Fuzzy linear regression analysis with trapezoidal coefficients

T. Razzaghnia 1,* , E. Pasha 2, E. Khorram 3, A. Razzaghnia 3

1 Department of Statistics, Islamic Azad University, Roudehen Branch, Tehran, Iran.
2 Department of Mathematics, Teacher Training University, Tehran, Iran.
3 Faculty of Mathematics and Computer Science, Amirkabir University of Technology, Tehran, Iran.
* E-mail: t_razaghnia@yahoo.com.

Abstract : In this paper, we aim to extended the constraints of Tanaka’s model. Applied coefficients of the fuzzy regression by them is the symmetric triangular fuzzy numbers, while we try to replace it by more general asymmetric trapezoidal one. Possibility of two asymmetric trapezoidal fuzzy numbers is explained by possibility distribution. Two different models is presented and a numerical example is given in order to compare the proposed models with previous one. Error values shows advantage of the presented models with respect to constraints of Tanaka’s model.

Keywords: Fuzzy numbers, Fuzzy linear regression, Possibility distribution, Mathematical programming.

1. Introduction

Fuzzy linear regression was proposed by Tanaka et al. [17] in 1982. Many different fuzzy regression approaches have been proposed by different researchers since then, and also this subject has drawn much attention from more and more people concerned. In general, there are two approaches in fuzzy regression analysis: linear programming-based method [6,11,16,17] and fuzzy least squares method [1,2,3,4,7,10,12,13,14]. The first method is based on minimizing fuzziness as an optimal criterion. The second method used least-square of errors as a fitting criterion. The advantage of first approach is its simplicity in programming and computation, while that the degree of fuzziness between the observed and predicted values is minimized by using fuzzy least squares method. Tanaka et al.[17] regarded fuzzy data as a possibility distribution, the deviations between the observed values and the estimated values were supposed to be due to fuzziness of system structure. This structure was represented as a fuzzy linear function whose parameters were given by fuzzy sets. They resorted to linear programming to develop their regression model.

In [17], the coefficients of the fuzzy regression are symmetric triangular fuzzy numbers. In this paper, we explain limitations of these methods and we extended the symmetric triangular fuzzy coefficients to asymmetric trapezoidal fuzzy numbers. Also we use possibility distribution for asymmetric trapezoidal fuzzy numbers and we prove a theorem about possibility of two asymmetric trapezoidal fuzzy numbers.

The paper is organized as follows. In section 2, some elementary properties of fuzzy numbers and fuzzy linear regression are described. The propose method is presented in section 3. A numerical example is illustrated to compare the proposed method with previous ones, in section 4.
2. Preliminaries

In this section, we describe fuzzy regression methods based on the linear fuzzy model with symmetric triangular fuzzy coefficient [15, 16]. The aim of fuzzy regression is to minimize the fuzziness of the linear fuzzy model that includes all the given data. So we need some definitions to describe fuzzy regression.

Recall that a fuzzy number \( \tilde{A} \) is a convex normalized fuzzy subset of the real line \( \mathbb{R} \) with an upper semi-continuous membership function of bounded support [5].

**Definition 2.1.** A symmetric fuzzy number \( \tilde{A} \), denoted by \( \tilde{A} = (\alpha, c) \), is defined as

\[
\tilde{A}(x) = L((x - \alpha)/c), \quad c > 0,
\]

Where \( \alpha \) and \( c \) are the center and spread of \( \tilde{A} \) and \( L(x) \) is a shape function of fuzzy numbers such that:

i) \( L(x) = L(-x) \),
ii) \( L(0) = 1, L(1) = 0 \),
iii) \( L \) is strictly decreasing on \( [0, \infty) \),
iv) \( L \) is invertible on \( [0,1] \).

The set of all symmetric fuzzy numbers is denoted by \( F_L(\mathbb{R}) \). If \( L(x) = 1 - |x| \) then the fuzzy number is a symmetric triangular fuzzy number.

**Definition 2.2.** Suppose \( \tilde{A} = (\alpha, c) \) is a symmetric fuzzy number and \( \lambda \in \mathbb{R} \), then \( \lambda \tilde{A} = (\lambda \alpha, |\lambda| c) \).

