Technium Vol. 4, No.10 pp.250-273 (2022)
ISSN: 2668-778X
www.techniumscience.com

Fuzzy Regression Analysis with a proposed model

Ayşe Tansu, Zarar Naeem

Cyprus International University, Department of Industrial Engineering, Haspolat, Nicosia, Cyprus
Ayşe Tansu, ayset@ciu.edu.tr

Abstract. Regression analysis refers to methods by which estimates are made for the model parameters from the knowledge of the values of a given input-output data set. The aim of this research this research is to find a suitable model and determine the 'best' values of the parameters of the model from the given data. In the statistical regression analysis, deviations between the observed output values and corresponding values predicted by the model are attributed to random errors. It is often assumed that the distribution of these random errors is Gaussian. On the other hand, in fuzzy regression analysis the deviations (errors) are attributed to the imprecision or the vagueness of the system structure or data. The research proposed a new fuzzy linear programming model. The new proposed model is compared with the models used in the literature which are Tanaka, Hojati and Tansu regression models. The results are presented and comparison has been done for each model. Eleven different applications have been mentioned. Then the comparison of results of all the application regarding each similarity measure of goodness of fit is stated in the paper.

Keywords. Statistical, fuzzy, regression analysis, linear programming.

1. Regression Analysis

Regression analysis is a statistical tool which is used to find the relationship between two or more quantitative variable. By the basis of the relationship found between the variables one variable can be predicted from others and other variable are predicted from others. The concept of relation can be dignified into functional relation and statistical relation. When a mathematical formula can express the relation between two variables than that relation is functional. Suppose you have a particular value of the independent variable the relation will give you the corresponding relating value of the dependent variable. A statistical relation unlike a functional relation cannot give a perfect corresponding value of the dependent variable, for the given value of the dependent variable. The term regression is referred to as a description of statistical relations between variable. A statistical regression model is generally based on the following two characteristics [1]. There is a probability distribution of a tendency of the dependent variable for each level of the independent variable. The means of this probability distribution varies in some systematic fashion with the values of the independent variables.

Fuzzy linear regression was first introduced by Tanaka in 1982 [2]. That is based on possibility distribution that reflects the membership values of the dependent rather than a probability distribution. The relationship between independent and dependent variables is defined by using fuzzy concept rather than statistical concept.

Regardless of the underlying assumptions, one of the most important objectives of a regression model is to estimate the value of the dependent variable associated with independent variable as close to the observed data as possible. In a fuzzy linear regression model, the degree of the fitting of the estimated fuzzy linear model to the given data was defined by h-level of the possibility distribution. However, since the dependent variable has a membership function, the estimated fuzzy output, which is also represented by a membership function, should be close to the membership function of the given data [3]. This means that a point having a membership of 1 from the estimated fuzzy membership function should be close to the point having a membership of 1 from the given data membership function. The difference of the membership values can be used to measure the degree of the fitting of the estimated fuzzy linear model to the given data.

The research proposed a new linear programming model in the context of fuzzy logic named as New Proposed Model. The aim of this study is to develop a mathematical based model for the regression analysis and modelling by using membership functions. Linear programming (LP) models for the fuzzy linear regression have been proposed when each of the fuzzy parameter has membership function. Using membership functions instead of commonly used statistical functions leads to improvement of the objective functions of LP models [4]. Numerical examples have been done to see the difference between new proposed model and other models and some performance measures of comparison have been done between the observed and predicted intervals of data. The objective of this study is to optimize the problem while the data is fuzzy.

2. Membership Functions

Membership function of a fuzzy set is a generalization of the indicator function in classical sets. In fuzzy logic it represents the degree of truth as an extension of valuation.

A fuzzy set $\boldsymbol{\tilde{A}}$ is characterized by its membership function $\boldsymbol{\mu_{\tilde{A}}}$, which maps each element of the universe X to the interval [0, 1]. This function indicates the degree of belonging to $\tilde{A}$ for each element of X. In the notion of fuzzy set theory one of the most important concepts is the concept of α-cut. Given a fuzzy set $\boldsymbol{\tilde{A}}$ defined on X andα $\in (0,1]$, the α-cut is defined as

$$^{\alpha}\tilde{A} = \{x \in X: \mu_{\tilde{A}}(x) \geq \alpha\} \quad (1)$$

A fuzzy set $\boldsymbol{\tilde{A}}$ is convex if and only if each its α-cuts is a convex set [5].

Approximation is well suitable for representing uncertainty in options during an engineering design process, and specifically in for computer-aided engineering design and analysis tools. To breed imprecise understanding through engineering tools, however, first the membership functions must be constructed based on the understandings of the engineering design. For this purpose, measurement theory offers an axiomatically based, easy to use method. For any given variable, the best and worst values are determined, and the remaining values are assigned a degree of membership by comparison with the best and the worst [6]. On real variables, however, this would require an infinite number of questions. Instead, a continuity assumption can be made, and the remaining membership values determined through interpolation. Traditional interpolation schemes, however, fail to satisfy the restrictions of a membership function. The 0 to 1 scale bounded conditions and the fuzzy-convex property in particular present difficulty [7]. A simple and efficient constrained interpolation scheme is developed for fitting a membership function to a finite number of known membership values.

The membership function is usually a convex fuzzy number we have to use mathematics of fuzzy sets to operate on these inaccuracy descriptions in design. We can use membership data in conjunction with the usual engineering computations encountered in design process.

The process is expanded not just to map on single values but the entire inaccurate set of variables specifying a design, such as geometric lengths. We map these calculations of

Technium Vol. 4, No.10 pp.250-273 (2022)
ISSN: 2668-778X
www.techniumscience.com

dependent variable through membership function. Further, dependent variable memberships can also be back-mapped onto the variables specifying a design [8].
A membership function is usually convex and monotonic, defined as a fuzzy number using the fuzzy convex property

$$\mu(\lambda x_i + 1 - \lambda)x_j \geq \min\{\mu(x_i), \mu(x_j)\} \quad (2)$$

Where $\lambda \in [0,1]$ and x_i and x_j are real numbers.
Least squares and simple splines may not hold the convexity and monotonicity. For example, the set of elicited membership values with the fuzzy convex data. When we will use least squares and cubic splines to interpolate, the results are not fuzzy convex.
Another restriction that any interpolation method must hold is to keep the membership function bounded within [0, 1]. Although we can see least square and simple splines do not guarantee these boundaries conditions.
For example, if two known membership values are closely spaced in the area but widely separated in "$\boldsymbol{\mu}$" then overshoots tend to occur, which may force the interpolation below 0 or above 1.

