ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ, СИСТЕМНИЙ АНАЛІЗ ТА КЕРУВАННЯ

DOI: 10.20535/kpi-sn.2019.5-6.176835
UDC 004.942:519

P.I. Bidyuk[1*], L.O. Korshevnyuk[1], O.P. Gozhyj[2], I.O. Kalinina[2]
T.I. Prosyankina-Zharova[1], O.M. Terentiev[1]
[1]Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine
[2]Petro Mohyla Black Sea National University, Mykolayiv, Ukraine
[*]corresponding author: pbidyuke_00@ukr.net

MODELING AND FORECASTING FINANCIAL AND ECONOMIC PROCESSES WITH DECISION SUPPORT SYSTEM

Background. The modern financial and economic processes and accompanying risks often exhibit sophisticated patterns, contain non-stationary and non-linear features that require development of special models for their description and forecasting. To solve the problem successfully it is helpful to construct appropriate decision support system using systemic principles.
Objective. Development of decision support system architecture and its functional layout for economic and financial processes model constructing with statistical data as well as financial risk estimation. The system should help coping with possible uncertainties and implemented on the basis of modern information technologies.
Methods. Mathematical modeling and forecasting techniques for financial and economic process; approaches to financial risks estimation using statistical data. The use of modern information technologies for practical implementation of decision support system.
Results. Information technology and decision support system as a practical tool for modeling nonlinear non-stationary processes in economy and finances, as well as financial risk estimation were developed. Experimental results of statistical data processing prove the correctness of the approaches proposed.
Conclusions. The systemic methodology is proposed and implemented for constructing decision support system for mathematical modeling and forecasting modern economic and financial processes as well as for financial 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 features, generating multiple decision alternatives, and tracking of computational processes at all the stages of data processing with appropriate sets of statistical quality criteria.
Keywords: information technologies; decision support system; nonlinear non-stationary financial processes; uncertainties; financial risks; modeling; forecasting.

Introduction

The problem of forecasting dynamics of the processes in economy, finances (EFP) and many other areas of human activities requires from researchers more and more efforts directed to reaching high quality of respective forecast estimates. It can be easily shown that most of the modern processes under study are nonlinear and non-stationary. They are characterized by availability of high order trends, transitions between different modes of operation, for example from post-soviet to market economy. Most of the processes in finances are heteroscedastic, i.e. their variance is changing in time what also relates them to nonlinear and nonstationary. When analyzing such processes, the main problem is to construct mathematical model capable to generate short-term forecasts of volatility and variance. Here special model structures are required that include dynamics of process variance as the main variable. A comprehensive treatment of the models and their constructing procedures can be found in [1—3].

High quality of forecasting dynamic processes could be reached with some market available data processing instruments. Well known SAS system contains modern data processing procedures that are capable of producing high quality final results of forecasting, risk estimation, etc. [3, 4]. However, very often such systems are very costly and poorly accessible; they require long special training for solving specific problems and need highly developed modern computers for implementation. All these specific features of the systems available create unfavorable conditions for their wide usage except for banks and large enterprises. Besides, many new modeling and forecasting procedures appearing in special publications need to be appropriately implemented and tested. In view of these problems it would be reasonable to develop and use in practice simpler and much less costly modeling respective decision support systems (DSS) constructed with application of modern system analysis principles. The modern systemic approach to modeling and forecasting is based on application of the system analysis methodology that provides an objective ground for constructing DSS possessing necessary functional and computational features [5].

