

# Probabilistic Fuzzy Regression Approach from the Point of View Risk

Nana Gao\*, Qiujun Lu

College of Science, University of Shanghai for Science and Technology, Shanghai, China

Email: \*nana\_gao1@outlook.com

**How to cite this paper:** Gao, N.N. and Lu, Q.J. (2018) Probabilistic Fuzzy Regression Approach from the Point of View Risk. *Journal of Data Analysis and Information Processing*, 6, 156-167.  
<https://doi.org/10.4236/jdaip.2018.64010>

**Received:** August 31, 2018

**Accepted:** November 9, 2018

**Published:** November 12, 2018

Copyright © 2018 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Fuzzy regression analysis is an important regression analysis method to predict uncertain information in the real world. In this paper, the input data are crisp with randomness; the output data are trapezoid fuzzy number, and three different risk preferences and chaos optimization algorithm are introduced to establish fuzzy regression model. On the basis of the principle of the minimum total spread between the observed and the estimated values, risk-neutral, risk-averse, and risk-seeking fuzzy regression model are developed to obtain the parameters of fuzzy linear regression model. Chaos optimization algorithm is used to determine the digital characteristic of random variables. The mean absolute percentage error and variance of errors are adopted to compare the modeling results. A stock rating case is used to evaluate the fuzzy regression models. The comparisons with five existing methods show that our proposed method has satisfactory performance.

## Keywords

Probabilistic Fuzzy Regression, Chaos Optimization Algorithm, Risk Preferences Models, Mean Absolute Percentage Error, Variance of Errors

## 1. Introduction

Fuzzy set theory was introduced by Zedah in 1964. On the basis, fuzzy set theory has been developed rapidly and applied in many fields, such as control system, artificial intelligence, and pattern recognition, etc. In the field of fuzzy regression analysis, two main methods are proposed by Tanaka and Diamond. The first approach; Tanaka [1] first proposed the fuzzy linear regression model in 1982; established the first fuzzy regression analysis model. Another, Diamond [2] proposed the fuzzy square method to determine the fuzzy parameters in 1988.

The latter scholars have developed new studies based on the two mentioned

above. Tanaka and Watada [3] introduced possibility measure into fuzzy linear regression. Savic and Pedrycz [4] proposed that the use of fuzzy linear regression is highly influenced by its purpose and the nature of the practical application. Kim and Bishu [5] studied fuzzy regression models by comparing membership functions. Modarres and Nasrabadi [6] proposed an fuzzy linear regression analysis from the point of view risk. Kwong and Chen [7] introduced fuzzy least squares regression approach to modelling relationships in Quality function deployment. Zhang [8] proposed an fuzzy linear regression analysis model based on the centroid method. Li and Zeng [9] introduced an fuzzy regression models based on least absolute deviation.

The previous models were studied to solve fuzziness, but some data not only have fuzziness but also have randomness. Only few previous studies have examined both fuzziness and randomness in empirical research. In this paper, according to the actual meaning of the random variables, the appropriate numerical characteristics are selected. Introduced chaos optimization algorithm (COA) determines the numerical characteristics of the random variable. Three mathematical programming models are proposed, called risk-neutral, risk-averse and risk-seeking. Based on this, the fuzzy regression coefficient of different risk models can be obtained, and model are determined.

## 2. Mathematical Preliminaries

**Definition 2.1** [10] let  $\tilde{A}$  be a mapping of domain  $x$  to  $[0,1]$ , that is  $\tilde{A}: X \rightarrow [0,1], x \mapsto \mu(x)$ ,  $\tilde{A}$  is called fuzzy set of  $x$ .  $\mu_A(x)$  is called the membership function of  $x$  in the fuzzy set  $\tilde{A}$ .

**Definition 2.2** [10] Fuzzy set  $\tilde{A}$  is called fuzzy number in the real number field  $\mathbb{R}$ , if  $\forall \alpha \in (0,1]$ ,  $A_\alpha$  is finite closed interval,  $\tilde{R}$  is a set of all fuzzy numbers.

**Definition 2.3** [5] when a fuzzy number  $\tilde{A} = (\alpha, a, \beta)_{LR}$  has a membership function  $\mu_A(x)$  is defined as

$$\mu_A(x) = \begin{cases} L\left(\frac{a-x}{\alpha}\right) & x \leq a, \alpha > 0 \\ R\left(\frac{x-a}{\beta}\right) & x > a, \beta > 0 \end{cases}$$

where  $a \in \mathbb{R}$  is the center of the function,  $\alpha, \beta$  are the left and right spread of the function, respectively. This fuzzy number is referred to as  $L-R$  fuzzy number. The shape of function satisfies 1)  $L(x) = L(-x)$ ,  $R(x) = R(-x)$ , 2)  $L(0) = 1$ ,  $R(0) = 1$ , and 3)  $R(x)$  are non-increasing on  $[0, \infty)$ .

