Different distance measures for fuzzy linear regression with Monte Carlo methods

Duygu İçen · Marco E.G.V. Cattaneo

Abstract The aim of this study is to determine the best distance measure for estimating the fuzzy linear regression model parameters with Monte Carlo (MC) methods. It is pointed out that only one distance measure is used for fuzzy linear regression with MC methods within the literature. Therefore, three different definitions of distance measure between two fuzzy numbers are introduced. Estimation accuracies of existing and proposed distance measures are explored with the simulation study. Distance measures are compared to each other in terms of estimation accuracy; hence this study demonstrates that the best distance measures to estimate fuzzy linear regression model parameters with MC methods are the distance measures defined by Kaufmann and Gupta [13], Heilpern-2 [12] and Chen and Hsieh [4]. One the other hand, the worst distance measure is the distance measure used by Abdalla and Buckley [1][2]. These results would be useful to enrich the studies that have already focused on fuzzy linear regression models.

Keywords Fuzzy linear regression · Fuzzy distance measure · Monte Carlo

Duygu İçen
Hacettepe Universtiy
Department of Statistics, 06800
Ankara,Turkey
Tel.: +90-312-2977900
Fax: +90-312-2977913
E-mail: duyguicn@hacettepe.edu.tr

Marco E.G.V. Cattaneo
University of Hull
Department of Physics and Mathematics
Hull, HU6 7RX, UK

1 Introduction

In many cases in real life, most of the data are approximately known. Fuzzy set theory introduced by Zadeh [28] has found important application areas in different field of science as well as in regression analysis, because fuzzy set theory helps to define the vague relationship between variables or the observations that are reported as imprecise quantities for regression analysis.

Fuzzy regression analysis has been introduced by Tanaka *et al.* [25]. After the first attempt of using fuzzy sets in regression analysis, there have been several different approaches for the parameter estimation of fuzzy regression analysis. Bardossy [3] introduced how the problem of fuzzy regression can be formulated as a mathematical problem. Tanaka and Lee [24] considered interval regression analysis. Another approach to fuzzy regression is fuzzy least squares approach which is proposed by Diamond [7]. Peters [18], Luczynski and Matloka [17], Tanaka *et al.* [23], and Yen *et al.* [27] are some of the authors who focused on crisp input and fuzzy output regression models. D'Urso [10] carried out fuzzy linear regression analysis for fuzzy/crisp input and fuzzy/crisp output data. Moreover, Roh *et al.* [19] presented a new estimation approach based on Polynomial Neural Networks for fuzzy linear regression. Recently, a generalized maximum entropy estimation approach to fuzzy regression model is introduced by Ciavolinoa and Calcagni [6]. Both of the above approaches to fuzzy regression require complex mathematical operations and long calculations.

Another approach to fuzzy regression is introduced by Abdalla and Buckley [1][2] using MC methods. In this method, several random crisp or fuzzy vectors are generated as regression coefficient vector. Then using these random vectors, the dependent variable is cal-

culated. Two error measures are obtained by the difference of observed and estimated values of dependent variable to decide the best random vector for parameter estimation. One of these error measures depends on the error measure defined by Kim and Bishu [14]. In this error measure, distance of two fuzzy numbers has to be calculated. Therefore, distance measure between two fuzzy numbers plays a key role in estimating fuzzy linear regression model parameters with MC methods. However, current studies about MC methods in fuzzy linear regression within the literature do not account for different definitions of distance measure between fuzzy numbers.

The main contribution of this study to literature is to figure out the appropriate distance measure between two fuzzy numbers for the estimation of fuzzy linear regression model parameters with MC methods. Therefore, different definitions of distance measure between two fuzzy numbers introduced by Kauffman and Gupta [13], Heilpern [12], Chen and Hsieh [4] are used in the error measure used by Abdalla and Buckley [1][2]. A simulation study is conducted to evaluate estimation accuracy of new and existing distance measures. The best distance measure and the one that should not be used for the estimation of fuzzy linear regression parameters with MC methods are identified without using any mathematical programming or heavy fuzzy arithmetic operations.

The rest of the paper is organized as follows: Sect. 2 contains some basic definitions and notations about fuzzy sets and fuzzy numbers. Sect. 3 includes the properties of fuzzy linear regression and the parameter estimation procedure in fuzzy linear regression with MC methods. Sect. 4 defines the distance measures for fuzzy numbers that are taken into account for parameter estimation in this study. The simulation study that compares the performances of distance measures is conducted in Sect. 5. After the decision of the best and the worst distance measures in MC methods for fuzzy linear regression models, these distance measures are used for the real data sets in Sect. 6. The paper concludes with some discussions about applications and some possible future researches in Sect. 7.

2 Preliminaries

In this section, some important definitions of fuzzy concepts which are used throughout the paper are recalled.

Definition 2.1. $\mu_{\widetilde{A}}(x)$ is the membership function of an element x belonging to a fuzzy set $\widetilde{A}$, where $0 \leq \mu_{\widetilde{A}}(x) \leq 1$. If $\mu_{\widetilde{A}}(x) = 1$ then x belongs to $\widetilde{A}$, on the other hand, if $\mu_{\widetilde{A}}(x) = 0$ than x does not belong to $\widetilde{A}$. Crisp sets are considered as special cases of fuzzy sets, whose membership values are always 0 or 1 [9].

Definition 2.2. A general fuzzy number $\widetilde{A}$ is a normal convex fuzzy set of $\Re$ with a piecewise continuous membership function [9]. The trapezoidal fuzzy number is the simplest form of fuzzy number. It is defined by four parameters $\widetilde{A} = [a_1, a_2, a_3, a_4]$ the inner borders a_2, a_3, and the spreads $a_2 - a_1$, $a_4 - a_3$. The left and right sides of fuzzy numbers are $L(x) = \frac{(a_2 - x)}{(a_2 - a_1)}$ and $R(x) = \frac{(x - a_3)}{(a_4 - a_3)}$ respectively. When $a_2 = a_3$, triangular fuzzy number is obtained. The conditions $a_1 = a_2$ and $a_3 = a_4$ imply closed interval. In the case $a_1 = a_2 = a_3 = a_4$, crisp number is obtained [12].

