STATISTICAL ANALYSIS OF FUZZY LINEAR REGRESSION MODEL BASED ON DIFFERENT DISTANCES

Jing Zeng* and Aiwu Zhang

School of Mathematics and Statistics ,
Yancheng Teachers University, Yancheng, P.R.China

ABSTRACT

Using fuzzy linear regression model, the least squares estimation for linear regression (LR) fuzzy number is studied by Euclidean distance, Y-K distance and D_k distance respectively. It is concluded that the three different distances have the same coefficient of the least squares estimation. The data simulation shows the correctness of this conclusion.

KEYWORDS:

Fuzzy linear regression ; Euclidean distance ; Y-K distance. ; D_k distance   ; Least-squares estimations
AMS Subject Classification: 62G68.

1. INTRODUCTION

Linear regression analysis is one of the most widely used content in mathematical statistics. The observational data is used to fit the real model in order to solve the relationship between variables in system, explore the changing trendency and forecast of the variables. The most commonly used method in linear regression analysis is least square method. Since Diamond et.al [1] proposed fuzzy least squares method in the 1980s, fuzzy regression analysis has been greatly developed. Savic and Pedrycz [2] put forward fuzzy least squares method which combined fuzzy minimum criteria with the traditional least squares method. Chang and Ayyub [3] discussed the reliability analysis of fuzzy least squares method with the standard of error, correlation coefficient and so on. Fuzzy least squares method is similar to traditional least square method, which both define the distance between two fuzzy numbers firstly and attribute to determine the regression coefficient by making the sum of squared deviations minimum. Since the late 20th century, least square method based on different distance of fuzzy numbers has been proposed, see [4-7]. Many paper only discussed least square method of simple fuzzy linear regression model. But least square method of multiple fuzzy linear regression model is less studied. And almost no paper involves the relationship investigation with fuzzy least squares method based on different distances[16-18].

This paper studies the least squares estimation and the error item of multiple fuzzy linear regression model parameter with clear numbers for inputs, linear regression-type (LR-type) fuzzy numbers for outputs and coefficients on Euclidean distance, Y-K distance, and D_k distance. And it also reveals the relationship of least squares estimations on different distances.

The paper is organized as follows: part 2 and part 3 are preliminaries of definition of fuzzy numbers, properties and distance of fuzzy numbers. Part 4 discusses the least squares estimation and the error item estimation of fuzzy linear regression model on different distances, and gets corresponding conclusions. By the way, a new method is proposed to obtain the parameter estimation of fuzzy linear regression model. Part 5 gives example analysis, and the conclusions are made at last.

2. BASIC CONCEPT

Define 2.1[8] The fuzzy set $\tilde{A}$ in real number field R can be called a fuzzy number, if it satisfies:

(1) there exists $x_0 \in R,$ such that $A(x_0) = 1;$

(2) for any $\alpha \in [0,1], \tilde{A}_\alpha = \{x | A(x) \geq \alpha\} = \left[\underline{A}_\alpha, \overline{A}_\alpha\right]$ is a closed interval.

where $A(x)$ is the membership function of $\tilde{A}$.

Let $\tilde{F}(R)$ denote all fuzzy numbers in real number set R. There is a kind of common fuzzy number among so many fuzzy numbers.

Define 2.2[9] Suppose L and R are decreasing functions from R^+ to [0,1], $L(0) = R(0) = 1,$ for any real number m, it has

$$\tilde{A}(x) = \begin{cases} L\left(\dfrac{m-x}{\alpha}\right), & x \leq m, \alpha > 0, \\ R\left(\dfrac{x-m}{\beta}\right), & x \geq m, \beta > 0, \\ 0 \quad , & otherwise. \end{cases}$$

$\tilde{A}$ is called LR-type fuzzy number, where m is called the center (or mean) of $\tilde{A}$. α, β are called left and right width (or spread) respectively, written as $\tilde{A} = (m, \alpha, \beta)_{LR}$.

