Bootstrap Statistical Inference about the Regression Coefficients Based on Fuzzy Data

M. G. Akbari, R. Mohammadalizadeh, and M. Rezaei

Abstract

Theory of regression methods is based on the crispness of the observations and the parameters of interest. But there can be many different situations in which the above mentioned concepts are imprecise. On the other hand, the theory of fuzzy sets is a well established tool for formulation and analysis of imprecise and subjective concepts. In these times we must use the fuzzy regression. In this paper, at first we use a well-known signed distance, then with using this signed distance we estimate the crisp regression coefficients based on fuzzy data using least square method. Finally, we exhibit confidence interval and hypothesis testing for these coefficients based on bootstrap theory and numerical examples are also provided to illustrate the approach. In case of the confidence interval and hypothesis testing problem, bootstrap techniques (Efron and Tibshirani, [10]) have empirically been shown to be efficient and powerful.

Keywords: Canonical fuzzy number, Yao-Wu signed distance, Bootstrap theory, Linear regression, Confidence interval, Hypothesis testing.

1. Introduction

Fuzzy linear regression models are used to obtain and appropriate linear relation between a dependent variable and several independent variables in a fuzzy environment. A fuzzy linear regression model was first introduced by Tanaka et al. [27]. They formulated a linear regression model with fuzzy response data, crisp predictor data and fuzzy parameters as a mathematical programming problem. Their approach was later improved by Tanaka et al. [25, 26, 28]. Diamond [9] proposed the fuzzy least square approach to determine fuzzy parameters by defining a metric between two fuzzy numbers. Xizhao and Minghu [30] introduced a new principle of estimating parameters, which was called Minmax principle. The Minmax estimation of the regression model parameters and the LS estimation of the approximation model parameters were discussed. Wu [29] presented the fuzzy estimates of regression parameters in linear regression models for imprecise input and output data with the help of “Resolution Identity”. Nasrabad et al. [20] considered fuzzy linear regression models with fuzzy-crisp output, fuzzy-crisp input, and proposed an estimated method along with a mathematical-programming-based approach. Nätter [21] exhibited different approaches to deal with regression analysis when the data were fuzzy. Bargielaa et al. [7] proposed an iterative algorithm for multiple regression with fuzzy variables. Ge and Wang [11] extended the fuzzy linear regression model first to its regularized version, i.e. regularized fuzzy linear regression model, so as to enhance its generalization capability; then regularized fuzzy linear regression model was explained as the corresponding equivalent maximum a posteriori problem. González-Rodríguez et al. [12] considered estimation of a simple linear regression model for fuzzy random variables, specifically, the least-squares estimation problem in terms of a versatile metric was addressed. Lina and Pai [18] studied a fuzzy support vector regression model for business cycle predictions. Azadeh et al. [6] presented an integrated fuzzy regression algorithm for energy consumption estimation with non-stationary data (a case study of Iran). Sener and Karsaka [24] presented a fuzzy multiple objective decision framework that includes not only fulfillment of engineering characteristics to maximize customer satisfaction, but also maximization of extendibility and minimization of technical difficulty of engineering characteristics as objectives subjected to a financial budget constraint to determine target levels of engineering characteristics in product design. The most literatures about the regression model in fuzzy environments are investigated based on the independent fuzzy random variables [11, 12, 21, 26] or using linear programming methods [6, 20, 24, 26]. These approaches for estimating of parameter of regression models in some cases are complex. In addition to, we can not use these
approaches, if some items of model do not exist. Hence, in this paper, we estimate the parameters of regression model based on Yao and Wu [32] signed distance. Also, some statistical inference about parameters of model is investigated using the bootstrap approaches. It should be mentioned that, the present work has some advantages as follows:

1. The proposed method has achieved a simpler procedure for estimating of parameters of model. Also, it is available and useful for many researchers in the different sciences [6, 24].

2. The present method in some cases is better than Yang and Lin [31] and Aragpour and Tata [5] (see Section 4).

We organize the matter in the following way: In section 2 we describe some basic concepts of canonical fuzzy numbers and Yao-Wu signed distance. Section 3 is devoted to estimate regression coefficients using least fuzzy numbers and Yao-Wu signed distance. Section 4).

Yang and Lin [31] and Aragpour and Tata [5] (see

Let \( \mathcal{F} \) be a binary operation \( \oplus \) or \( \ominus \) between two canonical fuzzy numbers \( \tilde{a} \) and \( \tilde{b} \). The membership function of \( \tilde{a} \oplus \tilde{b} \) is defined by

\[
\mu_{\tilde{a} \oplus \tilde{b}}(x) = \sup_{y \in \mathbb{R}} \min\{\mu_{\tilde{a}}(x), \mu_{\tilde{b}}(y)\}
\]

for \( \oplus = \ominus \) or \( \ominus \) and \( \ominus = + \) or \(-\).

