Supplementary Notebook (RTEP - Brazilian academic journal, ISSN 2316-1493)

MODELING GROSS REGIONAL PRODUCT BASED ON CRISP AND FUZZY REGRESSIONS

Ilyas Idrisovich Ismagilov [1],
Ghena Alsaied [2]

1Doctor of Technical Sciences, Professor of the Department of Economic Theory and Econometrics, Institute of Management, Economics and Finance, Kazan (Volga Region) Federal University; e-mail: iiismag@mail.ru.
[2]Ghena Alsaied, Post-graduate student of the Institute of Management, Economics and Finance, Kazan (Volga Region) Federal University; e-mail: Ghenaalsaied@gmail.com.

Abstract: Regression modeling is a recognized tool for the analysis and forecasting of socio-economic processes. In this case, a linear regression apparatus is often used, which shows a high efficiency in solving many applied problems. In some situations the use of classical linear regression is not sufficiently substantiated. These situations are quite typical and characteristic in constructing models of socio-economic processes, especially in the case of short samples. Fuzzy regression analysis is one of the most promising areas of scientific research in this field. The fuzzy regression model gives fuzzy functional relationships between dependent and independent variables, while they can be either crisp or fuzzy. The article presents the results of a study on modeling the gross regional product as an indicator of socio-economic development of the Republic of Tatarstan by using the methods of linear crisp and fuzzy regressions. Linear GRP models constructed of four different factors (volume of shipped products, agricultural products, investment in fixed capital, and volume of performed work by type of activity "construction"). Models were built by using a sample of statistical data for 1999-2018. The forecast properties of the crisp and fuzzy regression models are compared with dividing the source sample into the training and test subsamples. The results obtained show the promise of using fuzzy regression to forecast regional economic growth.

Keywords: Modeling, Forecasting, Gross Regional Product, Regression Model, Linear Regression, Fuzzy Linear Regression.

INTRODUCTION

The current stage of development of national economies is characterized by the intensification of globalization and digitalization processes with increasing their competitiveness. Achieving this goal is inextricably linked with increasing the efficiency of the country's regional economies. It is often assessed by the state of a number of macroeconomic indicators of the region, primarily in the level of gross regional product (GRP) (4). This indicator is a general indicator of regional economic growth. The choice of factors for evaluating and constructing forecast models of GRP leads to the activation of regional and federal policies and a decrease in the level of interregional disproportions. As the most widely used statistical method, regression analysis (RA) plays an increasingly important role in creating models of economic growth (8; 7; 1). Application of classical linear regression in some situations of modeling socio-economic processes is not sufficiently substantiated. For example, the estimation of linear regression by the least squares method (OLS) is correct only if its premises (Gauss-Markov conditions) are met. To construct a multifactorial regression model, we need often to have a large amount of sample data. Moreover, the data must be specified in a deterministic form (3).

Fuzzy regression modeling expands the capabilities of the classical and removes a number of its limitations. It was proposed as an alternative to the classical probabilistic approach and is based on the use of fuzzy linear regressions (FLR). The use of FLR turned out to be justified in those situations where there is no reason to assume that errors in specific observations which obey probabilistic regularities. Its applicability to both micro and macroeconomic analysis has been shown (11; 5). The fuzzy regression model gives fuzzy functional relationships between dependent and independent variables, where the input data can be crisp or fuzzy. Thus, the problem statements of fuzzy regression analysis can vary depending on the fuzziness of the source data and/or model parameters. In the general case, the methods of fuzzy regression analysis can be divided into two groups: first is based on the least-squares method and its generalizations; second is based on linear programming. Tanaka created a model of fuzzy linear regression with a parameter that was a fuzzy number, by using the linear programming method to evaluate parameters with the criterion of the minimum fuzziness index (11). Three formulations of fuzzy linear regression analysis are considered in (6). This method has been applied to ordinary data, obtaining fuzzy parameters.

