UDC 004.942+519.816

O. Kozhukhivska[1], P. Bidyuk[2], A. Kozhukhivskyi[1]
[1]Cherkassy state technological university,
[2]Institute of applied system analysis of National technical university of Ukraine “KPI”

ADAPTIVE DECISION SUPPORT SYSTEM FOR ESTIMATING FINANCIAL RISKS

© Kozhukhivska O., Bidyuk P., Kozhukhivskyi A., 2015

A computer based decision support system is proposed the basic tasks of which are adaptive model constructing and forecasting of financial risks. The DSS development is based on the system analysis principles, i.e. the possibility for taking into consideration of some stochastic and information uncertainties, forming alternatives for models and forecasts, and tracking of the computing procedures correctness during all stages of data processing. A modular architecture is implemented that provides a possibility for the further enhancement and modification of the system functional possibilities with new forecasting and parameter estimation techniques. A high quality of final result is achieved thanks to appropriate tracking of the computing procedures at all stages of data processing during computational experiments: preliminary data processing, model constructing, and forecasts estimation. The tracking is performed with appropriate set of statistical quality parameters. Examples are given for estimation of financial credit. The examples solved show that the system developed has good perspectives for the practical use. It is supposed that the system will find its applications as an extra tool for decision making when developing the strategies for financial companies and enterprises of various types.

Key words: mathematical model, system analysis principles, adaptive forecasting, decision support system, risk estimation.

Introduction

Financial risk analysis and management is an urgent problem not only for the active financial organizations and companies but for all industrial enterprises, small and medium business, investment and insurance companies etc. Adequate models of multidimensional risks and the loss forecasts based upon them help to take into consideration a set of various influencing risk factors and make objective quality managerial decisions. There are many types of financial risks that could be described with mathematical models in the form of appropriately constructed equations or probability distributions. The market and some other types of risks are estimated with different modifications of VaR methodology that provides a possibility to reach practically acceptable quality of risk estimates [1, 2]. One of the widely spread type of risks is credit risk that arises due to failures of clients to return loans to banks. To analyze credit risks in banks the following models are used as of today: linear and nonlinear regression (logit and probit), Bayesian networks, decision trees, fuzzy logic, factor analysis, support vector machine (SVM), neural networks and neuro-fuzzy techniques, and combinations of the approaches mentioned [3 – 5].

All types of mathematical modeling usually need to cope with various kinds of uncertainties related to data, structure of the process under study and its model, parameter uncertainty, and uncertainties relevant to the models and forecasts quality. To avoid or to take into consideration the uncertainties and improve this way the quality of final result (risk values forecasts and decisions based on them) it is necessary to construct appropriate computer based systems for solving specific problems.

Selection and application of a specific model for process description and forecasts estimation depends on application area, availability of statistical data, qualification of personnel, who work on the financial analysis problems, and availability of appropriate applied software. Better results for estimation of financial processes risks is usually achieved with application of ideologically different techniques combined in the frames of one computer based system. Such approach to solving the problems of quality risk forecasts estimation can be implemented in the frames of modern decision support systems (DSS). DSS is a powerful instrument for supporting user's (managerial) decision making as far as it combines a set of appropriately selected data and expert estimates processing procedures aiming to reach final result of high quality – objective high quality alternatives for a decision making person (DMP). Development of a DSS is based on modern theories and techniques of system analysis, information processing systems, estimation and optimization theories, mathematical and statistical modeling and forecasting, decision making theory as well as many other results of theory and practice of processing data and expert estimates [6, 7].

The paper considers the problem of DSS constructing for solving the problems of modeling and estimating selected types of financial risks with the possibility for application of alternative data processing techniques, modeling and estimation of parameters and states for the processes under study.

Problem formulation. The purpose of the study is as follows: 1) analysis and development of requirements to the modern decision support systems; 2) development of the system architecture for financial risk evaluation; 3) selection of mathematical modeling and forecasting techniques for selected financial risks; 4) illustration of the system application to solving selected problem of financial risk estimation using statistical data.

Requirements to modern DSS

Modern DSS are rather complex multifunctional (possibly distributed) highly developed computing systems of informational type with hierarchical architecture that corresponds to the nature of decision making by a human. To make their performance maximum useful and convenient for users of different levels (like engineering and managerial staff) they should satisfy some general requirements. Define DSS formally as follows:

$$DSS = \{DKB, PDP, DT, MSE, MPE, RGP, DQ, MQ, REQ, AQ\},$$

where DKB is data and knowledge base; PDP is a set of procedures for preliminary data processing; DT is a set of statistical tests for determining possible effects contained in data (like integration or heteroskedasticity); MSE is a set of procedures for estimation of mathematical model structure; MPE is a set of procedures for estimation of mathematical model parameters; RGP are risk estimates generating procedures; DQ, MQ, REQ, AQ are the sets of statistical quality criteria for estimating quality of data, models, risk estimates, and decision alternatives, accordingly.

