

A COMBINED APPROACH TO MODELING NONSTATIONARY HETEROSCEDASTIC PROCESSES

Tymoshchuk O. L. – PhD, Associate Professor at the Department of Mathematical Methods of System Analysis, Institute for Applied Systems Analysis at the NTUU «Igor Sikorsky Kyiv Polytechnic Institute», Kiev, Ukraine.

Huskova V. H. – Post-graduate student at the Department of Mathematical Methods of System Analysis, Institute for Applied Systems Analysis at the NTUU «Igor Sikorsky Kyiv Polytechnic Institute», Kiev, Ukraine.

Bidyuk P. I. – Dr. Sc., Professor at the Department of Mathematical Methods of System Analysis, Institute for Applied Systems Analysis at the NTUU «Igor Sikorsky Kyiv Polytechnic Institute», Kiev, Ukraine.

ABSTRACT

Context. Nonlinear nonstationary processes are observed today in various fields of studies: economy, finances, ecology, demography etc. Very often special approaches are required for model development and forecasts estimation for the processes mentioned. The modeling methodologies have to take into consideration possible uncertainties that are encountered during data processing and model structure and parameter estimation.

Objective. To develop a modified methodology for constructing models for nonlinear processes that allows for achieving high quality of forecasts. More specifically heteroscedastic processes are considered that create a wide class of nonlinear nonstationary processes and are considered in many areas of research.

Method. To reach the aim of the study mentioned the following methods are used: systemic approach to model building and forecasting, modified methodology for modeling nonlinear processes, methods for identification and taking into consideration possible uncertainties. To cope with the structural uncertainties following techniques: refinement of model order applying recurrent adaptive approach to modeling and automatic search for the “best” structure using complex statistical criteria; adaptive estimation of input delay time, and the type of data distribution with its parameters; describing detected nonlinearities with alternative analytical forms with subsequent estimation of the forecasts generated.

Results. The proposed modified methodology for modeling nonlinear nonstationary processes, adaptation scheme for model building, new model structures proposed. As a result of performing computational experiments, it was found that nonlinear models constructed provide a possibility for computing high quality forecasts for the process under study and their variance.

Conclusions. Application of the modeling methodology proposed provides a possibility for structural and parametric adaptation of the models constructed with statistical data. The models developed exhibit acceptable adequacy and quality of short-term forecasting.

KEYWORDS: nonlinear nonstationary processes, systemic approach to modeling, structural and parametric adaptation, combined models, uncertainties in modeling and forecasting.

ABBREVIATIONS

ACF is an autocorrelation function;
AIC is an Akaike information criterion;
AR is an autoregression;
ARCH is an autoregressive conditional heteroscedasticity;
ARMA is an autoregression with moving average;
DW is a Durbin-Watson statistic;
EGARCH is an exponential generalized autoregression with conditional heteroscedasticity;
GARCH is a generalized autoregressive conditionally heteroscedastic;
GMDH is a group method of data handling method;
GNP is a gross national product;
HSP is a heteroscedastic processes;
IDA is an intellectual data analysis;
KF is a Kalman filter;
MAE is a mean absolute error;
MAPE is a mean absolute percentage error;
MCMC is a Markov chain Monte Carlo;
ML is a maximum likelihood;
MLNSVM is a modified log-normal model of stochastic volatility;
NLS is a nonlinear least squares;
NNP is a nonlinear nonstationary processes;
SS is a state space;

U is a Theil coefficient.

NOMENCLATURE

$h(k)$ is a conditional variance;
 $y_1(k)$ is a gross national product logarithm;
 $y_2(k)$ is a logarithm of tax deductions;
 \bar{y} is a sample mean for the data window w selected for computing conditional variance;
 $\mathbf{x}(k)$ is a vector of state variables;
 $x_1(k)$ is a level of internal investment;
 $x_2(k)$ is a level of external investment;
 $\mathbf{z}(k)$ is a vector of time delayed values of dependent variable $y(k)$;
 $z_i(k), i = 1, 2, \dots, m$ is an explaining variables for $x(k)$;
 α_i, γ_i is a model parameters to be estimated with maximum likelihood or Monte Carlo Markov chain procedures;
 $\varphi_i(x)$ is a set of linear and nonlinear functions;
 $\varepsilon(k)$ is a random process that influences financial process under study;
 $\hat{\varepsilon}^2(k)$ is an account effects of absolute values.

INTRODUCTION

The study is focused on combined models development for forecasting nonlinear nonstationary heteroscedastic processes in economy, finances, ecology and other areas. Most of the processes taking place in these areas belong to the class of nonlinear and nonstationary due to many random factors influencing their evolution in time. Financial processes are influenced by various random shocks happening inside of countries (local inflation, high competition between companies, unstable market situations) where they are originated and by outside shocks in the form of general economic crisis, fast variability of prices on energy resources, raw materials and food etc. Processes in modern ecology and climatology also often exhibit unstable nonstationary development with nonlinearities. Thus, the problem of constructing forecasting models for the process mentioned is urgent and it is considered by many researchers [1–4].

A simplified classification of NNP in economy, finances, demography, ecology, is shown in Fig. 1. The figure provides information for a general situation in the area of study.

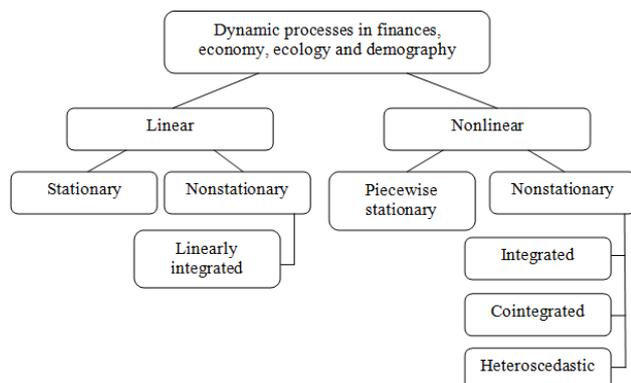


Figure 1 – A simplified classification of dynamic processes

The object of study is the linear process can be nonstationary only in a case when it contains linear trend. Nonlinear processes can be piecewise stationary mostly in their stable mode of operation. NNP are most commonly met in many areas of study. They include nonlinear integrated process with trends of order two and higher, cointegrated processes with the same degree of integration, and heteroscedastic processes. The latter suggest simultaneous constructing of the following two model types: first model describing evolution of the process (amplitude) itself, and the second model describing dynamics of conditional variance that is widely used in practice for solving the problems of diagnostics (technical, medical, financial and economic), risk analysis in various spheres including, stock trading, investments etc. The article is concentrated on the studying of nonlinear nonstationary processes with time varying variance and mean on the interval of study.