Fuzzy linear regression can be considered as an alternative to traditional econometric approaches for real estate valuation and both of the crisp and fuzzy methods lead to a certain extent consistent results, but at the same time, the use of fuzzy linear regression should be accompanied by an economic analysis of the results obtained, without which the model may lose its adequacy (13). A description of the fuzzy regression method by using triangular fuzzy numbers (fuzzy corridor) is presented in (2). In (2; 15; 16), small amounts of initial data were used and the obtained results say that the classical crisp and fuzzy linear regressions lead to similar results. In (10), the method of constructing a fuzzy linear regression led to the fact that the results of the regression significantly depended on the data scale. In (16), various approaches for the construction of models of fuzzy linear regression of socio-economic processes and their critical analysis are described.

In the present work, traditional multiple linear regression (MLR) is estimated by the OLS, and FLR is applied by crisp sample data, both are used for the analysis and modeling of GRP. When constructing a FLR, the coefficients for independent variables were estimated in the form of triangular fuzzy numbers. As a result, the predicted values of the dependent variable (GRP) are also presented in the form of fuzzy numbers. A comparison of two types of linear regressions by predictive properties is made.

METHODOLOGY

The following models are used for GRP modeling: MLR, is estimated by OLS (implementation in the environment econometric package Gretl); FLR which using triangular fuzzy numbers, is estimated by the fuzzy corridor method (implementation in the environment table processor MS Excel). In the general case, the theoretical equation of LMR has the form:

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \varepsilon \qquad (1)$$

According to statistical data, the estimated selective model has the form:

$$\hat{Y} = A_0 + A_1 x_1 + \cdots + A_p x_p \qquad (2)$$

Note that, in the case of FLR, the selective model parameters A_0 , $A_1, \ldots, A_p$ are fuzzy symmetric confidence triples of numbers describe the corresponding fuzzy triangular numbers. Each fuzzy coefficient will have the form:

$$A_i = (a_i - b_i, a_i, a_i + b_i),$$

where a_i – is the most probable coefficient value, and the value b_i – describes the width of the coefficient blur. The membership function of a fuzzy triangular number is shown in Fig.1.

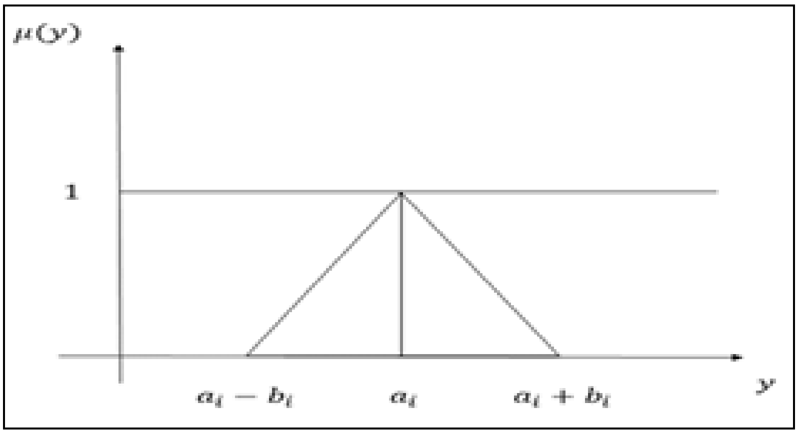

Figure 1. Membership function of fuzzy number.A_i.

Note that, in the setting of constructing a FLR under consideration, the predicted values of Y will be described by triangular symmetric fuzzy numbers, inside which the real values of the dependent variable should be located. A formal description of FLR can be presented as follows (Biryukov, 2010):

$$\sum_{j=1}^{n}\sum_{i=0}^{p}((a_i - b_i)x_{ij}) + (a_0 - b_0) \le y_j, \qquad j = \overline{1,n}$$