Such systems should satisfy the following general requirements: 1) – contain highly developed bases of data and knowledge with mathematical models, quality criteria for each type of computing, and model selection rules, as well as necessary computational procedures; 2) – to achieve high quality of the final result the hierarchy of a system functioning should correspond to the hierarchic process of making decision by a human; 3) – their interface should be based on the human factors principles, user friendly, convenient and simple for use, as well as adaptive to users of various levels (e.g., engineering and managerial staff); 4) – the system should possess an ability for learning in the process of its functioning, i.e. accumulate appropriate knowledge regarding possibilities of solving the problems of definite (selected) class; 5) – the organization and techniques for computing procedures should provide for appropriate rate of computing that corresponds to the human requirements with regard to the rate of alternatives generating and reaching the final result; 6) – computing (precision) quality should satisfy preliminary established requirements by a user and developer; 7) – intermediate and final results of computations should be controlled with appropriate sets of analytic quality criteria, what will allow to enhance significantly quality and reliability of the final result (decision alternatives); 8) – DSS should generate all necessary for a user forms and types of intermediate and final results representations with taking into consideration the users of various levels; 9) – the system should contain the means for exchanging with data and knowledge with other information processing systems via local and/or global computer nets; 10) – to make the system functionality flexible DSS should be easily expandable with new functions.

Satisfaction of all the requirements mentioned above provides a possibility for effective practical application of the system developed and enhancing general behavioristic effect of the DSS as a whole for a specific company or an enterprise within long periods of time [7].

Basic mathematical tools for DSS

All mathematical methods and techniques that are hired for development and implementation of DSS could be divided into two following groups: 1 – general purpose methods that provide for implementation of system functions; and 2 – special purpose methods that are necessary for solving specific problems regarding preliminary and basic data processing, model constructing, alternatives generating, selecting the best alternative for implementation and forecasting of the implementation consequences.

The group of the general purpose methods includes the following ones: – data and knowledge collecting and editing procedures; – preliminary data processing techniques such as digital filtering, normalization, imputation of missing values, detecting special effects (regime switching, seasonal effects, nonstationarity etc); – the methods for accumulating information regarding previous applications of DSS to problem solving for the retrospective use; – computer graphics techniques; – techniques for syntactic analysis to be used in a command interpreter; – methods for organizing communications with other information processing systems via local and global nets; – logical rules to control the system functioning. The set of the methods mentioned could be modified or expanded depending on specific practical application.

Selection of the application defined mathematical methods for a DSS depends on the specific system application area, possible problem statements regarding data processing, model building, processes forecasting, and alternatives generation. However, it is possible to state that in most cases of DSS development it is necessary to use the following mathematical methods: – methods and methodologies for mathematical (statistical and probabilistic) modeling using statistical/experimental data; – risk estimating and forecasting techniques on the basis of the models constructed with possibilities for combining the forecasts computed with different techniques; – operations research optimization techniques and dynamic optimization (optimal control) methods; – the methods for forecasting/foresight of decision implementation consequences; – the sets of special analytic criteria to control the processes of computations performed at each stage of data processing and alternatives generation aiming to reach high quality of a final result.

All the methods and methodologies mentioned are described well in special modern literature. For example, time series modeling and forecasting are presented in many references, more particularly in [8, 9]. The task for a DSS developer is in appropriate selection of model classes, modeling and optimization techniques, quality criteria as well as relevant methodologies for appropriate organization of computing procedures.

Coping with uncertainties

As it was mentioned above all types of mathematical modeling usually need to cope with various kinds of uncertainties linked with data, structure of the process under study and its model, parameter uncertainty, and uncertainties relevant to the models and forecasts quality. In many cases a researcher has to cope with the following types of uncertainties: structural, statistical and parametric. Structural uncertainties are encountered in the cases when structure of the process under study is unknown or not clearly enough defined. For example, when the functional approach to model constructing is applied usually we do not know object (or a process) structure, it is estimated with appropriate model structure estimation techniques: correlation analysis, estimation of mutual probabilities, lags estimation, testing for nonlinearities and nonstationarity etc. As far as we usually work with stochastic data, application of all the techniques mentioned provides a possibility for approximate estimation of an object (and its model) structure. To find "the best" model structure it is recommended to apply adaptive estimation schemes that provide automatic search in a wide range of model structure parameters (model order, time lags, and nonlinearities). Usually the search is performed in the class of regression type models with the use of integrated criterion of the following type [9]:

$$V_N(q, D_N) = e^{|1-R^2|} + \ln(1 + \frac{SSE}{N}) + e^{|2-DW|} + \ln(1 + MSE) + \ln(MAPE) + e^U , \quad (1)$$

where q is a vector of model parameters; N is a power of time series used; R^2 is a determination coefficient; DW is Durbin-Watson statistic; *MSE* is mean square error; *MAPE* is mean absolute percentage error; U is Theil coefficient. There are several possibilities for adaptive model structure estimation: (1) automatic analysis of partial autocorrelation for determining autoregression order; (2) automatic search for the exogeneous variable lag estimate (detection of leading indicators); (3) automatic analysis of residual properties; (4) analysis of data distribution type and its use for selecting correct model estimation method; (5) adaptive model parameter estimation with hiring extra data; (7) optimal selection of weighting coefficients for exponential smoothing, nearest neighbor and some other techniques; (6) the use of adaptive approach to model type selection. The use of a specific adaptation scheme depends on volume and quality of data, specific problem statement, requirements to forecast estimates, etc.

The adaptive estimation schemes also help to cope with the model parameters uncertainties. New data are used to compute model parameter estimates that correspond to possible changes in the object under study. In the cases when model can be nonlinear alternative parameter estimation techniques can be hired to compute alternative (though admissible) sets of parameters and to select the most suitable of them using statistical quality criteria.