Symmetric triangular membership function
Suppose the fuzzy linear regression model is:

$$\tilde{Y} = \tilde{A}_0 + \sum_{i=1}^{n} \tilde{A}_i x_i \quad (3)$$

Where the control variables x_i are assumed to be crisp. Let that the fuzzy parameter $\tilde{A}_i$ is symmetric triangular with center a_i^c and spread a_i^s for each i= 1, 2…n. Then $\tilde{Y}$ is a symmetric triangular fuzzy number with center [1]

$$\tilde{Y}_c = \sum_{i=1}^{n} a_i^c x_i \quad (4)$$

and spread

$$\tilde{Y}_s = \sum_{i=1}^{n} a_i^s |x_i|$$

The regression model is defined as, given a set of crisp data points (x_1, y_1), (x_2, y_2), (x_m, y_m), with $x_i = (x_{i1}, x_{i2}, \ldots, x_{in})$, we want to find a set of fuzzy parameters $\tilde{A}_0, \tilde{A}_1, \ldots, \tilde{A}_n$ for which the equation (1) express the best fit to the given data points, according to some criteria of goodness of fit. [2] intimate to minimize the sum spreads of the fuzzy parameters $\tilde{A}_i = i = 0,1,\ldots,n$, subject to the condition that y_i is contained in the h-cut of $\tilde{y}_i$ for each i=1,2,…,m, where $1 > h \geq 0$ is a prescribed number. It guides us to the following LP model:

(L1)

$$\min \quad z = \sum_{i=0}^{n} a_i^s \quad (5)$$

Subject to:

$$\sum_{i=0}^{n} (a_i^c x_{ji} + (1-h) a_i^s |x_{ji}|) \geq y_j$$

$$\sum_{i=0}^{n} (a_i^c x_{ji} - (1-h) a_i^s |x_{ji}|) \leq y_j$$

$a_i^s \geq 0$, a_i^c urs for all i=0, 1, 2…, n , j=1, 2… m

Here $x_{j0} = 1$ for all j=1, 2… m

With the help of [9], the author modified the objective function (2) of LP model (L1) and comes to an end of suggestion solving the LP problem:

(L2)

$$min \quad z = a_0^s + \sum_{i=1}^{n} a_i^s \left(\sum_{j=1}^{m} |x_{ji}|\right)$$

Subject to:

$$\sum_{i=0}^{n} (a_i^c x_{ji} + (1-h) a_i^s |x_{ji}|) \geq y_j$$

$$\sum_{i=0}^{n} (a_i^c x_{ji} - (1-h) a_i^s |x_{ji}|) \leq y_j$$

$a_i^s \geq 0,\ \ a_i^c\ urs$ for all i=0, 1,2…,n , j=1, 2… m

The constraints that are used in this (L2) are the same as we used in (L1) linear programming model. The change is in the objective function of the equation (5), the spreads of the control variables are weighted by the absolute values of their measurements. They wrongly point out that (5) is the sum of the absolute values of the spreads of predictors $\tilde{Y}_i (i = 1, 2, \dots, m)$. If this to be true, the coefficients of a_0^s in (5) should be m, m= number of observations. It seems more natural to take weight of a_0^s in (5) as m. Then, the linear programming model becomes:
(L3)

$$min \quad z = m a_0^s + \sum_{i=1}^{n} a_i^s \left(\sum_{j=1}^{m} |x_{ji}|\right)$$

Subject to:

$$\left.\begin{array}{l} \sum_{i=0}^{n} (a_i^c x_{ji} + (1-h) a_i^s |x_{ji}|) \geq y_j \\ \sum_{i=0}^{n} (a_i^c x_{ji} - (1-h) a_i^s |x_{ji}|) \leq y_j \end{array}\right\} , j=1, 2\dots m$$

$a_i^s \geq 0,\ \ a_i^c\ urs$ for all i=0, 1, 2…, n

Note that the constraints in LP models (L1), (L2), and (L3) are the same. Model (L3) is given in [10]. [9] has taken the influence of h value into consideration. It is observed that when the values of h is getting smaller, the values of a_0^c and z are getting smaller.
In literature symmetric triangular function is also used with the interval model. A non-linear polynomial model $\tilde{Y} = \tilde{A}_0 + \tilde{A}_1 x + \tilde{A}_2 x^2 + \cdots + \tilde{A}_n x^n$ is considered in [11], where x is the input crisp number and $\tilde{A}_i (i = 0, 1, 2, \dots, n)$ are the fuzzy parameters. The non-linear polynomial model represents the interval model. Suppose (a_i^c, a_i^s) be the center and spread of parameter A_i, respectively. By using the interval parameters A_i,a linear programming model problem is used to minimize the sum of spreads of the predicted intervals. Then model (L3) yields the linear programming model with the m observation as following:
(L3)

$$\text{Minimize} \quad z = m a_0^s + \sum_{i=1}^{n} a_i^s \left(\sum_{j=1}^{m} |x_j|^i\right)$$

Subject to:

$$\sum_{i=0}^{n} a_i^c\, x_j^i + \sum_{i=0}^{n} a_i^s\, |x_j^i| \geq y_j, \qquad j = 1, 2, \ldots\ldots, m$$

$$\sum_{i=0}^{n} a_i^c\, x_j^i - \sum_{i=0}^{n} a_i^s\, |x_j^i| \leq y_j, \qquad j = 1, 2, \ldots\ldots, m$$

$$a_i^c \quad urs, \quad a_i^s \geq 0, \quad i = 0, 1, \ldots\ldots, n$$

In [12], the fuzzy linear regression method based on fuzzy models (1) with symmetric triangular fuzzy coefficients is studied. To be precise the following three cases are studied.

- Fuzzy regression analysis for non-fuzzy data.
- Fuzzy regression analysis for non-fuzzy data with membership grades.
- Fuzzy regression analysis for fuzzy data.

In first case, the analysis leads exactly to LP model (L3).
In second case, there is a membership grade $h_p$that is assigned by human experts to each of the m non-fuzzy input-output pairs (x_p, y_p), p=1, 2, m where $x_p = (x_{p1}, \ldots, x_{pm})$ is an n-dimensional non-fuzzy input vector.
The usage of membership grade can be explained with an example. Suppose we try to analyse the characteristics feature of large cities using numerical data for each city. Since the definition of large cities is fuzzy, each city has its own membership grade for the fuzzy set "large cities". So, this is how a membership grade is utilized in fuzzy regression analysis. This leads to the LP model,
Minimize

$$z = \sum_{j=1}^{m} h_j\, [a_0^s + \sum_{i=1}^{n} a_i^s\, |x_{ji}|]$$

Subject to

$$\sum_{i=0}^{n} (\, a_i^c x_{ji} + (1-h) a_i^s |x_{ji}|) \geq y_j$$

, j=1, 2... m

$$\sum_{i=0}^{n} (\, a_i^c x_{ji} - (1-h) a_i^s |x_{ji}|) \leq y_j$$

$a_i^s \geq 0$, a_i^c urs For all i=0, 1, 2…, n
In third case, the study of fuzzy regression method for fuzzy data has been done. It is assumed that there is, m fuzzy input-output pairs ($x_p, \tilde{y}_p$), p= 1, 2, m where $x_p = (x_{p1}, \ldots, x_{pm})$ is an n-dimensional non-fuzzy input vector and $\tilde{y}_p$ is the corresponding fuzzy output. The fuzzy output $\tilde{y}_p$is symmetric triangular with centre $y_p$and spread as e_p.

The linear programming model in this case is the follow:
(L4)
Minimize

$$z = ma_0^s + \sum_{i=1}^{n} a_i^s \left(\sum_{j=1}^{m} |x_{ji}| \right)$$

Subject to

$$\left(a_0^c + \sum_{i=1}^{n} a_i^c \, x_{ji} \right) - (1-h) \left(a_0^s + \sum_{i=1}^{n} a_i^s \, |x_{ji}| \right) \leq y_p - (1-h)e_p \quad p = 1,2, \dots, m$$

$$y_p + (1-h)e_p \leq \left(a_0^c + \sum_{i=1}^{n} a_i^c \, x_{ji} \right) + (1-h) \left(a_0^s + \sum_{i=1}^{n} a_i^s \, |x_{ji}| \right), p = 1,2, \dots, m$$

$a_i^s \geq 0,\;\; a_i^c \; urs$ For all i=0, 1, 2…, n

The two inequalities represent the lower limits and upper limits of the h-level sets respectively, for given $0 \leq h \leq 1$.

