ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ, СИСТЕМНИЙ АНАЛІЗ ТА КЕРУВАННЯ

UDC 004.942:519.216.3
DOI: 10.20535/1810-0546.2018.1.120361

P.I. Bidyuk[1], S.P. Overmyer[2], T.I. Prosyankina-Zharova[3*], O.M. Terentiev[1]
[1]Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine
[2]Southern New Hampshire University, Hooksett, USA
[3]Institute of Telecommunications and Global Information Space of NASU, Kyiv, Ukraine

METHODOLOGY OF MODELING AND FORECASTING NONLINEAR PROCESSES IN FINANCES

Background. Most of the models of financial and economic processes are characterized by considerable computational complexity, and construction of predictions of acceptable quality for the required time horizon – by considerable efforts. Therefore, the development and implementation of effective tools for forecasting the modeling of financial and economic processes are one of the actual and practically meaningful tasks. The paper deals with the modeling and forecasting of nonlinear nonstationary processes in macroeconomics and finance using a methodology based on the principles of system analysis such as hierarchical modeling, consideration of the influence of uncertainties, optimization of the characteristics of models using complex criteria, structural and parametric adaptation. The application of the proposed methodology will improve the quality of forecasting by studying the features of the analyzed process and adapting models to new data, etc.
Objective. The purpose of this article is to develop a methodology for predictive modeling of nonstationary processes in finance and macroeconomics using statistical data, as well as its implementation in the corresponding computer system.
Methods. The methodology is based on the technologies of preliminary processing of statistical data intended to eliminate possible uncertainties, the use of correlation analysis to evaluate structure of the model and choice of methods for estimating its parameters, calculating forecast estimates and generating alternative solutions. This allows us to objectively evaluate the results obtained at each stage of solving the problem of modeling nonlinear nonstationary processes in macroeconomics and finance. The paper proposes an original methodology for determining the structure of the model and its implementation in the information system for decision support.
Results. Appropriate models were built for the selected financial and macroeconomic processes. High quality of the final result of data analysis and forecasting is achieved due to implementation of evaluation of the results obtained using statistical quality criteria at each stage of data processing, modeling and forecasting, and also due to the possibility of adapting models to new data through analysis of statistical characteristics of the processes under study and application of combined criteria for the adequacy of models and quality of estimates of forecasts, and the convenient presentation of intermediate and final results.
Conclusions. The proposed methodology is used for forecasting modeling of some macroeconomic and financial processes in Ukraine. The obtained results show that it can be successfully used to solve practical problems of constructing models and prediction of nonlinear nonstationary processes under conditions of uncertainties of various types, which, as a rule, have to be considered during modeling and forecasting on the basis of statistical data.

Keywords: nonlinear nonstationary process; uncertainties; mathematical modeling; forecasting; macroeconomic and financial processes.

Introduction

Most of the modern processes taking place in economy of transition and finances today are nonlinear and nonstationary or at best piecewise linear and stationary. Practically all the process analyzed previously exhibit trends of various order and/or their variance is not constant within the time period studied. The processes trends are stochastic or deterministic dependently on the set of specific internal and external factors influencing the processes under study, and heteroscedasticity is practically inherent to all financial process related to various prices evolution and return forming, exchange rates, etc. [1–3]. Mathematical modeling and forecasting of the processes dynamics based on the use of statistical and/or experimental data usually need to consider various kinds of uncertainties related to the statistical data, structure of the process (and consequently its model) under study, parametric uncertainty, and uncertainties relevant to forecasts estimates. To identify and take into consideration the uncertainties in relevant (already available and being developed) computational algorithms, and improve this way quality of intermediate and the final results (processes evolution forecasts and the decisions based upon them) it is necessary to analyze the reasons for the uncertainties to appear, the consequences of their influence and to

* corresponding author: t.pruman@gmail.com

construct appropriate computational algorithms for solving multiple related specific problems. Development and application of the methodologies possessing the necessary features mentioned is an important task that is being solved nowadays by many researchers [4, 5].

Today there exist methodologies developed for studying nonlinear nonstationary processes (NNP) and constructing mathematical models in various research fields using statistical procedures and state space representation. Another approach to development models for NNP is based upon intellectual data analysis (IDA) techniques such as artificial neural networks, the group method for data handling (GMDH) [5], and Bayesian networks (both static and dynamic) [6, 7]. On the other side these methodologies need some refinement so that to produce better results regarding models adequacy and quality of the forecasts based upon them. The refinement may touch preliminary data processing algorithms, aiming to improvement of statistical characteristics of data, model structure and parameter estimation procedures, as well as the forecast estimates. A very important point regarding modeling methodology development is hiring of appropriate sets of statistical quality criteria necessary to monitor all stages of computations: data quality analysis, model adequacy estimation, and determining the forecasts quality. Finally the quality criteria should analyze alternative decisions (alternatives) based on the forecasts generated. For example, in a case of computing the alternatives using optimization procedures popular quadratic criterion is used very often taking into consideration input control energy (or other equivalent of input control variables) and deviations of the controlled system states from prescribed trajectories [8, 9]. Many other quality criteria are available or could be constructed additionally for supporting specific applications if necessary. The basic requirements to them are easy interpretation and practical implementation, say in the frames of decision support systems that are very popular in the area of process modeling, diagnostic, forecasting and control.

This study is directed towards improvement of model constructing methodology, more specifically it is touching upon data preparing techniques for model constructing, as well as model structure and parameter estimation using multiple computational procedures.

Problem Statement

The purpose of the study is as follows: (1) to perform analysis and development of requirements to the preliminary data processing algorithms (preparing of data for model constructing); (2) development of the software system architecture for model constructing, process evolution forecasting for various dynamic processes in economy and finances including nonlinear and nonstationary ones; (3) identification of some uncertainties relevant to model structure and parameters estimating, and selection of mathematical techniques for minimizing influence of the uncertainties identified; (4) application illustration of the software developed to solving selected problems of modeling and forecasting using actual statistical data.

