Theory of Stochastic Processes
Vol. 12 (28), no. 3–4, 2006, pp. 114–122

MIKHAIL MOKLYACHUK AND ALEKSEY ZRAZHEVSKY

ANALYSIS AND FORECASTING OF SELF-SIMILAR FINANCIAL TIME SERIES

Time series of prices of MSFT ticker are considered. Results on self-similarity of this time series are presented. A method of prediction from FARIMA model for long-range dependent time series is described. This method is used for prediction of MSFT time series of prices that exhibits long-range dependence with the Hurst parameter close to 1.

1. Introduction

Many methods of analysis of time series of various types and origin are based on self-similarity and long-range dependence of time series. The long-range dependent processes provide a good description of many highly persistent financial time series. The predominant way to quantify long-range dependence and self-similarity is the value of the Hurst parameter. Methods of estimating the Hurst parameter are described in [1], [3], [4], [9], [10]. The problem of prediction of time series is one of the important problems of the theory of stochastic processes. Prediction the cash flows, constructing the arbitration strategy for prices of indexes on securities markets are examples of the problem. Different methods are proposed in literature for prediction of time series (see, for example, [5], [6]). Most of the proposed methods of prediction of long-range dependent time series are based on the choice of an appropriate stochastic model. Examples of such model are FARIMA, SEMIFAR, FIGARCH, GARCH [5].

The self-similarity of time series of prices of MSFT ticker and the corresponding derived series (the returns of MSFT ticker) is analyzed in [1]. It is shown that for prices of MSFT ticker the Hurst parameter is close to 1, so the prices may include trend and the data is predictable. This time series can be simulated with the help of self-similar processes. For the returns of MSFT ticker the Hurst parameter is close to 0.5, so the returns are uncorrelated, and it is impossible to predict them.

This paper deals with the prediction of long-range dependent time series based on the FARIMA model. On the basis long-range dependence of time series of prices of MSFT ticker analyzed in [1] the FARIMA model is used for description of MSFT time series of prices. The prediction from this model is constructed.

2000 *Mathematics Subject Classification.* Primary 60G52, 62M20, 93E10, 93E12.

Key words and phrases. Self-similarity, long-range dependence, MSFT time series, prediction from long-memory model, Toeplitz normal equations, FARIMA.

2. Self-similarity. Long-range dependence.

Let $\{x_t,\, t = 0, 1, \ldots\}$ be a wide-sense stationary time series with the autocorrelation function $r(k)$, $k \geq 0$. For each $m \in \mathrm{N}$ let $x^{(m)} = \{x_k^{(m)}, k = 0, 1, \ldots\}$ denote a new aggregated time series obtained by averaging of the original series $\{x_t,\, t = 0, 1, \ldots\}$ over non-overlapping blocks of size m, replacing each block by its sample mean. That is $x_k^{(m)} = \frac{1}{m}(x_{km-m+1} + \cdots + x_{km})$, $k \geq 1$. This aggregated time series is also wide-sense stationary. The corresponding autocorrelation function is denoted as $r^{(m)}(k), k \geq 0$. A time series $\{x_t,\, t = 0, 1, \ldots\}$ is called asymptotically second-order self-similar (further simply self-similar) if for large enough k:

$$r^{(m)}(k) \to r(k), m \to \infty. \tag{1}$$

A time series $\{x_t,\, t = 0, 1, 2, \ldots\}$ is said to be long-range dependent if:

$$r(k) \sim k^{-\beta}, k \to \infty \quad \text{and} \quad 0 < \beta < 1. \tag{2}$$

A long-range dependent time series is self-similar (see, for example, [4]) and the Hurst parameter is defined as $H = 1 - \beta/2$. For long-range dependent time series $1/2 < H < 1$. For $H = 1/2$ samples (observations) are uncorrelated. For $0 < H < 1/2$ time series is said to be short-range dependent [4].

