Fuzzy Regression based on nonparametric methods

Jinhee Yoon and Seunghoe Choi

Yonsei University
Korea Aerospace University
South Korea

2012 FEB. 2

Contents of Presentation

1.1 Ordinary Regression Model

- Ordinary regression model takes the following form:

$$y_i = f(x_i, \theta_o) + \epsilon_i, \quad i = 1, \cdots, n.$$

y_i : the dependent variable.
x_i: the independent variable.
θ_o : the true parameter.
ϵ_i: the observation error.

- In the ordinary regression, the deviation of the observed value from the estimated value is assumed to derive from the observation error and the estimating method is based on the values to minimize the difference between the actual and estimated value.

- Tanaka, H., Uejima, S., Asai, K., Linear regression analysis with fuzzy model, IEEE Trans. Systems, Man Cybernet 12 (1982), 903-907

LR-fuzzy numbers

- The membership function of the LR-fuzzy number A is

$$\mu_A(x) = \begin{cases} L_A\left((a-x)/l_a\right) & \text{if } \ 0 \le a-x \le l_a, \\ R_A\left((x-a)/r_a\right) & \text{if } \ 0 \le x-a \le r_a, \\ 0 & \text{otherwise.} \end{cases}$$

$\circ L_A(R_A)$: the monotonic decreasing function.
$\circ L_A(0) = R_A(0) = 1, L_A(1) = R_A(1) = 0.$

- The α-level set of the fuzzy number A can be represented as follows:

$$A(\alpha) = (a, l_a(\alpha), r_a(\alpha)),$$

where $l_a(\alpha) = l_a L_A^{-1}(\alpha)$ and $r_a(\alpha) = r_a R_A^{-1}(\alpha)$.

- Dubois, D. and Prade, H. Fuzzy real algebra and some results, Fuzzy Sets and Systems 2 (1979) 327-348.
H.-C. Wu. Fuzzy linear regression model based on fuzzy scalar product, Soft Computing 12 (2008) 469-477.

1.2 Fuzzy regression model

- The fuzzy regression model is expressed as follows:

$$Y(\mathbf{X}_i) = F(\mathbf{A}, \mathbf{X}_i)$$

$\mathbf{X}_i = (X_{ij})_{1\times(p+1)}$: the fuzzy input,

$\mathbf{A} = (A_j)_{1\times(p+1)}$: the fuzzy coefficients,

$F(\mathbf{A}, \mathbf{X}_i)$: the known linear function.

$Y(\mathbf{X}_i)$: the fuzzy output.

  - $A_k(\alpha) = (a_k, l_k(\alpha), r_k(\alpha))$,
  - $X_{ik}(\alpha) = (x_{ik}, l_{x_{ik}}(\alpha), r_{x_{ik}}(\alpha))$,
  - $Y_i(\alpha) = (y_i, l_{y_i}(\alpha), r_{y_i}(\alpha))$,

  where $(i = 1, 2, \cdots, N,\ k = 1, 2, \cdots, p)$

1.2 Fuzzy regression model

- The α-level set of the fuzzy linear regression model, denoted by $Y(\mathbf{X}_i)(\alpha)$, with fuzzy input and output, and crisp parameters are

$$\left(\sum_{k=0}^{p} a_k x_{ik}, \sum_{k=0}^{p} l_k(\alpha) l_{x_{ik}}(\alpha), \sum_{k=0}^{p} r_k(\alpha) r_{x_{ik}}(\alpha)\right)$$

where $0 < \alpha < 1$.

- L. A. Zadeh, The concept of linguistic variable and its application to approximate reasoning I, Inform. Sci. 8 (1975) 199-249.

