

# Multivariate Least Squares Regression using Interval-Valued Fuzzy Data and based on Extended Yao-Wu Signed Distance

Faezeh Torkian, Mohsen Arefi\*, Mohammad Ghasem Akbari

*Department of Statistics, Faculty of Mathematical Sciences and Statistics,  
University of Birjand, Birjand, Iran*

*E-mails: Fa\_Torkian@yahoo.com, Arefi@birjand.ac.ir, G\_Z\_Akbari@yahoo.com*

Received 18 November 2012

Accepted 16 June 2013

## Abstract

The purpose of this study is to introduce a new regression model, based on the least squares method, when the available data of both explanatory variable(s) and response variable are interval-valued fuzzy (IVF) numbers. The proposed method is based on a new metric on the space of IVF numbers, which is an extended version of the signed distance introduced by Yao and Wu (2000). In order to evaluate the goodness of fit of the proposed model, we introduce some new indices based on the similarity measure and the coefficient of multiple determination. Finally, the application of proposed approach is provided to model some real data.

**Keywords:** Coefficient of multiple determination, Goodness of fit, Interval-valued fuzzy set, Multivariate least squares regression, Similarity measure.

## 1 Introduction

Regression is a very powerful tool in statistic for analyzing data and finding the relationship between variables. Fuzzy regression in 1980 decade after presenting fuzzy set theory by Zadeh<sup>1</sup>, has been studied. In general, there are two approaches for modelling a linear regression in imprecise (fuzzy) environments. In first approach, the parameters of regression model are estimated based on the linear/goal programming methods and in second approach, they are estimated based on least absolute/least squares errors methods. The first studies on regression analysis in fuzzy environment initiated by Tanaka *et al.*<sup>2,3</sup> based on linear programming method and by Celmins<sup>4</sup> and Diamond<sup>5</sup> based on least squares errors method. For studying some other works on regression model based

on linear/goal programming methods, see Yen *et al.*<sup>6</sup>, Nasrabadi and Nasrabadi<sup>7</sup>, Hasanpour *et al.*<sup>8,9</sup>.

In this paper, we focus on the least squares regression model. Hence, some approaches in this topic can be presented as follows: Wünsche and Näther<sup>10</sup> investigated an approach to model the least squares fuzzy regressions using  $L_2$  metric and based on random fuzzy variables. Yang and Lin<sup>11</sup> studied the estimation of fuzzy parameters of a regression model based on least squares method when the input and output data are fuzzy. Wu<sup>12</sup> and Kao and Chyu<sup>13</sup> introduced a least squares regression model using the extension principle and based on fuzzy observation. Mohammadi and Taheri<sup>14</sup> studied a least squares fuzzy regression model with fuzzy parameters and crisp input-fuzzy output data. Coppi *et al.*<sup>15</sup> studied a new approach of a least squares regression model with the LR fuzzy re-

\*Corresponding author, Tel.: (+98) 561 2502301; Fax: (+98) 561 2502041.

sponse variables. Arabpour and Tata<sup>16</sup> presented a least squares method for estimating the parameters of fuzzy regression model based on the distance introduced by Diamond<sup>5</sup>. Choi and Yoon<sup>17</sup> introduced a general fuzzy regression model, which separates the response function on a mode and spreads of an  $\alpha$ -level set for an observed fuzzy number, to estimate a fuzzy relation between two fuzzy random variables. Ferraro and Giordani<sup>18</sup> investigated the multiple linear regression model in the presence of one or more imprecise (fuzzy) elements. Wu<sup>19</sup> presented a least squares fuzzy linear regression model with fuzzy parameter and imprecise (fuzzy) input and output data. In this approach, the  $\alpha$ -cuts of fuzzy linear regression model is constructed based on some statistical techniques. Taheri and Kelkinnama<sup>20</sup> and Kelkinnama and Taheri<sup>21</sup> also investigated some approaches to model the fuzzy linear regression based on least absolute methods. Roh *et al.*<sup>22</sup> studied an estimation approach to determine the parameters of the fuzzy linear regression model. In this study, a new methodology of fuzzy linear regression based on the design method of polynomial neural networks is proposed. For an overview on the various methods of regression models in imprecise environment, see Taheri<sup>23</sup>.

The regression models in imprecise environments can be used in other fields. For example, An *et al.*<sup>24</sup> studied some techniques for machine learning based on support vector regression when the available data are as the interval data, and Sentürk<sup>25</sup> investigated some fuzzy regression control charts for evaluating the process in which the average has a trend and the data represents a linguistic value. Also, we will need to introduce some new procedures (in future works) for analyzing the regression models based on soft computing methods<sup>26,27,28</sup> and/or using the methods of computing with words<sup>29,30</sup>.

Although the fuzzy sets theory provides the useful methods for modelling complex systems, but there are some situations that the evaluations of membership and non-membership values are not possible, and consequently, there remains an indeterministic value on which hesitation survives. Certainly, the fuzzy sets theory is not appropriated to deal with such problems. The interval-valued (intuitionistic) fuzzy sets theory<sup>31,32,33</sup> is a generalization of fuzzy sets theory which can an-

swer in such situations. This theory has been widely applied in various fields such as: decision making<sup>34</sup>, logic programming<sup>35,36</sup>, medical diagnosis<sup>37</sup>, pattern recognition<sup>38</sup>, and ...

Based on knowledge of the authors, there has not been any work in the problem of linear regression analysis in interval-valued fuzzy environment. Hence, in this paper, we want to model a least square regression based on interval-valued fuzzy data. For executing this idea, we first extend the Yao-Wu signed distance between interval-valued fuzzy numbers, and then, the parameters of regression model are estimated.

The paper is organized as follows: In Section 2, we review some preliminary concepts on interval-valued fuzzy sets. In Section 3, the Yao-Wu signed distance is extended based on interval-valued fuzzy numbers. In Section 4, we propose a least squares approach to analyze a multivariate regression model when the input and output data of model are as interval-valued fuzzy numbers. In Section 5, some new indices to evaluate the goodness of fit of proposed regression model are introduced. Application of the proposed approach to model some real data is studied in Section 6. Finally, in Section 7, a brief conclusion is provided.

## 2 Preliminary concepts

Let  $X$  be an universal set. A fuzzy set  $\tilde{A}$  is defined as  $\tilde{A} = \{(x, \mu_{\tilde{A}}(x)) : x \in X\}$ , where  $\mu_{\tilde{A}}(x) : X \rightarrow [0, 1]$  is the degree of membership of  $x$  into  $\tilde{A}$ . Thus, it is clear that the degree of non-membership of  $x$  into  $\tilde{A}$  is  $1 - \mu_{\tilde{A}}(x)$ . Note that, in some cases, the degree of non-membership is not always defined as  $1 - \mu_{\tilde{A}}(x)$ . For solving this problem, Atanassov<sup>31</sup> generalized the notion of fuzzy set theory to the concept of the intuitionistic fuzzy set (IFS) which was composed of the membership degree, non-membership degree and indeterminacy degree of  $x$  into  $\tilde{A}$ . Also, Gau and Buehrer<sup>39</sup> introduced the concept of vague sets (VS), which is another generalization of fuzzy sets. Another well-known generalization of ordinary fuzzy sets is the concept of interval-valued fuzzy set (IVFS) introduced by Gorzalczyński<sup>32</sup>, Atanassov and Gargov<sup>33</sup>. Note that these approaches are in general not independent and there exist relationships among them (see also, Bustince and Burl-

lio<sup>40</sup>).

Since, in this paper, we focus on the regression models in the interval-valued fuzzy environment, we recall some preliminary concepts about the interval-valued fuzzy set as follow (see Turksen<sup>41</sup>, Kumar and Biswas<sup>42</sup>, Grzegorzewski<sup>43</sup>, Guha and Chakraborty<sup>44</sup>).

**Definition 1.** An interval-valued fuzzy set  $\tilde{A}$  on the universal set  $X$  is defined as

$$\tilde{A} = \{ \langle x, \mu_{\tilde{A}}(x), \nu_{\tilde{A}}(x) \rangle | x \in X \}, \quad (1)$$

where  $\mu_{\tilde{A}} : X \rightarrow [0, 1]$  is the “degree of membership”,  $\nu_{\tilde{A}} : X \rightarrow [0, 1]$  is the “degree of nonmembership”, and  $0 \leq \mu_{\tilde{A}}(x) + \nu_{\tilde{A}}(x) \leq 1$  for all  $x \in X$ . Also, the value  $\tau_{\tilde{A}}(x) = 1 - \mu_{\tilde{A}}(x) - \nu_{\tilde{A}}(x)$  is called the “degree of indeterminacy” of the element  $x \in X$  to the IVFS  $\tilde{A}$ .

Note that in the above definition,  $\mu_{\tilde{A}}(x)$  is the lower bound for degree of membership of  $x$  into  $\tilde{A}$ , and  $\nu_{\tilde{A}}(x)$  is the lower bound for negation of membership of  $x$  into  $\tilde{A}$ . Therefore, the degree of membership of  $x$  into the interval-valued fuzzy set  $\tilde{A}$  is characterized by the interval  $[\mu_{\tilde{A}}(x), 1 - \nu_{\tilde{A}}(x)]$ .

**Definition 2.** An interval-valued fuzzy set  $\tilde{A}$  is called an interval-valued fuzzy number (IVFN), if

- (i) There exist  $m \in R$ , such that  $\mu_{\tilde{A}}(m) = 1$  and  $\nu_{\tilde{A}}(m) = 0$ .
- (ii) The membership and non-membership functions are the continuous mapping from  $R$  to  $[0, 1]$  as follows:

$$\mu_{\tilde{A}}(x) = \begin{cases} f_1(x) & m - s_1 \leq x < m, \\ 1 & x = m, \\ h_1(x) & m < x < m + s_2, \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

$$\nu_{\tilde{A}}(x) = \begin{cases} f_2(x) & m - s_3 \leq x < m, \\ 0 & x = m, \\ h_2(x) & m < x < m + s_4, \\ 1 & \text{otherwise,} \end{cases} \quad (3)$$

where,  $f_1$  and  $h_2$  are strictly increasing functions and  $h_1$  and  $f_2$  are strictly decreasing functions.

Also,  $s_1, s_2 \geq 0$  and  $s_3, s_4 \geq 0$  are the spreads of  $\mu_{\tilde{A}}(x)$  and  $\nu_{\tilde{A}}(x)$ , respectively. An IVFN is denoted by  $\tilde{A} = (m, s_1, s_2, s_3, s_4)$ .

**Example 1.** In the following, the membership and non-membership functions of an IVF number are given (Fig. 1)

$$\mu_{\tilde{A}}(x) = \begin{cases} e^{-(\frac{20-x}{5})^2}, & -\infty \leq x < 20 \\ e^{-(\frac{x-20}{6})^2}, & 20 \leq x < \infty \end{cases}$$

$$\nu_{\tilde{A}}(x) = \begin{cases} 1 - e^{-(\frac{20-x}{9})^2}, & -\infty \leq x < 20 \\ 1 - e^{-(\frac{x-20}{9})^2}, & 20 \leq x < \infty \end{cases}$$

These values represent the interval-valued fuzzy number “approximately 20”, in which the degree of membership in each point  $x$  is characterized by the interval  $[\mu_{\tilde{A}}(x), 1 - \nu_{\tilde{A}}(x)]$  (see Fig. 1). For instance, the degree of membership for  $x = 20$  is exactly equal to 1, and for  $x = 15$  is a value between 0.37 and 0.73.

