INFORMATION TECHNOLOGY

Information Technology for Analysing and Forecasting Time Series with Fractal Properties on the Basis of Linguistic Modelling

Y. A. Nedashkivskyi

National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute»
Corresponding author. E-mail: e.niedaszkowski@yandex.ua

Paper received 06.05.19; Accepted for publication 20.05.19

https://doi.org/10.31174/SEND-NT2019-200VII24-05

Abstract. The research paper proposes and determines the effectiveness of information technology for analyzing and forecasting time series with fractal properties based on linguistic modelling. The information technology for analyzing and forecasting time series with fractal properties based on linguistic modelling is superior to its competitors/analogues and confirms the high quality of development with the possibility of implementing the system when required.

Keywords: *information technology, analysis, forecasting, financial time series, linguistic modelling, fractal properties.*

Introduction. Modern economic theory, having undergone a complex evolution, differentiating into a multitude of schools and directions, consistently reproduced a significant difference in approaches to forecasting. From the 60’s of the 20th century to date, the structure of the economy of the developed countries and methods of econometric analysis have undergone fundamental changes [1]. The linear approach to the analysis of economic systems becomes irrelevant and ineffective due to the fact that modern economic processes are not in a state of strict determination, self-regulation, clear certainty. It is not able to reflect the behaviour of the economy in conditions of uncertainty, instability and imbalance. Therefore, the topical issue of the introduction of mathematical apparatus into the process of modelling arises.

The economic field is characterized by phenomena that evolve and fluctuate in time. Over time, prices, economic conditions, the flow of a particular production process change. The set of measurements of this kind of indicators over a period of time is a time series [2].

Since the beginning of the 21st century, the issues of the natural language and speech modelling have become topical in the studies of many scholars, which is due, first of all, to the rapid development of applied linguistics [3, 4]. There was also an increase in the number of linguistic models to solve the main practical issues in various spheres of application. The term “model”, in the context of linguistic modelling, is understood as the typical model construction, which represents the standard [4].

The scale of the concept of “model” is due to different branches of linguistics, as well as to the different volume of material that needs to be modelled. Analyzing the conducted research of A.F. Losev [3], 27 main linguistics-related meanings were singled out as of the beginning of the 60’s. In the modern scientific world, the concept of "model" is interpreted in different interpretations, which differ slightly from one another. It should be noted that it is necessary to agree with A.F. Losev, who noted that “any model is a structure, but not every structure is a model” [3]. This is a generalization of the fact that when exploring the structure, type, paradigm, an idea arises that a certain class of objects that is not always a model is behind each concept.

The possibility of modelling is based on the fact that the model, in some sense, reflects any properties of the original, therefore, the relevant theories that would indicate the limit of permissible simplifications is important in the process of modelling. As we see, the essence of modelling is to replace the original object with its model, while preserving only the essential properties of this object in it, which makes the modelling method relatively limited in its capabilities, but rather objective in obtaining results. Thus, the tasks of modelling include the construction of models, their study and verification in the process of proving the veracity of knowledge [9].

The mechanism of constructing a model of dynamic processes that may have fractal properties of the use of linguistic modelling is described in [10]. The very process of constructing a linguistic model based on the time series of a dynamic process is described in [11].

Fractal behaviour is revealed in time characteristics of processes and phenomena occurring in media with a self-similar structure. Fractal time series are a whole class of fractal curves, widely used in describing and modelling a variety of phenomena. They describe such phenomena that seemingly have nothing in common: the Brownian motion, the behaviour of the exchange rate in the financial markets, the change in water level in lakes and rivers, etc.

The use of fractals in the modelling of time series, in particular, such a characteristic of the time series as fractal dimension, allows determining the moment in which the system becomes unstable and ready to move to a new state.

Materials and methods of research. Financial time series demonstrate a high degree of nonlinear variability, especially at high frequencies, and often demonstrate fractal properties. When the fractal dimension of the time series equals zero, this is due to two characteristic features: fractal processes reveal heterogeneity - high probability of extreme or distant fluctuations, usually with irregular intervals; fractal processes also demonstrate the symmetry of exposure - the proportionality of the relations between the vibrations at different fluctuations.

The signs of fractality in the financial markets do not imply a chaotic behaviour that resembles a coincidence generated by a small number of deterministic equations. Fractality in large-scale multidimensional systems, such as financial markets, is stochastic [12]. This type of fractality usually occurs as a result of multiplicative interactions between two or more stochastic processes.