When $\alpha = \beta = 0$, LR-type fuzzy number is promised to be ordinary real number, i.e. $(m, 0, 0) = m.$

If $\alpha = \beta,$ $\tilde{A} = (m, \alpha, \alpha)_{LR}$ is called symmetrical LR-type fuzzy number, denoted as $\tilde{A} = (m, \alpha)_{LR}$. $\tilde{F}_{LR}(R)$ denotes the set of all LR-type fuzzy numbers.

Specially, when

$$\tilde{A}(x) = \begin{cases} 1 - \dfrac{m-x}{\alpha}, & m-\alpha \le x \le m, \alpha > 0, \\ 1 - \dfrac{x-m}{\beta}, & m \le x \le m+\beta, \beta > 0, \\ 0, & \text{otherwise.} \end{cases}$$

i . e. $L(x) = R(x) = 1 - x$, $\tilde{A}$ is called triangular fuzzy number.

For LR-type fuzzy numbers, there are the following properties:

Property[10]. Let $\tilde{A} = (a_m, a_l, a_r)_{LR}, \tilde{B} = (b_m, b_l, b_r)_{LR} \in \tilde{F}_{LR}(R)$.(Suppose $\tilde{A}$ and $\tilde{B}$ have the same L and R function), then

(1) $-\tilde{A} = (-a_m, a_l, a_r)_{LR}$;

(2) $\tilde{A} + \tilde{B} = (a_m + b_m, a_l + b_l, a_r + b_r)_{LR}$;

(3) $\tilde{A} - \tilde{B} = (a_m - b_m, a_l + b_l, a_r + b_r)_{LR}$;

(4) $t\tilde{A} = (ta_m, |t|a_l, |t|a_r)_{LR}, t \in R$

(5) $\tilde{A}\tilde{B} = (a_m b_m, |a_m|b_l + |b_m|a_l, |a_m|b_r + |b_m|a_r)_{LR}$.

3. THE DISTANCE BETWEEN LR-TYPE FUZZY NUMBERS

In order to discuss least squares problem of fuzzy linear regression model, the distance between fuzzy numbers must be introduced at first. There are many types of distances between fuzzy numbers, the most commonly used concepts are Euclidean distance, Y-K distance, and D_k distance.

Define 3.1[11] **(Euclidean distance)** Let $\tilde{A} = (a_m, a_l, a_r)_{LR}, \tilde{B} = (b_m, b_l, b_r)_{LR} \in \tilde{F}_{LR}(R)$, then the distance between $\tilde{A}, \tilde{B}$ is defined as

$$D_E^2(\tilde{A}, \tilde{B}) = (a_m - b_m)^2 w_m + (a_l - b_l)^2 w_l + (a_r - b_r)^2 w_r \qquad (3.1)$$

where $w_m > 0, w_l > 0, w_r > 0$ are arbitrary weights.

Specially, if $\tilde{A} = (a, \alpha)_{LR}, \tilde{B} = (b, \beta)_{LR}$ are symmetric fuzzy numbers. It has

$$D_E^2(\tilde{A}, \tilde{B}) = (a-b)^2 + (\alpha - \beta)^2 \qquad (3.2)$$

From (3.1) it is easy to test that $D_E^2(\tilde{A},\tilde{B})$ is nonnegative, symmetric and satisfied triangle inequality. So $(F_{LR}(R), D_E^2)$ is a complete metric space.

Define 3.2[12] (Y-K distance) Let $\tilde{A} = (a_m, a_l, a_r)_{LR}$, $\tilde{B} = (b_m, b_l, b_r)_{LR} \in \tilde{F}_{LR}(R)$, and the Y-K distance between $\tilde{A}, \tilde{B}$ is

$$D_{Yk}^2(\tilde{A},\tilde{B}) = (a_m - b_m)^2 + ((a_m - pa_l) - (b_m - pb_l))^2 + ((a_m + qa_r) - (b_m + qb_r))^2 \quad (3.3)$$

where $p = \int_0^1 L^{-1}(\omega)d\omega, \ q = \int_0^1 R^{-1}(\omega)d\omega.$

$Yang - Ko(1996)$ proved that $(F_{LR}(R), D_{YK})$ is a complete metric space.

