

Regression parameters prediction in data set with outliers using neural network

Tahereh Razzaghnia* † 

Abstract

Popular regression techniques often suffer at the presence of data outliers. The different methods have proposed to make smaller the effect of the outlier on the parameter estimates. In this study, an algorithm has been addressed based on Adaptive network based fuzzy inference system to define the unknown parameters of regression model where dependent variable has outlier. So, three numerical examples are solved to test the activity of the proposed algorithm in regression model estimation. Also, the obtained results from the different methods, such as linear programming (LP) and fuzzy weights with linear programming (FWLP) are compared together. The results show that the proposed method is not to be affected the outliers in the solving process.

Keywords: Fuzzy regression, Outlier, Linear programming, Fuzzy least squares, Adaptive neural networks.

Mathematics Subject Classification (2010): 62G08, 62M45, 68M10

Received : 16.06.2017 *Accepted :* 28.06.2018 *Doi :* 10.15672/hujms.544496

1. Introduction

Zadeh [40] proposed fuzzy logic and fuzzy inference systems (FIS) for the first time in 1965 and the concept of fuzzy regression analysis was introduced by Tanaka et al [37] in 1982. Tanaka et al. [34] regarded fuzzy data as a possibility distribution and the deviations between the observed values and the estimated values were supposed to be due to the fuzziness of the system structure. In general, fuzzy regression techniques can be classified into two distinct areas: linear programming-based method that minimizes the total spread of the output, is named possibility regression (see, e.g. [27, 28, 29, 30, 32, 34, 35, 36, 37, 38]) and fuzzy least squares method that minimizes the total square error of the

*Department of Statistics, Roudehen Branch, Islamic Azad University, Roudehen - Iran.
Email: razaghnia@riau.ac.ir

†Corresponding Author.

output is called the fuzzy least square method (FLSM) (see, e.g. [1, 6, 11, 12, 15, 26, 31])). In the fuzzy literature, several extensions of this method has been proposed [20, 29, 30], occasionally in a non-parametric context [4, 13, 32, 39]. In recent years, the prediction of the regression parameters has gained a great attention among the researchers of neural networks. Fausett [14] has proposed the fundamental concepts of neural networks such as architectures, algorithms, and applications. Also, Ishibuchi et al. [17] have introduced a learning algorithm of fuzzy neural networks with triangular fuzzy weights. James and Donald [19] studied fuzzy regression using neural networks. Fuzzy neural networks have been applied for the fuzzy regression (see, e.g. [7, 8, 10, 18, 23, 25]). Jang [21] proposed the adaptive networkbased fuzzy inference system (ANFIS) in 1993 and Cheng and Lee [3] established the ANFIS model for fuzzy regression analysis using linear programming, and studied on both fuzzy adaptive networks and the switching regression model in 1999. In a study of Takagi and Sugeno [33], the method was presented for identifying a system using its input-output data. Also in 2009 and 2014, Dalkilic and Apaydin [7, 8] used the ANFIS model to analyze switching regression and estimate the fuzzy regression parameters, and in 2016, Danesh et al. [9, 10] used the ANFIS model to predict fuzzy regression model. Generally for real-world applications, data sets often contain multiple variables as well as noise or outliers that are inconsistent with the other data. Outliers may occur for a variety of reasons, such as environment changes or erroneous measurements. Different methods have been proposed for reducing the influence of outliers (see, e.g. [2, 16, 24]).

So, this paper aimed to design the adaptive network fuzzy inference system model to predict the fuzzy regression model where exist outliers. So a new algorithm is applied based on adaptive neural fuzzy inference system structure. In this study, we use fuzzy least squares method (FLSM) for consequence parameters prediction in ANFIS method (FWLS) and show that if outliers exist in the data set, the proposed method can yield good results.

2. Basic Concepts

2.1. Fuzzy regression models. Fuzzy regression methods are described based on the linear fuzzy model with symmetric triangular fuzzy coefficient [34, 37]. The aim of fuzzy regression is to minimize the fuzziness of the linear fuzzy model that includes all the given data. Thereupon, to describe fuzzy regression some definitions are needed.

A fuzzy number \tilde{A} is a convex normalized fuzzy subset of the real line \mathbb{R} with an upper semi-continuous membership function of bounded support [40].

2.1. Definition. symmetric fuzzy number is shown by $\tilde{A} = (\alpha, c)_L$;

Where α and c are the center and spread of \tilde{A} and $L(x)$ is a shape function of fuzzy numbers such that:

- i) $L(x) = L(-x)$,
- ii) $L(0) = 1, L(1) = 0$,
- iii) L is strictly decreasing on $[0, \infty)$,
- iv) L is invertible on $[0, 1]$.

The set of all symmetric fuzzy numbers is denoted by $F_L(\mathbb{R})$. If $L(x) = 1 - |x|$ then the fuzzy number is a symmetric triangular fuzzy number.

2.2. Definition. Suppose $\tilde{A} = (\alpha, c)_L$ is a symmetric fuzzy number and $\lambda \in \mathbb{R}$, then $\lambda\tilde{A} = (\lambda\alpha, |\lambda|c)_L$.

In fuzzy regression, Deviation between observed values and estimated values are assumed to be due to system fuzziness or fuzziness of regression coefficients in fuzzy regression [37]. This assumption is shared by described fuzzy regression methods in the

present study. To find a regression model, fuzzy regression is analyzed. A fuzzy regression model fits all observed fuzzy data within a specified fitting criterion. Different fuzzy regression models are obtained depending on the fitting criterion used. The first linear regression analysis with a fuzzy model was proposed by Tanaka et al. [37]. According to this method, the regression coefficients are fuzzy numbers, which can be expressed as interval numbers with membership values. For this reason, the estimated dependent variable \hat{Y} is also a fuzzy number. A fuzzy regression analysis results in the following regression model:

$$(2.1) \quad \hat{Y}_i = \tilde{A}_0 X_{i0} + \tilde{A}_1 A_{i1} + \cdots + \tilde{A}_p X_{ip} = \tilde{A} X_i, \quad i = 1, 2, \dots, n$$

Where, $\tilde{A} = (\tilde{A}_0, \tilde{A}_1, \dots, \tilde{A}_p)$ is a vector of fuzzy parameters where $\tilde{A}_j = (\alpha_j, c_j)_L$ symmetric fuzzy number, which consists of fuzzy center α_j and fuzzy half-width c_j . Also, $\tilde{Y}_i = (\bar{y}_i, e_i)$ is the observed value in this model.