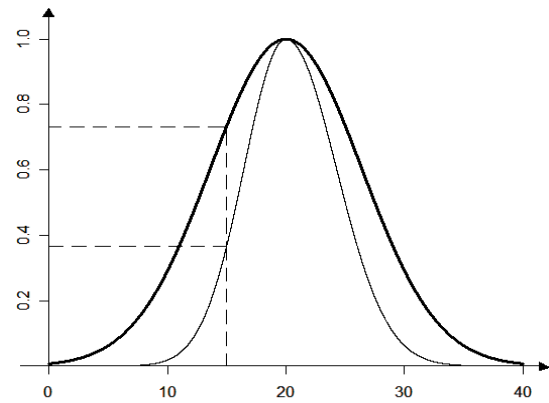


Fig. 1. Interval-valued fuzzy number in Example 1.

**Definition 3.** An interval-valued fuzzy number  $\tilde{A}$  is called a LR-IVFN, if the membership and nonmembership functions are as

$$\mu_{\tilde{A}}(x) = \begin{cases} L\left(\frac{m-x}{s_1}\right) & m - s_1 \leq x < m, \\ 1 & x = m, \\ R\left(\frac{x-m}{s_2}\right) & m < x < m + s_2, \\ 0 & \text{otherwise,} \end{cases} \quad (4)$$

$$v_{\tilde{A}}(x) = \begin{cases} 1 - L\left(\frac{m-x}{s_3}\right) & m - s_3 \leq x < m, \\ 0 & x = m, \\ 1 - R\left(\frac{x-m}{s_4}\right) & m < x < m + s_4, \\ 1 & \text{otherwise,} \end{cases} \quad (5)$$

where,  $L(\cdot)$  and  $R(\cdot)$  are strictly decreasing functions from  $R^+$  to  $[0,1]$ , and  $L(0) = R(0) = 1$ .  $L(\cdot)$  and  $R(\cdot)$  are called the reference functions. A LR-IVFN is denoted by  $\tilde{A} = (m; s_1, s_2, s_3, s_4)_{LR}$  (see Guha and Chakraborty<sup>44</sup>).

Note that based on the different functions for  $L(\cdot)$  and  $R(\cdot)$ , we can provide the wide kinds of LR-IVF numbers. A well-known case of LR-IVF numbers is the triangular interval-valued fuzzy number (TIVFN) that is given as follows.

**Remark 1.** The interval-valued fuzzy number  $\tilde{A}$  is called a triangular interval-valued fuzzy number (TIVFN), denoted by  $\tilde{A} = (m; s_1, s_2, s_3, s_4)_T$ , if  $L(x) = R(x) = \max\{0, 1 - x\}$  for all  $x \in [0,1]$ . Note that in such a case,  $s_3 > s_1$  and  $s_4 > s_2$  (for proof, see, Guha and Chakraborty<sup>44</sup>).

**Definition 4.** (Guha and Chakraborty<sup>44</sup>, Taheri and Zarei<sup>45</sup>) Let  $\tilde{A}$  be an IVFS on  $X$ . Then, the  $\alpha$ -cuts of  $\tilde{A}$  are defined by the following two crisp sets

$$\begin{cases} \tilde{A}_\mu[\alpha] &= \{x : \mu_{\tilde{A}}(x) \geq \alpha\}, \\ \tilde{A}_{1-\nu}[\alpha] &= \{x : 1 - \nu_{\tilde{A}}(x) \geq \alpha\}. \end{cases} \quad (6)$$

In a special case, if  $\tilde{A} = (m; s_1, s_2, s_3, s_4)_{LR}$  is a LR-IVFN on  $X$ , then the  $\alpha$ -cuts of  $\tilde{A}$  are as

$$\begin{cases} \tilde{A}_\mu[\alpha] &= [m - s_1 L^{-1}(\alpha), m + s_2 R^{-1}(\alpha)], \\ \tilde{A}_{1-\nu}[\alpha] &= [m - s_3 L^{-1}(\alpha), m + s_4 R^{-1}(\alpha)]. \end{cases} \quad (7)$$

In the following, the arithmetic operations on LR-IVFN's are defined based on the "Extension Principle" for IVF sets (see Taheri and Zarei<sup>45</sup>).

**Proposition 1.** Let  $\tilde{M} = (m; s_1, s_2, s_3, s_4)$  and  $\tilde{N} = (n; r_1, r_2, r_3, r_4)$  be two LR-IVFN's and  $\lambda \in R - \{0\}$ . Then,

$$\begin{aligned} \tilde{M} \oplus \tilde{N} &= (m; s_1, s_2, s_3, s_4)_{LR} \oplus (n; r_1, r_2, r_3, r_4)_{LR} \\ &= (m+n; s_1+r_1, s_2+r_2; s_3+r_3, s_4+r_4)_{LR}, \end{aligned} \quad (8)$$

$$\begin{aligned} \tilde{M} \ominus \tilde{N} &= (m; s_1, s_2, s_3, s_4)_{LR} \ominus (n; r_1, r_2, r_3, r_4)_{RL} \\ &= (m-n; s_1+r_2, s_2+r_1; s_3+r_4, s_4+r_3)_{LR}, \end{aligned} \quad (9)$$

$$\lambda \otimes \tilde{M} = \begin{cases} (\lambda m; \lambda s_1, \lambda s_2, \lambda s_3, \lambda s_4)_{LR} & \lambda > 0, \\ (\lambda m; -\lambda s_2, -\lambda s_1, -\lambda s_4, -\lambda s_3)_{RL} & \lambda < 0. \end{cases} \quad (10)$$

**Definition 5.** (Hung and Yang<sup>38</sup>) Let  $\tilde{A}$  and  $\tilde{B}$  be two IVF sets. Then

- (i)  $\tilde{A} \subseteq \tilde{B}$  if and only if for each  $x \in X$ ,  $\mu_{\tilde{A}}(x) \leq \mu_{\tilde{B}}(x)$  and  $\nu_{\tilde{A}}(x) \geq \nu_{\tilde{B}}(x)$ ,
- (ii)  $\tilde{A} = \tilde{B}$  if and only if for each  $x \in X$ ,  $\mu_{\tilde{A}}(x) = \mu_{\tilde{B}}(x)$  and  $\nu_{\tilde{A}}(x) = \nu_{\tilde{B}}(x)$ .

### 3 Extension of the Yao-Wu signed distance

In this section, we define a new distance between the interval-valued fuzzy numbers. This distance is an extended version of the Yao-Wu signed distance<sup>46</sup>. It should be mentioned that the distance between IVF sets introduced by some other authors, for instance, Atanassov<sup>31</sup>, Grzegorzewski<sup>43</sup>, Hung and Yang<sup>38</sup>, Guha and Chakraborty<sup>44</sup> and Li *et al.*<sup>47</sup>.

**Definition 6.** Let  $\tilde{A}$  and  $\tilde{B}$  be two interval-valued fuzzy numbers. The extension of Yao-Wu signed distance between  $\tilde{A}$  and  $\tilde{B}$  is defined as follows

$$\begin{aligned} d(\tilde{A}, \tilde{B}) &= \frac{1}{2} \int_0^1 (M_\alpha(\tilde{A}_\mu) - M_\alpha(\tilde{B}_\mu)) d\alpha \\ &\quad + \frac{1}{2} \int_0^1 (M_\alpha(\tilde{A}_{1-\nu}) - M_\alpha(\tilde{B}_{1-\nu})) d\alpha, \end{aligned} \quad (11)$$

where

$$\begin{aligned} M_\alpha(\tilde{A}_\mu) &= \frac{\tilde{A}_\mu^L[\alpha] + \tilde{A}_\mu^R[\alpha]}{2}, & M_\alpha(\tilde{A}_{1-\nu}) &= \frac{\tilde{A}_{1-\nu}^L[\alpha] + \tilde{A}_{1-\nu}^R[\alpha]}{2}, \\ M_\alpha(\tilde{B}_\mu) &= \frac{\tilde{B}_\mu^L[\alpha] + \tilde{B}_\mu^R[\alpha]}{2}, & M_\alpha(\tilde{B}_{1-\nu}) &= \frac{\tilde{B}_{1-\nu}^L[\alpha] + \tilde{B}_{1-\nu}^R[\alpha]}{2}. \end{aligned}$$

and,  $\tilde{A}_\mu^L[\alpha]$ ,  $\tilde{A}_\mu^R[\alpha]$ ,  $\tilde{A}_{1-\nu}^L[\alpha]$ ,  $\tilde{A}_{1-\nu}^R[\alpha]$  are the  $\alpha$ -cuts of  $\tilde{A}$ , and  $\tilde{B}_\mu^L[\alpha]$ ,  $\tilde{B}_\mu^R[\alpha]$ ,  $\tilde{B}_{1-\nu}^L[\alpha]$ ,  $\tilde{B}_{1-\nu}^R[\alpha]$  are the  $\alpha$ -cuts of  $\tilde{B}$ .

**Remark 2.** Note that if the IVF numbers  $\tilde{A}$  and  $\tilde{B}$  is reduced to fuzzy numbers, then, the distance introduced in Definition 6, is reduced to the Yao-Wu signed distance<sup>46</sup> as follows

$$d(\tilde{A}, \tilde{B}) = \int_0^1 (M_\alpha(\tilde{A}) - M_\alpha(\tilde{B})) d\alpha, \quad (12)$$

where

$$M_\alpha(\tilde{A}) = \frac{\tilde{A}^L[\alpha] + \tilde{A}^R[\alpha]}{2}, \quad M_\alpha(\tilde{B}) = \frac{\tilde{B}^L[\alpha] + \tilde{B}^R[\alpha]}{2}.$$

In a special case, if  $\tilde{A} = (a, s_1, s_2, s_3, s_4)_T$  and  $\tilde{B} = (b, r_1, r_2, r_3, r_4)_T$  are the triangular IVFN's, then the extension of Yao-Wu signed distance between  $\tilde{A}$  and  $\tilde{B}$  is presented as

$$d(\tilde{A}, \tilde{B}) = a - b + \frac{1}{8}[(s_2 - s_1 + s_4 - s_3)] - \frac{1}{8}[(r_2 - r_1 + r_4 - r_3)]. \quad (13)$$

**Definition 7.** Let  $\tilde{A}$  and  $\tilde{B}$  be two IVFN's. Then, the ranking of  $\tilde{A}$  and  $\tilde{B}$  is expressed as

$$\begin{aligned} d(\tilde{A}, \tilde{B}) > 0 &\Leftrightarrow d(\tilde{A}, 0) > d(\tilde{B}, 0) \Leftrightarrow \tilde{A} \succ \tilde{B}, \\ d(\tilde{A}, \tilde{B}) < 0 &\Leftrightarrow d(\tilde{A}, 0) < d(\tilde{B}, 0) \Leftrightarrow \tilde{A} \prec \tilde{B}, \\ d(\tilde{A}, \tilde{B}) = 0 &\Leftrightarrow d(\tilde{A}, 0) = d(\tilde{B}, 0) \Leftrightarrow \tilde{A} \approx \tilde{B}. \end{aligned} \quad (14)$$

**Lemma 2.** Let  $\tilde{A}, \tilde{B}, \tilde{C} \in IVFN(R)$ . Then, the signed distance introduced in Definition 6 satisfies the following properties:

- (i)  $d(\tilde{A}, \tilde{B}) = -d(\tilde{B}, \tilde{A})$ ,
- (ii)  $d(\tilde{A}, \tilde{B}) + d(\tilde{B}, \tilde{C}) = d(\tilde{A}, \tilde{C})$ ,
- (iii)  $\tilde{A} \approx \tilde{B} \Leftrightarrow \tilde{B} \approx \tilde{A}$ ,
- (iv)  $\tilde{A} \approx \tilde{B}, \tilde{B} \approx \tilde{C} \Rightarrow \tilde{A} \approx \tilde{C}$ .