The predominant way to determine self-similarity is to establish the long-range dependence, that is, quantify the value of the Hurst parameter. There are many methods of estimation of the Hurst parameter. The most well-known of them are: method of standard deviation of aggregates, rescaled adjusted range statistic (R/S) method, method of autocorrelation function, method of periodograms, Robinson's method (see [1], [3], [9], [10]).

If the self-similarity is detected, then it is possible to select the appropriate model for description and prediction of the considered time series.

3. Linear prediction for long-range dependent time series

Let $\{x_t,\, t = 0, 1, \ldots\}$ be a wide-sense stationary long-range dependent time series. Suppose, that we know values of this time series until some fixed moment of time T. The prediction problem is to estimate the value x_T, using the information for the past observations $x_0, x_1, \ldots, x_{T-1}$.

Consider the FARIMA process which is a model for financial long-range dependent time series (see [2], [5], [7]). This process is a solution of the equation:

$$\Phi_p(L)(1 - L)^d (y_t - \mu) = \Theta_q(L)\, \varepsilon_t, \quad t \in Z, \tag{3}$$

where L denotes the lag operator, d is difference parameter, Φ_p and Θ_q are polynomials:

$$\Phi_p(z) = 1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p,$$
$$\Theta_q(z) = 1 + \theta_1 z + \theta_2 z^2 + \ldots + \theta_q z^q,$$

with roots outside the unit circle, ε_t is assumed to be i.i.d. Gaussian random variables with zero mean and variance σ^2, $E[y_t] = \mu$. This process is usually referred to as the $ARIMA(p, d, q)$ model. By allowing d to be real number instead of a positive integer, ARIMA model becomes the autoregressive fractional integrated moving average (FARIMA) model. $FARIMA(p, d, q)$ model can be equivalently expressed as an AR(∞) model [5].

The prediction from FARIMA model can be considered as prediction from the $AR(p)$ representation with very large p, where the $AR(p)$ coefficients are the first p coefficients of the $AR(\infty)$ representation of the $FARIMA(p, d, q)$ model. This method is referred to as the truncation method (see [5]). The prediction, obtained by this method, is of the form:

$$\hat{x}_T = a_1 x_{T-1} + ... + a_p x_{T-p}, \tag{4}$$

where a_i, $i = 1, \ldots, p$ are weighting coefficients selected yield the best linear predictor, p denotes the model order. To estimate the vector of coefficients $\mathbf{a} = [a_1, a_2, \ldots, a_p]^\top$ let us form a vector process $\{\mathbf{x}(t),\, t \geq L-1\}$ from the time series $\{x_t,\, t = 0, 1, \ldots\}$ in such manner: $\mathbf{x}(t) = [x_t, \ldots, x_{t-L+1}]^T$, where $0 \leq L \leq T+1$. The vector $\mathbf{x}(T)$ includes the value x_T to be predicted. The problem of prediction of the time series $\{x_t,\, t = 0, 1, \ldots\}$ at the moment T can be reformulated to the problem of prediction of the process $\{\mathbf{x}(t),\, t \geq L-1\}$ at the moment T, using the past observations $\mathbf{x}(T-1), \ldots, \mathbf{x}(L-1)$. This process is stationary and its autocorrelation function has the same asymptotic behavior as the autocorrelation function for the initial time series. Construct the prediction for the process $\{\mathbf{x}(t),\, t \geq 0\}$ from FARIMA model. Denote the forward predicted process vector at the moment T as a weighted linear combination of previous measurements similarly to (4) with the vector of coefficients $\mathbf{a}$:

$$\mathbf{x}^f(T, \mathbf{a}) = \sum_{m=1}^{p} a_m\, \mathbf{x}(T-m), \tag{5}$$

where $p = T - L + 1$ denotes the model order.

