Analysis of Uncertainty Types for Model Building and Forecasting Dynamic Processes

Peter Bidyuk[1(✉)], Aleksandr Gozhyj[2], Irina Kalinina[2], and Victor Gozhyj[2]

[1] National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnik Institute", Kiev, Ukraine
pbidyke@gmail.com
[2] Black Sea National University named after Petro Mohyla, Nikolaev, Ukraine
alex.gozhyj@gmail.com, irina.kalininal612@gmail.com, gozhyi.v@gmail.com

Abstract. The article deals with the issues of handling uncertainties in the problems of modeling and forecasting dynamic systems within the framework of the dynamic planning methodology. To analyze and take into account possible structural, statistical and parametric uncertainties, the Kalman filter, various methods for calculating missing data, numerous methods for estimating the model parameters and the Bayesian approach to programming are used. The questions of an estimation of quality of predicted decisions are considered.

Keywords: Dynamic planning · Stochastic uncertainty · Kalman filter
Uncertainties of model parameters · Uncertainties of a level (amplitude)
Probabilistic uncertainties · Bayesian networks

1 Introduction

Analysis of dynamic processes in the planning and decision-making procedures is an urgent problem not only for financial organizations and companies but for all industrial enterprises, small and medium business, investment and insurance companies etc. This problem is solved by using dynamic programming methodology. Dynamic planning (DP) could be defined as the process of estimation by an enterprise of its current state on the market in comparison with other competing enterprises, and determining the further goals as well as sequences of actions and resources that are necessary for reaching the goals stated. This process of planning is performed continuously (or quasi-continuously) with acquiring new information (knowledge) about market, technologies, forecast estimates of necessary variables and situations. All this knowledge is used for correcting of actions and activities of the enterprise and supporting its competitiveness with flow of time.

Formally DP could be represented in the form:

$$DP = \{\mathbf{X}_0, \mathbf{G}, \mathbf{R}, \mathbf{D}(t), \mathbf{K}, \mathbf{T}, \mathbf{F}, \Delta\mathbf{D}(t), \Delta\mathbf{R}(t)\},$$

Where $\mathbf{X}_0$ is initial state of an enterprise; $\mathbf{G}$ are the goals stated by the enterprise management; $\mathbf{R}$ are resources that are necessary for reaching the goals stated. $\mathbf{D}(t)$ is a

N. Shakhovska and V. Stepashko (eds.), *Advances in Intelligent Systems and Computing II*, Advances in Intelligent Systems and Computing 689,
https://doi.org/10.1007/978-3-319-70581-1_5

sequence of actions that should be performed on the interval of planning; **K** is a new knowledge about environment; **T** are new technologies. **F** are results of forecasting and foresight; $\Delta\mathbf{D}(t)$ are corrections that are to be performed for reaching the goals; $\Delta\mathbf{R}(t)$ are necessary extra resources. One of the main problems that are to be solved within the DP paradigm is high quality forecasting of relevant processes.

Adequate models of the process and the forecasts generated with them help to take into consideration a set of various influencing factors and make objective planning managerial decisions. Another purpose of the studies is in estimating possible risks using forecasts of volatility. There are several types of processes that could be described with mathematical models in the form of appropriately constructed equations or probability distributions. Among them are the processes with deterministic and stochastic trends, and heteroscedastic processes. As of today the following mathematical models are widely used for describing nonlinear dynamics of processes relevant to planning: linear and nonlinear regression (logit and probit, polynomials, splines), autoregressive integrated moving average (ARIMA) models, autoregressive conditionally heteroscedastic models (ARCH), generalized ARCH (GARCH), dynamic Bayesian networks, support vector machine (SVM) approach, neural networks and neuro-fuzzy techniques as well as combinations of the approaches mentioned [1–5].

All types of mathematical modeling usually need to cope with various kinds of uncertainties related to statistical data, structure of the process under study and its model, parameter uncertainty, and uncertainties relevant to the models and forecasts quality. Reasoning and decision making are very often performed with leaving many facts unknown or rather vaguely represented in processing of data and expert estimates. To avoid or to take into consideration the uncertainties and improve this way quality of the final result (processes forecasts and the planning decisions based upon them) it is necessary to construct appropriate computer based decision support systems (DSS) for solving multiple 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 data analysis problems, and availability of appropriate applied software. Better results for estimation of processes forecasts is usually achieved with application of ideologically different techniques combined in the frames of one computer system. Such approach to solving the problems of quality forecasts estimation can be implemented in the frames of modern decision support systems (DSS). DSS today 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 theory and techniques of system analysis, data 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 [6–8].

