BUILDING NOISE IMMUNITY MODELS FOR GDP FORECAST BASED ON ELECTRICAL POWER CONSUMPTION[2]

Ksenia Terekhina, Mikhail Alexandrov, Oleksiy Koshulko

Abstract: *Forecasting GDP is one of the most popular applications of forecast methods in macro-economy. In this paper we build medium-term predictive GDP-models for 19 countries using Group Method of Data Handling (GHDH). The models are the hybrid ones: they include data of GDP itself and data of electrical power consumption. GMDH is realized by means of neuro- similar algorithm from the software package GMDH Shell. We studied properties of the models related to volume of teaching sample, ratio of teaching/control data, and noise immunity. The built models can be a useful addition to traditional models of GDP forecast.*

Keywords*: GDP forecast, inductive modeling, GMDH Shell.*

ACM Classification Keywords: *I.2 Artificial Intelligence.*

Introduction

Problem settings

Gross Domestic Product (GDP) is one of the most important macroeconomic indicators, which determines a place of country in the world. Dynamics of GDP is used in economic analysis, sociological prognosis, and in political battles. To build the forecast model one needs to define model parameters and to select method for building this model:

1) Predictive models usually use the values of GDP itself or other macroeconomic parameters, such as a price of exported gas or oil, volume of industrial and agricultural products related to export, etc. [Angelini, 2010]. Productions of goods and services, mineral extraction, as well as other economic activities require energy. So, it is very naturally to build GDP models using electrical power consumption as an integral variable.
2) By the moment there are many well-studied auto-regression and regression models for forecasting GDP [Kraay, 1999; Angelini, 2010]. These models are built under strong limitations concerning model structure and noise characteristics. Thus, applying GMDH technique, which allows to build models without mentioned limitations, is very promising. The models with higher noise immunity would be result of GMDH use.

These circumstances taken together define actuality of completed research. In this paper the models for GDP forecast in 19 countries with comprehensive amount of data are built and studied.

Related works

This moment there are many models for GDP forecast using indirect parameters. One of the most interesting indirect parameters is night lights. The model with this parameter was considered by Weil in 2009, the main conclusion of his work was direct correlation night lights' intensity and GDP [Weil, 2009]. Unfortunately, Weil's models proved to be invalid for Russia and some other northern countries due to territorial compression of economy and switching on business activity. Basing on the physics laws Stern claims that energy market has an impact on GDP, because the energy is required for production [Stern, 2010]. The researcher built the model of GDP using labor, capital and energy as input parameters. As for the GMDH application in economic forecast

[2] Work done under partial support of the British Petroleum grant (BP-RPANEPA 2013).

then one should mention here the large experience of Ukrainian researchers. In particularly they used GMDH for forecasting GDP and inflation of Ukraine [Ivakhnenko, 1997; Stepashko, 2010].

In section 2 of this paper the short description of GMDH and models to be built is given. Section 3 contains experimental study of model properties. Finally, section 4 is devoted to conclusions.

Tools and Models

Group Method of Data Handling

Although this method has long history it still is not well-known to researches dealing with modeling [Ivakhnenko, 1968; Ivakhnenko, 1971; Madala, 1994; Stepashko, 2013]. Particularly, the method demonstrates its advantages over the others, when:

- there is no or almost no a priori information about the structure of model and its parameter distribution,
- there are very limited quantity of initial data (data of observations) reflecting model behavior.

However, when a model or set of models are fixed, meaning have a known structure and unknown values of its parameters, and large volume of measured data is accessible, one can successfully use other approaches for model identification.

The GMDH generally follows 5-steps scheme:

1. Model class is given by an expert. Models are ordered with increasing complexity.
2. Learning data set is divided into training set and control set. Models are built on one set and tested on the other one.
3. Model parameters are determined using any internal criterion (the least squares, for example) on teaching set.
4. Model quality is determined using any external criterion (criterion of regularity, for example) on control set.
5. Model complexity is increased until the external criterion will reach its extreme.

