DOI: 10.18372/2310-5461.57.17439
UDC 519.766.4

Nick A. Vinogradov, Doctor of Sciences (Techn.), Prof.
National Aviation University
orcid.org/0000-0002-7971-9901
e-mail: vl43radar@gmail.com;

Anastasie V. Lesnaya, Post-graduate student,
Kyiv National Economic University named after Vadym Hetman
orcid.org/0000-0001-8536-7659
e-mail: prkom@kneu.edu.ua;

Iliya D. Savinov, Post-graduate student,
Open International University of Human Development "Ukraine"
orcid.org/0000-0001-7561-7756
e-mail: office@uu.edu.ua

PROBABILISTIC MODELS AND METHODS OF REGRESSION ANALYSIS OF VOLATILE FINANCIAL TIME SERIES

Introduction

The article analyses the interaction of the participants of the global financial and economic market, including individual, institutional producers, suppliers, traders, transnational financial institutions, as well as supranational regulators of the global market in conditions of uncertainty. The main factors of the uncertainty of the world trade market, which affect the interaction of market participants, as well as specific forms and methods of overcoming this uncertainty with the use of modern financial and information technologies, are studied.

Uncertainty is an immanent property of the market, which is cyclically manifested in unpredictable fluctuations in the world economic and financial situation. In conditions of growth in cross-border flows of loan capital, direct and portfolio investments, which form the total international turnover of financial resources, which is accompanied by an increase in the number and variety of financial instruments and the internationalisation of the activities of banks, investment and insurance companies, pension funds and other financial institutions, as well as stock exchanges, this feature of the financial market becomes a key factor in global financial instability.

Market uncertainty consists of two interrelated groups of factors. The first group of factors is determined by a stable state of information asymmetry inherent in trade relations, that is, by an uneven distribution of information, which may depend on the decision to enter into an agreement and its outcome, among market participants. The second group, related to the irrationality of the behaviour of market participants, represents the most important factor of market uncertainty, moreover, this behaviour does not have any proven forms of state regulation or self-regulation and is extremely difficult to analyse and measure.

As markets develop, both groups of factors expand, their influence increases, despite the development of financial institutions and tools designed not only to rationalize the decision-making process, but also to minimize the negative impact of information asymmetry on this process. At the same time, both groups of factors influence each other.

On the one hand, the study of information asymmetry together with the irrational behaviour of investors forms the basis of the analysis of the development of the modern world market. However, on the other hand, these factors create additional uncertainty, which, in turn, makes it impossible to solve the problem of decision-making under the conditions of not even complete, but at least acceptable a priori information.

Therefore, modern methods of financial and economic analysis are based on the principles of mathematical statistics. Currently, the methods of mathematical statistics are widespread in scientific and technical and scientific and economic research. The study of qualitative, and even more so, quantitative regularities of random processes and phenomena, as well as objects that depend on random factors, is impossible in principle. Real phenomena, processes and objects are subject to the influence of numerous external and internal actions. These actions in the vast majority of cases are random in nature and cannot be described or simply taken into account by deterministic methods.

Statistical relationships between variables studied in economics and finance are no exception. Usually, these connections are very complex and ambiguous, depending on many intermediate direct and reverse actions, which, in turn, are also random. With the use of mathematical statistics methods, it is possible to identify patterns of relationships and mutual impacts of individual economic indicators that appear for a given set of observations or measurements. Using the methods of non-parametric statistics, it is possible to obtain statistical regularities of distributions of values of economic indicators (probability densities, initial, central and mixed moments of statistical distributions).

In the economy, organization and production management, the problem of identifying the impact of key indicators on global technical and economic indicators (GTEI), such as profitability, labour productivity, cost of production, etc., is very relevant. At the same time, the number of key indicators is at least an order of magnitude greater than the number of global technical and economic indicators. In a certain sense, the problem of assessing the impact of key indicators on GTEIs is reduced to the processing of large arrays of random data. Therefore, with rare exceptions, in order to identify the impact of partial factors on global technical and economic indicators, it is necessary to solve three related problems:

- collection of statistics, i.e. observations, surveys, measurements and accumulation of a sample of observations;
- isolation of the so-called "sufficient statistics" - reduction (sometimes very significant) of the sample volume without loss of information that is of fundamental importance for statistical analysis;
- actual processing of statistical data with obtaining qualitative and quantitative assessments and justified statistical decision-making.

