Integrated Analysis Technology of Non-Stationary Time Series

Aydın Hasanov[1], Sayyar Abdullayev[2]
[1]Research and Training Center of Information Technologies and Systems
National Academy of Sciences and Ministry of Education, Kiev, Ukraine
[2]Institute of Information Technology of ANAS, Baku, Azerbaijan
[1]*ayding@rambler.ru*, [2,3]*depart5@iit.ab.az*

Abstract— The paper considers the issues of integrated time-series analysis, which are becoming increasingly important with the development and improvement of information processing tools. It also proposes a number of methods for modification and integration in determining and using the accuracy of these methods and the prediction quality obtained by using them in decision making.

Keywords - technology of time series analysis; knowledge base; calculation unit; interface; Matlab

I. INTRODUCTION

Advanced technology of time series analysis, based on an integrated approach is an effective means of improving the quality of models for the study of various processes and the phenomena of the world. Objectives of integrated time series analysis, today, are becoming increasingly important with the development and improvement of information processing tools. Solving financial and economical, technological, technical and other challenges presented by time series, an important step is to obtain the necessary knowledge about the analyzed process for the further development of modeling and predictive system for decision making. In such setting without the process of integration and the usage of interfaces it is impossible to solve these problems effectively. However, researchers are increasingly in need of an effective user interface systems for the analysis of non-stationary nonlinear time series. A special challenge in this case is the analysis of nonlinear and non-stationary dynamical processes. In this regard, a number of modification and integration methods have been proposed in determining and using the accuracy of these methods and the prediction quality obtained by using them for decision making.

At present, there are common approaches to the design of technologies for analysis of non-stationary time series. These include a variety of techniques, in particular, described elsewhere [1,2], various software packages, such as Statistics packet, StatGraf plus for Windows, Eviews, Matlab, etc. [3].

However, many issues related to research and study of basic process properties (it can also be control objects) remain unsolved or studied poorly. Some of these properties to be explored, in particular, include the following knowledge. Knowledge of stationarity (non-stationarity) allows one to select an appropriate model with the addition of a member of non-stationary time series; knowledge of heteroscedasticity, i.e. the changes in conditional and unconditional dispersion of time series allows to take into account prediction value of the model and thereby to increase the adequacy of the model; knowledge of cointegration properties of the analyzed processes (cointegration - when two or more processes are related to each other), i.e., when the long-term time series develop mutually and allow them to adjust the equilibrium state, to improve the accuracy of models that describe them. Modeling of these processes provides model design for two or more time series for long-term equilibrium state and development [4].

To identify and explore all of these properties in time series the tests are carried out in the first stage, which is an important step in the analysis of time series, giving the answer - which adequate model needs which process. The model includes the component in order to improve its statistical, predictive performance, which are adequate for the process. If the research process of the selected model does not provide answers to multiple questions, it is necessary to choose a different approach or method (today there are hundreds of them).

Processing of these methods and the choice of the appropriate method may be delayed in time, and it is not satisfactory. At the same time, during the complex (integrated) and automated analysis, as a rule, the researcher has numerous additional features, not to mentioning their comfort.

II. PROBLEM STATEMENT

To meet the growing requirements of the systems of time series analysis (modeling and prediction) and to make an justified decision, taking account modern information technology, it is necessary to build a comprehensive system of analysis for the problems expressed by non-linearity, non-stationarity and dynamics. An approach is proposed for an integrated analysis, in which the analyzed time series for various properties can be integrated, i.e. most processes of time series analysis end with an answer - which time series require is the best (in terms of adequacy) model.

The technique of integrated time-series analysis

In most cases, this analysis of time series it is necessary to create a knowledge base that includes at least a sufficient set of the investigated models (or their structures), techniques of testing time series for non-stationarity, heteroscedasticity and

co-integration, evaluation criteria, as well as newly acquired knowledge about the process and its properties.

With this approach, most procedures can be combined into a single module (procedural modules) to automate and comfort studies, and last but not least - to reduce the terms of the investigated processes. The system includes all necessary nodes for an integrated analysis of time series in order to make an justified decision on a particular process. After loading the data they are tested and the model for the calculation of statistical and predictive characteristics is chosen.

Applied methods and tools

Analysis of non-stationary time series is traditionally made by Box-Jenkins method, or probabilistic approaches are used for this purpose, Bayesian network, neural network are developed, or an apparatus of fuzzy sets, genetic and immune systems, artificial intelligence, etc. are used [5].

Despite the fact that Box-Jenkins model remains the most popular to this day, out of probabilistic methods for analysis of time series, Bayesian networks, neural and other networks, genetic and immune systems are widely used in recent years [6]. Visual form of analyzed data in the constructed Bayesian networks significantly simplifies the analysis of time series. In addition, Bayesian networks can ensure high-quality forecasts and provide versatility of applying a wide class of processes in the analysis. Logical and production models of knowledge representation, as well as elements of soft computing (fuzzy sets) are the most popular in connection with the development of artificial intelligence.

