Iranian Journal of Fuzzy Systems
Volume 17, Number 4, (2020), pp. 105-119
University of Sistan and Baluchestan

A robust least squares fuzzy regression model based on kernel function

A. H. Khammar[1], M. Arefi[2] and M. G. Akbari[3]

[1,2,3] *Department of Statistics, Faculty of Mathematical Sciences and Statistics, University of Birjand, Birjand, Iran*

amir.khmmar@birjand.ac.ir, arefi@birjand.ac.ir, g_z_akbari@birjand.ac.ir

Abstract

In this paper, a new approach is presented to fit a robust fuzzy regression model based on some fuzzy quantities. In this approach, we first introduce a new distance between two fuzzy numbers using the kernel function, and then, based on the least squares method, the parameters of fuzzy regression model is estimated. The proposed approach has a suitable performance to present the robust fuzzy model in the presence of different types of outliers. Using some simulated data sets and some real data sets, the application of the proposed approach in modeling some characteristics with outliers, is studied.

Keywords: Distance, kernel function, least squares method, outliers, robust fuzzy regression.

1 Introduction

In fuzzy regression analysis, we fit a regression model between the response variable and the independent variables when some quantities are as imprecise. In real world, we also may encounter to the types of outliers in between the observed data. In such situations, we need to introduce a regression model which supports the imprecise quantities and also is robust under outlier data. One of the usual methods for estimating the parameters of a fuzzy regression, is called the fuzzy least squares method (see Celmins [2] and Diamond [12]). This method is a generalization of classical least squares regression method based on the definitions of the distance between fuzzy sets. It is notable that this method is very sensitive under outlier data. In such cases, when there are some outliers in the data set, it is usually preferred to use some robust approaches. There are two general approaches for obtaining the robust estimations of parameters of regression model as follows:
1) First fit a regression model and detect the outlier data, and then, modify the regression model with deleting outliers.
2) Introduce a robust regression model that is not sensitive to outliers (such as quantile regression model).

In the following, we review some works on robust estimation procedures in fuzzy regression models. Chang and Lee [7] presented fuzzy least absolute deviations regression based on the ranking of fuzzy numbers (see also Kula and Apaydin [17]). Oussalsh and Schutter [21] proposed a fuzzy regression model with the combination of least trimmed squares (LTS) and least median squares (LMS). They studied the performance of the proposed model when data is contaminated by outliers. Sanli and Apaydin [23] investigated a robust estimation procedure for fuzzy linear regression model with fuzzy input-output data based on the least median squares method. A robust approach to model a fuzzy linear regression based on M-estimators is studied by Sohn [24]. Choi and Buckley [10] utilized the least absolute deviations (LAD) method for estimating the parameters of a fuzzy regression model, and investigated the performance of the proposed model under the outliers. A robust fuzzy regression analysis using multilayered feedforward neural networks is suggested by Nasrabadi and Hashemi [20]. D'Urso and Massari [13] and D'Urso et al. [14] proposed a robust fuzzy linear regression model with crisp inputs and fuzzy outputs based on the least median squares-weighted least squares (LMS-WLS) estimation procedure. Some approach to fit the fuzzy regression models based the least absolutes method are investigated by Chachi and Taheri [5], Chachi et al. [6], Taheri and Kelkinnama [25] and Zeng et al. [27]. Based on the least trimmed squares estimation, Chachi and Roozbeh [4] proposed a estimation procedure for determining the coefficients of a fuzzy regression model with crisp input-fuzzy output data. A weighted least-squares

Corresponding Author: M. Arefi
Received: September 2019; Revised: December 2019; Accepted: February 2020.

fuzzy regression model under crisp input-fuzzy output data and fuzzy coefficients is provided by Chachi [3]. Some approaches to fit the fuzzy regression models based on the quantile loss function are investigated by Arefi [1] and Hesamian and Akbari [15].

In this study, the main problem is to provide a suitable robust estimation method of parameters of a fuzzy regression model under fuzzy data and with the presentence of the different types of outliers. In this approach, we first introduce a new distance between LR fuzzy numbers using the kernel function, and then, this distance can be used in a new objective function for obtaining robust estimations of parameters of the fuzzy regression model.

This paper is organized as follows: In Section 2, some preliminary concepts about fuzzy sets and fuzzy numbers are provided. A new distance between LR fuzzy numbers based on kernel function is presented in Section 3. In Section 4, we present a new approach to fit a robust fuzzy regression model when the response variable and the independent variables are as fuzzy numbers. Section 5 exhibits some simulated and numerical examples to assess the effectiveness and robustness of the proposed method under outliers. In Section 6, we compare our method with some other works on fuzzy regression theory. Finally, in Section 7, some concluding remarks are provided.

2 Preliminary concepts

In this section, we recall some notations and preliminary concepts about fuzzy sets (for more details, see Zimmermann [28]).

Let X be an universal set. A fuzzy set $\tilde{N}$ of X is defined by its membership function $\tilde{N}: X \to [0,1]$. The α-cut of $\tilde{N}$ is defined by the set $\tilde{N}[\alpha] = \{x \in R : \tilde{N}(x) \geq \alpha\}$, $0 < \alpha \leq 1$.

Definition 2.1. *A fuzzy set $\tilde{N}$ on the universal set X is called a fuzzy number, if*
i) $\tilde{N}(x) = 1$ for some $x \in X$,
ii) $\tilde{N}[\alpha]$ is a closed bounded interval for $0 < \alpha \leq 1$.

Definition 2.2. *A fuzzy number $\tilde{N}$ is called a LR fuzzy number, if there are real numbers m, l and r with $l, r \geq 0$, and the strictly decreasing functions $L, R : R^+ \to [0,1]$ such that*

$$\tilde{N}(x) = \begin{cases} L\left(\frac{m-x}{l}\right), & x \leq m, \\ R\left(\frac{x-m}{r}\right), & x > m. \end{cases}$$

It is denoted by $\tilde{N} = (m, l, r)_{LR}$.

In a special case, if $L(x) = R(x)$ for all $x \in [0,1]$, then $\tilde{N}$ is called the LL fuzzy number. On the other hand, if $L(x) = R(x) = 1 - x$ for all $x \in [0,1]$, then $\tilde{N}$ is called the triangular fuzzy number and is denoted by $(m, l, r)_T$. Also, for $l = r$, $\tilde{N}$ is a symmetric triangular fuzzy number as $\tilde{N} = (m, l)_T$.

Based on Extension Principle, the arithmetic operations on LR fuzzy numbers are defined as follows (see Zimmermann [28]).

Proposition 2.3. *Let $\tilde{A} = (m_a, l_a, r_a)_{LR}$ and $\tilde{B} = (m_b, l_b, r_b)_{LR}$ be two LR fuzzy numbers and $\lambda \in R - \{0\}$, then*

$$\lambda \otimes \tilde{A} = \begin{cases} (\lambda m_a, \lambda l_a, \lambda r_a)_{LR}, & \lambda > 0, \\ (\lambda m_a, -\lambda l_a, -\lambda r_a)_{RL}, & \lambda < 0, \end{cases} \quad \text{and} \quad \tilde{A} \oplus \tilde{B} = (m_a + m_b, l_a + l_b, r_a + r_b)_{LR}.$$

Remark 2.4. *Since the LR fuzzy numbers are the favorite of many authors in various articles, and the arithmetic operations between these numbers are well-defined, we use these kinds of fuzzy numbers in this paper. On the other hand, they are also a natural extension of precise numbers.*

3 Distance between LR fuzzy numbers based on kernel function

Let $Z = \{z_1, z_2, ..., z_n\}$ be a non-empty set with $z_i \in \Re^d$. A function $K : Z \times Z \to \Re$ is called a positive definite (or Mercer) kernel [18] if and only if K is symmetric $K(z_i, z_k) = K(z_k, z_i)$ and $\sum_{i=1}^{n} \sum_{k=1}^{n} c_i c_k K(z_i, z_k) \geq 0$, $n \geq 2$, $c_r \in \Re$ for $r = 1, 2, ..., n$.

Let $\phi : Z \to F$ be a non-linear mapping from the input space Z to a high dimensional feature space F. By applying the mapping ϕ, the dot product $z_i^T z_k$ in the input space is mapped to $\phi(z_i)^T \phi(z_k)$ in the feature space. It is interesting that each Mercer kernel can be expressed as $K(z_i, z_k) = \phi(z_i)^T \phi(z_k)$ while the non-linear mapping ϕ does not need to be explicitly specified (see Mueller et al. [19]).

One of the most relevant aspects in applications is that it is possible to compute Euclidean distances in F without knowing explicitly ϕ. This can be done using the so called distance kernel trick as

$$\begin{aligned}\| \phi(z_i) - \phi(z_k) \|^2 &= (\phi(z_i) - \phi(z_k))^T(\phi(z_i) - \phi(z_k)) \\ &= \phi(z_i)^T\phi(z_i) - 2\phi(z_i)^T\phi(z_k) + \phi(z_k)^T\phi(z_k) \\ &= K(z_i, z_i) - 2K(z_i, z_k) + K(z_k, z_k).\end{aligned}$$

Using the distance kernel trick, we can define a new distance between two LR fuzzy numbers as follows.

Definition 3.1. *Suppose that $\tilde{A}$ and $\tilde{B}$ are two LR fuzzy numbers. Distance between $\tilde{A}$ and $\tilde{B}$ based on kernel function (kernel distance) is defined as*

$$\begin{aligned}D_K(\tilde{A},\tilde{B}) &= \Big[\frac{1}{3}[\| \phi(m_a) - \phi(m_b) \|^2 + \| \phi(m_a - \lambda l_a) - \phi(m_b - \lambda l_b) \|^2 + \| \phi(m_a + \rho r_a) - \phi(m_b + \rho r_b) \|^2]\Big]^{1/2} \\ &= \Big[\frac{1}{3}[K(m_a, m_a) - 2K(m_a, m_b) + K(m_b, m_b) \\ &\quad + K(m_a - \lambda l_a, m_a - \lambda l_a) - 2K(m_a - \lambda l_a, m_b - \lambda l_b) + K(m_b - \lambda l_b, m_b - \lambda l_b) \\ &\quad + K(m_a + \rho r_a, m_a + \rho r_a) - 2K(m_a + \rho r_a, m_b + \rho r_b) + K(m_b + \rho r_b, m_b + \rho r_b)]\Big]^{1/2}, \end{aligned} \tag{1}$$

where $\lambda = \int_0^1 L^{-1}(w)dw$ and $\rho = \int_0^1 R^{-1}(w)dw$.

