The Korean Communications
in Statistics Vol. 8, No. 2, 2001
pp. 319-325

A Note on a Fuzzy Linear Regression Model for Fuzzy Input-output Data Using Real Coefficients

Dug Hun Hong[1)]

Abstract

In this note, we propose a simple fuzzy linear regression model for fuzzy input-output data based on Tanaka's approach. Then an LP-based method to derive the satisfying solution of the decision making is developed.

Keywords : Fuzzy regression analysis, linear regression model, fuzzy input-output.

1. Introduction

Fuzzy linear regression provides a means for tackling regression problems lacking a significant amount of data for determining regression models and with vague relationship between the dependent variable and independent variables.

In fuzzy linear regression analysis, recently proposed by Tanaka et al. [6, 7], deviations between the observed values and the estimated values are assumed to depend on the fuzziness of the system structure, in contrast to the usual linear regression analysis where deviation are supposed to be caused by observation errors or variables not in the model. Linear programming based methods for obtaining fuzzy linear regression models for some fixed fuzzy threshold via the set inclusion relations among α-level sets for fuzzy numbers, or possibility and necessity measures have been proposed [6, 7].

In these fuzzy linear regression models [6, 7], however, it should be stressed that while output data is assumed to be a fuzzy number, input data is not a fuzzy number.

Sakawa and Yager [5], formulated three types of multiobjective programing problems for obtaining fuzzy linear regression models, where both input data and output data are fuzzy numbers by using three indices for equalities between two fuzzy numbers [2]. Recently, Hong et al. [3, 4] have presented two new method to evaluate fuzzy linear regression models using shape preserving fuzzy arithmetic operations where both input data and output data are fuzzy

1) Associate Professor, School of Mechanical and Automotive Engineering, Catholic University of Daegu, Kyungbuk, 712 - 702, Korea.
E-mail : dhhohg@cuth.cataegu.ac.kr

numbers.

The purpose of this work is to propose a simple fuzzy linear regression model based on Tanaka's approach where both input data and output data are represented as fuzzy numbers and parameters are real numbers.

2. Fuzzy linear systems

According to the definitions defined by Dubois and Prade [2], fundamental definitions and concepts of fuzzy numbers and fuzzy linear systems are briefly described in this section.

Definition 2.1. A fuzzy number is a fuzzy set A, whose membership function $u_A : R \rightarrow [0,1]$ which is upper semi-continuous, normal and convex, i.e. (i) $[A]_h = \{ x \mid u_A(x) \geq h \}$ is a closed set , (ii) $\exists x$ such that $u_A(x) = 1$, (iii) $u_A(\lambda x_1 + (1-\lambda)x_2) \geq u_A(x_1) \wedge u_A(x_2)$ for $\lambda \in [0,1]$.

A symmetrical fuzzy number A denoted by $A = (\alpha, c)_L$ is defined as

$$u_A(x) = L((x-\alpha)/c), \quad c>0,$$

where $L(x)$ is a shape function of fuzzy numbers such that (i) $L(x) = L(-x)$, (ii) $L(0) = 1$, (iii) L is strictly decreasing on $[0, +\infty)$. As examples of $L(x)$, $L(x) = \max(0, 1-|x|^P)$, $L(x) = e^{-|x|^P}$ and $L(x) = 1/(1+|x|^P)$ are considered.

Given a fuzzy information M defined by a fuzzy number $u_M(x)$, the possibility distribution $\pi_X(x)$ is defined as $\pi_x(x) \triangleq u_M(x)$. The possibility measure of fuzzy set A is defined as

$$\pi_X(A) = \sup_x u_A(x) \wedge \pi_X(x).$$

A possibility space is represented by $(X, P(X), \pi_X(\cdot))$, where $P(X)$ is the set of all fuzzy subsets of X. Let us consider two sets X and Y , and a function $f: X \rightarrow Y$. A possibility space $(X, P(Y), \pi_Y(\cdot))$ can be induced from the given possibility space as follows.

Denoting a set E_y as $E_y = \{x \mid y = f(x)\}$, a possibility distribution of Y is induced from $\pi_X(\cdot)$ as

$$\pi_Y(y) = \pi_X(E_y)$$

which can be rewritten by the definition of possibility measure as

$$\pi_Y(y) = \sup_{x \in E} \pi_X(x).$$

Hence, we have for $B \in P(Y)$,

$$\pi_Y(B) = \sup_x u_B(y) \wedge \pi_Y(y),$$

which leads to a possibility space $\{Y, P(Y), \pi_Y(\cdot)\}$.

Definition 2.2. Given a function $y = f(x_1, x_2)$, the fuzzy output $Y = f(A, B)$ is defined by the following membership function:

$$u_Y(y) = \sup_{y=f(x_1x_2)} u_A(x_1) \wedge u_B(x_2).$$

Since we consider only symmetrical fuzzy numbers defined by $L(x)$ in this paper, the following definition is more favorable in applications than the definition given by Zadeh, i.e. $A \supseteq B$ if and only if $u_A(x) \geq u_B(x)$.

