FUZZY OPTIMIZATION AND MODELLING 2(3) (2021) 17-23

E-ISNN: 2676-7007

Contents lists available at FOMJ

Fuzzy Optimization and Modelling

Journal homepage: http://fomj.qaemiau.ac.ir/

Paper Type: Research Paper

Fuzzy Regression Models Using the Least-Squares Method based on the Concept of Distance: Simplified Approach

Abdullah Al-Qudaimi [a,*], **Walid S. M. Yousef** [a]

[a] *Department of Information Technology, University of Science and Technology, Sana'a, Yemen*

ARTICLE INFO

Article history:
Received 15 June 2021
Revised 30 August 2021
Accepted 3 September 2021
Available online 3 September 2021

Keywords:
Fuzzy regression model
Fuzzy sets
Least squares method

ABSTRACT

Regression models have been tremendously studying with so many applications in the presence of imprecise data. The regression coefficients are unknown i.e., they cannot be restricted. To the best of our knowledge, there is no approach except Chen and Hsueh approach (IEEE Transactions on Fuzzy Systems, vol. 17, no. 6, December 2009 pp.1259-1272) which can be used to find the regression coefficients of a fuzzy regression model without considering the non-negative restrictions on the regression coefficients. Chen and Hsueh have used some mathematical assumptions which lead to limitations in their approach. Furthermore, Chen and Hsueh approach is inefficient regarding to computational complexity. This paper proposed a simplified approach overcoming the limitations and computational complexity of Chen and Hsueh approach which can be considered by the researchers who would like to use Chen and Hsueh approach in real life applications.

1. Introduction

Fuzzy regression analysis has been vastly used with the perspective of least-squares based methods, where finding the square deviation between two variables is used in regression. The optimal regression coefficients are estimated by minimizing the error comes from the deviation between an estimated and a given fuzzy response. Since the first fuzzy least-squares approaches initiated [1, 2, 6, 7], various fuzzy least squares regression have been presented [3]. The fuzzy least-squares approach proposed by Chen and Hsueh [5] categorized by Chukhrova and Johannssen in their systematic comprehensive review [3] as a development in least-squares method. Chen and Hsueh [4] pointed out the flaws of the existing methods for evaluating the regression coefficients of fuzzy regression models and proposed a mathematical programming method, based on distance criteria to determine the regression coefficients. Chen and Hsueh [5] pointed out that if the observation is large, then the efficiency of Chen and Hsueh approach [5] may reduce due to the increase in number of constraints. Chen and Hsueh [5] also pointed out that the Chen and Hsueh [4] approach and other existing approaches are proposed by considering the regression coefficients as positive real numbers. However, it is well-known fact

* Correspondig author
E-mail address: aalqudaimi@yahoo.com (Abduallah Al-Qudamini)

DOI: 10.30495/fomj.2021.1933334.1031

that some of regression coefficients may be positive real numbers and the remaining real numbers may be negative.

To overcome this limitation of Chen and Hsueh approach [4] as well as other existing approaches, Chen and Hsueh [5] proposed an approach that uses the least squares method to minimize the total estimation of the distance between the observed and estimated responses.

To the best of our knowledge, there is no other approach except Chen and Hsueh approach [5] which can be used to find the regression coefficients of a fuzzy regression model without considering the nonnegative restrictions on the regression coefficients. Motivated by the above discussion, this paper aims to propose a simplified approach which is sound and general without a need to predict the sign of regression coefficient(s) and consequently, reduces the computational complexity of constructing fuzzy regression models.

2. Chen and Hsueh approach

Chen and Hsueh [5] proposed the following approach to find the fuzzy regression model,

$$\hat{\tilde{Y}}_i = b_0 + b_1\tilde{X}_{i1} + b_2\tilde{X}_{i2} + \cdots + b_p\tilde{X}_{ip} + \tilde{\delta}, \quad i = 1, \dots, n.$$

where, $\tilde{\delta}$, $\hat{\tilde{Y}}_i$ and $\tilde{X}_{ij}, j = 1, \dots, p$, are fuzzy numbers for finding the regression coefficients.