While performing practical modeling very often we don't know statistical characteristics (covariance) of random external disturbances and measurement noise (errors). To eliminate this uncertainty optimal filtering algorithms are applied that provide for a possibility of simultaneous estimation of object (system) states and the covariance matrices. One of the possibilities hired is optimal Kalman filter. Kalman filter is used to find optimal estimates of system states on the bases of the system model represented in convenient state space form as follows:

$$\mathbf{x}(k) = \mathbf{A}(k,k-1)\mathbf{x}(k-1) + \mathbf{B}(k,k-1)\mathbf{u}(k-1) + \mathbf{w}(k), \quad (2)$$

where $\mathbf{x}(k)$ is *n*-dimensional vector of system states; $k = 0,1,2,...$ is discrete time; $\mathbf{u}(k-1)$ is *m*-dimensional vector of deterministic control variables; $\mathbf{w}(k)$ is *n*- dimensional vector of external random disturbances; $\mathbf{A}(k,k-1)$ is $(n \times n)$-matrix of system dynamics; $\mathbf{B}(k,k-1)$ is $(n \times m)$-matrix of control coefficients. The double argument $(k,k-1)$ means that the variable or parameter is used at the moment k, but its value is based on the former (earlier) data including moment ($k-1$). Usually the matrices $\mathbf{A}$ and $\mathbf{B}$ are written with one argument like $\mathbf{A}(k)$, and $\mathbf{B}(k)$, to simplify text. Obviously stationary system model is described with constant parameters like $\mathbf{A}$, and $\mathbf{B}$. As far as matrix $\mathbf{A}$ is a link between two consequent system states, it is also called state transition matrix. Discrete time k and continuous time t are linked via data sampling time T_s: $t = kT_s$. In the classic problem statement for optimal filtering the vector sequence of external disturbances $\mathbf{w}(k)$ is supposed to be zero mean white Gaussian noise with covariance matrix $\mathbf{Q}$, i.e. the noise statistics are as follows:

$$E[\mathbf{w}(k)] = 0, \quad \forall k, \qquad E[\mathbf{w}(k)\mathbf{w}^T(j)] = \mathbf{Q}(k)\delta_{kj}, \tag{3}$$

where d_{kj} is Kronecker delta-function: $\delta_{kj} = \begin{cases} 0, & k \neq j \\ 1, & k = j \end{cases}$; $\mathbf{Q}(k)$ is positively defined covariance ($n \times n$)-matrix. The diagonal elements of the matrix are variances for the components of disturbance vector $\mathbf{w}(k)$. Initial system state $\mathbf{x}_0$ is supposed to be known with the following statistics:

$$E[\mathbf{x}_0] = \bar{\mathbf{x}}_0; \quad E[\mathbf{x}_0\mathbf{x}_0^T] = \mathbf{M}; \quad E[\mathbf{w}(k)\mathbf{x}_0^T] = 0, \ \forall k.$$

The measurement equation for vector $\mathbf{z}(k)$ of output variables has the following form:

$$\mathbf{z}(k) = \mathbf{H}(k)\mathbf{x}(k) + \mathbf{v}(k), \tag{4}$$

where $\mathbf{H}(k)$ is $(r \times n)$ observation (coefficients) matrix; $\mathbf{v}(k)$ is r-dimensional vector of measurement noise with statistics:

$$E[\mathbf{v}(k)] = 0, \quad E[\mathbf{v}(k)\mathbf{v}^T(j)] = \mathbf{R}(k)\delta_{kj}, \tag{5}$$

where $\mathbf{R}(k)$ is $(r \times r)$ positively defined measurement noise covariance matrix, the diagonal elements of which represent variances of additive noise for each measured variable. The noise of measurements is also supposed to be zero mean white noise sequence that is not correlated with external disturbance $\mathbf{w}(k)$ and initial system state:

$$E[\mathbf{v}(k)\mathbf{w}^T(j)] = 0, \quad \forall\ k, j; \qquad E[\mathbf{v}(k)\mathbf{x}_0^T] = 0, \quad \forall\ k. \tag{6}$$

For system (2) – (6) with state vector $\mathbf{x}(k)$ it is necessary to find state estimate $\hat{\mathbf{x}}(k)$ at arbitrary moment k as a linear combination of estimate $\hat{\mathbf{x}}(k-1)$ at the previous moment $(k-1)$ and the last measurement available, $\mathbf{z}(k)$. The estimate of state vector $\hat{\mathbf{x}}(k)$ is computed as optimal with minimizing the expectation of the sum of squared errors, i.e.:

$$E[(\hat{\mathbf{x}}(k) - \mathbf{x}(k))^T(\hat{\mathbf{x}}(k) - \mathbf{x}(k))] = \min_K, \tag{7}$$

where $\mathbf{x}(k)$ is an exact value of state vector that can be found by deterministic part of the state equation (2); $\mathbf{K}$ is optimal matrix gain that is determined as a result of minimizing criterion (7).