**Proof.**

(i) It follows from the signed distance property

$$\begin{aligned} d(\tilde{A}, \tilde{B}) &= d(\tilde{A}, 0) - d(\tilde{B}, 0) \\ &= -(d(\tilde{B}, 0) - d(\tilde{A}, 0)) = -d(\tilde{B}, \tilde{A}), \end{aligned}$$

(ii) From item (i), we have

$$\begin{aligned} d(\tilde{A}, \tilde{B}) + d(\tilde{B}, \tilde{C}) &= d(\tilde{A}, 0) - d(\tilde{B}, 0) \\ &\quad + d(\tilde{B}, 0) - d(\tilde{C}, 0) \\ &= d(\tilde{A}, \tilde{C}), \end{aligned}$$

(iii) From item (i) and Definition 7, we have

$$\tilde{A} \approx \tilde{B} \Leftrightarrow d(\tilde{A}, \tilde{B}) = 0 = -d(\tilde{B}, \tilde{A}),$$

Hence,  $d(\tilde{B}, \tilde{A}) = 0$ , and  $\tilde{B} \approx \tilde{A}$ .

(iv) It is simple based on item (ii) and Definition 7.  $\square$

## 4 Multivariate least squares regression based on interval-valued fuzzy data

In this section, we introduce a linear regression model based on the extension of Yao-Wu signed distance between the interval-valued fuzzy input-output data. For simplicity, we assume that the input-output data are the triangular IVFN's. We want to fit a regression model with the crisp coefficients  $\beta_j$ ,  $j = 0, 1, \dots, k$ , and based on the triangular IVF observations  $(\tilde{x}_{i1}, \tilde{x}_{i2}, \dots, \tilde{x}_{ik}, \tilde{y}_i)$ ,  $i = 1, 2, \dots, n$ , as follows

$$\begin{aligned} \tilde{y}_i &= \beta_0 \oplus (\beta_1 \otimes \tilde{x}_{i1}) \oplus \dots \oplus (\beta_k \otimes \tilde{x}_{ik}) \\ &= \beta_0 \oplus \sum_{j=1}^k (\beta_j \otimes \tilde{x}_{ij}), \quad i = 1, \dots, n \end{aligned} \quad (15)$$

where,  $\tilde{y}_i = (y_i; r_{i1}, r_{i2}, r_{i3}, r_{i4})_T$  and  $\tilde{x}_{ij} = (x_{ij}, s_{ij1}, s_{ij2}, s_{ij3}, s_{ij4})_T$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, k$ . Based on the distance introduced in Definition 6 and the arithmetic operations on LR-IVFN's (Proposition 1), the sum of squares errors for the regression model (15) is obtained as follows

$$\begin{aligned} SSE &= \sum_{i=1}^n d^2[\tilde{y}_i, \beta_0 \oplus (\beta_1 \otimes \tilde{x}_{i1}) \oplus \dots \oplus (\beta_k \otimes \tilde{x}_{ik})] \\ &= \sum_{i=1}^n \left[ (y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij}) + \frac{1}{8} [R_i - \sum_{j=1}^k \beta_j S_{ij}] \right]^2, \end{aligned} \quad (16)$$

where,  $R_i = r_{i2} - r_{i1} + r_{i4} - r_{i3}$  and  $S_{ij} = s_{ij2} - s_{ij1} + s_{ij4} - s_{ij3}$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, k$ . To estimate the parameters of regression model (15), we obtain the following results:

$$\frac{\partial SSE}{\partial \beta_0} = \sum_{i=1}^n (y_i + \frac{1}{8}R_i) - n\beta_0 - \sum_{i=1}^n \sum_{j=1}^k \beta_j (x_{ij} + \frac{1}{8}S_{ij}), \quad (17)$$

$$\begin{aligned} \frac{\partial SSE}{\partial \beta_j} &= \sum_{i=1}^n (x_{ij} + \frac{1}{8}S_{ij})(y_i + \frac{1}{8}R_i) \\ &\quad - \sum_{i=1}^n (x_{ij} + \frac{1}{8}S_{ij})(\beta_0 + \sum_{j=1}^k \beta_j (x_{ij} + \frac{1}{8}S_{ij})). \end{aligned}$$

By taking  $\frac{\partial SSE}{\partial \beta_j} = 0$  for  $j = 0, 1, \dots, k$ , and using the matrix forms, we can rewrite the above relations as follows

$$(X + \frac{S}{8})'(X + \frac{S}{8})\beta = (X + \frac{S}{8})'(Y + \frac{R}{8}), \quad (18)$$

where,  $R = [R_1, R_2, \dots, R_n]'$ ,  $Y = [y_1, y_2, \dots, y_n]'$ ,  $\beta = [\beta_0, \beta_1, \dots, \beta_k]'$ , and

$$X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1k} \\ 1 & x_{21} & \dots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{nk} \end{bmatrix}_{n \times (k+1)},$$

$$S = \begin{bmatrix} 0 & S_{11} & \dots & S_{1k} \\ 0 & S_{21} & \dots & S_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & S_{n1} & \dots & S_{nk} \end{bmatrix}_{n \times (k+1)}.$$

If  $(X + \frac{S}{8})'(X + \frac{S}{8})$  is a nonsingular matrix (i.e.  $[(X + \frac{S}{8})'(X + \frac{S}{8})]^{-1}$  exists), then the parameters of regression model are estimated as follows

$$\hat{\beta} = \left[ (X + \frac{S}{8})'(X + \frac{S}{8}) \right]^{-1} \left[ (X + \frac{S}{8})'(Y + \frac{R}{8}) \right]. \quad (19)$$

**Theorem 3.** The proposed IVF regression model is estimated such that it can be partitioned as follows

$$SST = SSE + SSR,$$

where,  $SST = \sum_{i=1}^n d^2(\tilde{y}_i, \tilde{\tilde{y}})$ ,  $SSE = \sum_{i=1}^n d^2(\tilde{y}_i, \hat{\tilde{y}}_i)$ , and  $SSR = \sum_{i=1}^n d^2(\hat{\tilde{y}}_i, \tilde{\tilde{y}})$  are the total sum of squares, the

sum of squares errors, and the regression sum of squares, respectively.

**Proof.**

From  $\frac{\partial SSE}{\partial \beta_0} = 0$  in Equation (17), we can inference the following result

$$\begin{aligned} \hat{\beta}_0 &= \frac{1}{n} \sum_{i=1}^n (y_i + \frac{1}{8}R_i) - \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^k \beta_j (x_{ij} + \frac{1}{8}S_{ij}) \\ &= d(\tilde{\tilde{y}}, (\hat{\beta}_1 \otimes \tilde{\tilde{x}}_{.1}) \oplus \dots \oplus (\hat{\beta}_k \otimes \tilde{\tilde{x}}_{.k})), \end{aligned} \quad (20)$$

where  $\tilde{\tilde{x}}_{.j} = \frac{1}{n} \otimes (\tilde{x}_{1j} \oplus \dots \oplus \tilde{x}_{nj})$ . Hence, based on the properties of the proposed Yao-Wu signed distance (Lemma 2), we have

$$\begin{aligned} d(\tilde{\tilde{y}}, 0) &= \frac{1}{n} \sum_{i=1}^n d(\tilde{y}_i, 0) \\ &= \frac{1}{n} \sum_{i=1}^n d(\hat{\beta}_0 \oplus (\hat{\beta}_1 \otimes \tilde{x}_{i1}) \oplus \dots \oplus (\hat{\beta}_k \otimes \tilde{x}_{ik}), 0) \\ &= \frac{1}{n} \sum_{i=1}^n d[d(\tilde{\tilde{y}}, (\hat{\beta}_1 \otimes \tilde{\tilde{x}}_{.1}) \oplus \dots \oplus (\hat{\beta}_k \otimes \tilde{\tilde{x}}_{.k})) \\ &\quad \oplus (\hat{\beta}_1 \otimes \tilde{x}_{i1}) \oplus \dots \oplus (\hat{\beta}_k \otimes \tilde{x}_{ik}), 0] \\ &= \frac{1}{n} \sum_{i=1}^n [d(\tilde{\tilde{y}}, (\hat{\beta}_1 \otimes \tilde{\tilde{x}}_{.1}) \oplus \dots \oplus (\hat{\beta}_k \otimes \tilde{\tilde{x}}_{.k})) \\ &\quad + d((\hat{\beta}_1 \otimes \tilde{x}_{i1}) \oplus \dots \oplus (\hat{\beta}_k \otimes \tilde{x}_{ik}), 0)] \\ &= \frac{1}{n} \sum_{i=1}^n [d(\tilde{\tilde{y}}, 0) - d((\hat{\beta}_1 \otimes \tilde{\tilde{x}}_{.1}) \oplus \dots \oplus (\hat{\beta}_k \otimes \tilde{\tilde{x}}_{.k}), 0) \\ &\quad + d((\hat{\beta}_1 \otimes \tilde{x}_{i1}) \oplus \dots \oplus (\hat{\beta}_k \otimes \tilde{x}_{ik}), 0)] \\ &= \frac{1}{n} \sum_{i=1}^n d((\hat{\beta}_1 \otimes \tilde{x}_{i1}) \oplus \dots \oplus (\hat{\beta}_k \otimes \tilde{x}_{ik}) \\ &\quad , (\hat{\beta}_1 \otimes \tilde{\tilde{x}}_{.1}) \oplus \dots \oplus (\hat{\beta}_k \otimes \tilde{\tilde{x}}_{.k})) \\ &\quad + \frac{1}{n} \sum_{i=1}^n d(\tilde{\tilde{y}}, 0) = \frac{1}{n} \sum_{i=1}^n d(\tilde{\tilde{y}}, 0) + 0 \\ &= d(\tilde{\tilde{y}}, 0). \end{aligned} \quad (21)$$

Also, we have

$$\begin{aligned} \sum_{i=1}^n d(\tilde{y}_i, 0) d(\tilde{y}_i, \hat{\tilde{y}}_i) &= \sum_{i=1}^n \left( (\hat{\beta}_0 + \sum_{j=1}^k \hat{\beta}_j (x_{ij} + \frac{S_{ij}}{8})) \right. \\ &\quad \times (y_i + \frac{R_i}{8} - \hat{\beta}_0 - \sum_{j=1}^k \hat{\beta}_j (x_{ij} + \frac{S_{ij}}{8})) \Big) \\ &= ((X + \frac{S}{8})\hat{\beta})'(Y + \frac{R}{8} - (X + \frac{S}{8})\hat{\beta}) \\ &= \hat{\beta}'(X + \frac{S}{8})'(Y + \frac{R}{8}) - \hat{\beta}'(X + \frac{S}{8})'(X + \frac{S}{8})\hat{\beta} \quad (22) \\ &= \hat{\beta}'(X + \frac{S}{8})'(Y + \frac{R}{8}) - \hat{\beta}'(X + \frac{S}{8})'(Y + \frac{R}{8}) \\ &= 0, \end{aligned}$$

(By Equation (18))



and

$$\begin{aligned}
\sum_{i=1}^n d(\hat{y}_i, \tilde{y}) d(\tilde{y}_i, \hat{y}_i) &= \sum_{i=1}^n (d(\hat{y}_i, 0) - d(\tilde{y}, 0))(d(\tilde{y}_i, 0) - d(\hat{y}_i, 0)) \\
&= \sum_{i=1}^n [d(\hat{y}_i, 0)(d(\tilde{y}_i, 0) - d(\hat{y}_i, 0)) \\
&\quad - d(\tilde{y}, 0)d(\tilde{y}_i, 0) + d(\tilde{y}, 0)d(\hat{y}_i, 0)] \\
&= \sum_{i=1}^n d(\hat{y}_i, 0)d(\tilde{y}_i, \hat{y}_i) - nd^2(\tilde{y}, 0) \\
&\quad + nd(\tilde{y}, 0)d(\tilde{y}, 0) \\
&= \sum_{i=1}^n d(\hat{y}_i, 0)d(\tilde{y}_i, \hat{y}_i) - nd^2(\tilde{y}, 0) \\
&\quad + nd^2(\tilde{y}, 0) \quad (\text{By Equation (21)}) \\
&= \sum_{i=1}^n d(\hat{y}_i, 0)d(\tilde{y}_i, \hat{y}_i) \\
&= 0. \quad (\text{By Equation (22)})
\end{aligned}$$

Based on Lemma 2,  $d(\tilde{y}_i, \tilde{y}) = d(\tilde{y}_i, \hat{y}_i) + d(\hat{y}_i, \tilde{y})$ , and we obtain

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

From the above relations, we can inference the following result

$$\begin{aligned}
SST &= \sum_{i=1}^n d^2(\tilde{y}_i, \tilde{y}) \\
&= \sum_{i=1}^n d^2(\tilde{y}_i, \hat{y}_i) + \sum_{i=1}^n d^2(\hat{y}_i, \tilde{y}) \\
&\quad + 2 \sum_{i=1}^n d(\tilde{y}_i, \hat{y}_i)d(\hat{y}_i, \tilde{y}) \\
&= \sum_{i=1}^n d^2(\tilde{y}_i, \hat{y}_i) + \sum_{i=1}^n d^2(\hat{y}_i, \tilde{y}) + 0 \\
&= SSE + SSR.
\end{aligned}$$

Therefore, the proof is completed.  $\square$

## 5 Evaluation of the IVF regression model

In this section, we introduce some concepts and notations for evaluating the proposed IVF regression model.