Definition 2.3. The $\alpha - cut$ of a fuzzy number $\widetilde{A}$ is a non-fuzzy set defined as $\widetilde{A}(\alpha) = \{x \in \Re, \mu_{\widetilde{A}} \geq \alpha\}$. Hence $\widetilde{A}(\alpha) = [A^L(\alpha), A^U(\alpha)]$ where $A^L(\alpha) = inf\{x \in \Re, \mu_{\widetilde{A}} \geq \alpha\}$ and $A^U(\alpha) = sup\{x \in \Re, \mu_{\widetilde{A}} \geq \alpha\}$ [9].

Definition 2.4. $v_k = (v_{0k}, \ldots, v_{mk})$ is called random crisp vector where v_{ik} are all real numbers in intervals I_i, $i = 0, 1, \ldots, m$. Firstly, random crisp vectors $v_k = (x_{0k}, \ldots, x_{mk})$ with all x_{ik} in [0, 1], $k = 1, 2, \ldots, N$ are generated. Then all x_{ik} are put in the interval $I_i = [c_i, d_i]$ by $v_{ik} = c_i + (d_i - c_i)x_{ik}$, $i = 0, 1, \ldots, m$ [2].

Definition 2.5. $\widetilde{V}_k = (\widetilde{V}_{0k}, \ldots, \widetilde{V}_{mk})$ is called random fuzzy vector where $\widetilde{V}_{ik}$ are all triangular fuzzy numbers. First crisp vectors $v_k = (v_{1k}, \ldots, v_{3m+3,k})$ with all the x_{ik} in [0, 1], $k = 1, \ldots, N$ are generated. Then the first three numbers in v_k are chosen and ordered from smallest to largest. Let us assume that $x_{3k} < x_{1k} < x_{2k}$, then the first triangular fuzzy numbers is $\widetilde{V}_{0k} = (x_{3k}/x_{1k}/x_{2k})$. The other $\widetilde{V}_{ik}$ are generated with the next three numbers in v_k. In order to obtain $\widetilde{V}_{ik}$ in certain intervals, it is possible to put all x_{ik} into $I_i = [a_i, b_i]$ by computing $a_i + (b_i - a_i)x_{ik}$, $i = 1, 2, \ldots, m$ [1].

3 Fuzzy linear regression with Monte Carlo methods

One of the most important objectives of a regression model is to estimate the value of the dependent variable associated with independent variable(s) as close to the observed data as possible. Choi and Buckley [5] classified fuzzy regression models in three categories:

- Input and output data are both crisp.
- Input data is crisp and output data is fuzzy.
- Input and output data are both fuzzy.

In this paper only the second and the third categories are examined because the first category is an ordinary regression model.

The fuzzy linear regression model where input data is crisp and output data is fuzzy (Case-II) is expressed as follows:

$$\widetilde{Y}_l = \widetilde{A}_0 + \widetilde{A}_1 x_{1l} + \widetilde{A}_2 x_{2l} + \ldots + \widetilde{A}_m x_{ml}. \tag{1}$$

Here, $\widetilde{A}_i$ is a triangular fuzzy number with membership function $\mu_{\widetilde{A}_i}$ and x_{il} is a real number, $i = 0, 1, \ldots, m$ and $l = 1, 2, \ldots, n$. The membership function of $\widetilde{Y}_l$ is $\mu_{\widetilde{Y}_l}$.

The third category (Case-III) consists of fuzzy input variables and fuzzy outputs. It is given as follows:

$$\widetilde{Y}_l = a_0 + a_1 \widetilde{X}_{1l} + a_2 \widetilde{X}_{2l} + \ldots + a_m \widetilde{X}_{ml}. \tag{2}$$

In this model, $\widetilde{X}_{il}$ and $\widetilde{Y}_l$ for $i = 1, 2, \ldots, m$ and $l = 1, 2, \ldots n$ are triangular shaped fuzzy numbers and a_i is a crisp number.

In the parameter estimation process for fuzzy linear regression with MC methods, possible solutions are generated randomly and inferior solutions are discarded. This continues for N times, and N is usually 10000 or 100000. Predicted values are determined by using randomly generated $\widetilde{V}_k = (\widetilde{V}_{0k}, \widetilde{V}_{1k}, \ldots, \widetilde{V}_{mk})$ for Case-II with Eq. 3 and $v_k = (v_{0k}, v_{1k}, \ldots, v_{mk})$ for Case-III with Eq. 4 where $k = 1, 2, \ldots N$ [1][2].

$$\widetilde{Y}_{lk}^* = \widetilde{V}_{0k} + \widetilde{V}_{1k} x_{1l} + \widetilde{V}_{2k} x_{2l} + \ldots + \widetilde{V}_{mk} x_{ml} \tag{3}$$

$$\widetilde{Y}_{lk}^* = v_{0k} + v_{1k} \widetilde{X}_{1l} + v_{2k} \widetilde{X}_{2l} + \ldots + v_{mk} \widetilde{X}_{ml} \tag{4}$$

In Case-II and Case-III, given values $(\widetilde{Y}_l)$ and the predicted values $(\widetilde{Y}_{lk}^*)$ are triangular fuzzy numbers and have membership functions. Hence, the difference of these membership values should be used to measure the degree of the fitting of the estimated fuzzy linear regression model to the given data.

The sum of the differences between the observed and predicted fuzzy numbers is calculated as

$$D = \int \left| \mu_{\widetilde{Y}_l}(x) - \mu_{\widetilde{Y}_{lk}^*(x)} \right| dx, \tag{5}$$

where D represents the difference of membership values between two membership functions [14].

Since predicted fuzzy number is expected to have membership function close to the observed fuzzy membership function, the error of the fitting of the membership functions can be defined by the ratio of the difference of membership values to the observed membership values. This is given with Eq. 6 [14].

$$E = \frac{\int_{S_{\widetilde{Y}_l} \cup S_{\widetilde{Y}_{lk}^*}} \left| \mu_{\widetilde{Y}_l}(x) - \mu_{\widetilde{Y}_{lk}^*}(x) \right| dx}{\int_{S_{\widetilde{Y}_l}} \mu_{\widetilde{Y}_l}(x) dx} \tag{6}$$

If the difference of membership values between two membership functions D becomes zero, the error of fit E, becomes zero. This error measure is introduced by Kim and Bishu [14].

