**Bayesian data analysis in modeling and forecasting nonlinear nonstationary financial and economic processes**

*Trofymchuk O. M., Bidyuk P.I., Prosiankina-Zharova T. I., Terentiev O. M.*

*(Institute of Telecommunications and Global Information Space* [*of NAS of Ukraine*](https://www.nas.gov.ua/EN/About/Pages/museum.aspx) *Е-mail: Trofymchuk@nas.gov.ua, pbidyuke\_00@ukr.net,* [*t.pruman@gmail.com*](mailto:t.pruman@gmail.com)*, o.terentiev@gmail.com*

All Bayesian techniques of data analysis are very popular today thanks to their flexibility, high quality of final results, availability of possibilities for adaptation to new data and conditions of functioning. Besides, each of these approaches to data analysis is well supported by appropriate sets of statistical criteria that make possible thorough quality analysis of intermediate and final results.

***Key words:*** nonlinear nonstationary processes, Bayesian methods, modeling, forecasting, Generalized Linear Models

Many studies today are related to modeling and forecasting evolution of processes in various areas; they are mostly touching upon widely spread nonlinear non-stationary processes (NNP). To be more exact, definition of NNP means that such processes exhibit at least one type on non-stationarity (regarding trend or integration, and variance or heteroscedasticity) as well as nonlinearity regarding variables or model parameters. Such processes create majority in ecology, economy, finances, industrial technologies, engineering systems, hydrology, climate studies, in the problems of technical, medical and economic diagnostics, physical experiments of various type etc. For example, many processes in economy show availability of low (first or second) order trend, but transition to second order of integration automatically shifts the process from the class of linear to the nonlinear ones because quadratic and higher order dependence indicates relation to the class of non-linear characteristics [1-7].

The main goals of the study are as follows: (1) to provide a review of Bayesian data processing and model constructing methods for their further use in intellectual decision support system for modeling and forecasting nonlinear non-stationary processes in economy and finances; (2) to present illustrative examples of Bayesian techniques application to solving the problems mentioned; (3) to stress the necessity of development intellectual DSS for high quality solving the problems.

An important issue of the study is development and implementation of a system that would provide a substantial help in solving the tasks of selected process model building, forecast estimation, alternatives generation, and selection of the best one of them on the purpose of its further practical implementation. Usually this is specialized intellectual decision support system (DSS) containing all necessary computational procedures and sets of statistical criteria to estimate quality of intermediate and final results of process analysis. Thus, the main ideas related to constructing and practical application of the DSS are as follows [8]: (1) to get handy and useful instrument for carrying out necessary computations and get results within short period of time; (2) to test immediately quality of computations using various sets of quality analysis criteria in automatic mode of system operation; (3) to identify possible uncertainties and take countermeasures to reduce their negative influence on the computations; (4) to calculate possible alternatives and select the best one of them for further practical implementation. Availability of intellectual features in the system simplifies the tasks to be performed by the system user and provides substantial help for him in form of advices, extra information on the methods to be used for model structure and parameter estimation, statistical quality criteria to be applied in specific cases, on the quality of intermediate and final results of computations and some other features.

In the case of analysis statistical/experimental data possible uncertainties are touching upon data itself. They can be encountered in the form of state and measurement noise, missing observations, short samples, and multicollinearity. All these problems can be solved using appropriate techniques mentioned earlier [9,10].

As an example of actuarial process modeled by GLM the problem of estimating (forecasting) financial loss in car insurance had been selected. The experimental data includes one dependent variable, “*Loss*”, reflecting paid volume of insurance among the cars of the following brands: VAZ, Mitsubishi, and Toyota. The regions of selling (distribution) of the car policies include the cities of Kyiv, Donetsk, and Odesa. The data relates to the years starting from 2006 with the sample size of 9546 examples. The results of GLM constructing for different distribution laws are presented in Tables 1 – 3.

