Intern. J. Fuzzy Mathematical Archive
Vol. 13, No. 1, 2017, 69-76
ISSN: 2320 –3242 (P), 2320 –3250 (online)
Published on 22 September 2017
www.researchmathsci.org
DOI: http://dx.doi.org/10.22457/ijfma.v13n1a7

International Journal of
Fuzzy Mathematical
Archive

Error Estimation Using Fuzzy Linear Regression Analysis

***S.Ismail Mohideen*[1], *A.Nagoor Gani*[2] and *U.Abuthahir*[3]**

[1]PG & Research Department of Mathematics, Jamal Mohamed College (Autonomous)
Tiruchirappalli-620020, Tamil Nadu, India. E-mail: simohideen@yahoo.co.in
[2]PG & Research Department of Mathematics, Jamal Mohamed College (Autonomous)
Tiruchirappalli-620020, Tamil Nadu, India. E-mail: ganijmc@yahoo.co.in
[3]PG & Research Department of Mathematics, Jamal Mohamed College (Autonomous)
Tiruchirappalli-620020, Tamil Nadu, India. E-mail: abugous2004@yahoo.co.in

Received 7 September 2017; accepted 20 September 2017

Abstract. In this paper an attempt is made to reduce the error of non-fuzzy data using fuzzy linear regression. A numerical example has been analyzed. It is concluded from the example that the normal regression method is better compared to fuzzy linear regression method.

Keywords: Linear regression, symmetric triangular fuzzy number and fuzzy regression.

AMS Mathematics Subject Classification (2010): 03E72, 62J05, 62J12

1. Introduction

Regression analysis, including statistical regression analysis and fuzzy regression analysis, aims to determine the best-fit model for describing the functional relationship between dependent variables and independent variables by exploiting the knowledge from the given input-output data pairs. Some discrepancy between the observed values (from the data sets) and the estimated values (from a regression model) can occur due to measurement errors and/or modeling errors.

Regression analysis is one of the areas in which fuzzy set theory is used frequently, since Tanaka [4] initiated research on fuzzy linear regression (FLR) analysis. This area is widely developed and wide varieties of methods are proposed. One approach to deal with FLR is Linear Programming (LP). This approach was first introduced by *Tanaka* and developed by others, and next approach is least squares method, which was first introduced by Celmins [1] and developed by others [2].

The fuzzy set theory introduced by Zadeh [5] has derived meaningful applications in many field of studies. The idea of fuzzy set is welcomed because it handles uncertainty and vagueness. In fuzzy set theory, the membership of an element of a fuzzy set is a single value between zero and one.

This chapter presents linear regression analysis in a fuzzy environment using fuzzy linear models with symmetric triangular fuzzy number (STFN) coefficients. The aim of this fuzzy regression (FR) analysis is to find the coefficients of a proposed model for all given input–output data sets. Here, the basic idea is to minimize the fuzziness of the model by minimizing the total support of the fuzzy coefficients, subject to all data.

The rest of this chapter is organized as follows: In Section 2, the basic concept and definitions are presented. In Section 3, fuzzy linear model is given. In Section 4, fuzzy regression analysis is given and finally numerical example is given In Section 5.

2. Preliminaries

Definition 2.1. A fuzzy set (FS)$\tilde{A}$ is defined by $\tilde{A}$ = { x, $\mu_{\tilde{A}}(x)$: x∈ A , $\mu_{\tilde{A}}(x) \in [0,1]$ }. In the pair (x, $\mu_{\tilde{A}}(x)$), the first element x belong to the classical set A and the second element belong to the interval [0, 1] is called membership function.

Definition 2.2. A fuzzy set $\tilde{A}$ is convex if

$$\mu_{\tilde{A}}(\lambda\, x_1 + (1-\lambda)x_2\ ) \geq \text{Min}\ (\mu_{\tilde{A}}(x_1),\ \mu_{\tilde{A}}(x_2)), \forall x_1,\ x_2 \in \text{R and } \lambda \in [0,1].$$