Remark 3.2. *The kernel distance in Definition 3.1 is the average of errors that is computed not in the original space but in a high dimensional space F through a non-linear mapping ϕ applied to the centers and the left and right spreads of two fuzzy numbers $\tilde{A}$ and $\tilde{B}$. In order to implement the kernel distance, we have to determine the kernel function $K(.,.)$. Some common kernel functions are given as follows:*

1) *Gaussian kernel:* $K(z_i, z_k) = \exp\{-\frac{(z_i - z_k)^2}{2h^2}\}$,
2) *Triweight kernel:* $K(z_i, z_k) = \frac{35}{32}(1 - |z_i - z_k|^2)^3$ *for* $|z_i - z_k| \leq 1$,
3) *Epanecknikiv kernel:* $K(z_i, z_k) = \frac{3}{4}(1 - (z_i - z_k)^2)$ *for* $|z_i - z_k| \leq 1$,

where h is the smoothing (bandwidth) parameter. Because to some mathematical facilities provided by the Gaussian kernel function, we consider it to fit a suitable fuzzy regression model in this paper.

Using the Gaussian kernel function, the kernel distance given in Definition 3.1 can be rewritten as follows

$$D_K(\tilde{A},\tilde{B}) = \Big[2 - \frac{2}{3}[K(m_a, m_b) + K(m_a - \lambda l_a, m_b - \lambda l_b) + K(m_a + \rho r_a, m_b + \rho r_b)]\Big]^{1/2}. \tag{2}$$

Proposition 3.3. *Assume that $\tilde{A}$, $\tilde{B}$, and $\tilde{C}$ are three LR fuzzy numbers. Then, the kernel distance D_K satisfies the following items:*
1) $D_K(\tilde{A},\tilde{B}) \geq 0$,
2) *If* $\tilde{A} = \tilde{B}$ *then,* $D_K(\tilde{A},\tilde{B}) = 0$
3) $D_K(\tilde{A},\tilde{B}) = D_K(\tilde{B},\tilde{A})$,
4) $D_K(\tilde{A},\tilde{C}) \leq D_K(\tilde{A},\tilde{B}) + D_K(\tilde{B},\tilde{C})$.

Proof. Items 1, 2 and 3 are obviously hold. We prove item 4. Using the definition of kernel distance, we have

$$\begin{aligned}D_K(\tilde{A},\tilde{C}) &= \Big[\frac{1}{3}[\| \phi(m_a) - \phi(m_c) \|^2 + \| \phi(m_a - \lambda l_a) - \phi(m_c - \lambda l_c) \|^2 + \| \phi(m_a + \rho r_a) - \phi(m_c + \rho r_c) \|^2]\Big]^{1/2} \\ &= \Big[\frac{1}{3}[\| \phi(m_a) - \phi(m_b) + \phi(m_b) - \phi(m_c) \|^2 \\ &\quad + \| \phi(m_a - \lambda l_a) - \phi(m_b - \lambda l_b) + \phi(m_b - \lambda l_b) - \phi(m_c - \lambda l_c) \|^2 \\ &\quad + \| \phi(m_a + \rho r_a) - \phi(m_b + \rho r_b) + \phi(m_b + \rho r_b) - \phi(m_c + \rho r_c) \|^2]\Big]^{1/2} \\ &\leq \Big[\frac{1}{3}[\| \phi(m_a) - \phi(m_b) \|^2 + \| \phi(m_a - \lambda l_a) - \phi(m_b - \lambda l_b) \|^2 + \| \phi(m_a + \rho r_a) - \phi(m_b + \rho r_b) \|^2]\Big]^{1/2} \\ &\quad + \Big[\frac{1}{3}[\| \phi(m_b) - \phi(m_c) \|^2 + \| \phi(m_b - \lambda l_b) - \phi(m_c - \lambda l_c) \|^2 + \| \phi(m_b + \rho r_b) - \phi(m_c + \rho r_c) \|^2]\Big]^{1/2} \\ &= D_K(\tilde{A},\tilde{B}) + D_K(\tilde{B},\tilde{C}).\end{aligned}$$

We use Minkowski inequality to obtain the first inequality. Hence, the proof is completed. □

Example 3.4. *Consider the triangular fuzzy numbers as follows.*

$$\begin{array}{ll}\tilde{A}_1=(0.5,0.3,0.4)_T, & \tilde{B}_1=(0.6,0.5,0.2)_T,\\ \tilde{A}_2=(0,1,1)_T, & \tilde{B}_2=(1,4,3)_T,\\ \tilde{A}_3=(3,3,3)_T, & \tilde{B}_3=(5,2,2)_T,\\ \tilde{A}_4=(7,3,2)_T, & \tilde{B}_4=(7,2,3)_T.\end{array}$$

By kernel distance (2) for $h=1$, we have

$$D_K(\tilde{A}_1,\tilde{B}_1)=0.058,\quad D_K(\tilde{A}_2,\tilde{B}_2)=0.958,\quad D_K(\tilde{A}_3,\tilde{B}_3)=1.290,\quad D_K(\tilde{A}_4,\tilde{B}_4)=0.396.$$

4 Methodology of robust fuzzy regression model

In this section, we introduce a least squares regression model under the kernel distance based on the crisp input and fuzzy output data. In the proposed model, the parameters of model are also considered as fuzzy quantities.

4.1 Fuzzy regression model

Assume that we have a set of observed data $(x_{i1},x_{i2},...,x_{ip},\tilde{Y}_i)$, $i=,2,...,n$, where $\tilde{Y}_i=(y_i,s_{l_i},s_{r_i})_{LR}$ is as LR fuzzy numbers. The aim in this paper is to fit a fuzzy regression model with fuzzy coefficients to data set as follows:

$$\hat{\tilde{Y}}=\tilde{\beta}_0\oplus(\tilde{\beta}_1\otimes x_1)\oplus...\oplus(\tilde{\beta}_p\otimes x_p),$$

where $\tilde{\beta_j}=(\beta_j,\gamma_j,\theta_j)_{LR}$, $j=0,1,...,p$ are assumed to be LR fuzzy numbers. The values of estimated responses using Proposition 2.3 are obtained as

$$\begin{aligned}\hat{\tilde{Y}}_i &= \tilde{\beta}_0\oplus(\tilde{\beta}_1\otimes x_{i1})\oplus(\tilde{\beta}_1\otimes x_{i2})\oplus...\oplus(\tilde{\beta}_p\otimes x_{ip})\\ &= \begin{cases}(\sum_{j=0}^{p}\beta_jx_{ij},\sum_{j=0}^{p}\gamma_jx_{ij},\sum_{j=0}^{p}\theta_jx_{ij})_{LR}, & \text{if}\ \ x_{ij}\geq 0,\quad i=1,2,...,n,\\ (\sum_{j=0}^{p}\beta_jx_{ij},-\sum_{j=0}^{p}\theta_jx_{ij},-\sum_{j=0}^{p}\gamma_jx_{ij})_{RL}, & \text{if}\ \ x_{ij}<0,\quad i=1,2,...,n,\end{cases}\end{aligned}$$

where $x_{i0}=0$. In this study, we assume that the observations of input variables are positive $x_{ij}\geq 0$.

4.2 Estimations of parameters of model

In this section, based on the kernel distance introduced in section 3, we fit a least squares fuzzy regression model. Using the kernel distance (2), the sum of squared distances between $\tilde{Y_i}$ and $\hat{\tilde{Y_i}}$ is obtained as follows:

$$S=\sum_{i=1}^{n}D_K^2(\tilde{Y}_i,\hat{\tilde{Y}}_i)=2n-\frac{2}{3}\sum_{i=1}^{n}\big(K(y_i,\sum_{j=0}^{p}\beta_jx_{ij})+K(y_i-\lambda s_{l_i},\sum_{j=0}^{p}\beta_jx_{ij}-\lambda\sum_{j=0}^{p}\gamma_jx_{ij})+K(y_i+\rho s_{r_i},\sum_{j=0}^{p}\beta_jx_{ij}+\rho\sum_{j=0}^{p}\theta_jx_{ij})\big). \tag{3}$$

So, we have

$$\begin{aligned}\frac{\partial S}{\partial\beta_j} &= -\frac{2}{3h^2}\sum_{i=1}^{n}x_{ij}\Big[\big(y_i-\sum_{j=0}^{p}\beta_jx_{ij}\big)K(y_i,\sum_{j=0}^{p}\beta_jx_{ij})\\ &\quad+\big(y_i-\lambda s_{l_i}-\sum_{j=0}^{p}\beta_jx_{ij}+\lambda\sum_{j=0}^{p}\gamma_jx_{ij}\big)K(y_i-\lambda s_{l_i},\sum_{j=0}^{p}\beta_jx_{ij}-\lambda\sum_{j=0}^{p}\gamma_jx_{ij})\\ &\quad+\big(y_i+\rho s_{r_i}-\sum_{j=0}^{p}\beta_jx_{ij}-\rho\sum_{j=0}^{p}\theta_jx_{ij}\big)K(y_i+\rho s_{r_i},\sum_{j=0}^{p}\beta_jx_{ij}+\rho\sum_{j=0}^{p}\theta_jx_{ij})\Big],\\ \frac{\partial S}{\partial\gamma_j} &= \frac{2\lambda}{3h^2}\sum_{i=1}^{n}x_{ij}\big(y_i-\lambda s_{l_i}-\sum_{j=0}^{p}\beta_jx_{ij}+\lambda\sum_{j=0}^{p}\gamma_jx_{ij}\big)K(y_i-\lambda s_{l_i},\sum_{j=0}^{p}\beta_jx_{ij}-\lambda\sum_{j=0}^{p}\gamma_jx_{ij}),\\ \frac{\partial S}{\partial\theta_j} &= -\frac{2\rho}{3h^2}\sum_{i=1}^{n}x_{ij}\big(y_i+\rho s_{r_i}-\sum_{j=0}^{p}\beta_jx_{ij}-\rho\sum_{j=0}^{p}\theta_jx_{ij}\big)K(y_i+\rho s_{r_i},\sum_{j=0}^{p}\beta_jx_{ij}+\rho\sum_{j=0}^{p}\theta_jx_{ij}),\end{aligned}$$

