Fuzzy Correlation and Fuzzy Non-linear Regression Analysis

Murat Alper Basaran, Biagio Simonetti and Luigi D'Ambra

Abstract In this chapter, we will deal with fuzzy correlation and fuzzy non-linear regression analyses. Both correlation and regression analyses that are useful and widely employed statistical tools have been redefined in the framework of fuzzy set theory in order to comprehend relation and to model observations of variables collected as either qualitative or approximately known quantities which are no longer being utilized directly in classical sense. When fuzzy correlation and fuzzy non-linear regression are concern, dealing with several computational complexities emerging due to the nature of fuzzy set theory is a challenge. It should be noted that there is no well-established formula or method in order to calculate fuzzy correlation coefficient or to estimate parameters of the fuzzy regression model. Therefore, a rich literature will accompany with the readers. While extension principle based methods are utilized in the computational procedures for fuzzy correlation coefficient, the distance based methods preferred rather than mathematical programming ones are employed in parameter estimation of fuzzy regression models. That extension principle combined with either fuzzy arithmetic or non-linear programming is two different methods proposed in the literature will be examined with small but illustrative examples in detail for fuzzy correlation analysis. Fuzzy non-linear regression has been a relatively new studied method when compared to fuzzy linear regression. However, both employ similar tools. S-curve fuzzy regression and two types of quadratic fuzzy regression models in the literature will be discussed.

M.A. Basaran (✉)
Department of Management Engineering, Alanya Alaaddin Keykubat University,
Antalya, Turkey
e-mail: muratalper@yahoo.com

B. Simonetti
Department of Economics Juridical and Social System Studies,
University of Sannio, Benevento, Italy
e-mail: simonetti@unisannio.edu.it

L. D'Ambra
Department of Economics Management and Institution,
University of Naples Federico II, Naples, Italy
e-mail: dambra@unina.edu.it

1   Introduction

Both correlation analysis and regression analysis are two of the most applied statistical tools in several disciplines due to its applicability and interpretability. They allow certain types of measurements to be used in classical statistical theory which means that observation are supposed to follow certain distributions. However, encountering observations either described by linguistic terms such as "bad", "good" and "very good", or approximately known quantities such as "around 2" is possible. With the introduction of fuzzy set theory, uncertainty different than one defined by probabilistic framework being modeled with possibility distribution for data collected as either qualitative or approximately known quantities has been a research area for data analysts.

Extending both methods to fuzzy framework gives rise to several proposed methods utilizing different aspects of fuzzy set theory.

2   Fuzzy Correlation Analysis

Correlation coefficient is a statistical measure which determines both the direction and strength of the linear relation between two variables which is defined by

$$r_{XY} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{1}$$

where X and Y are variables whose values are denoted by $(x_i, y_i), i = 1, 2, \ldots, n$ and their corresponding arithmetic means are denoted by $\bar{x}$ and $\bar{y}$ respectively. Its range restricted in a closed interval $[-1, 1]$ tells how strong the linear dependence is between those variables with the knowledge of direction.

When the correlation coefficient is reconsidered in the fuzzy setting which means that observation values either are qualitative knowledge such as linguistic terms taking values of, for example, "bad" or "good" or "excellent", or are approximately known values, for instance, the value of the quantity can be defined around 2, measuring it is a need to quantify the relation. Both types of data are encountered when subjective or linguistic evaluations are provided by experts in the field of engineering, management or social sciences [1–3]. For example, the need for fuzzy correlation measure can arise when to quantify relation between the technology level and the management achievements of firms in management science or when to

partition images to determine similarity or dissimilarity is concern in the field of engineering. Indeed, these types of exemplifications can easily be extended to any disciplines. Therefore, measuring correlation coefficient between two variables involving fuzziness is a need and computational procedures are challenging than that given in (1).

Computing fuzzy correlation employs basically two different methods. The first of which is to rely on Zadeh's extension principle, which aims at finding the membership function of fuzzy correlation. In order to determine membership function of fuzzy correlation, some methods are available providing with both analytical and numerical solutions, for example, using weakest t-norm and non-linear programming. Before explaining the details of the methods that are utilized in the computation of fuzzy correlation as well as fuzzy non-linear regression, some preliminary notions and definitions are needed which are fuzzy numbers, *LR* type fuzzy numbers, α-cuts of a fuzzy set and triangular norm, namely, t-norm, Zadeh's extension principle, fuzzy arithmetic. More detailed treatment of the subjects mentioned above can be found in variety of books pertinent to fuzzy set theory or fuzzy logic [4].

A fuzzy number is a convex subset of the real line R with a normalized membership function. For example, an asymmetric triangular fuzzy number $\tilde{x} = (x, \alpha, \beta)$ is defined by

$$\tilde{x}(t) = \begin{cases} 1 - \frac{x-t}{\alpha}, & \textit{if } x - \alpha \leq t \leq x \\ 1 - \frac{t-x}{\beta}, & \textit{if } x \leq t \leq x + \beta \\ 0, & \textit{otherwise} \end{cases} \tag{2}$$

where the center value $x \in R$, left spread value $\alpha > 0$, and right spread value $\beta > 0$ are based on the definition of fuzzy number. When $\alpha = \beta$ is assumed, an asymmetric triangular fuzzy number is called a symmetric triangular fuzzy number and is denoted by $\tilde{x} = (x, \alpha)$. Other types of fuzzy numbers such as trapezoidal fuzzy number and Gaussian fuzzy number are also defined and utilized in various applications dependent upon the suitability, interpretability, and applicability.

A fuzzy number $\tilde{x} = (x, \alpha)_{LR}$ of type LR is a function from real numbers into the interval [0, 1] defined by

$$\tilde{x}(t) = \begin{cases} L\left(\frac{x-t}{\alpha}\right) & \textit{for } x - \alpha \leq t \leq x \\ R\left(\frac{t-x}{\beta}\right) & \textit{for } x \leq t \leq x + \beta \end{cases} \tag{3}$$

where L and R are non-increasing and continuous shape functions from [0,1] to [0,1] satisfying $L(0) = R(0) = 1$ and $L(1) = R(1) = 0$.