The final model has an optimal complexity having in view a) reflection of object behavior in data of observation b) stability to unknown factors, which is titled noise.

GMDH Shell

For modeling we used software package GMDH Shell here-in-after named GS [GS, http].It is a well-known tool for time series forecast, function approximation and object classification including extended possibilities for visualization of results. GS employs GMDH technique in all its algorithms. Actual GS version includes two algorithms:

- Combinatorial GMDH;
- GMDH-type neural networks;

The unique possibility of GS is its automatic adjustment to a given data set. Namely GS itself tests each of the mentioned algorithms on a part of initial data and then selects the algorithm giving the best results. Of course, such a selection is completed with the agreement of user. In our case GS suggested to use neuro-similar algorithm which was used in our research. The comparison of algorithms from GS is presented in [Koshulko, 2011]

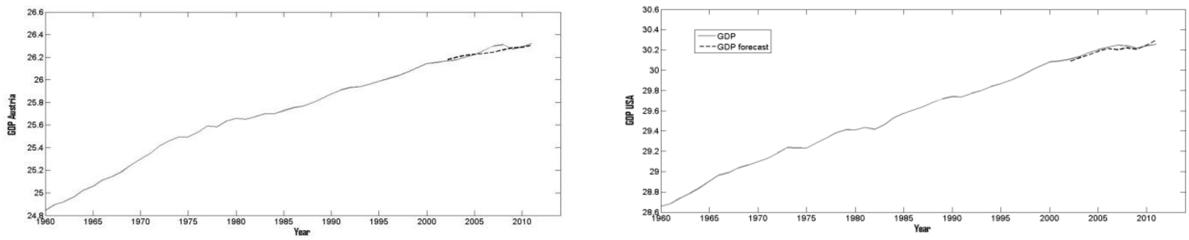
Building Models

Models to be built are the hybrid ones including data of GDP itself and data of electricity consumption. Class of models is polynomials including lags. GDP were measured in local currency and electricity consumption was measured in kW-hour. In our paper we built models for 10 year forecast in 19 countries, for which enough volume of initial data was acquired. The complete learning period was 1960-2011.

Special attention was paid on preprocessing procedures. We used:

- Scaling-1. All data were scaled using logarithmic transformation ln(1+x). After that all the data fall within the range of 20-40.
- Scaling-2. All variables were transformed using cubic root. It allowed to avoid extreme values of the variables under consideration.

The examples of models for Austria and USA are presented below in Table 1. The results of modeling are presented on Figure 1. The data about the best and the worst MAPE are presented in Table 2.



a) Austria b) USA

Figure 1. GDP and GDP forecast, solid line is raw data, dashed line is forecast.

Table 1. Model of GDP

Y1 = -3,657e^{-07}+ N4*0,489 + N3*0,510 N3 = -0,168*10^3 +$[Elec(t\text{-}14)_{Austria}]^{1/3}$*0,326*$10^3$+ N4 N4=-8,9+$[Elec(t\text{-}14)_{Austria}]^{1/3}$*8,776+ $[GDP(t\text{-}14)_{Austria}]^{1/3}$ *3,282 MAPE = 0,1239	Y1 = 4,836e^{-07} +$[GDP(t\text{-}9)_{USA}]^{1/3}$ *(-1,64e^{-07}) + N2 N2 = 4,759e^{-05} +$[GDP(t\text{-}9)_{USA}]^{1/3}$ *(-1,624e-05) + N3 N3 = 20,9229 + $[GDP(t\text{-}9)_{USA}]^{1/3}$*(-7,155) + N4*1.039 N4 = -39,578 + $[GDP(t\text{-}17)_{USA}]^{1/3}$*20,529 + N5*0,213 N5 = -61,426 + $[Elec(t\text{-}5)_{USA}]^{1/3}$*29,943 MAPE = 0,0699

MAPE stated for Mean Absolute Percent Error. It can be calculated by the formula

$$\mathbf{MAPE} = \frac{1}{h}\sum_{t=T+1}^{T+h}\left|\frac{\hat{y}_t - y_t}{y_t}\right| * 100\%$$

where y'_t, y_t are data of modeling and experimental data, h is number of data for forecasting

In the represented formulas, Yi are predicted values, Ni are neurons. Each neuron is an elementary transformator of entrance data

Table 2.

