Development of a T_w-norm based Novel Fuzzy Regression Model

B. Pushpa[1], S. Muruganandam[2]

[1] Research Scholar, Manonmaniam Sundaranar University, Tirunelveli, Tamilnadu, 627102, India.
Department of Mathematics, Panimalar Institute of Technology, Chennai, 600123, India.

[2] Prof. & Head, Department of Mathematics, TRP Engineering College, Trichy, Tamilnadu, 621105, India

---

ABSTRACT— ***The weakest t-norm (T_w-norm) based novel fuzzy regression model is developed using the concept of symmetric difference. The proposed model will be useful in a realistic environment and improve upon the traditional fuzzy regression. The traditional system usually adopts alpha cut operations for its calculations. Here the T_w- norm based operations are used, to reduce the width of the estimated responses which will give exact prediction. Fuzzy linear regression analysis can be seen as an optimization problem where the aim is to derive a model which fits the given dataset. The proposed fuzzy regression analysis uses the extended objective function which is insensitive to the outlier data and the performance of the method is illustrated with different examples.***

Keywords— Weakest t-norm, Fuzzy regression, Symmetric difference

---

1. INTRODUCTION

Regression analysis, including statistical regression analysis and fuzzy regression analysis, aims to determine the best-fit model for describing the functional relationship between dependent variables and independent variables by exploiting the knowledge from the given input-output data pairs. Some discrepancy between the observed values (from the data sets) and the estimated values (from a regression model) can occur due to measurement errors and/or modeling errors. With the modeling errors ignored, the deviations are supposed to be random in classical regression analysis. For the fuzzy regression analysis, the deviations are attributed to the imprecision of the observed values and/or the indefiniteness of model structure. In this case, the observed values can differ from the estimated values to a certain degree of belief. Fuzzy regression has been found to be more appealing than statistical regression in estimating the relationship between the dependent variable and independent variables when a high degree of fuzziness is involved and only a few data sets are available.

Tanaka et al. [1] proposed the formulation of possibilistic linear regression analysis and determined the fuzzy parameters by applying linear programming models. However, it is known that this method has several drawbacks [2,3,4,5,7]. It is very sensitive to outliers and when linear programming is used in possibilistic regression, some coefficients tend to become crisp due to the characteristics of linear programming. To tackle this problem, Nasrabedi et al.[2] proposed a mathematical-programming approach to fuzzy linear regression analysis. Another problem observed is that when the coefficients are fuzzy numbers the spread of the estimated response becomes wider as the magnitudes of the explanatory variables increase, even though the spreads of the observed responses decrease, or as more observations are included in the model. This contradicts intuition. To prevent this problem, Diamond [3], Wu and Tseng [4] and Kao and Chyu [6] considered numeric coefficients to describe the fuzzy relationship between the fuzzy response variable and fuzzy (or numeric) explanatory variables. All of them used the concept of least squares to determine the regression coefficients. Kao and Chyu[6] considered a two-stage solution procedure to determine the numeric coefficients using the criterion of Kim and Bishu et al.[7] which is a modification of the fuzzy linear regression analysis criterion proposed by Tanaka et al. [1]. Although this criterion has been used by several authors, it has the drawback that if the observed and estimated fuzzy responses do not intersect with each other, the error estimation remains constant.

Regression analysis based on the method of least -absolute deviation has been used as a robust method. When outlier exists in the response variable, the least absolute deviation is more robust than the least square deviations estimators. Some recent works on this topic are as follows: Lee and Chang [8] studied the fuzzy least absolute deviation regression based on the ranking method for fuzzy numbers. Kim and Bishu et al. [7] proposed a two stage method to construct the fuzzy linear regression models, using a least absolutes deviations method. Torabi and Behboodian et al.[9] investigated the usage of ordinary least absolute deviation method to estimate the fuzzy coefficients in a linear regression model with fuzzy input – fuzzy output observations. Considering a certain fuzzy regression model, Chen and Hsueh et al.[10] developed a mathematical programming method to determine the crisp coefficients as well as an adjusted term for a fuzzy regression model, based on L_1 norm (absolute norm) criteria. Choi and Buckley et al.[11] suggested two methods to obtain the least absolute deviation estimators for common fuzzy linear regression models using T_M based arithmetic operations. Taheri and Kelkinnama et al.[12] introduced some least absolute regression models, based on crisp input-fuzzy output and fuzzy input-fuzzy output data respectively.

In a regression model, multiplication of fuzzy numbers are done by arithmetic operations such as α-levels of multiplication of fuzzy numbers and the approximate formula for multiplication of fuzzy numbers. The α-cut arithmetic provides results such that the fuzziness of the model calculation was fuzzier than that of the T_w-fuzzy arithmetic due to the accumulation of fuzziness of the α-cut arithmetic and the α-cut arithmetic cannot effectively preserve the original shape of a membership function. Apart from these two, we know that using the weakest T – norm (T_w), the shape of fuzzy numbers in multiplication will be preserved. The T_w arithmetic gives a justifiable fuzziness/ fuzzy spread because it takes only the maximal fuzziness encountered and calculates that into the operation. In this regard, Hong, Lee and Do et al. [13] presented a method to evaluate fuzzy linear regression models based on a possibilistic approach, using T_w - norm based arithmetic operations. The objective of this study is to develop a fuzzy regression model to handle the functional dependence of crisp/ fuzzy inputs-fuzzy output variables using the symmetric difference between fuzzy numbers.

In this paper, section 2 focuses on some important preliminary definitions and basics on fuzzy arithmetic operations based on the weakest T-norm. In section 3, the new approach based on symmetric difference is presented using the shape preserving operations on fuzzy numbers and it is analyzed with crisp/ fuzzy input and fuzzy output with symmetric/ non-symmetric triangular fuzzy numbers and Trapezoidal fuzzy numbers. The goodness of fit of the proposed model is also discussed. In section 4, by using numerical examples some comparative studies are presented to show the performance of the proposed method.

2. THE WEAKEST T-NORM BASED ARITHMETIC OPERATIONS

Since our study concerns fuzzy arithmetic based on the weakest norm, this section will briefly introduce T_w arithmetic operation. The basic concepts and definition of the weakest t-norm arithmetic operations will be introduced in the following:

In fuzzy arithmetic approaches, Zadeh's sup-min operator [14] can be defined as $\left(\tilde{A}\circ\tilde{B}\right)(z)=\sup_{x\circ y=z}\min\left[\tilde{A}(x),\tilde{B}(y)\right]$ where $\circ$ denotes any fuzzy arithmetic operation which is performed in the equivalent manner by using α-cut of fuzzy numbers and interval arithmetic. The resulting fuzzy arithmetic may be called the α-cut (fuzzy) arithmetic. In α-cut arithmetic, addition/ subtraction, multiplication/division and others may be performed at each α on the intervals of confidence by interval arithmetic.

Definition 2.1: A triangular norm (t-norm) T is an increasing associative and commutative $[0,1]^2 \to [0,1]$ mapping that satisfies the boundary condition for every $x \in [0,1]$, $T(x,1) = x$.

