A goal programming model for computation of Fuzzy Linear Regression

with least error

Sohrab Abdollahzadeh
Industrial Eng. Dept
Urmia University of Technology
Uromiyeh, Iran
inassa@uut.ac.ir

Mina Rezaei
Computer Eng. Dept
Azad University of Arak
Arak, Iran
mi_rezaei@gmail.com

Mehdi Dabbaghian
Computer Eng. Dept
Roozbeh Institute of higher education
Zanjan, Iran
m.dabbaghian@engineer.com

Ali Norouzi
Computer Eng. Dept
Istanbul Technical University
Istanbul, Turkey
norouzi@ogr.istanbul.edu.tr

Abstract—The paper presents a simple method for computation of fuzzy linear regression with fuzzy output and crisp inputs. The proposed method is based on Goal Programming technique and for estimation upper and lower fuzzy bands, two separated linear programming models are applied. The proposed method is minimized the estimation error between observed and estimated values and has better performance in comparison with previous approaches. The performance of the method is evaluated by applying the considered data in the same example of other methods and the results illustrate decreasing in total error.

Keywords-component: Fuzzy numbers, Fuzzy Linear Regression, Fuzzy Linear Programming, Goal Programming

I. INTRODUCTION

In 1982, Tanaka et al. [1] introduced Fuzzy Linear Regression (FLR). In order to estimate of regression parameters, they applied linear programming and minimized the total spread of the fuzzy parameters subject to covering the observed values by estimated values. Although their approach is improved by many researchers [2, 3, 4, 5, 6], this approach is still one of the most frequently and simplest methods for estimating parameters of fuzzy regression.

The important problem associated with the Tanaka approach is the influence of outliers on the predicted upper and lower fuzzy bands. The Tanaka model is very sensitive to outliers and the outliers are caused that the fuzzy linear regression is not able to predict, correctly. To overcome this problem, several studies are done which discuss about handling the problem of outliers [2, 4, 7, and 8].

To evaluate the performance of a fuzzy regression model, Kim and Bishu [5] proposed an approach based on the criterion of minimization the different of membership values between the given data and estimated values. In this paper, to calculate of estimation error, the Kim and Bishu [5] approach is applied and the results are compared with Tanaka et al. [9], Savic and Pedrycz [10], Diamond [3] and Modarses et al [11] methods.

Although there are two approaches in fuzzy regression analysis: linear programming and least squares method, the first model due to simplicity in computation and programming is more common. In this paper, the first approach is applied for computation of fuzzy linear regression. The paper is organized as follows. In Section 2, some notations of fuzzy numbers are reviewed. Fuzzy linear regression proposed method is represented in sections 3. In section 4, to illustrate the performance method, one numerical example is solved and results are compared with previous methods.

II. NOTATION OF FUZZY NUMBERS

In this section, the definitions of need fuzzy data are review. Let $X$ is a classical set of objects. The membership of $x$ number is showed as $\mu_A(x) \in [0,1]$ and is defined as follows:

$$\mu_A(x) = \begin{cases} 1 & \text{if and only if } x \in A \\ 0 & \text{Otherwise} \end{cases}$$

Definition 1: If $\tilde{A}$ be a fuzzy set, and then is showed as follows [12]:

$$\tilde{A} = \{(x, \mu_{\tilde{A}}(x)) \mid x \in X\}$$

Definition 2 (\(\alpha\)-cut): The \(\alpha\)-cut of a fuzzy set \(\tilde{A}\) is a crisp subset of $X$ and define as follow [12]:

$$[\tilde{A}]_\alpha = \{x \mid \mu_{\tilde{A}}(x) \geq \alpha\}$$

Definition 3 (Triangular fuzzy numbers): The triangular fuzzy number can be defined as $\tilde{A} = (a, m, n)$, where $a$ is the central value, $m$ and $n$ are the left and right spread, respectively. The triangular fuzzy number is shown in Figure 1.
III. FUZZY LINEAR REGRESSION

Tanaka et al. [1] proposed the fuzzy linear regression (FLR) model in the case of crisp input and fuzzy output data set as follow:
\[ \hat{Y} = A_0 + A_1 x_1 + A_2 x_2 + \ldots + A_k x_k \]  

(1)

Where \( A_j = (a_j, c_j) \), \( j = 0, 1, \ldots, k \) is assumed to be a symmetric triangular fuzzy number with center \( a_j \) and half-width \( c_j \). To estimate \( A_j \), Tanaka et al. [1] applied the following model:

\[
\begin{align*}
\min & \sum_{j=0}^{k} c_j \\
\text{s.t.:} & \\
\sum_{j=0}^{k} (a_j + (1-H)c_j)x_{ij} & \geq \bar{y}_i + (1-H)e_i \quad i = 1, 2, \ldots, n \\
\sum_{j=0}^{k} (a_j - (1-H)c_j)x_{ij} & \leq \bar{y}_i - (1-H)e_i \quad i = 1, 2, \ldots, n \\
\alpha_j & = \text{free}, c_j \geq 0, j = 0, 1, \ldots, k
\end{align*}
\]