Requirements to the Modern Applied Software Systems

Modern applied software systems for modeling and forecasting (ASSMF) are rather complex multifunctional (very often possibly distributed) highly developed computing systems with hierarchical architecture that corresponds to the nature of decision making by a human being. To make functionality of the ASSMF maximum useful and convenient for users of different levels (like engineering personnel and managerial staff) they should satisfy some general requirements. In short, the requirements to ASSMF are as follows:

– hierarchical system architecture corresponding to the natural procedures of data and knowledge analysis and decision making by a human;

– availability of model adaptation features making possible models (structure and parameters) adaptation to new data, and possible changes in the modes of functioning of the system under study;

– application of optimization techniques (where possible and necessary) aiming to obtain optimal state and parameter estimates, and optimized forecast estimates as well;

– identification of possible uncertainties and availability of computational techniques directed towards elimination or minimization of negative influence of the uncertainties detected;

– availability of several sets of statistical quality criteria for estimating quality of data, models, forecasts, and decision alternatives, accordingly;

– functional completeness of the system providing to a user all necessary functions necessary for computer-human interaction, and solving specific problem statement and computing results representation in a convenient form;

– adaptation of the system to the needs and preferences of a user, for example, to the ways of representation of intermediate and final results of computations;

– providing high speed of the system functioning including implementation of parallel computing algorithms where possible and necessary.

Appropriate satisfaction of all the requirements to development of ASSMF provides good possibilities for effective practical usage of the system developed and for enhancing its general behavioral effect for specific applications in the area of model constructing with statistical data and forecasting of the relevant processes evolution for a given time horizon [5, 7, 9].

Coping with Possible Uncertainties

Practically all types of mathematical modeling (using functional and structural approaches) usually need to cope with various kinds of uncertainties associated with data, structure of the process under study and subsequently its model, parameters uncertainties, and the forecasts quality.

Here we consider uncertainties as the factors and events that influence negatively the whole process of data collecting, processing and mathematical model building, forecasting of processes evolution, and generating of alternative managerial decisions. They are practically always inherent to the most actual processes under study due to incompleteness or inexactness of our knowledge regarding the objects (systems) under study, incorrect selection or application of computational procedures etc. The uncertainties appear very often due to incompleteness of available data (due to missing observations or short samples), noisy measurements or they are invoked by stochastic external disturbances with unknown probability distribution, poor estimates of model structure or by a wrong selection of a model parameters estimation procedure.

In many cases, a researcher has to cope with the following generalized types of possible uncertainties: structural, statistical, parametric, probabilistic and amplitude uncertainties. The structural uncertainties are encountered in the cases when the structure of the process under study (as well as its model) is only partially known or not clearly defined. For example, in a case when the functional approach to model constructing is applied usually we do not know exactly structure of the system under study. It is determined with appropriate model structure estimation techniques such as correlation analysis, estimation of mutual information, time lags estimation, testing for nonlinearity and nonstationarity using appropriate statistical tests, identification of external disturbances types etc. The sequence of actions necessary for identification, processing and taking into consideration of possible uncertainties is given below.

Identification of data uncertainties: missing and extreme values, measurement errors, influence of external random disturbances, and short samples.

Reduction of influence of the data uncertainties using available or newly developed techniques (missing values imputation, normalizing the data, digital and optimal filtering of noisy data, outliers processing, etc.).

Estimation of model structure and parameters using available (collected) information, correlation analysis methods, statistical tests, and appropriate parameter estimation methods.

Reduction of uncertainties of model structure and its parameters hiring repeated estimation in a loop until the model adequacy is acceptable. Several alternative techniques are available for nonlinear models parameter estimation.

Computing forecasts and reduction of forecasts uncertainty thanks to the use of several alternative methods and combination schemes.

The use of the forecasts for generating alternative decisions according to the problem statement.

All the tasks and procedures mentioned are solved successfully with appropriately designed and implemented ASSMF. As far as usually we have to process stochastic data, application of available statistical techniques provides a possibility for approximate estimation of a system (and its model) structure. First of all, we apply correlation analysis techniques to separate and multiple time series and appropriate statistical tests. To find "the best" model structure it is recommended to apply adaptive estimation schemes that provide automatic search in a definite selected range of model structure parameters (type of distribution, model order, time lags, nonlinearities and nonstationarities). Very often the search is performed, for example, in the class of regression type models with the use of integrated criterion of the following type:

$$V_N(\theta, D_N) = |1-R^2| + |1-0.5DW| + \alpha\frac{SSE(\theta)}{\max SSE} + U, \quad (1)$$

$$V_N(\theta, D_N) \to \min_{\theta}$$

where θ is a vector of model parameters; D_N is statistical data in the form of time series (N is a power of time series used); R^2 is a determination coefficient; DW is Durbin–Watson statistic; α is adjustment coefficient that could be selected by a user or searched for automatically using optimization techniques; U is Theil coefficient characterizing quality of forecasts; $SSE(\theta)$ is actual SSE for a model being constructed; max $SSE(\theta)$ is maximum possible value of SSE for a given application. If $SSE(\theta)$ is greater than the value of max $SSE(\theta)$, then $SSE(\theta)$ = max $SSE(\theta)$. The first three terms in right hand side of criterion (1) characterize model adequacy, and Theil coefficient is used to take into account forecasts quality. The complex criterion (1) is minimized with parameter vectors of candidate models being estimated. The adaptive estimation scheme also helps to minimize model parameter uncertainties, first of all to get unbiased estimates. New coming data is used to compute model parameter estimates with alternative techniques that correspond to possible changes in the system under study. The alternative parameter estimation techniques include OLS, and nonlinear LS, maximum likelihood (ML), and Monte Carlo for Markov chains (MCMC). Availability of the techniques mentioned covers all possible special cases of nonlinear model parameters estimation.