Let \( \mathcal{O}_{\text{int}} \) be a binary operation \( \mathcal{O}_{\text{int}} \) or \( \ominus \) between two closed intervals \( \tilde{a}_a = [a^L_a, a^U_a] \) and \( \tilde{b}_a = [b^L_a, b^U_a] \). Then

\[
\tilde{a}_a \mathcal{O}_{\text{int}} \tilde{b}_a = \{z \in \mathbb{R} : z = x \cdot y, x \in \tilde{a}_a, y \in \tilde{b}_a\}.
\]

If \( \tilde{a} \) and \( \tilde{b} \) be two closed fuzzy numbers, then \( \tilde{a} \oplus \tilde{b} \) and \( \tilde{a} \ominus \tilde{b} \) are also closed fuzzy numbers. Furthermore, we have

\[
(\tilde{a} \oplus \tilde{b})_a = \tilde{a}_a \mathcal{O}_{\text{int}} \tilde{b}_a = [a^L_a + b^L_a, a^U_a + b^U_a];
\]

\[
(\tilde{a} \ominus \tilde{b})_a = \tilde{a}_a \mathcal{O}_{\text{int}} \tilde{b}_a = [a^L_a - b^U_a, a^U_a - b^L_a].
\]

### B. Yao-Wu signed distance

Now we define a distance between fuzzy numbers which is used later. Several ranking methods have been proposed so far, by Puri and Ralescu [23], Cheng [8], Modarres and Sadi-Nezhad [19], Nojavan and Ghazanfari [22], and Akbari and Rezaei [1].

In this paper we use another metric for canonical fuzzy numbers that is known as Yao and Wu signed distance.

**Definition 1:** For each \( a, b \in R \) define the signed distance \( d^*(a, b) = a - b \). Thus, we have the following way to define the rank of any two numbers on \( R \). For each \( a, b \in R \),

\[
d^*(a, b) > 0 \iff d^*(a, 0) > d^*(b, 0) \iff a > b,
\]

\[
d^*(a, b) < 0 \iff d^*(a, 0) < d^*(b, 0) \iff a < b.
\]

**Definition 2:** Let \( F(R) \) be all of canonical fuzzy numbers on the real line \( R \). For each \( \tilde{a}, \tilde{b} \in F(R) \) define the signed distance of \( \tilde{a} \) and \( \tilde{b} \) as follows:

\[
d(\tilde{a}, \tilde{b}) = \int_0^1 d^*(M_a(\tilde{a}), M_b(\tilde{b})) \, d\alpha
\]

\[
= \int_0^1 (M_a(\tilde{a}) - M_b(\tilde{b})) \, d\alpha
\]

where, \( M_a(\tilde{a}) = \frac{a^L_a + a^U_a}{2} \) and \( M_b(\tilde{b}) = \frac{b^L_b + b^U_b}{2} \).

**Definition 3 (Yao and Wu, [32]):** For each \( \tilde{a}, \tilde{b} \in F(R) \), define the ranking of \( \tilde{a} \) and \( \tilde{b} \) by

\[
d(\tilde{a}, \tilde{b}) > 0 \iff d(\tilde{a}, 0) > d(\tilde{b}, 0) \iff \tilde{a} > \tilde{b},
\]

\[
d(\tilde{a}, \tilde{b}) < 0 \iff d(\tilde{a}, 0) < d(\tilde{b}, 0) \iff \tilde{a} < \tilde{b},
\]
$d(\bar{a}, \bar{b}) = 0 \iff d(\bar{a}, 0) = d(\bar{b}, 0) \iff \bar{a} \approx \bar{b}$.

3. Least squares error

Let $(\bar{x}_i, \bar{y}_i), i = 1, 2, \ldots, n$ be $n$ pairs of canonical fuzzy numbers. For any fuzzy line

$$\hat{y}_i = \beta_1 \hat{x}_i + \hat{\beta}_0 \hat{\xi}_i, \text{ } i = 1, 2, \ldots, n$$

sum of squares error (SSE) is defined to be

$$\text{SSE} = \sum_{i=1}^{n} d^2(\hat{y}_i, \beta_1 \hat{x}_i + \beta_0)$$

where $d$ is Yao- Wu signed distance.

