Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 221 (2023) 1337–1342

Procedia Computer Science

www.elsevier.com/locate/procedia

Information Technology and Quantitative Management (ITQM 2023)

On modeling regression in full interval-valued fuzzy environment

Bogdana Stanojević[a,1], 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

We apply the general extension-principle-based approach to make predictions based on a regression model in a full interval-valued fuzzy environment. We use triangular interval-valued fuzzy numbers that model the uncertainty of the observed inputs and outputs to derive the predicted outputs in full accordance with Zadeh's extension principle. On one side, we enhance the Monte Carlo based algorithm introduced in the literature for simulating the output predictions of a fuzzy regression model by reducing the universe of random selections still keeping the accuracy of the empirical results; and on the other side, we solve quadratic models to derive the left endpoints of the $\alpha-$cut intervals of the exact results. We use one real-life problem from hydrology engineering with data recalled from the literature to carry out numerical experiments and illustrate our proposed methodology.

Keywords: interval-valued fuzzy number; fuzzy regression; least-squares method; extension principle; Monte Carlo simulation

1. Introduction

Fuzzy regression is widely studied nowadays. We refer the reader to [1], [2] and [3] for finding out the most recent ideas on fuzzy regression approaches. A systematic review is presented in [4].

This research is a continuation of the results published in [5], where a general approach in full accordance with the extension principle was proposed to a regression model with fuzzy-numbered observed data. Our first goal is to extend the approach to the full interval-valued fuzzy environment. The new proposed extension is described in Section 3.1.

Monte Carlo simulation algorithms were proposed in the literature in order to handle the complex fuzzy optimization problems. For instance, the approaches introduced in [3] and [7] to fuzzy regression analysis simulated the Zadeh's extension principle [6], and derived conformed empiric solutions. The importance of finding results in full compliance to the extension principle to fuzzy optimization problems was emphasized in [8].

Our next goal is to enhance the existing Monte Carlo simulation algorithm for fuzzy optimization by reducing the universe of random selections still keeping the accuracy of the empirical results. The new Monte Carlo based algorithm is described in Section 3.2.

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

10.1016/j.procs.2023.08.123

Within our experiments, we recall a real-life example from the literature. In Section 4 we report the numerical results derived by our new approach; and analyze them in comparison to the results reported in [9] and [10].

2. Notation and terminology

Type-2 fuzzy numbers were introduced in [6] as generalization to fuzzy numbers to enable the modeling of a higher level of uncertainty. A particular case of type-2 fuzzy numbers are the triangular interval-valued fuzzy numbers (TIVFNs) that can be represented by a quintuple $\left[\left(x_L^U, x_L^L\right), x, \left(x_R^L, x_R^U\right)\right]$, where $x_L^U, x_L^L, x, x_R^L, x_R^U \in R$, and $x_L^U \leq x_L^L \leq x \leq x_R^L \leq x_R^U$.

Generally, the membership function of a type-2 fuzzy set is defined with the help of a type-1 fuzzy set. Particularly, the membership degree of a value v in the TIVFN $\left[\left(x_L^U, x_L^L\right), x, \left(x_R^L, x_R^U\right)\right]$ is defined by an interval $[\mu_L(v), \mu_U(v)]$, where $\mu_L(v)$ is the membership degree of v in the triangular fuzzy number represented by the triple $\left(x_L^L, x, x_R^L\right)$, and $\mu_U(v)$ is the membership degree of v in the triangular fuzzy number represented by the triple $\left(x_L^U, x, x_R^U\right)$. The arithmetic operators on TIVFNs are defined with respect to the min operator applied separately to the lower and upper membership functions of the involved numbers.

Briefly, through a regression analysis a prediction function able to describe a general relation between inputs and outputs based on the observed inputs and outputs, is provided. In Section 3, X denotes the matrix of dimension $n \times (p+1)$ which contains p observed values of n inputs, and is bordered in front by a column of 1s; while the column matrix Y of dimension n contains the values describing the corresponding n observed outputs. Formula $X^T XA = X^T Y$ defines the column matrix A that contains the $p+1$ coefficients $a_p, \ldots, a_0$ of the prediction function derived by the least squared linear regression model.

The fuzzy regression models use the observed fuzzy inputs and outputs $\widetilde{x}_{ij}$ and $\widetilde{y}_i$, $i = 1, \ldots, n, j = 1, \ldots, p$, respectively. In the context of this paper, the fuzzy quantities are TIVFNs, and their components are used within Models (1) and (2).