Specially, if $\tilde{A} = (a,\alpha)_{LR}, \tilde{B} = (b,\beta)_{LR}$ are symmetric fuzzy numbers, then

$$p = q, D_{Yk}^2(\tilde{A},\tilde{B}) = 3(a-b)^2 + 2p^2(\alpha - \beta)^2 \quad (3.4)$$

When $\tilde{A}$ and $\tilde{B}$ both are triangle fuzzy numbers, namely, $L(x) = R(x) = 1 - x$, it has

$$p = q = \int_0^1 L^{-1}(x)dx = \int_0^1 (1-x)dx = \frac{1}{2}.$$

When $\tilde{A}$ and $\tilde{B}$ both are exponential fuzzy numbers, namely,

$$L(x) = R(x) = \exp(-x^2),$$

it has

$$p = q = \int_0^1 L^{-1}(x)dx = \int_0^1 (-\ln x)^{-\frac{1}{m}}dx = \Gamma\left(1 + \frac{1}{m}\right).$$

Define3.3[13](D_k distance) Let $\tilde{A} = (a_m, a_l, a_r)_{LR}, \tilde{B} = (b_m, b_l, b_r)_{LR} \in \tilde{F}_{LR}(R)$, and the D_k distance between $\tilde{A}, \tilde{B}$ is defined as

$$D_k^2 = (a_m - b_m)^2 + p_2(a_l - b_l)^2 + q_2(a_r - b_r)^2$$
$$- 2p_1(a_m - b_m)(a_l - b_l) + 2q_1(a_m - b_m)(a_r - b_r) \quad (3.5)$$

where

$$P_1 = \frac{1}{2}\int_0^1 |L^{-1}(\alpha)|d\alpha, \ p_2 = \frac{1}{2}\int_0^1 |L^{-1}(\alpha)|^2 d\alpha, \ r_1 = \frac{1}{2}\int_0^1 |R^{-1}(\alpha)|d\alpha, \ r_2 = \frac{1}{2}\int_0^1 |R^{-1}(\alpha)|^2 d\alpha.$$

Diamond et. al (1998) proved $(F_{LR}(R), D_K)$ is a complete metric space.

Specially, if $\tilde{A} = (a, \alpha)_{LR}$, $\tilde{B} = (b, \beta)_{LR}$ are both symmetric fuzzy numbers, then

$$D_k^2(\tilde{A}, \tilde{B}) = (a-b)^2 + (p_2 + q_2)(\alpha - \beta)^2 + 2(q_1 - p_1)(a-b)(\alpha - \beta) \qquad (3.6)$$

When $\tilde{A}$ and $\tilde{B}$ both are triangle fuzzy numbers, namely, $L(x) = R(x) = 1 - x$, it has

$$p_1 = q_1 = \frac{1}{2}\int_0^1 L^{-1}(x)dx = \frac{1}{2}\int_0^1 (1-x)dx = \frac{1}{4}, \ p_2 = q_2 = \frac{1}{2}\int_0^1 (1-x)^2 dx = \frac{1}{6}.$$

When $\tilde{A}$ and $\tilde{B}$ both are exponential fuzzy numbers, namely, $L(x) = R(x) = \exp(-x^2)$, it has

$$p_1 = q_1 = \frac{\sqrt{\pi}}{4}, \ p_2 = q_2 = \frac{1}{2}.$$

4. FUZZY LEAST SQUARES REGRESSION AND COMPARISON BASED ON DIFFERENT DISTANCES

4.1 Fuzzy least squares regression

Consider the following fuzzy linear regression model

$$\tilde{y}_i = \tilde{A}_0 + \tilde{A}_1 x_{i1} + \tilde{A}_2 x_{i2} + \cdots + \tilde{A}_p x_{ip}, \quad i = 1, 2, \cdots, n \qquad (4.1)$$

