Australian Journal of Basic and Applied Sciences, 5(3): 261-267, 2011
ISSN 1991-8178

Evaluation of Fuzzy Linear Regression Models by Parametric Distance

[1]Rahim Saneifard and [2]Rasoul Saneifard

[1]Department of Mathematics, Islamic Azad Unuversity, Oroumieh Branch, Oroumieh, Iran.
[2]Department of Engineering Technology, Texas Southern University, Houston, Texas, USA.

Abstract: Fuzzy linear regression models can provide an estimated fuzzy number that has a fuzzy membership function. If a point that has the highest membership value from the estimated fuzzy number is not within the support of the observed fuzzy membership function, a decision-maker can have high risk from the estimate. In this study a new distance, between fuzzy numbers is proposed. On the basis of this distance a fuzzy least square regression model is constructed for the case of polynomial-type dependent variable and ordinary input variables.

Key words: Fuzzy number; Defuzzification; Fuzzy regression; Weighted distance.

INTRODUCTION

Regression analysis is a statistical tool used to find a relation between two or more quantitative variables. Based on the relation, one variable can be predicted from the other, or others. The concept of a relation can be distinguished in to functional relation and statistical relation. If a mathematical formula can express the relation between two variables, the relation is functional. Given a particular value of the independent variable, the functional relation can indicate the corresponding value of the dependent variable. A statistical relation, unlike a functional relation, cannot give a perfect corresponding value of the dependent variable, given a value of independent variable. The term *regression* is referred to as a description of statistical relations between variables. A statistical regression model is generally based on the following two characteristics: (1) there is a probability distribution of a tendency of the dependent variable for each level of the independent variable; (2) the means of these probability distributions vary in some systematic fashion with the values of the independent variables. However, sometimes it is difficult to find a probability distribution of the dependent variable, especially when the dependent variable tends to be influenced by subjective judgment such as human decision-making. Fuzzy linear regression, introduced by (Tanaka *et al.* 1982), is based on a possibility distribution that reflects the membership values of the dependent variable rather than a probability distribution. Also, the relation between the dependent variable and independent(s) is defined by using fuzzy concept rather than statistical concept. Regardless of the underlying assumptions, one of the most important objectives of a regression model is to estimate the value of the dependent variable associated with independent variable(s) as close to the observed data as possible. In a fuzzy regression fuzzy model, the degree of the fitting of the estimate fuzzy linear model to the given data was defined by -level of the possibility distribution (Tanaka *et al.* 1982). Using a metric defined over triangular fuzzy number, the authors in (Diamond, 1987) suggested a model of fuzzy least squares. For Diomand, the objective function was minimization of the distance between the experimental and predicted data. In the latter study a fuzzy regression model with fuzzy/crisp free variables and fuzzy/crisp dependent variables using a least square approach was proposed. A fuzzy least squares regression model with fuzzy/crisp inputs and outputs was also considered in (D'urso, 2003). The case of non-symmetric triangular fuzzy numbers, a case that is a generalization of a linear regression model with coefficients in the form of symmetric fuzzy triangular type numbers, was analyzed in (Yen, *et al.* 1999). The least squares problem with fuzzy data and an -type regression model was considered in(Ruoning, 1997). In (Kim, *et al.* 1998) a fuzzy least squares regression model was proposed with minimization of the distance between the experimental and predicted fuzzy values of the dependent variable. But in the case in which the support of the fuzzy numbers do not intersect, the distance (Kim, *et al.* 1998) always yields the same result, regardless of any dependence on the relative position of the fuzzy numbers. The distance used in (Diamond, 1987) is not sensitive to fuzzy numbers that are not of triangular form. The distance

Corresponding Author: Rahim Saneifard, Department of Mathematics, Islamic Azad University, Oroumieh Branch, Oroumieh, Iran.
E-mail: srsaneeifard@yahoo.com

considered in the present study is similar to the distance in (Ma. M, *et al.* 2002), though the latter studies do not make use of the distribution function of the importance of the degrees of membership. In the present paper a parametric distance between fuzzy numbers is proposed. This distance is universal and flexible and is capable of expressing the strategy of a decision maker and is also sensitive to any form of membership function of fuzzy numbers. On the basis of this distance a fuzzy least squares regression model is constructed for the case of polynomial type dependent variable. The parametric distance used in the present paper makes it possible to express a strategy of estimating fuzziness for individuals responsible for making decision in the course of determining the distance between fuzzy individuals. In additions, such a distance is sensitive to any form of the membership function of a fuzzy number.