	Value	Country
The best MAPE	0,039%	Australia
Mean MAPE	0,186%	
The worst value	0,499%	Turkey

All values of MAPE mentioned above refer to transformed data that is to natural logarithm of GDP data. The values of MAPE related to real data are approximately 10-20 times greater (it depends on data). In spite of this note one can see that the accuracy keeps high level having in view 10 year forecast.

Studying Models

Sensibility to sample size

We studied dependence of the forecast models quality on the length of samples. On each sample (1960-2011, 1970-2011, 1980-2011) for each country optimal prediction model for the horizon of 5 years was built. The MAPE indicator was used to compare the qualitative characteristics of the models. After the experiment with 19 countries, only 3 of 19 models were found to have the best quality characteristics in the time interval 1960-2011.

Table 3. Sensibility to sample size.

	1990-2011	1980-2011	1970-2011	1960-2011
Country	**MAPE %**			
Australia	0,0493	**0,0430**	0,1142	0,0768
Austria	0,1280	**0,0652**	0,2690	0,1038
Belgium	**0,0256**	0,1035	0,1145	0,0732
Canada	**0,0254**	0,1085	0,0602	0,1595
Denmark	0,2158	0,2327	**0,2146**	0,3361
Finland	**0,1526**	0,2901	0,1675	0,2728
France	**0,0604**	0,1167	0,1207	0,1505
Greece	0,5412	0,7309	0,5843	**0,5312**
Italy	**0,1880**	0,2401	0,2083	0,2173
Japan	0,0892	**0,0732**	0,0978	0,1262
Luxemburg	0,2185	0,9944	0,1884	**0,1824**
Netherlands	**0,0734**	1,8250	0,1486	0,2527
Norway	0,1487	**0,0961**	0,2230	0,1795
Portugal	0,1466	**0,0924**	0,1441	0,5068
Spain	0,2447	**0,2241**	0,2274	0,3312
Sweden	1,2510	0,1239	**0,1238**	0,1284
Turkey	0,5046	0,3597	0,3220	**0,2069**
United Kingdom	0,2506	**0,1385**	0,2385	0,1955
USA	**0,1013**	0,1204	0,2457	0,1461

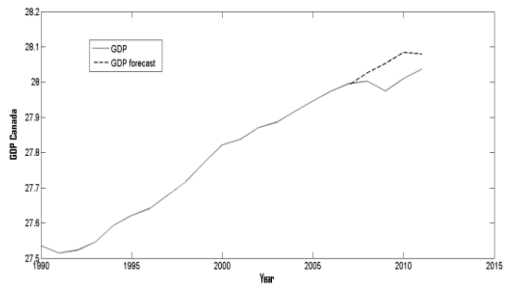


Figure 2. Model for Canadian GDP built on 1960-2011

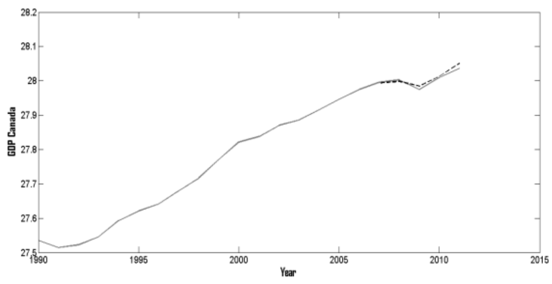


Figure 3. Model of Canadian GDP built on 1990-2011

The left graph represents model for Canadian GDP built on a sample 1960 - 2011, the right one represents model for Canadian GDP built on the sample 1990 – 2011.

