

## ADAPTIVE FORECASTING AND FINANCIAL RISK ESTIMATION

V. DANILOV, O. GOZHYJ, I. KALININA, A. BELAS, P. BIDYUK, O. JIROV

**Abstract.** The study is directed towards development of an adaptive decision support system for modeling and forecasting nonlinear nonstationary processes in economy, finances and other areas of human activities. The structure and parameter adaptation procedures for the regression and probabilistic models are proposed as well as the respective information system architecture and functional layout are developed. The system development is based on the system analysis principles such as adaptive model structure estimation, optimization of model parameter estimation procedures, identification and taking into consideration of possible uncertainties met in the process of data processing and mathematical model development. The uncertainties are inherent to data collecting, model constructing and forecasting procedures and play a role of negative influence factors to the information system computational procedures. Reduction of their influence is favourable for enhancing the quality of intermediate and final results of computations. The illustrative examples of practical application of the system developed proving the system functionality are provided.

**Keywords:** economic and financial processes, adaptive modeling, forecasting nonlinear nonstationary processes, uncertainties, system analysis, decision support system.

### INTRODUCTION

Modeling and forecasting financial, economic, ecological, climatology and many processes in other spheres of human activity is important problem that is to be solved by many companies and institutions in business, at the state and industrial enterprises, scientific and educational laboratories etc. The most distinctive common features of such processes today are non-stationarity and nonlinearity that require a lot of special attention for estimating respective model structure and its parameters. To improve the forecasts based upon mathematical models it is necessary to develop new appropriate model structures that would adequately describe the processes under study and provide a possibility for computing high quality forecasts. One of the most promising modern approaches to modeling and forecasting is based upon so called systemic approach that supposes application of system analysis principles in the frames of specialized decision support system (DSS) [1–3]. Usually the set of the principles includes the following ones: – constructing DSS according to the hierarchical strategy of decision making; – application of optimization and adaptation techniques for model building, forecasting and control; – identification of possible uncertainties (the factors of negative influence to the computational procedures used in DSS that are of various kind and origin) and application of algorithmic means helping to reduce their influence on the quality of intermediate and final results of data processing and decision making [4]. Some other systemic principles could be hired for constructing DSS, though perhaps not so important as those mentioned above. The most important for prac-

tical use are the principles of adaptation, optimization and minimization of uncertainty influence that are helpful for enhancing adequacy of the models being constructed and improving the quality of intermediate and final results.

There are many positive examples of adaptation and optimization techniques application in modeling, forecasting and control [5–7]. This is especially urgent task for analyzing non-stationary processes met practically in all the areas mentioned above. There are two basic directions of adaptation while solving the modeling problems: adaptation of model structure and parameters. According to our definition the notion of model structure includes the following elements: model dimension that is determined by the number of its equations; model order that is determined by the highest order of a model equation; output reaction delay time (or lag) for independent variables (regressors); system or process nonlinearity and its type (nonlinearity with respect to variables or to parameters); type of external stochastic disturbance (distribution and its parameters); system (process) initial conditions and possible restrictions on variables and/or model parameters [8]. Thus, we have many possibilities for model structure corrections and its adaptation to varying system operation modes and conditions of application.

The books [6–8] consider various possibilities for mathematical models adaptation and their further applications to short-term forecasting dynamics of specific processes under consideration. The set of possible model structures proposed is very wide, starting from linear regression equations and up to sophisticated probabilistic models in the form of Bayesian networks, various nonlinear structures and combined models. There also can be found some adaptation procedures illustrating possible changes of a model structure and re-computing of their parameters. It is stressed that application of adaptation schemes helps to increase model adequacy in changing conditions of random external influences, nonlinearity and non-stationarity of the process under study.

The study [9] describes procedures for constructing adaptive regression models on the basis of large datasets. The authors proposed development of decision rules in application to machine learning. They stress that model trees and regression rules are most expressive approaches for data mining procedures of model development. The adaptive model rules proposed in the study create a one-pass algorithm that can adapt available set of rules to the possible changes in the processes under consideration. The sets of rules generated can be ordered or unordered, and it was shown experimentally that unordered rules exhibited higher performance in the terms of statistical quality parameters of the models generated.

The results presented in [10–11] consider the problem of adaptive models constructing for nonstationary heteroscedastic processes widely known today in analysis of financial time series. The authors proposed a procedure for automated constructing and model selection in finance. The flexible procedure is general-to-specific modeling of the mean, variance and probabilistic distribution. The initial specification of a model starts from autoregressive terms and regressors (explanatory variables). The variance specification is based upon log-ARCH and log-GARCH terms, the term of asymmetry, Bernoulli jumps and other possible explanatory variables. The algorithm developed returns specifications of parsimonious mean and variance as well as standardized error distribution in cases when normality is rejected. The extensive Monte Carlo simulations were performed and three empirical applications studied that support usefulness of the method proposed for practical analysis of financial data.

The use of adaptive exponential smoothing for lumpy demand forecasting is considered in [12]. It showed substantial advantages over some conventional approaches used in practice due to appropriate selecting the model smoothing factor. Kalman filter is used to perform preliminary measurement data processing, and then forecasting models are constructed using adaptive smoothing factor based upon optimal filter weighting function. As a result the model performance with this weighting function managed to generate smaller forecasting errors than their counterparts used in demand prediction.

Adaptive forecasting of dynamic processes in conditions when recent and ongoing structural changes are present is considered in [13], and the nature of the changes is unknown. The authors used the method of down-weighting older data based on the tuning parameter found as a result of minimizing mean square error of time series forecasts. A detailed theoretical analysis of the forecasting method is presented as well as positive results of multiple computational experiments based upon macroeconomic data from US economy.

The problem of short-term forecasting in conditions of availability of structural breaks is considered in [14]. The optimal one step ahead forecasts are generated using known exponential smoothing techniques. Analytical expressions are derived for optimal weights in models with one and multiple regressors. The authors showed that the weight remains the same within a given operating regime of a system under study. The comparative study of the method proposed was performed using Monte Carlo simulations and the data from industrial economies. It was shown that robust optimal weights provide high quality short-term forecasts when information on structural breaks is uncertain.

A short review of adaptive approaches to modeling and forecasting processes in various areas of human activities presented above indicates that appropriate adaptation of the models constructed usually helps to construct adequate models and to enhance forecast quality. The study proposed is directed towards development of adaptive forecasting system providing a possibility for forecasting nonlinear nonstationary processes (NNPs) met in economy, finances, ecology etc.

## **PROBLEM STATEMENT**

The purpose of the study is in solving the following problems: to develop structure and parameter adaptation procedures for the regression and probabilistic models; to develop the system architecture for modeling and forecasting nonlinear nonstationary processes in economy, finances, ecology and other areas based on the system analysis principles; to consider possibilities for elimination of some uncertainties inherent in data collecting, model constructing and forecasting procedures; to develop the methodology for modeling and forecasting linear and nonlinear processes in the frames of the same system; providing illustrative examples of practical application of the system developed proving the system functionality.