If $\beta_0$ and $\beta_1$ are least squares estimators of $\beta_0$ and $\beta_1$, then we have

$$\min_{\beta_0, \beta_1} \sum_{i=1}^{n} d^2(\hat{y}_i, \beta_1 \hat{x}_i + \beta_0) = \sum_{i=1}^{n} d^2(\hat{y}_i, \beta_1 \hat{x}_i + \beta_0)$$

The minimized values of $\beta_1$ and $\beta_0$ are:

$$\beta_1 = \frac{1}{n} \sum_{i=1}^{n} d(\hat{x}_i, 0) d(\hat{y}_i, 0) - \frac{1}{n} \sum_{i=1}^{n} d(\hat{x}_i, 0)^2 \beta_0 = \frac{1}{n} \sum_{i=1}^{n} d(\hat{y}_i, 0) - \beta_1 \frac{1}{n} \sum_{i=1}^{n} d(\hat{x}_i, 0)$$

By using the properties of Yao- Wu signed distance we verify

$$\beta_1 = \frac{\sum_{i=1}^{n} d(\hat{x}_i, \hat{\bar{x}}) d(\hat{y}_i, \hat{\bar{y}})}{\sum_{i=1}^{n} d^2(\hat{x}_i, \hat{\bar{x}})} = \frac{S_{\bar{x}\bar{y}}}{S_{\bar{x}\bar{x}}} \beta_0 = d(\hat{\bar{y}}, 0) - \beta_1 d(\hat{\bar{x}}, 0)$$

where $S_{\bar{x}\bar{y}}$ and $S_{\bar{x}\bar{x}}$ are equal to $\sum_{i=1}^{n} d(\hat{x}_i, \hat{\bar{x}}) d(\hat{y}_i, \hat{\bar{y}})$ and $\sum_{i=1}^{n} d^2(\hat{x}_i, \hat{\bar{x}})$ respectively, furthermore $\hat{\bar{x}} = \frac{1}{n} \sum_{i=1}^{n} \hat{x}_i$.

Theorem 1: If $(\hat{x}_i, \hat{y}_i), i = 1, 2, \ldots, n$ are pairs canonical fuzzy numbers then

$$SSR = \sum_{i=1}^{n} d^2(\beta_1 \hat{x}_i + \beta_0, \hat{\bar{y}}) = \frac{S_{\bar{x}\bar{y}}^2}{S_{\bar{x}\bar{x}}}$$

Proof:

$$SSR = \sum_{i=1}^{n} d^2(\beta_1 \hat{x}_i + \beta_0, \hat{\bar{y}}) = \sum_{i=1}^{n} [d(\beta_1 \hat{x}_i, \hat{\bar{y}}) + d(\beta_0, 0)]^2 = \sum_{i=1}^{n} d(\beta_1 \hat{x}_i, \hat{\bar{y}}) + d(\beta_0, 0)^2 = \sum_{i=1}^{n} \beta_1 d(\hat{x}_i, 0) - d(\hat{\bar{y}}, 0) + d(\hat{\bar{y}}, 0) - \beta_1 d(\hat{\bar{x}}, 0)^2 = \beta_1 \sum_{i=1}^{n} d^2(\hat{x}_i, \hat{\bar{x}}) = \frac{S_{\bar{x}\bar{y}}^2}{S_{\bar{x}\bar{x}}}$$

Theorem 3: Let $(\hat{x}_i, \hat{y}_i), i = 1, 2, \ldots, n$ be pairs canonical fuzzy numbers. Hence we have

$$SST = SSR + SSE$$

where $SST = \sum_{i=1}^{n} d^2(\hat{y}_i, \hat{\bar{y}})$.

Proof: The facts that are used to prove the theorem are as follows:

\[a) \quad d(\hat{y}_i, 0) = d(\beta_1 \hat{x}_i + \beta_0, 0) = \beta_1 d(\hat{x}_i, 0) + \beta_0 = \beta_1 d(\hat{x}_i, 0) + \beta_1 d(\hat{\bar{y}}, 0) - \beta_1 d(\hat{\bar{y}}, 0) + \beta_1 d(\hat{\bar{y}}, 0) = \beta_1 d(\hat{x}_i, \hat{\bar{y}})\]

\[b) \quad d(\hat{y}_i, \hat{y}_i) = d(\hat{y}_i, 0) - d(\hat{\bar{y}}, 0) = d(\hat{y}_i, 0) - \beta_1 d(\hat{x}_i, 0) = \beta_1 d(\hat{x}_i, \hat{\bar{y}})\]

in view of (1) and (2) we simply have

$$\sum_{i=1}^{n} d(\hat{y}_i, 0) d(\hat{y}_i, 0) = \sum_{i=1}^{n} d(\hat{y}_i, 0) - \sum_{i=1}^{n} d^2(\hat{y}_i, 0) + \sum_{i=1}^{n} d(\hat{y}_i, 0) d(\hat{y}_i, 0) = \sum_{i=1}^{n} d(\hat{y}_i, 0)$$

Now by (3) and equality $d(\hat{y}_i, 0) = d(\hat{y}_i, 0)$ we have

$$\sum_{i=1}^{n} d(\hat{y}_i, \hat{y}_i) = \sum_{i=1}^{n} d(\hat{y}_i, \hat{y}_i)$$.  