Some well known continuous T- norms are the minimum operator T_M, the algebraic product T_p, and Lukasiewicz t-norm T_L defined by $T_L(x,y)=\max(x+y-1,0)$. The minimum operator T_M is the strongest (greatest) t-norm. The Weakest t-norm Tw is defined by $T_w(x,y)=\begin{cases}\min(x,y), \text{if } \max(x,y)=1\\ 0\text{ , elsewhere}\end{cases}$

Corollary (Dubois [15]): Consider an L-R fuzzy interval $\tilde{A}_i=(l_i,r_i,\alpha_i,\beta_i)_{LR}$, $i=1,2,...,n$, then the Tw- sum $A=\underset{i=1}{\overset{n}{\oplus_{Tw}}} A_i$, is given by, $A=\underset{i=1}{\overset{n}{\oplus_{Tw}}} A_i=\left(\sum_{i=1}^{n} l_i,\sum_{i=1}^{n} r_i,\max_{i=1}^{n}\alpha_i,\max_{i=1}^{n}\beta_i\right)$. Notice that for the addition based on the minimum operator, the resulting spreads are the sum of the incoming spreads, while for the addition based on weakest t-norm, resulting spreads are the greatest of the incoming spreads.

Moreover, each t-norm may be shown to satisfy the following inequalities, $T_w(a,b)\le T(x,y)\le T_M(a,b)=\min(a,b)$ where $T_w(a,b)=\begin{cases}a\text{ , if } b=1\\ b\text{ , if } a=1\\ 0\text{ , otherwise}\end{cases}$

T_w is the weakest t-norm. The importance of t-norms, e.g., $\min(a,b)$, $a\bullet b$, $\max(0,a+b-1)$ *and* $T_w(a,b)$ are shown in [15,16]. Two characteristics can be observed in the previous research. First, the addition/ subtraction of fuzzy numbers by T_M and T_w preserve the original shape of the fuzzy numbers. With the T_M in the multiplication/division, the shapes of the original FNs may not be preserved. However, for given shapes, in multiplication, the T_w preserves the original FN's shape. Second, the weakest t-norm operations can elicit more exacting performance, meaning smaller fuzzy spreads within uncertain environments [16]. This exact performance may successfully reduce accumulating phenomena of fuzziness in complex systems. The addition, subtraction and multiplication of T_w fuzzy arithmetic can be seen in the following: Let $\tilde{A}=(a,\alpha_A,\beta_A)_{LR}, \tilde{B}=(b,\alpha_B,\beta_B)_{LR}$ be two L-R fuzzy numbers. The fuzzy operations of T_w can be shown as follows:

(1) Addition:

$$\tilde{A}\oplus_W \tilde{B}=(a+b,\max(\alpha_A,\alpha_B),\max(\beta_A,\beta_B))_{LR}$$

(2) Subtraction:

$$\tilde{A}(-)_W \tilde{B}=(a-b,\max(\alpha_A,\alpha_B),\max(\beta_A,\beta_B))_{LR}$$

(3) Multiplication:

$$\tilde{A}\otimes_W \tilde{B}=\begin{cases}(ab,\max(\alpha_A b,\alpha_B a),\max(\beta_A b,\beta_B a))_{LR} & for\ a,b>0\\ (ab,\max(\beta_A|b|,\beta_B|a|),\max(\alpha_A|b|,\alpha_B|a|))_{RL}, & for\ a,b<0\\ (ab,\max(\alpha_A b,\beta_B|a|),\max(\beta_A b,\alpha_B|a|))_{LL}, & for\ a<0,b>0\\ \left(0,\alpha_A b,\beta_A b\right)_{LR}, & for\ a=0,b>0\\ \left(0,-\beta_A b,-\alpha_A b\right)_{RL}, & for\ a=0,\ b<0\\ (0,0,0)_{LR}, & for\ a=0,b=0\end{cases}$$

If $\tilde{A}_1=\left(l_1,r_1,\alpha_1,\beta_1\right)_{LR}$ and $\tilde{A}_2=\left(l_2,r_2,\alpha_2,\beta_2\right)_{LR}$ be two trapezoidal fuzzy numbers, then the T_w based arithmetic operations between $\tilde{A}_1$ *and* $\tilde{A}_2$ are defined as follows:

(i) $\tilde{A}_1\oplus_w \tilde{A}_2=\left(l_1+l_2,r_1+r_2,\max(\alpha_1,\alpha_2),\max(\beta_1,\beta_2)\right)$

(ii) $\tilde{A}_1(-)_w\tilde{A}_2=\left(l_1-r_2,r_1-l_2,\max(\alpha_1,\alpha_2),\max(\beta_1,\beta_2)\right)$

(iii) $$\tilde{A}\otimes_W \tilde{B}=\begin{cases}(l_1l_2,r_1r_2,\max(\alpha_1l_2,\alpha_2l_1),\max(\beta_1r_2,\beta_2r_1))_{LR}, & for\ l_1,l_2<0\\ (l_1l_2,r_1r_2,\max(\beta_1r_2,\beta_2r_1),\max(\alpha_1l_2,\alpha_2l_1))_{LR}, & for\ l_1,l_2>0\\ (l_1l_2,\max(\alpha_1l_2,-\beta_2r_1),\max(\beta_1r_2,-\alpha_2l_1))_{RR}, & for\ l_1<0,l_2>0\\ \left(0,\alpha_1l_2,\beta_1r_2\right)_{LR}, & for\ l_1=0,l_2>0\\ \left(0,-\beta_1n_2,-\alpha_1m_2\right)_{RL}, & for\ l_1=0,\ l_2<0\\ (0,0,0)_{LR}, & for\ l_1=0,l_2=0\end{cases}$$

For example, Let $\tilde{A}=(a,\alpha_A,\beta_A)=(4,2,2), \tilde{B}=(b,\alpha_B,\beta_B)=(8,3,3)$, then the triangular fuzzy numbers are $\tilde{A}=(2,4,6), \tilde{B}=(5,8,11)$. The Table 1 shows the results of the α-cut and T_w operations. In order to observe accumulating phenomena of fuzziness, the fuzzy spread is given in Table 1, the fuzzy spread is the distance of the left to right bounds. In the numerical example, this study has found that using multiplication operation of T_w produces higher reduction rates which is 42.85 %, and the T_w operations shows the average reduction rate of 40.95%. This is an evidence that T_w based arithmetic operations can provide a more conservative evaluation for a decision maker in uncertain environments, and comparatively, α-cut arithmetic operations produce more optimistic evaluations.