## **SOME COMMON FEATURES OF THE PROCESSES IN ECONOMY, FINANCES AND ECOLOGY**

A wide diversity of various processes exists in economy, finances, ecology, demography and other areas of human activity. However, there are some general common features of the process like linearity/nonlinearity, and stationar-

ity/nonstationarity that allow to divide them into practically understandable classes and select appropriate modeling and forecasting techniques. Fig. 1 shows simplified classification of the processes from which we could make a conclusion about wide variety of mathematical model structures that could be applied for formal description of the processes dynamics and solving the problem of forecasting their evolution in time.

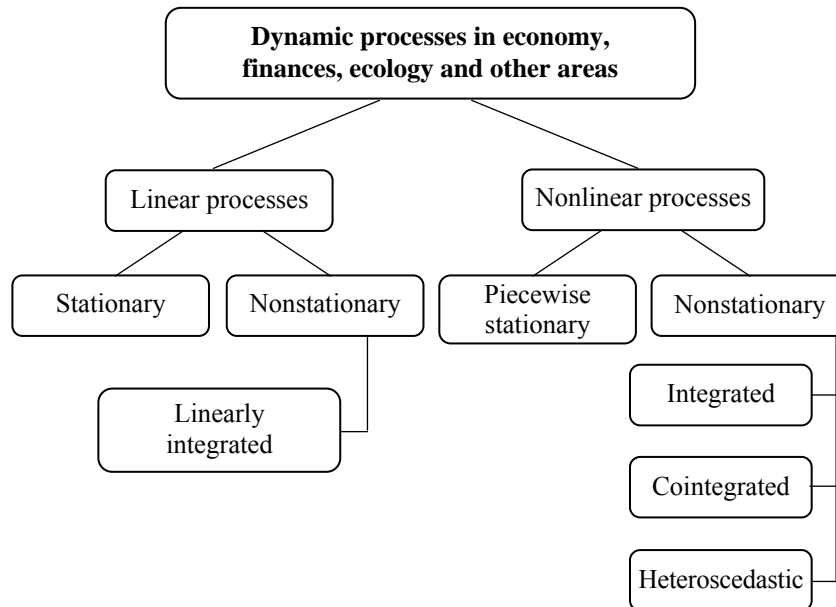


Fig. 1. A simplified classification of dynamic processes in economy and finances

Linear processes can be stationary without trend and nonstationary when they contain linear (first order) trend,  $I(1)$ , where  $I(1)$  means integrated of the first order. If variance (covariance) of stochastic linear process is time dependent then it is classified as heteroscedastic and requires nonlinear models for describing the process variance and possibly the process itself.

There also exists a wide variety of nonlinear processes though we selected only some of them that are more frequent in economy and finances. Generally the processes can be split into nonlinear regarding parameters and nonlinear regarding variables. The first type is more sophisticated with respect to modeling and parameter estimation and usually requires more efforts and time for their model development; it is not considered here. As an example of such a model could be mentioned widely used in practice logistic regression.

Some nonlinear processes can exhibit linear behavior in their stable (nominal) mode of operation. This feature allows for linear description of the process in the vicinity of operating point. Generally NNPs are very often met in the areas of study mentioned above. The set of the processes includes integrated processes (IP) that contain a trend of order two or higher as well as cointegrated processes with the trends of the same order, and the processes with time changing variance, i. e. heteroscedastic processes. Most of financial processes illustrating price evolution of stock instruments belong to this class [15, 16]. In engineering applications such processes are studied in diagnostic systems where appropriate decision is made regarding current system state.

**METHODOLOGY OF MODELING NONLINEAR NONSTATIONARY PROCESSES**

The methodology proposed for modeling NNPs illustrates Fig. 2. At the first step of the methodology the data collected is subjected to preliminary processing that may include the following basic operations: imputation of missing observations,

