On relationship among h value, membership function, and spread in fuzzy linear regression using shape-preserving operations

Dug Hun Hong

Department of Mathematics, Myongji University, Yongin Kyunggido 449-728, South Korea

Abstract

Fuzzy regression, a nonparametric method, can be quite useful in estimating the relationships among variables where the available data are very limited and imprecise. It can also serve as a sound methodology that can be applied to a variety of management and engineering problems where variables are interacting in an uncertain, qualitative, and fuzzy way. A close examination of the fuzzy regression algorithm reveals that the resulting possibility distribution of fuzzy parameters, which makes this technique attractive in a fuzzy environment, is dependent upon an h parameter value. The h value, which is between 0 and 1, is referred to as the degree of fit of the estimated fuzzy linear model to the given data, and is subjectively selected by a decision maker (DM) as an input to the model. The selection of a proper value of h is important in fuzzy regression, because it determines the range of the posibility ditributions of the fuzzy parameters. In this paper, we discuss the interdependent relationship among the h value, membership function shape, and the spreads of fuzzy parameters in fuzzy linear regression with fuzzy input-output using shape-preserving operations.

Key words : Regression analysis ; membership function ; spread ; h value

1. Introduction

Fuzzy linear regression, developed by Tanaka, Uejima, and Asai [12], aims to model vague and imprecise phenomena using the fuzzy functions defined by Zadeh's extension principle [14], which provides a general method for extending nonfuzzy mathematical concepts in order to deal with fuzzy quantities. Fuzzy regression is a nonparametric method, in the sense that the deviations between the observed values and estimated values are assumed to depend on the indefiniteness/vagueness of the parameters which govern the system structure, not on its measurement errors. A close examination of the fuzzy regression algorithm reveals that resulting possibility distribution of fuzzy parameters, which makes this technique attractive in a fuzzy environment, is dependent upon an h parameter value. The h values, which is between 0 and 1, is referred to as the degree of fit of the estimated fuzzy linear model to the given data, and is subjectively selected by a decision maker (DM) as an input to the model [12]. Conceptually, the h parameter denotes a desired level of credibility or confidence, i.e., for a given degree of credibility (viz. h), the fuzzy regression algorithm determines the spreads (and the center values as well) of the parameters that satisfy the h level specified. The selection of a proper value of h is important in fuzzy regression, because it determines the range of the possibility distributions of the fuzzy parameters. Tanaka and Watada [13] suggest that the selection of the h value be based upon the sufficiency of the data set available. They subjectively set $h = 0$ when the data set is sufficiently large, and use a higher h as the size of data set becomes smaller compared to some ideal size. h value used in previous research vary widely, ranging from 0 to 0.9; for example,$h = 0$ in Tanaka and Watada [13], $h = 0.5$ in Tanaka, Uejima, and Asai [12] $h = 0.7$ in Bardossy, Bogardi, and Kelly [3], Bardossy, Bogardi, and Duckstein [2], $h = 0.75$ in Bardossy [1], and $h = 0.9$ in Gharpuray, Tanaka, Fan, and Lai [4], Oh, Kim, and Lee [9]. Moskowitz and Kim [7] determined the independent relationship among the $h$value, membership function shape, and the spreads of fuzzy parameters in fuzzy linear regression for real input-output data.

Recently, Hong, Lee and Do [6] presented a new method to evaluate fuzzy linear regression models based on Tanaka's approach, where both input data output data are fuzzy numbers using shape preserving fuzzy arithmetic operations. They also proved scale-independent property and discussed the effects outliers.

In this paper, we study further on the relationship among the h value, membership function shape, and the spreads of fuzzy parameters in fuzzy linear regression for fuzzy input-output data using shape-preserving operations.

