TIME SERIES PROGNOSIS OF GDP WITH THE SYSTEM GMDH-SHELL (EXPERIMENTAL WORK)

Victor Lebedev

Abstract: *Time series prognosis of economical indexes is one of the main problems of econometrics. In the paper we study possibility to give an interval prognosis of time series using the set of the best prognostic models. Speaking 'model' we mean a combined model of regression and auto-regression. Speaking 'the best models' we mean the ordered series of models constructed by the well-known Group Method of Data Handling (GMDH). The proposed simple approach consists in the following: a) one generates the fixed numbers of models on the basis of experimental data b) these models give correspondent prognoses c) the real value is supposed to belong to min-max interval the models provide. We shortly describe the software tool GMDH-Shell (GS) that implements GMDH and the results of experiments with GS. The experimental data are time series of the Gross Domestic Products (GDP) of 100 countries given on the period 1980-2000.*

Keywords: *GMDH, GMDH Shell, time series prognosis, gross domestic product*

ACM Classification Keywords: *I.2 Artificial Intelligence*

Introduction

Econometrics offers many methods for time series prognosis of economical indexes. Almost all of these methods are based on certain statistical assumptions and use various variants of regression analysis [Kandler, 1981; Klayner, 2000]. Group Method of Data Handling (GMDH) is an alternative to the traditional statistical approach. This method determines the model of the optimal complexity from the given class of models on the basis of experimental data. GMDH uses two or more subsets from a given data set for model construction, selection, and verification. It allows automatically to take into account indefiniteness concerning features of source data.

GMDH is effective when: a) we have limited information about the structure of a model; b) we have limited value of observation data. Often we know nothing about the model and often the number of observation data is less then the number of parameters to be determined. GMDH was introduced at 80s by Ukrainian academician Alexey Ivakhnenko and now it is developed by his colleagues and pupils. We can mention here some publications in English [Ivakhnenko, 1981; Ivakhnenko, 1994]. The theoretical basis of GMDH is presented in the paper [Stepashko, 2008]. The list of publications related with GMDH and its applications is presented in [GMDH, http].

The traditional way of using GMDH consists in construction of one the best model and its further application for prognosis. To make the prognosis more reliable one uses assembling or ensembling. In the first case the prognosis is an average value all models provide, in the second case the prognosis is an averaged value based on only some models from a given set. This approach is enough new and it needs additional research. We can mention here the work [Zhi-Hua Zhou, 2002] **where ensembling neural network for prognosis is considered**. In statistics a prognosis is usually accompanied by so-called confidential interval. In inductive modeling an interval prognosis can be obtain, for example, with fuzzy GMDH [Zaychenko, 2008]. In this paper we propose to form this interval using minimum and maximum values, which first the best GMDH models provide.

The paper is organized by the following way. In the section 2 we shortly describe the tool GMDH Shell (GS). Section 3 presents the results of experiments with GS. Section 4 contains the conclusions.

GMDH Shell

The set of models for time series prognosis is created by the tool (GS) mentioned in the Introduction. This software implements algorithms of GMDH and has very friendly user interface. One can meet with GS possibilities and download it [GS, http]. GS has a Wizard for new users with a series of typical examples (see Fig.1).