Where x_{ij} are clear numbers, $\tilde{y}_i (i = 1, 2, \cdots, n)$ and $\tilde{A}_m (m = 1, 2, \cdots, p)$ are LR-type fuzzy numbers. Without loss of generality, let $x_{ij} > 0$ (if $x_{ij} < 0$, it can be transformed to the situation $x_{ij} > 0$). And suppose $\tilde{y}_i (i = 1, 2, \cdots, n), \tilde{A}_m (m = 1, 2, \cdots, p)$ are both symmetric LR-type fuzzy numbers for simplicity.

4.2 FUZZY LEAST SQUARES REGRESSION BASED ON EUCLIDEAN DISTANCE

4.2.1 Least squares estimation

Apply traditional least squares method to estimate the regression coefficient $\tilde{A}_m (m = 0, 1, \cdots, p)$ of (4.1). Suppose $\tilde{y}_i = (c_i, s_i)_{LR}, \tilde{A}_m = (a_m, r_m)_{LR}$ have the same membership function, then (4.1) can be expressed as

$$(c_i, s_i) = (a_0, r_0) + (a_1, r_1)x_{i1} + \cdots + (a_p, r_p)x_{ip}$$

According to Euclidean distance (3.1), square sum of error can be wrritten as

$$D_E^2 = \sum_{i=1}^{n}\left[\left(c_i - \left(a_0 + a_1 x_{i1} + \cdots + a_p x_{ip}\right)\right)^2 + \left(s_i - \left(r_0 + r_1 x_{i1} + \cdots + r_p x_{ip}\right)\right)^2\right]$$

Denote $a = \left(a_0, a_1, \cdots, a_p\right)^T$, $r = \left(r_0, r_1, \cdots, r_p\right)^T$, $c = \left(c_1, c_2, \cdots, c_n\right)^T$, $s = \left(s_1, s_2, \cdots, s_n\right)^T$,

$$X = \begin{pmatrix} 1 & x_{11} \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ 1 & x_{n1} \cdots & x_{np} \end{pmatrix} \tag{4.2}$$

Then $D_E^2 = \left(Xa - c\right)^T \left(Xa - c\right) + \left(Xr - s\right)^T \left(Xr - s\right)$, and

$$\begin{cases} \dfrac{\partial D_E^2}{\partial a} = 2X^T Xa - 2X^T c = 0 \\ \dfrac{\partial D_E^2}{\partial r} = 2X^T Xr - 2X^T s = 0 \end{cases},$$

It has

$$\begin{cases} \hat{a} = \left(X^T X\right)^{-1} X^T c \\ \hat{r} = \left(X^T X\right)^{-1} X^T s \end{cases} \tag{4.3}$$

4.2.2 The error item of estimation

The simplest method to evaluate performance of fuzzy regression model is to use residual or residual sum of square (4.4) as measuring index.

Denote $e_i = \tilde{y}_i - \hat{\tilde{y}}_i$ or $e_i^2 = \left|\tilde{y}_i - \hat{\tilde{y}}_i\right|^2$ and $SST = \sum_{i=1}^{n} e_i^2$, then on the Euclidean distance

$$SST_E = \sum_{i=1}^{n} D_E^2\left(\tilde{y}_i, \hat{\tilde{y}}_i\right) = \left(X\hat{a} - c\right)^T \left(Xa - c\right) + \left(X\hat{r} - s\right)^T \left(Xr - s\right)$$

Substitution (4.3) to the above formula, it gives

$$SST_E = c^T \left(I - P_X\right) c + s^T \left(I - P_X\right) s. \tag{4.4}$$

where I is a unit matrix of n -order and $P_X = X\left(X^T X\right)^{-1} X^T$. $c^T \left(I - P_X\right) c$ is often called as central residual sum of square while $s^T \left(I - P_X\right) s$ is often called as siding residual sum of square. Because residual is the difference between $\tilde{A}_m \left(m = 0, 1, \cdots, p\right)$ and its estimation $\hat{\tilde{A}}_m$,