The paper considers the problem analysis, accounting and handling of uncertainties for solving the problems of modeling and estimating forecasts for selected types of dynamic processes with the possibility for application of alternative data processing techniques, modeling and estimation of parameters and states for the processes under study.

2 Problem Formulation

The purpose of the study is as follows: (1) analysis of uncertainty types characteristic for model building and forecasting dynamic processes; (2) selection of techniques for taking into consideration of the uncertainties; (3) selection of mathematical modeling and forecasting techniques for nonstationary heteroscedastic processes.

3 Coping with Uncertainties

All types of mathematical modeling with the use of statistical experimental data usually need to consider various kinds of uncertainties caused by data, informational structure of a process under study and its model, parameter uncertainty, and uncertainties relevant to the quality of models and forecasts. In many cases a researcher has to cope with the following basic types of uncertainties: structural, statistical and parametric. Structural uncertainties are encountered in the cases when structure of the process under study (and respectively its model) is unknown or not clearly enough defined (known partially). 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 information, estimation of lags, testing for nonlinearity and nonstationarity, identification of external disturbances etc. Uncertainty could also be introduced by an expert who is studying the process and provides its estimates for model structure, parameter restrictions, selection of computational procedures etc. The sequence of actions necessary for identification, processing and taking into consideration of uncertainties could be formulated as follows: – identification and reduction of data uncertainty; – model structure and parameters estimation; – reduction of uncertainties related to the model structure and parameters estimation; – reduction of uncertainties relevant to expert estimates; – estimation of forecasts and reduction of respective uncertainties; – selection of the best final result. All the tasks mentioned above are usually solved sequentially (in an adaptive loop) with appropriately designed and implemented DSS.

We consider uncertainties as the factors that influence negatively the whole process of mathematical model constructing, forecasts and possible risk estimating and generating of alternative decisions. They are inherent to the process being studied due to incomplete or noise corrupted data, complex stochastic external influences, incompleteness or inexactness of our knowledge regarding the objects (systems) structure, incorrect application of computational procedures etc. The uncertainties very often appear due to incompleteness of data, noisy measurements or they are invoked by sophisticated stochastic external disturbances with complex unknown probability distributions, poor estimates of model structure or by a wrong selection of parameter estimation procedure. The problem of uncertainties identification is solved with application of special statistical tests and visual studying of available data.

As far as we usually work with stochastic data, correct application of existing statistical techniques provides a possibility for approximate estimation of a system (and its model) structure. To find “the best” model structure it is recommended to apply adaptive estimation schemes that provide automatic search in a pre-defined range of

model structures and parameters (model order, time lags, and possible nonlinearities). It is often possible to perform the search in the class of regression type models with the use of information criterion of the following type [2]:

$$N \log (FPE) = N \log (V_N(\hat{\theta})) + N \log \left(\frac{N+p}{N-p} \right), \quad (1)$$

where $\hat{\theta}$ is a vector of model parameters estimates; N is a power of time series used; FPE is final prediction error term; $V_N(\hat{\theta})$ is determined by the sum of squared errors; p is a number of model parameters. The value of the criteria (1) is asymptotically equivalent to the Akaike information criterion with $N \to \infty$. As the amount of data, N, may be limited, then an alternative, the minimum description length (MDL) criterion

$$MDL = \log (V_N(\hat{\theta})) + p \frac{\log (N)}{N}$$

could be hired to find the model that adequately represents available data with the minimum amount of available information.

There are several possibilities for adaptive model structure estimation: (1) application of statistical criteria for detecting possible nonlinearities and the type of nonstationarity (integrated or heteroskedastic process); (2) analysis of partial autocorrelation for determining autoregression order; (3) automatic estimation of the exogeneous variable lag (detection of leading indicators); (4) automatic analysis of residual properties; (5) analysis of data distribution type and its use for selecting correct model estimation method; (6) adaptive model parameter estimation with hiring extra data; (7) optimal selection of weighting coefficients for exponential smoothing, nearest neighbor and other techniques. The development and 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 re-compute model parameter estimates that correspond to possible changes in the object under study. In the cases when model is nonlinear alternative parameter estimation techniques (say, MCMC) could be hired to compute alternative (though admissible) sets of parameters and to select the most suitable of them using statistical quality criteria.