### 5.1 Goodness of fit of the IVF regression model

To evaluate the goodness of fit of the IVF regression model, we introduce a new similarity measure between two IVFN's (see also, Li and Cheng<sup>48,49</sup>, Liang and Shi<sup>50</sup> and Zhang et al.<sup>51</sup>).

**Definition 8.** Let  $\tilde{A} = (a_0, a_1, a_2, a_3, a_4)_{LR}$  and  $\tilde{B} = (b_0, b_1, b_2, b_3, b_4)_{LR}$  be two LR-IVFN's. Then, the similarity measure between  $\tilde{A}$  and  $\tilde{B}$  is defined as follows

$$S(\tilde{A}, \tilde{B}) = \frac{1}{1 + \frac{1}{5} \sum_{i=0}^4 |a_i - b_i|^p} \cdot \quad p \geq 1 \quad (23)$$

**Theorem 4.** The mapping  $S$  on  $IVFN \times IVFN$  satisfies the properties of a similarity measure as follows

- (i)  $S(\tilde{A}, \tilde{B}) \in [0, 1]$ ,
- (ii)  $S(\tilde{A}, \tilde{B}) = 1$  iff  $\tilde{A} = \tilde{B}$ ,
- (iii)  $S(\tilde{A}, \tilde{B}) = S(\tilde{B}, \tilde{A})$ ,
- (iv) if  $\tilde{A} \subset \tilde{B} \subset \tilde{C}$ , then  $S(\tilde{A}, \tilde{C}) \leq \min\{S(\tilde{A}, \tilde{B}), S(\tilde{B}, \tilde{C})\}$ .

**Proof.**

- (i) Since for each  $i = 0, \dots, 4$ ,  $0 < |a_i - b_i|^p < \infty$ , we have

$$0 < \frac{1}{1 + \frac{1}{5} \sum_{i=0}^4 |a_i - b_i|^p} < 1.$$

- (ii) If  $S(\tilde{A}, \tilde{B}) = 1$ , then  $1 + \frac{1}{5} \sum_{i=0}^4 |a_i - b_i|^p = 1$ , and  $\frac{1}{5} \sum_{i=0}^4 |a_i - b_i|^p = 0$ . Hence,  $|a_i - b_i|^p = 0$ ,  $i = 0, \dots, 4$ , and we have  $\tilde{A} = \tilde{B}$ .

- (iii)  $|a_i - b_i|^p = |b_i - a_i|^p$  iff  $S(\tilde{A}, \tilde{B}) = S(\tilde{B}, \tilde{A})$ .

- (iv) If  $\tilde{A} \subset \tilde{B} \subset \tilde{C}$ , then based on Definition 5,  $a_0 = b_0 = c_0$  and  $a_i \leq b_i \leq c_i, i = 1, \dots, 4$ .

Also,  $|a_i - b_i|^p \leq |a_i - c_i|^p$  and

$$1 + \frac{1}{5} \sum_{i=0}^4 |a_i - b_i|^p \leq 1 + \frac{1}{5} \sum_{i=0}^4 |a_i - c_i|^p.$$

Hence, we obtain  $S(\tilde{A}, \tilde{C}) \leq S(\tilde{A}, \tilde{B})$ . Similarity,  $S(\tilde{A}, \tilde{C}) \leq S(\tilde{B}, \tilde{C})$ .

Therefore, the proof is completed.  $\square$

**Definition 9.** To evaluate goodness of fit of IVF regression model, the mean of similarity measures between the observed values  $\tilde{y}_i, i = 1, \dots, n$ , and the estimated values  $\hat{\tilde{y}}_i, i = 1, \dots, n$ , is defined as

$$\overline{SM} = \frac{1}{n} \sum_{i=1}^n S(\tilde{y}_i, \hat{\tilde{y}}_i). \quad (24)$$

## 5.2 Detection of outliers

Sometimes in applications, the data set contains some elements that are outlying or extreme. The existence of outliers in a set of experimental data can cause the incorrect interpretation of the regression results. Hence, if we improve the data set with removing outliers, we can obtain the better results in regression model.

In this paper, we identify the outliers by the proposed similarity measure  $S(.,.)$  and the square of the signed distance  $d^2(.,.)$  (introduced in Section 3) between the response values  $\tilde{y}_i$  and the estimated values  $\hat{\tilde{y}}_i, i = 1, \dots, n$ . The point that has the minimum of degree of similarity in between data (or the maximum of  $d^2(.,.)$ ), can be regarded as possible outlier.

## 5.3 Variable selection

Variable selection is a fundamental topic for choosing a suitable regression model. In practice, some variables are available in an initial analysis, but many of them may not be significant and should be excluded from the final model in order to increase the accuracy of prediction. Traditional variable selection procedures such as stepwise regression and the best subset variable selection for linear regression models can be extended

to the interval-valued fuzzy regression model. In the following, we introduce and extend some methods in the interval-valued fuzzy environment.

### 5.3.1 Coefficient of multiple determination

A measure of the adequacy of a linear regression model that has been widely used is the coefficient of multiple determination  $R_p^2$  with  $p = k + 1$  terms. It is the proportion of variation in the response variable  $y$  explained by the  $k$  predictors. The extension of  $R_p^2$  for the proposed IVF regression model is defined as

$$R_p^2 = \frac{SSR_p}{SST} = \frac{\sum_{i=1}^n d^2(\hat{\beta}_0 \oplus (\hat{\beta}_1 \otimes \tilde{x}_{i1}) \oplus (\hat{\beta}_2 \otimes \tilde{x}_{i2}) \oplus \dots \oplus (\hat{\beta}_k \otimes \tilde{x}_{ik}), \tilde{y})}{\sum_{i=1}^n d^2(\tilde{y}_i, \tilde{y})}, \quad (25)$$

where,  $SSE_p$  and  $SSR_p$  are given in Theorem 3. We are intending to find the point where adding more predictors is not worthwhile because it leads to a very small increase in  $R_p^2$  (see Fig. 2). Also, this index makes sense to use for comparing the submodels that are in the same unites.

### 5.3.2 Adjusted coefficient of multiple determination

Since the number of parameters in the IVF regression model is not taken into account by  $R_p^2$  ( $R_p^2$  does not decrease as  $p$  increases), the adjusted coefficient of multiple determination  $\overline{R}_p^2$  has been suggested as an alternative criterion.  $\overline{R}_p^2$  method is similar to  $R_p^2$  method and it finds the best model with the highest  $\overline{R}_p^2$  within the range of sizes. We define  $\overline{R}_p^2$  on an IVF regression model as follows (see Fig. 2).

$$\overline{R}_p^2 = 1 - (1 - R_p^2) \frac{n-1}{n-p}. \quad (26)$$

### 5.3.3 Mean square error

Another criterion for variable selection is the mean squares errors. The mean squares errors for the IVF regression model is defined as

$$MSE_p = \frac{SSE_p}{n-p}. \quad (27)$$



Because  $SSE_p$  always decreases as  $p$  increases,  $MSE_p$  initially decreases, then stabilizes, and eventually may increase (see Fig. 3). Hence, we choose the suitable subset of variables based on  $MSE_p$  as follows:

- (i) the minimum  $MSE_p$ ,
- (ii) the value of  $p$  such that  $MSE_p$  is approximately equal to  $MSE$  for the full model,
- (iii) a value of  $p$  near the point where the smallest  $MSE_p$  turns upward.

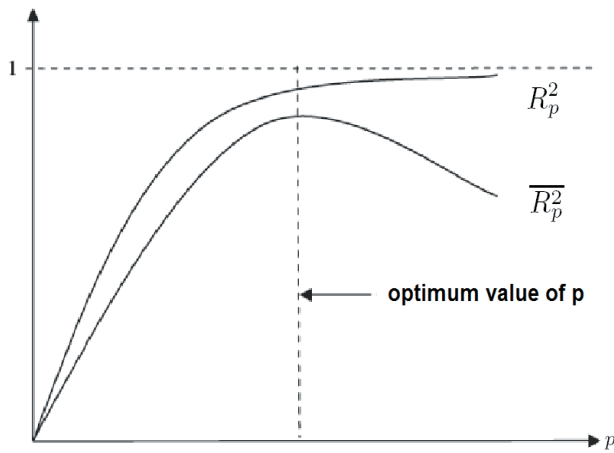


Fig. 2. Curve of  $R_p^2$  and  $\overline{R_p^2}$  in terms of  $p$ .

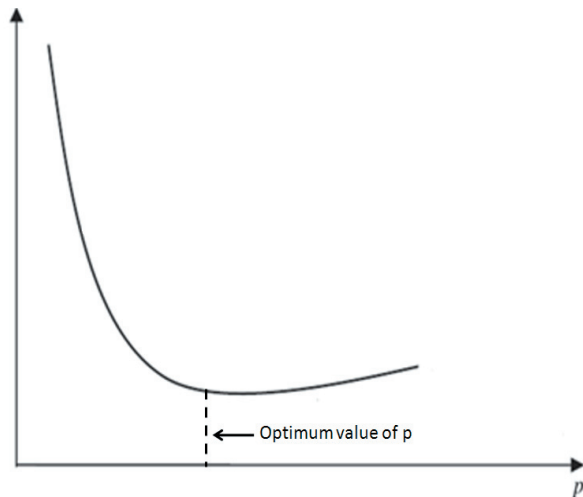


Fig. 3. Curve of  $MSE_p$  in term of  $p$ .

**Lemma 5.** The IVF regression submodel that minimizes  $MSE_p$ , will also maximize  $\overline{R_p^2}$ .

**Proof.** Based on  $\overline{R_p^2}$ , we have

$$\begin{aligned}\overline{R_p^2} &= 1 - \frac{n-1}{n-p}(1 - R_p^2) = 1 - \frac{n-1}{n-p} \frac{SSE_p}{SST} \\ &= 1 - \frac{n-1}{SST} \frac{SSE_p}{n-p} = 1 - \frac{n-1}{SST} MSE_p.\end{aligned}$$

Therefore, the proof is complete.  $\square$

Thus, the proposed results for selecting a submodel based on  $MSE_p$  and  $\overline{R_p^2}$  is similar.

