Fuzzy modelling for tasks of management of the agricultural-industrial complex

V.E. Parfenova, G.G. Bulgakova, Yu.V. Amagaeva, K.V. Evdokimov*
Saint-Petersburg State Agrarian University, Pushkin, Saint-Petersburg, 196601, Russia
Corresponding author *: ekvmanager@mail.ru

Abstract The problem of modeling and accounting for uncertainty in modern tasks of management is relevant. The effectiveness of the decisions depends significantly on the methods for describing the uncertainty in the problem. The greatest development in agrarian science received optimization and econometric models. However, these models are based on quantitative determined initial information and accounting for uncertainty as randomness, which is described by probabilistic and statistical methods. Meanwhile, many modern decision-making tasks in planning and managing agricultural production are characterized by the presence of uncertain factors, as well as the availability of high-quality, inaccurate or incomplete information. To account and describe such uncertainty, an alternative approach to the probabilistic approach is needed. Fuzzy set theory is one of the most effective mathematical tools aimed at formalizing and processing uncertain information. The econometrics section related to using fuzzy set theory in regression analysis is developing methods of fuzzy regression modeling. The possibilities of using fuzzy regression modeling tools for analyzing management processes of agricultural production are discussed in this article.

Keywords: fuzzy modeling, regression, analysis, agricultural production.

1. Introduction
The problem of accounting for uncertainty and incompleteness of data is known to occupy a key place in solving the problems of managing an economic object, including management in the agro-industrial complex (AIC). Let us see how this problem is solved when modeling agricultural processes. Optimization and econometric models have received the greatest application in agrarian science and practice.

The currently developed quantitative methods for making optimal decisions in the management of the agro-industrial complex are based on the use of quantitative clear information and make it possible to choose the best of many possible solutions in two cases: under conditions of full certainty (deterministic models) or under the conditions of one particular type of uncertainty, namely uncertainty of probabilistic nature (stochastic models). Meanwhile, the requirement of determinacy of models is a simplification of reality; since in real life, situations devoid of uncertainty are the exception rather than the rule. As for accounting for uncertainty, there are various types of uncertainty in the decision-making process, of-ten not of a probabilistic nature.

Many modern decision-making tasks in planning and managing agricultural production are characterized by the presence of uncertain factors, as well as the availability of high-quality, inaccurate or incomplete information. Using a probabilistic approach to describing uncertainty, in this case, is not adequate to the
nature of uncertainty. A different approach is needed. One of the most effective mathematical tools aimed at formalizing and processing uncertain information is fuzzy set theory. This theory allows considering various types of uncertainty from a unified point of view. The branch of mathematics related to the search for optimal solutions based on fuzzy information is called fuzzy mathematical programming (FMP). This device, as shown in [1, 2], is widely used in the formulation and solution of problems of economics and management. However, specialists in mathematical, informational and technical specialties mainly use it. The possibilities of using fuzzy tools in modeling and solving optimization problems of agricultural production were considered earlier [3].

2. The problem under study.

This article is devoted to the consideration of econometric methods, in particular, regression modeling and analysis methods, which, along with optimization, have been widely used in the study of agricultural processes. The task of the regression analysis is to identify the dependence of a certain indicator on various characteristics and display this relationship in the form of a model. In classical regression analysis, the data for both the dependent variables and the independent variables are numeric [10-12]. At the same time, accurate numerical statistical information is necessary to obtain a qualitative regression model, since uncertainty is interpreted as randomness. In this case, to obtain estimates of model parameters, using statistical methods is justified.

Only approximate and fuzzy data are available in many modern applied problems, or they are not enough to establish a reliable type of probability distribution. In this case, the uncertainty becomes fuzzy, for which traditional regression modeling methods are not applicable [13,14]. As mentioned above, the best way to formalize approximate and fuzzy information is tools of the theory of fuzzy sets. The section of econometrics related to building a regression in a fuzzy form is called fuzzy regression modeling. Methods of fuzzy regression modeling are quite developed today [4, 5, 6]. However, to solve the problems of the agricultural sector, they have not yet found proper use.