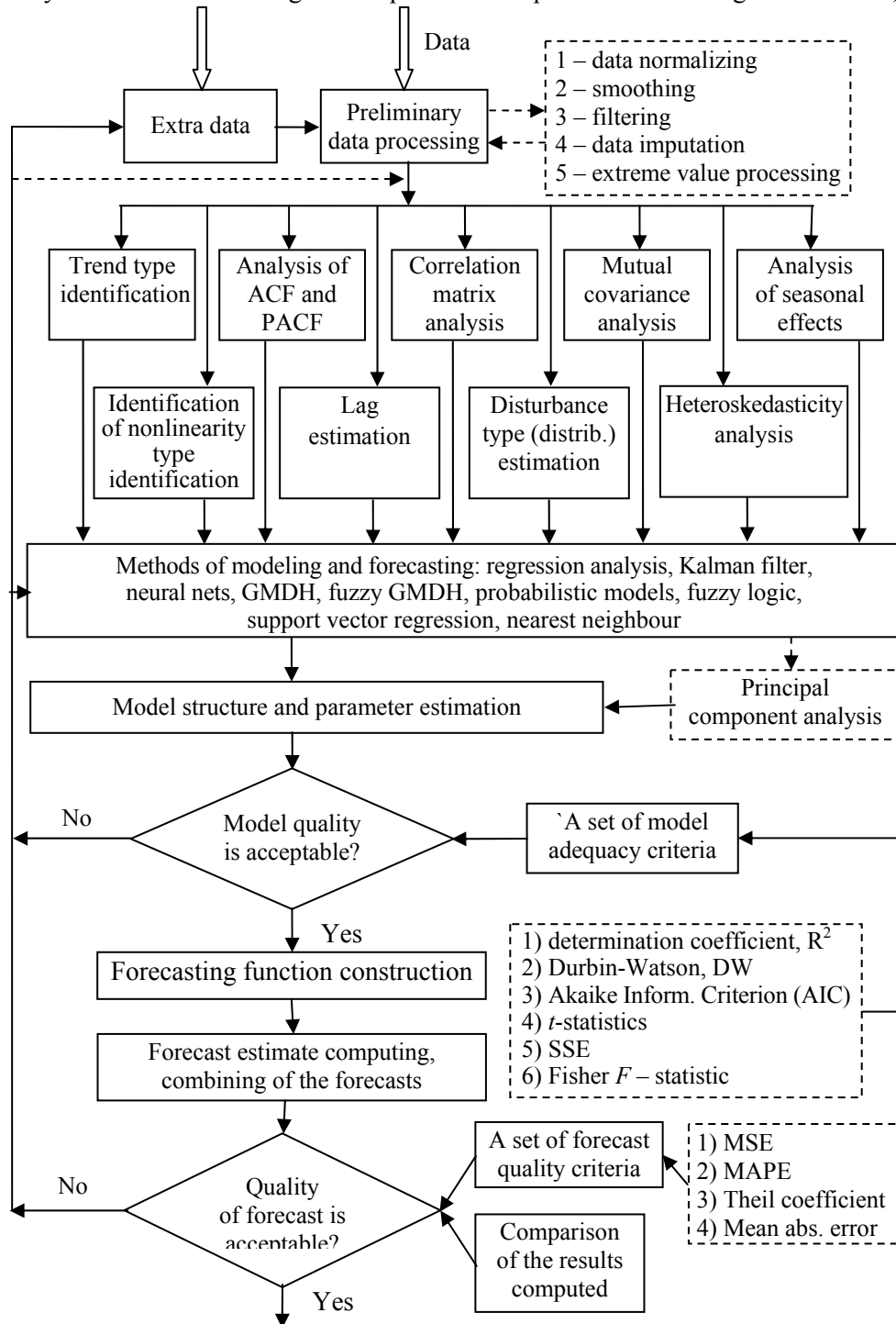


Fig. 2. Functional layout of the forecasting system proposed

normalization in a given range, digital or optimal filtering dependently on problem statement, principal component analysis, appropriate processing of outliers etc. Here it is also appropriate to perform identification and elimination (reduction) of data uncertainties that may touch the following: non-measurable value estimation; computing the general statistical parameters (variance, covariance, mean, median etc.); performing data structuring according to the problem statement; analysis of distribution types and their parameters; estimation of prior probabilities where necessary [17, 18].

Estimation of a model structure using statistical and probabilistic (mutual) information analysis that provides a possibility for estimation of the following elements of a model structure: dimension of a model – number of equations creating the model; model order (highest order of difference or differential equation of the model); nonlinearity and its type; estimate of input delay time, and type of probabilistic distribution for the model variables. It is always appropriate to perform structure estimation for several candidate-models so that to have a possibility for selecting the best one of the candidates estimated.

Formally, to detect nonlinearity in statistical data available statistical tests and techniques should be applied. Fig. 3 shows some known techniques for testing the data for nonlinearity.

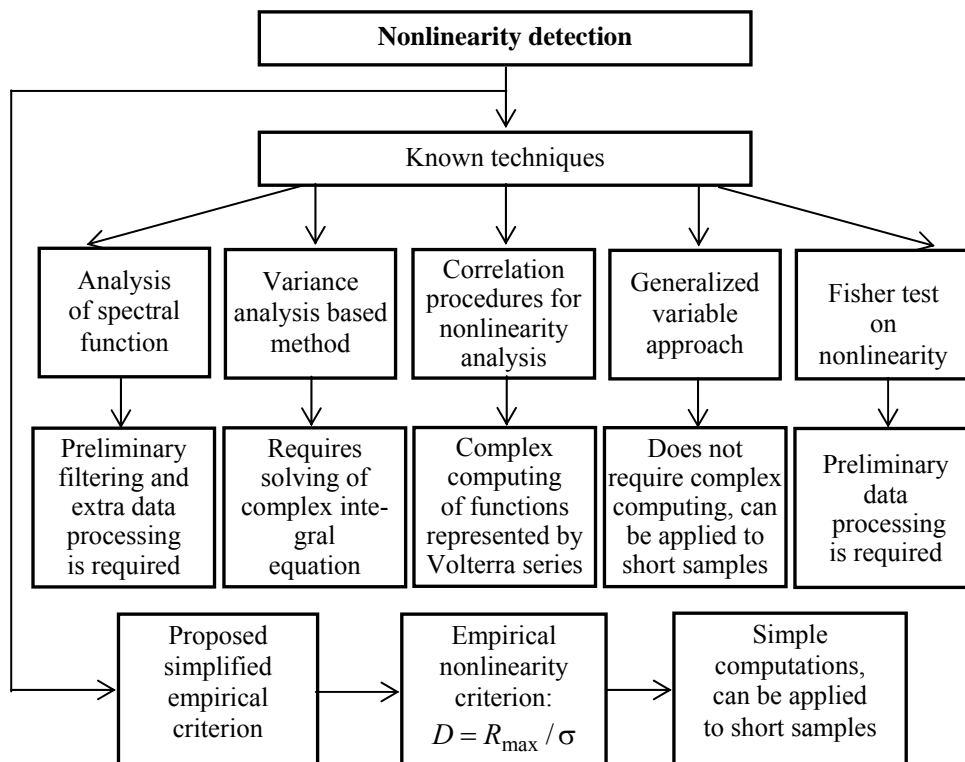


Fig. 3. Some techniques for testing data for nonlinearity

Along with application known technics we proposed a new simplified empirical criterion for detecting nonlinearity in data that is shown below in the Fig. 3: here  $R$  is maximum deviation of the process under study from its linear approximation;  $\sigma$  is sample standard deviation of the process. It does not require

sophisticated computations though provides for additional information about availability of nonlinearity.

The sequence of operations allowing for constructing nonlinear model illustrates Fig. 4; actually this is a part of general model constructing procedure given in Fig. 2.

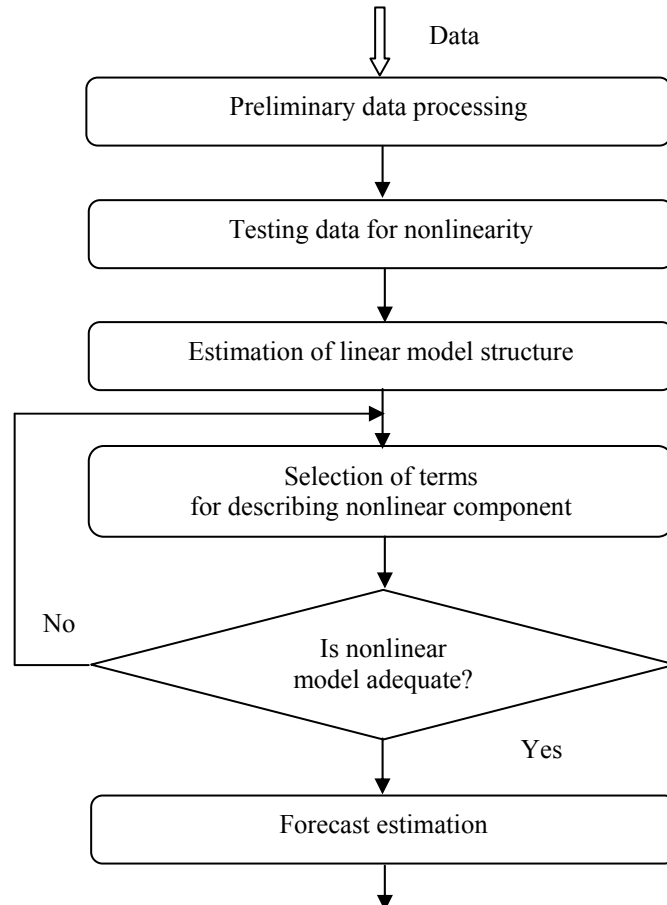


Fig. 4. Procedure for formal describing nonlinear process

Possible nonlinearities (with respect to model variables) could be taken in the following way: first linear part is estimated using known linear structures like autoregressive equations with moving average with linear trend, paired or multiple regression etc. Then nonlinear part of the model is added in the form of nonlinear trend, quadratic, bilinear or higher order terms, nonlinear terms describing cyclic changes of the main variable etc. The modeling practice shows that acceptable model adequacy can be often reached using combination of linear and nonlinear regression, linear regression and Bayesian networks, linear regression and special nonlinear functions like nonparametric kernels. Using this approach a set of candidate models could be constructed with subsequent selecting the best one using appropriate set of statistical adequacy criteria as shown in Fig. 2. Unfortunately formal possibilities for determining in a unique way the type of nonlinearity not always exist, especially when the data samples are short.

