УДК 658.0

L. Timashova I. Skachko
International Research and Training Centre for Information Technologies and Systems

FINANCIAL MARKET FORECASTING USING THE FRACTAL ANALYSIS

Розглянуто проблему прогнозування фінансового ринку, якому властива довгострокова пам'ять. Використовується метод фрактального аналізу для виявлення фундаментальних характеристик за допомогою алгоритму R / S-аналізу. Відповідно до алгоритму розроблено програмний продукт, що дає змогу виявити і числово оцінити фундаментальні характеристики часових рядів, такі як наявність та глибина довготривалої пам'яті, трендостійкість (персистентність) тощо.

Ключові слова: фінансовий ринок, фрактал, алгоритм R / S-аналізу, тренд.

In this paper, we examine the problem of forecasting of the financial market where long-term memory occurs. We apply the fractal analysis method to identify some fundamental characteristics. The basic element here is the R/S-analysis algorithm. Based on this algorithm, a software product was designed that allows identifying and computing numerically the fundamental characteristics of time series such as long-term memory availability and depth, stability of the trend (persistency) etc.

Key words: financial market, fractal, R/S-analysis algorithm, trend.

Introduction

Today financial markets arouse a great deal of public interest. Regular traders, analysts who work for global corporations and state agencies all deal with financial analytics.

There are many ways to analyze the events taking place in the markets. They include technical analysis, fundamental analysis, the Elliott Wave Principle and many other less known methods. However, one method is uniquely placed among all others due to its simplicity and originality. It is called fractal analysis.

1. The analysis of recent research and purpose of papers

The term fractal was coined by Benot Mandelbrot in 1975. The concept of fractals is that they have a large degree of self similarity within themselves. The fractal dimension D of a profile or surface is a measure of roughness, with D [n,n+1) for a surface in n-dimensional space and higher values indicating rougher surfaces. Long-memory dependence or persistence in time series or spatial data is associated with power-law correlations and often referred to as Hurst effect. Scientists in diverse fields observed empirically that correlations between observations that are far apart in time or space decay much slower than would be expected from classical stochastic models. Long-memory dependence is characterized by the Herst coefficient, H. In principle, fractal dimension and Hurst coefficient are independent of each other: fractal dimension is a local property, and long-memory dependence is a global characteristic. The basic tool for fractal analysis of time series is the algorithm of R/S analysis. The methodology of the R/S analysis was developed in the middle of the XX century by the British hydrologist Hurst, who was studying the time series of river flow volumes. While testing the hypothesis that these series were governed by the normal law, Hurst defined a new statistic - the Hurst exponent (H). In the course of his research, Hurst measured fluctuations of water in the reservoir relative to the average over time and introduced the dimensionless ratio by dividing the amplitude of R by the standard deviation S. This method of analysis was called the rescaled range method (R/S-analysis). Hurst found that most natural phenomena, including river flows, temperatures, precipitation, sun spots follow a “biased random walk” - a trend with noise. The strength of the trend and the noise level can be measured by how the rescaled range changes over time, or in other words, by how the H exponent exceeds the value of 0,5.

We describe the algorithm of the R/S-analysis as it is implemented in the modern methods of fractal analysis [1,2]. Given a time series:

$$Z = \{z_i\} \; i = 1,2, \dots n \,, \qquad (1)$$

where we consistently highlight its initial segments:

$$Z_\tau = z_1, z_2, \dots, z_\tau, \text{where } \tau = 3,4, \dots, n$$

for each we calculate the current average:

$$\overline{z_\tau} = \frac{1}{\tau}\sum_{i=1}^{\tau} z_i$$

Then for each fixed Z_τ, $\tau = 3,4, \dots, n$, we calculate the accumulated deviation for its segments of length t:

$$X_{\tau,t} = \sum_{i=1}^{t}(z_i - \overline{z_\tau}) \,, \text{where } t = \overline{1,\tau}.$$