**Definition 2.3.** An asymmetric trapezoidal fuzzy number \( \tilde{A} \), denote by \( \tilde{A} = (a^{(1)}, a^{(2)}, a^{(2)}, a^{(3)}) \) is defined as

\[
\tilde{A}(x) = \begin{cases} 
L(x - a^{(2)})/(a^{(2)} - a^{(1)}), & x < a^{(2)} \\
1, & a^{(2)} \leq x \leq a^{(2)} \\
R(x - a^{(3)})/(a^{(3)} - a^{(2)}), & x > a^{(2)}
\end{cases}
\]

Where \( a^{(1)}, a^{(2)}, a^{(2)}, a^{(3)} \) are four parameters of the asymmetric trapezoidal fuzzy number.

**Definition 2.4.** Dubois and Prade [5] proposed the possibility of equality between two fuzzy numbers \( \tilde{A} \) and \( \tilde{B} \) as

\[
POS(\tilde{A} = \tilde{B}) = \sup_{x \in \mathbb{R}} \{ \min \{ \tilde{A}(x), \tilde{B}(x) \} \}
\]

Where \( POS \) is short for possibility.

In some fuzzy regression models, deviation between observed values and estimated values are assumed to be due system fuzziness or fuzziness of regression coefficients [17], this assumption is shared by fuzzy regression methods described in this paper. The goal of fuzzy regression analysis is to find a regression model that fits all observed fuzzy data within a specified fitting criterion. Different fuzzy regression models are obtained depending on the fitting criterion used. Tanaka et al. [17] proposed the first linear regression analysis with a fuzzy model. According to this method, the regression coefficients are fuzzy numbers, which can be expressed as interval numbers with membership values. Since the regression coefficients are fuzzy numbers, the estimated dependent variable \( \hat{Y} \) is also a fuzzy number. A fuzzy regression analysis results in the following regression model:

\[
\hat{Y}_i = \tilde{A}_0 X_{i0} + \tilde{A}_1 X_{i1} + \ldots + \tilde{A}_p X_{ip} = \tilde{A} X_i
\]

\[i = 1, 2, \ldots, n\] (1)

In model (1), \( \tilde{A} = (\tilde{A}_0, \tilde{A}_1, \ldots, \tilde{A}_p) \) is a vector of fuzzy parameters where \( \tilde{A}_j = (\alpha_j, c_j) \) is a symmetric fuzzy number, which consists of fuzzy center \( \alpha_j \) and fuzzy half-width \( c_j \). Also, \( \hat{Y}_i = (\tilde{y}_i, c_i)_L \) is the observed value in this model.

In the fuzzy linear regression, the fuzzy parameters are estimated according to certain conditions. One of them is that, for a certain \( h \)-level, \( 0 \leq h \leq 1 \), the support of the estimated values should include the support of the observed values. To determine the fuzzy parameters while minimizing the total sum of the spreads of the estimated values for a certain \( h \)-
level. Tanaka et al. [15,16] formulated a linear programming problem, as follows:

Minimize \( \sum_{i=1}^{n} \sum_{j=0}^{p} c_j |x_{ij}| \)

s. t:
\[
\begin{align*}
\sum_{j=0}^{p} \alpha_j x_{ij} + |L^{-1}(h)| \sum_{j=0}^{p} c_j |x_{ij}| & \geq \bar{y}_j + |L^{-1}(h)| e_i \\
\sum_{j=0}^{p} \alpha_j x_{ij} - |L^{-1}(h)| \sum_{j=0}^{p} c_j |x_{ij}| & \leq \bar{y}_j - |L^{-1}(h)| e_i \\
\end{align*}
\]
i = 1,2, ..., n
\( \alpha_j \in R, \ c_j \geq 0 \)

In Tanaka’s model, the constraints guarantee the support of the estimated values from the model (2) includes the support of the observed values.

3. The proposed approach

In this section, we describe some criticisms in general fuzzy regression with triangular fuzzy numbers and propose a new method to rectify this shortcoming.

In equation (2), if \( h = 1 \), then \( L^{-1}(h) = 0 \). Hence the following equation must hold, 
\( \bar{y}_j = \alpha_0 + \alpha_1 x_{i1} + \cdots + \alpha_p x_{ip}, \ i = 1,2, \cdots, n. \)

Thus in general, there is no solution because the given data do not usually satisfy the equation. In order to handle this difficulty, we use asymmetric trapezoidal fuzzy coefficients.