Table 1: The results of α-cut and T_w operations

Operations	**α-cut (when α =1)**	**(1)Fuzzy spread (α-cut)**	**T_w**	**(2) fuzzy spread(T_w)**	**Reduction rate (%)** $=\frac{((1)-(2))}{(1)}\times 100$
+	(7,12,17)	10	(9,12,15)	6	40
_-	(-9,-4,1)	10	(-7,-4,-1)	6	40
×	(10,32,66)	56	(16,32,48)	32	42.85
Avg. of reduction cost					40.95

From the above calculation it is observed that the spread of the fuzzy set is increasing during calculation in the traditional arithmetic operation. But in the T_w based operation, the spread is controlled. From the above discussion it is concluded that T_w based multiplication can also preserve the shape of LR type fuzzy numbers**.**

3. FUZZY LINEAR REGRESSION USING THE PROPOSED APPROACH

Consider the set of observed data $\{(\tilde{X}_i, \tilde{Y}_i), i = 1,...,n\}$ where $\tilde{X}_i$ is the explanatory variable which may be with crisp/ fuzzy numbers and $\tilde{Y}_i$ is the response variable which is either with symmetric/non-symmetric triangular or trapezoidal fuzzy numbers. Our aim is to fit a fuzzy regression model with fuzzy coefficients to the aforementioned data set as follows: $\hat{\tilde{Y}}_i = \tilde{A}_0 \oplus_w (\tilde{A}_1 \otimes_w \tilde{X}_{i1}) \oplus_w ..... \oplus_w (\tilde{A}_p \otimes_w \tilde{X}_{ip}) = \tilde{A} \otimes_w \tilde{X}_i$, $\boldsymbol{i = 1,....n}$, where $\tilde{A}_j = (a_j, \alpha_j)$, $j = 1,...p$ are symmetric fuzzy numbers and the arithmetic operations are based on the weakest T_w norm. In FLR model, we are interested in finding a fuzzy function $\hat{\tilde{Y}}_i$ in the above given equation which fits a finite number of crisp input-fuzzy output data The parameters are optimized in such a way that the difference between the observed outputs $\tilde{Y}_i$ and the estimated ones $\hat{\tilde{Y}}_i$ are made as small as possible.

The most important part in an optimization problem is the selection of the objective function. The sought model could be varied based on the chosen objective function. Fuzzy linear regression analysis can be seen as an optimization problem where the aim is to derive a model which fits the given dataset. Another challenge in fuzzy regression analysis is to obtain a model which is insensitive of the outlier. Although the portion of the outlier is usually small compared to the rest of the data set, a model which fits all the data including the outliers will have an unpredictable behaviour. One of the challenges in fuzzy linear regression is that the fitness measure can be a trade secret which cannot be transparent to the fuzzy linear regression analyst. Another difficulty with the existing fuzzy linear regression analysis is that the mathematician is limited in selecting the fitness measure. For example, in least square approach the fitness measure needs to be differentiable and thus it must be continuous. The application of meta-heuristic approach relaxes these restrictions as this approach does not dictate any condition for the selection of the objective function. Therefore, the objective function is designed in such a way that it can tackle the issue of dealing with outliers.

In this section, we are discussing the objective function based on the symmetric difference along with similarity measure between two fuzzy sets, the observed and estimated values of the outputs using T_w norm, with crisp/fuzzy input-fuzzy output data, in which the coefficients of the models are also considered as fuzzy numbers.

The Least absolute deviation based on the symmetric difference between the observed and the estimated response variable is defined as follows.

$LAD = \left|\tilde{Y}_{i_L} - \hat{\tilde{Y}}_{i_L}\right| + \left|\tilde{Y}_{i_U} - \hat{\tilde{Y}}_{i_U}\right|$ where $\tilde{Y}_i$ is the observed response variable and $\hat{\tilde{Y}}_i$ is the estimated response variable. Using the T_w – norm based operations, the above distance can be defined as $LAD = \left|\left(a^T x_i + \sum_1^n \max_{1\le j\le p}\left(|a_j|\gamma_{ij}, |x_{ij}|\alpha_j\right)\right) - (y_i + e_i)\right| + \left|\left(a^T x_i - \sum_1^n \max_{1\le j\le p}\left(|a_j|\gamma_{ij}, |x_{ij}|\alpha_j\right)\right) - (y_i - e_i)\right|$. To show the fitness (performance) of the fuzzy linear regression model, we compare the fuzzy estimated response of the model $\hat{\tilde{Y}}_i$ with the observed one $\tilde{Y}_i$ where 'i' is the index of the data. There are different measures to determine the similarity between two fuzzy numbers. Here we have used the similarity measure based graded mean integration representation [10] which gives more accurate results. The similarity measure is just for comparison purposes, and any other type of objective function can be designed and then applied. $S(\tilde{Y}_i, \hat{\tilde{Y}}_i) = \frac{1}{1 + ABS(a^T x_i - y_i)}$. The similarity measure varies between 0 and 1, so the closer the value to 1, the better the model. However, for the sake of conformity, the dissimilarity measure [1-$S(\tilde{Y}_i, \hat{\tilde{Y}}_i)$] is considered in the objective function and the value is closer to 0, the better the model.

The extended objective function which combines the LAD using symmetric difference and $S(\tilde{Y}_i, \hat{\tilde{Y}}_i)$ to decrease the possibility of being trapped into local minima is given in the following form:

$$\min \sum_{i=1}^m \left(\left|\left(a^T x_i + \sum_1^n \max_{1\le j\le p}\left(|a_j|\gamma_{ij}, |x_{ij}|\alpha_j\right)\right) - (y_i + e_i)\right| + \left|\left(a^T x_i - \sum_1^n \max_{1\le j\le p}\left(|a_j|\gamma_{ij}, |x_{ij}|\alpha_j\right)\right) - (y_i - e_i)\right|\right) + \sum_{i=1}^m 1 - \left(\frac{1}{1 + abs(a^T x_i - y_i)}\right)$$

The optimization problem to find the parameters of the model

$$\min \sum_{i=1}^{m}\left(\left|\left(a^T x_i + \sum_{1}^{n} \max_{1\le j\le p}\left(\left|a_j\right|\gamma_{ij},\ \left|x_{ij}\right|\alpha_j\right)\right) - (y_i + e_i)\right| + \left|\left(a^T x_i - \sum_{1}^{n} \max_{1\le j\le p}\left(\left|a_j\right|\gamma_{ij},\ \left|x_{ij}\right|\alpha_j\right)\right) - (y_i - e_i)\right|\right) + \sum_{i=1}^{m} 1 - \left(\frac{1}{1+abs(a^T x_i - y_i)}\right)$$

subject to

$$\sum_{j=0}^{n} a_j x_{ij} + L^{-1}(h)\sum_{1}^{n}\max_{1\le j\le p}\left(\left|a_j\right|\gamma_{ij},\ \left|x_{ij}\right|\alpha_j\right) \ge y_i - L^{-1}(h)e_i$$

$$\sum_{j=0}^{n} a_j x_{ij} - L^{-1}(h)\sum_{1}^{n}\max_{1\le j\le p}\left(\left|a_j\right|\gamma_{ij},\ \left|x_{ij}\right|\alpha_j\right) \le y_i + L^{-1}(h)e_i$$

$$\max_{1\le j\le p}\left(\left|a_j\right|\gamma_{ij},\ \left|x_{ij}\right|\alpha_j\right) \ge 0, \quad \forall i = 1,2,\ldots m$$

where the variables are x_{ij} : value of the j^{th} independent variable for the i^{th} observation, yi : value of the dependent variable for the i^{th} observation, and the parameters are: ei : spread of the dependent variable for the i^{th} observation, h : target degree of belief, a_j : midpoint of the j^{th} regression coefficient, α_j : spread of the j^{th} regression coefficient, n : number of independent variables, m : number of observations. Solving this optimization problem using LINGO13.0, we can estimate the fuzzy coefficients of the model.