2.1 *k*-nearest neighborhood(*k*-NN) method

- The *k*-nearest neighborhood estimate is the weighted average in a varying neighborhood. The neighborhood is defined as the *k*-nearest neighbors of in Euclidean distance. The smoother is defined as

$$\hat{Y}_i = S(x = X_i) = \sum_{j=1}^{N} \omega_j(x) y_j,$$

where $\hat{Y}_i$ is the estimate of Y_i, and $\omega_j(x)$ $(j = 1, 2, \cdots, N)$ is the weight sequence defined through the set of indices

$J_{X_i} = \{j : X_j$ is the one of the *k*-nearest observations to $X_i\}$

2.1 *k*-nearest neighborhood(*k*-NN) method

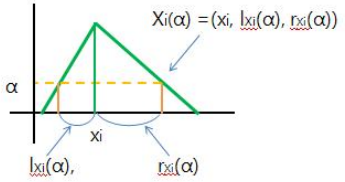
with this set of indexes for neighboring observations, the *k*-NN weight sequence is constructed as

$$\omega_j(x) = \begin{cases} \dfrac{1}{k} & \text{if} \quad j \in J, \\ 0 & \text{if} \quad \text{otherwise.} \end{cases}$$

- Let $X_{ik}(\alpha) = (x_{ik}, l_{x_{ik}}(\alpha), r_{x_{ik}}(\alpha))$ and $Y_i(\alpha) = (y_i, l_{y_i}(\alpha), r_{y_i}(\alpha))$ $(i = 1, 2, \cdots, N,\ k = 1, 2, \cdots, p)$ be non-symmetric fuzzy numbers.

2.1 *k*-nearest neighborhood(*k*-NN) method

- Let $l_{x_{ik}}(\alpha), l_{y_i}(\alpha)$ be the left(right) spreads of the $\alpha-$level sets of $X_{ik}(\alpha), Y_i(\alpha)$ $(i = 1, 2, \cdots, N,\ k = 1, 2, \cdots, p)$.



Then we have the intermediate regression equation:

$$\overline{l_{y_i}}(\alpha) = \sum_{j=1}^{N} \omega_j(l_{x_i k}(\alpha)) l_{y_j}(\alpha),$$

2.1 *k*-nearest neighborhood(*k*-NN) method

where

$$\omega_j(l_{x_{ik}}(\alpha)) = \begin{cases} \dfrac{1}{k} & \text{if } \; j \in l_{x_{ik}}(\alpha), \\ 0 & \text{if } \; \text{otherwise.} \end{cases}$$

Then the estimates of $\widehat{Y}_i(\alpha) = (\widehat{y}_i, \widehat{l_{y_i}}(\alpha), \widehat{r_{y_i}}(\alpha))$ are

$$\widehat{l_{y_i}}(\alpha) = \max\{\max_{\alpha \leqq s \leqq 1}\{\overline{l_{y_j}}(s), 0\}\},$$

$$\widehat{r_{y_i}}(\alpha) = \max\{\max_{\alpha \leqq s \leqq 1}\{\overline{r_{y_j}}(s), 0\}\},$$

$$\widehat{y}_i = \widehat{l_{y_i}}(1) = \max_{\alpha \leqq s \leqq 1} \overline{l_{y_j}}(s).$$

2.2 Median smoothing method

- The Median smoother is defined as

$$\hat{Y}_i = S(x = X_i) = \mathrm{med}_{j \in J_x}(Z_j,)$$

where Z_j is the observation at X_j and is real number. And

$J_{X_i} = \{j : X_j$ is the one of the k-nearest observations to $X_i\}$

- Let $X_{ik}(\alpha) = (x_{ik}, l_{x_{ik}}(\alpha), r_{x_{ik}}(\alpha))$ and $Y_i(\alpha) = (y_i, l_{y_i}(\alpha), r_{y_i}(\alpha))$ $(i = 1, 2, \cdots, N,\ k = 1, 2, \cdots, p)$ be non-symmetric fuzzy numbers.

2.2 Median smoothing method

- Then the estimates of $\widehat{Y_i}(\alpha) = (\widehat{y_i}, \widehat{l_{y_i}}(\alpha), \widehat{r_{y_i}}(\alpha))$ are

$$\widehat{l_{y_i}}(\alpha) = \max_{\alpha \leqq s \leqq 1} \{\text{med}\{l_{y_j}(s) : j \in J_{X_{ik}(s)}\}\},$$

$$\widehat{r_{y_i}}(\alpha) = \max_{\alpha \leqq s \leqq 1} \{\text{med}\{r_{y_j}(s) : j \in J_{X_{ik}(s)}\}\},$$