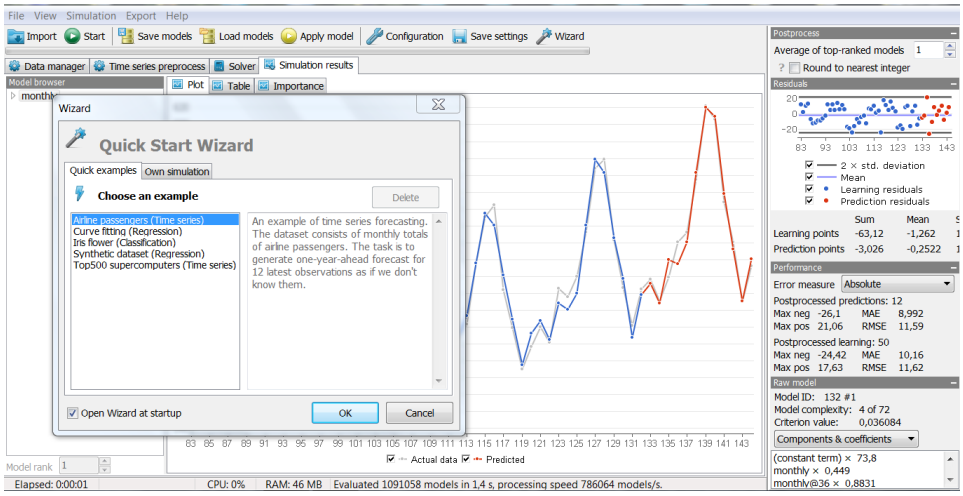

Fig. 1 Screenshot of GS interface

Traditionally GMDH offers 4 algorithms: combinatorial algorithm (COMBI), combinatorial selective algorithm (MULTI), multilayered Iterative algorithm (MIA), and relaxation iterative algorithm (RIA). GS uses modified COMBI and modified MIA. The modifications concern the limitations of models a user assigns. GS algorithms are described in [Koshulko, 2007; Koshulko, 2009]. GS includes many modes for testing model validity [Koshulko, 2011] that is very important for users [Latysh, 2012]. Figure 2 presents the procedures of GS.

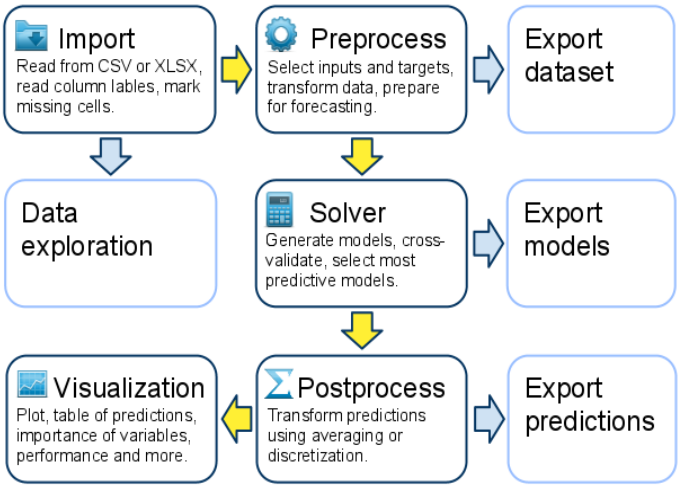

Fig. 2 Data processing in GS

In our research we used first of all the possibility to select the set of the most predictive models in the module Solver. These models allow to find the minimum and maximum values of the prognosis. On the stage of pre-processing we recovered missing data (not more then 5%). On the stage of post-processing we studied the most frequent variables used in models.

Experiments

The experimental data are time series of the Gross Domestic Products (GDP) of 100 countries given on the period 1980-2009 [WorldBank, http]. The part of this data is presented on Figure 3.

Variable	Gross Domestic Product, at current prices in national currency, in millions						
Year	1980	1981	1982	1983	1984	1985	1986
Country							
Australia	154648	178405	192066	216657	239161	264060	289214
Austria	76595,71	81596,949	87625,032	93331,881	98011,346	103419,239	108956,922
Belgium	90698,929	95093,801	102901,426	109016,03	117780,09	125266,165	131116,748
Canada	325910,25	365318	387740,75	420935	458615	492420,75	524143
Denmark	392875	430068	491088	541428	597727	648540	698783
Finland	33240	37568	42272	47074	52645	57271	61616
France	445233,05	500755,79	574445,4	636621,61	693087,94	743889,54	802364,52

Fig. 3 Source data (the part of full table)

In our experiment we consider 2 countries: US and Switzerland. The experiment consists in the following:

1 First 10 the best prognostic models of GDP for 2007 are generated. Here we use the data on the period 1980-2006

2 Prognoses with all 10 models are calculated and minimum and maximum values are founded

3 We test whether the real value belong to the interval of minimum-maximum values

The experiment is repeated for 2008 and 2009 using the data on the periods 1980-2007 and 1980-2008 respectively. Table 1 shows the results related with US, table 2 shows the results related with Switzerland.

