Journal of Mathematics and System Science 6 (2016) 112-122
doi: 10.17265/2159-5291

Features of SAS Enterprise Guide for probabilistic Modeling System, Macroeconomic Analysis and Forecasting

Prosyankina-Zharova Tetyana[1], Terentiev Oleksandr[2], Bidyuk Petro[3] and Makukha Mikhailo[4]

1. European University, Department of Information Systems and Technologies, Kyiv, Ukraine, 16-V Acad. Vernadskiy Blvd

2. National Technical University of Ukraine "KPI", Institute for applied system analysis, Kyiv, Ukraine, 37 st. Pobedy

3. National Technical University of Ukraine "KPI", Institute for applied system analysis, Kyiv, Ukraine, 37 st. Pobedy

4. National Technical University of Ukraine "KPI", Institute for applied system analysis, Kyiv, Ukraine, 37 st. Pobedy

Received: December 01, 2014 / Accepted: November 30, 2015 / Published: March 25, 2016.

Abstract: This paper addresses to the problem of using SAS Enterprise Guide 6.1 as a means for building probabilistic models and as optimum method of modeling gross domestic product in terms of the economic crisis and social threats is proposed. Today in a complex socio-political and economic situation growing influence of external factors, presence of uncertainties and risks there exists a problem of anticipating potential threats in the humanitarian and social spheres and ways to overcome them aiming to provide food security and controllability of ecological situation. All these problems, as reported in the NATO program “Science for Peace and Security”, are of high priority for the countries that need to take into account threats to security, including Ukraine. That is why in the framework of the project NUKR. SFPP G4877 “Modeling and Mitigation of Social Disasters Caused by Catastrophes and Terrorism” the problems of scientific prediction of national economy for the period to 2030 as one of the measures preventing growth of social tension in the country are disclosed.

Keywords: Bayesian networks, data-mining, forecasting, modeling of gross domestic product

1. Introduction

Among the tools of data-mining Bayesian networks do not enjoy great popularity among academic economists and business analysts yet, though they provide perspective probabilistic tools for modeling complex hierarchical process (static and dynamic) with uncertainties of arbitrary nature that provides the ability to describe accurately their functioning forecasts to assess and build control systems for various purposes. Due to presentation of the interaction between the factors in the form of cause-effect relationships, the highest level of visualization and, consequently, a clear understanding of the processes occurring in the system are achieved. Also among the advantages of Bayesian networks for data-mining there is the possibility to incorporate statistical uncertainty of structural and parametric nature which is especially important for scenario analysis as well as for inferencing through various methods - close and accurate.

The network is constructed in the form of directed acyclic graph that displays the causal relationships between nodes (variables) of the process investigated. Reduction of joint probability distribution as a product of conditional probabilities that depend on a small number of variables helps to avoid "combinatorial explosion" in the simulation.

Generally Bayesian methodology is actually much broader than family of manipulation means with

Corresponding author: Prosiankina-Zharova Tetyana Ivanivna, Ph.D., research field: modeling, macroeconomic analisys and forecasting. E-mail: t.pruman@gmail.com.

conditional probabilities in oriented graphs. These are models with symmetric connections (random fields and grilles), models of dynamic processes (Markov chains), as well as a wide class of models with hidden variables that allow to solve the probabilistic problems of classification, pattern recognition and forecasting.

However, despite the advantages of the possibility to process both statistical data and expert estimates, the use of probabilistic modeling apparatus in the study of socio-economic and socio-political processes of different nature in Ukraine is extremely limited. This is due primarily to objective reasons, including the fact that a significant amount of statistical data is either confidential or for some years nonexistent due to the adaptation of the national statistical methodology to international standards. Such restrictions impede formation of sufficiently long sets of time series of these indicators which in its turn is the cause of the inability of clear and complete definition of quantitative characteristics of the most economic processes, the study of inter-branch exchange, ensuring coherence of performers and coordinators of processes and complexity of the decision making process by the necessity of processing large number of alternatives.

Solving these problems requires a scientific and reasonable methodology of practical application of data-mining systems for modeling and forecasting of macroeconomic and socio-political processes in various scenarios under uncertainty, and an active proceeding of specialized software for data- mining, including such powerful tool as SAS Enterprise Guide and SAS Enterprise Miner [1] that meet modern requirements for handling large amounts of data through the implementation of complex machine learning algorithms such as neural networks, cluster analysis, decision trees, ensemble models, gradient busting, Bayesian networks, regression analysis and others. These solutions are scalable according to the size of solved business problems. The module of high-performance processing uses the technologies of multithreaded and distributed data processing, using available processor and memory capabilities of your computer. To build a Bayesian network in the set of procedures of the core system of SAS Foundation, a procedure HPBNET was implemented, with which you can build different types of networks: naive Bayesian network, tree-augmented, naïve; Bayesian network- augmented naïve, the structure of parent-child and Markov blanket [1]. The advantages of using the procedure HPBNET is that it provides a choice of variables based on the independent tests as well as the automatic selection of the best model.

