Journal of Statistics & Management Systems
ISSN 0972-0510 (Print), ISSN 2169-0014 (Online)
Vol. 27 (2024), No. 6, pp. 1095–1115
DOI : 10.47974/JSMS-1031

Possibilistic linear and quadratic regression analysis for fuzzy random data and application

Mrutyunjaya Sahoo *
S. Chakraverty †
Department of Mathematics
National Institute of Technology Rourkela
Rourkela 769008
Odisha
India

Abstract

Regression is the study of how the distribution of response varies as the values of its predictors change. Regression models are employed in almost every area of science and engineering and are essential tools for analyzing data. Sometimes obtained data may be in the form of fuzzy uncertainty along with the associated probabilities. As such, it may be interesting and required to develop the corresponding regression models to figure out the process. Accordingly, in this work, the perception of fuzzy random variables (FRV) and associated expectations are discussed in the regression models. Linear and quadratic fuzzy random regression models have been investigated based on fuzzy random data to approximate the fuzzy coefficients. Trapezoidal fuzzy numbers have been considered to express both fuzzy input and fuzzy output variables. Finally, some problems have also been solved to illustrate the application of the current approach.

Subject Classification: *62J86, 62M10, 03E72.*

Keywords: *Trapezoidal fuzzy number, Fuzzy random variable, Expectation value, Fuzzy random linear, Quadratic regression model.*

1. Introduction

One of the most widely used modeling and prediction techniques is regression analysis (RA). In fuzzy regression analysis (FRA), fuzzy numbers are used to serve as the elements of the model. Random errors

* *E-mail:* **mrutyunjay.sahoo716@gmail.com (Corresponding Author)**
† *E-mail:* **sne_chak@yahoo.com**

are thought to be the cause of differences between examined and approximated values in traditional regression models. In these situations, statistical approaches are frequently used. However, in practice, there is a lot of limited theory and inaccurate knowledge; frequently, data can't always be reported accurately, and the elemental model can't always be precisely specified. In such instances, Zadeh [1] introduced the concept of fuzzy numbers and related arithmetic processes for the first time in 1965. Also, fuzziness is advantageous in practical systems where human estimation is influential.

The fuzzy set theory led to the development of two primary techniques for fuzzy regression. The first method is possibilistic regression, and another technique is the fuzzy least squares regression. Zadeh [2] was the first to introduce the concept of possibility. Nahmias [3] and Sugeno [4] also have done extensive research on the possibility space. With this approach, to handle fuzzy data, the first fuzzy linear regression model (FLRM) was proposed by Asai et al. They investigated FLR with fuzzy coefficients and input variables as crisp, whose objective is to reduce the model's fuzziness and convert the optimization task of estimation to a linear programming problem (LPP). Tanaka and Watada [6], Tanaka and Hayashi [7] proposed possibilistic regression analysis built on the idea of possibility measure in place of fuzziness. They addressed some features of possibilistic linear systems and constructed a model that may be implemented as an interval analysis model. Bisserier et al. [8] introduced an improved technique to FLR with trapezoidal fuzzy intervals. Chan et al. [9] researched fuzzy regression based on genetic programming. "Fuzzy Logistic Regression" is suggested and discussed by Pourahmad et al. [10] to model fuzzy binary observations. Aliev et al. [11] integrated FRA into genetic algorithms. Wang [12] suggested a new approximate Bayesian computation-based FLR approach. Their method uses a Bayesian posterior distribution to obtain independent samples of unknown model coefficients using the likelihood-free inference procedure.

The fuzzy least square method (FLSM) was invented by Diamond [13] in 1988 for estimating fuzzy parameters. Its objective is to reduce the average squared error between the investigated and estimated output. In terms of prediction, FLSM exceeds Tanaka's model, but Tanaka's model outperforms FLSM in terms of computational efficiency. Many researchers have revised and applied this approach. Celmins [14] extended the least-squares method to accommodate triangular fuzzy number values in both the dependent and independent variables. Papadopoulos [15] uses a multiple FLR-based model to integrate rainfall, streamflow, and potential evapotranspiration to groundwater. Khammar and Akbari [16] proposed a robust FRM using the least-squares method.

In 1938, Watada [17] suggested a heuristics-based FRM that should apply heuristic approaches to solve production between fuzzy numbers and to solve FRM for fuzzy input and fuzzy output data. Model switching FRM was developed by Yabuuchi and Watada [18] and Yabuuchi [19] to assess mixed data acquired from various systems. Fuzzy logic theory may also play a vital role in constructing an FRM [6]. Few researchers have also looked at fuzzy input-output data using nonlinear membership functions, which can lead to different programming problems. For instance, Tanaka and Lee [20] explore interval regression analysis using quadratic programming rather than linear programming. Kwakernaak [21] proposed the idea of a FRV. It's a function that maps from a probability space to a set of fuzzy integers based on particular measurability criteria. In the area of FRV, various researchers have published different works. In this regard, Kruse and Meyer [22], Liu and Liu [23], and Puri et al. [24] have modified and expanded this concept for various uses. Hesamian and Akbari [25] suggested a new multiple regression model with non-fuzzy regression coefficients and fuzzy intercepts. They also introduced a novel procedure using FRV and LR-fuzzy to estimate the model's components. Yoon and Grzegorzewski [26] considered a least squares estimator with a fuzzy-input and fuzzy-output linear regression model (LRM). A dual neural network technique based on confidence intervals has been presented by Hang Yu [27] for fuzzy random regression problems. Appadoo et al [28] stated the correct form of Mellin transformations $M_x(\tau)$, mean, $\mu_T(\tau)$ and, variance as these are incorrect in Rezvani [29]. Further, they stated that the probability density function of any generalized exponential trapezoidal fuzzy number $T=(p,q,r,s)_{Trfn}$ does not depend on the weight function ω.

In one way, statistical data could comprise both fuzzy and stochastic data at the same time. As such, in a given environment, the lifespan of a particular animal might be expressed as: "around 2 years" with a probability of 0.5, "around 3 years" with a probability of 0.4, "around 5 years" with a probability of 0.6, and "around 1 year" with a probability of 0.3, where "around 2 years", "around 3 years", "around 5 years" and "around 1 year" are all linguistic values that may be defined by fuzzy numbers or variables. In this situation, the animals' lifespans are distributed as follows.

$$\tilde{Y} \approx \begin{pmatrix} \tilde{2} & \tilde{3} & \tilde{5} & \tilde{1} \\ 0.5 & 0.4 & 0.6 & 0.3 \end{pmatrix}$$

It can't be represented by just one random variable. As a result, to assess such two-fold problematic data, we must integrate the two and design a new technique.