$$\frac{\partial^2 S}{\partial\beta_j\partial\beta_k} = -\frac{2}{3h^2}\sum_{i=1}^{n} x_{ij}x_{ik}\Big[K(y_i,\sum_{j=0}^{p}\beta_j x_{ij})\Big(\Big(\frac{y_i-\sum_{j=0}^{p}\beta_j x_{ij}}{h}\Big)^2-1\Big)$$
$$+K(y_i-\lambda s_{l_i},\sum_{j=0}^{p}\beta_j x_{ij}-\lambda\sum_{j=0}^{p}\gamma_j x_{ij})\Big(\Big(\frac{y_i-\lambda s_{l_i}-\sum_{j=0}^{p}\beta_j x_{ij}+\lambda\sum_{j=0}^{p}\gamma_j x_{ij}}{h}\Big)^2-1\Big)$$
$$+K(y_i+\rho s_{r_i},\sum_{j=0}^{p}\beta_j x_{ij}+\rho\sum_{j=0}^{p}\theta_j x_{ij})\Big(\Big(\frac{y_i+\rho s_{r_i}-\sum_{j=0}^{p}\beta_j x_{ij}-\rho\sum_{j=0}^{p}\theta_j x_{ij}}{h}\Big)^2-1\Big)\Big],$$
$$\frac{\partial^2 S}{\partial\beta_j\partial\gamma_k} = -\frac{2}{3h^2}\sum_{i=1}^{n} x_{ij}x_{ik}K(y_i-\lambda s_{l_i},\sum_{j=0}^{p}\beta_j x_{ij}-\lambda\sum_{j=0}^{p}\gamma_j x_{ij})\Big(1-\Big(\frac{y_i-\lambda s_{l_i}-\sum_{j=0}^{p}\beta_j x_{ij}+\lambda\sum_{j=0}^{p}\gamma_j x_{ij}}{h}\Big)^2\Big),$$
$$\frac{\partial^2 S}{\partial\beta_j\partial\theta_k} = -\frac{2}{3h^2}\sum_{i=1}^{n} x_{ij}x_{ik}K(y_i+\rho s_{r_i},\sum_{j=0}^{p}\beta_j x_{ij}+\rho\sum_{j=0}^{p}\theta_j x_{ij})\Big(\Big(\frac{y_i+\rho s_{r_i}-\sum_{j=0}^{p}\beta_j x_{ij}-\rho\sum_{j=0}^{p}\theta_j x_{ij}}{h}\Big)^2-1\Big),$$
$$\frac{\partial^2 S}{\partial\gamma_j\partial\beta_k} = \frac{2\lambda}{3h^2}\sum_{i=1}^{n} x_{ij}x_{ik}K(y_i-\lambda s_{l_i},\sum_{j=0}^{p}\beta_j x_{ij}-\lambda\sum_{j=0}^{p}\gamma_j x_{ij})\Big(\Big(\frac{y_i-\lambda s_{l_i}-\sum_{j=0}^{p}\beta_j x_{ij}+\lambda\sum_{j=0}^{p}\gamma_j x_{ij}}{h}\Big)^2-1\Big),$$
$$\frac{\partial^2 S}{\partial\gamma_j\partial\gamma_k} = \frac{2\lambda}{3h^2}\sum_{i=1}^{n} x_{ij}x_{ik}K(y_i-\lambda s_{l_i},\sum_{j=0}^{p}\beta_j x_{ij}-\lambda\sum_{j=0}^{p}\gamma_j x_{ij})\Big(1-\Big(\frac{y_i-\lambda s_{l_i}-\sum_{j=0}^{p}\beta_j x_{ij}+\lambda\sum_{j=0}^{p}\gamma_j x_{ij}}{h}\Big)^2\Big),$$
$$\frac{\partial^2 S}{\partial\gamma_j\partial\theta_k} = 0,$$
$$\frac{\partial^2 S}{\partial\theta_j\partial\beta_k} = -\frac{2\rho}{3h^2}\sum_{i=1}^{n} x_{ij}x_{ik}K(y_i+\rho s_{r_i},\sum_{j=0}^{p}\beta_j x_{ij}+\rho\sum_{j=0}^{p}\theta_j x_{ij})\Big(\Big(\frac{y_i+\rho s_{r_i}-\sum_{j=0}^{p}\beta_j x_{ij}-\rho\sum_{j=0}^{p}\theta_j x_{ij}}{h}\Big)^2-1\Big),$$
$$\frac{\partial^2 S}{\partial\theta_j\partial\gamma_k} = 0,$$
$$\frac{\partial^2 S}{\partial\theta_j\partial\theta_k} = -\frac{2\rho}{3h^2}\sum_{i=1}^{n} x_{ij}x_{ik}K(y_i+\rho s_{r_i},\sum_{j=0}^{p}\beta_j x_{ij}+\rho\sum_{j=0}^{p}\theta_j x_{ij})\Big(\Big(\frac{y_i+\rho s_{r_i}-\sum_{j=0}^{p}\beta_j x_{ij}-\rho\sum_{j=0}^{p}\theta_j x_{ij}}{h}\Big)^2-1\Big).$$

Considering $\frac{\partial S}{\partial\beta_j}=0$, $\frac{\partial S}{\partial\gamma_j}=0$, and $\frac{\partial S}{\partial\theta_j}=0$ for $j=0,1,...,p$ and using the following algorithm, parameters are estimated. In this algorithm, we use of the Newton-Raphson method for solving nonlinear equations.

Algorithm: Estimation of parameters of model.

Step 1: Assume that $\mathbf{P}=\left[\beta_0,...,\beta_p,\gamma_0,...,\gamma_p,\theta_0,...,\theta_p\right]^T$, $\frac{\partial S}{\partial\beta_j}=f1_j^\beta$, $\frac{\partial S}{\partial\gamma_j}=f2_j^\gamma$, $\frac{\partial S}{\partial\theta_j}=f3_j^\theta$, $\frac{\partial^2 S}{\partial\beta_j\partial\beta_k}=g1_{jk}^\beta$, $\frac{\partial^2 S}{\partial\beta_j\partial\gamma_k}=g2_{jk}^\beta$, $\frac{\partial^2 S}{\partial\beta_j\partial\theta_k}=g3_{jk}^\beta$, $\frac{\partial^2 S}{\partial\gamma_j\partial\beta_k}=g1_{jk}^\gamma$, $\frac{\partial^2 S}{\partial\gamma_j\partial\gamma_k}=g2_{jk}^\gamma$, $\frac{\partial^2 S}{\partial\gamma_j\partial\theta_k}=g3_{jk}^\gamma$, $\frac{\partial^2 S}{\partial\theta_j\partial\beta_k}=g1_{jk}^\theta$, $\frac{\partial^2 S}{\partial\theta_j\partial\gamma_k}=g2_{jk}^\theta$, $\frac{\partial^2 S}{\partial\theta_j\partial\theta_k}=g3_{jk}^\theta$, $j=0,1,...,p$, $k=0,1,...,p$. Also, we define $\mathbf{F_1}=\left[f1_0^\beta,...,f1_p^\beta,f2_0^\gamma,...,f2_p^\gamma,f3_0^\theta,...,f3_p^\theta\right]^T$ and

$$\mathbf{F_2}=\begin{bmatrix} g1_{00}^\beta & \cdots & g1_{0p}^\beta & g1_{00}^\gamma & \cdots & g1_{0p}^\gamma & g1_{00}^\theta & \cdots & g1_{0p}^\theta \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ g1_{p0}^\beta & \cdots & g1_{pp}^\beta & g1_{p0}^\gamma & \cdots & g1_{pp}^\gamma & g1_{p0}^\theta & \cdots & g1_{pp}^\theta \\ g2_{00}^\beta & \cdots & g2_{0p}^\beta & g2_{00}^\gamma & \cdots & g2_{0p}^\gamma & g2_{00}^\theta & \cdots & g2_{0p}^\theta \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ g2_{p0}^\beta & \cdots & g2_{pp}^\beta & g2_{p0}^\gamma & \cdots & g2_{pp}^\gamma & g2_{p0}^\theta & \cdots & g2_{pp}^\theta \\ g3_{00}^\beta & \cdots & g3_{0p}^\beta & g3_{00}^\gamma & \cdots & g3_{0p}^\gamma & g3_{00}^\theta & \cdots & g3_{0p}^\theta \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ g3_{p0}^\beta & \cdots & g3_{pp}^\beta & g3_{p0}^\gamma & \cdots & g3_{pp}^\gamma & g3_{p0}^\theta & \cdots & g3_{pp}^\theta \end{bmatrix}.$$

Step 2: Given a tolerance limit ϵ, a maximum number of iterations T and bandwidth h.

Step 3: Set $t=1$.

Step 4: Compute the starting value of $\mathbf{P}$ from the LS solution. Let the starting value be denoted by $\mathbf{P}^{(1)}$.

Step 5: Compute the starting value $\mathbf{F1}^{(1)}$ and $\mathbf{F2}^{(1)}$ based on the values of step 4.

Step 6**:** Compute $S^{(1)}$ by using (3).
Step 7**:** Set $t = t + 1$.
Step 8**:** Compute $\mathbf{P}^{(t)} = \mathbf{P}^{(t-1)} - \left(\mathbf{F2}^{(t-1)}\right)^{-1}\mathbf{F1}^{(t-1)}$.
Step 9**:** Compute $S^{(t)}$ by using (3).
Step 10**:** Until $\mid S^{(t)} - S^{(t-1)} \mid \leq \epsilon$ or $t \geq T$, repeat steps $7 - 9$.

Remark 4.1. *Note that for solving optimization problems, we also can use the command "NMinimize" in "Mathematica software"* [26].

Remark 4.2. *The bandwidth h in the kernel function $K(.,.)$ is a smoothing parameter. To select the optimal value of bandwidth h (h_{opt}), we use a heuristic tool. We plot the minimum values attained by the objective function* (3) *for the different values of h. So we can see an approximate monotonicity in the chart, after the value of point m ($m > 0$). Therefore, we suggest the value of m as a possible sensible choice for h, ($h_{opt} = m$).*

Remark 4.3. *The proposed method in this paper is an extended version of classical method introduced by Carvalho et al.* [11]. *In Carvalho et al.'s method, the kernel distance between two crisp numbers x and y is as $D_K(x, y) = [2(1 - K(x, y))]^{1/2}$. Also, the coefficients of robust crisp regression model are obtained based on the following objective function $S = \sum_{i=1}^{n} 2[1 - K(y_i, \sum_{j=0}^{p} \beta_j x_{ij})]$.*

4.3 Goodness of fit of the model

In order to evaluate goodness of fit of the proposed fuzzy regression model, we introduce two indices as follows.