After that, we calculate the difference between the maximum and minimum accumulated deviations:

$$R = R(\tau) = \max_{1<t<\tau}(X_{\tau,t}) - \min_{1<t<\tau} X_{\tau,t},$$

which is usually defined by the term «range of R». This range is rescaled, i.e. represented as a fraction $\frac{R}{S}$, where $S = S(\tau) = \sqrt{\frac{1}{\tau}\sum_{j=1}^{\tau}(z_j - \overline{z_\tau})^2}$ - is a standard deviation for the time series segment $Z_\tau, 3 \le \tau \le n$, . The Hurst exponent $H = H(\tau)$ that characterizes the fractal dimension of the time series in question and the corresponding noise color is derived from the following equation $\frac{R}{S} = (\alpha\tau)^H$[1]. By taking the logarithm of both sides of this equation and assuming that $= \frac{1}{2}$, we get the Cartesian coordinates(x_τ, y_τ), the H–trajectory points, whose ordinates and abscissas are respectively:

$$y_\tau = H(\tau) = \frac{\log(R(\tau)/S(\tau))}{\log(\tau/2)} \,, \;\; x_\tau = \tau. \qquad (2)$$

The R/S-trajectory required for the fractal analysis (1) is presented in logarithmic Cartesian coordinates as a sequence of points, whose abscissas are $x_\tau = \log(\tau/2)$, and the ordinates are $y_\tau = \log(R(\tau)/S(\tau))$. By connecting the neighboring points (x_τ, y_τ) and $(x_{\tau+1}, y_{\tau+1})$, where $\tau = 3,4, \dots, n-1$ with a line segment, we get a graphical description of the R/S-trajectory / (H-trajectory) in logarithmic coordinates (in Cartesian coordinates). One of the main fractal characteristics of a time series is noise color, which corresponds to this series at one or another point in time. The values of $H \ge 0,6$ define a black noise. The higher the H value is, the greater stability of the trend this particular segment of the time series has.

The H values within the range of $\sim 0,5 \pm 0,1$ define a white noise, which is characterized by the "chaotic behavior of a time series," and therefore, implies the lowest accuracy and reliability of the forecast.

The H values within the range of $\sim 0,3 \pm 0,1$ define a pink noise. The pink noise tells us that the segment of the time series in question is characterized by anti-persistency, meaning the time series reverses more often than a random series.

As for the occurrence of long-term memory of the time series in question (1), it is impossible to make a definite conclusion, unless its H-trajectory stays in the black noise area for a long time and if the behavior of the R/S-trajectory is chaotic, starting from its initial points.

The basis for claiming that the time series (1) has long-term memory is the fulfillment of the following conditions:

1. The *H*-trajectory through some of its initial points goes to the black noise area, and for the R/S-trajectory, the above-mentioned "black noise" entry points demonstrate that there is a trend showing. The number, for which the following condition is met, determines the depth of this memory: at the point the trajectory gets decremented, the R/S-trajectory at this point demonstrates a sudden (dramatic) change in the trend.

2. If we randomly shuffle the elements of this time series and then present the resulting series to the input of the R/S-analysis algorithm, the output maximum value of the Hurst exponent and R/S-trajectory will be much lower when compared to the values for the initial time series, should this time series have long-term memory.

2. The main material of research

A software product was prepared for carrying out the R/S analysis. The system allows conducting a fractal analysis of the dynamic data and creating a forecast of the ARFIMA-class model, based on the previously obtained results. As an example of a time series, we will examine the dynamics of the Microsoft (MSFT) stock. The time series being used in this work is a sequential daily sample (of the n volume) for the period from January 3, 2005 to November 20, 2015 of the market statistics. Each element of the time series corresponds to the trading result for this specific financial instrument over one trading day. For each time series $Z \in \{Z_1, Z_2, ..., Z_\tau\}$ a sequential R/S-analysis was conducted. As a result, the Hurst exponent ($H(\tau)$) was calculated for each sequential segment of the time series Z of the length τ, and the H – and R/S-trajectories of the corresponding time series were built.

