2.2 Fuzzy and interval approaches in inductive modelling

The Investigations of Fuzzy Group Method of Data Handling with Fuzzy Inputs in the Problem of Forecasting in the Financial Sphere

Yuriy Zaychenko

Institute for Applied System Analysis, National Technical University of Ukraine "KPI"
03056, Kiev-56, Peremogy Avenue, 37, Ukraine

baskervil@voliacable.com zaych@i.com.ua

Abstract. *The problem of forecasting models constructing using experimental data in terms of fuzziness, when input variables are not known exactly and determined as intervals of uncertainty is considered in this paper. The fuzzy group method of data handling is proposed to solve this problem. The theory of this method was suggested and researched in [1-3]. As it is well known, fuzzy GMDH allows to construct fuzzy models and has the following advantages:*

1)The problem of optimal model finding is transformed to the problem of linear programming, which is always solvable;

2)There is interval regression model built as the result of method work out;

3)There is a possibility of adaptation of the obtained model;

The mathematical model of the problem mentioned above is built and fuzzy GMDH with fuzzy inputs is elaborated in the paper. The corresponding software kit, which uses the suggested algorithm, was developed. And also the experimental investigations and comparison of FGMDH with GMDH and neural nets in problems of stock prices forecasting was carried out and presented..

Keywords

Group method of Data Handling, fuzzy, stock prices, forecasting

1 Math Model of Group Method of Data Handling with Fuzzy Input Data

1.1General view of FGMDH model with fuzzy input data

Let's consider a linear interval regression model:

$$Y = A_0Z_0+A_1Z_1+\ldots+A_nZ_n , \qquad (1.1)$$

where Ai are fuzzy numbers, which are described by threes of parameters $A_i = (\underline{A_i}, \breve{A}_i, \overline{A_i})$, where $\breve{A}_i$ – interval center, $\overline{A_i}$ – upper border of the interval, $\underline{A_i}$ - lower border of the interval,

and Zi – also fuzzy numbers, which are determined by parameters $(\underline{Z_i}, \breve{Z}_i, \overline{Z_i})$, $\underline{Z_i}$ - lower border, $\breve{Z}_i$ - center, $\overline{Z_i}$ - upper border of fuzzy number.

Then Y – output fuzzy number, which parameters are defined as follows (in accordance with L-R numbers multiplying formulas):

Center of interval:

$$\breve{y} = \sum \breve{A}_i * \breve{Z}_i ,$$

And lower border of the interval

$$\underline{y} = \sum (\breve{A}_i * \breve{Z}_i - |\breve{A}_i| * (\breve{Z}_i - \underline{Z}_i) - (\breve{A}_i - \underline{A}_i) * |\breve{Z}_i|)$$

Thus upper border of the interval

$$\overline{y} = \sum (|\breve{A}_i| * (\overline{Z}_i - \breve{Z}_i) + |\breve{Z}_i| * (\overline{A}_i - \breve{A}_i) + \breve{A}_i * \breve{Z}_i).$$

For the interval model to be correct, the real value of input variable Y is needed to lay in the interval got by the method workflow.

So, the general requirements to estimation linear interval model are to find such values of parameters $(\underline{A_i}, \breve{A}_i, \overline{A_i})$ of fuzzy coefficients, which allow:

a) Observed values y_k lay in estimation interval for Y_k;

b) Total width of estimation interval is minimal.

Input data for this task is Zk=[Zki]i - input training sample, and also yk – known output values, $k = \overline{1, M}$, M – the number of observation points.

There are two cases of fuzzy values membership function described in this work:

Triangular membership functions
Gaussian membership functions.

1.2. FGMDH with fuzzy input data for triangular membership functions

Let's consider the linear interval regression model:

$$Y = A_0 Z_0 + A_1 Z_1 + \ldots + A_n Z_n \quad , \tag{1.1}$$

where Ai – fuzzy number of triangular shape, which is described by threes of parameters $A_i = (\underline{A_i}, a_i, \overline{A_i})$, where a_i – center of the interval, $\overline{A_i}$ – its upper border, $\underline{A_i}$ - its lower border.

Current task contains the case of symmetrical membership function for parameters Ai, so they can be described via pair of parameters (a_i , c_i).

$$\underline{A}_i = a_i - c_i, \ \overline{A}_i = a_i + c_i, \ c_i \text{ – interval width, } c_i \geq 0,$$

Zi – also fuzzy numbers of triangular shape, which are defined by parameters $(\underline{Z_i}, \breve{Z}_i, \overline{Z_i})$, $\underline{Z_i}$ - lower border, $\breve{Z}_i$ - center, $\overline{Z_i}$ - upper border of fuzzy number.