Statistical data processing is accompanied very often by uncertainties of various kind and origin. Here the uncertainties are considered as the factors of negative influence on the analysis processes that result in lower quality of intermediate and final results of data and expert estimates analysis. In particular, in modeling problems the uncertainties mostly exist in statistical, structural and parametric form. The uncertainties of statistical type appear due to impossibility to determine in a unique way the type of probability distribution for external stochastic disturbances and collected data. Knowledge of the data distribution type is very important from the point of view of selecting the method for model parameters estimation. The structural uncertainties are encountered in those cases when analysis of statistical (time series) data doesn't exhibit clear structure that should be used for constructing respective model. Let's remind that the notion of model structure includes the following basic elements: (1) dimensionality determined by the number of equations in a model; (2) model order that is determined by the highest order of a model equation; (3) input delay time (or lag) for independent (input) variables; (4) process nonlinearity and its type: nonlinearity in variables (very often) or in parameters; (5) external stochastic disturbance(s) and its type: type of probability distribution and its parameters; (6) initial conditions and restrictions on variables and/or model parameters [2]. It is known that model order and input delay time (time lags for independent variables) can only be approximately determined using the statistical correlation analysis or appropriate parameter estimation algorithms. Unfortunately, also the possibility for determining in a unique way the type of nonlinearity, especially when the data samples are short, doesn't always exist.

As a consequence for existence of the two previous uncertainty types it is possible to trace the reason for appearing parametric uncertainties in the following form: estimates of a model parameters computed with statistical data can be not consistent, they may contain bias, and also can be inefficient. All this results in poor adequacy of the model constructed.

Thus, availability of the data uncertainties mentioned, and the necessity for hierarchical construction of data processing system with the features of adaptation and optimization require development and application of the modern systemic approach, that provides a possibility for successful and high quality solving of many problems encountered during statistical data analysis, 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.

Problem statement

The purposes of the study are as follows: to consider some principles of the system analysis that are appropriate for solving the problem of successful modeling and short-term forecasting; to develop an efficient data processing scheme for implementation in decision support systems based on these principles; to analyze some uncertainties inherent in data collecting, model building and forecasting process; to show with examples advantages of the system analysis approach developed.

System analysis principles used in the DSS implementation

In this study it is proposed to use the following systems analysis principles for implementing the DSS: the systemic function coordination principle; the principle of procedural completeness; the functional orthogonality principle; the principle of dependence of mutual information between functions implementation; the principle of goal directed correspondence; the principle of functional rationality; the principle of multipurpose generalization; the principle of multifactor adaptation, and the principle of rational supplement [4, 5].

According to the principle of systemic function coordination all the techniques, approaches, and algorithms (functions) used in the system should be structurally and functionally coordinated, and should be mutually dependent, i.e. linked and coordinated with each other. This way it is possible to create and practically implement a unique systemic methodology for statistical data analysis in the frames of modern DSS constructed, 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 estimating quality of the model and the forecast estimates based on it.

Development and implementation of all computational procedures in the DSS using mutually independent functions corresponds to the principle of functional 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 expansion. 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 (or modules). This feature is implemented with respective project development solutions for the system created.

Application of the system principle of goal directed correspondence to computational procedures and functions provides a good possibility for reaching of a unique set goal: high (acceptable) quality of the final result in the form of forecast estimates for the process under study, as well as alternative decisions based on the forecasts. The next principle of functional rationality prevents duplication of separate DSS functions. As a result of following this principle it is quite possible to reach substantial economy of respective computational resources.

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 development 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 necessary mathematical models; constructing forecasting functions on the models developed and computing of appropriate forecasts; selecting the best computing results using appropriate sets of quality criteria.

