Fuzzy Statistics

Dr. Mohammed Jasim Mohammed

Fuzzy Regression Input &output data are fuzzy

FUZZY LEAST SQUARES METHOD

When input and output data are fuzzy, we will use fuzzy least squares method to estimate the parameters of the fuzzy regressing .

Triangular fuzzy numbers is defined as $X = (l_x, m_x, r_x)_{LR}$ where m_x is modal value of X, l_x is left spreads and r_x is right spreads. For fuzzy least squares method, when $x = (l_x, m_x, r_x)_{LR}$, $y = (l_y, m_y, r_y)_{LR}$ triangular fuzzy number is taken, following model will be considered

$$y_t = A_0 + A_1 x_{1t} + \varepsilon_t$$

Fuzzy Regression Input &output data are fuzzy

FUZZY LEAST SQUARES METHOD

where a, b are crisp numbers. When parameters are crisp, least squares optimization problem is defined as:

$$\min(S) = \sum d(A_0 + A_1 x_{1t}, y_t)^2$$

In this case d is defined as

$$d(A_0 + A_1 x_{1t}, y_t)^2 = [A_0 + A_1 x_{1t} - y_t - (A_1 l_x - l_y)]^2 + [A_0 + A_1 x_{1t} - y_t + (A_1 r_x - r_y)]^2 + (A_0 + A_1 x_{1t} - y_t)^2$$

Solving

$$\frac{\partial d}{\partial A_0} = (3A_0 + 3A_1 x_{1t} - 3y_t - A_1 l_x - l_y + A_1 r_x - r_y) = 0$$

$$\frac{\partial d}{\partial A_1} = [2(A_0 + A_1 x_{1t} - y_t - A_1 l_x + l_y)(x_{1t} - l_x) + 2(A_0 + A_1 x_{1t} - y_t + A_1 r_x - r_y)(x_{1t} + r_x) + (A_0 + A_1 x_{1t} - y_t) x_{1t}] = 0$$

Fuzzy Regression Input &output data are fuzzy

FUZZY LEAST SQUARES METHOD

We get

$$a_0 = \frac{[\sum_t^n x_t^2 + \sum_t^n (x_t - l_x)^2 + \sum_t^n (x_t + r_x)^2][\sum_t^n y_t + \sum_t^n (y_t - l_y) + \sum_t^n (y_t + r_y)] -}{3n[\sum_t^n x_t^2 + \sum_t^n (x_t - l_x)^2 + \sum_t^n (x_t + r_x)^2] - [\sum_t^n x_t + \sum_t^n (x_t - l_x) + \sum_t^n (x_t + r_x)]} \times$$

$$\frac{[\sum_t^n x_t + \sum_t^n (x_t - l_x) + \sum_t^n (x_t + r_x)][\sum_t^n x_t y_t + \sum_t^n (x_t - l_x)(y_t - l_y) + \sum_t^n (x_t + r_x)(y_t + r_y)]}{1}$$

Fuzzy Regression Input &output data are fuzzy

FUZZY LEAST SQUARES METHOD

We get

$$a_1 = \frac{3n[\sum_t^n x_t y_t + \sum_t^n (x_t - l_x)(y_t - l_y) + \sum_t^n (x_t + r_x)(y_t + r_y)] - [\sum_t^n x_t + \sum_t^n (x_t - l_x) + \sum_t^n (x_t + r_x)]}{3n[\sum_t^n x_t^2 + \sum_t^n (x_t - l_x)^2 + \sum_t^n (x_t + r_x)^2] - [\sum_t^n x_t + \sum_t^n (x_t - l_x) + \sum_t^n (x_t + r_x)]} \times$$

$$\frac{[\sum_t^n y_t + \sum_t^n (y_t - l_y) + \sum_t^n (y_t + r_y)]}{1}$$

Fuzzy Regression Input &output data are fuzzy

FUZZY LEAST SQUARES METHOD

For multi linear regression the model has been

$$y_t = A_0 + A_1 x_{1t} + A_2 x_{2t} \cdots + A_k x_{kt} + \varepsilon_t$$

least squares optimization problem is defined as:

$$\min(S) = \sum d(A_0 + A_1 x_{1t} + A_2 x_{2t} \cdots + A_k x_{kt}, y_t)^2$$

$$d(A_0 + A_1 x_{1t} + A_2 x_{2t} \cdots + A_k x_{kt}, y_t)^2 = [A_0 + A_1 x_{1t} + A_2 x_{2t} \cdots + A_k x_{kt} - y_t - (A_1 l_{x_1} + A_2 l_{x_2} \cdots + A_k l_{x_k} - l_y)]^2 +$$

$$[A_0 + A_1 x_{1t} + A_2 x_{2t} \cdots + A_k x_{kt} - y_t + (A_1 r_{x_1} + A_2 r_{x_2} \cdots + A_k r_{x_k} - r_y)]^2 +$$

$$(A_0 + A_1 x_{1t} + A_2 x_{2t} \cdots + A_k x_{kt} - y_t)^2$$

Fuzzy Regression Input &output data are fuzzy

FUZZY LEAST SQUARES METHOD

For multi linear regression the model has been

$$y_t = A_0 + A_1 x_{1t} + A_2 x_{2t} \cdots + A_k x_{kt} + \varepsilon_t$$

least squares optimization problem is defined as:

$$\min(S) = \sum d(A_0 + A_1 x_{1t} + A_2 x_{2t} \cdots + A_k x_{kt}, y_t)^2$$

$$d(A_0 + A_1 x_{1t} + A_2 x_{2t} \cdots + A_k x_{kt}, y_t)^2 = [A_0 + A_1 x_{1t} + A_2 x_{2t} \cdots + A_k x_{kt} - y_t - (A_1 l_{x_1} + A_2 l_{x_2} \cdots + A_k l_{x_k} - l_y)]^2 +$$

$$[A_0 + A_1 x_{1t} + A_2 x_{2t} \cdots + A_k x_{kt} - y_t + (A_1 r_{x_1} + A_2 r_{x_2} \cdots + A_k r_{x_k} - r_y)]^2 +$$

$$(A_0 + A_1 x_{1t} + A_2 x_{2t} \cdots + A_k x_{kt} - y_t)^2$$

Fuzzy Regression Input &output data are fuzzy

FUZZY LEAST SQUARES METHOD

When P problem is minimized, $\hat{\beta}$ is obtained as

$$\hat{\beta} = (X'X + C'C + D'D)^{-1}(X'\underline{Y} + C'\underline{E} + D'\underline{F})$$

Where

$$X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix}$$

$$\underline{Y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \underline{E} = \begin{bmatrix} y_1 - l_{y_1} \\ \vdots \\ y_n - l_{y_n} \end{bmatrix}, \underline{F} = \begin{bmatrix} y_1 + r_{y_1} \\ \vdots \\ y_n + r_{y_n} \end{bmatrix}$$

Fuzzy Regression Input &output data are fuzzy

FUZZY LEAST SQUARES METHOD

$$C = \begin{bmatrix} 1 & (x_{11} - l_{x_{11}}) & \cdots & (x_{1k} - l_{x_{1k}}) \\ 1 & (x_{21} - l_{x_{21}}) & \cdots & (x_{2k} - l_{x_{2k}}) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (x_{n1} - l_{x_{n1}}) & \cdots & (x_{nk} - l_{x_{nk}}) \end{bmatrix} \quad , D = \begin{bmatrix} 1 & (x_{11} + r_{x_{11}}) & \cdots & (x_{1k} + r_{x_{1k}}) \\ 1 & (x_{21} + r_{x_{21}}) & \cdots & (x_{2k} + r_{x_{2k}}) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (x_{n1} + r_{x_{n1}}) & \cdots & (x_{nk} + r_{x_{nk}}) \end{bmatrix}$$

Fuzzy Robust Regression Input &output data are fuzzy

In regression solution, many studies were done about parameter estimation in the event that outlier and were defined the Robust estimators.

We will found estimator by using the weighted fuzzy least squares method :

$$\hat{\beta} = (X^{'}WX + C^{'}WC + D^{'}WD)^{-1}(X^{'}W\underline{Y} + C^{'}W\underline{E} + D^{'}W\underline{F}))$$

Where W can found as follows:

Fuzzy Robust Regression Input &output data are fuzzy

Step 1: Estimation of regression parameters is obtained from above equation.

Step 2: $\hat{y}_t$ are estimated and residuals (e_t) are determined.

Step 3: According to absolute residual value, median is determined and distances are calculated $d_t = \left\| \, |e_t| - \text{Median}\, |e_t| \, \right\|$

Where $\|.\|$ is Euclide Distance .

Step 4: According to distance, membership function has been defined

$$M_{(d)} = \begin{cases} 1 & d \le Median(d) \\ \dfrac{\max(d) - |e|}{\max(d) - Median(d)} & Median(d) < d < \max(d) \\ 0 & o.w \end{cases}$$

Fuzzy Robust Regression Input &output data are fuzzy

Step 5: From above equation has been defined membership function, membership values are determined and weighted matrix (W) is constituted. Weighted matrix is diagonal matrix which diagonal elements are consist of membership value. weighted fuzzy least squares parameters coefficient is defined as

$$\hat{\beta} = (X'WX + C'WC + D'WD)^{-1}(X'W\underline{Y} + C'W\underline{E} + D'W\underline{F}))$$

As this parameters coefficient is used, regression parameters will be estimated.

Fuzzy Robust Regression Input &output data are fuzzy

Step 6: If $\left|\hat{\beta}^{k+1} - \hat{\beta}^{k}\right| < \varepsilon$ then stop. Otherwise is go to Step 2. Where $\hat{\beta}$ is estimates of regression model coefficients, k is iteration number and $\varepsilon > 0$ is a small number.

Fuzzy Robust Regression Input &output data are fuzzy

To comparison between models, we use mean square error.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n} d^2\left(Y_i, \widehat{Y_i}\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\left(l_{yi} - \hat{l}_{yi}\right)^2 + \left(m_{yi} - \widehat{m}_{yi}\right)^2 + \left(u_{yi} - \hat{u}_{yi}\right)^2\right)$$