Most of the modern literature on nonlinear variability [13-15] in the financial markets was based on the diffusion of volatility with a multitude of random factors. Volatility diffusers with a multiplicative relationship between stochastic factors tend to generate fractality.

At large distances, the financial series can be modelled using structural equations, which are not fractal, as a rule. In this sense, it is reasonable to characterize economic time series as a manifestation of fractal properties on short horizons, but with asymptotic similarity to equilibrium. In some cases, the basic structural equations also give rise to fractality [16].

The clearest example is the exchange rate which is most likely conditioned by differentials in real rates of return. Let X_t be the exchange rate, I_t be the interest rate, e - waiting count, ε_t - the residual component of the financial time series, F - the foreign currency. The structural equation has the following form:

$$X_t = \omega_0 + \omega_1 X_{t-1} + \omega_2[(I_t - \pi_1^e) - (I_{Fl} - \pi_{Ft}^e)] + \varepsilon_l. \quad (1)$$

As noted in [17], the adoption of coefficients of stochastic processes can lead to nonlinear variability. On this basis, the expression $[(I_t - \pi_1^e) - (I_{Fl} - \pi_{Ft}^e)]$ involves fractal behaviour. In addition, nominal and real interest rates can also be fractal, so the difference in the real rates of return is in itself the difference between two independent fractal processes.

The methodology of time series modelling is based on the decomposition of a time series into components and modelling of the values of each component separately. In the framework of this study, it is proposed to apply aggregation to each row, that is, the decomposition of a series into short intervals whose values are similar by any sign. From the standpoint of mathematical science, aggregation is considered as the transformation of the original model into a model with a smaller number of variables and constraints, which gives an approximate (in comparison with the original) description of the investigated process or object.

The main factor in the study of the fractal structure of financial time series is the finding of fractal indicators within the series under study. In [18], it is proved that the fractal dimension is close to 1.5 ($\mu = 0{,}5$) is valid only for 15-25% of the financial time series, which is an unsatisfactory indicator; the remainder of the time behavior of the series differs significantly from the Brownian motion. This fact is due to the fact that the sections of abnormal behaviour have a short duration. At the same time, large-scale data arrays (from several hundred to several thousand points) are required for normality validation by standard methods. There will be a large number of sections with different behaviours within the interval used to validate normality by standard methods. When calculating the test values, the characteristics of persistent and anti-persistent sections are compensated and the final values will be close to normal.

On the basis of the foregoing, let N be the number of events, L - a characteristic length, and D - the fractal dimension; let's fix the value for the financial time series $D = 1$.

Probabilistic measure of dimension is determined using the formula:

$$[N(|Y_t - Y_{t-1}|) > L/N(|Y_t - Y_{t-1}|)]/L.$$

The proportion of observations lying outside the L threshold for the total sample varies depending on:

$$[N(|Y_t - Y_{t-1}|) > L/N(|Y_t - Y_{t-1}|)] \approx L^d, \quad (2),$$

where asymptotic equality is designated as $\approx$. In this sense, dimension is a measure of entropy or randomness.

The measurement varies according to the threshold, and therefore is often measured as an asymptotic limit as the threshold approaches zero. The modification of this method is the estimation of the average heterogeneity of the process, using the codimension C, the result of which is the difference between the dimension of the enclosure d and the fractal dimension D:

$$C = D - d. \quad (3)$$

If $C \neq 0$, the process is called fractal.

At low degrees of fractality (close to zero) the process is more homogeneous: there are few extreme fluctuations. Conversely, more extreme events or fluctuations beyond the threshold are characteristic of higher values of C. The process becomes less homogeneous, shorter, rarer, and more volatile.

The relations between codimension and symmetric scaling are set by the following equations in which τ is a time scale from 1 to T, where T is the largest time scale; μ is the index; q is a series of scaling indicators. The scaling symmetry has the following form:

$$\mu(|Y_t - Y_{t-1}|^q) \approx [\mu(|Y_t - Y_{t-1}|^q)][(\tau/T)^{\zeta(q)}], \quad (4)$$

where ζ – a function that includes three parameters C_1, H, α:

$$\zeta(q) = qH - \{[C_1/(\alpha - 1)](q^\alpha - q)\}, where\ \alpha \neq 1, \quad (5)$$