As a rule, correlation and regression analysis is used to solve these problems. It should be noted that due to the complexity of relationships, the large number of active factors and their interdependence, widely used methods of regression analysis often turn out to be unacceptable and require improvement.

In economics, the task of compact presentation of information, selection of main factors and connections, and screening of secondary ones also often arises. In mathematical statistics, there is a set of methods widely used and known as factor analysis. This direction of statistical research quickly gained recognition due to its advantages compared to other methods and the effectiveness of work where other methods were unacceptable. The solution of the mentioned tasks in the presence of a large number of interrelated factors turns out to be the most effective due to the application of factor analysis based on the method of principal components and orthogonal functions.

The limitations of most known methods and approaches to assessing the value of internationally traded assets became particularly evident after impossibility of the largest transnational financial institutions to predict financial crises and take preventive measures to protect against growing risks. The crisis events exacerbated the problem of the ratio of rational and irrational motivations driven by financial market participants. The danger of underestimation of the irrational economic behaviour of institutional and individual investors in the financial markets had been predicted still in Adam Smith's fundamental works [1,2]. However, the most theories and concepts of the so-called economic mainstream (primarily neoclassical economic thought) raised the question of the demand for research into non-economic factors that shape the uncertainty of the trade and financial market.

The article discusses the method of building mathematical models of heteroscedastic processes and its application to the description of the dynamics of discrete time series [3,4].

In the transitional periods of the state's life, during the development of new economic and financial models, radical changes in the economic and social strategy, global changes of a critical nature may take place, which will affect the fundamental principles of the functioning of entire industries. The peculiarities of such events are, firstly, their unexpected appearance against the background of the seemingly relatively smooth and slow flow of observed processes. Secondly, it is the transience of such processes, in fact, a sudden disturbance, and an outburst.

Problem Statement and relevance

The interaction of the participants of the global financial and economic markets in conditions of full certainty actually is ideal, non-realistic construction, which never can be reached. As a fact, the process of interaction in the sphere of human relations is accompanied by many random impact factors, both internal and external ones. Firstly, these impact factors have complex nature: some factors are pure economical, another factors may have mixed humanitarian and economical nature. Secondly, some impact factors are mutually dependent at that the dependences are subject to sporadic variations. At last, we often face unexpected sudden appearances of new factors and disappearances of existing ones.

All these considerations lead to necessity of application statistical approach to analysis and estimation processes in economics, i.e., to econometric methods [3]. Real samples of observed series values, including financial time series [3], are sequences of random values with appropriate statistical characteristics.

Financial time series, as a sequence of sample values obtained by collecting statistics, most often represent non-stationary random processes. As has been repeatedly noted, in economics and econometrics, the concept of heteroscedastic, i.e. weakly non-stationary, process is often used [4]. However, the concepts of non-stationarity in general and weak non-stationarity in particular also need to be specified. Moreover, financial time series are a typical representative of discrete time series. Heteroscedastic processes as processes with time-varying dispersion, in the mathematical sense, belong to the class of non-stationary processes. This is a class of financial and economic processes that are widely used, especially in an unstable economy. Non-stationary processes belong to the class of processes with slow changes in statistical characteristics over the observation interval. Therefore, the description of their dynamics, the development of methods of approximation, interpolation and extrapolation has certain specificity in comparison with stationary (both in the broad and narrow sense) processes.

Thus, the research and development of mathematical models of non-stationary financial time series are useful and relevant for many branches of economics, econometrics, and financial analysis.

Analysis of recent research and publications

The exclusive relevance of methods of statistical analysis of financial time series with essentially non-stationary (heteroscedastic) nature appeals the attention and great effort of specialists in theory and practice of researching and estimation financial and economical time series [3, 4, 7 – 9, 10]. In reality, deterministic time series of any origin (physical, economic, climatic, seismic, etc.) are only some ideal construction that gives a very approximate idea of the actual state of the system.

Determining the stationarity of a random process is one of the key points in probability theory. It is given in many monographs on statistical theory. We will only remind that in applied statistics and in econometrics, the concept of stationarity is most often used in a broad sense [3]. Also, a process that is stationary in the narrow sense is a process that is stationary in the broad sense (but not vice versa).