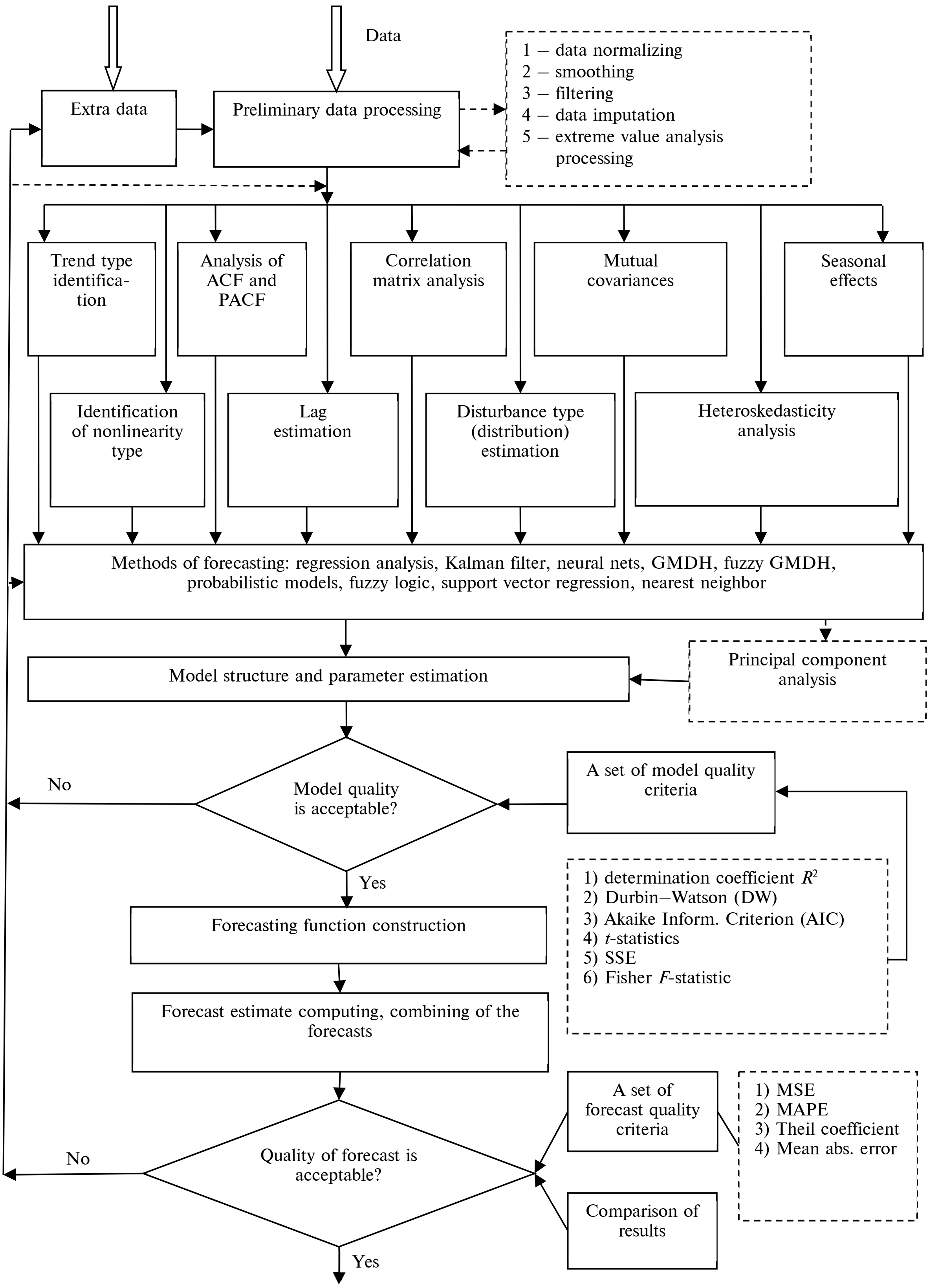
Application of the systemic principle of multifactor adaptation is directed towards the possibility of solving the problems of computing procedures adaptation to the problems of modeling various processes of different complexity depending on the completeness of available information and user requirements. The adaptation is performed during the process of model structure and parameters estimation, i.e. the whole identification process of a process or object under study is formed by 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 data types, computing procedures and criteria. 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. Hierarchical structure of the DSS proposed is given in the Figure. Implementation in the DSS of the system analysis principles mentioned above favors its functional flexibility, computational reliability, enhancement of quality for the final results, prolongation of system life span, and simplification of possible drawbacks elimination and modification procedures.

Now consider some functional elements of the general hierarchical structure of the DSS proposed that is given in the Figure. The preliminary data processing is directed towards data preparation for further model building. Then statistical data analysis is performed aiming to determining type of the process under consideration (linear/nonlinear, and stationary/non-stationary) as well as model structure estimation. An important and useful feature of the system is that it uses separate sets of statistical parameters for analyzing quality of data, adequacy of model constructed and quality of forecast estimates generated with the model. If the forecasts are used for generating alternative decisions then we include another set of criteria for testing decision quality. This way high quality forecasts are generated and the decision quality based on these forecasts is usually quite acceptable.