Step 1: Assuming the fuzzy adjustment term $\tilde{\delta}$ as triangular fuzzy number $(\delta^L, \delta^M, \delta^U)$, the considered fuzzy regression model may be rewritten as:

$\hat{\tilde{Y}}_i = b_0 + b_1\tilde{X}_{i1} + b_2\tilde{X}_{i2} + \cdots + b_p\tilde{X}_{ip} + (\delta^L, \delta^M, \delta^U), i = 1,2, \dots, n.$

Step 2: Replacing the fuzzy numbers $\tilde{X}_{ij}$, $\hat{\tilde{Y}}_i$ and $\tilde{\delta}$ with the corresponding k^{th} $\alpha-$cut (say, α_k), $\left[\left(\tilde{X}_{ij}\right)^L_{\alpha_k}, \left(\tilde{X}_{ij}\right)^U_{\alpha_k}\right]$, $\left[\left(\hat{\tilde{Y}}_i\right)^L_{\alpha_k}, \left(\hat{\tilde{Y}}_i\right)^U_{\alpha_k}\right]$ and $[\delta^L + \alpha_k(\delta^M - \delta^L), \delta^U - \alpha_k(\delta^U - \delta^M)]$ respectively, the fuzzy regression model, obtained in Step 1, may be transformed into the following interval regression model.

$\left[\left(\hat{\tilde{Y}}_i\right)^L_{\alpha_k}, \left(\hat{\tilde{Y}}_i\right)^U_{\alpha_k}\right] = b_0 + b_1\left[\left(\tilde{X}_{i1}\right)^L_{\alpha_k}, \left(\tilde{X}_{i1}\right)^U_{\alpha_k}\right] + b_2\left[\left(\tilde{X}_{i2}\right)^L_{\alpha_k}, \left(\tilde{X}_{i2}\right)^U_{\alpha_k}\right] + \cdots + b_p\left[\left(\tilde{X}_{ip}\right)^L_{\alpha_k}, \left(\tilde{X}_{ip}\right)^U_{\alpha_k}\right] +$ $[\delta^L + \alpha_k(\delta^M - \delta^L), \delta^U - \alpha_k(\delta^U - \delta^M)], i = 1, 2, \dots, n\,; k = 0,1, \dots, m-1$, where m represents the number of $\alpha-$cuts.

Step 3: Assuming $\alpha_k = \frac{k}{m-1}$, the interval regression model, obtained in Step 2, may be transformed into the following interval regression model.

$\left[\left(\hat{\tilde{Y}}_i\right)^L_{\alpha_k}, \left(\hat{\tilde{Y}}_i\right)^U_{\alpha_k}\right] = b_0 + b_1\left[\left(\tilde{X}_{i1}\right)^L_{\alpha_k}, \left(\tilde{X}_{i1}\right)^U_{\alpha_k}\right] + b_2\left[\left(\tilde{X}_{i2}\right)^L_{\alpha_k}, \left(\tilde{X}_{i2}\right)^U_{\alpha_k}\right] + \cdots + b_p\left[\left(\tilde{X}_{ip}\right)^L_{\alpha_k}, \left(\tilde{X}_{ip}\right)^U_{\alpha_k}\right] +$ $\left[\delta^L + \left(\frac{k}{m-1}\right)(\delta^M - \delta^L), \delta^U - \left(\frac{k}{m-1}\right)(\delta^U - \delta^M)\right], i = 1,2, \dots, n\,; k = 0,1, \dots, m-1.$

Step 4: Find the correlation coefficient $\rho_{X_jY} = \frac{\sum_i(\dot{X}_{ij} - \bar{X}_j)(\dot{Y}_i - \bar{Y})}{\sqrt{\sum_i(\dot{X}_{ij} - \bar{X}_j)^2(\dot{Y}_i - \bar{Y})^2}}$, where $\dot{X}_{ij} = \frac{X^L_{ij} + X^M_{ij} + X^U_{ij}}{3}$ and $\dot{Y}_i = \frac{Y^L_{ij} + Y^M_{ij} + Y^U_{ij}}{3}$ represents the defuzzified values of $\tilde{X}_{ij} = (X^L_{ij}, X^M_{ij}, X^U_{ij})$ and $\tilde{Y}_i = (Y^L_{ij}, Y^M_{ij}, Y^U_{ij})$ respectively and $\bar{X}_j = \sum_i \frac{\dot{X}_{ij}}{n}$, and $\bar{Y} = \sum_i \frac{\dot{Y}_i}{n}$.

Case 1: If ρ_{X_jY} is negative then consider the j^{th} regression coefficient as negative real number and hence, replace $b_j\left[\left(\tilde{X}_{ij}\right)^U_{\alpha_k}, \left(\tilde{X}_{ij}\right)^L_{\alpha_k}\right]$ with $\left[b_j\left(\tilde{X}_{ij}\right)^U_{\alpha_k}, b_j\left(\tilde{X}_{ij}\right)^L_{\alpha_k}\right]$ in the interval regression model obtained in Step 3.