and also

$$d(\hat{y}_i, \hat{\bar{y}}) = d(\hat{y}_i, \hat{\bar{y}}) = d(\hat{y}_i, \hat{\bar{y}}) = d(\hat{\bar{y}}, \hat{\bar{y}})$$

which results in:

$$d^2(\hat{\bar{y}}, \hat{\bar{y}}) = d^2(\hat{y}_i, \hat{\bar{y}}) + d(\hat{y}_i, \hat{\bar{y}}) + 2 d(\hat{\bar{y}}, \hat{\bar{y}})$$

Finally, with use (4) and (5) we have

$$SST = \sum_{i=1}^{n} d^2(\hat{y}_i, \hat{\bar{y}}) = \sum_{i=1}^{n} d^2(\hat{y}_i, \hat{\bar{y}}) + \sum_{i=1}^{n} d^2(\hat{y}_i, \hat{\bar{y}}) + 2 \sum_{i=1}^{n} d(\hat{\bar{y}}, \hat{\bar{y}})$$

A statistic that is used to show how well the fuzzy line is fitted to fuzzy data is the coefficient of determination. It is defined as the ratio of the regression sum of squares (SSR) to the total sum of squares (SST) also It is shown by $R^2$ and is written as follows:

$$R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^{n} d^2(\beta_1 \hat{x}_i + \beta_0, \hat{\bar{y}})}{\sum_{i=1}^{n} d^2(\hat{y}_i, \hat{\bar{y}})}$$

Definition 4: The correlation of $\bar{x}$ and $\bar{y}$ is defined as follows:

$$\rho(\bar{x}, \bar{y}) = \frac{\text{cov}(\bar{x}, \bar{y})}{\sigma_{\bar{x}} \sigma_{\bar{y}}}$$

where cov$(\bar{x}, \bar{y})$, $\sigma_{\bar{x}}$ and $\sigma_{\bar{y}}$ are equal to $\frac{1}{n} \sum_{i=1}^{n} d(\hat{x}_i, \hat{\bar{x}}) d(\hat{y}_i, \hat{\bar{y}})$, $\frac{1}{n} \sum_{i=1}^{n} d^2(\hat{x}_i, \hat{\bar{x}})$ and $\frac{1}{n} \sum_{i=1}^{n} d^2(\hat{y}_i, \hat{\bar{y}})$ respectively, furthermore $\hat{\bar{x}} = \frac{1}{n} \sum_{i=1}^{n} \hat{x}_i$ and d is Yao-Wu signed distance.

Now, we define or verify some of the properties of correlation as follows:

\[a) \quad \rho(\bar{x}, \bar{y}) = 1.\]

\[b) \quad \rho(\bar{x}, \bar{y}) = \rho(\bar{y}, \bar{x}).\]

\[c) \quad |\rho(\bar{x}, \bar{y})| \leq 1.\]

Remark 1: $|\rho(\bar{x}, \bar{y})| \leq 1$.

Proof: We have
Proof: thus we verify \( \rho(\bar{x}, \bar{y}) \geq -1 \) and similarly

\[
\frac{1}{n} \sum_{i=1}^{n} \left( \frac{\tilde{x}_i}{\sigma_x} \right)^2 + \frac{1}{n} \sum_{i=1}^{n} \left( \frac{\tilde{y}_i}{\sigma_y} \right)^2 + 2 \frac{\text{cov}(\tilde{x}, \tilde{y})}{\sigma_x \sigma_y} = \frac{\sigma^2_x + \sigma^2_y + 2 \text{cov}(\tilde{x}, \tilde{y})}{\sigma_x \sigma_y}
\]

so that the traditional least-squares method could be applied to find a crisp regression line showing the general trend of the data. In the second stage, the error term of the fuzzy regression model, which represents the fuzziness of the data in a general sense, was determined to give the regression model the best explanatory power for the data.

Remark 2: \( \text{cov}(\tilde{x}, \tilde{y}) \leq \sqrt{\frac{\sigma^2_x \sigma^2_y}{n}} \).

Theorem 4: If \( \rho(\bar{x}, \bar{y}) \) and \( R^2 \) are correlation and coefficient of determination respectively, then

\[ |R| = |\rho(\bar{x}, \bar{y})| \]

Remark 3: The above procedures of regression reduce to ordinary procedures of regression, if the data set is crisp (non-fuzzy).

Example 1: Suppose that we have taken a fuzzy random sample of size \( n = 8 \) from a population and we have observed the following triangular fuzzy data:

Table 1. Fuzzy random sample of size \( n = 8 \) from a population.