An α-cut of a fuzzy set is a crisp set defined by

$$A_\alpha = \{x \in A | \mu_A(x) \geq \alpha\} \tag{4}$$

A binary operation T on unit interval is said to be a triangular norm or t-norm if and only if T is associative, commutative, non-decreasing and $T(x, 1) = x$ for each $x \in [0, 1]$.

Extending ordinary arithmetic into fuzzy number setting is possible by employing Zadeh's extension principle defined by

$$\mu_B(y) = \underbrace{Sup}_{\substack{(x_1, \ldots, x_n) \in U_1 x \ldots x U_n \\ y = f(x_1, \ldots, x_n)}} \min(\mu_{A_1}(x_1), \ldots \mu_{A_n}(x_n)) \qquad (5)$$

where $A = A_1 x \ldots x A_n$ and $U = U_1 x \ldots x U_n$ are Cartesian product of the fuzzy sets $A_i, (i = 1, \ldots, n)$ and universal sets $U_i, (i = 1, \ldots, n)$ of fuzzy sets respectively.

2.1 Fuzzy Correlation Coefficient Based on the Weakest t-Norm (T_w) and Fuzzy Arithmetic

When Zadeh's extension principle is rewritten using one of union operators such as t-norm instead of minimization, the arithmetic operators are defined by

$$(\tilde{A} \oplus \tilde{B}) = \underbrace{Sup}_{x+y=z} T(\tilde{A}(x), \tilde{B}(y)) \qquad (6a)$$

$$(\tilde{A} \otimes \tilde{B}) = \underbrace{Sup}_{x \cdot y=z} T(\tilde{A}(x), \tilde{B}(y)) \qquad (6b)$$

$$(\tilde{A} \o \tilde{B}) = \underbrace{Sup}_{x/y=z} T(\tilde{A}(x), \tilde{B}(y)) \qquad (6c)$$

where $\tilde{A}$ and $\tilde{B}$ are fuzzy numbers and $\oplus$, $\otimes$, $\oslash$ are fuzzy arithmetic operators for addition, multiplication and division, respectively.

When fuzzy correlation is being computed, applying the extension principle based on the weakest t-norm denoted by T_w for a sample of n independent pairs of *LR* type fuzzy numbers is the method using the classical definition of the correlation coefficient given in (1) [5]. Instead of using the union operator *Sup*, T_w based fuzzy addition and multiplication are preferred in order to preserve the shape of the resultant *LR* type fuzzy numbers since it is the fact that fuzzy multiplication and division operators lead to resultant fuzzy numbers different than *LR* types except fuzzy addition and subtraction.

When the observations are fuzzy, the sample correlation coefficient given in (1) is rewritten.

$$\tilde{r}_{\widetilde{X,Y}} = \frac{\sum_{i=1}^{n}(\tilde{x}_i - \frac{1^{\circ}}{n}\sum_{i=1}^{n}\tilde{x}_i)(\tilde{y}_i - \frac{1^{\circ}}{n}\sum_{i=1}^{n}\tilde{y}_i)}{\sqrt{\sum_{i=1}^{n}(\tilde{x}_i - \frac{1^{\circ}}{n}\sum_{i=1}^{n}\tilde{x}_i)^2 \sum_{i=1}^{n}(\tilde{y}_i - \frac{1^{\circ}}{n}\sum_{i=1}^{n}\tilde{y}_i)^2}} \tag{7}$$

where $\tilde{x}_i = (x_i, \gamma_i)$ and $\tilde{y}_i = (y_i, \delta_i)$, $i = 1, 2, \ldots, n$, are symmetric triangular fuzzy numbers and $\bar{\tilde{x}} = \frac{1^{\circ}}{n}\sum_{i=1}^{n}\tilde{x}_i$ and $\bar{\tilde{y}} = \frac{1^{\circ}}{n}\sum_{i=1}^{n}\tilde{y}_i$ are the average values of fuzzy numbers $\tilde{X}$ and $\tilde{Y}$, respectively. Then the average values of fuzzy numbers $\tilde{X}$ and $\tilde{Y}$ are calculated based on T_w as follows:

$$\bar{\tilde{x}} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i, \max_{1 \leq i \leq n} \gamma_i\right) \tag{8a}$$

$$\bar{\tilde{y}} = \left(\frac{1}{n}\sum_{i=1}^{n} y_i, \max_{1 \leq i \leq n} \delta_i\right) \tag{8b}$$

The expressions given in (8a) and (8b) can be written for just some observation using T_w as follows:

$$(\tilde{x} - \bar{\tilde{x}}_i) = \left(x_i - \frac{1}{n}\sum_{i=1}^{n} x_i, \max_{1 \leq i \leq n} \gamma_i\right)_L \tag{9a}$$

$$(\tilde{y} - \bar{\tilde{y}}_i) = \left(y_i - \frac{1}{n}\sum_{i=1}^{n} y_i, \max_{1 \leq i \leq n} \delta_i\right)_L \tag{9b}$$

Then the product of (9a) and (9b) is obtained as follows:

$$\left(\left(x_i - \frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(y_i - \frac{1}{n}\sum_{i=1}^{n} y_i\right), \max\left|x_i - \frac{1}{n}\sum_{i=1}^{n} x_i\right| \max_{1 \leq k \leq n} \delta_k, \left(y_i - \frac{1}{n}\sum_{i=1}^{n} y_i\right) \max_{1 \leq k \leq n} \gamma_k\right)_L \tag{10}$$