This work deals with the fuzzy random regression model (FRRM), which is built on FRV and its expectations. As a result of the aforesaid explanation, the present work is a generalization of [30], where they have formulated a LRM using TFNs as fuzzy random input-output data. In this paper, linear as well as quadratic regression models are developed for the trapezoidal fuzzy random data with trapezoidal fuzzy numbers as parameters.

The rest of this paper is structured as follows. Section 2 considers some background information on fuzzy numbers followed by a brief introduction about fuzzy regression and related studies in Section 1. Section 3 delves into fuzzy variables, FRV, and expectations of FRV. The regression model built on FRV was explored in Section 4. In Section 5, the proposed fuzzy random regression analysis approach is explained with an example. Finally, in Section 6, there are some closing observations.

2. Basic Preliminaries

Some basic concepts and notations related to fuzzy set theory are presented in this section that will be used in subsequent discussions.

2.1 *Definition [31] Fuzzy number*

A fuzzy set $\tilde{A}$ is said to be a fuzzy number, which satisfies the conditions of convexity and normality defined on real line $\boldsymbol{R}$, such that $\{\mu_{\tilde{A}}(x): \boldsymbol{R} \to [01], \forall x \in \boldsymbol{R}\}$.

Here $\mu_{\tilde{A}}(x)$ is called as membership function, which is piecewise continuous.

2.2 *Definition [32] Trapezoidal fuzzy number (TrFN)*

A TrFN $\tilde{A}$ can be interpreted as $(\alpha, \beta, \gamma, \eta)_{Trfn}$, which has the membership function $\mu_{\tilde{A}}(x)$ (Figure 1) as follows:

$$\mu_{\tilde{A}}(x)=\begin{cases} 0, & x \le \alpha \\ \frac{x-\alpha}{\beta-\alpha}, & \alpha < x \le \beta \\ 1, & \beta \le x \le \gamma \\ \frac{\eta-x}{\eta-\gamma}, & \gamma < x \le \eta \\ 0, & x \ge \eta \end{cases}$$

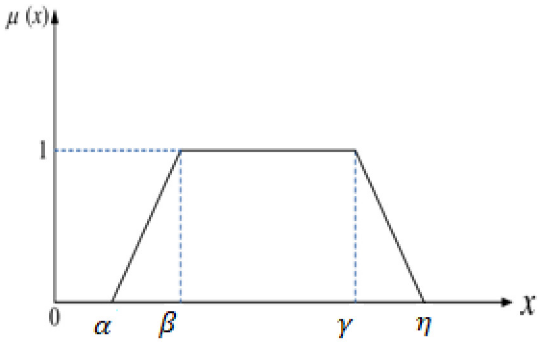

Figure 1

Membership function of TrFN

3. Fuzzy Variables

In 1965, Zadeh [2] introduced the possibility theory as a method for studying ambiguous situations. Uncertainty possibility theory has become the fundamental structure for studying possibilistics. Kaufmann [33] coined the phrase "fuzzy variable", which has since been adopted by others. Nahmias [3] and Wang [34], respectively, extended the idea of a fuzzy variable to sample space and pattern space. First of all, let us discuss the basics of fuzzy variables.

Probability space is related to crisp variables, whereas possibility space is associated with fuzzy variables. Liu and Liu [35] proposed a different version of credibility theory with a weighted mean based on the principles of possibility and necessity measure to investigate the behavior of fuzzy occurrence.

3.1 *Possibility Space [23], [36]*

A possibility space can be interpreted as the 3-tuple, i.e. (©, P(©), Pos), where

- The sample space is © = (θ1, θ2,…, θ_N),
- The power set of © is P (©)
- The possibility measure defined on © is Pos.

3.2 *Necessity Measure of a set [23], [36]*

If B is a set on a possibility space (©, P(©), Pos), the necessity measure Nec{B} of *B* is defined as the impossibility of the B^C, described by Zadeh, 1978.

$$\mathrm{Nec}\{B\} = 1 - \mathrm{Pos}\{B^C\} \text{ for any fuzzy event B.}$$

3.3 *Credibility Measure in a fuzzy environment [23], [36]*

In the fuzzy context, Liu and Liu [23] defined credibility as the average of the possibility and necessity measure.

$$\mathrm{Cr}\{B\} = \frac{1}{2}\,[\mathrm{Pos}\{B\}+\mathrm{Nec}\{B\}] \text{ for any fuzzy event B.}$$

3.4 *Possibility, Necessity, and Credibility of a TrFN [23], [36]*

Let $\tilde{A}$ be a TrFN defined on a possibility space (©, P(©), Pos). Then $\mu_{\tilde{A}}(x)$ be its membership function, where μ and x are real numbers. For a Trapezoidal fuzzy variable, the explicit form of the membership function is provided by Zadeh. (1965).

$$\mu_{\tilde{A}}(x) = \begin{cases} 0, & x \le \alpha \\ \frac{x-\alpha}{\beta-\alpha}, & \alpha < x \le \beta \\ 1, & \beta \le x \le \gamma \\ \frac{\eta-x}{\eta-\gamma}, & \gamma < x \le \eta \\ 0, & x \ge \eta \end{cases} \tag{1}$$

Then, the possibility, necessity, and credibility of this fuzzy occurance are characterized in following manner:

- The explicit expression for the possibility of a TrFN is given by:

$$Pos\{\tilde{A} \le a_0\} = Sup_{x \le a_0} \mu_{\tilde{A}}(x) \begin{cases} 0 & a_0 \le \alpha \\ \frac{a_0-\alpha}{\beta-\alpha} & \alpha \le a_0 \le \beta \\ 1 & a_0 \ge \beta \end{cases} \tag{2}$$

- The explicit expression for the necessity of a TrFN is given by:

$$Nec\{\tilde{A} \le a_0\} = 1 - \text{Pos }\{\tilde{A} > a_0\} = 1 - Sup_{x>a_0}\mu_{\tilde{A}}(x)$$

$$= \begin{cases} 0 & \beta \le a_0 \le \gamma \\ \frac{\eta - a_0}{\gamma - \eta} & \gamma \le a_0 \le \eta \\ 1 & a_0 \ge \eta \end{cases} \tag{3}$$

