Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 214 (2022) 391–396

Procedia Computer Science

www.elsevier.com/locate/procedia

9th International Conference on Information Technology and Quantitative Management

Quadratic least square regression in fuzzy environment

Bogdana Stanojević[a,*], Milan Stanojević[b]

[a]*Mathematical Institute of the Serbian Academy of Sciences and Arts, Kneza Mihaila 36, 11000 Belgrade, Serbia*
[b]*University of Belgrade, Faculty of Organizational Sciences, Jove Ilića 154, 11000 Belgrade, Serbia*

Abstract

The role of the regression analysis is crucial in many disciplines. Addressing the fuzzy quadratic least square regression for observed data modeled by fuzzy numbers, we aim to emphasize how a methodology that does not fully comply to the extension principle may fail to predict fuzzy valued numbers. We also propose a solution approach that functions in full accordance to the extension principle, thus overcoming the shortcomings arisen from the practice of splitting the optimization of a fuzzy number in independent optimizations of its components.

Keywords: Fuzzy regression; Fuzzy numbers; Extension principle

1. Introduction

The regression analysis has a crucial role in many research fields (see for instance [1], [9] and [10]), and is widely applied in various areas. According to He et al. [6], *the fuzzy regression analysis is a powerful method for forecasting the fuzzy outputs of an uncertain system*.

Zeng et al. [16] introduced a new distance between triangular fuzzy numbers, and proposed a fuzzy regression model based on the least absolute deviation method and the new formulated distance. They used linear programming to develop the algorithm for deriving the coefficients of their fuzzy linear regression model. We will use one of their examples to illustrate our theoretical findings.

Recently, Kashani et al. [7] proposed a penalized estimation method to derive the coefficients of a linear regression model in order to analyze the dependence of a LR fuzzy response (output) variable on a set of crisp explanatory (input) variables. They carried out a simulation study under three scenarios of multi-co-linear and sparse data in order to emphasize the performances of their novel model.

* Corresponding author. Tel.: +381-11-2630170 ; fax: +381-11-2186105 .
E-mail address: bgdnpop@mi.sanu.ac.rs

10.1016/j.procs.2022.11.190

Chen and Nien [2] proposed a new operator, namely the fuzzy product core, and used it to provide a new formulation to the fuzzy regression model. They increased the model's performance by reducing the unnecessary embedded information.

Chukhrova and Johannssen [3] provided a critical discussion of the most relevant methods and approaches on the topic of fuzzy regression analysis.

He et al. [6] used a random weight network in their fuzzy nonlinear regression analysis where both the inputs and outputs were triangular fuzzy numbers. Stanojevic et al. [11] reviewed the main ideas in dealing with mathematical programming in fuzzy environment, aiming to rehabilitate the position of the extension principle within the fuzzy optimization stage.

Kupka [8] and Diniz et al. [4] introduced the main theoretical foundations related to the optimization of fuzzy-valued functions using Zadeh's extension principle.

Ghanbari et al. [5] reviewed the most relevant methods and solution approaches to fuzzy linear programming problems. Majority of the approaches were based on ranking functions and α-cuts; and used duality results or penalty functions to solve the crisp formulations of the fuzzy problems. In this way such methodologies unfortunately deviate from the extension principle and the derived results loose their relevance.

2. Preamble

Generally, a regression model describes a relationship between a scalar response, and one or more explanatory variables. It uses a predictor function whose unknown model coefficients are estimated from the observed data. The regression model aims to quantify the correlation between observed data, or to make a prediction of the response in the case of new values of the explanatory variables.

Many variants of regression models are widely studied, and their differences are mainly due to various norms they are based on: the least square regression uses the l_2 norm, RIDGE regression uses a penalized l_2 norm, the LASSO regression uses a penalized l_1 norm.