The numerator of (7) is the summation of the product of (9a) and (9b) using T_w based fuzzy arithmetic denoted by

$$\left(\sum_{i=1}^{n}\left(x_i - \frac{1}{n}\sum_{k=1}^{n} x_k\right)\sum_{i=1}^{n}\left(y_i - \frac{1}{n}\sum_{k=1}^{n} y_k\right), \max_{1 \leq i \leq n}\left|x_i - \frac{1}{n}\sum_{k=1}^{n} x_k\right| \max_{1 \leq k \leq n} \delta_k, \left|x_i - \frac{1}{n}\sum_{k=1}^{n} x_k\right| \max_{1 \leq k \leq n} \gamma_k\right) \tag{11}$$

In order to compute the denominator of (7), we will follow the similar steps. The summation of the square of the differences between fuzzy observations and its fuzzy arithmetic mean for each variable is denoted using T_w based fuzzy arithmetic in (12) and (13).

$$\left(\sum_{i=1}^{n} \left(x_i - \frac{1}{n} \sum_{i=1}^{n} x_k \right)^2, \quad \max_{1 \le i \le n} \left| x_i - \frac{1}{n} \sum_{i=1}^{n} x_k \right| \max_{1 \le k \le n} \gamma_k \right)_L \tag{12}$$

$$\left(\sum_{i=1}^{n} \left(y_i - \frac{1}{n} \sum_{i=1}^{n} y_k \right)^2, \quad \max_{1 \le i \le n} \left| y_i - \frac{1}{n} \sum_{i=1}^{n} x_k \right| \max_{1 \le k \le n} \delta_k \right)_L \tag{13}$$

The product of (12) and (13) yields (14) and (15)

$$\sqrt{ \sum_{i=1}^{n} \left(x_i - \frac{1}{n} \sum_{i=1}^{n} x_k \right)^2 \sum_{i=1}^{n} \left(y_i - \frac{1}{n} \sum_{i=1}^{n} y_k \right)^2 }, \tag{14}$$

$$\frac{\max\{ \sum_{i=1}^{n} (x_i - \frac{1}{n} \sum_{i=1}^{n} x_k)^2 \max_{1 \le i \le n} \left| y_i - \frac{1}{n} \sum_{i=1}^{n} y_k \right| \max_{1 \le k \le n} \delta_k, \\ \sum_{i=1}^{n} (y_i - \frac{1}{n} \sum_{i=1}^{n} y_k)^2 \max_{1 \le i \le n} \left| x_i - \frac{1}{n} \sum_{i=1}^{n} x_k \right| \max_{1 \le k \le n} \gamma_k,}{\sum_{i=1}^{n} (x_i - \frac{1}{n} \sum_{i=1}^{n} x_k)^2 \sum_{i=1}^{n} (y_i - \frac{1}{n} \sum_{i=1}^{n} y_k)^2} \tag{15}$$

where expressions in (14) and (15) are center and the spread part of the fuzzy number in denominator of (7), respectively.

Hence, both numerator and denominator are obtained. The last step is to divide those two fuzzy numbers. Its division is simply based on the implementation of the expression given in (6c). It is denoted by

$$(\tilde{A} \oslash \tilde{B})(z) = \begin{cases} L\left[\frac{\left(\frac{a}{b} - z \right)}{\left(\left(\frac{1}{b} \right) \max(\alpha, z\beta) \right)} \right], & z \ge \min\left\{ \left(\frac{a - \alpha}{b}, \frac{a}{b + \beta} \right) \right\} \\ R\left[\frac{\left(z - \frac{a}{b} \right)}{\left(\frac{1}{b} \right) \max(\alpha, z\beta)} \right], & z \le \max\left\{ \frac{(a + \alpha)}{b}, \frac{a}{(b - \beta)} \right\} \end{cases} \tag{16}$$

where $a, b > 0$ and it is assumed that $L = R$, also $\tilde{A} = (a, \alpha)_{LL}$ and $\tilde{B} = (b, \beta)_{LL}$ are fuzzy numbers. Also, other cases including the different signs of two fuzzy numbers are easily defined and given in [5]. It should be noted that expression given in (16) holds for LL types fuzzy numbers.

A small data set presented in Table 1 will be used in order to exemplify calculations.

Table 1 Data set for both fuzzy numbers written in the form of symmetric triangular fuzzy numbers

$\tilde{x}_i = (x_i, \gamma_i)$	$\tilde{y}_i = (y_i, \delta_i)$
(2.5,0.10)	(2.0,0.2)
(3.0,0.4)	(2.6,0.3)
(3.2,0.3)	(2.9,0.5)
(3.5,0.2)	(3.8,0.4)
(4.1,0.5)	(6.0,0.60)

Fuzzy arithmetic for both variables are obtained as $\bar{\bar{X}} = (3.26, 0.5)$ and $\bar{\bar{Y}} = (3.46, 0.6)$. Then using expression (11) results in (3.582, 1.27) which is the enumerator of (7). The denominator is calculated using (14) and (15) leading to (3.72, 2.71). When former value one is divided by the latter one, the membership function for correlation coefficient is denoted by

$$\tilde{r}_{\widetilde{XY}} == \frac{(3.58, 1.27)}{(3.72, 2.71)} = \begin{cases} 1 - \frac{0.96-z}{\max(0.341, 0.728z)} & \text{if } 0.341 \le z \le 0.96 \\ 1 - \frac{z-0.96}{\max(0.341, 0.728z)} & \text{if } 0.96 \le z \le 1.304 \end{cases}$$