6

i.e., the difference between $\tilde{y}_i = \tilde{A}_0 + \tilde{A}_1 x_{i1} + \tilde{A}_2 x_{i2} + \cdots + \tilde{A}_p x_{ip}$ and

$\hat{\tilde{y}}_i = \hat{\tilde{A}}_0 + \hat{\tilde{A}}_1 x_{i1} + \hat{\tilde{A}}_2 x_{i2} + \cdots + \hat{\tilde{A}}_p x_{ip}$ Residual reflects the goodness of fit, the less $|e_i|$ is, the

better fitting of the i th point. While residual sum of square $\sum_{i=1}^{n}$ reflects the overall

$$SST = \sum_{i=1}^{n} e_i^2$$

condition of fitting, the less RRS is ,the better fitting of the overall condition.

4.3 Fuzzy least squares regression based on Y-K distance

4.3.1 Least squares estimation

Suppose $\tilde{y}_i = (c_i, s_i)_{LR}$, $\tilde{A}_m = (a_m, r_m)_{LR}$ have the same membership function, according to the formula of D_{YK} (3.4), residual sum of square can be expressed as

$$D_{Yk}^2 = 3\sum_{i=1}^{n}\left[c_i - \left(a_0 + a_1 x_{i1} + \cdots + a_p x_{ip}\right)\right]^2 + 2p^2\sum_{i=1}^{n}\left[s_i - \left(r_0 + r_1 x_{i1} + \cdots + r_p x_{ip}\right)\right]^2$$

$$= 3(Xa - c)^T (Xa - c) + 2p^2 (Xr - s)^T (Xr - s),$$

where X, a, c, r, s are the same as (4.2). From

$$\begin{cases} \dfrac{\partial D_{Yk}^2}{\partial a} = 6X^T Xa - 6X^T c = 0, \\ \dfrac{\partial D_{Yk}^2}{\partial r} = 4p^2 X^T Xr - 4p^2 X^T s = 0. \end{cases}$$

It has

$$\begin{cases} \hat{a} = \left(X^T X\right)^{-1} X^T c \\ \hat{r} = \left(X^T X\right)^{-1} X^T s \end{cases} \tag{4.5}$$

4.3.2 The error item of estimation

In accordance with the method used in 4.2.2, residual sum of square of model (4.1) on Y-K distance can be got

$$SST_{YK} = 3c^T (I - P_X) c + 2p^2 s^T (I - P_X) s. \tag{4.6}$$

4.4 Fuzzy least squares regression based on D_K distance

4.4.1 Least squares estimation

Suppose $\tilde{y}_i = (c_i, s_i)_{LR}$, $\tilde{A}_m = (a_m, r_m)_{LR}$ have the same membership function, according to the formula of D_K (3.6), residual sum of square can be expressed as

$$
\begin{aligned}
D_K^2 = & \sum_{i=1}^{n} \Big[c_i - \big(a_0 + a_1 x_{i1} + \cdots + a_p x_{ip}\big) \Big]^2 + (p_2 + q_2) \sum_{i=1}^{n} \Big[s_i - \big(r_0 + r_1 x_{i1} + \cdots + r_p x_{ip}\big) \Big]^2 \\
& + 2(q_1 - p_1) \sum_{i=1}^{n} \Big[c_i - \big(a_0 + a_1 x_{i1} + \cdots + a_p x_{ip}\big) \Big] \Big[s_i - \big(r_0 + r_1 x_{i1} + \cdots + r_p x_{ip}\big) \Big] \\
= & (Xa - c)^T (Xa - c) + (p_2 + q_2)(Xr - s)^T (Xr - s) \\
& + 2(q_1 - p_1)(Xa - c)^T (Xr - s),
\end{aligned}
$$

where X, a, c, r, s are the same as (4.2). From

$$
\begin{cases}
\dfrac{\partial D_K^2}{\partial a} = 2 X^T X a - 2 X^T c + 2(q_1 - p_1) X^T (Xr - s) = 0, \\[2mm]
\dfrac{\partial D_K^2}{\partial r} = (p_2 + q_2)\big(2 X^T X r - 2 X^T s\big) + 2(q_1 - p_1) X^T (Xa - c) = 0.
\end{cases}
$$