Definition 2.3. A triangular fuzzy number (TFN) with left spread and right spread is denoted by$\tilde{A}$ and corresponding membership function is given,

$$\mu_{\tilde{A}}(\text{x}) = \begin{cases} \frac{x-(a-\alpha)}{\alpha} & for\ x\ \in [a-\alpha, a] \\ \frac{\alpha+\beta-x}{\beta} & for\ x \in [a,\ \ a+\beta] \\ 0 & otherwise \end{cases}$$

where a∈R; α, β>. The symbolic representation of TFN is $\tilde{A}_{TFN}$ = [a:α, β]. Here α and β are called left and right spreads of membership function $\mu_{\tilde{A}}$(x) respectively The diagrammatic representation of a TFN is as following

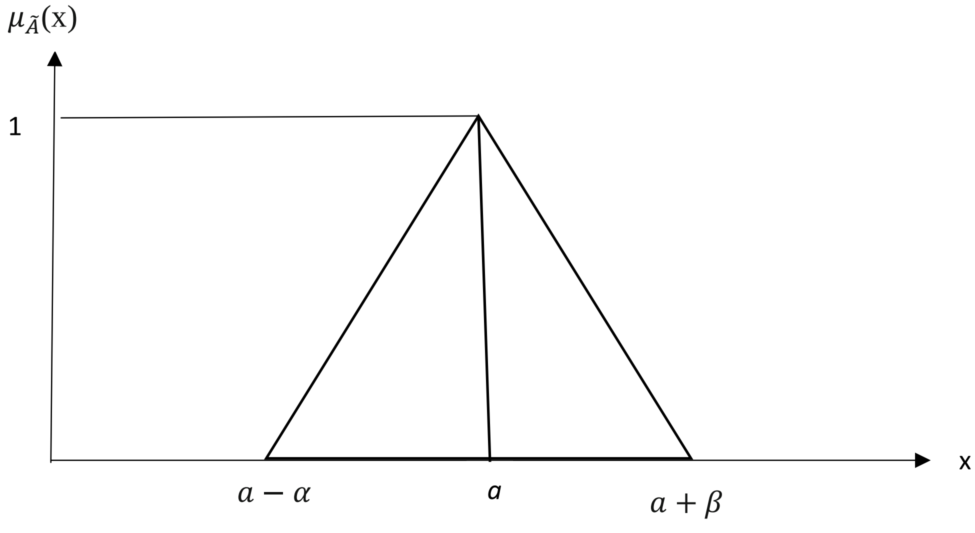

Figure 1: Triangular fuzzy number

Definition 2.4. A Linear Programming Problem (LPP) is defined as:

Maximize z= $c\ x$

Subject to $Ax = b$, $x \geq 0$

where c = ($c_1, c_{2,} \dots, c_n$), b = $(b_1 b_2, \dots ,\ b_m)^T$ and A= $[a_{ij}]_{m\times n}$where all the parameters are crisp.

2.5. Graded mean integration representation method-defuzzification

If $\tilde{A}$ = $(a_1$, a_2, $a_3)$ is a triangular fuzzy number then the graded mean integration representation of $\tilde{A}$ is given by $P(\tilde{A}) = [a_1 + 4a_2 + a_3]/6$

3. Fuzzy linear regression model (FLR)

Fuzzy functions and fuzzy linear models [3] are presented, here for continuation.
The general form of FR model is given by

$$\tilde{y} = \text{f}(\text{x}, \tilde{A}) = \tilde{A}_0 + \tilde{A}_1 x_1 + \tilde{A}_2 x_2 + \dots + \tilde{A}_n x_n \qquad (1)$$

where $\tilde{y}$ is the fuzzy output, $\tilde{A}_i$, *i= 1,2,…,n* is an fuzzy coefficient and X = $(x_1, x_{2,} \dots, x_n)$ is an *n* dimensional non-fuzzy input vector. Each Triangular Fuzzy Number (TFN) coefficient $\tilde{A}_i$ can be defined by $\tilde{A}_{TFN}$= [a; α_i, β_i] whereα_i, $\beta_i$are called the left and right spreads of membership function $\mu_{\tilde{A}}$(x) respectively. When two spreads are equal, the TFN is known as symmetric triangular fuzzy number (STFN). Hence a TFN $\tilde{A}_{TFN}$ = [a;α, β] is said to be STFN if α = β (say *m*), this concept gives the definition of STFN as follows:

A fuzzy set$\tilde{A}$ in R is said to be a STFN if there exist real number *a* and *m* where *m>0* such that the membership functions are derived from Figure 2.