Given n observed input k-tuples $(x_{i1}, x_{i2}, \ldots, x_{ik})_{i=1,\ldots,n}$ and n output scalars $(y_i)_{i=1,\ldots,n}$ one searches for a set of coefficients A of the predictor function $f_A(x_1, x_2, \ldots, x_k)$ able to give a good approximation $f_A(x_{i1}, x_{i2}, \ldots, x_{ik})$ to y_i, for all $i = 1, \ldots, n$. The parametric expression of the quadratic predictor function in its matrix form is

$$f_{Q,P,p_0}(X) = \frac{1}{2}X^T QX + P^T X + p_0, \tag{1}$$

where Q is an $n \times n$ matrix with real components, P is an n-column vector with real components and p_0 is a real scalar.

The quadratic predictor function in the particular case of scalar input observations (i.e. $k = 1$) is

$$f_A(x) = a_0 + a_1 x + a_2 x^2. \tag{2}$$

The method we propose in the next section addresses this particular case but in fuzzy environment. The parameters a_0, a_1, a_2 of the least square regression model (2) are computed using the following formula.

$$\begin{pmatrix} a_2 \\ a_1 \\ a_0 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n x_i^4 & \sum_{i=1}^n x_i^3 & \sum_{i=1}^n x_i^2 \\ \sum_{i=1}^n x_i^3 & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i & n \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^n x_i^2 y_i \\ \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n y_i \end{pmatrix} \tag{3}$$

Zadeh [14] defined the fuzzy sets and proposed the extension principle aiming to provide a powerful tool for modeling the uncertainty. Later on, in [15], Zadeh introduced the linguistic variables emphasizing their application to approximate reasoning. Based on the fuzzy set theory, the fuzzy linear regression model was proposed by Tanaka et al. [13] and [12].

Formally, by analogy, the fuzzy quadratic regression model aims to provide fuzzy values to the parameters $\widetilde{a}_0, \widetilde{a}_1, \widetilde{a}_2$ of the predictor function

$$f_{\widetilde{A}}(\widetilde{x}) = \widetilde{a}_0 + \widetilde{a}_1\widetilde{x} + \widetilde{a}_2\widetilde{x}^2. \tag{4}$$

Many solution approaches derive the fuzzy parameters $\widetilde{a}_0, \widetilde{a}_1, \widetilde{a}_2$ using simple algebraic formula and the fuzzy arithmetics based on the extension principle. However, the optimization behind the algebraic formula in crisp environment cannot be analogically extended to fuzzy environment. The approach we introduce in the next section incorporates the extension principle in the optimization stage, thus deriving more realistic results. The numerical results reported in Section 4 shows that our approach is able to overcome the shortcomings of some approaches from the literature that do not fully comply to the extension principle.

3. Our approach

The main idea of our approach is in simulating the effect of applying the quadratic fuzzy function (4) with the help of the crisp functions (2) defined for each set of crisp inputs that belong to their corresponding α-cut intervals.

The models proposed in this section concern the crisp inputs x_i, and the trapezoidal fuzzy outputs

$$\widetilde{y}_i = \left(y_i^1, y_i^2, y_i^3, y_i^4\right), i = 1, \ldots, n, \tag{5}$$

under the assumption that the output $\widetilde{y}_i$ is always the same for the same input x_i, $i = 1, \ldots, n$. Let $G = \left(g_{ij}\right)_{i=\overline{1,3}}^{j=\overline{1,3}}$ be the inverted matrix which appears in (3).

To derive the left end-point of the α-cut interval of the predicted fuzzy value at the given point u, Model (6)

$$\begin{aligned}
&\min\ a_0 + a_1 u + a_2 u^2 \\
&\text{s.t.} \\
&\quad a_0 = \sum_{i=1}^{n} \left(g_{11}x_i^2 + g_{12}x_i + g_{13}\right) v_i, \\
&\quad a_1 = \sum_{i=1}^{n} \left(g_{21}x_i^2 + g_{22}x_i + g_{23}\right) v_i, \\
&\quad a_2 = \sum_{i=1}^{n} \left(g_{31}x_i^2 + g_{32}x_i + g_{33}\right) v_i, \\
&\quad a_j \text{ free}, \qquad j = 0, 1, 2 \\
&\quad v_i \in \left[\alpha\left(y_i^2 - y_i^1\right) + y_i^1, \alpha\left(y_i^3 - y_i^4\right) + y_i^4\right], \ i = 1, \ldots, n
\end{aligned} \tag{6}$$