Case 2: If ρ_{X_jY} is positive then consider the j^{th} regression coefficient as positive real number and hence, replace $b_j\left[\left(\tilde{X}_{ij}\right)^L_{\alpha_k}, \left(\tilde{X}_{ij}\right)^U_{\alpha_k}\right]$ with $\left[b_j\left(\tilde{X}_{ij}\right)^L_{\alpha_k}, b_j\left(\tilde{X}_{ij}\right)^U_{\alpha_k}\right]$ in the interval regression model obtained in Step 3.

Case 3: If neither ρ_{X_jY} is positive nor ρ_{X_jY} is negative then assume the j^{th} regression coefficient b_j as negative real number and find the value of $b_j$by using the remaining steps.

Case (3a): If the obtained value of b_j is negative then it is the required value of b_j.

Case (3b): If the obtained value of b_j is positive then apply the remaining steps again to find the value of b_j by considering b_j as positive real number.

Step 5: For simplicity, let us assume that all the regression coefficients b_j are positive and hence, the interval regression model, obtained in Step 3, may be transformed into the following interval regression model.

$$\left[\left(\hat{\tilde{Y}}_i\right)^L_{\alpha_k}, \left(\hat{\tilde{Y}}_i\right)^U_{\alpha_k}\right] = b_0 + \left[b_1\left(\tilde{X}_{i1}\right)^L_{\alpha_k}, b_1\left(\tilde{X}_{i1}\right)^U_{\alpha_k}\right] + \cdots + \left[b_p\left(\tilde{X}_{ip}\right)^L_{\alpha_k}, b_p\left(\tilde{X}_{ip}\right)^U_{\alpha_k}\right] + \left[\delta^L + \left(\frac{k}{m-1}\right)(\delta^M - \delta^L), \delta^U - \left(\frac{k}{m-1}\right)(\delta^U - \delta^M)\right] \; i = 1,2,\ldots,n\,; k = 0,1,\ldots,m-1.$$

Step 6: Using the arithmetic operations, $[a^L, a^U] + [b^L, b^U] = [a^L + b^L, a^U + b^U]$ and $\gamma + [a^L, a^L] = [\gamma + a^L, \gamma + a^U]$, the interval regression model, obtained in Step 5, may be transformed into the following interval regression model.

$$\left[\left(\hat{\tilde{Y}}_i\right)^L_{\alpha_k}, \left(\hat{\tilde{Y}}_i\right)^U_{\alpha_k}\right] = \left[b_0 + b_1\left(\tilde{X}_{i1}\right)^L_{\alpha_k} + \cdots + b_p\left(\tilde{X}_{ip}\right)^L_{\alpha_k} + \delta^L + \left(\frac{k}{m-1}\right)(\delta^M - \delta^L), b_0 + b_1\left(\tilde{X}_{i1}\right)^U_{\alpha_k} + \cdots + b_p\left(\tilde{X}_{ip}\right)^U_{\alpha_k} + \delta^U - \left(\frac{k}{m-1}\right)(\delta^U - \delta^M)\right], i = 1,2,\ldots,n\,; k = 0,1,\ldots,m-1.$$

Step 7: Using the relation, $[a^L, a^U] = [b^L, b^U] \Rightarrow a^U = b^L, a^U = b^U$, the interval regression model, obtained in Step 6, may be transformed into the following two crisp regression models.

$$\left(\hat{\tilde{Y}}_i\right)^L_{\alpha_k} = b_0 + b_1\left(\tilde{X}_{i1}\right)^L_{\alpha_k} + \cdots + b_p\left(\tilde{X}_{ip}\right)^L_{\alpha_k} + \delta^L + \left(\frac{k}{m-1}\right)(\delta^M - \delta^L), \; i = 1,2,\ldots,n\,; k = 0,1,\ldots,m-1.$$

and

$$\left(\hat{\tilde{Y}}_i\right)^U_{\alpha_k} = b_0 + b_1\left(\tilde{X}_{i1}\right)^U_{\alpha_k} + \cdots + b_p\left(\tilde{X}_{ip}\right)^U_{\alpha_k} + \delta^U - \left(\frac{k}{m-1}\right)(\delta^U - \delta^M), i = 1,2,\ldots,n\,; k = 0,1,\ldots,m-1.$$

Step 8: Find the mean squared error E_i between the estimated and the observed responses.