2.2 Fuzzy Correlation Based on Zadeh's Extension Principle

Another approach in the computation of fuzzy correlation coefficient is to use the α-cuts of fuzzy numbers in order to derive the membership function proposed by [6]. This method relies on the application of the extension principle aiming at finding the α-cuts of $\tilde{r}_{\widetilde{X,Y}}$. The α-cuts of $\tilde{X}_i$ and $\tilde{Y}_i$ are denoted by

$$(X_i)_\alpha = \left[(X_i)^L_\alpha, (X_i)^L_\alpha \right] = [\min_x \{x_i | \mu_{\tilde{x}_i}(x_i) \ge \alpha\}, \max_x \{x_i | \mu_{\tilde{x}_i}(x_i) \ge \alpha\}] \qquad (17a)$$

$$(Y_i)_\alpha = \left[(Y_i)^L_\alpha, (Y_i)^L_\alpha \right] = [\min_y \{y_i | \mu_{\tilde{y}_i}(y_i) \ge \alpha\}, \max_y \{y_i | \mu_{\tilde{y}_i}(y_i) \ge \alpha\}] \qquad (17b)$$

Also, its interval form containing the values of both variables are denoted by

$$[(X_i)_\alpha, (Y_i)_\alpha] = [\min_x \{x_i | \mu_{\tilde{x}_i}(x_i) \ge \alpha\}, \max_x \{x_i | \mu_{\tilde{x}_i}(x_i) \ge \alpha\}] \qquad (18)$$

where the α-cuts of $\tilde{X}_i$ and $\tilde{Y}_i$ are both crisp sets.

Then as mentioned in [6], a pair of non-linear mathematical programs are introduced in order to find the lower and upper bounds of the α-cuts of $\tilde{r}_{\widetilde{X,Y}}$. Those are denoted as follows:

$$(r_{XY})^L_\alpha = \min \left[\left(\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \right) \bigg/ \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2} \right]$$
$$s.t \quad (X_i)^L_\alpha \le x_i \le (X_i)^U_\alpha, \quad \forall i$$
$$(Y_i)^L_\alpha \le x_i \le (Y_i)^U_\alpha, \quad \forall i$$

$$(19a)$$

$$(r_{XY})_\alpha^U = \max\left[\left(\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})\middle/ \sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}\right)\right]$$

$$s.t \quad (X_i)_\alpha^L \le x_i \le (X_i)_\alpha^U, \quad \forall i$$
$$(Y_i)_\alpha^L \le x_i \le (Y_i)_\alpha^U, \quad \forall i$$

(19b)

In the case of nonexistence of analytic solutions of non-linear programming problems, it is possible to obtain the numeric solutions for $(r_{XY})_\alpha^L$ and $(r_{XY})_\alpha^U$ at different α levels, which leads to the approximate shape of $L(r)$ and $R(r)$. A small data set which is given in Table 1 will be used to exemplify.

In order to work with a pair of non-linear programming problems, the α-cuts of variables for specified values (α = 0.0, 0.1, …, 0.9, 1.0) are tabulated in Tables 2 and 3, respectively.

For each α value, while the first column shows the left end point, the second column denotes the right end point. Similar construction is made for the fuzzy $\tilde{Y}$ variable in Table 3.

Based on those values presented in Tables 2 and 3, a pair of non-linear programming problem is solved in order to calculate correlation values for each corresponding α-cut values. Those are tabulated in Table 4.

Table 2 The α-cuts values for $\tilde{X}$

α = 0.0		α = 0.1		α = 0.2		α = 0.3		α = 0.4		α = 0.5	
2.4	2.6	2.41	2.59	2.42	2.58	2.4	2.6	2.41	2.59	2.42	2.58
2.6	3.4	2.64	3.36	2.68	3.32	2.6	3.4	2.64	3.36	2.68	3.32
2.9	3.5	2.93	3.47	2.96	3.44	2.9	3.5	2.93	3.47	2.96	3.44
3.3	3.7	3.32	3.68	3.34	3.66	3.3	3.7	3.32	3.68	3.34	3.66
3.6	4.6	3.65	4.55	3.70	4.50	3.6	4.6	3.65	4.55	3.70	4.50

α = 0.6		α = 0.7		α = 0.8		α = 0.9		α = 1.0			
2.46	2.54	2.47	2.53	2.48	2.46	2.54	2.47	2.53	2.48		
2.84	3.16	2.88	3.12	2.92	2.84	3.16	2.88	3.12	2.92		
3.08	3.32	3.11	3.29	3.14	3.08	3.32	3.11	3.29	3.14		
3.42	3.58	3.44	3.56	3.46	3.42	3.58	3.44	3.56	3.46		
3.90	4.30	3.95	4.25	4.00	3.90	4.30	3.95	4.25	4.00		

Table 3 The α-cuts values for $\tilde{Y}$

α = 0.0		α = 0.1		α = 0.2		α = 0.3		α = 0.4		α = 0.5	
1.8	2.2	1.82	2.18	1.84	2.16	1.86	2.14	1.88	2.12	1.9	2.1
2.3	2.9	2.33	2.87	2.36	2.84	2.39	2.81	2.42	2.78	2.45	2.75
2.4	3.4	2.45	3.35	2.5	3.30	2.55	3.25	2.6	3.2	2.65	3.15
3.4	4.2	3.44	4.16	3.48	4.12	3.52	4.08	3.56	4.04	3.6	4.0
5.4	6.6	5.46	6.54	5.52	6.48	5.58	6.42	5.64	6.36	5.7	6.3
α = 0.6		α = 0.7		α = 0.8		α = 0.9		α = 1.0			
1.92	2.08	1.94	2.06	1.96	2.04	1.98	2.02	2.0	2.0		
2.48	2.72	2.51	2.69	2.54	2.66	2.57	2.63	2.6	2.6		
2.7	3.1	2.75	3.05	2.8	3	2.85	2.95	2.9	2.9		
3.64	3.96	3.68	3.92	3.72	3.88	3.76	3.84	3.8	3.8		
5.76	6.24	5.82	6.18	5.88	6.12	5.94	6.06	6.0	6.0		

3 Fuzzy Non-linear Regressions

Fuzzy linear regression has been utilized as a modeling technique since the first introduction by [7] when one encounters different settings such as linguistically defined values, small data sets, unknown structure between dependent variable and independent variables, approximate measurements like intervals. Modeling endeavor covers several applications in many disciplines ranging from quality function deployment to determining claiming reserves [8, 9]. Also, it allows crisp numbers to be utilized in the modeling. Therefore, several types of fuzzy linear regression models and their parameter estimation methods have been proposed. The generic form of it can be denoted by

$$\tilde{Y} = \tilde{A}_0 + \tilde{A}_1 \otimes \tilde{X}_1 + \cdots + \tilde{A}_n \otimes \tilde{X}_k, \quad i = 1, \ldots, n \qquad (20)$$

where parameters, dependent and independent variables are all fuzzy numbers represented as one of the types such as triangular, trapezoidal and Gaussian fuzzy numbers.