According to this approach, the fuzzy coefficients \tilde{A}_i are determined such that the estimated fuzzy output \hat{Y}_i has the minimum fuzzy width, while satisfying a target degree of belief h . The term h is referred to as a measure of goodness of fit or a measure of compatibility between data and a regression model. Each of the observed data sets, which can be fuzzy \tilde{Y}_i or crisp datum Y_i , must fall within the estimated \hat{Y}_i at h level as shown in Fig (1). To determine the fuzzy coefficients $\tilde{A}_j = (\alpha_j, c_j)_L$, Tanaka et al. [37] formulated the fuzzy regression objective as the following linear programming problem.

$$(2.2) \quad \begin{aligned} & \sum_{i=1}^n \sum_{j=0}^p c_j |x_{ij}| \\ \text{s.t.} \quad & \sum_{j=0}^p \alpha_j x_{ij} + |L^{-1}(h)| \sum_{j=0}^p c_j |x_{ij}| \geq \bar{y}_i + |L^{-1}(h)| e_i, \quad i = 1, 2, \dots, n \\ & \sum_{j=0}^p \alpha_j x_{ij} + |L^{-1}(h)| \sum_{j=0}^p c_j |x_{ij}| \leq \bar{y}_i - |L^{-1}(h)| e_i, \quad i = 1, 2, \dots, n \\ & \alpha_j \in R, \quad c_j \geq 0, \quad j = 0, 1, \dots, p \end{aligned}$$

In Tanaka's model, the constraints warranty the support of the estimated values from the model (2.2) includes the support of the observed values.

2.2. outlier detection in symmetric triangular fuzzy numbers. Outliers occur when human error is involved. By using general regression models, the predicted values become too large when outliers exist in the data. In order to handle the outlier problem, Chen (2001) in [2] proposed a method for fuzzy linear regression using triangular fuzzy numbers. In this paper, the width between the spread of predicted and dependent values have to be below a certain specified value K for the outlier detection. Should this difference be larger than K , no feasible solution can be obtained. Thus, the following equation must be added to the constraints function of regression model.

$$(2.3) \quad \alpha |x_i| - e_i \leq K, \quad i = 0, 1, \dots, n$$

In Eq. (2.3), if the value of K is too small, normal values may become abnormal. On the other hand, if it is too big, abnormal values may become normal or, abnormal values will go undetected.

Hung & Yang [16] proposed an omission approach for fuzzy regression model, in which they changed the objective function of fuzzy outliers, to compensate for the flaw in Chen's method. An omission approach was applied by them to detect a single outlier in a set of

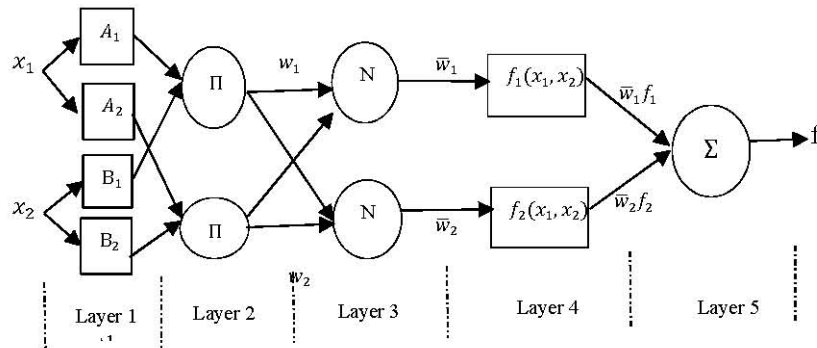


Figure 1. ANFIS architecture.

data. A ratio by objective function with a deleted observation was defined. It was also claimed that the larger the value of this ratio, the greater the impact of such observation.

Maleki et al. [24] proposed a new method in trapezoidal fuzzy data when the outlier is detected. They defined a new parameters called " H " and replaced it to the " h " in main fuzzy regression model. H is defined by spread of the observed values and replaced in the constraints of the fuzzy regression model.

2.3. Adaptive neuro fuzzy inference system (ANFIS). Takagi-Sugeno type fuzzy system [20, 21] is used by the ANFIS architecture. Precisely, one of the most popular neural fuzzy systems is this type. A fuzzy inference system is comprised of three main parts: fuzzy rules, membership functions and a reasoning mechanism. There are three types of fuzzy inference systems: the Mamdani system, in which the fuzzy output has to be defuzzified, the Takagi-Sugeno system, in which a real number as its output is produced, and the Tsukamoto system, in which monotonous functions are utilized.

The ANFIS structure is shown in Fig. 1. For simplicity, it is considered a system that has two inputs x_1 , x_2 and one output y . In Fig. 1, a circle indicates a fix node without parameters; where as a square indicates an adaptive node with parameters. A common rule set with two fuzzy if-then rules is defined as follows:

$$(2.4) \quad \text{Rule 1: IF } x_1 \text{ is } A_1 \text{ and } x_1 \text{ is } B_1 \text{ then } f_1 = p_0^1 + p_1^1 x_1 + p_2^1 x_2$$

$$(2.5) \quad \text{Rule 1: IF } x_1 \text{ is } A_2 \text{ and } x_2 \text{ is } B_2 \text{ then } f_2 = p_0^2 + p_1^2 x_1 + p_2^2 x_2$$

where x_1 , x_2 and $y \in R$ are input and output variables, respectively. A_r and A_s are fuzzy sets, μ_{A_r} , μ_{A_s} are appropriate membership function that are defined as follows:

$$\mu_{A_r}(x) = \exp \left[- \left(\frac{x_1 - \tau_r}{\sigma_r} \right)^2 \right], r = 1, 2.$$

$$\mu_{A_s}(x) = \exp \left[- \left(\frac{x_2 - \tau_s}{\sigma_s} \right)^2 \right], s = 1, 2.$$

and f_j represents system output due to rule R_j where $j = 1, 2$. The typical ANFIS consists of five layers which are explained below:

The five layers of system have one two-dimensional input and one output. In the first layer, all the nodes are adaptive. They generate membership grades of the inputs. $o_{i,j}$

is the output of the j^{th} node of the layer 1. The node function is given by:

$$(2.6) \quad o_{1,i} = \mu_{A_r}(x_1), \quad r = 1, 2$$

$$(2.7) \quad o_{1,j} = \mu_{A_s}(x_2), \quad s = 1, 2$$

In the second layer, the nodes are also fixed which multiply the inputs base on incoming output of the first layer and send the product as the output of this layer which can be calculated as:

$$(2.8) \quad o_{2,j} = w_j = \mu_{A_r}(x_1) \cdot \mu_{A_s}(x_2), \quad r, s, j = 1, 2$$

Where this layer output is the information premise section of the fuzzy if-then rule.