2. Basic Definitions and Notations:

The basic definitions of a fuzzy number are given in (Zimmermann, 1991; Saneifard, 2009) as follows:

Definition 1:

A fuzzy number A is a mapping $A(x)$: $\Re \to [0,1]$ with the following properties:

1. A is an upper semi-continuous function on $\Re$,
2. A (x) = 0 outside of some interval $[a_1,b_2] \subset \Re$,
3. There are real numbers a_1,b_2 such as $a_1 \leq a_2 \leq b_1 \leq b_2$ and

- A (x) is a monotonic increasing function on $[a_1,a_2]$
- A (x) is a monotonic decreasing function on $[b_1,b_2]$
- A (x) = 1 for all in $[a_2,b_1]$.

Let $\Re$ be the set of all real numbers. The researchers assume a fuzzy number A that can be expressed for all $x \in \Re$ in the form

$$A(x) = \begin{cases} g(x) & when \quad x \in [a,b), \\ 1 & when \quad x \in [b,c], \\ h(x) & when \quad x \in (c,d], \\ 0 & otherwise. \end{cases}$$

where a,b,c,d are real numbers such as $a < b \leq c < d$ and g is a real valued function that is increasing and h right continuous and is a real valued function that is decreasing and left continuous. Support function is defined as follows:

$$\text{supp}(A) = \overline{\{x|A(x) > 0\}}$$

where $\overline{\{x|A(x) > 0\}}$ is closure of set $\{x|A(x) > 0\}$.

3. Concept of Parametric Distance:

Suppose F denotes the space of fuzzy numbers. We will assume that the fuzzy number A is represented by means of the following -representation:

$$A = \bigcup_{\alpha \in [0,1]} (\alpha, A_\alpha)$$

where

$$\forall \alpha \in [0,1]: A_\alpha = [L_A(\alpha), R_A(\alpha)] \subset (-\infty, \infty)$$

Here, $L:[0,1] \rightarrow (-\infty,\infty)$s a monotonically non-decreasing $R:[0,1] \rightarrow (-\infty,\infty)$d is a monotonically non-increasing left-continuous functions. The functions L (.) and R (.) express the left and right sides of a fuzzy number, respectively. In other words,

$$L(\alpha) = \mu_{\uparrow}^{-1}(\alpha), R(\alpha) = \mu_{\downarrow}^{-1}(\alpha),$$

where $L(\alpha) = \mu_{\uparrow}^{-1}(\alpha)$ and $R(\alpha) = \mu_{\downarrow}^{-1}(\alpha)$ denote quasi-inverse functions of the increasing and decreasing parts of the membership functions μ (t), respectively.

Definition 2:
The following values constitute the weighted averaged representative and weighted width, respectively, of the fuzzy number A:

$$I(A) = \int_0^1 (cL_A(\alpha) + (1-c)R_A(\alpha))d\alpha,$$

and

$$D(A) = \int_0^1 (R_A(\alpha) - L_A(\alpha))f(\alpha)d\alpha.$$

Here $0 \leq c \leq 1$ denotes an "optimism/pessimism" coefficient in conducting operations on fuzzy numbers. The function f (α) is nonnegative and increasing function on [0,1] with f (0) = 0, f (1) = 1and $\int_0^1 f(\alpha)d\alpha = \frac{1}{2}$ The function is also called weighting function. In actual applications, function can be chosen according to the actual situation. In practical cases, it may be assume that

$$f(\alpha) = \frac{k+1}{2} \alpha^k, k = 1,2,\cdots \tag{3.1}$$

Definition 3:
For arbitrary fuzzy numbers and the quantity

$$TRD(A,B) = \sqrt{[I(A) - I(B)]^2 + [D(A) - D(B)]^2} \tag{3.2}$$

is called the TRD distance between the fuzzy numbers A and B.

