A Hybrid Method for Estimating Fuzzy Regression Parameters Data

Ali Maleki[1*]

Abstract – In this study, a hybrid fuzzy regression method for trapezoidal fuzzy data is proposed. Initially, a new definition of a weighted fuzzy arithmetic for trapezoidal fuzzy data is introduced. Subsequently, a method for hybrid fuzzy least squares linear regression based on this definition is developed. Additionally, methods for bivariate and multivariate regression models are obtained. Finally, reliability measures for the hybrid regression model using the new definition of weighted fuzzy arithmetic for trapezoidal fuzzy data are calculated.

Keywords: Trapezoidal fuzzy numbers, hybrid fuzzy regression, reliability.

1. Introduction

Fuzzy linear regression was first introduced by Tanaka et al. in 1982 [1]. They formulated linear regression models with fuzzy parameters, non-fuzzy inputs, and fuzzy outputs as linear programming problems. Their objective was to minimize the ambiguity of the fuzzy linear regression model such that the support of the estimated values would cover the support of the observed values at a certain level. Although this method was later improved by Tanaka and Watada [2] and Tanaka et al. [3], as noted by Redden and Woodall [4], it was highly sensitive to outliers. Additionally, this method could produce an infinite number of solutions, and the width of the estimated values increased with more data, limiting its practical application.

Subsequently, several improved fuzzy regression methods were proposed using the criterion of minimizing model ambiguity. Tanaka and Ishibushi [5] introduced quadratic membership functions to obtain fuzzy coefficients. Tanaka et al. [3] proposed possibility regression. In these improved methods, minimizing model ambiguity was used as the fitting criterion. The main drawback of all these methods is that the concept of least squares was not considered.

In 1988, Diamond [6] defined a metric on the set of fuzzy numbers and proposed a fuzzy least squares method for fuzzy linear regression models based on it. Diamond considered the case where both the inputs and outputs are fuzzy, and determined the fuzzy parameters such that the sum of the distances between the observed and estimated values was minimized. Diamond only discussed univariate models.

Sakawa and Yano [7] considered fuzzy linear regression models with fuzzy inputs and outputs, and formulated three multi-objective programming problems to determine the fuzzy parameters using three indices for equality between two fuzzy numbers. They used minimizing (maximizing) model ambiguity and the degree of equality between observed and estimated values as fitting criteria.

Kim and Bishu [8] proposed another method based on the criterion of minimizing the discrepancy between the membership function values of the observed and estimated values to determine fuzzy linear regression models. They also used the discrepancy between the membership function values of the observed and estimated values as a criterion for evaluating the performance of fuzzy linear regression models, demonstrating with two numerical examples that their method outperforms Tanaka et al.'s method [3].

Hong et al. [9] considered fuzzy linear regression models with fuzzy inputs and outputs and formulated a programming problem to determine the fuzzy parameters using the extension principle and the criterion of minimizing model ambiguity.

Also, in recent years, Razzaghnia et. al. [10,11,12] had studied and investigated regression methods based on ANFIS.

This paper introduces a weighted fuzzy arithmetic for trapezoidal fuzzy data. This arithmetic defines arithmetic operations between two fuzzy numbers as two corresponding values in each fuzzy set at the same membership level, merging each weighted operation level with the membership level for fuzzy sets, and dividing the

1 * **Corresponding Author :** Department of Statistics, West Tehran Branch, Islamic Azad University, Tehran, Iran. maleki.ali@wtiau.ac.ir

Received: 2024.11.30; Accepted: 2025.01.06

weighted integral by the total integral of the membership function. The weighted fuzzy arithmetic transforms the set of fuzzy values obtained from a weighted arithmetic operation into a real constant using the concept of defuzzification. The resulting value can then be interpreted as an average from a weighted fuzzy arithmetic, in contrast to ordinary fuzzy arithmetic, which represents all possible values as a fuzzy set. The next section presents the necessary definitions and concepts for the paper.

2. Preliminary Concepts and Definitions:

Definition 1: If X is a non-empty set, then a fuzzy set $\tilde{A}$ of X is specified by a set of ordered pairs as follows:

$$\tilde{A} = \{(x, \mu_{\tilde{A}}(x)) \mid x \in X)\} \tag{1}$$

or in other words, $\mu_{\tilde{A}}(x): X \to [0,1]$ where $x \in X \quad \forall x$. $\mu_{\tilde{A}}(x)$ denotes the membership degree of x in $\tilde{A}$ within the range $[0,1]$.

Definition 2: The h-level set (h-cut) of a fuzzy set $\tilde{A}$: Suppose $\tilde{A} \in F(X)$ is a fuzzy set. For any $h \in [0,1]$, the h-level set of $\tilde{A}$, denoted by $[\tilde{A}]^h$, is defined as:

$$[\tilde{A}]^h = \tilde{A}_h = \{x \in X \mid \tilde{A}(x) \geq h\} \tag{2}$$

The set $[\tilde{A}]^h = \{x \in X \mid \tilde{A}(x) > h\}$ is called the strong h-level set (strong cut).