This high-performance tool can solve most of the problems related to the choice of the optimal Bayesian network structure, that is the one that describes the modeled process as exactly as possible. The need for powerful software tools for building Bayesian networks is associated with large amounts of computation to be performed in the construction of the network structure. In particular, the widespread use of evaluation methods of Bayesian network structure use test for conditional independence without using the ordered set of nodes, which exponentially increases the number of such tests. In addition, often in the analysis of socio-economic processes at the macro level there is a need to build the structure of a network with hundreds or even thousands of nodes, which involves processing training samples from hundreds thousands or even millions of records.

In general, the development of Bayesian algorithms is paid much attention to in the works of C. K. Chow and C. Liu [2], G. Rebane and J. Pearl [3], G. Cooper and E.A. Herskovits [4, 5], S. Wong and Y. Xiang [6], W. Lam and F. Bacchus [7], N. Friedman and M. Goldszmit [8], J. Suzuki [9, 10], who studied the methods based on evaluation functions. In addition, a number of algorithms based on tests for conditional independence are offered in the works of N. Wermuth [11], S. Lauritzen [11, 12] and J. Pearl [13].

Methods that can solve the problems of modeling processes domain area of which is not fully defined or there are no separate data sets in time series become currently important. These methods of building network structures with hidden variables, limits compression methods, algorithms that use the maximum expectation are described in works [14-26]. The reason to use these methods is a lack of full volume of statistical data to identify patterns and establish an adequate structure and calculation estimates of the model parameters. In addition, a significant amount of statistical data is either confidential or non-existent for some years, affecting the adequacy of the results of the economic analysis obtained by traditional methods. Such restrictions on the confidentiality of certain socio-economic indicators, statistical methodology adaptation to international standards complicate the formation of sufficiently long time series sets of these indicators. This situation caused by the inability of clear and complete definition of quantitative characteristics of most economic processes, cross-industry exchange of research and decision-making process is complicated by the large number of alternatives. Therefore, the use of traditional methods of economic-mathematical and statistical analysis does not give acceptable results. This is a major and significant disadvantage of majority of methods used in the development of state programs and strategies when studies are performed mainly on the basis of short time intervals (1- 3 years) and are based solely on numerical characteristics. This prevents not only the possibility of fully track cycling and trends in macroeconomic indicators changes, but also considering multifactor dependence on short-term, identifying development trends for the future.

Therefore, this article suggests the use of decision support systems designed to macroeconomic forecasting which are based on the use of Bayesian networks in combined with morphological analysis, cognitive and economic mathematic modeling. To determine the prospects of Ukraine's economy development in terms of socio-political and economic crisis, the gross domestic product and gross fixed capital formation were forecasted by different scenarios (pessimistic, moderate, optimistic). To form scenario sets it was suggested to perform means of morphological analysis on the basis of the set of factors identified by the SWOT and PEST-analysis, and to determine the relations with the external environment and within the system – to use cognitive modeling. As a result of cognitive modeling a set of factors which are the most significant for the changes of the gross domestic product (the resulting rate of economic growth) was selected. It allowed reducing the number of input variables.

Since the investigated system contains a number of uncertainties for building Bayesian network, a method of construction with hidden nodes was chosen. Basing on Bayesian network topology the most significant variables affecting the intentional were defined, after that a multiple regression equation with forced inclusion in the model identified variables was built. Evaluation of the parameters of this model was performed basing on recursive least squares method. The undertaken study allowed grounding the most probable scenario of Ukraine's economic development in the short and long term. The feature of the suggested method is that the solution of problems of this kind belongs to the range of problems of system analysis on the one hand and the problems of decision-making on the other, so the use of a composition of different approaches to solve formalized and weakly formalized problems taking into account the uncertainty allows to optimize the existing modeling system, forecasting and control in the presence of various kinds of uncertainties, creates conditions for wider introduction of modern highly efficient decision support systems.