$$\widehat{y_i} = \widehat{l_{y_i}}(1) = \text{med}\{y_j : j \in J_{X_{ik}}\}.$$

2.3 Kernel method

- The kernel estimate $S(x)$ is defined as a weighted average of the response variable in a fixed neighborhood around x, determined in a shape by the kernel function K and the bandwidth h.
- Kernel estimate, $S(x)$, is defined as

$$\hat{Y}_i = S(s = X_{ik}) = \sum_{j=1}^{N} \omega_j(x) Y_j$$

and weight sequence is

$$\omega_j(x) = \frac{K_h(x - X_j)}{P_h(x)},$$

2.3 Kernel method

where $P_h(x) = \sum_{j=1}^{N} K_h(x - X_j)$ and $K_h(u) = \frac{1}{h}K(\frac{u}{h})$. $K_h(u)$ is called the kernel with scale factor h.

- Then the estimates of $\widehat{Y_i}(\alpha) = (\widehat{y_i}, \widehat{l_{y_i}}(\alpha), \widehat{r_{y_i}}(\alpha))$ are

$$\widehat{y_i} = \frac{\sum_{j=1}^{n} \frac{1}{h}K(\frac{|x_j - x_i|}{h})y_j}{\sum_{j=1}^{n} \frac{1}{h}K(\frac{|x_j - x_k|}{h})},$$

$$\widehat{l_{y_i}}(\alpha) = \frac{\sum_{j=1}^{n} \frac{1}{h}K(\frac{|l_{x_j}(\alpha) - l_{x_i}(\alpha)|}{h})l_{y_j}(\alpha)}{\sum_{j=1}^{n} \frac{1}{h}K(\frac{|l_{x_j}(\alpha) - l_{x_k}(\alpha)|}{h})},$$

$$\widehat{r_{y_i}}(\alpha) = \frac{\sum_{j=1}^{n} \frac{1}{h}K(\frac{|r_{x_j}(\alpha) - r_{x_i}(\alpha)|}{h})r_{y_j}(\alpha)}{\sum_{j=1}^{n} \frac{1}{h}K(\frac{|r_{x_j}(\alpha) - r_{x_k}(\alpha)|}{h})},$$

2.3 Kernel method

- The kernel function is continuous, bounded, and symmetric real function which integrates to one, $\int K(v)dv = 1$.
- The first common kernel is of the parabolic shape function, which is called the Epanechnikov's kernel,

$$K_1(v) = \begin{cases} \frac{3}{4}(1-v^2) & \text{if } \ |v| <= 1, \\ 0 & \text{otherwise.} \end{cases}$$

and the second is a Gaussian function

$$K_2(v) = (2\pi)^{-\frac{1}{2}} exp(-\frac{v^2}{2}).$$

3. Theil's Method

- We use Theil's method to estimate the model

$$l_{y_i}(\alpha) = c_0 + c_1 l_{x_{i1}}(\alpha) + \cdots + c_p l_{x_{ip}}(\alpha).$$

- Procedure for Theil's method on the basis of $(l_{x_{ik}}(\alpha), l_{y_i}(\alpha))$ $(i = 1, 2, \cdots, N,\ k = 1, 2, \cdots, p)$.

Step1] Apply Gram-schmidts process to $l_{x_{ik}}(\alpha)$ $(i = 1, 2, \cdots, N,\ k = 1, 2, \cdots, p)$.

$$l_{Z_k}(\alpha) = \begin{cases} l_{X_k}(\alpha) & \text{if} \quad k = 1, \\ l_{X_k}(\alpha) - \sum_{m=1}^{k-1} \frac{< l_{X_k}(\alpha), l_{Z_m}(\alpha) >}{< l_{Z_m}(\alpha), l_{Z_m}(\alpha) >} l_{Z_m}(\alpha) & \text{if} \quad k > 1, \end{cases}$$

where $l_{Z_k}(\alpha) = (l_{z_{1k}}(\alpha), \cdots, l_{z_{nk}}(\alpha))$.