(3)

$$\sum_{j=1}^{n}\sum_{i=0}^{p}((a_i + b_i)x_{ij}) + (a_0 + b_0) \ge y_j, \qquad j = \overline{1,n};$$

$$b_i \ge 0 \;, i = \overline{0,p}.$$

Conditions (3) are conditions for including the values of the dependent variable Y in the range of possible values of F. The statement of the problem of FLR is to find such values of the parameters of the fuzzy coefficients of the model a_i and b_i, $i = \overline{0,p}$, minimizing the width of the fuzzy corridor, covering the actual values of the dependent variable:

$$F = \sum_{j=1}^{n}[\sum_{i=0}^{p}((a_i + b_i)x_{ij}) + (a_0 + b_0) - \qquad \sum_{i=0}^{p}((a_i + b_i)x_{ij})] \longrightarrow \min$$

Moreover, the found FLR equation includes three components that describe the fuzzy corridor: function Y1 (lower bound); function Y3 (upper bound); function Y2 (middle of the corridor). Note that between the functions Y1 and Y3 all actual values of the dependent variable will be located. In order the tasks to be solved, samples of indicators of the socio-economic processes of the Republic of Tatarstan from 1999 to 2018 were used (data is taken from the website of the Federal State Statistics Service for the Republic of Tatarstan http://tatstat.gks.ru). In this case, the dependent variable represents the gross regional product, mln.rub. (Y). The following indicators are selected as independent variables: volume of shipped products, mln.rub. (X1); agricultural products, mln.rub. (X2); investment in fixed capital, mln.rub. (X3); volume of performed work by type of activity "construction", mln.rub. (X4). The selected sectors of the economy of the Republic of Tatarstan are among the main ones by contribution to GRP. Regression models were built using variables at current prices.

RESULTS AND DISCUSSION

As a result of the research, a number of regression models were constructed and analyzed. Model (4) is constructed on a full sample of statistical data for the period 1999-2018. By applying OLS to the initial data in the environment of the Gretl package, a model of the form is obtained:

$$\hat{Y} = 44172{,}8 + 0{,}643X_1 + 0{,}172X_2 + 0{,}572X_3 + 0{,}596X_4 \quad (4)$$

The model is characterized by the following confidence indicators: coefficient of multiple determination R^2=0,9995, mean absolute percentage error MAPE=2,5%, the model is statistically significant according to the Fisher F-test at a significance level of α = 0.01, however, only the free term and the regression coefficient for X_1 are significant for α=0.01, the other regression coefficients are insignificant even at α=0.1. This situation is a characteristic of multicollinearity of factors (4). We will improve the model through eliminating non-essential variables by using the joint F-test. The improved model has the following format:

$$\hat{Y} = 47320{,}8 + 0{,}638X_1 + 0{,}971X_3 \qquad (5)$$

The adequacy and significance of this model are confirmed by the following: R^2 =0.999, MAPE=2,65%, regression coefficients and the model as a whole are significant at the level of α = 0.01. The analysis of the residuals of the model showed the following: all built-in Gretl tests for heteroskedasticity and autocorrelation showed their absence at α = 0.01. The normality of the distribution law is also confirmed at this level α. It should be noted here that the checks for the normality of the residues are correct only in the case of large samples. The results of the analysis allow us to conclude that the model has high quality indicators and is suitable for solving applied tasks. We will construct a fuzzy linear regression using triangular fuzzy numbers on the full set of initial factors. In the used method of constructing a FLR, instead of minimizing the quadratic functional of the OLS, the problem of minimizing the width of the fuzzy interval of forecast Y values is solved. In this regard, this method does not require the fulfilment of the premises of the OLS (7). Estimated FLR has the following form:

$$\hat{Y} = (32294{,}78; 46887{,}06; 61479{,}36) + 0{,}617X_1 + (-0{,}055; 0; 0{,}055)X_2 + 0{,}739X_3 + 0{,}551X_4 \qquad (6)$$

For a comparative evaluation of the quality of regression models, we present a number of their main indicators in table 1. Indicators for FLR are calculated by using the defuzzified values of fuzzy predicted values of the dependent variable through using the center of gravity method.