2. Relationship among h value, membership function, and spread of real input-output fuzzy linear regression

Consider a fuzzy linear function

$$Y = A_1x_1 + A_2x_2 + \cdot + A_px_p = AX,$$

where $A = (A_1, A_2, \cdots, A_p)$ are fuzzy parameters, and can be denoted in the vector form of $A = \{\mathbf{a}, \boldsymbol{\alpha}\}, \mathbf{a} = (a_1, a_2, \cdots, a_p), \boldsymbol{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_p)$. Here, a_j is the mean or center value and α_j the spread of a_j. The problem in the fuzzy linear regression model is to determine fuzzy parameters A^* such that the fuzzy estimate $y_j^* = A^*x_j$ contains y_j with more than h degree, for all j. The fuzzy linear regression problem can be posed as an equivalent linear program as follows:

Find $\mathbf{a}$ and $\boldsymbol{\alpha}$ with

Minimize

$$J(\mathbf{a}, \boldsymbol{\alpha}) = \sum_{i=1}^{n} \alpha^t |x_i| = \sum_{i=1}^{n}\sum_{j=1}^{p} \alpha_j |x_{ij}| \quad (1)$$

Subject to

$$\sum_{j=1}^{p} a_j x_{ij} + |L^{-1}(h)| \sum_{j=1}^{p} \alpha_j |x_{ij}| \geq y_i \quad (2)$$

$$\sum_{j=1}^{p} a_j x_{ij} - |L^{-1}(h)| \sum_{j=1}^{p} \alpha_j |x_{ij}| \leq y_i \quad (3)$$

where $\alpha_j \geq 0$ for all $i = 1, 2, \cdots, n,\ j = 1, 2, \cdots, p$.

where L represent the membership function of a standardized parameter.

We note that if the optimal solution to a fuzzy regression with a level of credibility (say, h_1) is known, then it is possible to identify how much the optimal solution is affected by employing a different level of credibility (say, h_2) without solving the problem. This property is proved by Tanaka and Watada [13]. We restate their theorem using our notation to support the development to follow.

Theorem 2.1. ([13]) Let $A^*_{h_1,L} = (\boldsymbol{\alpha}^*, \mathbf{c}^*)$ denote the optimal solution to a fuzzy regression given in (1)-(3) with a level of credibility h_1 and membership function L. Then the optimal solution with h_2 and L is

$$A^*_{h_2,L} = \left(\boldsymbol{\alpha}^*, \frac{|L^{-1}(h_1)|}{|L^{-1}(h_2)|}\mathbf{c}^*\right). \quad (4)$$

A high h implies a high degree of confidence in the fuzzy estimates associated with each model parameter, with more confidence associated with a greater spread analogous in classical statistics to a wider confidence interval for a higher confidence coefficient.

If the optimal solution to a fuzzy regression with a membership function (say, L_1) is known, then one can readily determine how much the optimal solution is affected by employing a different membership function (say, L_2) at the same level of credibility h, which derives from the following theorem by Moskowitz and Kim [8].

Theorem 2.2. ([8]) Let $A^*_{h,L_1} = (\boldsymbol{\alpha}^*, \mathbf{c}^*)$ denote the optimal solution to a fuzzy regression given in (1)-(3) with a level of credibility h and membership function L_1. Then the optimal solution with h and L_2 is

$$A^*_{h,L_2} = \left(\boldsymbol{\alpha}^*, \frac{|L_1^{-1}(h)|}{|L_2^{-1}(h)|}\mathbf{c}^*\right). \quad (5)$$

Moskowitz and Kim[8] also showed their joint effects on the spreads of fuzzy parameters with the following theorem.

Theorem 2.3. ([8]) Let $A^*_{h_1,L_1} = (\boldsymbol{\alpha}^*, \mathbf{c}^*)$ denote the optimal solution to a fuzzy regression given in (1)-(3) with a level of credibility h_1 and membership function L_1. Then the optimal solution with h_2 and L_2 is

$$A^*_{h_2,L_2} = \left(\boldsymbol{\alpha}^*, \frac{|L_1^{-1}(h_1)|}{|L_2^{-1}(h_2)|}\mathbf{c}^*\right). \quad (6)$$

Remark 2.4. We note that it the input as output in fuzzy regression model given in (1)-(3) are fuzzy numbers(not real) then we don't have any similar types of results as Theorem 1,2 and 3, i.e., there is no relationship among h value, membership function shape, and spread.

