Journal of Korean
Data & Information Science Society
2007, Vol. 18, No. 2, pp. 515 ~ 524

Fuzzy Local Linear Regression Analysis

Dug Hun Hong[1] · **Jong-Tae Kim**[2]

Abstract

This paper deals with local linear estimation of fuzzy regression models based on Diamond(1998) as a new class of non-linear fuzzy regression. The purpose of this paper is to introduce a use of smoothing in testing for lack of fit of parametric fuzzy regression models.

Keywords : Fuzzy Linear Regression, Fuzzy Regression Models, Local Linear Estimation

1. Introduction

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

The concept of fuzzy regression analysis was introduced by Tanaka et al. (1982), where an LP-based method with symmetric triangular fuzzy parameters was proposed. Similar to traditional least-squares, Diamond (1998) defined a distance on a triangular fuzzy number space to measure the best fit for the regression model to observed data, and then derived regression parameters based on the distance. Diamond's model is well corresponding to traditional least squares regression. Diamond and Körner (1997) discussed the L_2 optimization of LR-fuzzy numbers and applied this concept to least squares estimation of linear models. Hong and Hwang(2003, 2004) studied for support vector regression machines and the extended fuzzy regression models using regularization method, respectively. Hong et al (2006) suggested the regression method of predicting fuzzy multivariable nonlinear regression models using triangular fuzzy numbers.

---

1) Professor, Department of Mathematics Myongji University, Kyunggido 449-728, Korea
2) Corresponding Author : Professor, Department of Computer Science & Statistics Daegu University, Kyungbook 712-714 Korea.
E-mail : jtkim@daegu.ac.kr

In this paper, we concentrate on the model of Diamond (1998). Diamond proposed the so-called fuzzy least squares. This paper introduces a new class of non-linear fuzzy regression which deals with local linear estimation of fuzzy regression models.

2. preliminaries

Let R be the set of real numbers and let $X=(m,\alpha,\beta)$ be a triangular fuzzy number where m is the modal value of X and α and β are the left and right spreads, respectively. Diamond(1998) gave a metric d on the space $T(R)$ of all triangular fuzzy numbers by

$$d(X,Y)^2=(m_X-m_Y)^2+((m_X-\alpha_X)-(m_Y-\alpha_Y))^2 \tag{2.1}$$
$$+((m_X+\beta_X)-(m_Y+\beta_Y))^2,$$

where $X=(m_X,\alpha_X,\beta_X)$ and $Y=(m_Y,\alpha_Y,\beta_Y)$ are any two triangular fuzzy numbers in $T(R)$. A linear structure is defined on $T(R)$ by

$$(m_X,\alpha_X,\beta_X)+(m_Y,\alpha_Y,\beta_Y)=(m_X+m_Y,\alpha_X+\alpha_Y,\beta_X+\beta_Y),$$

$$t(m,\alpha,\beta)=(tm,t\alpha,t\beta),\ \text{if}\ t\geq 0,\quad t(m,\alpha,\beta)=(tm,|t|\beta,|t|\alpha),\quad \text{if}\ t<0.$$

There are three simple fuzzy regression models considered in (2.1):

$$(\mathrm{F1}):\quad Y=a+bX,\ \ a,b\in R,\ X\in T(R),$$

$$(\mathrm{F2}):\quad Y=A+bX,\ \ b\in R,\ A,X\in T(R),$$

$$(\mathrm{F3}):\quad Y=A+Bx,\ \ x\in R,\ A,B\in T(R).$$

The corresponding least-squares optimization problems are:

$$(\mathrm{M1}):\ \min\{r(a,b)\}=\sum_{i=1}^{n}d(a+bX_i,Y_i)^2,$$
$$(\mathrm{M2}):\ \min\{r(A,b)\}=\sum_{i=1}^{n}d(A+bX_i,Y_i)^2,$$
$$(\mathrm{M3}):\ \min\{r(A,B)\}=\sum_{i=1}^{n}d(A+Bx_i,Y_i)^2.$$

The models are rigorously justified by a projection-type theorem for cones on a Banach space containing the cone of triangular fuzzy numbers. Here, we briefly review the local linear estimator of real function $r(x)$.

If the function r has two continuous derivatives, then for each $x \in [0,1]$ we have

$$r(u) \approx r(x) + (u - x) r'(x)$$

for all u in a small neighborhood of x. In other words, r is approximately linear in a neighborhood of x.