The next step is model parameter estimation by making use of alternative techniques; in linear case these are the following ones: ordinary least squares

(OLS) and its clones, maximum likelihood (ML) and many others. In a case of nonlinear model estimation the following methods are useful: ML, Markov Chain Monte Carlo (MCMC) procedures [19], nonlinear least squares (NLS) and other suitable approaches able to provide unbiased parameter estimates under specific probabilistic distributions of model variables and model structures. Correct application of alternative parameter estimation techniques provides a possibility for further comparison of the candidate models and selection of the best one. It is also possible to trace the reasons for existing parametric uncertainties in the following form: parameter estimates computed with statistical data cannot be consistent, they may contain bias, and can be inefficient. All these effects finally result in poor adequacy of the model constructed.

At the next stage is computed a set of statistical parameters characterizing model quality (adequacy) and selecting the most suitable model out of the set of candidate models. There is no need to leave only one model for computing forecasts (or solving control problem). Again, it can be a set of the “best” models constructed on different ideologies. The final choice is always made after models application for solving the problem according to the initial problem statement.

After computing the process (under study) forecasts using candidate models another set of forecast quality criteria is applied to select the best result, say mean absolute percentage error (MAPE), Theil coefficient, mean absolute error (MAE), minimum and maximum errors of forecasting etc. The models constructed should also be tested on similar process, i.e. model calibration process performed.

At this point we can conclude that availability of the data uncertainties mentioned, and the necessity for hierarchical construction of the data processing system with the features of adaptation and optimization (structural and parametric) require application of the modern systemic approach that provides a possibility for successful and high quality solving the problems encountered during statistical data processing, mathematical model construction, forecast estimation and generating the decision alternatives. In this study we propose some practical possibilities for constructing data processing procedures based on modern principles of systemic approach.

**Dealing with model structure uncertainties.** When using DSS, model structure should practically always be estimated using data. It means that elements of the model structure accept almost always only approximate values. When a model is constructed for forecasting we build several candidates and select the best one of them using a set of model quality (adequacy) statistics. Generally we could define the following techniques to fight structural uncertainties: gradual refinement of model order (for  $AR(p)$  or  $ARMA(p, q)$  structures) applying adaptive approach to modeling and automatic search for the “best” structure using complex statistical quality criteria; adaptive estimation of input delay time (lag) and the type of data distribution with its parameters; formal description of detected process nonlinearities using alternative analytical forms with subsequent estimation of model adequacy and forecast quality. A simple example of the complex model and forecast criterion may look as follows:

$$J = \left| 1 - R^2 \right| + \left| 2 - DW \right| + \beta \ln(MAPE) \rightarrow \min_{\theta_i};$$

or in more complicated form:



$$J = \left| 1 - R^2 \right| + \alpha \ln \left[ \sum_{k=1}^N e^2(k) \right] + |2 - DW| + \beta \ln(MAPE) + U \rightarrow \min_{\hat{\theta}_i},$$

where  $R^2$  is determination coefficient;  $\sum_{k=1}^N e^2(k) = \sum_{k=1}^N [y(k) - \hat{y}(k)]^2$  is a sum of squared model errors;  $DW$  is Durbin-Watson statistic;  $MAPE$  is mean absolute percentage error for one step-ahead forecasts;  $U$  is Theil coefficient that characterizes forecasting capability of a model;  $\alpha, \beta$  are appropriately selected weighting coefficients;  $\hat{\theta}_i$  is parameter vector for  $i$ -th candidate model. A combined criterion of this type is used for automatic selection of the best candidate model constructed. The criteria presented also allow operation of DSS in adaptive mode. Obviously, other forms of the combined criteria are possible dependently on specific purpose of model building. What is important while constructing the criterion: not to overweigh separate members in right hand side that would suppress other components.

**Coping with uncertainties of a level (amplitude) type.** The availability of random and/or non-measurable variables results in the necessity of hiring fuzzy sets for describing processes in such situations. The variable with random amplitude can be described with some probability distribution if the measurements are available or when they come for analysis in acceptable time span. However, some variables cannot be measured in principle, say amount of shadow capital that “disappears” every month in offshore zones, or amount of shadow salaries paid at some company, or a technology parameter that cannot be measured on-line due to absence of appropriate gauge or in-situ physical difficulties. In such situations it is possible to assign to the variable a set of characteristic values in linguistic form, say as follows: *capital amount* = { *very low, low, medium, high, very high* }. There exists a complete set of necessary mathematical operations to be applied to such fuzzy variables. Finally fuzzy value can be transformed into exact non-fuzzy form using known transformation techniques.

**Probabilistic uncertainties and their description.** The use of random variables leads to the necessity of estimating actual probability distributions and their application in inference computing procedures. Usually observed value is known only approximately though we know the limits for the actual values. Appropriate probability distributions are very useful for describing the processes under study in such situations. When dealing with discrete outcomes, we assign probabilities to specific outcomes using a mass function. It shows how much “weight” (or mass) to assign to each observation or measurement. An answer to the question about the value of a particular outcome will be its mass. The Kolmogorov’s axioms of probability are helpful for deeper understanding of what is going on. If two or more variables are analyzed simultaneously it is necessary to construct and use joint distributions. Joint distributions allow estimation of conditional probabilities using renormalization procedures when necessary.

Very helpful for performing probabilistic computations is a notion of conditional independence:  $P(x, y | z) = P(x | z) P(y | z)$ , where  $x$  and  $y$  are independent events. Such identities are very handy though one should be careful when using them, i.e. the events should be actually independent. The remarkable intui-

tive meaning of discrete Bayes' law,  $P(A/B)=P(B/A)P(A)/P(B)$ , is that it allows to ask the reverse questions: "Given that event  $A$  happened, what is the probability that a particular event  $B$  evoked it?" The marginal probability,  $P(B)$ , can be computed using appropriate conditionals. The probability that event  $B$  will occur in general,  $P(B)$ , could be obtained from the following condition:  $P(B)=P(B/A)P(A)+P(B/\bar{A})P(\bar{A})$ .

