FORECAST OF FORRESTER'S VARIABLES USING GMDH TECHNIQUE[1]

Olga Proncheva, Mikhail Alexandrov, Volodymyr Stepashko, Oleksiy Koshulko

***Abstract**: In distant 1970 the MIT professor J. Forrester proposed a model of world dynamics in the form of five differential equations for 5 basic variables, namely: population, capital assets, agricultural assets ratio, pollution and natural resources. In the paper, we show the possibility to build a predictive model based on non-lineal difference equations using technique of inductive modeling. Unlike Forrester's model, this model provides a quantitative prediction. The model was tested on real data presented by World Bank and showed its high accuracy when forecasting for 1997-2012 years.*

***Keywords**: inductive modeling, world dynamics.*

***ACM Classification Keywords**: 1.6.4. Model validation and analysis*

Introduction

Forrester´s macroeconomic variables are as follows:

- population (P);
- capital assets(K);
- agricultural assets ratio(X);
- pollution (Z);
- natural resources (R).

They relate to overpopulation of our planet, lack of basic resources, critical level of pollution, food shortages and industrialization as well as the related industrial growth. For these variables a differential model has built, named "classical J.Forrester model" [Forrester, 1979] presented here in a generalized form:

$$\frac{dP}{dt} = f_1(P,K,X,Z,R) \quad (1)$$

$$\frac{dK}{dt} = f_2(P,K,X,Z,R) \quad (2)$$

$$\frac{dX}{dt} = f_3(P,K,X,Z,R) \quad (3)$$

$$\frac{dZ}{dt} = f_4(P,K,X,Z,R) \quad (4)$$

$$\frac{dR}{dt} = f_5(P,K,X,Z,R) \quad (5)$$

Many other researchers also have built their own models of world dynamics [Egorov, 1980; Matrosov, 1999; Matrosov, 1999; Makhov, 2010; Meadows, 2007]. All of them were based on differential equations and have the following two properties:

- they provide long-term qualitative prediction because they take into account only tendencies;
- they are very subjective because model forms are specified and based only on author´s vision.

[1]The work done under support of the British Petroleum grant (BP-RANEPA 2013).

A topical problem is to build models which could

- provide quantitative predictionfor short-term and middle-term period;
- be free from subjective preferences of authors.

An adequate tool to solve this problem is well-known Group Method of Data Handling (GMDH) which is one of the most successful techniques of Inductive Modeling (IM). This method was proposed in 1968 by Prof. A. Ivakhnenko and now it is developed by his disciples and followers [Stepashko, 2013]. Now GMDH has numerous applications in natural sciences, economical, technical and social areas [Bulakh, 2013; Bulgakova, 2013; Kovalchuk, 2013; Pavlov, 2013; Samoilenko, 2013; Tutova, 2013; Zubov, 2013]. In the paper we build IM-models for middle-term forecast of Forrester's variables and test their accuracy according to the data for the period 1998-2012. By IM-models we mean here the models built using GMDH.

The paper structure is the following. Section 2 contains a short description of GMDH and here we build IM-model. In section 3 we present the results of experiments. Section 4 includes short discussion and directions of future research.

GMDH technique

General description

GMDH has long enough history [Ivakhnenko, 1968; Ivakhnenko, 1971; Madala, 1994; Stepashko, 1983] and presented in many publications. In what follows, we only remind its generalized scheme:

1. Certain model class is chosen. Models of gradually increasing complexity are generated in this class.
2. Learning data set is divided into training set and checking set to form an external criterion. Models are built on the first set and tested in predictive mode on the second one.
3. Model parameters are estimated using any internal criterion (e.g. the least squares) on the checking set.
4. Model quality is determined using any external criterion (e.g. the regularity criterion) on the checking set.
5. Model complexity is increased until the external criterion would reach its minimum.

It is easy to see that:

- GMDH belongs to the data-driven technique of inductive modeling, which means the models to be built are based on given instances; they are not theory-driven ones.
- GMDH realizes so-called model self-organization approach that is the models change their structure trying to be adjusted according to a given data of observations.
- GMDH is an evolutionary algorithm because model complexity is increased to reach the best one reflecting properties of observation data.

The final IM-model has an optimal complexity with regard to a) reflection of object properties contained in observation data and b) robustness to unknown factors which we call noise.