It has

$$
\begin{cases}
\hat{a} = (X^T X)^{-1} X^T c \\[2mm]
\hat{r} = (X^T X)^{-1} X^T s
\end{cases}
\tag{4.7}
$$

4.4.2 The error item of estimation

Similarly, we can get residual sum of square of model (4.1) on D_K distance

$$
SST_{D_K} = c^T (I - P_X) c + (p_2 + q_2) s^T (I - P_X) s + 2(q_1 - p_1) c^T (I - P_X) s. \tag{4.8}
$$

From (4.3), (4.5) and (4.7) we can see that for fuzzy linear regression model (4.1) whether we use Euclidean distance, Y-K distance or D_k distance, least squares estimations of regression coefficient $\tilde{A}_m$ are the same in the case when response variable $\tilde{y}_i$ is symmetrical LR-type fuzzy number. Anymore, when $\tilde{y}_i$ reduces to clear number (in this case, model (4.1) is traditional linear regression model), namely, $s_i = 0$, it has

$$\hat{a} = \left(X^T X \right)^{-1} X^T c, \qquad \hat{r} = 0.$$

It is in accordance with least squares estimation of traditional linear regression model.

At the same time, from (4.4), (4.6) and (4.8) we can see residual sum of square on Euclidean distance is always less than that on Y-K distance for model (4.1), which illustrates that the fitting effect of parameter estimation on Euclidean distance is better. Residual sum of square on Euclidean distance is always equal to that on D_k distance for model (4.1) when fuzzy number is triangle fuzzy number or exponential fuzzy number (in this case, $p_1 = q_1, p_2 = q_2 = \dfrac{1}{2}$), which illustrates that the fitting effect of parameter estimation on Euclidean distance and D_k distance is the same. Anymore, residual sum of square on Euclidean distance and Y-K distance consists of two parts, one is central residual sum of square and the other is siding residual sum of square. While there is cross term of center and siding for residual sum of square on D_k distance except for the above two parts. Therefore in solving the least squares estimator that priority should be given to the Euclidean distance.

4.5 Further discussion of least squares regression estimations on three distances

After comparing the conclusion of (4.3) or (4.5) or (4.7) with the least squares estimation of traditional linear regression model, a new method to get least squares estimation of fuzzy linear regression (4.1) is obtained, i.e., converting solving least squares estimation of fuzzy linear regression (4.1) into solving traditional least squares estimation two times.

Considering central observation data set $\left\{ \left(y_i, x_{i1}, x_{i2}, \cdots, x_{ip} \right) \middle| i = 1, 2, \cdots, n \right\}$ and siding data set $\left\{ \left(s_i, x_{i1}, x_{i2}, \cdots, x_{ip} \right) \middle| i = 1, 2, \cdots, n \right\}$, we can apply traditional linear regression model

$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$ to get parameter estimations of the following models:

$$y_i = a_0 + a_1 x_{i1} + \cdots + a_p x_{ip} \qquad i = 1, 2, \cdots, n,$$
$$s_i = r_0 + r_1 x_{i1} + \cdots + r_p x_{ip} \qquad i = 1, 2, \cdots, n.$$

And with the estimator $\hat{a}_m, \hat{r}_m, m = 1, 2, \cdots, p$, the estimator $\hat{\tilde{A}}_m = \left(\hat{a}_m, \hat{r}_m \right), m = 1, 2, \cdots, p.$ for $\tilde{A}_m$ is obtained.