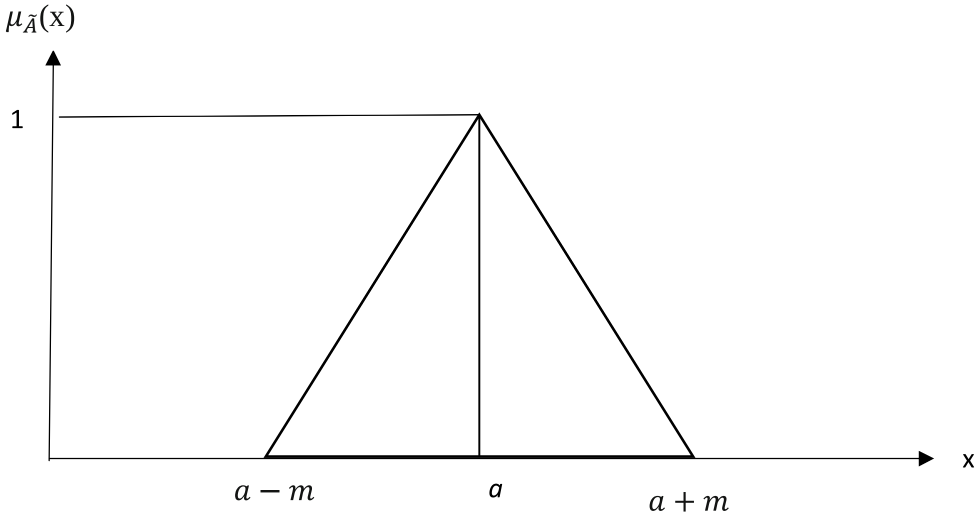

Figure 2: Symmetric triangular fuzzy number

$$\mu_{\tilde{A}}(x) = \begin{cases} \frac{x-(a-m)}{m} & for\ x\ \in [a-m, a] \\ \frac{a+m-x}{m} & for\ x \in [a,\ \ a+m] \\ 0 & otherwise \end{cases}$$

STFN can be represented as $\tilde{A}_{STFN}$ = [a; m,m], where a is the center, m is the spread of membership function $\mu_{\tilde{A}}$(x).

The fuzzy components are assumed to be STFNs. The fuzzy output from the linear model f(x, $\tilde{A}$) in (1) can be expressed as $\tilde{y}$=f(x, $\tilde{A}$) = ($f^c(x)$, $f^l(x)$)where $f^c(x)$ is the center of the of linear model f(x, $\tilde{A}$) and has the form$f^c(x)$ = $a_0 + a_1x_1 + \cdots + a_nx_n$ and $f^l(x)$is the spreads of membership functions of f(x, $\tilde{A}$).

$f^l(x) = m_0 + m_1 \mid x_1 \mid + \cdots + m_n \mid x_n \mid$

Then the membership of $\tilde{y}$ defined in (1) can be represented as,

$\mu_{\tilde{A}}$(y)=

$$\begin{cases} \frac{y-[(a_0+\sum_i a_ix_i)-(m_0+\sum_i m_i|x_i|)]}{(m_0+\sum_i m_i|x_i|)} for\ y \in [(a_0 + \sum_i a_ix_i) - (m_0 + \sum_i m_i|x_i|), (a_0 + \sum_i a_ix_i)] \\ \frac{[(a_0+\sum_i a_ix_i)+(m_0+\sum_i m_i|x_i|)]-y}{(m_0+\sum_i m_i|x_i|)} for\ y \in [(a_0 + \sum_i a_ix_i), (a_0 + \sum_i a_ix_i) + (m_0 + \sum_i m_i|x_i|)] \\ 0 \qquad otherwise \end{cases}$$

The diagramatic representation of fuzzy output function of an FN $\tilde{A}$ with h-level set is presented in the following figure.