$$\zeta(q) = qH - (C_1 q ln q), where\ \alpha = 1. \quad (6)$$

The parameter C_1 is the encoding associated with sample size scaling. When $C_1 = 0$, $\zeta(q)$ is a linear trend. When $C_1 \neq 0$, the curvature of the axis $\zeta(q)$ depends on the codimension and probability distribution. The coefficient α characterizes the probability distribution. The case $\alpha = 2$ corresponds to the Gaussian distribution, while $\alpha = 1$ corresponds to the Cauchy distribution. Most economic processes show $1 < \alpha < 2$. In this case, the distribution has heavier tail areas than the standard norm, and variance changes over time. An interesting property of the series, both $< \alpha < 2$ and $0 < C_1 < 1$, is that integration usually does not lead to smoothing. Instead, the integral will show discrete jumps.

The coefficient H characterizes the fractality index, that is, the decreasing value when the delay between two identical pairs of values in the time series increases. The designation H actually derives from the Hurst exponent, or the recalled range factor [17]. However, in this framework, H is rated as one of a series of scaling factors. This statistics is related to the additive constant integration order. For the process $I(0)H = 0{,}5$. For a nonstationary process, the integration order can be restored by evaluating the pace of change [19]. This method is reliable both for nonlinearity and for fractional integration orders.

Results. The object of research is a time series of the following type: $\{\tilde{X}(x)\}(x = 1 \div t)$, (7)

where $\tilde{X}(x)$ is a time series with fractal properties, which is characterized by the tuple:

$$\left\{ {k_j^x}/{\mu(k_j^x)} \right\}, \mu(k_j^x) \to [0,1], j = 1 \div J, \quad (8)$$

The initial data is given in Figure 1. The total data volume is 53,208 items.

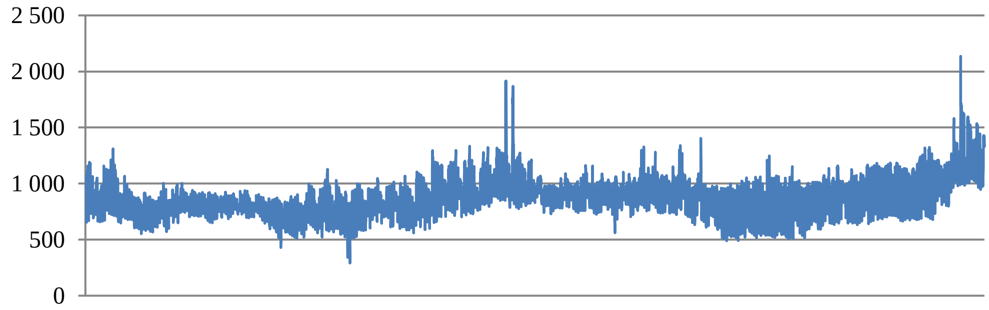

Fig. 1. Distribution chart of the time series with fractal properties

Let the value of the linguistic variable (term) according to the set be:

$$U = \{0; 0{,}1; 0{,}2; 0{,}3; 0{,}4; 0{,}5; 0{,}6; 0{,}7; 0{,}8; 0{,}9; 1\},$$

consequently:

V – high, then: $\mu_v(u) = u, u \in U$;

BV – higher, then: $\mu_{bv}(u) = \sqrt{u}, u \in U$;

DV – very high, then: $\mu_{dv}(u) = u^2, u \in U$;

ZV – too high, then: $\mu_{zv}(u) = \begin{cases} 1, u = 1, \\ 0, u < 1, \end{cases} u \in U$;

N – low, then: $\mu_n(u) = 1 - u, u \in U$;

BN – lower, then: $\mu_{bn}(u) = \sqrt{1-u}, u \in U$;

DN – very low, then: $\mu_{dn}(u) = (1-u)^2, u \in U$;

ZN – too low, then: $\mu_{zn}(u) = \begin{cases} 0, u = 1, \\ 1, u < 1, \end{cases} u \in U$.

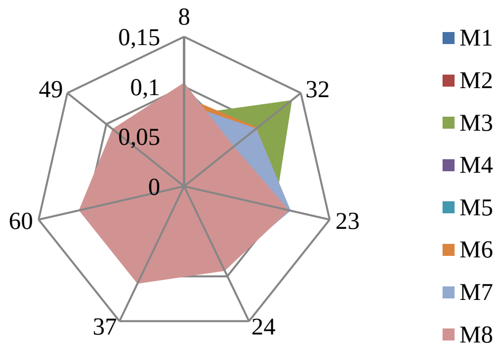

Fig. 2. Distribution chart of consolidated sets in relation to the weight of the time series with fractal properties

On the basis of the formation of consolidated sets, with respect to weight of the time series with fractal properties within the time distribution, according to the expressions $e_k(k = 1 - 7)$, the corresponding values will have the following form (Figure 2).