$$E_i = \frac{1}{2m}\sum_{k=0}^{m-1}\left[\left(\left(b_0 + b_1\left(\tilde{X}_{i1}\right)^L_{\alpha_k} + \cdots + b_p\left(\tilde{X}_{ip}\right)^L_{\alpha_k} + \delta^L + \left(\frac{k}{m-1}\right)(\delta^M - \delta^L)\right) - \left(\tilde{Y}_i\right)^L_{\alpha_k}\right)^2 - \left(\left(b_0 + b_1\left(\tilde{X}_{i1}\right)^U_{\alpha_k} + \cdots + b_p\left(\tilde{X}_{ip}\right)^U_{\alpha_k} + \delta^U - \left(\frac{k}{m-1}\right)(\delta^U - \delta^M)\right) - \left(\tilde{Y}_i\right)^U_{\alpha_k}\right)^2\right], \; i = 1,2,\ldots,n.$$

Step 9: Find the total mean squared error

$$\sum_{i=1}^n E_i = \sum_{i=1}^n \frac{1}{2m}\sum_{k=0}^{m-1}\left[\left(\left(b_0 + b_1\left(\tilde{X}_{i1}\right)^L_{\alpha_k} + \cdots + b_p\left(\tilde{X}_{ip}\right)^L_{\alpha_k} + \delta^L + \left(\frac{k}{m-1}\right)(\delta^M - \delta^L)\right) - \left(\tilde{Y}_i\right)^L_{\alpha_k}\right)^2 - \left(\left(b_0 + b_1\left(\tilde{X}_{i1}\right)^U_{\alpha_k} + \cdots + b_p\left(\tilde{X}_{ip}\right)^U_{\alpha_k} + \delta^U - \left(\frac{k}{m-1}\right)(\delta^U - \delta^M)\right) - \left(\tilde{Y}_i\right)^U_{\alpha_k}\right)^2\right].$$

Step 10: Find the three linear equations by putting $\frac{\partial}{\partial \delta^L}\left(\sum_{i=1}^n E_i\right) = 0$, $\frac{\partial}{\partial \delta^M}\left(\sum_{i=1}^n E_i\right) = 0$ and $\frac{\partial}{\partial \delta^U}\left(\sum_{i=1}^n E_i\right) = 0$.

Step 11: Find the values of $b_0 + \delta^L$, $b_0 + \delta^M$ and $b_0 + \delta^U$ in terms of b_j with the help of the linear equations obtained by putting $\frac{\partial}{\partial \delta^L}\left(\sum_{i=1}^n E_i\right) = 0$, $\frac{\partial}{\partial \delta^M}\left(\sum_{i=1}^n E_i\right) = 0$ and $\frac{\partial}{\partial \delta^U}\left(\sum_{i=1}^n E_i\right) = 0$ respectively.

Step 12: Find a system of linear equations by putting $\frac{\partial}{\partial b_j}\left(\sum_{i=1}^n E_i\right) = 0$.

Step 13: Put the values of $b_0 + \delta^L$, $b_0 + \delta^M$ and $b_0 + \delta^U$, obtained in Step 11, in the system of linear equations obtained in Step 12 by putting $\frac{\partial}{\partial b_j}\left(\sum_{i=1}^n E_i\right) = 0$.

Step 14: Solve the system of linear equations, obtained in Step 13, to obtain the values of b_j.

Step 15: Find the values of $b_0+\delta^L$, $b_0+\delta^M$ and $b_0+\delta^U$ by putting the values of b_j in the expressions of $b_0+\delta^L$, $b_0+\delta^M$ and $b_0+\delta^U$, obtained in Step 11.

Step 16: Find the fuzzy regression model $\hat{\tilde{Y}}_i = b_0 + b_1\tilde{X}_{i1} + b_2\tilde{X}_{i2} + \cdots + b_p\tilde{X}_{ip} + \tilde{\delta} = (b_0+\delta^L, b_0+\delta^M, b_0+\delta^U) + b_1\tilde{X}_{i1} + b_2\tilde{X}_{i2} + \cdots + b_p\tilde{X}_{ip}$ by putting the obtained values of $b_0+\delta^L, b_0+\delta^M, b_0+\delta^U$ and b_j.

3. Limitation of Chen and Hsueh approach

It is obvious from Step 4 of Chen and Hsueh approach [5] that this approach can be used only if the obtained correlation coefficient is negative or positive or 0 . However, if the obtained correlation coefficient is an indeterminate value, this approach cannot be used. For example, it can be easily verified that the correlation coefficient between the response variable $\tilde{y}_1 = (-9,2,7)$, $\tilde{y}_2 = (-10,0,10)$ and the explanatory variable $\tilde{x}_{11} = (-5,-2,7)$ and $\tilde{x}_{21} = (-3,-1,4)$, obtained by using the expression mentioned in Step 4 of Chen and Hsueh approach [5], the obtained value of correlation coefficient is $\frac{0}{0}$, which is an indeterminate value. Therefore, Chen and Hsueh approach [5] cannot be used to find the fuzzy regression model for fitting this data.