Ratio training/control

According to GMDH model, parameters are determined on a training set and model quality is tested on a control set. Obviously:

- when training set is too small then model proves to be too simple,
- when training set is too large then model proves to be too complex.

In both cases the quality of the models are low. The different ratios of volumes for training set and control set in order to find the best ratio were tested. In the experiment we built models for 5 year forecast using learning set 1960-2011. The results of this testing are presented in Table 4.

Table 4. Ratio training/control

Country	40/60	60/40	90/10
Australia	0,0495	**0,0266**	0,0266
Austria	0,1901	**0,0466**	0,1247
Canada	0,1236	**0,0546**	0,1822
Denmark	0,3131	**0,098**	0,1617
France	0,1191	0,1465	**0,1091**
Norway	**0,1099**	0,1928	0,1603
Portugal	0,3463	**0,2973**	0,41
Turkey	0,6703	**0,3141**	0,4541
United Kingdom	0,191	**0,1749**	0,1791
USA	0,2507	**0,1716**	0,2541

According to the testing, the ratio 60/40 is the best one for the majority of the cases. Thus, this ratio was used in our experiments described in section 2.3. One should say that this ratio is recommended by GMDH developers [Koshulko, 2011].

Noise immunity

Noise immunity is one of the principal characteristics of models built with GMDH. Such an effect is reached when model is simplified under high level of noise. Experimental study of this effect is presented in many publications, for example [Ponomareva, 2008]. Theoretical justification is done in [Stepashko, 2008]. Our experiment consisted in the following:

Data (electricity and GDP) were taken for both developed and developing countries. These data were supplemented with gauss noise. The levels of noise were 10%, 20% and 50%. We built models for 5 year forecast on the sample 1990-2011. To compare results the values of MAPE were used.

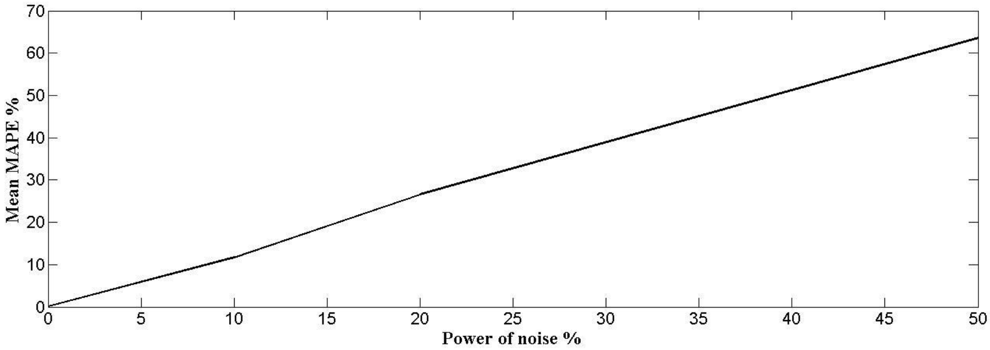


Figure 4. Dependence of the average MAPE values for 10 countries on the noise level

The Figure 4 shows the dependence of averaged MAPE on the noise level. The averaged MAPE was calculated using data of 10 countries. The linear dependence is evidence for the noise immunity of the built hybrid models.

Conclusions

In this paper we built models for 10 year forecast for 19 countries The neuro-similar algorithm was used to build models, which provided high quality of forecast. Namely, MAPE varied from 0,039% to 0,499% with a middle value 0,186 % (all values refer to the transformed data).

We studied characteristics of built models and obtained the following results:

- Long time series do not guarantee better model quality. It proves that better models refer to learning period 20-30 years for the majority of countries.
- The best ratio teaching/testing is equal 3:2. It corresponds to the well-known recommendations concerning GMDH techniques.
- All built models have high level of noise immunity that is one of the main advantages of GMDH technique.

In the nearest future we plan to build models with GMDH technique for other macroeconomic parameters.