For these and other reasons, statistical modelling faces a common problem: in many, if not most, situations, they are forced to exclude variables that are relevant to forecast frequency and loss size. Since these "missing variables" are correlated with both the target variables and one or more other modelling variables, the authors [6, 8, 10] investigate the estimates of the corresponding model parameters. This phenomenon, as a rule, is known as "exclusion of a shifted variable" as a factor in creating specific non-stationarity – volatility of the market, currency pairs, shares, etc. This circumstance forces the development of special evaluation and analysis methods, each of which is acceptable only for a separate (and rather narrow) class of non-stationary random processes [4]. Before everything, it is necessary to introduce a hypothesis about the specific nature of non-stationarity of the process under study.

The material of the article is taught in accordance with the modern requirements of the scientific and experimental approach in the field of economics, information transfer and related disciplines. Another task that often occurs in economics, organisation and production management is the classification of objects according to their characteristics in order to identify "bad" and "good" ones. As a rule, such a classification consists in the grouping of objects in the space of features. Then, based on the involvement of additional information or previously known information, groups of "good" and "bad" objects are distinguished. This stage is called training, and it allows you to develop an algorithm for determining whether other similar objects belong to the "bad" or "good" class.

So the goal of article is the analyses the interaction of the participants of the global financial and economic market, including individual, institutional producers, suppliers, traders, transnational financial institutions, as well as supranational regulators of the global market in conditions of uncertainty. The main factors of the uncertainty of the world trade market, which affect the interaction of market participants, as well as specific forms and methods of overcoming this uncertainty with the use of modern financial and information technologies, are studied.

Forecasting non-stationary processes in the economy

Methods of specifying econometric models.

The classic "material" for creating dynamic (deterministic) mathematical models are differential equations. If it is simplified to consider financial time series as deterministic, then the equation of dynamics is an actual standard of a dynamic system – a mathematical model that allows you to predict the evolution of an object based on a given initial state.

Obviously, the comprehensive characteristic of any random process is the multidimensional law of probability distribution [5]:

$$F_n\left[(x_1,t_1),(x_2,t_2),\ldots,(x_n,t_n)\right]= \\ =P\left[X(t_1)<x_1,X(t_2)<x_2,\ldots,X(t_n)<x_n\right]. \quad (1)$$

However, it is not so obvious that the results of the analysis of probability distributions above the first or second order cannot serve as a more or less reliable basis for obtaining convincing statistical conclusions. First, the determination of multivariate probabilistic characteristics is associated with great computational difficulties. Secondly, very uninformative characteristics of the function $F_n(x_1,x_2,\ldots,x_n)$ are described by the central or initial moments of orders above the fourth; in practical problems they can be excluded from consideration without any loss of commonality. The increase in the amount of information when using probabilistic characteristics of a higher order turns out to be insignificant (for engineering applications – very insignificant), primarily due to the rapid deterioration of accuracy. The accuracy of the calculation of the higher moments of the distribution is worse, the higher the moment. It may happen that with a limited sample size, the informational contribution of higher moments will generally become negative, starting from some *i*-th moment. In addition, the definition of multidimensional probabilistic characteristics is associated with great difficulties in the hardware implementation of algorithms for their calculation. Finally, for samples of a small volume, it is impossible to obtain reliable estimates for moments above the second at all [6].

Therefore, in practice, as a rule, one-dimensional or two-dimensional distribution laws of a random process are considered. Let us give an expression for the two-dimensional distribution law of a random process $X(t)$:

$$F_n(x_1,t_1;x_2,t_2)= \\ =P\left[X(t_1)<x_1,X(t_2)<x_2\right], \quad (2)$$

where x_1, x_2 are numerical values of the process at moments of time t_1, t_2, respectively. They provide sufficient information about the parameters and state of the process, as they contain a sufficient amount of information about the properties of the random process. The increase in the amount of information when using probabilistic characteristics of a higher order turns out to be insignificant and does not justify the efforts spent on collecting and processing the experimental data. Moreover, possible calculation errors that lead to loss of information can often compensate for the increase in information that the researcher expects [6].