2. Experimental Section

The problem of integrated system development for macroeconomic planning and forecasting through the application of modern mathematical tools and information technologies remains unresolved. Solution of this problem is constrained by inertia as the executive authorities and the lack of established system of scientifically grounded methodological approaches to the analysis and modeling of national economy and its constituents. Therefore, the substantiation of the choice of methods and tools for analysis and forecasting of national economy under the conditions of potential socio-political and economic threats is an urgent task.

Development of forecasting system of the national economy which would avoid subjective expert assessments, process large volumes of both quantitative and qualitative indicators in a short period of time under changeable conditions of the present, which are characterized by social and economic threats, is one of the priorities of scientific prediction.

Traditionally, the use of the multiplier- accelerator model, in which a mathematical relationship between income growth and performance multiplier and accelerator in terms of the function of consumption and investment is optimal for the quantitative evaluation of the cause-consequence relations between the volume of investment and gross domestic product and is as in (1) and (2) [27, p.362-363]:

$$C_t = C_0 + MPC \cdot Y_{t-1}, \quad (1)$$

where C_t is the consumption for the period of time t; C_0 - autonomous consumption; MPC - marginal propensity to consume; Y_{t-1} - output and income in the period preceding time t.

$$I_t = I_0 + b(Y_{t-1} - Y_{t-2}), \quad (2)$$

where I_t - investments in period t; I_0 - autonomous investments; b - coefficient of acceleration; $(Y_{t-1} - Y_{t-2})$ - change in income for the previous period.

It is believed that the use of the multiplier-accelerator model allows both to quantify the change in the gross domestic product under the influence of changes in investments, and the impact of changes in the value of independent demand for the cyclical fluctuations of the economy.

Calculations of multiplier and accelerator investments during 2006-2014 years, made on the basis of the State Statistics Service of Ukraine [28-29] showed that at the state level multiplier-accelerator effect is negligible. One reason is inflation; another is the reduction of national income, accompanied by a decrease in depreciation so that even simple reproduction becomes almost impossible. In addition, the ratio of gross fixed capital formation and gross domestic product remains far from that which ensures the reproduction of the national economy for a long time. This became one of the causes of the financial-economic crisis in 2009 and paved the way for the decline of the national economy in 2014 (fig. 1).

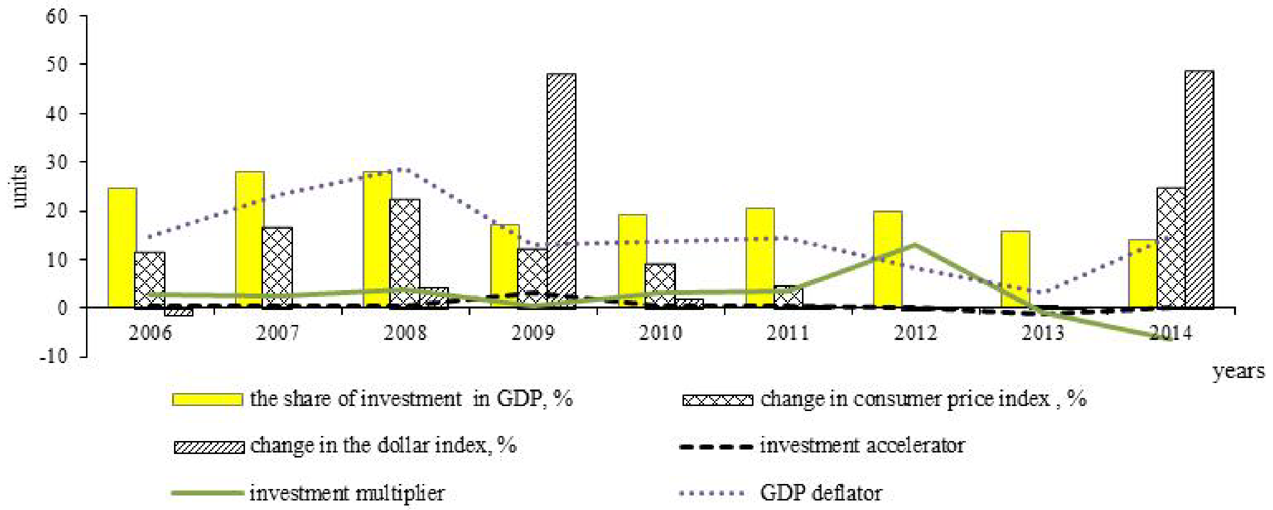

Fig. 1 Graphical representation of certain indicators of national economic development.

Jumping change in the dollar exchange rate and the consumer price index in 2014 with a simultaneous increase in the gross domestic product deflator and the investments accelerator, unlike in 2009, which also casts doubt on the use of traditional methods of macroeconomic forecasting, applied mainly to regression analysis and so on should be noted.