Thus, the filter is constructed to compute optimal state vector $\hat{\mathbf{x}}(k)$ in conditions of influence of random external system disturbances and measurement noise. Here uncertainty arises when we don't know estimates of covariance matrices $\mathbf{Q}$ and $\mathbf{R}$ in (3) and (4), respectively. To solve the problem an adaptive Kalman filter is constructed that allows to find estimates $\hat{\mathbf{Q}}$ and $\hat{\mathbf{R}}$ together with the state vector $\hat{\mathbf{x}}(\mathrm{k})$. Another choice is in constructing separate algorithm for computing $\hat{\mathbf{Q}}$ and $\hat{\mathbf{R}}$. Other instruments to fight uncertainties are fuzzy logic, neuro-fuzzy models, Bayesian networks and appropriate types of distributions.

Other statistical data uncertainties such as skipped measurements, extreme values and high level jumps of unknown origin could be processed with appropriately selected statistical procedures. There exist a number of data imputation procedures that help to complete the data collected. For example, very often skipped measurements for time series can be generated with appropriately selected distributions.

Appropriate processing of jumps and extreme values helps with adjusting data stationarity and to estimate correctly probability distribution.

Generation and implementation of alternatives with DSS

Decision making process includes rather sophisticated procedures that could be partially or completely iterative, i.e. executed repeatedly when the alternative found is not satisfactory for a decision making person (DMP). DSS can return automatically (or on DMP initiative) to the previous stages of data and knowledge analysis.

The whole process of making and implementing decision could be considered as consisting of the stages given below.

1 – A thorough analysis of the decision problem using all available sources of information, collection of data and knowledge relevant to the problem. At this stage it is also important to consider and use former solutions to the problem if such are available. The information regarding former solutions of similar problem can be helpful for correcting problem statement, to select appropriate techniques for data analysis, to speed up alternative generation, and to decline the alternatives that turned out to be ineffective in the past.

2 – Selection of a class (classes) of mathematical models for the problem description, and analysis of a possibility for the use of available (previously developed) models. The models could belong to different classes as far as they can be formulated in continuous or discrete time, be linear or nonlinear, they can be developed according to the structural or functional approach etc. In some cases it is necessary to construct complex simulative model that would include a set of simpler models of different classes.

3 – Development of new models for the problem (process, object, system) under study what includes structure and parameter estimation for candidate models using available data (and possibly expert estimates) and knowledge of various types. The alternative structures of candidate models provide a possibility for selecting the best one of them for generating alternative decisions (loss estimates, forecasts, control actions, risk estimates etc) on their bases.

4 – Analysis of the candidate models constructed and selecting of the best one of them with application of a set of statistical quality criteria and expert estimates. At this stage again more than one model can be selected for the further use as far as the best model (for a particular application) can be found only after application of the candidates for solving particular problem, i.e. after alternatives generating and estimating possible consequences of their implementation.

5 – Application of the model (models) selected to solving risk estimation and/or control problem (when necessary). If the forecasts or controls computed are not satisfactory we should return back to stage one or stage three, and repeat the process of model constructing. At this stage another set of statistical quality criteria should be applied to the analysis of risk estimates, forecasts or controls determined.

6 – Generating of a set of alternatives with the use of the model (models) constructed and various admissible initial conditions and constraints on variables. In a case of controls generating the alternatives could be built with different optimality criteria, utility functions or other criteria.

7 – Analysis of the alternatives generated with the experts of an enterprise or a company, and final selection of the best one for practical implementation. In a case when no alternative is acceptable we should return back to the model constructing or alternative generating stages. New knowledge or data can be required for the next iteration of computing new decision alternatives.

8 – Planning of actions and estimation of financial, material and human resources that are necessary for implementation of the alternative selected. Determining of a time horizon (horizon of control) necessary for implementing the decision made.

9 – Implementation of the decision made: current monitoring of availability and spending the necessary resources, estimation of necessary time frames, registering and quality estimation of intermediate and final results.

10 – Application of possible analytic and expert quality criteria to estimation of final results.

11 – Analysis of the final results by the company experts, and final elucidation of advantages and disadvantages of the alternative implemented; analysis of the decision making and implementing process, and forming forecasts (foresights) for the future.

12 – Writing the final report on the tasks performed.

Architecture of DSS for estimation of financial risks

DSS architecture is a generalized large-scale representation of basic system elements with links between them. Architecture gives a notion for the general purpose of system constructing and its basic functions (Fig. 1).

DSS functionality is controlled by user commands, correctness of which is monitored by the command interpreter which constitutes a part of user interface. The user commands are implemented by the main operation module that coordinates functioning of all system elements. Specific commands and actions can be as follows: expanding and modification of bases available in the system; initiation and starting of data and knowledge processing procedures; model constructing, risks and forecasts estimation, alternative generating; viewing intermediate and final results of computing; retrospective analysis of previous results of decision making; comparing of current results with the previous ones.