The most common class of traditional IM-models being considered when using GMDH is so called Kolmogorov-Gabor polynomial (Ivakhnenko, 1971):

$$Y(x_1, x_2, \dots, x_n) = a_0 + \sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} x_i x_j + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} a_{ijk} x_i x_j x_k + \cdots \quad (6)$$

During preprocessing stageone can consider not only natural powers of variables but also factional powers and functions from initial variables, for example exponential or trigonometric.

The classes of predictive models

The models to be built belong to the class of nonlinear difference equations. We decided to check predictive properties of two model types. The first one contains separate variables in different powers without their combinations. The second one additionally contains pairwise multiplications to take into account their mutual influence. Consequently, these structures are "simple model" and "model with pairwise multiplications".

The first experiments did not give any satisfied results, so we decided to make an additional preprocessing:

- Normalization: the values of variables were changed to belong to the interval [0; 1] to have the same scale. To do this, the numerical values of the population size, the capital assets and agricultural assets ratio in each year were divided by 10^{10}, the quantity of remaining natural resources –by10^{12}.
- Transformations: the variables were allowed to have not only natural powers but also fractional ones to provide that any small change in the variables did not lead to the sharp jump in the model parameters.

Thus, the new model type had the following form:

$$A_{i,t+1}=f(A_{j,t}^{-3/2}, A_{j,t}^{-1}, A_{j,t}^{-1/2}, A_{j,t}^{1/2}, A_{j,t}, A_{j,t}^{3/2}, A_{j,t-1}, ..., A_{j,t-T}), \quad (7)$$

$$A_{i,t+1}=f(A_{j,t}^{-3/2}, A_{j,t}^{-1}, A_{j,t}^{-1/2}, A_{j,t}^{1/2}, A_{j,t}, A_{j,t}^{3/2}, A_{j,t-1}, ..., A_{j,t-T}, A_{j,t}^{1/2}\cdot A_{l,t}^{1/2}, A_{j,t}^{-1/2}\cdot A_{l,t}^{-1/2}) \quad (8)$$

here: $i=1,...,5$; $j=1,...,5$; $l=1,...,5$; $l\neq j$; T –the needed forecast horizon; $A_{i,t}$– the value of i-th variable in a year t; f() – a function linear in parameters.

Using these models we produced forecast for the period 15 years from 1997 to 2012. Then the result of prediction was compared with the real data provided by the World Bank. The World Bank provides only data on population, the capital assets and agricultural assets ratio, so the data on pollution and natural resources have been restored. Therefore, we could compare real and forecasted dynamics only for these three variables.

To build IM-models we used the program package GMDH Shell (GS) developed by GEOS company to solve problems of time series forecast, function approximation and data classification [http://www.gmdhshell.com/]. It contains several GMDH-based algorithms and broad possibilities for visualization as well as pre- and post-processing. In our research we used the iterative GMDH algorithm of neural network type.

Experimental study of two models

Selection of best models

We completed two series of experiments, in each series the best models from classes (7) and (8) were chosen. The limitations on the models were: powers no more than 3/2, number of lags no more than 15. The models were trained on the interval 1900-1998 and then tested on the data 1998-2012. Model of each class with the greatest accuracy of forecast was considered as "the best model".

For the experiments we prepared data for training and checking:

a) Period 1900-1960: we used data from Forrester's model.

b) Period 1961-1998: we used data from World Bank for population, capital assets and agricultural assets ratio. The other variables (pollution and resources) were restored with Inductive Modeling technique. Data from Forrester model were used up to 1970 and then forecast was made for one year. Thereafter, the numerical values of the population, capital assets and agricultural assets ratio were replaced with real data and then forecast for 1 year was made and so on.

c) Period 1999-2013: we restored data with Inductive Modeling technique and then compared the results with real data provided by the World Bank.

Forecast for 15 years with the best simple model

In the next experiments the forecast horizon was equal to 15 years. The best IM-model in the class of "simple" models is:

$P_{t+1} = -\,0.00603058 + 0.00521011\cdot Z_t^{-3/2} + 0.0255463\cdot X_{t-7} + 1.22781\cdot P_t - 0.182304\cdot P_{t-2}$