The main and a significant disadvantage of most methods used in the development of state programs and strategies is studied mainly by means of regression analysis based on the short time intervals (1- 3 years) [27, 30, 31]. This prevents not only the possibility to fully track cycling and trends in macroeconomic indicators changes, but to consider multifactor dependence over a long period, to reveal uncertainty. Another problem is the lack of a complete set of data to describe the subject area of a studied problem since a large number of statistical data is either confidential or non-existent for some years due to changes in statistical methodology according to international standards. This complicates formation of sufficiently long time sets of these indicators, makes it impossible to identify patterns by means of traditional econometric methods of analysis, and prevents the formation of adequate structure and calculating estimates of the model parameters affecting the adequacy of the economic analysis results.

Traditional approaches to decision-making, based exclusively on deductive approach using rules of inference, such as "what-if" implied that the investigated problem would be formalized. However, there are many weakly formalized problems which are difficult though possible to find the optimal solution for, and there are problems which are impossible to find an optimal solution by traditional methods based exclusively on formalized methods. The best option of decision support system for macroeconomic forecasting, in our opinion, is a combined decision support system, which combines algorithmic procedures with the decision support system based on decision rules. Therefore, to forecast the development of the national economy the decision support system consisting of subsystems for building Bayesian networks, statistical calculations, regression models, morphological analysis and cognitive modeling is suggested in this work. (table 1).

The analysis of the condition and dynamics of the national economy, SWOT and PEST-analysis showed that the features of the economy of Ukraine is the variability of events development, the presence of uncertainty, a significant impact of a number of countervailing factors that prevents the development of macroeconomic forecasting by one scenario only. Therefore, by means of morphological analysis on the plane with the axes of uncertainty configuration of influence states of the driving forces in the key variables was determined and their influence was set, their consistence was checked through the matrix of coherence. Given the need to evaluate the possibility of occurrence of one of the unforeseen scenarios, probabilistic modeling was used in the process of modeling gross domestic product.

Table 1 Scheme of methods application.

Stage	Methodical approach
1. Information collection, forming a diagnosis model	Empirical research methods, method of statistical observation, analysis and synthesis, system analysis SWOT-analysis, PEST-analysis
2. Review of the condition and dynamics of the studied object, identification of features	Statistical analysis, data-mining, factor analysis, multivariate analysis, principal components method, correlation and regression analysis, typological and structural grouping, RFM-analysis, cluster analysis
3. Causal analysis	Correlation-regression analysis, probabilistic modeling (Bayesian Networks)
4. Scenarios development	Scientific assessment, morphological analysis, scenario modeling
5. Grounding variants of management solutions	Data-mining, neural networks, econometric modeling, cognitive modeling
6. Forecasting	Econometric analysis, data mining methods, probabilistic modeling
7. Results analysis	Scientific assessment, Delphi method, graphic

Since the formation of gross domestic product is affected by a significant number of factors that not only promote its growth, but slow it, it was important to identify causal relationships between factors taking into account different scenarios of events development.

3. Results and Discussion

As the main tool of analytical data processing such powerful analytical tool like SAS Enterprise Guide 6.1 [1] was used, which allowed to process large volumes of statistics. Overall over a thousand time series in the annual cutover of Ukrainian regions and economic activities according to current classifier of economic activities was processed by means of ETL procedures in the programming language SAS Base. The study was performed in two stages. Bayesian network topology that provides the information needed to identify causal relationships between variables and power relations between them was built on the first stage. When building a Bayesian network 56 indicators of socio-economic development of Ukraine for the period 2010-2014 years [28, 29] were used, from which 18 were used to build the network.

Bayesian Network is a pair of values $<G,P>$, where G - directed acyclic graph consisting of a set of nodes U, and P- the set of values of probabilities. Combined probability values distribution for the nodes U are described as:

$$p(U) = \prod_{X \in U} p(X|\pi(X)), \qquad (3)$$

where $\pi(X)$ – set of parental nodes for X.

It is believed that the network topology G and probabilities distribution P corresponding each other.

For the target variable T Markov blanket is defined a subset of input variables $MB \subseteq U - \{T\}$, such that T is relatively independent of each input variable $X \subseteq U - MB - \{T\}$ on the set of Bayesian network. It is believed that the target variable T corresponding the only unique Markov blanket structure.

The procedure HPBNET supports two approaches to variables selection. The first approach is using tests for independence between each input variable and the target (the case when given option PRESCREENING = 1). The second approach is using conditional independence tests between each target and input variable for a given set of input variables (the case when given option VARSELECT = 1, 2 or 3).