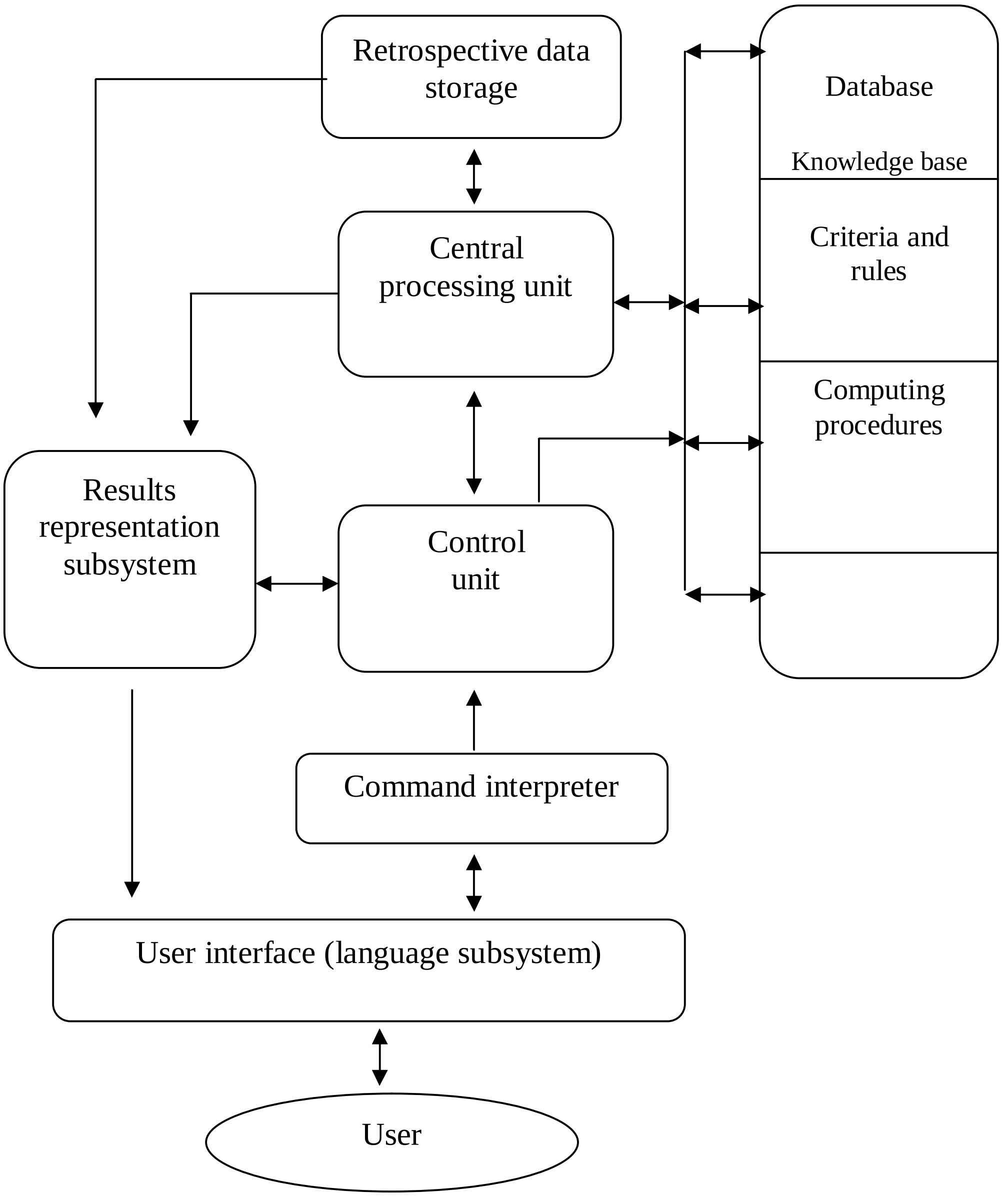

Fig. 1. DSS architecture for estimation of financial risks

The system interface is considered as its most important element from the point of view of its presentation to user. This is justified by the fact that interface construction influences substantially convenience as well as rate and effectiveness of user interaction with the system. The principles of interface constructing and its implementation create a separate special task that is not considered here.

It is clear that architecture, given in Fig. 1, is highly generalized. It means that practically the same type of architecture could be used to construct DSS for solving rather wide class of problems that require statistical/experimental data processing, mathematical modeling, optimal state and parameter estimation for dynamic systems, possible loss estimation, forecasting the future process evolution and making decisions on this basis.

Data, model and forecasts quality criteria

To achieve reliable high quality final result of risk estimation and forecasting at each stage of computational hierarchy separate sets of statistical quality criteria have been used. Data quality control is performed with the following criteria:

- database analysis for missing values using developed logical rules, and imputation of missed values with appropriate techniques;
- analysis of data for availability of outliers with special statistical tests, and processing of outliers to reduce their negative influence on statistical properties of data;
- normalizing of data in a case of necessity;
- application of low-order digital filters (usually that's low-pass filters) for separation of observations from measurement noise;
- application of principal component method to achieve desirable level of orthogonalization between the variables selected;
- computing of extra indicators for the use in regression and other models.

It is also useful to test how informative is the data collected. Very formal indicator for data being informative is its sample variance. It is considered formally that the higher is the variance the richer is the data with information. Another criterion is based on computing derivatives with a polynomial that describes data in the form of a time series. For examples, such polynomial may describe rather complex process trend as follows:

$$y(k)=a_0+\sum_{i=1}^{p}a_i y(k-i)+c_1 k+c_2 k^2+...+c_m k^m+e(k), \qquad (8)$$

where $y(k)$ is basic dependent variable; a_i, c_i are model parameters; $k=0,1,2,...$ is discrete time; $e(k)$ is a random process that integrates the influence of external disturbances to the process being modeled as well as model structure and parameters errors. Autoregressive part of model (1) describes the deviations that are imposed on a trend, and the trend itself is described with the *m*-th order polynomial. In this case maximum number of derivatives can be *m*, though in practice actual number of derivatives is defined by the largest number i of parameter c_i, that is statistically significant. To select the best models constructed the following statistical criteria are used: determination coefficient (R^2); Durbin-Watson statistic (DW); Fisher *F*-statistic; Akaike information criterion (*AIC*), and residual sum of squares (*SSE*). The forecasts quality is estimated with hiring the following criteria: mean squared error (*MSE*); mean absolute percentage error (*MAPE*); and Theil inequality coefficient (U). To perform automatic model selection the above mentioned combined criterion (1) could be hired. The power of the criterion was tested experimentally and proved with a wide set of models and statistical data. Thus, the three sets of quality criteria are used to insure high quality of final result.