5. EXAMPLE ANALYSIS

The following $Tanaka^{[14]}$ data (1987) (see Table 1) is used to illustrate the above conclusions. The data set which has ten pairs of data, concludes three clear explanatory variables and one fuzzy response variable. Consider only exponential fuzzy number. Now use least squares method to estimate parameters of fuzzy linear regression model on Euclidean distance, Y-K distance and

D_k distance. The corresponding error terms are also calculated. The estimation results are shown in Table 2-4.

NO.	Explanatory variable			Response variable
	x_1	x_2	x_3	$\tilde{y}_i = (c_i, r_i)_{LR}$
1	3	5	9	(96,42)
2	14	8	3	(120,47)
3	7	1	4	(52,33)
4	11	7	3	(106,45)
5	7	12	15	(189,79)
6	8	15	10	(194,65)
7	3	9	6	(107,42)
8	12	15	11	(216,78)
9	10	5	8	(108,52)
10	9	7	4	(103,44)

Table 1 Tanaka (1987) data

Table 2 Fuzzy least squares estimations and errors on Euclidean distance

NO.	(c_i, s_i)	$(\hat{c}_i, \hat{s}_i)$	e_i^2
1	(96, 42)	(93.20, 44.62)	14.69
2	(120, 47)	(122.48, 49.13)	10.67
3	(52, 33)	(49.36, 32.11)	7.75
4	(106, 45)	(104.82, 43.01)	5.35
5	(189, 79)	(191.79, 76.71)	13.06
6	(194, 65)	(193.64, 67.67)	7.25
7	(107, 42)	(109.77, 40.85)	9.08
8	(216, 78)	(211.65, 77.08)	19.73
9	(108, 52)	(110.89, 53.24)	9.91
10	(103, 44)	(103.36, 42.58)	2.14

$\hat{a} = (-1.39, 3.25, 7.92, 5.03)^T, \hat{r} = (8.01, 1.64, 1.20, 2.85)^T, SST_E = 99.63.$

Table 3 Fuzzy least squares estimations and errors on Y-K distance

NO.	(c_i, s_i)	$(\hat{c}_i, \hat{s}_i)$	e_i^2
1	(96, 42)	(93.20, 44.62)	34.24
2	(120, 47)	(122.48, 49.13)	25.53
3	(52, 33)	(49.36, 32.11)	22.10
4	(106, 45)	(104.82, 43.01)	10.38
5	(189, 79)	(191.79, 76.71)	31.66
6	(194, 65)	(193.64, 67.67)	11.57
7	(107, 42)	(109.77, 40.85)	25.35
8	(216, 78)	(211.65, 77.08)	57.99
9	(108, 52)	(110.89, 53.24)	27.53
10	(103, 44)	(103.36, 42.58)	3.55

$$\hat{a} = (-1.39, 3.25, 7.92, 5.03)^T, \hat{r} = (8.01, 1.64, 1.20, 2.85)^T, SST_{YK} = 249.91.$$

Table 4 Fuzzy least squares estimations and errors on D_K distance

NO.	(c_i, s_i)	$(\hat{c}_i, \hat{s}_i)$	e_i^2
1	(96, 42)	(93.20, 44.62)	14.69
2	(120, 47)	(122.48, 49.13)	10.67
3	(52, 33)	(49.36, 32.11)	7.75
4	(106, 45)	(104.82, 43.01)	5.35
5	(189, 79)	(191.79, 76.71)	13.06
6	(194, 65)	(193.64, 67.67)	7.25
7	(107, 42)	(109.77, 40.85)	9.08
8	(216, 78)	(211.65, 77.08)	19.73
9	(108, 52)	(110.89, 53.24)	9.91
10	(103, 44)	(103.36, 42.58)	2.14

$$\hat{a} = (-1.39, 3.25, 7.92, 5.03)^T, \hat{r} = (8.01, 1.64, 1.20, 2.85)^T, SST_{D_K} = 99.63.$$

Table 2, 3 and 4 show that least squares estimations of fuzzy linear regression model (4.1) on Euclidean distance, Y-K distance and D_k distance are the same. They also illustrate that the Y-K distance estimation error term is greater than the Euclidean distance estimation error term and the Euclidean distance estimation error term is equal to D_k distance estimation error because of the exponential fuzzy number.