Further on the statistical analysis of data is performed on the following purposes: testing for heteroskedasticity, for integration (availability of trend), nonlinearity, lag estimation and determining type of data distribution. After this step it is possible to determine the class of the process under study and estimate the elements of model structure mentioned above in the introduction.

The forecasting models and methods used in the system are as follows: 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.



Hierarchical structure of the DSS for modeling EFP and generating forecasts

The set of modeling techniques used covers linear and nonlinear non-stationary process. 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 uncertainties. In the process of model structure estimation an appropriate principal component analysis technique is applied if necessary.

The information regarding model structure provides a possibility for correct selection of parameter estimation techniques for candidate models among which are the following: ordinary and nonlinear LS (NLS), recursive LS (RLS), maximum likelihood (ML), and some versions of Markov Chain Monte Carlo (MCMC) techniques. Some special optimization techniques are applied to compute model parameters in a case of estimating fuzzy GMDH structures.

The candidates models estimated are tested with a set of statistical quality criteria some of which are shown in the Figure. One or (usually) more acceptable model structures are used for computing necessary forecasts which quality is also tested using another set of criteria. Usually two or more modeling and forecasting techniques are used for the same dataset to get a possibility for combining the forecasts so that to further improve the final estimate.

Uncertainties identification and processing

Processing possible statistical uncertainty. The most often considered statistical uncertainties related to model development and forecast estimation are provoked by the following factors [2, 5, 11]:

– availability of measurement errors (or noise) that are present practically in all cases of data collection independently on the data origin (including economy, finances, ecology etc.);

– random external disturbances that negatively influence flow of the process under study and shift it from desired mode (say, offshore capital transfer from some country, low quality of administering in business and economy, unstable legislative system);

– missed (lost) measurements (observations) and outliers;

– multicollinearity of data vectors that requires application of special data processing techniques to reduce degree of mutual correlation between separate time series.

The most often instruments hired to fight measurement noise and external random disturbances are digital and optimal filters, among which is widely used optimal Kalman filter [5]. A little simpler digital filters (DF) help to select for subsequent processing the frequency band of interest by processing the time series data with linear expressions that could be represented by autoregression (AR) or autoregression with moving average (ARMA) equations of the type:

$$y(k) = \sum_{i=1}^{p} a_i y(k-i) + \sum_{j=1}^{q} b_j x(k-j),$$

where $x(k)$ is input sequence of measurements for a filter; $y(k)$ is filtered output sequence; $\theta=[a_1 \ldots a_p \, b_1 \ldots b_q]^T$ is filter parameter vector that actually determines frequency response of a filter; p and q are orders of AR and MA parts, respectively. There exist many handy instruments for DF design based on optimization procedures, say, in Matlab.

Correctly designed adaptive optimal Kalman filter provides a possibility for unknown covariance estimation, for suppressing stochastic disturbances and measurement noise as well as estimation of short-term forecasts [5, 6, 10, 11]. An optimal filter development requires constructing mathematical model of the process (system) under study in the state space form as follows:

$$\mathbf{x}(k) = \mathbf{A}(k, k-1)\mathbf{x}(k-1) + \mathbf{B}(k, k-d)\mathbf{u}(k-1) + \mathbf{w}(k), \quad (1)$$

$$\mathbf{z}(k) = \mathbf{H}\,\mathbf{x}(k) + \mathbf{v}(k), \quad (2)$$

where $\mathbf{x}(k)$ is system state vector including n elements; $\mathbf{u}(k)$ is m – dimensional vector of control inputs; $\mathbf{w}(k)$ is a vector of stochastic external disturbances; $\mathbf{A}(k,k-1)$ is $(n \times n)$ – matrix of system dynamics; $\mathbf{B}(k)$ is $(n \times m)$ matrix of control coefficients; k is discrete time that is linked to continuous time t using sampling time T_s: $t = kT_s$; d is system input delay time; $\mathbf{z}(k)$ is vector of measurements containing r elements; $\mathbf{H}(k)$ is matrix of measurement coefficients that very often has diagonal form); $\mathbf{v}(k)$ is a vector of measurement noise values. Here double time argument $(k,\ k-1)$ means that variable with the argument is used at time, k, though its value is based on the previous measurements up to $(k-1)$. Usually such double argument is not used in text to avoid symbols overload for the mathematical expressions hired.

