Sample division with sliding interval as the method of accuracy increase for time series forecast

Nina Kondrashova[1], Andriy Pavlov[1] , Yaroslav Pavlov[2]

[1] *International Research and Training Center of Information Technologies and Systems of National Academy of Sciences of Ukraine, 03680, Kiev, av. Glushkova, 40, Ukraine*
[2]*National technical university of Ukraine «Kiev polytechnic institute», faculty of informatics and computing technique*

e-mail: Nkondrashova@ukr.net

Abstract. *The forecast of heteroskedastic time series is considered at the example of data about the change of world price of indicative petroleum sorts of Brent and Urals. Two methods of forecast models construction are considered: firstly as the sum of trend model and the remainder model, secondly with help of model of first differences. inductive modeling algorithm for sliding window with optimization of division (AMSWOD) is offered, which uses the two-stage division in the sliding window of variable length. The window length is determined from the condition of accuracy maximization at the examinations subsamples. A comparison of the forecast obtained by AMSWOD and of forecast obtained by well known methods is made.*

Keywords

forecast, heteroskedastic time series, sliding window, two-stage division

5 Introduction

Empiric research in macroeconomics also as in a financial economy is based on time series largely. Unstationaryness is general property of macroeconomic and financial of the time series, which means that a variable does not have the clear tendency of return to the constant or to linear trend. A possibility of forecast of unstationary process and his variable volatility is researched in the paper with help of the models obtained of sliding window data limited by the time interval. A change of world price is forecast for Brent and Urals of the indicative petroleum sorts. Methods of stochastic approximation and group methods of data handling (GMDH) are compared in tests. Models of all methods depend on the history of the predicted values of time series exceptionally.

In the paper we will be limited to the forecasting of prices of Brent petroleum as showed the analysis of the time series of real values of month-averaged that the Brent and Urals are identical (almost complete coincidence of tendencies).

6 Problem statement and tested forecast methods

The purpose is to build forecasts of future petroleum prices to 3 months ahead using the data y_k, $k = 1,2,...,N$ of Brent prices for period 1999 - 2008. For this purpose are needed:

a) to construct models and $\hat{y}_{k+i}$ forecasts of time series by one of the most widespread methods, where a model is built not for a complete sample, but for the subsample shortened on the number of n_D examinations records. Number of examinations records is equaled the number of records, which a forecast will be built on. Forecast will be built on Q ahead, that is 3 months. Length of examination subsample will be equal to three records consequently.

b) To check the values of *FE* criterion which is the relation of the model mean deviation module to the spread of examination sample values and of *NMSE* criterion of the normalized mean-square error for the best models of an examination subsample.

$$FE = \sum_{i=1}^{n_D} \frac{|\hat{y}_{k+i} - y_{k+i}|}{(y_{\max} - y_{\min})n_D}, \quad NMSE = \sqrt{\frac{\sum_{i=1}^{n_D} (\hat{y}_{k+i} - y_{k+i})^2}{\sum_{i=1}^{n_D} (y_{k+i} - \bar{y})^2}}; \qquad (1)$$

c) to carry out the analysis of the obtained forecast results which were obtained by well known methods and to choose the best forecasting time series method;

d) to get new forecasts for the time series of Brent by the best forecast methods.

Forecast of heteroskedastic time series are built with help of the algorithms of GMDH [1], and also, for comparison, with autoregressive moving average (ARIMA) models of Box and Jenkins method [2], and with exponential smoothing models [3].

Holt-Winter's method, Holt's method, and method of smoothing with exponential trend were applied for the exponential smoothing. The parameters were estimated by quasi Newton's method.

STATISTICA programmatic package was used for realization of forecast construction in the exponential smoothing methods and in method of Box and Jenkins.