Definition 3: Suppose $T: [0,1] \times [0,1] \to [0,1]$ is a function that satisfies the following conditions for any $a, b, c \in [0,1]$:

(i) $T(a,1) = a$ and $T(0,a) = 0$ (boundedness),

(ii) $T(a,c) \leq T(b,c)$ for $a \leq b$ (monotonicity),

(iii) $T(a,b) = T(b,a)$ (commutativity),

(iv) $T(T(a,b),c) = T(a,T(b,c))$ (associativity).

Then T is called a triangular norm (t-norm).

Definition 4: A fuzzy quantity $\tilde{A}$ is called a fuzzy number if it satisfies the following three conditions [13]:

(i) There exists at least one $x \in R$ such that $\tilde{A}(x) = 1$.

(ii) $\tilde{A}$ is upper semicontinuous.

(iii) The support of $\tilde{A}$ is bounded.

Definition 5: A fuzzy quantity $\tilde{A}$ is called a $L-R$ fuzzy number if its membership function $\tilde{A}$ is of the form:

$$\tilde{A}(x) = \begin{cases} L(\dfrac{\underline{a}^{(2)} - x}{\alpha}); & x \leq \underline{a}^{(2)}, (\alpha > 0) \\ 1 \;\; ; & \underline{a}^{(2)} \leq x \leq \overline{a}^{(2)} \\ R(\dfrac{x - \overline{a}^{(2)}}{\beta}); & x \geq \overline{a}^{(2)}, (\beta > 0) \end{cases} \tag{3}$$

where $L, R: [0, \infty) \to [0,1]$

are continuous, decreasing, and invertible on the interval $[0,1]$. $L(0) = R(0) = 1$ *and* $L(1) = R(1) = 0$. $\underline{a}^{(2)}, \overline{a}^{(2)}, \alpha, \beta$ are the centers, left spread, and right spread of the fuzzy number $\tilde{A}$, respectively. L and R are called the reference functions.

Definition 6: (Trapezoidal Fuzzy Numbers): If $L(x) = R(x) = \max(0, 1-x)$ are the reference functions, $\tilde{A}$ is a trapezoidal fuzzy number, denoted as $\tilde{A} = (a^{(1)}, a^{(2)}, a^{(3)}, a^{(4)})$.

3. Hybrid Fuzzy Least Squares Regression with Trapezoidal Fuzzy Data:

In this section, methods for obtaining hybrid fuzzy least squares linear regression using a weighted fuzzy arithmetic are presented. The weighted fuzzy arithmetic is used to formulate the sum of squared errors between the predicted and observed variables using fuzzy numbers. Initially, a bivariate regression model is obtained. Then, it is extended to a multivariate regression model, and reliability measures for hybrid fuzzy linear regression are determined. In this part, we need to calculate the weighted fuzzy arithmetic with trapezoidal data.

Definition 7: If $\tilde{A} = (a^{(1)}, a^{(2)}, a^{(3)}, a^{(4)})$ and $\tilde{B} = (b^{(1)}, b^{(2)}, b^{(3)}, b^{(4)})$ are two trapezoidal fuzzy numbers, then at the membership level μ, the intervals $\tilde{A}$ and $\tilde{B}$ are expressed as follows:

$$\mu_{\tilde{A}} = [\mu_{A_L}, \mu_{A_R}] = [ha^{(2)} + (1-h)a^{(1)}, ha^{(3)} + (1-h)a^{(4)}] \tag{4}$$

$$\mu_{\tilde{B}} = [\mu_{B_L}, \mu_{B_R}] = [hb^{(2)} + (1-h)b^{(1)}, hb^{(3)} + (1-h)b^{(4)}] \tag{5}$$

In this case, the sum of two trapezoidal fuzzy numbers with the weighted operator is defined as:

$$\tilde{A} + \tilde{B} = \frac{([\int_h (\mu_{A_L} + \mu_{B_L})hdh]_L + [\int_h (\mu_{A_R} + \mu_{B_R})hdh]_R)}{\int_h hdh} \tag{6}$$

where in the denominator we have:

$$\int hdh = 2\int_0^1 hdh = 1 \tag{7}$$

By $\mu_{A_L}, \mu_{A_R}, \mu_{B_L}$ and μ_{B_R} into formula (6), we have:

$$\begin{aligned}[\int_h (\mu_{A_L} + \mu_{B_L})hdh]_L &= \int_0^1 \{[ha^{(2)} + (1-h)a^{(1)}] \\ &+[hb^{(2)} + (1-h)b^{(1)}]\}hdh = \int_0^1 [(a^{(1)} + b^{(1)}) + \\ &h(a^{(2)} - a^{(1)} + b^{(2)} - b^{(1)})]hdh \\ &= \frac{1}{6}(a^{(1)} + b^{(1)}) + \frac{1}{3}(a^{(2)} + b^{(2)})\end{aligned} \tag{8}$$

and similarly,

$$\begin{aligned}[\int_h (\mu_{A_R} + \mu_{B_R})hdh]_L &= \int_0^1 \{[ha^{(3)} + (1-h)a^{(4)}] \\ &+[hb^{(3)} + (1-h)b^{(4)}]\}hdh \\ &= \int_0^1 [(a^{(4)} + b^{(4)}) + h(a^{(3)} - a^{(4)} + b^{(3)} - b^{(4)})]hdh \\ &= \frac{1}{6}(a^{(4)} + b^{(4)}) + \frac{1}{3}(a^{(3)} + b^{(3)})\end{aligned} \tag{9}$$