Despite of the fact that several estimation methods have been defined, they are actually being grouped into two different methods that have been utilized and evolved during the research. The first of which is based on mathematical programming methods such as linear programming, goal programming, and non-linear programming and so on. The second one is to rely on the minimization of distance between two fuzzy sets so-called fuzzy least squares, which are the squares of the differences between the observed and estimated values of dependent variable.

When fuzzy non-linear regression is concern, the same variety pertinent to model types and their estimation methods are encountered. In this chapter, two different types of fuzzy non-linear regression models that are available in the literature will be presented with small but illustrative examples. The first one is called S-shaped curve fuzzy regression whose crisp version is widely utilized in the modeling of complex

Table 4 Upper and lower correlation values for each corresponding α-cuts

α-cuts	α = 0.0	α = 0.1	α = 0.2	α = 0.3	α = 0.4	α = 0.5	α = 0.6	α = 0.7	α = 0.8	α = 0.9	α = 1
Min	0.4977	0.5663	0.6288	0.6891	0.7460	0.8024	0.8457	0.8865	0.9113	0.9418	0.9653
Max	1.00	1.00	1.00	1.00	1.00	1.00	0.9999	0.9980	0.9926	0.9821	0.9653

systems such as biology, agriculture and social economy. Both input variable and output variable in this model are fuzzy numbers. Its parameter estimation method is based upon minimizing the distance between fuzzy observed values and fuzzy estimated values which are represented by the pre-defined model. The parameter estimates are obtained as crisp values. The second one is called quadratic fuzzy regression model which appears to be two different types. While the first quadratic fuzzy regression allows quadratic term to be included in the model, the second one has interaction terms. While input variables have crisp values, the output variable and the parameters are fuzzy values. Its parameter estimation method uses the distance based methods aiming at minimizing the difference between observed values and estimated values proposed by [10, 12, 13].

3.1 S-Shaped Curve Fuzzy Regression

S-shaped curve fuzzy regression was proposed by [11] in order to model observations that are encountered in complex systems such as biology, social economy and agricultural sciences where the trend of growing is experienced slowly at the beginning, rapid increments are observed during process and it finishes with the saturation at the last phase.

Suppose that $\tilde{x}_i$ and $\tilde{y}_i, (i = 1, \ldots, n)$ are observations that are tried to be modeled defined by

$$\tilde{y} = (a + b \cdot \exp(-\tilde{x}))^{-1}, \quad a, b \in R \tag{21}$$

It is assumed that least squares based metrics between fuzzy numbers has better estimation ability when parameter estimation in fuzzy non-linear regression is concern. Therefore, metric defined in (22) will be utilized to determine parameters of the model given in (21).

$$\tilde{d}(\tilde{A}, \tilde{B}) = \left[\int_0^1 w^2(\alpha) d^2(A_\alpha, B_\alpha) dt \right]^{\frac{1}{2}}, \quad \tilde{A}, \tilde{B} \in F(R) \tag{22}$$

where $w^2(\alpha)$ should be chosen as a monotone increasing function in [0, 1], and $\tilde{A}$ and $\tilde{B}$ are fuzzy numbers defined on real line denoted by $F(R)$.

The motivation behind choosing monotone increasing function is based on the desire of having higher degree of membership level set when determining the distance between fuzzy numbers.

The distance based on the α-cuts of $\tilde{A}$ and $\tilde{B}$ given in (22) is denoted by

$$d^2 = (\tilde{A}_\alpha, \tilde{B}_\alpha) = [l(\alpha) - p(\alpha)]^2 + [r(\alpha) - q(\alpha)]^2 \tag{23}$$

where $\tilde{A}_\alpha = [l(\alpha), r(\alpha)]$ and $\tilde{B}_\alpha = [p(\alpha), q(\alpha)]$.

Utilizing the metric and the model given in (21) and (22) respectively, the least squares optimization problem is written in (24).

$$Minimize \; M(a,b) = \sum_{i=1}^{n} \tilde{d}^2 \left(a + b \exp(-\tilde{x}_i), \frac{1}{\tilde{y}_i} \right) \tag{24}$$

The α-cuts of functions of $\tilde{X}$ and $\tilde{Y}$ are represented as follows:

$$(\tilde{Y}_i)_\alpha = [f_i(\alpha), \; g_i(\alpha)], \; (\tilde{X}_i)_\alpha = [u_i(\alpha), \; v_i(\alpha)],$$

$$\left(\frac{1}{\tilde{Y}_i} \right)_\alpha = \left[\frac{1}{g_i(\alpha)}, \frac{1}{f_i(\alpha)} \right] (\exp(-\tilde{x}_i))_\alpha = [\exp(-v_i), \exp(-u_i)], \tag{25}$$

$$(\exp(\tilde{x}_i))_\alpha = [\exp(u_i), \exp(v_i)], \quad \alpha \in (0, 1]$$

where expression in (25) holds for positive fuzzy numbers.

Two different minimization functions are defined with respect to the sign of b, which are for $b \geq 0$ and $b < 0$, respectively.

In order to simplify the notations, the α-cut in parenthesis is removed. Also, w^2 is adapted instead of using $w^2(\alpha)$ in (22) and (25).