This suggests that we estimate r by fitting straight lines locally to the data. Suppose that we have data $(x_1, Y_1), \cdots, (x_n, Y_n)$ from the model (F3). Let K be a probability density function that is unimodal, symmetric about 0 and supported on $(-1,1)$, and define

$$D(b_0, b_1; x) = \sum_{i=1}^{n} (Y_i - b_0 - b_1 (x_i - x))^2 K\left(\frac{x - x_i}{h}\right),$$

where $h > 0$. Now choose the values of b_0 and b_1 that minimize $D(b_0, b_1; x)$. Calling these values $\hat{b}_0(x)$ and $\hat{b}_1(x)$, respectively, the local linear estimator of $r(x)$ is $\hat{b}_0(x)$.. The slope, $\hat{b}_1$, may be used to estimate the derivative $r'(x)$.

The local linear estimate of $r(x)$ solves a weighted least squares problem. The only data used in this problem are those for which x_i is within a bandwidth h of x. Since K is unimodal and symmetric about 0, the squared

$$\frac{\sum_{i=1}^{n} w_i(x) Y_i}{\sum_{i=1}^{n} w_i(x)}, \tag{2.2}$$

where

$$w_i(x) = K\left(\frac{x - x_i}{h}\right)(m_{n,2}(x) - (x - x_i) m_{n,1}(x))$$

and

$$m_{n,k}(x) = \sum_{j=1}^{n} K\left(\frac{x - x_j}{h}\right)(x - x_j)^k, \quad k = 1, 2.$$

The local linear estimator has a noteworthy property not possessed by kernel estimators. This property is explained by noting that (2.2) may be written as

$\sum_{i=1}^{n} \widetilde{w}_i(x)\, Y_i$ and that the weights $\widetilde{w}_i(x)$ satisfy

$$\sum_{i=1}^{n} \widetilde{w}_i(x) = 1 \quad \text{and} \quad \sum_{i=1}^{n} (x - x_i)\widetilde{w}_i(x) = 0$$

for each x in $[0,1]$.

In this paper, we will modify this idea for the purpose of estimating fuzzy local linear regression model.

3. Fuzzy local linear regression model

3.1 The model (F1) and (F2)

We first consider the model (F1) and (F2). We assume, throughout this section, that A, B, Y are symmetric triangular fuzzy numbers for computational simplicity. Suppose that observations consist of data pairs (X_i, Y_i), $i = 1,2,\cdots,n$ where $X_i = (x_i, \delta_i)$ and $Y_i = (y_i, \eta_i)$. We define

$$H_2(A, b; x) = \sum_{i=1}^{n} d\,(A + b(X_i - x), Y_i)^2\, K\left(\frac{x - x_i}{h}\right),$$

where $h > 0$ and $A = (a, \alpha)$. Now choose the value of A and b that minimize $H_2(A, b; x)$.

Calling these value $A^* = (a^*, \alpha^*)$ and b^*. We note that

$$d(A + b(X_i - x), Y_i)^2 = 3(a + b(x_i - x) - y_i)^2 + 2(\alpha + |b|\delta_i - \eta_i)^2$$

and hence

$$H_2(A, b; x) = 3\sum_{i=1}^{n} (a + b(x_i - x) - y_i)^2 K\left(\frac{x - x_i}{h}\right)$$
$$+ 2\sum_{i=1}^{n} (\alpha + |b|\delta_i - \eta_i)^2 K\left(\frac{x - x_i}{h}\right).$$

To minimize H_2, we consider

$$\frac{dH_2}{da} = 6\sum_{i=1}^{n}(a+b(x_i-x)-y_i)K\left(\frac{x-x_i}{h}\right) = 0,$$

$$\frac{dH_2}{db} = 6\sum_{i=1}^{n}(x_i-x)(a+b(x_i-x)-y_i)K\left(\frac{x-x_i}{h}\right)$$

$$+4\sum_{i=1}^{n}\mathrm{sgn}(b)\delta_i(\alpha+|b|\delta_i-\eta_i)K\left(\frac{x-x_i}{h}\right) = 0$$

$$\frac{dH_2}{d\alpha} = 4\sum_{i=1}^{n}(\alpha+|b|\delta_i-\eta_i)^2K\left(\frac{x-x_i}{h}\right) = 0.$$

For convenience, we denote $K\left(\frac{x-x_i}{h}\right)$ by K_i. Let

$$D(x) = \begin{bmatrix} \sum_{i=1}^{n}K_i & \sum_{i=1}^{n}(x_i-x)K_i & 0 \\ 3\sum_{i=1}^{n}(x_i-x)K_i & 3\sum_{i=1}^{n}(x_i-x)^2K_i+2\sum_{i=1}^{n}K_i\delta^{2_i} & 2\sum_{i=1}^{n}\mathrm{sgn}(b)\delta_iK_i \\ 0 & \sum_{i=1}^{n}\mathrm{sgn}(b)\delta_iK_i & \sum_{i=1}^{n}K_i \end{bmatrix}.$$