In the third layer, the nodes are also fixed nodes. It calculates the ratio of a rule's firing of all the rules. The output of this layer can be calculated as:

$$(2.9) \quad o_{3,j} = \bar{w}_j = \frac{w_j}{w_1 + w_2}, \quad j = 1, 2$$

In the fourth layer, the node is an adaptive one. The node function associated in the level 4 is a linear function in the ANFIS structure. The output of this layer can be represented as below:

$$(2.10) \quad o_{4,j} = \bar{w}_j f_j = \bar{w}_j (p_0^j + p_1^j x_1 + p_2^j x_2), \quad j = 1, 2$$

In this work, p_i^j will be assumed to be a triangular fuzzy number for $i = 0, 1, 2$ and $j = 1, 2$. Parameters in this layer are referred to as consequent parameters.

In the fifth layer, the single node carries out the sum of inputs of all the layers which stand for the outcomes of learning rates. The overall output of the structure is expressed as:

$$(2.11) \quad o_{5,j} = \sum_{j=1}^2 \bar{w}_j f_j = \frac{w_1}{w_1 + w_2} f_1 + \frac{w_2}{w_1 + w_2} f_2 = \bar{w}_1 f_1 + \bar{w}_2 f_2$$

Substituting the fuzzy if-then rules in Eq. (2.11) gives the following:

$$(2.12) \quad \begin{aligned} o_{5,j} = \hat{Y} &= \bar{w}_1 (p_0^1 + p_1^1 x_1 + p_2^1 x_2) + \bar{w}_2 (p_0^2 + p_1^2 x_1 + p_2^2 x_2) \\ &= \bar{w}_1 p_0^1 + (\bar{w}_1 x_1) p_1^1 + (\bar{w}_1 x_2) p_2^1 + \bar{w}_2 p_0^2 + (\bar{w}_2 x_1) p_1^2 + (\bar{w}_2 x_2) p_2^2. \end{aligned}$$

The form of this equation is the same as the following linear equation:

$$(2.13) \quad Y = p_0 + p_1 x_1 + p_2 x_2.$$

That p_i is the fuzzy parameters for $i = 0, 1, 2$.

3. Methodology of the proposed method

In Eq. (2.12), assume that consequence parameter p_i^j is a symmetric triangular fuzzy number and is represented by $p_i^j = (a_i^j, \alpha_i^j)$, $i = 0, \dots, p$, $j = 1, \dots, m$. Also, Y_k and \hat{Y}_k are symmetric triangular fuzzy numbers and are represented by $Y_k = (c^{y_k}, \beta^{y_k})$ and $\hat{Y}_k = (\hat{c}^{y_k}, \hat{\beta}^{y_k})$, $k = 1, \dots, n$ where n is the number of data points, c^{y_k} is center value and β^{y_k} is spread value of Y_k and \hat{c}^{y_k} is center value and $\hat{\beta}^{y_k}$ is spread value of \hat{Y}_k . From the above definitions, using fuzzy arithmetic and substituting p_i^j into Eq. (2.12), the output \hat{Y} , for two inputs x_1 and x_2 , can be expressed as:

$$(3.1) \quad \begin{aligned} \hat{Y} &= (a_0^1, \alpha_0^1) \bar{w}_1 + (a_1^1, \alpha_1^1) \bar{w}_1 x_1 + (a_2^1, \alpha_2^1) \bar{w}_1 x_2 + (a_0^2, \alpha_0^2) \bar{w}_2 \\ &\quad + (a_1^2, \alpha_1^2) \bar{w}_2 x_1 + (a_2^2, \alpha_2^2) \bar{w}_2 x_2 \\ &= \sum_{j=1}^2 \sum_{i=0}^2 \alpha_i^j \bar{w}_j x_i + \sum_{j=1}^2 \sum_{i=0}^2 \alpha_i^j \bar{w}_j x_i, \end{aligned}$$

where \bar{w}_j is known.

Consider the following fuzzy regression model:

$$(3.2) \quad Y_k = p_0 + p_1 x_{k1} + p_2 x_{k2} + \dots + p_p x_{kp} = p x_k, \quad k = 1, \dots, n,$$

where n is the number of data points, $x_k = (1, x_{k1}, x_{k2}, \dots, x_{kp})$ is vector of values of the independent variables at the k^{th} observations. Also, $p = (p_0, p_1, \dots, p_p)$ is vector of unknown fuzzy parameters to be estimated and Y_k is the k^{th} observed value of the dependent variables. P can be denoted in vector form as $p = \{a, \alpha\}$ where $a = (a_0^j, a_1^j, \dots, a_p^j)$ and $\alpha = (\alpha_0^j, \alpha_1^j, \dots, \alpha_p^j), j = 1, \dots, m$, where a_i^j is center value and α_i^j is spread value of $p_i, i = 0, \dots, p$. So from the above definitions, using fuzzy arithmetic and with Eq. (3.1), \hat{c}^{y_k} and $\hat{\beta}^{y_k}$ it can be expressed as:

$$(3.3) \quad \hat{c}^{y_k} = \sum_{j=1}^m \sum_{i=0}^p a_i^j \bar{w}_j x_{ki},$$

and

$$(3.4) \quad \hat{\beta}^{y_k} = \sum_{j=1}^m \sum_{i=0}^p \alpha_i^j \bar{w}_j x_{ki}.$$

So

$$(3.5) \quad \hat{Y} = \sum_{j=1}^m \sum_{i=0}^p a_i^j \bar{w}_j x_{ki} + \sum_{j=1}^m \sum_{i=0}^p \alpha_i^j \bar{w}_j x_{ki}.$$

A hybrid algorithm is used in the ANFIS method. The hybrid algorithm is composed of a forward pass and a backward pass. The least square method (forward pass) is used to optimize the consequent parameters. Once the optimal consequent parameters are found, the backward pass starts immediately. The gradient descent is used to optimize the adjustment of the premise parameters. For more details, see [20, 21]. In the following, the fuzzy least squares based on Diamond's distance is used to optimize the consequent parameters for univariate crisp input and symmetric fuzzy output.