For $b \geq 0$, the α-cut of $(a + b \exp(-\tilde{x}_i))_\alpha$ is denoted by

$$[a + b \exp(-v_i), a + b \exp(-u_i)], \quad (i = 1, \ldots, n) \tag{26}$$

Then, its least squares optimization function given in (24) is rewritten

$$\min_{a,b} M_+ (a, b) = \sum_{i=1}^{n} \tilde{d}^2 \left(a + b \; exp(-x_i), \frac{1}{\tilde{y}_i} \right)$$

$$= \int_0^1 w^2 \left[\sum_{i=1}^{n} \left(a + b \; \exp(-v_i) - \frac{1}{g_i} \right)^2 + \left(a + b \exp(-u_i) - \frac{1}{f_i} \right)^2 \right] d\alpha \tag{27}$$

By taking derivatives of optimization function given in (27) with respect to parameters a and b, an equation system consisting two equations are obtained. The first equation system is denoted by ES1

$$ES1 = \begin{cases} 2na \int_0^1 w^2 d\alpha + b \int_0^1 w^2 \sum_{i=1}^{n} (\exp(-v_i) + \exp(-u_i)) d\alpha \\ \qquad = \int_0^1 w^2 \sum_{i=1}^{n} \left(\frac{1}{f_i} + \frac{1}{g_i} \right) d\alpha \\ a \int_0^1 w^2 \sum_{i=1}^{n} (\exp(-v_i) + \exp(-u_i)) d\alpha + b \int_0^1 w^2) \sum_{i=1}^{n} (\exp(-2v_i) + \exp(-2u_i)) d\alpha \\ \qquad = \int_0^1 w^2 \sum_{i=1}^{n} \left(\frac{1}{f_i} \exp(-2v_i + \frac{1}{g_i} \exp(-2u_i) \right) d\alpha \end{cases} \tag{28}$$

For $b < 0$, the α-cut of $(a + b \exp(-\tilde{x}_i))_\alpha$ is denoted by

$$[a + b \exp(-u_i), a + b \exp(-v_i)], \qquad i = 1, \ldots, n) \tag{29}$$

Then, its least squares optimization function given in (24) is rewritten for this case.

$$
\min_{a,b} M_-(a,b) = \sum_{i=1}^{n} \tilde{d}^2 \left(a + b \, exp(-x_i), \frac{1}{\tilde{y}_i} \right)
$$
$$
= \int_0^1 w^2 \sum_{i=1}^{n} \left[\left(a + b \, \exp(-u_i) - \frac{1}{g_i} \right)^2 + \left(a + b \exp(-v_i) - \frac{1}{f_i} \right)^2 \right] d\alpha
\tag{30}
$$

By taking derivatives of optimization function given in (30) with respect to parameters a and b, an equation system consisting two equations are obtained. The second equation system is denoted by ES2

$$
ES2 = \begin{cases}
2na \int_0^1 w^2 d\alpha + b \int_0^1 w^2 \sum_{i=1}^{n} (\exp(-v_i) + \exp(-u_i)) d\alpha \\
\qquad = \int_0^1 w^2(\alpha) \sum_{i=1}^{n} \left(\frac{1}{f_i(\alpha)} + \frac{1}{g_i(\alpha)} \right) d\alpha \\
a \int_0^1 w^2 \sum_{i=1}^{n} (\exp(-v_i) + \exp(-u_i)) d\alpha + b \int_0^1 w^2 \sum_{i=1}^{n} (\exp(-2v_i) + \exp(-2u_i)) d\alpha \\
\qquad = \int_0^1 w^2 \sum_{i=1}^{n} \left(\frac{1}{f_i(\alpha)} \exp(-v_i) + \frac{1}{g_i(\alpha)} \exp(-2u_i) \right) d\alpha
\end{cases}
\tag{31}
$$

In order to find the parameters of fuzzy non-linear regression defined in the form of S-curve fuzzy model, ES1 and ES2 needs to be solved. For this purpose, criterion is defined by [11], which is denoted by (32) and (33) are utilized.

$$
\begin{aligned}
D_b =\ & 2n \int_0^1 w^2 d\alpha \int_0^1 \left(\frac{1}{f_i} \exp(u_i) + \frac{1}{g_i} \exp(v_i) \right) d\alpha \\
& - \int_0^1 w^2 \sum_{i=1}^{n} (\exp(-v_i) + \exp(u_i)) d\alpha \int_0^1 w^2 \sum_{i=1}^{n} \left(\frac{1}{f_i} + \frac{1}{g_i} \right) d\alpha
\end{aligned}
\tag{32}
$$

$$
\begin{aligned}
D_{b_-} =\ & 2n \int_0^1 w^2 d\alpha \int_0^1 w^2 \sum_{i=1}^{n} \left(\frac{1}{g_i} \exp(u_i) + \frac{1}{f_i} \exp(v_i) \right) d\alpha \\
& - \int_0^1 w^2 \sum_{i=1}^{n} (\exp(-u_i) + \exp(-v_i)) d\alpha \int_0^1 w^2 \sum_{i=1}^{n} \left(\frac{1}{f_i} + \frac{1}{g_i} \right) d\alpha
\end{aligned}
\tag{33}
$$

It is proved by [11] that $D_b \geq D_{b_-}$.