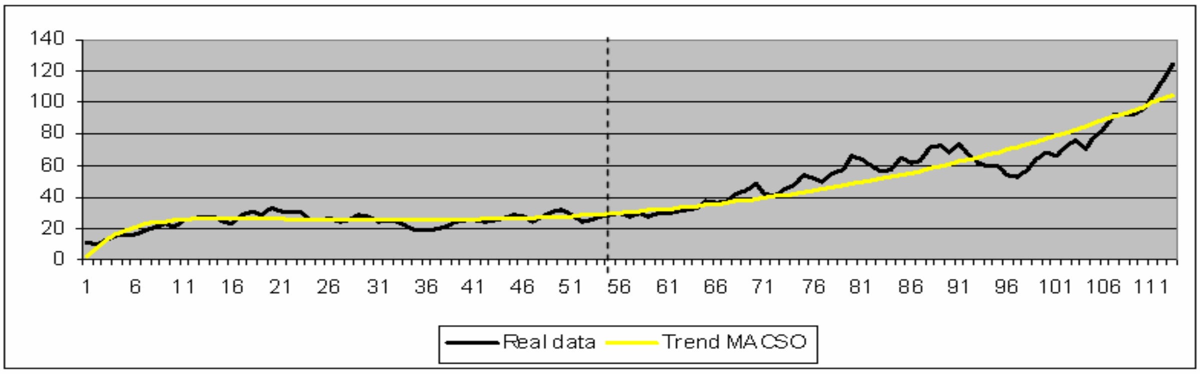

Fig. 1. Time series of Brent petroleum prices: real data and obtained trend by one of GMDH algorithms.

Initial data of *W* sample containing *N*=113 records of world petroleum price (see figure 1) of values of month-averaged was divided into parts: *U* training subsample and *D* examination subsample, see schema of division on figure 2a.

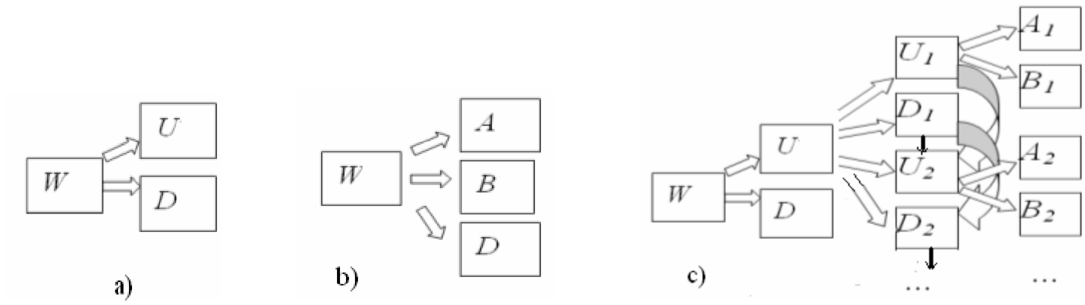

Fig. 2. Schemata of samples divisions: a) and b) onestage division; with) series-parallel multistage division.

Last 3 records of time series (March, April and May of 2008) belong to first *D* subsample for all methods of models construction.

The program of COMBI under the operating system of Windows was written for the realization of forecast construction by the combinatorial algorithm of GMDH.

In addition for the models construction the improved version of Sheludko's algorithm of GMDH [4] was used of multistage algorithm with a combinatorial selection and orthogonalization of variables (MACSO).

Initial *W* sample of petroleum prices of Brent brand [8] (see figure1) was divided into subsamples: *A* learning, *B* testing and *D* examination at the construction of GMDH models. $U = A \cup B$ subsample (n_U =110) was divided in the relation: $n_A : n_B$ =70:40 into *A* learning subsample and *B* testing subsample (see schema of division on figure 2b) in order of the prices values temporal sequence.

Model of forecast is as the sum of two models: of trend model (2) and remainder of model trend (4).

a) At first we get time trend *v(k)* as function of *k* . For COMBI *v(k)* is power series of *k*, for MAKSO *v(k)* is like:

$$v(k) = \boldsymbol{\theta}^{\mathrm{T}} \mathbf{f}(k) \qquad (2)$$

where: $\boldsymbol{\theta}$ is vector of parameters; $\mathbf{f}(k)$ is vector of functional transformations, elements of which belong to the set:

$$f_i(k) \in \left\{\frac{1}{k}, \sqrt[3]{k}, \frac{}{\sqrt[3]{k}}, 1, k, k^2, k^3, \ldots\right\} \quad (3)$$

b) we search part of forecast model for a remainder $\Delta_k = y_k - v(k)$ taking into account the effect of autocorrelation:

$$\Delta_{k+i} = f(\boldsymbol{\theta}\ , \Delta_k, \Delta_{k-1}, \Delta_{k-2}, \Delta_{k-3}), \quad i = \overline{1, n_D}\ . \quad (4)$$

A vector $\boldsymbol{\theta}$ is determined by a method of least-squares (MLS) of remainders errors of a learning subsample with the n_A number of records, and model structure have been selected by criterion of regularity at a subsample *B* with the n_B number of records.
From the tables 1 evidently that for Brent forecast of time series the best method is method of the smoothing with exponential trend.

