Journal of Korean
Data & Information Science Society
2004, Vol. 15, No. 3, pp. 617~624

The Rank Transform Method in Nonparametric Fuzzy Regression Model

Seung Hoe Choi[1] · Myung Sook Lee[2]

Abstract

In this article the fuzzy number rank and the fuzzy rank transformation method are introduced in order to analyse the non-parametric fuzzy regression model which cannot be described as a specific functional form such as the crisp data and fuzzy data as a independent and dependent variables respectively. The effectiveness of fuzzy rank transformation methods is compared with other methods through the numerical examples.

Keywords : Fuzzy rank transform method, Kernel smoothing, $k-$nearest neighbor smoothing, Nonparametric fuzzy regression model

1. Introduction

The regression models are used for explanation of the relationship between the response and explanatory variables in many fields such as the natural science, social science and the engineering. In the ordinary regression model, the difference between the estimated value and the observed one is assumed to be derived from the observation error and the method of estimation is based on the values to minimize the difference between actual value and estimated value.

However, sometimes it is difficult to find a statistical relation between the response and explanatory variable, especially when the dependent variable tends to be influenced by subjective judgment. Tanaka et al.(1982) introduced the parametric fuzzy regression model to present the relation between the dependent variable and the independent variable defined by fuzzy concept rather than statistical concept. The parametric fuzzy regression model with crisp coefficients

---

1) First Author : Associate Professor, Dept. of General Studies, Hankuk Aviation University, Koyang 411, Korea.
E-mail: shchoi@mail.hau.ac.kr

2) Lecturer, Dept. of Mathematics, Yonsei University, Seoul 120-749, Korea.

can be written as

$$Y_i = \beta_0 + \beta_1 X_{i1} + \ldots + \beta_p X_{ip},\ i = 1, 2, \ldots n$$

where β_is are unknown crisp parameters, $X_i = (X_{i1}, \cdots, X_{ip})$ is the ith fuzzy independent variable, and Y_i is the ith fuzzy response variable. The method of estimation in the parametric fuzzy regression model is based on the values to minimize the fuzziness in the predicted value of the dependent variable. The parametric fuzzy regression model has been studied a lot by many authors. See, Chang and Ayyub(2001), Cheng(1998), Kim and Chen(1997) and Savic and Pedrycz(1991). Since the spread of the estimated response variable becomes wide as the magnitude of the explanatory variables increases, even though the spreads of the observed responses are roughly constant or decreasing, the parametric fuzzy regression model could be a poor model.

To improve this problem, Cheng and Lee(1999) proposed the nonparametric fuzzy regression model with the independent variable x_i that is a crisp number and the dependent variable Y_i that is a fuzzy number. The nonparametric fuzzy regression model can be stated as follows;

$$Y_i = m(x_i) + \epsilon_i,\ i = 1, 2, \ldots, n$$

where m is a smooth function and ϵ_i is observational fuzzy error which is the difference between the observed value and the estimated ones. Cheng and Lee(1999) investigated several nonparametric approaches at the fuzzy regression such as the k-Nearest Neighbor ($k-NN$) and the kernel smoothing. These techniques are based on the concept of local weighted average. In case a function is smooth, local averaging can provide a good approximation for this function. The outlier in the some neighborhood, however, affects the calculation of center and also changes the spreads of the fuzzy coefficients. So, it needs to consider the rank transform in fuzzy data to reduce the influence of outlying observations.

In section 2 the fuzzy number rank and the fuzzy rank transformation methods are introduced for the nonparametric fuzzy regression model. The effectiveness of fuzzy rank transformation methods is compared with other methods through the simulations in the following section.

2 Fuzzy Rank Transformation Method

In the first place, we define on fuzzy set and fuzzy number to introduce the nonparametric fuzzy estimation.

Definition 2.1 Let X be a collection of objects x. A fuzzy set A in X is a set of ordered pairs:

$$A = \{ (x, \mu_A(x)) \mid x \in X \},$$

where $\mu_A : X \rightarrow [0,1]$ is the membership function of A.