The figures below show the H and R/S – trajectories at the output of the R/S-analysis. For the H – trajectory diagrams, there are segments of length of the τ series along the abscissa axis. For the R/S – trajectory diagrams, there are values $\ln\frac{\tau}{2}$ along the abscissa axis.

Fig. 1. MSFT behavior of the prices from 03.01.2005 to 20.11.2015

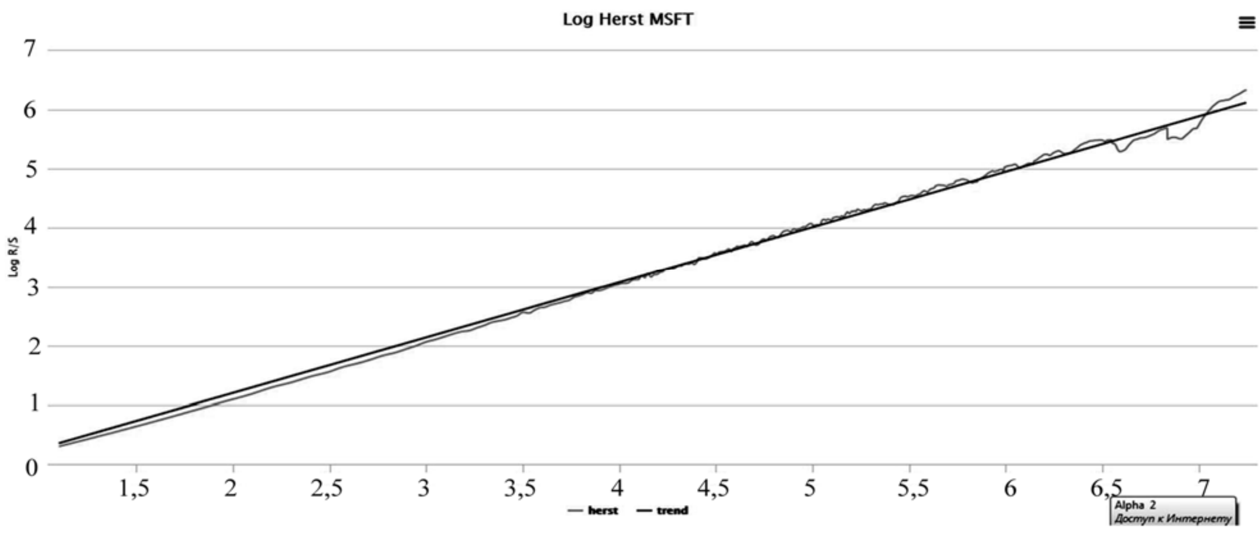

Fig. 2. R/S-trajectory

The Hurst exponent equals to 0.938537, this result shows that the stock is in the persistent interval. Let us change the interval length of the segments and calculate the Hurst exponent for each of them.

Table 1

The Hurst exponent on each interval segment

Interval	Hurst exponent
3–103	0.982368626898
103–203	0.990205696499
203–303	1.02330704858
303–403	0.966817092986
403–503	1.05207625507
503–603	0.90270971573
603–703	0.0617441128309
703–803	1.61274469297
803–903	1.40252095788

To solve the problem of time series forecasting, applying the Box-Jenkins methodology is essential. This methodology was first introduced back in the 1970s of the past century [2]. Is was based on the *ARIMA(p,d,q)*-class parametric model and focused on identifying the model and evaluating its parameters. When working on the time series analysis, the first step is to determine the order of integration of the series, i.e. to choose the parameter value *d* of the *ARIMA(p,d,q)* process.