4. Fuzzy Least Squares Regression:
The following is a mathematical description of the classical regression model:

$$Y_i = f(X_{ij}, \beta_j) + \varepsilon_i, i = 1,\cdots,N; j = 0,\cdots,n.$$

Where X_{ij} is the input value of the j-th variable of the *i-th* experiment and β_j, regression coefficient for the input *j*. f (X_{ij} , β_j) is a predicted value for the experimental output Y_i. For the case of a linear regression with n independent variables, the mathematical model assumes the form

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_n X_{in} + \varepsilon_i$$

It is well known that the following minimization laws represent the method of least squares for the classical regression model:

$$Q(\beta_0, \cdots, \beta_n) = \sum_{i=1}^{n} [Y_i - (\beta_0 + \beta_1 X_{i1} + \cdots + \beta_n X_{in})]^2 \quad \rightarrow \quad Min$$

A linear regression model with fuzzy regression coefficients and fuzzy value of the errors may be written as

$$Y_i = A_0(+)A_1(\times)x_{i1}(+)\cdots(+)A_n(\times)x_{in}(+)\varepsilon_i, \qquad i = 1, \cdots, N, \quad (*)$$

Where Y_i is a dependent variable, A_i, i = 1,..., n are unknown coefficient in the form of fuzzy numbers, X_{ij}, i = 1,... N; j = 1,.. n, are ordinary (nonfuzzy) independent variables and ε_i are errors. The two operations (+) and (*) generalized arithmetic operation on fuzzy numbers.

The fuzzy numbers that are considered in the present article are of polynomial form. For a concrete fuzzy numbers A_j the membership function of such numbers may be written as

$$\mu_{A_j}(t) = \begin{cases} 1 - \left(\dfrac{A_j^M - t}{A_j^M - A_j^-}\right)^s, & A_j^- \le t \le A_j^M \\ 1 - \left(\dfrac{t - A_j^M}{A_j^+ - A_j^M}\right)^s, & A_j^M \le t \le A_j^+ \\ 0, & otherwise. \end{cases}$$

Here the parameter $s > 0$ specifies the type of convexity of the sides of a fuzzy number, A, A_j^-, and A^+_j are the values of the mode, left point and right point of the boundary of the support of the fuzzy number, i.e.

$$A_j^- = inf\left\{t \middle| \mu_{A_j}(t) > 0\right\} \quad \text{and} \quad A_j^+ = sup\left\{t \middle| \mu_{A_j}(t) > 0\right\}$$

Setting

$$d_i^l = Y_i^m - Y_i^- \quad \text{and} \quad d_i^R = Y_i^+ - Y_i^m.$$

Formula (*) may be rewritten in the following form:

$$Y_i \equiv (Y_i^m, d_i^L, d_i^R) = (A_0^M, d_0^L, d_0^R)(+)(A_1^M, d_1^L, d_1^R)(\times)x_{i1}(+)\cdots(+)(A_n^M, d_n^L, d_n^R)(\times)x_{in}(+)\varepsilon_i,$$

$$i = 1, \cdots, N.$$

The predicted fuzzy value of the dependent variable Y_i^* may be written thus:

$$Y_i^* = \left(\sum_{j=0}^{n} \xi_j x_{ij}, \sum_{j=0}^{n} \delta_j^L x_{ij}, \sum_{j=0}^{n} \delta_j^R x_{ij} \right), \quad i = 1, \cdots, N,$$

Where $\sum_{j=0}^{n} \xi_j x_{ij}$ is the value of the mode and $\sum_{j=0}^{n} \delta_j^L x_{ij}$ and $\sum_{j=0}^{n} \delta_j^R x_{ij}$ are respectively, the left and right deviation, moreover, $x_{i0} = 1, \ i = 1, \cdots, N.$

The deviation between the experimental and predicted fuzzy values will also have a fuzzy value,

$\varepsilon_i = Y_i(-)Y_i^*$, $i = 1, \ldots, N$.