Therefore, we have

$$\begin{bmatrix} a^* \\ b^* \\ \alpha^* \end{bmatrix} = (D(x))^{-1}\begin{bmatrix} \sum_{i=1}^{n}y_iK_i \\ 3\sum_{i=1}^{n}(x_i-x)y_iK_i+2\sum_{i=1}^{n}\mathrm{sgn}(b)\delta_i\eta_iK_i \\ \sum_{i=1}^{n}\eta_iK_i \end{bmatrix}.$$

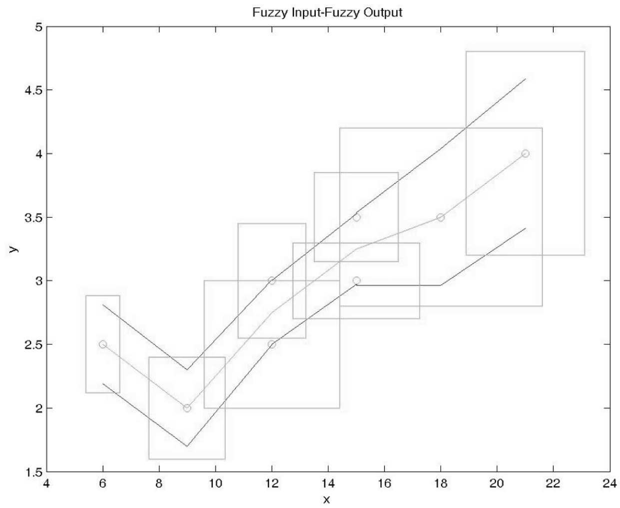
For the case of model (F1), by putting $\alpha=0$ in the model (F2), we have

$$\begin{bmatrix} a^* \\ b^* \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n}K_i & \sum_{i=1}^{n}(x_i-x)K_i \\ 3\sum_{i=1}^{n}(x_i-x)K_i & 3\sum_{i=1}^{n}(x_i-x)^2K_i+2\sum_{i=1}^{n}K_i\delta_i^2 \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{n}y_iK_i \\ 3\sum_{i=1}^{n}(x_i-x)y_iK_i+2\sum_{i=1}^{n}\mathrm{sgn}(b)\delta_i\eta_iK_i \end{bmatrix}$$

<Table1>. Fuzzy input-fuzzy output data

$X=(x,\delta)$	$Y=(y,\eta)$
(21.0, 2.1)	(4.0, 0.8)
(15.0, 2.25)	(3.0, 0.3)
(15.0, 1.5)	(3.5, 0.35)
(9.0, 1.35)	(2.0, 0.4)
(12.0, 1.2)	(3.0, 0.45)
(18.0, 3.6)	(3.5, 0.7)
(6.0, 0.6)	(2.5, 0.38)
(12.0, 2.4)	(2.5, 0.5)



<Figure.1 Fuzzy Input , Fuzzy Output>

3.2 The model (F3)

Suppose that observations consist of data pairs $(x_i, Y_i=(y_i,\eta_i))$, $i=1,2,\cdots,n$. Let K be a probability density function that is unimodal, symmetric about 0 and supported on $(-1,1)$. In association with the model (F3), consider the least-square optimization problem, choosing the value of A and B that minimize

$$H_3(A,B;x)=\sum_{i=1}^{n} d(A+B(x_i-x),\, Y_i)^2 K\left(\frac{x-x_i}{h}\right),$$

where $h>0$. Let $A=(a,\alpha)$ and $B=(b,\beta)$. Then we have

$$d(A+B(x_i-x), Y_i)^2 = 3(a+b(x_i-x)-y_i)^2 + 2(\alpha+\beta|x_i-x|-\eta_i)^2$$

and hence

$$H_3(A,B;x) = 3\sum_{i=1}^{n}(a+b(x_i-x)-y_i)^2 K\left(\frac{x-x_i}{h}\right)$$
$$+2\sum_{i=1}^{n}(\alpha+\beta|x_i-x|-\eta_i)^2 K\left(\frac{x-x_i}{h}\right).$$

To minimize $H_{3,}$ we consider

$$\frac{dH_3}{da} = 6\sum_{i=1}^{n}(a+b(x_i-x)-y_i)K\left(\frac{x-x_i}{h}\right)=0,$$

$$\frac{dH_3}{db} = 6\sum_{i=1}^{n}(x_i-x)(a+b(x_i-x)-y_i)K\left(\frac{x-x_i}{h}\right)=0$$

$$\frac{dH_3}{d\alpha} = 4\sum_{i=1}^{n}(\alpha+\beta|x_i-x|-\eta_i)K\left(\frac{x-x_i}{h}\right)=0.$$

$$\frac{dH_3}{d\beta} = 4\sum_{i=1}^{n}|x_i-x|(\alpha+\beta|x_i-x|-\eta_i)K\left(\frac{x-x_i}{h}\right)=0.$$