By adding two equations (8) and (9), the formula for the weighted sum of two trapezoidal fuzzy numbers is obtained as follows:

$$\tilde{A} + \tilde{B} = \frac{1}{6}(a^{(1)} + a^{(4)} + b^{(1)} + b^{(4)}) + \frac{1}{3}(a^{(2)} + a^{(3)} + b^{(2)} + b^{(3)}) \tag{10}$$

Additionally, the difference between two weighted fuzzy numbers $\tilde{A}$ and $\tilde{B}$ is defined as follows:

$$\tilde{A} - \tilde{B} = \frac{([\int_h (\mu_{A_L} - \mu_{B_L})hdh]_L + [\int_h (\mu_{A_R} - \mu_{B_R})hdh]_R)}{\int_h hdh} \tag{11}$$

By substituting $\mu_{A_L}, \mu_{A_R}, \mu_{B_L}$ and μ_{B_R} into equation (11), the following relationship is obtained:

$$\tilde{A} - \tilde{B} = \frac{1}{3}(a^{(2)} - b^{(2)} + a^{(3)} - b^{(3)}) + \frac{1}{6}(a^{(1)} - b^{(1)} + a^{(4)} - b^{(4)}) \tag{12}$$

Additionally, the multiplication and division of two weighted fuzzy numbers $\tilde{A}$ and $\tilde{B}$ are calculated as follows:

$$\begin{aligned}\tilde{A}\tilde{B} &= \frac{1}{4}(a^{(2)}b^{(2)} + a^{(3)}b^{(3)}) + \frac{1}{12}(b^{(1)} + b^{(4)} + a^{(1)}b^{(2)} \\ &+a^{(4)}b^{(3)} + a^{(1)}b^{(1)} + a^{(4)}b^{(4)}) \\ \frac{\tilde{A}}{\tilde{B}} &= \int_0^1 \frac{ha^{(2)} + (1-h)a^{(1)}}{hb^{(2)} + (1-h)b^{(1)}} hdh + \\ &\int_0^1 \frac{ha^{(3)} + (1-h)a^{(4)}}{hb^{(3)} + (1-h)b^{(4)}} hdh\end{aligned} \tag{13}$$

On the condition that $\tilde{B}$ in the denominator of formula (13) does not become zero.

3.1 Bivariate Regression Model:

In this section, we use asymmetric trapezoidal fuzzy numbers to present a hybrid linear regression model. The use of trapezoidal fuzzy numbers reduces the ambiguity of the model and decreases the regression error (for more details, refer to the works of Razzaghnia et al. [10, 11, 12]).

A bivariate regression model is expressed as follows:

$$\hat{\tilde{Y}}_i = \tilde{A}_0 + \tilde{A}_1 X_i = (a_0^{(1)}, a_0^{(2)}, a_0^{(3)}, a_0^{(4)}) + (a_1^{(1)}, a_1^{(2)}, a_1^{(3)}, a_1^{(4)})X \tag{14}$$

where X_i is the independent variable with a crisp value, and $\tilde{A}_0$ and $\tilde{A}_1$ are trapezoidal fuzzy parameters. $\hat{\tilde{Y}}_i$ is the predicted value of the dependent variable, which is given by:

$$\widehat{\widetilde{Y}}_i = (a_0^{(1)} + a_1^{(1)} X_i, a_0^{(2)} + a_1^{(2)} X_i, a_0^{(3)} + a_1^{(3)} X_i, a_0^{(4)} + a_1^{(4)} X_i) \quad (15)$$

Let $i = 1, 2, ..., n$ where n is the sample size. Each observed value of the trapezoidal fuzzy dependent variable is represented as $\widetilde{Y}_i = (Y_i^{(1)}, Y_i^{(2)}, Y_i^{(3)}, Y_i^{(4)})$. Here, $\widehat{\widetilde{Y}}_{i,L}$ and $\widehat{\widetilde{Y}}_{i,R}$ denote the left and right bounds of the predicted $\widehat{\widetilde{Y}}_i$ at level h, and $\widetilde{Y}_{i,L}$ and $\widetilde{Y}_{i,R}$ denote the left and right bounds of the observed $\widetilde{Y}_i$ at level h. These quantities are expressed as follows:

$$\begin{aligned}
\widehat{\widetilde{Y}}_{i,L} &= ha_0^{(2)} + ha_1^{(2)} X_i + (1-h)a_0^{(1)} + (1-h)a_1^{(1)} X_i \\
\widehat{\widetilde{Y}}_{i,R} &= ha_0^{(3)} + ha_1^{(3)} X_i + (1-h)a_0^{(4)} + (1-h)a_1^{(4)} X_i \\
\widetilde{Y}_{i,L} &= hY_i^{(2)} + (1-h)Y_i^{(1)} \\
\widetilde{Y}_{i,R} &= hY_i^{(3)} + (1-h)Y_i^{(4)}
\end{aligned} \quad (16)$$