4. PERFORMANCE OF THE PROPOSED MODEL

To verify the performance of the proposed fuzzy linear regression method, we apply the method on crisp input - fuzzy output and fuzzy input - fuzzy output data sets from the literature and the results are compared with similarity measure [10], Goodness of fit defined by Xu and Li et al.[18] and the Mean absolute percentage error (MAPE) are used to check the performance of the proposed method.

4.1 Similarity measure based on Graded mean integration representation :

Let $\tilde{C} = (a_1, a_2, a_3)_T$ be a triangular fuzzy number, then the graded mean integration representation of $\tilde{C}$ is $P(\tilde{C}) = \dfrac{a_1 + 4a_2 + a_3}{6}$. Based on the idea of graded mean integration representation distance, Chen et al.[10] proposed a similarity measure $S(\tilde{A},\tilde{B}) = \dfrac{1}{1+d(\tilde{A},\tilde{B})}$ *where* $d(\tilde{A},\tilde{B}) = P(\tilde{A}) - P(\tilde{B})$ where $P(\tilde{A})$ and $P(\tilde{B})$ are the graded mean integration representations of $\tilde{A}$ *and* $\tilde{B}$ respectively.

4.2. Goodness of fit [18]:

Let $\tilde{A}$ *and* $\tilde{B}$ be fuzzy numbers and $\tilde{A} \circ \tilde{B} = \vee_{x\in R} [\tilde{A}(x) \wedge \tilde{B}(x)]$, $\tilde{A}\square\ \tilde{B} = \wedge_{x\in R} [\tilde{A}(x) \vee \tilde{B}(x)]$, then $(\tilde{A},\tilde{B}) = (\tilde{A} \circ \tilde{B}) \wedge (\tilde{A}\square\ \tilde{B})^c$ is called goodness of fit of $\tilde{A}$ *and* $\tilde{B}$. Let $\tilde{A}$ *and* $\tilde{B}$ be fuzzy numbers, then

(1) $0 \le (\tilde{A},\tilde{B}) \le 1$; (2) $(\tilde{A},\tilde{B})$ =1; (3) $(\tilde{A},\tilde{B}) = (\tilde{B},\tilde{A})$; (4) $\tilde{A} \subseteq \tilde{B} \subseteq \tilde{C} \Rightarrow (\tilde{A},\tilde{C}) \le (\tilde{A},\tilde{B}) \wedge (\tilde{B},\tilde{C})$

The above conditions indicate that $(\tilde{A},\tilde{B})$ is a measure of $\tilde{A}$ close up to $\tilde{B}$, and $(\tilde{A},\tilde{B})$ =1 when $\tilde{A} = \tilde{B}$

If $\tilde{A} = (a,\sigma)$ and $\tilde{B} = (b,\tau)$ be two normal fuzzy numbers, then $(\tilde{A},\tilde{B}) = \exp\left(-\left(\dfrac{a-b}{\sigma+\tau}\right)^2\right)$ is the goodness of fit of observed and estimated fuzzy numbers $\tilde{A}$ and $\tilde{B}$. The goodness of fit of observed value $\tilde{y}_i$ and the estimated value $\tilde{Y}_i = A_0 + A_1 x_{i1} + \ldots + \tilde{A}_n x_{in}$ is given by $(\tilde{y}_i, \tilde{Y}_i) = \exp\left[-\left(\dfrac{a_0 + a_1 x_{i1} + \ldots + a_{in} - y_i}{(\sigma_0 + \sigma_1 x_{i1} + \ldots + \sigma x_{in} + s_i)}\right)^2\right]$

4.3. Mean absolute percentage error (MAPE):

MAPE -it is the most popular relative error estimator method. It is estimated as follows: $MAPE = \dfrac{1}{n}\sum_{i=1}^{n}\left|\dfrac{\tilde{y}_i - \tilde{Y}_i}{\tilde{y}_i}\right|$ where observed value is $\tilde{y}_i$ and the estimated value is $\tilde{Y}_i$.

(i) When Crisp input and fuzzy output data is given

As an illustration, a part of data given in Singh et al.[19] concerned with yield of pearl Millet Crop at block levels of Bhiwani district is considered here to develop a fuzzy estimate of Pearl Millet yield. From the Table 2, it is very clear that the proposed method gives effective result in case of Crisp input and fuzzy output practical data and it is with the similarity measure of 96% with the observed response variable. The Estimated fuzzy regression model is $\tilde{Y} = (8.232, 6.741) \oplus_W (0.2447, 0) \otimes_W X$ with the confidence level of h =0.695. Table 2 shows that the proposed method has 46% of the similarity measure. Fig. 1 illustrates the behavior of the estimated fuzzy linear model for the data set given in Table 2. The estimated regression model of Singh et al. [19] method is $\tilde{Y} = (6.04, 7.53) \oplus (0.41, 0) \otimes X$. The dotted lines show the estimated fuzzy output, the vertical lines show the observed outputs and the continuous lines show the result of estimated output given by [19]. The data set using the proposed approach has a high goodness of fit value and mean similarity measure of 46% with the observed output in comparison with [19].

Table 2: Pearl Millet yield as triangular fuzzy numbers with farmer's estimates for Bhiwani District [19].

		Observed -Y		Proposed method			Comparison with [19]		
Blocks	**Farmer's estimated x**	**Lower limit of yield**	**Upper limit of yield**	**Similarity measure**	**Goodness of fit**	**MAPE**	**similarity measure**	**Goodness of fit**	**MAPE**
1	13.36	10	15	0.50025	0.988381	0.0799	0.50443	0.99046	0.0785
2	19.69	12.5	20	0.238095	0.911158	0.1969	0.31876	0.964745	0.1315
3	10.01	6	12.42	0.404531	0.978356	0.1598	0.51703	0.992466	0.1014
4	10.66	5	10.8	0.253743	0.911142	0.3722	0.28485	0.943688	0.3178
5	9.98	6.25	12.01	0.393082	0.974574	0.1691	0.49955	0.990778	0.1097
6	11.93	9.09	14.51	0.40833	0.972335	0.1150	0.37471	0.977405	0.1324
7	11.96	7.33	15.01	0.98912	0.999999	0.0009	0.81539	0.999605	0.0202
8	10.08	8.75	13.75	0.644745	0.996451	0.0489	0.48141	0.988536	0.0957
9	9.75	11.43	15.01	0.277624	0.911168	0.1968	0.23909	0.889972	0.2406
			Avg	**0.456613**	**0.960396**	**0.1488**	**0.44836**	**0.970851**	**0.1364**

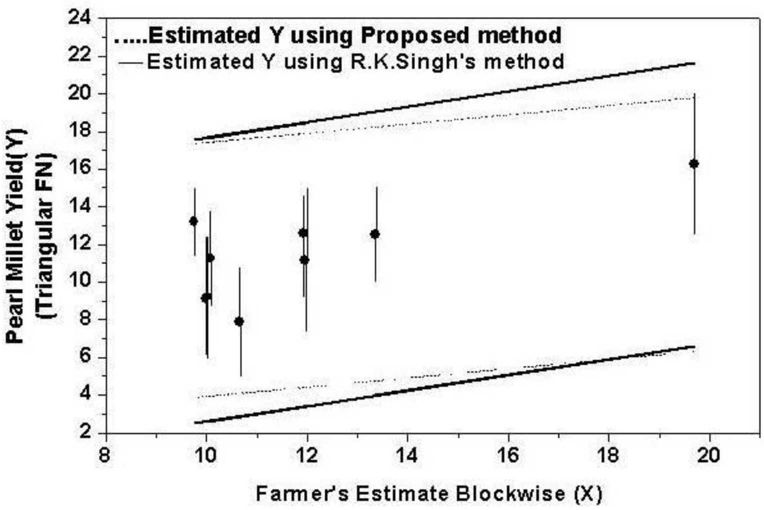

Fig. 1: Fuzzy linear regression model using the proposed method and Singh et al.[19] method for the data set in Table 2.