Table 1. Quality Indicators of models (5), (6).

Indicators	Crisp regressions	Fuzzy regressions
Coefficient of Multiple Determination R^2	0,999	0,999
Root Mean Square Error (RMSE)	16135	16364
Mean Absolute Percentage Error (MAPE)	2,648	2,583

Note that, the value of R^2 in crisp and fuzzy regression is close to one, and then both models have high explanatory properties. The RMSE value of both models is almost the same, the MAPE of crisp regression is higher. This indicates to higher predictive properties of fuzzy regression. We study the effect of sample dimension on the quality of models. To do this, we will conduct a modeling on the shortened sample from 2010 to 2018. This choice was made by considering the structural shifts in the Russian economy was caused by the crisis phenomena of 1998 and 2008. In the Gretl environment, the following model for shortened sampling is built:

$$\hat{Y} = 34346{,}9 + 0{,}643X_1 - 0{,}291X_2 + 0{,}925X_3 + 0{,}282X_4 \qquad (7)$$

The model adequately describes the initial data, since for it R^2 = 0.998. The model is statistically significant as a whole according to the F-test at the significant level α =

0.01, however, only the regression coefficient at X_1 is significant at α = 0.01, the remaining regression coefficients are insignificant even at α = 0.1. The obvious situation of multicollinearity factors. Due to the exclusion of non-essential variables, by using the joint F-test, the improved model has the following form:

$$\hat{Y} = 50805{,}2 + 0{,}639X_1 + 0{,}960X_3 \qquad (8)$$

The model is described as follows: R^2 = 0.998, both regression coefficients and the model as a whole are significant at α = 0.01. According to models 5 and 8, it should be concluded that in the economy of Tatarstan 1999-2018. The volume of shipped products and, investment in fixed capital positively (but not very significantly) affect the increase in GRP. On a shortened sample, Fuzzy linear regression using triangular fuzzy numbers has the form:

$$\hat{Y} = 30726{,}08 + 0{,}617X_1 + (0{,}091;0{,}215;0{,}339)X_2 + 0{,}628X_3 + 0{,}663X_4. \quad (9)$$

The high accuracy of the obtained model was confirmed, since the real GRP values fell into the corridor, which has formed, by fuzzy regression (Fig.2).

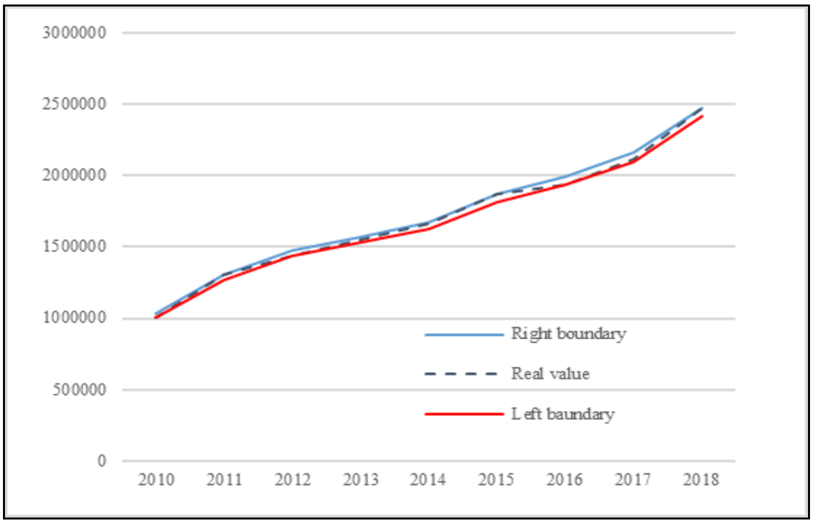

Figure 2. GRP and fuzzy corridor for the sample 2010-2018

For a comparative assessment of the quality of regression models, we present a number of their main indicators in table 2.

Table 2. Quality Indicators of models (8), (9).