The subject of study is the known mathematical models and construct the new ones using statistical data. An approach to modeling the processes mentioned is proposed providing a possibility for constructing the models of acceptable adequacy.

The purpose of the work is in development of methodology and system for constructing mathematical models of nonlinear nonstationary (heteroscedastic) processes of acceptable adequacy and providing a possibility for computing high quality of forecasts. Special software was developed on the purpose and illustrative examples are given.

1 PROBLEM STATEMENT

The purpose of the study is as follows: (1) development of some generalized methodology for modeling and forecasting nonlinear non-stationary heteroscedastic processes both for the process itself and its conditional variance; (2) development of combined model based upon linear and nonlinear regression; (3) performing of computational experiments directed towards model constructing for NNP and computing short term forecasts for the process amplitudes and their conditional variances; (4) performing a comparative analysis of the results achieved with various models constructed with the computational experiments.

The formal problem statement is as follows: statistical data are given characterizing nonlinear nonstationary processes with arbitrary probabilistic distribution $\{y(k)\} \sim Dist(\mu(k), \sigma_y^2(k))$, $k=1, \dots, N$, where

$\mu(k) \neq \text{const}$ is time varying mean; $\sigma_y^2(k) \neq \text{const}$ is process variance that is varying on the interval of the process study. The statistical data parameters are subject to the following restrictions: $\mu(k) < \infty$; $0 < \sigma_y^2(k) < \infty$

on the interval of studying. It is necessary to construct mathematical models for the process mentioned of the general structure: $y(k) = F[y(k), x(k), \theta, \varepsilon(k)]$, where $x(k)$ is possible independent variable; θ is model vector parameters; $\varepsilon(k)$ is stochastic process induced by external disturbances and measurement errors. The constructed model selection criteria are based upon determination coefficient, R^2 , DW , $MAPE$, Theil coefficient, U , and combined criterion:

$J = |1 - R^2| + |2 - DW| + U \rightarrow \min_{\theta}$. The models will be constructed according to the methodology proposed.

2 REVIEW OF THE LITERATURE

A comprehensive study of nonlinear nonstationary HSP is given in [5]. Here methodology of modeling and forecasting HSP using ARCH models is provided based on statistical data. Several parameter estimation methods are considered including maximum likelihood. A set of model misspecification tests is provided and fractionally

integrated models are analyzed. The volatility forecasting problem is considered with one-step-ahead and multistep forecasting using actual statistical data and known computer based systems for time series analysis.

A comparative study of VaR estimation methods for structured products is considered in [6]. The author studies forecast accuracy of GARCH and EGARCH models by means of multivariate extension of the Diebold and Mariano [7] test proposed for non-nested models. It was shown that GARCH and EGARCH models often produce higher capital requirements than ARCH models in order to avoid default risks.

The article is concentrated on the problem of modeling and forecasting heteroscedastic processes using systemic approach to solving the problems. Some special software was developed for identification and taking into consideration possible uncertainties relevant to statistical data processing.

Thus, among possible data uncertainties are the following: missing measurements, random external influences (state noise), short low informative samples, possible outliers, noise corrupted measurements etc. These uncertainties are relatively easy identifiable and easy to correct with available data processing techniques for the lost measurements imputation, bootstrap analysis and filtering techniques. The model structure uncertainties aforementioned are provoked by poor data structure that does not contain enough information for model the structure and parameter estimation.

The parametric uncertainties are closely related to the quality of available data and their influence is usually related to biased parameter estimates and low model adequacy. The remedy for avoiding the bias is in application of alternative parameter estimation techniques, among which are the following: maximum likelihood and MCMC procedures. A substantial improvement of forecasts estimates can be reached with simultaneous hiring of alternative forecasting methods and forecasts combining techniques [8–12].

3 MATERIALS AND METHODS

Fig. 2 illustrates the scheme of the generalized approach proposed that is suitable for modeling both linear and nonlinear processes as well as financial risk estimation.

First exponential smoothing, optimal Kalman or elliptical filtering is applied to input data to perform smoothing and prepare the data to model constructing procedures. After smoothing the model can be linear or nonlinear dependently on the specific effects contained in the input data. Generally data may contain linear and nonlinear part what will result in linear and nonlinear models (or components of a single model) simultaneously. In a case of heteroscedastic processes analysis usually the following problems are solved: (1) constructing a model that describes formally the process itself; (2) constructing a model describing dynamics of conditional variance; (3) estimation of short-term forecasts using both types of models.

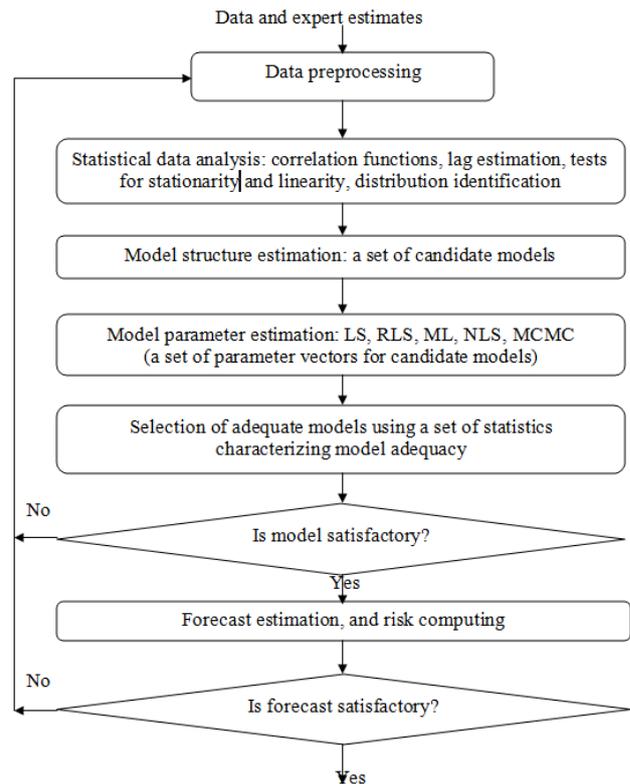


Figure 2 – An approach to modeling heteroscedastic processes and financial risk estimation

The methodology proposed for modeling nonlinear non-stationary processes includes the steps formulated below.

– Data pre-processing before modeling that includes the following operations: imputation of missing observations, measurements normalization, digital and/or optimal filtering, statistical processing of outliers etc.

– Identification and elimination (or decreasing of influence) of data uncertainties using the following data processing procedures: estimation of non-measurable values, estimation of statistical parameters for observations (variance, covariance, mean, median), appropriate data structuring, analysis of random external influences and their distributions, coping with measurement errors, taking into consideration probabilistic uncertainties etc.