Non-stationary time series, in particular, time series of the processes of the real economy, are characterized by a two-parameter impulse characteristic $h(t,\tau)$, and in the case of time series with time-constant parameters, simply $h(t-\tau)$. A distinctive feature of this description is that the input message is considered known on the interval $-\infty<t<\infty$. Often this method of description is the most convenient. (Note that we mean by the term "message" any input action on the system: a request, numerical reference data, a change in the conditions of interaction of business entities, the results of making a decision, etc.)

The task of forecasting the future values of a time series based on its current and previous values is the basis for planning in finance, economics and trade, management and optimisation of production volumes, warehouse control.

The task of forecasting a time series is solved as a task of mathematical statistics based on the development of a forecast model that adequately describes the researched process. At the same time, it is necessary to make the following observations of a principled nature.

1. In modern statistical science, there is a division into two main schools: the most numerous classical school - the followers of Fisher and his students, as well as the subjectivist, or Bayesian, school. And although at the level of applied statistics, the results obtained within the framework of these different scientific schools agree quite well, on a wide range of theoretical and philosophical issues, these two directions often diverge, offering different approaches to solving problems, including in the field of econometrics. Briefly, the main difference in the approaches could be characterized as follows: supporters of the classical approach consider the frequentist interpretation of probability to be the only possible one (that is why such an approach is also called the "frequentist school"). The essence of their approach is that they start solving the problem with the selection of a model and check whether this model can "explain" the obtained (or even more "extreme") data.

The difference of the Bayesian approach is that even before the data are obtained, the statistician considers the degree of confidence in various possible models based on the obtained data (prior probabilities). Once the data is obtained, Bayes' theorem allows the calculation of a new set of probabilities that represent revised degrees of confidence in the possible models based on the obtained data (posterior probabilities). The assessment of a priori probabilities is subjective; therefore, this approach is called subjectivist.

The subjectivist, or Bayesian, theory of statistical decisions can be applied in those tasks in which the uncertainty or information about the parameters at any moment in time can be given by means of a probability distribution on a set of its possible values. Therefore, we will consider only those decision-making tasks that satisfy the following two requirements.

a) The conditions of the task can be described using such a number of parameters that can be taken into account.

b) Although we don't assume the value of these parameters to be known with complete accuracy, the degree of uncertainty of these values can be described using the appropriate probability distribution.

2. In connection with these requirements, the following two reasons should be kept in mind. First, the number of parameters that can be taken into account in this situation depends to a large extent on the current state of computing equipment and technology (obviously, today this number is more than ever). Secondly, among statisticians and other specialists who study the basics of probability theory, there are significant disagreements about whether it is possible to characterize the uncertainty in the values of a given parameter using a probability distribution. Some believe that one can talk about such a distribution only in the case when the relative frequencies of occurrence of the investigated values of the parameters are clearly observed. Others take the view that probability is a logical concept that can be related to parameters in a much wider class of cases. Moreover, these scientists believe that in each problem there is a single distribution described in a certain parameter, which should be assigned to it. Finally, others believe that, in fact, all probability distributions are subjective by definition and that in statistical studies, with unknown parameter values, the statistician can always characterise the range of his lack of knowledge of truthful value through conditional probabilistic distribution of $p(x|a)$ type, where a is unknown parameter.

By the way, mentioned differences in opinions are scholastic in the valuable range. Actually all the reasonable methods works good (and almost with the same efficiency) when we have sufficient large stored information. In other case different methods give, respectfully, different recommendations, but all are poor. So we can choose the most comfortable method, e.g., Bayesian, and acceptable probabilistic distribution without any responsibility, and compare it with other probabilistic distributions, e.g., by information an entropy criterion. As it's known, when the finite interval of random values the uniform distribution has the maximal entropy. It is fully logical that even uniform distribution has satisfactory result of estimation then the other distributions will be not worse as minimum. Uniform distribution, as it were, is of usual minimax property.

We can check the fact of smoothing different methods in the frame of Bayesian theory by researching the tolerance of results (asymptotic invariance) to the choice of prior distribution. "Prior problems" are overcome dialectically. The prior data uncertainty is not argument for refuse from prior distribution. Moreover, it's argument in favour of multiple dissolving problems under different prior distributions. The same reasons are fair for uncertainty dynamic equations, quality criteria etc.