There are two groups among the currently available methods of fuzzy regression modeling. One is based on the least squares method, the other - on linear programming. Building a fuzzy regression model for the analysis of the agricultural production process using the linear programming method is proposed in this article [15].

Different variants of fuzzy regression models are considered in the scientific literature. Independent variables (regressors) of regression can be both clear, with fuzzy coefficients and fuzzy with clear coefficients, or with fuzzy and regressors and coefficients of the model. In this case, the data for fuzzy regression can be specified interval, fuzzy or deterministic. Most modern fuzzy regression models have a linear structure [7, 8]. We consider fuzzy linear multiple regression modeling with fuzzy coefficients, clear regressors, and deterministic input data.

3. Problem statement and algorithm of its decision.

Let us set the following problem of fuzzy linear regression. We have \( n \) observational results of the dependent variable \( y_j, j = 1 \div n \). and \( m \ast n \) values of independent variables \( x_{ij}, i = 1 \div m; j = 1 \div n \) (Table 1).

| № observation | \( y \) | \( x_1 \) | \( x_2 \) | ... | \( x_m \) |
|---------------|-------|-------|-------|-----|-------|
| 1             | \( y_1 \) | \( x_{11} \) | \( x_{21} \) | ... | \( x_{m1} \) |
| 2             | \( y_2 \) | \( x_{12} \) | \( x_{22} \) | ... | \( x_{m2} \) |
| ...           | ...   | ...   | ...   | ... | ...   |
| \( n \)       | \( y_n \) | \( x_{1n} \) | \( x_{2n} \) | ... | \( x_{mn} \) |
We search for the resulting variable \( y \) in the form of a fuzzy multiple regression equation:

\[
\hat{y} = \hat{A}_0 + \hat{A}_1 x_1 + \ldots + \hat{A}_m x_m,
\]

where \( x_i \) — deterministic values, and \( \hat{y} \) — fuzzy values (~fuzziness symbol).

The task of building a fuzzy linear regression model, in this case, is to select fuzzy parameters \( \hat{A}_j, j=0+ m \) so that the following two conditions are satisfied:

1) The overall fuzziness of the model should be minimal.

2) The task of building a fuzzy linear regression model, in this case, is to select fuzzy parameters \( \hat{A}_j, j=0+ m \) that ensure the following two conditions are satisfied:

\[
y_i \text{ is a fuzzy number with } \mu_\text{F} = \mu(y_i) \geq h,
\]

where \( h \) is the specified threshold of reliability, and \( y_i \) is the dependent numeric value in the fuzzy regression equation:

\[
\text{equality } \sum_{i=1}^{n} b_j x_{ij},
\]

where \( b_j \) are fuzzy parameters of the fuzzy linear programming problem with the criterion (\text{V}1) and constraints (\text{V}2).

We will look for the coefficients \( \hat{A}_j \) in the form of symmetric triangular fuzzy numbers: \( \hat{A}_j = (a_j – b_j, a_j, a_j + b_j) \), \( b_j \geq 0 \). Then the dependent quantities \( \hat{y}_i \) are also symmetric fuzzy numbers, which are defined as

\[
\hat{y}_i = (c_i – d_i, c_i, c_i + d_i),
\]

where \( c_i = \hat{a}_0 + \sum_{j=1}^{m} b_j x_{ij} \). The total measure of fuzziness is equal to \( d = b_0 + \sum_{j=1}^{m} b_j x_{ij} \).

The task of determining the coefficients of fuzzy regression is reduced to the formulation and solution of the linear programming problem with the criterion (\text{V}2) and constraints (\text{V}1).
Formally, the task is to find the unknown coefficients $a_j$, $b_j$, $j=0÷m$, at which:

$$d=\sum_{i=1}^{n} \sum_{j=1}^{m} b_j \left| x_{ij} \right| \rightarrow \min,$$

under constraints:

$$\sum_{j=1}^{m} a_j x_{ij} - (1-h)(\sum_{i=1}^{n} b_j \left| x_{ij} \right|) \leq y_i, \quad i=1÷n,$$

$$\sum_{j=1}^{m} a_j x_{ij} + (1-h)(\sum_{i=1}^{n} b_j \left| x_{ij} \right|) \geq y_i, \quad i=1÷n,$$

$$b_j \geq 0, \quad j=0÷m.$$

4. Experiment

Based on the information data set, 15 agricultural enterprises of the Leningrad region form the dependence of annual revenue from sales per 1 hectare of agricultural land $y$ (thousand rubles) on the cost of goods sold per 1 hectare of agricultural land $x_1$ (thousand rubles) and average annual value of fixed assets per 1 hectare of agricultural land $x_2$ (thousand rubles). Using the method analyzed in the article, we construct a fuzzy equation of a multiple (two-factor) model.