– Model structure estimation using statistical (correlation) and probabilistic (mutual information) data analysis techniques providing a possibility for determining the following elements of a model structure: model dimension (number of equations); model order (order of autoregression and moving average parts); nonlinearity and its type (nonlinearity with respect to variables or with respect to parameters); delay time estimate, and type of probabilistic distribution for the variables under study; it is always desirable to estimate structures for several candidate models to get a possibility for further selection of the best one.

– Taking into consideration possible nonlinearities. It is recommended to construct first nonlinear part of a data model using various possibilities for describing nonlinear parts: polynomials, exponents, squared harmonic func-

tions, bilinear constructions etc. Then linear part of the model is fitted using nonlinear model residuals. Good results were achieved with combining linear and nonlinear regression; linear regression and Bayesian networks; linear regression and special nonlinear functions like non-parametric kernels etc.

– Model parameters estimation using alternative methods such as NLS, ML, MCMC procedures and others providing unbiased estimates of parameters under specific distributions of variables and model structures. Application of alternative parameter estimation techniques provides a possibility for further comparison of the estimates and selection of the best ones after comparison of the models constructed.

– Computing statistical parameters characterizing candidate models adequacy and determining the most suitable (adequate) one(s) among them. It is not necessarily to leave only one model for computing forecasts, it can be a set of the “best” models based on different ideologies. The final choice is always made after model application for solving the following specific problems: forecasting, control, constructing simulation model or deep analysis of the process under study.

Computing model based forecasts for the process (under study) evolution and their quality estimation for the final selection of the best forecasting model. Here another set of statistical quality criteria is used: MAPE, Theil coefficient, mean absolute error, minimum and maximum forecasting errors etc. On the purpose of automatized model selection the following combined criteria is proposed:

$$J = \left| 1 - R^2 \right| + \left| 2 - DW \right| + U \rightarrow \min_{\theta}$$

– Testing the model(s) constructed on similar processes (model calibration).

To timely and reliably monitor the methodology implementation it is necessary to utilize in modeling system at least three sets of statistical quality parameters: data quality parameters, model adequacy and forecasts quality statistics. Also the quality criteria selected/developed should analyze alternative decisions based on the forecasts generated. Correct practical application of the methodology proposed is an important task aiming constructing adequate models suitable for solving the problems of forecasting, automatic (or semi-automatic) control, financial risk estimation, business decision making, complex system simulation etc.

Today there exist some known methodologies for studying linear and nonlinear nonstationary processes, and constructing adequate mathematical models in the aforementioned fields (including financial processes) using statistical procedures for regression analysis and SS representation [8, 9, 12]. Another popular approach to development linear and NNP models is based on IDA techniques: artificial neural networks, fuzzy sets, neuro-fuzzy models, static and dynamic Bayesian networks, complex multivariate distributions, non-parametric models, decision trees etc.

When constructing linear and nonlinear time series models it is convenient to use the proposed here unified notion of mathematical model structure:

$$S = \{r, p, m, n, d, w, l\} \quad (1)$$

where r is model dimension (number of equations); p is model order (maximum order of differential or difference equation used for process description); m is a number of independent variables; n is nonlinearity and its type (with respect to variables or parameters); d is input delay time; w is external stochastic disturbance and its type of distribution; l represents possible restrictions on variables and parameter values. The aforementioned elements of a model structure are estimated using available statistical tests and correlation analysis techniques: correlation matrix, ACF, partial ACF, bi-correlations, and higher order correlation functions.

Some nonlinear models that were utilized successfully in practice resulted from former studying of econometric time series. Nonlinear regression of the type shown below is used to describe GNP and tax deductions [10]:

$$y_1(k) = a_0 + a_1 y_1(k-1) + b_{12} \exp(y_2(k)) + a_2 x_1(k) x_2(k) + \varepsilon_1(k) \quad (2)$$

$$y_2(k) = c_0 + c_1 y_2(k-1) + b_{21} \exp(y_1(k)) + c_2 x_1(k) x_2(k) + \varepsilon_2(k) \quad (3)$$

Also widely used and convenient model structure is generalized bilinear model:

$$y(k) = a_0 + \sum_{i=1}^p a_i y(k-i) + \sum_{j=1}^q b_j v(k-j) + \sum_{i=1}^m \sum_{j=1}^s c_{i,j} y(k-i) v(k-j) + \varepsilon(k), \quad (4)$$

where p, q, m and S are positive numbers that represent the model order [8].

Very often modeling nonlinear processes is based on linear combination of linear and nonlinear components:

$$y(k) = \beta^T \mathbf{z}(k) + \sum_{i=1}^p \alpha_i \varphi_i(\theta_i^T \mathbf{z}(k)) + \varepsilon(k), \quad (5)$$

where $\mathbf{z}(k)$ is a vector of time delayed values of dependent variable $y(k)$, as well as former and current values of independent explaining variables $\mathbf{x}(k)$ with appropriate time shift. Here $\varphi_i(x)$ is a set of (linear and nonlinear) functions that include the following components: power function $\varphi_i(x) \equiv x^i$; trigonometric functions $\varphi_i(x) = \sin x$ or $\varphi_i(x) = \cos x$ etc.; this equation can be

expanded with quadratic form of the type $\mathbf{z}^T(k) \mathbf{A} \mathbf{z}(k)$; $\varphi_i(x) = \varphi(x)$, $\forall i$, where $\varphi(x)$ is a link function, for example appropriate probability density function or logistic function of the type:

$$\varphi(x(k, z)) = \frac{1}{1 + \exp(-x(k, z))}. \quad (6)$$

$$x(k) = \alpha_0 + \alpha_1 z_1(k) + \dots + \alpha_m z_m(k) + \varepsilon(k), \quad (7)$$

where $z_i(k)$, $i = 1, 2, \dots, m$ are explaining variables for $x(k)$, and $\varphi(x(k, z))$ respectively.

Another general class of nonlinear models can be presented in the following form:

$$\mathbf{y}(k) = \sum_{j=1}^p \varphi_j(\mathbf{x}(k-1)) \mathbf{y}(k-j) + \mu(\mathbf{x}(k-1)) + \varepsilon(k), \quad (8)$$

where $\mathbf{y}(k)$ is $[n \times 1]$ a vector of dependent variables; $\mathbf{x}(k) = [\mathbf{y}(k), \mathbf{y}(k-1), \dots, \mathbf{y}(k-n+1)]$ is a vector of state variables; dynamics of the variables is described by the model:

$$\mathbf{x}(k) = h(\mathbf{x}(k-1)) + \mathbf{F}(\mathbf{x}(k-1)) \mathbf{x}(k-1) + \mathbf{v}(k). \quad (9)$$

The model based forecasts can be computed, for example, with six selected (or more) techniques given in fig. 3. Regression model (autoregression or with moving average) is used for generating forecast as well as its transformed version into SS form for further use by optimal KF. Adaptive version of KF provides a possibility for forecasting and on-line (off-line) estimation of state disturbances and measurement noise covariance. The GMDH generates models in the general form of Kolmogorov-Gabor polynomial, and the last three methods are related to the intellectual data analysis techniques. Thus, here we have the combination of classic regression (statistical) approach with intellectual data analysis techniques. The best result of combining the forecasts is achieved when variances of forecasting errors for selected forecasting techniques do not differ substantially. Some other possibilities for hiring nonlinear models for formal description of NNP are shown in table 1. Most of them have been used successfully for modeling and forecasting economic and financial processes.