Recently, Hong, Lee and Do [6] presented a new method to evaluate fuzzy linear regression models based on Tanaka's approach, where both input data output data are fuzzy numbers using shape preserving fuzzy arithmetic operations. They also proved scale-independent property and discussed the effects outliers.

In the next section, we discuss further on the relationship among the h value, membership function shape, and the spreads of fuzzy parameters in fuzzy linear regression for fuzzy input-output data using shape-preserving operations which are similar to Theorem 1, 2, and 3, and are same when the input and output are real.

3. Relationship among h value, membership function, and spread in fuzzy linear regression using shape-preserving operations

In this section, we first consider fuzzy linear regression model based on Tanaka's approach, where both input data and output data are fuzzy numbers, under T_W-based fuzzy arithmetic operations where $T_W(a,b) = \min\{a,b\}$ it $\max\{a,b\} = 1$ and, 0, otherwise. It is noted that T_W is the only t-norm which induces a shape preserving multiplications of LR-fuzzy numbers [7].
The followings are definitions and results of T_W-based fuzzy arithmetic operations.
A fuzzy number is a convex subset of the real line R with a normalized membership function.

A triangular fuzzy number $\tilde{a}$ denoted by (a,α,β) is defined as

$$\tilde{a}(t) = \begin{cases} 1 - \frac{|a-t|}{\alpha} & \text{if } a-\alpha \le t \le a, \\ 1 - \frac{|a-t|}{\beta} & \text{if } a \le t \le \alpha+\beta, \\ 0 & \text{otherwise,} \end{cases}$$

where $a \in R$ is the center and $\alpha > 0$ is the left spread, $\beta > 0$ is the right spread of $\tilde{a}$.

If $\alpha = \beta$, then the triangular fuzzy number is called a symmetric triangular fuzzy number and denoted by (a,α).

An $L-R$ fuzzy number $\tilde{a} = (a,\alpha,\beta)_{LR}$ is a function from the reals into the interval $[0,1]$ satisfying

$$\tilde{a}(t) = \begin{cases} R(\frac{t-a}{\beta}) & \text{for } a \le t \le \alpha+\beta, \\ L(\frac{a-t}{\alpha}) & \text{for } a-\alpha \le t \le a, \\ 0 & \text{else} \end{cases}$$

where L and R and non-decreasing and continuous function from $[0,1]$ to $[0,1]$ satisfying $L(0) = R(0) = 1$ and $L(1) = R(1) = 0$. If $L = R$ and $\alpha = \beta$, then the symmetric $L-L$ fuzzy number is denoted $(a,\alpha)_L$.

Let A, B be fuzzy numbers of the real line R. The fuzzy number arithmetic operations are summarized as follows:

Fuzzy number addition $\oplus$:

$$(A \oplus B)(z) = \sup_{x+y=z} T_W(A(x), B(y)).$$

Fuzzy number multiplication $\otimes$:

$$(A \otimes B)(z) = \sup_{x \cdot y=z} T_W(A(x), B(y)).$$

Then we have the following two lemmas [5].

Lemma 3.1. Let $A_i = (a_i,\alpha_i,\beta_i)_{LR}, \tilde{X}_i = (x_i,\gamma_i,\delta_i)_{LR}, i = 1,2,\cdots,p$ and let $a_i > 0, x_i > 0, i = 1,2,\cdots,p$. Then, the possibilistic linear function with fuzzy parameter A_i and fuzzy variables $\tilde{X}_i, i = 1,2,\cdots,p$, is given by

$$\begin{aligned}&\tilde{Y}\\ &= (A_1 \otimes \tilde{X}_1) \oplus (A_2 \otimes \tilde{X}_2) \oplus \cdots \oplus (A_p \otimes \tilde{X}_p)\\ &= \left(\sum_{i=1}^{p} a_i x_i, \max_{1\le i\le p}(a_i\gamma_i, x_i\alpha_i), \max_{1\le i\le p}(a_i\delta_i, x_i\beta_i)\right)_{LR}.\end{aligned}$$

