The 6th International Conference on Control and Optimization with Industrial Applications, 11-13 July, 2018, Baku, Azerbaijan

METHODOLOGY OF MODELING OF NONLINEAR NONSTATIONARY PROCESSES

P.I. BIDYUK[1], S.H. ABDULLAYEV[2], A.S. GASANOV[3], T.I. PROSYANKINA-ZHAROVA[4]

[1,3,4]National Technical University of Ukraine, I. Sikorsky KPI, Kiev
[2]Institute of Information Technology of ANAS, Baku, Azerbaijan
e-mail: pbidyuke_00@ukr.net, depart5@iit.ab.az, ayding44@rambler.ru, t.pruman@gmail.com

1. Introduction

The process we meet in various fields of human activities today are mostly nonlinear and nonstationary or piecewise linear and stationary. Those are the processes of environmental pollution, weather forecasting, stock prices evolution etc. Such processes are characterized by stochastic or deterministic trends dependently on a set of specific random factors influencing them. Many processes under study are heteroscedastic by nature, i.e. their variance is changing in time [1]. Nonstationary processes usually exhibit nonlinearities of various kind. Say, trend can be described by quadratic or cubic function or even higher order what can be observed when studying, for example, financial processes. Heteroscedastic processes models include an equation describing the process itself (its amplitudes), and an equation describing dynamics of its variance. The nonlinear model of variance is very useful in many applications. Stock trading rules usually include variance as a parameter because there are exists so called leverage effect that says: stock return is inversely proportional to the variance of its price. Also variance is a very useful parameter for diagnostic systems in economy, technical and medical applications [2–4]. One of the important points in modeling is identification and taking into consideration of possible uncertainties: factors of negative influence for the modeling process that result in various errors decreasing the quality of final results. This study is directed towards refinement of nonlinear model constructing methodology. It is touching on quality of data improvement before model constructing, as well as model structure and parameter estimation using multiple statistical analysis procedures.

2. Modeling methodology proposed

The proposed methodology of modeling nonlinear nonstationary (NNP) processes includes the following steps:

- identification and elimination of data uncertainties using appropriate preliminary data processing procedures;
- model structure estimation using statistical and probabilistic data analysis techniques providing a possibility for determining the following elements of a model structure: model dimension (number of equations) model order, nonlinearity and its type (nonlinearity with respect to variables and parameters), delay time estimate, and type of probabilistic distribution for the variables under study; it is desirable to estimate structures for several candidate models;
- model parameters estimation using alternative methods such as nonlinear least squares (NLS), maximum likelihood (ML), Markov chain Monte Carlo (MCMC) and others providing unbiased estimates of parameters under specific conditions;

- computing statistical parameters characterizing model adequacy and their application to constructed candidate models aiming to determine the most suitable one among them;
- computing model based forecasts for the process (under study) evolution and their quality estimation for the final selection of the best forecasting model.

When speaking about data uncertainties we mean missing measurements, short non-informative samples, noise corrupted measurements etc. Such uncertainties are relatively easy identifiable and corrected with appropriate data processing techniques for the lost measurements imputation, bootstrap analysis and filtering techniques. Model structure uncertainties are usually provoked by poor structure of statistical/experimental data that does not contain enough information for the model structure parameter identification. The parametric uncertainties are also closely related to the quality of available statistical data and their influence is usually related to biased parameter estimates. To avoid the bias it is advisable to apply alternative parameter estimation techniques such as ML and MCMC. Improvement of forecasts estimates could be reached with hiring alternative forecasting methods and forecasts combining techniques [4,5]. To monitor the methodology implementation it is necessary to apply at least three sets of statistical quality parameters: data quality statistics, model adequacy statistics and forecasts quality statistics. Also the quality criteria should analyze alternative decisions based on the forecasts generated. Development and practical application of the methodology is an important task aiming constructing adequate models suitable for solving the problems of forecasting, automatic control, business decision making, complex system simulation etc.