The value of k of the given time A_k $(k_{2012 \div 2018} = 1 \div 7)$ can be found using the following compositional rule:

$$A_k = G_k{}^{\circ} \square R \; (k_{2012 \div 2018} = 1 - 7), \quad (9)$$

where G_k is a reflection of the decomposition in the form of a fuzzy subset.

A_k is a fuzzy interpretation of the kth given time series with fractal properties in accordance with the vector $(0;\ 0{,}1;\ 0{,}2;\ \ldots;\ 1)$.

As a result of the analysis of time series with fractal properties, an analogue of the considered time series is obtained, which is presented in Table. 1

The next step is to determine a set of each individual level:

$$A_a = \{u | \mu_A(u) \geq \alpha, u \in U\}, \quad (10)$$

and the average number of elements in the time series with fractal properties:

$$M(A_a) = {}^{1}/_{n} \sum_{k=1}^{n} u_k,\ u_k \in A_a.$$

For comparison, let's list the values obtained in the study of the developed system and the analogues accepted for consideration (Table 2) and carry out the analysis of the obtained data.

Table 1. The model of the time series with fractal properties

Value of the membership function	Year of the study						
	2012	2013	2014	2015	2016	2017	2018
0	0.01336	0.06154	0.06605	0.06197	0.05911	0.02490	0.09856
0.1	0.02943	0.07104	0.13332	0.08033	0.08209	0.10539	0.09838
0.2	0.04637	0.06262	0.06776	0.07003	0.09687	0.11207	0.10367
0.3	0.04224	0.0382	0.00388	0.03881	0.08680	0.08037	0.09490
0.4	0.06240	0.06202	0.07227	0.07266	0.08378	0.10092	0.10067
0.5	0.07584	0.04944	0.05809	0.05728	0.07753	0.08920	0.12964
0.6	0.04709	0.06536	0.06150	0.05959	0.09358	0.10662	0.12987
0.7	0.05853	0.05927	0.06629	0.07139	0.09045	0.10769	0.05897
0.8	0.03217	0.05248	0.08441	0.06972	0.05911	0.08540	0.06209
0.9	0.02905	0.04812	0.05445	0.06071	0.11357	0.09258	0.06308
1.0	0.06128	0.11871	0.07055	0.04853	0.10423	0.1089	0.10084
Analogue A_k	A_1	A_2	A_3	A_4	A_5	A_6	A_7
Evaluation	0.452549	0.626222	0.071477	0.482427	0.086108	0.92193	0.04611

Discussion. The use of information technology for analyzing and forecasting time series with fractal properties will allow for a high level of forecasting, and consequently, the fullest possible realization of the capabilities of the analytical system. The study of the methods of adaptive type showed that the Brown model only works with a small time horizon. The trend and seasonal changes are not taken into account. Information technology for analyzing and forecasting time series with fractal properties is universal and allows adapting the process of planning the forecast of financial time series to the level of development of the initial series.

Table 2. Results of the mathematical analysis

Factor	Automated system of analysis and forecasting of the financial time series linguistic modelling		Time series forecasting system Brown method		Time series forecasting system Holt method		Time series forecasting system Winters method		Automated system of analysis and forecasting of the financial time series	
	Average deviation	Variance	Average deviation	Variance	Average deviation	Variance	Average deviation	Variance	Average deviation	Variance
Efficiency	0.84	0.81	0.48	0.24	0.48	0.24	0.54	0.49	0.36	0.29
Innovation	0.42	0.21	0.48	0.24	0.48	0.24	0.48	0.24	0.42	0.21
Data and forecasting scope	0.84	0.96	1.4	3.6	1.12	2.09	1.2	2.4	1.44	2.61
Data categories	0.84	0.96	0.54	0.49	0.4	0.4	0.9	1.05	0.76	0.81
Priority	2.1	5.69	1.04	1.64	1.44	2.81	1.28	2.09	1.5	2.69
Ultimate solution level	0.84	0.96	0.6	0.6	0.4	0.4	0.82	1.01	0.84	0.96
State	0.42	0.21	0.48	0.24	0.42	0.21	0.5	0.25	0.42	0.21
Forecast quality factor	0.56	0.44	0.7	0.65	0.6	0.6	0.54	0.49	0.6	0.44
Client assessment	0.32	0.16	0.72	0.64	0.9	0.89	0.72	0.64	0.32	0.16

Conclusions. The paper identifies the effectiveness of information technology for analyzing and forecasting time series with fractal properties based on linguistic modelling.