3. The choice of "the best" method of statistical estimation depends both from presence of prior information and from cost function (criterion function, utility function). The minimum risk of error decisions with quadratic cost function corresponds to least mean square (LMS), i.e. it is the average of conditional variations on all observations. Minimal cost function by the module is the median of posterior probability density. For all that with Gaussian prior density all cost functions lead to the same estimations.

From the other side, minimal risk is achieved with estimation by the maximum likelihood method (MLM). MLM is the limit case of estimation by the method of maximum posterior density when the value of prior information aspires to zero.

4. MLM estimates are consistent (the decisions converge to precise value with aspiring of observation number to infinity). MLM also are asymptotically efficient and asymptotically Gaussian. All these properties are concerned with estimate behaviour with unlimited enlarging of observation number.

Thus MLM is the most appeal with the reasons of simplicity and comfortable application low sensibility to variations of prior data and observed current parameters.

Let consider the problem of application MLM to researched problem. Now the most known models of prognosis time series are:

– regression and auto regression models;
– neuro network models;
– exponential smoothing models;
– Markov chains models;
– classification models etc.

The most popular and widely using models are exponential smoothing models and auto regression models [7,8]. Let's consider the case of auto regression model with absence external impact factors. This simple case may serve as basis for researching more perfect and realistic methods.

V. Forecasting without taking in account external impact factors.

Let the values of time series are available in discrete moments of time $t_i = t_1, t_2, \dots t_n, \dots, t_N$. We mean time series as $\xi(n), n_i = 1,2,\dots,N$. It's necessary to define in moment t_N future values of process $N+1, \dots, N+M$. Moment N means the beginning of prognosis, and M is time of prediction.

1) We need to define the functional dependence of communication between fast and future values of time series for calculation of future values of this series

$$\xi(n) = \psi\left[\xi(n-1), \xi(n-2), \dots, \xi(n-M)\right] + \varepsilon(n). \quad (3)$$

The dependence (3) has name "the forecasting model". We need to choose from K available models prognosis model $\xi_j(n), j = 1,2,\dots,K$, with minimal module of average deviation of true value from predicted one for established M:

$$\overline{\mathrm{E}} = \sum_{n=N+1}^{N+M} \left|\varepsilon(n)\right| \to \min_{\xi_j(n)}.$$

We can write the expression (3) as

$$\hat{\xi}(n) = \psi\left[\xi(n-1), \xi(n-2), \dots, \xi(n-M)\right], \quad (4)$$

where time series $\hat{\xi}(n)$ represents predicted (calculated) values of $\xi(n)$ series. Here and later we use the symbol "$\wedge$" for meaning calculating values of time series.

2) We also need to calculate the confidence interval of possible deviations of values of the series

$$\hat{\xi}(n) = \psi\left[\hat{\xi}(n+1), \hat{\xi}(n+2), \dots, \hat{\xi}(n+M)\right]. \quad (5)$$

The task of forecasting time series $\xi(n)$ is illustrated on fig. 1.

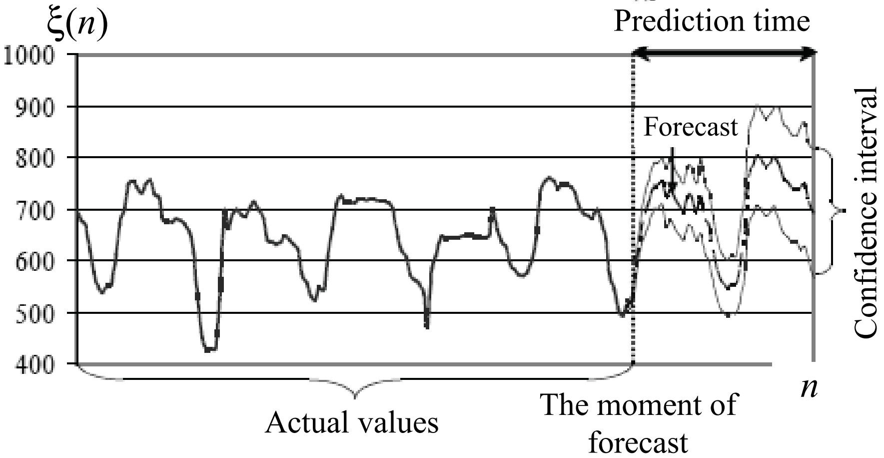

Fig. 1. The illustration of the problem of prediction of time series $\xi(n)$ without consideration impact factors