4. Computational complexity of Chen and Hsueh approach

It can be easily verified that on applying the proposed approach as well as Chen and Hsueh approach [5] the obtained fuzzy regression models will be same. However, it is better to apply the proposed approach as compared to Chen and Hsueh approach [5] due to the following reasons:

To apply Chen and Hsueh approach [5], there is firstly a need to find the correlation coefficient between each explanatory variable and response variable.

Let suppose that there are n such explanatory variables for which the correlation coefficient is 0 then according to Step 4 of Chen and Hsueh approach [5]; $b_1, b_2, \ldots, b_n$ will be initially obtained by considering these as positive real numbers. If all the values $b_1, b_2, \ldots, b_n$ are positive then there is no need to repeat Chen and Hsueh approach [5]. However, if out of the n, m are negative then there is need to repeat Chen and Hsueh approach [5] to calculate new values of $b_1, b_2, \ldots, b_n$ by considering these m regression coefficients as negative real numbers. Hence, high computational efforts are required to apply Chen and Hsueh approach [5].

On applying Step 1 to Step 11 of Chen and Hsueh approach [5] for $\tilde{y}_1 = (-8,1,7)$, $\tilde{y}_2 = (-5,0,5)$, $\tilde{x}_{11} = (-4,-2,6)$ and $\tilde{x}_{21} = (-2,-1,3)$, the following values of $b_0+\delta^L$, $b_0+\delta^M, b_0+\delta^U$ in terms of b_1 are obtained.

$$b_0+\delta^L = -\frac{13}{2} - \frac{9}{2}b_1,\ b_0+\delta^M = \frac{1}{2} + \frac{3}{2}b_1,\ \ b_0+\delta^U = 6 + 3b_1 .$$

Furthermore, on applying Step 12 of Chen and Hsueh approach [5] for the considered data, the following linear equation is obtained.

$$70b_1 + 9\delta_l - 3\delta_m - 6\delta_u = -103.$$

According to Step 13 of Chen and Hsueh approach [5], there is need to put $b_0+\delta^L = -\frac{13}{2} - \frac{9}{2}b_1$, $b_0 + \delta^M = \frac{1}{2} + \frac{3}{2}b_1$, $b_0+\delta^U = 6+3b_1$ in $70b_1+9\delta_l - 3\delta_m - 6\delta_u = -103$. However, a one may confuse when doing the same there is a need to think that the equation $70b_1+9\delta_l - 3\delta_m - 6\delta_u = -103$ to be transformed into the equation $70b_1 + 9(b_0+\delta^L) - 3(b_0+\delta^M) - 6(b_0+\delta^U) = -103$ by adding and subtracting $9b_0$.

5. Proposed approach

In this section, a new approach is proposed to find the regression coefficients of the fuzzy regression model considered by Chen and Hsueh [5]. The proposed approach uses a multiplication of unrestricted real number (a regression coefficient) by a known fuzzy number (a given input value) without a need to predict the sign of the regression coefficient.

The steps of the proposed approach are as follows:

Step 1: Assuming $\tilde{X}_{ij} = \left(X_{ij}^L, X_{ij}^M, X_{ij}^U\right)$ and $\hat{\tilde{Y}}_i = \left(\hat{Y}_i^L, \hat{Y}_i^M, \hat{Y}_i^U\right)$ the fuzzy regression model, obtained in Step 1, can be transformed into the following fuzzy regression model.

$$\left(\hat{Y}_i^L, \hat{Y}_i^M, \hat{Y}_i^U\right) = (a^L, a^M, a^U) + \sum_{j=1}^{p} b_j\left(X_{ij}^L, X_{ij}^M, X_{ij}^U\right),\ i = 1, \dots, n..$$

Step 2: Using the multiplication, $(x^L, x^M, x^U) = (min(\gamma x^L, \gamma x^U), \gamma X^M, max(\gamma x^L, \gamma x^U))$, where γ is a real number, the fuzzy regression model obtained in Step 1 can be transformed into the following interval regression model.