Testing for Linearity and Some Models for Nonlinear Nonstationary Processes

The problem of testing the processes for linearity (nonlinearity) is considered in many studies [10—16]. Usually combination of several tests helps to detect existing nonlinearity and to select appropriate model structure. To detect nonlinearity the following test was applied:

— construct linear regression model for dependent variable $y(k)$ and right hand side (RHS) vector $w(k)$ using LS [11]:

$$y(k) = \beta^T \mathbf{w}(k) + u(k)$$

where $\mathbf{w}(k) = [1, y(k-1),..., y(k-p); x_1(k),..., x_l(k)]^T$ is a vector of measurements for dependent variable and regressors; $\mathbf{v}(k) = [u(k-1),..., u(k-q)]^T$ is a vector of random variables; and $u(k) = g(\beta, \theta, \mathbf{w}(k), \mathbf{v}(k))\,\varepsilon(k)$; $\varepsilon(k)$ is martingale process with the following statistical characteristics: $E[\varepsilon(k)|\mathbf{I}(k)] = 0$, $\mathrm{cov}[\varepsilon(k)|\mathbf{I}(k)] = \sigma_\varepsilon^2$; $\mathbf{I}(k) = \{y(k-j),\ j>0;\ \mathbf{x}(k-i),\ i \ge 0\}$ is available observation information;

— compute residuals of the model $\hat{u}(k)$ and the sum of squared errors SSR_0 for the model constructed;

— construct regression model for $\hat{u}(k)$ with regressors $w(k)$, and compute sum of squared errors for the model SSR_1;

— compute the test statistics:

$$F(m,\ N-n-m) = \frac{SSR_0 - SSR_1/m}{SSR_1/(N-n-m)}$$

where $n = l + p + 1$; m is dimension of parameter vector θ; the value computed has F-distribution with $\theta = 0$. The use of F — statistics instead of χ^2 test is recommended for short samples.

When constructing mathematical models for time series it is convenient to use proposed unified notion of model structure which we define as follows:

$$S = \{r,\ p,\ m,\ n,\ d,\ w,\ l\}$$

where r is model dimension (number of equations that constitute a model); p is model order (maximum order of differential or difference equation); m is a number of independent variables in the right hand side of model equation; n is a nonlinearity and its type (nonlinearity in variables or in parameters); d is output reaction delay time (or lag); w is stochastic external disturbance and its probability distribution; l represents possible restrictions for variables and/or model parameters. All the elements of a model structure are estimated with appropriate statistical tests and correlation analysis procedures, such as correlation matrix, autocorrelation function (ACF), partial ACF (PACF), bi-correlations, and higher order correlation functions.

Consider some types of models nonlinear in variables that are widely used today. Some nonlinear models result from studying econometric time series. For example, nonlinear regression of the following type was used to describe gross domestic product (GDP) and tax income:

$$\begin{aligned} y_1(k) &= a_0 + a_1 y_1(k-1) + b_{12}\exp(y_2(k)) \\ &\quad + a_2 x_1(k) x_2(k) + \varepsilon_1(k), \\ y_2(k) &= c_0 + c_1 y_2(k-1) + b_{21}\exp(y_1(k)) \\ &\quad + c_2 x_1(k) x_2(k) + \varepsilon_2(k) \end{aligned}$$

where $y_1(k)$ is logarithm of GDP; $y_2(k)$ is logarithm of tax income; $x_1(k)$ — internal investments; $x_2(k)$ external investments. Another structure that is used often is generalized bilinear model:

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

where p, q, m and s are positive numbers characterizing model order. The model can also be represented in state space form where the states are presented in the form of a product of former innovations and vectors of random coefficients [12].

Some models of nonlinear processes result directly from economic theory and formally describe specific financial or economic processes. They can take into consideration possibilities for development optimization methods for systems under study using appropriate cost functional or utility function. One of possible model structures is as follows:

$$y(k) = \min(\beta^T \mathbf{w}(k), \theta^T \mathbf{w}(k)) + \varepsilon(k) \qquad (2)$$

where estimate of dependent variable $\hat{y}(k)$ is determined as a smaller one of two possible values computed via alternative functions: $\beta^T \mathbf{w}(k)$ or $\theta^T \mathbf{w}(k)$. If "min" in the model (2) is replaced by another variable, say $z(k\text{-}d)$, that could be an element of vector $w(k)$, though not equal one, then we get so called switching model:

$$y(k) = \beta^T \mathbf{w}(k) + \theta^T \mathbf{w}(k) F(z(k-d)) + \varepsilon(k)$$

where

$$F(z(k-d) = \begin{cases} 0 \text{ if } z(k-d) \le c; \\ 1 \text{ if } z(k-d) > c; \end{cases}$$

c is some threshold value that is used for switching from one model to another; $d = 0, 1, 2, \ldots$ is discrete delay time.

A scalar version of the model is called threshold autoregression with two modes. It can be generalized to the set of possible functioning modes using nonlinear function of the type:

$$F(z(k-d)) = \frac{1}{1 + \exp[-\gamma(z(k-d) - c)]}, \quad \gamma > 0.$$

This is a model of logistic smooth transition regression (LSTR). The function F can also be used in the form of probability density function (PDF). In a scalar case the model will correspond to the exponential smooth transition autoregressive (ESTAR) model.

A convenient approach to modeling nonlinear processes is based on the models that contain linear and nonlinear components, or flexible models:

$$y(k) = \beta^T \mathbf{z}(k) + \sum_{i=1}^{p} \alpha_i \varphi_i(\theta_i^T \mathbf{z}(k)) + \varepsilon(k) \qquad (3)$$

where $z(k)$ is a vector of time delayed values for dependent variable $y(k)$, as well as former and current values of the explaining variables vector $x(k)$ plus shift constant. The first component of the model is linear, and $\varphi_i(x)$ is a set of functions that could include the following components:

– power function $\varphi_i(x) \equiv x^i$, where variable x can be delayed in time value of y or some other variable;

– trigonometric function $\varphi_i(x) = \sin x$ or $\varphi_i(x) = \cos x$;

– equation (3) can be expanded with quadratic function $\mathbf{z}^T(k)\mathbf{A}\mathbf{z}(k)$, that will result in a flexible functional form;

– $\varphi_i(x) = \varphi(x), \forall i$, where $\varphi(x)$ is a link function, for example appropriate PDF or logistic function of the following type:

$$\varphi(x) = \frac{1}{1 + \exp(-x)};$$

– $\varphi(x)$ can also be represented by appropriately selected nonparametric function.