The extension of symmetric triangular fuzzy coefficient to asymmetric trapezoidal fuzzy coefficients is also effective for avoiding unnecessary fuzziness of the linear fuzzy model [8]. When large membership grades are given to many input-output pairs, the linear fuzzy model with the symmetric triangular fuzzy coefficients tends to have large fuzziness. In some cases, such large fuzziness can be avoided by using asymmetric trapezoidal fuzzy numbers as fuzzy coefficients.

Also, in fuzzy linear regression with asymmetric trapezoidal fuzzy coefficients, we use possibility distribution in constraints.

For the possibility distribution with asymmetric trapezoidal fuzzy numbers, we prove the following theorem.

Theorem 3.1. If \( \tilde{A} = (a^{(1)}, \tilde{a}^{(2)}, a^{(3)}) \) and \( \tilde{B} = (b^{(1)}, \tilde{b}^{(2)}, b^{(3)}) \), then

\[
\begin{align*}
pos(\tilde{A} = \tilde{B}) = \begin{cases} 
\frac{\tilde{a}^{(2)} - \tilde{b}^{(2)}}{(a^{(2)} - a^{(1)}) + (b^{(3)} - \tilde{b}^{(2)})} & \text{if} \quad a^{(2)} \geq \tilde{b}^{(2)} \\
\frac{\tilde{a}^{(2)} - \tilde{b}^{(2)}}{(a^{(2)} - a^{(1)}) + (\tilde{b}^{(3)} - \tilde{b}^{(2)})} & \text{otherwise} \\
\end{cases}
\end{align*}
\]

Proof: i) set 
\[
x^* = \frac{a^{(2)} [b^{(3)} - \tilde{b}^{(2)}] + \tilde{b}^{(2)} [a^{(2)} - a^{(1)}]}{[a^{(2)} - a^{(1)}] + [b^{(3)} - \tilde{b}^{(2)}]}
\]

then 
\[
\tilde{A}(x^*) = \tilde{B}(x^*) = L \left[ \frac{\tilde{a}^{(2)} - \tilde{b}^{(2)}}{(a^{(2)} - a^{(1)}) + (b^{(3)} - \tilde{b}^{(2)})} \right]
\]

therefore,
\[
pos(\tilde{A} = \tilde{B}) = \sup_{x \in R} \left\{ \min \{\tilde{A}(x), \tilde{B}(x)\} \right\} = \sup_{x \in R} \left\{ \min \{\tilde{A}(x^*), \tilde{B}(x^*)\} \right\} = \sup_{x \in R} \left\{ \min \{\tilde{A}(x), \tilde{B}(x)\} \right\} = \sup_{x \in \text{supp}\tilde{A} \cap \text{supp}\tilde{B}} \left\{ \min \{\tilde{A}(x), \tilde{B}(x)\} \right\}
\]

now assume,
\[
pos(\tilde{A} = \tilde{B}) > L \left[ \frac{\tilde{a}^{(2)} - \tilde{b}^{(2)}}{(a^{(2)} - a^{(1)}) + (b^{(3)} - \tilde{b}^{(2)})} \right].
\]

Since \( \tilde{A}(x) = 0 \) for all \( x \in R - \text{supp}\tilde{A} \) and \( \tilde{B}(x) = 0 \) for all \( x \in R - \text{supp}\tilde{B} \), we have
\[
pos(\tilde{A} = \tilde{B}) = \sup_{x \in R} \left\{ \min \{\tilde{A}(x), \tilde{B}(x)\} \right\} = \sup_{x \in \text{supp}\tilde{A} \cap \text{supp}\tilde{B}} \left\{ \min \{\tilde{A}(x), \tilde{B}(x)\} \right\}
\]