Using the definition of the weighted fuzzy operator, the sum of the squared residual errors between the predicted $\widehat{\widetilde{Y}}_i$ and the observed $\widetilde{Y}_i$ is calculated as follows:

$$\begin{aligned}
&\sum (residualerrors)^2 = \sum_{i=1}^{n} (\widehat{\widetilde{Y}}_i - \widetilde{Y}_i)^2 \\
&= \sum_{i=1}^{n} \frac{[\int_0^1 (\widehat{\widetilde{Y}}_{i,L} - \widetilde{Y}_{i,L})^2 hdh]_L + [\int_0^1 (\widehat{\widetilde{Y}}_{i,R} - \widetilde{Y}_{i,R})^2 hdh]_R}{\int_h hdh}
\end{aligned} \quad (17)$$

The denominator is the integral of the membership function, which is $\int hdh = 1$. Also:

$$\begin{aligned}
&[\int_0^1 (\widehat{\widetilde{Y}}_{i,L} - \widetilde{Y}_{i,L})^2 hdh]_L \\
&= \int_0^1 [h(a_0^{(2)} + a_1^{(2)} X_i - Y_i^{(2)}) + (1-h)(a_0^{(1)} + a_1^{(1)} X_i - Y_i^{(1)})]^2 hdh \\
&= \int_0^1 [h^2 (a_0^{(2)} + a_1^{(2)} X_i - Y_i^{(2)})^2 + 2h(1-h)(a_0^{(2)} \\
&+ a_1^{(2)} X_i - Y_i^{(2)})(a_0^{(1)} + a_1^{(1)} X_i - Y_i^{(1)}) \\
&+ (1-h)^2 (a_0^{(1)} + a_1^{(1)} X_i - Y_i^{(1)})^2]hdh \\
&= \frac{1}{4}(a_0^{(2)} + a_1^{(2)} X_i - Y_i^{(2)})^2 \\
&+ \frac{1}{6}(a_0^{(2)} + a_1^{(2)} X_i - Y_i^{(2)})(a_0^{(1)} + a_1^{(1)} X_i - Y_i^{(1)}) \\
&+ \frac{1}{12}(a_0^{(1)} + a_1^{(1)} X_i - Y_i^{(1)})^2
\end{aligned}$$

Similarly:

$$\begin{aligned}
&[\int_0^1 (\widehat{\widetilde{Y}}_{i,R} - \widetilde{Y}_{i,R})^2 hdh]_L \\
&= \int_0^1 [h(a_0^{(3)} + a_1^{(3)} X_i - Y_i^{(3)}) + (1-h)(a_0^{(4)} + a_1^{(4)} X_i - Y_i^{(4)})]^2 hdh \\
&= \int_0^1 [h^2 (a_0^{(3)} + a_1^{(3)} X_i - Y_i^{(3)})^2 \\
&+ 2h(1-h)(a_0^{(3)} + a_1^{(3)} X_i - Y_i^{(3)})(a_0^{(4)} + a_1^{(4)} X_i - Y_i^{(4)}) \\
&+ (1-h)^2 (a_0^{(4)} + a_1^{(4)} X_i - Y_i^{(4)})^2]hdh \\
&= \frac{1}{4}(a_0^{(3)} + a_1^{(3)} X_i - Y_i^{(3)})^2 \\
&+ \frac{1}{6}(a_0^{(3)} + a_1^{(3)} X_i - Y_i^{(3)})(a_0^{(4)} + a_1^{(4)} X_i - Y_i^{(4)}) \\
&+ \frac{1}{12}(a_0^{(4)} + a_1^{(4)} X_i - Y_i^{(4)})^2
\end{aligned}$$

Therefore:

$$\begin{aligned}
R &= \sum (residualerrors)^2 \\
&= \sum_{i=1}^{4} [\frac{1}{4}(a_0^{(2)} + a_1^{(2)} X_i - Y_i^{(2)})^2 \\
&+ \frac{1}{6}(a_0^{(2)} + a_1^{(2)} X_i - Y_i^{(2)})(a_0^{(1)} + a_1^{(1)} X_i - Y_i^{(1)}) \\
&+ \frac{1}{12}(a_0^{(1)} + a_1^{(1)} X_i - Y_i^{(1)})^2 \\
&+ \frac{1}{4}(a_0^{(3)} + a_1^{(3)} X_i - Y_i^{(3)})^2 \\
&+ \frac{1}{6}(a_0^{(3)} + a_1^{(3)} X_i - Y_i^{(3)})(a_0^{(4)} + a_1^{(4)} X_i - Y_i^{(4)})
\end{aligned} \quad (18)$$