$\tilde{R}_{LR}$  is a set of all  $L-R$  fuzzy numbers,  $L-R$  fuzzy number  $\tilde{A}$  has the left and right spreads from a real number whose membership is 1. The spreads represent the fuzziness of the fuzzy number could be symmetric or non-symmetric. If the spreads are null, there is no fuzziness of the number, and it is a real number. The choice of the  $L$  and  $R$  functions is dependent upon the subjective judgement.

**Proposition 2.1**  $(\alpha, a, \beta)_{LR}, (\gamma, b, \delta) \in \tilde{R}_{LR}, k \in R.$

1)  $(\alpha, a, \beta)_{LR} + (\gamma, b, \delta)_{LR} = (\alpha + \gamma, a + b, \beta + \delta)_{LR}$

2)  $k \cdot (\alpha, a, \beta)_{LR} = \begin{cases} (k\alpha, ka, k\beta)_{LR}, & k \geq 0 \\ (-k\beta, ka, -k\alpha)_{RL}, & k < 0 \end{cases}$

$-(\alpha, a, \beta)_{LR} = (\beta, -a, \alpha)_{RL}$ , especially.

**Definition 2.4**  $\tilde{A} \in \tilde{R}$ , if the membership function of  $\tilde{A}$  is defined as

$$\mu_A(x) = \begin{cases} \frac{x-a+l}{l}, & a-l \leq x \leq a \\ 1, & a \leq x \leq b \\ \frac{b+r-x}{r}, & b \leq x \leq b+r \\ 0, & \text{else} \end{cases}$$

$\tilde{A} = (l, a, b, r)_T$  is trapezoidal fuzzy number. If the fuzzy number is symmetric fuzzy numbers,  $l = r$ . The fuzzy number is triangular fuzzy number when  $a = b$ .

The set of trapezoid fuzzy numbers is denoted by  $\tilde{R}_T$ , and  $\tilde{R}_T \subseteq \tilde{R}_{LR} \subseteq \tilde{R}$ .

**Proposition 2.2**  $(l_1, a_1, b_1, r_1), (l_2, a_2, b_2, r_2) \in \tilde{R}_T, k \in R$ , then

1)  $(l_1, a_1, b_1, r_1)_T + (l_2, a_2, b_2, r_2)_T = (l_1 + l_2, a_1 + a_2, b_1 + b_2, r_1 + r_2)_T$

2)  $k \cdot (l_1, a_1, b_1, r_1)_T = \begin{cases} (kl_1, ka_1, kb_1, kr_1)_T, & k \geq 0 \\ (-kr_1, kb_1, ka_1, -kl_1)_T, & k < 0 \end{cases}$

### 3. Probabilistic Fuzzy Regression

The fuzzy linear regression model can be state as

$$\tilde{Y}_i = \tilde{A}_0 + \tilde{A}_1 x_{i1} + \dots + \tilde{A}_k x_{ik} = \tilde{A} x_i, \quad i = 1, \dots, n \tag{1}$$

where  $\tilde{Y}_i = (l_{y_i}, a_{y_i}, b_{y_i}, r_{y_i})_T$  is fuzzy output value of the  $i^{\text{th}}$  observation,  $x_{ij}$  is crisp value of the  $j^{\text{th}}$  independent variable in the  $i^{\text{th}}$  observation,  $j = 1, \dots, k$ ,  $k$  is the number of independent variables; and  $x_{i0} = 1$ .  $\tilde{A}_j = (l_j, a_j, b_j, r_j)$  is the fuzzy coefficients.  $\tilde{Y}_i$  is the observed value.

#### 3.1. Introduction of Chaos Optimization Algorithm

As the characteristics of the sample directly reflect the overall population, the sample's numerical characteristics can be used to evaluate the overall population. The common numerical characteristics are expectation, variance, etc. Select the corresponding numerical characteristics according to the actual needs, determine the numerical characteristics to be determined using COA.

The chaos optimization algorithm proposed by Li [11], chaos is introduced into the design variable of the optimization problem using a similar carrier method, and the ergodic range of the chaotic motion is extended to the range of value of the design variables. Then, search by chaotic variables. COA employs chaotic dynamics to solve optimization problems and it has been applied successfully in various areas such as function optimization and supply chain

optimization [12]. Compared with conventional optimization methods, COA has faster convergence and can search for better solutions [13]. This algorithm also has an improved capacity to seek for the global optimal solution of an optimization problem and can escape from a local minimum. The characteristic of randomness ensures the capability for a large-scale search. Ergodicity allows COA to traverse all possible states without repetition and overcome the limitations caused by ergodic searching in general random methods. COA uses the carrier wave method to linearly map the selected chaos variables onto the space of optimization variables and then searches for the optimal solutions based on the ergodicity of the chaos variables.

### 3.2. Determination of Numerical Characteristic

The processes of applying COA in this study are described as follows.