According to the study, it is worth noting that the developed information technology for analyzing and forecasting time series with fractal properties based on linguistic modelling surpasses its competitors/analogues in terms of all indicators, indicating the high quality of the development and the possibility of implementing the system in real work when required.

The verification of the efficiency of information technology for analyzing and forecasting time series with fractal properties has confirmed, based on real data, the possibility of ensuring objectivity during forecasting. The use of information technology for analysis and forecasting of time series with fractal properties will allow ensuring a high level of forecasting with the fullest possible realization of the analytical system. The developed information technology for analysing and forecasting time series with fractal properties is universal and allows adapting the process of planning the forecast of financial time series to the level of the development of the initial series.

REFERENCES

1. Valeev, S.G., Kuvayskova, Yu.E., Gubaidullina, S.A. (2008). Application of Multifractal Analysis in the Description of Time Series in Technology and Economics. *Bulletin of the Ulyanovsk State Technical University*, 2(42), 23-27.
2. Peters, E. (2004). Fractal Analysis of Financial Markets: Application of Chaos Theory in Investments and Economics, Internet-Trading, Moscow, 304 p.
3. Losev, A.F. (2010). Introduction to the General Theory of Language Models. 3rd ed., Editorial URSS, Moscow, 296 p.
4. Corpus Linguistics: Monograph / V.A. Shyrokov, O.V. Buhakov, T.O. Hriaznukhina, O.M. Kostyshyn, M.Yu. Kryhin; NAS of Ukraine, Dovira, Kyiv, 2005, 472 p.
5. Selivanova, O.O. (2008). Modern Linguistics: Directions and Problems, Dovkillia-K, Poltava, 712 p.
6. Povetkina, Yu.V. (2012). Modelling as a Method of Linguistic Research. Philological Sciences. Problems of Theory and Practice, 6.6, 132-137.
7. Lukach, M.O. (2013). Types of Linguistic Models and Their Application for Solving Linguistic Problems. Bulletin of the Lviv Polytechnic National University, 770: Information Systems and Networks, 143-153.
8. Povetkina, Yu.V. (2012). Modelling as a Method of Linguistic Research. *Philological Sciences. Problems of Theory and Practice*, 6, 132-137.
9. Kravtsova, Yu.V. (2014). Modelling in Modern Linguistics. *Bulletin of the Zhytomyr Ivan Franko State University*, 5, 181-189.
10. Baklan, I.V. (2015). Fractal Properties in the Linguistic Modelling of Dynamic Processes. *Bulletin of the Kherson National Technical University*, KhNTU, Kherson, 3(54), 210-213.
11. Baklan, I.V., Nedashkivskyi, Ye.A. (2016). Linguistic Modelling of Time Series. NTUU “KPI”, 11-14.
12. Svetlov, K.V. (2016). Stochastic Methods of Analysis of the Market of Borrowings: Thesis for ... the Candidate of Economic Sciences: 08.00.13. St. Petersburg, 143 p.
13. Liuu, Yu. (2007). Methods and Algorithms of Financial Mathematics, Binom, Laboratory of Knowledge, Moscow, 751 p.
14. Borodin, A.N. (2012). Random Processes, Lan, Saint Petersburg, 640 p.
15. Gibson, R., Lhabitant, F., Talay, D. (2010). Modeling the Term Structure of Interest Rates: A Review of the Literature. *Foundations and Trends in Finance*, 5, 1-2.
16. Yarushkina, N.G., Afanasyeva, T.V., Prefileva, I.G. (2010). Intelligent Analysis of Time Series: Tutorial, Ulyanovsk: UlGTU, 324 p.
17. Mandelbrot, B. (2004). Fractals, Case and Finance, Izhevsk: SIC Regular and Chaotic Dynamics”, Moscow.
18. Starchenko, N.V. (2006). Index of Fractality and Local Analysis of Chaotic Time Series. : Thesis for Candidate of Physics and Mathematics: 01.01.03: Defended on February 15, 2006.
19. Atkins, P.W., Beran, J.A. (1992), General Chemistry Scientific American Books, N.Y., P. 922.