- The credibility of a TrFN can be obtained as the mean of its possibility and necessity.

$$Cr\{\tilde{A} \le a_0\} = \tfrac{1}{2}[Pos\{\tilde{A} \le a_0\} + Nec\{\tilde{A} \le a_0\}]$$

$$= \frac{1}{2}\begin{cases} 0 & a_0 \le \alpha \\ \frac{a_0 - \alpha}{\beta - \alpha} & \alpha < a_0 \le \beta \\ 1 & \beta \le a_0 \le \gamma \\ 1 + \frac{a_0 - \eta}{\eta - \gamma} & \gamma < a_0 \le \eta \\ 1 & a_0 \ge \eta \end{cases} \tag{4}$$

3.5 *Expectation of Fuzzy variable [35]:*

The expectation of a fuzzy variable $\tilde{A}$ can be interpreted as:

$$E[\tilde{A}] = \int_0^\infty Cr\{\tilde{A} \ge a\}\,da - \int_{-\infty}^0 Cr\{\tilde{A} \le a\}\,da \tag{5}$$

with one of the two integrals is finite

where $Cr\{\tilde{A} \ge a_0\} = 1 - Cr\{\tilde{A} \le a_0\}$

$$= \begin{cases} 1 & a_0 \le \alpha \\ 1 - \frac{1}{2}\left(\frac{a_0 - \alpha}{\beta - \alpha}\right) & \alpha < a_0 \le \beta \\ \frac{1}{2} & \beta \le a_0 \le \gamma \\ \frac{1}{2} - \frac{1}{2}\left(\frac{a_0 - \eta}{\gamma - \eta}\right) & \gamma < a_0 \le \eta \\ \dfrac{1}{2} & a_0 \ge \eta \end{cases} \tag{6}$$

By using (6), we can obtain the expectation value of $\tilde{A}$ to be

$$E[\tilde{A}] = \tfrac{\alpha+\beta+\gamma+\eta}{4} \tag{7}$$

3.6 *Fuzzy Random Variables (FRV)*

Kwakernaak [21] proposed the concept of a FRV. It is a function that maps from a probability space to a set of fuzzy integers based on particular measurability criteria. Mathematically, for a probability space $(\Omega, \sum, P)$ F_v is an assortment of fuzzy variables designated on a possibility space (©, P(©), Pos). A FRV can be defined as a function $Y:\Omega \to F_v$ such that for any Borel subset B of $\Re$, Pos $Y(\varpi) \in B$ is a measurable function of ϖ.

Assuming Y as a FRV defined on Ω, for each $\varpi \in \Omega$, $Y(\varpi)$ is a fuzzy variable and $E[Y(\varpi)]$ is the expectation of $Y(\varpi)$, which has been established as a measurable function of ϖ [23]. Based on this concept, the expected value of the FRV 'Y' can be treated as the expectation of the random variable $E[Y(\varpi)]$ [23].

3.7 *Definition [23]*

If Y be a FRV defined on probability space $(\Omega, \sum, P)$. A definition of the expectation of Y would be as follows:

$$E[Y] = \int_{\Omega} [\int_0^{\infty} Cr\{Y(\varpi) \geq a\} da - \int_{-\infty}^{0} Cr\{Y(\varpi \leq a)\} da] P(d\varpi). \tag{8}$$

Example 1 [30]: Let W be a random variable defined on probability space $(\Omega, \sum, P)$. Let us define for each $\varpi \in \Omega$,

$$Y(\varpi) = (W(\varpi)+k_1, W(\varpi)-k_2, W(\varpi)+k_3, W(\varpi)+k_4)_{TrFV}.$$

Here Y is a trapezoidal FRV.

4. FRV regression model

Fuzzy arithmetic procedures have been examined in [3],[4] based on possibility. Sanchez [37] in 1984 explored the solution of fuzzy equations. According to Tanaka and Watada [6], Sanchez's fuzzy equations can be considered as possibilistic equations. The linear regression analysis has been done using a probabilistic approach in [5], [38]. In this research, we focus on the features of probabilistic linear and quadratic models and develop corresponding FRMs.

Table 1

Fuzzy Random Input and Output Data

Sl. no.	Input data	Output data
1	$\overline{X}_1$	$\overline{Y}_1$
2	$\overline{X}_2$	$\overline{Y}_2$
.	.	.
.	.	.
.	.	.
j	$\overline{X}_j$	$\overline{Y}_j$
.	.	.
.	.	.
.	.	.
n	$\overline{X}_n$	$\overline{Y}_n$

Table 1 represents the data to be dealt with in this investigation. $\overline{X}_i$ and $\overline{Y}_i$ for $i=1,2,3,\ldots,n$ are the FRV which represents the input and output data respectively, which may be defined as;

$$\overline{Y}_i = \bigcup_{t=1}^{k}\{(y_1{}^t, y_2{}^t, y_3{}^t, y_4{}^t)_{Trfn}, p_i{}^t\},\ \overline{X}_i = \bigcup_{t=1}^{k}\{(x_1{}^t, x_2{}^t, x_3{}^t, x_4{}^t)_{Trfn}, q_i{}^t\}.$$

It may be noted that all data are given as fuzzy numbers with their probability, where $(y_1^t, y_2^t, y_3^t, y_4^t)_{Trfn}$ and $(x_1^t, x_2^t, x_3^t, x_4^t)_{Trfn}$ are trapezoidal fuzzy variables that are obtained with probabilities p_i^t and q_i^t respectively, for $i=1,2,3,\ldots,n,$ and $t=1,2,3,\ldots,k.$

Let us consider the fuzzy linear and quadratic models using symmetric coefficients $\overline{P}$, $\overline{Q}$ and $\overline{R}$ as follows:

$$\overline{Y}_i^* = \overline{P}\overline{X}_i + \overline{Q}, \tag{9}$$

and,

$$\overline{Y}_i^* = \overline{P}\overline{X}_i^2 + \overline{Q}\overline{X}_i + \overline{R},\ \text{for}\ i=1,2,3,\ldots,n \tag{10}$$

Where $\overline{Y}_i^*$ shows the estimation and $\overline{P} = \left(\overline{P}^l, \frac{3\overline{P}^l+\overline{P}^r}{4}, \frac{\overline{P}^l+3\overline{P}^r}{4}, \overline{P}^r\right)_{Trfn}$ is the symmetric trapezoidal fuzzy coefficient when trapezoidal fuzzy random data X_i are given for $i=1,2,3,\ldots,n$ as shown in table 1. ($\overline{Q}$, $\overline{R}$ can also be defined similarly as $\overline{P}$).