In the light of the error measure defined by Kim and Bishu [14], two different error measures are defined by Abdalla and Buckley [1][2] for fuzzy linear regression with MC methods. These measures are used to assess the accuracy of candidate vectors $\widetilde{V}_k$ and v_k. They use two different error measures based on the given values $\widetilde{Y}_l$ and predicted values $\widetilde{Y}_{lk}^*$. First one is

$$E_{1k}(E_1) = \sum_{l=1}^{n} \frac{\left[\left| \int_{-\infty}^{\infty} |\widetilde{Y}_l(x) - \widetilde{Y}_{lk}^*(x)| \, dx \right| \right]}{\left[\int_{-\infty}^{\infty} \widetilde{Y}_l(x) dx \right]} \tag{7}$$

where the integrals are really only over intervals containing the support of the fuzzy numbers. Then second error measure is

$$E_{2k}(E_2) = \sum_{l=1}^{n} \left[|y_{l1} - y_{lk1}| + |y_{l2} - y_{lk2}| + |y_{l3} - y_{lk3}| \right] \tag{8}$$

where $\widetilde{Y}_l = (y_{l1}/y_{l2}/y_{l3})$ and $\widetilde{Y}_{lk}^* = (y_{lk1}/y_{lk2}/y_{lk3})$ are all triangular fuzzy numbers.

So $\widetilde{V}_k$ and v_k are obtained for the regression models given with Eq. 1 and Eq. 2, respectively. Then E_{1k} and E_{2k} are calculated for $k = 1, 2, \ldots, N$. The best solutions are $V_k \in \{V_1, \ldots, V_N\}$ for Case-II and $v_k \in \{v_1, \ldots, v_N\}$ for Case-III that minimizes E_{1k} and E_{2k}. Hence, there are two best solutions one with respect to E_{1k} and another for E_{2k}.

In this study, we consider the first error measure (E_1). Because the difference of two membership functions $(\widetilde{Y}_l$ and $\widetilde{Y}_{lk}^*)$ is calculated in the first error measure.

4 Distance measures for fuzzy numbers

The methods of measuring the distance between fuzzy numbers have become important due to the significant applications in diverse fields like data mining [20], pattern recognition [26], regression analysis [1][2] and so on. However, only one definition of the distance measure between two fuzzy numbers is used in the process

of estimating fuzzy linear regression model parameters with MC methods in the literature. This section explains the new distance measures that are proposed for fuzzy linear regression with MC methods.

In 1991, Kaufmann and Gupta [13] considered a distance measure between two fuzzy numbers. It is combined by the interval of $\alpha-$cuts of fuzzy numbers and is given with Eq. 9.

$$d(\widetilde{A}, \widetilde{B}) = \int_0^1 \Big(|A^L(\alpha) - B^L(\alpha)| + |A^U(\alpha) - B^U(\alpha)| \Big) d\alpha \quad (9)$$

In this equation, $[A^L(\alpha), A^U(\alpha)]$ and $[B^L(\alpha), B^U(\alpha)]$ are the closed interval of $\alpha-$cuts of fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$.

In 1997, Heilpern [12] proposed three definitions of the distance between two fuzzy numbers.

1. The mean distance method is generated by the expected values of fuzzy numbers (Heilpern-1). The lower and the upper expected value of a fuzzy number is given with Eq. 10 and Eq. 11 respectively.

$$E_*(\widetilde{A}) = a_2 - (a_2 - a_1) \int_0^\infty L(x) dx \quad (10)$$

$$E^*(\widetilde{A}) = a_3 + (a_4 - a_3) \int_0^\infty R(x) dx \quad (11)$$

By using the values above, the expected value of fuzzy number is calculated as

$$EV(\widetilde{A}) = \frac{1}{2} \left[E_*(\widetilde{A}) + E^*(\widetilde{A}) \right]. \quad (12)$$

Thus, Heilpern [12] defined the difference of two fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$ with respect to the expected values of these fuzzy numbers with Eq. 13.

$$\sigma(\widetilde{A}, \widetilde{B}) = |EV(\widetilde{A}) - EV(\widetilde{B})| \quad (13)$$

2. The second distance method is generated by combining Minkowski distance and the $\alpha-$levels of the closed intervals of fuzzy numbers (Heilpern-2). Let $\widetilde{A}$ and $\widetilde{B}$ be two fuzzy numbers and the distance between these two fuzzy numbers is given with Eq. 14.

$$d_p(\widetilde{A}, \widetilde{B}) = \int_0^1 d_p\big(\widetilde{A}(\alpha), \widetilde{B}(\alpha)\big) d\alpha \quad (14)$$

Here $\widetilde{A}(\alpha)=[A^L(\alpha), A^U(\alpha)]$ and $\widetilde{B}(\alpha)=[B^L(\alpha), B^U(\alpha)]$ are closed interval of $\alpha-$cut of a fuzzy number $\widetilde{A}$ and $\widetilde{B}$ and $d_p\big(\widetilde{A}(\alpha), \widetilde{B}(\alpha)\big)$ is given as follows:

$$d_p(\widetilde{A}(\alpha), \widetilde{B}(\alpha)) =$$
$$= \begin{cases} (0.5(|A^L(\alpha) - B^L(\alpha)|^p + |A^U(\alpha) - B^U(\alpha)|^p))^{(1/p)}, \\ \quad 1 \le p < \infty \\ \max\{|A^L(\alpha) - B^L(\alpha)|, |A^U(\alpha) - B^U(\alpha)|\}, \\ \quad p = \infty \end{cases}$$
$$\quad (15)$$

This function is generated by Minkowski distance. In many situations, the distance is calculated with $p = 1$ [12].

3. The third distance method is based on the geometrical operation of fuzzy numbers (Heilpern-3). Let $\widetilde{A} = (a_1, a_2, a_3, a_4)$ and $\widetilde{B} = (b_1, b_2, b_3, b_4)$ be two fuzzy numbers, then the geometrical distance is given as

$$\delta_p(\widetilde{A}, \widetilde{B}) = \begin{cases} 0.25(\sum_{i=1}^4 |a_i - b_i|^p)^{(1/p)} & 1 \le p < \infty \\ \max_i(|a_i - b_i|) & p = \infty \end{cases}$$
$$\quad (16)$$

This distance is called geometric distance between two fuzzy numbers. In many situations, the distance is calculated with $p = 1$. This distance method is only used in the case of trapezoidal fuzzy numbers [12]. Hence, this distance is excluded for this study.