Definition 2.3. The inclusion of fuzzy numbers with degree $0 \leq h < 1$ denoted $A \supseteq_h B$ is defined by $[A]_h \supseteq [B]_h$.

It follow from Definition 2.3 that $[A]_h \supseteq [B]_h$ is equivalent to

$$a_1 \leq a_2 + L^{-1}(h)(c_1 - c_2), \quad a_1 \geq a_2 - L^{-1}(h)(c_1 - c_2) \tag{1}$$

where $u_A(x) = L((x-a_1)/c_1)$ and $u_B(x) = L((x-a_2)/c_2)$.

Let us consider a linear system with fuzzy data by $u_{X_i}(x) = L((x-x_i)/\delta_i), i=,\cdots,n$, where x_i is the center and δ_i is the width of fuzzy number.

A fuzzy linear system is

$$Y = a_1X_1 + \cdots + a_nX_n \triangleq \mathbf{aX}$$

where X_i is fuzzy number and a_i is non-fuzzy. The following result is show in [8].

Theorem 1. Let $X_i = (x_i, \delta_i)_L$, $i = 1, \ldots, n$, then the fuzzy linear function

$$Y = a_1X_1 + \cdots + a_nX_N = (\sum a_ix_i, \sum |a_i|\delta_i)_L.$$

3. Fuzzy linear regression with fuzzy data

To formulate a fuzzy linear regression model with fuzzy input-output data , the following are assumed :

(i) the data can be represented by a fuzzy linear model :

$$Y_i^* = a_1^* X_{i1} + \cdots + a_n^* X_{in} = \boldsymbol{a}^* X_i \tag{2}$$

where $\boldsymbol{X}_i = (X_{i1} = (x_{i1}, \delta_{i1}), \ldots, X_{in} = (x_{in}, \delta_{in})_L)^t$ and $a^* = (a_1^*, \cdots, a_n^*)$.

(ii) Given input-output relation $(\boldsymbol{X}_i, Y_i)$, $Y_i = (y_i, e_i)_L$, $i = 1, \ldots, N$, and a threshold h, it must hold that

$$Y_i \subseteq_h Y_i^* \qquad i = 1, \ldots, N. \tag{3}$$

(iii) The index of fuzziness of fuzzy linear model is

$$J(\boldsymbol{a}) = \sum_{i=1}^{N} \sum_{j=1}^{n} |a_j| \delta_{ij} \tag{4}$$

Under the above assumptions our problem is to obtain real parameter a_i, $i = 1, \ldots, n$, that minimize $J(\boldsymbol{a})$ in (4) subject to the constraint (3). Using (1), this problem can be reduced as the following LP problem :

$$\operatorname*{Min}_{a} \; J(\boldsymbol{a}) = \sum_{i=1}^{N} \sum_{j=1}^{n} |a_j| \delta_{ij} \tag{5}$$

subject to

$$\begin{aligned} & y_i + L^{-1}(h) e_i \le \sum_{j=1}^{n} a_j x_{ij} + L^{-1}(h) \sum_{j=1}^{n} |a_j| \delta_{ij}, \\ & y_i - L^{-1}(h) e_i \ge \sum_{j=1}^{n} a_j x_{ij} - L^{-1}(h) \sum_{j=1}^{n} |a_j| \delta_{ij}, \\ & i = 1, \ldots, N, \end{aligned} \tag{6}$$

where $L^{-1}(h)$ is supposed to be finite in this paper.

If output data Y_i are real. i.e., $e_i = 0$, $i = 1, \ldots, N$, then the problem for obtaining fuzzy linear regression model is reduced to the LP problem :

$$\operatorname{Min} J(\boldsymbol{a}) = \sum_{i=1}^{N} \sum_{j=1}^{n} |a_j| \delta_{ij}$$

subject to

$$y_i \le \sum_{j=1}^{n} a_j x_{ij} + L^{-1}(h) \sum_{j=1}^{n} |a_j| \delta_{ij} \ ,$$
$$y_i \ge \sum_{j=1}^{n} a_j x_{ij} - L^{-1}(h) \sum_{j=1}^{n} |a_j| \delta_{ij} \ , \tag{7}$$

It is note that for given the data (X_i, Y_i), $i=1,\dots,N$, there does not always exist an optimal solution for $0 \le h < 1$ in the LP problem (5) and (6).

If y_i is close to $\sum_{j=1}^{n} a_j x_{ij}$, $i=1,\dots,N$, for some $\boldsymbol{a}$, i.e.,

$$|y_i - \sum_{j=1}^{n} a_j x_{ij}| \le L^{-1}(h) \sum_{j=1}^{n} |a_j| \delta_{ij} - L^{-1}(h) e_i \tag{8}$$

for some $\boldsymbol{a}$ and h, $i=1,\cdots,N$, then there can exist an optimal solution.

Theorem 2. Given the data (X_i, Y_i), $i=1,\cdots,N$, if there exists an $\boldsymbol{a}$ satisfying (8) for some h_0, then there is an optimal solution $\boldsymbol{a}^*$ for $h \le h_0$ in the LP problem (5,6).