It is well known that in the least squares method it is necessary to minimize the sum of the squares of the deviation for experiments, i.e.

$$\sum_{i=1}^{N} \varepsilon_i^2 = \sum_{i=1}^{N} (Y_i(-)Y_i^*)^2 \ \rightarrow Min \qquad (**)$$

In the case where the experimental and predicted values of the dependent variable are fuzzy numbers, formula (**) may be rewritten in the following form using the TRD distance:

$$\left\{ \sum_{i=1}^{N} ([I(Y_i) - I(Y_i^*)]^2 + [D(Y_i) - D(Y_i^*)]^2) \right\} \ \rightarrow Min$$

In the latter formula I (Y_i) and I (Y^*_i) are the weighted averaged representatives while D (Y_i) and D (Y^*_i) are the weighted averaged width of the experimental and predicted fuzzy values, respectively.

Let us suppose that the distribution function of the importance of the degrees $f(a)$ has the form as in (3.1). Then the fuzzy coefficient in the form of polynomial-type fuzzy numbers will have the form

$$L_{Y_i}(\alpha) = Y_i^m - (1-\alpha)^{1/s} d_i^L \ , \ R_{Y_i}(\alpha) = Y_i^m - (1-\alpha)^{1/s} d_i^R.$$

Hence,

$$I(Y_i) = c\int_0^1 \left[Y_i^m - (1-\alpha)^{\frac{1}{s}} d_i^L \right] d\alpha + (1-c)\int_0^1 \left[Y_i^m - (1-\alpha)^{\frac{1}{s}} d_i^R \right] d\alpha$$

$$= cY_i^m - cd_i^L \int_0^1 (1-\alpha)^{\frac{1}{s}} d\alpha + (1-c)Y_i^m + (1-c)d_i^R \int_0^1 (1-\alpha)^{\frac{1}{s}} d\alpha \ \cdot$$

There is,

$$I(Y_i) = Y_i^m - [(1-c)d_i^R - cd_i^L] \int_0^1 (1-\alpha)^{\frac{1}{s}} d\alpha, \qquad (3.3)$$

where the integral $\int_0^1 (1-\alpha)^{\frac{1}{s}} d\alpha$ is the beta function $\mathrm{B}\left(1,1+\frac{1}{s}\right)$. Using the gamma function we may write

$$\mathrm{B}\left(1,1+\frac{1}{s}\right)=\frac{\Gamma(1)\Gamma(1+\frac{1}{s})}{\Gamma(2+\frac{1}{s})}$$

And, finally, recalling that $(1+k)\Gamma(1+k)=\Gamma(2+k)$ and setting

$$\Phi(k,s)=\frac{\Gamma(2+k)\Gamma(1+\frac{1}{s})}{\Gamma(2+k+\frac{1}{s})} \tag{3.4}$$

Formula (3.3) may be written in the following form:

$$I(Y_i)=Y_i^m+\ \Phi(0,s)[(1-c)d_i^R-cd_i^L].$$

The weighted averaged representative of a dependent variable may be represented as follows:

$$L_{Y_i^*}=\sum_{j=0}^{n}\left[\xi_j x_{ij}-(1-\alpha)^{\frac{1}{s}}\delta_i^L x_{ij}\right], R_{Y_i^*}=\sum_{j=0}^{n}\left[\xi_j x_{ij}-(1-\alpha)^{\frac{1}{s}}\delta_i^R x_{ij}\right]\ , i=1,\cdots,N,$$

Where $x_{i0}=1, i=1,..,N$. In view of the normalization condition (3.1) for the distribution function of the importance of degree, we may write

$$I(Y_i^*)=c\int_0^1\left[\sum_{j=0}^{n}\left(\xi_j x_{ij}-(1-\alpha)^{\frac{1}{s}}\delta_i^L x_{ij}\right)\right]d\alpha \quad +(1-c)\int_0^1\left[\sum_{j=0}^{n}\left(\xi_j x_{ij}-(1-\alpha)^{\frac{1}{s}}\delta_i^R x_{ij}\right)\right]d\alpha$$

$$=c\sum_{j=0}^{n}\xi_j x_{ij}-c\sum_{j=0}^{n}\delta_i^L x_{ij}\int_0^1(1-\alpha)^{\frac{1}{s}}d\alpha+(1-c)\sum_{j=0}^{n}\xi_j x_{ij}\ -(1-c)\sum_{j=0}^{n}\delta_i^R x_{ij}\int_0^1(1-\alpha)^{\frac{1}{s}}d\alpha.$$