The models (No. 1–8) presented in Table 1 have known structure though it can be modified in the process of adaptation using specific statistical data. Model 1 was successfully applied for trend modeling of various orders together with short-term deviations from conditional mean. Models 2, 4 can describe bilinear and exponential nonlinearities or nonlinearity with saturation (model 3). Models 5 and 6 are used for description of conditional variance dynamics while modeling heteroscedastic process. The last one turned out to be the best model for short term forecasting of variance

Table 1 – Some linear and nonlinear models for describing process dynamics

No	Model description Formal model structure
1	AR + polynomial of time $y(k) = a_0 + \sum_{i=1}^p a_i y(k-i) + b_1 k + \dots + b_m k^m + \varepsilon(k)$ $k = 0, 1, 2, \dots$ is discrete time; $t = kT_s$; T_s is sampling time.
2	Generalized bilinear model $y(k) = a_0 + \sum_{i=1}^p a_i y(k-i) + \sum_{j=1}^q b_j v(k-i) = \sum_{i=1}^m \sum_{j=1}^s c_{ij} y(k-i) v(k-j) + \varepsilon(k)$
3	Logistic regression $\varphi(x(k, z)) = \frac{1}{1 + \exp(-x(k, z))}$, $x(k) = \alpha_0 + \alpha_1 z_1(k) + \dots + \alpha_m z_m(k) + \varepsilon(k)$
4	Nonlinear extended econometric autoregression $y_1(k) = a_0 + a_1 y_1(k-1) + b_{12} \exp(y_2(k)) + a_{21} x_1 x_2 + \varepsilon_1(k)$, $y_2(k) = c_0 + c_1 y_2(k-1) + b_{21} \exp(y_1(k)) + c_2 x_1 x_2 + \varepsilon_2(k)$
5	Generalized autoreg-ression with conditional heteroscedasticity $h(k) = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon^2(k-i) + \sum_{i=1}^p \beta_i h(k-i)$.
6	Exponential general-ized autoregression with conditional heteroscedasticity $\log[h(k)] = \alpha_0 + \sum_{i=1}^p \alpha_i \frac{ \varepsilon(k-i) }{\sqrt{h(k-i)}} + \sum_{i=1}^p \beta_i \frac{\varepsilon(k-i)}{\sqrt{h(k-i)}} + \sum_{i=1}^q \gamma_i \log[h(k-i)] + v(k)$
7	Nonparametric model with functional coefficients $y(k) = \sum_{i=1}^p \{\alpha_i + (\beta_i + \gamma_i y(k-d)) \cdot \exp(-\theta_i y^m(k-d))\} + \varepsilon(k)$
8	Radial basis function $f_\theta(x(k)) = \sum_{i=1}^M \lambda_i \exp\left(-\frac{(x(k) - \mu_i)^2}{2\sigma_i^2}\right) + \varepsilon(k)$, $\theta = [\mu_i, \sigma_i, \lambda_i]^T$; $M = 2, 3, \dots$
9	State-space representation $\mathbf{x}(k) = \mathbf{F}[\mathbf{a}(k), \mathbf{x}(k-1)] + \mathbf{B}[\mathbf{b}(k), \mathbf{u}(k-d)] + \mathbf{w}(k)$
10	Neural networks – Selected (constructed) network structure
11	Fuzzy sets and neuro-fuzzy models – Combination of fuzzy variables and neural network model
12	Dynamic Bayesian networks – Probabilistic Bayesian network structure constructed with data and/or expert estimates
13	Multivariate distributions – Like copula application for describing multivariate distribution
14	Immune systems – Immune algorithms and combined models

in about 90% of applications performed by the authors. Models 7, 8, 9 can describe arbitrary nonlinearities with respect to variables of order 3–5 or higher. Fuzzy sets based approach to modeling supposes generating of a set of

rules that could describe with acceptable quality functioning of selected processes and formulate appropriate logical inference. Neural networks and fuzzy neural networks are suitable for modeling sophisticated nonlinear functions in conditions of availability of some unobservable variables. Dynamic Bayesian networks and multivariate distributions are statistical/probabilistic models that could describe complex multivariate processes (systems) with generating final result of their application in the form of conditional probabilities (probabilistic inference).

The model based forecasts for nonstationary process itself can be computed, for example, with six selected (or more) techniques given in Fig. 3. Regression model (autoregression or with moving average) is used for generating forecast as well as its transformed version into SS form for further use by optimal Kalman filter. Adaptive version of KF provides a possibility for forecasting and on-line (off-line) estimation of state disturbances and measurement noise covariance. The GMDH generates models in the general form of Kolmogorov-Gabor polynomial, and the last three methods are related to the intellectual data analysis techniques. Thus, here we have the combination of classic regression (statistical) approach with intellectual data analysis techniques. The best result of combining the forecasts is achieved when variances of forecasting errors for selected forecasting techniques do not differ substantially. Some other possibilities for hiring linear and nonlinear models are shown in table 1.

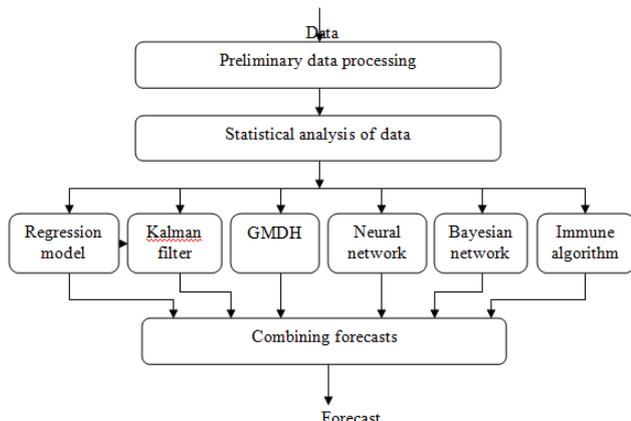


Figure 3 – Illustration of the combining principle for alternative forecasts

Models of conditional variance. Today there exist a wide variety of models describing dynamic of conditional variance. Here we will mention only those that were tested during experimental studies using statistical data.