In the fuzzy regression model (2.2), the error measurement is defined as:

$$(3.6) \quad e_k = Y_k \{-\} \hat{Y}_k,$$

where Y_k is the k^{th} output, \hat{Y}_k is the network output of the k^{th} input vector, $x_k = (1, x_{k1}, x_{k2}, \dots, x_{kp})$, and $\{-\}$ is an operator, whose definition depends on the used fuzzy ranking method. The calculation of the distance or difference between two fuzzy numbers determines the error measurement. To obtain the difference between fuzzy numbers, various fuzzy ranking methods can be used [39]. To optimize consequence parameters, this study uses the fuzzy least squares based on Diamond's distance. It can be considered that the observed values $Y_k = (l^{y_k}, c^{y_k}, r^{y_k})$ and the predicted values $\hat{Y}_k = (\hat{l}^{y_k}, \hat{c}^{y_k}, \hat{r}^{y_k})$ are asymmetric triangular fuzzy numbers for $k = 1, \dots, n$. where l^{y_k}, c^{y_k} and r^{y_k} are lower, center, and upper limits of the observed fuzzy outputs. Also, $\hat{l}^{y_k}, \hat{c}^{y_k}$ and \hat{r}^{y_k} are the estimated lower, center, and upper limits of the predicted fuzzy outputs. The fuzzy least squares problem based on Diamond's distance is defined as:

$$(3.7) \quad e_k = \sum_{i=1}^n (Y_k - \hat{Y}_k)^2 = \sum_{i=1}^n ((l^{y_k} - \hat{l}^{y_k})^2 + (c^{y_k} - \hat{c}^{y_k})^2 + (r^{y_k} - \hat{r}^{y_k})^2).$$

Suppose $Y_k = (c^{y_k}, \beta^{y_k})$ and $\hat{Y}_k = (\hat{c}^{y_k}, \hat{\beta}^{y_k})$ are two symmetric fuzzy numbers, where c^{y_k} and β^{y_k} are center, and spread of the observed fuzzy outputs. Also, \hat{c}^{y_k} and $\hat{\beta}^{y_k}$ are the estimated center, and spread of the predicted fuzzy outputs and $l^{y_k} = c^{y_k} - \beta^{y_k}$,

$r^{y_k} = c^{y_k} + \beta^{y_k}$ and $\widehat{l}^{y_k} = \widehat{c}^{y_k} - \widehat{\beta}^{y_k}$, $\widehat{r}^{y_k} = \widehat{c}^{y_k} + \widehat{\beta}^{y_k}$. By substituting l^{y_k} , r^{y_k} , \widehat{l}^{y_k} and \widehat{r}^{y_k} in Eq. (3.7), the fuzzy least squares problem can be rewritten as:

$$\begin{aligned}
 e_k &= \sum_{k=1}^n (Y_k - \widehat{Y}_k)^2 \\
 &= \sum_{k=1}^n \left((c^{y_k} - \beta^{y_k}) - (\widehat{c}^{y_k} - \widehat{\beta}^{y_k}) \right)^2 + (c^{y_k} - \widehat{c}^{y_k})^2 \\
 &\quad + \left((c^{y_k} + \beta^{y_k}) - (\widehat{c}^{y_k} + \widehat{\beta}^{y_k}) \right)^2 \\
 &= \sum_{k=1}^n \left(3(c^{y_k} - \widehat{c}^{y_k})^2 + 2(\beta^{y_k} - \widehat{\beta}^{y_k})^2 \right) \\
 (3.8) \quad &= \sum_{k=1}^n \left(3 \left(c^{y_k} - \sum_{j=1}^m \sum_{i=0}^1 \alpha_i^j \bar{w}_{kj} x_{ki} \right)^2 + 2 \left(\beta^{y_k} - \sum_{j=1}^m \sum_{i=0}^1 \alpha_i^j \bar{w}_{kj} x_{ki} \right)^2 \right)
 \end{aligned}$$

It is observed that the objective function in Eq. (3.8) is the summation of two parts with two different groups of unknown parameters. The consequent parameters can be determined by minimizing e_k respect to the unknown parameters α_i^j and α_i^j . In order to derive the error function ERROR respect to the unknown parameters, set the derivations to zero and solve for the unknown parameters. By solving these two groups of linear equations, the estimates of these parameters can be obtained for the univariate fuzzy nonparametric regression model as follows:

$$(3.9) \quad (\widehat{a}_i^j)' = (X'X)^{-1} X' C^Y,$$

$$(3.10) \quad (\widehat{\alpha}_i^j)' = (X'X)^{-1} X' \alpha^Y,$$

where,

$$X = \begin{pmatrix} \bar{w}_{11} \bar{w}_{12} \cdots \bar{w}_{1m} x_{11} \bar{w}_{12} x_{11} \cdots \bar{w}_{1m} x_{11} \\ \bar{w}_{21} \bar{w}_{22} \cdots \bar{w}_{2m} x_{21} \bar{w}_{22} x_{21} \cdots \bar{w}_{2m} x_{21} \\ \vdots \\ \bar{w}_{n1} \bar{w}_{n2} \cdots \bar{w}_{nm} x_{n1} \bar{w}_{n2} x_{n1} \cdots \bar{w}_{nm} x_{n1} \end{pmatrix},$$

$$C^Y = \begin{pmatrix} c^{y_1} \\ c^{y_2} \\ \vdots \\ c^{y_n} \end{pmatrix} \alpha^Y = \begin{pmatrix} \alpha^{y_1} \\ \alpha^{y_2} \\ \vdots \\ \alpha^{y_n} \end{pmatrix}$$

and $X_{k0} = 1$, the symbol ($'$) is the mean transpose of a matrix. Also, one of the following two constraints must be established:

$$(3.11) \quad \sum_{j=1}^m \sum_{i=0}^p \alpha_i^j \bar{w}_{kj} x_{ki} - (1 - \alpha) \sum_{j=1}^m \sum_{i=0}^p \alpha_i^j \bar{w}_{kj} x_{ki} \leq b^{y_k} + (1 - \alpha) \beta^{y_k},$$

or

$$\sum_{j=1}^m \sum_{i=0}^p \alpha_i^j \bar{w}_{kj} x_{ki} + (1 - \alpha) \sum_{j=1}^m \sum_{i=0}^p \alpha_i^j \bar{w}_{kj} x_{ki} \geq b^{y_k} - (1 - \alpha) \beta^{y_k},$$

and $\sum_{j=1}^m \sum_{i=0}^p \alpha_i^j \bar{w}_{kj} \geq 0$, $i = 0, \dots, p$, $j = 1, \dots, m$, $k = 1, \dots, n$.