$$\left(\hat{Y}_i^L, \hat{Y}_i^M, \hat{Y}_i^U\right) = (a^L, a^M, a^U) + \sum_{j=1}^{p}\left(min\left(b_jX_{ij}^L, b_jX_{ij}^U\right), b_jX_{ij}^M, max\left(b_jX_{ij}^L, b_jX_{ij}^U\right)\right),\ i = 1, \dots, n..$$

Step 3: Using the expression $min(ax^L, ax^U) = \frac{a(x^L+x^U)}{2} - \frac{|a|(x^U-x^L)}{2}$ and $max(ax^L, ax^U) = \frac{a(x^L+x^U)}{2} + \frac{|a|(x^U-x^L)}{2}$ the fuzzy regression model, obtained in Step 2, can be transformed into the following interval regression model.

$$\left(\hat{Y}_i^L, \hat{Y}_i^M, \hat{Y}_i^U\right) = (a^L, a^M, a^U) + \sum_{j=1}^{p}\left(\frac{b_j(X_{ij}^L + X_{ij}^U)}{2} - \frac{|b_j|(X_{ij}^U - b_jX_{ij}^L)}{2},\ b_jX_{ij}^M, \frac{b_j(X_{ij}^L + X_{ij}^U)}{2} + \frac{|b_j|(X_{ij}^U - X_{ij}^L)}{2}\right),\ i = 1, \dots, n..$$

Step 4: Using the addition $\sum_{j=1}^{p}\left(a_j^L, a_j^M, a_j^U\right) = \left(\sum_{j=1}^{p} a_j^L, \sum_{j=1}^{p} a_j^M, \sum_{j=1}^{p} a_j^U\right)$, the fuzzy regression model, obtained in the Step 3, can be transformed into the following interval regression model.

$$\left(\hat{Y}_i^L, \hat{Y}_i^M, \hat{Y}_i^U\right) = (a^L, a^M, a^U) + \left(\sum_{j=1}^{p}\left(\frac{b_j(X_{ij}^L + X_{ij}^U)}{2} - \frac{|b_j|(X_{ij}^U - b_jX_{ij}^L)}{2}\right), \sum_{j=1}^{p} b_jX_{ij}^M,\ \sum_{j=1}^{p}\left(\frac{b_j(X_{ij}^L + X_{ij}^U)}{2} + \frac{|b_j|(X_{ij}^U - X_{ij}^L)}{2}\right)\right),\ i = 1, \dots, n.$$

Step 5: Using the addition $(x^L, x^M, x^U) + (y^L, y^M, y^U) = (x^L + y^L, x^M + y^M, x^U + y^U)$,the fuzzy regression model obtained in Step 4 can be transformed into the following fuzzy regression model:

$$\left(\hat{Y}_i^L, \hat{Y}_i^M, \hat{Y}_i^U\right) = \left(a^L + \sum_{j=1}^{p}\left(\frac{b_j(X_{ij}^L + X_{ij}^U)}{2} - \frac{|b_j|(X_{ij}^U - b_jX_{ij}^L)}{2}\right), a^M + \sum_{j=1}^{p} b_jX_{ij}^M,\ a^U + \sum_{j=1}^{p}\left(\frac{b_j(X_{ij}^L + X_{ij}^U)}{2} + \frac{|b_j|(X_{ij}^U - X_{ij}^L)}{2}\right)\right),\ i = 1, \dots, n.,$$

Step 6: Using the relation $(x^L, x^M, x^U) = (y^L, y^M, y^U) \Rightarrow x^L = y^L, x^M = y^M$ and $x^U = y^U$ the fuzzy regression model, obtained in Step 5, can be transformed into the following three crisp regression models:

$$\hat{Y}_i^L = a^L + \sum_{j=1}^{p}\left(\frac{b_j(X_{ij}^L + X_{ij}^U)}{2} - \frac{|b_j|(X_{ij}^U - b_jX_{ij}^L)}{2}\right),$$

$$\hat{Y}_i^M = a^M + \sum_{j=1}^{p} b_jX_{ij}^M \text{ and } \hat{Y}_i^U = a^U + \sum_{j=1}^{p}\left(\frac{b_j(X_{ij}^L + X_{ij}^U)}{2} + \frac{|b_j|(X_{ij}^U - X_{ij}^L)}{2}\right)$$

Step 7: Find the mean squared error E_i between the estimated and the observed responses.