Lemma 3.2. Let $A_i = (a_i,\alpha_i)_L, \tilde{X}_i = (x_i,\gamma_i)_L, i = 1,2,\cdots,p$. Then the membership function of $\tilde{Y} = (A_1 \otimes \tilde{X}_1) \oplus (A_2 \otimes \tilde{X}_2) \oplus \cdots \oplus (A_p \otimes \tilde{X}_p)$ is given by

$$\mu_{\tilde{Y}}(y) = L((y - \sum_{i=1}^{p} a_i x_i) / \max_{1\le i\le p}(|a_i|\gamma_i|x_i|\alpha_i)).$$

Let us consider fuzzy output $\tilde{Y}_i = (y_i, e_i)_L$ and fuzzy inputs $\tilde{X}_i = (\tilde{X}_{i1},\cdots,\tilde{X}_{ip}), i = 1,\cdots,n$ where $\tilde{X}_{ij} = (x_{ij},\gamma_{ij})_L, j = 1,\cdots,p$. In the case of non-fuzzy output data, to formulate a possibilistic linear model, the following conditions are assumed.

(i) The data can be represented by a possibilistic linear model:

$$\begin{aligned}\tilde{Y}_i^* &= (A_1^* \otimes \tilde{X}_{i1}) \oplus \cdots \oplus (A_p^* \otimes \tilde{X}_{ip})\\ &= A * \tilde{X}_i,\ i = 1,2,\cdots,n. \qquad (7)\\ &\text{where } A_j^* = (a_j,\alpha_j)_L, j = 1,2,\cdots,p.\end{aligned}$$

(ii) Given input-output relations $(\tilde{X}_i,\tilde{Y}_i), i = 1,2,3,\cdots,n$ and a threshold h, it must hold that

$$\mu_{\tilde{Y}_i}^{-1}([h,1]) \subset \mu_{\tilde{Y}_i^*}^{-1}([h,1]) i = 1,2,\cdots,n. \qquad (8)$$

(iii) The index of fuzziness of the possiblitic linear model,

$$J(\mathbf{a},\boldsymbol{\alpha}) = \sum_{i=1}^{n} \max_{1\le j\le p}(|a_j|\gamma_{ij}, |x_{ij}|\alpha_j). \qquad (9)$$

Under the above assumption, our problem is to obtain fuzzy parameters $A_1^*,\cdots,A_p^*$ that minimize $J(\mathbf{a},\boldsymbol{\alpha})$ in (9) subject to constraint (8). This problem can be solved as the following mixed LP problem:

Minimize

$$J(\mathbf{a},\boldsymbol{\alpha}) = \sum_{i=1}^{n} \max_{1\le j\le p}(|a_j|\gamma_{ij}, |x_{ij}|\alpha_j),$$

Subject to

$$\begin{aligned}&\left| y_i - \sum_{j=1}^{p} a_j x_{ij} \right|\\ &\le |L^{-1}(h)| \max_{1\le j\le p}(|a_j|\gamma_{ij}, |x_{ij}|\alpha_j) - |L^{-1}(h)| e_i,\end{aligned} \qquad (10)$$

where $\alpha_j \geq 0$ for all $i = 1, 2, \cdots, n, j = 1, 2, \cdots, p$.

4. Fuzzy input-Fuzzy output case

Firstly, we consider the following mixed LP problem which can be obtained from (10) by multiplying the spread of input and output data by some constant C.
Minimize

$$J_c(\mathbf{a}, \boldsymbol{\alpha}) = J(c\mathbf{a}, \boldsymbol{\alpha})$$

Subject to

$$|y_i - \sum_{j=1}^{p} a_j x_{ij}| \leq |L^{-1}(h)| \max_{1\leq j\leq p} (|a_j|\delta_{ij}, |x_{ij}|\alpha_j) - |L^{-1}(h)| f_i \quad (11)$$

where $\alpha_j \leq 0$, $\delta_{ij} = c\gamma_{ij}$ and $f_i = ce_i$ for all $i = 1, 2, \cdots, n, j = 1, 2, \cdots, p$.