3. Theil's Method

Step2] Estimate the coefficients θ_k in the regression model

$$l_{y_i}(\alpha) = \theta_o + \theta_1 l_{z_{i1}}(\alpha) + \cdots + \theta_p l_{z_{ip}}(\alpha).$$

1. $\hat{\theta}_o = \text{med}\{y_1, \cdots, y_n\}$.
2. Let $\theta_k^{(0)} = 0 (k = 1, \cdots, p)$ and $y_i^{(0)} = y_i (i = 1, \cdots, n)$.
3. $\delta\theta_k^{(0)} = \text{med}\{\frac{y_j^{(0)} - y_i^{(0)}}{l_{z_{jk}}(\alpha) - l_{z_{ik}}(\alpha)} : l_{jk}(\alpha) > l_{z_{ik}}(\alpha), i < j\}$.
4. $\theta_k^{(1)} = \theta_k^{(0)} + \delta\theta_k^{(0)}$ and $y_i^{(1)} = y_i^{(0)} - \delta\theta_k^{(0)} l_{z_{ki}}(\alpha)$.
5. Repeat the above process until the sequence $\{\theta_k^{(j)} : j = 1, \cdots\}$ converges.
6. Let the estimator $\hat{\theta}_k$ denotes the limit point of the sequence.

3. Theil's Method

Step3] Apply Gram-schmidts process to $\{\hat{\theta}_1, \cdots, \hat{\theta}_p\}$.

$$\hat{c}_{p-k} = \begin{cases} \hat{\theta}_p & \text{if } k = 0, \\ \hat{\theta}_{p-k} - \sum_{m=0}^{k-1} \frac{< l_{X_{p-m}}(\alpha), l_{Z_{p-k}}(\alpha) >}{< l_{Z_{p-k}}(\alpha), l_{Z_{p-k}}(\alpha) >} \hat{\theta}_{p-m} & \text{if } 1 \leq k, \\ \text{med}\{y_i - \sum_{k=1}^{p} \hat{\theta}_k l_{x_{ik}}(\alpha) : i = 1, \cdots, n\} & \text{if } k = p. \end{cases}$$

∘ We can obtain the predicted value of the left(right) spread $\widehat{l_{y_i}}(\alpha)(\widehat{r_{y_i}}(\alpha))$, and mode $\hat{y}_i = \widehat{l_{y_i}}(1)$ by applying the these procedures to $\{l_{X_{i1}}(\alpha), \cdots, l_{X_{ip}}(\alpha), l_{y_i}(\alpha)\}$ and $\{r_{X_{i1}}(\alpha), \cdots, r_{X_{ip}}(\alpha), r_{y_i}(\alpha)\}$.

Theil, H. A rank invariant of linear and polynomial regression analysis, Pro. Kon. Ned. Akad. Wetensch. A 53 (1950), 386-392.

4. Rank Transform Method

We use the RTM to estimate the value $Y(\mathbf{X}_i)(\alpha)$ based on $\{X_{i1}(\alpha), \cdots, X_{ip}(\alpha), Y_i(\alpha)\}$.

- Procedure for the RTM on the basis of $(l_{x_{ik}}(\alpha), l_{y_i}(\alpha))$ $(i = 1, 2, \cdots, N,\ k = 1, 2, \cdots, p)$.
  ◦ Determine the rank $R(l_{x_{ik}}(\alpha))$ of the observed value $_{x_{ik}}(\alpha)$ among $\{l_{x_{ik}}(\alpha), \cdots, l_{x_{nk}}(\alpha)\}$ and $R(l_{y_i}(\alpha))$ of $l_{y_i}(\alpha)$. ◦ Find the best regression equation using the sample $\{(R(l_{x_{i1}}(\alpha)), \cdots, R(l_{x_{ip}}(\alpha)), R(l_{y_i}(\alpha)))\}$ and the method of variable selection.

$$\hat{R}(l_{y_i}(\alpha)) = \beta_0 + \sum_{k=1}^{p_1} \beta_k R(l_{x_{ik}}(\alpha)) + \sum_{j=1}^{p_2} \sum_{k=1}^{p_3} \beta_{kj} R(l_{x_{ik}}(\alpha)) R(x_{ij}),$$

  where p_1, p_2 andp_3 are equal and less than p.
  ◦ Predict the output corresponding to the input $l_{X_i}(\alpha)$ using the rank $\hat{R}(l_{y_i}(\alpha))$.