Then Y – fuzzy number, which parameters are defined as follows:

Center of the interval:

$$\breve{y} = \sum a_i * \breve{Z}_i,$$

Deviation in the left part of the membership function:

$$\breve{y} - \underline{y} = \sum (a_i * (\breve{Z}_i - \underline{Z}_i) + c_i |\breve{Z}_i|)$$

Lower border of the interval:

$$\underline{y} = \sum (a_i * \underline{Z}_i - c_i |\breve{Z}_i|)$$

Deviation in the right part of the membership function:

$$\overline{y} - \breve{y} = \sum (a_i * (\overline{Z}_i - \breve{Z}_i) + c_i |\breve{Z}_i|) = \sum a_i \overline{Z}_i - a_i \breve{Z}_i + c_i |\breve{Z}_i|$$, so

Upper border of the interval:

$$\overline{y} = \sum (a_i * \overline{Z}_i + c_i |\breve{Z}_i|)$$

For the interval model to be correct, the real value of input variable Y should lay in the interval got by the method workflow.

It can be described in such a way:

$$\begin{cases} \sum (a_i * \underline{Z}_{ik} - c_i |\breve{Z}_{ik}|) \le y_k \\ \sum (a_i * \overline{Z}_{ki} + c_i |\breve{Z}_{ik}|) \ge y_k, k = \overline{1,M} \end{cases}$$

Where Zk=[Zk]i is input training sample, yk –known output values, $k = \overline{1,M}$, M – number of observation points.

So, the general requirements to estimation linear interval model are to find such values of parameters (a_i, c_i) of fuzzy coefficients, which enable:

a) Observed values y_k lay in estimation interval for Y_k;

b) Total width of estimation interval is minimal.

These requirements can be redefined as a task of linear programming:

$$\min_{a_i, c_i} \sum_{k=1}^{M} (\sum (a_i * \overline{Z}_i + c_i |\breve{Z}_i|) - \sum (a_i * \underline{Z}_i - c_i |\breve{Z}_i|)) \quad (1.2)$$

under constraints:

$$\begin{cases} \sum (a_i * \underline{Z}_{ik} - c_i |\breve{Z}_{ik}|) \le y_k \\ \sum (a_i * \overline{Z}_{ki} + c_i |\breve{Z}_{ik}|) \ge y_k, k = \overline{1,M} \end{cases} \quad (1.3)$$

Let's consider partial description

$$f(x_i, x_j) = A_0 + A_1 x_i + A_2 x_j + A_3 x_i x_j + A_4 x_i^2 + A_5 x_j^2 \quad (1.4)$$

Rewriting it in accordance with the model (1.1) needs such substitution $z_0 = 1$, $z_1 = x_i$,

$$z_2 = x_j,\ z_3 = x_i x_j,\ z_4 = x_i^2,\ z_5 = x_j^2.$$

Then math model (1.2)-(1.3) will take the form

$$\begin{aligned} \min_{a_i, c_i} \ & (2Mc_0 + a_1 \sum_{k=1}^{M} (\overline{x}_{ik} - \underline{x}_{ik}) + 2c_1 \sum_{k=1}^{M} |\breve{x}_{ik}| + a_2 \sum_{k=1}^{M} (\overline{x}_{jk} - \underline{x}_{jk}) + 2c_2 \sum_{k=1}^{M} |\breve{x}_{jk}| + \\ & + a_3 \sum_{k=1}^{M} (|\breve{x}_{ik}| (\overline{x}_{jk} - \underline{x}_{jk}) + |\breve{x}_{jk}| (\overline{x}_{ik} - \underline{x}_{ik})) + 2c_3 \sum_{k=1}^{M} |\breve{x}_{ik} \breve{x}_{jk}| + 2a_4 \sum_{k=1}^{M} |\breve{x}_{ik}| (\overline{x}_{ik} - \underline{x}_{ik}) + \\ & + 2c_4 \sum_{k=1}^{M} \breve{x}_{ik}^2 + 2a_5 \sum_{k=1}^{M} |\breve{x}_{jk}| (\overline{x}_{jk} - \underline{x}_{jk}) + 2c_5 \sum_{k=1}^{M} \breve{x}_{jk}^2) \end{aligned}$$

with the following conditions:

$$a_0 + a_1 \underline{x}_{ik} + a_2 \underline{x}_{jk} + a_3(-|\breve{x}_{ik}|(\breve{x}_{jk} - \underline{x}_{jk}) - |\breve{x}_{jk}|(\breve{x}_{ik} - \underline{x}_{ik}) + \breve{x}_{ik}\breve{x}_{jk}) +$$
$$+ a_4(-2|\breve{x}_{ik}|(\breve{x}_{ik} - \underline{x}_{ik}) + \breve{x}_{ik}{}^2) + a_5(2|\breve{x}_{jk}|(\breve{x}_{jk} - \underline{x}_{jk}) + \breve{x}_{jk}{}^2) - c_0 - c_1|\breve{x}_{ik}| -$$
$$- c_2|\breve{x}_{jk}| - c_3|\breve{x}_{ik}\breve{x}_{jk}| - c_4\breve{x}_{ik}^2 - c_5\breve{x}_{jk}^2 \le y_k$$
$$a_0 + a_1\bar{x}_{ik} + a_2\bar{x}_{jk} + a_3(|\breve{x}_{ik}|(\bar{x}_{jk} - \breve{x}_{jk}) + |\breve{x}_{jk}|(\bar{x}_{ik} - \breve{x}_{ik}) - \breve{x}_{ik}\breve{x}_{jk}) + a_4(2|\breve{x}_{ik}|(\bar{x}_{ik} -$$
$$- \breve{x}_{ik}) - \breve{x}_{ik}{}^2) + a_5(2|\breve{x}_{jk}|(\bar{x}_{jk} - \breve{x}_{jk}) - \breve{x}_{jk}{}^2) + c_0 + c_1|\breve{x}_{ik}| + c_2|\breve{x}_{jk}| + c_3|\breve{x}_{ik}\breve{x}_{jk}| +$$
$$c_4\breve{x}_{ik}^2 + c_5\breve{x}_{jk}^2 \ge y_k$$
$$c_l \ge 0, \quad l = \overline{0,5}.$$