[12] has proposed another model which is based on the goal programming approach [13]. Let the linear programming model be:

$$\tilde{Y} = \tilde{A}_0 + \sum_{i=1}^{n} \tilde{A}_i x_i \qquad [6]$$

Where is $\tilde{A}_i$ is a symmetric triangular fuzzy number with centre $a_i^c$and spread (or half spread)$a_i^s; i = 0,1, \dots, n$. With this also suppose that the data $(x, \tilde{y})$ available is $(x_{j1}, x_{j2} \dots, x_{jn}; y_j, e_j), j = 1,2, \dots, m$ where each response $\tilde{y}_j$is a symmetric triangular fuzzy number with centre at y_j and half spread ase_j. After this, if the h-cut of the predictor $\tilde{Y}_i$ contains the h-cut of response $\tilde{y}_i$ for each i=1, 2, m, the following constraints must be satisfied.

$$\sum_{i=0}^{n} (a_i^c x_{ji} + (1-h) a_i^s |x_{ji}|) \geq y_j + (1-h)e_j \qquad [7]$$

j=1, 2... m

$$\sum_{i=0}^{n} (a_i^c x_{ji} - (1-h) a_i^s |x_{ji}|) \leq y_j - (1-h)e_j \qquad [8]$$

j=1, 2... m

$a_i^c \geq 0,\;\; a_i^s \; urs$ For all i=0, 1, 2…, n

Where $x_{j0} = 1$ for each j and $0 \leq h \leq 1$ is a prescribed level. The goal programming approach requires that constraints (7) and (8) must be satisfied as equalities as closely as possible. This leads to the following Linear Programming problem:

(L5)

Min

$$\sum (d_{jU}^+ + d_{jU}^- + d_{jL}^+ + d_{jL}^-)$$

Subject to

$$\sum_{i=0}^{n} (a_i^c x_{ji} + (1-h) a_i^s |x_{ji}|) + d_{jU}^+ - d_{jU}^- = y_j + (1-h)e_j$$

j=1, 2... m

$$\sum_{i=0}^{n} (a_i^c x_{ji} - (1-h) a_i^s |x_{ji}|) + d_{jL}^+ - d_{jL}^- = y_j - (1-h)e_j$$

j=1, 2... m

$$d_{jU}^+, d_{jU}^-, d_{jL}^+, d_{jL}^- \geq 0 \qquad j = 1,2, \dots, m$$

$a_i^s \geq 0, \quad a_i^s \; urs$ For all i=0, 1, 2…, n

It may be remarked here that in [14]. The coefficients of a_i^s in (L5) is erroneously written as x_{ji} instead of $|x_{ji}|$.

2.1.1. Tanaka Model. Tanaka model was introduced by [2] according to him fuzziness must be taken into account in systems where human estimation is important.

$$\min z = m a_0^s + \sum_{i=1}^{n} a_i^s \left(\sum_{j=1}^{m} |x_j|^i \right)$$

Subject to

$$\sum_{i=0}^{n} a_i^c x_j^i + \sum_{i=0}^{n} a_i^s |x_j^i| \geq y_j \qquad j = 1,2, \ldots, m$$

$$\sum_{i=0}^{n} a_i^c x_j^i - \sum_{i=0}^{n} a_i^s |x_j^i| \leq y_j \qquad j = 1,2, \ldots, m$$

$$a_i^c \; urs \; a_i^s \geq 0$$

2.1.2. *Hojati Model.* He proposed a new method for computation of fuzzy regression that is simple and also gives good solution [2]. In his model he considered two different cases and these two different cases are:

- Case 1: When independent variables (*x*) are crisp numbers and response variable (*y*) is fuzzy.
- Case 2: When independent variables (*x*) are fuzzy and response variable (*y*) is also fuzzy.

The basic Hojati Model is as follows:

Case 1: When independent variables (*x*) are crisp numbers and response variable (*y*) is fuzzy. In this case we propose the following simple goal programming-like approach. Choose the fuzzy regression coefficients so that the total deviation of upper points of H-certain predicted and associated observed intervals and deviation of lower points of H-certain predicted and associated observed intervals are minimized. This can be achieved by using the following linear program:

Minimize

$$\sum_{i=1}^{n} (d_{iU}^{+} + d_{iU}^{-} + d_{iL}^{+} + d_{iL}^{-})$$

Subject to:

$$\sum_{j=0}^{k} (\alpha_j + (1-H).c_j) x_{ij} + \; d_{iU}^{+} - d_{iU}^{-} = \overline{y_i} + (1-H) e_i, \quad i = 1, \ldots, n,$$

$$\sum_{j=0}^{k}(\alpha_j-(1-H).c_j)x_{ij}+\ d_{iL}^{+}-d_{iL}^{-}=\overline{y_i}-(1-H)e_i,\quad i=1,\ldots,n,$$

$d_{iL}^{+}, d_{iL}^{-}, d_{iU}^{+}, d_{iU}^{-} \geq 0$, i = 1…*n*,
α_j = free, $c_j \geq 0$, j =0… *k*,

Note that for each i, i = 1,…,n, at most one of d_{iU}^{+} and d_{iU}^{-} will be positive and at most one of d_{iL}^{+} and d_{iL}^{-} will be positive. In fact, $|d_{iU}^{+}-d_{iU}^{-}|$ is the distance between upper point of H-certain predicted interval and the upper point of the H-certain observed interval, and $|d_{iL}^{+}-d_{iL}^{-}|$ is the distance between lower point of H-certain predicted interval and the lower point of the H-certain observed interval. The objective function is equivalent to minimizing the sum of these two distances.
If the values of the dependent variable are crisp, then it can be proved (by contradiction) that the solution to case 1 problem will also be crisp and it will be close to the solution of the Least Squares method. In this case, if a fuzzy solution is desired, one can use the approach of [2] & [14] to expand this solution to have fuzzy values.
Case 2: When independent variables (*x*) are crisp numbers and response variable (*y*) is fuzzy.
We choose the fuzzy regression coefficients such that the total deviation of upper points of predicted and associated observed intervals and deviation of lower points of predicted and associated observed intervals are minimized at both lower points ("left") and upper points ("right") of each of the independent variable values (except x_0). For simplicity, the following LP is formulated for the case when there is only one independent variables (in addition to x_0):

Minimize

$$\sum_{i=1}^{n}(d_{ilU}^{+}+\ d_{ilU}^{-}+d_{ilL}^{+}+d_{ilL}^{-}+d_{irU}^{+}+d_{irU}^{-}+d_{irL}^{+}+d_{irL}^{-})$$

Subject to

$$\sum_{j=0}^{1}(\alpha_j+(1-H).c_j)\left(\overline{x_{ij}}\ -(1-H).f_{ij}\right)+\ d_{ilU}^{+}-d_{ilU}^{-}=\overline{y_i}+\ (1-H)e_i,$$
$$i\ =\ 1,\ldots,n,$$

$$\sum_{j=0}^{1}(\alpha_j+(1-H).c_j)\left(\overline{x_{ij}}+(1-H).f_{ij}\right)+\ d_{irU}^{+}-d_{irU}^{-}=\overline{y_i}+\ (1-H)e_i,$$
$$i\ =\ 1,\ldots,n,$$

$$\sum_{j=0}^{1}(\alpha_j-(1-H).c_j)\left(\overline{x_{ij}}-(1-H).f_{ij}\right)+\ d_{ilL}^{+}-d_{ilL}^{-}=\overline{y_i}-\ (1-H)e_i,$$
$$i\ =\ 1,\ldots,n$$

$$\sum_{j=0}^{1}(\alpha_j-(1-H).c_j)\left(\overline{x_{ij}}+(1-H).f_{ij}\right)+\ d_{irL}^{+}-d_{irL}^{-}=\overline{y_i}-\ (1-H)e_i,$$
$$i\ =\ 1,\ldots,n$$

$$d_{ilU}^{+}, d_{ilU}^{-}d_{irU}^{+}, d_{irU}^{-}, d_{ilL}^{+}, d_{ilL}^{-}, d_{irL}^{-}, d_{irL}^{-}\ \geq 0,\qquad i\ =\ 1,\ldots,n$$

α_j = free, $c_j \geq 0$, j = 0, 1,

Where in the indices of the deviation variables "l" refers to the left (lower) point and "r" refers to the right (upper) point of the independent variable intervals, and "U" refers to the upper points and "L" refers to the lower points of the observed and predicted intervals.