Some general class of nonlinear models is also given by the form:

$$\mathbf{y}(k) = \sum_{j=1}^{p} \varphi_j(\mathbf{x}(k-1))\mathbf{y}(k-j) + \mu(\mathbf{x}(k-1)) + \varepsilon(k) \qquad (4)$$

where $y(k)$ is $[n \times 1]$ stochastic vector of dependent variables; $x(k) = [y(k), y(k-1), \ldots, y(k-n+1)]$ is a vector of state variables dynamics of which is described by the state space model:

$$\mathbf{x}(k) = h(\mathbf{x}(k-1)) + \mathbf{F}(\mathbf{x}(k-1))\mathbf{x}(k-1) + v(k). \qquad (5)$$

Equation (4) can also include the moving average members. It means that in this case to describe selected process we use two models simultaneously what may result in some difficulties with the model structure estimation. Equation (5) is a state space model that can be supplemented with the measurement equation. The elements of matrix $\mathbf{F}(\cdot)$ could be linear functions or low order polynomials. The models (4), (5) can also contain the members that reflect availability of long memory what takes place very often when we study ecological, financial and economic processes.

In the process of constructing forecasting models we build several candidates and select the best one of them with a set of model adequacy statistics. The following techniques are used to fight structural uncertainties: improvement of model order (for example, NAR(p) or NARMAX(p, q)) by applying adaptive approach in a loop to modeling with automatic search for the "best" structure using complex statistical adequacy criteria mentioned above; adaptive estimation of delay time (lag) and the type of data probability distribution with its parameters; describing detected process nonlinearities with alternative analytical forms with subsequent estimation of model adequacy and quality of the forecasts generated.

A wide subclass of nonlinear models is created today by the models describing dynamics of conditional variance for heteroscedastic process (HSP). HSP are nonlinear by definition as far as variance description is based on quadratic variables and functions. Popularity of variance forecasts is explained by their wide possibilities for practical applications such as follows: volatility is a parameter used in stock trading systems for supporting decisions regarding operations with various type of stocks; variance characterizes evolution of prices for many market goods; there is no engineering or medical diagnostic system that does not use variance as a parameter incorporated in decision making rules. The simplest mathematical model of conditional variance is autoregressive conditionally heteroscedastic (ARCH) equation of the form:

$$E_k[\hat{\varepsilon}^2(k+1)] = \alpha_0 + \alpha_1\hat{\varepsilon}^2(k-1) + \alpha_2\hat{\varepsilon}^2(k-2) + \ldots + \alpha_q\hat{\varepsilon}^2(k-q) \qquad (6)$$

where $\hat{\varepsilon}(k)$ is a stochastic part of equation describing HSP under study. It can be estimated by hiring low order autoregressive equation (such as AR(1) or AR(2)) for description of goal variable $y(x)$ in LHS; $E_k(\cdot)$ is a symbol of conditional mathematical expectation computed for specific moment of time k. Usually equation (6) does not allow compute acceptable results of short term variance forecasting.

The structure of equation (6) was improved by introducing another variable into right hand side (RHS) as follows:

$$h_S(k) = \alpha_0 + \sum_{i=1}^{q}\alpha_i\varepsilon^2(k-i) + \sum_{i=1}^{p}\beta_i h_S(k-i) \qquad (7)$$

where sample conditional variance $h_S(k)$ is computed as follows:

$$h_S(k) = \frac{1}{w-1}\sum_{i=k-\frac{w-1}{2}}^{k+\frac{w-1}{2}}[y(i) - \bar{y}_S(i)]^2,\ k = 2,3,\ldots,N,$$

where $\bar{y}_S(k)$ is sample mean computed for each window w; w is a size of moving window for computing conditional variance which is usually selected as an odd number for convenience. Equation (7) is generalized ARCH (GARCH) which is usually much more efficient for describing and short term forecasting conditional variance (and volatility).

Very good results of short term volatility forecasting can be achieved with the exponential GARCH (EGARCH) model that has the following structure:

$$\log[h(k)] = \alpha_0 + \sum_{i=1}^{p}\alpha_i\frac{|\varepsilon(k-i)|}{\sqrt{h(k-i)}} + \sum_{i=1}^{p}\beta_i\frac{\varepsilon(k-i)}{\sqrt{h(k-i)}} + \sum_{i=1}^{q}\gamma_i\log[h(k-i)] + \upsilon(k).$$

This equation contains so called “standard” part that takes into account the innovations $|\varepsilon(k)|$, and another part that takes into consideration sign of the innovation. The values of $\varepsilon(k)$ are normalized by volatility what leads to reduction of possible high values, and the logarithm function is applied for smoothing the volatility. Definitely there exists a wider set of conditional variance models that need separate considerations.

Processing Some Types of Stochastic Uncertainties

Performing modeling of economic, financial and ecological processes usually we don’t have enough information about statistical characteristics (covariance matrices) of measurement errors and stochastic external disturbances (so called state noise). To minimize influence of these uncertainties on data quality digital and optimal filtering techniques are applied. Procedures for digital filters development are well known, and we only have to determine pass-bands and stop-bands. Optimal filters provide for a possibility of simultaneous estimation of system states and the covariance matrices though they require mathematical models in state space form.

Generally optimal Kalman filter can solve the following problems: generating optimal states, estimation of non-measurable state vector components, estimation of statistical parameters for the state and measurement noise, short-term forecasting of state variables. Here we hire nonlinear version of optimal filter that uses approximate (after linearization) state space model of processes and systems under study. System state estimation via optimal filter is performed on the basis of discrete form system model represented by two equations in state space as follows:

$$\mathbf{x}(k+1) = \Phi(k+1,k)\mathbf{x}(k) + \Gamma(k+1)\mathbf{u}(k) + \mathbf{w}(k), \qquad (8)$$

$$\mathbf{z}(k+1) = \mathbf{H}(k+1)\mathbf{x}(k+1) + \mathbf{v}(k+1) \qquad (9)$$

where $x(k)$ is n-dimensional vector of system states; $k = 0, 1, 2, \ldots$ is discrete time; $u(k)$ is m-dimensional vector of control variables; $w(k)$ is n-dimensional vector of external random disturbances; $\Phi(k, k)$ is $[n \times n]$ matrix of system dynamics; $\Gamma(k+1)$ is $[n \times m]$ matrix of control coefficients. The double argument $(k, k-1)$ means that the variable or parameter is computed at the moment k, but its value is based on the previous data processing procedure including moment $(k-1)$. Discrete time k and continuous time t are linked via data sampling time T_s:$t = kT_s$. The classic problem statement supposes that vector sequence of external disturbances $w(k)$ is zero mean white Gaussian noise with covariance matrix Q. In the measurement equation (7) $z(k)$ is a vector of measured output variables; $H(k+1)$ is $[r \times n]$ observation matrix; $v(k)$ is r-dimensional vector of measurement errors (noise) with covariance matrix $\mathbf{R}$. The filter is constructed to compute optimal state vector $\hat{x}(k)$, in conditions of influence of random external system disturbances and measurement noise.