Equation (18) contains 8 unknown parameters. To derive formulas for the unknown regression coefficients based on minimizing the error, we differentiate equation (18) with respect to the 8 unknown parameters and set the derivatives to zero:

$$\frac{\partial R}{\partial a_0^{(1)}}=\frac{1}{6}\sum_{i=1}^{n}(a_0^{(2)}+a_1^{(2)}X_i-Y_i^{(2)})$$
$$+\frac{1}{6}\sum_{i=1}^{n}(a_0^{(1)}+a_1^{(1)}X_i-Y_i^{(1)})=0$$

$$\frac{\partial R}{\partial a_1^{(1)}}=\frac{1}{6}\sum_{i=1}^{n}(a_0^{(2)}+a_1^{(2)}X_i-Y_i^{(2)})X_i$$
$$+\frac{1}{6}\sum_{i=1}^{n}(a_0^{(1)}+a_1^{(1)}X_i-Y_i^{(1)})X_i=0$$

$$\frac{\partial R}{\partial a_0^{(2)}}=\frac{1}{2}\sum_{i=1}^{n}(a_0^{(2)}+a_1^{(2)}X_i-Y_i^{(2)})$$
$$+\frac{1}{6}\sum_{i=1}^{n}(a_0^{(1)}+a_1^{(1)}X_i-Y_i^{(1)})=0$$

$$\frac{\partial R}{\partial a_1^{(2)}}=\frac{1}{2}\sum_{i=1}^{n}(a_0^{(2)}+a_1^{(2)}X_i-Y_i^{(2)})X_i$$
$$+\frac{1}{6}\sum_{i=1}^{n}(a_0^{(1)}+a_1^{(1)}X_i-Y_i^{(1)})X_i=0$$

$$\frac{\partial R}{\partial a_0^{(3)}}=\frac{1}{2}\sum_{i=1}^{n}(a_0^{(3)}+a_1^{(3)}X_i-Y_i^{(3)})$$
$$+\frac{1}{6}\sum_{i=1}^{n}(a_0^{(4)}+a_1^{(4)}X_i-Y_i^{(4)})=0$$

$$\frac{\partial R}{\partial a_1^{(3)}}=\frac{1}{2}\sum_{i=1}^{n}(a_0^{(3)}+a_1^{(3)}X_i-Y_i^{(3)})X_i$$
$$+\frac{1}{6}\sum_{i=1}^{n}(a_0^{(4)}+a_1^{(4)}X_i-Y_i^{(4)})X_i=0$$

$$\frac{\partial R}{\partial a_0^{(4)}}=\frac{1}{6}\sum_{i=1}^{n}(a_0^{(4)}+a_1^{(4)}X_i-Y_i^{(4)})$$
$$+\frac{1}{6}\sum_{i=1}^{n}(a_0^{(3)}+a_1^{(3)}X_i-Y_i^{(3)})=0$$

$$\frac{\partial R}{\partial a_1^{(4)}}=\frac{1}{6}\sum_{i=1}^{n}(a_0^{(4)}+a_1^{(4)}X_i-Y_i^{(4)})X_i$$
$$+\frac{1}{6}\sum_{i=1}^{n}(a_0^{(3)}+a_1^{(3)}X_i-Y_i^{(3)})X_i=0$$

These equations are formulated as follows to obtain the unknown coefficients:

$$\begin{cases} na_0^{(1)}+\left(\sum_{i=1}^{n}X_i\right)a_1^{(1)}=\sum_{i=1}^{n}y_i^{(1)} \\ \left(\sum_{i=1}^{n}X_i\right)a_0^{(1)}+\left(\sum_{i=1}^{n}X_i^2\right)a_1^{(1)}=\sum_{i=1}^{n}X_iy_i^{(1)} \end{cases} \quad (19)$$

By solving the system(19), the parameters $a_0^{(1)}$ and $a_1^{(1)}$ are calculated,

$$\begin{cases} na_0^{(2)}+\left(\sum_{i=1}^{n}X_i\right)a_1^{(2)}=\sum_{i=1}^{n}y_i^{(2)} \\ \left(\sum_{i=1}^{n}X_i\right)a_0^{(2)}+\left(\sum_{i=1}^{n}X_i^2\right)a_1^{(2)}=\sum_{i=1}^{n}X_iy_i^{(2)} \end{cases} \quad (20)$$

By solving the system(20), the parameters $a_0^{(2)}$ and $a_1^{(2)}$ are calculated,

$$\begin{cases} na_0^{(3)}+\left(\sum_{i=1}^{n}X_i\right)a_1^{(3)}=\sum_{i=1}^{n}y_i^{(3)} \\ \left(\sum_{i=1}^{n}X_i\right)a_0^{(3)}+\left(\sum_{i=1}^{n}X_i^2\right)a_1^{(3)}=\sum_{i=1}^{n}X_iy_i^{(3)} \end{cases} \quad (21)$$