<table>
<thead>
<tr>
<th>( x )</th>
<th>( \hat{x} )</th>
<th>( y )</th>
<th>( \hat{y} )</th>
<th>( \hat{y} = \hat{\beta}_1 \hat{x} + \hat{\beta}_0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(1.5, 2.0, 2.5)</td>
<td>(3.5, 4.0, 4.5)</td>
<td>(4.35, 4.61, 4.87)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>(3.0, 3.5, 4.0)</td>
<td>(3.0, 5.5, 6.0)</td>
<td>(5.13, 5.39, 5.65)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>(4.5, 5.5, 6.5)</td>
<td>(6.5, 7.5, 8.5)</td>
<td>(5.91, 6.43, 6.96)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>(6.5, 7.0, 7.5)</td>
<td>(6.0, 6.5, 7.0)</td>
<td>(6.95, 7.21, 7.47)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>(8.0, 8.5, 9.0)</td>
<td>(8.0, 8.5, 9.0)</td>
<td>(7.73, 7.99, 8.25)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>(9.5, 10.5, 11.5)</td>
<td>(7.0, 8.0, 9.0)</td>
<td>(8.51, 9.03, 9.55)</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>(10.5, 11.0, 11.5)</td>
<td>(10.0, 10.5, 11.0)</td>
<td>(9.03, 9.29, 9.55)</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>(12.0, 12.5, 13.0)</td>
<td>(9.0, 9.5, 10.0)</td>
<td>(9.81, 10.07, 10.33)</td>
<td></td>
</tr>
</tbody>
</table>

We obtain \( S_{xx} = 97.72, S_{yy} = 31.5, S_{xy} = 50.75, R = 0.915, \beta_1 = 0.52 \) and \( \beta_0 = 3.57 \).

4. Performance of fuzzy regression model

To evaluate the performance of fuzzy regression model, Kim and Bishu [16] studied a modification of fuzzy linear regression analysis based on a criterion of minimizing the difference of the fuzzy membership values between the observed and estimated fuzzy numbers. Kao and Chyu [14] proposed a two-stage approach to construct the fuzzy linear regression model. In the first stage, the fuzzy observations were defuzzified so that the traditional least-squares method could be applied to find a crisp regression line showing the general trend of the data. In the second stage, the error term of the fuzzy regression model, which represents the fuzziness of the data in a general sense, was determined to give the regression model the best explanatory power for the data. Yang and Lin [31] proposed two estimation methods along with a fuzzy least-squares approach. These proposed methods can be effectively used for the parameter estimation.

Kratsehmer [17] was a contribution to the estimation of parameter in fuzzy regression models with random fuzzy sets. Here models with crisp parameters and fuzzy observations of the variables was investigated. Two estimation methods along with a fuzzy least squares approach were proposed. These proposed methods could be effectively used for the parameter estimation. Comparisons are also made between them. Arabpour and Tata [5] used the absolute difference between the membership value of the observed and estimated fuzzy response, as a measure of error \( (a + b) \) area in Figure 1.

In these papers this measure is also adopted as the criterion for computing the performance of different methods; smaller values of this measure indicate a better fit.
By applying the fuzzy least squares method described in section 3 we estimated the parameters \( \beta_1 \) and \( \beta_0 \) for this model. In order to compare our method with that of Yang-Lin and Arabpour-Tata, we assume that \( \tilde{x}_i \) and \( \tilde{y}_i \), \( i = 1, 2, \ldots, n \) are triangular fuzzy numbers.

**Example 2:** consider Table 1. To show the effectiveness of our method with Yang-Lin and Arabpour-Tata, we apply this data.

The fuzzy regression model estimated by our method is as follows:

\[
\tilde{y}_i = 0.52\tilde{x}_i \oplus 3.57
\]

and the fuzzy regression model estimated by Yang and Lin is as follows:

\[
\tilde{y}_{Y-L} = (0.5251, 0.5293, 0.5335)\tilde{x} \oplus (3.2052, 3.4967, 3.7882)
\]

and the fuzzy regression model estimated by Arabpour and Tata is as follows:

\[
\tilde{y}_{A-T} = (0.518804, 0.519348, 0.523524)\tilde{x} \oplus (3.27580, 3.57243, 3.83865)
\]

Table 2 lists these errors. Again the sum of errors of our method (6.66069) is less than that of Yang-Lin method and that of Arabpour-Tata method. This implies that our fit is slightly better than that of Yang-Lin method and that of Arabpour-Tata method.

Table 2. The data and error in Example 1.