Auto-regression model

The background of auto-regression models is the proposition about linear or at least parametrical dependence of process $\xi(n)$ current value from some number of previous values that process $\xi(n-1), \xi(n-2), \dots, \xi(n-s)$. Let consider the most popular and useful model of auto-regression and moving average (ARMA). This model expresses current value of time series as finite sum of previous values and discrete Gaussian white (δ-correlated) noise:

$$\xi(n) = a_0 + b_1\xi(n-1) + b_2\xi(n-2) + \dots + b_p\xi(n-p) + \varepsilon(n). \quad (6)$$

Expression (6) describes p-order auto-regression $AR(p)$ process; here a_0 is constant component, $b_1, b_2, \dots b_p$ are auto-regression coefficients, $\xi(n)$ is model discrepancy conditioned by observation noises. They use the least square method or maximal likelihood method for definition coefficients a_0 and $b_i, i = 1,2,\dots,p$ [7].

Another model type is q-order moving average MA (q) model; the MA equation is

$$\xi(n) = c_0\eta(n) + c_1\eta(n-1) + c_2\eta(n-2) + \dots + c_q\eta(n-q) + \varepsilon(n), \quad (7)$$

where $c_0, c_1, \dots, c_q$ are moving average coefficients, $\varepsilon(n)$ is model discrepancy.

We combine auto-regression and moving average models (ARMA models) for achieving more flexibility in model matching. Resulting model ARMA (p, q) includes q-order MA filter and p-order AR filter.

If not the values of the time series themselves are used as input data, but their d-order difference (in practice in most cases $d \le 2$), then the model is called auto-regression and integrated moving average – ARIMA (*p*, *q*, *d*) [7].

A development of the ARIMA (*p*, *q*, *d*) model is the ARIMAX (*p*, *q*, *d*) model, which is described by the equation

$$\xi(n) = AR(p) + a_1\phi_1(n) + \ldots + a_q\phi_q(n).$$

A development of the ARIMA (*p*, *q*, *d*) model is the ARIMAX (*p*, *q*, *d*) model, which is described by the equation

$$\xi(n) = AR(p) + a_1\phi_1(n) + \ldots + a_q\phi_q(n).$$

Here $a_1, a_2, \ldots, a_q$ are the coefficients of external impact factors $\phi_1(t_i), \phi_2(t_i), \ldots, \phi_q(t_i)$. In this model, the process $\xi(n)$ is the result of processing by the MA(*q*) filter, that is, the filtered values of the original process. For forecasting $\xi(n)$, an autoregressive model is used, in which additional regressors of external factors $\phi_1(n), \phi_2(n), \ldots, \phi_q(n)$ are introduced.

The autoregressive conditional heteroskedasticity (GARCH) model is the residual model for the AR(*p*) model. At the first stage, the AR(*p*) model (6) is determined for the initial time series. It is assumed that the discrepancy of model (4) has two components:

$$\varepsilon(n) = \sigma(n) \times \varsigma_n,$$

where $\sigma(n)$ is the time-dependent standard deviation (SD); ς_n is a normally distributed random variable with zero average and standard deviation equal to 1.

The dispersion $\sigma^2(n)$ is described by the equation

$$\sigma_n^2 = \alpha_0 + \alpha_1\varepsilon_{n-1}^2 + \ldots + \alpha_q\varepsilon_{n-q}^2 + \\ + \beta_1\sigma_{n-1}^2 + \ldots + \beta_p\sigma_{n-p}^2. \quad (8)$$

Here $\alpha_0, \alpha_1, \ldots, \alpha_q$ and $\beta_1, \beta_2, \ldots, \beta_p$ are the coefficients of the conditional heteroscedastic model equation.

This model has received the most frequent use in financial and economic analysis, where it is used to model the volatility of discrete time series.

It should be additionally noted that for none of the considered groups of models (and methods) the accuracy of forecasting is not indicated in the merits. This is due to the fact that the accuracy of forecasting a particular process depends not only on the type of model, but also on the experience of the researcher, on the availability of data, on the available hardware resources, and many other factors [9]. Forecasting accuracy will be evaluated for specific tasks to be solved as part of the work.