4. Rank Transform Method

If $R(l_{y_{(i)}}(\alpha)) < \hat{R}(l_{y_i}(\alpha)) < R(l_{y_{(i+1)}}(\alpha))$, then
$\overline{l_{y_i}}(\alpha) = l_{y_{(i)}}(\alpha) + (l_{y_{(i+1)}}(\alpha) - l_{y_{(i)}}(\alpha)) \cdot \frac{\hat{R}(l_{y_i}(\alpha)) - R(l_{y_{(i)}}(\alpha))}{R(y_{(i+1)}) - R(y_{(i)})}$,
where $l_{y_{(i)}}(\alpha)$ denotes the i-th largest value among
$\{l_{y_1}(\alpha), \cdots, l_{y_n}(\alpha)\}$.
Otherwise, we have

$$\overline{l_{y_i}}(\alpha) = \begin{cases} l_{y_{(1)}}(\alpha) & \text{if } \hat{R}(l_{y_i}(\alpha)) < R(l_{y_{(1)}}(\alpha)), \\ l_{y_{(n)}}(\alpha) & \text{if } \hat{R}(l_{y_i}(\alpha)) > R(l_{y_{(n)}}(\alpha)), \\ l_{y_{(i)}}(\alpha) & \text{if } \hat{R}(l_{y_i}(\alpha)) = R(l_{y_{(i)}}(\alpha)). \end{cases}$$

Then, $\widehat{l_{y_i}}(\alpha) = \max_{\alpha \leqq s \leqq 1}\{\overline{l_{y_j}}(s)\}$,
∘ We can obtain the predicted value of the left(right) spread $l_{\hat{y}_i}(\alpha)(r_{\hat{y}_i}(\alpha))$ by applying the same procedure to $\{l_{X_{i1}}(\alpha), \cdots, l_{X_{ip}}(\alpha), l_{y_i}(\alpha)\}$ $(\{r_{X_{i1}}(\alpha), \cdots, r_{X_{ip}}(\alpha), r_{y_i}(\alpha)\})$.

■ Iman, R. L. and Conover, W, J. The use of the rank transform in regression, Technomerics 21 (1979) 499-509.

5. Numerical Example

- **Performance Measures**

  Consider a measure of performance for the estimated fuzzy regression model:

$$M_{\widehat{Y}} = \sum_{i=1}^{n} m(Y_i, \widehat{Y}_i)$$

where

$$m(Y_i, \widehat{Y}_i) = \frac{d(Y_i, \widehat{Y}_i)}{\int_{-\infty}^{\infty} \mu_{Y_i}(x)dx}$$

and $d(Y_i, \widehat{Y}_i) = \int_{-\infty}^{\infty} |\mu_{Y_i}(x) - \mu_{\widehat{Y}_i}(x)|dx.$

- Kim B., Bishu R. R. (1998) Evaluation of fuzzy linear regression models by comparing membership functions. Fuzzy Sets and Systems. **100**: 343-352.

5. Numerical Example

Table. Estimation Errors

Method	Error								Sum
Theil's Method	1.7564	1.1552	1.1099	1.5503	0.8106	1.0079	0.7764	1.7910	9.9578
Rank Transform Method	2.1235	1.2000	1.4343	1.4229	1.1894	1.4118	1.1543	1.9071	12.1152
Median	2.0688	1.2000	1.4000	1.6000	1.5000	1.8000	1.2875	1.6740	12.5284
3-NN Method	2.2060	1.2000	1.3185	1.1904	1.5000	1.8867	1.2862	1.9071	12.4949
Kernel Method	2.2259	1.0054	1.2974	1.5643	0.8129	1.6978	1.3188	1.6218	11.5442

- Iman, R. L. and Conover, W, J. The use of the rank transform in regression, Technomerics 21 (1979) 499-509.

6. Conclusions

- In this paper, we introduced introduced three smoothing methods(k-nearest neighborhood method, Median smoothing method, Kernel method), also introduced the rank transform method and the Theil's method, which are known for the distribution free method and does not depend on the error distribution, to construct the fuzzy regression model with predetermined response function. This paper has shown that the proposed fuzzy model using Theil's method superior than the other fuzzy regression models studied by many authors through some examples.