Here uncertainty arises due to unknown covariance matrices $\mathbf{Q}$ and $\mathbf{R}$. The problem is solved with adaptive filter constructing that provides a possibility for computing the estimates of $\hat{\mathbf{Q}}$ and $\hat{\mathbf{R}}$ simultaneously with state vector $\mathbf{x}$. Another possibility is in constructing separate algorithm for computing the values of the covariance matrices mentioned. A convenient statistical data based algorithm for computing the covariance matrices can be found for example in [17]. The algorithm was successfully applied to estimating the values of covariance matrices in many practical applications. The experiments showed that the values of estimates become stationary after about 20–25 of time sampling periods in a scalar case, though this figure may increase substantially with the growth of dimensionality of the model (8), (9).

In a case of hiring nonlinear models for the processes description a discrete nonlinear transformed version of Kalman filter was applied that also uses fundamental matrix $\mathbf{A}(\cdot)$ of continuous time model [18]. Here the sequence of estimates $\hat{\zeta}(k)$, $\hat{\mathbf{x}}(k)$, and state transition matrix are computed as follows:

$$\hat{\mathbf{x}}(k+1) = \zeta(k) + \mathbf{A}(\zeta(k))\,\zeta(k)\,T_s + \sqrt{T_s}\mathbf{B}(k+1)\mathbf{u}(k+1),$$

$$\Phi(k+1,k) = \left[\mathbf{I} - \frac{T_s}{2}\mathbf{A}(\mathbf{x}(k+1))\right]^{-1}\left[\mathbf{I} + \frac{T_s}{2}\mathbf{A}(\zeta(k)\right],$$

$$\hat{\zeta}(k+1) = \Phi(k+1,k)\hat{\zeta}(k) + \sqrt{T_s}\Phi(k+1,k)\mathbf{B}(k+1)\mathbf{u}(k+1)$$

where $\mathbf{A}(\cdot)$ is fundamental matrix for continuous time model; $\zeta(k)$ is approximating estimate for $\hat{\mathbf{x}}(k)$. The prior covariance matrix of state estimates errors and optimal matrix coefficient are computed in the following way:

$$\Gamma(k) = \sqrt{T_s}\Psi(k+1,k)\mathbf{B}(k+1),$$

$$\mathbf{P}(k+1,k) = \Psi(k+1,k)\ \mathbf{P}(k,k)\Psi^T(k+1,k) + \Gamma(k)\mathbf{Q}(k)\Gamma^T k),$$

$$\mathbf{K}(k+1) = \mathbf{P}(k+1,k)\mathbf{H}^T[\mathbf{HP}(k+1,k)\mathbf{H}^T + \mathbf{R}]^{-1},$$

$$\hat{\zeta}(k+1,k+1) = \hat{\zeta}(k+1,k) + \mathbf{K}(k+1)[\mathbf{z}(k+1) - \mathbf{H}\hat{\zeta}(k+1,k)].$$

Among other instruments for processing uncertainties are fuzzy sets and logic, static and dynamic Bayesian networks, appropriate types of conditional probability distributions etc. Bayesian networks represent powerful probabilistic tools for modeling processes and systems with hiring statistical data, expert estimates, information on PDFs for continuous and discrete variables. They can cope successfully with probabilistic and amplitude type uncertainties. Some uncertainties such as missing observations, extreme values and high level jumps of stochastic origin are processed with appropriately selected known statistical procedures. Existing data imputation procedures help to complete the sets of data collected. For example, very often missing measurements of time series could be generated with correctly selected distributions or in the form of short term forecasts. Processing of jumps and extreme values helps with adjusting data non-stationarity and to estimate the probability distribution functions for the stochastic processes under study.

Results and Discussion

The developed procedure for linear and nonlinear model constructing using statistical data includes the steps given below.

1. Preliminary processing of data and expert estimates with application of data quality criteria such as missing values counters, parameters of information content (computing of variance, and number of derivatives for approximating polynomials), power of samples. Filtering of data and imputation of missing values where necessary; estimation of non-measurable components.

2. Application of statistical tests aiming to discovering nonlinearity and nonstationarity; correlation data analysis is giving the grounds for estimation of model structure.

3. Estimation of candidate models structure and their parameters. To reduce the influence of pos-

sible parametric uncertainties three parameter estimation techniques are applied: LS (NLS), maximum likelihood (ML), and Markov chain Monte Carlo (MCMC) procedures.

4. Application of model adequacy statistics and selection of the “best” models. If model quality is not acceptable we return to the step 2 to get more information regarding the model structure and repeat the procedures of structure and parameter estimation.

5. The model(s) selected is used for computing forecasts that are analyzed with another set of quality criteria. Among them are mean absolute percentage error and Theil coefficient.

6. The final step is practical application of the model constructed. If the model is not satisfactory for practical usage the process of model constructing is repeated with extra statistical data.

Very often uncertainties of model parameter estimates such as bias and inconsistency result from low informative data, or data do not correspond to selected type of distribution, what is required for correct application of parameter estimation method. Such situation may also take place in a case of multicollinearity of independent variables and substantial influence of process nonlinearity that for some reason has not been taken into account when model structure was estimated. When power (size) of data sample is not satisfactory for model construction it can be expanded by special techniques, or Monte Carlo simulation is hired, or special model constructing techniques, such as GMDH, are applied. GMDH produces very often results of acceptable quality with short samples. If data does not correspond to normal distribution, then ML technique could be used or appropriate MCMC procedures [10]. The last techniques can be applied with quite acceptable computational expenses when the number of parameters is not large.