The procedure of time series analysis starts with the selection of mathematical and predictive models. Statistical characteristics of the model and prediction obtained at this stage are compared with the characteristics of the previous model, which results in the selecting the most appropriate model and corresponding prediction. This procedure, as shown by the analysis of the set of processes described by time series [1,2,4] is the most effective. If it is found out that the chosen model unsatisfactorily describes original time series, the system selects the next model for the analysis. The procedure ends with a satisfactory evaluation of the quality of the forecast.

Training and decision-making system constructed in this way for the analysis includes data base and knowledge base to accommodate modeling and predictive mathematical models, and to make recommendations on further use of obtained results and make justified decisions.

Various software tools, written on the problematic, object-oriented programming languages, such as, Delphi, VBA, Pascal, C + +, Matlab, Eviews, etc. are used for time-series analysis, which can also be used to build the user interface. Using the package Eviews the user interface, unlike the other packages listed above, is a relatively uncomplicated tool of time-series analysis. The package lets you explore the time series for nonstationarity, heteroscedasticity, co-integration. Afterwards, obtained results need to be transferred (e.g. into Excel) in order to display the results in a readable form. The disadvantages of the package is the fact that it is autonomous (closed) system, and does not allow presenting the results in vivid form (in the form of graphs, diagrams, etc.), this procedural part is not provided in the package.

Use of languages Delphi, VBA, Pascal, C + +, Matlab for research of time series for non-stationary, heteroskedasticity and co-integration entails a long and complicated process of writing the corresponding user program and their debug.

The principles of integration of software systems

With an integrated approach it is possible to combine different procedures, including the same type of procedures that can be combined into a single software module. Many procedures can be distinguished in the structure and be automated.

The basic principles of integration are:

- the possibility of complementarity, in which after the completion of one module the results can be transferred to another module (in a different package) for further processing;
- call of subroutines from the total (integrated) user interface system, which can simultaneously involve two or more packages or modules;
- availability of service features of user interface (flexible menus, prompts subsystems, window editor, graphic interface, etc.);
- ability to add and use modern technologies and methods for time-series analysis;
- insurance of high adaptation, adequacy and comfort of the investigated processes;
- insurance of high automation and accelerated development of analysis procedures of poorly trained user contingent.

The software provides: processing of parameter changes in real-time; operational solution output; presentation of information and results in a convenient form for analysis. Integrated environment contains description of all functional features of all modules and modes of interaction.

Logical and production models of knowledge representation, as well as elements of soft computing (fuzzy sets) are the most popular in connection with the development of artificial intelligence systems.

Integrated system can provide the following functional (basic) blocks:

The database (DB) may include: initial and intermediate data of the solved problem, a list of processes and objects, and other information.

The information derived from the experimental study of objects, processes and phenomena of studied problem field is called the data. The data passes through the following types of processing: experimental data of measurements and observations; data on material carriers; data models (structures) in the form of diagram, graphs, functions; computer data, described in the system language.

Knowledge Base (KB) is a warehouse of knowledge about the properties and regularities of the processes, and it forms the basis of an integrated system. It contains arranged set of studied models and methods derived from the use of professional expertise of specialists (experts); redeveloped process models; the tests studied on the proper models of the analyzed processes and their models; estimation methods in search of the best model structures through the variety of statistical quality criteria of the models; new knowledge about the properties and regularities derived from experimental studies, as well as from the publications and other sources. KB is periodically added extracted knowledge about that which process needs which adequate model, and which method works best in which circumstances.

Computing unit (CU) is the major component of the system. CU draws time series from the DB and the tests and sequence of corresponding models from the KB, and it analyzes data in off-line mode, and transfers the data to the user interface if necessary. It allows to get solution of the given problem or to initiate new activities or user to be in standby mode. Integrated CU, along with traditional tools, uses the tools of soft computing, fuzzy knowledge, genetic algorithms, neural and probabilistic (e.g., Bayesian) networks. The combination of different methods of problem solution and models improves the mobility of CU and provides required quality of problem solution and decision making.

Editors of KB and DB are computer programs that provide conditions to create databases in a dialogue mode. The editors contain a system of submenus, templates, system prompts and other service facilities that make it easier to work with databases.

User Interface (UI) is a software device that implements the user dialogue with a system, both at the stage of information input and at the stage of obtaining results and decision making. The dialogue is realized in a convenient form using a variety of window menus of selection mode, and enabling a specific software package to perform data analysis. Depending on the operation mode, UI can be used as follows: an interface with an unmanaged scenario assumes the user interaction with an integrable system in the framework of directives; hard-coded script, determined by the integrable system itself, and the user gets the right to choose only in certain fragments of dialogue; a flexible script allows to create own interaction mechanism with an integrable system or to adjust the system script.