Definition 4.4. *(Chen and Dang* [8]*) Suppose that $\tilde{Y}_i$ and $\hat{\tilde{Y}}_i$ are the values of the observed and estimated fuzzy responses. Based on the related error measure $E(\hat{\tilde{Y}}_i, \tilde{Y}_i)$, a measure for goodness of fit of the model is defined as $G = \frac{1}{n}\sum_{i=1}^{n} S_{CD}(\hat{\tilde{Y}}_i, \tilde{Y}_i)$, where $S_{CD}(\hat{\tilde{Y}}_i, \tilde{Y}_i) = \frac{1}{1+E(\hat{\tilde{Y}}_i, \tilde{Y}_i)}$ and $E(\hat{\tilde{Y}}_i, \tilde{Y}_i) = \frac{\int_{S_{\tilde{Y}_i} \cup S_{\hat{\tilde{Y}}_i}} |\hat{\tilde{Y}}_i(y) - \tilde{Y}_i(y)| dy}{\int_{S_{\tilde{Y}_i}} \tilde{Y}_i(y) dy}$. Also, $\hat{\tilde{Y}}_i(y)$ and $\tilde{Y}_i(y)$ with the supports $S_{\hat{\tilde{Y}}_i}$ and $S_{\tilde{Y}_i}$ are the membership functions of $\hat{\tilde{Y}}_i$ and $\tilde{Y}_i$, respectively.*

Definition 4.5. *(Pappis and Karacapilidis* [22]*) Suppose that $\tilde{Y}_i$ and $\hat{\tilde{Y}}_i$ are the values of the observed and estimated fuzzy responses. The mean of similarity measures (MSM) is defined as $MSM = \frac{1}{n}\sum_{i=1}^{n} S_{PK}(\hat{\tilde{Y}}_i, \tilde{Y}_i)$, where $S_{PK}(\hat{\tilde{Y}}_i, \tilde{Y}_i) = \frac{Card(\hat{\tilde{Y}}_i \cap \tilde{Y}_i)}{Card(\hat{\tilde{Y}}_i \cup \tilde{Y}_i)}$, and $Card(\tilde{A})$ denotes the cardinal number of $\tilde{A}$ defined as $Card(\tilde{A}) = \int_R \tilde{A}(x)dx$.*

Remark 4.6. *The indices G and MSM are in the interval* $[0, 1]$. *The optimal model is the model with maximum value of G or MSM.*

4.4 Outlier

Among data set, there are some points with the small value of $S_{CD}(\hat{\tilde{Y}}_i, \tilde{Y}_i)$ or $S_{PK}(\hat{\tilde{Y}}_i, \tilde{Y}_i)$ for optimal selected model. These points can be regarded as possible outliers. Since the proposed fuzzy regression model is suggested to be robust under the outliers, we introduce different types of outliers as follows (see D'Urso and Massari [13]):
1) Outlier is in the independent variables (**X-axis**),
2) Outlier is in the centers of the dependent variables (**Y-axis**),
3) Outlier is in the spreads of the dependent variables (**S-axis**),
4) Outlier is as a combination of the items 2 and 3 (**YS-axis**),
5) Outlier is as a combination of the items 1 and 2 (**XY-axis**).
Figure 1 shows the different types of outliers in a sample of observations. For instance, in Figure 1 (**X-axis**), there are two outliers with respect to X.

5 Examples

In this section, we present four examples. In Example 5.1, we simulate some data set from the fuzzy regression models, and we evaluate the performance of the proposed method. In Examples 5.2 and 5.3, we present two data sets and compare our method with some other methods which have used these data sets. In Example 5.4, we present a real data set and we fit a robust fuzzy regression model to this data set.

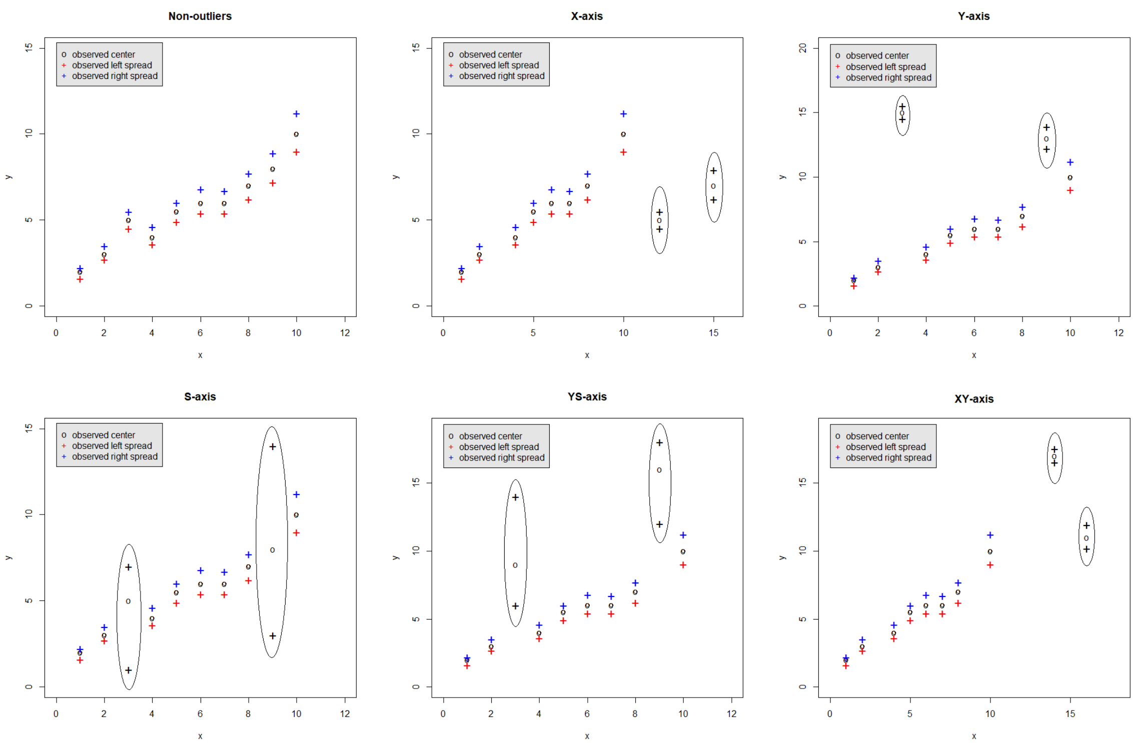

Figure 1: The different types of outliers.

Example 5.1. *Let we simulate some data from two fuzzy regression models as follows.*

$$\text{model } I: \quad \hat{\hat{Y}} = (2.00, 0.20) \oplus ((3.50, 0.40) \otimes X) \oplus \epsilon; \quad X \sim U(10, 30),$$

$$\text{model } II: \quad \hat{\hat{Y}} = (1.50, 1.25) \oplus ((2.50, 0.50) \otimes X) \oplus ((4.50, 0.75) \otimes X^2) \oplus \epsilon; \quad X \sim U(-3, 4),$$

where $\epsilon \sim N(0,1)$. *We simulate 35 data based on these models, such that 30 data are original data and 5 data are outliers. In Table 1, the different types of outliers are generated according to the bivariate Gaussian distribution* $N_2(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix})$. *Figures 2 and 3 show the different types of data simulated based on model I and model II, respectively. We repeat the designed models 20 times, to investigate the robust performance of our proposed models in presence of anomalous data sets.*

In each simulation, the estimations of parameters are computed by solving the minimization problem (3). The best fitted fuzzy regression models are selected based on the goodness of fit indices G *and* MSM *(see Table 2). The performance of the best fuzzy regression models are shown in Figures 2 and 3. Also, the values of optimal bandwidth* h_{opt} *in kernel functions are given in Table 2 (for instance, see Figure 4 from the original data sets).*

Based on the presented results in Tables 2-4, we can infer the following cases:

1. *Based on Table 2, the proposed fuzzy regression models are robust under the different types of outliers (the estimations of parameters are approximately similar in all cases for each model, separately).*

2. *The values of the goodness of fit measures under* G *and* MSM *are approximately similar based on original data set and while outliers are removed from data set.*

3. *The mean values of the estimated parameters from all points in data sets are approximately similar.*

4. *The mean values of goodness of fit measures* ($\overline{G}$ *and* $\overline{MSM}$) *are approximately similar based on original data set and while outliers are removed from data set (see Table 4).*

Table 1: Simulation based on the different types of outliers in Example 5.1.

Outlier's type	Model	Parameters
X-axis	Model I	$(X,Y) \sim N_2(\begin{bmatrix} max(x)+4 \\ \bar{y} \end{bmatrix}, \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix})$
	Model II	$(X,Y) \sim N_2(\begin{bmatrix} max(x)+3.5 \\ \bar{y} \end{bmatrix}, \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix})$
Y-axis	Model I	$(X,Y) \sim N_2(\begin{bmatrix} \bar{x} \\ max(y)+4 \end{bmatrix}, \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix})$
	Model II	$(X,Y) \sim N_2(\begin{bmatrix} \bar{x} \\ max(y)+3.5 \end{bmatrix}, \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix})$
S-axis	model I	$S_l = 0.2 + 0.4X; \quad X \sim U(60,80)$
	model II	$S_l = 1.25 + 0.5X + 1.75X^2; \quad X \sim U(4,5)$
YS-axis	Model I	$(X,Y) \sim N_2(\begin{bmatrix} \bar{x} \\ max(y)+4 \end{bmatrix}, \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix})$ $S_l = 0.2 + 0.4X; \quad X \sim U(60,80)$
	Model II	$(X,Y) \sim N_2(\begin{bmatrix} \bar{x} \\ max(y)+3.5 \end{bmatrix}, \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix})$ $S_l = 1.5 + 0.5X + 1.75X^2; \quad X \sim U(4,5)$
XY-axis	Model I	$(X,Y) \sim N_2(\begin{bmatrix} max(x)+8 \\ max(y)+6 \end{bmatrix}, \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix})$
	Model II	$(X,Y) \sim N_2(\begin{bmatrix} max(x)+6 \\ max(y)+5 \end{bmatrix}, \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix})$

Table 2: Optimal robust fuzzy regression models with the different types of outliers in Example 5.1.

Data	Model	Optimal fitted model	h_{obt}
Non-outliers	model I	$\hat{\tilde{Y}} = (2.075, 0.200) \oplus (3.493, 0.400) \otimes x_i$	3.1
	model II	$\hat{\tilde{Y}} = (1.475, 1.250) \oplus (2.729, 0.500) \otimes x_i \oplus (4.460, 0.750) \otimes x_i^2$	1.3
X-axis	model I	$\hat{\tilde{Y}} = (2.099, 0.200) \oplus (3.490, 0.400) \otimes x_i$	4.8
	model II	$\hat{\tilde{Y}} = (1.421, 1.250) \oplus (2.687, 0.500) \otimes x_i \oplus (4.480, 0.750) \otimes x_i^2$	1.5
Y-axis	model I	$\hat{\tilde{Y}} = (2.099, 0.199) \oplus (3.492, 0.400) \otimes x_i$	4.8
	model II	$\hat{\tilde{Y}} = (1.425, 1.250) \oplus (2.681, 0.500) \otimes x_i \oplus (4.445, 0.750) \otimes x_i^2$	1.5
S-axis	model I	$\hat{\tilde{Y}} = (2.103, 0.322) \oplus (3.490, 0.399) \otimes x_i$	3.9
	model II	$\hat{\tilde{Y}} = (1.434, 1.246) \oplus (2.692, 0.510) \otimes x_i \oplus (4.475, 0.756) \otimes x_i^2$	1.5
YS-axis	model I	$\hat{\tilde{Y}} = (2.105, 0.210) \oplus (3.493, 0.397) \otimes x_i$	7.4
	model II	$\hat{\tilde{Y}} = (1.415, 1.250) \oplus (2.687, 0.504) \otimes x_i \oplus (4.481, 0.753) \otimes x_i^2$	1.5
YX-axis	model I	$\hat{\tilde{Y}} = (2.012, 0.199) \oplus (3.497, 0.404) \otimes x_i$	2.0
	model II	$\hat{\tilde{Y}} = (1.429, 1.250) \oplus (2.696, 0.500) \otimes x_i \oplus 4.476, 0.750) \otimes x_i^2$	1.7

Table 3: Goodness of fit indices for models I and II in Example 5.1.