In general HPBNET procedure uses two approaches to build network topology based on the evaluation function and on the basis of restrictions. The approach based on evaluation function uses functional which makes clear the quality of the structure obtained on training data, and seeks to build a structure with the best value of the functional. Restrictions based approach uses tests to determine the independence of branches and their directions [1]. To implement an approach based on building topology based on evaluation function in the procedure HPBNET Bayesian information criterion (BIC - Bayesian Information Criterion) was implemented and is calculated using the formula:

$$BIC(G,D) = N \cdot \sum_{i=1}^{n}\sum_{j=1}^{q_i}\sum_{k=1}^{r_i} p(\pi_{ij}) \cdot p(X_i = v_{ik}|\pi_{ij}) \cdot \ln\left(p(X_i = v_{ik}|\pi_{ij})\right) - , \\ -\frac{M}{2} \cdot \ln(N) \qquad (4)$$

where G – topology of network, D – training data set, N– number of observations in D, n – number of variables, X_i – i-variable, r_i– number of variable levels X_i, π_{ij} – j parents variable combination X_i, $M = \sum_{i=1}^{n}(r_i - 1) \times q_i$ – the number of parameters of probability distribution.

The procedure HPBNET uses tests to determine the independence of branches and their directions. We assume that three variables X, Y, Z exist. After completing the test for independence, it was determined that there is a connection between X and Z, Y and Z, but there is no connection between X and Y. If X is conditionally independent of Y for a given subset of variables $S = \{Z\} \cup S', S' \subseteq U - \{X,Y,Z\}$, direction between X and Y and Z, is designated as $X \to Z$ and $Y \to Z$ respectively.

It should be pointed out that the use of tests for independence cannot show the orientation relationships between variables because some patterns are equivalent and have the same test values for conditional independence. For example, $X \leftarrow Y \leftarrow Z$, $X \rightarrow Y \rightarrow Z$ and $X \leftarrow Y \rightarrow Z$ belong to equivalent classes. In that case, the procedure HPBNET uses the value function of Bayesian information criterion for determining the direction of bonds.

HPBNET procedure allows to build Bayesian networks of different types, naive Bayes (NB), tree-augmented naive (TAN), Bayes network augmented naive (BAN), parent-child network (PC) and Markov Blanket (MB). Depending on the chosen type of network different algorithms are used. For example, for the case STRUCTURE = TAN algorithm for constructing the widest tree is used, where value of mutual information is used as the weight value of relations between peaks. When forming the topology values BESTONE or BESTSET are used for option PARENTING.

The procedure HPBNET organizes input variables based on the value of Bayesian information criterion of the function and is calculated by the formula (5):

$$BIC(X, T) = \max(BIC(X \rightarrow T), BIC(T \rightarrow X)), \tag{5}$$

where X – input variable, a T – target variable, $BIC(T \rightarrow X)$ is the value Bayesian information criterion function when X father towards T (ignoring all other variables), and $BIC(T \rightarrow X)$ is the case where X is a descendant in relation to T (ignoring all other variables).

HPBNET procedure primarily defines parental nodes in relation to the target structures "parent-child" and Markov blanket. After that identifies parents for input variables that have the highest value for Bayesian information criterion of the function in relation to the target. This continues until there are input variables with large values of the function of Bayesian information criterion. When determining parental nodes in relation to the input, the relationship is based on tests for independence. After that the orientation (direction) of relations is made on the basis of tests on the independence and function of information criterion. HPBNET procedure uses the function of Bayesian information criterion for targeting not only relations, but also to control the complexity of the network because a complicated network has many fathers, for which the function of Bayesian information criterion is imposed a penalty value.

Values Bestone and Bestset of the option PARENTING try to identify local best structure for each peak. BESTONE adds the best candidate - variable to parents at each iteration, while BESTSET tries to choose the best candidates among the plurality of sets of variables.

If there are too many input variables, the structure training may take a long time because the number of possible combinations of variables increases exponentially. Therefore, selection of the most important variables of the analysis was made at a previous stage.

Parameters study consists in determining the probability distribution for each peak of the network structure. Probability distribution has a discrete character for procedures HPBNET because interval variables are divided into groups (intervals).

According to the received results, $(x_1, x_2, \ldots, x_{n-1})$ notation by formula 6 is reflected in a table of probability distribution:

$$\begin{aligned} &\arg\max_c p(T = c \mid x_1, x_2, \ldots, x_{n-1}) = \\ &p(x_1, x_2, \ldots, x_{n-1} \mid T = c) \times K = \\ &\prod_i p(x_i \mid = \pi(X_i)) \times K \end{aligned} \tag{6}$$

where c – level of a target variable, $\pi(X_i)$ – parent variable of X_i, K – constant, $X_n = T$ – target variable.