$$E_i = \frac{1}{3}\left[\left(Y_i^L - a^L - \sum_{j=1}^{p}\left(\frac{b_j(X_{ij}^L + X_{ij}^U)}{2} - \frac{|b_j|(X_{ij}^U - b_jX_{ij}^L)}{2}\right)\right)^2 + \left(Y_i^M - a^M - \sum_{j=1}^{p} b_jX_{ij}^M\right)^2 + \left(Y_i^U - a^U - \sum_{j=1}^{p}\left(\frac{b_j(X_{ij}^L + X_{ij}^U)}{2} + \frac{|b_j|(X_{ij}^U - X_{ij}^L)}{2}\right)\right)^2\right] i = 1,2, \dots, n.$$

Step 8: Find the total mean squared error

$$\sum_{i=1}^{n} E_i = \frac{1}{3}\sum_{i=1}^{n}\left[\left(Y_i^L - a^L - \sum_{j=1}^{p}\left(\frac{b_j(X_{ij}^L + X_{ij}^U)}{2} - \frac{|b_j|(X_{ij}^U - X_{ij}^L)}{2}\right)\right)^2 + \left(Y_i^M - a^M - \sum_{j=1}^{p} b_jX_{ij}^M\right)^2 + \left(Y_i^U - a^U - \sum_{j=1}^{p}\left(\frac{b_j(X_{ij}^L + X_{ij}^U)}{2} + \frac{|b_j|(X_{ij}^U - X_{ij}^L)}{2}\right)\right)^2\right].$$

Step 9: Find the equations by putting $\frac{\partial}{\partial a^L}(\sum_{i=1}^{n} E_i) = 0$, $\frac{\partial}{\partial a^M}(\sum_{i=1}^{n} E_i) = 0, \frac{\partial}{\partial a^U}(\sum_{i=1}^{n} E_i) = 0$ and $\frac{\partial}{\partial b_j}(\sum_{i=1}^{n} E_i) = 0$.

Step 10: Use an appropriate optimization technique/software (e.g. Maple, Mathematica etc.) to obtain the values of a^L, a^M, a^U and b_j by solving the system of equations, obtained in Step 9, with restrictions $a^L \leq a^M \leq a^U$

Step 11: Find the fuzzy regression model

$\hat{\tilde{Y}}_i = (a^L, a^M, a^U) + \sum_{j=1}^{p} b_j \tilde{X}_{ij}$, $i = 1, \ldots, n$. by putting the obtained values of a^L, a^M, a^U and b_j.

6. Illustrative example

In this section, to illustrate the proposed approach a fuzzy regression model $\hat{\tilde{Y}}_i = (a^L, a^M, a^U) + b_1 \tilde{X}_{i1}$, $i = 1,2$. is obtained by considering $\tilde{y}_1 = (-9, 2, 7)$, $\tilde{y}_2 = (-10, 0, 10)$, $\tilde{x}_{11} = (-5, -2, 7)$ and $\tilde{x}_{21} = (-3, -1, 4)$.

Step 1: Since, $X_{11}^L = -5$, $y_1^L = -9, X_{11}^M = -2, y_1^M = 2, X_{11}^U = 7, X_{21}^L = -3, y_2^L = -10, X_{21}^M = -1, y_2^M = 0, X_{21}^U = 4$ and $y_2^U = 10$. So, the crisp regression models, obtained Step 6, are transformed into the following equations.

$\hat{Y}_1^L = a^L + b_1 - 6|b_1|, \hat{Y}_2^L = a^L + \frac{1}{2} b_1 - \frac{7}{2}|b_1|,$

$\hat{Y}_1^M = a^M - 2b_1, \hat{Y}_2^M = a^M - b_1$ and $\hat{Y}_1^U = a^U + b_1 + 6|b_1|, \hat{Y}_2^U = a^U + \frac{1}{2} b_1 + \frac{7}{2}|b_1|$

Step 2: Using Step 7 of the proposed approach, the mean squared error E_1 and E_2 between the estimated and the observed responses are:

$$E_1 = \frac{1}{3}[(-9 - a^L - b_1 + 6|b_1|)^2 + (2 - a^M + 2b_1)^2 + (7 - a^U - b_1 - 6|b_1|)^2].$$

$$E_2 = \frac{1}{3}\left[\left(-10 - a^L - \frac{1}{2} b_1 + \frac{7}{2}|b_1|\right)^2 + (0 - a^M + b_1)^2 + \left(10 - a^U - \frac{1}{2} b_1 - \frac{7}{2}|b_1|\right)^2\right]$$