Data	Model	G	MSM	G without outliers	MSM without outliers
Non-outliers	model I	0.802	0.860	-	-
	model II	0.805	0.734	-	-
X-axis	model I	0.716	0.737	0.801	0.860
	model II	0.686	0.627	0.792	0.731
Y-axis	model I	0.714	0.737	0.801	0.860
	model II	0.766	0.666	0.792	0.731
S-axis	model I	0.759	0.781	0.797	0.858
	model II	0.707	0.627	0.798	0.733
YS-axis	model	0.729	0.734	0.798	0.858
	model II	0.748	0.629	0.792	0.731
YX-axis	model I	0.714	0.743	0.801	0.860
	model II	0.688	0.630	0.796	0.732

Example 5.2. *Consider the data set in Table 5 (Chang and Lee [7]). In this data set, the input observations are as crisp numbers but the output observations are as triangular fuzzy numbers. We change some data according to the different types of outliers as Table 6.*

We compare in this example our method with the methods introduced by Arefi [1], Diamond [12], and Zeng et al. [27] based on some goodness of fit indices. The results are given in Table 7. Also, Figure 5 shows the robustness of the proposed fuzzy regression model under the different types of outliers. By comparing the results given in Table 7, some

Table 4: Performance of fuzzy regression models in 20 times of simulationsfrom different types of outliers in Example 5.1.

Data	Model	Mean of $\tilde{\beta}_0 = (\beta_0, \gamma_0)$	Mean of $\tilde{\beta}_1 = (\beta_1, \gamma_1)$	Mean of $\tilde{\beta}_2 = (\beta_2, \gamma_2)$	$\bar{G}$ without outliers	$\overline{MSM}$ without outliers
Non-outliers	model I	$(2.175, 0.201)$	$(3.493, 0.399)$	-	0.768	0.828
	model II	$(1.540, 0.250)$	$(2.496, 0.500)$	$(4.489, 0.750)$	0.701	0.675
X-axis	model I	$(2.156, 0.199)$	$(3.494, 0.400)$	-	0.767	0.827
	model II	$(1.536, 0.248)$	$(2.495, 0.509)$	$(4.491, 0.750)$	0.700	0.673
Y-axis	model I	$(2.157, 0.189)$	$(3.493, 0.402)$	-	0.767	0.827
	model II	$(1.542, 0.249)$	$(2.495, 0.500)$	$(4.880, 0.749)$	0.697	0.673
S-axis	model I	$(2.212, 0.307)$	$(3.491, 0.397)$	-	0.765	0.828
	model II	$(1.539, 0.248)$	$(2.502, 0.501)$	$(4.487, 0.742)$	0.700	0.671
YS-axis	model I	$(2.118, 0.237)$	$(3.484, 0.409)$	-	0.765	0.828
	model II	$(1.535, 0.263)$	$(2.601, 0.512)$	$(4.490, 0.755)$	0.698	0.673
XY-axis	model I	$(2.284, 0.133)$	$(3.487, 0.404)$	-	0.768	0.828
	model II	$(1.536, 0.244)$	$(2.488, 0.500)$	$(4.485, 0.746)$	0.695	0.672

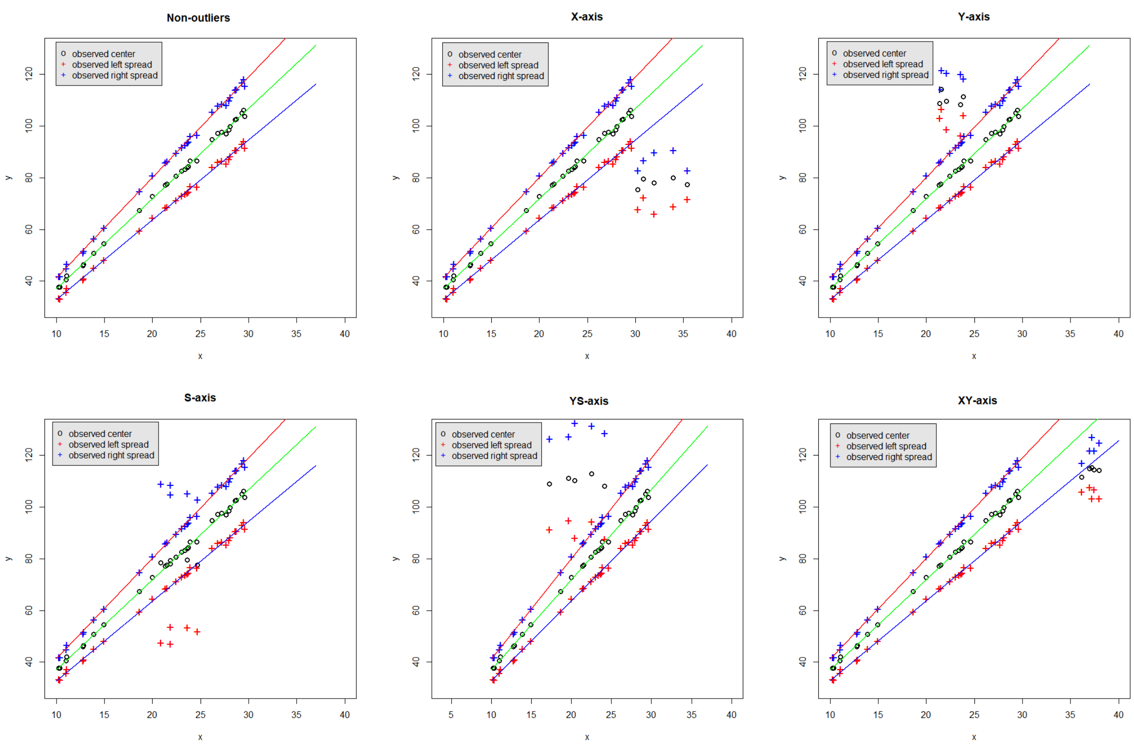

Figure 2: The robust fuzzy regression model (in model I) in the presence of different types of outliers in Example 5.1.

cases can be suggested as follows:

1. *In original data, the performance of each four models are approximately similar and good.*

2. *In the presence of outlier data, the Diamond's model is heavily impacted by outliers. Based on goodness of fit indices, our proposed model has a suitable performance rather than other models.*

Example 5.3. *Consider the data set in Table 8 (Kim and Bishu [16]). To compare the effect of outlier (No. 3), we obtain the fuzzy regression models with and without the presence of outlier. The results of optimal models are given in Table 9. By comparing the results provided in Table 9 based on the goodness of fit indices, our method provides a model with better performance and more robustness than the models given by Chen and Hsueh [9], Diamond [12], Taheri and Kelkinnama [25], and Zeng et al [27] (the difference between each goodness of fit measure for our model with and without the presence of outlier is less than other models).*

Example 5.4. *One of the important problem in meteorology is the measurement of average annual temperature. The average annual temperature Y (as the response variable) depends on the annual rainfall x_1 (in millimeter), the percent*

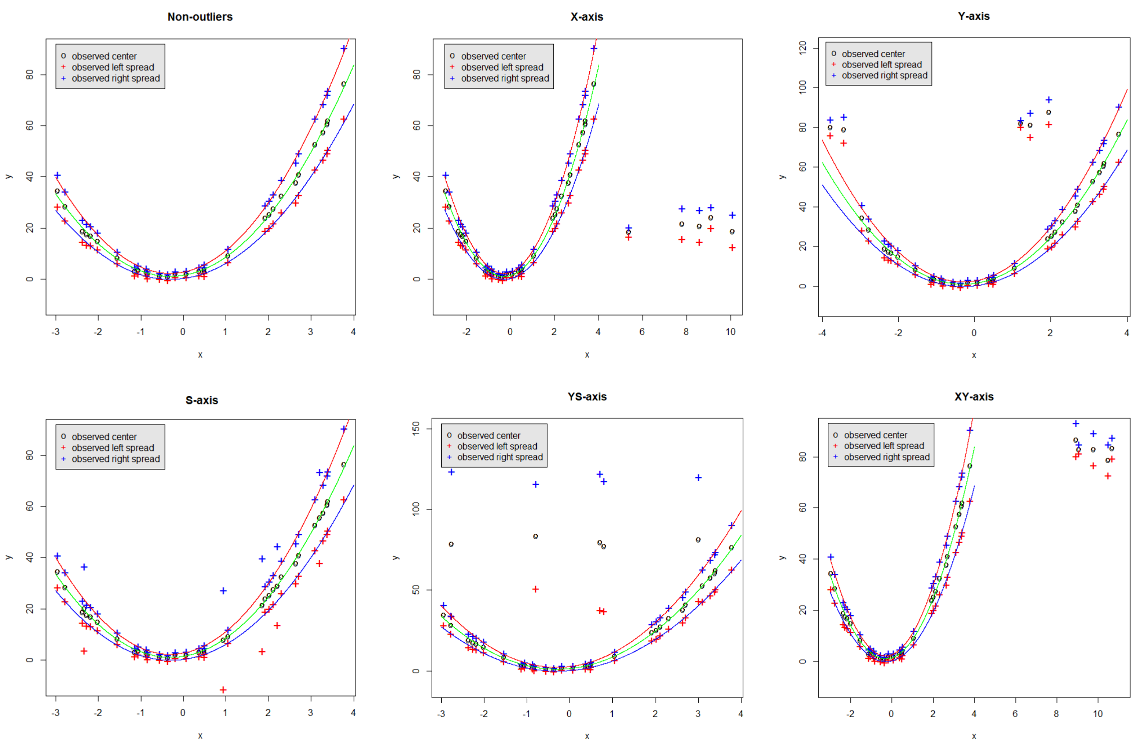

Figure 3: The robust fuzzy regression model (in model II) in the presence of different types of outliers in Example 5.1.