has to be used. Parameters x_i and $\widetilde{y}_i$, $i = 1, \ldots, n$ are the observed inputs and outputs respectively; parameter u is the given scalar representing the crisp input whose prediction $a_0 + a_1u + a_2u^2$ is optimized; while the model's variables are a_0, a_1, a_2, and v_i, $i = 1, \ldots, n$. Similarly, to derive the right end-point of the same α-cut interval one has to maximize the same objective function over the same feasible set as in Model (6).

Using Model (6) and its corresponding maximization model we predict fuzzy values that are in strict accordance to the extension principle. To illustrate our methodology we recall an example from Zeng et al. [16] and used it in two ways: (i) as it is given in the literature, using the fuzzy quadratic regression; and (ii) adapted to the purpose of showing how a methodology that do not fully comply to the extension principle may fail to predict fuzzy valued numbers.

Zeng et al.'s approach [16] applied separately the least absolute linear regression to the left, center and right components respectively of the fuzzy outputs. In this way they obtained very narrow predicted fuzzy values that is highly desirable in fuzzy environment. However, for certain combinations of observed fuzzy outputs such class of methodologies may derive right components of predicted fuzzy values that are smaller than the left components, hence making the approach unable to predict anything.

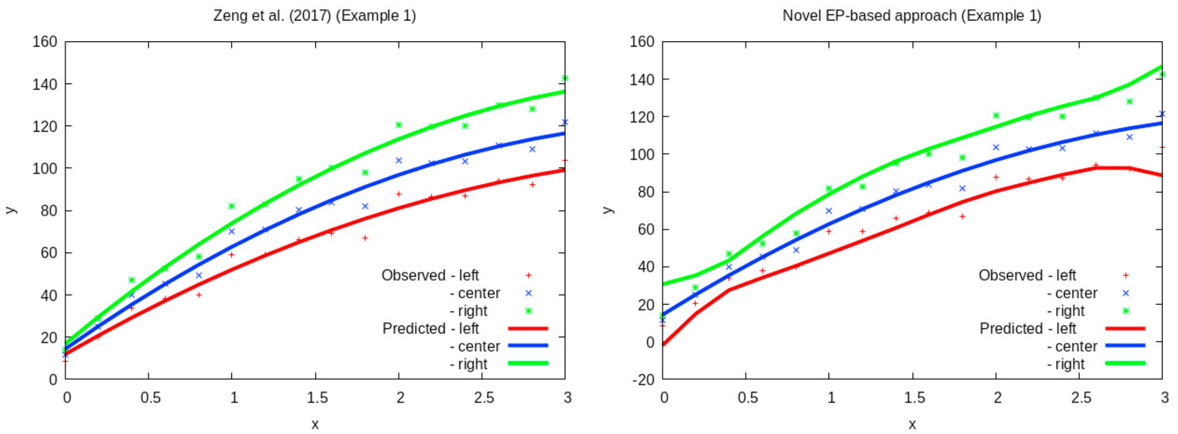

Fig. 1. The numerical results obtained for Scenario I. Left: using Zeng et al.'s approach adapted to fuzzy least square quadratic regression. Right: using our approach that is in full accordance to the extension principle