## 6 Application examples

**Example 2.** The data in Table 1 show a coloration process in loom industrial (see Tavanai *et al.* <sup>52</sup>). The variables  $\tilde{x}_1$  and  $\tilde{x}_2$  are the color density(g/l) and the time of process(m), respectively, and the variable  $\tilde{y}$  is the value of color suction. Because of some impreciseness in experimental environment, the observed data are reported as triangular IVFN's. Based on these data, we want to model a relation between  $\tilde{y}$  (as the response variable) and  $\tilde{x}_1$  and  $\tilde{x}_2$  (as explanatory variables) as

$$\tilde{y}_i = \beta_0 \oplus (\beta_1 \otimes \tilde{x}_{i1}) \oplus (\beta_2 \otimes \tilde{x}_{i2}), \quad i = 1, \dots, 24$$

Using the matrix forms introduced in Section 4, we have

$$X = \begin{bmatrix} 1 & 0.75 & 24 \\ 1 & 1.50 & 24 \\ \vdots & \vdots & \vdots \\ 1 & 4.50 & 48 \end{bmatrix}, \quad Y = \begin{bmatrix} 1.014 \\ 1.104 \\ \vdots \\ 7.288 \end{bmatrix},$$

$$S = \begin{bmatrix} 0 & 0.19 & 6 \\ 0 & 0.38 & 6 \\ \vdots & \vdots & \vdots \\ 0 & 1.12 & 12 \end{bmatrix}, \quad R = \begin{bmatrix} 0.26 \\ 0.28 \\ \vdots \\ 1.81 \end{bmatrix}.$$

Since  $(X + \frac{S}{8})'(X + \frac{S}{8})$  is a nonsingular matrix, the parameters of model are estimated as

$$\begin{aligned}\hat{\beta} &= \left[ \left( X + \frac{S}{8} \right)' \left( X + \frac{S}{8} \right) \right]^{-1} \left[ \left( X + \frac{S}{8} \right)' \left( Y + \frac{R}{8} \right) \right] \\ &= \begin{bmatrix} 24.00 & 60.33 & 891.00 \\ 60.33 & 204.57 & 2239.75 \\ 891.00 & 2239.75 & 35528.63 \end{bmatrix}^{-1} \begin{bmatrix} 90.48 \\ 238.35 \\ 3481.93 \end{bmatrix} \\ &= \begin{bmatrix} 1.3905 \\ 0.2061 \\ 0.0501 \end{bmatrix}.\end{aligned}$$

Hence, the optimal model is obtained as

$$\tilde{y} = 1.3905 \oplus (0.2061 \otimes \tilde{x}_1) \oplus (0.0501 \otimes \tilde{x}_2).$$

For example, suppose that the value of color density and the time of process are reported as “approximately 3.15 g/l” and “approximately 30 m” with  $\tilde{x}_1 = (3.15; 0.32, 0.16, 0.63, 1.56)_T$  and  $\tilde{x}_2 = (30.00; 3.00, 1.50, 6.00, 15.00)_T$ , respectively. Then the value of color suction is predicted as

$$\begin{aligned}\tilde{y} &= 1.3905 \oplus (0.2061 \otimes (3.15; 0.32, 0.16, 0.63, 1.56)_T) \\ &\quad \oplus (0.0501 \otimes (30.00; 3.00, 1.50, 6.00, 15.00)_T) \\ &= (3.5427; 0.2163, 0.0859, 0.4304, 1.0730)_T.\end{aligned}$$

It means that the predicted value of color suction is “approximately 3.5427”. Therefore, for example, the value of color suction is 3.5427 with the degree of membership 1 and it is 3.40 with the degree of membership between 0.35 and 0.67.

The estimated values  $\hat{\tilde{y}}_i$  and observed values  $\tilde{y}_i$ ,  $i = 1, 2, \dots, 24$  of color suction are shown in Table 2. To evaluate the goodness of fit of proposed model, the estimated values assessed using  $S(\tilde{y}_i, \hat{\tilde{y}}_i)$  and  $d^2(\tilde{y}_i, \hat{\tilde{y}}_i)$ ,  $i = 1, 2, \dots, 24$ . Among 24 data points in Table 2, the data point with number 23 has the smallest similarity measure or the largest distance (see Table 3). It can be regarded as possible outlier. To investigate the effects of this outlier on model performance, it was removed and then a new model was fitted to the remained data as follows

$$\tilde{y} = 1.9839 \oplus (0.1686 \otimes \tilde{x}_1) \oplus (0.0329 \otimes \tilde{x}_2).$$

Note that the result of the model performance has been improved after removing outlier (see the averages of similarity measures and of square errors in Table 3).

For example the average value of  $d^2(\tilde{y}_i, \hat{\tilde{y}}_i)$  decrease from 5.1262 for the original model to 4.8962 after removing outliers.

Table 1. Some measured color characteristics in Example 2.

$i$	$(x_{i1}, s_{i11}, s_{i12}, s_{i13}, s_{i14})_T$	$(x_{i2}, r_{i21}, r_{i22}, r_{i23}, r_{i24})_T$	$(y_i, 0.1y_i, 0.05y_i, 0.2y_i, 0.5y_i)_T$
1	(0.75, 0.08, 0.04, 0.15, 0.38) <sub>T</sub>	(24.00, 2.40, 1.20, 4.80, 12.00) <sub>T</sub>	(1.014, 0.10, 0.05, 0.20, 0.51) <sub>T</sub>
2	(1.50, 0.15, 0.08, 0.30, 0.75) <sub>T</sub>	(24.00, 2.40, 1.20, 4.80, 12.00) <sub>T</sub>	(1.104, 0.11, 0.06, 0.22, 0.55) <sub>T</sub>
3	(3.00, 0.30, 0.15, 0.60, 1.50) <sub>T</sub>	(24.00, 2.40, 1.20, 4.80, 12.00) <sub>T</sub>	(1.148, 0.11, 0.06, 0.23, 0.57) <sub>T</sub>
4	(4.50, 0.45, 0.23, 0.90, 2.25) <sub>T</sub>	(24.00, 2.40, 1.20, 4.80, 12.00) <sub>T</sub>	(1.178, 0.12, 0.06, 0.24, 0.59) <sub>T</sub>
5	(0.75, 0.08, 0.04, 0.15, 0.38) <sub>T</sub>	(36.00, 3.60, 1.80, 7.20, 18.00) <sub>T</sub>	(1.421, 0.14, 0.07, 0.28, 0.71) <sub>T</sub>
6	(1.50, 0.15, 0.08, 0.30, 0.75) <sub>T</sub>	(36.00, 3.60, 1.80, 7.20, 18.00) <sub>T</sub>	(1.518, 0.15, 0.08, 0.30, 0.76) <sub>T</sub>
7	(3.00, 0.30, 0.15, 0.60, 1.50) <sub>T</sub>	(36.00, 3.60, 1.80, 7.20, 18.00) <sub>T</sub>	(1.651, 0.17, 0.08, 0.33, 0.83) <sub>T</sub>
8	(4.50, 0.45, 0.23, 0.90, 2.25) <sub>T</sub>	(36.00, 3.60, 1.80, 7.20, 18.00) <sub>T</sub>	(1.741, 0.17, 0.09, 0.35, 0.87) <sub>T</sub>
9	(0.75, 0.08, 0.04, 0.15, 0.38) <sub>T</sub>	(48.00, 4.80, 2.40, 9.60, 24.00) <sub>T</sub>	(1.610, 0.16, 0.08, 0.32, 0.81) <sub>T</sub>
10	(1.50, 0.15, 0.08, 0.30, 0.75) <sub>T</sub>	(48.00, 4.80, 2.40, 9.60, 24.00) <sub>T</sub>	(1.790, 0.18, 0.09, 0.36, 0.90) <sub>T</sub>
11	(3.00, 0.30, 0.15, 0.60, 1.50) <sub>T</sub>	(48.00, 4.80, 2.40, 9.60, 24.00) <sub>T</sub>	(1.928, 0.19, 0.10, 0.39, 0.96) <sub>T</sub>
12	(4.50, 0.45, 0.23, 0.90, 2.25) <sub>T</sub>	(48.00, 4.80, 2.40, 9.60, 24.00) <sub>T</sub>	(1.867, 0.19, 0.09, 0.37, 0.93) <sub>T</sub>
13	(0.75, 0.08, 0.04, 0.15, 0.38) <sub>T</sub>	(24.00, 2.40, 1.20, 4.80, 12.00) <sub>T</sub>	(4.459, 0.45, 0.22, 0.89, 2.23) <sub>T</sub>
14	(1.50, 0.15, 0.08, 0.30, 0.75) <sub>T</sub>	(24.00, 2.40, 1.20, 4.80, 12.00) <sub>T</sub>	(4.799, 0.48, 0.24, 0.96, 2.40) <sub>T</sub>
15	(3.00, 0.30, 0.15, 0.60, 1.50) <sub>T</sub>	(24.00, 2.40, 1.20, 4.80, 12.00) <sub>T</sub>	(5.023, 0.50, 0.25, 1.00, 2.51) <sub>T</sub>
16	(4.50, 0.45, 0.23, 0.90, 2.25) <sub>T</sub>	(24.00, 2.40, 1.20, 4.80, 12.00) <sub>T</sub>	(5.422, 0.54, 0.27, 1.08, 2.71) <sub>T</sub>
17	(0.75, 0.08, 0.04, 0.15, 0.38) <sub>T</sub>	(36.00, 3.60, 1.80, 7.20, 18.00) <sub>T</sub>	(5.797, 0.58, 0.29, 1.16, 2.90) <sub>T</sub>
18	(1.50, 0.15, 0.08, 0.30, 0.75) <sub>T</sub>	(36.00, 3.60, 1.80, 7.20, 18.00) <sub>T</sub>	(4.974, 0.50, 0.25, 0.99, 2.49) <sub>T</sub>
19	(3.00, 0.30, 0.15, 0.60, 1.50) <sub>T</sub>	(36.00, 3.60, 1.80, 7.20, 18.00) <sub>T</sub>	(6.025, 0.60, 0.30, 1.21, 3.01) <sub>T</sub>
20	(4.50, 0.45, 0.23, 0.90, 2.25) <sub>T</sub>	(36.00, 3.60, 1.80, 7.20, 18.00) <sub>T</sub>	(6.687, 0.67, 0.33, 1.34, 3.34) <sub>T</sub>
21	(0.75, 0.08, 0.04, 0.15, 0.38) <sub>T</sub>	(48.00, 4.80, 2.40, 9.60, 24.00) <sub>T</sub>	(5.268, 0.53, 0.26, 1.05, 2.63) <sub>T</sub>
22	(1.50, 0.15, 0.08, 0.30, 0.75) <sub>T</sub>	(48.00, 4.80, 2.40, 9.60, 24.00) <sub>T</sub>	(6.702, 0.67, 0.34, 1.34, 3.35) <sub>T</sub>
23	(3.00, 0.30, 0.15, 0.60, 1.50) <sub>T</sub>	(48.00, 4.80, 2.40, 9.60, 24.00) <sub>T</sub>	(7.325, 0.73, 0.37, 1.47, 3.66) <sub>T</sub>
24	(4.50, 0.45, 0.23, 0.90, 2.25) <sub>T</sub>	(48.00, 4.80, 2.40, 9.60, 24.00) <sub>T</sub>	(7.288, 0.73, 0.36, 1.46, 3.64) <sub>T</sub>

Table 2. The estimated values and observed values of color suction in Example 2.