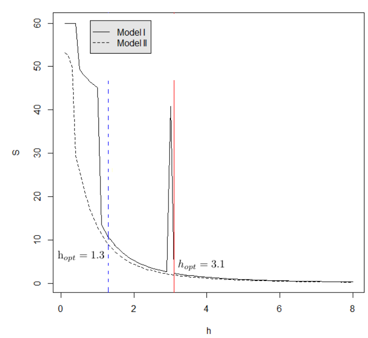

Figure 4: The optimal values of bandwidth h from the original data sets in Example 5.1.

Table 5: Data set in Example 5.2.

No.	x	$\tilde{Y}$	*No.*	x	$\tilde{Y}$
1	0.0	$(11.5, 3.0, 2.5)_T$	9	1.6	$(84.0, 15.0, 16.0)_T$
2	0.2	$(24.8, 4.5, 4.0)_T$	10	1.8	$(82.0, 15.0, 16.0)_T$
3	0.4	$(40.0, 6.0, 7.0)_T$	11	2.0	$(103.7, 16.0, 17.0)_T$
4	0.6	$(45.2, 7.0, 7.0)_T$	12	2.2	$(102.6, 16.0, 17.0)_T$
5	0.8	$(49.1, 9.0, 9.0)_T$	13	2.4	$(103.1, 16.0, 17.0)_T$
6	1.0	$(70.0, 11.0, 12.0)_T$	14	2.6	$(111.0, 17.0, 19.0)_T$
7	1.2	$(70.9, 12.0, 12.0)_T$	15	2.8	$(109.1, 17.0, 19.0)_T$
8	1.4	$(80.1, 14.0, 15.0)_T$	16	3.0	$(121.7, 18.0, 21.0)_T$

of average of relative humidity $x_2 = [x_{21}, x_{22}]$*, the number of ice days* x_3*, the number of days of dust* x_4*, the sum of sunny hours* x_5*, and the maximum of wind speed* x_6 *(in meter/second). In this example, the study area is based on 31 provinces of Iran in year 2015 (see Table 10). Among the data, the average annual temperature and the percent of*

Table 6: The different types of outliers in Example 5.2.

Outlier's type	No.	Outlier points	
		x	$\tilde{Y}$
X-axis	4	3.5	$(45.2, 7.0, 7.0)_T$
	8	2.5	$(80.1, 14.0, 15.0)_T$
	16	1.0	$(121.7, 18.0, 21.0)_T$
Y-axis	3	2.0	$(140.0, 6.0, 7.0)_T$
	14	2.0	$(40.0, 17.0, 19.0)_T$
	16	3.0	$(160.0, 8.0, 30.0)_T$
S-axis	2	0.2	$(24.8, 20.0, 60.0)_T$
	12	2.2	$(102.6, 30.0, 80.0)_T$
	14	2.6	$(111.0, 90.0, 60.0)_T$
YS-axis	11	2.2	$(180.0, 30.0, 60.0)_T$
	12	2.4	$(150.0, 80.0, 30.0)_T$
	16	3.0	$(160.0, 8.0, 30.0)_T$
XY-axis	9	3.5	$(190.0, 15.0, 16.0)_T$
	15	4.0	$(180.0, 17.0, 19.0)_T$

Table 7: Comparison between the proposed models in Example 5.2.

Data	Method	Optimal fitted Model	G	MSM
Non-outliers	Diamond	$\hat{\tilde{Y}} = (14.529, 4.853, 4.460)_T \oplus (55.356, 4.952, 5.798)_T \otimes x_i \oplus (-7.101, 0.000, 0.000)_T \otimes x_i^2$	0.593	0.607
	Zeng et al.	$\hat{\tilde{Y}} = (13.738, 4.513, 4.818)_T \oplus (56.800, 5.008, 5.455)_T \otimes x_i \oplus (-7.458, 0.000, 0.000)_T \otimes x_i^2$	0.602	0.614
	Arefi	$\hat{\tilde{Y}} = (13.523, 3.609, 2.981)_T \oplus (57.410, 5.632, 6.366)_T \otimes x_i \oplus (-7.691, 0.000, 0.000)_T \otimes x_i^2$; $\tau = 0.5$	0.625	0.624
	Our method	$\hat{\tilde{Y}} = (13.340, 3.963, 3.197)_T \oplus (58.032, 5.557, 6.588)_T \otimes x_i \oplus (-8.019, 0.000, 0.000)_T \otimes x_i^2$; $h_{opt} = 4.7$	0.615	0.625
X-axis	Diamond	$\hat{\tilde{Y}} = (8.470, 7.836, 8.203)_T \oplus (89.340, 2.736, 3.048)_T \otimes x_i \oplus (-21.626, 0.000, 0.000)_T \otimes x_i^2$	0.368	0.294
	Zeng et al.	$\hat{\tilde{Y}} = (11.500, 5.444, 4.818)_T \oplus (73.024, 4.444, 5.455)_T \otimes x_i \oplus (-14.524, 0.000, 0.000)_T \otimes x_i^2$	0.479	0.440
	Arefi	$\hat{\tilde{Y}} = (11.500, 3.006, 2.518)_T \oplus (75.394, 5.703, 6.594)_T \otimes x_i \oplus (-15.450, 0.002, 0.000)_T \otimes x_i^2$; $\tau = 0.6$	0.511	0.452
	Our method	$\hat{\tilde{Y}} = (13.405, 4.402, 4.067)_T \oplus (58.869, 5.248, 5.985)_T \otimes x_i \oplus (-8.687, 0.000, 0.000)_T \otimes x_i^2$; $h_{opt} = 11.7$	0.508	0.479
Y-axis	Diamond	$\hat{\tilde{Y}} = (42.959, 4.853, 4.459)_T \oplus (24.721, 4.952, 5.798)_T \otimes x_i \oplus (-0.091, 0, 0)_T \otimes x_i^2$	0.444	0.365
	Zeng et al.	$\hat{\tilde{Y}} = (13.528, 3.777, 4.818)_T \oplus (58.067, 5.556, 5.455)_T \otimes x_i \oplus (-8.548, 0, 0)_T \otimes x_i^2$	0.556	0.520
	Arefi	$\hat{\tilde{Y}} = (13.664, 3.349, 2.687)_T \oplus (57.204, 5.751, 7.052)_T \otimes x_i \oplus (-7.627, 0.000, 0.000)_T \otimes x_i^2$; $\tau = 0.65$	0.560	0.514
	Our method	$\hat{\tilde{Y}} = (11.885, 4.565, 4.026)_T \oplus (60.827, 5.220, 6.030)_T \otimes x_i \oplus (-9.271, 0, 0)_T \otimes x_i^2$; $h_{opt} = 10.6$	0.559	0.536
S-axis	Diamond	$\hat{\tilde{Y}} = (16.623, 6.298, 15.163)_T \oplus (51.028, 8.259, 0)_T \otimes x_i \oplus (-5.682, 0, 2.578)_T \otimes x_i^2$	0.483	0.479
	Zeng et al.	$\hat{\tilde{Y}} = (13.739, 4.281, 4.500)_T \oplus (56.799, 5.945, 6.250)_T \otimes x_i \oplus (-7.458, 0.000, 0.000)_T \otimes x_i^2$	0.575	0.500
	Arefi	$\hat{\tilde{Y}} = (13.978, 7.956, 1.708)_T \oplus (56.108, 0.627, 9.335)_T \otimes x_i \oplus (-7.228, 2.286, 0.020)_T \otimes x_i^2$; $\tau = 0.5$	0.567	0.518
	Our method	$\hat{\tilde{Y}} = (13.732, 6.938, 4.814)_T \oplus (57.121, 5.116, 5.724)_T \otimes x_i \oplus (-7.673, 0.000, 0.000)_T \otimes x_i^2$; $h_{opt} = 9$	0.573	0.510
YS-axis	Diamond	$\hat{\tilde{Y}} = (10.445, 5.044, 3.658)_T \oplus (64.837, 7.658, 9.040)_T \otimes x_i \oplus (-7.105, 0.000, 0.000)_T \otimes x_i^2$	0.374	0.299
	Zeng et al.	$\hat{\tilde{Y}} = (13.695, 4.383, 2.500)_T \oplus (57.03, 5.304, 8.125)_T \otimes x_i \oplus (-7.542, 0.000, 0.000)_T \otimes x_i^2$	0.556	0.498
	Arefi	$\hat{\tilde{Y}} = (13.703, 7.389, 2.568)_T \oplus (56.980, 1.693, 7.522)_T \otimes x_i \oplus (-7.754, 0.306, 0.002)_T \otimes x_i^2$; $\tau = 0.45$	0.525	0.479
	Our method	$\hat{\tilde{Y}} = (13.887, 4.879, 4.178)_T \oplus (57.315, 4.200, 5.893)_T \otimes x_i \oplus (-8.240, 0.000, 0.000)_T \otimes x_i^2$; $h_{opt} = 10.4$	0.566	0.511
XY-axis	Diamond	$\hat{\tilde{Y}} = (21.523, 6.136, 5.983)_T \oplus (34.325, 3.628, 4.235)_T \otimes x_i \oplus (1.647, 0.000, 0.000)_T \otimes x_i^2$	0.405	0.358
	Zeng et al.	$\hat{\tilde{Y}} = (20.441, 5.500, 4.846)_T \oplus (42.736, 4.375, 5.385)_T \otimes x_i \oplus (-2.450, 0.000, 0.000)_T \otimes x_i^2$	0.489	0.484
	Arefi	$\hat{\tilde{Y}} = (14.171, 6.336, 4.132)_T \oplus (45.235, 2.991, 2.483)_T \otimes x_i \oplus (-3.131, 0.300, 1.069)_T \otimes x_i^2$; $\tau = 0.25$	0.470	0.411
	Our method	$\hat{\tilde{Y}} = (14.309, 4.410, 3.892)_T \oplus (55.470, 5.109, 5.998)_T \otimes x_i \oplus (-6.880, 0.000, 0.000)_T \otimes x_i^2$; $h_{opt} = 9.1$	0.543	0.523

Table 8: Data set in Example 5.3.