Indicators	Crisp regressions	Fuzzy regressions
Coefficient of Multiple Determination R^2	0,997	0,997
Root Mean Square Error (RMSE)	18890	19325
Mean Absolute Percentage Error (MAPE)	1,01	1,02

Note that in case of a shortened sample from 2010 to 2018: the R^2 value in crisp and fuzzy regression is close to unity, and then both models have high explanatory properties. The RMSE value of the fuzzy regression is higher. The MAPE of both models are almost identical. Therefore, both models (crisp and fuzzy) can be used for analyzing and predicting GRP. To compare models based on forecast properties, we will divide the available sample into two parts: a training and a test sub-sample. The training sub-sample includes the main part of observations by using data from 1999 to 2017. For the test sub-sample, select data for 2018. In the Gretl environment, the built adequate crisp model for 1999-2017 has the following form:

$$\hat{Y} = 51572{,}06 + 0{,}577X_1 + 1{,}160X_3 \qquad (10)$$

Fuzzy model for 1999-2017. has the form:

$$\hat{Y} = (33717{,}74; 49726{,}66; 65735{,}57) + 0{,}565X_1 + (-0{,}036; 0; 0{,}036)X_2 + 0{,}914X_3 + 0{,}513X_4 \qquad (11)$$

The forecast properties of crisp and fuzzy regression models were compared in Table 3.

Table 3. Forecasts and their relative errors.

Model	Types of forecasts		Relative Error
	Point forecast	Interval forecast	
Crisp	2408512,1	[2317285,1; 2499739,1] at the level α=0.05	2,5%
Fuzzy	2398305	[2374326; 2422283]	2,9%

The forecasts of models (10) and (11) are quite close to the actual values. Note that, the accuracy of the forecast according to fuzzy regression model gives a narrower interval forecast of GRP for 2018. Table 4 presents the forecast data of the GRP of the Republic of Tatarstan for 2019-2021 using models (5) and (6).

Table 4. Forecast data of the GRP of the Republic of Tatarstan.

Model	Year	$\hat{Y}$	Y	Relative Error	Interval forecast
Crisp	2019	2557489	2552548,2	0,19%	[2505341,9; 2609636,9]
	2020	2706096,5	2702310	0,13%	[2652028,6; 2760164,3]
	2021	2856090,6	2865842,8	0,34%	[2800556,8; 2911624,4]
Fuzzy	2019	2544674,1	2552548,2	0,31%	[2517541,1; 2571807,2]
	2020	2692450,5	2702310	0,36%	[2664670,3; 2720230,8]
	2021	2841637	2865842,8	0,85%	[2813149,2; 2870124,7]

For the executed GRP forecasts, the relative error is less than 1%. The forecast values (Y) previously announced by the Ministry of economy of the Republic of Tatarstan are included in all the interval forecasts we received. Thus, the constructed models (5) and (6) have proved their high predictive ability. Interval forecasts are narrow in the case of fuzzy regression, so they are more preferable, especially they give the uncertainty regarding the distributions of random errors of linear clear regression.

SUMMARY

The model of crisp and fuzzy linear regression gives a better accuracy of predicting GRP in case of a shortened sample than in case of a full sample, and well explained the behavior of GRP in both cases (full and shortened). The accuracy of forecasts for 2018 is not very satisfactory by using actual data, the reason for this is that the volume of GRP of the Republic of Tatarstan at the end of 2018 amounted to 2469217.4. million rubles, (in 2017 - 2114176.1 million rubles), and in 2018 reduced investment in fixed capital in Tatarstan, and this differs from previous years (https://prav.tatarstan.ru). Regression models, were based on the theory of fuzzy sets, have significant advantages over traditional models because they allow predicting the result within a given corridor. Therefore, the results of fuzzy modeling of GRP can be considered satisfactory.