| № enterprise | Revenue, thousand rubles $y$ | Costs, thousand rubles $x_1$ | Average annual value of fixed assets, thousand rubles $x_2$ |
|--------------|-----------------------------|-------------------------------|-------------------------------------------------|
| 1            | 71                          | 52.5                          | 90.3                                           |
| 2            | 46.4                        | 38.5                          | 105.7                                          |
| 3            | 42.5                        | 35.9                          | 32.7                                           |
| 4            | 47.6                        | 39.6                          | 93.4                                           |
| 5            | 35.4                        | 29.9                          | 27.2                                           |
| 6            | 42.5                        | 42.1                          | 30.3                                           |
| 7            | 36.9                        | 25.9                          | 25.8                                           |
| 8            | 34.1                        | 29.7                          | 38.1                                           |
| 9            | 21.3                        | 11                            | 16.2                                           |
| 10           | 23.7                        | 15.8                          | 30                                             |
| 11           | 54.8                        | 32.6                          | 37.4                                           |
| 12           | 23.3                        | 15.3                          | 23.6                                           |
| 13           | 25.6                        | 23.5                          | 20.2                                           |
| 14           | 37.8                        | 37.6                          | 44.9                                           |
| 15           | 50.7                        | 43.6                          | 34.3                                           |

Fuzzy linear two-factor regression has the following form:

$$\tilde{y}=(\tilde{y}_1, \tilde{y}_2)=A_0 + A_1 x_1 + A_2 x_2.$$  (2)

Building a regression is reduced to the determination of its coefficients. Let the regression coefficients (2) are symmetric triangular fuzzy numbers: $A_1=(a_1 - b_1, a_1, a_1 + b_1)$, $A_2=(a_2 - b_2, a_2, a_2 + b_2)$. Then for each $y_i$ ($i=1÷15$) a fuzzy value $\tilde{y}_i=(c_i - d_i, c_i, c_i + d_i)$ is defined, where
\[
\sum_{i=1}^{2} a_i x_i \quad \sum_{i=1}^{2} b_i x_i \quad \sum_{i=1}^{15} d_i \quad \sum_{i=1}^{15} b_i x_i
\]

\[
c_i = a_0 + a_1 x_1 + a_2 x_2 = a_0 + \sum_{j=1}^{2} a_j x_j ; \quad d_i = b_0 + b_1 x_1 + b_2 x_2 = b_0 + \sum_{j=1}^{2} b_j x_j .
\]

General fuzziness

\[
d = a_0 + a_1 x_1 + a_2 x_2 = b_0 + b_1 x_1 + b_2 x_2 = a_0 + \sum_{i=1}^{15} d_i = \sum_{i=1}^{15} b_i x_i \quad i = 1, 15 .
\]

We have the following linear programming problem:

\[
\begin{align*}
\sum_{i=1}^{15} \sum_{j=1}^{2} b_j x_i & \to \text{min}, \\
\sum_{i=1}^{15} \sum_{j=1}^{2} a_j x_i & - (1 - h)[\sum_{j=1}^{2} b_j x_i] & \leq y_i, \quad i = 1, 15; \\
\sum_{i=1}^{15} \sum_{j=1}^{2} a_j x_i & + (1 + h)[\sum_{j=1}^{2} b_j x_i] & \geq y_i, \quad i = 1, 15; \\
(\sum_{i=1}^{2} b_j x_i) & \geq 0 .
\end{align*}
\]

Problem (3) has 6 unknowns and 33 limitations. The task was solved using the built-in function “Search for a solution” in MS Excel.