For the detailed terminology related to least squared fuzzy regression we refer the reader to [5].

3. Theoretical results

In this section we propose two modifications to the existing methods from the literature, aiming to provide improved solutions to regression optimization problems. Firstly, we adjust the Extention-principle-based regression optimization process EPBRO proposed in [5] to work in full interval-valued fuzzy environment, and present the new formulation in Section 3.1. Secondly, we enhance the Monte Carlo simulation algorithm [3] by reducing the universe of random selections, and present the new algorithm in Section 3.2.

3.1. Extended EPBRO method

In this section we aim to adjust the Extention-principle-based regression optimization process EPBRO [5] to work in full interval-valued fuzzy environment. Interpreting an TIVFN $X_{\text{TIVFN}} = \left[\left(x_L^U, x_L^L\right), x, \left(x_R^L, x_R^U\right)\right]$ as two TFNs $\widetilde{X}_{\text{FN}}^L = \left(x_L^L, x, x_R^L\right)$ and $\widetilde{X}_{\text{FN}}^U = \left(x_L^U, x, x_R^U\right)$, we may apply EPBRO twice, on the observed data $\left(\widetilde{X}_{i\text{FN}}^L, \widetilde{Y}_{i\text{FN}}^L\right)_{i=1,\ldots,n}$ and $\left(\widetilde{X}_{i\text{FN}}^U, \widetilde{Y}_{i\text{FN}}^U\right)_{i=1,\ldots,n}$, respectively, using Model (1) recalled form [5].

$$
\begin{array}{lll}
\min (\max) & a_0 + a_1 v_1 + a_2 v_2 + \cdots + a_p v_p & \\
\text{s.t.} & & \\
 & X^T XA = X^T Y, & \\
 & (1-\alpha)\left(\widetilde{x}_{ij}\right)^L + \alpha\left(\widetilde{x}_{ij}\right)^C \leq x_{ij} \leq \alpha\left(\widetilde{x}_{ij}\right)^C + (1-\alpha)\left(\widetilde{x}_{ij}\right)^U, & i = 1, \ldots, n, \\
 & & j = 1, \ldots, p, \\
 & (1-\alpha)\left(\widetilde{v}_j\right)^L + \alpha\left(\widetilde{v}_j\right)^C \leq v_j \leq \alpha\left(\widetilde{v}_j\right)^C + (1-\alpha)\left(\widetilde{v}_j\right)^U, & j = 1, \ldots, p, \\
 & (1-\alpha)\left(\widetilde{y}_i\right)^L + \alpha\left(\widetilde{y}_i\right)^C \leq y_i \leq \alpha\left(\widetilde{y}_i\right)^C + (1-\alpha)\left(\widetilde{y}_i\right)^U, & i = 1, \ldots, n, \\
 & a_j \text{ free variable}, & j = 0, \ldots, p,
\end{array}
\tag{1}
$$

where α is fixed and $\widetilde{v}$ is the analyzed fuzzy input. The optimization is made with respect to the variables v, a_0, a_j, x_{ij} and y_i, $i = 1, \ldots, n$, $j = 1, \ldots, p$. The objective function of Model (1) is minimized (maximized) to derive the left (right) side of the membership function of the estimated fuzzy output.

First run of EPBRO would derive the lower membership function of the estimated TIVFN outputs, i.e. an TFN $\widehat{Y}_i^L = \left(\widehat{y}_{iL}^L, \widehat{y}_i, \widehat{y}_{iR}^L\right)$ for each $i = 1, \ldots, n$, while the second run would derive the upper membership function of the same estimations, i.e. an TFN $\widehat{Y}_i^U = \left(\widehat{y}_{iL}^U, \widehat{y}_i, \widehat{y}_{iR}^U\right)$ for each $i = 1, \ldots, n$. In this way, the estimated $\widehat{Y}^i_{\text{TIVFN}}$ in full interval-valued fuzzy environment would be $\left[\left(\widehat{y}_{iL}^U, \widehat{y}_{iL}^L\right), \widehat{y}_i, \left(\widehat{y}_{iR}^L, \widehat{y}_{iR}^U\right)\right]$, $i = 1, \ldots, n$.