First, the number of iterations of COA is defined. Each chaos variable represents numerical characteristics to be estimated, and the number of elements in a chaos variable is equal to the number of parameters to be determined. The chaos variable is initialized in which the values are selected randomly in the range  $[0,1]$ . The ranges of parameters  $[a,b]$  are initialized, in which  $a$  and  $b$  are the lower and upper limits of the optimization variable, respectively.

Second, the iteration number is set as  $m=1$ . Based on the initialized chaos variable, the logistic model used in COA in Equation (2), and logistic mapping can generate chaos variables through iteration.

$$c_m = f(c_{m-1}) = \mu c_{m-1} (1 - c_{m-1}) \quad (2)$$

where  $\mu$  is control parameter;  $c_m \in [0,1]$  is the  $m^{\text{th}}$  iterations value of the chaos variable  $c$ ; and  $c_0$  is the initialized chaos variable.

The linear mapping for converting chaos variables into optimization variables is formulated as follows:

$$q_m = a + (b - a)c_m \quad (3)$$

where  $q_m$  is the optimization variable and the value of  $q_m$  is the parameter settings to be determined. Based on the iteration, the chaos variable traverse between  $[0,1]$ , and the corresponding optimization variables traverse in the corresponding range  $[a,b]$ . In this case, the optimal solution can be identified in the area of feasible solutions. Based on the values of  $q_m$ , the numerical characteristics can be obtained. The model can be developed based on numerical characteristics and fuzzy coefficients by which the predicted output of  $\tilde{Y}_i = (l_{y_i}, a_{y_i}, b_{y_i}, r_{y_i})$  can be obtained. The predicted crisp output of  $\tilde{Y}_i$  is denoted as  $\hat{y}_i$ , which is the centroid of  $\tilde{Y}_i$ .

Third, the mean absolute percentage error (*MAPE*) is defined as the average of percentage errors, which is scale-independent and is a popular measure for evaluating accuracy [8]. Thus, *MAPE* was adopted in this study as the fitness function in COA, which is defined as follows:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{y_i} \cdot 100 \tag{4}$$

where  $\hat{y}_i$  is the centroid of  $\tilde{Y}_i$ ,  $y_i$  is the centroid of  $\tilde{Y}_i$ . The value of  $MAPE$  and  $q_m$  in the first iteration are recorded as the best fitness value  $fv^* = MAPE_1$  and the best solution  $q^* = q_1$ , respectively.

Fourth, the iteration continues by  $m+1 \rightarrow m$ . The chaos variable and optimization variable are updated by Equation (2) and Equation (3), respectively. The  $MAPE$  in the  $m+1$  iteration,  $MAPE_{m+1}$  is obtained using Equation (4). The process is as follows:

$$\begin{cases} fv^* = MAPE_{m+1}, q^* = q_{m+1} & MAPE \leq fv^* \\ fv^* = MAPE_m, q^* = q_m & \text{else} \end{cases} \tag{5}$$

### 3.3. Mathematical-Programming Model

Considering the random variable, the model in Equation (1) can be rewritten as follows:

$$(l_{y_i}, a_{y_i}, b_{y_i}, r_{y_i}) = (l_0, a_0, b_0, r_0) + (l_1, a_1, b_1, r_1)x'_{i1} + \dots + (l_k, a_k, b_k, r_k)x'_{ik} \tag{6}$$

$$a_{y_i} = \sum_{j=0, x_{ij} \geq 0}^k a_j x'_{ij} + \sum_{j=1, x_{ij} < 0}^k b_j (-x'_{ij}) \tag{7}$$

$$b_{y_i} = \sum_{j=0, x_{ij} \geq 0}^k b_j x'_{ij} + \sum_{j=1, x_{ij} < 0}^k a_j (-x'_{ij}) \tag{8}$$

$$l_{y_i} = \sum_{j=0, x_{ij} \geq 0}^k l_j x'_{ij} + \sum_{j=1, x_{ij} < 0}^k r_j (-x'_{ij}) \tag{9}$$

$$r_{y_i} = \sum_{j=0, x_{ij} \geq 0}^k r_j x'_{ij} + \sum_{j=1, x_{ij} < 0}^k l_j (-x'_{ij}) \tag{10}$$

where  $x'_{ij}$  is a certain numerical characteristic of  $x_{ij}$ ,  $i = 1, \dots, n; j = 1, \dots, k$ .

## 4. Risk in Model

In this section, three different risk preference are introduced, called risk-neutral, risk-averse and risk-seeking problems, to determine the fuzzy regression model.