The task (1.5)-(1.7) can be solved using simplex-method. Having optimal values of dual variables $\{\delta_k\}$, $\{\delta_{k+M}\}$, we easily obtain the optimal values of desired variables c_i, a_i, $i = \overline{0,5}$, and also a desired fuzzy model for given partial description.

Additionally the corresponding model with fuzzu inputs for Gaussian membership functions (MF) was elaborated which has a similar form like tha model with triangular MF.

3 Experimental Results of FGMDH with Fuzzy Input Data in RTS Index Forecasting

For estimation of efficiency of the suggested FGMDH method with fuzzy inputs the corresponding software kit was elaborated and numerous experiments of financial markets forecasting were carried out. Some of them are presented below.

3.1 Forecasting of RTS index

In this experiment we used 5 fuzzy input variables, which represent stock prices of leading Russian energetic companies, which are included to the list of computations of RTS index:

LKOH – shares of “LUKOIL” joint-stock company, EESR – shares of “РАО ЕЭС России“joint-stock company, YUKO – shares of “ЮКОС” joint-stock company, SNGSP – privileged shares of “Сургутнефтегаз” joint-stock company,SNGS – common shares of “Сургутнефтегаз” joint-stock company.

Output variable is the RTS (opening price) index value of the same period (03.04.2006 – 18.05.2006). Sample size – 32 values.

Training sample size – 18 values (optimal size of training sample for current experiment).

The following results were obtained:

Tab.1. For normalized data

	Triangular MF	Gaussian MF
MSE	0.055557	0.028013

Tab.2. For non-normalized data

	Triangular MF	Gaussian MF
MSE	18.48657	9.321461
MAPE	0.8%	0.4%

As we can see from the results of experiment 1, forecasting using triangular and Gaussian membership functions gives good results. Results of experiments with Gaussian MF are better than results of experiments with triangular MF.

Experiment 2. Forecasting of RTS index (closing price)

This experiment uses the same input variables as the experiment 1 does.

Output variable is the value of RTS index (closing price) for the same period (03.04.2006 – 18.05.2006). Sample size – 32 values.

Training sample size – 18 values (optimal size of training sample for current experiment).

The following results were obtained:

Tab. 3. For normalized data

	Triangular MF	Gaussian MF
MSE	0.057379	0.029582

Tab.4. For non-normalized data

	Triangular MF	Gaussian MF
MSE	18.04394	9.302766
MAPE	0.78%	0.37%

As we can see from the results of experiment 2, forecasting using triangular and Gaussian membership functions gives good results. Results of experiments with Gaussian MF are better than results of experiments with triangular MF.

4. Conclusion

In this paper a new method of inductive modeling FGMDH with fuzzy inputs was suggested. This method represents the development of fuzzy GMDH when information is fuzzy and given in the form of uncertainty intervals. The mathematical model was constructed and corresponding algorithm was elaborated. The experimental results of application of the suggested method in the forecasting of market index and stock prices are presented and discussed. The main advantages of the suggested method are following:

- It operates with fuzzy and uncertain input information and constructs the fuzzy model;
- The constructed model has minimal possible total width and in this sense is optimal;
- For finding optimal model we solve corresponding linear programming problem which is always solvable for this task;
- We should not a priori set the form of a model the algorithm finds it itself using the ideas of evolution.

References

[1] Zaychenko Yu. “The Fuzzy Group Method of Data Handling and Its Application for Economical Processes Forecasting” - Scientific Inquiry, - Vol. 7, No.1, June, 2006 - p.83-96.

[2] Zaychenko Yu. “Fuzzy method of inductive modeling in problems of macroeconomic indexes forecasting.” System researches and informational technologies, #3 of 2003, p. 25-45.

[3] Zaychenko Yu. P., Zayetz I.O., O.V. Kamotsky, O.V. Pavlyuk. Research of different kinds of membership functions of fuzzy forecasting models parameters in fuzzy group method of data handling. USiM, 2003, #2, p.56-67.