According to the error of back propagation, the gradient decent method updates the premise parameters. In order to optimize the adjustment of the position and the shape of the associated membership function, the premise parameters are trained so as to

represent the density of input functions. This training does not focus on the spread of the membership function. To calculate the back propagation error, only the first part of the ERROR function, which is the center, is considered, and the influences of the spread are ignored. The back propagation error for each layer can be calculated as the method of Jang [20] and Cheng and Lee [5]. For training data, when *ERROR* is smaller than a predefined small number and one of the relationships (3.11) is established, the training of network terminates.

In this investigation, for evaluation of the accuracy of ANFIS model is defined a quantity that is called goodness of fit (GOF). It measures the bias between observed, $Y_k = (l^{y_k}, c^{y_k}, r^{y_k})$, and predicted values, $\hat{Y}_k = (\hat{l}^{y_k}, \hat{c}^{y_k}, \hat{r}^{y_k})$, for all X_k s, based on Diamond's distance (2.2), where l^{y_k} , c^{y_k} and r^{y_k} are lower, center, and upper limits of the observed fuzzy outputs. Also, \hat{l}^{y_k} , \hat{c}^{y_k} and \hat{r}^{y_k} are the estimated lower, center, and upper limits of fuzzy regression function. Error rate based on Diamond's distance can be defined as [3]:

$$(3.12) \quad \text{GOF} = \frac{1}{n} \sum_{k=1}^n \left((l^{y_k} - \hat{l}^{y_k})^2 + (c^{y_k} - \hat{c}^{y_k})^2 + (r^{y_k} - \hat{r}^{y_k})^2 \right)$$

where n is number of the observation's pairs. Large value of this quantity indicates lack-of-fit and too small value reflects over-fit for the observed fuzzy outputs. Because of the error term in model (2.2), GOF value cannot reflect the closeness between the underlying fuzzy nonparametric regression function $f(x)$ and its estimate efficiently. Thus, for measuring the bias between the underlying fuzzy regression function and its estimate, a quantity that is called BIAS is defined. This quantity can be expressed as [4]:

$$(3.13) \quad \begin{aligned} \text{BIAS} &= \frac{1}{n} \sum_{k=1}^n d^2(f(x), \hat{f}(x)) \\ &= \frac{1}{n} \sum_{k=1}^n \left((l(x) - \hat{l}(x))^2 + (c(x) - \hat{c}(x))^2 + (r(x) - \hat{r}(x))^2 \right) \end{aligned}$$

BIAS is not computable in practical applications because function $f(x)$ is certainly unknown. This quantity will be reported for examining the performance of two methods in our simulations. So in the present study, the method of Kim and Bishu [22] will be used for evaluation of the performance of ANFIS model. In this method, the absolute difference between the membership observed values and estimated values is calculated as:

$$(3.14) \quad E_k = \int_{s(Y_k) \cup s(\hat{Y}_k)} |Y_k - \hat{Y}_k| dy$$

where $s(Y_k)$ and $s(\hat{Y}_k)$ are support of Y_k and \hat{Y}_k , respectively. In other words, E_k is the error in estimation. If E_k trend to zero, then the fitting is the best.

3.1. The algorithm for forecasting model. In order to predict model parameters, the steps taken can be summarized as follows:

- Step 1: The input and the output variables are defined.
- Step 2: Value V is inputted.
- Step 3: The initial values of the fuzzy weights are determined.
- Step 4: The consequent parameters by Eq. (3.9) and (3.10) are identified.
- Step 5: When GOF in Eq. (3.12) is smaller than a predefined small number and one of the relationships (3.11) is established, the training of network terminates, otherwise the values of the fuzzy weights are updated.
- Step 6: The values of E_k , GOF for the evaluation of the method are determined.

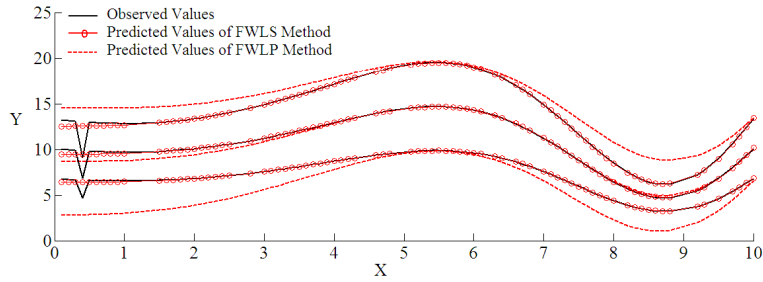


Figure 2. The observed values and the predicted values of FWLS and FWLP methods for train data in Example 4.1.

Table 1. The estimated error values E_k for Example 4.1 ($\alpha = 0.5$)

	LP method	FWLP method	FWLS method
ERROR train	434.4931	6.6832	0.3140
ERROR test	496.6499	8.8678	1.6311
e_k train	993.4188	121.8532	11.1796
e_k test	266.3852	66.6992	3.3251

In this paper, the algorithm was executed with written program in MATLAB.

4. Numerical examples

4.1. Example. Consider the following function:

$$g(x) = 10 + 5 \sin(0.25\pi(1 - x^2))$$

and let $x_k = 0.1k, k = 1, 2, \dots, 100$, on $[0, 10]$. 100 pairs of sample data are generated from $g(x)$. Let $Y_k = (b^{y_k}, \beta^{y_k})$ be a symmetric triangular fuzzy number such that

$$\begin{cases} b^{y_k} = g(x_k) + \text{rand}[-0.5, 0.5] \\ \beta^{y_k} = (1/3)g(x_k) + \text{rand}[-0.25, 0.25] \end{cases} \quad k = 1, \dots, 100.$$

For outlier generation, two variables x_k are randomly selected and $Y_k = (b^{y_k}, \beta^{y_k})$ are produced as follows:

$$\begin{cases} b^{y_k} = g(x_k) + \text{rand}[-4, 4] \\ \beta^{y_k} = (1/3)g(x_k) + \text{rand}[-0.25, 0.25] \end{cases} \quad k = 1, 2.$$

At first, we divided data set to train and test data. Then we fitted the regression model for this data set via the different methods for $\alpha = 0.5$. We displayed the error values of different methods in Table 1. In addition, the estimated values and the observed values of FWLS and FWLP methods have been depicted in Fig. 2. We use Fig. 2 and Table 1 to compare the obtained results. It is seen, model related to the FWLS method provides the best predictions. We can observe that the outlier is not influence on the estimated values in proposed method. So in outlier cases, the FWLS method is a better candid than the other methods.