Table 1. Input and output data grouped in Scenarios I and II for numerical illustrations

i	Inputs x_i	Scenario I Outputs $\widetilde{y}_i$	Estimations $\widetilde{y}_i^*$	Scenario II Outputs $\widetilde{y}_i$	Estimations $\widetilde{y}_i^*$
1	0	(8.5, 11.5, 14)	(−1.91, 14.44, 30.67)	(10.5, 11.5, 12)	(−0.72, 14.44, 28.26)
2	0.2	(20.3, 24.8, 28.8)	(14.99, 25.26, 35.48)	(23.3, 24.8, 25.8)	(16.60, 25.26, 31.67)
3	0.4	(34, 40, 47)	(27.63, 35.51, 43.41)	(39, 40, 42)	(29.01, 35.51, 38.99)
4	0.6	(38.2, 45.2, 52.2)	(34.22, 45.19, 56.27)	(38.2, 45.2, 46.2)	(33.91, 45.19, 52.81)
5	0.8	(40.1, 49.1, 58.1)	(40.50, 54.29, 68.29)	(40.1, 49.1, 51.1)	(38.69, 54.29, 65.75)
6	1	(59, 70, 82)	(47.21, 62.82, 78.77)	(59, 70, 72)	(44.10, 62.82, 77.05)
7	1.2	(58.9, 70.9, 82.9)	(53.91, 70.78, 88.12)	(58.9, 70.9, 73.9)	(49.78, 70.78, 87.08)
8	1.4	(66.1, 80.1, 95.1)	(60.74, 78.16, 96.21)	(60.1, 80.1, 95.1)	(55.86, 78.16, 95.70)
9	1.6	(69, 84, 100)	(67.83, 84.97, 102.92)	(55, 84, 110)	(62.43, 84.97, 102.82)
10	1.8	(67, 82, 98)	(74.58, 91.21, 108.82)	(60, 82, 108)	(68.99, 91.21, 108.96)
11	2	(87.7, 103.7, 120.7)	(80.29, 96.88, 114.64)	(67.7, 103.7, 125.7)	(74.93, 96.88, 114.70)
12	2.2	(86.6, 102.6, 119.6)	(84.81, 101.97, 120.52)	(86.6, 102.6, 119.6)	(80.16, 101.97, 120.16)
13	2.4	(87.1, 103.1, 120.1)	(89.06, 106.48, 125.54)	(87.1, 103.1, 120.1)	(85.34, 106.48, 124.64)
14	2.6	(94, 111, 130)	(92.69, 110.43, 130.04)	(94, 111, 135)	(90.29, 110.43, 128.36)
15	2.8	(92.1, 109.1, 128.1)	(92.64, 113.80, 137.09)	(92.1, 109.1, 118.1)	(92.68, 113.80, 133.63)
16	3	(103.7, 121.7, 142.7)	(88.69, 116.60, 146.92)	(103.7, 121.7, 135.7)	(90.18, 116.60, 142.80)

4. Computation results

The first example is recalled from [16]. The observed crisp inputs, fuzzy outputs and estimated fuzzy values, that also describe our Scenario I, are shown in Table 1.

Zeng et al. [16] used a fuzzy least absolute deviation regression to derive the quadratic predictor function in fuzzy environment. Their idea was to generalize the crisp least absolute deviation method to a similar fuzzy method. To compare our methodology to a formal-similar methodology that fails to fully comply with the extension principle, we followed Zeng et al.'s idea, replace the crisp least absolute deviation regression method by the least squares regression method, and derive results in fuzzy environment.

The numerical results for this scenario are graphed in Figure 1: Figure 1 (left) reports the results obtained by Zeng et al.'s approach [16] adapted to the least square regression, while our results are shown in Figure 1 (right).

We created the second scenario to explain the shortcomings of Zeng et al.'s class of approaches [16], and emphasize that our approach overcomes them. The observed inputs and outputs are also provided in Table 1.