(ii) When Fuzzy input and fuzzy output data(both are non symmetric triangular fuzzy numbers)

Consider an example studied by Diamond [3]. In this example there are 8 pairs of $(\tilde{X}_i, \tilde{Y}_i)$ observations, as shown in Table 3, where both the response and explanatory variables are non-symmetric triangular fuzzy numbers in the form (y, α, β) where' y' is the center, α , β are the left and right spread respectively. Using the proposed approach for the data given in Table 3, the estimated fuzzy regression model is $\tilde{Y} = (1.733, 0, 0.7422) \oplus_w (0.0995, 0, 0) \otimes_w \tilde{X}$ with the confidence level of h = 0.45. Fig. 2 illustrates the behavior of the proposed approach for the non-symmetric fuzzy input and fuzzy output data set in which the dotted rectangle and solid rectangle show the estimated dependent variable and observed dependent variable values respectively. Fig. 3 illustrates the S. P. Chen et al. [20] fuzzy regression model. From the Fig. 3, it is clear that the spreads of estimated response is wider than the model by the proposed method. Among the general fuzzy models, the proposed model in which T_w- norm based arithmetic operations are used has a larger mean of similarity measure than the method proposed by S.P. Chen et al.

The least absolute regression model has more predictive ability than the least square regression model of S.P. Chen et al. The data set using the proposed approach has a high goodness of fit value and similarity measure of 82% with the observed output in comparison with [20].

Table 3: Non- symmetric triangular fuzzy number dataset taken from [20] using the proposed approach

X			**Observed -Y**			**Proposed method**		**Comparison with [20]**	
Lower	upper	center	lower	center	upper	Similarity measure	Goodness of Fit	Similarity measure	Goodness
16.8	23.10	21.00	3.4	4	4.8	0.849257	0.99901	0.768049	0.998574
12.75	17.25	15.00	2.7	3	3.3	0.81606	0.996906	0.658816	0.994336
13.5	17.25	15.00	3.15	3.5	3.85	0.78456	0.995469	0.962927	0.999968
7.65	10.35	9.00	1.6	2	2.4	0.614175	0.969385	0.631263	0.989042
10.80	13.20	12.00	2.7	3	3.45	0.931793	0.999628	0.998502	1
14.40	19.8	18.00	2.97	3.5	4.2	0.976563	0.999978	0.737191	0.997661
5.4	7.2	6.00	2.25	2.5	2.88	0.854482	0.996824	0.746269	0.99532
10.2	14.4	12.00	2	2.5	3	0.700869	0.989254	0.617284	0.99053
					Average	**0.81597**	**0.993307**	**0.765038**	**0.995679**

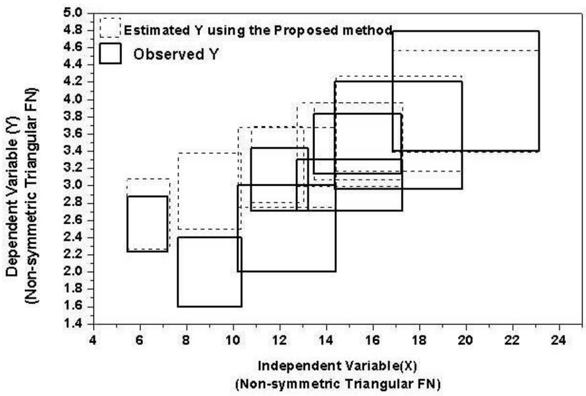

Fig. 2: The estimated fuzzy function using the proposed method for the given data in Table 3.

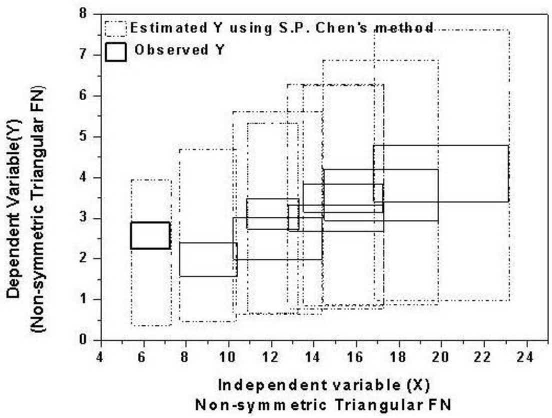

Fig. 3: The estimated fuzzy function using the S.P.Chen et al.[20] method for the given data in Table 3.

(iii) With multiple crisp input and fuzzy output data

The data set in this example given in Table 4 is related to cognitive response times of the nuclear power plant control room crew to an abnormal event and has been introduced by Kao and Chyu et al.[6]. This is a benchmark data set for the multiple linear regressions.

Table 4: Kao and Chyu et al.[6] data set using the proposed approach

Obs.	**Response time**	**Indep. variable (Inside control room experience)**	**Indep. variable(outside control room experience)**	**Indep. variable (Education)**	**Similarity measure**	**Goodness of fit**
Team 1	(5.83,3.56)	2	0	15.25	0.812942	0.999256
Team 2	(0.85,0.52)	0	5	14.13	0.215703	0.634848
Team 3	(13.93,8.5)*	1.13	1.5	14.13	0.099846	0.634827
Team 4	(4, 2.44)	2	1.25	13.63	0.479478	0.978211
Team 5	(1.65,1.01)	2.19	3.75	14.75	0.205626	0.649822
Team 6	(1.58,0.96)	0.25	3.5	13.75	0.257712	0.783691
Team 7	(8.18,4.99)	0.75	5.25	15.25	0.246676	0.9086
Team 8	(1.85,1.13)	4.25	2	13.5	0.198126	0.634831
	* indicates the outlier data			Average	**0.314514**	**0.778011**

The fuzzy regression model is obtained by the proposed approach, $\tilde{Y}=(0,4.8741)\oplus_w(0.3791,0)\otimes_w X_1\oplus_w(0,0)\otimes_w X_2\oplus_w(0.3175,0)\otimes_w X_3$ with optimum value h= 0.3259. In the above Table 4, the third data is an abnormal data. Using the proposed approach which is insensitive to the outlier yielded a better result in comparison with existing methods as per the literature, given in Table 5. The Table 5 explains that the mean similarity measure for the proposed model based on T_w- norm arithmetic operations is 0.3145 which has effective performance with other existing methods.