5. Results analysis

The values \( h = 0, 0.25, 0.5, 0.75 \) were considered (Table 3). Depending on the given value \( h \), various solutions of the desired parameters of the fuzzy regression were found. The greater \( h \), the more reliable we want to get the values of the value of the effective index. The fuzziness of the effective index increases.

| \( h \)  | 0    | 0.25 | 0.5  | 0.75 | 0     |
|---------|------|------|------|------|-------|
| \( d \) | 153.7| 204.9| 307.3| 614.7| 197.6 |
| \( a_0 \) | 8,036| 8,036| 8,036| 8,036| 0     |
| \( a_1 \) | 0,958| 0,958| 0,958| 0,958| 0,851 |
| \( a_2 \) | 0,132| 0,132| 0,132| 0,132| 0,433 |
| \( b_0 \) | 0    | 0    | 0    | 0    | 0     |
| \( b_1 \) | 0,324| 0,433| 0,649| 1,298| 0     |
| \( b_2 \) | 0    | 0    | 0    | 0    | 0,304 |

The last column of the table shows the coefficients of the fuzzy regression without taking into account unaccounted factors \( a_0 = 0 \). In this case, the general uncertainty of the model with the same value \( h = 0 \) turned out to be higher than when taking into account unaccounted factors \( a_0 \neq 0 \). Therefore, a model with \( a_0 \neq 0 \) is preferable.

It can be seen from the table (3) that the fuzzy coefficient in the regression (2) associated with the fuzziness of the dependent indicator (revenue) for all values \( h \) is the degree of confidence, the coefficient \( A_1 \) with the explanatory variable “costs”.

Let us write the regression model at \( h = 0, a_0 \neq 0 \). We have

\[
\tilde{y} = 8,036 + 0,634; 0,958; 1,282 > x_1 + 0,132 x_2 .
\]

The resulting model (4) can be used to predict the amount of revenue. For this, the predicted values of
the factors should be substituted into the resulting equation (4) and the left, average, and right values of the effective index should be calculated from the predicted values of the factors. Table 4 shows the results of the evaluation function of the effective indicator (revenue), obtained on the basis of fuzzy regression (2), as well as the predicted values for the control sample (given in the last three lines).

| h=0          | Revenue, thousand rubles | Lower bound of range | Middle of the fuzzy range | Upper bound of range | Width of the fuzzy range |
|--------------|--------------------------|-----------------------|---------------------------|----------------------|--------------------------|
| Data for forecasting | fact                     | 71                    | 53,2                      | 70,3                 | 87,3                     | 17                       |
|              | 46,4                     | 46,4                  | 58,9                      | 71,4                 | 12,5                     |
|              | 42,5                     | 35,1                  | 46,8                      | 58,4                 | 11,6                     |
|              | 47,6                     | 45,5                  | 58,3                      | 71,2                 | 12,8                     |
|              | 35,4                     | 30,6                  | 40,3                      | 50                   | 9,7                      |
|              | 42,5                     | 38,7                  | 52,4                      | 66                   | 13,7                     |
|              | 36,9                     | 27,9                  | 36,3                      | 44,7                 | 8,4                      |
|              | 34,1                     | 31,9                  | 41,5                      | 51,1                 | 9,6                      |
|              | 21,3                     | 17,1                  | 20,7                      | 24,3                 | 3,6                      |
|              | 23,7                     | 22                    | 27,1                      | 32,3                 | 5,1                      |
|              | 54,8                     | 33,6                  | 44,2                      | 54,8                 | 10,6                     |
|              | 23,3                     | 20,8                  | 25,8                      | 30,8                 | 5                        |
|              | 25,6                     | 25,6                  | 33,2                      | 40,8                 | 7,6                      |
|              | 37,8                     | 37,8                  | 50                        | 62,2                 | 12,2                     |
|              | 50,7                     | 40,2                  | 54,3                      | 68,5                 | 14,1                     |
| Test        | 35,9                     | 27,4                  | 34,8                      | 42,3                 |                          |
|              | 47,7                     | 40                    | 52,6                      | 65,2                 |                          |
|              | 31,3                     | 29,5                  | 39,6                      | 49,8                 |                          |

The data in the table shows that the entire initial and predicted values y fell into the range of their values $\bar{y}$.