\( \text{supp}\tilde{A} \cap \text{supp}\tilde{B} \) is a compact subset of \( R \) with upper semi-continuous membership function of bounded support. Therefore, there exists an \( x_0 \in \text{supp}\tilde{A} \cap \text{supp}\tilde{B} \) such that
From (3) and (4) it is implied that
\[
L\left(\frac{x^0 - a^{(2)}}{a^{(2)} - a^{(1)}}\right) > L\left[\frac{a^{(2)} - \bar{b}^{(2)}}{[a^{(2)} - a^{(1)}] + [b^{(3)} - \bar{b}^{(2)}]}\right]
\]
and
\[
L\left(\frac{x^0 - \bar{b}^{(2)}}{b^{(3)} - \bar{b}^{(2)}}\right) > L\left[\frac{a^{(2)} - \bar{b}^{(2)}}{[a^{(2)} - a^{(1)}] + [b^{(3)} - \bar{b}^{(2)}]}\right]
\]
or equivalently
\[
\left|\frac{x^0 - a^{(2)}}{a^{(2)} - a^{(1)}}\right| < \left|\frac{a^{(2)} - \bar{b}^{(2)}}{[a^{(2)} - a^{(1)}] + [b^{(3)} - \bar{b}^{(2)}]}\right| \quad (5)
\]
and
\[
\left|\frac{x^0 - \bar{b}^{(2)}}{b^{(3)} - \bar{b}^{(2)}}\right| < \left|\frac{a^{(2)} - \bar{b}^{(2)}}{[a^{(2)} - a^{(1)}] + [b^{(3)} - \bar{b}^{(2)}]}\right| \quad (6).
\]

From (5) and (6) it is implied that
\[
\left|x^0 - a^{(2)}\right| + \left|x^0 - \bar{b}^{(2)}\right| < \left|a^{(2)} - \bar{b}^{(2)}\right| \quad (7).
\]
Since (7) contradicts the triangular inequality concept, (3) is concluded.
Part (ii) and (iii) of theorem are proved similarly. □

In model (1), we assume \(\widehat{A}_j = (a^{(1)}_j, a^{(2)}_j, \bar{a}^{(2)}_j, a^{(3)}_j)\) is an asymmetric trapezoidal fuzzy coefficient for \(j = 0, 1, \ldots, p\) and \(\widehat{Y}_i = (y^{(1)}_i, y^{(2)}_i, \bar{y}^{(2)}_i, y^{(3)}_i)\) is an observed values for \(i = 1, 2, \ldots, n\). Now, by using theorem 3.1 and model (1) with trapezoidal fuzzy numbers, we have the following lemma.

**Lemma 3.1:** The degree of fitness of the fuzzy linear regression model with trapezoidal parameters \(f_i\) is calculated as follows:

\[
f_i = \text{POS}(\widehat{Y}_i - \bar{Y}_i) = L\left[\frac{a^{(2)}_i}{[a^{(2)}_i - a^{(1)}] + [b^{(3)}_i - \bar{b}^{(2)}_i]}\right] \quad (8)
\]

By these constraints, we propose two models for fuzzy linear regression and compare these models with Tanaka’s constraints in trapezoidal fuzzy coefficients.

In model (1), the objective function is to minimize the total spreads of the estimated values and in model (2) the objective function is to minimize the square of total difference between the observed spread and the estimated spread.

**The constraints of two models:** The problem in the fuzzy linear regression models is to determine fuzzy parameters \(\widetilde{A}\) such that \(f_i \geq h, \forall i\).

Therefore we have two following mathematical models:

**Model 1:** Minimize
\[
Z = \sum_{j=0}^{p} (a^{(3)}_j - a^{(1)}_j + \bar{a}^{(2)}_j - a^{(2)}_j) \sum_{i=1}^{n} |y_i|
\]
s.t:
\[
h\sum_{j=0}^{p} (a^{(2)}_j x^+_i - \bar{a}^{(2)}_j x^-_i) + (1 - h)\sum_{j=0}^{p} (a^{(1)}_j x^+_i - \bar{a}^{(3)}_j x^-_i) \leq h \bar{y}^{(2)}_i + (1 - h)\bar{y}^{(3)}_i, \quad i = 1, \ldots, n \quad (9)
\]
\[
h\sum_{j=0}^{p} (a^{(2)}_j x^+_i - \bar{a}^{(2)}_j x^-_i) + (1 - h)\sum_{j=0}^{p} (a^{(1)}_j x^+_i - \bar{a}^{(3)}_j x^-_i) \geq h y^{(2)}_i + (1 - h)y^{(3)}_i, \quad i = 1, \ldots, n \quad (10)
\]
\[
a^{(1)}_j \leq a^{(2)}_j \leq \bar{a}^{(2)}_j \leq a^{(3)}_j, \quad j = 0, 1, \ldots, p \quad (11)
\]