Statistical characteristics of the constructed model: Misclassification Rate - 22%, Roc Index - 0.84.

On the basis of the constructed topology it was found that the highest growth of the gross domestic product (6%) can be achieved by increasing such factors as tax revenues (more than 30%), volume of deposits in the economy (more than 20%), capital investment in agriculture (more than 36%), growth in volume of gross value added of agricultural products (more than 25%), growth in volume of capital investment in food industry (more than 20%). The probability of this scenario implementation is quite low that is 26%.

If tax revenues (more than 14%), volume of deposits in the economy (not less than 17%) and investment in capital assets (more than 36%), gross value added of agricultural products (more than 15%), growth in volume of capital investment in food industry (more than 10%) are increased, the gross domestic product will grow by 2.8%. The probability of such developments is rather low that is 51%. In other words the use of probabilistic modeling allowed to move from quality values applied in morphological analysis to numerical characteristics of the development of national economy.

The following models were constructed during the further research: autoregressive model of the first order, autoregressive model of the first order including seven exogenous variables, regression model where input variables were used as a combination of products and polynomials, neural network with previous selection of input variables which are the most correlated with the input variable, neural network with previous selection of input variables, neural network with the previous reduction of the number of input variables based on cluster analysis, regression model with clusters as input variables, combined model as a Bayesian network using multiple regression.

In particular, mathematical model of the gross domestic product of the first order is (7):

$$Y=11873.2+0.7*Y(-1)-0.29*Y(-2)+0.41*X-0.35*X(-1) \tag{7}$$

where Y – the gross domestic product of Ukraine, million of UAH; X – volume of investment in capital assets, million of UAH.

The model used for a long-term forecasting of the gross fixed capital formation is as follows (8):

$$V=-57510.57+1.02*Y(-1)+0.48*Y-0.46*Y(-1) \tag{8}$$

where Y – value range of the gross domestic product, of Ukraine, million of UAH; V – volume of the gross fixed capital formation of Ukraine, million of UAH.

Forecasting results are presented in table 2.

Table 2 The results of long-term forecasting of gross domestic product and gross fixed capital formation by different scenarios

Year	Predictive value of gross domestic product by the scenario, millions of UAH			Forecasting of gross fixed capital formation, millions of UAH	Ratio of gross fixed capital formation and gross domestic product by optimistic scenario, %
	Moderate	*Pessimistic*	*Optimistic*		
2015	1 684 854	1 128 852.058	2 240 856	336 128.37	15
2016	1 802 980	1 207 996.356	2 397 963	359 694.51	15
2017	1 921 105	1 287 140.655	2 555 070	408 811.14	16
2018	2 039 231	1 366 284.953	2 712 177	433 948.36	16
2019	2 157 357	1 445 429.251	2 869 285	573 856.96	20
2020	2 275 483	1 524 573.549	3 026 392	665 806.33	22
2021	2 393 609	1 603 717.847	3 183 500	764 039.99	24
2022	2 511 735	1 682 862.145	3 340 608	901 964.04	27
2023	2 629 860	1 762 006.444	3 497 714	1 084 291.3	31
2024	2 747 986	1 841 150.742	3 654 821	1 206 091.1	33
2025	2 866 112	1 920 295.04	3 811 929	1 257 936.6	33
2026	2 984 238	1 999 439.338	3 969 037	1 349 472.4	34
2027	3 102 364	2 078 583.636	4 126 144	1 444 150.4	35
2028	3 220 489	2 157 727.935	4 283 250	1 584 802.6	37
2029	3 338 615	2 236 872.233	4 440 358	1 731 739.6	39
2030	3 456 741	2 316 016.531	4 597 466	1 838 986.2	40

Table 3 Comparison of forecasting results of gross domestic product by different investment patterns

Year	Forecasting result of gross domestic product by:		Deviations of predicting value,%
	moderate scenario	*existence of investment pattern in agriculture production branch*	
2015	1 684 854	1 747 523.44	3.6
2016	1 802 980	1 871 757.49	3.7
2017	1 921 105	1 995 991.56	3.8
2018	2 039 231	2 120 225.60	3.8
2019	2 157 357	2 244 459.65	3.9
2020	2 275 483	2 368 693.71	3.9

Built economic and mathematical model has acceptable characteristics and ensures proper quality of the forecasting. Based on the forecasting it was found that it is necessary to attract not less than 1 084 291.3 millions UAH of investments to ensure simple reproduction of the national economy in 2022-2023 by the positive scenario and 1 838 986.2 millions UAH for advanced reproduction (2030).