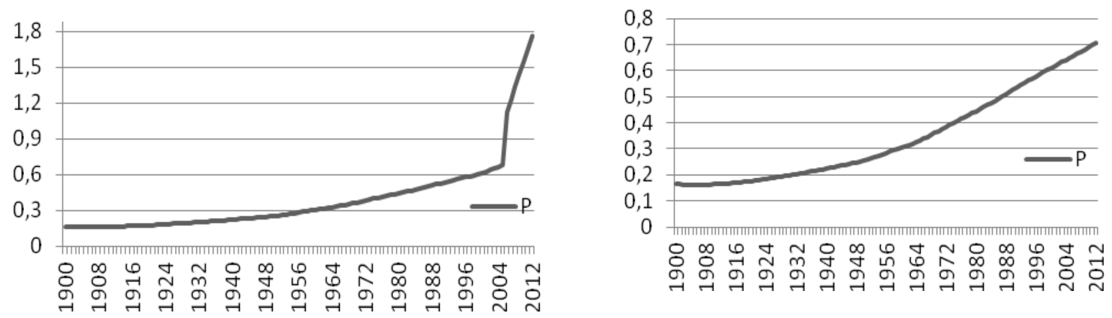
$K_{t+1} = 0.00361846 + 1.38141\cdot K_t - 0.372398\cdot K_{t-1}$

$X_{t+1} = -\,0.01487678 - 0.2118857\cdot X_t + 0.943634\cdot X_{t-5} + 1.6358\cdot X_t^{3/2}$

$Z_{t+1} = -\,0.00551411 + 0.787061\cdot P_t^{-1/2}$

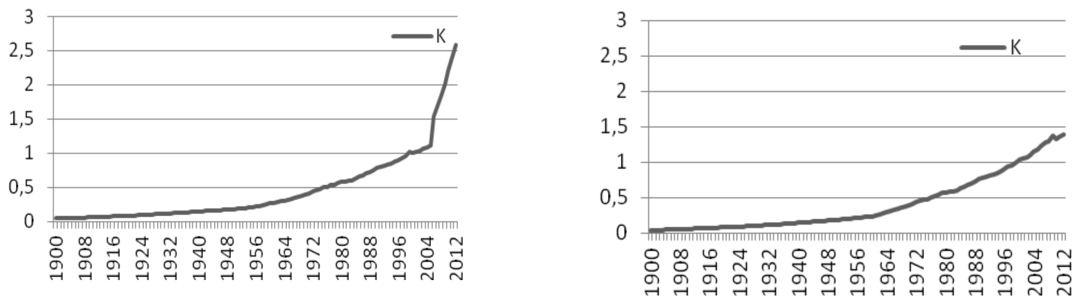
$R_{t+1} = -\,0.852927 - 0.0110949\cdot P_{t-10} - 1.94371\cdot Z_t^{1/2} + 0.965339\cdot P_t^{1/2}$

The real and forecast dynamics is compared below(fig. 7-8). The accuracy results of modeling are: R^2=0.98, RMSE=1.54%. Here R^2 is the coefficient of determination; RMSE is the root-mean-square-error.



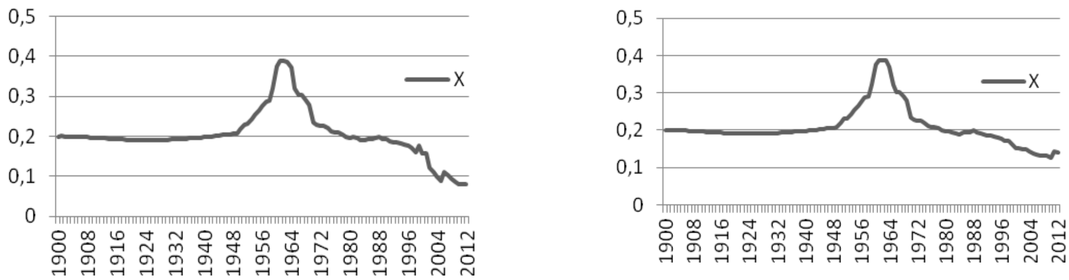
Forecast Real data

Fig. 7. Comparison of real and forecast dynamics of population



Forecast Real data

Fig. 8. Comparison of real and forecast dynamics of capital assets



Forecast Real data

Fig. 9. Comparison of real and forecast dynamics of the agricultural assets ratio

Forecast on 15 years with the best model having pairwise multiplications

The best IM-model in the class of models with pairwise multiplications is:

$$P_{t+1} = -\,0.001603058 - 0.0468069 \cdot P_{t-2} + 0.0255463 \cdot X_{t-7} + 0.09195 \cdot Z_t^{-3/2}$$

