

## A NOTE ON EVALUATION OF FUZZY LINEAR REGRESSION MODELS BY COMPARING MEMBERSHIP FUNCTIONS

H. HASSANPOUR, H. R. MALEKI AND M. A. YAGHOOBI

ABSTRACT. Kim and Bishu (Fuzzy Sets and Systems 100 (1998) 343-352) proposed a modification of fuzzy linear regression analysis. Their modification is based on a criterion of minimizing the difference of the fuzzy membership values between the observed and estimated fuzzy numbers. We show that their method often does not find acceptable fuzzy linear regression coefficients and to overcome this shortcoming, propose a modification. Finally, we present two numerical examples to illustrate efficiency of the modified method.

### 1. Introduction

Since Zadeh [20] introduced fuzzy set theory, it has been widely developed in theory and application (e.g. see [2], [3], [5], [15], [16], [17], [19]) and in particular, there has been much research in the area of fuzzy regression analysis (e.g. see [1],[4], [6], [7], [8], [9], [10], [12], [13], [14], [18]). Regression analysis is a methodology for analyzing phenomena in which a variable (output or response) depends on other variables called input (independent or explanatory) variables. A function is fitted to a set of given data, to predict the value of dependent variable for a specified value(s) of the independent variable(s). However, the phenomena in the real world cannot be analyzed exactly, because they depend on some uncertain factors and in some cases, it may be appropriate to use fuzzy regression analysis.

Kim and Bishu [9] proposed a model called the fuzzy membership function least-squares (FMLS) regression model. This paper shows that FMLS often does not work properly and a modification is suggested to improve it.

Suppose the data  $(x_{1j}, x_{2j}, \dots, x_{mj}, y_j), j = 1, \dots, n$ , are given. For each  $j$ ,  $x_{1j}, x_{2j}, \dots, x_{mj}$  are the values of  $m$  crisp input variables  $x_1, x_2, \dots, x_m$ , and  $y_j$  is the corresponding value of fuzzy output variable  $y$ . The purpose of fuzzy linear regression (FLR) is to fit a fuzzy linear relation to this data as follows [1], [4], [6], [7], [8], [9], [10], [14], [18]:

$$Y = A_0 + A_1x_1 + A_2x_2 + \dots + A_mx_m = \sum_{i=0}^m A_ix_i, \quad (1)$$

where  $x_0 = 1$ . In model (1),  $A_0, A_1, \dots, A_m$  are fuzzy numbers, called the coefficients (parameters) of the model. These coefficients must be estimated such that

---

Received: October 2007; Revised: March 2008; Accepted: December 2008

*Key words and phrases:* Fuzzy linear regression, Fuzzy number, Least-squares method.

This paper is supported in part by Fuzzy Systems and Applications Center of Excellence, Shahid Bahonar University of Kerman, Kerman, I.R. of Iran.

the estimated responses  $Y_j$ ,

$$Y_j = A_0 + A_1x_{1j} + A_2x_{2j} + \dots + A_mx_{mj}, \quad j = 1, \dots, n, \quad (2)$$

are the best fit to the observed responses  $y_j$ ,  $j = 1, \dots, n$ , with respect to a given criterion.

## 2. The FMLS Method

In this section, we briefly review the method of Kim and Bishu (FMLS) [9]. In [9], it is assumed that the coefficients of model (1) and the observed responses  $y_j$  are all triangular fuzzy numbers. In their paper, Kim and Bishu have attempted to bring the membership functions of the estimated responses as close as possible to those of the corresponding observed values. Before describing their method, we recall the following definitions.

**Definition 2.1.** A fuzzy number  $A$  is said to be an L-R fuzzy number if its membership function has the following form:

$$\mu_A(x) = \begin{cases} L\left(\frac{c-x}{\alpha}\right) & x \leq c \quad \alpha > 0 \\ R\left(\frac{x-c}{\beta}\right) & x \geq c \quad \beta > 0 \end{cases} \quad (3)$$

where  $c$  is the mean value of  $A$ ,  $\alpha$  and  $\beta$  are respectively its left and right spreads, and  $L(\cdot)$  is a left shape function satisfying

- (1)  $L(x) = L(-x)$
- (2)  $L(0) = 1$
- (3)  $L(x)$  is nondecreasing on  $[0, \infty)$ .

The right shape function  $R(\cdot)$  is defined similarly.