By constant sectoral structure of capital investments during 2015-2030, the increase of capital investment is possible in volumes close to predicted by realistic scenario that confirms the forecasting quality. The model built using the method of principal components has the best characteristics among the variety of built models (9):

$$Y=1312704.25+47724.60*Z \qquad (9)$$

where Y – is a volume of gross domestic product in millions of UAH; Z - the main component which is described by the formula: $Z= 0.28 * I_{01} + 0.27 * I_{02} + 0.23 * I_{03} + 0.17 * I_{04} + 0.22 * I_{05} + 0.14 * I_{06} +0.27 * I_{07} -0.29 * I_{08} +0.29 * I_{09} +0.28 * I_{10}+0.25 * I_{11} + 0.17 * I_{12} + 0.27 * I_{13} – 0.26 * I_{14}$; $I_{01} – I_{14}$ – the volumes of capital investments by kinds of economic activities, millions of UAH.

In general, the following models were built during the study of prospects of the national economy development: autoregressive model of first order, autoregressive model of first order with the inclusion of seven exogenous variables, regression model where a combination in the form of products and polynomials were used as its input variables, combined model in form of Bayesian network using multiple regression

Analyzing the predicting results of financial crisis consequences by developed model, it should be noted that the decline in gross domestic product of Ukraine was predicted by 18.5% in 2015 with high probability (51%) against the value of the corresponding index in 2014, which fully corresponds to previous statistical data.

4. Conclusions

The problem of preventing and response to potential threats in humanitarian and public spheres, promoting the growth of national economy, ensuring the national food security, controllability of ecological situation etc. is characterized by complexity, the presence of significant amounts of quantitative and qualitative information, uncertainty, and existence of complex causal relationships between factors.

Therefore, research of problems of forecasting the national economy development in a complex socio-political and economic situation, the growing influence of external factors, the presence of uncertainties and risks requires a systematic multidisciplinary approach, the use of complex analysis instruments and modeling the complex systems. The proposed scheme of the research provides an integrated analysis and forecasting, allows to analyze the extent and the character of impact of various factors on the development of the situation in the national economy, both on short and long term and to build appropriate models of acceptable quality.

Using the developed approach allows us to consider various aspects of the research problem, to structure the subject field using conceptual schemes, to explore possible scenarios of the development, to simulate and predict the development of the national economy according to the most probable scenarios, to define the conditions and factors of formation of the best possible solutions for the national economy development and prevention of the growth of social tension in the country.

References

[1] SAS Enterprise Miner 13.2: Reference Help. SAS Documentation, SAS Institute Inc., Cary, NC, USA, (2014), 1740 p.

[2] C. K. Chow, C. N. Liu Approximating discrete probability distributions with dependence trees // IEE Transactions on information theory, May 1968. Vol. IT-14, №3. P. 462-467.

[3] G. Rebane, J. Pearl The recovery of causal poly-trees from statistical data. // International journal of approximatereasoning, July 1988. Vol. 2, №3. P. 175-182

[4] E. Herskovits and G. Cooper Kutato: an entropy-driven system for construction of probabilisticexpert systems from databases / Proceedings of the sixth international conference on Uncertainty in Artificial Intelligence (UAI'90), Cambridge, Massachusetts, USA, 27-29 July, NY.: Elsevier science, 1991. P 54-62.

[5] G. Cooper, E. Herskovits A Bayesian method for the induction of probabilistic networks from data. // Machine Learning, 1992. 9. P. 309-347.

[6] S. Wong, Y. Xiang Construction of a Markov network from data for probabilistic inference / Third International workshop on rough sets and soft computing (RSSC'94), San Jose (California). 1994. P. 562-569.

[7] W. Lam, F. Bacchus Learning Bayesian belief networks: an approach based on the MDL principle. // Computational Intelligence, July 1994. Vol. 10, № 4. P. 269-293.

[8] N. Friedman, M. Goldszmidt Learning Bayesian networks with local structure. / Proceedings of the twelfth international conference on Uncertainty in Artificial Intelligence (UAI'96), Portland, Oregon, USA, 1-4 August, 1996. SF.: Morgan Kaufmann, 1996. P. 252-262.

[9] J. Suzuki Learning Bayesian belief networks based on the MDL principle: an efficient algorithm using the branch and bound technique. // IEICE Transaction on information and systems, February 1999. P. 356-367.