### 4.1. Degree of Fitness from the Point of View Risk

Dobois and Prade [10] proposed the following equality indices to compare two fuzzy numbers.

$$Pos(\tilde{A} = \tilde{B}) = \sup_{x \in R} \{ \min \{ \tilde{A}(x), \tilde{B}(x) \} \} \tag{11}$$

$$Nes(\tilde{A} \subseteq \tilde{B}) = \inf_{x \in R} \{ \max \{ 1 - \tilde{A}(x), \tilde{B}(x) \} \} \tag{12}$$

where  $Pos$  and  $Nes$  are short for Possibility and Necessity. Observe that index Equation (11) means the degree of possibility that  $\tilde{A}$  is equal to  $\tilde{B}$  and index Equation (12) means the degree of necessity that  $\tilde{A}$  includes  $\tilde{B}$ . Let  $\tilde{Y}_i$  and  $\tilde{Y}_i$  represent the estimate value and the observed value, respectively. The degree

of fitness of the risk-neutral, risk-averse, risk-seeking denoted by  $f_i^{RN}$ ,  $f_i^{RA}$ ,  $f_i^{RS}$ , respectively and defined as,  $f_i^{RN} = Pos(\tilde{Y}_i = \tilde{Y}_i)$ ,  $f_i^{RA} = Nes(\tilde{Y}_i \subseteq \tilde{Y}_i)$ ,  $f_i^{RS} = Nes(\tilde{Y}_i \supseteq \tilde{Y}_i)$ . By considering  $\tilde{Y}_1, \tilde{Y}_2, \dots, \tilde{Y}_m$ , the risk-neutral, risk-averse and risk-seeking degree of fitness the estimate model, denoted by  $f^{RN}$ ,  $f^{RA}$ ,  $f^{RS}$ , respectively, are defined by  $f^{RN} = \min_i \{f_i^{RN}\}$ ,  $f^{RA} = \min_i \{f_i^{RA}\}$ ,  $f^{RS} = \min_i \{f_i^{RS}\}$ . From the properties of the two indices Equation (11) and Equation (12),  $f^{RN} \geq h$ ,  $f^{RA} \geq h$  and  $f^{RS} \geq h$  can be transformed to the usual inequalities for some fixed  $h \in [0, 1]$ .

**Proposition 4.1** [6] 1)  $f^{RN} \geq h$  if and only if

$$\tilde{Y}_{i,L}(h) \leq \tilde{Y}_{i,R}(h), \tilde{Y}_{i,R}(h) \geq \tilde{Y}_{i,L}(h), \quad i = 1, \dots, n$$

2)  $f^{RN} \geq h$  if and only if

$$\tilde{Y}_{i,L}(h) \leq \tilde{Y}_{i,R}(1-h), \tilde{Y}_{i,R}(h) \geq \tilde{Y}_{i,L}(1-h), \quad i = 1, \dots, n$$

3)  $f^{RS} \geq h$  if and only if

$$\tilde{Y}_{i,L}(1-h) \leq \tilde{Y}_{i,R}(h), \tilde{Y}_{i,R}(1-h) \geq \tilde{Y}_{i,L}(h), \quad i = 1, \dots, n$$

where  $\tilde{Y}_{i,L}(h)$ ,  $\tilde{Y}_{i,R}(h)$ ,  $\tilde{Y}_{i,L}(h)$ ,  $\tilde{Y}_{i,R}(h)$  are the left and right end point of fuzzy estimated output and observed output in the  $h$  level, respectively. Proposition 4.1 can be understood as

1)  $f^{RN} \geq h$  if and only if

$$\tilde{Y}_i(h) \cap \tilde{Y}_i(h) \neq \emptyset, \quad i = 1, \dots, n$$

2)  $f^{RA} \geq h$  if and only if

$$\tilde{Y}_i(1-h) \subseteq \tilde{Y}_i(h), \quad i = 1, \dots, n$$

3)  $f^{RS} \geq h$  if and only if

$$\tilde{Y}_i(1-h) \subseteq \tilde{Y}_i(h), \quad i = 1, \dots, n$$

**Corollary 4.1** If  $\max\{f^{RA}, f^{RS}\} \geq h$ ,  $\frac{1}{2} \leq h \leq 1$ , then  $f^{RN} \geq h$ .

### 4.2. Parameter Estimation of the Model

The objective function of the model is to minimize the difference between the total spread of observed and estimated values it is given by

$$J = \sum_{i=1}^n \left\{ \left( \tilde{Y}_{i,R}(0) - \tilde{Y}_{i,R}(0) \right)^2 + \left( \tilde{Y}_{i,L}(0) - \tilde{Y}_{i,L}(0) \right)^2 \right\} \quad (13)$$

Considering the above assumption, the problem is to obtain fuzzy parameters of risk-neutral, risk-averse and risk-seeking model in order to minimize  $J$  in Equation (13) subject to  $f^{RN} \geq h$ ,  $f^{RA} \geq h$  and  $f^{RS} \geq h$ , respectively. Consequently, the mathematical models lead to the following three quadratic programming problems:

RNP( $h$ ) (Risk—neutral problem)

$$\min : J = \sum_{i=1}^n \left\{ \left( \tilde{Y}_{i,R}(0) - \tilde{Y}_{i,R}(0) \right)^2 + \left( \tilde{Y}_{i,L}(0) - \tilde{Y}_{i,L}(0) \right)^2 \right\} \quad (14)$$

$$\begin{aligned} \text{s.t. } & \tilde{Y}_{i,L}(h) \leq \tilde{Y}_{i,R}(h), \tilde{Y}_{i,R}(h) \geq \tilde{Y}_{i,L}(h) \\ & \tilde{Y}_{i,R}(0) - \tilde{Y}_{i,L}(0) \geq 0, i = 1, \dots, n \end{aligned} \tag{15}$$