In 2000, Chen and Hsieh [4] have defined the distance of two generalized fuzzy numbers by graded mean integration representation (GMIR). GMIR of generalized fuzzy number $\widetilde{A}$ is based on the integral value of graded mean $\alpha-$level. It is given with Eq. 17 where $0 < \alpha < w$ and $0 < w \le 1$.

$$P(\widetilde{A}) = \frac{\int_0^w \alpha \left(\frac{L^{-1}(\alpha) + R^{-1}(\alpha)}{2} \right) d\alpha}{\int_0^w \alpha dw} \quad (17)$$

Let $\widetilde{A} = (a_1, a_2, a_3, a_4)$ be a fuzzy number, Chen and Hsieh [4] have already formulated the GMIR of this fuzzy number as follows:

$$P(\widetilde{A}) = \frac{a_1 + 2a_2 + 2a_3 + a_4}{6}. \quad (18)$$

Generalized triangular fuzzy number is the special case of trapezoidal fuzzy number when $a_2 = a_3$. Hence GMIR of triangular fuzzy number is

$$P(\widetilde{A}) = \frac{a_1 + 4a_2 + a_4}{6}. \quad (19)$$

Then the distance of two fuzzy number based on GMIR is defined as

$$|P(\widetilde{A}) - P(\widetilde{B})|. \tag{20}$$

All of the above methods use crisp real number to calculate the distance of two fuzzy numbers [11].

5 Simulation

We conduct a simulation study to compare the estimation performances of the distance measures mentioned in the previous section. Simulation scenarios depend on the intervals, which are used to generate candidate solutions in the MC method, given with Table 1. Both sides of the real line and interval widths are considered in the determination of these intervals.

In Table 1, the interval I_0 excludes negative numbers also it is a short interval. The interval I_1 is short and includes negative numbers. The interval I_2 is long and excludes negative numbers. The interval I_3 is long and includes both negative and positive numbers. The interval I_4 is short and exclude positive numbers. The interval I_5 is long and include only negative numbers.

We use intervals I_0 to I_5 to estimate regression parameters $\widetilde{A}_0$, $\widetilde{A}_1$ and $\widetilde{A}_2$ for Case-II. Then we use the same intervals to estimate regression parameters a_0, a_1 and a_2 for Case-III.

We obtain parameter estimates of fuzzy linear regression models by using the MC method for both Case-II and Case-III for each interval given in Table 1 for 10^3 times.

In both Case-II and Case-III dependent variable is triangular fuzzy number. Hence it is possible to measure the deviation between observed and estimated values by using triangular fuzzy numbers' left, center and right values. On this point, we use Mean Absolute Error (MAE) given with Eq. 21 to measure the deviation between observed and estimated values.

$$MAE = \frac{1}{3} \sum_{j=1}^{3} |y_{lj} - y_{lkj}| \tag{21}$$

We apply each scenario given in Table 1 for the simulation study, generate 10^4 candidate solutions and apply the described MC method in order to obtain estimates of parameters of the fuzzy linear regression model by using distance measures given in Sec.4. Afterwards, based on the minimum error given with Eq. 7 the differences between the estimated values and observed values are calculated using MAE with Eq. 21.

5.1 Simulation study for Case-II

In each of 10^3 replication, we randomly generate data sets of size 10 from Normal(15,9) distribution for the the first variable (x_1) and from Normal(-3,2) distribution for the second (x_2) independent variable. Fuzzy numbers for the value of each parameter are randomly generated sequentially from Normal distribution with mean 1, standard deviation 0.04 for $\widetilde{A}_0$; from Normal distribution with mean 4, standard deviation 0.9 for $\widetilde{A}_1$; from Normal distribution with mean -5, standard deviation 0.1 for $\widetilde{A}_2$. The corresponding values of dependent variable are obtained over the model given in Eq. 1.

We apply each scenario given in Table 1 for the simulation study. We generate 10^4 vectors and applied MC method in order to obtain estimates of parameters of the fuzzy linear regression model. Afterwards, we consider different definitions of distance measures for fuzzy numbers given in Sec. 4 for the error measure given with Eq. 7. Then, based on the minimum error, the differences between the estimated values and the observed values are calculated using MAE.

Table 2 gives simulation results of Case-II for the error measure MAE. The results are presented as follows (minimum values of MAE are written as bold in Table 2):

- Minimum values of MAE is reached with considering the distance measures described by Kaufmann and Gupta [13], Heilpern-1 [12], Heilpern-2 [12] and Chen and Hsieh [4] for interval I_0 and I_5. On the other hand maximum value of MAE is calculated when the distance measure described by Abdalla and Buckley [1] is handled for the same intervals.
- According to interval I_1, I_2 and I_4, minimum values of MAE is reached with considering the distance measures described by Kaufmann and Gupta [13], Heilpern-1 [12], Heilpern-2 [12]. However, maximum value of MAE is calculated when the distance measure described by Abdalla and Buckley [1] is taken into account.
- Minimum values of MAE is reached with considering the distance measures described by Kaufmann and Gupta [13] and Heilpern-2 [12] for interval I_3. Nevertheless, maximum MAE value is calculated when the distance measure described by Abdalla and Buckley [1] is used for the same interval.

As a result, it is proven to be the best distance measure to estimate regression model parameters with MC method for Case-II is the distance measure described by Kaufmann and Gupta [13] and Heilpern-2 [12]. One the other hand, distance measure used by Abdalla and

Buckley [1] is not appropriate to estimate fuzzy linear regression model parameters with MC method for Case-II.

5.2 Simulation study for Case-III

We randomly generate 10 triangular fuzzy numbers that have normal distribution with mean 0 and with standard deviation 2 for the first independent variable ($\widetilde{X}_1$), with mean -3 and with standard deviation 0.01 for the second independent variable ($\widetilde{X}_2$). Also three crisp numbers for the value of each parameters are randomly generated from normal distribution with mean -1, standard deviation 0.02 for a_0, from normal distribution with mean 2, standard deviation 0.01 for a_1 and from normal distribution with mean 3, standard deviation 0.04, for a_2. The corresponding values for dependent variable are obtained over the model given in Eq. 4.

We apply each scenario given in Table 1 for the simulation study, generate 10^4 vectors and apply the described MC method in order to obtain estimates of parameters of the fuzzy linear regression model. Afterwards based on the minimum errors, the differences between estimated values and the actual values are calculated using MSE.