<table>
<thead>
<tr>
<th>Obs.</th>
<th>( \tilde{x}_i )</th>
<th>( \tilde{y}_i )</th>
<th>( \tilde{E}_{Y-L} )</th>
<th>( \tilde{E}_{A-T} )</th>
<th>( \tilde{E}_{Our} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(1.5, 2.0, 2.5)</td>
<td>(3.5, 4.0, 4.5)</td>
<td>0.82247</td>
<td>0.85856</td>
<td>0.73039</td>
</tr>
<tr>
<td>2</td>
<td>(3.0, 3.5, 4.0)</td>
<td>(5.0, 5.5, 6.0)</td>
<td>0.27822</td>
<td>0.21580</td>
<td>0.27419</td>
</tr>
<tr>
<td>3</td>
<td>(4.5, 5.5, 6.5)</td>
<td>(6.5, 7.5, 8.5)</td>
<td>1.53498</td>
<td>1.51221</td>
<td>1.38975</td>
</tr>
<tr>
<td>4</td>
<td>(6.5, 7.0, 7.5)</td>
<td>(8.0, 8.5, 9.0)</td>
<td>0.09641</td>
<td>0.94164</td>
<td>0.76971</td>
</tr>
<tr>
<td>5</td>
<td>(8.0, 8.5, 9.0)</td>
<td>(9.0, 9.5, 10.0)</td>
<td>0.77937</td>
<td>0.77724</td>
<td>0.67774</td>
</tr>
<tr>
<td>6</td>
<td>(9.5, 10.5, 11.5)</td>
<td>(7.0, 8.0, 9.0)</td>
<td>1.51546</td>
<td>1.47083</td>
<td>1.36204</td>
</tr>
<tr>
<td>7</td>
<td>(10.5, 11.0, 11.5)</td>
<td>(10.0, 10.5, 11.0)</td>
<td>1.10235</td>
<td>1.06797</td>
<td>0.76000</td>
</tr>
<tr>
<td>8</td>
<td>(12.0, 12.5, 13.0)</td>
<td>(9.0, 9.5, 10.0)</td>
<td>0.88847</td>
<td>0.83765</td>
<td>0.71250</td>
</tr>
<tr>
<td>Total error</td>
<td></td>
<td></td>
<td>7.89673</td>
<td>7.09090</td>
<td>6.00999</td>
</tr>
</tbody>
</table>

**5. Bootstrap confidence interval for \( \beta_1 \) and \( \beta_0 \)**

In this section we introduce bootstrap confidence interval based on \( t \)-bootstrap theory for \( \beta_1 \) and \( \beta_0 \).

We generate \( B \) bootstrap fuzzy random samples \( (\tilde{x}_1^b, \tilde{y}_1^b), (\tilde{x}_2^b, \tilde{y}_2^b), \ldots, (\tilde{x}_n^b, \tilde{y}_n^b) \) where each \( (\tilde{x}_i^b, \tilde{y}_i^b) \) is a fuzzy sample of size \( n \) drawn randomly and with replacement from \( (\tilde{x}_i, \tilde{y}_i); i = 1, 2, \ldots, n \) we compute

\[
t_b(\tilde{x}_i^b, \tilde{y}_i^b) = \frac{\hat{\beta}_0^b - \hat{\beta}_0}{\sqrt{\frac{S^b}{n} + \frac{d_{2b}(\tilde{x}_0, \tilde{y}_0)}{S_{xy}}} \oplus b = 1, 2, \ldots, B,
\]

where \( S^b \) is equal to \( \frac{SE^b}{\sqrt{n-2}} \).

The \( \gamma \) th percentile of \( t_b(\tilde{x}_i^b, \tilde{y}_i^b) \), \( i = 0,1 \), is estimated by the value \( \tilde{t}_b^{\gamma} \) that is in:

\[
\frac{\# \{ b: t_b(\tilde{x}_i^b, \tilde{y}_i^b) \leq \tilde{t}_b^{\gamma} \}}{B} = \gamma
\]

where \( b = 1, 2, \ldots, B \) and \( i = 0,1 \).

The bootstrap confidence interval by using fuzzy data is

\[
\prod_1^n = \left[ \beta_i - t_1(1 - \frac{1}{2})S_{\beta_0}, \beta_i + t_1(\frac{1}{2})S_{\beta_0} \right]
\]

where \( S_{\beta_1} \) and \( S_{\beta_0} \) are equal to \( S = \frac{SE}{\sqrt{n-2}} \) and \( S_{x_0} = \frac{d_{x_0}(\tilde{x}_0, \tilde{y}_0)}{S_{x_0}} \) respectively, furthermore \( S = \frac{SE}{\sqrt{n-2}} \).

**Remark 4:** If \( B \times \frac{1}{2} \) is not an integer, the following procedure can be used.

Assuming \( \frac{1}{2} \leq \gamma < 1 \), let \( k = [(B+1)\frac{1}{2}] \), the largest integer \( \leq (B+1)\frac{1}{2} \). then we define the empirical \( \gamma \) and \( 1 - \gamma \) quantizes by the \( k \)th largest and \( B + 1 - k \)th largest values of \( t_b(\tilde{x}_i^b, \tilde{y}_i^b) \), \( i = 0,1 \), respectively.