The probabilistic types of uncertainties regarding whether or not some event will happen can be taken into consideration with probabilistic models. To solve the problem of describing and taking into account such uncertainties a variety of Bayesian models could be hired that are considered as Bayesian Programming formalism. The set of the models includes Bayesian networks (BN), dynamic Bayesian networks (DBN), Bayesian filters, particle filters, hidden Markov models, Kalman filters, Bayesian maps etc. The structure of Bayesian program includes the following elements: problem description and statement formulation with a basic question of the form:  $P(\text{Searched}/\text{Known})$  or  $P(X_i/D,Kn)$ , where  $X_i$  defines one variable only, i.e. what should be estimated using specific inference engine; the use of prior knowledge  $Kn$  and experimental data  $D$  to perform model structure and parameters identification; selection and application of pertinent inference technique to answer the question stated before; testing quality of the final result. Such approach also works well in adaptation mode aiming to adjusting structure and parameters of a model being developed to new experimental data or a new system operation mode, for example, for estimation of prior distributions or BN structure.

#### **SOME SYSTEM ANALYSIS PRINCIPLES USED IN DSS IMPLEMENTATION**

In our study we propose to use the following system analysis principles for implementing specialized DSS for modeling and forecasting: the systemic function coordination principle; the principle of procedural completeness; the functional orthogonality principle; the principle of dependence of mutual information between the functions being implemented; the principle of functional rationality; the principle of multipurpose generalization; the principle of multifactor adaptation, and the principle of rational supplement [20–22].

The principle of systemic functions coordination supposes that all the techniques, approaches, and algorithms (functions) implemented in the system should be structurally and functionally coordinated, and should be mutually dependent. This way it is possible to create and practically implement a unique systemic methodology for statistical data analysis in the frames of modern DSS, and to improve substantially quality of intermediate and final results. The next systemic principle of procedural completeness guaranties that the system developed will provide the possibility for timely and in place execution of all necessary computing functions directed towards data collection (editing, normalizing, filtering and renewing), formalization of a problem statement, model constructing, computing forecasts, and for performing estimation quality of the model and the forecast estimates based upon it.

Development and implementation of all computational procedures in the DSS using mutually independent functions corresponds to the principle of func-

tional orthogonality. Such approach to the DSS constructing is directed towards substantial enhancement of computational stability of the system and simplification of its further possible modifications and functional enhancement. According to the principle of mutual informational dependence the results of computing, generated by each procedure, should correspond to the formats and requirements of other procedures. This feature is easily implemented with respective project development solutions for the system created.

Application of the systemic principle of goal directed correspondence to computational procedures and functions provides a good possibility for reaching of a unique goal set in advance: high (acceptable) quality of the final result in the form of forecast estimates for the process under study as well as alternative decisions based upon the forecasts.

According to the systemic principle of multipurpose generalization all functional modules for the system developed should possess necessary degree of generalization that provides a possibility for reaching high quality solution results for a set of possible problems that belong to the same class (it can be high quality forecasting and decision alternative generation regarding future evolution of linear or nonlinear non-stationary processes). Among these problems could be the following: accumulating necessary data and their preliminary processing; estimation of structure and parameters for a set of candidate mathematical models; constructing forecasting functions on the models developed and computing of appropriate forecasts; selecting the best results of computing using appropriate sets of quality criteria.

The systemic principle for multifactor adaptation is directed towards the possibility of solving the problems of computational procedures adaptation to the problems of modeling various processes of different complexity depending on the completeness of available information and user requirements. The adaptation is performed within the process of model structure and parameters estimation, i.e. the whole identification procedure of a process under study is compiled from a set of adaptive procedures directed towards reaching the main goal of a study: constructing adequate model and computing high quality forecasts.

Hiring the rational supplement principle provides a possibility for expanding the sphere of application of the DSS constructed by adding new processes types, computational procedures and criteria sets. These new procedures could be directed towards implementation of additional preliminary data processing procedures, model structure and parameter estimation as well as selection of the best result for its further use aiming generating of appropriate decision. Implementation in the frames of the constructed DSS of the systemic principles mentioned above favors its functional flexibility, computational reliability, quality enhancement for the intermediate and final results, prolongation of system life span, and simplification of possible drawbacks elimination and modification procedures.

Finally, the forecasting models and methods used in the system are the following: regression analysis, the group method for data handling (GMDH), fuzzy GMDH, fuzzy logic, appropriate versions of the optimal Kalman filter (KF), neural nets, support vector regression, nearest neighbor and probabilistic type techniques like Bayesian networks and regression. The set of modeling techniques used covers linear and many types of nonlinear non-stationary processes. The nearest neighbor technique is hired for generating long term forecasts in a case of

availability long data samples with periodical patterns. All the techniques are implemented in adaptive versions what makes the system more flexible for newly coming data and capable to fight some types of possible uncertainties mentioned above. During the process of model structure estimation an appropriate principal component analysis technique is applied when necessary.

## **BAYESIAN NETWORK ADAPTATION**

Bayesian networks (BN) create one of the powerful modern probabilistic instruments for solving the problems of mathematical modeling, forecasting, classification, control and decision support [23–26]. To estimate BN model structure the algorithms are used on the basis of statistical data that characterize evolution of the network variables. It is possible to develop and use the algorithms that allow for adaptation of the network structure to the new data coming in real time. This is a choice used in the DSS with adaptation features.

The adaptation procedure could be explained using the following notation:  $Z = \{X_1, \dots, X_n\}$  is a set of BN model nodes that is determined by the number of variables hired to construct appropriate directed graph;  $E = \{(X_i, X_j) | X_i, X_j \in Z\}$  is a set of BN arcs;  $X_i$  is a BN node that corresponds to the observations of one variable;  $n = |Z|$  is a total number of BN nodes;  $r_i$  is a number of values that could be accepted by the node  $X_i$ ;  $v_{ik}$  is the  $k$ -th value of variable  $X_i$ ;  $\Pi_i$  is the set of parent nodes for the variable  $X_i$ ;  $\varphi$  is the set of possible initializations  $\Pi_i$  for the node  $X_i$ ;  $q_i = |\varphi_i|$  is the number of possible initializations  $\Pi_i$ ;  $\varphi_{ij}$  is  $j$ -th initialization for the set of parent nodes  $\Pi_i$  for  $X_i$ ;  $B_S$  is structure of BN;  $B_P$  is probabilistic specification of BN, i.e. the part of BN description that represents its probabilistic characteristics,  $\theta_{ijk} = p(X_i = v_{ik} | \varphi_{ij}, B_P)$  under condition that the sum of the probabilities  $\sum_k \theta_{ijk} = 1$ ;  $f(\theta_{ij1}, \dots, \theta_{ijr_i})$  is the probability density for the node  $X_i$  and initialization  $\varphi_{ij}$ ;  $D_0$  is database;  $S_0$  is preliminary estimate of BN structure computed on the basis of available data  $D_0$ ;  $D_1$  is database of observations that were not used for estimating preliminary structure  $S_0$ ;  $S_1$  is BN structure found after  $S_0$  adaptation to the new data  $D_1$ . The problem is to construct algorithm for adaptation of initial Bayesian network  $G = \langle Z, E \rangle$  having the structure,  $S_0$ , to the new observations  $D_1$ .