$i$	$\tilde{y}_i = (y_i; r_{i1}, r_{i2}, r_{i3}, r_{i4})_T$	$\hat{\tilde{y}}_i = (\hat{y}_i; \hat{s}_{i1}, \hat{s}_{i2}, \hat{s}_{i3}, \hat{s}_{i4})_T$
1	(1.014, 0.10, 0.05, 0.20, 0.51) <sub>T</sub>	(2.7484, 0.14, 0.07, 0.27, 0.68) <sub>T</sub>
2	(1.104, 0.11, 0.06, 0.22, 0.55) <sub>T</sub>	(2.9030, 0.15, 0.08, 0.30, 0.76) <sub>T</sub>
3	(1.148, 0.11, 0.06, 0.23, 0.57) <sub>T</sub>	(3.2122, 0.18, 0.09, 0.36, 0.91) <sub>T</sub>
4	(1.178, 0.12, 0.06, 0.24, 0.59) <sub>T</sub>	(3.5213, 0.21, 0.11, 0.43, 1.07) <sub>T</sub>
5	(1.421, 0.14, 0.07, 0.28, 0.71) <sub>T</sub>	(3.3501, 0.20, 0.10, 0.39, 0.98) <sub>T</sub>
6	(1.518, 0.15, 0.08, 0.30, 0.76) <sub>T</sub>	(3.5047, 0.21, 0.11, 0.42, 1.06) <sub>T</sub>
7	(1.651, 0.17, 0.08, 0.33, 0.83) <sub>T</sub>	(3.8138, 0.24, 0.12, 0.48, 1.21) <sub>T</sub>
8	(1.741, 0.17, 0.09, 0.35, 0.87) <sub>T</sub>	(4.1230, 0.27, 0.14, 0.55, 1.37) <sub>T</sub>
9	(1.610, 0.16, 0.08, 0.32, 0.81) <sub>T</sub>	(3.9517, 0.26, 0.13, 0.51, 1.28) <sub>T</sub>
10	(1.790, 0.18, 0.09, 0.36, 0.90) <sub>T</sub>	(4.1063, 0.27, 0.14, 0.54, 1.36) <sub>T</sub>
11	(1.928, 0.19, 0.10, 0.39, 0.96) <sub>T</sub>	(4.4155, 0.30, 0.15, 0.61, 1.51) <sub>T</sub>
12	(1.867, 0.19, 0.09, 0.37, 0.93) <sub>T</sub>	(4.7246, 0.33, 0.17, 0.67, 1.67) <sub>T</sub>
13	(4.459, 0.45, 0.22, 0.89, 2.23) <sub>T</sub>	(2.7484, 0.14, 0.07, 0.27, 0.68) <sub>T</sub>
14	(4.799, 0.48, 0.24, 0.96, 2.40) <sub>T</sub>	(2.9030, 0.15, 0.08, 0.30, 0.76) <sub>T</sub>
15	(5.023, 0.50, 0.25, 1.00, 2.51) <sub>T</sub>	(3.2122, 0.18, 0.10, 0.34, 0.91) <sub>T</sub>
16	(5.422, 0.54, 0.27, 1.08, 2.71) <sub>T</sub>	(3.5213, 0.21, 0.11, 0.43, 1.07) <sub>T</sub>
17	(5.797, 0.58, 0.29, 1.16, 2.90) <sub>T</sub>	(3.3501, 0.20, 0.10, 0.39, 0.98) <sub>T</sub>
18	(4.974, 0.50, 0.25, 0.99, 2.49) <sub>T</sub>	(3.5047, 0.21, 0.11, 0.42, 1.06) <sub>T</sub>
19	(6.025, 0.60, 0.30, 1.21, 3.01) <sub>T</sub>	(3.8138, 0.24, 0.12, 0.48, 1.21) <sub>T</sub>
20	(6.687, 0.67, 0.33, 1.34, 3.34) <sub>T</sub>	(4.1230, 0.27, 0.14, 0.55, 1.37) <sub>T</sub>
21	(5.268, 0.53, 0.26, 1.05, 2.63) <sub>T</sub>	(3.9517, 0.26, 0.13, 0.51, 1.28) <sub>T</sub>
22	(6.702, 0.67, 0.34, 1.34, 3.35) <sub>T</sub>	(4.1063, 0.27, 0.14, 0.54, 1.36) <sub>T</sub>
23	(7.325, 0.73, 0.37, 1.47, 3.66) <sub>T</sub>	(4.4155, 0.30, 0.15, 0.61, 1.51) <sub>T</sub>
24	(7.288, 0.73, 0.36, 1.46, 3.64) <sub>T</sub>	(4.7246, 0.33, 0.17, 0.67, 1.67) <sub>T</sub>

Table 3. Goodness of fit for the color characteristics in Example 2.

$i$	$S(\tilde{y}_i, \hat{\tilde{y}}_i)$	$S(\tilde{y}_i, \hat{\tilde{y}}_i)^*$	$d^2(\tilde{y}_i, \hat{\tilde{y}}_i)$	$d^2(\tilde{y}_i, \hat{\tilde{y}}_i)^*$
1	0.7111	0.7179	3.0430	3.5381
2	0.6997	0.7168	3.2812	3.6823
3	0.6542	0.6887	4.3466	4.5556
4	0.6140	0.6530	5.6337	5.5936
5	0.6759	0.7187	3.7845	3.4897
6	0.6671	0.7173	4.0164	3.6015
7	0.6400	0.7093	4.7843	4.0924
8	0.6080	0.6799	5.8190	4.7957
9	0.6126	0.6976	5.6256	4.3291
10	0.6176	0.7092	5.4993	4.1000
11	0.5939	0.6878	6.3605	4.5959
12	0.5488	0.6364	8.4344	6.0910
13	0.5351	0.5224	3.2646	2.7898
14	0.5160	0.5016	3.9944	3.5761
15	0.4669	0.5048	3.6534	3.4665
16	0.5617	0.4913	4.0168	4.0508
17	0.4868	0.4417	6.5888	6.9917
18	0.4578	0.5228	2.4317	2.7747
19	0.4868	0.4530	5.3961	6.1866
20	0.4578	0.4240	7.2159	8.4599
21	0.5809	0.5208	1.9598	2.8598
22	0.4551	0.4156	7.4022	9.2471
23	0.4322	—	9.2644	—
24	0.4579	0.4107	7.2121	9.7453
Mean	0.5681	0.5888	5.1262	4.8962

\*: After removing outliers

**Example 3.** The amount and status of water in soil is described by different constants like SP, which shows soils saturated by water. Both mineral and organic colloids; i.e., percentage of sand and soil organic matter, can increase water capacity of soils (see Donahue *et al.*<sup>53</sup>). These different properties were measured using standard procedures (see Mohammadi and Taheri<sup>14</sup>). But due to some impreciseness in related experimental environment, the observed data were reported as IVFN's (see Table 4). Based on these data, we wish to model a relation between SP (as the response variable) and the silt of sand and soil (SILT), percentage of sand content (SAND), and organic matter content (OM) (as the exploratory variables) as follows

$$\tilde{y}_i = \beta_0 \oplus (\beta_1 \otimes \tilde{x}_{i1}) \oplus (\beta_2 \otimes \tilde{x}_{i2}) \oplus (\beta_3 \otimes \tilde{x}_{i3}), \quad i = 1, \dots, 24$$

Table 4. Some measured soil protection in Example 3.

$i$	OM	SAND
1	(0.88, 0.09, 0.04, 0.18, 0.44) <sub>T</sub>	(35, 3.50, 1.75, 7.40, 17.50) <sub>T</sub>
2	(1.13, 0.11, 0.06, 0.22, 0.57) <sub>T</sub>	(37, 3.70, 1.85, 7.40, 18.50) <sub>T</sub>
3	(1.31, 0.13, 0.07, 0.26, 0.66) <sub>T</sub>	(27, 2.70, 1.35, 5.40, 13.50) <sub>T</sub>
4	(1.98, 0.20, 0.10, 0.37, 0.99) <sub>T</sub>	(29, 2.90, 1.45, 5.80, 14.50) <sub>T</sub>
5	(1.02, 0.10, 0.05, 0.20, 0.51) <sub>T</sub>	(38, 3.80, 1.90, 7.60, 19.00) <sub>T</sub>
6	(1.29, 0.13, 0.06, 0.26, 0.65) <sub>T</sub>	(32, 3.20, 1.60, 6.40, 16.00) <sub>T</sub>
7	(1.52, 0.15, 0.08, 0.30, 0.76) <sub>T</sub>	(29, 2.90, 1.45, 5.80, 14.50) <sub>T</sub>
8	(1.33, 0.13, 0.07, 0.27, 0.67) <sub>T</sub>	(18, 1.80, 0.90, 3.60, 9.00) <sub>T</sub>
9	(1.71, 0.17, 0.09, 0.34, 0.86) <sub>T</sub>	(40, 4.00, 2.00, 8.00, 20.00) <sub>T</sub>
10	(2.00, 0.20, 0.10, 0.40, 1.00) <sub>T</sub>	(28, 2.80, 1.40, 5.60, 14.00) <sub>T</sub>
11	(1.68, 0.17, 0.08, 0.33, 0.84) <sub>T</sub>	(13, 1.30, 0.65, 2.60, 6.50) <sub>T</sub>
12	(2.15, 0.22, 0.11, 0.43, 1.08) <sub>T</sub>	(19, 1.90, 0.95, 3.80, 9.50) <sub>T</sub>
13	(3.52, 0.35, 0.18, 0.70, 1.76) <sub>T</sub>	(31, 3.10, 1.55, 6.20, 15.50) <sub>T</sub>
14	(2.33, 0.23, 0.12, 0.47, 1.17) <sub>T</sub>	(31, 3.10, 1.55, 6.20, 15.50) <sub>T</sub>
15	(1.71, 0.17, 0.09, 0.34, 0.86) <sub>T</sub>	(17, 1.70, 0.85, 3.40, 8.50) <sub>T</sub>
16	(1.14, 0.11, 0.06, 0.23, 0.57) <sub>T</sub>	(14, 1.40, 0.70, 2.80, 7.00) <sub>T</sub>
17	(0.99, 0.10, 0.05, 0.20, 0.50) <sub>T</sub>	(19, 1.90, 0.95, 3.80, 9.50) <sub>T</sub>
18	(1.14, 0.11, 0.06, 0.23, 0.57) <sub>T</sub>	(28, 2.80, 1.40, 5.60, 14.00) <sub>T</sub>
19	(1.46, 0.15, 0.07, 0.29, 0.73) <sub>T</sub>	(26, 2.60, 1.30, 5.20, 13.00) <sub>T</sub>
20	(1.81, 0.18, 0.09, 0.36, 0.91) <sub>T</sub>	(32, 3.20, 1.60, 6.40, 16.00) <sub>T</sub>
21	(1.38, 0.14, 0.07, 0.28, 0.69) <sub>T</sub>	(10, 1.00, 0.50, 2.00, 5.00) <sub>T</sub>
22	(0.84, 0.08, 0.04, 0.17, 0.42) <sub>T</sub>	(38, 3.80, 1.90, 7.60, 19.00) <sub>T</sub>
23	(1.48, 0.15, 0.07, 0.30, 0.74) <sub>T</sub>	(49, 4.90, 2.45, 9.80, 24.50) <sub>T</sub>
24	(1.08, 0.11, 0.05, 0.22, 0.54) <sub>T</sub>	(42, 4.20, 2.10, 8.40, 21.00) <sub>T</sub>

"Table 4. (Continued)"