Tab.1. Value of criteria for different methods.

	FE Brent	FE Urals	NMSE Brent	NMSE Urals
ARIMA	0,731025	0,803978	1,98235917	2,11396537
Holt's method	0,665735	0,664249	1,804646862	1,757806852
Holt-Winter's method	0,76445	0,965947	2,085621707	2,489740819
Method of smoothing with exponential trend	**0,579998**	0,554378	**1,576475667**	1,474579277
COMBI	0,827213	0,73623	2,160335869	1,918524493
recursive COMBI	0,779904	0,742736	2,03923394	1,94325782
MAKSO	0,603186	**0,530797**	1,637136114	**1,398610872**
recursive MAKSO	0,599118	0,56061	1,646099532	1,530100871

The numerical experimentation shows that the models obtained by these methods are not such, which well forecast, it follow from the *NMSE > 1* value of criterion*,* it means that dispersion of models error of the real data is greater than dispersion of the real data, but the at the same time obtained models are not quite bad, because the value of *FE* criterion is less than unit, this is evidence of that the middle module of models error from the real data is less than the size of variation of the real data.

We will apply for forecasting of the Brent time series in future Q (June, July and August of 20008) with the use of all records and of the method which gave the best values of criteria namely method of smoothing with exponential trend. Results are represented on a figure 5.
For research of methods of forecasts accuracy enhancement the next algorithm was created.

7 Algorithm of modeling for sliding window with optimization of division

The program of algorithm was written for realization of model construction for a sliding interval and with the quasi-optimal division of data utilizing the sample division [5] and algorithm model construction MAKSO [6].

The data from time series of length n_U is put into a sliding window consistently. The data are divided into U_j and D_j subsamples in sequence of values for the time in the first stage of division in a slide window, and in a second stage the data are divided into A_j and B_j subsamples quasi-optimally, see schema of division on figure 2c. A lengths of all examinations subsamples of *D* and D_j are identical $n_D = n_{D_j}, \forall j$ and can be set. Lengths of subsamples n_{U_j}, n_{A_j}, n_{B_j} and n_w length of window are variable in the set limits.

In this work the results of construction of linear autoregressive models of the ∇_k monthly first differences of initial variable are represented (see data on figure 3) as the modeling of the remains Δ_k not raised accuracy of forecasts. The content of subsamples is being updated, at the change of the window position per unit:

$$\nabla_{k+i} = f(\boldsymbol{\theta}, \nabla_k, \nabla_{k-1}, \nabla_{k-2}, \nabla_{k-3}), \quad i = \overline{1, n_D} \quad (5)$$

process of division and of models construction is being repeatable oneself for (*j*+1)th window.

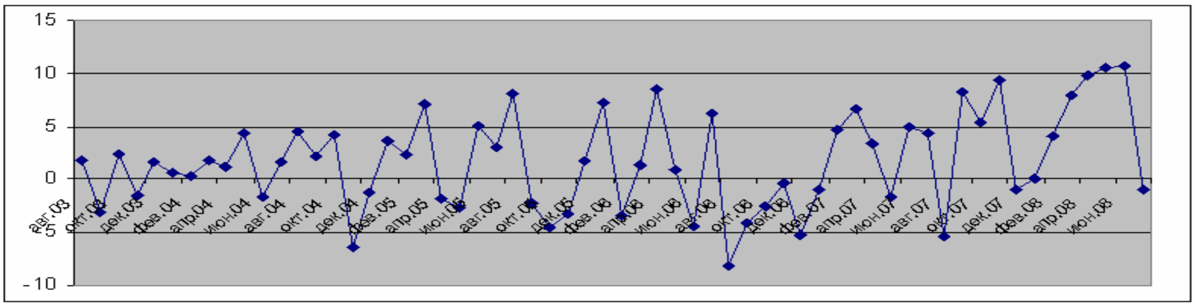

Fig. 3. Real data (monthly first differences of Brent time series, 61 records)