This way a new (or modified) model structure will be formed:  $S_1 \Leftrightarrow D_1$ . The statistical data used could exhibit arbitrary probability distribution, and the processes described by the data could be of nonlinear non-stationary nature i.e. their mathematical expectation  $E[X_i] \neq \text{const}$  and variance,  $E\{X_i - E[X_i]\}^2 \neq \text{const}$ .

Adaptation of the BN to new data is implemented in the following way:

- implementation of the procedure for refining the model structure: here the model arcs can be eliminated or added;

- correcting the probabilistic part of the model (conditional probability tables or CPTs).

At the initial stage of learning BN the probabilistic part of the model is represented in the form of CPTs that are computed on the basis of the frequency analysis of available statistical data. Consider the procedure of correcting this probabilistic part of the model. For this purpose it is more convenient to save (and use) the values of  $N_{ijk}$  instead of the CPTs themselves, where  $N_{ijk}$  is a number of values corresponding to the,  $\theta_{ijk}$ . This way it is possible to perform renewing the data faster regarding conditional distributions and the values themselves could be computed using the Dirichlet expression:

$$p(X_i = v_{ik} | \Pi_i = \varphi_{ij}) = \frac{N_{ijk} + 1}{N_{ij} + r_i}.$$

When correcting the BN structure the order of the nodes analysis will be determined by the value that each node provides for the following conditional probability [27]:

$$p(D_1 | D_0, S_0) = \prod_{i=1}^n \frac{\prod_{s=1}^{R_i} \prod_{t=1}^{Q_i} \prod_{u=1}^{m_{its}} (N_{its} + u)}{\prod_{t=1}^{Q_i} \prod_{u=1}^{M_{it}} (N_{it} + r_i - 1 + u)}.$$

An informational importance of the model arcs is performed as follows. To determine the necessity of deleting a node the following value is computed:  $K_{\text{delete}}(S_0)$  for the current configuration of the parent nodes set. Also the value of  $K_{\text{delete}}(S_{-1}^m)$  is computed for the directed graph configurations that represent the result of deleting one of  $M$  ( $1 \leq m \leq M$ ) input arcs for the current node. Under condition  $K_{\text{delete}}(S_{-1}^m) \leq K_{\text{delete}}(S_0)$  the  $m$ -th arc continues to belong to the model structure because its elimination results in decreasing of the local quality functional (i.e. for the current node). Otherwise the arc is registered in the list of arcs that should be tested further on for elimination. The further testing is based upon computing the value of the local functional for initially set configuration (structure) and for the configurations that result from eliminating of one of the arcs that still are left in the list.

As far as BN model constructing strategy is based upon the general functional

$$P(S_1 | D_1, D_0, S_0) = \arg \max_S \frac{P(S | D_0) P(D_1 | S, D_0)}{P(D_1 | S_0, D_0)},$$

the arcs elimination and adding procedure is of optimization type and is performed in the following way. The arc elimination should result in decreasing the value of the first member in the nominator,  $P(S | D_0)$ , because it reaches maximum with  $S = S_0$  when initial structure  $S_0$  is formed. Generally, to get a positive effect of adaptation it is necessary to compensate the loss due to arc elimination by the effect of adding new arc. That is why the search for the arc to be added to the graph is performed as mentioned above. Estimation of effect due to adding the arc is also based on the local quality functional, its value should increase [27, 28].

## EXAMPLES OF THE DSS APPLICATION

**Example 1.** Numerous examples of model constructing and forecasting have been solved with the DSS developed. In this example bank client's solvency is analyzed, i.e. application scoring is estimated. The database used consisted of 4700 records that were divided into learning sample (4300 records), and test sample (400 records). The default probabilities were computed and compared to actual data, and also errors of the first and second type were computed using various levels of cut-off value. It was established that maximum model accuracy reached for Bayesian network was 0,787 with the cut-off value 0,3. The Bayesian network is "inclined to over insurance", i.e. it rejects more often the clients who could return the credit. The model accuracy and the errors of type I and type II depend on the cut-off level selected. The cut-off value determines the lowest probability limit for client's solvency, i.e. below this limit a client is considered as such that will not return the credit. Or the cut-off value determines the lowest probability limit for client's default, i.e. below this limit a client is considered as such that will return the credit. As far as the cut-off value of 0,1 or 0,2 is considered as not important, in practice it is reasonable to set the cut-off value at the level of about 0,25–0,30. Statistical characteristics characterizing quality of the models constructed are given in Table 1.

**Table 1.** Adequacy of the models constructed

Model type	Gini index	AUC	Common accuracy	Model quality
Bayesian network	0,719	0,864	0,787 (0,806)	Very high
Logistic regression	0,685	0,858	0,813 (0,828)	Very high
Decision tree	0,597	0,798	0,775	Acceptable
Linear regression	0,396	0,657	0,631 (0,639)	Unacceptable

Thus, the best models for estimation of probability for credit return are logistic regression and Bayesian network. The best common accuracy showed logistic regression, 0,813, though Bayesian network exhibited higher Gini index, 0,719 (the values in parenthesis show improvement due to application of adaptive mode of modeling). The decision tree hired is characterized by Gini index of about 0,597, and  $CA = 0,775$ . It should be stressed that acceptable values of Gini index for developing countries like Ukraine are located usually in the range between 0,4–0,6. The Bayesian network constructed and nonlinear regression showed high values of Gini index that are acceptable for the Ukrainian economy in transition.

**Example 2.** In this case the following four types of scoring were studied:

- *application scoring* that is based on the data given by clients during the process of analyzing the possibility for providing them with a loan;
- *behavioral scoring* or scoring analysis within the period of loan usage; this study was directed to monitoring of a loan keeper account state, in this case we estimated the probability of timely return of the loan by clients, optimal loan limit for the loans etc.;
- *strategic scoring* that is directed towards determining the strategy regarding non-reliable loan keepers violating the rules established;
- *fraud scoring* the purpose of which is to determine the probability of potential fraud on behalf of clients.

The database used in this case consisted of 96000 records with 30 tokens for each client. Some results of computational experiments are presented in Table 2.

**Table 2.** Results of computational experiments for application and behavior scoring

Model used	Application scoring			Behavior scoring		
	Mean AUC	Common accuracy	Learning time	Mean AUC	Common accuracy	Learning time
Logistic regression	0,917	0,873	3,47	0,905	0,854 (0,876)	2,66
Bayesian network	0,922	0,862	2,98	0,913	0,851 (0,864)	2,86
Gradient boosting	0,974	0,925	148,64	0,971	0,911 (0,929)	150,78

The table contains common accuracy values for the computational experiments without adaptation and with adaptation in parenthesis. The adaptation mode has always generated better results than the mode without adaptation feature. For the purpose of simulating adaptation mode the data were divided into parts of equal size (3000 records in each part) and then after model constructing and usage the new data portion was fed into model constructing algorithm.