Table 1. Prognoses for US

	2007	***2008***	***2009***
Model 1	14114673	14506527	14462123
Model 2	13975186	14660353	15242288
Model 3	13819792	14505804	14462021
Model 4	13938927	14702221	13934209
Model 5	14005305	14661333	14491801
Model 6	13900404	14684275	15156285
Model 7	13935342	14641953	15205031
Model 8	13914712	14619275	14610805
Model 9	13913198	14521542	15069660
Model 10	13914228	14714699	14677642
Minimum	13819792	14505804	13934209
Maximum	14114673	14714699	15242288
Real	**13830300**	**14221425**	**14107150**

Table 2. Prognoses for Switzerland

	2007	***2008***	***2009***
Model 1	515726	541035	530527
Model 2	515418	546639	530458
Model 3	507785	550349	530458
Model 4	506546	556283	561952
Model 5	507064	547059	568298
Model 6	503074	537962	530983
Model 7	503658	539430	586720
Model 8	506366	539430	531862
Model 9	507165	539374	532408
Model 10	508011	555281	528059
Minimum	503074	537962	528059
Maximum	515726	556283	586720
Real	**521100**	**544195**	**535282**

One can see that in 2 cases from the 6 ones the real data are outside the interval of minimum-maximum values.

Conclusions

The main results of the paper are:

- We proposed the way of prognosis based on the set of the best models generated by GMDH
- We shortly described the software tool GMDH Shell, which implements GMDH
- The results proved not to be completely perfect

In the future we suppose to study more detail the proposed way having in view both the number of models and the algorithm of modeling

Acknowledgement

The author is very grateful to Dr. Oleksiy Koshulko from Gluskov Institute of Cybernetics (Ukraine), for his numerous consultations and help on GMDH Shell.

Bibliography

[GMDH, http] GMDH: www.gmdh.net

[GS, http] GS: www.gmdhshell.com

[Ivakhnenko, 1971] A. Ivakhnenko. Polynomial theory of complex systems. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-1(4), 1971, pp. 364-378

[Ivakhnenko, 1994] A. Ivakhnenko, H. Madala. Inductive learning algorithms for complex systems modeling. CRC Press, NY, 1994

[Kandle, 1981] Kandle M. Time series. M.: Finances and statistics, 1981 (in Russian, translated from English)

[Klayner, 2000] Klayner G.B., Smolyak S.A. Econometrician relationships: principles and building up methods. M.: Nauka, 2000 (in Russian, translated from English)

[Koshulko, 2007] O. Koshulko, A. Koshulko. Adaptive parallel implementation of the Combinatorial GMDH algorithm. In: Proc. of the 2-nd Intern. Workshop on Inductive Modeling (IWIM-2007). Czech Rep., Prague, 2007, pp. 71-74

[Koshulko, 2009] O. Koshulko, A. Koshulko, Multistage combinatorial GMDH algorithm for parallel processing of high-dimensional data. In: Proc. of the 3-rd Intern. Workshop on Inductive Modeling (IWIM-2009). Poland, Rzeshov, 2009, pp. 114-116

[Koshulko, 2011] O. Koshulko, G. Koshulko. Validation strategy selection in Combinatorial and Multilayered Iterative GMDH algorithms. In: Proc. of the 4-th Intern. Workshop on Inductive Modeling (IWIM-2011), Ukraine, Kyev, 2011, 4 pp.

[Latysh, 2012] E. Latysh, O. Koshulko. Testing k-value in k-fold cross validation of forecasting models for time series analysis of G-spreads of top-quality RUB bonds. In: Proc. of Intern. Conf on Inductive Modeling (IWIM-2012), Kyiv, 2012, 4 pp.

[Stepashko, 2008] V. Stepashko. Method of critical variances as analytical tool of theory of inductive modeling. Journal of Information and Automation Sciences, Publ.: Begell House Inc, 2008, Vol. 40, N_3, pp. 4-22

[WorldBank, http] WorldBank: http://www.worldbank.org/eca/russian/

[Zaychenko, 2008] Y. Zaychenko. The Investigations of Fuzzy Group Method of Data Handling with Fuzzy Inputs in the problem of forecasting in financial sphere. In: Proc. of 2-nd Intern. Conf. on Inductive Modeling (ICIM-2008), pp.129-133

[Zhi-Hua Zhou, 2002] Zhi-Hua Zhou. Et Al. Ensembling neural networks: many could be better than all. Journ. of Artificial. Intelligence. 2002, vol.137, No,1-2, pp.239-263

Authors' Information

Victor Lebedev *– student, the Russian Presidential Academy of national economy and public administration; Prosp. Vernadskogo 82, bld. 1, Moscow, 119571, Russia; Moscow Institute of Physics and Technology (State University); Institutsky pereulok, bld. 9, Dolgoprudny, 141700, Moscow region, Russia;*

e-mail: viktor.lebedev@phystech.edu

Major Fields of Scientific Research: inductive modeling, business intelligence