The main advantage of the model (1), (2) is that it includes explicitly two random process $\mathbf{w}(k)$ and, $\mathbf{v}(k)$; consequently it is a priory more adequate to reality than say linear regression. The main task of optimal filter is in computing optimal according to quadratic criterion state vector estimates with taking into consideration covariances of the two stochastic processes mentioned above. Such approach provides a possibility for improving the state estimates and to compute estimates of non-measurable components of state vector $\mathbf{x}(k)$ if such are available in a problem statement. The main equation for the filter is as follows:

$$\hat{\mathbf{x}}(k) = \mathbf{A}(k)\hat{\mathbf{x}}(k-1) + \mathbf{B}(k)\mathbf{u}(k-1) + \mathbf{K}(k)[\mathbf{z}(k) - \mathbf{H}(k)\mathbf{A}(k)\hat{\mathbf{x}}(k-1)], \qquad (3)$$

where $\mathbf{K}(k)$ is optimal coefficient of the filter in matrix form. The coefficient is computed by minimizing the following quadratic functional:

$$J = \min_{\mathbf{K}} E\{[\hat{\mathbf{x}}(k) - \mathbf{x}(k)]^T [\hat{\mathbf{x}}(k) - \mathbf{x}(k)]\}.$$

where $\mathbf{x}(k)$ is exact value for state vector that could be found using deterministic part of model (1). In linear discrete case the coefficient is rather easily computed by solving Riccati equation.

Thus, correct application of optimal filter provides a possibility for reducing uncertainties in the form of influence of the two stochastic processes $\mathbf{w}(k)$ and, $\mathbf{v}(k)$, and estimation of non-measurable components of a state vector when respective components of covariance matrices are known (we mean covariance matrix of filtering

$$\mathbf{P}(k) = E\{[\hat{\mathbf{x}}(k) - \mathbf{x}(k)]\,[\hat{\mathbf{x}}(k) - \mathbf{x}(k)]^T\}$$

for the system state vector estimation errors). Especially useful are adaptive versions of the filters that are most suitable for practical applications in online and off-line modes of operation. They are suitable for repeated estimation of system (object) matrices $\mathbf{A}(k)$ and $\mathbf{B}(k)$ as well as covariances of the two stochastic processes mentioned [5, 7, 11].

Coping with uncertainties caused by missing observations. For the data in the time series form the most suitable imputation techniques are:

– simple averaging when it is possible (when only a few values are missing);

– generation of forecast estimates with the model constructed using available measurements;

– generation of missing (lost) estimates from distributions the form and parameters of which are again determined using available part of data;

– the use of optimization techniques, say appropriate forms of EM-algorithms (expectation maximization);

– exponential smoothing etc. The simplest model that could be hired for generating short-term forecasts is AR(1):

$$y(k) = a_0 + a_1\, y(k-1) + \varepsilon(k),$$

where a_0, a_1 are model parameters; random process $\varepsilon(k)$ takes into consideration model structure uncertainties: say, lack of appropriate independent variables, external stochastic disturbances, systemic errors of parameter computing etc., and possible measurement errors. If parameters a_0, a_1 are known, we could compute one step-ahead prediction as conditional expectation of the form:

$$\begin{aligned}\hat{y}(k+1) &= E_k[y(k+1)] \\ &= E_k[y(k+1)\,|\,y(k), y(k-1), \ldots, \varepsilon(k), \varepsilon(k-1), \ldots] \\ &= a_0 + a_1\, E_k\,[y(k)] = a_0 + a_1\, y(k),\end{aligned}$$

as far as $y(k)$ at the moment k takes known value. It is possible to derive iteratively the forecasting function for s steps-ahead [2, 9, 11]:

$$\begin{aligned}\hat{y}(k+s) &= E_S[y(k+s)] \\ &= a_0\left(\sum_{i=0}^{S-1} a_1{}^i\right) + a_1{}^S y(k) = a_0 \sum_{i=0}^{S-1} a_1{}^i + a_1{}^S y(k).\end{aligned}$$