RNP( $h$ ) (Risk—averse problem)

$$\min : J = \sum_{i=1}^n \left\{ \left( \tilde{Y}_{i,R}(0) - \tilde{Y}_{i,R}(0) \right)^2 + \left( \tilde{Y}_{i,L}(0) - \tilde{Y}_{i,L}(0) \right)^2 \right\} \tag{16}$$

$$\begin{aligned} \text{s.t. } & \tilde{Y}_{i,L}(h) \leq \tilde{Y}_{i,R}(1-h), \tilde{Y}_{i,R}(h) \geq \tilde{Y}_{i,L}(1-h) \\ & \tilde{Y}_{i,R}(0) - \tilde{Y}_{i,L}(0) \geq 0, i = 1, \dots, n \end{aligned} \tag{17}$$

RNP( $h$ ) (Risk—seeking problem)

$$\min : J = \sum_{i=1}^n \left\{ \left( \tilde{Y}_{i,R}(0) - \tilde{Y}_{i,R}(0) \right)^2 + \left( \tilde{Y}_{i,L}(0) - \tilde{Y}_{i,L}(0) \right)^2 \right\} \tag{18}$$

$$\begin{aligned} \text{s.t. } & \tilde{Y}_{i,R}(1-h) \leq \tilde{Y}_{i,R}(h), \tilde{Y}_{i,L}(h) \geq \tilde{Y}_{i,L}(h) \\ & \tilde{Y}_{i,R}(0) - \tilde{Y}_{i,L}(0) \geq 0, i = 1, \dots, n \end{aligned} \tag{19}$$

### 4.3. Algorithm of Model

The algorithm of the proposed model is summarized below.

Step 1: The parameters are initialized, including the number of iterations, initial value of numerical characteristics, initialized chaos variables, and ranges of parameters.

Step 2: The structure of the proposed model is generated using Equation (8). The number of terms in the model is  $1+k$ , where  $k$  is the number of independent variables.

Step 3: The iteration begins from  $m=1$ . The chaos variables  $c_m$  are generated based on the logistic model in Equation (2) and transformed into optimization variables  $q_m$  using Equation (3).

Step 4: The interval of a random variable is defined based on the experimental data, and the corresponding numerical characteristics is selected. The numerical characteristic value of random variables are then generated based on the values of  $q_m$ . The random variables are substituted by their corresponding numerical characteristics, and the probabilistic terms of the proposed models are generated.

Step 5: The fuzzy coefficient of each term of the proposed model is determined by solving the quadratic programming problems shown in Equations (14) to (19).

Step 6: Predicted output  $\hat{y}_i$  is calculated with the developed proposed models. *MAPE* between estimate value  $\hat{y}_i$  and actual value  $y_i$  for all data sets can then be obtained using Equation (4) as the fitness value of the iteration  $m$ .

Step 7: The iteration is continued by  $m+1 \rightarrow m$  and stops after the number of iterations reaches the predefined value. The values of *MAPE* are obtained for each iteration and compared. The solution with the smallest fitness value is selected based on step 3 in Section 3.2. A model with the smallest error is then generated.

## 5. Empirical Studies

A stocking rating case was conducted to evaluate the effectiveness of the proposed approach. Securities analysts collect data to study macro economy and company, then forecast the earning and stocking value of the company being tracked. And provide five investment rating, which are buying, increasing, neutral, reducing, and selling for investors. Stock rating coefficient: 1.00 - 1.09 buying; 1.10 - 2.09 increasing; 2.10 - 3.09 neutral; 3.10 - 4.09 decreasing; 4.10 - 5.00 selling. Introducing fuzzy number to deal with the interval of different ratings. Therefore, the stock rating can be expressed as trapezoid fuzzy number as follows:

$$\text{“buying”} = (0.00, 1.00, 1.09, 0.01)_T$$

$$\text{“increasing”} = (0.01, 1.10, 2.09, 0.01)_T$$

$$\text{“neutral”} = (0.01, 2.10, 3.09, 0.01)_T$$

$$\text{“decreasing”} = (0.01, 3.10, 4.09, 0.01)_T$$

$$\text{“selling”} = (0.01, 4.10, 5.00, 0.00)_T$$

There are many factors that affect stock rating. However, based on the three principles of comprehensiveness, comparability and feasibility, we choose the following indicators to form an evaluation index system: profitability ( $x_1$ ), operational capacity ( $x_2$ ), short-term debt paying ability ( $x_3$ ) and volatility of the stock ( $x_4$ ).