Table 3 gives simulation results of Case-III for the error measure MAE. The results are presented as follows:

- Minimum MAE value is calculated with considering the distance measures Kaufmann and Gupta [13], Heilpern-1 [12], Heilpern-2 [12] and Chen and Hsieh [4] for interval I_0, I_2, I_3 and I_5. On the other hand maximum value of MAE is calculated when the distance measure described by Abdalla and Buckley [2] is handled for the same intervals.
- According to interval I_1 and I_4, minimum values of MAE is reached with considering the distance measure described by Chen and Hsieh [4]. However, maximum value of MAE is calculated when the distance measure described by Abdalla and Buckley [2] is taken into account.

As a result, it is proven to be the best distance measure to estimate regression model parameters with MC method for Case-III is the distance measure described by Chen and Hsieh [4]. One the other hand, distance measure used by Abdalla and Buckley [2] is not appropriate to estimate fuzzy linear regression model parameters with MC method for Case-III.

6 Application

In this section, there are two different applications. First application is for the second fuzzy regression model category (Case-II) and the other one is for the third fuzzy regression model category (Case-III).

We consider different distance measures for fuzzy numbers given in Sec. 4 in the error measure E_1 with Eq. 7 for fuzzy linear regression models with MC approach.

6.1 Application for Case-II

The data for this application is taken from Kim and Bishu [14] and is given with Table 4. There are eight items and three independent variables in the data set. This data set is studied by Tanaka [22], Abdalla and Buckley [1], Savic and Pedryzc [21], Choi and Buckley [5] for comparing the proposed methods with their works. We use this data to apply different distance measures in the error measure E_1 and to compare our new results with Abdalla and Buckley [1].

Before the application we have to decide the intervals for I_i, $i = 0, 1, 2, 3$ to obtain the model coefficients as explained in Definition 2.5. We use same intervals in order to compare the results we have with the results from Abdalla and Buckley [1] in the literature. Four separate intervals ($MCI, MCII, MCIII, MCIV$) that they studied are given with Table 5. For more information about why using these intervals see Abdalla and Buckley [1].

We apply different definitions of distance measure between two fuzzy numbers for estimating fuzzy linear regression model parameters in Case-II by using the data set given in with Table 4. For this purpose $N = 10^5$ random vectors ($\widetilde{V}_k = (\widetilde{V}_{0k}, \widetilde{V}_{1k}, \widetilde{V}_{2k}, \widetilde{V}_{3k})$) which defines model parameters ($\widetilde{A}_0, \widetilde{A}_1, \widetilde{A}_2, \widetilde{A}_3$)) are generated. Values of these parameters which gives minimum E_1 value is recorded according to each definition of distance measure used in this error measure. Results for the $\widetilde{A}_i$, $i = 0, 1, 2, 3$ are shown in Table 6 according to each interval.

Optimal solutions for $\widetilde{A}_i$ ($i = 0, 1, 2, 3$) are stated in Abdalla and Buckley [1] as $\widetilde{A}_0 = (-0.710/-0.539/-0.524)$, $\widetilde{A}_1 = (-0.610/-0.473/-0.472)$, $\widetilde{A}_2 = (-1.090/-1.089/-1.088)$ and $\widetilde{A}_3 = (0.459/0.487/0.680)$.

The value of Error measure E_1 is computed by using different distance measure definitions given in Sect. 4. The results are shown in Table 7.

It is seen from Table 7 that the smallest Error value for Interval MCI is obtained with Heilpern-1[12] distance measure. In addition Chen and Hsieh [4] gives minimum error value for the interval $MCII$. Besides,

Abdalla and Buckley [1] calculate minimum error value for $MCIII$ and $MCIV$. Biggest error values are calculated when Kaufman and Gupta [13] and Heilpern-2 [12] distance measures are taken into account in the error measure E_1.

6.2 Application for Case-III

The data for this application is taken from Choi and Buckley [5] and is shown in Table 8. There are ten items and two independent variables in the data set. This data set is studied by Choi and Buckley [5], Dimond and Korner [8], Abdalla and Buckley [2] for comparing the proposed methods with their works. We use this data to apply different distance measures in the error measure E_1 and compare our new results with Abdalla and Buckley [2].

Before the application we have to decide the intervals for I_i, $i = 0, 1, 2$ to obtain the model coefficients as explained in Definition 2.4. We use same intervals in order to compare the results we have with the results from Abdalla and Buckley [2] in the literature. Four separate intervals ($MCI, MCII, MCIII, MCIV$) that they studied are given with Table 9. For more information about why using these intervals, see Abdalla and Buckley [2].

We apply different definitions of distance measure between two fuzzy numbers for estimating fuzzy linear regression model parameters for Case-III by using the data set given in with Table 8. For this purpose $N = 10^5$ random vectors ($v_k = (v_{0k}, v_{1k}, v_{2k})$) which defines model parameters (a_0, a_1, a_2) are generated. Values of these parameters which gives minimum E_1 value is recorded according to each definition of distance measure used in this error measure. Results for the a_i, $i = 0, 1, 2$ are shown in Table 10 according to each interval.

Optimal solutions for a_i ($i = 0, 1, 2$) are stated in Abdalla and Buckley [2] as $a_0 = 4.19$, $a_1 = 4.97$ and $a_2 = 3.11$. According to these results, Heilpern-1 [12] gives the closest parameter estimations for interval MCI, Kaufmann and Gupta [13] and Heilpern-2 [12] gives the closest parameter estimations for interval $MCII$ and $MCIII$, Abdalla and Buckley [2] gives the closest parameter estimations for interval $MCIV$.

Error value E_1 for each distance measure is given according to the defined intervals ($MCI, MCII, MCIII, MCIV$) in Table 11 for Case-III.

It is seen from Table 11 that smallest error value of MSE for interval MCI, $MCII$ and $MCIV$ is obtained with Abdalla and Buckley [2]. Besides, distance measure defined by [12] (Heilpern-1) gives minimum MSE value for interval $MCIII$.

7 Conclusion

In this study we use different definitions of distance measure between two fuzzy numbers for estimating the parameters of fuzzy linear regression models with Monte Carlo method. The reasons of this research are summarized below:

– Firstly, MC methods in fuzzy regression is a very new and potential area that is easy to calculate model parameters without any long and complex mathematical equations, also no need for any regression assumptions. Model parameters are obtained according to the random vector which makes the error measure (E_1) minimum for fuzzy regression model. In this error measure, distance between two fuzzy numbers ($\widetilde{Y}_l$ and $\widetilde{Y}_{lk}^*$) has to be calculated.