Long-memory processes are stationary processes whose autocorrelation functions decay more slowly than short-memory processes. Because the autocorrelations die out so slowly, long-memory processes display a type of long-run dependence. The autoregressive fractionally integrated moving-average (ARFIMA) model provides a parsimonious parameterization of long-memory processes. This parameterization nests the autoregressive moving-average (ARMA) model, which is widely used for short-memory processes. The ARFIMA model also generalizes the autoregressive integrated moving-average (ARIMA) model with integer degrees of integration. ARFIMA models provide a solution for the tendency to overdifference stationary series that exhibit long-run dependence. In the ARIMA approach, a nonstationary time series is differenced d times until the differenced series is stationary, where d is an integer. Such series are said to be integrated of order d, denoted I (d), with not differencing, I(0), being the option for stationary series. Many series exhibit too much dependence to be I(0) but are not I(1), and ARFIMA models are designed to represent these series. The ARFIMA model allows for a continuum of fractional differences, $-0,5 < d < 0,5$. The generalization to fractional differences allows the ARFIMA model to handle processes that are neither I(0) nor I(1), to test for overdifferencing, and to model long-run effects that only die out at long horizons. An ARIMA model for the series y_t is given by

$$p(L)(1-L)^d y_t = \theta(L)\varepsilon_t.$$

where $p(L) = (1 - p_1 L - p_2 L^2 - \cdots - p_p L^p)$ is the autoregressive (AR) polynomial in the lag operator L; $Ly_t = y_{t-1}$; $\theta(L) = (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_p L^p)$ is the moving-average (MA) lag polynomial; ε_t is the independent and identically distributed innovation term; and d is the integer number of differences required to make the y_t stationary. An ARFIMA model is also specified by (3) with the generalization that $0,5 < d < 0,5$. Series with $d \geq 0,5$ are handled by differencing and subsequent ARFIMA modeling. An ARFIMA model specifies a fractionally integrated ARMA process. Formally, the ARFIMA model specifies that

$$y_t = (1-L)^{-d}\{p(L\}^{-1}\theta(L)\varepsilon_t.$$

The short-run ARMA process $p(L)^{-1}\theta(L)\varepsilon_t$ captures the short-run effects, and the long-run effects are captured by fractionally integrating the short-run ARMA process. Essentially, the fractional-integration parameter d captures the long-run effects, and the ARMA parameters capture the short-run effects. Having separate parameters for short-run and long-run effects makes the ARFIMA model more flexible and easier to interpret than the ARMA model. After estimating the ARFIMA parameters, the short-run effects are obtained by setting d = 0, whereas the long-run effects use the estimated value for d. The short-run effects describe the behavior of the fractionally differenced process $(1-L)^d y_t$ whereas the long-run effects describe the behavior of the fractionally integrated y_t [4].

The model allows simultaneous occurrence of long-term memory in the price/earnings ratio series and in the volatility series, using various error distribution types and including some additional explanatory variables. The d value for our model equals H-0,5 = 0,439, that is why we needed to calculate the Hurst exponent. To build such a model, we first need to fractionally differentiate the initial series of the stock with respect to the d power. For a start, we need to expand the 1-0,439L difference operator in the Taylor series. This difference will account for the values for some previous periods. Before using the Taylor series, we need to prove that at the d power, the numerical series of coefficients at lag operators coincides. For this purpose, we shall use the Leibniz criterion:

1) let us prove that $a_1 > a_2 > ... > a_n$;

2) let us prove that a_n tends to 0.

Despite the fact that stocks have infinite long-term memory, in my opinion, the most logical and optimal solution would be to limit the number of the Taylor series terms for differencing. Therefore, we will show 36 previous days to calculate each of the differences.

Table 2

Shows computational results of coefficient values for each lag

Lag	d value	Lag	d value	Lag	d value
1	0,43900	13	0.00757	25	0.00299
2	0,12313	14	0.00681	26	0.00284
3	0,06407	15	0.00618	27	0.00269
4	0,04102	16	0.00563	28	0.00256
5	0,02922	17	0.00517	29	0.00243
6	0,02221	18	0.00477	30	0.00232
7	0,01764	19	0.00442	31	0.00224
8	0,01447	20	0.00410	32	0.00212
9	0,01216	21	0.00384	33	0.00202
10	0,01216	22	0.00356	34	0.00194
11	0,01041	23	0.00337	35	0.00187
12	0,00905	24	0.00317	36	0.00179

After that, we obtained the resulting series on all the d values and weights that allowed us to see the memory length. The result can be seen in the diagram, which shows the number of successful forecasts along the Y axis and the number of days along the X axis. In the process, the system calculated 200 passings with different series length.