$No.$	$\tilde{Y}$	x_1	x_2	x_3
1	$(5.83, 3.56)_T$	2.00	0.00	15.25
2	$(0.85, 0.52)_T$	0.00	5.00	14.13
3	$(13.93, 8.50)_T$	1.13	1.50	14.13
4	$(4.00, 2.44)_T$	2.00	1.25	13.63
5	$(1.65, 1.01)_T$	2.19	3.75	14.75
6	$(1.58, 0.96)_T$	0.25	3.50	13.75
7	$(8.18, 4.99)_T$	0.755	5.25	15.25
8	$(1.85, 1.13)_T$	4.25	2.00	13.50

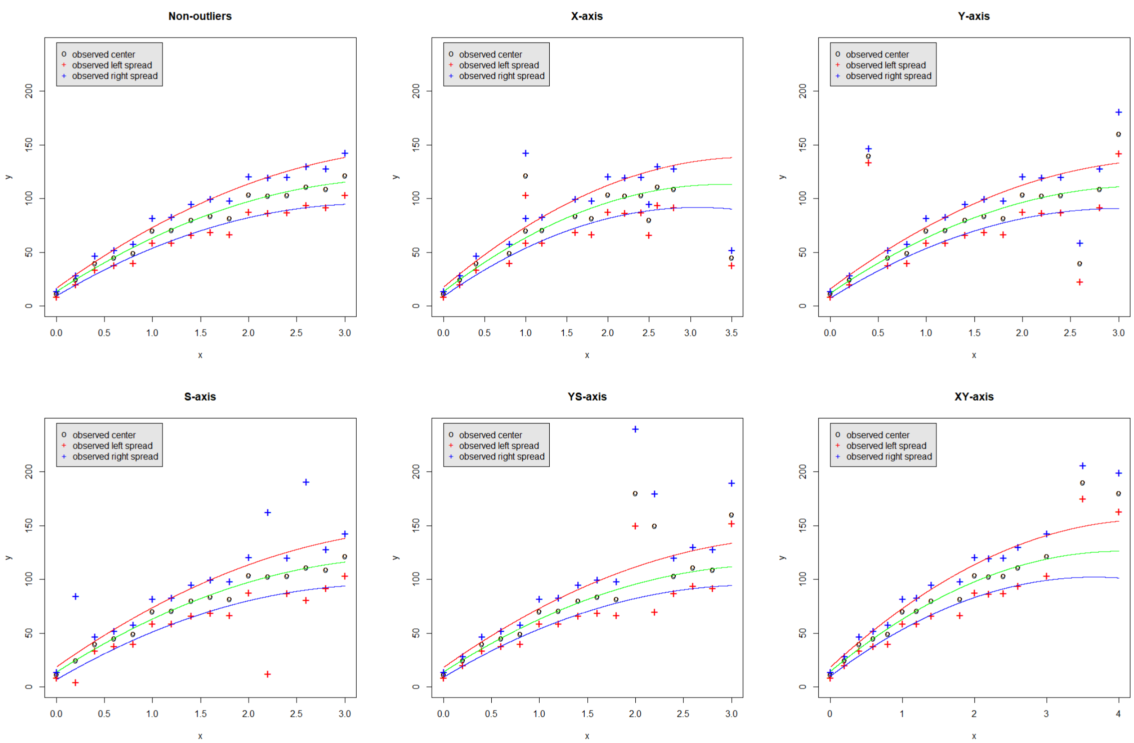

Figure 5: The robust fuzzy regression models with the different types of outliers in Example 5.2.

Table 9: Comparison between the different models in Example 5.3.

Method	Optimal fitted Model	G	MSM	G without outlier	MSM without outlier
Diamond	$\hat{\tilde{Y}} = (-16.796, 0)_T \ominus ((1.099, 0)_T \otimes x_{1i}) \ominus ((1.180, 0)_T \otimes x_{2i}) \oplus ((1.856, 0.204)_T \otimes x_{3i})$	0.421	0.185	0.498	0.297
Chen and Hsueh	$\hat{\tilde{Y}} = -16.796 \ominus 1.099x_{1i} \ominus 1.180x_{2i} \oplus 1.856x_{3i} \oplus (0, 2.289)_T$	0.420	0.181	0.469	0.290
Taheri and Kelkinnama	$\hat{\tilde{Y}} = -15.557 \ominus 0.244x_{1i} \ominus 0.997x_{2i} \oplus 1.514x_{3i} \oplus (0, 1.130)_T$	0.647	0.391	0.686	0.485
Zeng et al.	$\hat{\tilde{Y}} = (-2.827, 0)_T \ominus ((0.388, 0)_T \otimes x_{1i}) \ominus ((1.013, 0)_T \otimes x_{2i}) \oplus ((0.619, 0.179)_T \otimes x_{3i})$	0.531	0.363	0.707	0.502
Our method	$\hat{\tilde{Y}} = (-4.868, 0)_T \ominus ((0.321, 0.114)_T \otimes x_{1i}) \ominus ((1.018, 0)_T \otimes x_{2i}) \oplus (0.751, 0.064)_T \otimes x_{3i})$; $h_{opt} = 0.6$	0.683	0.445	0.704	0.508

average of relative humidity is reported as intervals $Y = [Y_1, Y_2]$ and $x_2 = [x_{21}, x_{22}]$. We suppose that these two cases of data are formulated by the symmetric triangular fuzzy numbers as follows:

$$\tilde{Y} = (y, s_l, s_r)_T, \quad with \quad y = \frac{Y_1 + Y_2}{2}, \quad s_l = y - Y_1, \quad s_r = Y_2 - y,$$

$$\tilde{x}_2 = (x_2, s_{l_2}, s_{r_2})_T, \quad with \quad x_2 = \frac{x_{21} + x_{22}}{2}, \quad s_{l_2} = x_2 - x_{21}, \quad s_{r_2} = x_{22} - x_2.$$

Now, we wish to fit a robust fuzzy regression model as follows:

$$\hat{\tilde{Y}} = \beta_0 \oplus (\beta_1 x_1) \oplus (\beta_2 \otimes \tilde{x}_2) \oplus (\beta_3 x_3) \oplus (\beta_4 x_4) \oplus (\beta_5 x_5) \oplus (\beta_6 x_6).$$

Based on Section 4, the optimal fuzzy regression model with $h_{opt} = 5.7$ is obtained as

$$\hat{\tilde{Y}} = -22.8219 \oplus (-0.0001x_1) \oplus (0.2710 \otimes \tilde{x}_2) \oplus (-0.0937x_3) \oplus (0.0118x_4) \oplus (0.0115x_5) \oplus (-0.1192x_6).$$

Also, based on the indices of goodness of fit, we have $G = 0.679$ and $MSM = 0.612$. Hence, it can be concluded that there is a good performance of the fitness on this real data set.

Table 10: Real data set in Example 5.4.

$No.$	provincial capital	Y		x_1	x_2		x_3	x_4	x_5	x_6
		Y_1	Y_2		x_{21}	x_{22}				
1	Arak	8.3	22.0	284.8	26.6	60.5	78	84	2968.2	27
2	Ardabil	4.7	17.1	296.5	46.5	86.2	101	11	2671.0	27
3	Urmia	5.7	19.3	277.3	37.4	76.7	108	16	2902.0	23
4	Isfahan	10.5	24.8	96.3	17.7	49.6	62	50	3430.3	22
5	Ahvaz	20.4	34.3	269.7	28.2	65.3	0	79	3155.5	16
6	Ilam	11.3	24.6	842.4	25.2	56.5	33	70	3122.0	30
7	Bojnourd	8.0	21.6	227.4	35.4	78.3	82	23	2953.0	19
8	Bandar Abbas	22.4	33.1	152.2	45.0	81.9	0	139	3269.3	22
9	Bushehr	22.5	30.4	272.5	52.9	79.2	0	52	3063.0	40
10	Birjand	9.3	25.5	144.3	14.6	46.9	67	61	3435.3	16
11	Tabriz	8.2	19.8	286.9	34.7	69.4	88	28	2784.2	21
12	Tehran	14.2	23.9	209.3	21.7	47.7	13	24	2983.2	22
13	Khorramabad	9.9	26.0	535.6	23.6	63.8	49	69	2929.0	20
14	Rasht	13.3	21.4	1388.3	64.4	96.2	9	4	1885.8	18
15	Zahedan	11.8	27.7	103.7	17.5	45.7	45	85	3442.3	16
16	Zanjan	6.8	21.1	283.1	35.6	69.7	82	31	2850.4	26
17	Sari	14.0	23.2	724.7	61.1	93.1	3	1	2134.3	27
18	Semnan	14.0	25.0	107.5	21.5	45.8	31	27	3204.9	16
19	Sanandaj	7.3	23.4	444.4	25.6	66.5	92	53	2970.7	15
20	Shahrekord	2.3	21.3	309.7	22.9	66.2	133	42	3258.3	32
21	Shiraz	11.0	26.8	271.5	18.4	58.9	43	73	3360.1	16
22	Qazvin	8.4	23.0	313.7	3.5	72.1	65	20	3019.1	26
23	Qom	11.6	27.6	111.6	24.1	56.8	49	113	3211.3	25
24	Karaj	10.5	22.9	220.5	28.5	64.2	42	45	3050.8	27
25	Kerman	8.3	26.1	109.8	13.1	48.1	71	23	3486.3	25
26	Kermanshah	8.7	24.3	512.8	21.3	56.9	63	48	2980.0	17
27	Gorgan	12.6	25.0	477.8	53.4	91.5	21	6	2289.1	26
28	Mashhad	10.6	23.8	183.4	26.7	63.1	40	11	3101.3	25
29	Hamadan	5.2	21.0	215.7	30.2	66.6	107	33	3013.6	29
30	Yasuj	7.9	23.6	611.1	19.9	63.0	79	34	3230.6	14
31	Yazd	14.6	28.1	38.4	13.6	37.3	21	53	3490.7	24

6 Comparison studies

In this section, we compare our approach with some other approaches on fuzzy regression model as follows.