Example 1. Consider the problem of modeling return $y(k)$ for a selected stock on the basis of monthly data including 300 observations. According to partial autocorrelation function computed the model of the process should include lags 1—3. Thus the bilinear model selected may look as follows:

$$y(k) = \mu + a_1 y(k-1) + a_2 y(k-2) + a_3 y(k-3) + (1+\beta_1 v(k-1) + \beta_2 v(k-2) + \beta_3 v(k-3))\cdot \varepsilon(k)$$

where $v(k)$ is moving average process; it was suggested that $\{\varepsilon(k)\} \sim N(0, 1)$. The model parameters were estimated with conditional maximum likelihood:

$$y(k) = 0.0117 + 0.173y(k-1) + 0.115y(k-2) - 0.089y(k-3) + 0.077\cdot(1.0 + 0.383v(k-1) + 0.103v(k-2) - 0.551v(k-3))\cdot\varepsilon(k).$$

Adequacy of the model is rather high: $R^2 = 0.89$, $DW = 1.92$, with mean absolute percentage error for one-step-ahead prediction on test sample: $MAPE = 5.2$ %. All parameter estimates are statistically significant at the confidence level of 5 %.

The model for random process $\hat{\varepsilon}(k)$ from the last equation is as follows:

$$\hat{\varepsilon}(k) = \frac{y(k) - 0.0117 - 0.173y(k-1) - 0.115y(k-2) + 0.089y(k-3)}{0.077\cdot(1.0 + 0.383v(k-1) + 0.103v(k-2) - 0.551v(k-3))},$$

where $\hat{\varepsilon}(k) = 0$ for $k \le 3$. The sample autocorrelation function for the process $\hat{\varepsilon}(k)$ shows that it does not contain statistically significant correlations.

Example 2. A set of models were developed for forecasting direction of price evolution for some stocks. As far as price evolution trajectories contain nonlinearities of arbitrary form nonlinear models are needed for description of the processes. One of possible model types is logistic regression. If at the moment of time $t + 1$ the price is higher than at the moment t, this growth is identified as “1”, and decrease of price value is designated as “0”. As input variables (regressors) for logistic regression were selected the following values of *Pivot Point* indicator: $\hat{S}1$, $\hat{S}2$, $\hat{S}3$, $\hat{P}$, $\hat{R}1$, $\hat{R}2$, $\hat{R}3$. The following forecasting model was constructed for maximum price using statistical data:

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

$$x_1(k) = -0.993 + 1.604\cdot\hat{S}1(k) - 0.649\cdot\hat{S}2(k) + 0.363\cdot\hat{S}3(k) + 0.355\cdot\hat{P}(k) - 0.298\cdot\hat{R}1(k) - 0.217\cdot\hat{R}2(k) - 0.278\cdot\hat{R}3(k).$$

Threshold value of probability for logistic regression was selected 0.29, in this case it minimizes the first and second type errors. The error values were 18 % (first type), and 57 % (second type); the number of correctly forecasted directions of price evolution was 69.76 %.

For the minimum price the following model was constructed:

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

$$x_2(k) = -0.139 + 1.21\cdot\hat{S}1(k) - 0.979\cdot\hat{S}2(k) - 0.472\cdot\hat{S}3(k) - 0.22\cdot\hat{P}(k) + 0.423\cdot\hat{R}1(k) + 0.577\cdot\hat{R}2(k) - 0.01\cdot\hat{R}3(k).$$

Threshold value of probability for logistic regression was selected 0.29. The error values were 21 % (first type), and 63 % (second type); the number of correctly forecasted directions of price evolution was 66.13 %.

Using the classification tree (CHAID algorithm) for the maximum prices and threshold value 0.25 we got the first type error 27 %, and the second type error of about 50 %. The number of correctly forecasted directions of the process development was about 68.95 %. For the minimum prices the following results were achieved: threshold value 0.31 we got the first type error 26 %, and the second type error of about 62 %. The number of correctly forecasted directions of the process development was about 64.66 %.

To improve the forecasts quality logistic regression model and decision tree were extended with the price forecasts, computed with regression models for $y(k)$. The model for maximum price looks as follows:

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

$$x_1(k) = -1.267 + 1.071 \cdot \hat{S}1(k) - 0.384 \cdot \hat{S}2(k) + 0.142 \cdot \hat{S}3(k) + 0.206 \cdot \hat{P}(k) + 0.305 \cdot \hat{R}1(k) - 0.181 \cdot \hat{R}2(k) - 0.269 \cdot \hat{R}3(k) + 1.075 \cdot \hat{y}(k)$$

where $\hat{y}(k)$ is regression model output that accepts the value of "1" when the price is growing, and "0" when the price is decreasing. Using the threshold value of 0.46 we got the first type error 28 %, and the second type error of about 36 %. The relative number of correctly forecasted direction of price movement was about 74.19 %. Using the classification tree (CHAID) with the threshold value 0.25, we got the first type error 28 %, and the second type error of about 48 %. The relative number of correctly forecasted direction of price movement was about 69.48 %.

To forecast minimum price the following model was constructed:

$$g_{\min 2}(x_2) = \frac{e^{x_2(k)}}{1+e^{x_2(k)}},$$

$$x_2(k) = -0.124 + 0.364 \cdot \hat{S}1(k) - 0.642 \cdot \hat{S}2(k) - 0.55 \cdot \hat{S}3(k) - 0.549 \cdot \hat{P}(k) - 0.282 \cdot \hat{R}1(k) + 0.599 \cdot \hat{R}2(k) - 0.451 \cdot \hat{R}3(k) + 2.974 \cdot \hat{y}(k).$$

Using the threshold value of 0.47 we got the first type error 46 %, and the second type error of about 17 %. The relative number of correctly forecasted direction of price movement was about 74.77 %. Using the classification tree (CHAID) with the threshold value 0.47, we got the first type error 50 %, and the second type error of about 13 %. The relative number of correctly forecasted direction of price movement was about 74.68 %. The results of maximum and minimum price direction forecasting are given in Tables 1 and 2 respectively.