Knowing the fuzzy random output as $\overline{Y}_i = \bigcup_{t=1}^{k}\{(y_1{}^t, y_2{}^t, y_3{}^t, y_4{}^t)_{Trfn}, p_i{}^t\}$, the following relation should hold for linear as well as quadratic cases:

$$\overline{Y}_i^* = \overline{P}\overline{X}_i + \overline{Q} \underset{FR}{\supset} \overline{Y}_i, \tag{11}$$

and,

$$\overline{Y}_i^* = \overline{P}\overline{X}_i^2 + \overline{Q}\overline{X}_i + \overline{R} \underset{FR}{\supset} \overline{Y}_i, \text{ for } i = 1,2,3,\ldots,n \tag{12}$$

Where $\underset{FR}{\supset}$ is a fuzzy inclusion relation (FIR). It can be expressed in a diffetent ways, including chance-based inclusion, expected value-based inclusion, etc. In this study, we deal with inclusion relations based on the expected values of FRV.

The task of establishing a fuzzy linear and quadratic regression model under fuzzy arithmetic operations yields the following mathematical programming problem.

4.1 *Fuzzy random data based Linear Regression model*

The model may be written as,

$$\min_{\overline{P},\overline{Q}} J(\overline{P},\overline{Q}) = \overline{P}^r - \overline{P}^l + \overline{Q}^r - \overline{Q}^l \tag{13}$$

Subject to

$$\overline{P}^r \geq \overline{P}^l, \; \overline{Q}^r \geq \overline{Q}^l$$

$$\overline{Y}_i^* = \overline{P}\overline{X}_i + \overline{Q} \underset{FR}{\supset} \overline{Y}_i, \text{ for } i = 1,2,3,\ldots,n$$

4.2 *Fuzzy random data based Quadratic Regression model*

In this case, the model may be presented as,

$$\min_{\overline{P},\overline{Q},\overline{R}} J(\overline{P},\overline{Q},\overline{R}) = \overline{P}^r - \overline{P}^l + \overline{Q}^r - \overline{Q}^l + \overline{R}^r - \overline{R}^l$$

Subject to

$$\begin{gathered} \overline{P}^r \geq \overline{\mathrm{P}}^{\mathrm{l}}, \; \overline{Q}^r \geq \overline{\mathrm{Q}}^{\mathrm{l}}, \; \overline{R}^r \geq \overline{\mathrm{R}}^{\mathrm{l}} \\ \overline{Y}_i^* = \overline{P}\overline{X}_i^2 + \overline{Q}\overline{X}_i + \overline{R} \underset{FR}{\supset} \overline{Y}_i \text{ for } i = 1,2,3,\ldots,n \end{gathered} \tag{14}$$

The above formulations are defined using only the FRV.

Table 2

Expectation of fuzzy random input and output data

Sl. no.	Input data		Output data
i	$E[\overline{X}_i]$	$E[\overline{X}_i^2]$	$E[\overline{Y}_i]$
1	$E[\overline{X}_1]$	$E[\overline{X}_1^2]$	$E[\overline{Y}_1]$
2	$E[\overline{X}_2]$	$E[\overline{X}_2^2]$	$E[\overline{Y}_2]$
.	.	.	.
.	.	.	.
.	.	.	.
j	$E[\overline{X}_j]$	$E[\overline{X}_j^2]$	$E[\overline{Y}_j]$
.	.	.	.
.	.	.	.
.	.	.	.
n	$E[\overline{X}_n]$	$E[\overline{X}_n^2]$	$E[\overline{Y}_n]$

Further, if the expectation of these data is taken, then we may have alternative models for both linear and quadratic cases, which may be called expected fuzzy random data regression models.

In this respect, the expectations of these data are shown in Table 2.

4.3 *LRM for Expected fuzzy random data*

Considering the expectation $E[\overline{X}_i]$ and $E[\overline{Y}_i]$ (Table 2) of a FRV, the following FLRM may be obtained:

$$\min_{\overline{P},\overline{Q}} J(\overline{P}) = \overline{P}^r - \overline{P}^l + \overline{Q}^r - \overline{Q}^l$$

Subject to

$$\begin{gathered} \overline{P}^r \geq \overline{P}^l,\ \overline{Q}^r \geq \overline{Q}^l \\ \overline{Y}_i^* = \overline{P}E[\overline{X}_i] + \overline{Q} \supset_{\tilde{h}} E[\overline{Y}_i] \text{ for } i = 1,2,3,\ldots,n \end{gathered} \tag{15}$$

Where $\supset_{\tilde{h}}$ represents the FIR at level h.

4.4 *Quadratic regression model for Expected fuzzy random data*

Similarly following fuzzy Quadratic regression model may be written:

$$\min_{\bar{P},\bar{Q},\bar{R}} J(\bar{P},\bar{Q},\bar{R}) = \bar{P}^r - \bar{P}^l + \bar{Q}^r - \bar{Q}^l + \bar{R}^r - \bar{R}^l$$

Subject to

$$\begin{gathered} \bar{P}^r \geq \bar{\mathrm{P}}^1, \bar{Q}^r \geq \bar{\mathrm{Q}}^1, \bar{R}^r \geq \bar{\mathrm{R}}^1 \\ \bar{Y}_i^* = \bar{P}E[\bar{X}_i^2] + \bar{Q}E[\bar{X}_i] + \bar{R} \underset{h}{\supset} E[\bar{Y}_i], \text{ for } i = 1,2,3,\ldots,n \end{gathered} \tag{16}$$

Where $\underset{\tilde{h}}{\supset}$ represents the FIR at level h.

There is a challenge in solving mathematical programming (13) and (14) for FRM, since the product of two fuzzy quantities may not necessarily equal a fuzzy number. This problem leads to a heuristics algorithm, as discussed in Watada et al. [39]. The mathematical programming (15) and (16) are FRM, and these may be challenging to solve. However, if those are simple lpp, then they can easily be solved by the traditional procedure.

5. Fuzzy Regression Model (FRM) for Expected Fuzzy Random Data

In this section, the FRM based on the expectation of FRV using the TrFN (table 2) is discussed. Based on (8), we consider the FIR at level h=0, which is the expected values of fuzzy random data.