$i$	SILT	SP
1	(45, 4.50, 2.25, 9.00, 22.50) <sub>T</sub>	(38, 3.80, 1.90, 7.60, 19.00) <sub>T</sub>
2	(42, 4.20, 2.10, 8.40, 21.00) <sub>T</sub>	(41, 4.10, 2.05, 8.20, 20.50) <sub>T</sub>
3	(43, 4.30, 2.15, 8.60, 21.50) <sub>T</sub>	(47.5, 4.75, 2.38, 9.50, 23.75) <sub>T</sub>
4	(41, 4.10, 2.05, 8.20, 20.50) <sub>T</sub>	(51, 5.10, 2.55, 10.20, 25.50) <sub>T</sub>
5	(39, 3.90, 1.95, 7.80, 19.50) <sub>T</sub>	(35, 3.50, 1.75, 7.00, 17.50) <sub>T</sub>
6	(39, 3.90, 1.95, 7.80, 19.50) <sub>T</sub>	(43, 4.30, 2.15, 8.60, 21.50) <sub>T</sub>
7	(37, 3.70, 1.85, 7.40, 18.50) <sub>T</sub>	(54, 5.40, 2.70, 10.80, 27.00) <sub>T</sub>
8	(45, 4.50, 2.25, 9.00, 22.50) <sub>T</sub>	(52, 5.20, 2.60, 10.40, 26.00) <sub>T</sub>
9	(38, 3.80, 1.90, 7.60, 19.00) <sub>T</sub>	(45, 4.50, 2.25, 9.00, 22.50) <sub>T</sub>
10	(46, 4.60, 2.30, 9.20, 23.00) <sub>T</sub>	(50, 5.20, 2.50, 10.00, 20.00) <sub>T</sub>
11	(40, 4.00, 2.00, 8.00, 20.00) <sub>T</sub>	(58.6, 5.86, 2.93, 11.72, 29.3) <sub>T</sub>
12	(41, 4.10, 2.05, 8.20, 20.50) <sub>T</sub>	(62, 6.20, 3.10, 12.40, 31.00) <sub>T</sub>
13	(41, 4.10, 2.05, 8.20, 20.50) <sub>T</sub>	(60, 6.00, 3.00, 12.00, 30.00) <sub>T</sub>
14	(42, 4.20, 2.10, 8.40, 21.00) <sub>T</sub>	(52, 5.20, 2.60, 10.40, 26.00) <sub>T</sub>
15	(50, 5.20, 2.50, 10.00, 20.00) <sub>T</sub>	(52, 5.20, 2.60, 10.40, 26.00) <sub>T</sub>
16	(53, 5.30, 2.65, 10.60, 26.50) <sub>T</sub>	(49, 4.90, 2.45, 9.80, 24.50) <sub>T</sub>
17	(44, 4.40, 2.20, 8.80, 22.00) <sub>T</sub>	(49, 4.90, 2.45, 9.80, 24.50) <sub>T</sub>
18	(43, 4.30, 2.15, 8.60, 21.50) <sub>T</sub>	(44, 4.40, 2.20, 8.80, 22.00) <sub>T</sub>
19	(44, 4.40, 2.20, 8.80, 22.00) <sub>T</sub>	(49, 4.90, 2.45, 9.80, 24.50) <sub>T</sub>
20	(42, 4.20, 2.10, 8.40, 21.00) <sub>T</sub>	(50.3, 5.03, 2.52, 10.06, 25.15) <sub>T</sub>
21	(49, 4.90, 2.45, 9.80, 24.50) <sub>T</sub>	(52, 5.20, 2.60, 10.40, 26.00) <sub>T</sub>
22	(43, 4.30, 2.15, 8.60, 21.50) <sub>T</sub>	(42, 4.20, 2.10, 8.40, 21.00) <sub>T</sub>
23	(35, 3.50, 1.75, 7.00, 17.50) <sub>T</sub>	(40, 4.00, 2.00, 8.00, 20.00) <sub>T</sub>
24	(44, 4.40, 2.20, 8.80, 22.00) <sub>T</sub>	(37, 3.70, 1.85, 7.40, 18.50) <sub>T</sub>

Based on Section 4, and using the matrix forms, we

have

$$X = \begin{bmatrix} 1 & 0.88 & 35 & 45 \\ 1 & 1.13 & 37 & 42 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1.08 & 42 & 44 \end{bmatrix}, \quad Y = \begin{bmatrix} 38 \\ 41 \\ \vdots \\ 37 \end{bmatrix},$$

$$S = \begin{bmatrix} 0 & 0.21 & 8.75 & 11.25 \\ 0 & 0.30 & 9.25 & 10.50 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0.26 & 10.50 & 11.00 \end{bmatrix}, \quad R = \begin{bmatrix} 9.50 \\ 10.25 \\ \vdots \\ 9.25 \end{bmatrix}.$$

The parameters of IVF regression model are estimated as

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \left[ \left( X + \frac{S}{8} \right)' \left( X + \frac{S}{8} \right) \right]^{-1} \left[ \left( X + \frac{S}{8} \right)' \left( Y + \frac{R}{8} \right) \right]$$

$$= \begin{bmatrix} 24.00 & 38.03 & 703.31 & 1058.06 \\ 38.03 & 68.64162 & 1102.96 & 1666.56 \\ 703.31 & 1102.96 & 23047.66 & 30393.10 \\ 1058.06 & 1666.56 & 30393.10 & 47054.58 \end{bmatrix}^{-1} \begin{bmatrix} 1188.31 \\ 1953.65 \\ 33662.86 \\ 52467.25 \end{bmatrix}$$

$$= \begin{bmatrix} 75.4846 \\ 7.0818 \\ -0.5608 \\ -0.4709 \end{bmatrix}.$$

Hence, the optimal IVF regression model is obtained as follows

$$\tilde{y} = 75.4846 \oplus (7.0818 \otimes \tilde{x}_1) \oplus (-0.5608 \otimes \tilde{x}_2) \oplus (-0.4709 \otimes \tilde{x}_3).$$

The estimated values  $\hat{\tilde{y}}_i$  and observed values  $\tilde{y}_i$ ,  $i = 1, 2, \dots, 24$  of soil protection listed in Table 5.

Table 5. The estimated values and observed values of soil protection in Example 3.

$i$	$\tilde{y}_i = (y_i; r_{i1}, r_{i2}, r_{i3}, r_{i4})_T$	$\hat{\tilde{y}}_i = (\hat{y}_i; \hat{r}_{i1}, \hat{r}_{i2}, \hat{r}_{i3}, \hat{r}_{i4})_T$
1	(38, 3.80, 1.90, 7.60, 19.00) <sub>T</sub>	(40.90; 2.68, 4.36, 21.68, 11.28) <sub>T</sub>
2	(41, 4.10, 2.05, 8.20, 20.50) <sub>T</sub>	(42.96; 2.80, 4.48, 21.88, 12.07) <sub>T</sub>
3	(47.5, 4.75, 2.38, 9.50, 23.75) <sub>T</sub>	(49.37; 2.69, 4.03, 19.53, 11.75) <sub>T</sub>
4	(51, 5.10, 2.55, 10.2, 25.5) <sub>T</sub>	(53.94; 3.19, 4.26, 20.61, 14.12) <sub>T</sub>
5	(35, 3.50, 1.75, 7.00, 17.50) <sub>T</sub>	(43.03; 2.69, 4.32, 21.25, 11.54) <sub>T</sub>
6	(43, 4.30, 2.15, 8.60, 21.50) <sub>T</sub>	(48.31; 2.74, 4.05, 19.99, 11.89) <sub>T</sub>
7	(54, 5.40, 2.70, 10.80, 27.00) <sub>T</sub>	(52.56; 2.75, 3.93, 18.96, 12.12) <sub>T</sub>
8	(52, 5.20, 2.60, 10.40, 26.00) <sub>T</sub>	(53.62; 2.48, 3.62, 17.55, 11.00) <sub>T</sub>
9	(45, 4.50, 2.25, 9.00, 22.50) <sub>T</sub>	(47.27; 3.22, 4.67, 22.56, 14.08) <sub>T</sub>
10	(50, 5.20, 2.50, 10.00, 20.00) <sub>T</sub>	(52.28; 3.28, 4.44, 21.51, 14.55) <sub>T</sub>
11	(58.6, 5.86, 2.93, 11.72, 29.3) <sub>T</sub>	(61.25; 2.51, 3.18, 15.46, 11.17) <sub>T</sub>
12	(62, 6.20, 3.10, 12.40, 31.00) <sub>T</sub>	(60.75; 2.98, 3.77, 18.02, 13.57) <sub>T</sub>
13	(60, 6.00, 3.00, 12.00, 30.00) <sub>T</sub>	(63.72; 4.31, 4.94, 23.30, 19.80) <sub>T</sub>
14	(52, 5.20, 2.60, 10.40, 26.00) <sub>T</sub>	(54.82; 3.49, 4.56, 21.90, 15.72) <sub>T</sub>
15	(52, 5.20, 2.60, 10.40, 26.00) <sub>T</sub>	(54.52; 2.86, 3.94, 18.94, 12.63) <sub>T</sub>
16	(49, 4.90, 2.45, 9.80, 24.50) <sub>T</sub>	(50.75; 2.42, 3.70, 18.02, 10.59) <sub>T</sub>
17	(49, 4.90, 2.45, 9.80, 24.50) <sub>T</sub>	(51.12; 2.28, 3.49, 17.10, 9.74) <sub>T</sub>
18	(44, 4.40, 2.20, 8.80, 22.00) <sub>T</sub>	(47.61; 2.58, 4.02, 19.60, 11.22) <sub>T</sub>
19	(49, 4.90, 2.45, 9.80, 24.50) <sub>T</sub>	(50.52; 2.83, 4.02, 19.70, 12.23) <sub>T</sub>
20	(50.3, 5.03, 2.52, 10.06, 25.15) <sub>T</sub>	(50.58; 3.16, 4.41, 21.40, 13.99) <sub>T</sub>
21	(52, 5.20, 2.60, 10.40, 26.00) <sub>T</sub>	(56.58; 2.42, 3.36, 16.32, 10.62) <sub>T</sub>
22	(42, 4.20, 2.10, 8.40, 21.00) <sub>T</sub>	(39.87; 2.64, 4.44, 21.97, 11.28) <sub>T</sub>
23	(40, 4.00, 2.00, 8.00, 20.00) <sub>T</sub>	(42.00; 3.26, 4.89, 24.10, 14.03) <sub>T</sub>
24	(37, 3.70, 1.85, 7.40, 18.50) <sub>T</sub>	(38.86; 2.99, 4.78, 23.69, 12.68) <sub>T</sub>

For evaluating the goodness of fit of IVF regression model, the similarity measures  $S(\tilde{y}_i, \hat{\tilde{y}}_i)$ , and the distance  $d^2(\tilde{y}_i, \hat{\tilde{y}}_i)$  are obtained in Table 6.

Among 24 data points, the values of similarity measures are approximately similar, but the values of index  $d^2(.,.)$  for data points with numbers 5, 7, 12, and 22 are more than others. They can be regarded as possible outliers. To investigate the effects of possible outliers on model performance, they were removed and a new model was fitted to the remained data as follows

$$\tilde{y} = 69.5331 \oplus (6.9213 \otimes \tilde{x}_1) \oplus (-0.5123 \otimes \tilde{x}_2) \oplus (-0.3699 \otimes \tilde{x}_3).$$

The updated results of  $d^2(\tilde{y}_i, \hat{\tilde{y}}_i)$  are given in column 4 of Table 6. As seen in this table, the IVF regression model is improved after removing outliers. Particularly, the average value of  $d^2(.,.)$  decreases from 4.6467 for the original model to 1.1216 after removing outliers.

Table 6. Goodness of fit for the soil protection in Example 3.

$i$	$S(\tilde{y}_i, \hat{\tilde{y}}_i)$	$d^2(\tilde{y}_i, \hat{\tilde{y}}_i)$	$d^2(\tilde{y}_i, \hat{\tilde{y}}_i)^*$
1	0.0624	0.3853	0.9241
2	0.0596	0.1159	0.0645
3	0.0619	0.0096	0.0584
4	0.0535	0.4420	0.0114
5	0.0582	35.1506	---
6	0.0575	9.6424	6.7753
7	0.0574	14.6891	---
8	0.0574	0.4786	1.6281
9	0.0621	0.0001	0.0815
10	0.0602	0.0000	0.0088
11	0.0593	9.799	0.1788
12	0.0578	13.3670	---
13	0.0528	2.2112	0.5983
14	0.0449	0.3119	0.0214
15	0.0521	0.0610	0.0050
16	0.0646	0.3029	0.2209
17	0.0612	0.0287	0.5649
18	0.1088	1.8632	1.2184
19	0.0689	0.6273	1.2139
20	0.0582	4.2982	5.5191
21	0.0615	5.5486	3.1675
22	0.0637	20.7148	---
23	0.0671	0.0904	0.1609
24	0.0606	0.2027	0.0106
Mean	0.0616	4.6467	1.1216

\*:After removing outliers

In an IVF regression model, we have  $2^k$  submodels for each subset of  $p \in \{0, 1, \dots, k\}$  explanatory variables. Based on variables proposed in Example 3, we have presented  $2^3$  IVF regression submodels and also, some indices of variable selection ( $R^2$ ,  $\bar{R}^2$  and  $MSE_p$ ) in Table 7. Since  $\bar{R}^2$  in full model with  $p = 4$  regressors has the highest values of  $\bar{R}^2$ , the full regression model is the best selection between presented IVF regression submodels. Also, based on  $MSE_p$ , the result of variable selection is similar. (i.e.,  $MSE_p$  in the full regression model is the minimum value).