3.1 Processing Some Types of Stochastic Uncertainties

While performing practical modeling very often statistical characteristics (covariance matrix) of stochastic external disturbances and measurement noise (errors) are unknown. To eliminate this uncertainty optimal filtering algorithms are usually applied that provide for a possibility of simultaneous estimation of object (system) states and the covariance matrices. One of the possibilities to solve the problem is optimal Kalman filter. Kalman filter is used to find optimal estimates of system states on the bases of the system model represented in a convenient state space form as follows:

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

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; $\boldsymbol{A}(k, k-1)$ is $(n \times m)$ - matrix of system dynamics; is $\boldsymbol{B}(k, k-1)$ $(n \times m)$ 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 processing including moment $(k-1)$. Usually the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are written with one argument like, $\boldsymbol{A}(k)$ and $\boldsymbol{B}(k)$, to simplify the text. Besides the main task, optimal state estimation, Kalman filter can be used to solve the following problems: computing of short-term forecasts, estimation of unknown model parameters including statistics of external disturbances and measurement errors (adaptive extended Kalman filter), estimation of state vector components that cannot be measured directly, and fusion of data coming from various external sources.

Obviously stationary system model is described with constant parameters like $\boldsymbol{A}$, and $\boldsymbol{B}$. As far as matrix $\boldsymbol{A}$ is a link between two consequent system states, it is also called state transition matrix. Discrete time k and continuous time t are linked to each other via data sampling time: $\boldsymbol{T}_s : t = k\,\boldsymbol{T}_s$. 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:

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

where δ_{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 and the measurement equation for vector $\mathbf{z}(k)$ of output variables is described by the equation:

$$\mathbf{z}(k) = \mathbf{H}(k)\mathbf{x}(k) + \mathbf{v}(k), \tag{3}$$

where $\mathbf{H}(k)$ is $(r \times n)$ observation (coefficients) matrix; $\mathbf{v}(k)$ is r-dimensional vector of measurement noise with statistics: $E[\mathbf{v}(k)] = 0, E[\mathbf{v}(k)\mathbf{v}^T(j)] = \mathbf{R}(k)\delta_{kj}$,

where $R(k)$ is $(r \times r)$ positively defined measurement noise covariance matrix, the diagonal elements of which represent variances of additive noise for each measurable 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. For the system (2), (3) with state vector $\mathbf{x}(k)$ it is necessary to find optimal 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 one with minimizing the expectation of the sum of squared errors, i.e.:

$$E[(\hat{\mathbf{x}}(k) - \mathbf{x}(k))^T(\hat{\mathbf{x}}(k) - \mathbf{x}(k))] = \min_K, \tag{4}$$

where $\mathbf{x}(k)$ is an exact value of state vector that can be found as deterministic part of the state Eq. (2); $\boldsymbol{K}$ is optimal matrix gain that is determined as a result of minimizing quadratic criterion (4).

Thus, the filter is constructed to compute optimal state vector $\hat{\mathbf{x}}(k)$ in conditions of influence of external random system disturbances and measurement noise. Here one of possible uncertainties arises when we don't know estimates of covariance matrices $\mathbf{Q}$ and $\mathbf{R}$. To solve the problem an adaptive Kalman filter is to be constructed that allows to compute estimates of $\hat{\mathbf{Q}}$ and $\hat{\mathbf{R}}$ simultaneously with the state vector $\hat{\mathbf{x}}(k)$. Another choice is in constructing separate algorithm for computing the values of $\hat{\mathbf{Q}}$ and $\hat{\mathbf{R}}$. A convenient statistical algorithm for estimating the covariance matrices was proposed [11]:

$$\hat{\mathbf{R}} = \frac{1}{2}\left[\hat{\mathbf{B}}_1 + \mathbf{A}^{-1}(\hat{\mathbf{B}}_1 - \hat{\mathbf{B}}_2)(\mathbf{A}^{-1})^T\right],$$
$$\hat{\mathbf{Q}} = \hat{\mathbf{B}}_1 - \hat{\mathbf{R}} - \mathbf{A}\hat{\mathbf{R}}\mathbf{A}^T,$$

where $\hat{\mathbf{B}}_1 = E\left\{[\mathbf{z}(k) - \mathbf{A}\,\mathbf{z}(k-1)]\,[\mathbf{z}(k) - \mathbf{A}\,\mathbf{z}(k-1)]^T\right\}$; $\hat{\mathbf{B}}_2 = E\left\{[\mathbf{z}(k) - \mathbf{A}^2\,\mathbf{z}(k-2)]\,[\mathbf{z}(k) - \mathbf{A}^2\,\mathbf{z}(k-2)]^T\right\}$.