Algorithms (COMBI, MAKSO and AMSWOD) were used for the models construction with a recursive forecast and without its [7]. Nonrecursive forecast, when for the every of forecast record $i=\overline{1,n_D}$ is built the model (4) or depending on method of presentation of initial data the model (5). Recursive forecast, when the every value obtained of forecasting per one step ahead take part in further forecast and the model is being recalculated taking into account this forecast:

$$\hat{x}_{k+1}=\varphi(\mathbf{\theta},x_k,x_{k-1}...),\ \hat{x}_{k+2}=\varphi(\mathbf{\theta},x_k,x_{k-1}...,\hat{x}_{k+1}),\ldots,\hat{x}_{k+n_D}=\varphi(\mathbf{\theta},x_k,x_{k-1}...,\hat{x}_{k+1},...,\hat{x}_{k+n_D-1}).$$

Thus on every step of recursive forecasting, actually, the model of forecast is built for one step, which is being a result of chain substitution of the $\hat{x}_{k+i},\ i=\overline{1,(n_D-1)}$ forecast values.

For the construction of models the *U* initial data subsample was shorted and contained n_U =55 records, as in this algorithm it is necessary to make the enormous number of search variants of different parameters and types of models. In addition, for the models construction of current forecast data of current window is really used only n_w of direct predecessors of the subsample to the point of forecast ($n_w < n_U$). How many records need to be sifted from, and how many to take into account as to raise at the average the accuracy of forecast? It was one of aims of this algorithm creation: to define dependence of forecast accuracy on window length. As the researches in [7] have showed that optimization of division will raise forecast accuracy, the optimization of division was realized for window data in AMSWOD.

Double sliding is used in this algorithm: firstly by a window along up of all initial data; secondly by a difference pattern[1] in a window of data. The first sliding allows to be more grounded to make conclusions about forecast accuracy at the set length of window; the second sliding is needed for the parameters estimation

How to measure forecast accuracy on the average at the set window?

The criteria of accuracy (1) make sense for comparison of obtained forecasts by different methods and algorithms, when the set of examination records identically and more than one record.

In AMSWOD at the change of position of window the set of records of examination is varied, in addition, as the forecasting model is being explored for one step, the criteria of (1) can not be calculated. It is therefore examined the following four possible criteria: module of forecast error for last record of sliding window; *PRT* the criterion of forecasting accuracy of tendency; number of allowable forecasts, when a model has the possible module of error in records of a forecast at different positions of sliding window; criterion of invariability and of simplicity of model for the possibly greater number of positions and lengths of window.

The *PRT* criterion of forecasting of tendency calculates probability of correct forecast of tendency, as: $PRT=\dfrac{n}{n_D}$, where *n* is number of forecast steps, on which direction of change of the forecast value coincides with direction of change of real value. We will designate $\hat{\alpha}_{k+1}=\hat{y}_{k+1}-y_k$, $\hat{\alpha}_{k+i}=\hat{y}_{k+i}-\hat{y}_{k+i-1}$, $i=\overline{2,n_D}$. Then *PRT* was calculated as follows: $n=0$, $n\overset{\Delta}{=}n+1$ if the $sign\,\alpha_{k+i}=sign\,\hat{\alpha}_{k+i}$, and $n\overset{\Delta}{=}n$ if the $sign\,\delta_{k+i}\neq sign\,\hat{\delta}_{k+i}$, $i=\overline{1,n_D}$. This index was averaged for the various maximal number of windows positions with the change of window length in order to get PRT_O value in the recalculation for one position of window. The index ω which is the part of models with acceptable of forecast was recalculated also for one position of window. Criterion of invariability and of simplicity of type of model is not formalized and was determined by sight as invariability of type of model for the change of window and as number of variables entering to the forecast model of the last record.

At the change of window length $11\le n_w\le 27$ abovementioned PRT_O, δ, ω indexes (see figure 4) were estimated, $n_w^*=15$ of window length is selected, which is the best result of the sum of the conflicting indexes.

[1] difference pattern is concrete type of difference equation (5).

After limitation on linearity of structures class, second, by what else it was possible to be limited in research it was a single-stage model. It is possible also taking into account the recursive characteristic of forecasting. Thus, the length of window, obtained for a single-stage model, was chosen for the models construction per three steps.