Table 7. Variable selection for the soil protection in Example 3.

Number of regressors	$p$	Regressors in model	$SSE_p$	$R_p^2$	$\bar{R}_p^2$	$MSE_p$
None	1	None	1224.3660	0	0	53.2333
1	2	$x_1$	629.8066	0.4856	0.4622	28.6276
1	2	$x_2$	672.1173	0.4510	0.4261	30.5508
1	2	$x_3$	1208.9010	0.0126	-0.0322	54.9501
2	3	$x_1 x_2$	163.4566	0.8665	0.8538	7.7836
2	3	$x_1 x_3$	561.2320	0.5416	0.4980	26.7253
2	3	$x_2 x_3$	495.1245	0.5956	0.5571	23.5774
3	4	$x_1 x_2 x_3$	111.5211	0.9089	0.8953	5.5761

## 7 Conclusion

The new approach proposed in this paper for estimating the parameters a regression model in interval-valued fuzzy environment, has certain merits as follows:

- I) It is established upon a new signed distance, which is an extended version of the Yao-Wu signed distance<sup>46</sup>.
- II) The available data of both explanatory variable(s) and response variable are triangular interval-valued fuzzy numbers.
- III) To evaluate the proposed regression model, we introduce some new indices on the basis of the concepts of the similarity measure, the square errors, and (adjusted) coefficient of multiple determination.

The extension of results for modelling a multivariate IVF regression model using the least absolute approach and also, testing parameters of proposed IVF regression model, can be investigated in the future researches.

## References

1. L. A. Zadeh, "Fuzzy sets," *Information and Control*, **8**, 338–356 (1965).
2. H. Tanaka, S. Uejima, and K. Asai, "Fuzzy linear regression model," *IEEE Trans System Man Cybernet*, **12**, 903–907 (1982).
3. H. Tanaka, S. Vegima, and K. Asai, "Linear regression analysis with fuzzy model," *IEEE Trans Systems Man Cybernet* **12**, 903–907 (1982).
4. A. Celmins, "Least squares model fitting to fuzzy vector data," *Fuzzy Sets and Systems*, **22**, 260–269 (1987).
5. P. Diamond, "Least squares of fitting several fuzzy variables," *Proceeding of the Second IFSA Congress*, Tokyo, 20–25 (1987).

6. K. K. Yen, S. Ghoshray, and G. Roig, "A linear regression model using triangular fuzzy number coefficients," *Fuzzy Sets and Systems*, **106**, 167–177 (1999).
7. M. M. Nasrabadi and E. Nasrabadi, "A mathematical-programming approach to fuzzy linear regression analysis," *Applied Mathematics and Computation*, **155**, 873–881 (2004).
8. H. Hasanpour, H. R. Maleki, and M. A. Yaghoubi, "Fuzzy linear regression model with crisp coefficients: A goal programming approach," *Iranian Journal of Fuzzy Systems*, **7**, 19–39 (2010).
9. H. Hasanpour, H. R. Maleki, and M. A. Yaghoubi, "A goal programming approach to fuzzy linear regression with fuzzy input-output data," *Soft Computing*, **15**, 1569–1580 (2011).
10. A. Wünsche and W. Näther, "Least squares fuzzy regression with fuzzy random variables," *Fuzzy Set and Systems*, **130**, 43–50 (2002).
11. M. S. Yang and T. S. Lin, "Fuzzy least-squares linear regression analysis for fuzzy input-output data," *Fuzzy Sets and Systems*, **126**, 389–399 (2002).
12. H. Ch. Wu, "Linear regression analysis for fuzzy input and output data using the extension principle," *Computers and Mathematics with Applications*, **45**, 1849–1859 (2003).
13. Ch. Kao and Ch. L. Chyu, "Least square estimates in fuzzy regression analysis," *European Journal of Operational Research*, **148**, 426–435 (2003).
14. J. Mohammadi and S. M. Taheri, "Pedomodels fitting with fuzzy least squares regression," *Iranian Journal of Fuzzy System*, **1**, 45–62 (2004).
15. R. Coppi, P. D'Urso, and A. Santoro, "Least squares estimation of a linear regression model with LR fuzzy response," *Computational Statistics & Data Analysis*, **51**, 267–286 (2006).
16. A. R. Arabpour and M. Tata, "Estimating the parameters of a fuzzy linear regression model," *Iranian Journal of Fuzzy Systems*, **5**, 1–19 (2008).
17. S. H. Choi and J. H. Yoon, "General fuzzy regression using least squares method," *International Journal of Systems Science*, **41**, 477–485 (2010).
18. B. M. Ferraro and P. Giordani, "A multiple linear regression model for imprecise information," *Metrika*, **75**(8), 1049–1068 (2012).
19. H. Ch. Wu, "The construction of fuzzy least squares estimators in fuzzy linear regression models," *Expert Systems with Applications*, **38**, 13632–13640 (2011).
20. S. M. Taheri and M. Kelkinnama, "Fuzzy linear regression based on least absolute deviations," *Iranian Journal of Fuzzy Systems*, **9**, 121–140 (2012).
21. M. Kelkinnama and S. M. Taheri, "Fuzzy least-absolutes regression using shape preserving operations," *Information Sciences*, **214**, 105–120 (2012).
22. S. B. Roh, T. Ch. Ahn, and W. Pedrycz, "Fuzzy linear regression based on Polynomial Neural Networks," *Expert Systems with Applications*, **39**, 8909–8928 (2012).
23. S. M. Taheri, "Trends in fuzzy statistics," *Austrian Journal of Statistics*, **32**, 239–257 (2003).
24. W. An, C. Angulo, and Y. Sun, "Support Vector Regression with Interval-Input Interval-Output," *International Journal of Computational Intelligence Systems*, **1**(4), 299–303 (2008).
25. S. Sentürk, "Fuzzy Regression Control Chart Based on  $\alpha$ -cut Approximation," *International Journal of Computational Intelligence Systems*, **3**(1), 123–140 (2010).
26. P. P. Bonissone, "Soft computing: the convergence of emerging reasoning technologies," *Soft Computing*, **1**, 6–18 (1997).
27. L. Magdalena, "What is Soft Computing? Revisiting possible answers," *International Journal of Computational Intelligence Systems*, **3**(2), 148–159 (2010).
28. R. Seising, "What is Soft Computing? – Bridging gaps for 21st Century science!," *International Journal of Computational Intelligence Systems*, **3**(2), 160–175 (2010).
29. J. Liu, L. Martinez, H. Wang, R. M. Rodriguez, and V. Novozhilov, "Computing with words in risk assessment," *International Journal of Computational Intelligence Systems*, **3**(4), 396–419 (2010).
30. L. A. Zadeh, "From computing with numbers to computing with words - from manipulation of measurements to manipulation of perceptions," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, **45**, 105–119 (1999).
31. K. Atanassov, *Intuitionistic Fuzzy Sets: Theory and Applications*, Physica-Verlag Heidelberg, Germany (1999).
32. M. B. Gorzalczany, "An Interval-valued fuzzy inference method some basic properties," *Fuzzy Sets and Systems*, **31**, 243–251 (1989).
33. K. Atanassov and G. Gargov, "Interval valued intuitionistic fuzzy sets," *Fuzzy Sets and Systems*, **31**, 343–349 (1986).
34. E. Szmidt, J. Kacprzyk, "Intuitionistic fuzzy sets in group decision making," *Notes IFS*, **2** (1), 11–14 (1996).
35. K. Atanassov, G. Gargov, "Intuitionistic fuzzy logic," *CR Acad. Bulg. Soc.*, **43** (3), 9–12 (1990).
36. K. Atanassov, G. Gargov, "Intuitionistic fuzzy prolog," *Fuzzy Sets Systems*, **53**, 121–128 (1993).
37. S. K. De, R. Biswas, A. R. Roy, "An application of intuitionistic fuzzy set in medical diagnosis," *Fuzzy Sets Systems*, **117**, 209–213 (2001).
38. W. L. Hung and M. Sh. Yang, "Similarity measures of intuitionistic fuzzy sets based on  $L_p$  metric," *International Journal of Approximate Reasoning*, **46** 120–136 (2007).
39. W. L. Gau and D. J. Buehrer, "Vague sets," *IEEE Transactions on Systems Man and Cybernetics*, **23**, 610–614 (1993).
40. H. Bustince and P. Burillo, "Vague sets are intuitionistic fuzzy sets," *Fuzzy Sets and Systems*, **79**, 403–405 (1996).
41. B. Turksen, "Interval valued fuzzy sets based on normal



- forms," *Fuzzy Sets and Systems*, **20**, 191–210 (1986).
42. S. Kumar, R. Biswas, and A. R. Roy, "Some operations on intuitionistic fuzzy sets," *Fuzzy Set and Systems*, **114**, 477–484 (2000).
43. P. Grzegorzewski, "Distances between intuitionistic fuzzy sets and/or interval-valued fuzzy sets based on the Hausdorff metric," *Fuzzy Sets and Systems*, **148**, 319–328 (2004).
44. D. Guha and D. Chakraborty, "A theoretical development of distance measure for intuitionistic fuzzy numbers," *International Journal of Mathematics and Mathematical Sciences*, **2010**, Article ID 949143, 25 pages, Doi:10.1155/2010/949143 (2010).
45. S. M. Taheri and R. Zarei, "Extension principle for vague sets and its applications," *Advances in Fuzzy Mathematics*, **6**, 17–26 (2011).
46. J. SH. Yao and K. Wu, "Ranking fuzzy numbers based on decomposition principle and signed distance," *Fuzzy Sets and Systems*, **116**, 275–288 (2000).
47. D. F. Li, J. X. Nan, and M. J. Zhang, "A ranking method of triangular intuitionistic fuzzy numbers and application to decision making," *International Journal of Computational Intelligence Systems*, **3**(5), 522–530 (2010).
48. D. Li and C. Cheng, "New similarity measure of intuitionistic fuzzy sets and application to pattern recognition," *Pattern Recognition Letters*, **23**, 221–225 (2002).
49. D. Li and C. Cheng, "New similarity measure of intuitionistic fuzzy sets," *Pattern Recognition Letters*, **24**, 2687–2693 (2003).
50. Z. Liang and P. Shi, "Similarity measure on intuitionistic fuzzy sets," *Pattern Recognition Letters*, **24**, 2687–2693 (2003).
51. Q. Zhang, H. Xing, and L. Wu, "Entropy, Similarity Measure, Inclusion Measure of Intuitionistic Fuzzy Sets and Their Relationships," *International Journal of Computational Intelligence Systems*, **5**(3), 519–529 (2012).
52. H. Tavanai and S. M. Taheri and M. Nasiri, "Modelling of colour yield in polyethylene terephthalate dyeing with statistical and fuzzy regression," *Iranian Polymer Journal*, **14**, 954–967 (2005).
53. R. L. Donahue and R. W. Miller and J. C. Shickluna, *Soils, an introduction to soils and plant growth*, Prentice-Hall (1983).