(2)

In model (2), \( n \) is the number of observations and \( H \in [0,1] \) is the threshold level to be chosen by decision maker. Later, model (2) has been modified by other researchers [9, 13]. They suggested the objective function should be as

\[
\min \sum_{i=0}^{n} \sum_{j=0}^{k} c_j x_{ij}
\]

To prevent of being \( c_i \)'s = 0. Modarres et al [12] proposed a mathematical programming model to calculate parameters of fuzzy linear regression and showed the proposed method have better performance than the previous studies. In this paper, the method based on goal programming is suggested which has least error.

A. Proposed Model

In this section, following Hojati et al [14], the proposed model is explained. This model applies Goal Programming (GP) for estimating the linear regression parameters.

Let \( \hat{y}_{iL} \), \( \bar{y}_i \) and \( \hat{y}_{iU} \) be the upper, center and lower points of \( i^{th} \) observed interval, and \( \bar{y}_{iL} \) , \( \bar{y}_{iU} \) be the upper and lower point of the \( i^{th} \) predicted interval. Moreover, \( \hat{y}_{iL} \) and \( \hat{y}_{iU} \) are considered as predicted fuzzy bands, too. In this model, it is allowed \( \hat{y}_{iU} \) to be larger than \( \hat{y}_{iL} \), but smaller than \( \bar{y}_i \), and \( \hat{y}_{iL} \) to be larger than \( \bar{y}_i \), but smaller than \( \hat{y}_{iU} \). In fact, the sum of deviations of \( \hat{y}_{iU} \) from \( \bar{y}_i \) and \( \hat{y}_{iU} \) and the sum of deviations of \( \hat{y}_{iL} \) from \( \bar{y}_i \) and \( \bar{y}_{iL} \) is minimized. In other worth, to obtain \( \hat{y}_{iU} \), \( \bar{y}_i \) is selected as upper point and to obtain \( \hat{y}_{iL} \), \( \bar{y}_i \) is selected as lower point. Note that the fuzzy bands \( \hat{y}_{iL} \) and \( \hat{y}_{iU} \) are calculated, separately. First, a goal programming model is solved for obtained lower fuzzy band, \( \hat{y}_{iL} \) and then, another model is run for obtained upper fuzzy band, \( \hat{y}_{iU} \). Furthermore, these bands are less affected by outliers.

Because, in estimation of \( \hat{y}_{iL} \) and \( \hat{y}_{iU} \), \( \bar{y}_i \) is used which is less sensitive other than the lower and upper points. Then, the difference between the proposed model and previous models is in estimating of \( \hat{y}_{iL} \) and \( \hat{y}_{iU} \) fuzzy bands.

To obtained \( \hat{y}_{iL} \) band, the model (3) is solved.

\[
\begin{align*}
\min \sum_{i=1}^{n} (d_{iL}^+ + d_{iU}^- + d_{iL}^- + d_{iU}^+)
\end{align*}
\]

(3)

s.t.:

\[
\begin{align*}
\sum_{j=0}^{k} (a_j + (1-H)c_j)x_{ij} & + d_{iL}^- - d_{iU}^+ = \bar{y}_i - (1-H)e_i \\
\sum_{j=0}^{k} (a_j - (1-H)c_j)x_{ij} & + d_{iL}^- + d_{iU}^- = \bar{y}_i - (1-H)e_i \\
d_{iL}^+, d_{iU}^+, c_j & \geq 0, a_j = \text{free}
\end{align*}
\]

In the first constraint, the \( \bar{y}_i + (1-H)e_i \) is replaced by \( \bar{y}_i (e = 0) \) in order the upper point to be the \( \bar{y}_i \) values when predicting \( \hat{y}_{iL} \). In the model (3), \( |d_{iL}^+ - d_{iU}^-| \) is the distance between \( \bar{y}_i \) and \( \hat{y}_{iL} \) and \( |d_{iL}^- - d_{iU}^+| \) is the distance between lower point of \( H \)-certain observed interval and \( \hat{y}_{iL} \). Thus, the sum of two deviations to be minimized.