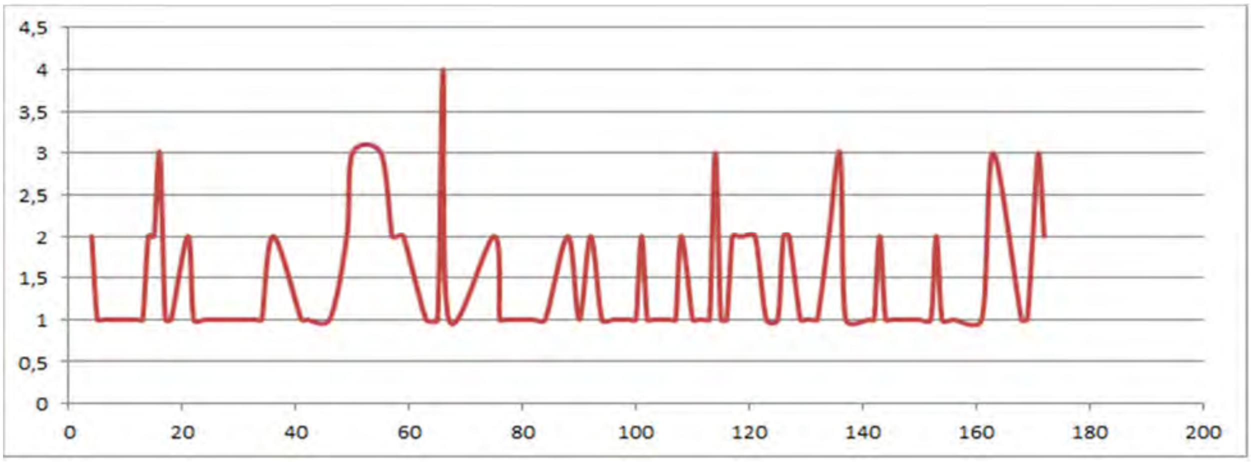

Fig. 3. Successful forecasting results are shown

As appears from the above diagram, in our case, the memory is 49 days. And it will allow us to create a model capable of making short-term stock forecasts. In the end, the cumulative error for the ARFIMA-model was 6 %.

Conclusions

In this work, we examined building a prognostic model of a fractal process. We described the general concept of building an ARFIMA-model and analyzed the interaction of the properties of a fractal process. Moreover, when making a forecast using the ARFIMA-model, we firstly conducted a full-fledged R/S-analysis of the initial time series and then used it to calculate the Hurst exponent. Based on these intermediate results, we can make some conclusions regarding the nature of the input data. Thus, it can be argued that with the help of this program system, we can obtain a reliable and effective forecast for any initial time series. So, as a result of this work, we were able to identify the main weaknesses and challenges of linear models of stationary time series with short-term memory. Using the R/S-analysis, we proved that time series with long-term correlation structures existed, which allowed us to rationalize the use of synergic methods for forecasting time series of such kind.

1. Peters E. Chaos and Order in the Capital Markets: A New View of Cycles, Prices, and Market Volatility / Edgar E. Peters. – John Wiley & Sons, Inc., 1996. – 274 p. 2. Mandelbrot B. The Fractal Geometry of Nature/B. Mandelbrot.-New Yor: W. H. Freeman, 1983. 3. Sowel F. Modelling long-run behavior with fractional ARIMA-models / F.Sowel // Journal of Monetary Economics. – 1992. – P. 277–302. 4. Contreras-Reyes J. E., Goerg G., Palma W. (2011). Estimation, Diagnostic and Forecasting Functions for ARFIMA models. R package version 0.1.7, http://cran.r-project.org/web/packages/afmtools/index.html.