Table 5: Comparison of different models available in the literature for the data set in Table 4

Method	**Estimated fuzzy regression function**	**Similarity measure using** [18]
Proposed method	$\tilde{Y} = (0, 4.8755) \oplus_w (0.3791, 0) \otimes_w X_1 \oplus_w (0,0) \otimes_w X_2 \oplus_w (0.3175, 0) \otimes_w X_3$	**0.315**
Choi and Buckley et al. [11]	$\tilde{Y} = -2.8273 \oplus 0.3878 \otimes X_1 \oplus 1.0125 \otimes X_2 \oplus 0.6185 \otimes X_3 \oplus (0, 1.0696, 2.0042)$	**0.2155**
Chen and Hsueh et al.[10]	$\tilde{Y} = -16.7956 \oplus 1.0989 \otimes X_1 \oplus 1.1798 \otimes X_2 \oplus 1.8559 \otimes X_3 \oplus (0, 2.8888)$	**0.1222**
Hassanpour [22]	$\tilde{Y} = (-2.8273, 0.0000) \oplus (0.3877, 0.0000) \otimes X_1 \oplus (1.0125, 0.000) \otimes X_2 \oplus (0.6185, 0.1790) \otimes X_3$	**0.1630**
Shakouri and Nadimi et al.[23]	$\tilde{Y} = (-20.08, 0) \oplus (-0.16, 0.07) \otimes X_1 \oplus (-0.9, 0.32) \otimes X_2 \oplus (1.81, 0.15) \otimes X_3$	**0.5704**
Taheri and Kelkinnama et al. [12]	$\tilde{Y} = -15.5578 \oplus (0.2444, 0) \otimes X_1 \oplus (0.9976, 0) \otimes X_2 \oplus (1.5142, 0) \otimes X_3 + (0, 1.13)$	**0.2019**

(iv) With Trapezoidal fuzzy numbers

The proposed method with extended objective function is illustrated with trapezoidal fuzzy number given in Table 6. The fuzzy regression model obtained by the proposed method is $\tilde{Y} = (1.267, 7.166, 0, 0) \oplus_w (2.3867, 1.833, 0.2, 0) \otimes_w X$ with the confidence level h=0.4 and the proposed fuzzy regression model is given in Fig 4. The dotted lines represent the proposed fuzzy regression, the vertical lines represent the observed fuzzy trapezoidal fuzzy numbers and the continuous lines represent the estimated fuzzy regression model by Maleki et al.[25]. From the Fig.4, it is clear that the model proposed by Maleki et al.[25] has narrow spread, it could not cover all the observed responses. But the estimated response using the proposed method includes all the observed responses having 71% of goodness of fit with the observed response and a similarity measure of 36% with observed response.

Table 6: Trapezoidal fuzzy data using the proposed method and the method used by[25]

	OBSERVED				**Proposed**			Maleki et al.[25] method	
X	**LOWER**	**Y1**	**Y2**	**UPPER**	**Similarity measure**	**MAPE**	**Goodness of fit**	**Similarity measure**	**Goodness of fit**
1	7	8	9	10	0.2463	0.2401	0.4006	0.3677	0.4843
2	5.4	6.4	7.4	8.4	0.2895	0.2372	0.5938	0.3690	0.4884
3	8.5	9.5	10.5	11.5	0.5207	0.0614	0.9367	0.6944	0.9537
4	12.5	13.5	14.5	15.5	0.3372	0.0936	0.7653	0.3663	0.4802
5	12.2	13.2	14.2	15.2	0.3884	0.0766	0.8564	0.7143	0.9616
*				Average	**0.3564**	**0.1418**	**0.7106**	**0.5024**	**0.6736**

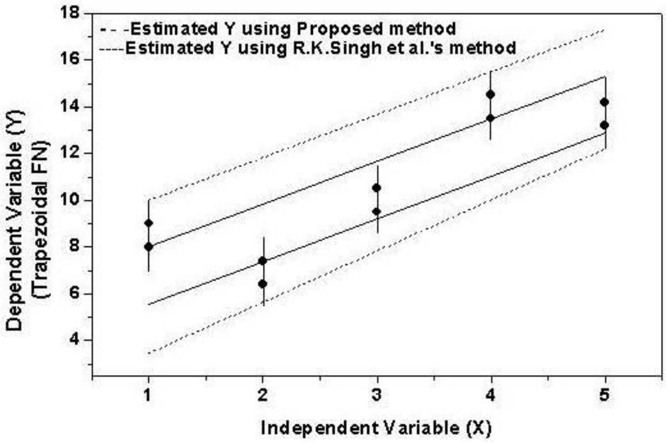

Fig. 4: The Estimated fuzzy linear regression models using the proposed method and Maleki et al.[25]method for the dataset in Table 6.

5. PROPOSED METHOD WITH OUTLIER DATASET

When outlier exists with spreads of the response variable, the treatment of outliers using Hung et al.[26] omission method and Chen [27] outlier treatment gives effective results with the dataset. In this section, examples are illustrated that the proposed method is insensitive to outlier data.

(i) Crisp input –symmetric fuzzy output:

The dataset is originally proposed by Tanaka et al.[1] and modified by Lee and Chang et al.[8] to introduce outliers. There is only one outlier in the dataset given in Table 7 and the outlier happens only in the spread of the 8th observed response. This example explains that the proposed approach is insensitive to outliers. The comparison of the results obtained from our approach to the one given by Chen [27] and Lee and Chang et al.[8] is given in Table 8. The comparison results show that the proposed model has high similarity measure, better goodness of fit and MAPE value nearest to zero. The estimated models from this work and the one proposed by Chen [27] and Lee et al.[8] are given in the Fig.5.

Table 7: The effect of the proposed approach on the Crisp input fuzzy output data set

X	**Output – Y** (Y, e)	**Similarity measure**	**Goodness of fit**	**MAPE**
1	(8.0,1.8)	0.378301	0.917736	0.205425
2	(6.4,2.2)	0.362201	0.917709	0.275141
3	(9.5,2.6)	0.682361	0.994738	0.049
4	(13.5,2.6)	0.366314	0.929735	0.128141
5	(13.0,2.4)	0.635042	0.991469	0.044208
6	(15.2,2.3)	0.847961	0.999139	0.011796
7	(17.0,2.2)	0.844666	0.999064	0.010818
8	(19.3,4.8)*	0.762486	0.998692	0.01614
9	(20.1,1.9)	0.590633	0.985369	0.034483
10	(24.3,2.0)	0.370055	0.917708	0.070053
Average		**0.584002**	**0.965136**	**0.08452**

* Outlier

Table 8: Comparison between different methods exists in the literature for the dataset given in Table 7

Approach	Model	Similarity measure	Goodness of fit	MAPE
Proposed work	$\tilde{Y} = (4.552, 3.8098) \oplus_w (1.8046, 0) \otimes_w X$	0.584002	0.965136	0.08452
Chen's [27] method	$\tilde{Y} = (4.75, 4.55) + (1.85, -0.15)X$	0.50314	0.972654	0.10085
Lee's [8] method	$\tilde{Y} = (4.43, 3.76) + (1.86, 0.14)X$	0.553184	0.968988	0.092012

