corresponding to the new variable $\tilde{\Xi}$ superimposed by the fitting function (1). These results are typical almost for all analyzed data.

Thus, this double normalization scheme allows one to establish a third important property of the stock market data - the distributions of the de-trended close prices normalized by the de-trended volumes fit well the log-normal form (1).

5. Conclusion. The statistical analysis of a wide spectrum of stock market data has demonstrated that for most stocks the distribution of the closing prices normalized by

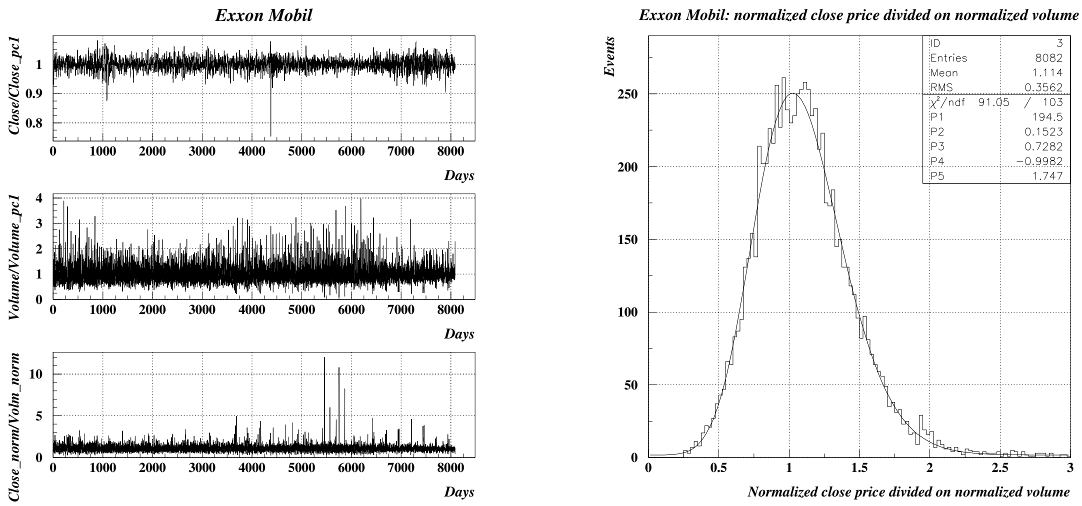

Figure 5: Left plots (from top to bottom): a) Exxon Mobil daily closing prices normalized by trend, b) Exxon Mobil volumes normalized by trend, c) ratio of normalized closing prices and normalized volumes. Right plot: the distribution of ratio values and its approximation by function (1)

traded volumes (Ξ) fits well the log-normal function (1). The stocks of some companies are of more complicated character – such data need preliminary application of the de-trending procedure. We have found that the distributions of traded volumes normalized by their trends fit the log-normal function (1) with a reliable accuracy.

For some stocks there still exists a noticeable disagreement between the distribution of the variable Ξ ("Price/Volume") and the log-normal distribution. In most cases this disagreement could be overcome by applying the double normalization scheme – the usage of the variable $\tilde{\Xi}$ (ratio of de-trended prices on de-trended volumes).

Besides, the statistical distributions of a few stocks and for all analyzed (by us) market indices have a significantly more complicated character and can not be approximated by a single log-normal function. On the other hand, they are well approximated by a weighted sum of two log-normal distributions. The latter may mean that the structural transitions in the dynamics of the analyzed series take place.

References

[1] J.C. Hull: *Options, Futures and Other Derivatives*, Third edition, Prentice-Hall, Inc., 1997.

[2] I. Antoniou, Vi.V. Ivanov, Va.V. Ivanov and P.V. Zrelov: *On a Log-Normal Distribution of Stock Market Data* , Physica A 331 (2004) 617-638.

[3] N. Golyandina, V. Nekrutkin, and A. Zhigljavsky: *Analysis of time series structure: SSA and related techniques*, Chapman & Hall/CRC, 2001.

[4] G.V. Martinov: *Omega-squared criteria*, Moscow, "Nauka", 1978 (in Russian).