6. Conclusions
Modeling methods based on the theory of fuzzy sets provide wider opportunities for analyzing economic processes. Fuzzy regression allows getting a range of possible values, while providing coverage of the source data. The presence of interval variation of the resultant characteristic allows taking into account the uncertainty of the initial information and can replace the confidence interval of the average line of the classical regression. Fuzzy regression analysis can be used when setting the raw data in the form of clear, interval or fuzzy numbers.

In the above model, the membership function of fuzzy parameters of the estimated variable had a symmetrical triangular shape for the convenience and simplicity of calculating, but different membership function can be used for this.

References
[1] Zak YU.A. Prinyatiye resheniy v usloviyakh nechetkikh i razmytykh dannykh. Izd.: Librokom. 2018. 352 s.
[2] Ptuskin A.S. Nechotkiye modeli i metody v menedzhmente. – M.: Izd. MGTU im. Bauman.
[3] Parfenova V.Ye. Netchetkiye modeli prinyatiya optimalk'nykh resheniy v upravlenii agrarnym proizvodstvom.//Innovatsii. №10. – 2018. S.88 – 92.

[4] Lin J.-G., Zhuang Q.-Y., Huang C. Fuzzy Statistical Analysis of Multiple Re-gression with Crisp and Fuzzy Covariates and Applications in Analyzing Economic Data of China // Computational Economics. 2012. 39. P. 29–49.

[5] Yun-Hsi O. Chang, Bilal M. Ayyub Fuzzy regression methods-a comparative assessment.//Fuzzy Sets and Systems, Vol. 119 (2), 2001. p. 187–203.

[6] Sakawa M. Multiobjective fuzzy linear regression analysis for fuzzy input–output data / M. Sakawa, H. Yano // Fuzzy Sets and Systems. 1992. V. 47. P. 173-181.

[7] Seraya, O. V. Linear regression analysis of a small sample of fuzzy input data [Text] / O. V. Seraya, D. A. Demin // Journal of Automation and Information Sciences. – V.. 44, Issue 7. – P. 34–48.

[8] Tanaka H., Watada J. Possibilistic linear systems and their application to the linear regression model //Fuzzy sets and systems. 1988. T. 27. № 3. P. 275–289.

[9] Zade L.A. Ponyatiye lingvisticheskoy peremennoy i yego primeneniye k prinya-tiyu priblizhennykh resheniy. – M.: Mir. - 1976. 168 s.

[10] Korshunov, G. I., Nazarevich, S. A., & Smirnov, V. A. (2018). Fuzzy classification of technical condition at life cycle stages of responsible appointment systems. Paper presented at the CEUR Workshop Proceedings, 2258 427-437

[11] Dunaieva, I., Mirschel, W., Popovych, V., Pashtetsy, V., Golovastova, E., Vecherkov, V., Melnichuk, A., Terleev, V., Nikonorov, A., Ginevsky, R., Lazarev, V. Topaj, A. (2019). GIS services for agriculture monitoring and forecasting: Development concept. Advances in Intelligent Systems and Computing, Volume 983, 2019, Pages 236-246 doi:10.1007/978-3-030-19868-8_24

[12] Babkin, A. V., Karlina, E. P., & Epifanova, N. S. (2016). Neural networks as a tool of forecasting of socioeconomic systems strategic development. Paper presented at the Proceedings of the 28th International Business Information Management Association Conference - Vision 2020: Innovation Management, Development Sustainability, and Competitive Economic Growth, 11-17.

[13] Solopchenko, G. N. (2016). Fields of effective application of statistical methods of processing the results of repeated measurements. Measurement Techniques, 59(5), 476-484. doi:10.1007/s11018-016-0993-2

[14] Volkov, M., Kibkalo, A., Vodolagina, A., & Murgul, V. (2016). Existing models residual life assessment of structures and their comparative analysis. Paper presented at the Procedia Engineering, 165 1801-1805. doi:10.1016/j.proeng.2016.11.925

[15] Egorova, D., Ivanov, K., Lebedev, D., & Nikulina, S. (2018). Recommendations for improving the accuracy of virtual testing. Journal of Advanced Research in Dynamical and Control Systems, 10(12)