Proof. The condition in (8) yield (6) for some $\boldsymbol{a}$ and $h \le h_0$. Then $\boldsymbol{a}$ is a feasible solution. Hence, there is an admissible set of (8) and there is an optimal solution $\boldsymbol{a}^*$.

If $\lim_{h \to 0} L^{-1}(h) = \infty$ and $\sum_{j=1}^{n} |a_j| \delta_{ij} - e_i > 0$, $i=1,\cdots,N$, for some $\boldsymbol{a}$, then there exist $h > 0$ satisfying (8). By similar argument, we have the following result.

Theorem 3. Given the data (X_i, Y_i), $i=1,\cdots,N$, if $\lim_{h \to 0} L^{-1}(h) = \infty$ and $\sum_{j=1}^{n} |a_j| \delta_{ij} - e_i > 0$, $i=1,\cdots,N$, for some $\boldsymbol{a}$, then there exists an optimal solution of LP problem (5,6).

Example. We use the same example provided by Sakawa and Yano [5], given in Table 1. The fuzzy linear model is

$$Y = a_0 + a_1 X$$

with $L(x) = 1 - x$. We prove the LP problem by using the conventional linear program approach for $h =$ 0.3, 0.5, 0.6 and 0.7 .

Table 1

sample number i	$X_i=(x_i, \gamma_i)$	$Y_i=(y_i, e_i)$
1	(2.0, 0.5)	(4.0, 0.5)
2	(3.5, 0.5)	(5.5, 0.5)
3	(5.5, 1.5)	(7.5, 1.0)
4	(7.0, 0.5)	(6.5, 0.5)
5	(8.5, 0.5)	(8.5, 0.5)
6	(10.5, 1.5)	(8.0, 1.0)
7	(11.0, 0.5)	(10.5, 0.5)
8	(12.5, 0.5)	(9.5, 0.5)

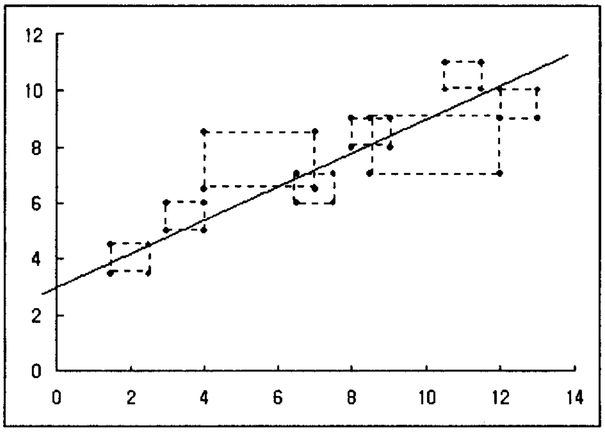

Fig. 1. The fuzzy linear regression model with $h=0.5$

Table 2

h	0.3	0.5	0.6	0.7
a_0^*	2.37	2.92	3.30	4.18
a_1^*	0.67	0.60	0.56	0.47
$J(a^*)$	22.34	26.36	29.18	35.77

In this example, at the fourth iteration, the satisfying solution of the LP problem is obtained as shown in Table 2. We don't need any algorithm to prove the optimization problem and just a simple LP based method is need.

4. Conclusion

In this note, we have presented a simple fuzzy linear regression model based on Tanaka's approach where both input data and output data are fuzzy number and coefficients are reals. Then just a simple *LP*-base method is need to derive the satisfying solution of the decision making.

Acknowledgements

This work was supported by grant No. 2000-2-10400-002-3 from the Basic Research Program of the Korea Science & Engineering Foundation.

References

[1] D. Dubiois, (CRC Press, Boca Raton, 1987). Linear programming with fuzzy data, in : J. C Bezdek, Ed., *Analysis of Fuzzy information* 3, 241-263.

[2] D. Dubiois and H. Prade, (Academic Press, New York, 1980). *Fuzzy sets and Systems : Theory and Application.*

[3] D. H. Hong, S. Lee and H. Y. Do, (2001) Fuzzy linear regression analysis for fuzzy input-output data using shape-preserving operation, *Fuzzy Sets and Systems* 122, 157-170.

[4] D. H. Hong, J. K. Song and H. Y. Do, (2001) Fuzzy least squares linear regression analysis using shape preserving operations, *Information sciences* 138, 77-85.

[5] M. Sakawa and H. Yano, (1992). Multiobjective fuzzy linear regression analysis for fuzzy input-output data, *Fuzzy Sets and Systems* 47, 173-181.

[6] H. Tanaka, J. Watada and I, Hayashi, (1986). On three formulations of fuzzy linear regression analysis, *Trans. Soc. Instr. and Control Engrs.* 22, 1051-1057 (in Japanese).

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

[8] L. A. Zadeh, (1978). Fuzzy sets as a basis for theory of possibility, *Fuzzy Sets and Systems* 1, 3-28.