CONCLUSION

The paper considers an approach to modeling and forecasting GRP by using linear models of fuzzy and crisp regression. As a result of the analysis of the regional economy of the Republic of Tatarstan for the period 1999-2018. Adequate crisp and fuzzy models of GRP are obtained. An analysis of the qualitative indicators of crisp and fuzzy regression shows that both models have similar results in modeling GRP. However, at the prediction by using the fuzzy regression, we get results in narrower interval forecasts. Thus, to construct models of GRP which depend on a number of production factors, it is advisable to use fuzzy regression analysis. This is because the fuzzy regression model, unlike the standard regression, does not require the assumptions of the estimation method to be met. This fact allows us to build models from short samples with more reliable interval forecasts. In this case, the obtained point forecasts differ slightly. All of the above, according to the authors, indicates the possibility and effectiveness of using fuzzy regression analysis for modeling and forecasting regional economic growth.

ACKNOWLEDGEMENTS

The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

REFERENCES

1. Alsaied Ghena, (2019). Modeling the gross regional product based on crisp and fuzzy regressions. *Economy in a Changing World, Kazan Federal University, Russia,* 235-238.

2. Biryukov, A. N, (2010). Multivariate fuzzy regression predictive model for an applied problem. *Management of economic systems: electronic scientific journal, 2*(22), , http://uecs.ru/uecs-22-222010/item/156-2011-03-/(27-04-2020).

3. Draper, N., & Smith, G. (1973). *Applied regression analysis*. M.: Statistics.

4. Gorshenina, E.V, (2009). *Regional economic researches: theory and practice*, Tver, Tver State University Publ.

5. Heshmaty, B., & Kandel, A. (1985). Fuzzy linear regression and its applications to forecasting in uncertain environment. *Fuzzy sets and systems*, *15*(2), 159-191.

6. Ismagilov, I.I, & Alsaied Ghena, (2019). Fuzzy regression modeling in economic research. *International Symposium on Management, Economics and Finance (ISMEF), Kazan Federal University, Russia, 1*, 10-13.

7. Miroliubova, T.V., & Voronchikhina, E.N. (2015). Determination of key focus areas of regional economic policy on the basis of regional economic growth econometric modeling and forecasting. *Vestnik Permskogo universiteta. Seria Ekonomika = Perm University Herald. Economy, 4*(27), 80–91.

8. Serga, L.K., Simonova, E.Y., & Zaykov, K.A, (2016). Methodological approach to modeling of economic growth of innovative economy of the region. *Vestnik NSUEM, 4,* 301-314.

9. Summary report of the Republic of Tatarstan on the results of monitoring the effectiveness of local authorities in urban districts and municipal regions of the Republic of Tatarstan according the results of 2018, https://prav.tatarstan.ru/rus/file/pub/pub_2046726.pdf.

10. Tanaka, H., & Ishibuchi, H, (1991). Identification of possibilistic linear systems by quadratic membership functions of fuzzy parameters. *Fuzzy sets and Systems, 41*(2), 145–160.

11. Tanaka, H., Hayashi, I., & Watada, J. (1989). Possibilistic linear regression analysis for fuzzy data. *European Journal of Operational Research*, *40*(3), 389-396.

12. Tanaka, H., Uejima, S., & Asai, K., (1982). Linear regression analysis with fuzzy model. IEEE Trans. *Systems Man Cybern, 12*(6), 903–907.

13. Vishniakova, E.V., Ivanova, E.V., Kamalov, S.M., Kolodyazhnaya, U.A., Khamidullina, L.F. (2015). Fuzzy linear regression in assessment. *Scientific notes of young scientists, 5,* 14-29.

14. Volkova, E.S., & Gisin, V.B. (2015). Fuzzy linear regression in real estate valuation. *The Appraisal Issues, 1*(79), 26–33.

15. Volkova, E.S., & Gisin, V.B. (2015). The use of fuzzy linear regression in the model of technological knowledge growth. *Finance: Theory and Practice, 5,* 97-104.

16. Volkova, E.S., & Gisin, V.B. (2018). Fuzzy linear regression in models of socio economical systems. *Economics and Management: Problems, Solutions, 5,* 30-33.