Bibliography

[Angelini, 2010] E. Angelini, G. Camba-Mendez , D. Giannone, L. Reichlin and G. Runstler. Short-term forecasts of euroarea GDP growth. Econometrics Journal (2011), volume 14, pp. C25–C44. http://dx.doi.org/10.1111/j.1368-423X.2010.00328.x

[Kraay, 1999] A. Kraay, G. Monokroussos. Growth Forecasts Using Time Series and Growth Models. Policy Research Working Papes, 1999, p.38 http://dx.doi.org/10.1596/1813-9450-2224

[Weil, 2009] D. Weil, H. Vernon, A. Storeygard. Measuring Economic Growth from Outer Space. NBER Working Paper Series,2009, Vol. w15199.

[Stern, 2010] Stern D. I. The Role of Energy in Economic Growth. USAEE-IAEE, 2010, w10-055 http://dx.doi.org/10.2139/ssrn.1715855.

[Ivakhnenko, 1997] A.G. Ivakhnenko, G.A. Ivakhnenko, N.M. Andrienko. Inductive Computer Advisor for Current Forecasting of Ukraine Macroeconomy. http://www.gmdh.net/articles/papers/advisor.pdf

[Stepashko, 2010] V.S. Stepashko, Y.V. Koppa. Comparison of predictive properties of regression models and GMDH. Kiev, 2010. (rus)

[Ivakhnenko, 1968] A. Ivakhnenko. Group Method of Data Handling as a Rival of Stochastic Approximation Method. Journal "Soviet Automatic Control", No. 3 , 1968, pp. 58-72

[Ivakhnenko, 1971] A. Ivakhnenko. Polynomial theory of complex systems. IEEE Trans. Sys., Man and Cyb., No 4, 1971, pp. 364-378.

[Madala, 1994] H.R. Madala, A.G. Ivakhnenko. Inductive learning algorithms for complex systems modeling. CRC Press, 1994, 368 pp.

[Koshulko, 2011] O., Koshulko. Validation Strategy Selection in Combinatorial and Multilayered Iterative GMDH Algorithms. In: Proc. Intern. Workshop on Inductive Modeling (IWIM-2011), NAS of Ukraine and Prague Tech. Univ, 2011, pp. 51-54

[Ponomareva, 2008] N. Ponomareva, M. Alexandrov, A. Gelbukh. Performance of Inductive Method of Model Self-Organization with Incomplete Model and Noisy Data. Published in Gelbukh.A., Morales, E. (Eds.): Journal of IEEE Computer Society, N_3441, pp.101-108

[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

[GS, http] Program GMDH Shell, http:// gmdhshell.com

[Stepashko, 2013] V. Stepashko, Ideas of academician O. Ivakhnenko in Inductive Modeling field from historical perspective. In: Proc. of 4th Intern. Conf. on Inductive Modeling (ICIM-2013), 31-37. NAS of Ukraine, Prague Tech. University, Kyev, 2013.

Authors' Information

Ksenia Terekhina – *M.Sc student, 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); Institutskii per 9., Dolgoprudny, Moscow Region, 141700, Russia*

e-mail: kseniya.terekhina@phystech.edu

Major Fields of Scientific Research: mathematical modeling, world economy

Mikhail Alexandrov – *Professor, Russian Presidential Academy of national economy and public administration; Prosp. Vernadskogo 82, bld. 1, Moscow, 119571, Russia; fLexSem Research Group, Autonomous University of Barcelona, 08193 Bellaterra (Barcelona), Spain;*

e-mail: MAlexandrov@ mail.ru

Major Fields of Scientific Research: data mining, text mining, mathematical modelling

Oleksiy Koshulko – *Senior Research Associate, Glushkov Institute of Cybernetics of NAS of Ukraine; Glushkov str. 40, Kyiv, 03680, Ukraine;*

e-mail: koshulko@ gmail.com

Major Fields of Scientific Research: data science, parallel processing