– Secondly, there are several different definitions of distance measure between two fuzzy numbers in the literature. However only one definition has been used in fuzzy linear regression with MC methods until now. Hence, we investigate using different definitions of distance measure between fuzzy numbers in estimating the parameters of fuzzy linear regression with MC methods.

– Thirdly, distance measure between fuzzy numbers have gained importance due to the widespread applications in diverse fields like decision making, machine learning and market prediction. We explore the effect of this importance on estimating fuzzy linear regression model parameters.

A simulation study is conducted to compare the estimation performances of the considered distance measures. Considering the overall statement of the simulation results, we reached minimum MAE values with taking into account the distance measure described by Kaufmann and Gupta [13] and Heilpern-2 [12] for Case-II. Besides, the distance measure described by Chen and Hsieh [4] gives minimum MAE values for Case-III. It is demonstrated that the distance measure used by Abdalla and Buckley [1][2] is not convenient to estimate fuzzy linear regression model parameters with MC methods. Since all maximum values of MAE are calculated with the distance measure that is considered by Abdalla and Buckley [1][2] for both Case-II and Case-III.

Obtained results can and will be used to enrich the studies that have already focused on fuzzy linear regression models. For example, extreme learning machine [16] can be enhanced by the help of fuzzy linear regression with MC methods according to the best distance measures determined in this study. Fuzzy regression model based on least absolute deviation studied by Li et al. [15] can be improved using MC methods by the

help of the distance measures described by Kaufmann and Gupta [13] or Heilpern-2 [12].

Using fuzzy distance measures in fuzzy linear regression models with Monte Carlo methods is a potential area for the future works. Since, all the distance measures discussed in this paper use the real number to calculate the distance between two fuzzy numbers. Moreover, other future research will be concern with investigating different definitions of distance measure between fuzzy numbers in different types of regression models, such as nonparametric regression, exponential regression or considering different types of fuzzy numbers, such as trapezoidal, Gaussian in these regression models.

Compliance with Ethical Standards:

Conflict of Interest: *Author Duygu İçen declares that she has no conflict of interest. Author Marco Cattaneo declares that he has no conflict of interest.*

Ethical approval: *This article does not contain any studies with human participants or animals performed by any of the authors.*

References

1. Abdalla A, Buckley JJ (2007) Monte Carlo methods in fuzzy linear regression. Soft Comput 11:991-996
2. Abdalla A, Buckley JJ (2008) Monte Carlo methods in fuzzy linear regression II. Soft Comput 12:463-468
3. Bardossy, A (1990) Note on fuzzy regression. Fuzzy Sets Syst 37:65-75
4. Chen SH, Hsieh CH (2000) Representation, Ranking, Distance, and Similarity of L-R type fuzzy number and Application. Australia Journal of Intelligent Information Processing Systems, 6(4):217-229
5. Choi HS, Buckley JJ (2007) Fuzzy regression using least absolute deviation estimators. Soft Comput 12:257-263
6. Ciavolinoa E, Calcagni A (2016) A Generalized Maximum Entropy (GME) estimation approachto fuzzy regression model. Appl Soft Comput 38:51-63
7. Diamond P (1987) Least squares fitting of several fuzzy variables. In Proc of Second IFSA Congress. IFSA, Tokyo, p. 20-25
8. Diamond P, Korner, R (1997) Extended Fuzzy linear models and least squares estimate. Comput Math Appl 33:15–32
9. Dubois D, Prade, H (1978) Operations on fuzzy numbers. International Journal of Systems Science, 9(6):613–626
10. D'Urso P (2003) Linear regression analysis for fuzzy/crisp input and fuzzy/crisp output data. Comput Stat Data Anal 42:47-72.
11. Hajjari T (2010) Measuring Distance of Fuzzy Numbers by Trapezoidal Fuzzy Numbers, AIP Conference Proceedings 1309, 346
12. Heilpern S (1997) Representation and application of fuzzy numbers. Fuzzy sets and Systems, 91(2):259–268
13. Kaufmann A, Gupta MM (1991) Introduction to fuzzy arithmetic theory and applications, Van Nostrand Reinhold
14. Kim B, Bishu RR (1998) Evaluation of fuzzy linear regression models by comparing membership functions. Fuzzy sets and systems, 100:343–352
15. Li J, Zeng W, Xie J, Yin Q (2016) A new fuzzy regression model based on least absolute deviation. Engineering Applications of Artificial Intelligence, 52:54–64
16. Liu H, Wang J, He Y, Ashfaq RAR (2016) Extreme learning machine with fuzzy input and fuzzy output for fuzzy regression. Neural Comput & Applic, DOI:10.1007/s00521-016-2232-9
17. Luczynski W, Matloka M (1995) Fuzzy regression models and their applications. J Fuzzy Math 3:583–589
18. Peters G (1994) Fuzzy linear regression with fuzzy intervals. Fuzzy Sets Syst 63(1):45–55
19. Roh SB, Ahn TC, Pedrycz W (2012) Fuzzy linear regression based on Polynomial Neural Networks. Expert Systems with Applications 39:8909-8928
20. Sadi-Nezhad S, Khalili Damghani K (2010) Application of a fuzzy TOPSIS method base on modified preference ratio and fuzzy distance measurement in assessment of traffic police centers performance. Applied Soft Computing 10:1028-1039
21. Savic DA, Pedryzc W (1991) Evaluation of fuzzy linear regression models. Fuzzy Sets Syst 39:51–63
22. Tanaka H (1987) Fuzzy Data analysis by possibilistic linear regression models. Fuzzy Sets Syst 24:363–375
23. Tanaka H, Ishibuchi H, Yoshikawa S (1995) Exponential possibility regression analysis. Fuzzy Sets Syst 69:305–318
24. Tanaka H, Lee H (1999) Exponential possibility regression analysis by identification method of possibilistic coefficients. Fuzzy Sets Syst 106:155–165
25. Tanaka H, Uejima S, Asai K. (1982) Linear regression analysis with fuzzy model. IEEE Trans Systems Man Cybernet 12:903–907
26. Wang W, Xin X (2005) Distance measure between intuitionistic fuzzy sets. Pattern Recognition Letters, 26(13):2063-2069
27. Yen KK, Ghoshray S, Roig G (1999) A linear regression model using triangular fuzzy number coefficients. Fuzzy Sets Syst 106:167–177
28. Zadeh LA (1965) Fuzzy Sets. Information and control, 8:338–353