***Table* 1.** Results of forecasting maximum price direction

Model type	Relative number of correctly forecasted directions
Regression model with indicators	68.95 %
Logistic regression with indicators	69.76 %
Classification tree with indicators	68.95 %
Logistic regression with indicators + forecasts via regression model	**74.19 %**
Classification tree with indicators + forecasts via regression model	69.48 %

***Table* 2.** Results of forecasting maximum price direction

Model type	Relative number of correctly forecasted directions
Regression model with indicators	73.79 %
Logistic regression with indicators	66.13 %
Classification tree with indicators	64.66 %
Logistic regression with indicators + forecasts via regression model	**74.7 %**
Classification tree with indicators + forecasts via regression model	74.6 %

Thus, in both cases the best results were achieved with the logistic regression model that uses forecasts computed by linear regression. The statistical characteristics of forecasts quality used indicate high quality of the forecasts and the possibility for their practical applications in trading rules.

Conclusions

The general methodology was proposed for developing automatized software system for mathematical modeling and forecasting nonlinear nonstationary economic and financial processes using statistical data. The software system development is based on the following system analysis principles: hierarchical system architecture, taking into consideration possible probabilistic and statistical uncertainties, availability of model adaptation features, generating of multiple decision alternatives (multiple models and forecasts), and tracking of computational processes at all the stages of data processing and model constructing with appropriate sets of statistical quality criteria. As instrumentation for fighting possible uncertainties the following techniques were used: optimal Kalman filter, missing data imputation techniques, multiple methods for model parameter estimation, and probabilistic Bayesian programming approach. A short review of models for nonlinear nonstationary process was presented which shows rapidly growing popularity of the model structures considered.

The examples of the software system application show that it can be used successfully for solving practical problems of mathematical model building and forecasts estimation. The system proposed could be used for support of decision making in multiple areas of human activities including strategy development for the state government, various financial institutions, industrial and agricultural enterprises, investment companies etc. Further extension of the system functions is planned with new forecasting techniques based on probabilistic models, neural networks, and fuzzy sets and rules.

References

[1] O.M. Trofymchuk *et al.*, “Probabilistic and statistical uncertainty decision support systems”, *Visnyk Natsional’noho Universytetu “L’vivs’ka politekhnika”. Ser. Komp’yuterni Nauky ta Informatsiyni Tekhnolohiyi*, no. 826, pp. 237–248, 2015 (in Ukrainian).

[2] F.X. Diebold. *Elements of Forecasting.* Ohio: Thomson South-Western, 2007.

[3] *Econometrics Theory and Practice: Frontiers of Analysis and Applied Research*, D. Corbae *et al.*, eds. Cambridge, UK: Cambridge University Press, 2010.

[4] R.S. Tsay, *Analysis of Financial Time Series.* Hoboken: John Wiley & Sons, Inc., 2010.

[5] S.O. Dovhii *et al.*, *Decision Support Systems Based on Statistical Probabilistic Methods.* Kyiv, Ukraine: Logos, 2014 (in Ukrainian).

[6] J. Pearl. *Causality: Models, Reasoning, and Inference.* Cambridge, UK: Cambridge University Press, 2009.

[7] *Bayesian Networks: A Practical Guide to Applications*, O. Pourret *et al.*, eds., Chichester, UK: John Wiley & Sons Ltd, 2008.

[8] *Optimization, Simulation and Control*, A. Chinchuluun *et al.*, eds. New York: Springer-Verlag New York, 2013. doi: 10.1007/978-1-4614-5131-0

[9] R. Baldick, *Applied Optimization: Formulation and Algorithms for Engineering Systems.* Cambridge, UK: Cambridge University Press, 2006. doi: 10.1017/CBO9780511610868

[10] J.B. Ramsey, “Tests for specification errors in classical linear least squares regression analysis”, *J. Royal Statistical Soc. B*, vol. 31, pp. 350–371, 1969.

[11] T. Teräsvirta *et al.*, “Aspects of modeling nonlinear time series”, *Handbook of Econometrics*, vol. 4, pp. 2919–2957, 1994. doi: 10.1016/S1573-4412(05)80017-0

[12] R.S. Tsay, “Nonlinearity tests for time series”, *Biometrika,* vol. 73, pp. 461–466, 1986. doi: 10.1093/biomet/73.2.461

[13] D. Tjøstheim, “Some doubly stochastic time series models”, *J. Time Series Analysis*, vol. 1, no. 7, pp. 51–72, 1986. doi: 10.1111/j.1467-9892.1986.tb00485.x

[14] D. Tjøstheim, “Nonlinear time series: a selective review”, *Scandinavian J. Statistics*, vol. 2, no. 21, pp. 97–130, 1994.

[15] B.K. Stensholt and D.Tjøstheim, “Multiple bilinear time series models”, *J. Time Series Analysis*, vol. 2, no. 8, pp. 221–233, 1987. doi: 10.1111/j.1467-9892.1987.tb00434.x

[16] R.S. Tsay, “Testing and modeling threshold autoregressive processes”, *J. Am. Statistical Association*, vol. 84, no. 405, pp. 231–240. 1989. doi: 10.1080/01621459.1989.10478760

[17] Jr.W.N. Anderson *et al.*, “Consistent estimates of the parameters of a linear system”, *Annals Math. Stat.*, vol. 40, no. 6, pp. 2064–2075, 1969. doi: 10.1016/B978-0-12-012711-5.50007-9

[18] Y. Yavin, “A discrete Kalman filter for a class of non-linear stochastic systems”, *Int. J. System Sci.*, vol. 13, no. 11, pp. 1233–1246, 1982. doi: 10.1080/00207728208926425