Based on values of D_b, The solution set of (28) or (31) is searched using the computational procedure

If $D_b \geq 0$, expression (28) has unique solution which is denoted by the form of parameter estimates

$$a = \frac{p_1}{D} \text{ and } b = \frac{D_b}{D} \tag{34}$$

where p_1 and D are determinant values which are defined by

$$p_1 = \begin{vmatrix} \int_0^1 w^2 \sum_{i=1}^n \left(\frac{1}{f_i} + \frac{1}{g_i}\right) d\alpha & \int_0^1 w^2 \sum_{i=1}^n (\exp(-v_i) + \exp(-u_i)) d\alpha \\ \int_0^1 w^2 \sum_{i=1}^n \left(\frac{1}{f_i} \exp(u_i) + \frac{1}{g_i} \exp(v_i)\right) d\alpha & \int_0^1 w^2 \sum_{i=1}^n (\exp(-2v_i) + \exp(-2u_i)) d\alpha \end{vmatrix} \tag{35}$$

$$D = \begin{vmatrix} 2n \int_0^1 w^2 d\alpha & \int_0^1 w^2 \sum_{i=1}^n (\exp(-v_i) + \exp(-u_i)) d\alpha \\ \int_0^1 w^2 \sum_{i=1}^n (\exp(-v_i) + \exp(-u_i)) d\alpha & \int_0^1 w^2 \sum_{i=1}^n (\exp(-2v_i) + \exp(-2u_i)) d\alpha \end{vmatrix} \tag{36}$$

If $D_b < 0$, then $D_{b-} \leq D_b \leq 0$. Hence expression (31) has a unique solution which is expressed in the form of parameter estimates

$$a = \frac{p_2}{D} \text{ and } b = \frac{D_{b-}}{D} \tag{37}$$

where p_2 is a determinant value which are defined by

$$p_2 = \begin{vmatrix} \int_0^1 w^2 d\alpha \sum_{i=1}^n \left(\frac{1}{f_i} + \frac{1}{g_i}\right) d\alpha & \int_0^1 w^2 \sum_{i=1}^n (\exp(-v_i) + \exp(-u_i)) d\alpha \\ \int_0^1 w^2 \sum_{i=1}^n \left(\frac{1}{g_i} \exp(u_i) + \frac{1}{f_i} \exp(v_i)\right) d\alpha & \int_0^1 w^2 \sum_{i=1}^n (\exp(-2v_i) + \exp(-2u_i)) d\alpha \end{vmatrix} \tag{38}$$

A small data set is tabulated in Table 5. Our aim is to determine parameters of fuzzy non-linear regression model defined in (21).

$$a = \frac{p_1}{D} = \frac{-6.53}{1.41} = -4.63 \text{ and } b = \frac{D_b}{D} = \frac{56.46}{1.41} = 40.04$$

Table 5 Data set for S-curve fuzzy regression

$\tilde{x}_i = (x_i, \gamma_i)$	$\tilde{y} = (y_i, \delta_i)$
(1.0,0.3)	(3.0,0.2)
(2.0,0.7)	(8.0,0.4)
(3.0,0.4)	(22.0,0.6)
(4.0,1.9)	(25.0,1.3)
(5.0,1.3)	(30.0,1.5)

The model given in (21) is denoted by

$$\tilde{y} = -4.63 + 40.04 \exp(-\tilde{x}) \tag{39}$$

3.2 Quadratic Fuzzy Regression

The second type of fuzzy non-linear regression model is quadratic fuzzy regression expressed in two different models. While the first of which is the one including a quadratic term, the second one contains a term consisting of the interaction of the independent variables. They are proposed by [12] and denoted by (40) and (41).

$$\tilde{Y}_i = A_0 + A_1 X_{i1} + A_2 X_{i1}^2, \quad i = 1, 2, \ldots, n \tag{40}$$

$$\tilde{Y}_i = A_0 + A_1 X_{i1} + A_2 X_{j1} + A_3 X_{i1} X_{j1}, \quad i, j = 1, 2, \ldots, n \tag{41}$$

In both models, it is assumed that input variables are non-negative crisp values and output variable is normal and convex fuzzy numbers with either symmetric or non-symmetric triangular membership functions. The parameter estimation method so-called fuzzy least squares which aim to minimizing the squares of the differences between the observed fuzzy dependent variable and the estimated fuzzy outputs are widely applied to estimation of the parameters. In order to define the difference between the observed and the estimated fuzzy numbers, some methods transforming those fuzzy numbers into crisp numbers are proposed in [10, 12, 13]. One of the methods called Overall Existence Ranking Index (OERI) was proposed in [13]. It is based on the usage of the inverse membership function which is simply a ranking method developed for fuzzy sets. For a given existence level w, the inverse image in terms of membership function, $\mu(x)$, is defined as

$$\mu^{-1}(w) = \{x : \mu(x) = w\} \tag{42}$$

Then for any two arbitrary fuzzy numbers A and B, if A is said to be larger than B at w where $w \in (0, 1], \{\mu_A^{-1}(w)\} > \{\mu_B^{-1}(w)\}$ holds. The inverse is not generally true. The OERI for a fuzzy number $A = (x, \alpha, \beta)$ is a crisp number defined as

$$OM(A) = x - \frac{1}{2}X_1(w)\alpha + \frac{1 - X_1(w)}{2}\beta \tag{43}$$

where $X_1(w)$ is a weighting function determined by decision makers subjectively. The more realistic weighting function is the linear one mentioned in [13]. Then for fuzzy numbers A and B, the distance is defined as

$$D(A, B) = OM(A) - OM(B) \tag{44}$$

When OERI is adapted into regression problem, its distance function can be written as follows:

$$MIN \sum_{i=1}^{n} [Y_i - (\widehat{Y}_i)]^2 = \sum_{i=1}^{n} [D(Y_i, \widehat{Y}_i)]^2 = \sum_{i=1}^{n} [OM(Y_i) - OM(\widehat{Y}_i)]^2 \tag{45}$$

The minimization function and its constraints employing OERI can be written as

$$MIN \sum_{i=1}^{n} \left\{ \left[y_i^m - ax_i - \frac{X_1(w)}{2}(y_i^L - \gamma x_i) + \frac{1 - X_1(w)}{2}(y_i^R - \delta x_i) \right]^2 + (y_i^L - \gamma x_i)^2 + (y_i^R - \delta x_i)^2 \right\} \tag{46}$$

$$y_i^m - (1 - \alpha)y_i^L \geq ax_i - (1 - \alpha)\gamma x_i$$

$$y_i^m + (1 - \alpha)y_i^R \leq ax_i + (1 - \alpha)\delta x_i$$

$$\gamma, \delta \geq 0$$

where fuzzy number $A = (a, \gamma, \delta)$ is the parameter of the regression and fuzzy number $Y = (y^m, y^L, y^R)$ is the observed dependent variable and $0 < \alpha \leq 1$. The formulation given in (46) is the case having one independent variables and its formulation can be easily extended to multiple cases of independent variables easily.