This case study the relationship between stock rating ( $y$ ) and profitability ( $x_1$ ), operational capacity ( $x_2$ ), short-term debt paying ability ( $x_3$ ) and volatility of the stock ( $x_4$ ). In the stock rating forecast, the input is the crisp value and the output is the trapezoid fuzzy number. The value of stock price volatility is the standard deviation of the closing price of the stock from Sep 18th, 2017 to Oct 18th.

The sample data collected are shown in **Table 1**, **Table 2**, especially, the  $x_4$  in **Table 1** is the standard deviation of the sample. Then the confidence interval of volatility of the stock is calculated by MATLAB shown in **Table 3**.

### The Construction of the Model

After the data collection, based on the initialized stock volatility, fuzzy coefficient of different risks can be obtained by Equations (14) to (19). Generating new stock volatility by COA, Equation (4) are considered as the objective function. In the end, stock volatility based on different risks are shown in **Table 4** and **Table 5**. The results of the proposed models were compared with those of statistical regression (SR), Tanaka *et al.* [14] (denoted by TH), Diamond [1] (denoted by DM), Chiang Kao *et al.* [15] (denoted by KC) and Junhong Li *et al.* [9] (denoted by JL) to evaluate the proposed method effectiveness. *MAPE* and the variance of error (*VoE*) defined in Equations (4) and (20), respectively, were adopted to compare the modeling results of these approaches.



**Table 1.** The financial data of the stock.

ID	$x_1$	$x_2$	$x_3$	$x_4$	ID	$x_1$	$x_2$	$x_3$	$x_4$
1	29.98	0.98	0.37	0.12	14	43.12	0.84	2.21	0.08
2	22.88	1.01	0.86	0.08	15	37.49	2.22	1.50	1.03
3	48.57	1.50	0.49	0.54	16	1.50	0.63	2.12	1.19
4	26.20	2.53	0.22	0.95	17	10.05	0.73	1.89	0.03
5	149.13	1.07	0.64	0.11	18	40.87	0.65	3.32	0.24
6	99.59	1.75	0.96	1.15	19	9.63	0.35	3.20	0.13
7	76.79	0.91	1.22	0.56	20	3.08	0.43	0.72	0.04
8	1.72	2.02	0.06	0.07	21	23.10	1.35	1.78	1.31
9	6.46	0.56	0.23	0.12	22	20.97	1.25	1.43	0.39
10	65.67	0.17	0.11	0.60	23	21.99	1.07	0.67	0.11
11	58.65	0.91	0.09	0.19	24	19.87	2.54	0.23	0.22
12	197.61	2.64	0.43	0.25	25	35.34	2.22	1.37	0.32
13	95.72	1.33	0.79	0.33					

**Table 2.** Stocking rating.

ID	Rating	ID	Rating	ID	Rating
1	(0.00 1.00 1.09 0.01)	10	(0.00 1.00 1.09 0.01)	19	(0.01 3.10 4.09 0.01)
2	(0.01 2.10 3.09 0.01)	11	(0.00 1.00 1.09 0.01)	20	(0.01 3.10 4.09 0.01)
3	(0.00 1.00 1.09 0.01)	12	(0.00 1.00 1.09 0.01)	21	(0.01 3.10 4.09 0.01)
4	(0.00 1.00 1.09 0.01)	13	(0.01 1.10 2.09 0.01)	22	(0.01 2.10 3.09 0.01)
5	(0.00 1.00 1.09 0.01)	14	(0.01 1.10 2.09 0.01)	23	(0.01 3.10 4.09 0.01)
6	(0.01 1.10 2.09 0.01)	15	(0.00 1.00 1.09 0.01)	24	(0.01 1.10 2.02 0.01)
7	(0.01 1.10 2.09 0.01)	16	(0.01 3.10 4.09 0.01)	25	(0.01 3.10 4.09 0.01)
8	(0.00 1.00 1.09 0.01)	17	(0.01 2.10 3.09 0.01)		

$$VoE = \frac{1}{n-1} \sum_{i=1}^n \left( \frac{|\hat{y}_i - y_i|}{y_i} - MAPE \right)^2 \tag{20}$$

The same survey data was utilized to develop models based on SR, TH, DM, KC, JL. For SR, the centroid of fuzzy number are considered as fuzzy number. The models are produced by different methods as follows in **Table 6**.

**Table 7** shows that the values of *MAPE* and *VoE* based on different approaches, it shows that RNP outperforms the other approaches in modeling in terms of the *MAPE* and *VoE*.