To analyze the quality of credit borrowers classification model the following quality criteria were used: common accuracy, errors of type I and type II, ROC-curve, and Gini index. Common accuracy is computed as follows [3]:

$$CA = \frac{Correct\ Forecast}{N},$$

where $Correct\ Forecast$ is a number of correctly forecasted cases; N is general number of cases (clients) considered. To some extent this criteria is subjective because it depends on a number of defaults as well as on the cut-off threshold value. ROC-curve (Receiver Operation Characteristic) shows relation between the number of correctly classified positive cases (positives) and the number of incorrectly classified negative cases (negatives). The first ones are called true positive set, and the second one – negative set (specificity). Obviously, the cut-off threshold value also influences the errors of type I and type II. Among other criteria are the following: True Positives Rate (TPR), False Positives Rate (FPR), sensitivity (Se), specificity (Sp), and Gini index. The last one is determined by the area under ROC-curve [10]. Table 1 shows relation between area under curve (AUC) and Gini index.

Table 1

Relation between AUC and GINI index

AUC interval	Gini index	Model quality
0,9 – 1,0	0,8 – 1,0	Excellent
0,8 – 0,9	0,6 – 0,8	Very high
0,7 – 0,8	0,4 – 0,6	Acceptable
0,6 – 0,7	0,2 – 0,4	Medium
0,5 – 0,6	0 – 0,2	Unacceptable

The ROC-curve can be used to find optimum cut-off value as a compromise between sensitivity and specificity of a model. The following criteria can be used for cut-off value selection: 1 – the requirement of minimum sensitivity, Se, (or specificity, Sp); 2 – the requirement of maximum total sensitivity and specificity of a model, $cutoff = \max_k (Se_k + Sp_k)$, where $k = 1, 2, 3, ...$ is a number of client; the requirement of a balance between sensitivity and specificity, i.e. when $Se \approx Se$: $cutoff = \min_k |Se_k - Sp_k|$.

Some mathematical models used in DSS

When considering mathematical models it is important to use a unified notion of model structure which we define as follows:

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

where r is model dimensionality (number of equations); p is model order (maximum order of differential or difference equation in a model); m is a number of independent variables in the right hand side; n is a nonlinearity and its type; d is a lag or output reaction delay time; z is external disturbance and its type; l are possible restrictions for the variables and/or parameters.

Generalized linear models (GLM). GLM can be considered as further enhancement of multiple linear regression (MLR) model. It is distinguished from MLR with the following features: – distribution of dependent variable can be non-Gaussian and not necessarily continuous, say binomial; – predicted values of dependent variable are computed as linear combination of predictors that are linked to dependent variable via selected link function. Generally, GLM create a class of statistical models that includes linear regression, variance analysis relations, nonlinear models like logit and probit, Poisson regression and some others [11]. In a general linear model independent variable is supposed to be normally distributed and the link function is called identity function, i.e. linear combination of independent variables is not subjected to any transform. Thus, GLM is a model of the following type:

$$y = g^{-1}\left(\sum_{i=1}^{m} \mathrm{b}_i\, g_i(x) \right),$$

where m is a number of independent (explaining) variables; $g(\cdot)$ is a link function. It is usually supposed that dependent variable y belongs to the class of exponential distributions. Thus, characteristics of GLM suppose the knowledge of dependent variable distribution, characteristics and parameters of the link function $g(\cdot)$, and of linear predictor $\mathbf{X\,b}$, where $\mathbf{X}$ is a measurement matrix for independent variables; $\mathbf{b}$ is parameter vector. The class of exponential distributions includes the following distribution types: normal, gamma, and beta, and the discrete families – binomial, Poisson, and negative binomial. General representation of PDFs or PMFs for them is as follows:

$$f(x|q) = h(x)\,c(q)\exp\left(\sum_{i=1}^{k} w_i(q)\,l_i(x)\right),$$

where $h(x) \geq 0$ and $l_1(x),...,l_k(x)$ are real-valued functions of the observation x (they cannot depend on q); $c(q) \geq 0$ and $w_1(x),...,w_k(x)$ are real-valued functions of the possibly vector-valued parameter q (they cannot depend on x).

Nonlinear models logit and probit. To solve the problem of classifying credit borrowers into two groups it is quite logically to use appropriately transformed CDF. CDF belongs to the class of monotonous functions that monotonously decrease or increase on some interval. Suppose that for determining probability of crediting a client p_c it is chosen a normal distribution:

$$p_c = \Phi(\mathrm{b}^T\mathrm{x}) = \int_{-\infty}^{u} j\,(z)\,d\,z\,,$$

where $\varphi(z)$ is a density for standard normal distribution; $u=\mathbf{b}^T\mathbf{x}$ is upper integration limit. This way so called probit model is constructed.