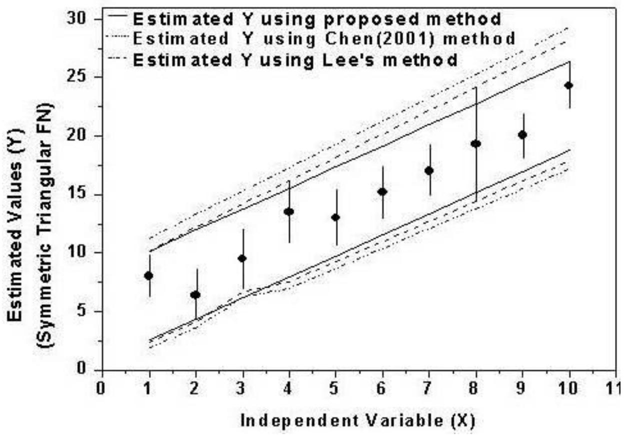

Fig. 5: Comparison between Chen [27] method, Lee and Chang et al.[8] method and our approach when outlier exists in Crisp input and fuzzy output.

From the above Fig. 5, the estimated responses have very good accuracy in results, which results the narrow spread for the estimated responses. The estimated responses have central tendency which has the mean similarity measure of 58% and the mean absolute percentage error value 0.08, which is almost nearer to zero.

(ii) Fuzzy input and fuzzy output

Sakawa and Yano [28] introduced a fuzzy input and fuzzy output data set, which is given in Table 9. Many approaches present in the literature in [2,3,6,20,28,29,30,31] have used this data without the outliers. Our approach is insensitive to outlier which illustrates the behavior of the estimated fuzzy linear model for the data set in Table 9. The dashed and continuous lines show the estimated fuzzy output using the proposed approach and the observed fuzzy outputs, respectively. From the Fig. 6, it is very clear that the proposed approach is not sensitive to outliers. The estimated fuzzy function for the above data with outlier is $\tilde{Y} = (2.7898, 0.7803) \oplus_w (0.6136, 0) \otimes_w \tilde{X}$ with h= 0.25. Fig.6 also shows that the estimated responses have the central tendency.

Table 9: Sakawa and Yano[28] Data set using the proposed approach

Obs.	Indep. variable (x, α)	Dependent variable (y, e)	Similarity measure	Goodness of fit	MAPE
1	(2.0,0.5)	(4.0,0.5)	0.983284	0.999824	0.00425
2	(3.5,0.5)	(5.5,0.5)	0.640000	0.824458	0.102273
3	(5.5,1.0)	(7.5,1.0)	0.428229	0.569795	0.178027
4	(7.0,0.5)	(6.5,0.5)	0.630915	0.811574	0.090000
5	(8.5,0.5)	(8.5,2.5)*	0.669344	0.977576	0.058118
6	(10.5,1.0)	(8.0,1.0)	0.447848	0.619038	0.154113
7	(11.0,0.5)	(10.5,0.5)	0.510152	0.5698	0.091448
8	(12.0,0.5)	(9.5,0.5)	0.510204	0.569933	0.101053
		Average	**0.602497**	**0.74275**	**0.09741**

*outlier

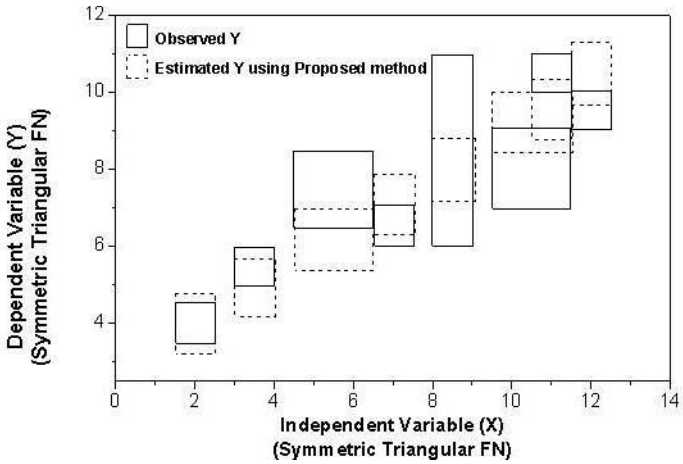

Fig. 6: Estimated fuzzy regression function for the given data set in Table 9.

(iii) when crisp input and trapezoidal fuzzy number output is given along with the outlier in the 8th data

In this example, the proposed approach on outliers with trapezoidal fuzzy numbers is discussed. The Table 10 lists the data with trapezoidal fuzzy numbers and similarity measure, MAPE and the goodness of fit of the estimated with the observed response variable using the proposed method. The proposed model for the above data is given by $\tilde{Y}_i = (1.571, 7.44, 0.594, 0) \oplus_w (1.974, 1.76, 0, 0) \otimes_w X_i$ with the confidence level h= 0.9. The proposed model has 50% mean similarity measure with the observed values and the goodness of fit value 96%.

Table 10: Crisp input and Trapezoidal output data with proposed approach taken from [25]

X	C1	C2	Lower	Upper	Similarity measure	MAPE	Goodness of fit
1	7.4	8.6	6.2	9.8	0.423855	0.169913	0.933867
2	5.6	7	4.2	8.6	0.315023	0.343321	0.868743
3	8.6	10.3	6.9	12.1	0.523963	0.095972	0.979396
4	12.6	14.3	10.9	16.1	0.449546	0.090926	0.96289
5	12.2	13.8	10.6	15.4	0.477509	0.084169	0.967253
6	14.4	15.9	12.9	17.5	0.552476	0.053409	0.981306
7	16.3	17.8	14.8	19.2	0.552374	0.047575	0.979967
8	17.7*	20.9*	14.5*	24.1*	0.708868	0.02128	0.998008
9	19.4	20.6	18.2	22	0.393025	0.07709	0.921368
10	23.6	24.9	22.3	26.3	0.549008	0.033852	0.977912
				Average	**0.494565**	**0.101751**	**0.957071**

*outlier

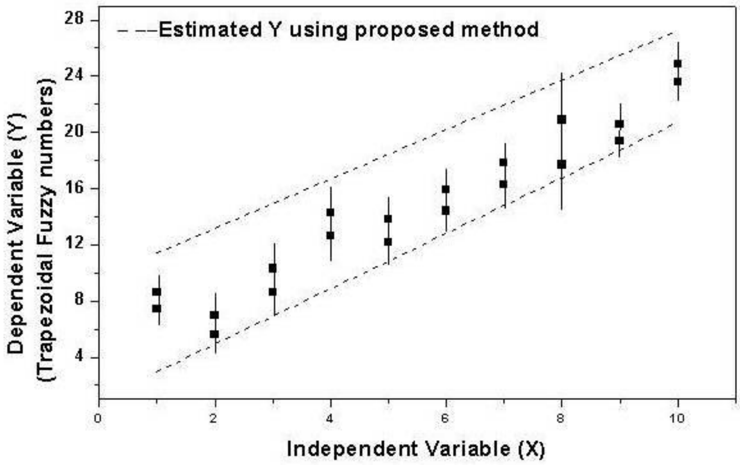

Fig. 7: The estimated fuzzy function for the given data in Table 10.