To analyze strategic scoring the subset of data was used that characterizes annual income of active clients and their total expenditures according to their credit cards within a year. The purpose of the study is to divide clients into clusters and to apply a unique management strategy to each cluster using *K*-means technique. The basic parameter for using *K*-means clustering technique is a number of clusters *K*. The parameter is estimated using the concept of minimizing sum of squares criterion within a cluster (WCSS). It was established that six clusters provide for an acceptable clustering of the clients:

- *K*1: an average income and low expenses;
- *K*2: low income and low expenses;
- *K*3: high income and high expenses;
- *K*4: low income and high expenses;
- *K*5: an average income and high expenses;
- *K*6: very high income and high expenses.

The fraud analysis was performed with the highly unbalanced data: 187 operations out of the total number of operations 86754 were classified as the fraud. The positive class of the data (fraud) included 0,215% of all the operations performed. The Bayesian network constructed on the data showed AUC = 0,863. After the data was corrected with expanding the smaller class of data (oversampling approach) the result of classification was improved to the following: AUC = 0,896. Finally a combined approach was applied to solving the problem that supposes application of oversampling, elimination of “noise” from the observations, and gradual improvement of balance between the classes to about 40 : 60 and 50 : 50. The result of classification was improved to the AUC = 0,928, and in adaptation mode to the value of about AUC = 0,935.

**Example 3.** As an example of methodology application a time series was studied, the values of which were gold prices within the period between the years

2005–2006 (sample contains 504 values). The statistical characteristics showing constructed models and forecasts quality are given in Table 3. Here the case is considered when adaptive Kalman filter was not used for preliminary data processing smoothing.

**Table 3.** Models and forecasts quality without adaptive Kalman filter application

Model type	Model quality			Forecast quality			
	$R^2$	$\sum e^2(k)$	DW	MSE	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

Thus, the best model turned out to be AR(1) + trend of 4<sup>th</sup> order. It provides a possibility for one step ahead forecasting with mean absolute percentage error of about 3,19%, and Theil coefficient is  $U = 0,024$ . The Theil coefficient shows that this model is generally good for short-term forecasting. Statistical characteristics of the models and respective forecasts computed with adaptive Kalman filter application for data smoothing are given in Table 4. Here optimal filter played positive role what is supported by respective statistical quality parameters.

**Table 4.** Models and forecasts quality with application of adaptive 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) + 1 <sup>st</sup> order trend	0,99	23964,73	2,08	31,15	22,11	3,27	0,029
AR(1) + 4 <sup>th</sup> order trend	0,99	22396,83	2,04	21,35	13,52	2,71	0,019

Again the best model turned out to be AR(1) + trend of 4<sup>th</sup> 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.

**Example 4.** 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 5 contains statistical charac-



teristics of the models constructed as well as quality of short-term variance forecasting. To solve the problem we used generalized autoregressive conditionally heteroscedastic (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. The values of MAPE (adapt.) given in the 6th column for the mode of operation with adaptation show improvement of short term forecasting for conditional variance when modeling system operated in the mode with model adaptation.

**Table 5.** Results of modeling and forecasting conditional variance

Model type	Model quality			Forecast quality			
	$R^2$	$\sum e^2(k)$	DW	MSE	MAPE (adapt.)	MAPE	Theil
GARCH(1,7)	0,99	153639	0,113	972,5	515,3	517,6	0,113
GARCH (1,10)	0,99	102139	0,174	458,7	208,2	211,3	0,081
GARCH (1,15)	0,99	80419	0,337	418,3	118,7	121,6	0,058
EGARCH (1,7)	0,99	45184	0,429	67,8	7.85	8,74	0,023

Thus, the best model constructed was exponential GARCH(1,7). The achieved value of MAPE = 8,74% (and 7,85% in adaptation mode) comprises very good result for forecasting conditional variance. Further improvements of the forecasts were achieved with application of the adaptation scheme proposed. An average improvement of the forecasts was in the range between 0,8–1,5%, 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.

**DISCUSSION**

The results of computational experiments achieved lead to the conclusion that today the family of scoring models used including logistic regression, Bayesian networks and gradient boosting belong to the family of the best current instruments for banking system due to the fact they provide a possibility for detecting “bad” clients and to reduce financial risks caused by the clients. It also should be stressed that DSS developed creates very useful instrument for a decision maker that helps to perform quality processing of client’s statistical data using various techniques, generate alternatives and to select the best one relying upon a set of appropriate statistical criteria. An important role in the computational experiments performed played the possibility of model adaptation to available and new data. The adaptation mode has always generated better results than the mode without this adaptation feature. The extra model variables can be created by combining available statistical data, and nonlinearities can be introduced into a model by inserting appropriate polynomial members. The system proposed performs tracking of the whole computational process using separate sets of statistical quality criteria at each stage (each level of the system hierarchy) of decision making: quality

of data, adequacy of models constructed and quality of the forecasts (or risk estimates).

Thus, the systemic approach to modeling and forecasting proposed is definitely helpful for constructing the DSS possessing the features of directed search for the best forecasting model in respective spaces of model structures and parameters, and consequently to enhance its adequacy. The computational experiments with actual data showed high usefulness of the systemic approach to modeling and forecasting. It is necessary to perform its further refinement in the future studies and applications. And it is also important to improve formal descriptions for the uncertainties mentioned and to use them for reducing the degree of uncertainty in model building procedures and forecast estimation. It was found that influence of statistical and probabilistic uncertainties can be reduced substantially by making use of respective data filtering techniques, imputation of missing values, orthogonal transforms, and the models of probabilistic type; first of all those are Bayesian programming models and techniques.

## **CONCLUSIONS**

The systemic methodology was proposed for constructing decision support system for adaptive mathematical modeling and forecasting modern economic and financial processes as well as for credit risk estimation that is based on the following system analysis principles: hierarchical system structure, taking into consideration probabilistic and statistical uncertainties, availability of model adaptation procedures, generating multiple decision alternatives, and tracking of computational processes at all the stages of data processing with appropriate sets of statistical quality criteria (known or newly introduced).