[10] J. Suzuki Learning Bayesian belief networks based on the minimum description length principle: basic properties. // IEICE Transaction. on fundamentals, September 1999. Vol. E82-A, № 9. 9 p.

[11] Wermuth N. and Lauritzen S. Graphical and recursive models for contingency tables // Biometrika, December 1983. Vol. 70, №3. P. 537-552.

[12] S. L. Lauritzen The EM algorithm for graphical association models with missing data [Text] / Computational Statistics and Data Analysis. 1995. Vol. 19. P. 191-201

[13] J. Pearl Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, San Francisco, Morgan Kaufmann (1988), 552 p.

[14] T. Chu, Y. Xiang Exploring parallelism in learning belief networks / Proceedings of the thirteenth international conference on Uncertainty in Artificial Intelligence (UAI'97), Providence, Rhode Island, USA, 1-3 August, 1997. SF.: Morgan Kaufmann, 1997. P. 90-98.

[15] S. J. Russell Artificial Intelligence: A Modern Approach (2nd Edition) [Text] / Stuart J. Russel, Peter Norvig. New Jersey.: Prentice Hall, 2002. 1132 p.

[16] J.-L. Golmard, A. Mallet Learning probabilities in causal trees from incomplete databases [Text] / Revued' Intelligence Arti?cielle, 5. 1991. P. 93-106.

[17] D. J. Spiegelhalter, R. G. Cowell Learning in probabilistic expert systems [Text] / Bayesian Statistics, vol. 4. Oxford, UK:Oxford University Press, 1992. P. 447-465.

[18] D. Spiegelhalter, P. Dawid, S. Lauritzen, R. Cowell Bayesian analysis in expert systems [Text] / Statistical Science, 8. 1993. P.219-282.

[19] K. G. Olesen, S. L. Lauritzen, F.V. Jensen. aHUGIN: A system for creating adaptive causal probabilistic networks [Text] / Uncertainty in Arti?cial Intelligence: Proceedings of the Eighth Conference. CA: MorganKaufmann, 1992. P. 223-229/

[20] A. P. Dempster, N. M. Laird, D. B. Rubin Maximum likelihood from incomplete data via the EM algorithm // Journal of the Royal Statistical Society. 1977. Vol. 39, №1. P. 1-38.

[21] N. Friedman Learning belief networks in the presence of missing values and hidden variables [Text] / Fourteenth International Conference on Machine Learning (ICML 1997), Nashville, Tennessee, USA, July 8-12, 1997. SF.: Morgan Kaufmann, 1997. P. 125-133.

[22] W. R. Gilks, S. Richardson, D. J. Spiegelhalter Markov chain Monte Carlo methods in practice [Text] / New York: Chapman Hall. 1996.

[23] D. J. Spiegelhalter, S.L. Lauritzen Sequential updating of conditional probabilities on directed graphical structures [Text] / Networks, 20. 1990. P.579-605.

[24] J. Binder, D. Koller, S. Russell, K. Kanazawa Adaptive probabilistic networks with hidden variables [Text] / Machine Learning. 1997. Vol. 29. P. 213-244.

[25] S. Russell, J. Binder, D. Koller, K. Kanazawa Local learning in probabilistic networks with hidden variables

[Text] / Fourteenth International Joint Conference on Artificial Intelligence, Montreal, Canada, August 20-25 1995. 1995. P.1146-1152.

[26] B. Thiesson Accelerated quantification of Bayesian networks with incomplete data [Text] / In Proceedings of First International Conference on Knowledge Discovery and Data Mining, Montreal, QU. Morgan Kaufmann, 1995. P.306-311.

[27] M. P. Talavyrya, O. V. Pashchenko. Macroeconomics: teaching aid for students of higher educational institutions Kyiv: Lysenko M. M., 2012. 551 p.

[28] The official website of the State Statistics Service of Ukraine. Statistical information. Change of gross domestic product volume for 2014. CHP 2008 [Electronic resource]. Access: http://www.ukrstat.gov.ua/

[29] Statistical Annual of Ukraine for 2014. / Ed. by O.G. Osaulenko. - Kyiv, 2015. - 533 p.

[30] M. Z. Zgurovsky, N.D. Pankratova System strategy of technological forecasting in innovation activity// System Research and Information Technologies. 2003. № 3. P. 7-24

[31] M. Z. Zgurovsky, V.A. Pankratov Strategy of innovative development of the region based on the synthesis of prediction methodologies and cognitive modeling // System Research and Information Technologies. 2014. № 2 - P. 7-17