Evaluation of the "closeness" of the predicted process $\mathbf{x}^f(T, \mathbf{a})$ to the true observation $\mathbf{x}(T)$ makes it necessary to define a prediction error vector $\varepsilon(T, \mathbf{a}) = \mathbf{x}(T) - \mathbf{x}^f(T, \mathbf{a})$. We will determine the estimate of $\mathbf{a}$ as a vector $\hat{\mathbf{a}}$ that minimizes the prediction error energy $\mathbf{E}(T, \mathbf{a})$:

$$\min_a \mathbf{E}(T, \mathbf{a}) = \min_a \varepsilon^\top(T, \mathbf{a})\, \varepsilon(T, \mathbf{a}) = \mathbf{E}(T, \hat{\mathbf{a}}). \tag{6}$$

A solution of the least-squares problem (6) results in the normal equations of linear prediction:

$$\mathbf{X}^\top(T)\, \mathbf{X}(T)\, \hat{\mathbf{a}} = \mathbf{X}^\top(T)\mathbf{x}(T). \tag{7}$$

where $\mathbf{X}(T) = [\mathbf{x}(T-1), \mathbf{x}(T-2), \ldots, \mathbf{x}(T-p)]^\top$ is the matrix of past observations.

Let us define the covariance matrix $\mathbf{\Phi}(T) = [\Phi_{i,j}(T)]$, $\Phi_{i,j}(T) = \mathbf{x}(T-i)\, \mathbf{x}(T-j)$, $\quad 0 \leq i,\, j\ \leq p$ for the vector process $\mathbf{x}(t)$. Then the normal equations (7) can alternatively be expressed as (see [7]):

$$\Phi(T)\begin{bmatrix} -1 \\ \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{x}^\top(T)\,\mathbf{X}(T)\,\hat{\mathbf{a}} - \mathbf{x}^\top(T)\mathbf{x}(T) \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{8}$$

The observed vector process is stationary: $\Phi_{0,j}(T-k) = \Phi_{0,j}(T)$, $0 \le k \le p$ and the covariance matrix can be replaced with the Toeplitz matrix - matrix of autocorrelation coefficients of the form $c_i = \Phi_{0,i}/\Phi_{0,0}$, $0 \le i \le p$. The system (8) can be replaced by the Toeplitz Normal Equations:

$$\begin{bmatrix} 1 & c_1 & \dots & c_{p-1} \\ c_1 & 1 & & \vdots \\ \vdots & & \ddots & c_1 \\ c_{p-1} & \dots & c_1 & 1 \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \vdots \\ \hat{a}_{p-1} \\ \hat{a}_p \end{bmatrix} = \begin{bmatrix} c_1 \\ \vdots \\ c_{p-1} \\ c_p \end{bmatrix}. \tag{9}$$

The system of linear equations (9) is sometimes called the Yule-Walker equations. Note, that the Toeplitz matrix and matrix from the Normal Equations $\mathbf{A} = \mathbf{X}^{\top}{}'(T)\,\mathbf{X}(T)$ are symmetric. The equation (9) (or, if the process is not stationary, (7)) can be solved using classical algorithms for symmetric linear systems. Most important of them that are used in linear prediction theory are Cholevsky decomposition, QR decomposition, singular value decomposition [7].

A solution of the equation (9) provides the least-squares estimation of $\mathbf{a}$ and the forward predicted process vector $\mathbf{x}^f(T, \mathbf{a})$ can be obtained by (5). The first element of this vector is the prediction x_T^f of the time series $\{x_t,\, t = 0, 1, \dots\}$ at the moment T.

Equations (9) may be solved only if we know the vector of autocorrelation coefficients $\mathbf{c} = [c_1, c_2, \dots, c_p]^\top$. Coefficients $c_1, c_2, \dots, c_{p-1}$ can be computed using the past observations $\mathbf{x}(T), \dots, \mathbf{x}(L-1)$. The value of c_p is unknown. We propose the linear regression model based on the long range dependence of the observed time series for predicting unknown autocorrelation coefficients.