To assure that $\left[\left(\widehat{y}_{iL}^U, \widehat{y}_{iL}^L\right), \widehat{y}_i, \left(\widehat{y}_{iR}^L, \widehat{y}_{iR}^U\right)\right]$ are TIVFNs, $i = 1, \ldots, n$ the following two statements should be true: (i) the central values of $\widehat{Y}_i^L$and $\widehat{Y}_i^U$ are the same; and (ii) $\widehat{y}_{iL}^U \le \widehat{y}_{iL}^L$ and $\widehat{y}_{iR}^L \le \widehat{y}_{iR}^U$. Statement (i) is evident, since the central values of both estimations $\widehat{Y}_i^L$and $\widehat{Y}_i^U$ are derived based on a crisp regression optimization applied to the crisp observed data $(x_i, y_i)_{i=1,\ldots,n}$, where x_i, y_i are the central values of $X^L_{i\text{FN}}, Y^L_{i\text{FN}}$, $i = 1, \ldots, n$, respectively.

For proving the inequalities included in Statement (ii) we propose Model (2) that optimizes the membership degree α with respect to the fixed real value y^* and the variables α, v, a_0, a_j, x_{ij} and y_i, $i = 1, \ldots, n$, $j = 1, \ldots, p$.

$$
\begin{array}{lll}
\max & \alpha & \\
\text{s.t.} & & \\
& a_0 + a_1 v_1 + a_2 v_2 + \cdots + a_p v_p = y^* & \\
& X^T X A = X^T Y, & \\
& (1-\alpha)\left(\widetilde{x}_{ij}\right)^L + \alpha\left(\widetilde{x}_{ij}\right)^C \le x_{ij} \le \alpha\left(\widetilde{x}_{ij}\right)^C + (1-\alpha)\left(\widetilde{x}_{ij}\right)^U, & i = 1, \ldots, n, \\
& & j = 1, \ldots, p, \\
& (1-\alpha)\left(\widetilde{v}_j\right)^L + \alpha\left(\widetilde{v}_j\right)^C \le v_j \le \alpha\left(\widetilde{v}_j\right)^C + (1-\alpha)\left(\widetilde{v}_j\right)^U, & j = 1, \ldots, p, \\
& (1-\alpha)\left(\widetilde{y}_i\right)^L + \alpha\left(\widetilde{y}_i\right)^C \le y_i \le \alpha\left(\widetilde{y}_i\right)^C + (1-\alpha)\left(\widetilde{y}_i\right)^U, & i = 1, \ldots, n, \\
& a_j \text{ free variable}, & j = 0, \ldots, p.
\end{array}
\tag{2}
$$

Let $\mu_L(y^*)$ and $\mu_U(y^*)$ denote the optimal values obtained by solving (2) for the lower and upper descriptions of the observed data $\left(\widetilde{X}^L_{i\text{FN}}, \widetilde{Y}^L_{i\text{FN}}\right)_{i=1,\ldots,n}$and $\left(\widetilde{X}^U_{i\text{FN}}, \widetilde{Y}^U_{i\text{FN}}\right)_{i=1,\ldots,n}$, respectively. Due to the box constraints of Model (2), the feasible set defined by the observed data $\left(\widetilde{X}^L_{i\text{FN}}, \widetilde{Y}^L_{i\text{FN}}\right)_{i=1,\ldots,n}$ is a subset of the feasible set defined by $\left(\widetilde{X}^U_{i\text{FN}}, \widetilde{Y}^U_{i\text{FN}}\right)_{i=1,\ldots,n}$. Therefore, $\mu_L(y^*) \le \mu_U(y^*)$ for any real value y^*.

Models (1) and (2) provide different perspectives on the same interval-valued fuzzy number that is the prediction of the regression model at input $\widetilde{v}$.

3.2. Enhanced Monte Carlo simulation algorithm

Any simulation provided by a Monte Carlo method is based on random selections of relevant values from their corresponding universes. The Monte Carlo simulation algorithm proposed for deriving extension-principle-based empiric solutions to fuzzy regression [3] selected random values form the $\alpha-$cut intervals of the involved fuzzy quantities.

To downsize the selection spaces we randomly select either the left or right endpoint of each $\alpha-$cut interval. In this way, the simulation faster provides a better spread of the relevant results.

Many fuzzy optimization problems were solved in the literature by using only the left (right) endpoints of the $\alpha-$cut intervals of all fuzzy coefficients in deriving the left (right) side of the membership function of the fuzzy solution. However, our experiments showed that such approaches yield improper fuzzy solutions.