Definition 2.2 A fuzzy number A is a fuzzy subset of the real line R with membership function $\mu_A(x)$ satisfying the following criteria:

(i) α-cut set of A, $A_\alpha = \{x \in X \mid \mu_A(x) \geq \alpha\}$, is a closed interval.

(ii) There exist a number x such that $\mu_A(x) = 1$.

(iii) $\mu_A(\lambda x_1 + (1-\lambda)x_2) \geq \min\{\mu_A(x_1),\ \mu_A(x_2)\}$ for $\lambda \in [0,1]$.

For a special fuzzy number, let L and R decrease in the interval $[0, \infty)$ and satisfy $L(-x) = L(x)$, $R(-x) = R(x)$, $L(0) = R(0) = 1$ and $L(1) = R(1) = 0$.

Definition 2.3 The fuzzy number A is said to be the LR type fuzzy number with the following membership function:

$$\mu_A(x) = \begin{cases} L\left(\dfrac{m-x}{e_A^l}\right) & \text{if } x \leqq m,\ e_A^l \geqq 0, \\ R\left(\dfrac{x-m}{e_A^r}\right) & \text{if } x > m,\ e_A^r \geqq 0, \\ 0 & otherwise, \end{cases}$$

where $A = (m, e_A^l, e_A^r)_{LR}$ with m as the center and e_A^l and e_A^r as the left and right spread of A. Specially, the LR type fuzzy number A is called triangular fuzzy number in case of $L(x) = R(x) = 1 - x$. Since the nonsymmetric LR-fuzzy number can be changed into the symmetric LR-fuzzy number, in this paper we always assume that the LR fuzzy numbers are symmetric and normal.

Now, to construct the nonparametric fuzzy regression analysis we consider several smoothing techniques such as the $k-NN$ smoothing method and kernel smoothing method which are developed by Cheng and Lee(1999).

The $k-NN$ smoothing estimate $\widehat{Y^*}$ of the new dependent value Y^* corresponding to a given value x^* is defined as

$$\widehat{Y^*} = \sum_{j=1}^{n} w_j(x) Y_j$$

where $w_j(x)$ is the weight sequence defined through the set of indexes

$$J_x=\{j : x_j \text{ is one of the } k\text{-nearest observations to } x^*\}.$$

With this set of indexes for neighbor observations, the $k-NN$ weight sequence $w_j(x)$ is constructed as

$$w_j(x) = \begin{cases} \frac{1}{k}, & \text{if } j \in J_x \\ 0, & otherwise \end{cases}.$$

Then, from the given data (Y_i, x_i) we have the regression equation

$$\widehat{Y^*}=(\sum_{j=1}^{n} w_j(x) m_j, \ \sum_{j=1}^{n} w_j(x) e_j)_{LR},$$

where $(m_i, e_i)_{LR}$ represents fuzzy number with center m_i and spread e_i.

On the other hand, we know that the Kernel function is a continuous, bounded, and symmetric real function which integrates to one. The probability density function is known as the kernel function.

The Kernel estimate $\widehat{Y^*}$ of the response value Y^* for the independent value x^* is defined as

$$\widehat{Y^*}= \sum_{j=1}^{n} w_j(x^*) Y_j$$

where $w_j(x^*) = \dfrac{K_h(x^* - x_j)}{P_h(x^*)}$, $P_h(x) = \sum_{j=1}^{n} K_h(x - x_j)$ and $K_h(u)=\frac{1}{h}K(\frac{u}{h})$ is the kernel with scale factor h.

Since $k-NN$ smoothing method and kernel smoothing method are based on the concept of local weighted average, these smoothing techniques are affected by the outliers sensitively. In order to reduce the influence of outlying observations we consider the rank transform procedure which was given by Iman and Conover(1979).

We define the rank of the fuzzy number before we apply the rank transform procedure to the fuzzy regression model.

Definition 2.4 Let $R(A)$ be the rank of the fuzzy number of A and $\underline{A}_\alpha \equiv inf A_\alpha$ and $\overline{A_\alpha} \equiv sup A_\alpha$ for the α-level set $A_\alpha = \{x | \mu_A(x) \geqq \alpha\}$.

(1) If $\underline{A}_\alpha \leqq \underline{B}_\alpha$ and $\overline{A_\alpha} \leqq \overline{B_\alpha}$ then $R(A) \leqq R(B)$.