At the end of this Hojati model we found some limitations because we are estimating the predicted band using the endpoints of the observed intervals, our approach only works when the fuzzy regression coefficients are assumed to be symmetric triangular numbers (or intervals). Thus, non- symmetric and or non-linear fuzzy numbers need a more complicated treatment.

*2.1.3 Tansu Model.*Consider the fuzzy linear regression model:

$$\tilde{Y} = \tilde{A}_0 + \tilde{A}_1\tilde{x}_1 + \tilde{A}_2\tilde{x}_2, \ldots., +\tilde{A}_n\tilde{x}_n$$

Where the control variables $x_1, x_2, \ldots., x_n$ are assumed to be crisp and the parameters $\tilde{A}_0, \tilde{A}_1, \tilde{A}_2, \ldots\ldots, \tilde{A}_n$ are assumed to be symmetric triangular with $\tilde{A}_i$ having center at a_i^c and spread a_i^s. Then the predictor $\tilde{Y}$ is a symmetric triangular fuzzy number, with center at

$$a_0^c + \sum_{i=0}^{n} a_i^c \, x_i$$

And spread as

$$a_0^s + \sum_{i=0}^{n} a_i^s \, |x_i|$$

Suppose that we are given the data $\{\left(x_{1j}, x_{2j}, \ldots., x_{nj}; \left(\bar{y}_j, e_j\right)\right), j = 1,2, \ldots., m\}$ where each observed response $\tilde{y}_j$ is taken as fuzzy triangular with center at $\bar{y}_j$ and spread ase_j. The idea behind our approach is the following:

The predictor $\tilde{Y}_j$ should be as close to its response $\tilde{y}_j$ as possible, for each$j = 1,2, \ldots., m$. Since $\tilde{Y}_j$ and $\tilde{y}_j$ are both symmetric triangular fuzzy numbers, the centers and spread of $\tilde{Y}_j$ should be close to that of$\tilde{y}_j$. This means, the differences

$$a_0^c + \sum_{i=0}^{n} a_i^c \, x_{ij} - \bar{y}_j, j = 1,2, \ldots, m$$

should be as close to zero as possible. This can be formulated as a goal programming problem named as the part 1 LP model:

Part 1 LP model:

$$\min \quad \sum_{j=1}^{m} (u_{1j} + v_{1j})$$

Subject to:

$$a_0^c + \sum_{i=1}^{n} a_i^c \, x_{ij} + v_{1j} - u_{1j} = \bar{y}_j, \qquad j = 1,2, \ldots, m$$
$$a_i^c \qquad urs, \qquad i = 0,1,2, \ldots., n$$
$$v_{1j}, u_{1j} \geq 0, \quad j = 1,2, \ldots., m$$

Similarly; we require the difference also.

$$a_0^s + \sum_{i=0}^{n} a_i^s \, |x_{ij}| - e_j, j = 1,2, \ldots, m$$

This difference in spread values must be as close to zero as possible.

This can be formulated as a goal programming formulation called part 2 LP model below:

Part 2 LP Model:

$$\min \quad \sum_{j=1}^{m} (u_{2j} + v_{2j})$$

Subject to

$$a_0^s + \sum_{i=1}^{n} a_i^s |x_{ij}| + v_{2j} - u_{2j} = e_j, \qquad j = 1,2, \dots, m$$

$$a_i^s, v_{2j}, u_{2j} \geq 0, \text{for all} \quad i = 0,1,2, \dots, n \text{ and } j = 1,2, \dots, m$$

The variables v_{ij} and $u_{ij}$are upper and lower deviation variables. Note that because the deviation variables are required to be non-negative v_{ij}only measures the difference between upper values of predicted and observed interval when the upper value of the predicted interval is smaller than the upper value of the observed interval, and u_{ij} only measures the difference between lower values of predicted and observed intervals when the lower value of the predicted interval is larger than the lower value of the observed interval. Point here to be noted is that the difference of Tansu Model from Hojati and Tanaka model is that this LP formulation model is independent of the degree of level h.

3. Applications to the Models

In this section some 11 applications are considered with New Proposed Model, Tansu Model, Hojati Model and Tanaka Model.

Application 1:

The fuzzy linear regression model $\tilde{Y} = \tilde{A}_0 + \tilde{A}_{1x}$is taken into account. Table 6 shows the data set for the application 1.

Table **Error! No text of specified style in document.** Data set for Application 1

I	x_{ij}	$\tilde{Y}_{i=}(\bar{y}_j, e_j)$	Observed interval
1	1	(8.0,1.8)	(6.2,9.8)
2	2	(6.4,2.2)	(4.2,8.6)
3	3	(9.5,2.6)	(6.9,12.1)
4	4	(13.5,2.6)	(10.9,16.1)
5	5	(13.0,2.4)	(10.6,15.4)

The values of center and spread of the function is obtained by using L1 programming model using lingo with the data set of application 1.

Figure 1 LP model for Application 1

For center and spread values of application 1 the linear programming model is:

```
Model:
Sets:
Variable/1..2/:ac,as;
ConstU/1..5/:LimitU;
ConstL/1..5/:LimitL;
Const/1..5/:limit,u,v;
CarpU(Variable,ConstU):cofU;
CarpL(Variable,ConstL):cofL;
Endsets
Min=@Sum(Const(j):u(j)+v(j));
@For(ConstU(j):@Sum(Variable(i):cofU(i,j)*ac(i)+cofU(i,j)*as(i)+u(j)-
v(j))>=LimitU(j));
@For(ConstL(j):@Sum(Variable(i):cofL(i,j)*ac(i)-cofL(i,j)*as(i)+u(j)-
v(j))<=LimitL(j));
@For(Variable(i):@Free(ac(i)));
Data:
LimitU=8.0,6.4,9.5,13.5,13.0;
LimitL=1.8,2.2,2.6,2.6,2.4;
cofU=1,1,1,1,1,
           1,2,3,4,5;
cofL=1,1,1,1,1,
```

Technium Vol. 4, No.10 pp.250-273 (2022)
ISSN: 2668-778X
www.techniumscience.com

Here ac01-ac02= a_0^c and ac11-ac12 = a_1^c
And as01-as02= a_0^s and as11-as12 = a_1^s

The values that are obtained after running through lingo software are: for center at a_0^c are 3.908 and at a_1^c is 0.991 and for spread at a_0^s are 2.258 and at a_1^s is 0.841.
These center and spread values that are obtained are used to find out $\tilde{Y}_c$ and $\tilde{Y}_s$ and these values are used for finding the predicted intervals. The new predicted intervals that are found are shown in Table 2. The measures of goodness of fit and similarities for Tansu model, Hojati, Model, Tanaka model and proposed models are also shown in Table 2