4.2. Example. In this example, there are ten pair's observations as shown in Table 4 which 8th observation of the dependent variable is outlier. We divided data set to train and test data that test data has been shown with star in Tables 4 and 5. Suppose

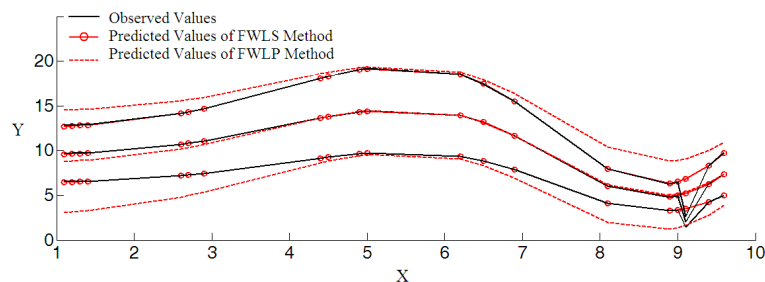


Figure 3. The observed values and the predicted values of FWLS and FWLP methods for test data in Example 4.1.

Table 2. The obtained parameters of the method FWLS for Example 4.2.

k	(τ_k, σ_k)	(a_0^k, α_0^k)	(a_1^k, α_1^k)
1	(-1.0396, 1.0741)	(-3.5495, 1.5615)	(4.3497, 0.2924)
2	(9.2937, 1.0413)	(3.3024, 2.7088)	(1.9575, -0.0095)

Table 3. The obtained parameters of the method FWLP for Example 4.2.

k	(τ_k, σ_k)	(a_0^k, α_0^k)	(a_1^k, α_1^k)
1	(-1.0396, 1.0741)	(-3.4353, 1.6465)	(4.3143, 0.3467)
2	(9.2937, 1.0413)	(2.5499, 0.0838)	(2.0935, 0.5923)

Eq. (3.5). We fitted the regression model for this data set via the different methods for $\alpha = 0.5$, and the obtained parameters for FWLS and FWLP are shown in Tables 2 and 3. Also in the LP method, the obtained regression model is shown as follows:

$$y = (1.1353, 7.2855) + (2.4562, 0.1)x$$

We displayed the obtained predictions and the errors related to these predictions in Table 4. In addition, the values of the estimated error X are shown in Table 5. Also, the estimated values and the observed values of different methods for test data have been showed in Fig. 4. Moreover, estimations have been obtained using the different methods are used for comparison. We use Fig. 4, Tables 4 and 5 to compare the obtained results. Like previous example, model related to the proposed method provides the best predictions. We can observe that the outlier is not influence on the estimated values in proposed method. So in outlier cases, the proposed method is a candid better than the other methods.

4.3. Example. In this example, there are ten pair's observations as shown in Table 8 which 7th observation of the dependent variable is outlier. Suppose Eq. (3.5). We divided data set to train and test data that test data has been shown with star in Tables 8 and 9. We applied different methods to fit fuzzy regression model and, displayed the obtained predictions and the errors related to these predictions in Table 8 and the values of the estimated error E_k are shown in Table 9. We obtained the regression model and estimations for this data set that obtained parameters of the method are shown in Tables 6 and 7. In the following, the obtained regression model using the LP method is shown:

$$y = (4.3895, 1.9348) + (5.5098, 2.4514).$$

Table 4. The estimated fuzzy outputs by using different methods for Example 4.2.

x_j	$Y_j = (b_j, \beta_j)$	$\hat{f}(x_j) = (\hat{b}_j, \hat{\beta}_j)$ LP method	$\hat{f}(x_j) = (\hat{b}_j, \hat{\beta}_j)$ FWLP method	$\hat{f}(x_j) = (\hat{b}_j, \hat{\beta}_j)$ proposed method
1	(0.8,1.8)	(3.5916,7.3856)	(0.8790,1.9932)	(0.8002,1.8539)
*2	(6.4,2.2)	(6.0478,7.4856)	(5.1932,2.3398)	(5.1499,2.1462)
3	(9.5,2.6)	(8.5040,7.5857)	(9.5075,2.6865)	(9.4996,2.4386)
4	(13.5,2.6)	(10.9602,7.6857)	(13.4495,2.9587)	(13.5002,2.7232)
5	(13.0,2.4)	(13.4164,7.7858)	(13.0208,3.0458)	(13.0933,2.6617)
6	(15.2,2.1)	(15.8726,7.8858)	(15.1108,3.6379)	(15.0473,2.6520)
*7	(17.0,2.0)	(18.3288,7.9859)	(17.2043,4.2303)	(17.0048,2.6425)
8	(19.3,4.8)	(20.7850,8.0859)	(19.2978,4.8226)	(18.9623,2.6330)
9	(20.1,1.9)	(23.2412,8.1860)	(21.3913,5.4150)	(20.9198,2.6236)
10	(23.3,2.0)	(25.6974,8.2860)	(23.4847,6.0073)	(22.8773,2.6141)
e_{train}	—————	71.1877	8.4870	1.8772
e_{test}	—————	66.6035	7.2408	2.7600

Table 5. The estimated error values E_k for Example 4.2 ($\alpha = 0.5$).

x_j	LP	FWLP	Proposed method
1	6.1324	0.2239	0.0539
*2	5.2963	2.0928	2.1406
3	5.0873	0.0871	0.1614
4	5.7269	0.3654	0.1232
5	5.4010	0.6464	0.2932
6	6.1041	2.2423	0.5893
*7	5.8187	1.5417	0.6425
8	3.7859	0.0228	2.2043
9	6.8774	3.7614	1.4910
10	6.6416	4.0116	0.0539
e_k	56.8716	14.9954	7.6994

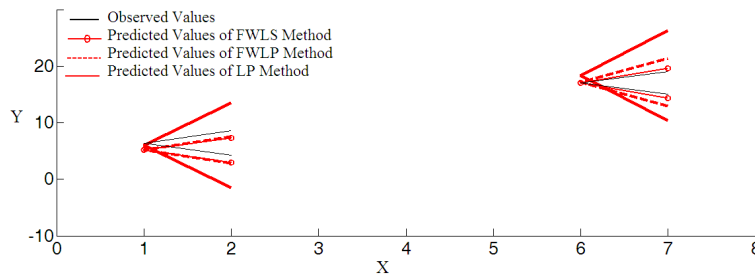


Figure 4. The observed values and the predicted values of the different methods for test data in Example 4.2.