(2) If $\underline{A}_\alpha \leqq \underline{B}_\alpha$ and $\overline{A_\alpha} \geqq \overline{B_\alpha}$ then $R(A) \geqq R(B)$.

(3) If $\underline{A}_\alpha = \underline{B}_\alpha$ and $\overline{A_\alpha} = \overline{B_\alpha}$ then it's average value is the rank of the fuzzy number.

The fuzzy rank transformation method used at finding predicted fuzzy number $\widehat{Y^*}$ of the dependent fuzzy number Y^* corresponding to the given independent value x^* follows:

1. Determine the rank of Y_i and x_i from Definition 2.4 and the given data (Y_i, x_i).

2. Determine the coefficient of simple regression model based on $R(u_{Y_i}^{-1}(\alpha))$ and $R(x_i)$ where $u_{Y_i}^{-1}(\alpha)$ is a inverse image of α under the membership function μ_{Y_i}. That is, find β such that $R\left(u_{Y_i}^{-1}(\alpha)\right) = \frac{(n+1)}{2} + \beta\left(R(x_i) - \frac{(n+1)}{2}\right)$.

3. Determine the rank of x^* from

$$R(x^*) = R(x_i) + (R(x_{i+1}) - R(x_i)) \times \frac{(x^* - x_i)}{(x_{i+1} - x_i)} \quad \text{when} \quad x_i < x^* < x_{i+1}.$$

4. Determine the rank of $R(u_{\widehat{Y^*}}^{-1}(\alpha))$ using $R(x_i^*)$ and the second step.

5. Determine the predict number $u_{\widehat{Y^*}}^{-1}(\alpha)$ from

$$u_{\widehat{Y^*}}^{-1}(\alpha) = Y_i + (Y_{i+1} - Y_i) \times \frac{R(u_{\widehat{Y^*}}^{-1}(\alpha)) - R(Y_i)}{R(Y_{i+1}) - R(Y_i)}$$

when $R(Y_i) < R(u_{\widehat{Y^*}}^{-1}(\alpha)) < R(Y_{i+1})$.

3. Effectiveness of the Fuzzy Rank Transformation Method

To investigate the performance the fuzzy rank transformation method in the nonparametric fuzzy regression model we examine two numerical examples given by Cheng and Lee(1999).

In the numerical examples we assume that the independent variable x_i is created 10 times larger than the uniform distribution of $U(0, 10)$. The center m_i, and the spread α_i of a dependent variable $Y_i = (m_i, \alpha_i)_{LR}$ are generated from the following two models. In the following models we suppose that η_i and ζ_i come from the $U(-0,5, 0.5)$, k_i comes from the uniform distribution $U(0, 5)$, and w_i comes from $U(0, 5)$.

Model 1. Consider $m_i = \frac{x_i^2}{5} + 2e^{\frac{x_i}{10}} + \eta_i$ and $\alpha_i = \frac{x_i^2}{20} + \frac{1}{2}e^{\frac{x_i}{10}} + \zeta_i$.

Model 2. Consider $m_i = \sin(x_i) + 2e^{\frac{x_i}{10}} + \omega_i$ and $\alpha_i = k_i + \zeta_i$.

Table1 shows the original data sets and the estimate data sets using the proposed methods with K=3, $h = \frac{3}{4}$ and kernel function

$$k(x) = \begin{cases} \frac{1}{4}x + \frac{1}{2}, & if -2 \leqq x \leqq 0 \\ -\frac{1}{4}x + \frac{1}{2}, & if -2 \leqq x \leqq 0 \end{cases}.$$

On the other hand, since the aim of fuzzy regression analysis is to minimize the fuzziness in the predicted value of the dependent variable we can take advantage of the value to minimize the difference of membership values between the observed and estimated fuzzy numbers as a measure for the effectiveness of the estimation method proposed in Section 2. For this, let $M(A,B)$ denotes the area bounded below by the difference of membership function of A and membership function of B and above by x-axis. Then $M(A,B)$ can be calculated as

$$M(A,B) = \int_{S_A \cup S_B} |\mu_A(x) - \mu_B(x)|\, dx,$$

where S_A and S_B are the supports of μ_A and μ_B, respectively.