**Example 3:** Consider Table 1. Suppose we are interested in bootstrap confidence interval using those fuzzy data.

We use the package "MINITAB 13" to deal with complicated fuzzy data structure for this data.

For example, if \( B = 10000 \), the estimate of the %5 point is the 500th largest value of the \( t_b(\tilde{x}_i^b, \tilde{y}_i^b) \)s and the estimate of the %95 point is the 9500th largest value of the \( t_b(\tilde{x}_i^b, \tilde{y}_i^b) \)s.

The last line of Table 3 shows the percentiles of \( t_b(\tilde{x}_i^b, \tilde{y}_i^b) \) for \( \hat{\beta}_1 \) and \( \hat{\beta}_0 \) computed using 10000 bootstrap samples.

Table 3. Percentiles of the \( t \) distribution with 6 degree of freedom, the bootstrap distribution of \( t_b(\tilde{x}_i^b, \tilde{y}_i^b) \) and \( t_0(\tilde{x}_i^b, \tilde{y}_i^b) \).

<table>
<thead>
<tr>
<th>Percentile (( \gamma ))</th>
<th>( t_{\gamma,0} )</th>
<th>( t_{\gamma,1} )</th>
<th>( t_{\gamma,2} )</th>
<th>( t_{\gamma,3} )</th>
<th>( t_{\gamma,4} )</th>
<th>( t_{\gamma,5} )</th>
<th>( t_{\gamma,6} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>-3.143</td>
<td>-2.992</td>
<td>-3.147</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.025</td>
<td>-2.417</td>
<td>-2.398</td>
<td>-2.123</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td>-1.943</td>
<td>-1.939</td>
<td>-1.589</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>-1.440</td>
<td>-1.401</td>
<td>-1.164</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>1.440</td>
<td>1.226</td>
<td>1.201</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.95</td>
<td>1.943</td>
<td>1.845</td>
<td>1.689</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.975</td>
<td>2.447</td>
<td>2.736</td>
<td>2.140</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.99</td>
<td>3.143</td>
<td>3.982</td>
<td>2.718</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The bootstrap confidence intervals (\( \gamma = 0.05 \)) using fuzzy data are
\[
\Pi_i = [0.52 - 2.736 \times 0.0937, 0.52 + 2.398 \times 0.0937] = [0.26, 0.75]
\]
\[
\Pi_0 = [3.57 - 2.140 \times 0.7806, 3.57 + 2.123 \times 0.7806] = [1.90, 5.23]
\]
We show the distribution of \( t_i(\tilde{x}^b, \tilde{y}^b) \), \( i = 0, 1 \), computed using 10000 bootstrap samples as follows:

6. Bootstrap testing statistical hypotheses

In this section we give hypothesis testing on coefficients based on bootstrap theory.

We describe bootstrap method that is designed directly for hypothesis testing for fuzzy data based on Yao-Wu signed distance.

Computation of the bootstrap test statistics for testing
\( H_0: \beta_1 = \beta_{10} \)
1. Draw \( B \) samples of size \( n \) with replacement from \( (\tilde{x}_i, \tilde{y}_i; i = 1, 2, ..., n) \),
2. Evaluate \( t_1(.) \) on each sample, where
\[
t_1(\tilde{x}, \tilde{y}) = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{s_k}}, \quad t_1(\tilde{x}^b, \tilde{y}^b) = \frac{\hat{\beta}_{1b} - \beta_{1b}}{\sqrt{s_{kb}}}, b = 1, 2, ..., B
\]
\( 1, 2, ..., B \) also \( S \) and \( S^b \) are equal to \( \sqrt{\frac{SSE}{n-2}} \) and \( \sqrt{\frac{SSE^b}{n-2}} \), respectively.

Computation of the bootstrap test statistics for testing
\( H_0: \beta_0 = \beta_{00} \)
1. Draw \( B \) samples of size \( n \) with replacement from \( (\tilde{x}_i, \tilde{y}_i; i = 1, 2, ..., n) \),
2. Evaluate \( t_0(.) \) on each sample, where
\[
t_0(\tilde{x}, \tilde{y}) = \frac{\hat{\beta}_0 - \beta_0}{\sqrt{\frac{1}{n} \frac{d(\tilde{x}^b, \tilde{y}))}{s_k}}}, \quad t_0(\tilde{x}^b, \tilde{y}^b) = \frac{\hat{\beta}_{0b} - \beta_{0b}}{\sqrt{\frac{1}{n} \frac{d(\tilde{x}^b, \tilde{y}^b))}{s_{kb}}}}
\]
\( b = 1, 2, ..., B \), also \( S \) and \( S^b \) are equal to \( \sqrt{\frac{SSE}{n-2}} \) and \( \sqrt{\frac{SSE^b}{n-2}} \), respectively.