3. Processes models

Today there exist known methodologies for studying NNP and constructing appropriate mathematical models in various research fields using statistical procedures for regression analysis and state space (SS) representation. Another popular approach to development NNP models is based on intellectual data analysis (IDA) techniques: artificial neural networks, fuzzy sets, neuro-fuzzy models, Bayesian networks (both static and dynamic), decision trees etc. When building time series (TS) models it is convenient to use proposed unified notion of a model structure [5]:

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

where r is model dimension; p is model order (maximum order of differential or difference equation); m is a number of independent variables; n is a nonlinearity and its type; d is delay time; w is external stochastic disturbance and its type; l represents restrictions for variables and parameters. All the elements of a model structure are estimated with appropriate statistical tests and correlation analysis procedures: correlation matrix, autocorrelation function (ACF), partial ACF, bi-correlations, and higher order correlation functions. Some nonlinear models came from studying econometric TS. Nonlinear regression of the following type is used to describe gross national product (GNP) and tax income [6]:

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

where $y_1(k)$ is logarithm of GNP; $y_2(k)$ is logarithm of tax income; $x_1(k)$ is internal investment; $x_2(k)$ is external investment. Another useful structure is generalized bilinear model:

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

where p, q, m and s are positive numbers that represent model order [6,7]. Modeling nonlinear processes can be based on combination of linear and nonlinear components:

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

where $\mathbf{z}(k)$ is a vector of time delayed values for dependent variable $y(k)$, as well as former and current values of explaining variables $\mathbf{x}(k)$ plus shift. Here $\varphi_i(x)$ is a set of functions that can include the following components: power function $\varphi_i(x) \equiv x^i$; trigonometric functions $\varphi_i(x) = \sin x$ or $\varphi_i(x) = \cos x$; the equation can be expanded with quadratic form $\mathbf{z}^T(k)\mathbf{A}\mathbf{z}(k)$; $\varphi_i(x) = \varphi(x), \forall i$, where $\varphi(x)$ is a link function, for example appropriate probability density function or logistic function of the type:

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

A general class of nonlinear models is given in the form:

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

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

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

When constructing forecasting models we build several candidates and select the best one with a set of model adequacy statistics. Here the following techniques are used to fight structural uncertainties: refinement of model order applying adaptive approach in a loop to modeling and automatic search for the best structure using complex statistical criteria; adaptive estimation of delay time (lag), and the type of data distribution with its parameters; describing detected nonlinearities with alternative analytical forms with subsequent estimation of the forecasts generated. Also a wide subclass of nonlinear processes is created today by the models describing dynamics of conditional variance for heteroscedastic process (HP). HP are nonlinear by definition as far as variance description is based on quadratic variables and functions. Thus, the general methodology was proposed for mathematical modeling and forecasting nonlinear nonstationary processes using statistical data. The model development procedure is based on the following principles: taking into consideration possible probabilistic and statistical uncertainties, availability of model adaptation features, generating of decision alternatives, and tracking of computational procedures at all stages of data processing with appropriate sets of quality statistics. 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 Bayesian programming approach.

Keywords: Modeling, nonlinear nonstationary processes, uncertainties, finances.

AMS Subject Classification: 6062Q99.

References

[1] Trofymchuk O., et al., Probabilistic and statistical uncertainty processing using decision support systems, *Visnyk of Lviv Polytechnic National University*, 2015, No.826, pp.237–248.

[2] Diebold F.X., *Forecasting in Economics, Business, Finance and Beyond*, University of Pennsylvania, 2015, 607 p.

[3] Hansen B.E., *Econometrics*, University of Wisconsin, 2017, 427 p.

[4] Tsay R.S., *Analysis of Financial Time Series*, New York, John Wiley & Sons, Inc., 2010, 715 p.

[5] Dovgij S.O., Trofymchuk O.M., Bidyuk P.I., *DSS Based on Statistical and Probabilistic Procedures*, Kyiv, Logos, 2014, 420 p.

[6] Terasvirta T., Tjostheim D., Granger C.W.J., *Aspects of Modeling Nonlinear Time Series*, in Handbook of Econometrics, Vol.4, Eds.: R.F. Engle, D.L. McFadden, New York: Elsevier Science B.V., 1994, pp.2919-2957.

[7] Tjostheim D., Nonlinear time series: a selective review, *Scandinavian Journal of Statistics*, 1994, Vol.21, No.2, pp.97-130.