The matrices $\hat{\mathbf{Q}}$ and $\hat{\mathbf{R}}$ are used in the optimal filtering procedure as follows:

$$\mathbf{S}(k) = \mathbf{A}\mathbf{P}(k-1)\mathbf{A}^T + \hat{\mathbf{Q}};$$
$$\Delta(k) = \mathbf{S}(k)[\mathbf{S}(k) + \hat{\mathbf{R}}]^{\#};$$
$$\mathbf{P}(k) = [\mathbf{I} - \Delta(k)]\mathbf{S}(k), \quad k = 0, 1, 2, \cdots,$$

where $\mathbf{S}(k)$ and $\mathbf{P}(k)$ are prior and posterior covariance matrices of estimates errors respectively; the symbol "#" denotes pseudo-inverse; $\mathbf{A}^T$ means matrix transposition; $\Delta(k)$ is a matrix of intermediate covariance results. The algorithm was successfully applied to the covariance estimating in many practical applications. The computation experiments showed that the values of $\Delta(k)$ become stationary after about 20–25 periods of time (sampling periods) in a scalar case, though this figure is growing substantially with the growth of dimensionality of the system under study. It was also determined that the parameter estimators are very sensitive to the initial conditions of the system. The initial conditions should differ from zero enough to provide stability for the estimates generated.

Other appropriate instruments for taking into consideration the uncertainties are fuzzy logic, neuro-fuzzy models, Bayesian networks, appropriate types of distributions etc. Some of statistical data uncertainties such as missing measurements, extreme values and high level jumps of stochastic origin could be processed with appropriately selected statistical procedures. There exists a number of data imputation schemes that help to complete the sets of the data collected. For example, very often missing measurements for time series could be generated with appropriately selected distributions or in the form of short term forecasts. Appropriate processing of jumps and

extreme values helps with adjusting data nonstationarity and to estimate correctly the probability distribution for the stochastic processes under study.

3.2 Processing Data with Missing Observations (Data Are in the Form of Time Series)

As of today for the data in the time series form the most suitable imputation techniques are as follows: 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 estimates from distributions the form and parameters of which are again determined using available part of data and expert estimates; the use of optimization techniques, say appropriate forms of EM-algorithms (expectation maximization); exponential smoothing etc. It should also be mentioned that optimal Kalman filter can also be used for imputation of missing data because it contains "internal" forecasting function that provides a possibility for generating quality short-term forecasts [12]. Besides, it has a feature of fusion the data coming from various external sources and improving this way the quality of state vector and its forecasts.

Further reduction of this uncertainty is possible thanks to application of several forecasting techniques to the same problem with subsequent combining of separate forecasts using appropriate weighting coefficients. The best results of combining the forecasts is achieved when variances of forecasting errors for different forecasting techniques do not differ substantially (at any rate the orders of the variances should be the same).

3.3 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 required in a case of LS application for parameter estimation. This situation may also take place in a case of multicollinearity of independent variables and substantial influence of process nonlinearity that for some reason has not been taken into account when model was constructed. When power of the 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 group method for data handling (GMDH), are applied. Very often GMDH produces results of acceptable quality with rather 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) [13]. The last techniques could be applied with quite acceptable computational expenses when the number of parameters is not large.

3.4 Dealing with Model Structure Uncertainties

When considering mathematical models it is convenient to use proposed here a unified notion of a model structure which we define as follows: $S = \{r, p, m, n, d, w, l\}$, 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 of a model; n is a nonlinearity and its type; d is a lag or output reaction delay time; is stochastic external disturbance and its type; l are possible restrictions for the variables and/or parameters. When using DSS, the 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 with a set 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 (improvement) of input delay time (lag) and data distribution type with its parameters; describing detected process nonlinearities with alternative analytical forms with subsequent estimation of model adequacy and forecast quality. As another example of complex statistical model adequacy and forecast quality criterion could be the following:

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

where R^2 is a determination coefficient; DW is Durbin-Watson statistic; $MAPE$ is mean absolute percentage error for forecasts;

$\sum_{k=1}^{N} e^2(k) = \sum_{k=1}^{N} [y(k) - \hat{y}(k)]^2$ is sum of squared model errors; U is Theil coefficient that measures forecasting characteristic of a model; α, β are appropriately selected weighting coefficients; $\hat{\theta}_i$ is parameter vector for the *i-th* candidate model. A criterion of this type is used for automatic selection of the best candidate model. The criterion also allows operation of DSS in the automatic adaptive mode. Obviously, other forms of the complex criteria are possible. While constructing the criterion it is important not to overweigh separate members in the right hand side.