Table 2 Results for Application 1

Observation ,i	Observed Interval	Tanaka predict interval	Hojati predict interval	Tansu predict interval	Proposed model
1 2 3 4 5	(6.2,9.8) (4.2,8.6) (6.9,12.1) (10.9,16.1) (10.6,15.4)	(2.7,9.8) (4.2,11.9) (6.3,14) (8.4,16.1) (10.5,18.2)	(4,8.8) (5.65,10.45) (7.3,12.1) (8.95,13.75) (10.6,15.4)	(4,8) (5.55,9.95) (7.1,11.9) (8.65,13.85) (10.2,15.8)	(1.8,7.99) (1.95,9.83) (2.1,11.66) (2.25,13.494) (2.4,15.326)
(a) Average percentage of observed in predict		60.26%	75%	71.35%	38.85%
(b) Average percentage of predict in observed		100%	77.27%	73.67%	77.96%
(c) Average similarity measure		47.28%	50.9%	48.29%	
(d) $\sum_{i=0}^{n} a_i^s$		3.85	2.4	2	3.099%
(e) $ma_0^s + \sum_{i=0}^{n} a_i^s \left(\sum_{j=1}^{m} \lvert x_{ij} \rvert \right)$		19.25	12	12	23.905%

Technium Vol. 4, No.10 pp.250-273 (2022)
ISSN: 2668-778X
www.techniumscience.com

Application: 2

In this application there are 15 observations. This dataset shows the input and output data concerning house prices. The variables are:

y_i: i^{th} fuzzy house price (1000 yen),
x_{i1}: rank of material,
x_{i2}: first floor space (m^2)
x_{i3}: second floor space (m^2),
x_{i4}: number of rooms,
x_{i5}: number of Japanese-style rooms.

Table 3 shows the results for application 2. The observed intervals are also calculated.

Table 1 Results for Application 2

Observation, i	Observed Interval	Tanaka predict interval	Hojati predict interval	Tansu predict interval	Proposed model
1	(5510,6610)	(5244.359,7067.125)	(5349.541,5853.9)	(5211.488,5941.477)	(-854.67,6145.57)
2	(7050,7150)	(6438.066,8260.833)	(6331.473,6937.042)	(6295.768,6941.794)	(-50.85,7099.29)
3	(7680,8480)	(7678.804,9501.571)	(7725.13,8480)	(7680.001,8480.001)	(162.63,8487.17)
4	(8110,8410)	(7800.214,9622.981)	(7842.739,8647.148)	(7887.988,8632.013)	(150.66,8408.72)
5	(7900,9400)	(8814.25,10637.02)	(8488.262,9317.754)	(8260.711,9039.29)	(159.84,8649.06)
6	(8070,8970)	(8801.417,10624.18)	(8401.247,8970.001)	(8337.844,8990.353)	(-742.6,9665.90)
7	(8470,9870)	(9384.25,11207.02)	(8997.081,9645.224)	(8997.288,9649.796)	(-462.73,10261.07)
8	(10110,10510)	(10656.88,12479.64)	(10110,10867.33)	(9952.433,10667.57)	(-263.66,11157.52)
9	(10320,11520)	(11234.25,13057.02)	(10987.28,11831.8)	(11048.15,11840.77)	(-303.07,11859.83)
10	(11930,12130)	(12634.63,14457.4)	(12167.67,13122.82)	(12228.64,13017.7)	(100.78,12686.32)
11	(13590,14290)	(14104.17,15926.93)	(13433.41,14290)	(13576.81,14276.81)	(-317.99,14974.17)
12	(13590,14450)	(14771.75,16594.52)	(13950,14862.76)	(14107.3,14776.1)	(-1.03,15367.39)
13	(15710,16310)	(15858.61,17681.38)	(14870.43,15868.72)	(15026.51,15695.31)	(300.78,16008.92)
14	(15820,16820)	(16368.61,18191.38)	(15820,16820)	(15820,16820)	(-954.27,16318.85)
15	(16340,17640)	(17711.38,19534.14)	(16500.41,17640)	(16636.11,17343.89)	(650.9,17103.58)
(a) Average percentage of observed in predict		43.89%	64.92%	64.08%	5.71%
(b) Average percentage of predict in observed		100%	63.26%	57.89%	80.99%

Technium Vol. 4, No.10 pp.250-273 (2022)
ISSN: 2668-778X
www.techniumscience.com

(c) Average similarity measure		47.28%	32.87%	40.94%	
(d) $\sum_{i=0}^{n} a_i^s$		911.3834	6.740337	215.205774	1820.63
(e) $ma_0^s + \sum_{i=0}^{n} a_i^s \left(\sum_{j=1}^{m} \right.$		13670.751	6107.929381	5522.925916	88356.18

Application: 3

In this application there are 5 observations. The data set has no half-width e_i value. This value is usually approximately 2*standard deviation of y_i. The standard deviation depends on the nature of y_i. So it is chosen randomly or it is usually between $\sqrt{y_i}$ and y_i . The e_1 value is selected randomly between square root of y_i and y_i. The values of center and spread are also in the figure. Table 4 summarizes the results for application 3.

Table 2 Results for Application 3

Observation, *i*	Observed Interval	Tanaka predict interval	Hojati predict interval	Tansu predict interval	Proposed model
1 2 3 4 5	(2.14,4.94) (2.53,5057) (2.85,6.17) (1.57,3.69) (1.08,2.72)	(2.14,6.482468) (1.551863,5.894332) (1.827531,60169999) (1.57,5.912468) (1.08,5.422468)	(2.14,4.940001) (1.551863,3.6820051) (1.827531,4.257035) (1.57,3.690001) (1.08,2.720001)	(2.14,4.940001) (1.551863,3.682051) (1.827531,4.257035) (1.57,690001) (1.08,2.720001)	(1.18,5.06) (0.914,3.9) (1.06,4.58) (0.98,4.22) (0.604,2.6)
(a) Average percentage of observed in predict		59.50%	76%	76%	61.81%
(b) Average percentage of predict in observed		100%	82.40%	82.40%	77.93%
(c) Average similarity measure		39.15%	44.69%	44.69%	
(d) $\sum_{i=0}^{n} a_i^s$		2.171234	1.637951	1.637951	2.31

(e) $ma_0^s + \sum_{i=0}^{n} a_i^s \left(\sum_{j=1}^{m} \lvert x_{ij} \right.$		10.85617	5.55984784	5.55984784	7.19

Application 4:
In this application there are 5 observations. The data set pertain to the number of jobs per day and the central processing unit time required. The regression of number of jobs on central processing unit time is studied. The variables are:
x_{i1}: number of jobs per day,
y_i: fuzzy central processing unit time required.
Table 5 summarizes the results for application 4.