5.1 *Linear Fuzzy Regression Model (LFRM)*

$$\min_{\bar{P},\bar{Q}} J(\bar{P}) = \bar{P}^r - \bar{P}^l + \bar{Q}^r - \bar{Q}^l$$

Subject to

$$\begin{gathered} \bar{P}^r \geq \bar{P}^l, \bar{Q}^r \geq \bar{Q}^l \\ \left. \begin{array}{l} E[\bar{Y}_i] \leq \bar{P}^r(E[\bar{X}_i]) + \bar{Q}^r, \\ E[\bar{Y}_i] \geq \bar{P}^l(E[\bar{X}_i]) + \bar{Q}^l, \end{array} \right. \text{ for } i = 1,2,3,\ldots,n \end{gathered} \tag{17}$$

Model (17) is a simple LPP that may be easily solved to obtain solutions.

Results and discussions

Let us assume the following fuzzy input and output random data:

$\overline{X}_1 = ((1,2,3,4)_{Trfn}, 0.3; (2,3,4,5)_{Trfn}, 0.7),$ $\overline{Y}_1 = ((10,12,14,16)_{Trfn}, 0.2; (12,14,16,18)_{Trfn}, 0.8),$
$\overline{X}_2 = ((2,3,4,5)_{Trfn}, 0.5; (3,4,5,6)_{Trfn}, 0.5),$ $\overline{Y}_2 = ((10,14,16,20)_{Trfn}, 0.4; (16,18,20,22)_{Trfn}, 0.6),$
$\overline{X}_3 = ((4,6,8,10)_{Trfn}, 0.5; (6,8,10,12)_{Trfn}, 0.5),$ $\overline{Y}_3 = ((16,17,18,19)_{Trfn}, 0.8; (18,20,22,24)_{Trfn}, 0.2),$
$\overline{X}_4 = ((6,8,10,12)_{Trfn}, 0.4; (8,10,12,14)_{Trfn}, 0.6),$ $\overline{Y}_4 = ((20,22,24,26)_{Trfn}, 0.4; (24,26,28,30)_{Trfn}, 0.6),$

$\overline{X}_5 = ((8,10,12,14)_{Trfn}, 0.3; (10,12,14,16)_{Trfn}, 0.7),$ $\overline{Y}_5 = ((24,26,28,30)_{Trfn}, 0.5; (26,28,30,32)_{Trfn}, 0.5),$
$\overline{X}_6 = ((12,13,14,15)_{Trfn}, 0.5; (13,14,15,16)_{Trfn}, 0.5),$ $\overline{Y}_6 = ((28,30,32,34)_{Trfn}, 0.4; (30,32,34,36)_{Trfn}, 0.6),$
$\overline{X}_7 = ((10,12,14,16)_{Trfn}, 0.25; (12,14,16,18)_{Trfn}, 0.75),$ $\overline{Y}_7 = ((30,32,36,38)_{Trfn}, 0.2; (32,36,38,40)_{Trfn}, 0.8),$
$\overline{X}_8 = ((13,14,15,16)_{Trfn}, 0.3; (14,15,16,17)_{Trfn}, 0.7),$ $\overline{Y}_8 = ((32,36,38,42)_{Trfn}, 0.4; (354,36,38,40)_{Trfn}, 0.6),$
$\overline{X}_9 = ((12,14,16,18)_{Trfn}, 0.5; (14,16,18,20)_{Trfn}, 0.5),$ $\overline{Y}_9 = ((34,36,40,44)_{Trfn}, 0.2; (36,40,42,46)_{Trfn}, 0.8),$
$\overline{X}_{10} = ((14,16,18,20)_{Trfn}, 0.25; (16,18,20,22)_{Trfn}, 0.75),$ $\overline{Y}_{10} = ((36,38,40,42)_{Trfn}, 0.5; (38,40,42,44)_{Trfn}, 0.5),$

The LFRM of expected fuzzy random data for the considered problem is

$$\overline{Y}_i^* = \overline{P}E[\overline{X}_i] + \overline{Q} \tag{18}$$

Here our main target is to find the coefficients of the expectation of the fuzzy variable, such that it satisfies the FIR,

$$\overline{Y}_i^* = \overline{P}E[\overline{X}_i] + \overline{Q} \underset{\tilde{h}}{\supset} E[\overline{Y}_i].$$

Case 1: Intercept as zero in (18)

In this case

if $\overline{Q} = 0,$ then equation (18) becomes $\overline{Y}_i^* = \overline{P}E[\overline{X}_i].$

The expected value of the fuzzy input-output data can be evaluated by using (8) as:

$$E[\overline{X}_1] = 3.2, E[\overline{X}_2] = 4, \dots, E[\overline{X}_{10}] = 18.5$$

and

$$E[\overline{Y}_1] = 14.6, E[\overline{Y}_2] = 17.4, \dots, E[\overline{Y}_{10}] = 40$$

Based on (17), we can write the LFRM as:

$$\min_{\overline{P}} j(\overline{P}) = \overline{P}^r - \overline{P}^l$$

Subject to

$$\overline{P}^l \le \overline{P}^r$$
$$3.2\overline{P}^l \le 14.6 \le 3.2\overline{P}^r,$$
$$4\overline{P}^l \le 17.4 \le 4\overline{P}^r,$$
$$8\overline{P}^l \le 18.2 \le 8\overline{P}^r,$$
$$10.2\overline{P}^l \le 25.4 \le 10.2\overline{P}^r,$$

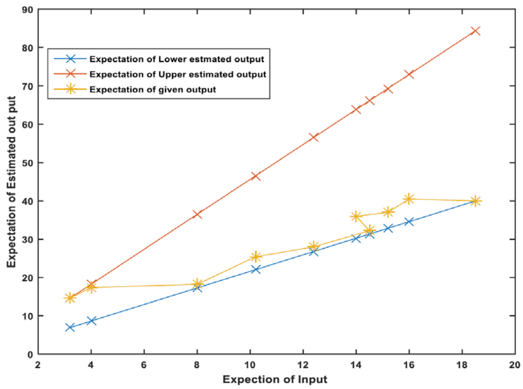

Figure 2

Plot of $E[\overline{Y}_i]$ vs. $\overline{Y}_i^*$

$$12.4\overline{P}^l \leq 28 \leq 12.4\overline{P}^r,$$
$$14.5\overline{P}^l \leq 32.2 \leq 14.5\overline{P}^r,$$
$$14\overline{P}^l \leq 36 \leq 14\overline{P}^r,$$
$$15.2\overline{P}^l \leq 37 \leq 15.2\overline{P}^r,$$
$$16\overline{P}^l \leq 40.5 \leq 16\overline{P}^r,$$
$$18.5\overline{P}^l \leq 40 \leq 18.5\overline{P}^r.$$

The best possible solution may be found by solving the aforementioned LPP, which are as follows: $\overline{P}^l = 2.16$ and $\overline{P}^r = 4.56$.