If the probability for successful crediting is determined with logistic distribution function then logit model is constructed. In this case we have:

$$p_c = \Phi(\mathbf{b}^T\mathbf{x}) = \int_{-\infty}^{u} \varphi(z)\,d\,z = \frac{1}{1+\exp(-\mathbf{b}^T\mathbf{x})}, \tag{9}$$

or

$$p_c = \frac{\exp(b_1x_1 + ... + b_mx_m)}{1+\exp(b_1x_1 + ... + b_mx_m)}.$$

In contrast to the normal distribution logistic function has so called closed form that provides a possibility for simplified computations in comparison to probit. Parameter estimates for both models can be found with maximum likelihood technique. An alternative possibility is Markov chain Monte Carlo (MCMC) approach that is based on correct generation of pseudorandom sequences that satisfy certain conditions. Due to availability of multiple alternative techniques for generating pseudorandom sequences MCMC has found wide applications [12]. Classification results achieved with logit and probit are usually acceptable in most cases of application.

Bayesian networks (BN). Bayesian networks are probabilistic and statistical models represented in the form of directed acyclic graphs (DAG) with vertices as variables of an object (system) under study, and arcs showing existing causal relations between the variables. Each variable of BN is characterized with complete finite set of mutually excluding states. The relations between the variables are established via expert estimates or applying special statistical and probabilistic tests to statistical data (when available) characterizing variables dynamics. The process of constructing BN is generally the same as for models of other types, say regression models. For example, as model parameters for BN are unconditional and conditional probabilities for specific values of variables, that are stored in respective tables. For parent variables these are unconditional probabilities and for daughter variables – conditional probability tables (CPT). Unconditional and conditional probabilities are determined by experts (in simpler cases), and by special computational algorithms when appropriate sets of statistical (or experimental) data are available. Thus to each node of DAG is assigned CPT that is used for computing probabilistic inference over the BN [13, 14].

The process of constructing a model in the form of BN can be represented with the following steps: 1) – a thorough analysis of the process (object) under study aiming to detecting of its special functioning features and identification of parent and daughter variables; 2) – search and analysis of existing process models and determining the possibility of their usage in DSS; 3) – determining degree of relations between the process variables using special tests and expert estimates; 4) – reduction of the process dimensionality whenever this is possible; 5) – scaling and discretization of the data available when necessary; 6) – determining semantic restrictions on the future model; 7) – estimation of candidate model (directed acyclic graphs) structures using appropriate optimization procedures and score functions; 8) – candidate models analysis and selection of the best one using model quality criteria (including values of score functions); 9) – application of the model(s) constructed to solve the problem stated; 10) – computing inference with the model(s) constructed with regards to the variables selected, quality analysis of the result. In our case the final result of the model application is computing of client default probability with the conditions established by other model variables. According to alternative problem statement BM can be constructed for estimation of operational or other type of financial risks.

Example of DSS application

In this example we used the database consisting of 4700 records that was divided into learning sample (4300 records), and test sample (400 records). The default probabilities were computed and compared to actual data, also errors of the first and second type were computed using different values of cut-off value. It was established for Bayesian network that maximum model accuracy reached was 0.764 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 value. 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 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 2.

Table 2

Quality of the models constructed

Model type	Gini index	AUC	Common accuracy	Model quality
Bayesian network	**0.689**	0.845	0.764	Very high
Logistic regression	0.678	0.847	**0.798**	Very high
Decision tree	0.583	0.791	0.763	Acceptable
Linear regression	0.386	0.647	0.616	Unacceptable

It follows from the table 1 that the best models for estimation of credit return probability turned out to be logistic regression and Bayesian network. The best common accuracy showed logistic regression (0.798) though Bayesian network showed higher Gini index (0.689). The decision tree used is characterized by Gini index of about 0.583, and CA = 0.763. It should be stressed that acceptable values of Gini index for developing countries like Ukraine are in the range 0.4 – 0.6. Bayesian network constructed and nonlinear regression showed rather high values of Gini index that are acceptable for the Ukrainian economy in transition.

The results of computing experiments lead to the conclusion that today scoring models and Bayesian networks are the best instruments for banking system due to the fact that BN provide a possibility for detecting "bad" clients and to reduce financial risks caused by the clients. It also should be stressed that DSS constructed is 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 with a set of appropriate criteria. The system performs tracking of the whole computational process using separate sets of statistical quality criteria at each stage of decision making: quality of data, models and forecasts (or risk estimates).

Conclusions

The methodology was proposed for constructing DSS for mathematical modeling of economic and financial processes, and credit risk estimation that is based on the following system analysis principles: hierarchical system structure, taking into consideration of probabilistic and statistical uncertainties, features of adaptation, generating of multiple decision alternatives, and tracking of computational processes at all the stages of data processing with appropriate sets of statistical quality criteria.

The system proposed 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 alternative generation. 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) as well as thanks to convenient for a user intermediate and final results representation. The system is based on the ideologically different techniques of modeling and risk forecasting what creates a good base for combination of various approaches to achieve the best results. The examples of the system application show that it can be used successfully for solving practical problems of risk estimation. The results of computing experiments lead to the conclusion that today scoring models and 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 with a set of appropriate criteria. The system performs tracking of the whole computational process using separate sets of statistical quality criteria at each stage of decision making: quality of data, models and forecasts or risk estimates.