We consider the long-range dependent time series $\{x_t,\, t = 0, 1, \dots\}$ with the autocorrelation function $r(k)$ that satisfies [4]:

$$\frac{r(k)}{H(2H-1)k^{2H-2}} \to 1, \quad k \to \infty. \tag{10}$$

Note, that (10) is the extended version of (2) and is correct for any long-range dependent time series.

The vector process $\{\mathbf{x}(t),\, t \ge 0\}$, constructed above, has the same asymptotic behavior of the autocorrelation function as the origin time series. So we will suppose that:

$$c_k \sim H(2H-1)k^{2H-2}, k \to \infty \quad \text{and} \quad 1/2 < H < 1, \tag{11}$$

where c_k denotes the autocorrelation coefficients for the vector process $\{\mathbf{x}(t), t \geq 0\}$, H is the Hurst parameter.

We will use the linear regression model for fitting values $c_m, c_{m+1}, \ldots, c_{p-1}$ on the basis of (11):

$$c_k = \alpha_1 H(2H-1)\, k^{2H-2} + \alpha_2 + \xi_k, \quad m \leq k \leq p-1, \tag{12}$$

where α_1, α_2 are coefficients, ξ_k, $m \leq k \leq p$, are random errors. The standard assumptions of linear regression model are: $M[\xi_k] = 0$, $M[\xi_k] = \sigma^2$, $M[\xi_k\, \xi_l] = 0$ for $k \neq l$. Estimates of the coefficients α_1, α_2 are calculated to minimize the residual sum of squares $\sum\limits_k \xi_k^2$ [5]:

$$(\hat{\alpha}_1,\, \hat{\alpha}_2) = \underset{(\alpha_1,\, \alpha_2)}{\arg\min} \sum_k (c_k - \alpha_1\, H\, (2H-1)\, k^{2H-2} - \alpha_2)^2 \tag{13}$$

Estimates of autocorrelation coefficients c_k are:

$$\hat{c}_k = \hat{\alpha}_1 H(2H-1)\, k^{2H-2} + \hat{\alpha}_2, \quad m \leq k \leq p. \tag{14}$$

Estimated residuals for the fitted model are $\hat{\xi}_k = c_k - \hat{c}_k$. Several residuals diagnostics are used to evaluate the validity of the underlying assumptions of the model. The most common diagnostic for serial correlation is Durbin-Watson statistics

$$DW = \frac{\sum\limits_{k=m+1}^{p} (\hat{\xi}_k - \hat{\xi}_{k-1})^2}{\sum\limits_{k=m}^{p} \hat{\xi}_k^2}. \tag{15}$$

Values of DW around 2 indicate no serial correlation in the errors: $M[\xi_k\, \xi_l] \approx 0$ for $k \neq l$ ([5]).

The linear regression (12) is not used for first m values because (11) gives the asymptotic property for autocorrelation function. We obtain the estimate $\hat{c}_p$ from (14) necessary for prediction.

4. Practical results

The practical use of the described above method is shown on the example of MSFT time series - time series of prices on ticker of firm Microsoft. This time series is long-range dependent (and self-similar) with the Hurst parameter close to 1, that implies the possibility for "good" prediction. This is shown in paper [1] on the basis of four methods. The corresponding estimates of the Hurst parameter are shown in Table 1.

Table 1. Estimate of the Hurst parameter (see [1]).

		Period.	Robins.	Autocor.	R/S
MSFT prices	estim.	0.962	0.977	0.953	0.993
	C.I.	0.94-0.98	0.93-1.01	0.94-0.96	0.96-1.02