Table 3 Results for Application 4

Observation, *i*	Observed Interval	Tanaka predict interval	Hojati predict interval	Tansu predict interval	Proposed model
1	(1.15,2.85)	(0.36,11.72)	(1.15,2.85)	(1.15,2085)	(0.85,3)
2	(3.19,6.81)	(2.32,12.87)	(2.766666,4.9325)	(3.0675,4.9325)	(0.925,5)
3	(3,5)	(4.21,14.12)	(4.383332,7.015)	(4.985,7.015)	(1,7)
4	(6,12)	(6.10,16.83)	(5.999998,9.0975)	(6.9025,9.0975)	(1.075,9)
5	(8.82,11.18)	(8.19,19.73)	(7.616664,11.18)	(8.82,11.18)	(1.15,11)
(a) Average percentage of observed in predict		48.96%	74%	78.83%	43.356 %
(b) Average percentage of predict in observed		100%	66.12%	57.09%	78.47%
(c) Average similarity measure		38.79%	40.51%	48.5%	
(d) $\sum_{i=0}^{n} a_i^s$		3.025	0.85	0.85	1.075
(e)		16.0125	6.57917	5.075	15

Technium Vol. 4, No.10 pp.250-273 (2022)
ISSN: 2668-778X
www.techniumscience.com

$ma_0^s + \sum_{i=0}^{n} a_i^s \left(\sum_{j=}^{m} \right.$					

Application 5:
In this application there are 6 observations. The data set shows how many weeks a sample of six persons have worked at an automobile inspection station and the number of cars each one inspected between noon and 2 P.M. on a given day. The variables are;
x_{i1}: number of weeks employed,
y_i: fuzzy number of cars inspected
The results for application 5 are summarized in Table 6.

Table 4 Results for Application 5

Observation , i	Observed Interval	Tanaka predict interval	Hojati predict interval	Tansu predict interval	Proposed model
1	(10.9,15.1)	(10.31429,20.4)	(12.58636,16.68636)	(12.6,16.67273)	(2.036,15.17)
2	(18.2,23.8)	(14.45714,24.54285)	(15.5181820.11818,)	(15.6,20.03637)	(2.218,20.99)
3	(19.8,26.2)	(16.11428,26.2)	(16.69091,21.49091)	(16.8,21.38182)	(2.29,23.33)
4	(12,16)	(9.485714,19.57143)	(12,16)	(12,16)	(2,13.998)
5	(12.8,17.2)	(12.8,22.88571)	(14.34546,18.7456)	(14.4,18.69091)	(2.145,18.665)
6	(18.6,23.4)	(18.6,28.68571)	(18.45,23.55)	(18.6,23.40001)	(2.4,26.832)
(a) Average percentage of observed in predict		48.58%	67.09%	66.2048%	21.095%
(b) Average percentage of predict in observed		100%	63.44%	64.23%	75.85%
(c) Average similarity measure		30.77%	43.34%	42.6%	
(d) $\sum_{i=0}^{n} a_i^s$		5.042857	2	2	5.999
(e) $ma_0^s + \sum_{i=0}^{n} a_i^s \left(\sum_{j=1}^{m} \right.$		30.257142	13.09	13.5	52.9529

Application 6:
In this application there are 6 observations. That is x_{i1} is the tensile force applied to a steel specimen in thousands of pounds, and y_i is the resulting elongation in thousands of an inch. The regression of y_i on x_{ij} is studied.
Table 7 summarizes the results for application 6.

Table 5 Results for Application 6

Observation, *i*	Observed Interval	Tanaka predict interval	Hojati predict interval	Tansu predict interval	Proposed model
1 2 3 4 5 6	(10.5,17.5) (30,36) (36,44) (58.8,67.2 (71.5,80.5) (80,90)	(6.666674,23.19999) (21.33334,37.86666) (36.00001,52.53333) (50.66668,67.2) (65.33334,81.86666) (80.00001,96.53333)	(10.15,17.5) (25.75,33.25) (41,49) (56.25,64.75) (71.5,80.5) (86.75,96.25)	(10.5,17.5) (25.75,33.25) (41,49) (56.25,64.75) (71.5,80.5) (86.75,96.25)	(2.5,18) (3,33) (3.5,48) (4,63) (4.5,78) (5,93)
(a) Average percentage of observed in predict		48.79%	64.17%	64.17%	16.74%
(b) Average percentage of predict in observed		100%	65.83%	65.83%	78.70%
(c) Average similarity measure		29.86%	51.9%	51.9%	
(d) $\sum_{i=0}^{n} a_i^s$		8.26666	3.5	3.5	7.75
(e) $ma_0^s + \sum_{i=0}^{n} a_i^s \left(\sum_{j=1}^{m} \lvert x_{ij} \rvert \right.$		49.599996	24.75	24.75	155.25

Application 7:
In this application there are 5 observations. The data set shows the improvement (gain in reading speed) of five students in a speed-reading program, and the number of weeks they have been in the program. The variables are;
x_{i1}: number of weeks,
y_i: fuzzy speed gain (words per minute)
Table 8 shows the results for application 7.

Table 6 Results for Application 7

Observation, *i*	Observed Interval	Tanaka predict interval	Hojati predict interval	Tansu predict interval	Proposed model
1 2 3 4 5	(2.4,5.6) (7.8,12.2) (6.2,9.8) (15.5,20.5) (17.2,22.8)	(1.4,7.5) (5.55,12.25) (9.7,17) (13.85,21.75) (18,26.5)	(2.4,5.6) (6.1,9.9) (9.8,14.2) (13.5,18.5) (17.2,22.8)	(2.4,5.6) (6.1,9.9) (9.8,14.2) (13.5,18.5) (17.2,22.8)	(1.6,6) (1.7,10) (1.8,14) (1.9,18) (2.0,22)
(a) Average percentage of observed in predict		42.95%	63.10%	63.10%	33.64%
(b) Average percentage of predict in observed		100%	61.55%	61.55%	77.14%
(c) Average similarity measure		28.52%	47.41%	47.41%	
(d) $\sum_{i=0}^{n} a_i^s$		5.075	1.6	1.6	2.2
(e) $ma_0^s + \sum_{i=0}^{n} a_i^s \left(\sum_{j=1}^{m} \lvert x_{ij} \rvert \right)$		23.375	11	11	30.5

Application 8:
In this application there are 16 observations. The data on the number of twists required to break a certain kind of forged alloy bar and the percentage of two alloying elements present in the metal.
The variables are;
x_{i1}: percentage of element A,
x_{i2}: percentage of element B,
y_i: fuzzy number of twists.
Table 9 shows the result for application 8.

Technium Vol. 4, No.10 pp.250-273 (2022)
ISSN: 2668-778X
www.techniumscience.com

Table 7 Results for Application 8

Observation, *i*	Observed Interval	Tanaka predict interval	Hojati predict interval	Tansu predict interval	Proposed model
1	(36.9,45.1)	(35.3125,63.25)	(41,54.25)	(40.86667,50.46667)	(4.1,48.998)
2	(44,54)	(41.1875,69.125)	(47.41667,62.25)	(47.21111,57.01111)	(4.43,58.99)
3	(63,75)	(47.0625,75)	(53.83334,70.25)	(53.55556,63.55556)	(4.75,68.96)
4	(59.9,70.1)	(52.9375,80.875)	(60.25,78.25)	(59.9,70.1)	(5.08,78.95)
5	(34,46)	(26.875,54.8125)	(32.25,45.5)	(32.33333,44.33333)	(-4.9,39.98)
6	(42,58)	(32.75,60.6875)	(38.6667,53.5)	(38.67778,50.87778)	(-4.573,49.97)
7	(51.8,64.2)	(38.625,66.5625)	(45.08334,61.5)	(45.02222,57.42222)	(-4.243,59.96)
8	(44.5,69.5)	(44.5,72.4375)	(51.5,69.5)	(51.36667,63.96667)	(-3.913,69.95)
9	(23.5,38.5)	(18.4375,46.375)	(23.5,36.75)	(23.8,38.2)	(-15.9,28.98)
10	(31,41)	(24.3125,52.25)	(29.91667,44.75)	(30.14444,44.74444)	(-15.573,38.97)
11	(33,55)	(30.1875,58.125)	(36.33343,52.75)	(36.48889,51.28889)	(-15.243,48.96)
12	(50,64)	(36.0625,64)	(42.75,60.75)	(42.83333,57.83333)	(-14.913,58.96)
13	(10,28)	(10,37.9375)	(14.75,28)	(15.26666,32.06666)	(-22.9,21.99)
14	(26.7,35.3)	(15.875,43.8125)	(21.16667,36)	(21.61111,38.61111)	(-22.573,31.98)
15	(26.9,39.1)	(21.75,49.6875)	(27.58334,44)	(27.95555,45.15555)	(-22.243,41.97)
16	(34,52)	(27.625,55.5625)	(34,52)	(34.3,51.7)	(-21.913,51.96)
(a) Average percentage of observed in predict		50.02%	72.06%	70.4%	16.59%
(b) Average percentage of predict in observed		100%	81.16%	71.22%	74.75%
(c) Average similarity measure		41.42%	39.43%	45.54%	
(d) $\sum_{i=0}^{n} a_i^s$		13.96875	6.62499988	3.8399999	22.449
(e)		223.5	125.000008	107.999996	475.056