By solving the system(21), the parameters $a_0^{(3)}$ and $a_1^{(3)}$ are calculated,

$$\begin{cases} na_0^{(4)}+\left(\sum_{i=1}^{n}X_i\right)a_1^{(4)}=\sum_{i=1}^{n}y_i^{(4)} \\ \left(\sum_{i=1}^{n}X_i\right)a_0^{(4)}+\left(\sum_{i=1}^{n}X_i^2\right)a_1^{(4)}=\sum_{i=1}^{n}X_iy_i^{(4)} \end{cases} \quad (22)$$

By solving the system (22), the parameters $a_0^{(4)}$ and $a_1^{(4)}$ are calculated,

3.2 Multivariate Regression Model:

In this section, the bivariate regression model is generalized to a multivariate regression model. The hybrid multivariate regression model includes multiple predictor variables and is expressed as follows:

$$\tilde{\tilde{Y}}_i=\tilde{A}_0+\tilde{A}_1X_{i1}+\tilde{A}_2X_{i2}+\ldots+\tilde{A}_pX_{ip} \quad (23)$$

where $\tilde{A}_j = (a_j^{(1)}, a_j^{(2)}, a_j^{(3)}, a_j^{(4)})$ and $i = 1, 2, ..., n$ and $j = 1, 2, ..., p$ represent the sample size and the number of parameters, respectively.

Using the hybrid fuzzy univariate regression model obtained in the previous section, the normal equations for the multivariate case are expressed as follows:

Normal equations to obtain $a_0^{(k)}, a_1^{(k)}, ..., a_p^{(k)}$, where $k = 1, 2, 3, 4$

For a set of data $\left(X_{ij}, \tilde{Y}_i\right)$, the normal equations for the trapezoidal fuzzy coefficients are as follows:

$$\begin{aligned}
&na_0^{(k)} + \left(\sum_{i=1}^{n} X_{i1}\right) a_1^{(k)} + ... + \left(\sum_{i=1}^{n} X_{ip}\right) a_p^{(k)} = \sum_{i=1}^{n} y_i^{(p)} \\
&\left(\sum_{i=1}^{n} X_{i1}\right) a_0^{(k)} + \left(\sum_{i=1}^{n} X_{i1}^2\right) a_1^{(k)} + ... + \left(\sum_{i=1}^{n} X_{i1} X_{ip}\right) a_p^{(k)} = \sum_{i=1}^{n} X_i y_i^{(k)} \\
&\quad\vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\
&\left(\sum_{i=1}^{n} X_{ip}\right) a_0^{(k)} + \left(\sum_{i=1}^{n} X_{ip} X_{i1}\right) a_1^{(k)} + ... + \left(\sum_{i=1}^{n} X_{ip}^2\right) a_p^{(k)} = \sum_{i=1}^{n} X_i y_i^{(k)}
\end{aligned} \tag{24}$$

These normal equations can be solved to obtain the coefficients $a_0^{(k)}, a_1^{(k)}, ..., a_p^{(k)}$ for $k = 1, 2, 3, 4$.

To further elucidate the proposed methodology, we have introduced a comprehensive framework that leverages the unique properties of weighted fuzzy arithmetic for trapezoidal fuzzy data. Specifically, this section delineates the step-by-step process of applying the weighted fuzzy arithmetic to formulate the sum of squared errors between predicted and observed variables. By incorporating this methodology into both bivariate and multivariate regression models, we not only enhance the robustness of the regression analysis but also significantly reduce model ambiguity and regression error. This meticulous approach ensures that the proposed hybrid fuzzy regression models are both reliable and accurate, thereby substantiating the practical applicability of our method in various real-world scenarios.

4. Reliability Measures for the Hybrid Regression Model:

After obtaining the regression coefficients, we calculate the reliability of the hybrid regression equation in this section. The mean of the observed fuzzy numbers $\overline{\tilde{Y}_i}$ is calculated as a constant using the weighted fuzzy operator and is given by:

$$\left(\overline{\tilde{Y}_i}\right) = \frac{\sum_{i=1}^{n} \tilde{Y}_i}{n} = \frac{1}{n}\left[\frac{1}{3}\sum_{i=1}^{n}\left(y_i^{(2)} + y_i^{(3)}\right) + \frac{1}{6}\sum_{i=1}^{n}\left(y_i^{(1)} + y_i^{(4)}\right)\right] \tag{25}$$

To evaluate reliability, we need to calculate the hybrid standard error $\left(HS_e\right)$ and the dispersion of the hybrid regression data $\left(S_{\tilde{y}}\right)$. The hybrid standard error $\left(HS_e\right)$ indicates the goodness of fit between the hybrid regression model and the observed fuzzy values, while $\left(S_{\tilde{y}}\right)$ shows the data dispersion. Two indices, $\left(HS_e\right)$ and $\left(\frac{HS_e}{S_{\tilde{y}}}\right)$, are used to evaluate the performance of different regression models [14].