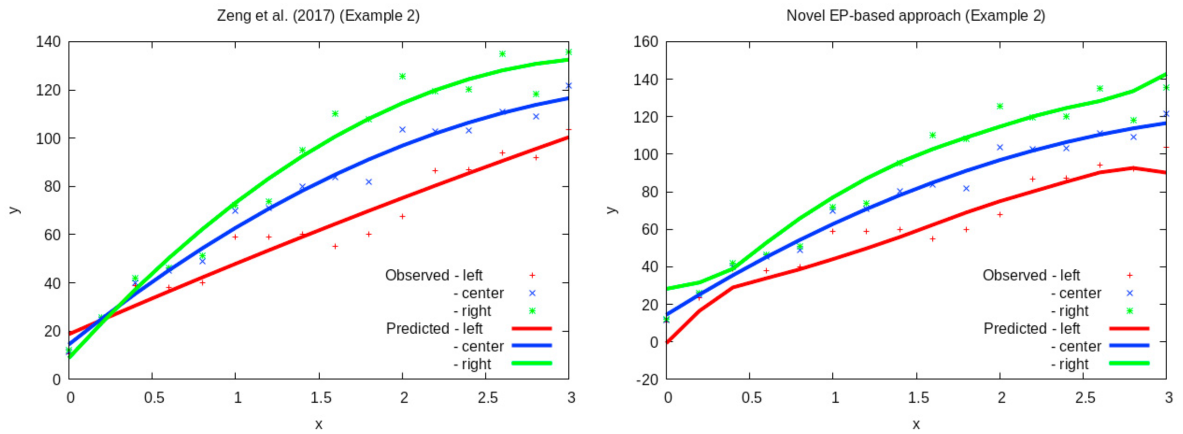

Fig. 2. The numerical results obtained for Scenario II. Left: using Zeng et al.'s approach adapted to fuzzy least square quadratic regression (note how the left and right estimations cross themselves). Right: using our approach that is in full accordance to the extension principle

The numeric results are shown in Figure 2. Note that the curves which describe the left and right endpoints of the estimated fuzzy outputs derived by the methodology introduced in [16] (Figure 2 (left)) cross each other, hence are not able to describe fuzzy number estimated outputs for certain inputs. On the other side, the curves derived by our methodology (Figure 2 (right)) being constructed in full accordance with the extension principle properly describe the estimated fuzzy outputs for any input.

Analyzing the results obtained for Scenario I (Figure 1) one can see as an advantage the narrower prediction outputs obtained when following Zeng et al.'s idea. The shapes of the regression lines obtained by our approach are obviously wider than in the literature on the intervals $[0, 0.5]$ and $[2.5, 3]$. However the advantage of our approach becomes clear when solving Scenario II (Figure 2). Our regression lines are similarly spread, as those obtained in Scenario I; while both the left and right regression lines of Zeng et al.'s approach cross their central regression line and also each other in the vicinity of $x = 0.2$. That crossover makes deriving proper fuzzy outputs impossible at least on the interval $[0, 0.2]$.

5. Final conclusion

The fuzzy quadratic least square regression for crisp observed inputs and fuzzy observed outputs was addressed in this study. We proposed a family of crisp optimization models that are able to derive the endpoints of the α-cut intervals of the estimated fuzzy values at any given crisp input. The estimated values were yielded in accordance to the extension principle. We carried out several experiments aiming to show that a fuzzy regression approach that does not comply to the extension principle may fail to estimate proper fuzzy values in certain contexts. A numerical example recalled from the literature was adapted to emphasize the main advantage of our approach compared to a class of approaches that split the involved fuzzy entities in their components instead of treating them in a unified manner as demanded by the extension principle.

We have also carried out some experiments on data-set having multiple-inputs X. However, in such a case the optimization models became non-linear, thus hard to be solved. Finding a solution approach that uses proper approximations still keeping focus on the extension principle within optimization is the next step in our research.

The role of the extension principle within the optimization problems is essential. The results derived by methodologies that ignore Zadeh's principle are potentially ineffective. Testing the relevance of such results in the context they were obtained, as well as analyzing their generality in wider contexts are of great importance in the field of fuzzy optimization. Extending the current methodology from fuzzy regression prediction to other related fields adapted to fuzzy environment is the main direction of our future research.

Acknowledgements

This work was supported by the Serbian Ministry of Education, Science and Technological Development through Mathematical Institute of the Serbian Academy of Sciences and Arts and Faculty of Organizational Sciences of the University of Belgrade.