Table 6. The obtained parameters of the method FWLS for Example 4.3.

k	(τ_k, σ_k)	(a_0^k, α_0^k)	(a_1^k, α_1^k)
1	(1.3338,0.4542)	(10.7319,2.4405)	(-0.1466,-0.3956)
2	(9.5500,3.8042)	(10.6522,0.2857)	(4.3043,1.2857)

Table 7. The obtained parameters of the method FWLP for Example 4.3.

k	(τ_k, σ_k)	(a_0^k, α_0^k)	(a_1^k, α_1^k)
1	(0.3881,1.3341)	(8.8511,1.9827)	(2.2152,0.0872)
2	(4.2786,1.0993)	(0.3885,5.3245)	(6.3649,1.4238)
3	(10.1165,0.7472)	(-6.5345,3.5800)	(6.0591,0.9354)

Table 8. The estimated fuzzy outputs using different methods for Example 4.3.

x_j	$Y_j = (b_j, \beta_j)$	$\hat{f}(x_j) = (\hat{b}_j, \hat{\beta}_j)$ LP method	$\hat{f}(x_j) = (\hat{b}_j, \hat{\beta}_j)$ FWLP method	$\hat{f}(x_j) = (\hat{b}_j, \hat{\beta}_j)$ proposed method
1	(11,2)	(9.8993,4.3862)	(11.0109,2.1300)	(11.0000,2.0000)
2	(13,2)	(15.4091,6.8376)	(13.2497,3.3297)	(13.0000,2.0000)
*3	(21,4)	(20.9188,9.2891)	(18.5886,7.9460)	(23.4956,4.1277)
4	(29,4)	(26.4286,11.7405)	(25.6386,10.7958)	(27.8696,5.4286)
5	(29,6)	(31.9384,14.1919)	(32.1746,12.4119)	(32.1739,6.7143)
*6	(34,6)	(37.4482,16.6433)	(18.5886,7.9460)	(36.4783,8.0000)
7	(45,15)	(42.9580,19.0948)	(44.9087,15.2713)	(40.7826,9.2857)
8	(44,8)	(48.4678,21.5462)	(43.3628,11.9223)	(45.0870,10.5714)
9	(48,12)	(53.9776,23.9976)	(48.0000,12.0001)	(49.3913,11.8571)
10	(54,12)	(59.4874,26.4490)	(54.0561,12.9336)	(53.6957,13.1429)
e_{train}	—————	219.3639	34.5487	22.9162
e_{test}	—————	159.1000	117.4260	22.5713

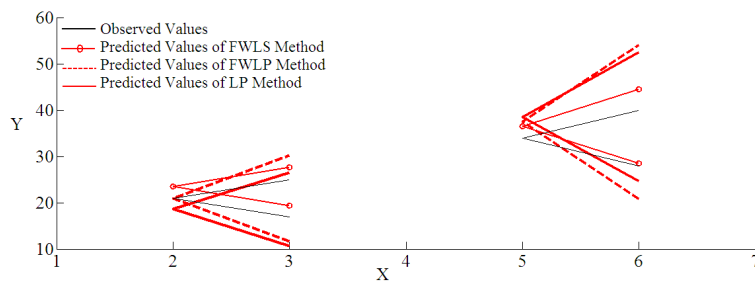


Figure 5. The observed values and the predicted values of the different methods for test data in Example 4.3.

Also, the estimated values and the observed values of different methods have depicted in Fig. 5. Previous example like, the proposed method is a candid better than the other methods in outlier cases.

Table 9. The estimated error values E_k for Example 4.3 ($\alpha = 0.5$).

x_j	LP	FWLP	Proposed method
1	2.7042	0.1309	0
2	5.3806	1.3649	0
*3	5.2899	4.9328	4.2249
4	8.1747	7.6948	2.1875
5	8.8183	7.4363	5.5555
*6	11.2353	9.4667	4.5179
7	4.9908	0.3017	8.0945
8	14.3442	4.0054	2.9673
9	13.9832	1.0000e-04	2.7015
10	15.7498	0.9368	1.2202
e_k	90.6710	36.2704	31.4693

Conclusions

In this paper, a novel combining fuzzy weights and fuzzy least square was applied for regression model prediction where dependent variable has outlier and compared the performance of the proposed algorithm with different methods, such as linear programming (LP), linear programming and fuzzy weights (FWLP). As it was seen the proposed method has increased the prediction accuracy where dependent variable has outlier. We observed that the outlier was not influence on the estimated values in proposed methods. So in outlier cases, the proposed method is a best candid than the other methods. As it can be seen in numerical examples, error related to estimations obtained via the network according to error criterion is lower than errors obtained via all the other methods.

References

- [1] A. Celmins. *Multidimensional least-squares fitting of fuzzy models*, Fuzzy sets and systems **22**, 1997.
- [2] Y. S. Chen. *Outliers detection and confidence interval modification in fuzzy regression*. Fuzzy sets and systems **119**, 259-272, 2001.
- [3] C. -B. Cheng and E. S. Lee. *Applying Fuzzy Adaptive Network to Fuzzy Regression Analysis*. Computers and Mathematics with Applications **38**, 123-140, 1999.
- [4] C. -B. Cheng and E. S. Lee. *Nonparametric fuzzy regression k-NN and kernel smoothing techniques*. Computers and Mathematics with Applications **38**, 239-251, 1999.
- [5] C. -B. Cheng and E. S. Lee. *Fuzzy regression with radial basis function networks*. Fuzzy Sets and Systems **119**, 291-301, 2001.
- [6] P. D'Urso and T. Gastaldi. *A least-squares approach to fuzzy linear regression analysis*. Computational Statistics and Data Analysis **34**, 427-440, 2000.
- [7] T. E. Dalkilic and T. Apaydin. *A fuzzy adaptive network approach to parameter estimation in cases where independent variables come from an exponential distribution*. Journal of Computational and Applied Mathematics **233**, 36-45, 2009.
- [8] T. E. Dalkilic and T. Apaydin. *Parameter Estimation by ANFIS in Cases Where Outputs are Non-Symmetric Fuzzy Numbers*. International Journal of Applied Science and Technology, 92-103, 2014.
- [9] S. Danesh. *Fuzzy Parameters Estimation via Hybrid Methods*. Hacettepe Journal of Mathematics and Statistics, inpress, Doi: 10.15672/HJMS.201614621831.
- [10] S. Danesh, R. Farnoosh and T. Razzaghnia. *Fuzzy nonparametric regression based on adaptive neuro fuzzy inference system*, Neurocomputing, **173**, 1450-1460, 2016.
- [11] P. Diamond. *Fuzzy least squares*. Information Sciences **46**, 141-157, 1988.