Now the following LFRM of fuzzy random data can be obtained using the aforementioned optimal solution:

$$\overline{Y}_i^* = \overline{P}E[\overline{X}_i]. \tag{19}$$

Applying the expected values of input fuzzy random data in (19), we may have the following output:

$$\overline{Y}_1^* = [6.91, 14.59], \overline{Y}_2^* = [8.64, 18.24], \overline{Y}_3^* = [17.28, 36.48], \ldots, \overline{Y}_{10}^* = [39.96, 84.36]$$

Now, one may easily check that the FIR (15) is satisfied for (19).

Also, from Figure 2, one may conclude that the FIR holds for the obtained optimal solutions.

Case 2 : If $\overline{Q}_i \neq 0$ in (18) that is $\overline{Y}_i^* = \overline{P}E[\overline{X}_i] + \overline{Q}$.

This can be solved by formulating the LFRM as follows:

$$\min_{\overline{P}} j(\overline{P}) = \overline{P}^r - \overline{P}^l + \overline{Q}^r - \overline{Q}^l$$

Subject to

$$\overline{P}^l \le \overline{P}^r,$$
$$\overline{Q}^l \le \overline{Q}^r,$$
$$3.2\overline{P}^l + \overline{Q}^l \le 14.6 \le 3.2\overline{P}^r + \overline{Q}^r,$$
$$4\overline{P}^l + \overline{Q}^l \le 17.4 \le 4\overline{P}^r + \overline{Q}^r,$$
$$8\overline{P}^l + \overline{Q}^l \le 18.2 \le 8\overline{P}^r + \overline{Q}^r,$$
$$10.2\overline{P}^l + \dot{Q}^l \le 25.4 \le 10.2\overline{P}^r + \overline{Q}^r,$$
$$12.4\overline{P}^l + \overline{Q}^l \le 28 \le 12.4\overline{P}^r + \overline{Q}^r,$$
$$14.5\overline{P}^l + \overline{Q}^l \le 32.2 \le 14.5\overline{P}^r + \overline{Q}^r,$$
$$14\overline{P}^l + \overline{Q}^l \le 36 \le 14\overline{P}^r + \overline{Q}^r,$$
$$15.2\overline{P}^l + \overline{Q}^l \le 37 \le 15.2\overline{P}^r + \overline{Q}^r,$$
$$16\overline{P}^l + \overline{Q}^l \le 40.5 \le 16\overline{P}^r + \overline{Q}^r,$$
$$18.5\overline{P}^l + \overline{Q}^l \le 40 \le 18.5\overline{P}^r + \overline{Q}^r.$$

The optimal solution can be obtained by solving the above LPP as $\overline{P}^l = 1.0625$, $\overline{P}^r = 1.9250$ and $\overline{Q}^l = 9.7 = \overline{Q}^r$.

We can get the following result utilizing the above mentioned optimal solution and the expected values of input fuzzy random data in equation (19):

$$\overline{Y}_1^* = [13.1, 15.86], \overline{Y}_2^* = [13.95, 17.4], \overline{Y}_3^* = [18.2, 25.1], \ldots, \overline{Y}_{10}^* = [29.35, 45.31]$$

It's now easy to check whether the FIR (15) holds for (18).

Also, one may conclude from Figure 3, that the FIR holds for the observed optimal solutions.

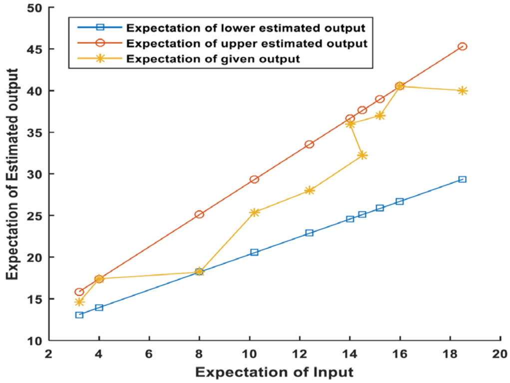

Figure 3

Plot of $E[\overline{Y}_i]$ vs. $\overline{Y}_i^*$

5.2 *Quadratic Fuzzy Regression Model (QFRM)*

$$\min_{\bar{P},\bar{Q},\bar{R}} J(\bar{P}) = \bar{P}^r - \bar{P}^l + \bar{Q}^r - \bar{Q}^l + \bar{R}^r - \bar{R}^l$$

Subject to

$$\begin{aligned} &\bar{P}^r \geq \bar{\mathrm{P}}^1,\ \bar{Q}^r \geq \bar{\mathrm{Q}}^1,\ \bar{R}^r \geq \bar{\mathrm{R}}^1 \\ &E[\bar{Y}_i] \leq \bar{P}^r E[\bar{X}_i^2] + \bar{Q}^r E[\bar{X}_i] + \bar{R}^r, \\ &E[\bar{Y}_i] \geq \bar{P}^l E[\bar{X}_i^2] + \bar{Q}^l E[\bar{X}_i] + \bar{R}^l,\ \text{for}\ i = 1,2,3,\ldots,n \end{aligned} \tag{20}$$

Model (20) is a LPP that can be solved to obtain the solution.

From the given fuzzy random input data that has been taken in LFRM, we can easily find the square of the given TrFN [40].