**Definition 2.2.** If, in Definition 2.1,  $L(x) = R(x) = 1 - |x|$  for  $0 \leq x \leq 1$  and 0 elsewhere, then we say that  $A$  is a triangular fuzzy number and denote it by  $A = (l, c, r)$ .  $c$  is called the center of  $A$ , and  $l = c - \alpha$  and  $r = c + \beta$  are respectively called the left and right end points. Also, if  $\alpha = \beta$ , then  $A$  is called a symmetric triangular fuzzy number. Clearly, the membership function of a triangular fuzzy number  $A$  is as follows:

$$\mu_A(x) = \begin{cases} \frac{x-l}{c-l} & l \leq x \leq c, \\ \frac{r-x}{r-c} & c \leq x \leq r, \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

Based on Definition 2.2, Kim and Bishu suggested a fuzzy membership function least-squares regression model. In their model, the regression function is defined as [9]:

$$\begin{aligned} Y &= (y - |L^{-1}(\alpha)|e, y, y + |L^{-1}(\alpha)|e) \\ &= (l_0, c_0, r_0) + (l_1, c_1, r_1)x_1 + \dots + (l_m, c_m, r_m)x_m \\ &= \left( \sum_{i=0}^m l_i x_i, \sum_{i=0}^m c_i x_i, \sum_{i=0}^m r_i x_i \right), \end{aligned} \quad (5)$$

where  $x_0 = 1$ ,  $e$  is the spread of the symmetric fuzzy response  $Y$ , and  $\alpha \in [0, 1]$ . The regression coefficients  $A_i = (l_i, c_i, r_i)$  are also (not necessarily symmetric) triangular fuzzy numbers. When minimizing the difference of the membership values of the observed and estimated fuzzy numbers, the regression model is decomposed into three ordinary least-squares regression models as follows [9]:

$$y - |L^{-1}(\alpha)|e = \sum_{i=0}^m l_i x_i, \tag{6}$$

$$y = \sum_{i=0}^m c_i x_i, \tag{7}$$

$$y + |L^{-1}(\alpha)|e = \sum_{i=0}^m r_i x_i. \tag{8}$$

The least-squares estimators for  $l_i$ ,  $c_i$  and  $r_i$  are obtained by solving the least-squares normal equations for (6), (7), and (8). In fact in [9], it is assumed that  $\alpha = 0$ .

### 3. Examples and a Modification

As mentioned above, FMLS has been studied by many authors. However, as the following example illustrates, it sometimes finds FLR coefficients  $A_i = (l_i, c_i, r_i)$  that are unacceptable.

**Example 3.1.** Consider the data of Table 1.

j	$x_j$	$y_j = (l_{y_j}, c_{y_j}, r_{y_j})$
1	2	(8, 10, 12)
2	4	(6, 8, 10)
3	6	(6, 7, 8)
4	8	(5, 6.5, 8)
5	10	(2, 4, 6)

Table 1. The given data

Taking  $\alpha = 0$  and  $L(x) = 1 - |x|$  (as in [9]), the least-squares estimators for the coefficients of model (5) are obtained as:

$$\begin{aligned} l_0 &= 9.3, & c_0 &= 11.15, & r_0 &= 13, \\ l_1 &= -0.65, & c_1 &= -0.675, & r_1 &= -0.7. \end{aligned}$$

Note that  $(l_0, c_0, r_0) = (9.3, 11.15, 13)$  is a triangular fuzzy number. However,  $(l_1, c_1, r_1) = (-0.65, -0.675, -0.7)$  is not a triangular fuzzy number, since it does not satisfy the conditions of Definition 2.2. In particular, we have the following inequalities:  $l_1 = -0.65 > c_1 = -0.675 > r_1 = -0.7$ .

At first, it may seem that such a problem may rarely appear in real world problems. However, the following simulation study refutes this idea.

1000 samples were generated, each containing 100  $x_j$  values, uniformly distributed on the interval (0,10). Also, for each  $j$ , the observed response  $y_j = (l_{y_j}, c_{y_j}, r_{y_j})$  was chosen as:

$$c_{y_j} = A_0 + A_1x_j + \varepsilon_j, \quad l_{y_j} = c_{y_j} - \alpha_j, \quad r_{y_j} = c_{y_j} + \beta_j, \quad (9)$$

where  $A_0 = 1$ ,  $A_1$  varies over 0.6, 0.8, 1, 1.2, 1.4, and  $\varepsilon_j$  is a standard normal variable. Also,  $(\alpha_j, \beta_j)$  corresponding to  $y_j$  was taken to be a random point in the unit square. The procedure was replicated 10 times i.e. 10000 random problems were generated and solved and it was found that for approximately 75% of the instances, FMLS obtained unacceptable FLR coefficients. This is certainly not negligible.