We propose modified log-normal model of stochastic volatility in the following form:

$$\begin{aligned}
 y(k) &= e^{-\frac{h(k)}{2}} \varepsilon(k), \quad \varepsilon(k) \sim N(0,1), \quad k = 1, \dots, N \\
 h(k) &= \alpha + \beta_1 h(k-1) + \dots + \beta_p h(k-p) + \\
 &+ \gamma_1 y(k) + \gamma_2 \varepsilon^2(k-q) + \eta(k), \quad \eta(k) \sim N(0, \sigma_\eta^2).
 \end{aligned} \tag{10}$$

where $h(k) = \frac{1}{w} \sum_{l=k-\frac{w-1}{2}}^{k+\frac{w-1}{2}} [y(l) - \bar{y}]^2$ is conditional

variance that characterizes dynamics of the financial time series under study. The model proposed also takes into consideration necessary depth of memory for the process formed by the values of conditional variance $h(k)$, as well as the values of basic variable $y(k)$, that provides a possibility for taking into consideration influence of positive and negative disturbances to conditional variance. The performed testing of the model proved the possibility of its practical application for forecasting conditional volatility and financial risk estimation.

One of the best results so far with forecasting the variance was achieved with exponential generalized autoregression with conditional heteroscedasticity (EGARCH) shown below [2]:

$$\begin{aligned}
 \log[h(k)] &= \alpha_0 + \sum_{i=1}^p \alpha_i \frac{|\varepsilon(k-i)|}{\sqrt{h(k-i)}} + \\
 &+ \sum_{i=1}^p \beta_i \frac{\varepsilon(k-i)}{\sqrt{h(k-i)}} + \sum_{i=1}^q \gamma_i \log[h(k-i)] + v(k)
 \end{aligned} \tag{11}$$

where $h(k)$ is conditional variance; $\varepsilon(k)$ is random process that influences financial process under study; α_i, γ_i are model parameters to be estimated with maximum likelihood or MCMC procedures; $v(k)$ are model residuals.

4 EXPERIMENTS

The analysis included three experiments. In the first experiment time series was applied for studying of the gold prices within the period between 2005–2006 years. The statistical characteristics showing constructed models and forecasts quality are given in Table 2. Here the case is considered when optimal KF was not used for preliminary data processing smoothing.

The second experiment was statistical analysis of the time series selected with application of Goldfeld-Quandt test proved that gold prices data create heteroscedastic process with time varying conditional variance. As far as the variance is one of the key parameters that is used in the rules for performing trading operations it is necessary to construct appropriate forecasting models. Table 4 contains statistical characteristics of the models constructed as well as quality of short-term forecasting. To solve the problem we used GARCH models together with description of the processes trend which is rather sophisticated (high order process). The models of this type GARCH demonstrated low quality of short-term forecasts, and quite acceptable EGARCH one-step ahead forecasting properties.

Forecasting financial process (stock prices) using combined (linear + nonlinear) model – third experiment. The combined model includes optimal and digital filters, linear regression models and nonlinear logit model (Fig. 4).

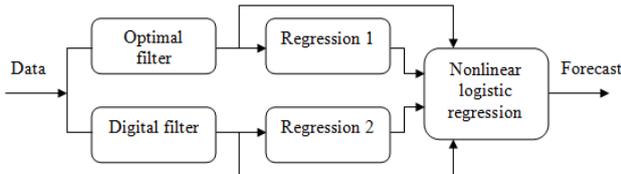


Figure 4 – Combined model: filtering + linear regression + nonlinear regression

The purpose of using the filters is to perform data smoothing (suppressing high frequency components) and this way prepare it for modeling. Besides, application of optimal KF additionally provides a possibility for solving the following problems: estimation of non-measurable state vector components; variance (covariance) estimation for observations and the state vector; and short-term forecasting in a case of necessity.

The data used reflect dynamics of stock prices together with technical analysis indicators. The indicators computed on the basis of actual dynamics of prices create useful data for modeling and forecasting financial processes. Among widely used indicators are the following: Pivot Point, Woodie’s Pivot Points, Fibonacci’s Pivot Points, Camarilla’s Pivot Points.

5 RESULTS

Experiment 1 shown that the best model turned out to be AR(1) + trend of 4th order. It provides a possibility for one step ahead forecasting with mean absolute percentage error of about 3.19%, and Theil coefficient is $U = 0.024$.

Table 2 – Models and forecasts quality without Kalman filter application

Model type	Model quality			Forecast quality			
	R^2	$\sum e^2(k)$	DW	MS	MAE	MAPE	Theil
AR(1)	0.99	25644.67	2.15	49.82	41.356	8.37	0.046
AR(1,4)	0.99	25588.10	2.18	49.14	40.355	8.12	0.046
AR(1) + 1st order trend	0.99	25391.39	2.13	34.39	25.109	4.55	0.032
AP(1,4) + 1st order trend	0.99	25332.93	2.18	34.51	25.623	4.67	0.032
AR(1) + 4th order trend	0.99	25173.74	2.12	25.92	17.686	3.19	0.024

The Theil coefficient shows that this model is generally good for short-term forecasting. Statistical characteristics of the models and respective forecasts computed with KF application for data smoothing are given in table 3. Here optimal filter (with random walk model) played positive role what is supported by respective statistical quality parameters.

Again the best model turned out to be AR(1) + trend of 4th order. It provides a possibility for one step ahead forecasting with mean absolute percentage error of about 2.71%, and Theil coefficient is: $U = 0.019$. Thus, in this case the results achieved are better than in previous modeling and short-term forecasting without filter application.

Table 3 – Models and forecasts quality with application of Kalman filter

Model type	Model quality			Forecast quality			
	R^2	$\sum e^2(k)$	DW	MSE	MAE	MAPE	Theil
AR(1)	0.99	24376.32	2.11	45.21	39.73	7.58	0.037
AR(1,4)	0.99	24141.17	2.09	47.29	38.75	7.06	0.035
AR(1) + 1st order trend	0.99	23964.73	2.08	31.15	22.11	3.27	0.029
AR(1) + 4th order trend	0.99	22396.83	2.04	21.35	13.52	2.71	0.019

Experiment 2. Statistical analysis of the time series selected with application of Goldfeld-Quandt test proved that gold prices data create heteroscedastic process with time varying conditional variance. As far as the variance is one of the key parameters that is used in the rules for performing trading operations it is necessary to construct appropriate forecasting models. Table 4 contains statistical characteristics of the models constructed as well as quality of short-term variance forecasting. To solve the problem we GARCH models together with description of the processes trend which is rather sophisticated (high order process). The models of this type GARCH demonstrated low quality of short-term forecasts, and quite acceptable EGARCH one-step ahead forecasting properties.