The most common forecasting models for homoscedastic processes are autoregressive and moving average (ARMA) models, and for heteroscedastic processes are autoregressive and integrated moving average (ARIMA) models [4].

Also we note that unlimited enlarging of models order require the significant complication of models but don't give noticeable accuracy of process interpolation. Moreover, the errors of extrapolation grow very quickly with enlarging model order more than 2…3. Thus the principal useless of tendency to non-critical enlarging of order of moving average and regression equations takes place.

Conclusions

Empirical research in the field of economics increasingly uses data at the individual or household level obtained from surveys. Unlike aggregate data based on surveys, where errors can be hoped to "cancel out", the transition to micro data requires constant concern about measurement errors as a possible source of bias. Some variables (transfer income, wealth funds, health care use, and expenditures) are difficult enough to measure that such problems arise even when estimating simple bivariate regressions; others (unionisation, schooling, and perhaps earnings) that appear to be messages with reasonable precision become candidates for concern when panel data are used in ways that effectively distinguish much of the true change while adding to the noise.

Many of the parameter validation studies of financial and economic time series examined in this article are based on under-sampled statistics. Even when validation data refer to a significant proportion of a larger survey, concerns about representativeness are difficult to dismiss.

The article attempts a preliminary survey and analysis of time series data for the specification of a model of the interrelationship of variables. It should be recognized that the practical implementation of the above rules is not trivial. In particular, it is obvious that it is possible to obtain satisfactory estimates of the spectrum of financial and economic time series, but at the moment it is not clear how to quantitatively estimate volatility values, cooperation processes in conflict conditions, etc. Only further analysis, both theoretical and empirical, can provide answers to these questions.

Many important economic variables show non-stationary behaviour. The difficulty lies in the correct interpretation of the form of non-stationarity. After extending the linear approach, it is necessary to solve the problem of specifying non-linear or generalized linear models. As shown in the article, there are many non-linear models, and there is no clear way to decide which non-linear specification is best. This problem is quite important and controversial, since using an insufficiently correct non-linear specification may be more appropriate than ignoring the non-stationarity. In addition, linear model can always be considered as a local approximation of a non-linear process. There are some standard guidelines for evaluating a non-linear process. The most important thing is to use a specific general modelling strategy.

REFERENCES

[1] Smith Adam, Cannan Edwin. The Wealth of Nations. Modern Library; Sixth Printing edition, 1994. Reprint of 1880 George Bell Publishers. – 1200 pp.

[2] Smith Adam. The Theory of Moral Sentiments. 2011. Reprint of 1790 London Edition Gutenberg Publishers. 538 pp.

[3] Tsay R. S., Analysis of Financial Time Series. 3rd ed. John Wiley & Sons, Inc., Hoboken, New Jersey, 2010. 677 p.

[4] Korolyuk, V., Limnios, N., Mishura, Y., Sakhno, L., & Shevchenko, G. (Eds.). (2014). Modern Stochastics and Applications. Springer Optimization and Its Applications. – Springer International Publishing Switzerland 2014. XVII, 349 p.

[5] Papoulis A. Probability, Random Variables, and Stochastic Processes, 4th Ed. / Athanasios Papoulis, Unnikrishna Pillai. – McGraw-Hill, New York, NY, 2002. 852 p.

[6] Bendat J.S. Random Data: Analysis and Measurement Procedures, 4th Ed.\ Julius S. Bendat, Allan G. Piersol. – John Wiley & Sons, Inc., Hoboken, New Jersey, 2010. 640 p.

[7] George.E.P. Box, G.M. Jenkins, G.C. Reinsel, G.M. Ljung Time series analysis: forecasting and control, Fifth ed. John Wiley & Sons, Inc., Hoboken, New Jersey, 2016. 669 pp.

[8] Mosteller F., Tukey J. W. Data Analysis and Regression: A Second Course in Statistics. Pearson, 1977. 588 p.

[9] Anderson T.W. The Statistical Analysis of Time Series. Wiley, Hoboken, 2011. 704 p.

[10] Afifi A. A., Azen S. P. Statistical Analysis, Second Ed.: A Computer Oriented Approach 2nd Edition. Academic Press; 2nd ed., 1979. 442 pp.