## 6. Conclusion

In this paper, we developed three mathematical programming models, called risk-neutral, risk-averse, and risk-seeking, to studied fuzzy linear regression

**Table 3.** 95% confidence interval of stock volatility.

ID	Confidence Interval	ID	Confidence Interval	ID	Confidence Interval
1	[0.0940 0.2010]	10	[0.4539 0.9276]	19	[0.0734 0.1034]
2	[0.0577 0.1179]	11	[0.1431 0.2925]	20	[0.0276 0.0564]
3	[0.4031 0.8238]	12	[0.1924 0.3932]	21	[0.9000 1.8391]
4	[0.7527 1.4829]	13	[0.2474 0.5057]	22	[0.2895 0.5916]
5	[0.0758 0.1548]	14	[0.0651 0.1330]	23	[0.0865 0.1767]
6	[0.8851 1.8087]	15	[0.6015 1.2292]	24	[0.1550 0.3167]
7	[0.4293 0.8772]	16	[0.9011 1.8414]	25	[0.2229 0.4555]
8	[0.0553 0.1131]	17	[0.0293 0.0294]		
9	[0.0825 0.1686]	18	[0.1644 0.2844]		

**Table 4.** Stock volatility obtained by COA (RNP).

ID	Variance	ID	Variance	ID	Variance	ID	Variance
1	0.1737	8	0.0603	15	0.7842	22	0.5819
2	0.0996	9	0.1086	16	0.9935	23	0.1307
3	0.6695	10	0.8233	17	0.0293	24	0.2934
4	0.9967	11	0.1974	18	0.1844	25	0.2609
5	0.0813	12	0.2061	19	0.0934		
6	1.3950	13	0.4853	20	0.0292		
7	0.6433	14	0.0799	21	1.2637		

**Table 5.** Stock volatility obtained by COA (RAP).

ID	Variance	ID	Variance	ID	Variance	ID	Variance
1	0.1977	8	0.1024	15	0.7519	22	0.3215
2	0.1045	9	0.1090	16	1.5833	23	0.1449
3	0.4042	10	0.5274	17	0.0293	24	0.2303
4	1.1250	11	0.1826	18	0.1844	25	0.3179
5	0.1009	12	0.2223	19	0.0934		
6	1.1709	13	0.2794	20	0.0347		
7	0.8414	14	0.1223	21	1.4748		

analysis, and meanwhile considered the randomness caused by independent variables in the modeling. In the proposed approach, the numerical characteristics of random variables is decided by COA. Fuzzy regression analysis is then conducted to determine the fuzzy coefficients for all the terms of the

**Table 6.** Results from different approaches<sup>1</sup>.

Approaches	Developed models
SR	$\tilde{y}_{SR} = 1.8487 - 0.0062x_1 - 0.0557x_2 + 0.5399x_3 - 0.0385x_4$
TH	$\tilde{y}_{TH} = (1.9286 \times 10^{-8}, -0.0417, 1.0065, 6.5089 \times 10^{-9})$ $+ (2.0434 \times 10^{-8}, -0.0254, 1.0415, 2.1557 \times 10^{-8})x_1$ $+ (1.8247 \times 10^{-8}, -0.1749, 0.8416, 9.7891 \times 10^{-9})x_2$ $+ (1.9099 \times 10^{-8}, 0.0139, 0.9926, 6.8315 \times 10^{-9})x_3$ $+ (2.4991 \times 10^{-8}, 0.3568, 0.6446, 4.3585 \times 10^{-8})x_4$
DM	$\tilde{y}_{DM} = (0.0044, 1.8221, 2.3881, 0.0083)$ $+ (1.3791 \times 10^{-5}, 0.0063, 0.0068, 1.0424 \times 10^{-5})x_1$ $+ (0.0010, 0.0064, 0.0069, 0.0006)x_2$ $+ (1.9002 \times 10^{-6}, -0.420391, -0.420390, 1.9229 \times 10^{-6})x_3$ $+ (0.0016, -0.0632, -0.0148, 0.0012)x_4$
KC <sup>2</sup>	$\tilde{y}_{KC} = 0.6673 - 0.1242x_1 + 0.1517x_2 + 0.8083x_3 + 0.5304x_4$ $+ (0.2784, 0.2852, 0.9872, 0.9972)$
JL	$\tilde{y}_{JL} = (0.00016, 0.9514, 0.9537, 0.0016)$ $+ (3.4895 \times 10^{-6}, -0.0474, -0.0471, 3.4895 \times 10^{-6})x_1$ $+ (0.0001, 0.5746, 0.5760, 0.0001)x_2$ $+ (0.0001, 0.2136, 0.2155, 0.0001)x_3$ $+ (0.0004, 0.7414, 0.7454, 0.0001)x_4$
RNP	$\tilde{y}_{RNP} = (0.3066, 1.6750, 2.1384, 1.0213)$ $+ (6.8889 \times 10^{-10}, -0.007399967, -0.007399966, 6.8890 \times 10^{-10})x_1$ $+ (0.0106, -0.1169, -0.1067, 0.0106)x_2$ $+ (2.7728 \times 10^{-7}, 0.3528503, 0.3528505, 2.6929 \times 10^{-7})x_3$ $+ (6.6698 \times 10^{-8}, -0.30650488, -0.30650481, 6.7342 \times 10^{-8})x_4$
RAP	$\tilde{y}_{RAP} = (4.0794 \times 10^{-7}, 1.2521, 1.9009, 2.0041)$ $+ (9.8956 \times 10^{-9}, -0.01046018, -0.01046014, 3.4631 \times 10^{-8})x_1$ $+ (3.1431 \times 10^{-7}, 0.0188, 0.1190, 0.1158)x_2$ $+ (3.5577 \times 10^{-7}, 0.2115, 0.2508, 0.0335)x_3$ $+ (1.1538 \times 10^{-6}, -0.76236, -0.76235, 2.4335 \times 10^{-6})x_4$