The system developed has a modular architecture that provides a possibility for easy extension of its functional possibilities with new parameter estimation techniques, forecasting methods, financial risk estimation, and generation of decision alternatives. High quality of the final result is achieved thanks to appropriate tracking of the computational processes at all data processing stages: preliminary data processing, model structure and parameter estimation, computing of short- and middle-term forecasts, and estimation of risk variables/parameters. The system is based on the ideologically different methods of dynamic processes modeling and risk forecasting (regression analysis and probabilistic approach) what creates appropriate basis for hiring various approaches to achieve the best results. The illustrative examples of the system application show that it can be used successfully for solving practical problems of forecasting dynamic processes evolution and risk estimation. The results of computational experiments lead to the conclusion that today scoring models, nonlinear regression and Bayesian networks are the best instruments for banking system due to the fact that they provide a possibility for detecting “bad” clients and to reduce financial risks caused by the clients. It also should be stressed that the DSS constructed turned out to be very useful instrument for a decision maker that helps to perform quality processing of statistical data using ideologically different techniques, appropriate sets of statistical quality criteria, generate alternatives and select the best one. The DSS can be used for supporting decision making process in various areas of human activities including development of strategy for banking system regarding risk management and industrial enterprises, investment companies etc.

Further extension of the system functions is planned with new forecasting and decision making techniques based on probabilistic methodology, fuzzy sets and other artificial intelligence methods. An appropriate attention should also be paid to constructing user friendly adaptive interface based on the human factors principles.

## REFERENCES

1. *Jao C.S.* Efficient decision support systems – practice and challenges from current to future / C.S. Jao. — Rijeka (Croatia): Intech, 2011. — 556 p.
2. *Fernandez G.* Data mining using SAS applications / G. Fernandez. — New York: CRC Press LLC, 2003. — 360 p.
3. *Dovgyj S.O.* DSS on the basis of statistical and probabilistic methods / S.O. Dovgyj, P.I. Bidyuk, O.M. Trofymchuk. — Kyiv: Logos, 2014. — 419 p.
4. *Zgurowskii M.Z.* System analysis: problems, methodology, applications / M.Z. Zgurowskii, N.D. Pankratova. — Kyiv: Naukova Dumka, 2005. — 745 p.
5. *Harris C.* Adaptive modeling, estimation and fusion from data / C. Harris, X. Hong, Q. Gan. — Berlin: Springer, 2002. — 322 p.
6. *Harvey A.C.* Forecasting, structural time series models and the Kalman filter / A.C. Harvey. — Cambridge: The MIT Press, 1990. — 554 p.
7. *Rasmussen C.E.* Gaussian processes for machine learning / C.E. Rasmussen, C.K.I. Williams. — Cambridge (Massachusetts), The MIT Press, 2006. — 248 p.
8. *Bidyuk P.I.* Time series analysis / P.I. Bidyuk, V.D. Romanenko, O.L. Tymoshchuk. — Kyiv: Polytechnika, NTUU «KPI», 2013. — 600 p.
9. *Almeida E.* Adaptive model rules from data streams / E. Almeida, C. Ferreira, J. Gama // Machine Learning and Knowledge Discovery in Data bases. ECML PKDD-2013. Lecture Notes in Computer Science. — Springer, Berlin. — Vol. 8188. — P. 480–492.
10. *Succarat G.* Automated model selection in finance: general-to-specific modeling of the mean, variance and density / G. Succarat, A. Escribano // Oxford Bulletin of Economics and Statistics. — 2012. — Vol. 74, Issue 5. — P. 716–735.
11. *Pretis F.* Automated general-to-specific (GETS) regression modeling and indicator saturation for outliers and structural breaks / F. Pretis, J.J. Reade, G. Succarat // Journal of Statistical Software. — 2018. — Vol. 86, Issue 3. — P. 1–44.
12. *Quintana R.* Adaptive exponential smoothing versus conventional approaches for lumpy demand forecasting: case of production planning for a manufacturing line / R. Quintana, M.T. Leung // International Journal of Production Research. — 2007. — Vol. 45, Issue 21. — P. 4937–4957.
13. *Giraitis L.* Adaptive forecasting in the presence of recent and ongoing structural change / L. Giraitis, G. Kapetanias, S. Price // Journal of Econometrics. — 2013. — Vol. 177, Issue 2. — P. 153–170.
14. *Pesaran M.H.* Optimal forecasts in the presence of structural breaks / M.H. Pesaran, A. Pick, M. Pranovich // Journal of Econometrics. — 2013. — Vol. 177, Issue 2. — P. 134–152.
15. *Watsham T.J.* Quantitative Methods in Finance / T.J. Watsham, K. Parramore. — London: International Thomson Business Press, 1997. — 395 p.
16. *Xekalaki E.* ARCH Models for Financial Applications / E. Xekalaki, S. Degiannakis. — New York: John Wiley & Sons, Ltd, Publication, 2010. — 535 p.
17. *Gibbs B.P.* Advanced Kalman filtering, least squares and modeling / B.P. Gibbs. — Hoboken: John Wiley & Sons, Inc., 2011. — 627 p.
18. *Haykin S.* Adaptive filter theory / S. Haykin. — Upper Saddle River (New Jersey): Prentice Hall, 2002. — 922 p.

19. 19. *Gilks W.R.* Markov Chain Monte Carlo in practice / W.R. Gilks, S. Richardson, D.J. Spiegelhalter. — New York: CRC Press LLC, 2000. — 486 p.
20. *Zgurowskii M.Z.* System analysis: problems, methodology, applications / M.Z. Zgurowskii, N.D. Pankratova. — Kyiv: Naukova Dumka, 2005. — 743 p.
21. *Anfilatov V.S.* System analysis in control engineering / V.S. Anfilatov, A.A. Emelyanov, A.A. Kukushkin. — Moscow: Finansy i Statistika, 2002. — 368 p.
22. *Zgurowskii M.Z.* Analytical technics of Kalman filtering / M.Z. Zgurowskii, V.N. Podladchikov. — Kyiv: Naukova Dumka, 1995. — 285 p.
23. *Cowell R.G.* Probabilistic networks and expert systems / R.G. Cowell, A.Ph. Dawid, S.L. Lauritzen, D.J. Spiegelhalter. — Berlin: Springer, 1999. — 321 p.
24. *Jensen F.V.* Bayesian networks and decision graphs / F.V. Jensen, Th.D. Nielsen. — Berlin: Springer, 2007. — 427 p.
25. *Koski T.* Bayesian networks / T. Koski, J.M. Noble. — New York: John Wiley and Sons, Ltd., Publication, 2009. — 347 p.
26. *Zgurowskii M.Z.* Methods of constructing Bayesian networks based on scoring functions / M.Z. Zgurowskii, P.I. Bidyuk, O.M. Terentyev // Cybernetics and System Analysis. — 2008. — Vol. 44, N 2. — P. 219–224.
27. *Ng B.M.* Adaptive dynamic Bayesian networks / B.M. Ng. // Joint Statistical Meetings. — 2007. — 9 p.
28. *Corriveau G.* Bayesian network as an adaptive parameter setting approach for genetic algorithms / G. Corriveau, R. Guilbault, R. Tahan, R. Sabourin // Complex Intelligent Systems. — 2016. — N 1. — P. 1–23.

*Received 04.12.2019*

---

From the Editorial Board: the article corresponds completely to submitted manuscript.