- [12] S. Donoso, N. Marin and M. Amparo Vila. *Quadratic programming models for fuzzy regression*. International Conference on Mathematical and Statistical Modeling in Honor of Enrique Castillo, 2006.
- [13] R. Farnoosh, J. Ghasemian and O. Solaymani fard. *A modification on ridge estimation for fuzzy nonparametric regression*. Iranian Journal of Fuzzy System Vol. 9, No. 2, 75-88 (2012)
- [14] L. V. Fausett. *Fundamentals of neural networks: architectures, algorithms, and applications*. Prentice Hall (1994)
- [15] D. H. Hong, J. -K. Song and H. Young. *Fuzzy least-squares linear regression analysis using shape preserving operations*. Information Sciences **138**, 185-193 (2001)
- [16] W. L. Hung and M. S. Yang. *An omission approach for detecting outliers in fuzzy regression models*. Fuzzy sets and systems **157**, 3109-3122 (2006)
- [17] H. Ishibuchi, K. Kwon and H. Tanaka. *A learning algorithm of fuzzy neural networks with triangular fuzzy weights*. Fuzzy Sets and Systems **71**, 277-293 (1995)
- [18] H. Ishibuchi and H. Tanaka. *Fuzzy regression analysis using neural networks*. Fuzzy Sets and Systems **50**, 257-265 ,1992.
- [19] P. James and W. Donald. *Fuzzy regression by fuzzy number neural networks*, Journal Fuzzy Sets and Systems 112 Issue **3**, 371-380 (2000)
- [20] J. S. R. Jang. *Self-learning fuzzy controllers based on temporal back-propagation*. IEEE Transactions on Neural Network **3**, 714-723 (1992)
- [21] J. S. R. Jang. *ANFIS: adaptive-network-based fuzzy inference system*. IEEE Trans Syst Man Cyber **23**(3) 665-685 (1993)
- [22] B. Kim and R. R. Bishu. *Evaluation of fuzzy linear regression models by comparing membership functions*. Fuzzy Sets and Systems **100**, 343-351 (1998)
- [23] A. Z. Lotfi. *Fuzzy logic toolbox for use with MATLAB*. Math Works **38**, 109-12 ,1995.
- [24] A. Maleki, E. Pasha, G. H. Yari and T. Razzaghnia. *Detecting outliers in fuzzy regression analysis with asymmetric trapezoidal fuzzy data*. SPIE 8349,221-226 (2012)
- [25] M. B. Menhaj. *Fundamentals of neural networks*. Tehran: Professor Hesabi Publications, ,1998.
- [26] M. Ming, M. Friedman and A. Kandel. *General fuzzy least squares*. Fuzzy Sets and Systems **88**, 107-118 ,1997.
- [27] M. Modarres, E. Nasrabadi and M. M. Nasrabadi. *A mathematical-programming approach to fuzzy linear regression analysis*. Applied Mathematics and Computation **163** (2), 977-989,2005.
- [28] A. Mottaghi, A. Ezzat and R. E. Khorram. *A New Method For Solving Fuzzy Linear Programming Problems Based on The Fuzzy Linear Complementary Problem (FLCP)*. International Journal of Fuzzy Systems, 2015.
- [29] M .M Nasrabadi and E. Nasrabadi. *A mathematical-programming approach to fuzzy linear regression analysis*. Applied Mathematics and Computation **155** (3), 873-881 ,2004.
- [30] E. Pasha, T. Razzaghnia, T. Allahviranloo, G. H. Yari and H. R. Mostafaei. *A new mathematical programming approach in fuzzy linear regression models*. Applied Mathematical Sciences **35**, 2007.
- [31] T. Razzaghnia, S. Danesh and A. Maleki. *Hybrid fuzzy regression with trapezoidal fuzzy data*. Proc. SPIE 8349,211-216,2012.
- [32] T. Razzaghnia and S. Danesh. *Nonparametric Regression with Trapezoidal Fuzzy Data*. International Journal on Recent and Innovation Trends in Computing and Communication (IJRITCC), 3826-3831 ,2015.
- [33] T. Takagi and M. Sugeo. *Fuzzy identification of system and its application to modeling and control*. IEEE Trans. Syst. Man Cybern **15**, 116-132 ,1985.
- [34] H. Tanaka, I. Hayashi and J. Watada. *Possibilistic linear regression analysis for fuzzy data*. European Journal of Operational Research **40**, 389-396 ,1989.
- [35] H. Tanaka and H. Ishibushi. *Identification of possibilistic linear systems by quadratic membership functions of fuzzy parameters*. Fuzzy Sets and Systems **41**, 145-160 ,1991.
- [36] H. Tanaka and H. Lee. *Interval regression analysis by quadratic programming approach*. IEEE Transactions on Fuzzy Systems **6**, 473-481 ,1998.
- [37] H. Tanaka, S. Uejima and K. Asia. *Linear regression analysis with fuzzy model*. IEEE Transactions on Systems, Man and Cybernetics **12**, 903-907,1982.

- [38] H. Tanaka. *Fuzzy data analysis by possibilistic linear models*. Fuzzy Sets and Systems **24**, 363-375 ,1987.
- [39] N. Wang, W. X. Zhang and C. L. Mei. *Fuzzy nonparametric regression based on local linear smoothing technique*. Information Sciences **177**, 3882-3900 ,2007.
- [40] L. A. Zadeh. *Fuzzy sets*. Information and Control **8**, 338-353 ,1965.