References

[1] Chen, F., Chen, Y., Zhou, J., Liu, Y., 2016. Optimizing h value for fuzzy linear regression with asymmetric triangular fuzzy coefficients. Engineering Applications of Artificial Intelligence 47, 16–24. doi:10.1016/j.engappai.2015.02.011. artificial Intelligence Techniques in Product Engineering.

[2] Chen, L., Nien, S., 2020. A new approach to formulate fuzzy regression models. Applied Soft Computing 86. doi:10.1016/j.asoc.2019.105915.

[3] Chukhrova, N., Johannssen, A., 2019. Fuzzy regression analysis: Systematic review and bibliography. Applied Soft Computing 84, 105708. URL: https://www.sciencedirect.com/science/article/pii/S1568494619304892, doi:10.1016/j.asoc.2019.105708.

[4] Diniz, M., Gomes, L., Bassanezi, R., 2021. Optimization of fuzzy-valued functions using zadeh's extension principle. Fuzzy Sets and Systems 404, 23 – 37. URL: http://www.sciencedirect.com/science/article/pii/S0165011420302761, doi:https://doi.org/10.1016/j.fss.2020.07.007. fuzzy Numbers in Analysis.

[5] Ghanbari, R., Ghorbani-Moghadam, K., De Baets, B., 2019. Fuzzy linear programming problems: models and solutions. Soft Computing , 1433–7479doi:10.1007/s00500-019-04519-w.

[6] He, Y., Wang, X., Huang, J.Z., 2016. Fuzzy nonlinear regression analysis using a random weight network. Inf. Sci. 364-365, 222–240.

[7] Kashani, M., Arashi, M., Rabiei, M., D'Urso, P., De Giovanni, L., 2021. A fuzzy penalized regression model with variable selection. Expert Systems with Applications 175, 114696. URL: https://www.sciencedirect.com/science/article/pii/S0957417421001378, doi:10.1016/j.eswa.2021.114696.

[8] Kupka, J., 2016. On approximations of zadeh's extension principle. Fuzzy Sets and Systems 283, 26 – 39. URL: http://www.sciencedirect.com/science/article/pii/S0165011415001505, doi:https://doi.org/10.1016/j.fss.2015.03.014. theme: Fuzzy Intervals.

[9] Lu, K.P., Chang, S.T., 2018. A fuzzy classification approach to piecewise regression models. Applied Soft Computing 69, 671–688. doi:10.1016/j.asoc.2018.04.046.

[10] Skrabanek, P., Marek, J., Pozdilkova, A., 2021. Boscovich fuzzy regression line. Mathematics 9. URL: https://www.mdpi.com/2227-7390/9/6/685, doi:10.3390/math9060685.

[11] Stanojevic, B., Stanojevic, M., Nadaban, S., 2021. Reinstatement of the extension principle in approaching mathematical programming with fuzzy numbers. Mathematics 9. URL: https://www.mdpi.com/2227-7390/9/11/1272, doi:10.3390/math9111272.

[12] Tanaka, H., Hayashi, I., Watada, J., 1989. Possibilistic linear regression analysis for fuzzy data. European Journal of Operational Research 40, 389–396. URL: https://www.sciencedirect.com/science/article/pii/0377221789904311, doi:10.1016/0377-2217(89)90431-1.

[13] Tanaka, H., Uejima, S., Asai, K., 1982. Linear regression analysis with fuzzy model. IEEE Transactions on Systems Man and Cybernetics 12, 903–907. doi:10.1109/TSMC.1982.4308925.

[14] Zadeh, L., 1965. Fuzzy sets. Information and Control 8, 338 – 353. doi:10.1016/S0019-9958(65)90241-X.

[15] Zadeh, L., 1975. The concept of a linguistic variable and its application to approximate reasoning i. Information Sciences 8, 199 – 249. doi:10.1016/0020-0255(75)90036-5.

[16] Zeng, W., Feng, Q., Li, J., 2017. Fuzzy least absolute linear regression. Applied Soft Computing 52, 1009–1019. URL: https://www.sciencedirect.com/science/article/pii/S1568494616304884, doi:10.1016/j.asoc.2016.09.029.