4. Computation results

Within our experiments we use the real-life example from hydrology engineering described in [9] and recalled in [10]. The example consists of $n = 51$ observed pairs of inputs $X^i_{\text{TIVFN}} = \left[\left(x_L^U, x_L^L\right), x, \left(x_R^L, x_R^U\right)\right]_i$ and outputs $Y^i_{\text{TIVFN}}.\left[\left(y_L^U, y_L^L\right), y, \left(y_R^L, y_R^U\right)\right]_i$, $i = 1, \ldots, n$, expressed by TIVFNs.

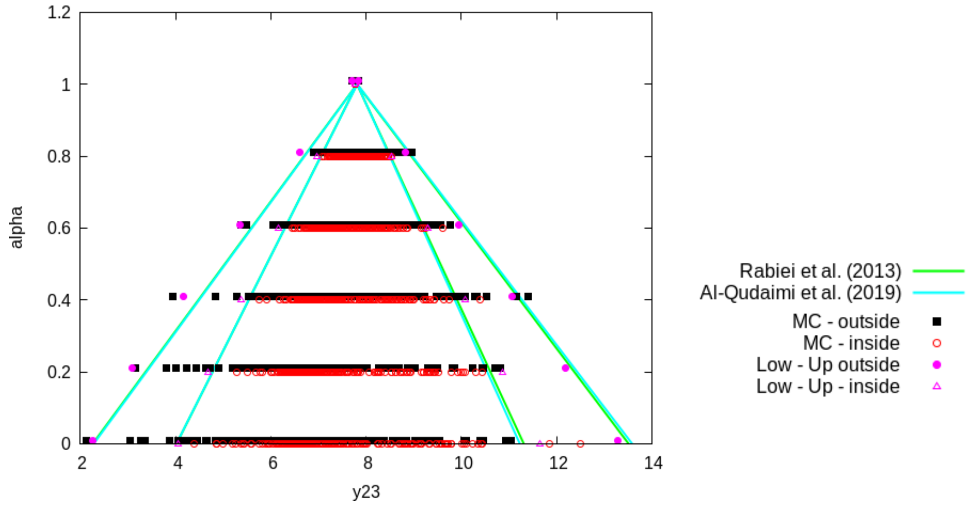

Fig. 1. Comparative analysis of the results derived by our enhanced Monte Carlo simulation algorithm on one side and the results reported in [9] and [10] on the other side, all for input X^{23}_{TIVFN}

The estimated outputs of the methodology proposed in [9] can be computed using the regression function

$$[(2.34, 3.96), 5.54, (7.79, 9.04)] + ([(0.034, 0.034), 0.216, (0.216, 0.216)] + [(-0.004, -0.004), 0, (0, 0)]) \times X_{\text{TIVFN}}.$$

The results reported in [10] can be computed using the regression function

$$[(2.28, 3.98), 5.54, (7.70, 9.11)] + [(0.01, 0.01), 0.216, (0.216, 0.216)] \times X_{\text{TIVFN}}.$$

Our empirical results can be derived using the enhanced Monte Carlo simulation method. A graphic representation of the estimated outputs obtained for the observed input X^{23}_{TIVFN} are shown in Figure 1. We named "Low - Up" the approach that derives the solutions using the left endpoints of the $\alpha-$cut intervals for all fuzzy coefficients; and separately, the right endpoints of the $\alpha-$cut intervals for all fuzzy coefficients. Both simulations were performed twice, for deriving the inside and outside descriptions of the TIVFN solutions, respectively.

There are several conclusions that can be made based on the representation given in Figure 1:

- both representations of [9] and [10] are very similar; there are no relevant differences between them;
- the estimations made by using the "Low - Up" approach are very close to the results derived in the literature. However, they do not describe estimated outputs that comply to the extension principle, since there are Monte Carlo simulated smaller/greater values with the same membership degree (note, for instance, the representation for $\alpha = 0.4$ and $\alpha = 0.8$);
- the enhanced Monte Carlo simulation (named "MC" in Figure 1) shows that there exist relevant values of the output $\widetilde{y}_{23}$ which do not fit between the borders provided in [9] and [10] (not all red circles are within the borders of the inside triangle; and not all black squares are within the borders of the outside triangle).

Figure 2 shows the left sides of both inside and outside membership functions of the exact TIVFN representing the estimated output derived for input X^{1}_{TIVFN} using the extended EPBRO algorithm and the algorithms described in [9] and [10]. The estimation provided by our approach is wider than those provided in the literature but more relevant for a regression analysis fully complying to the extension principle. The left sides of the extended EPBRO were derived using Model (1), while the right ones were derived using Model (2).