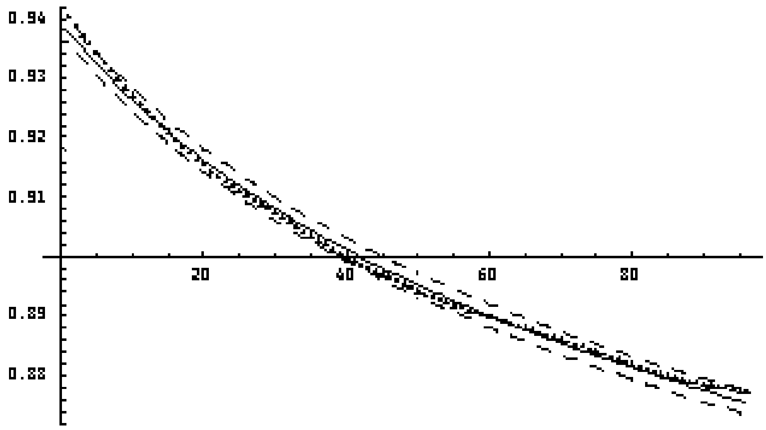

FIGURE 1. Predicted and observed autocorrelation coefficients with confidence intervals

According to results of paper [1] we may use the prediction from FARIMA model for MSFT time series of prices. The prediction is computed in two steps. On the first step the autocorrelation coefficients are fitted and predicted. On the second step, based on predicted autocorrelation coefficient, the Toeplitz Normal Equations are solved and prediction for the vector process is constructed.

We construct a linear regression model (12) for autocorrelation coefficients $c_m, c_{m+1}, \ldots, c_{p-1}$ changing the Hurst parameter with its estimate shown in the Table 1. The observed autocorrelation coefficients are evaluated using vectors $\mathbf{x}(t)$ with length $L = 3970$. A linear regression is used for $p - m = 128 - 32 = 96$ values of observed autocorrelation coefficients with $\hat{H} = 0.971$ (mean value of estimates computed using different methods). Estimates of the coefficients α_1, α_2 are calculated using (13). Estimates of coefficients are $\hat{\alpha}_1 = 1.0195$, $\hat{\alpha}_2 = 0.1979$ with the standard errors 0.0059 and 0.004 correspondingly. The value of Durbin-Watson statistic is 2.0139 that indicates no serial correlation in the errors. These results show that the chosen regression model is suitable. Estimates of autocorrelation coefficients $\hat{c}_k$ together with the observed values c_k and 95%-confidence intervals evaluated from the linear regression model (12) using the the Student-t distribution with $p - m - 2$ degrees of freedom are shown on Fig.1. The prediction for the next autocorrelation coefficient was found from (14): $\hat{c}_{128} = 0.8747$ with the standard error 0.013312. All calculations were made in *Mathematica* 5.0 with the help of function "Regress" [8].

An estimate for the parameter vector $\mathbf{a} = [a_1, a_2, \ldots, a_p]^T$ is found as a solution of the Toeplitz Normal Equations (9), where the autocorrelation coefficients c_p are replaced with estimates $\hat{c}_p$. This system of equations was