Table 4 – Results of modeling and forecasting conditional variance

Model type	Model quality			Forecast quality			
	R^2	$\sum e^2(k)$	DW	MSE	MAE	MAPE	Theil
GARCH (1,7)	0.99	153639	0.113	972.5	–	517.6	0.113
GARCH (1,10)	0.99	102139	0.174	458.7	–	211.3	0.081
GARCH (1,15)	0.99	80419	0.337	418.3	–	121.6	0.058
MLNSVM (3, 7)	0.99	61377	0.405	79.5	–	9.97	0.027
EGARCH (1, 7)	0.99	45184	0.429	67.8	–	8.74	0.023

Thus, the best model constructed was exponential GARCH(1,7). The achieved value of MAPE = 8.74% comprises very good result for forecasting conditional variance. The second was the model proposed MLNMSV that has two order parameters: p and q , with MAPE = 9.97%. Further improvements of the forecasts were achieved with application of the adaption scheme [1, 7]. An average improvement of the forecasts was in the range between 0.5–1.5%, what justifies advantages of the approach proposed. Combination of the forecasts generated with different forecasting techniques helped to further decrease mean absolute percentage forecasting error for about 0.5 – 0.8% in this particular case. It should be stressed that analysis of heteroscedastic processes is very popular today due to multiple engineering, economic and financial applications of the models and forecasts based upon them. The methodology developed also supposes constructing of combined models including classical statistical regression and probabilistic models in the form of Bayesian networks.

Experiment 3. The following notations were used for the indicators: Pivot is “turning point”; S1 is first level of support; S2 is second level of support; S3 is third level of support; R1, R2 and R3 are first, second and third levels of resistance. All three indicators work in a similar way. If open price is higher than, Pivot, and the price starts moving down, then from the view of technical analysis it is probable that after reaching Pivot value, the price will go up. In the case when this did not happen and the price continues moving down then, from the point of view of technical analysis, it is probable that the price may perform a turn or stop moving down when it approaches the level of, S1. The same situation takes place for S2 and, S3. However, price reaches the level of S3 very seldom. Even when the price gets down lower than, S3, then probability of its next moving down is very low. The same situation takes place in the case of a price moving up but here instead of indicators S1, S2, S3 should be used R1, R2, R3.

To construct regression models first were used minimum daily exchange rate for the pair USD/CHF (257 values for 2007). The first multiple regression model was constructed for minimum price with independent variables, S1, S2, S3:

$$y(k) = -0.0164 + 0.1724 \cdot S1(k) - 5.3885 \cdot S2(k) + 4.2251 \cdot S3(k), \quad (12)$$

where k is discrete time as before. This model does not include autoregression part because this variant results in degeneration of measurement matrix. The model constructed has the following statistical characteristics of adquacy: $R^2 = 0.9899$; $SSR = 0.0038$; $AIC = -8.277$;

$DW = 1.9542$ All statistics exhibit quite acceptable values. Statistical characteristics that characterize quality of one-step-ahead prediction are as follows: $RMSE = 0.0038$; $MAE = 0.003$; $MAPE = 0.2525\%$; $U = 0.0016$. Number of correct forecasts for direction of price evolution was 187 or 73.05%.

As far as evolution of stock prices very often takes place along arbitrary nonlinear trajectories, to describe correctly the processes it is necessary to use nonlinear models. In some cases rather simple approach to solving the problem is application of nonlinear logistic regression. It can be helpful for forecasting direction of a price moving during the next step of its evolution. If at the moment of time $t+1$ price will be higher than at the moment, t , denote this situation (price growth) as “1”, and price decreasing denote as “0”. These values were used as inputs (for left-hand-side) of logistic regression and classification tree. The same notation was used for growth and decreasing of the *Pivot Point* indicator with the following additional notations for respective variables: $\hat{S}1$, $\hat{S}2$, $\hat{S}3$, \hat{P} , $\hat{R}1$, $\hat{R}2$, $\hat{R}3$. These variables were used as independent ones in logistic regression and classification tree. The following model was constructed for the minimum price:

$$g_{\min}(x_2) = \frac{e^{x_2(k)}}{1 + e^{x_2(k)}}, \quad (13)$$

$$x_2(k) = -0.683 + 0.033 \cdot \hat{S}1(k) + 0.055 \cdot \hat{S}2(k) + 0.055 \cdot \hat{S}3(k) - 0.4 \cdot \hat{P}(k) + 1.627 \cdot \hat{R}1(k) + 0.133 \cdot \hat{R}2(k) + 0.264 \cdot \hat{R}3(k), \quad (14)$$

$$g_{\max 2}(x_1) = \frac{e^{x_1(k)}}{1 + e^{x_1(k)}}, \quad (15)$$

$$x_1(k) = -0.871 + 0.127 \cdot \hat{S}1(k) + 0.404 \cdot \hat{S}2(k) - 0.36 \cdot \hat{S}3(k) - 0.247 \cdot \hat{P}(k) + 0.616 \cdot \hat{R}1(k) + 0.079 \cdot \hat{R}2(k) - 0.009 \cdot \hat{R}3(k) + 2.159 \cdot \hat{y}(k). \quad (16)$$

Results of forecasting direction for evolution of minimum price are presented in Table 5.

Table 5 – Results of forecasting direction for evolution of minimum price

Model type	Probability of correct direction forecast
Linear regression with indicators	73.05%
Logistic regression with Pivot Point	68.36%
Classification tree with Pivot Point	68.36%
Logistic regression with Pivot Point + forecast by linear model	75.03%
Classification tree with Pivot Point + forecast by linear model	73.83%
Logistic regression with Woodie’s Pivot Point	68.75%
Classification tree with Woodie’s Pivot Point	68.36%
Logistic regression with Woodie’s Pivot Point + forecast by linear model	75.39%
Classification tree with Woodie’s Pivot Point + forecast by linear model	73.83%
Logistic regression with Fibonacci’s Pivot Point	65.23%
Classification tree with Fibonacci’s Pivot Point	65.23%
Logistic regression with Fibonacci’s Pivot Point + forecast by linear model	74.22%
Classification tree with Fibonacci’s Pivot Point + forecast by linear model	73.83%

Thus, in both cases (logistic regression and classification tree) the best results were achieved with the use of additional forecast achieved by the linear model. The statistical quality characteristics of the forecasts achieved show high quality of the forecasts and possibility of their use in trading rules.

6 DISCUSSION

Based on table 5 we got the following results: for the threshold value of probability 0.44 the first type error accepted the value of about 53, and second type error was 28. The number of correctly forecasted directions for price evolution was 175 or 68.36%. The use of the classification tree (CHAID algorithm) with threshold value 0.35 practically led to the same result.