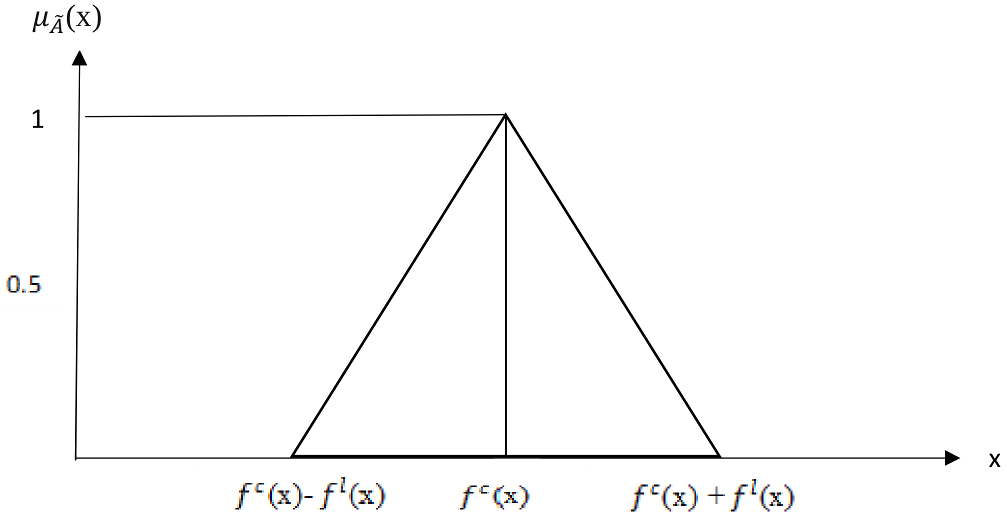

Figure 3: Fuzzy output function

where f^c(x) = $(a_0 + \sum_i a_ix_i)$

f^c(x)-f^l(x) $=(a_0 + \sum_i a_ix_i) - (m_0 + \sum_i m_i|x_i|)$

f^c(x)+f^l(x) $=(a_0 + \sum_i a_ix_i) + (m_0 + \sum_i m_i|x_i|)$

4. Analysis of fuzzy linear regression

The objective of the Fuzzy Linear Regression (FLR) method with non-fuzzy data is to determine the parameters $\tilde{A}$such that the fuzzy output set $\{y_j\}$ is associated with $\mu_{\tilde{A}}(y_j) \geq h$ where h∈[0,1] and 'h' is chosen for the purpose of generating the best-fitting model.

Now, the problem is to minimize the fuzziness of the output. Since, the values of the membership function of the fuzzy output are the function of the spread of the membership function, minimizing the spread of the membership function and hence it leads to the minimization of the fuzziness of the output.

Minimize z = Min $\{f^l(x)\}$ where $f^l(x)$is defined in (2)

Minimize z = Min $\{f^l(x)\}$ = Min $\{(m_0 + \sum_{i=1}^{n} m_i \sum_{j=1}^{m} |x_j|\}$ (3)

Subject to the set of constraints y_j ∈ [f(x_j)]$_{h,}$

where [f(x_j)]$_h$ = [$\tilde{A}_0$]$_h$ + [$\tilde{A}_1$]$_h x_j$+ [$\tilde{A}_2$]$_h x_j$+ . . . + [$\tilde{A}_n$]$_h x_j$ such that [..]$_h$ denotes the h-level set of an Fuzzy Number. By using the fuzzy membership function for the output, the two constraints of the FLR model are given by

$$\frac{y-[(a_0+\sum_i a_i x_i)-(m_0+\sum_i m_i|x_i|)]}{(m_0+\sum_i m_i|x_i|)} \geq h \quad (4)$$

and

$$\frac{[(a_0+\sum_i a_i x_i)+(m_0+\sum_i m_i|x_i|)]-y}{(m_0+\sum_i m_i|x_i|)} \geq h \quad (5)$$

Simplifying (4) and (5), we have

$$y - [(a_0 + \sum_i a_i x_i) - (m_0 + \sum_i m_i|x_i|)] \geq h(m_0 + \sum_i m_i|x_i|) \quad (6)$$

and

$$[(a_0 + \sum_i a_i x_i) + (m_0 + \sum_i m_i|x_i|)] - y \geq h(m_0 + \sum_i m_i|x_i|) \quad (7)$$

Again simplifying, we get

$$(a_0 + \sum_i a_i x_i) - (1-h)(m_0 + \sum_i m_i|x_i|) \leq y \quad (8)$$

and $$[(a_0 + \sum_i a_i x_i) + (1-h)(m_0 + \sum_i m_i|x_i|)] \geq y \quad (9)$$

Therefore, the problem is reduced to the following form:

Minimize z = Min $\{f^l(x)\}$

Min $\{f^l(x)\}$ = Min $\{m_0 + \sum_{i=1}^{n} m_i \sum_{j=1}^{m} |x_j|\}$ and

Subject to the constraints (8),(9) and m_0 ≥0 and a_i, m_i ≥0 *for i=1,2,…,n*

By using the software TORA, the values of the center and spread of membership can be estimated. This gives the FN coefficients as follows:

$\tilde{A}_0$ = $[a_0;m_0,m_0]$

$\tilde{A}_1$ = $[a_1;m_1,m_1]$

⋮

$\tilde{A}_n$ = $[a_n;m_n,m_n]$

By using the values of $\tilde{A}_0$, $\tilde{A}_1$…… $\tilde{A}_n$ the FLR model becomes

$\tilde{y}$ = $\tilde{A}_0$ + $\tilde{A}_1 x_1$ + $\tilde{A}_2 x_2$+ … + $\tilde{A}_n x_n$

The value of $\tilde{y}$ can be estimated by substituting the values of x_1, x_2 …. x_n, the estimated values of $\tilde{y}$are actually STFNs which can be defuzzified to crisp number by using the function principle

A = $[a_1 + 4a_2 + a_3]/6$(10)

5. Numerical illustration

Numerical problem is considered with the value of h = 0.7

Example 5.1. Consider the values for *X* and *Y* in the following table to calculate the coefficients of the FLR model for *i=1*

j	1	2	3	4	5	6
x_j	1	2	4	6	7	8
y_j	3010	4500	4400	5400	7295	8195

Let us consider the value of $h = 0.7$
Minimize z = Min $\{f^{l}(x)\}$
Min $\{f^{l}(x)\}$ = Min $\{(m_0 + \ m_1 \sum_{j=1}^{6} |x_j|)\}$
Subject to the constraints
$a_0 + x_j a_1 + 0.3m_0 + 0.3m_1 x_j \geq y$[from (9)]
$a_0 + x_j a_1 - 0.3m_0 - 0.3m_1 x_j \leq y$ [from (8)]

where x_j represents the component of x in the j^th^ entry value. The number of functional constraints depend on the number of data sets available. From the example six different sets of values (x, y) will generate 12 functional constraints respectively. Substituting the given values, the LPP becomes,

Min $\{f^{l}(x)\}$ = Minimize $\{m_0 + 28\ m_1\}$
Subject to the constraints
$a_0 + 1a_1 + 0.3m_0 + 0.3m_1 \geq 3010$
$a_0 + 1a_1 - 0.3m_0 - 0.3m_1 \leq 3010$
$a_0 + 2a_1 + 0.3m_0 + 0.6\ m_1 \geq 4500$
$a_0 + 2a_1 - 0.3m_0 - 0.6\ m_1 \leq 4500$
$a_0 + 4a_1 + 0.3m_0 + 1.2\ m_1 \geq 4400$
$a_0 + 4a_1 - 0.3m_0 - 1.2\ m_1 \leq 4400$
$a_0 + 6a_1 + 0.3m_0 + 1.8\ m_1 \geq 5400$
$a_0 + 6a_1 - 0.3m_0 - 1.8\ m_1 \leq 5400$
$a_0 + 7a_1 + 0.3m_0 + 2.1\ m_1 \geq 7295$
$a_0 + 7a_1 - 0.3m_0 - 2.1\ m_1 \leq 7295$
$a_0 + 8a_1 + 0.3m_0 + 2.4\ m_1 \geq 8195$
$a_0 + 8a_1 - 0.3m_0 - 2.4m_1 \geq 8195$
where $a_0, a_1, m_0, m_1 \geq 0$.