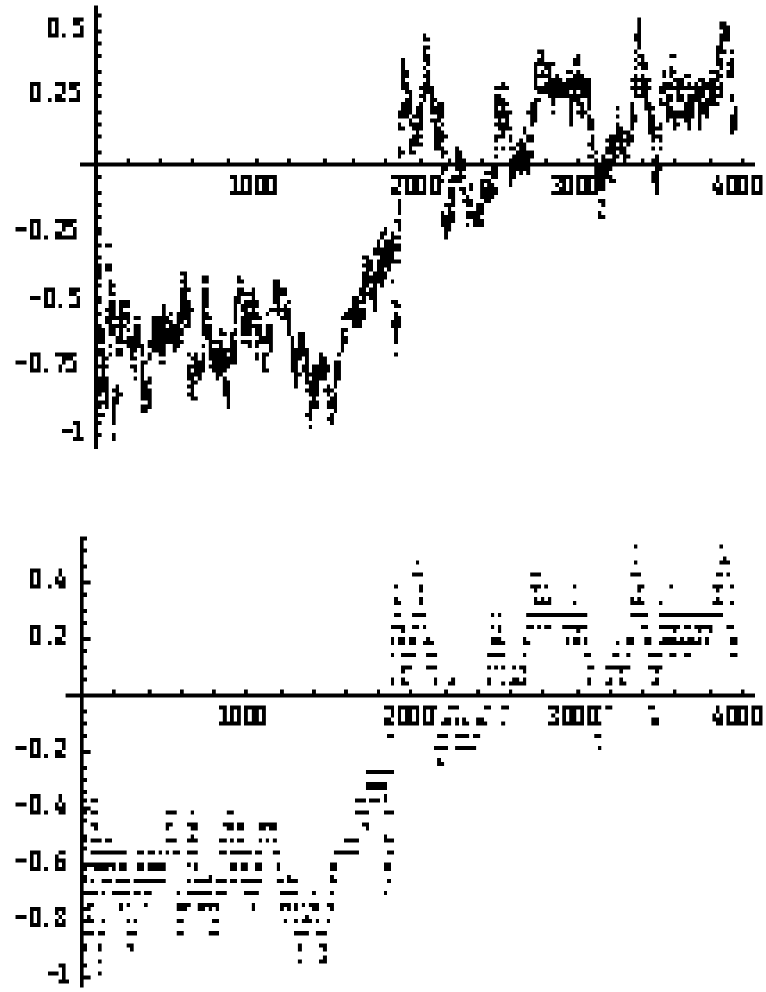

Figure 2. Predicted and observed vectors for MSFT time series

Table 2. Analysis of errors for prediction

	Mean	Median	Variance	Stand.Dev.	Stand.Error
$\varepsilon_{\mathbf{1}}$	0.00021	0.00027	0.00023	0.015	0.00024
$\varepsilon_{\mathbf{2}}$	-0.00056	-0.00012	0.00012	0.01096	0.00028

solved using function “Solve” in *Mathematica* 5.0 [8]. The forward predicted vector $\mathbf{x}^f(T)$ at the moment $T = 4097$ is found from (11). This is shown on Fig.2 along with the corresponding observed vector. The predicted vector includes all trends that are observed for the initial one. That implies the goodness of our method for predicting this time series.

Consider the prediction error vector $\varepsilon_{\mathbf{1}} = \mathbf{x}(T) - \mathbf{x}^f(T)$. The basic statistic characteristics for the data of errors are shown in Table 2. Note,