6. CONCLUSION

Fuzzy linear regression analysis is modelled as an optimization problem where the goal is to minimize the model fitting measure. In the objective function of the optimization problem for comparing the observed and estimated fuzzy responses, a similarity measure using graded mean integration representation of fuzzy numbers is considered. The proposed extended fuzzy regression model is illustrated with crisp input/ fuzzy output, non-symmetric fuzzy input and fuzzy output, data with crisp multiple inputs and also with trapezoidal fuzzy data. The results show that the model obtained by solving the optimization problem using LINGO13.0 is either superior or Pareto-equivalent to the ones in the literature. All the methods for the detection of outliers in fuzzy linear regression in the literature are the improvements on the possibilistic approaches. The goal of the possibilistic approaches is to cover the spreads of the outputs as much as possible. Using the proposed method without deleting the outlier data, the estimated fuzzy regression model is insensitive to the outlier in the case of spreads of the observed dataset. The extended objective function can be changed with different types of least absolute deviation between the observed and estimated fuzzy numbers along with the similarity measure. The future work is to get the best model which fits the negative observed data.

7. REFERENCES

[1] H. Tanaka, S. Uejima, and A. Asai, "Linear regression analysis with fuzzy model," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 12, pp. 903–907, 1982.

[2] M. M. Nasrabadi and E. Nasrabadi, "A mathematical-programming approach to fuzzy linear regression analysis", *Applied Mathematics and Computation*, vol. 55, no. 3, pp. 873–881, 2004.
[3] P. Diamond, "Fuzzy least squares", Information *Sciences*, vol. 46, no. 3, pp. 141–157, 1988.
[4] Wu B. & Tseng N. F, (2002) "A new approach to fuzzy regression models with application to business cycle analysis", Fuzzy Sets and Systems, Vol. 130, pp. 33–42.
[5] C. Kao and C.-L. Chyu, "A fuzzy linear regression model with better explanatory power," *Fuzzy Sets and Systems*, vol. 126, no. 3, pp. 401–409, 2002.
[6] C. Kao and C.-L. Chyu, "Least-squares estimates in fuzzy regression analysis", *European Journal of Operational Research*, vol. 148, no. 2, pp. 426–435, 2003
[7] B. Kim and R. R. Bishu, "Evaluation of fuzzy linear regression models by comparing membership functions," *Fuzzy Sets and Systems*, vol. 100, no. 1-3, pp. 343–352, 1998
[8] E. S. Lee and P.-T. Chang, "Fuzzy linear regression analysis with spread un- constrained in sign," *Computers and Mathematics with Applications*, vol. 28, no. 4, pp. 61–70, 1994
[9] Torabi H, & Behboodian J, (2007) "Fuzzy least absolutes estimates in linear regression models", Communi. Stat. – Theory methods, Vol. 36, pp. 1935-1944L.-H.
[10] Chen and C.-C. Hsueh, "A mathematical programming method for for- mulating a fuzzy regression model based on distance criterion," *IEEE Trans- actions on Systems, Man, and Cybernetics-part B: Cybernetics*, vol. 37, no. 3, pp. 705–712, 2007.
[11] Choi S. H. & Buckley J. J, (2008) "Fuzzy regression using least absolute deviation estimators", Soft Comput. Vol. 12, pp. 257-263.
[12] Taheri S. M. & Kelkinnama M, (2012) "Fuzzy linear regression based on least absolute deviations",Iranian J. fuzzy system Vol. 9, pp. 121-140.
[13] Hong D.H., Lee S. & Do D.Y, (2001) "Fuzzy linear regression analysis for fuzzy input-output data using shape preserving operations", Fuzzy sets Syst. Vol.122, pp. 513-526.
[14] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, pp. 338–353, 1965.
[15] D.Dubois and H. Prade, Fuzzy sets and systems: theory and applications, Academic press, Newyork , 1980
[16] D.H. Hong and H.Y Do, Fuzzy system reliability analysis by the use of (the weakest t-norm) on the fuzzy number arithmetic operations, Fuzzy sets and systems, 90(1997)307-316.
[17] Hsieh C.H. & Chen. S.H, (1999) "Similarity of generalized fuzzy numbers with graded mean integration representation", in Proceed. of 8th intn. fuzzy system association world congr., vol2, Taiwan, Republic of China, pp. 551-555.
[18] Xu R. & Li C, (2001) "Multidimensional least- squares fitting with a fuzzy model", Fuzzy sets and systems, Vol. 119, pp. 215-223.
[19] R.K. Singh, Himadri Ghosh, Possibility and necessity measure for fuzzy linear regression analysis: An Application, Journal of Indian Soc. Agricultural statistics, 62(1), 2008, pp. 19-25
[20] Shih-Pin Chen, Jr-Fong Dang , "A variable spread fuzzy linear regression model with higher explanatory power and forecasting accuracy, Information Sciences, 178 (2008) 3973–3988
[21] L.-H. Chen and C.-C. Hsueh, "Fuzzy regression models using the least squares method based on the concept of distance," *IEEE Transations on Fuzzy Systems*, vol. 17, no. 6, pp. 1259–1272, 2009.
[22] H. Hassanpour, *A Goal Programming Approach to Fuzzy Linear Regression*. PhD thesis, Shahid Bahonar University of Kerman, Kerman, Iran, 2008.
[23] H. Shakouri and R. Nadimi, "A novel fuzzy linear regression model based on a non-equality possibility index and optimum uncertainty," *Applied Soft Computing*, vol. 9, no. 2, pp. 590–598, 2009
[24] Hong D.H and C.Hwang, "Support vector Fuzzy regression machines, Fuzzy sets and systems, vol.138, No.2, pp.271-281, 2003.
[25] A.Maleki , E.Pasha and T.Razzaghnia, "Possibility Linear regression analysis with Trapezoidal fuzzy data", World applied sciences Journal, 18,(1), pp. 37-42, 2012.
[26] W.-L. Hung and M.-S. Yang, "An omission approach for detecting ouliers in fuzzy regression models," *Fuzzy Sets and Systems*, vol. 157, pp. 3109–3122, 2006.
[27] Y.-S. Chen, "Outliers detection and confidence interval modification in fuzzy regression," *Fuzzy Sets and Systems*, vol. 119, pp. 252–279, 2001
[28] M. Sakawa and H. Yano, *Fuzzy linear regression and its applications*, pp. 61–80.Fuzzy Regression Analysis, Heidelberg: Physica-Verlag, 1992.
[29] M.Hojati, C.R.Bector and K.Smimou, A simple method for computation of fuzzy linear regression, European Journal of operational Research, Vol.166, No.1, pp. 172-185, 2005.
[30] J.Lu and R. Wang, An enhanced Fuzzy linear regression model with more flexible spreads, Fuzzy sets and systems, Vol. 160, No.17, pp. 2505-2523, 2009.
[31] M.S.Yang and T.S.Lin, Fuzzy least squres linear regression analysis for fuzzy input and fuzzy output data, Fuzzy sets and systems, Vol.126, No.3, pp.389-399, 2002.