Technium Vol. 4, No.10 pp.250-273 (2022)
ISSN: 2668-778X
www.techniumscience.com

$ma_0^s + \sum_{i=0}^{n} a_i^s \left(\sum_{j=1}^{m} \lvert x \right.$					

Application 9:
In this application there are 5 observations. The ages and incomes of five executives working for the same company and the number of years they went to college. The variables are;
x_{i1}: age,
x_{i2}: Years College,
y_i: income (dollars)
Table 10 shows the result for application 9.

Table 8 Results for Application 9

Observation, *i*	Observed Interval	Tanaka predict interval	Hojati predict interval	Tansu predict interval	Proposed model
1	(71050,71350)	(70902.86,71907.62)	(70902.86,71268.57)	(70902.86,71268.57)	(-2105.14,71200.01)
2	(66600,67000)	(66600,67604.76)	(66600,67000)	(66600,67000)	(199.96,67485.674)
3	(74810,75190)	(74810,75814.14)	(74810,75190)	(74810,75190)	(190,75000.01)
4	(69900,70700)	(69695.24,70700)	(69695.24,70085.71)	(69695.24,70085.71)	(4.5716,70300,01)
5	(65240,65560)	(65240,66244.76)	(65240,65560)	(65240,65560)	(-7847.99,65457.15)
(a) Average percentage of observed in predict		43.79%	81.47%	81.47%	0.379%
(b) Average percentage of predict in observed		100%	79.21%	79.21%	63.57%
(c) Average similarity measure		44.58%	66.52%	66.52%	
(d) $\sum_{i=0}^{n} a_i^s$		8.26666	35.714285	35.714285	34395.29
(e) $ma_0^s + \sum_{i=0}^{n} a_i^s \left(\sum_{j=1}^{m} \lvert x_{ij} \rvert \right)$		49.599996	928.095	928.095	179500.73

Application 10:
In this application there are 8 observations. The data set shows credit card holdings of eight families, according to the family size and their incomes. The variables are;
x_{i1} : family size,
x_{i2}: family income,

y_i: fuzzy number of credit cards.
Table 11 shows the results for application 10.

Table 9 Results for Application 10

Observation, *i*	Observed Interval	Tanaka predict interval	Hojati predict interval	Tansu predict interval	Proposed model
1	(2.5,5.5)	(2.092106,7.536842)	(3.275,6.625)	(3.25,6.25)	(1.45,5.19)
2	(4,8)	(2.555264,8)	(3.625,7.075)	(3.6,6.9)	(2.01,5.99)
3	(4,8)	(3.25,8.694736)	(4.475,7.625)	(4.4625,7.5375)	(0.27,5.99)
4	(5.4,8.6)	(3.944737,9.389473)	(5,8.6)	(4.9875,8.5125)	(1.05,7.19)
5	(5.8,10.2)	(4.755263,10.2)	(5.775,9.525)	(5.76875,9.48125)	(0.72,7.99)
6	(5.45,8.55)	(5.45,10.89474)	(6.3,10.5)	(6.29375,10.45625)	(1.55,9.19)
7	(6.2,9.8)	(5.102631,10.54737)	(6.2,9.8)	(6.2,9.8)	(-0.16,7.99)
8	(7.6,12.4)	(6.9555263,12.4)	(7.6,12.4)	(7.6,12.4)	(2.044,11.19)
(a) Average percentage of observed in predict		69.10%	87.69%	88.06%	39.68%
(b) Average percentage of predict in observed		100%	85.88%	84.76%	64.99%
(c) Average similarity measure		62.07%	41.68%	66.38%	
(d) $\sum_{i=0}^{n} a_i^s$		2.722368	0.6	0.50625	0.567
(e) $ma_0^s + \sum_{i=0}^{n} a_i^s \left(\sum_{j=1}^{m} \lvert x_{ij} \rvert \right)$		21.778944	14.85	14.8575	25.958

Application 11:
In this application there are 10 observations. The data set pertain to the amount of hydrogen present in core drillings made at 1-foot intervals along the length of a vacuum-cast ingot. The variables are;
x_{i1}: core location in feet from base,
y_i: fuzzy amount of hydrogen present.
Table 12 summarizes the results for application 11.

Table 10 Results for Application 11

Observation, *i*	Observed Interval	Tanaka predict interval	Hojati predict interval	Tansu predict interval	Proposed model
1	(0.705,1.855)	(0.565,2.3025)	(0.81074,1.969286)	(1.015,2.145)	(0.51,1.53)
2	(0.78,2.28)	(0.5425,2.28)	(0.78,1.938572)	(0.955,2.085)	(0.51,1.53)
3	(0.52,1.54)	(0.52,2.2575)	(0.749286,1.907858)	(0.895,2.025)	(0.51,1.53)
4	(0.87,2.17)	(0.4975,2.235)	(0.718572,1.877144)	(0.835,1.965)	(0.51,1.53)
5	(0.59,1.65)	(0.475,2.2125)	(0.687858,1.84643)	(0.775,1.905)	(0.51,1.53)
6	(0.74,1.9)	(0.4525,2.19)	(0.657144,1.815716)	(0.715,1.845)	(0.51,1.53)
7	(0.6555,1.785)	(0.43,2.1675)	(0.62643,1.785002)	(0.655,1.785)	(0.51,1.53)
8	(0.895,2.145)	(0.4075,2.145)	(0.595716,1.754288)	(0.595,1.725)	(0.51,1.53)
9	(0.565,1.635)	(0.385,2.1225)	(0.565002,1.723574)	(0.535,1.665)	(0.51,1.53)
10	(0.52,1.54)	(0.3625,2.1)	(0.534288,1.69286)	(0.475,1.605)	(0.51,1.53)
(a) Average percentage of observed in predict		67.11%	87.20%	86.19%	82.93%
(b) Average percentage of predict in observed		100%	87.39%	84%	74.56%
(c) Average similarity measure		55.29%	58.91%	53.93%	
(d) $\sum_{i=0}^{n} a_i^s$		0.86875	0.579286	0.565	0.51
(e) $ma_0^s + \sum_{i=0}^{n} a_i^s \left(\sum_{j=1}^{m} \lvert x\right.$		49.599996	5.79286	5.65	5.1

4.Comparison of Results

In this section all four models are compared according to the (a), (b), and (d) and (e) values for 11 different applications. Table 13-Table 16 shows the comparison of four different models and every table shows the different results as (a), (b), and (d) and (e) values. The number is given to each model from 1 to 4 where 1 stands for the best result and 4 stands for worst result.