1) Diamond [12] studied a least squares approach to fit a fuzzy linear regression model with fuzzy coefficients $\tilde{\beta_j} = (\beta_j, \gamma_j, \theta_j)_T$ and based on crisp inputs x_{ij} and fuzzy output $\tilde{Y_i} = (y_i, s_{l_i}, s_{r_i})_T$. In this approach, a distance on the set of fuzzy numbers is introduced by invoking the Hausdorff distance, and then an objective function is defined based on this distance. The parameters of model are estimated by minimizing this function. The objective function is presented as follows:

$$S_D = \sum_{i=1}^{n} d^2(\tilde{Y_i}, \hat{\tilde{Y_i}}) = \sum_{i=1}^{n} \big[(y_i - \sum_{j=0}^{p} \beta_j x_{ij})^2 + ((y_i - \sum_{j=0}^{p} \beta_j x_{ij}) - (s_{l_i} - \sum_{j=0}^{p} \gamma_j x_{ij}))^2 + ((y_i - \sum_{j=0}^{p} \beta_j x_{ij}) + (s_{r_i} - \sum_{j=0}^{p} \theta_j x_{ij}))^2\big].$$

2) Chen and Hsueh [9] proposed a least squares fuzzy regression model with crisp coefficients. They used the least squares method to minimize the squares sum of errors between a few α-cuts of the observed and estimated fuzzy responses. Hence, the objective function is drawn as follows:

$$S_{CH} = \sum_{i=1}^{n} \frac{1}{2m} \sum_{k=0}^{m-1} \Big[((\tilde{Y_i})^L_{\alpha_k} - (\hat{\tilde{Y_i}})^L_{\alpha_k})^2 + ((\tilde{Y_i})^R_{\alpha_k} - (\hat{\tilde{Y_i}})^R_{\alpha_k})^2\Big].$$

3) Taheri and Kelkinnama [25] investigated an approach to fit a fuzzy linear regression based on the least absolutes method under the fuzzy inputs $\tilde{x_{ij}} = (x_{ij}, s'_{l_{ij}}, s'_{r_{ij}})_{LL}$, fuzzy output $\tilde{Y_i} = (y_i, s_{l_i}, s_{r_i})_{LL}$ and crisp coefficients β_j. They first introduced a distance based on the absolutes between the centers and the bounds of fuzzy numbers, and then the objective function is as

$$\begin{aligned} S_{TK} &= \textstyle\sum_{i=1}^{n} d(\tilde{Y_i}, \hat{\tilde{Y_i}}) = \frac{1}{3} \sum_{i=1}^{n} \Big[|y_i - \sum_{j=0}^{p} \beta_j x_{ij}| + |(y_i - s_{l_i}) - (\sum_{j=0}^{p} \beta_j x_{ij} - \lambda(\sum_{j=0}^{p} |\beta_j| s'_{l_{ij}} + a))| \\ &\quad + |(y_i + s_{r_i}) - (\textstyle\sum_{j=0}^{p} \beta_j x_{ij} + \lambda(\sum_{j=0}^{p} |\beta_j| s'_{r_{ij}} + b))|\Big]. \end{aligned}$$

4) Zeng et al. [27] first introduced an absolute distance between triangular fuzzy numbers, and then the least absolute fuzzy linear regression model is constricted based on this distance under the crisp inputs x_{ij}, the fuzzy outputs $\tilde{Y_i} = (y_i, s_{l_i}, s_{r_i})_T$ and the fuzzy coefficients $\tilde{\beta_j} = (\beta_j, \gamma_j, \theta_j)_T$. The parameters of the model are obtained to minimize the objective function given as follows

$$S_Z = \sum_{i=1}^{n} d(\tilde{Y_i} - \hat{\tilde{Y_i}}) = \sum_{i=1}^{n} \Big[|y_i - \sum_{j=0}^{p} \beta_j x_{ij}| + |s_{l_i} - \sum_{j=0}^{p} \gamma_j x_{ij}| + |s_{r_i} - \sum_{j=0}^{p} \theta_j x_{ij}|\Big],$$

where $\gamma_j \geq 0$, $\theta_j \geq 0$, $\sum_{j=0}^{p} \gamma_j x_{ij} \geq 0$, and $\sum_{j=0}^{p} \theta_j x_{ij} \geq 0$.

5) Arefi [1] studied a quantile approach to model a fuzzy linear regression based on the crisp inputs, the fuzzy outputs $\tilde{Y}_i = (y_i, s_{l_i}, s_{r_i})_{LL}$, and the fuzzy coefficients $\tilde{\beta}_j = (\beta_j, \gamma_j, \theta_j)_{LL}$. In this approach, a loss function between fuzzy numbers is first introduced which it can present some quantiles of fuzzy data. Then, a quantile fuzzy regression model is constricted based on this loss function. The objective function is as

$$S_A = \sum_{i=1}^{n} \psi_\tau(\tilde{Y}_i - \hat{\tilde{Y}}_i) = \frac{1}{3} \sum_{i=1}^{n} \big[\psi_\tau(y_i - \sum_{j=0}^{p} \beta_j x_{ij}) + \psi_\tau((y_i - \sum_{j=0}^{p} \beta_j x_{ij}) - \lambda(s_{l_i} - \sum_{j=0}^{p} \gamma_j x_{ij})) + \psi_\tau((y_i - \sum_{j=0}^{p} \beta_j x_{ij}) + \lambda(s_{r_i} - \sum_{j=0}^{p} \theta_j x_{ij}))\big],$$

where, $0 < \tau < 1$ and $\psi_\tau(e) = e \times (\tau - I(e < 0))$ is the classical quantile loss function.

Based on our approach and the above proposed approaches, some comments are notable as follows:

A1) All these approaches are proposed on the fuzzy linear regression model with some fuzzy quantities in the available data.

A2) Unlike the approaches given by Chen and Hsueh [9] and Diamond [12] (which they are based on the least squares approach), other approaches are the robust under outlier data.

A3) Since our method and the Arefi's method [1] are based on bandwidth parameter h (in kernel function) and quantile level τ (in the loss function), respectively, these two approaches are more flexibility and robustness rather than other approaches.

A4) An important feature of proposed method in this paper is its dependence on the kernel function. We can extend this approach based on the other kernel functions. Also, note that one of the reasons, which this method is robust under outliers, is because of the kernel function.

7 Conclusion

A new approach is presented to fit the robust fuzzy regression model when the input variables are as crisp/fuzzy and output variables are as fuzzy numbers. This approach has certain merits as follows:

1) It is a new approach on fuzzy regression model under crisp/fuzzy input, fuzzy output, and crisp/fuzzy parameters of model.

2) A new distance between fuzzy numbers is introduced based on the kernel function for obtaining the optimal robust fuzzy regression procedure.

3) The proposed fuzzy regression model is the robust under the different types of outliers.

Acknowledgement

The authors wish to express their appreciation for several excellent suggestions for improvements in this paper made by the referees.

References

[1] M. Arefi, *Quantile fuzzy regression based on fuzzy outputs and fuzzy parameters*, Soft Computing, **24** (2020), 311-320.

[2] A. Celmins, *Least squares model fitting to fuzzy vector data*, Fuzzy Sets and Systems, **22** (1987), 260-269.

[3] J. Chachi, *A weighted least-squares fuzzy regression for crisp input-fuzzy output data*, IEEE Transactions on Fuzzy Systems, **27** (2019), 739-748.

[4] J. Chachi, M. Roozbeh, *A fuzzy robust regression approach applied to bedload transport data*, Communications in Statistics-Simulation and Computation, **47** (2017), 1703-1714.

[5] J. Chachi, S. M. Taheri, *Multiple fuzzy regression model for fuzzy input-output data*, Iranian Journal of Fuzzy Systems, **13** (2016), 63-78.

[6] J. Chachi, S. M. Taheri, S. Fattahi, S. A. H. Ravandi, *Two robust fuzzy regression models and their application in predicting imperfections of cotton yarn*, Journal of Textiles and Polymers, **4** (2016), 60-68.

[7] P. T. Chang, C. H. Lee, *Fuzzy least absolute deviations regression based on the ranking of fuzzy numbers*, In: Proc. of the Third IEEE World Congress on Computaional Intelligence, Orlando, FL, **2** (1994), 1365-1369.

[8] S. P. Chen, J. F. Dang, *A variable spread fuzzy linear regression model with higher explanatory power and forecasting accuracy*, Information Sciences, **178** (2008), 3973-3988.

[9] L. H. Chen, C. C. Hsueh, *Fuzzy regression models using the least-squares method based on the concept of distance*, IEEE Transactions on Fuzzy Systems, **17** (2009), 1259-1272.

[10] S. H. Choi, J. J. Buckley, *Fuzzy regression using least absolute deviation estimators*, Soft Computing, **12** (2008), 257-263.

[11] F. A. T. De Carvalho, E. A. Lima Neto, M. R. P. Ferreira, *A robust regression method based on exponential-type kernel functions*, Neurocomputing, **234** (2019), 58-74.

[12] P. Diamond, *Fuzzy least squares*, Information Sciences, **46** (1988), 141-157.

[13] P. D'Urso, R. Massari, *Weighted least squares and least median squares estimation for the fuzzy linear regression analysis*, Metron, **71** (2013), 279-306.

[14] P. D'Urso, R. Massari, A. Santoro, *Robust fuzzy regression analysis*, Information Sciences, **181** (2011), 4154-4174.

[15] G. Hesamian, M. G. Akbari, *Fuzzy quantile linear regression model adopted with a semi-parametric technique based on fuzzy pedictors and fuzzy responses*, Expert Systems with Applications, **118** (2019), 585-597.

[16] B. Kim, R. R. Bishu, *Evaluation of fuzzy linear regression models by comparison membership function*, Fuzzy Sets and Systems, **100** (1998), 343-352.

[17] K. S. Kula, A. Apaydin, *Fuzzy robust regression analysis based on the ranking of fuzzy sets*, International Journal Uncertainty, Fuzziness and Knowledge-Based Systems, **16** (2008), 663-681.

[18] J. Mercer, *Functions of positive and negative type and their connection with the theory of integral equations*, Philosophical Transactions of the Royal Society A, **209** (1909), 441-458.

[19] K. R. Mueller, S. Mika, G. R. Raetsch, K. Tsuda, F. Schloelkopf, *An introduction to kernel-based learning algorithms*, IEEE Transactions on Neural Networks, **12** (2001), 181-202.

[20] E. Nasrabadi, S. M. Hashemi, *Robust fuzzy regression analysis using neural networks*, International Journal Uncertainty Fuzziness and Knowledge-Based Systems, **16** (2008), 579-598.

[21] M. Oussalah, J. De Schutter, *Robust fuzzy linear regression and application for contact identification*, Intelligent Automation and Soft Computing, **8** (2002), 31-39.

[22] C. P. Pappis, N. I. Karacapilidis, *A comparative assessment of measure of similarity of fuzzy values*, Fuzzy Sets and Systems, **56** (1993), 171-174.

[23] K. Sanli, A. Apaydin, *The fuzzy robust regression analysis, the case of fuzzy data sethas outlier*, Gazi University Journal of Science, **17** (2004), 71-84.

[24] B. Y. Shon, *Robust fuzzy linear regression based on M-estimators*, Applied Mathematics and Computation, **18** (2005), 591-601.

[25] S. M. Taheri, M. Kelkinnama, *Fuzzy linear regression based on least absolute deviations*, Irannian Journal of Fuzzy Systems, **9** (2012), 121-140.

[26] Inc. Wolfram Research, *Mathematica* 10.2, Wolfram Research, Inc. Place of publication: Champaign, Illinois Date of publication, (2015).

[27] W. Zeng, Q. Feng, J. Lia, *Fuzzy least absolute linear regression*, Applied Soft Computing, **52** (2016), 1009-1019.

[28] H. J. Zimmermann, *Fuzzy set theory and its applications*, Boston: 4th ed., Kluwer Nihoff, (2001).