Step 3: Using Step 8 of the proposed approach, the total mean squared error is

$$\sum_{i=1}^{2} E_i = \frac{1}{3}\left[(-9 - a^L - b_1 + 6|b_1|)^2 + (2 - a^M + 2b_1)^2 + (7 - a^U - b_1 - 6|b_1|)^2 + \left(-10 - a^L - \frac{1}{2} b_1 + \frac{7}{2}|b_1|\right)^2 + (0 - a^M + b_1)^2 + \left(10 - a^U - \frac{1}{2} b_1 - \frac{7}{2}|b_1|\right)^2\right].$$

Step 9: Putting $\frac{\partial}{\partial a^L}\left(\sum_{i=1}^{2} E_i\right) = 0$, $\frac{\partial}{\partial a^M}\left(\sum_{i=1}^{2} E_i\right) = 0$ and $\frac{\partial}{\partial a^U}\left(\sum_{i=1}^{2} E_i\right) = 0$ and $\frac{\partial}{\partial b_1}(\sum_{i=1}^{n} E_i) = 0$ the following equations are obtained

$a^L + \frac{3}{4} b_1 - \frac{15}{4}|b_1| = -\frac{13}{2}$, $a^M - \frac{3}{2} b_1 = \frac{1}{2}$, $a^U + \frac{3}{4} b_1 + \frac{15}{4}|b_1| = 6$ and $\frac{3}{2} a^L + \frac{30}{4} b_1 - 3a^M + \frac{3}{2} a^U + \frac{|b_1|}{b_1}\left(-100 - \frac{15}{2} a^L + \frac{15}{2} a^U + \frac{125}{2}|b_1|\right) = -3.$

Step 10: On solving the system of equations, obtained in Step 9 , the obtained values of a^L, a^M, a^U and b_1 are -2 , $1, 3$ and -1 respectively.

Step 11: Putting the values of a^L, a^M, a^U and b_1 in the fuzzy regression model

$\hat{\tilde{Y}}_i = (a^L, a^M, a^U) + b_1 \tilde{X}_{i1}$, $i = 1, 2$, the obtained fuzzy regression model is

$\hat{\tilde{Y}}_i = (-2, 1, 3) - \tilde{X}_{i1}$.

However, there is no fuzzy regression model can fit the dataset given in this example using Chen and Hsueh approach [5] as its shown in Section 3.

7. Conclusion

It is shown that many computational efforts are required to apply Chen and Hsueh approach [5]. Also, it is pointed out that Chen and Hsueh approach [5] is not applicable if the correlation coefficient between the response and explanatory variable is an indeterminate value The. It is pertinent to mentioned that, if there are n explanatory variables then for applying Chen and Hsueh approach [5], there is need to evaluate n correlation coefficients due to which high computational efforts are required on applying Chen and Hsueh approach [5]. Furthermore, a simplified approach is proposed to find the fuzzy regression model considered by Chen and Hsueh [5] without a need to predict the sign of unknown regression coefficients. The illustrative example shows the advantages of the proposed simplified approach. Constructing a full fuzzy regression approach using optimal distance against outliers better than least-squared based method can be considered as future research scope.

Conflict of interest: The authors declare that they *have no known* competing financial interests or personal relationships that could have appeared to in*flu*ence the work reported in this paper.

References

1. Celmins, A. (1987). Least squares model fitting to Fuzzy vector data. *Fuzzy Sets and Systems*, 22, 245–269.
2. Celmins, A. (1987). Multidimensional least-squares fitting of Fuzzy models. *Matematical Modelling*, 9, 669–690.
3. Chukhrova, N., & Johannssen, A. (2019). Fuzzy regression analysis: Systematic review and bibliography. *Applied Soft Computing*, 84, 10578.
4. Chen, L.H., & C. C. Hsueh, C. C. (2007). A mathematical programming method for formulating a fuzzy regression model based on distance criterion. *IEEE Transaction Systems*, 37, 705–712.
5. Chen, L.H., & C. C. Hsueh, C. C. (2009). Fuzzy regression models using the least-squares method based on the concept of distance. *IEEE Trans. Fuzzy Syst*., 17, 1259-1272.
6. Diamond, P. (1987). *Least squares of fitting several fuzzy variables*, in: Proceedings of the 2nd IFSA World Congress. Tokyo, Japan, 329–331.
7. Diamond, P. (1988). Fuzzy least squares. *Information Sciences*, 46 (3), 141–157.

Al-Qudaimi, A., & Yousef, W. (2021). Fuzzy Regression Models Using the Least-Squares Method based on the Concept of Distance: Simplified Approach. *Fuzzy Optimization and Modelling Journal*, 2(3), 17-23.

https://doi.org/10.30495/fomj.2021.1933334.1031

Received: 15 June 2021 Revised: 30 August 2021 Accepted: 3 September 2021