Then by using (8), we may have

$$E[\bar{X}_1^2] = 11.35, E[\bar{X}_2^2] = 19.35, E[\bar{X}_3^2] = 71, \ldots, E[\bar{X}_{10}^2] = 349.$$

For the given problem, the FRM of expected fuzzy random data is

$$\bar{Y}_i^* = \bar{P}E[\bar{X}_i^2] + \bar{Q}E[\bar{X}_i] + \bar{R}.$$

Let us find the coefficients of the expectation of a fuzzy variable, such that it satisfies the FIR,

$$\bar{Y}_i^* = \bar{P}E[\bar{X}_i^2] + \bar{Q}E[\bar{X}_i] + \bar{R} \supset E[\bar{Y}_i].$$

Now from (20), the QFRM can be formulated in following manner:

$$\min_{\bar{P},\bar{Q},\bar{R}} j(\bar{P},\bar{Q},\bar{R}) = \bar{P}^r - \bar{P}^l + \bar{Q}^r - \bar{Q}^l + \bar{R}^r - \bar{R}^l$$

Subject to

$$\begin{gathered} \bar{P}^l \leq \bar{P}^r, \\ \bar{Q}^l \leq \bar{Q}^r, \\ \bar{R}^l \leq \bar{R}^r, \\ 11.35\bar{P}^l + 3.2\bar{Q}^l + \bar{R}^l \leq 14.6 \leq 11.35\bar{P}^r + 3.2\bar{Q}^r + \bar{R}^r, \\ 19.35\bar{P}^l + 4\bar{Q}^l + R^l \leq 17.4 \leq 19.35\bar{P}^r + 4\bar{Q}^r + \bar{R}^r, \\ 71\bar{P}^l + 8\bar{Q}^l + \bar{R}^l \leq 18.2 \leq 71\bar{P}^r + 8\bar{Q}^r + \bar{R}^r, \\ 119\bar{P}^l + 10.2\bar{Q}^l + \bar{R}^l \leq 25.4 \leq 119\bar{P}^r + 10.2\bar{Q}^r + \bar{R}^r, \\ 160.6\bar{P}^l + 12.4\bar{Q}^l + \bar{R}^l \leq 28 \leq 160.6\bar{P}^r + 12.4\bar{Q}^r + \bar{R}^r, \\ 217\bar{P}^l + 14.5\bar{Q}^l + \bar{R}^l \leq 32.2 \leq 217\bar{P}^r + 14.5\bar{Q}^r + \bar{R}^r, \\ 197.75\bar{P}^l + 14\bar{Q}^l + \bar{R}^l \leq 36 \leq 197.75\bar{P}^r + 14\bar{Q}^r + \bar{R}^r, \end{gathered}$$

$$232.75\overline{P}^l + 15.2\overline{Q}^l + \overline{R}^l \le 37 \le 232.75\overline{P}^r + 15.2\overline{Q}^r + \overline{R}^r,$$
$$263\overline{P}^l + 16\overline{Q}^l + \overline{R}^l \le 40.5 \le 263\overline{P}^r + 16\overline{Q}^r + \overline{R}^r,$$
$$349\overline{P}^l + 18.5\overline{Q}^l + \overline{R}^l \le 40 \le 349\overline{P}^r + 18.5\overline{Q}^r + \overline{R}^r.$$

On solving the above LPP, we may have the optimal solution as

$$\overline{P}^l = -0.1624,\ \overline{P}^r = -0.0460,\ \overline{Q}^l = 2.8596 = \overline{Q}^r,\ \overline{R}^l = 6.8525 = \overline{R}^r.$$

In this case, we may get the following QFRM of fuzzy random data by using the above optimal solution;

$$\overline{Y}_i^* = \overline{P}E[\overline{X}_i^2] + \overline{Q}E[\overline{X}_i] + \overline{R} \tag{21}$$

Based on the expected values of (21), we have the following output:

$$\overline{Y}_1^* = [14.133, 15.454], \overline{Y}_2^* = [15.149, 17.401], \overline{Y}_3^* = [18.198, 26.449], \ldots,$$
$$\overline{Y}_{10}^* = [3.043, 43.701].$$

It is simple to verify that the FIR (16) is satisfied for (21).

Also, from Figure 4, one may conclude that the fuzzy inclusion relation holds for the obtained optimal solutions.

The above approximating fuzzy coefficients method is based on possibilistic regression analysis. However, the discussed method is associated with the formulation of LFRM and QFRM. By using these formulations, we have calculated the models' parameters. Further, with the help of these aforesaid explanations, the FIR for both linear as well as quadratic cases are satisfied. From Figure 2, one may observe that the expectation of the given output fuzzy variable is lying in between the estimated output i.e., the expectation of lower estimated output and the expectation of upper estimated output. Similarly, from Figure 3 and Figure

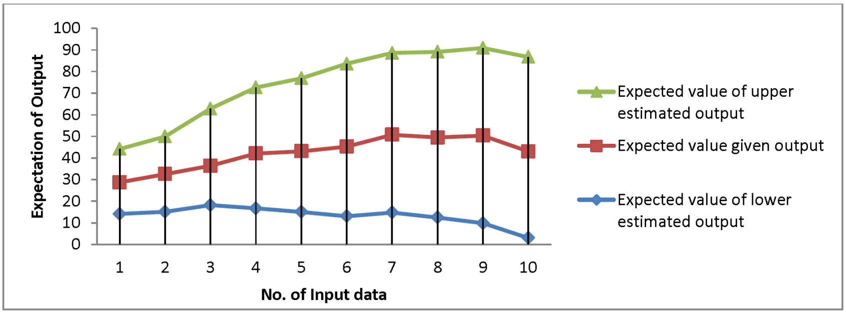

Figure 4

Plot of $E[\overline{Y}_i]$ vs. $\overline{Y}_i^*$

4, one may also observe that the expectation of the given output fuzzy variable is satisfying the FIR.

6. Conclusion

In this paper, linear and quadratic regression models for fuzzy random data have been developed using fuzzy variables and their expected values. We have successfully estimated the model coefficients in such a way that they have satisfied the given fuzzy inclusion relation (FIR). The discussed models based on linear and quadratic regression models have also been verified with respect to fuzzy random data, which satisfy the inclusion relation.

Acknowledgments

The first author would like to express gratitude to the University Grants Commission (UGC), New Delhi, India, for providing the fellowship support to carry out this research.

References

[1] Zadeh, L. A. Fuzzy sets. Information and control, 8(3), 338-353 (1965).

[2] Zadeh, L. A. Fuzzy sets as a basis for a theory of possibility. Fuzzy sets and systems, 1(1), 3-28 (1978).

[3] Nahmias, S. Fuzzy variables. Fuzzy sets and systems, 1(2), 97-110 (1978).

[4] Sugeno, M. Fuzzy theory IV (Lecture note). J. of SICE, 22, 554-559 (1983).

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

[6] Tanaka, H., & Watada, J. Possibilistic linear systems and their application to the linear regression model. Fuzzy sets and systems, 27(3), 275-289 (1988).