**3.1. A Modification.** Consider model (1) and the three least-squares models (6), (7) and (8). For a triangular fuzzy number it is necessary to have  $l_i \leq c_i \leq r_i$ . However, these inequalities are not guaranteed in the solution of the normal equations used in the FMLS method. To overcome this problem, we first obtain  $c_0, c_1, \dots, c_m$  by solving the usual least-squares normal equations for (7). Then to obtain the left-end points  $l_0, l_1, \dots, l_m$ , we add the constraints  $l_i \leq c_i \forall i$ , to (6) and to obtain the right-end points  $r_0, r_1, \dots, r_m$ , we add the constraints  $r_i \geq c_i \forall i$ , to (8). In other words, we first obtain  $c_0, c_1, \dots, c_m$ , as in [9]. Then, we obtain  $l_0, l_1, \dots, l_m$  and  $r_0, r_1, \dots, r_m$  as solutions to the following non-linear programming problems:

$$\begin{aligned} \min \quad & \sum_{j=1}^n \left( \sum_{i=0}^m l_i x_{ij} - (y_j - |L^{-1}(\alpha)|e_j) \right)^2 \\ \text{s.t.} \quad & l_i \leq c_i, \quad i = 0, 1, \dots, m, \end{aligned} \quad (10)$$

$$\begin{aligned} \min \quad & \sum_{j=1}^n \left( \sum_{i=0}^m r_i x_{ij} - (y_j + |L^{-1}(\alpha)|e_j) \right)^2 \\ \text{s.t.} \quad & r_i \geq c_i, \quad i = 0, 1, \dots, m. \end{aligned} \quad (11)$$

Indeed, (10) and (11) are classical least-squares problems with additional conditions.

**Example 3.2.** The FLR model for the data of Table 1 obtained using the modified method is as follows:

$$Y = A_0 + A_1x = (9.45, 11.15, 12.85) + (-0.675, -0.675, -0.675)x.$$

It has been suggested that one may replace negative spreads by zero. However, we note that although the spreads of  $A_1$  in the above example are zero, the coefficient  $A_0$  is not the same as that obtained by the FMLS method. Therefore, to

obtain the best acceptable minimizing solutions of least-squares models it is necessary to solve problems (10) and (11). Moreover, these problems can be solved easily, for example using MATLAB software [11].

When the FMLS method obtains acceptable coefficients, the coefficients obtained by the modified method and FMLS are identical. Indeed, in this case, since they minimize the objective functions of (10) and (11) as well as satisfy the additional constraints, the FLR coefficients of FMLS are the best solutions of problems (10) and (11). This matter is illustrated by the following example.

**Example 3.3.** Consider the data of Table 2 which have been used in [9].

j	$x_j$	$y_j = (l_{y_j}, c_{y_j}, r_{y_j})$
1	1	(6.2 , 8 , 9.8)
2	2	(4.2 , 6.4 , 8.6)
3	3	(6.9 , 9.5 , 12.1)
4	4	(10.9 , 13.5 , 16.1)
5	5	(10.6 , 13 , 15.4)

Table 2. The given data

Using the modified method, the FLR model is:

$$Y = A_0 + A_1x = (3.11, 4.95, 6.79) + (1.55, 1.71, 1.87)x.$$

which is the same as the FMLS model.

#### 4. Concluding Remarks

In this short note the method of Kim and Bishu [9] is briefly reviewed and it is shown by an example that their method may obtain unacceptable fuzzy linear regression coefficients. Also, by a simulation study it is shown that this shortcoming appears in about 75% of simulated samples, which is not negligible. To overcome this problem, a modification is proposed and the ability of the modified method to rectify the shortcoming is illustrated by a numerical example.