**Table 7.** Means and variance of the validation errors.

Validation error	SR	TH	DM	KC	JL	RNP	RAP	RSP
<i>MAPE</i>	35	1623	73	40	135	32	37	* <sup>3</sup>
<i>VoE</i>	121	1,845,500	5223	1669	18,542	1120	1369	*

proposed models. The generated model can address the fuzziness caused by human subjective judgement and the randomness caused by independent

<sup>1</sup> $h = 0$  for all approaches.

<sup>2</sup>In the KC model, (0:2784; 0:2852; 0:9872; 0:9972) is all coordinate points.

<sup>3</sup>RSP is infeasible for each  $0 \leq h \leq 1$ .

variables. The results from the stock rating case indicates that the proposed models have better performance than the others methods.

### Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

### References

- [1] Tanaka, H. (1982) Linear Regression Analysis with Fuzzy Model. *IEEE Transactions on Systems, Man, and Cybernetics*, **12**, 903-907. <https://doi.org/10.1109/TSMC.1982.4308925>
- [2] Diamond, P. (1988) Fuzzy Least Squares. *Information Sciences*, **46**, 141-157. [https://doi.org/10.1016/0020-0255\(88\)90047-3](https://doi.org/10.1016/0020-0255(88)90047-3)
- [3] Tanaka, H. and Watada, J. (1988) Possibilistic Linear Systems and Their Application to the Linear Regression Model. *Fuzzy Sets & Systems*, **27**, 275-289. [https://doi.org/10.1016/0165-0114\(88\)90054-1](https://doi.org/10.1016/0165-0114(88)90054-1)
- [4] Dragon, A. (1991) Evaluation of Fuzzy Linear Regression Models. *Fuzzy Sets & Systems*, **35**, 51-63.
- [5] Kim, B. and Bishu, R.R. (1998) Evaluation of Fuzzy Linear Regression Models by Comparing Membership Functions. *Fuzzy Sets & Systems*, **100**, 343-352. [https://doi.org/10.1016/S0165-0114\(97\)00100-0](https://doi.org/10.1016/S0165-0114(97)00100-0)
- [6] Modarres, M., Nasrabadi, E. and Nasrabadi, M.M. (2004) Fuzzy Linear Regression Analysis from the Point of View Risk. *International Journal of Uncertainty, Fuzziness & Knowledge-Based Systems*, **12**, 635-649. <https://doi.org/10.1142/S0218488504003120>
- [7] Wong, C.K. and Chen, K.Y. (2010) A Generalized Fuzzy Least-Squares Regression Approach to Modeling Relationship in QFD. *Journal of Engineering Design*, **21**, 601-603. <https://doi.org/10.1080/09544820802563234>
- [8] Zhang, A.W. (2016) Statistical Analysis of Fuzzy Linear Regression Model Based on Centroid Method. *Applied Mathematics*, **7**, 579-586.
- [9] Li, J.H., Zeng, W.Y. and Xie, J.J. (2016) A New Fuzzy Regression Model Based on Least Absolute Deviation. *Engineering Applications of Artificial Intelligence*, **52**, 54-64. <https://doi.org/10.1016/j.engappai.2016.02.009>
- [10] Hu, B.Q. (2010) Fuzzy Theory Basis. 2nd Edition, Wuhan University Press, Wuhan.
- [11] Li, B. and Jiang, W.S. (1997) Chaos Optimization Algorithm and Its Application. *Control Theory and Application*, **4**, 613-615.
- [12] Nishikant, M., Choudhary, K.A. and Tiwari, M.K. (2008) Modeling the Planning and Scheduling across the Outsourcing Supply Chain: A Chaos-Based Fast Tabu—SA Approach. *International Journal of Production Research*, **46**, 3683-3715.
- [13] Nanba, R., Hasegawa, M. and Nishita, T. (2002) Optimization Using Chaotic Neural Network and Its Application to Lighting Design. *Control & Cybernetics*, **2**, 249-269.
- [14] Tanaka, H. and Hayashi, I. (1989) Possibilistic Linear Regression Analysis for Fuzzy Data. *European Journal of Operational Research*, **40**, 389-396. [https://doi.org/10.1016/0377-2217\(89\)90431-1](https://doi.org/10.1016/0377-2217(89)90431-1)
- [15] Kao, C.A. and Chyu, C.-L. (2008) Least-Squares Estimates in Fuzzy Regression Analysis. *European Journal of Operational Research*, **148**, 426-435.