Table 1 Intervals for Case-II and Case-III for the simulation study

Parameters	I_0	I_1	I_2	I_3	I_4	I_5
$\widetilde{A}_o$ (or a_o)	[0,1]	[-1,1.2]	[1,10]	[-15,24]	[-3,-2]	[-22,-4.2]
$\widetilde{A}_1$ (or a_1)	[0,0.5]	[-0.5,1.4]	[3,24]	[-20,40]	[-1.76,0]	[-28,-1.5]
$\widetilde{A}_2$ (or a_2)	[1,1.3]	[-0.9,1.8]	[8,46]	[-37,60]	[-4.8,3.7]	[-45,-3]

Table 2 Simulation results for Case-II

Distance Measures	I_0	I_1	I_2	I_3	I_4	I_5
Abdalla and Buckley [1]	0.7453	0.5458	1.1532	3.6649	1.0706	1.9653
Kaufmann and Gupta [13]	**0.6684**	**0.4563**	**0.2617**	**0.5516**	**0.6903**	**1.1401**
Heilpern-1 [12]	**0.6684**	**0.4563**	**0.2617**	1.7381	**0.6903**	**1.1401**
Heilpern-2 [12]	**0.6684**	**0.4563**	**0.2617**	**0.5516**	**0.6903**	**1.1401**
Chen and Hsieh [4]	**0.6684**	0.4738	0.3059	2.2460	0.7170	**1.1401**

Table 3 Simulation results for Case-III

Distance Measures	I_0	I_1	I_2	I_3	I_4	I_5
Abdalla and Buckley [2]	0.0534	0.0427	0.1208	0.0498	0.0406	0.1633
Kaufmann and Gupta [13]	**0.0527**	0.0394	**0.1092**	**0.0316**	0.0343	**0.1533**
Heilpern-1 [12]	**0.0527**	0.0403	**0.1092**	**0.0316**	0.0312	**0.1533**
Heilpern-2 [12]	**0.0527**	0.0394	**0.1092**	**0.0316**	0.0343	**0.1533**
Chen and Hsieh [4]	**0.0527**	**0.0388**	**0.1092**	**0.0316**	0.0311	**0.1533**

Table 4 Data for the application (Case-II)

Fuzzy output	x_{1l}	x_{2l}	x_{3l}
(2.27/5.83/9.39)	2.00	0.00	15.25
(0.33/0.85/1.37)	0.00	5.00	14.13
(5.43/13.93/22.43)	1.13	1.50	14.13
(1.56/4.00/6.44)	2.00	1.25	13.63
(0.64/1.65/2.66)	2.19	3.75	14.75
(0.62/1.58/2.54)	0.25	3.50	13.75
(3.19/8.18/13.17)	0.75	5.25	15.25
(0.72/1.85/2.98)	4.25	2.00	13.50

Table 5 Intervals for I_i, $i = 0, 1, 2, 3$ for Case-II

Interval	MCI	$MCII$	$MCIII$	$MCIV$
I_0	[-1,0]	[0,1]	[-18.174,-18.174]	[28.000,47.916]
I_1	[-1,0]	[-1,0]	[-1.083,-1.083]	[-2.542,-2.542]
I_2	[-1.5,-0.5]	[-1.5,-0.5]	[-1.150,-1.150]	[-2.323,-2.323]
I_3	[0,1]	[0,1]	[1.733,2.149]	[-1.354,-1.354]

Table 6 Estimates of coefficients according to different distance measures under *MCI-MCII-MCIII* and *MCIV* settings for Case-II

Distance Measures	Parameters	*MCI*	*MCII*	*MCIII*	*MCIV*
Abdalla and Buckley [1]	$\tilde{A}_0$	(-0.4953/-0.4306/-0.3393)	(0.2464/0.4892/0.7266)	(-18.1740/-18.1740/-18.1740)	(35.8420/36.0300/36.0300)
	$\tilde{A}_1$	(-0.5005/-0.4656/-0.0059)	(-0.4815/-0.2852/-0.1398)	(-1.0830/-1.0830/-1.0830)	(-2.5420/-2.5420/-2.5420)
	$\tilde{A}_2$	(-0.7965/-0.7864/-0.7165)	(-0.8760/-0.8303/-0.7575)	(-1.1500/-1.1500/-1.1500)	(-2.3330/-2.3330/-2.3330)
	$\tilde{A}_3$	(0.3335/0.3540/0.3920)	(0.3174/0.3361/0.3398)	(1.7360/1.7521/1.7920)	(-1.3540/-1.3540/-1.3540)
Kaufmann and Gupta [13]	$\tilde{A}_0$	(-0.8530/-0.5900/-0.2935)	(0.0607/0.3163/0.3414)	(-18.1740/-18.1740/-18.1740)	(31.0713/31.5636/32.1763)
	$\tilde{A}_1$	(-0.6934/ -0.6033/-0.3096)	(-0.2712/-0.2684/-0.1293)	(-1.0830/-1.0830/-1.0830)	(-2.5420/-2.5420/-2.5420)
	$\tilde{A}_2$	(-1.4064/-1.3966/-1.3162)	(-0.8220/-0.7265/-0.7210)	(-1.150/-1.150/-1.150)	(-2.3230/-2.3230/-2.3230)
	$\tilde{A}_3$	(0.5474/0.5727/0.5923)	(0.2591/0.2938/0.3359)	(1.7337/1.7519/1.8307)	(-1.3540/-1.3540/-1.3540)
Heilpern-1 [12]	$\tilde{A}_0$	(-0.8472/-0.7690/-0.1782)	(0.0653/0.3254/0.3424)	(-18.174/-18.174/-18.174)	(28.6932/30.4576/35.6408)
	$\tilde{A}_1$	(-0.8527/-0.3606/-0.0810)	(-0.8627/-0.4147/-0.0858)	(-1.0830/-1.0830/-1.0830)	(-2.5420/-2.5420/-2.5420)
	$\tilde{A}_2$	(-1.4198/-1.1616/-0.5778)	(-1.4075/-1.2370/-0.6181)	(-1.1500/-1.1500/-1.1500)	(-2.3230/-2.3230/-2.3230)
	$\tilde{A}_3$	(0.0251/0.6431/0.7575)	(0.1463/0.4066/0.7275)	(1.7339/1.7583/1.7678)	(-1.3540/-1.3540/-1.3540)
Heilpern-2 [12]	$\tilde{A}_0$	(-0.8530/-0.5900/-0.2935)	(0.0607/0.3163/0.3414)	(-18.1740/-18.1740/-18.1740)	(31.0713/31.5636/32.1763)
	$\tilde{A}_1$	(-0.6934/-0.6033/-0.3096)	(-0.2712/-0.2684/-0.1293)	(-1.0830/-1.0830/-1.0830)	(-2.5420/-2.5420/-2.5420)
	$\tilde{A}_2$	(-1.4064/-1.3966/-1.3162)	(-0.8220/-0.7265/-0.7210)	(-1.1500/-1.1500/-1.1500)	(-2.3230/-2.3230/-2.3230)
	$\tilde{A}_3$	(0.5474/0.5727/0.5923)	(0.2591/0.2938/0.3359)	(1.7337/1.7519/1.8307)	(-1.3540/-1.3540/-1.3540)
Chen and Hsieh [4]	$\tilde{A}_0$	(-0.7617/-0.7454/-0.5821)	(0.0716/0.4464/0.5536)	(-18.1740/-18.1740/-18.1740)	(28.9831/31.8476/33.2103)
	$\tilde{A}_1$	(-0.6857/-0.4063/-0.3824)	(-0.9107/-0.4521/-0.0816)	(-1.0830/-1.0830/-1.0830)	(-2.5420/-2.5420/-2.5420)
	$\tilde{A}_2$	(-1.3294/-1.1576/-0.5469)	(-1.3458/-1.1448/-0.6135)	(-1.1500/-1.1500/-1.1500)	(-2.3230/-2.3230/-2.3230)
	$\tilde{A}_3$	(0.2521/0.4794/0.8036)	(0.2596/0.3323/0.9166)	(1.7443/1.7445/1.7981)	(-1.3540/-1.3540/-1.3540)