3.5 Coping with Uncertainties of a Level (Amplitude) Type

The use of random (i.e. with random amplitude or a level) and/or non-measurable variables leads to 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 they come for analysis in acceptable time span. However, some variables cannot be measured (registered) 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. In such situations we could assign to the variable a set of possible values in the linguistic form as follows: *capital amount* = {*very low*, *low*, *medium*, *high*, *very high*}. There exists a complete necessary set of mathematical operations to be applied to such fuzzy variables. Finally fuzzy value could be transformed into usual exact form using known techniques.

3.6 Processing Probabilistic Uncertainties

To fight probabilistic uncertainties it is possible to hire Bayesian approach that helps to construct models in the form of conditional distributions for the sets of random variables. Usually such models represent the process (under study) variables themselves, stochastic disturbances and measurement errors or noise. The problem of distribution type identification also arises in regression modeling. Each probability distribution is characterized by a set of specific values that random variable could take and the probabilities for these values. The problem is in the distribution type identification and estimating its parameters. The probabilistic uncertainty (will some event happen or not) could be solved with various models of Bayesian type. This approach is known as Bayesian programming or paradigm. The generalized structure of the Bayesian program includes the following steps: (1) problem description and statement with putting the question regarding estimation of conditional probability in the form: $p\,(X_i|D,\, Kn)$, where X_i - is the main (goal) variable or event; the probability p should be found as a result of application of some probabilistic inference procedure; (2) statistical (experimental) data D and knowledge K_n are to be used for estimating model and parameters of specific type; (3) selected and applied probabilistic inference technique should give an answer to the question put above; (4) analysis of quality of the final result. The steps given above are to some extent "standard" regarding model constructing and computing probabilistic inference using statistical data available. This sequence of actions is naturally consistent with the methods of cyclic structural and parametric model adaptation to the new data and operating modes (and possibly expert estimates).

One of the most popular Bayesian approaches today is created by the models in the form of static and dynamic 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 the arcs showing existing causal relations between the variables. Each variable of BN is characterized with complete finite set of mutually excluding states. Formally BN could be represented with the four following components: $\mathbf{N} = <\mathbf{V}, \mathbf{G}, \mathbf{P}, \mathbf{T}>$, where $\mathbf{V}$ stands for the set of model variables; $\mathbf{G}$ represents directed acyclic graph; $\mathbf{P}$ is joint distribution of probabilities for the graph variables (vertices), $\mathbf{V} = \{X_1, \ldots, X_n\}$; and $\mathbf{T}$ denotes conditional and unconditional probability tables for the graphical model variables [14, 15]. The relations between the variables are established via expert estimates or applying special statistical and probabilistic tests to statistical data (when available) characterizing dynamics of the variables.

The process of constructing BN is generally the same as for models of other types, say regression models. The set of the model variables should satisfy the Markov condition that each variable of the network does not depend on all other variables but for the variable's parents. In the process of BN constructing first the problem is solved of computing mutual information values between all variables of the net. Then an optimal BN structure is searched using acceptable quality criterion, say well-known minimum description length (MDL) that allows analyzing and improving the graph (model) structure at each iteration of the learning algorithm applied. Bayesian networks provide the following advantages for modeling: the model may include qualitative and quantitative variables simultaneously as well as discrete and continuous ones; number

of the variables could be very large (thousands); the values for conditional probability tables could be computed with the use of statistical data and expert estimates; the methodology of BN constructing is directed towards identification of actual causal relations between the variables hired what results in high adequacy of the model; the model is also operable in conditions of missing data.