By using the software Tora, the estimated values of the center and spread of membership function is given by
a_0= 2486.6667, a_1= 615.8333, m_0= 2605.5556, m_1= 0.
where the minimized value of the objective function on the spread is
Min $\{f^{l}(x)\}$ = Minimize $\{m_0 + 28\ m_1\}$= 2605.5556
Now the FN coefficients are as follows:
$\tilde{A}_0$ = [2486.6667; 2605.5556, 2605.5556]
$\tilde{A}_1$ = [615.8333; 0, 0]
The FLR model is given by
$\tilde{y} = \ \tilde{A}_0 + \tilde{A}_1 x_1$
$\tilde{y}$= [2486.6667; 2605.55,2605.55]+[615.8333; 0, 0](1)
$\tilde{y}$ = [3102.4997;2605.55,2605.55]
The transformed crisp value of $\tilde{y}$ is given by
$Y = [a_1 + 4a_2 + a_3]/6$ = [3102.4997 + 4(2605.55) + 2605.55] /6
Y = 2688.3805
The next value of Y can be calculated here, $\tilde{y} = \ \tilde{A}_0 + \tilde{A}_1 x_2$
$\tilde{y}$= [2486.6667; 2605.55,2605.55]+ [615.833; 0, 0] (2)
$\tilde{y}$ = [3718.3333; 2605.55,2605.55]

The transformed crisp value of $\tilde{y}$ is given by

$Y = [a_1 + 4a_2 + a_3]/6$ = [3718.3333 + 4(2605.55)+ 2605.55] /6 = 2791.0188

Similarly, the other values of Y are estimated and are presented in the following table

x_j	Observed y_j	$\widehat{y_j}$	$e_j = \widehat{y_j} - y_j$	e_j^2
1	3010	2688.381	-321.619	103438.7812
2	4500	2791.018	-1708.98	2920619.476
4	4400	2996.297	-1403.7	1970382.112
6	5400	3201.57	-2198.43	4833094.465
7	7295	3304.21	-3990.79	15926404.82
8	8195	3406.85	-4788.15	22926380.42
				48680320.08

Table 1

5.1. Determination of error using ordinary regression method

The linear regression equation is $Y = a + b\,X$ (A)

The value of a and b can be determined by using the following normal equation.

$\sum X + b\sum X^2 = \sum XY$(B)

$n\,a + b\sum X \qquad = \sum Y$ (C)

The values of X and Y becomes,

X	Y	X^2	XY
1	3010	1	3010
2	4500	4	9000
4	4400	16	17600
6	5400	36	32400
7	7295	49	51065
8	8195	64	65560
28	**32800**	**170**	**178635**

Table 2

From (C) ⇒ 6a + 28b = 32800

From (B) ⇒28a + 170b = 178635

Solving (B) and (C) we have, a = 2433.1356 and b = 650.0423

Therefore the regression model (A) becomes, $\hat{Y}$ = 2433.1356+ 650.0423X

The other values of Y are estimated and are presented in the following table

x_j	Observed y_j	$\widehat{y_j}$	$e_j = \widehat{y_j} - y_j$	e_j^2
1	3010	3083.1779	73.1779	5355.005048
2	4500	3733.2202	-766.78	587951.2617
4	4400	5033.3048	633.3048	401074.9697
6	5400	6333.3894	933.3894	871215.772
7	7295	6983.4317	-311.568	97074.80556
8	8195	7633.474	-561.526	315311.4487
				2277983.263

Table 3

6. Conclusion

In this paper, fuzzy linear regression with symmetrical triangular number is used for prediction of values instead of normal linear regression. Error values is also found out for both fuzzy linear regression and normal linear regression. The error value of normal regression analysis is efficient than that of fuzzy linear regression from the above example.

REFERENCES

1. A.Celmins, A least squares model fitting to fuzzy vector data, *Fuzzy Sets and Systems*, 22 (1987 245-269.
2. H.Hassanpour, H.R.Maleki and M.A.Yaghoobi, Fuzz linear regression model with crisp coefficients: A goal programming approach, *Iranian Journal of Fuzzy System*, 7 (2) (2010)19-39.
3. R.Parvathi, C.Malathi, M.Akram and K.T.Atanassov, Intuitionistic fuzzy linear regression analysis, *Fuzzy Optimization and Decision Making*, 12(2) (2012) 215-229.
4. H.Tanakaand H.Lee, Interval regression analysis by quadratic programming approach, *IEEE Trans. Systems Man Cybrnet.*, 6(4) (1988) 473-481.
5. L.A.Zadeh, Fuzzy sets, *Information and Control*, 8 (1965) 338-353.