We now have an interesting relationship among h value, membership function, and spread in the fuzzy linear regression using shape-preserving operation.

Theorem 4.1. Let $A^*_{h_1,L_1} = (\mathbf{a}^*, \boldsymbol{\alpha}^*)$ denote the optimal solution to a fuzzy regression in (10) with a level of credibility h_1 and membership function L_1. Then the optimal solution to a fuzzy regression in (11) with a level of credibility h_2, membership function L_2 and $c = \left|\frac{L_1^{-1}(h_1)}{L_2^{-1}(h_2)}\right|$ is

$$B^*_{h_2,L_2} = (\mathbf{a}^*, \left|\frac{L_1^{-1}(h_1)}{L_2^{-1}(h_2)}\right| \boldsymbol{\alpha}^*).$$

5. Real input-fuzzy output case

We note that if all input data are real, i.e., $\gamma_{ij} = 0$ for all i, j, then $J(\mathbf{a}, \boldsymbol{\alpha}) = J_c(\mathbf{a}, \boldsymbol{\alpha})$ in (11).

Assume all input data are real and consider the following mixed LP problem which can be obtained from (10) by rescaling spread of output data;

Minimize

$$J(\mathbf{a}, \boldsymbol{\alpha})$$

Subject to

$$|y_i - \sum_{j=1}^{p} a_j x_{ij}| \leq |J^{-1}(h)| \max_{1\leq j\leq p} |x_{ij}|\alpha_j - |L^{-1}(h)| f_i \quad (12)$$

where $\alpha_j \geq 0$, $f_i = ce_i$ for all $i = 1, 2, \cdots, n, j = 1, 2, \cdots, p$.

We will have the following result from Theorem 4.

Theorem 5.1. Suppose that all input data are real, i.e., $\gamma_{ij} = 0$ for $i = 1, 2, \cdots, n$, $j = 1, 2, \cdots, p$. Let $A^*_{h_1,L_1} = (\mathbf{a}^*, \boldsymbol{\alpha}^*)$ denote the optimal solution to a fuzzy regression given in (10) with a level of credibility h_1 and membership function L_1. Then the optimal solution to a fuzzy regression in (12) with a level of credibility h_2, membership function L_2 and $c = \left|\frac{L_1^{-1}(h_1)}{L_2^{-1}(h_2)}\right|$ is

$$C^*_{h_2,L_2} = (\mathbf{a}^*, \left|\frac{L_1^{-1}(h_1)}{L_2^{-1}(h_2)}\right| \boldsymbol{\alpha}^*).$$

6. Real input-output

It all input-output data are real, i.e., $\gamma_{ij} = e_i = 0$ for all i, j, then the optimal solution $B^*_{h,L}$ to a fuzzy regression in (11) with a level of credibility h and membership function L is same as $A^*_{h,L}$. Therefore, we have the following result which is same version of Moskowitz and Kim [8].

Theorem 6.1. Suppose that all input-output data are real, i.e., $\gamma_{ij} = e_i = 0$ for $i = 1, 2, \cdots, n$, $j = 1, \cdots, p$. Let $A^*_{h_1,L_1} = (\mathbf{a}^*, \boldsymbol{\alpha}^*)$ denote the optimal solution to a fuzzy regression given in (3.4) with a level of credibility h_1 and membership function L_1. Then the optimal solution with h_2 and L_2 is

$$A^*_{h_2,L_2} = (\mathbf{a}^*, \left|\frac{L_1^{-1}(h_1)}{L_2^{-1}(h_2)}\right| \boldsymbol{\alpha}^*).$$

The following results are immediate.