To reduce an influence of probabilistic and statistical uncertainties on models quality and the forecasts based upon them it is also possible to use the models in the form of Bayesian regression based on analysis of actual distributions of model variables and parameters. Consider a simple two variables regression

$$y(k)|x(k) = \beta_1 + \beta_2 x(k) + u(k), \quad k = 0, 1, \ldots, n.$$

It is supposed that of random values $u_1, \ldots, u_n$ are independent and belong, for example, to normal distribution, $\{u(k)\} \sim N(0, \sigma_u^2)$; here vector of unknown parameters includes three elements, $\theta = (\beta_1, \beta_2, \sigma_u^2)^T$. The likelihood function for dependent variable $\mathbf{y} = (y_1, \ldots, y_n)^T$ and predictor $\mathbf{x} = (x_1, \ldots, x_n)^T$ without proportion coefficient is determined as follows:

$$L(\mathbf{y}|\mathbf{x}, \beta_1, \beta_2, \sigma_u) = \frac{1}{\sigma_u^N} \exp\left\{ -\frac{1}{2\sigma_u^2} \sum_{k=1}^{N} [y(k) - \beta_1 - \beta_2 x(k)]^2 \right\}$$

Using simplified (non-informative) distributions for the model parameters:

$$g(\beta_1, \beta_2, \sigma_u) = g_1(\beta_1) g_2(\beta_2) g_3(\sigma_u),$$

$$g_1(\beta_1) \propto const,$$

$$g_2(\beta_{21}) \propto const,$$

$$g_3(\sigma_u) \propto 1/\sigma_u,$$

and Bayes theorem it is possible to find joint posterior distribution for the parameters in the form [16]:

$$h(\beta_1, \beta_2, \sigma_u|x, y) \propto \frac{1}{\sigma} \frac{1}{\sigma^N} \exp\left[-\frac{1}{2\sigma^2} \sum_{k=1}^{N} (y(k) - \beta_1 - \beta_2 x(k))^2 \right],$$
$$-\infty < \beta_1, \beta_2 < +\infty, \quad 0 < \sigma_u < \infty$$

Maximum likelihood estimates for the model parameters are determined as follows:

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}; \quad \hat{\beta}_2 = \frac{\sum_{k=1}^{N} [x(k) - \bar{x}][y(k) - \bar{y}]}{\sum_{k=1}^{N} [x(k) - \bar{x}] \sum_{k=1}^{N} [y(k) - \bar{y}]},$$

where $\bar{x} = N^{-1}\sum_{k=1}^{N} x(k)$, $\bar{y} = N^{-1}\sum_{k=1}^{N} y(k)$, with unbiased sample estimate of variance:

$$\hat{\sigma}_u^2 = s^2 = \frac{1}{N-2}\sum_{k=1}^{N}[y(k) - \hat{\beta}_1 - \hat{\beta}_2 x(k)]$$

Joint posterior density for the model parameters corresponds to two dimensional Student distribution:

$$h_1(\beta_1, \beta_2|\mathbf{y}, \mathbf{x}) \propto \left\{(N-2)s^2 + N(\beta_1 - \hat{\beta}_1)^2 + (\beta_2 - \hat{\beta}_2)^2 \sum_{k=1}^{N} x(k)^2 + 2(\beta_1 - \hat{\beta}_1)(\beta_2 - \hat{\beta}_2)\sum_{k=1}^{N} x(k)\right\}^{-0,5N}.$$

This way we get a possibility for using more exact distributions of models variables and parameters what helps to enhance model quality. Using new observation x^* and prior information regarding particular model it is possible to determine the forecast interval for the dependent variable, y^*, as follows:

$$p(y^*|x^*) = \iiint L(y^*|x^*, \beta_1, \beta_2, \sigma)h(\beta_1, \beta_2, \sigma)|\mathbf{x}, \mathbf{y})d\beta_1, d\beta_2, d\sigma.$$

Another useful Bayesian approach is in hierarchical modeling that is based on a set of simple conditional distributions comprising one model. The approach is naturally combined with the theory of computing Bayesian probabilistic inference using modern computational procedures [17]. The hierarchical models belong to the class of marginal models where the final result is provided in the form of a distribution $\boldsymbol{P}(\mathbf{y})$, where $\mathbf{y}$ is available data vector. The models are formed from the sequence of conditional distributions for selected variables including the hidden ones. The hierarchical representation of parameters usually supposes that data, $\mathbf{y}$, is situated at the lower (first) level, model parameters (second level) $\theta = (\theta_i, i = 1, 2, \ldots, n)$, $\theta_i \sim N(\mu, \tau^2)$, determine distributions of dependent variables $y_i \sim N(\theta_i, \sigma^2), i = 1, 2, \ldots, n$, and parameters $\{\theta_i\}$ are determined by the pair, (μ, τ^2), of the third level. Supposing the parameters σ^2 and τ^2 accept known finite values, and parameter μ is unknown with the prior π_μ, then joint prior density for (θ, μ) could be presented in the form: $\pi_\mu(\mu)\prod_i \pi_\theta(\theta_i|\mu)$, and the prior for parameter vector θ will be defined by the integral: $p(\theta) = \int \pi_\mu(\mu)\prod_i \pi_\theta(\theta_i|\mu)d\mu$.