Therefore, by (2.2) the solution for $A=(a,\alpha)$ and$B=(b,\beta)$ is given by the solution $A^*=(a^*,\alpha^*)$ and $B^*=(b^*,\beta^*)$to the equation

$$b^* = \frac{\sum_{i=1}^{n}y_i(x_i-x)K\left(\frac{x-x_i}{h}\right)-a^*\sum_{i=1}^{n}(x_i-x)K\left(\frac{x-x_i}{h}\right)}{\sum_{i=1}^{n}(x_i-x)^2K\left(\frac{x-x_i}{h}\right)}$$

$$\beta^* = \frac{\sum_{i=1}^{n}y_i|x_i-x|K\left(\frac{x-x_i}{h}\right)-\alpha^*\sum_{i=1}^{n}|x_i-x|K\left(\frac{x-x_i}{h}\right)}{\sum_{i=1}^{n}(x_i-x)^2K\left(\frac{x-x_i}{h}\right)}$$

and

$$a^* = \frac{\sum_{i=1}^{n} w_i(x) y_i}{\sum_{i=1}^{n} w_i(x)}$$

where

$$w_i(x) = \sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right)\left(\sum_{i=1}^{n}(x_i-x)^2 K\left(\frac{x-x_i}{h}\right)\right.$$

$$\left. -(x_i-x)\sum_{i=1}^{n}(x_i-x)K\left(\frac{x-x_i}{h}\right)\right)$$

$$\alpha^* = \frac{\sum_{i=1}^{n} v_i(x) y_i}{\sum_{i=1}^{n} v_i(x)}$$

and

$$v_i(x) = \sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right)\left(\sum_{i=1}^{n}(x_i-x)^2 K\left(\frac{x-x_i}{h}\right)\right.$$

$$\left. -|x_i-x|\sum_{i=1}^{n}(x_i-x)K\left(\frac{x-x_i}{h}\right)\right).$$

$$\alpha^* = \frac{\sum_{i=1}^{n} v_i(x) \eta_i}{\sum_{i=1}^{n} v_i(x)},$$

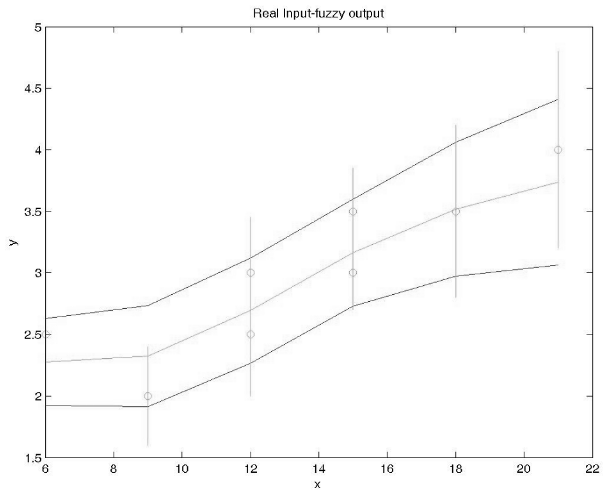
where

$$v_i(x) = \sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right)\left(\sum_{i=1}^{n}(x_i-x)^2 K\left(\frac{x-x_i}{h}\right)\right.$$

$$\left. -|x_i-x|\sum_{i=1}^{n}(x_i-x)K\left(\frac{x-x_i}{h}\right)\right).$$

<Table2>. Real input-fuzzy output data

x	$Y=(y,\eta)$
21.0	(4.0, 0.8)
15.0	(3.0, 0.3)
15.0	(3.5, 0.35)
9.0	(2.0, 0.4)
12.0	(3.0, 0.45)
18.0	(3.5, 0.7)
6.0	(2.5, 0.38)
12.0	(2.5, 0.5)



<Figure.2 Real Input , Fuzzy Output>

References

1. Diamond P.(1998), Fuzzy least squares, *Information Sciences* 46, 141-157.
2. Diamond P. and. Körner R.(1997), Extended fuzzy linear models and least squares estimates, *Computers & Mathematics with Applications* 33, 15-32.
3. Hong D. H. and Hwang C.(2003). Support vector fuzzy regression machines, *Fuzzy Sets and Systems* 138, 271-281.
4. Hong D. H. and Hwang C.(2004). Extended fuzzy regression models using regularization method, *Information Sciences*, 164, 31-46.
5. Hong D. H., Hwang C., Shim, J. and Seok, K.(2006). Locally weighted LS-SVM for fuzzy nonlinear regression with fuzzy input-output,

Proceeding of International Conference on Computational Intelligence and Security, 28-32,

6. Tanaka H,. Uejima S and Asia K.(1982), Linear regression analysis with fuzzy models, *Systems, Man and Cybernetics, IEEE Transactions on* 12, 903-907.

[received date : Feb. 2007, accepted date : May. 2007]