#### REFERENCES

- [1] A. Abdalla and J. J. Buckley, *Monte Carlo methods in fuzzy linear regression*, Soft Computing, **11** (2007), 991-996.
- [2] S. Abbasbandy and M. Alavi, *A method for solving fuzzy linear systems*, Iranian Journal of Fuzzy Systems, **2(2)** (2005), 37-44.
- [3] S. M. Abu Nayeem and M. Pal, *The p-center problem on fuzzy networks and reduction of cost*, Iranian Journal of Fuzzy Systems, **5(1)** (2008), 1-26.
- [4] A. R. Arabpour and M. Tata, *Estimating the parameters of a fuzzy linear regression model*, Iranian Journal of Fuzzy Systems, **5(2)** (2008), 1-19.
- [5] J. M. Cadenas and J. L. Verdegay, *A primer on fuzzy optimization models and methods*, Iranian Journal of Fuzzy Systems, **3(1)** (2006), 1-22.
- [6] S. H. Choi and J. J. Buckley, *Fuzzy regression using least absolute deviation estimators*, Soft Computing, **12(3)** (2008), 257-263.
- [7] C. Kao and C. L. Chyu, *A fuzzy linear regression model with better explanatory power*, Fuzzy Sets and Systems, **126** (2002), 401-409.

- [8] C. Kao and C. L. Chyu, *Least-squares estimates in fuzzy regression analysis*, European Journal of Operational Research, **148** (2003), 426-435.
- [9] B. Kim and R. R. Bishu, *Evaluation of fuzzy linear regression models by comparing membership functions*, Fuzzy Sets and Systems, **100** (1998), 343-352.
- [10] H. Lee and H. Tanaka, *Upper and lower approximation models in interval regression using regression quantile techniques*, European Journal of Operational Research, **116** (1999), 653-666.
- [11] MATLAB 7.0, The Mathworks, ([www.mathworks.com](http://www.mathworks.com)).
- [12] M. Modarres, E. Nasrabadi and M. M. Nasrabadi, *Fuzzy linear regression models with least square errors*, Applied Mathematics and Computation, **163** (2005), 977-989.
- [13] J. Mohammadi and S. M. Taheri, *Pedomodels fitting with fuzzy least squares regression*, Iranian Journal of Fuzzy Systems, **1(29)** (2004), 45-62.
- [14] M. Sakawa and H. Yano, *Fuzzy linear regression and its applications*, In: J. Kacprzyk and M. Fedrizzi (Eds.), *Studies in fuzziness, fuzzy regression analysis*, Omnitech Press, Warsaw, Poland, (1992), 61-80.
- [15] M. R. Safi, H. R. Maleki and E. Zaeimazad, *A note on Zimmermann method for solving fuzzy linear programming problems*, Iranian Journal of Fuzzy Systems, **4(2)** (2007), 31-46.
- [16] E. Shivanian, E. Khorram and A. Ghodousian, *Optimization of linear objective function subject to fuzzy relation inequalities constraints with Max-average composition*, Iranian Journal of Fuzzy Systems, **4(2)** (2007), 15-30.
- [17] S. M. Vaezpour and F. Karimi, *t-Best approximation in fuzzy normed spaces*, Iranian Journal of Fuzzy Systems, **5(2)** (2008), 93-99.
- [18] H. C. Wu, *Fuzzy linear regression model based on fuzzy scalar product*, Soft Computing, **12** (2008), 469-477.
- [19] M. A. Yaghoobi and M. Tamiz, *A Short note on the relationship between goal programming and fuzzy programming for vectormaximum problems*, Iranian Journal of Fuzzy Systems, **2(2)** (2005), 31-36.
- [20] L. A. Zadeh, *Fuzzy sets*, Information and Control, **8** (1965), 338-353.

H. HASSANPOUR, DEPARTMENT OF MATHEMATICS, UNIVERSITY OF BIRJAND, BIRJAND, IRAN  
*E-mail address:* [hassanhassanpur@yahoo.com](mailto:hassanhassanpur@yahoo.com)

H. R. MALEKI\*, FACULTY OF BASIC SCIENCES, SHIRAZ UNIVERSITY OF TECHNOLOGY, SHIRAZ, IRAN  
*E-mail address:* [maleki@sutech.ac.ir](mailto:maleki@sutech.ac.ir)

M. A. YAGHOOBI, DEPARTMENT OF STATISTICS, SHAHID BAHONAR UNIVERSITY OF KERMAN, KERMAN, IRAN  
*E-mail address:* [Yaghoobi@mail.uk.ac.ir](mailto:Yaghoobi@mail.uk.ac.ir)

\*CORRESPONDING AUTHOR