6. CONCLUSION

The paper studies the least squares estimation and the error terms of fuzzy linear regression model with clear numbers for inputs, LR-type fuzzy numbers for outputs and regression coefficients on Euclidean distance, Y-K distance, and D_k distance. The results show that the least squares estimations are the same on the above three distances and the priority should be given to the Euclidean distance in solving least squares estimator. When the outputs and regression coefficients are clear numbers, the estimation will be traditional least squares estimation. At the same time, a new method to get least squares estimation for fuzzy linear regression model is proposed. At last, the experimental results show the correctness of the conclusion.

ACKNOWLEDGEMENT

REFERENCE

[1] Diamond,P., Fuzzy least squares, Information Science, 46(1988), 141—157.

[2] D.Savic, W.Pedrycz, Evaluation of fuzzy regression models, Fuzzy Sets and Systems,39(1991), 51—63.

[3] Y-H.O. Chang ,B.M.Ayyub, Reliability analysis in fuzzy regression,Proc.Annual of the North American Fuzzy Information Processing Society, Allentown,PA,USA,1993, 93—97.

[4] Ma,M,Menahem,F et., General Fuzzy least squares, Fuzzy Sets and Systems, 88(1997), 107—118.

[5] Renato,C,Pierpaolo,D.U., Least squares estimation of a linear regression model with LR fuzzy response,Computational Statistics & Data Analysis,51(2006), 267—286.

[6] Pierpaola,D.U.,Tommaso,G.., A least-squares approach to fuzzy linear regression analysis, Computational Statistics & Data Analysis,34(2000), 427—440.

[7] Chang,P.T.,Lee,E.S., A generalized fuzzy weighted least-squares regression, Fuzzy Sets and Systems,82(1996), 289—298.

[8] Zadeh, L.A., The concept of linguistic variable and its application to approximate reasoning, Inform Sci., 8(1975) 99—244,301—357.

[9] Zimmermann,H.J.,Fuzzy Set Theory and Its Applications,Kluwer Academic Press Dordrecht,1996.

[10] D.Dubois,H.Prade, Fuzzy Sets and Systems:Theory and Applications, Academic Publishers, New York,1980.

[11] Arabpour, A.R., Tata,M., Estimating the Parameters of a fuzzy linear regression model, Iranian Journal of Fuzzy Systems,5(2)(2008), 1—19.

[12] Yang,M.S.,Ko,C.H., On a Class of c-numbers Clustering Procedures for fuzzy Data, Fuzzy Sets and Systems,84(1996), 49—60.

[13] Diamond,P., Korner.R.,Extended fuzzy linear r model and least squares estimation, Computers and Mathmatics with Applications, 33(9)(1998), 15—32.

[14] Tanak,H., Fuzzy Data Analysis by Possibilistic Linear Model, Fuzzy Sets and Systems,24(1987), 363—375.

[15] Zhang Aiwu, Contrast Between Fuzzy Linear Least-squares Regression Based on Different Distances, Journal of Jiangsu University of Science and Technology(Natural Science Edition), 26 (5) (2012):509-513

[16] Pierpaolo D'Urso • Riccardo Massari, Weighted Least Squares and Least Median Squares estimation for the fuzzy linear regression analysis, METRON,71 (2013):279–306.

[17] Jin-Guan Lin. Qing-Yun Zhuang. Chao Huang, Fuzzy Statistical Analysis of Multiple Regression with Crisp and Fuzzy Covariates and Applications in Analyzing Economic Data of China, Comput Econ , 39 (2012):29–49.

[18] S. Muzzioli • B. De Baets, A comparative assessment of different fuzzy regression methods for volatility forecasting, Fuzzy Optim Decis Making , 12 (2013):433–450.