\( \cdot \) Decision rule
We choose a small probability \( \gamma \) (significant level), like 0.01, 0.05 or 0.1, and
\( \cdot \) We reject \( H_0 \) in error level \( \gamma \) if \( t_i(\tilde{x}, \tilde{y}) < t_{\gamma}^i \) or \( t_i(\tilde{x}, \tilde{y}) > t_{1-\gamma}^i \).
\( \cdot \) We accept \( H_0 \) in error level \( \gamma \) if \( t_{\gamma}^i \leq t_i(\tilde{x}, \tilde{y}) \leq t_{1-\gamma}^i \).

Example 4: Consider Table 1. Suppose that we are interested for testing
\( a) \ H_0: \beta_1 = \beta_{10} \) versus \( H_1: \beta_1 \neq \beta_{10} \),
\( b) \ H_0: \beta_0 = \beta_{00} \) versus \( H_1: \beta_0 \neq \beta_{00} \).

We tested those hypothesis testing that are shown in tables 4 and 5.

Table 4. \( \beta_{10}, t_1(\tilde{x}, \tilde{y}) \) and results.
<table>
<thead>
<tr>
<th>( \beta_{10} )</th>
<th>( t_1(\tilde{x}, \tilde{y}) )</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>16.2922</td>
<td>Reject ( H_0 )</td>
</tr>
<tr>
<td>0</td>
<td>5.5451</td>
<td>Reject ( H_0 )</td>
</tr>
<tr>
<td>0.2</td>
<td>3.4097</td>
<td>Reject ( H_0 )</td>
</tr>
<tr>
<td>0.4</td>
<td>1.2743</td>
<td>Accept ( H_0 )</td>
</tr>
<tr>
<td>0.5</td>
<td>0.2066</td>
<td>Accept ( H_0 )</td>
</tr>
<tr>
<td>0.6</td>
<td>-0.8611</td>
<td>Accept ( H_0 )</td>
</tr>
<tr>
<td>0.8</td>
<td>-2.9965</td>
<td>Reject ( H_0 )</td>
</tr>
<tr>
<td>1</td>
<td>-5.1319</td>
<td>Reject ( H_0 )</td>
</tr>
<tr>
<td>2</td>
<td>-15.8000</td>
<td>Reject ( H_0 )</td>
</tr>
</tbody>
</table>

Table 5. \( \beta_{00}, t_0(\tilde{x}, \tilde{y}) \) and results.
<table>
<thead>
<tr>
<th>( \beta_{00} )</th>
<th>( t_0(\tilde{x}, \tilde{y}) )</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.300</td>
<td>Reject ( H_0 )</td>
</tr>
<tr>
<td>2</td>
<td>2.015</td>
<td>Accept ( H_0 )</td>
</tr>
<tr>
<td>3</td>
<td>0.734</td>
<td>Accept ( H_0 )</td>
</tr>
<tr>
<td>3.5</td>
<td>0.003</td>
<td>Accept ( H_0 )</td>
</tr>
<tr>
<td>4</td>
<td>-0.548</td>
<td>Accept ( H_0 )</td>
</tr>
<tr>
<td>5</td>
<td>-1.830</td>
<td>Accept ( H_0 )</td>
</tr>
<tr>
<td>6</td>
<td>-3.110</td>
<td>Reject ( H_0 )</td>
</tr>
</tbody>
</table>
7. Conclusions

In this paper, we proposed a technique in order to get bootstrap statistical inference about the regression coefficients based on fuzzy data according to Yao-Wu signed distance. It sounds the introduced method is simpler and more accurate than that of Yang-Lin method and that of Arabpour-Tata method and furthermore this metric is very realistic because

- which implies very good statistical properties in connection with regression coefficients;
- it involves distances between extreme points;
- it is distance with convenient statistical features.

This method is especially attractive since the estimators are very similar to those of the classical linear regression model and when the fuzzy data becomes crisp, the estimators and the fuzzy regression line are identical to those in the classical case. Extension of the proposed method to linear models, such as regression models, design of experiment is a potential area for the future work. Furthermore for the testing hypothesis $H_0: \beta_i = \beta_0$, $i = 0,1$ as a fuzzy hypothesis (fuzzy coefficient), we can apply the methods that are in the papers Akbari and Rezaei [2], Akbari et al. [3], and Akbari and Rezaei [4]. These methods are a potential area for the future work.

Acknowledgment

The authors wish to express their thanks to the referees for valuable comments which improve the paper. Also, the authors would like to thanks the Editor-in-Chief Professor Wang for useful helps.

References