4 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 appropriately selected 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 the data available;
- normalizing of data in the selected range in a case of necessity;
- application of low-order digital filters (usually low-pass filters) for separation of observations from measurement noise;
- application of optimal (usually Kalman) filters for optimal state estimation and fighting stochastic uncertainties;
- 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 (say, moving average processes based upon measurements of dependent variables).

It is also useful to test how informative is the data collected. Very formal indicator for the 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 example, the equation given below can describe rather complex process with nonlinear trend and short-term variations imposed on the trend curve:

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

Where $y(k)$ is basic dependent variable; a_i, c_i are model parameters; $k = 0, 1, 2, \ldots$ is discrete time; $\varepsilon(k)$ 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 of discrete time k. In this case maximum number of derivatives could 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 model 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 criteria mentioned above in (1) and (2). To perform automatic model selection the above mentioned combined criteria (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.

5 Conclusions

The general methodology was proposed for mathematical modeling and forecasting dynamics of economic and financial processes that is based on the system analysis principles. As instrumentation for fighting possible structural, statistic and parametric uncertainties the following techniques are used: Kalman filter, various missing data

imputation techniques, multiple methods for model parameter estimation, and Bayesian programming approach. The issues of estimating the quality of forecasted solutions are considered.

References

1. Tsay, R.S.: Analysis of Financial Time Series. Wiley, Chicago (2010). 715 pages
2. Harris, L., Hong, X., Gan, Q.: Adaptive Modeling, Estimation and Fusion from Data. Springer, Heidelberg (2002). 323 pages
3. Congdon, P.: Applied Bayesian Modeling. Wiley, Chichester (2003). 472 pages
4. De Lurgio, S.M.: Forecasting Principles and Applications. McGraw-Hill, Boston (1998). 802 pages
5. Taylor, S.J.: Modeling stochastic volatility: a review and comparative study. Math. Fin. **2**, 183–204 (1994)
6. Burstein, F., Holsapple, C.W.: Handbook of Decision Support Systems. Springer-Verlag, Heidelberg (2008). 908 pages
7. Hollsapple, C.W., Winston, A.B.: Decision Support Systems. West Publishing Company, Saint Paul (1996). 860 pages
8. Bidyuk, P.I., Gozhyj, A.P.: Computer Decision Support Systems. Black Sea State University named after Petro Mohyla, Mykolaiv (2012). 380 pages
9. Xekalaki, E., Degiannakis, S.: ARCH Models for Financial Applications. Wiley, Chichester (2010). 550 pages
10. Chatfield, C.: Time Series Forecasting. Chapman & Hall, CRC, Boca Raton (2000). 267 pages
11. Anderson, W.N., Kleindorfer, G.B., Kleindorfer, P.R., Woodroofe, M.B.: Consistent estimates of the parameters of a linear system. Ann. Math. Stat. **40**, 2064–2075 (1969)
12. Gibbs, B.P.: Advanced Kalman Filtering, Least-Squares and Modeling, Wiley, Hoboken (2011). 267 pages
13. Gilks, W.R., Richardson, S., Spiegelhalter, D.J.: Markov Chain Monte Carlo in Practice. Chapman & Hall, CRC, New York (2000). 486 pages
14. Jensen, F.V., Nielsen, T.D.: Bayesian Networks and Decision Graphs. Springer, New York (2007). 457 pages
15. Zgurovsky, M.Z., Bidyuk, P.I., Terentyev, O.M., Prosyankina-Zharova, T.I.: Bayesian Networks in Decision Support Systems. Edelweis, Kyiv (2015). 300 pages
16. Bernardo, J.M., Smith, A.F.M.: Bayesian Theory. Wiley, New York (2000). 586 pages
17. Bolstad, W.M.: Understanding Computational Bayesian Statistics. Wiley, Hoboken (2010). 334 pages