The sequence of forecast estimates, $\{\hat{y}(k+i)\}$, $i = 1, \ldots, s$ is convergent if $|a_1| < 1$:

$$\lim_{s\to\infty} E_k\,[\,y(k+s)\,] = \frac{a_0}{1 - a_1}.$$

The last expression means that for stationary AR or ARMA processes the estimates of conditional forecasts asymptotically $(s \to \infty)$ converge to unconditional mean that is considered as a long-term forecast. It should also be mentioned that optimal filter can also be hired for missing data imputation because it contains “internal” forecasting engine providing a possibility for generating quality short-term forecasts (usually one-step-ahead).

Further reduction of this uncertainty type is possible thanks to application of several forecasting techniques to the same problem with subsequent combining of separate forecasts using appropriate weights. It is known that the best results of combining the forecasts are achieved when variances of forecasting errors for different forecasting techniques do not differ substantially.

Coping with uncertainties of model parameters estimates. Usually uncertainties of model parameter estimates such as bias and inconsistency result from low informative data, or data do not correspond to normal distribution, what is often required for correct parameter estimation. This situation may also take place in a case of independent variables multicollinearity and substantial influence of process nonlinearity that for some reason has not been taken into account when model was constructed. When the power of data sample is not satisfactory for model construction it could be expanded by applying special techniques, or simulation is hired, or special model building techniques, such as GMDH, are applied. Very often GMDH produces results of acceptable quality with short samples. If data do not correspond to normal distribution, then ML technique could be used or appropriate Monte Carlo procedures for Markov Chains (MCMC) [8, 9, 11]. The last techniques could be applied with quite acceptable computational expenses when the number of parameters is not large.

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 statistics. Generally we could define the following techniques to fight structural uncertainties: gradual improvement of model order (AR(p) or ARMA(p, q)) 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 data distribution type with its parameters; formal description of detected process nonlinearities using alternative analytical forms with subsequent estimation of model adequacy and forecast quality. An example of complex model and forecast criterion may look as follows:

$$J = |1 - R^2| + |2 - DW| + \beta \ln(MAPE) \to \min_{\hat{\theta}_i};$$

or in more complicated form:

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

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; α, β 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 the right hand side that would suppress other components.

Coping with uncertainties of a level (amplitude) type. The use of random and/or non-measurable variables results in the necessity of hiring fuzzy sets for describing 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, or amount of shadow salaries paid at some company, or a technology parameter that cannot be measures on-line due to absence of appropriate gauge or in-situ physical difficulties. In such situations it is possible assign to the variable a set of characteristic values in linguistic form, say as follows: *capital amount* = {*very low*, *low*, *medium*, *high*, *very high*}. Today there exist complete sets of necessary mathematical operations to be applied to such fuzzy variables. Finally, fuzzy value can be transformed into exact non-fuzzy form using known 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 procedures. Usually observed value is known only approximately though we know the limits for the values. Appropriate probability distributions are very useful for describing 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: $P(x,y \mid z) = P(x \mid z)\ P(y \mid z)$, 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 intuitive meaning of discrete Bayes' law, $P(A/B) = P(B/A)\,P(A)/P(B)$, is that it allows to ask reverse questions: "Given that event A happened, what is the probability that a particular event B evoked it?" [9]. The marginal probability, $P(B)$, can be computed from 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) [10], dynamic Bayesian networks (DBN), Bayesian filters, particle filters, hidden Markov models, Kalman filters, Bayesian maps etc. The structure of the Bayesian program includes the following elements: (1) problem description and statement formulation with a basic question of the form: $P(Searched/Known)$ or $P(X_i/D, Kn)$, where X_i defines one variable only, i.e. what should be estimated using specific inference engine; (2) the use of prior knowledge Kn and experimental data D to perform model structure and parameters identification; (3) selection and application of pertinent inference technique to answer the question stated before; (4) 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 new system operation mode, for example, for estimation of prior distributions or BN structure.