**Model 2:** Minimize
\[
Z = \sum_{j=0}^{p} \sum_{i=1}^{n} \left(\frac{x^+_i}{x^+_i} \frac{y_i}{y_i} + \frac{\bar{a}^{(2)}_j - a^{(2)}_j}{x^+_i} \frac{y_i}{y_i} + \frac{a^{(2)}_j}{x^+_i} - y_i^2 - y_i^2\right)^2
\]
s.t: (9),(10),(11), where
\[
x^+_i = \begin{cases} x_i & x_i \geq 0 \\ 0 & \text{otherwise} \end{cases}, \quad x^-_i = \begin{cases} -x_i & x_i < 0 \\ 0 & \text{otherwise} \end{cases}
\]
To evaluate the performance of a fuzzy regression model, Kim and Bishu in [9] used the absolute difference between the membership values of the observed and estimated values as:

\[ E_i = \int_{\text{Supp}(\tilde{Y}_i) \cup \text{Supp}(\tilde{\tilde{Y}}_i)} | \tilde{Y}_i(y) - \tilde{\tilde{Y}}_i(y) | \, dy \]  

(13)

In other words, \( E_i \) is the estimation error. If \( E_i \) tends to zero, then the fitting is the best.

4 Numerical example

In this section, we use an example to compare our proposed methods with Tanaka’s constraints method. In this example, there are five pairs of observations \((x_i, y_i)\) as shown in Table 1.

By using the proposed methods, estimated parameters are shown in Table 2. To compare the performance of the fuzzy linear regression models, (13) is applied to calculate the errors in estimation the observed responses. The total error of the proposed method for model 1 is 26.71 (see Table 2) and for model 2 is 19.19, which are better than the total error of 31.03 calculated by Tanaka’s constraints.

**Table 1: Numerical data**

<table>
<thead>
<tr>
<th>( x_i )</th>
<th>( \tilde{y} = (y^{(1)}, y^{(2)}, y^{(3)}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(7,8,9,10)</td>
</tr>
<tr>
<td>2</td>
<td>(5.4,6.4,7.4,8.4)</td>
</tr>
<tr>
<td>3</td>
<td>(8.5,9.5,10.5,11.5)</td>
</tr>
<tr>
<td>4</td>
<td>(12.5,13.5,14.5,15.5)</td>
</tr>
<tr>
<td>5</td>
<td>(12.2,13.2,14.2,15.2)</td>
</tr>
</tbody>
</table>

The observed values and estimated values for Tanaka’s constraints model and two proposed models are shown in Figure 1, 2, 3. The dependents values in Figure 2, 3 are estimated better than the values in Figure 1.
Table 2: Estimated coefficients and estimated errors (h=0.5)

<table>
<thead>
<tr>
<th>Estimated coefficients</th>
<th>Tanaka’s constraints</th>
<th>Proposed Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a_i^{(1)}, a_i^{(2)}, \overline{a}_i^{(2)}, a_i^{(3)})</td>
<td>(0.32,4.15,5.75,9.58)</td>
<td>Model 1: (3.72,4.74,5.16,6.17)</td>
</tr>
<tr>
<td>(\overline{a}_i^{(1)}, a_i^{(2)}, \overline{a}_i^{(2)}, a_i^{(3)})</td>
<td>(1.83,1.83,1.83,1.83)</td>
<td>Model 2: (2.67,4.16,4.83,5.53)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Error</th>
<th>1</th>
<th>1.47</th>
<th>1.77</th>
<th>3.38</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>7.32</td>
<td>6.26</td>
<td>4.83</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>6.78</td>
<td>5.90</td>
<td>3.64</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4.41</td>
<td>3.23</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>11.04</td>
<td>9.56</td>
<td>6.67</td>
</tr>
<tr>
<td>Total error</td>
<td>31.03</td>
<td>26.71</td>
<td>19.19</td>
<td></td>
</tr>
</tbody>
</table>

References