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

[8] Bisserier, A., Boukezzoula, R., & Galichet, S. A revisited approach to linear fuzzy regression using trapezoidal fuzzy intervals. *Information Sciences*, 180(19), 3653-3673 (2010).

[9] Chan, K. Y., Kwong, C. K., & Fogarty, T. C. Modeling manufacturing processes using a genetic programming-based fuzzy regression with detection of outliers. *Information Sciences*, 180(4), 506-518 (2010).

[10] Pourahmad, S., Taghi Ayatollahi, S. M., & Taheri, S. M. Fuzzy logistic regression: a new possibilistic model and its application in clinical vague status. *Iranian Journal of Fuzzy Systems*, 8(1), 1-17 (2011).

[11] Aliev, R. A., Fazlollahi, B., & Vahidov, R. Genetic algorithms-based fuzzy regression analysis. Soft Computing, 6, 470-475 (2002).

[12] Wang, N., Reformat, M., Yao, W., Zhao, Y., & Chen, X. Fuzzy Linear regression based on approximate Bayesian computation. *Applied Soft Computing*, 97, 106763 (2020).

[13] Diamond, P. Fuzzy least squares. *Information sciences*, 46(3), 141-157 (1988).

[14] Celmiņš, A. Least squares model fitting to fuzzy vector data. *Fuzzy sets and systems*, 22(3), 245-269 (1987).

[15] Papadopoulos, C., Spiliotis, M., Gkiougkis, I., Pliakas, F., & Papadopoulos, B. Relating hydro-meteorological variables to water table in an unconfined aquifer via fuzzy linear regression. Environments, 8(2), 9 (2021).

[16] Khammar, A. H., Arefi, M., & Akbari, M. G. A robust least squares fuzzy regression model based on kernel function. *Iranian Journal of Fuzzy Systems*, 17(4), 105-119 (2020).

[17] Watada, J. Fuzzy quantification theory type I. The Japanese Journal of Behaviormetrics (The Behaviormetric Society of Japan), 11(1), 66-73 (1983).

[18] Yabuuchi, Y. Fuzzy switching regression model based on genetic algorithm. In Proceedings of the 7th International Fuzzy Systems Association World Congress, 113-118 (1997).

[19] Yabuuchi, Y., & Watada, J. Formulation of fuzzy switching regression model. *Official Journal of Japan Association of Intteligent Information and Fuzzy Systems*, 16(1), 53-59 (2004).

[20] Tanaka, H., & Lee, H. Interval regression analysis by quadratic programming approach. *IEEE Transactions on Fuzzy Systems*, 6(4), 473-481 (1998).

[21] Kwakernaak, Huibert. "Fuzzy random variables—I. Definitions and theorems." Information sciences 15.1 : 1-29 (1978).

[22] Kruse, R., Meyer, K. D., Kruse, R., & Meyer, K. D. On a software tool for statistics with vague data. Statistics with Vague Data, 231-245 (1987).

[23] Liu, Y. K., & Liu, B. Fuzzy random variables: A scalar expected value operator. *Fuzzy Optimization and decision making*, 2, 143-160 (2003).

[24] Puri, M. L., Ralescu, D. A., & Zadeh, L. Fuzzy random variables. In Readings in fuzzy sets for intelligent systems, 265-271 (1993).

[25] Hesamian, G., & Akbari, M. G. A robust multiple regression model based on fuzzy random variables. *Journal of Computational and Applied Mathematics*, 388, 113270 (2021).

[26] Yoon, J. H., & Grzegorzewski, P. On Optimal and Asymptotic Properties of a Fuzzy L 2 Estimator. Mathematics, 8(11), 1956 (2020).

[27] Yu, H., Lu, J., Zhang, G., & Wu, D. (2018, July). A dual neural network based on confidence intervals for fuzzy random regression problems. In 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (pp. 1-8). IEEE.

[28] Appadoo, S. S., Kumar, A., & Gajpal, Y. Generalized exponential trapezoidal fuzzy numbers based on variance. *Journal of Information and Optimization Sciences*, 42(7), 1409-1424 (2021).

[29] Rezvani, S. Ranking generalized exponential trapezoidal fuzzy numbers based on variance. *Applied Mathematics and Computation*, 262, 191-198 (2015).

[30] Watada, J., & Wang, S. Regression model based on fuzzy random variables. Views on Fuzzy Sets and Systems from Different Perspectives: Philosophy and Logic, Criticisms and Applications, 533-545 (2009).

[31] Chakraverty, S., Sahoo, D. M., Mahato, N. R., Chakraverty, S., Sahoo, D. M., & Mahato, N. R. Fuzzy numbers. Concepts of soft computing: Fuzzy and ANN with programming, 53-69 (2019).

[32] Chakraverty, S., Sahoo, D. M., Mahato, N. R., Chakraverty, S., Sahoo, D. M., & Mahato, N. R. Fuzzy Sets. Concepts of Soft Computing: Fuzzy and ANN with Programming, 25-51 (2019).

[33] Kaufmann, A. Introduction to the theory of fuzzy subsets. Academic press (1975).

[34] Wang, P. Z. Fuzzy contactibility and fuzzy variables. *Fuzzy Sets and Systems*, 8(1), 81-92 (1982).

[35] Liu, B., & Liu, Y. K. Expected value of fuzzy variable and fuzzy expected value models. IEEE transactions on Fuzzy Systems, 10(4), 445-450 (2002).

[36] Koissi, M. C., & Shapiro, A. F. Credibility theory in a fuzzy environment. In The 47th Actuarial Research conference Proc., University of Manitoba, Canada (2012).

[37] Sanchez, E. Solution of fuzzy equations with extended operations. Fuzzy sets and Systems, 12(3), 237-248 (1984).

[38] Tanaka, H., Shimomura, T., Watada, J., & Asai, K. Fuzzy Linear Regression Analysis Of The Number Of Staff In Local Government (1987).

[39] Watada, J. Applications in business, multiattribute decision—making. Applied fuzzy system, 244-252 (1994).

[40] Mateos, A., & Jiménez, A. A trapezoidal fuzzy numbers-based approach for aggregating group preferences and ranking decision alternatives in MCDM. In Evolutionary Multi-Criterion Optimization: 5th International Conference, EMO 2009, Nantes, France, Proceedings 5 (pp. 365-379). Springer Berlin Heidelberg (2009).

Received June, 2022

Revised September, 2022