A user is a person making justified decision. With the help of mode selection unit it selects research mode, estimates the intermediate and final results. If necessary, it passes (starts) data (task) through the UI to the software packages for the further analysis for performance.

Software Packages (SP) contain standard software packages that are essential for time series analysis. They receive start directives of SP selection to accept imported data. After the problem solution, the data or the results are exported back into the Matlab system.

Let's examine some practical solutions of user problems arising from the integrated (complex) approach.

Interactions between Matlab and Eviews in the analysis of non-stationary time series.

The package Eviews, as we mentioned, unlike the Matlab package does not support the possibility of cross-program interaction with the use of ActiveX (com) or DDE (Dynamic Data Exchange) [7]. On the other hand, Eviews and Matlab have the facility of data import and export in various formats, such as text, Microsoft Excel, CSV formats, etc. The main program package Matlab starts to perform required procedures for the data analysis, produces the necessary calculations and renders the results. At the calculation moment, when it is required to transfer a part of the data to Eviews for further processing, Matlab program generates an ASCII file with a text of processing program in a dynamic mode. Afterwards, it records the data to a text file with a name that is specified to be imported into the developed Eviews program. Then, using the command dos (path "Eviews\eviews6.exe\program name"), Matlab enables Eviews to perform the developed program. After completion of Eviews data processing, if necessary, using the export it returns the needed results to the Matlab program in a similar way. Hereby, the main program in Matlab runs Eviews periodically, giving it instructions for processing in a program file and the data in text file format, after which waits for the end of Eviews and result export of the operation. Then, it continues the operation, possibly again, sometimes in accordance with the above described algorithm, referring to Eviews.

The procedure of process analysis continues as long as the knowledge is obtained, which indicates the model selection adequate to the process and satisfying the user. For the integration of multiple tasks of processing and analysis in a convenient interactive shell, it was used a powerful graphical Matlab interface (GUI). The foremost processing tasks were implemented as separate modules that are created using the programming language Matlab. The main program created a graphical interface with the buttons to call processing modules, and with the fields of input and changes in key parameters, and with the windows and fields for entering graphical and numerical information. After that, the main program enters a standby cycle of user action. Clicking on the button or inputting the parameter, corresponding function of event processing was called (Callback function). Eventually, convenient integrated environment for decision making in the analysis of non-stationary time series is achieved, and unlike the other systems it allows automating the analysis procedure, improving the efficiency and comfort.

Most of the procedures of data analysis from the Eviews software package are performed manually. For example, significant coefficients should be selected on the autocorrelation table and be included in the regression model. Then, for the resulting model it is necessary to calculate the statistical and forecast characteristics for decision-making in accordance with the appropriate criteria [2].

III. CONCLUSION

The main difference of the proposed system lies in the fact, that to give the answer to recalled task, what is the most effective method to be applied to analyze the task, is providing

the greatest effect from the point of automating the analysis process and comfort.

REFERENCES

[1] Bidyuk P.I., Hasanov A.S. Modelling and forecasting of heteroscedastic processes // Cybernetics and Computing technique.-2005, Vol.146, pp. 61-80. (Russian)

[2] Bidyuk P.I., Balkan I.B., Hasanov A.S. Systematic approach to the analysis of the adequacy of nonlinear time series models // Cybernetics and System analysis., 2003, № 3. pp. 147-158. (Russian)

[3] Smolentsev N.K. MATLAB: Programming in Visual C #, Borland, Jbulder, VBA: training course. - M.: DMK Press, St. Petersburg. Peter, 2009. pp. 464. (Russian)

[4] Hasanov A.S Co-integration modeling of time series // Cybernetics and Computing technique.-2003, Vol. 140, pp. 54-62. (Russian)

[5] ZaychenkoY.P., GritsenkoV.I., Gasanov A.S. Fuzzy group method of data handling in the problem of structural identification of process rolling of pipes // Proc. 1 International Conf. "Control and Optimization with Industrial Applications" (COIA –2005) / Baku (Azerbaijan), 2005.– P. 94-95.

[6] Bidyuk P.I., Povoroznyuk A.N., Hasanov A.S., Slobodenyuk O.V. Баєсівська ідентифікація динамічних систем // Proceedings of the Int. Conf. "Automation-2008" Odessa 23-26 September, 2008. pp. 697-700.

[7] Bidyuk P.I., Hasanov A.S., Vavilov S.E. International Workshop "Problems of decision making under uncertainties" (PDMU-2008).- Crimea (Novy Svit), Ukraine. pp. 56-58.

[8] Hasanov A.S., Vavilov S.E., Belyaev S.V. International Workshop "Problems of decision making under uncertainties" (PDMU-2009).- Skhidnytsia, Ukraine. pp. 80-81.