5. Conclusions and further researches

In this paper we explained how to apply the general extension-principle-based approach in order to make predictions based on a regression model in a full interval-valued fuzzy environment. We used triangular interval-

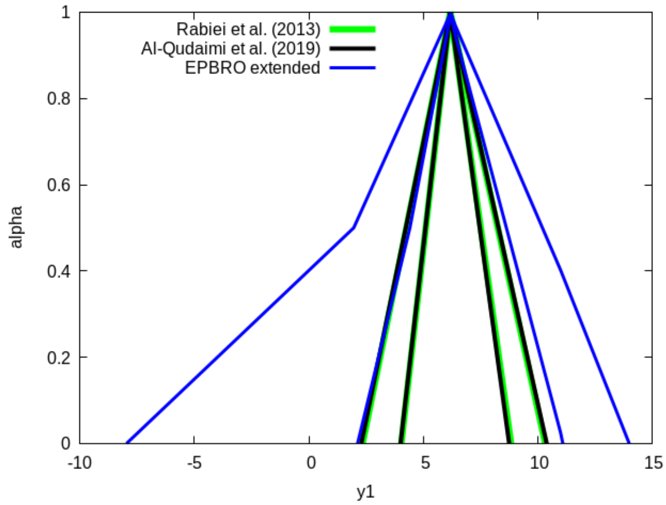

Fig. 2. Comparative analysis of the results derived by the extended EPBRO algorithm on one side and the results reported in [9] and [10] on the other side, all for input X^1_{TIVFN}

valued fuzzy numbers that modeled the uncertainty of the observed inputs and outputs; and derived the predicted outputs in full accordance with Zadeh's extension principle.

On one side, we enhanced the Monte Carlo based algorithm introduced in the literature for simulating the output predictions of a fuzzy regression model by reducing the universe of random selections still keeping the accuracy of the empirical results; and on the other side, we solved quadratic models to derive the left endpoints of the $\alpha-$cut intervals of the exact results.

We used one real-life problem from hydrology engineering with data recalled from the literature to carry out numerical experiments and illustrate our proposed methodology.

Further researches on this theme might be fruitful using more general regression functions, aggregating operators and/or other types of fuzzy numbers in modeling the uncertain data. Employing heuristics to solve more complex optimization models is also desirable.

Acknowledgements

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

References

[1] E. Bas, Robust fuzzy regression functions approaches, Information Sciences 613 (2022) 419–434. doi:https://doi.org/10.1016/j.ins.2022.09.047.
URL `https://www.sciencedirect.com/science/article/pii/S0020025522010921`

[2] J. Chachi, S. M. Taheri, P. D'Urso, Fuzzy regression analysis based on m-estimates, Expert Systems with Applications 187 (2022) 115891. doi:https://doi.org/10.1016/j.eswa.2021.115891.
URL `https://www.sciencedirect.com/science/article/pii/S0957417421012471`

[3] B. Stanojević, M. Stanojević, Extension-principle-based approach to least square fuzzy linear regression, in: S. Dzitac, D. Dzitac, F. G. Filip, J. Kacprzyk, M.-J. Manolescu, H. Oros (Eds.), Intelligent Methods Systems and Applications in Computing, Communications and Control, Springer International Publishing, Cham, 2023, pp. 219–228.

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

[5] B. Stanojević, M. Stanojević, Optimization-based fuzzy regression in full compliance with the extension principle, International Journal of Computers, Communications, Control 18(2) (2023) 5320. doi:https://doi.org/10.15837/ijccc.2023.2.5320.

[6] L. Zadeh, The concept of a linguistic variable and its application to approximate reasoning i, Information Sciences 8 (3) (1975) 199 – 249. doi:10.1016/0020-0255(75)90036-5.

[7] B. Stanojević, M. Stanojević, Quadratic least square regression in fuzzy environment, Procedia Computer Science 214 (2022) 391–396, 9th International Conference on Information Technology and Quantitative Management. doi:https://doi.org/10.1016/j.procs.2022.11.190. URL `https://www.sciencedirect.com/science/article/pii/S1877050922019007`

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

[9] M. Rabiei, N. Arghami, S. Taheri, B. Gildeh, Least-squares approach to regression modeling in full interval-valued fuzzy environment, Soft Computing 18 (2014) 2043–2059.

[10] A. Al-Qudaimi, A. Kumar, Comment on "Least-squares approach to regression modeling in full interval-valued fuzzy environment", Soft Computing 23 (2019) 10019–10027. doi:10.1007/s00500-018-3556-4.