The \( \hat{y}_{iU} \) band is obtained by solving GP model following:

\[
\begin{align*}
\min \sum_{i=1}^{n} (d_{iL}^+ + d_{iU}^- + d_{iL}^- + d_{iU}^+)
\end{align*}
\]

(4)

s.t.:

\[
\begin{align*}
\sum_{j=0}^{k} (a_j + (1-H)c_j)x_{ij} & + d_{iL}^- - d_{iU}^+ = \bar{y}_i + (1-H)e_i \\
\sum_{j=0}^{k} (a_j - (1-H)c_j)x_{ij} & + d_{iL}^- + d_{iU}^- = \bar{y}_i + (1-H)e_i \\
d_{iL}^+, d_{iU}^+, c_j & \geq 0, a_j = \text{free}
\end{align*}
\]
In each models (3) and (4), there is one estimated band for lower and upper points. The upper line of model (3) (lower band) and the lower line of model (4) (upper band) are located around \( \tilde{y}_i \)'s and could be eliminated to decrease predicted error (see Fig. 2). Thus, the lower and upper fuzzy bands are:

\[
\hat{y}_L = (\alpha_{0L} - c_{0L}) + (\alpha_{1L} - c_{1L})x \\
\hat{y}_U = (\alpha_{0U} + c_{0U}) + (\alpha_{1U} + c_{1U})x
\]  

(5) (6)

Where the \( \alpha_{jL} \) and \( c_{jL} \) are the estimated value for \( \hat{y}_L \), and \( \alpha_{jU} \) and \( c_{jU} \) are the estimated value for \( \hat{y}_U \).

Since, in the proposed model, it is probable that the slops of \( \hat{y}_L \) and \( \hat{y}_U \) lines to be different, the means of \( \alpha_j \)'s and \( c_j \)'s \( (j \neq 0) \) can be used as the slops of \( \hat{y}_L \) and \( \hat{y}_U \) lines. The means of \( \alpha_j \)'s and \( c_j \)'s \( (j \neq 0) \) are calculated as follows:

\[
\alpha_j = \frac{1}{2}(\alpha_{jL} + \alpha_{jU}) \quad j = 1,\ldots,k
\]  

(7)

\[
c_j = \frac{1}{2}(c_{jL} + c_{jU}) \quad j = 1,\ldots,k
\]  

(8)

Where the \( \alpha_{jL} \) and \( c_{jL} \) are the estimated value for \( \hat{y}_L \), and \( \alpha_{jU} \) and \( c_{jU} \) are the estimated value for \( \hat{y}_U \).

IV. NUMERICAL EXAMPLE

To illustrate the proposed model, in this section, we use the same example which is used by Kim and Bishu [5] to illustrate how the proposed method performs. We compare the results of our method with Modarres et al. [11] method and previous methods. To evaluate the performance of a fuzzy regression model, Kim and Bishu [5] used the ratio of the difference between the membership values to the observed membership values as follows:

\[
Error = \frac{\int S_{\hat{y}_i} S_{Y_i} |\tilde{Y}(y) - Y(y)| dy}{\int S_{\hat{y}_i} Y_i dy}
\]  

(9)

Where and \( S_{\hat{y}_i} \) and \( S_{Y_i} \) are the support of \( \tilde{Y}_i \) and \( Y_i \), respectively. To compare the performance of the FLR models, Eq. (9) is applied to calculate the errors in estimation the observed responses. The data given for this example was used by Tanaka et al [1]. These values list in Table 1. By solving models (3) and (4) (at H=0) and modification the values \( \alpha_i \)'s and \( c_i \)'s, two below fuzzy bands are calculated:

\[
\hat{y}_L = (5.55,1.2) \oplus (1.325,0)x
\]  

(10)

\[
\hat{y}_U = (7.20,1.2) \oplus (1.325,0)x
\]  

(11)

The right half of Table 1 shows the errors of the five observations for the different methods. The total error of the proposed method is 3.82 which obviously better than the other total errors.

V. CONCLUSION

In this paper, a simple method presented for computing of fuzzy linear regression. This method is based on Goal Programming technique. The proposed method has the least error and very easy in computation of fuzzy linear regression parameters. The same example by previous studies is solved by using the proposed model and the results are compared with previous approaches. The results shows that our method has the goodness fit depend on both the observation and fuzzy bands with the minimum error.
Table 1: Numerical data and the estimation errors

<table>
<thead>
<tr>
<th>i</th>
<th>x_i</th>
<th>(y_i,e_i)</th>
<th>Errors in estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>(8.0,1.8)</td>
<td>1.86</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>(6.4,2.2)</td>
<td>1.30</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>(9.5,2.6)</td>
<td>0.58</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>(13.5,2.6)</td>
<td>0.86</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>(13.0,2.4)</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>Total error</td>
<td></td>
<td>5.6</td>
</tr>
</tbody>
</table>

REFERENCES