$$\begin{aligned}
S_{\tilde{y}} &= \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(\tilde{Y}_i - \overline{\tilde{Y}}\right)^2} \\
&= \sqrt{\frac{1}{n-1}}\left\{\frac{1}{2}\sum_{i=1}^{n}\left(y_i^{(1)} - \overline{\tilde{Y}}\right)^2 + \frac{2}{3}\sum_{i=1}^{n}\left(y_i^{(1)} - \overline{\tilde{Y}}\right)\left(y_i^{(2)} - y_i^{(1)}\right)\right. \\
&\quad + \frac{1}{4}\sum_{i=1}^{n}\left(y_i^{(2)} - y_i^{(1)}\right)^2 + \frac{1}{2}\sum_{i=1}^{n}\left(y_i^{(4)} - \overline{\tilde{Y}}\right)^2 \\
&\quad \left. + \frac{2}{3}\sum_{i=1}^{n}\left(y_i^{(4)} - \overline{\tilde{Y}}\right)\left(y_i^{(3)} - y_i^{(4)}\right) + \frac{1}{4}\sum_{i=1}^{n}\left(y_i^{(3)} - y_i^{(4)}\right)^2\right\}^{\frac{1}{2}}
\end{aligned} \tag{26}$$

$$HS_e = \sqrt{\frac{1}{n-p-1}\sum_{i=1}^{n}\left(\widehat{\widetilde{Y}}_i - \widetilde{Y}_i\right)^2}$$

$$= \sqrt{\frac{1}{n-p-1}\sum_{i=1}^{n}\left[\frac{1}{4}(a_0^{(2)} + a_1^{(2)}X_i - Y_i^{(2)})^2\right.}$$
$$+\frac{1}{6}(a_0^{(2)} + a_1^{(2)}X_i - Y_i^{(2)})(a_0^{(1)} + a_1^{(1)}X_i - Y_i^{(1)}) \qquad (27)$$
$$+\frac{1}{12}(a_0^{(1)} + a_1^{(1)}X_i - Y_i^{(1)})^2 + \frac{1}{4}(a_0^{(3)} + a_1^{(3)}X_i - Y_i^{(3)})^2$$
$$+\frac{1}{6}(a_0^{(3)} + a_1^{(3)}X_i - Y_i^{(3)})(a_0^{(4)} + a_1^{(4)}X_i - Y_i^{(4)})$$
$$\left.+\frac{1}{12}(a_0^{(4)} + a_1^{(4)}X_i - Y_i^{(4)})^2\right\}^{\frac{1}{2}}$$

Therefore, the performance of the regression model can be obtained using formulas (26) and (27).

5.Numerical Example:

To demonstrate the effectiveness of the proposed model, we use the data from Chang [14], which has been converted into asymmetric trapezoidal fuzzy numbers. In this example, we have 8 data pairs where the input variable (independent) is crisp and non-fuzzy, and the output variable (dependent) consists of asymmetric trapezoidal fuzzy numbers.

$$\left[X_i, \tilde{Y}_i\right]$$
$$= [(2:(12.8, 13.7, 14.6, 15.6)), (4:(14.4, 15.57, 16.74, 17.9))$$
$$,(6:(12.6, 13.6, 14.6, 15.7)), (8:(16.8, 17.8, 18.8, 19.8)) \qquad (28)$$
$$,(10:(15.7, 17.1, 18.5, 19.9)), (12:(20.5, 21.5, 22.5, 23.7))$$
$$,(14:(15.3, 16.8, 18.3, 19.9)), (16:(20.9, 21.9, 22.9, 23.9))]$$

We use the formulas from the previous section to calculate the estimated values, which are presented in the table below.

Table 1: Observed values and estimated values using the proposed model.

i	X_i	$\tilde{Y}_i = (Y_i^{(1)}, Y_i^{(2)}, Y_i^{(3)}, Y_i^{(4)})$	$\widehat{\widetilde{Y}}_i = (\widehat{\widetilde{Y}}_i^{(1)}, \widehat{\widetilde{Y}}_i^{(2)}, \widehat{\widetilde{Y}}_i^{(3)}, \widehat{\widetilde{Y}}_i^{(4)})$
1	2	$(12.8, 13.7, 14.6, 15.6)$	$(12.63, 13.61, 14.66, 15.7)$
2	4	$(14.4, 15.57, 16.74, 17.9)$	$(13.63, 14.65, 15.72, 16.8)$
3	6	$(12.6, 13.6, 14.6, 15.7)$	$(14.63, 15.69, 16.78, 17.9)$
4	8	$(16.8, 17.8, 18.8, 19.8)$	$(15.63, 16.73, 17.84, 19)$
5	10	$(15.7, 17.1, 18.5, 19.9)$	$(16.63, 17.77, 18.9, 20.1)$
6	12	$(20.5, 21.5, 22.5, 23.7)$	$(17.63, 18.81, 19.96, 21.2)$
7	14	$(15.3, 16.8, 18.3, 19.9)$	$(18.63, 19.85, 21.02, 22.3)$
8	16	$(20.9, 21.9, 22.9, 23.9)$	$(19.63, 20.89, 22.08, 23.4)$

To evaluate the performance of the proposed model, the values of $\left(S_{\tilde{y}}\right)$ and $\left(HS_e\right)$ are calculated as follows:

$$S_{\tilde{y}} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(\tilde{Y}_i - \overline{\tilde{Y}}\right)^2} = 3.31$$

$$HS_e = \sqrt{\frac{1}{n-p-1}\sum_{i=1}^{n}\left(\widehat{\widetilde{Y}}_i - \tilde{Y}_i\right)^2} = 1.96$$

The estimated values for Chang's model [14] and the proposed model are shown in Table 2.