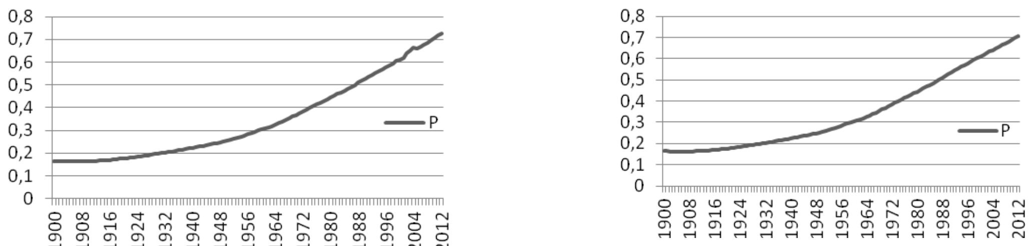
$$K_{t+1} = -\,0.00187356 - 0.165445 \cdot K_t \cdot X_t + 1.00087 \cdot K_t$$

$$X_{t+1} = -\,0.000409618 - 0.132082 \cdot t^{1/2} - 3.34978 \cdot X_t^{1/2} + 1.08047\, X_t^{3/2}$$

$$Z_{t+1} = -\,1.98521e\text{-}16 + 0.787061 \cdot Z_t \cdot P_t^{-1/2}$$

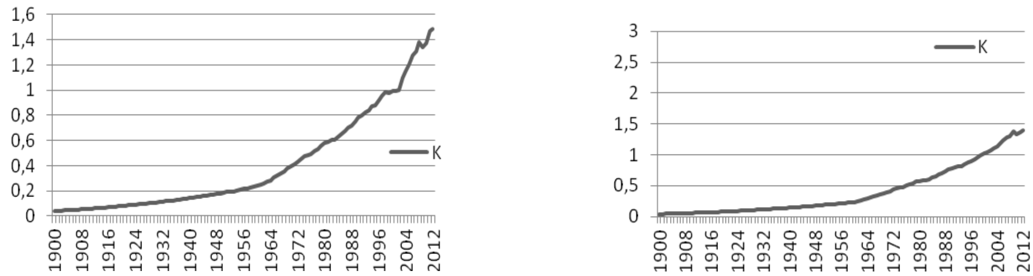
$$R_{t+1} = 0.852927 - 0.0110949 \cdot P_{t-10} - 1.94371 \cdot Z_t^{1/2} + Z_t^{1/2} \cdot P_t^{1/2}$$

The real and forecast dynamics is compared below(fig. 10-12). The accuracy results of modeling are: R^2=0.99, RMSE = 1,08%.



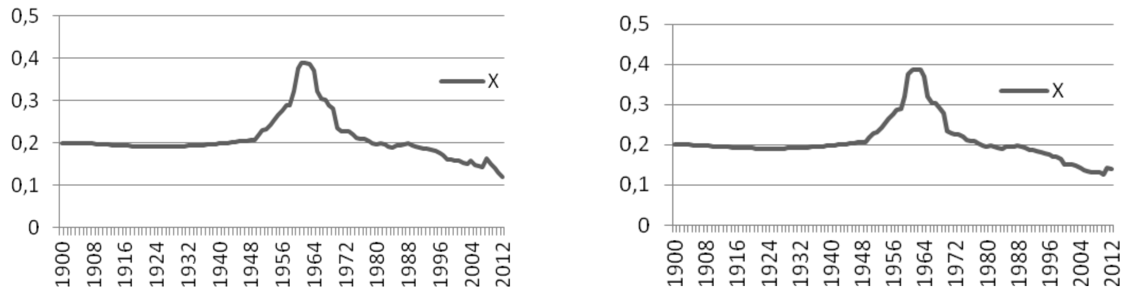
Forecast Real data

Fig. 10. Comparison of real and forecast dynamics of population



Forecast Real data

Fig. 11. Comparison of real and forecast dynamics of capital assets



Forecast Real data

Fig. 12. Comparison of real and forecast dynamics of agricultural assets ratio

One can see that IM-model with pairwise multiplications imitates real data much better.

Conclusion

This paper presents some middle-term forecasting models of world dynamics with Forrester´s variables, which were built using GMDH technique. The main results are the followings:

- There is a fundamental possibility to obtain enough accurate forecasts using GMDH with a polynomial-type class of nonlinear difference models;
- Successful modeling needs careful preprocessing including normalization and variable transformation;
- Models with pairwise multiplications of variables prove to be essentially better than the models without them.