Results and discussion

***Example* 1.** Numerous examples of model constructing and forecasting have been solved with the DSS developed. In this example a 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	Very high
Logistic regression	0.685	0.858	**0.813**	Very high
Decision tree	0.597	0.798	0.775	Acceptable
Linear regression	0.396	0.657	0.631	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 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 selected characteristics for each client. Some results of the computational experiments carried out are presented in Table 2.

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: K1: an average income and low expenses; K2: low income and low expenses; K3: high income and high expenses; K4: low income and high expenses; K5: an average income and high expenses; K6: 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 fraud. The positive class of the data (fraud) includes 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 AUC = 0.928.

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. Extra 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 (level of the system hierarchy) of decision making: quality of data, models and forecasts (or risk estimates).

Thus, the systemic approach to modeling and forecasting proposed is 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

***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	2.66
Bayesian network	0.922	0.862	2.98	0.913	0.851	2.86
Gradient boosting	0.974	0.925	148.64	0.971	0.911	150.78

to enhance its quality. 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.

Conclusions

The systemic methodology was proposed for constructing decision support system for 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 features, generating multiple decision alternatives, and tracking of computational processes at all the stages of data processing with appropriate sets of statistical quality criteria.

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 techniques of modeling and risk forecasting what creates an appropriate basis for combining various approaches to achieve the best results. The example of the system application shows that it can be used successfully for solving practical problems of forecasting and risk estimation. The results of computational experiments lead to the conclusion that today the 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 DSS constructed turned out to be very useful instrument for a decision maker that helps to perform quality processing of statistical data using different techniques, generate alternatives and to 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 and industrial enterprises, investment companies etc.

Further extension of the system functions is planned with new forecasting 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] G. Fernandez, *Data Mining Using SAS Applications.* New York: CRC Press LLC, 2003.

[2] P.I. Bidyuk *et al.*, *Time Series Analysis.* Kyiv, Ukraine: Polytechnika, 2013.

[3] M.Z. Zgurowskii *et al.*, *System Analysis: Problems, Methodology, Applications.* Kyiv, Ukraine: Naukova Dumka, 2005.

[4] V.S. Anfilatov *et al.*, *System Analysis in Control Engineering.* Moscow, Russia: Finansy i Statistika, 2002.

[5] M.Z. Zgurowskii *et al.*, *Analytical Technics of Kalman Filtering.* Kyiv, Ukraine: Naukova Dumka, 1995.

[6] B.P. Gibbs, *Advanced Kalman Filtering, Least Squares and Modeling.* Hoboken: John Wiley & Sons, 2011.

[7] S. Haykin, *Adaptive Filter Theory.* Upper Saddle River (New Jersey): Prentice Hall, 2002.

[8] W.R. Gilks *et al.*, *Markov Chain Monte Carlo in Practice.* New York: CRC Press LLC, 2000.

[9] C.S. Jao, *Efficient Decision Support Systems.* Rijeka, Croatia: Intech, 2011. doi: 10.5772/3371

[10] M.Z. Zgurowskii *et al.*, "Methods of constructing Bayesian networks based on scoring functions", *Cybernetics and System Analysis*, vol. 44, no. 2, pp. 219–224, 2008.

[11] O.M. Trofymchuk *et al.*, *DSS on the Basis of Statistical and Probabilistic Methods.* Kyiv: Logos, 2014.

П.І. Бідюк, Л.О. Коршевнюк, О.П. Гожий, І.О. Калініна, Т.І. Просянкіна-Жарова, О.М. Терентьєв

МОДЕЛЮВАННЯ І ПРОГНОЗУВАННЯ ФІНАНСОВИХ ТА ЕКОНОМІЧНИХ ПРОЦЕСІВ ІЗ ВИКОРИСТАННЯМ СИСТЕМИ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ

Проблематика. Сучасні фінансові й економічні процеси та ризики, що супроводжують їх розвиток, часто характеризуються складними елементами, містять нестаціонарні та нелінійні компоненти, які потребують створення спеціальних математичних моделей для їх адекватного опису і прогнозування. Для успішного розв'язання таких задач необхідно створювати відповідні системи підтримки прийняття рішень із використанням принципів системного аналізу.