Table1. The original fuzzy numbers and the estimated fuzzy numbers.

Model 1					Model 2				
X	Original	K-NN	Kernel	Rank	X	Original	K-NN	Kernel	Rank
8	(15,12,19)	(20,13,27)	(16,10,22)	(18,15,22)	8	(5,4,7)	(5,4,7)	(5,4,7)	(7,3,11)
3	(5,4,7)	(8,3,12)	(7,4,9)	(8,6,11)	4	(2,1,4)	(5,-3,13)	(4,-2,10)	(6,3,8)
10	(24.18,30)	(20,13,27)	(24,18,30)	(23,18,28)	10	(5,4,6)	(5,4,7)	(6,5,8)	(8,3,10)
6	(12,9,15)	(14,12,15)	(13,8,15)	(13,9,15)	5	(2,-2,7)	(6,13,2)	(3,3,2)	(6,3,8)
5	(9,7,11)	(9,6,3)	(9,5,14)	(10,8,12)	7	(4,2,6)	(4,-4,12)	(5,3,7)	(6,3,9)
7	(13,9,16)	(16,16,17)	(18,17,20)	(15,11,19)	1	(3,2,4)	(5,4,6)	(5,4,6)	(5,3,7)
1	(2,1,3)	(6,3,9)	(4,2,6)	(3,2,4)	2	(3,2,4)	(5,6,6)	(7,6,9)	(5,3,7)
2	(4,3,5)	(6,7,6)	(8,8,9)	(7,5,8)					

Since smaller value of the sum of the difference of the observed and the estimated membership function indicates that the nonparametric fuzzy regression model fits the data better, we can compare the fuzzy rank transformation methods with the other methods using the sum of the difference of the observed and the estimated membership function by $SAE = \sum_{i=1}^{n} M(Y_i, \widehat{Y}_i)$.

Table 2 shows SAEs for the $k-$nearest neighbor smoothing method, the kernel smoothing method and the rank transform method.

Table 2. Effectiveness of the estimation.

	Model1			Model2		
	k-NN	Kernel	Rank	k-NN	Kernel	Rank
SAE	25	19	14	28	19	17

In the Table 2, we know that the fuzzy rank transformation method is more effective than the kernel smoothing and $k-NN$ method in Model 1 and Model 2.

4. Conclusion

The fuzzy rank transform method to construct the nonparametric fuzzy regression model was developed in this article. Based on the results of the numerical examples in this paper, we know that the rank transform approach would do better perform than the two approach to nonparametric fuzzy regression model, the $k-$nearest neighbor smoothing method and the kernel smoothing method.

However, since the rank transform is frequently not enough to produce a good estimate in case that if the fuzzy data sets have spreads which are extremely changable, in future work we intend to derive the numerical results for the large data sets and asymptotic efficiency for developing the nonparametric fuzzy regression model.

Acknowledgements

We thank to the referees for constructive comments and careful readings of this paper.

References

1. Chang, Y. O., and Ayyub, B. M. (2001). Fuzzy regression methods-a comparative assessment, Fuzzy Sets and Systems, 119, 187-203.
2. Cheng, C. H. (1998). A new approach for ranking fuzzy numbers by distance methods, Fuzzy Sets and Systems, 95, 307-317.
3. Cheng, C. B., and Lee, E. S. (1999). Nonparametric fuzzy regression, An International Journal Computers and Mathematics with Applications 38, 239-251.
4. Iman, R. L., and Conver, W. J. (1979). The use of the rank transform in regression, Technometrics, 21, 499-509.
5. Kim, K. J., and Chen, H. R. (1997). A comparison of fuzzy and nonparametirc linear regression, Computers and Operations Research, 6, 505-519.
6. Savic, D., and Pedrycz, W. (1991). Evaluation of fuzzy regression models, Fuzzy Sets and Systems, 39, 51-63.
7. Tanaka, H., Uejima, S., and Asai, K. (1982), Linear regression analysis with fuzzy model, IEEE Trans. Systems Man Cybernet, 12, 903-907.

[received date : Mar. 2004, accepted date : Aug. 2004]