In the future we plan to study in detail various classes of models, various GMDH algorithms, and give forecasts of Forrester's variables for future periods.

Bibliography

[Bulakh, 2013] Bulakh V., Perep'olkina L., Sinelnikova O. Models and Estimation Methods for the city in the real estate market. ICIM 2013, 4th international conference on inductive modeling, pp. 251-253.

[Bulgakova, 2013] Bulgakova O. Modeling the Ukraine Black Sea economic region GRP dependence on socio-economic indicators. ICIM 2013, 4th international conference oninductive modeling, pp. 254-260.

[Egorov, 1980] Egorov V., Kallistov N., Mitrofanov V., Piontkovskii A. Mathematical models of global development: a critical analysis of the patterns of nature. - Gidrometeoizdat, 1980. - 192p.

[Forrester, 1979] Forrester J. World Dynamics. Productivity Pr; 2nd edition, 1979, 142 p.

[Ivakhnenko, 1982] Ivakhnenko A.: The inductive method of self-organizing models of complex systems. Kiev, Naukova Dumka, 1982, 296 p.(In Russian)

[Kovalchuk, 2013] Kovalchuk P.I., Gerus A.V., Kovalchuk V.P. Perceptron model of system environmental assessment of water quality in river basins. ICIM 2013, 4th international conference on inductive modeling, pp. 279-284.

[Makhov, 2010] Makhov S. Long-term trends and forecasts from the standpoint of the new global dynamics models / simulations and forecast crises and global dynamics / Ed. Akayev, Korotaev, Malinetskii / Future Russia. - Moscow: LKI, 2010. - P. 262 - 276. (In Russian)

[Matrosov, 1999] Matrosov V., Matrosov I. Global modeling taking into account biomass dynamics and scenarios for sustainable development / New development paradigm forRussia (Comprehensive study of the problems of sustainable development). - Moscow: Academia, MGUK, 1999. - P. 18 - 24. (In Russian)

[Matrosova, 1999] Matrosova K. Sustainable development in the modified mathematical model "World Dynamics" / New development paradigm for Russia (Comprehensive study of the problems of sustainable development). - Moscow: Academia, MGUK, 1999. - P. 344-353. (In Russian)

[Meadows, 1973] Donella M. Meadows, Donella H. Meadows, Dennis L. Meadows, Tzonis. Toward Global Equilibrium: Collected Papers. Productivity Press Inc, 1973, 358 p.

[Pavlov, 2013] Pavlov A., Kondrashova N. Application of generalized relaxational-iterative algorithm for solving the "space weather" forecast problem. ICIM 2013, 4th international conference on inductive modeling, pp. 302-306.

[Samoilenko, 2013] Samoilenko O.A. Combinatorial GMDH Algorithm Application to predict the Dynamic of Demographic Characteristics in Ukraine. ICIM 2013, 4th international conference on inductive modeling, pp. 312-314.

[Stepashko, 2013] Stepashko, V. Ideas of academician O. Ivakhnenko in Inductive Modeling field from historical perspective / Proc. of 4th Intern. Conf. on Inductive Modeling (ICIM-2013), IRTC of NAS of Ukraine, Kyiv, 2013, pp. 31-37.

[Tutova, 2013] Tutova O.,Savchenko Ie. Modeling of impact of macroeconomic indicators on the growth of national income. ICIM 2013, 4th international conference on inductive modeling, pp. 337-343.

[Zubov, 2013] Zubov D. Average daily temperature's long-range forecast using inductive modeling and satellite datasets. ICIM 2013, 4th international conference on inductive modeling, pp. 348-353.

Authors' Information

Olga Proncheva – M.Sc, 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, MoscowRegion, 141700, Russia

e-mail: olga.proncheva@gmail.com

Major Fields of Scientific Research: macroeconomics, mathematical modelling, applied mathematics

Mikhail Alexandrov – Professor, Academy of national economy and civil service under the President of Russia; 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

Volodymyr Stepashko –Head of Department for Information Technologies ofInductive Modeling of IRTC ITS, Professor, Dr Sci, P.A.: 40, Akademik Glushkov Prospect,Kyiv, Ukraine, 03680;

e-mail: stepashko@irtc.org.ua

Main Fields of Scientific Research: data mining, knowledge discovery, information technologies,inductive modelling, Group Method of Data Handling (GMDH)

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

e-mail: koshulko@gmail.com

Major Fields of Scientific Research: data science, parallel processing.