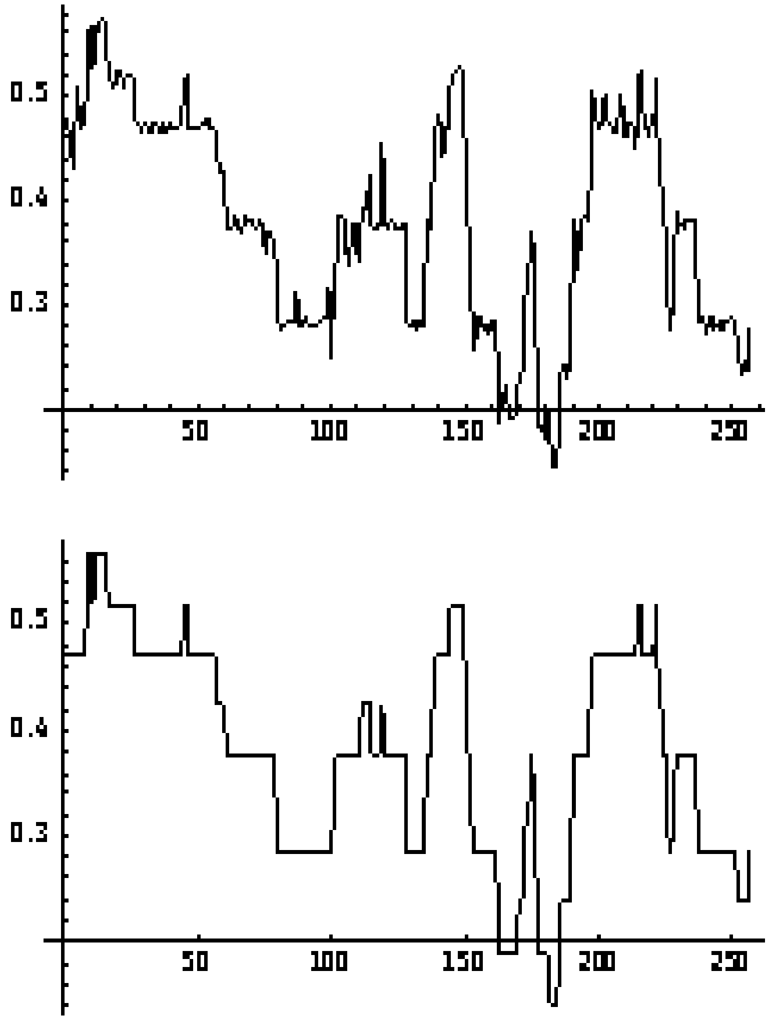

FIGURE 3. Predicted and observed MSFT time series

that predicted vector have no practical interest except one element x_{4097} that have to be predicted.

We repeat the described procedure for 256 times. Every time we were interested in the last element of the predicted vector that was unknown before prediction. The corresponding data along with the true values of time series are shown on Fig.3. The errors $\varepsilon_{\mathbf{2}} = x_t - x_t^f$ for $t = 4097, \ldots, 4353$ are obtained and basic statistic characteristics for them are shown in Table 2.

Note, that values p, T, L, m where chosen from the practical reasons.

5. CONCLUSIONS

Time series of prices of MSFT (a ticker of firm Microsoft) is analized in this paper. Main results from paper [1] on self-similarity of these time series are reviewed. On the basis of these results a regression model for the autocorrelation function of the vector process constructed from MSFT time series of prices is proposed and used for prediction. A prediction problem from FARIMA long-memory model is considered. A described method of prediction is applied to MSFT time series of prices. The corresponding results along with the errors are presented. The predicted time series has all trends observed at the origin one. Errors are insignificant.

References

1. Moklyachuk, Mikhail P., Zrazhevsky, Aleksey G., *Long-range dependence of time series for MSFT data of shares and returns,* Random Operators and Stochastic Equations, Vol. 14, No. 4, (2006), 393–403.
2. Zhurbenko, I. G.; Kozhevnikova, I. A., *Stochastic modelling of processes,* Moskva: Izdatel'stvo Moskovskogo Universiteta. 148 p. (1990).
3. Robinson, P. M., *Gaussian semi-parametric estimation of long range dependence,* The Annals of Statistics, **23**, (1995), 1630-1661.
4. Beran, J., *Statistics for Long-Memory Processes,* Monographs on Statistics and Applied Probability, Chapman and Hall, New York, (1994).
5. Zivot, Eric, Wang, Jiahui, *Modeling Financial Time Series with S-PLUS,* New York: Springer, (2003).
6. Strobach, Peter, *Linear Prediction Theory - A Mathematical Basis for Adaptive Systems,* Springer-Verlag, (1990).
7. Stoev, S., Taqqu, M.S. *Simulation methods for Linear fractional stable motion and FARIMA using the fast Fourier transform,* Fractals, **12(1)**, (2004), 95-121.
8. Wolfram Research, Inc., *Mathematica, Version 5.0,* Champaign, IL (2003).
9. Deane, J. H. B., Smythe, C., Jefferies, D. J., *Self-similarity in a deterministic model of data transfer,* International Journal of Electronics, **80**, 5, (1996), 677-691.
10. Teverovsky, V., Taqqu, M., *Testing for long-range dependence in the presence of shifting mean or a slowly declining trend, using a variance-type estimator,* Time Series Analysis, **18**, 3, (1997), 279-304.

Department of Probability Theory and Mathematical Statistics, Kyiv National Taras Shevchenko University, Kyiv, Ukraine
E-mail address: mmp@univ.kiev.ua

Department of Probability Theory and Mathematical Statistics, Kyiv National Taras Shevchenko University, Kyiv, Ukraine
E-mail address: zalex@univ.kiev.ua