In view of (3.4), the latter formula may be rewritten thus:

$$I(Y_i^*)=\sum_{j=0}^{n}\xi_j x_{ij}+\ \Phi(0,s)\sum_{j=0}^{n}\left(c\delta_i^R x_{ij}-(1-c)\delta_i^L x_{ij}\right).$$

The weighted averaged width of the experimental D (Y_i) and predicted D (Y^*_i) values of the dependent variable maybe calculated in the following way:

$$D(Y_i)=\int_0^1\left(Y_i^m+(1-\alpha)^{\frac{1}{s}}d_i^R-\left(Y_i^m-(1-\alpha)^{\frac{1}{s}}d_i^L\right)\right)\frac{k+1}{2}\alpha^k d\alpha$$

$$=d_i^R\frac{k+1}{2}\int_0^1(1-\alpha)^{\frac{1}{s}}\alpha^k d\alpha+d_i^L\frac{k+1}{2}\int_0^1(1-\alpha)^{\frac{1}{s}}\alpha^k d\alpha \quad =\frac{1}{2}\Phi(k,s)(d_i^R+d_i^L).$$

Similarly,

$D(Y_i^*)$ = $\Phi(k,s)\sum_{j=0}^{n}\left(\delta_i^R x_{ij} + \delta_i^L x_{ij}\right)$.

Thus, the fuzzy least squares regression model based on TRD distance may be represented in the form

$$\left\{\sum_{i=1}^{N}\left[\begin{array}{c}\left(Y_i^m - \sum_{j=0}^{n}\xi_j x_{ij} + \Phi(0,s)\sum_{j=0}^{n}\left[(1-c)\left(d_i^R - \delta_i^R x_{ij}\right) + c\left(-d_i^L + \delta_i^L x_{ij}\right)\right]\right)^2 \\ +\left(\Phi(k,s)\sum_{j=0}^{n}\left[\left(d_i^R - \delta_i^R x_{ij}\right) + (d_i^L - \delta_i^L x_{ij}\right)^2\end{array}\right]\right\} \rightarrow min$$

5. *Conclusion:*

In this article a distance between fuzzy numbers is proposed. On the basis of this distance a fuzzy regression model is constructed for the case of polynomial type dependent variable and ordinary input variables. It should be noted that fuzzy numbers possessing such a polynomial form are quite universal and encompass the majority of cases that are considered in the literature of fuzzy numbers.

REFERENCES

Diamond, E., 1987. Fuzzy Least Squares Fitting of Several fuzzy Variables, in Analysis of Information, CRC Press, Tokyo., pp: 329-331.

D'urso, P., 2003. Linear Regression Analysis for fuzzy/Crisp Input and Fuzzy/crisp Output Data, Computational Statistics and data Analysis., 42: 47-72.

Kim., B., R.R. Bishu., 1998. Evaluation of Fuzzy Linear Regression Models by Comparing Membership Functions, Fuzzy Sets and Systems., 100: 343-352.

Ma, M., M. Kandel, M. Friedman, 2002. A New Approach for Defuzzification, Fuzzy Sets and Systems., 128: 351-356.

Ruoning., X., 1997. S-curve Regression Model in Fuzzy Environments, Fuzzy Sets and Systems., 90: 317-326.

Saneifard., R., 2009. Ranking L-R fuzzy numbers with weighted averaging based on levels, Int. J. Industrial Mathematics., 2: 163-173.

Saneifard., R., 2009. A method for defuzzification by weighted distance, Int. J. Industrial mathematics., 3: 209-217.

Tanaka, H., H. Ishibuchi, 1992. Possibilistic Regression Analysis Based on Linear Programming, in: J.Kacprzyk. Fuzzy Regression Analysis, Omnitech Press, Heidelberg., pp: 47-60.

Tanaka, H., S. Uejima, K. Asay, 1982. Linear Regression Analysis With Fuzzy Model, IEEE Trans. System, Man Cybernet., 12: 903-907.

Yen. K.K., S. Ghosray., G. Roig., 1999. A Linear Regression Model Using Triangular Fuzzy Number Coefficient, Fuzzy Sets and Systems., 106: 167-177.

Zimmermann, H.J., 1991. Fuzzy sets theory and its applications, Kluwer Academic Press, Dordrecht.