Similarly, Diamond [10] proposed another distance function defined as

$$d^2(A, B) = (x - y)^2 + [(x - y) - (\alpha - \gamma)]^2 + [(x - y) - (\beta - \delta)]^2 \tag{47}$$

The formulations denoted based on the method proposed by Diamond is given as follows:

$$(X^T X)\boldsymbol{\alpha}_L = X^T Y_L \tag{48a}$$

$$(X^T X)\boldsymbol{\alpha}_U = X^T Y_R \tag{48b}$$

Table 6 Data set for quadratic regression models

$\tilde{y}_i$	x_1	x_2
(4,0.5)	1	2.1
(3,0.3)	2	3.3
(2,0.2)	3	4.6
(4,0.7)	4	1.8
(6,0.9)	5	3.7

where Y_L and Y_R are vectors denoting the left end points and the right end points of the response values and α_L and α_R are vectors denoting the left end points and the right end points of the center values of the predicted parameters.

where X is the data matrix denoted by

$$X = \begin{bmatrix} 1 & X_{11} & X_{11}^2 \\ \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n1}^2 \end{bmatrix} \tag{49}$$

Similar construction was also proposed by [12] for the quadratic fuzzy regression containing the interaction term of independent variables.

The generic data matrix for the model given in (41) is given as follows:

$$X = \begin{bmatrix} 1 & X_{11} & X_{12} & X_{11}X_{12} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{m2} & X_{n1}X_{m2} \end{bmatrix} \tag{50}$$

A small data set is used in order to illustrate the models given in (40) and (41) (Table 6).

The parameter estimates for model (40) and (41) are denoted respectively by

$$\tilde{y} = (6.37, 0.86) + (-3.12, 0.48)x_1 + (0.59, 0.10)x_1^2$$
$$\tilde{y} = (10.05, 1.53) + (-1.62, 0.18)x_1 + (-2.90, 0.51)x_2 + (0.76, 0.12)x_1x_2$$

4 Conclusion

Fuzzy correlation measure is an important fuzzy statistics that helps comprehend the relation between two variables that are collected as either linguistically defined values or approximately known quantities. In classical statistical theory, the correlation of these types of variables can no longer be calculated without losing information included. With the help of fuzzy set theory providing mathematical tools allowing to model uncertainty different than one defined by probabilistic approach, the relation between those variables can be quantified using fuzzy

correlation measure. Despite of the fact several methods are available in the literature, two different methods are chosen due to having utilized fuzzy concepts directly in computational procedures and their reliable results. Both of them using basically Zadeh's extension principle with the combination of either fuzzy arithmetic and the weakest t-norm or non-linear programming problem are employed. Both methods with same small data set are run.

Fuzzy non-linear regression is a method fully benefiting from methodological developments used in fuzzy linear regression when it is defined in a form different than linear structure distance. They are called S-curve regression and quadratic fuzzy regression. It is a fact that distance based parameter estimation methods has better ability than mathematical programming ones do when parameter estimation is concern in fuzzy non-linear regression. Two data sets for S-curve fuzzy regression and quadratic fuzzy regression are employed respectively.

References

1. Ngan, S.-L.: Correlation coefficient of linguistic variables and its applications to quantifying relations in imprecise management data. Eng. Appl. Artif. Intell. **26**, 347–356 (2013)
2. Tang, Y., Mu, W., Zhao, L., Zhao, G.: An image segmentation method based on maximizing fuzzy correlation and its fast recursive algorithm. Comput. Electr. Eng. **40**, 833–843 (2014)
3. Huyn, V.-N., Kreinovich, V., Sriboonchitta, S., Suriya, K. (Eds): Econometrics of Risk. Springer (2015)
4. Zimmermann, H.J.: Fuzzy set theory and its applications, 3rd edn. Kluwer-Nijhoff, Boston (1996)
5. Hong, D.H.: Fuzzy measure for a correlation coefficient of fuzzy numbers under T_w (the weakest t-norm) based fuzzy arithmetic operation. Inf. Sci. **176**, 150–160 (2006)
6. Lin, S.-T., Kao, L.: Fuzzy measures for correlation coefficient of fuzzy numbers. Fuzzy Sets Syst. **128**, 267–275 (2002)
7. Tanaka, H., Uejima, S., and Asai, K.: Linear regression analysis with fuzzy model. IEEE Trans. Syst. Man Cyber. **12**, 903–907 (1982)
8. Liu, Y., Zhou, J., Chen, Y.: Using fuzzy non-linear regression to identify degree of compensation among customer requirements in QFD. Neuro Comput. **142**, 115–124 (2014)
9. Sanches, J.A.: Calculating insurance claim reserves with fuzzy regression. Fuzzy Sets Syst. **157**(23), 3091–3108 (2006)
10. Diamond, P.: Fuzzy least squares. Inf. Sci. **46**, 141–157 (1988)
11. Ruoning, X.: S-curve regression model in fuzzy environment. Fuzzy Sets Syst. **97**, 317–326 (1990)
12. Chen, Y.-S.: Fuzzy ranking and quadratic fuzzy regression. Comput. Math. Appl. **38**, 265–279 (1999)
13. Chang, P.T., Lee, E.S.: Ranking of fuzzy sets based on the concept of existence. Comput. Math. Appl. **27**(9/10), 1–21 (1994)