Віноградов Н. А., Лєсна А. В., Савінов І. Д.

ІМОВІРНІСНІ МОДЕЛІ ТА МЕТОДИ РЕГРЕСІЙНОГО АНАЛІЗУ ВОЛАТИЛЬНИХ ФІНАНСОВИХ РЯДІВ

У статті розглянуто фінансово-економічні часові ряди у виробничій, банківській та інвестиційній галузях. Емпіричні дослідження в галузі економіки все частіше використовують дані на індивідуальному рівні або на рівні домогосподарств, отримані в результаті опитувань. Деякі змінні настільки важко виміряти, що такі проблеми виникають навіть при оцінці простих двофакторних регресій; коли панельні дані використовуються таким чином, що ефективно виділяють більшу частину справжніх змін, додаючи шуму. Результати аналізу реальної інформації дають підстави припускати, що найбільш адекватними математичними моделями нестаціонарних фінансових часових рядів є гомоскедастичні та гетероскедастичні ймовірнісні моделі з частково невідомими імпакт-факторами. Ми пропонуємо моделі авторегресії та ковзного середнього (ARMA) для аналізу гомоскедастичних рядів та моделі авторегресії та інтегрованого ковзного середнього (ARIMA) для аналізу гетероскедастичних рядів. Ці моделі охоплюють досить широкий клас випадкових процесів, нестаціонарних у широкому та вузькому розумінні. Правильний вибір порядку моделей дозволяє отримувати результати з допустимими похибками (розбіжністю) за досить простими моделями. Ми показали принципову марність тенденції до некритичного розширення порядку ковзного середнього та рівнянь регресії. Крім того, модель значно ускладнюється, а похибки екстраполяції, відповідні прогнозу, зростають дуже швидко. У статті зроблено спробу попереднього огляду та аналізу даних часових рядів для специфікації моделі взаємозв'язку змінних. Слід визнати, що практична реалізація наведених вище правил не є тривіальною. Зокрема, очевидно, що можна отримати задовільні оцінки спектру фінансово-економічних часових рядів, але наразі незрозуміло, як кількісно оцінити значення волатильності, процеси взаємодії в умовах конфлікту тощо. Лише подальший аналіз , як теоретичні, так і емпіричні, можуть дати відповіді на ці запитання.

Ключові слова: фінансовий часовий ряд, гомоскедастичність, гетероскедастичність, регресія, авторегресія та ковзне середнє.

Vinogradov N., Lesnaya A., Savinov I.

PROBABILISTIC MODELS AND METHODS OF REGRESSION ANALYSIS OF VOLATILE FINANCIAL TIME SERIES

The article considers financial and economic time series in production, banking, and investment branches. Empirical research in the field of economics increasingly uses data at the individual or household level obtained from *surveys. Some variables are difficult enough to measure that such problems arise even when estimating simple bivariate regressions; when panel data are used in ways that effectively distinguish much of the true change while adding to the noise. Results of analysis of real information give the reasons to suppose that the most adequate mathematic models of non-stationary financial time series are homoscedastic and heteroscedastic probabilistic models with partially unknown impact factors. We propose the auto regression and moving average (ARMA) models for analysis of homoscedastic series and auto regression and integrated moving average (ARIMA) models for analysis of heteroscedastic series. These models cover rather wide class of random processes, which are non-stationary in wide and narrow sense. Correct choice of model order allows getting the results with acceptable errors (discrepancy) using rather simple models. We showed the principal useless of tendency to non-critical enlarging of order of moving average and regression equations. Moreover, the model gets much more complicated, and the errors of extrapolation, corresponding with forecasting, grow very quickly. The article attempts a preliminary survey and analysis of time series data for the specification of a model of the interrelationship of variables. It should be recognized that the practical implementation of the above rules is not trivial. In particular, it is obvious that it is possible to obtain satisfactory estimates of the spectrum of financial and economic time series, but at the moment it is not clear how to quantitatively estimate volatility values, cooperation processes in conflict conditions, etc. Only further analysis, both theoretical and empirical, can provide answers to these questions.*

Keywords: financial time series, homoscedasticity, heteroscedasticity, regression, auto regression and moving average.

Стаття надійшла до редакції 19.01.2023 р.
Прийнято до друку 22.03.2023 р.