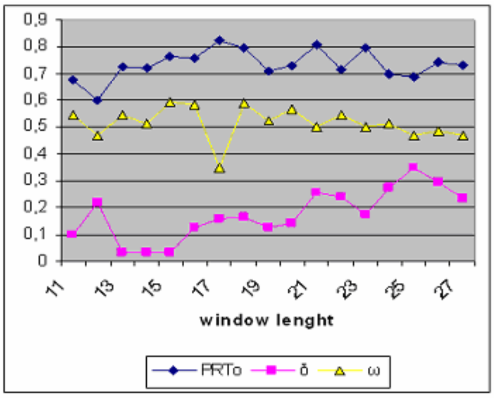

Fig. 4

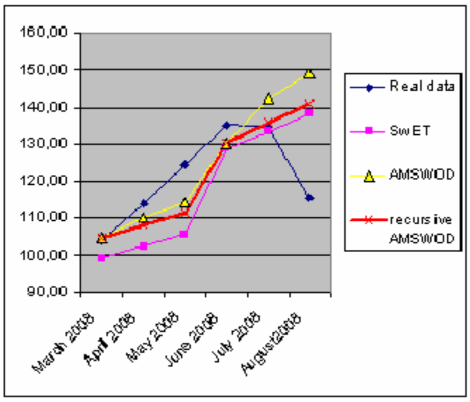

Fig. 5

Fig. 4. *PRT* the criterion of tendency forecasting; δ module of forecast error for last record of sliding window; ω part of models with allowable of forecast

Fig. 5. Real data and double forecasting of time series of Brent for 3 steps by different algorithms

4 Comparison of forecasts obtained by different methods

A model has guessed about the tendency of variable in 76% cases and in 60% gives allowable of forecast as is obvious from figure 4. This result of forecasting of AMSWOD is obtained at the limited search, on a quite short sample from 55 records and by linear autoregressive models of the first differences of variable. It requires further researches. As compared to the best forecasting of method by the smoothing with exponential trend (SwET) of AMSWOD on the first temporal series of forecast (March, April and the May of 2008) gives more than in two times more exact forecast (criterion *NMSE =0.61* at recursive AMSWOD against *NMSE=1.65* at SwET). Forecast of AMSWOD on the first of three records occupies intermediate position of *NMSE=0.75*. Taking into account a forecast in the second series of forecast for three records (June, July and August of 2008) the SwET is more exact *NMSE=1.50* (see figure 5), although it together with other methods does not reflect the tendency of prices change in July and in August.

5 Conclusion

At forecasting of the time series of petroleum prices by the use of sliding window with recursive forecast and optimization of data division was provided more accurate the forecasting on the areas of monotonous change of time series, but does not reflect the tendency of change at unmonotonous of change of variable. The further improvement of forecast accuracy of AMSWOD would be possibly due to application of nonlinear models of autoregression.

References

[1] Ivakhnenko A. G.: The group methods of data handling is the concurrent of the stochastic approximation method. *Avtomatika*, 1968, №3. (In Russian)

[2] G. Box , Gw. M. Jenkins, Gr. Reinsel : Time Series Analysis: Forecasting & Control (3rd Edition), *Prentice Hall*, 1994, 592 p.

[3] Kantorovich G.G.: Time Series Analysis. *Economic magazine of High school of economics*. Vol. 7. №1, 79-103, 2003. (In Russian)

[4] Sheludko O. I.: The GMDH algorithm with complete orthogonalized description for models' synthesis using the results of the design of experiments. *Avtomatica*, 1974, №5, 32-41. (In Russian).

[5] Stepashko V.S., Kondrashova N.V.: Research of methods of sample division in the GMDH algorithms, *Works of I International Conference on Inductive Modeling, Lviv* , Vol. 1, P.1, 90-94, 2002. (in Russian).

[6] Vanin V. V., Pavlov Alex. V.: Development and application of self-organization algorithms for modeling of complex processes and objects which are represented by the point former, *Proceedings of Tavria State agrotechnical academy*. Pub. 4, Vol. 24, Melitopol, 2004, 51-56. (In Ukrainian).

[7] Kondrashova N.V.: Influence of method of sample division in GMDH algorithms on accuracy of forecasting // *Control Systems and Computers*, 2003, №2, 128-133 (In Russian).

[8] http://trade.tatcenter.ru