П.І. Бідюк, С. Овермайєр, Т.І. Просянкіна-Жарова, О.М. Терентьєв

МЕТОДИКА МОДЕЛЮВАННЯ І ПРОГНОЗУВАННЯ НЕСТАЦІОНАРНИХ ПРОЦЕСІВ У ФІНАНСАХ

Проблематика. Більшість моделей фінансових та економічних процесів вирізняються значною обчислювальною складністю, а побудова прогнозів прийнятної якості для цих процесів на потрібний часовий горизонт – великою трудомісткістю. Тому розробка і впровадження ефективних інструментів прогнозного моделювання фінансових та економічних процесів є однією з актуальних і практично значимих задач. У роботі розглядається питання моделювання та прогнозування нелінійних нестаціонарних процесів у макроекономіці та фінансах із використанням методики, в основу якої покладені принципи системного аналізу, такі як ієрархічне моделювання, врахування невизначеностей, оптимізація характеристик за допомогою комплексних критеріїв, а також структурно-параметрична адаптація. Застосування запропонованої методики дає змогу покращити якість прогнозування за рахунок виявлення особливостей досліджуваного процесу й адаптації моделі до нових даних тощо.

Мета дослідження. Розроблення методики для прогнозного моделювання нестаціонарних процесів у фінансах і макроекономіці із використанням статистичних даних та її реалізація у відповідній комп'ютерній системі.

Методика реалізації. Методика базується на технологіях попередньої обробки статистичних даних, спрямованих на усунення можливих невизначеностей, застосуванні кореляційного аналізу для оцінювання структури моделі та виборі методів оцінювання параметрів моделей, обчисленні оцінок прогнозів і генеруванні альтернативних рішень. Це дає можливість об'єктивно оцінювати результати, одержані на кожному етапі розв'язку задачі моделювання нелінійних нестаціонарних процесів у макроекономіці та фінансах. У дослідженні пропонується оригінальна методика визначення структури моделі з подальшим її впровадженням в інформаційну систему підтримки прийняття рішень.

Результати дослідження. Для вибраних фінансових і макроекономічних процесів було побудовано відповідні моделі. Висока якість кінцевого результату аналізу даних і прогнозування досягнута завдяки реалізації оцінювання одержаних результатів із використанням статистичних критеріїв якості на кожному етапі обробки даних, побудови моделей і прогнозування, а також завдяки можливості адаптації моделей до нових даних через повторний аналіз статистичних характеристик досліджуваних процесів і застосування комбінованих критеріїв адекватності моделей та якості оцінок прогнозів і зручному представленню проміжних та кінцевих результатів.

Висновки. Запропонована методика використана для прогнозного моделювання окремих макроекономічних та фінансових процесів України. Отримані результати свідчать про те, що її можна успішно використовувати для розв'язання практичних задач побудови моделей і прогнозів нелінійних нестаціонарних процесів в умовах невизначеностей різних типів, які зазвичай доводиться розглядати під час моделювання та прогнозування, використовуючи статистичні дані.

Ключові слова: нелінійний нестаціонарний процес; невизначеності; математичне моделювання; прогнозування; макроекономічні та фінансові процеси.

П.И. Бидюк, С. Овермайер, Т.И. Просянкина-Жарова, А.Н. Терентьев

МЕТОДИКА МОДЕЛИРОВАНИЯ И ПРОГНОЗИРОВАНИЯ НЕСТАЦИОНАРНЫХ ПРОЦЕССОВ В ФИНАНСАХ

Проблематика. Большая часть моделей финансовых и экономических процессов отличается значительной вычислительной сложностью, а построение прогнозов приемлемого качества на необходимый временной горизонт – значительной трудоемкостью. Поэтому разработка и внедрение эффективных инструментов прогнозного моделирования финансовых и экономических процессов являются одной из актуальных и практически значимых задач. В работе рассматриваются вопросы моделирования и прогнозирования нелинейных нестационарных процессов в макроэкономике и финансах с использованием методики, в основу которой положены принципы системного анализа, такие как иерархическое моделирование, учет влияния неопределенностей, оптимизация характеристик моделей с помощью комплексных критериев, а также структурно-параметрическая адаптация. Применение предложенной методики позволит улучшить качество прогнозирования за счет исследования особенностей анализируемого процесса и адаптации моделей к новым данным и т.п.

Цель исследования. Разработка методики прогнозного моделирования нестационарных процессов в финансах и макроэкономике с использованием статистических данных, а также ее реализация в соответствующей компьютерной системе.

Методика реализации. Методика основывается на технологиях предварительной обработки статистических данных, предназначенных для устранения возможных неопределенностей, применении корреляционного анализа для оценивания структуры модели и выбора методов оценивания ее параметров, вычислении оценок прогнозов и генерировании альтернативных решений. Это позволяет объективно оценивать результаты, полученные на каждом этапе решения задачи моделирования нелинейных нестационарных процессов в макроэкономике и финансах. В работе предлагается оригинальная методика определения структуры модели и внедрение ее в информационную систему поддержки принятия решений.

Результаты исследования. Для выбранных финансовых и макроэкономических процессов были построены соответствующие модели. Высокое качество конечного результата анализа данных и прогнозирования достигнуто благодаря реализации оценивания полученных результатов с использованием статистических критериев качества на каждом этапе обработки данных, построения моделей и прогнозирования, а также благодаря возможности адаптации моделей к новым данным через повторный анализ статистических характеристик исследуемых процессов и применение комбинированных критериев адекватности моделей и качества оценок прогнозов и удобному представлению промежуточных и конечных результатов.

Выводы. Предложенная методика использована для прогнозного моделирования некоторых макроэкономических и финансовых процессов Украины. Полученные результаты свидетельствуют о том, что ее можно успешно использовать для решения практических задач построения моделей и прогнозов нелинейных нестационарных процессов в условиях неопределенностей разных типов, которые, как правило, приходится рассматривать во время моделирования и прогнозирования на основе статистических данных.

Ключевые слова: нелинейный нестационарный процесс; неопределенности; математическое моделирование; прогнозирование; макроэкономичесие и финансовые процессы.

Рекомендована Радою
Інституту прикладного системного аналізу
КПІ ім. Ігоря Сікорського

Надійшла до редакції
21 грудня 2017 року

Прийнята до публікації
08 лютого 2018 року