To improve quality of the forecasts the models of logistic regression and classification tree were augmented with forecasts of price evolution generated by linear regression using the same notation for the growth and decreasing, where linear regression output that accepts the value of “1” if forecast shows growth of the price, and “0”

if forecast shows decreasing of the price. For the threshold value of probability 0.39 the first type error accepted the value of about 39, and second type error was 25. The number of correctly forecasted directions for price evolution was 192 or 75.0%. When classification tree was used for the threshold value of probability 0.32 the first type error accepted the value of about 54, and second type error was 13. The number of correctly forecasted directions for price evolution was 189 or 73.83%.

CONCLUSIONS

In this article we proposed an approach for obtaining high-quality results of the adequacy of the final model was obtained to describe formally linear and nonlinear parts of a process under study using the software developed on the purpose.

The scientific novelty of obtained results is that the original definition of model structure was given and a short review of mathematical models for nonlinear nonstationary processes was presented. Several application examples were provided that exhibit high quality of final results namely model adequacy and quality of the forecasts. A methodology was proposed for mathematical modeling and forecasting nonlinear nonstationary in economy and finances using statistical data, though the methodology is applicable to constructing models in demography, ecology as well as in many other spheres of human activity where data is available in the time series form. The methodology is based on general system analysis principles that suppose hierarchical structure of data analysis procedure, identification and processing of possible uncertainties. The systemic approach to modeling also supposes development and implementation of adaptive schemes for model structure and parameter estimation, application of statistical, probabilistic and other criteria for model constructing procedure, and to selection of the best model for specific application.

The practical significance of obtained results is that the performed tracking the computational procedures at all stages of data processing and model development with appropriate sets of quality statistics provides a possibility for achieving high quality of intermediate and final results. As instrumentation for fighting possible uncertain-

УДК 004.942: 519.216.3

КОМБІНОВАНИЙ ПІДХІД ДО МОДЕЛЮВАННЯ НЕСТАЦІОНАРНИХ ГЕТЕРОСКЕДАСТИЧНИХ ПРОЦЕСІВ

Тимошук О. Л. – канд. техн. наук, доцент кафедри математичних методів системного аналізу, Інститут прикладного системного аналізу, НТУУ «Київський політехнічний інститут ім. Ігоря Сікорського», Київ, Україна.

Гуськова В. Г. – аспірант кафедри математичних методів системного аналізу, Інститут прикладного системного аналізу, НТУУ «Київський політехнічний інститут ім. Ігоря Сікорського», Київ, Україна.

Бідюк П. І. – д-р техн. наук, професор кафедри математичних методів системного аналізу, Інститут прикладного системного аналізу, НТУУ «Київський політехнічний інститут ім. Ігоря Сікорського», Київ, Україна.

АНОТАЦІЯ

Актуальність. Нелінійні нестационарні процеси сьогодні спостерігаються в різних областях досліджень: економіка, фінанси, екологія, демографія і т. д. Дуже часто для розробки моделей і прогнозів згаданих процесів потрібні спеціальні підходи. Методології моделювання повинні враховувати можливі невизначеності, що виникають при обробці даних, структурі моделі і оцінці параметрів.

Мета роботи. Розробити модифіковану методологію побудови моделей нелінійних процесів, яка дозволяє досягти високої якості прогнозів. Більш конкретно розглянути гетероскедастичні процеси, які створюють широкий клас нелінійних нестационарних процесів і розглядаються в багатьох областях досліджень.

Метод. Для досягнення мети згаданого дослідження використовуються наступні методи: системний підхід до побудови моделей і прогнозування, модифікована методологія моделювання нелінійних процесів, методи ідентифікації та врахування можливих невизначеностей. Для подолання структурних невизначеностей використовуються методи: уточнення порядку моделей з викорис-

© Tymoshchuk O. L., Huskova V. H., Bidiuk P. I., 2019
DOI 10.15588/1607-3274-2019-2-9

ties the following techniques were used: optimal filtering procedures, missing data imputation techniques, multiple methods for model parameter estimation, and Bayesian programming approach, decision trees etc.

Prospects for further research will be focused on refinement of the methodology proposed namely model structure and parameter estimation procedures, development of new model structures for NNP and active use of ideologically different approaches to modeling and forecasting in the frames of one modeling and forecasting system.

REFERENCES

1. Bidiuk P., Prosyankina-Zharova T., Terentiev O. Modeling nonlinear nonstationary processes in economy and finances, *Advances in Intelligent Systems and Computing (Springer)*, 2018, Vol. 754, pp. 735–745. DOI: 10.1007/978-3-319-91008-6_72
2. De Gooijer J. G. Elements of nonlinear time series analysis and forecasting. Cham (Switzerland): Springer, 2017, 618 p. DOI: 10.1007/978-3-319-43252-6
3. Tsay R. S. Analysis of financial time series. New York, John Wiley & Sons, Inc., 2010, 715 p. DOI: 10.2307/4128199
4. Cheng Ch., Sa-Ngasoongsong, Beyka O. F. et al. Time series forecasting for nonlinear and nonstationary process: a review and comparative study, *IIE Transactions*, 2016, Vol. 47, pp. 1053–1071. DOI: 10.1080/0740817X.2014.999180
5. Xekalaki E., Degiannakis S. ARCH models for financial applications. Chichester, John Wiley & Sons Ltd., 2010, 520 p. DOI: 10.1002/9780470688014
6. Chen F. Y. A Comparative study of VaR estimation for structured products, *Economics Research International*, 2010, Article ID 838469, pp. 1–16. DOI: 10.1155/2010/838469
7. Diebold F. X., Mariano R. S. Comparing predicting, *Journal of Business and Economic Statistics*, 1995, Vol. 13, pp. 253–263. DOI: 10.1080/07350015.1995.10524599
8. Diebold F.X. Elements of forecasting. Ohio, Thomson South-Western, 2007, 458 p. DOI: 10.1016/j.ijforecast.2008.05.004
9. Bidiuk P. I., Trofymchuk O.M., Kozhukhivska O. A. Probabilistic and statistical uncertainty processing using decision support systems, *Visnyk of Lviv Polytechnic National University*, 2015, No. 826, pp. 237–248.
10. Diebold F.X. Forecasting in economics, business, finance and beyond. Pennsylvania, University of Pennsylvania, 2015, 607 p.
11. Hansen B.E. Econometrics. University of Wisconsin, 2017, 427 p. DOI: 10.1080/00220485.2017.1320610
12. Bidiuk P. I., Dovgij S. O., Trofymchuk O. M. DSS based on statistical and probabilistic procedures. Kyiv, Logos, 2014, 420 p.

Received 13.12.2018.

Accepted 27.02.2019.