Based on the values in Table 2, using trapezoidal fuzzy data in the proposed model reduces the data dispersion $\left(S_{\tilde{y}}\right)$. To evaluate reliability, we use the values of $\left(HS_e\right)$ and $\left(\frac{HS_e}{S_{\tilde{y}}}\right)$. As shown in Table 2, both of these indices are greater in the proposed model compared to the method proposed by Chang [14].

6.Conclusion:

In this paper, we propose a novel regression method using trapezoidal fuzzy data, which significantly reduces the fuzziness of the data for hybrid fuzzy least squares regression. The introduction of weighted fuzzy arithmetic allows for a more precise formulation of the sum of squared errors between predicted and observed variables, enhancing the accuracy of the regression models. This method has been validated through both bivariate and multivariate regression models, demonstrating its applicability across different types of data sets. Furthermore, when comparing the reliability measures with other conventional regression methods, our approach yields superior predictive values, showcasing its robustness and effectiveness. The reduced ambiguity and improved accuracy make this hybrid fuzzy regression method a valuable tool for researchers and practitioners dealing with uncertain and imprecise data.

Overall, the proposed methodology not only advances the field of fuzzy regression but also opens new avenues for its application in various real-world scenarios, from economics to engineering. Future research can explore the integration of this method with other fuzzy and non-fuzzy techniques to further enhance its capabilities and applications.

Table 2: Reliability measures for two regression models

Regression Method	$S_{\tilde{y}}$	HS_e	$\frac{HS_e}{S_{\tilde{y}}}$
Chang's Method	3.37	1.85	0.55
Proposed Method	3.31	1.96	0.59

References

[1] H. Tanaka, S. Uejima, K. Asia, “Linear regression analysis with fuzzy model”, *IEEE Transactions on Systems, Man, and Cybernetics,*Vol. 12, pp. 903-907,(1982).

[2] H. Tanaka, J. Watada, “Possibility linear systems and their application to the linear regression model” ,*Fuzzy Sets and Systems*,Vol. 27,pp. 275-289(1988).

[3] H. Tanaka, I. Hayashi, J. Watada, “Possibility linear regression analysis for fuzzy data” ,*European Journal of Operational Research*Vol. 40,pp.389-396, (1989).

[4] D.T. Redden and W. H Woodall,“Properties of certain fuzzy linear regression methods”, Fuzzy Sets and SystemsVol. 64,pp. 361-375, (1994).

[5] Tanaka, H., Ishibuchi, H., “Identification of possibility linear sytsems by quadratic Membership functions of Fuzzy parmeters”, Fuzzy Sets and Systems,Vol. 69,pp. 305-318 , (1995).

[6] P. Diamond, “Fuzzy least squares”, Inform.Sci, Vol. 46, pp. 669-690, (1988).

[7] M. Sakawa, H. Yano, “Multi objective fuzzy linear regression analysis for fuzzy input-output data”, Fuzzy Sets and Systems,Vol. 47,pp. 173-181, (1992).

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

[9] D.Hun Hong, Ch.Hwang, Ch.Ahn, “Ridge estimation for regression models with crisp inputs and Gaussian fuzzy output”, *Fuzzy Sets and Systems,*Vol. 142,pp. 307-319, (2004)

[10] T. Razzaghnia, “Regression parameters Prediction in data set with outlier using neural network”, *Hacettepe Journal of Mathematics and Statistics*, Vol. 48,pp. 1170-1184, (2019).

[11] T. Razaghnia. ,R.Naderkhani, M. H. Behzad, R. Farnoosh, “Fuzzy Regression Analysis Based on Fuzzy Neural Networks Using Trapezoidal Data”,*Int. J. Fuzzy Syst.*, Vol. 23, No. 5,pp. 1267-1280, (2021).

[12] T. Razzaghnia, S. Danesh, A. Maleki, M. Danesh, M, “An Adaptive Fuzzy Inference System Model to Analyze Fuzzy Regression with Quadratic Programming and Fuzzy Weights Incorporating Uncertainty in the Observed Data”, *Fuzzy Information and Engineering*, Vol. 15, No. 2,pp. 87-99, (2023).

[13] D. Dubois, H. Prade, “Fuzzy Sets and Systems: Theory and Application”, *Academic Press, New York*, (1980).

[14] Y.H.O. Chang, “Hybrid fuzzy least-squares regression analysis and its reliability measures”, *Fuzzy Sets and Systems,*Vol. 119,pp. 225-246, (2001).