Table 7 Comparison of $E1$ values in the application for Case-II

Distance Measures	*MCI*	*MCII*	*MCIII*	*MCIV*
Abdalla and Buckley [1]	6.1690	5.8120	**7.1250**	**8.2010**
Kaufmann and Gupta [13]	32.6313	31.0182	24.1279	110.6466
Heilpern-1 [12]	**4.5126**	6.8999	12.202	50.9251
Heilpern-2 [12]	16.3156	15.5091	12.0639	55.3233
Chen and Hsieh [4]	6.1242	**4.8169**	11.7306	58.7061

Table 8 Data for the application (Case-III)

Fuzzy output	$\widetilde{X}_{1l}$	$\widetilde{X}_{2l}$
(55.4/61.6/64.7)	(5.7/6.0/6.9)	(5.4/6.3/7.1)
(50.5/53.2/58.5)	(4.0/4.4/5.1)	(4.7/5.5/5.8)
(55.7/65.5/75.3)	(8.6/9.1/9.8)	(3.4/3.6/4.0)
(61.7/64.9/74.7)	(6.9/8.1/9.3)	(5.0/5.8/6.7)
(69.1/71.7/80.0)	(8.7/9.4/11.2)	(6.5/6.8/7.1)
(49.6/52.2/57.4)	(4.6/4.8/5.5)	(6.7/7.9/8.7)
(47.7/50.2/55.2)	(7.2/7.6/8.7)	(4.0/4.2/4.8)
(41.8/44.0/48.4)	(4.2/4.4/4.8)	(5.4/6.0/6.3)
(45.7/53.8/61.9)	(8.2/9.1/10.0)	(2.7/2.8/3.2)
(45.4/53.5/58.9)	(6.0/6.7/7.4)	(5.7/6.7/7.7)

Table 9 Intervals for I_i, $i = 0, 1, 2$ for Case-III

Interval	*MCI*	*MCII*	*MCIII*	*MCIV*
I_0	[0,5]	[0,37]	[16.528,16.528]	[33.808,36.601]
I_1	[0,6]	[0,6]	[3.558,3.982]	[1.294,3.756]
I_2	[0,4]	[0,6]	[2.575,2.575]	[0.423,0.423]

Table 10 Estimates of coefficients according to different distance measures under *MCI-MCII-MCIII* and *MCIV* settings for Case-III

Distance Measures	Parameters	*MCI*	*MCII*	*MCIII*	*MCIV*
Abdalla and Buckley [2]	a_0	3.9855	35.3251	16.528	33.8196
	a_1	0.0060	3.9498	3.9820	3.7559
	a_2	0.0096	0.0063	2.5750	0.4730
Kaufmann and Gupta [13]	a_0	1.9138	1.8114	16.5280	33.8108
	a_1	4.7655	4.7820	3.5733	3.1333
	a_2	3.6687	3.6775	2.5750	0.4730
Heilpern-1 [12]	a_0	2.4841	0.3650	16.5280	33.8106
	a_1	4.9058	4.8024	3.5580	2.7181
	a_2	3.4424	3.9099	2.5750	0.7430
Heilpern-2 [12]	a_0	1.9138	1.8114	16.5280	33.8108
	a_1	4.7655	4.7820	3.5733	3.1333
	a_2	3.6687	3.6775	2.5750	0.4730
Chen and Hsieh [4]	a_0	2.1047	0.5538	16.5280	33.8086
	a_1	5.0605	5.0276	3.5580	3.0994
	a_2	3.3305	3.6148	2.5750	0.4730

Table 11 Comparison of $E1$ values in the application for Case-III

Distance Measures	*MCI*	*MCII*	*MCIII*	*MCIV*
Abdalla and Buckley [2]	**10.017**	**9.3890**	12.7267	**9.5933**
Kaufmann and Gupta [13]	52.7943	83.9582	19.0558	24.3161
Heilpern-1 [12]	26.2680	42.0170	**9.4604**	13.4241
Heilpern-2 [12]	26.3971	41.9791	9.5279	12.1581
Chen and Hsieh [4]	26.3563	41.9412	9.4544	11.6395