Мета дослідження. Розробити архітектуру інформаційних технологій системи підтримки прийняття рішень та її функціональну схему для побудови математичних моделей економічних і фінансових процесів на основі статистичних даних та оцінювання фінансових ризиків. Така система повинна виконувати функції обробки невизначеностей і має бути реалізована на основі сучасних інформаційних технологій.

Методика реалізації. Для розв'язання поставленої задачі використані методи математичного моделювання і прогнозування фінансових та економічних процесів; методи оцінювання фінансових ризиків на основі статистичних даних. Для практичної реалізації системи підтримки прийняття рішень використано сучасні інформаційні технології.

Результати. Розроблено інформаційну технологію та систему підтримки прийняття рішень як практичний інструмент для моделювання нелінійних нестаціонарних процесів у економіці та фінансах і оцінювання фінансових ризиків. Експериментальні результати обробки статистичних даних підтверджують коректність запропонованих рішень.

Висновки. Запропоновано і реалізовано системну методологію для побудови систем підтримки прийняття рішень для математичного моделювання і прогнозування сучасних економічних і фінансових процесів та оцінювання фінансових ризиків, яка базується на таких принципах системного аналізу: ієрархічність системи, ідентифікація та врахування можливих ймовірнісних і статистичних невизначеностей, адаптивність моделей до даних, генерування альтернативних рішень та використання належної критеріальної бази на всіх етапах виконання обчислень.

Ключові слова: інформаційні технології; система підтримки прийняття рішень; нелінійні нестаціонарні процеси в економіці та фінансах; невизначеності; фінансові ризики; моделювання; прогнозування.

П.И. Бидюк, Л.О. Коршевнюк, А.П. Гожий, И.А. Калинина, Т.И. Просянкина-Жарова, А.Н. Терентьев

МОДЕЛИРОВАНИЕ И ПРОГНОЗИРОВАНИЕ ФИНАНСОВЫХ И ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ С ИСПОЛЬЗОВАНИЕМ СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ

Проблематика. Современные финансовые и экономические процессы часто характеризуются сложными элементами, содержат нестационарные и нелинейные компоненты, которые требуют создания специальных математических моделей для их адекватного математического описания и прогнозирования. Для успешного решения таких задач необходимо строить соответствующие системы поддержки принятия решений с использованием принципов системного анализа.

Цель исследования. Разработать архитектуру информационных технологий системы поддержки принятия решений и ее функциональную схему для построения математических моделей экономических и финансовых процессов на основе статистических данных и оценивания финансовых рисков. Такая система должна выполнять функции обработки неопределенностей и реализоваться на основе современных информационных технологий.

Методика реализации. Для решения поставленной задачи использованы методы математического моделирования и прогнозирования финансовых и экономических процессов; методы оценивания финансовых рисков на основе статистических данных. Для практической реализации системы поддержки принятия решений использованы современные информационные технологии.

Результаты. Разработаны информационная технология и система поддержки принятия решений как практический инструмент для моделирования нелинейных нестационарных процессов в экономике и финансах и оценивания финансовых рисков. Экспериментальные результаты обработки статистических данных подтверждают корректность предложенных решений.

Выводы. Предложена и реализована системная методология построения систем поддержки принятия решений для математического моделирования и прогнозирования современных экономических и финансовых процессов, а также оценивания финансовых рисков, которая базируется на таких принципах системного анализа: иерархичность системы; идентификация и учет возможных вероятностных и статистических неопределенностей; адаптивность моделей к данным, генерирование альтернативных решений и использование соответствующей критериальной базы на всех этапах выполнения вычислений.

Ключевые слова: информационные технологии; система поддержки принятия решений; нелинейные нестационарные процессы в экономике и финансах; неопределенности; финансовые риски; моделирование; прогнозирование.

Рекомендована Радою
Інституту прикладного системного аналізу
КПІ ім. Ігоря Сікорського

Надійшла до редакції
05 червня 2019 року

Прийнята до публікації
31 жовтня 2019 року