Table 1 Distributions and link functions used

|  |  |  |
| --- | --- | --- |
| № | Model | |
| Distribution of dependent variable | Link function |
| 1. | Gamma | LOG |
| 2. | Normal | LOG |
| 3. | Poisson | LOG |
| 4. | Normal | Identity |

Table 2 Numerical characteristics of the models constructed

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Total loss, mln grn | Mean | Std. deviance | Min | Max | Mean of standard error | Vari-ance, % |
| **102.008** | **11805.69** | **15358.12** | **6273.87** | **18549.82** | **0.075** | **130.09** |
| 18.111 | 1897.46 | 939.91 | 4010.98 | 634.05 | 0.120 | 49.54 |
| **17.921** | **1877.53** | **1027.57** | **4234.95** | **558.35** | **0.176** | **54.73** |
| 17.921 | 1877.53 | 999.30 | 3535.40 | 118.00 | 224.348 | 53.22 |

The Table 3 shows that the risk of financial loss for the models constructed varies approximately between 40-60% what is marginally acceptable but requires undertaking of some measures regarding its minimization.

Table 3. Statistical characteristics of the models

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Total loss, grn | Log-likelihood | Actual total loss | Difference | Risk |
| **102008320.905** | **-15742.754** | 17921032.58 | **84087288.32** | **1.301** |
| 18111231.380 | -98700.167 | 190198.799 | 0.495 |
| **17921032.574** | **-42173677.24** | **0.007** | **0.547** |
| 17921032.589 | -98700.167 | 0.009 | 0.532 |

Comparison of the models hiring normal distribution with logarithmic and identity link functions leads to results that information Akaike criterion accepts approximately the same value of about 20.78. It means that better model selecting should be based upon the total forecast of possible loss. Thus, adequate model in this study is the one based upon Poisson distribution and exponential link function. This choice is supported by the minimum loss estimation error, maximum approach of forecast to actual data, and true estimate of possible risk of loss. Normal model with identity link function provides for a small loss error of about 1.65% but “weak” forecasts of loss and erroneous risk estimate.

It was shown that generalized linear models can serve as effective instrument of analysis financial and economic processes that helps to take into consideration actual complicated factor interactions and their influence on dependent variable. There also exists a possibility for model and forecasts quality analysis (of the results achieved) using a set of appropriate statistical criteria.

This approach to financial processes analysis will help to minimize financial risks in insurance as well as in many other spheres of human activities. Finally such studies will positively influence macro economy as a whole.

**References**

1. De Gooijer, J. G. (2017) *Elements of Nonlinear Time Series Analysis and Forecasting*. – Cham, Switzerland: Springer, 618 p.
2. Tsay R.S. (2010) Analysis of Financial Time Series. – Hoboken (NJ, USA): Wiley & Sons, Inc., 715 p.
3. Hansen B.E. (2017) Econometrics. – University of Wisconsin, 427 p.
4. Rossi P.E., Allenby G.M., McCulloch R. (2005) Bayesian Statistics and Marketing. – Hoboken (New Jersey): John Wiley & Sons, Ltd, 348 p.
5. Bidyuk P., Terentyev A., Konovalyuk M., (2010) Bayesian networks in data mining technologies // *Artificial Intelligence*, № 2, рр. 104 – 113. Available at: <<http://dspace.nbuv.gov.ua/bitstream/handle/123456789/56141/13-Bidyuk.pdf?sequence=1>>.
6. Zgurovsky M.Z., Bidyuk P.I., Terentiev O.M., Prosyankina-Zharova T.I. (2014) *Bayesian Networks in Decision Support Systems*. – Kyiv: NTUU KPI, 320 p.
7. De Jong P., Heller G.Z. (2008) Generalized Linear Models for Insurance Data. – New York: Cambridge University Press, 197 p.
8. Sugumaran V. (2002) *Intelligent Support Systems Technology: Knowledge Management*. – London: IRM Press, 318 p.
9. Kuznietsova N., Bidyuk P. (2020) Intelligence information technologies for financial data processing in risk management. *Communications in Computer and Information Science* №1158. P. 539–558.
10. Trofymchuk O., et al. (2019) Decision Support Systems for Modelling, Forecasting and Risk Estimation, LAP Lambert Academic Publishing, Riga. 160 p.