Corollary 6.2. Suppose that all input-output data are real. Let $A^*_{h,L_1} = (\mathbf{a}^*, \boldsymbol{\alpha}^*)$ denote the optimal solution to a fuzzy regression in(3,4) with a level of credibility h and membership function L_1. Then the optimal solution with h and L_2 is

$$A^*_{h,L_2} = (\mathbf{a}^*, \left|\frac{L_1^{-1}(h)}{L_2^{-1}(h)}\right| \boldsymbol{\alpha}^*).$$

Corollary 6.3. Suppose that all input-output data are real. Let $A^*_{h,L_1} = (\mathbf{a}^*, \boldsymbol{\alpha}^*)$ denote the optimal solution to a fuzzy regression in (3,4) with a level of credibility h_1 and membership function L. The optimal solution with h_2 and L is

$$A^*_{h_2,L} = (\mathbf{a}^*, \left|\frac{L^{-1}(h_1)}{L^{-1}(h_2)}\right| \boldsymbol{\alpha}^*).$$

7. Conclusion

We determined the relationship among h value, membership function shape L, and spread of fuzzy parameters α in fuzzy linear regression using shape-preserving operations, in order that the mechanism of fuzzy regression can be understood more insightfully. We also examined the sensitivity of the spread with respect to the $h$value and membership function shape L for the cases of fuzzy input-output, real input-fuzzy output and real input-output. For the case of real input-output, we have exactly same version of results in Moskowitz and Kim [8] and Tanaka,Uejima and Asai [12].

References

[1] . Bardossy, *Note on fuzzy regression*, Fuzzy Sets and Systems **37** (1990) 65-75.

[2] . Bardossy, I. Bogardi and L. Duckstein, *Fuzzy regression in hydrology*, Water Resources Research **26(7)** (July 1990) 1497-1508.

[3] . Bardossy, I. Bogardi and W. Kelly, *Fuzzy regression for electrical resistivity-Hydraulic conductivity relationships*, in:J.Chameau and J.T.P. Yao, Eds, Proc. North American Fuzzy Information Processing Society Workshop (1987) 333-352.

[4] . Gharpuray, H. Tanaka, L. Fan and F. Lai, *Fuzzy linear regression analysis of cellulose hydrolysis*, Chem. Eng. Commum. **41**(1986) 299-314.

[5] . Hayashi and H. Tanaka, *The fuzzy GMDH algorithm by possibility models and its application*, Fuzzy Sets and System **36** (1990) 245-258.

[6] . H. Hong, S Lee and H. Y. Do, *Fuzzy linear regression analysis for fuzzy input-output data using shape-preserving operations*, Fuzzy Sets and Systems **122** (2001) 513-526.

[7] . H. Hong, *Shape preserving multiplications of fuzzy numbers*, Fuzzy Sets and Systems **123**(2001) 81-84.

[8] . Moskowitz and K. Kim, *On assessing the Hvalue in fuzzy linear regression*, Fuzzy Sets and Systems **58**(1993) 303-327.

[9] . Oh, W. Kim and J. Lee, *An approach to causal modelling in fuzzy environment and its application*, Fuzzy Sets and Systems **35**(1990) 43-55.

[10] . Suzuki, K. Furukawa, Y. Inoue and K. Nakagawa, *A proposal of rock mass classification for tunnels by the fuzzy regression model*, Jap.J.Civil Eng. **418**(June 1990)181-190.

[11] . Tanaka, *Fuzzy data analysis by possibilistic linear models*, Fuzzy Sets and Systems **24**(1987) 363-375.

[12] . Tanaka, S. Uejima and K. Asai, *Linear regression analysis with fuzzy model*, IEEE Transactions on Systems, Man, and Cybernetics SMC-12 (1982) 903-907.

[13] . Tanaka and J. Watada, *Possibilistic linear systems and their applications to the linear regression model*, Fuzzy Sets and Systems **27**(1988) 275-289.

[14] . A. Zadeh, *The concept of a linguistic variable and its application to approximate reasoning-I*, Inf. Sci. **8** (1975) 199-249.

Dug Hun Hong
Professor of Muongji University
Research Area:
E-mail :