танням періодичного адаптивного підходу до моделювання і автоматичний пошук «найкращої» структури з використанням складних статистичних критеріїв; адаптивна оцінка часу затримки введення і типу розподілу даних з їх параметрами; опис виявлених нелінійностей за допомогою альтернативних аналітичних форм з подальшою оцінкою отриманих прогнозів.

Результати. Запропоновано модифіковану методологію моделювання нелінійних нестационарних процесів, схему адаптації для побудови моделей, запропоновані нові модельні структури. В результаті виконання обчислювальних експериментів було виявлено, що побудовані нелінійні моделі дають можливість обчислювати прогнози високої якості для досліджуваного процесу та їх дисперсії.

Висновки. Застосування запропонованої методології моделювання дає можливість структурної та параметричної адаптації моделей, побудованих за статистичними даними. Розроблені моделі демонструють прийнятну адекватність і якість короткострокового прогнозування.

КЛЮЧОВІ СЛОВА: нелінійні нестационарні процеси, системний підхід до моделювання, структурна і параметрична адаптація, комбіновані моделі, невизначеність в моделюванні та прогнозуванні.

УДК 004.942: 519.216.3

КОМБИНИРОВАННЫЙ ПОДХОД К МОДЕЛИРОВАНИЮ НЕСТАЦИОНАРНЫХ ГЕТЕРОСКЕДАСТИЧЕСКИХ ПРОЦЕССОВ

Тимошук О.Л. – канд. техн. наук, доцент кафедры математических методов системного анализа, Институт прикладного системного анализа, НТУУ «Киевский политехнический институт им. Игоря Сикорского», Киев, Украина.

Гуськова В. Г. – аспирант кафедры математических методов системного анализа, Институт прикладного системного анализа, НТУУ «Киевский политехнический институт им. Игоря Сикорского», Киев, Украина.

Бидюк П.И. – д-р техн. наук, профессор кафедры математических методов системного анализа, Институт прикладного системного анализа, НТУУ «Киевский политехнический институт им. Игоря Сикорского», Киев, Украина.

АННОТАЦИЯ

Актуальность. Нелинейные нестационарные процессы наблюдаются сегодня в различных областях исследований: экономика, финансы, экология, демография и т. д. Очень часто для разработки моделей и прогнозов упомянутых процессов требуются специальные подходы. Методологии моделирования должны учитывать возможные неопределенности, возникающие при обработке данных, структуре модели и оценке параметров.

Цель работы. Разработать модифицированную методологию построения моделей нелинейных процессов, которая позволяет достичь высокого качества прогнозов. Более конкретно рассмотрены гетероскедастичные процессы, которые создают широкий класс нелинейных нестационарных процессов и рассматриваются во многих областях исследований.

Метод. Для достижения цели упомянутого исследования используются следующие методы: системный подход к построению моделей и прогнозированию, модифицированная методология моделирования нелинейных процессов, методы идентификации и учет возможных неопределенностей. Для преодоления структурных неопределенностей используются методы: уточнение порядка моделей с использованием периодического адаптивного подхода к моделированию и автоматический поиск «наилучшей» структуры с использованием сложных статистических критериев; адаптивная оценка времени задержки ввода и типа распределения данных с их параметрами; описание обнаруженных нелинейностей с помощью альтернативных аналитических форм с последующей оценкой полученных прогнозов.

Результаты. Предложена модифицированная методология моделирования нелинейных нестационарных процессов, схема адаптации для построения моделей, предложены новые модельные структуры. В результате выполнения вычислительных экспериментов было обнаружено, что построенные нелинейные модели дают возможность вычислять прогнозы высокого качества для исследуемого процесса и их дисперсии.

Выводы. Применение предложенной методологии моделирования дает возможность структурной и параметрической адаптации моделей, построенных со статистическими данными. Разработанные модели демонстрируют приемлемую адекватность и качество краткосрочного прогнозирования.

КЛЮЧЕВЫЕ СЛОВА: нелинейные нестационарные процессы, системный подход к моделированию, структурная и параметрическая адаптация, комбинированные модели, неопределенности в моделировании и прогнозировании.

ЛИТЕРАТУРА / LITERATURA

1. Modeling nonlinear nonstationary processes in economy and finances / [P. Bidyuk, T. Prosyankina-Zharova, O. Terentiev] // *Advances in Intelligent Systems and Computing* (Springer). – 2018. – Vol. 754. – P. 735–745. DOI: 10.1007/978-3-319-91008-6_72
2. De Gooijer J. G. Elements of nonlinear time series analysis and forecasting / J. G. De Gooijer. – Cham (Switzerland): Springer, 2017. – 618 p. DOI: 10.1007/978-3-319-43252-6
3. Tsay R.S. Analysis of financial time series / R. S. Tsay. – New York: John Wiley & Sons, Inc., 2010. – 715 p. DOI: 10.2307/4128199
4. Time series forecasting for nonlinear and nonstationary process: a review and comparative study / [Ch. Cheng, Sa-Ngasoongsong, O. F. Beyka et al.] // *IEE Transactions*. – 2016. – Vol. 47. – P. 1053–1071. DOI: 10.1080/0740817X.2014.999180
5. Xekalaki E. ARCH models for financial applications / E. Xekalaki, S. Degiannakis. – Chichester: John Wiley & Sons Ltd., 2010. – 520 p. DOI: 10.1002/9780470688014
6. Chen F. Y. A Comparative study of VaR estimation for structured products / F. Y. Chen // *Economics Research International*. – 2010, Article ID 838469. – P. 1–16. DOI: 10.1155/2010/838469
7. Diebold F. X. Comparing predicting accuracy / F. X. Diebold, R.S. Mariano // *Journal of Business and Economic Statistics*. – 1995. – Vol. 13. – P. 253–263. DOI: 10.1080/07350015.1995.10524599
8. Diebold F. X. Elements of forecasting / F. X. Diebold. – Ohio: Thomson South-Western, 2007. – 458 p. DOI: 10.1016/j.ijforecast.2008.05.004
9. Bidyuk P. I. Probabilistic and statistical uncertainty processing using decision support systems / P. I. Bidyuk, O. M. Trofymchuk, O. A. Kozhukhivska // *Visnyk of Lviv Polytechnic National University*. – 2015. – No. 826. – P. 237–248.
10. Diebold F. X. Forecasting in economics, business, finance and beyond / F. X. Diebold. – Pennsylvania: University of Pennsylvania, 2015. – 607 p.
11. Hansen B. E. Econometrics / B. E. Hansen. – University of Wisconsin, 2017. – 427 p. DOI: 10.1080/00220485.2017.1320610
12. Bidyuk P. I. DSS based on statistical and probabilistic procedures / P. I. Bidyuk, S. O. Dovgij, O. M. Trofymchuk. – Kyiv: Logos, 2014. – 420 p.