The DSS can be used for support of decision making in various areas of human activities including strategy development for banking system and industrial enterprises, investment companies etc. Further extension of the system functions is planned with new forecasting techniques based on probabilistic technologies and fuzzy sets.

1. McNeil A. J. Quantitative Risk Management / A. J. McNeil, R. Frey, P. Embrechts. – Princeton (New Jersey): Princeton University Press, 2005. – 538 p. 2. International Convergence of Capital Measurement and Capital Standards. A Revised Framework. Comprehensive Version. – Basel Committee on Banking Supervision, Bank for International Settlements. – Basel, 2006. – 158 p. 3. Mays E. (Ed.) Handbook of Credit Scoring / E. (Ed.) Mays. – Chicago: Glenlake Publishing Company, Ltd., 2001. – 460 p. 4. Neil M. Using Bayesian networks to model expected and unexpected operational losses / Neil M., Fenton N. E., Tailor M. // Risk Analysis. – 2005. – P. 34-57. 5. Shakhov V. V. Theory and management of insurance riks / Shakhov V. V., Medvedev V. G., Millerman A. S. – M.: Finances and Statistics, 2002. – 224 p. 6. Burstein F. Handbook of Decision Support Systems / F. Burstein, C. W. Holsapple. – Berlin:

Springer-Verlag, 2008. – 908 p. 7. Hollsapple C. W. Decision support systems / C. W. Hollsapple, A. B. Winston. – Saint Paul: West Publishing Company, 1996. – 860 p. 8. Tsay R. S. Analysis of financial time series / R. S. Tsay. – Hoboken: Wiley & Sons, Inc., 2010. – 715 p. 9. Bidyuk P.I Methods of Forecasting / P. I. Bidyuk, O. S. Menyailenko, O. V. Polovcev. – Lugansk: Alma Mater, 2008. – 608 p. 10. Bidyuk P. I. Models for credit risk estimation / P. I. Bidyuk, Ye. O. Matros // Cybernetics and Computing Techniques, 2008. – Vol. 153. – P. 87–95. 11. De Jong P. Generalized Linear Models for Insurance Data / P. De Jong, G. Z. Heller. – New York: Cambridge University Press, 2008. – 197 p. 12. Gilks W. R. Markov chain Monte Carlo in practice /W. R. Gilks, S. Richardson, D. J. Spiegelhalter. – New York: Chapman & Hall/CRC, 2000. – 486 p. 13. Jensen F. V. Bayesian Networks and Decision Graphs / F. V. Jensen, Th. D. Nielsen. – New York: Springer, 2007. – 457 p. 14. Zgurovsky M. Z. Method of constructing Bayesian networks based on scoring functions / M. Z. Zgurovsky, P. I. Bidyuk, O. M. Terentyev // Cybernetics and System Analysis, 2008. – Vol. 44.– No. 2.– P. 219-224.

УДК 004.4, УДК 621.317.089.6

О. Олеськів, І. Микитин
Національний університет “Львівська політехніка”,
кафедра інформаційно-вимірювальних технологій

МЕТРОЛОГІЧНА ПЕРЕВІРКА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ЗАСОБІВ ВИМІРЮВАННЯ З РІЗНИМИ СТРУКТУРАМИ

© Олеськів О., Микитин І., 2015

Розглянуто класифікацію програмного забезпечення засобів вимірювання. Проведено класифікацію структур засобів вимірювання. Проаналізовано можливість проведення метрологічної перевірки певного типу програмного забезпечення засобів вимірювання відповідно до методів метрологічної перевірки програмного забезпечення.

Ключові слова: засіб вимірювання, програмне забезпечення, метрологічна перевірка, структурна схема, вбудована система керування, кіберфізична система.

We consider the classification of software measuring instruments. Classification of their structures is performed. The possibility of verification a certain type of the metrological software for measuring instruments is analyzed.

Key words: measuring instrument, software, verification, block diagram, embedded system control, cyber-physical system.

Вступ

Застосування програмного забезпечення (ПЗ) та мікроконтролерів дозволило зменшити аналогову та цифрову частину засобів вимірювання (ЗВ). Основне опрацювання результатів вимірювання, а саме усереднення, апроксимація, фільтрація, інтерполяція, перетворення Фур’є тощо, реалізуються переважно програмним способом. Некоректна програмна реалізація алгоритмів розрахунку та опрацювання результатів вимірювання, невідповідність ПЗ вимірювальній задачі приладу, випадкова або навмисна зміна функцій ПЗ можуть призвести до виникнення додаткової похибки вимірювання. Тому доцільно здійснювати перевірку програмного забезпечення засобів вимірювання для визначення його впливу на метрологічні характеристики ЗВ.

Структура засобу вимірювання може впливати на процес, а інколи і на можливість проведення метрологічної перевірки ПЗ. Не завжди є доступ до ПЗ, що ускладнює, а в деяких випадках і унеможливлює перевірку ПЗ. З іншого боку, відсутність доступу до ПЗ покращує його захищеність, оскільки немає можливості його навмисної зміни або пошкодження. Якщо доступ до