Table 11 Comparison of (a) Values

	Applications	Tanaka Model	Hojati Model	Tansu Model	Proposed Model
Data Generated from Sample	Application 1	3	1	2	4
	Application 2	3	1	2	4
	Application 3	3	1	1	2
	Application 4	3	2	1	4
	Application 5	3	1	2	4
	Application 6	2	1	1	3
	Application 7	2	1	1	3
	Application 8	3	1	2	4
	Application 9	2	1	1	3
	Application 10	3	2	1	4
	Application 11	4	1	2	3
	Total	31	13	16	38

Table 13 shows the comparison of all the four models regarding the (a) part of goodness of fit. It can been seen that proposed model does not perform well in part (a) but the results are close to that of Tanaka model and others performs well as their total sum is less and Hojati Model performs best in this case.

Table12 Comparison of (b) Values

		Tanaka Model	Hojati Model	Tansu Model	Proposed Model
Data Generated from sample	Application 1	1	3	4	2
	Application 2	1	3	4	2
	Application 3	1	2	2	3
	Application 4	1	3	4	2
	Application 5	1	4	3	2
	Application 6	1	3	3	2
	Application 7	1	3	3	2
	Application 8	1	2	4	3
	Application 9	1	2	2	3
	Application 10	1	2	3	4
	Application 11	1	2	3	4
	Total	11	29	35	29

From table14 it can be seen that proposed model perform equally to Hojati model and performs better than the Tansu model regarding (b) similarity measure and Tanaka Model performs best in this case as all the predicted interval lies inside the observed interval.

Table 13 Comparison of (d) Values

		Tanaka Model	Hojati Model	Tansu Model	Proposed Model
Data Generated from Sample	Application 1	4	2	1	3
	Application 2	3	1	2	4
	Application 3	2	1	1	3
	Application 4	3	1	1	2
	Application 5	2	1	1	3
	Application 6	3	1	1	2
	Application 7	3	1	1	2
	Application 8	3	2	1	4
	Application 9	2	1	1	3
	Application 10	4	3	1	2
	Application 11	4	3	2	1
	Total	33	17	13	29

Table 15 shows the comparison of (d) measure of goodness of fit, it can be seen that proposed model performs better than Tanaka model and Tansu model performs the best in this case.

Table 14 Comparison of (e) Values

		Tanaka Model	Hojati Model	Tansu Model	Proposed Model
Data generated from Sample	Application 1	2	1	1	3
	Application 2	3	2	1	4
	Application 3	3	1	1	2
	Application 4	4	2	1	3
	Application 5	3	1	2	4
	Application 6	2	1	1	3
	Application 7	2	1	1	3
	Application 8	3	2	1	4
	Application 9	2	1	1	3
	Application 10	3	2	1	4
	Application 11	4	3	2	1
	Total	31	17	13	34

From table 16 we can see that Tansu model performs the best in this case, proposed model is close to that of Tanaka model.

From the above results based on 11 different applications it can be seen that the proposed model performs better than other models in (b) and (d) parts of similarity measure so, it must be taken into consideration. Although Tanaka and Tansu model performs a way better throughout the analysis.

5.Conclusion

In this research, a new linear programming model is proposed for the aim of getting better results and different measures in fuzzy linear regression. Fuzzy linear regression models are also investigated in tis thesis paper. In this paper proposed model is compared with the different LP models proposed by Tanaka, Hojati and Tansu and different measures of goodness of fit are taken into account as (a) average percentage of observed in predicted interval (b) average percentage of predicted in observed interval (d) sum of spread values and (e) sum of spread of output values and these measures are based on 11 different data sets from literature as well as collected from real data. The membership function is assumed to be symmetric triangular that plays an important role in formulation of linear programming models in fuzzy regression.

It is concluded from results that not on the all-similarity measures aspects but at some points in the similarity measure the proposed model perform better than Tanaka, Hojati and Tansu model as in (b) and (d) parts. In (b) part Tanaka gave the best results since in this model the constraints make sure that the predicted interval must always lies inside the observed interval. In addition to that we can say that proposed model must be taken into consideration as it is independent of the *h*-value as the Tansu model which is advantageous to use while measuring the goodness of fit for fuzzy linear regression.

References

[1] Wong, Y. J., Arumugasamy, S. K., Chung, C. H., Selvarajoo, A., & Sethu, V. (2020). Comparative study of artificial neural network (ANN), adaptive neuro-fuzzy inference system (ANFIS) and multiple linear regression (MLR) for modeling of Cu (II) adsorption from aqueous solution using biochar derived from rambutan (Nephelium lappaceum) peel. *Environmental monitoring and assessment*, *192*(7), 1-20.

[2] Hosseinzadeh, E., & Hassanpour, H. (2021). Estimating the parameters of fuzzy linear regression model with crisp inputs and Gaussian fuzzy outputs: A goal programming approach. *Soft Computing*, *25*(4), 2719-2728.

[3] Kashani, M., Arashi, M., Rabiei, M. R., D'Urso, P., & De Giovanni, L. (2021). A fuzzy penalized regression model with variable selection. *Expert Systems with Applications*, *175*, 114696.

[4] TOPUZ, A. P. D. D. (2021). FUZZY LINEAR REGRESSION AND FUZZY PEARSON CORRELATION ANALYSIS. *DIFFERENT STATISTICAL APPLICATIONS IN AGRICULTURE*, 167.

[5] Topuz, D. (2021). Jackknife Resampling Method for Estimation of Fuzzy Regression Parameters and Revised Tanaka Method. *PROGRESS IN NUTRITION*, *23*.

[6] Kayan, G., Türker, U., & Erten, E. (2022). A fuzzy logic framework to handle uncertainty in remote sensing-based hydrological data for water budget improvement across mid-and small-scale basins. *Hydrological Processes*, *36*(11), e14740.

[7] Asai, H. T. S. U. K., Tanaka, S., & Uegima, K. (1982). Linear regression analysis with fuzzy model. *IEEE Trans. Systems Man Cybern*, *12*, 903-907.

[8] Medaglia, A. L., Fang, S. C., Nuttle, H. L., & Wilson, J. R. (2002). An efficient and flexible mechanism for constructing membership functions. *European Journal of Operational Research*, *139*(1), 84-95.

[9] Chen, J. E., & Otto, K. N. (1995). Constructing membership functions using interpolation and measurement theory. *Fuzzy Sets and systems*, *73*(3), 313-327.

[10] Yen, K. K., Ghoshray, S., & Roig, G. (1999). A linear regression model using triangular fuzzy number coefficients. *Fuzzy sets and systems*, *106*(2), 167-177.

[11] Tanaka, H., Hayashi, I., & Watada, J. (1989). Possibilistic linear regression analysis for fuzzy data. *European Journal of Operational Research*, *40*(3), 389-396.

[12] Luczynski, W., & Matloka, M. (1996). Fuzzy regression models and their applications. *JOURNAL OF FUZZY MATHEMATICS*, *4*, 513-518.

[13] Hojati, M., Bector, C. R., & Smimou, K. (2005). A simple method for computation of fuzzy linear regression. *European Journal of Operational Research*, *166*(1), 172-184.

[14] Kambo, N. S. (1984). Mathematical Programming Techniques, Affiliated East.

[15] Ishibuchi, H., & Nii, M. (2001). Fuzzy regression using asymmetric fuzzy coefficients and fuzzified neural networks. *Fuzzy Sets and Systems*, *119*(2), 273-290.