Fuzzy Semi-Parametric Logistic Quantile Regression Model

Ayad H. shemail[a] Ahmed Razzaq Abed[b]

[a] Department of Statistics, College of Administration & Economics, University of Diyala, Iraq.

[b] Department of Statistics, College of Administration & Economics, University of Wasit, Iraq.

mailto:ayadstatistic@uodiyala.edu.iq[a] mailto:ahmedrazzaq@uowasit.edu.iq[b]

Abstract

In this paper, the fuzzy semi-parametric logistic quantile regression model was studied in the absence of special conditions in the classical regression models. This model becomes more flexible to deal with data for outlier's values and in the absence of a linear regression condition when the data of the dependent variable are restricted and fuzzy data and some of the variables are nonparametric, and the dependent variable in this model represents the fuzzy triangular number.

The estimation of the semi-parametric logistic quantile regression model has been applied to simulated data when the sample size is (25, 50, 75, 100) and with a repetition of 1000. Where the model is estimated in two steps, the first step is to estimate the parametric part, and the second step is to estimate the non-parametric part by the Nadaraya-Watson estimator through different kernel functions. Depending on the mean squares error and the measure of goodness of fit, the results indicated the best estimate of a model with the Kernel-Cassian function when the quantile of the fuzzy conditional distribution equals 0.2 for all sample sizes $\tau = 0.2$.

Key words: the fuzzy semi-parametric logistic quantile regression model, fuzzy triangular number, Nadaraya-Watson estimator, the quantile of the fuzzy conditional distribution.

1. Introduction

Regression analysis is generally defined as a mathematical measure of the average relationship between two or more variables, one of which is the dependent variable (response variable) and the other variables are the independent variables that are known as explanatory variables in terms of explanatory variables, and this relationship is often called linear regression .

Several researchers have recently resorted to using the quantile function (which is the inverse function of the cumulative function), in linear and non-linear regression such as quantile regression because it provides the ability to estimate the values of the dependent variable. Based on the measurement parameter (the standard deviation of the data), and the order of the error values by giving a rank for each value of the error and get the lowest error values. **(Koenker & Hallock ,2001)**

Quantile regression is a type of regression used in statistics and economic measurement, as it is known through the method of least squares. The conditional mean of the response variable is estimated at specific values of the explanatory variables, while quantile regression aims to estimate the conditional distribution of the response variable at different points, and basically, it is an extension of linear regression when the conditions for linear regression are not applicable. Quantile regression estimates are more immune to outliers in the response variable. However, the main attraction of quantile regression goes beyond that. Different measures of central tendency and statistical dispersal may be useful for obtaining a more comprehensive analysis of relationships between variables. **(Koenker & Bassett ,1978).**

The researchers Koenker and Bassett (1978) were the first to study the quantitative regression model by relying on the inverse of the cumulative function. The researcher Koenker & Pin (2005) presented a comprehensive description of quantitative regression instead of traditional linear regression, as quantitative regression is a powerful tool for comparison in a more comprehensive way to deal with data for anomalous values and in the absence of the condition for linear regression.

When the data of the dependent variable is restricted data and is limited to a limited period, it can be closed, open, or half-closed, and it is often confined between zero and unity. There are many examples of restricted data in medical research, and drawing data in such a type is more like a letter s. To analyze specific results, traditional statistical methods, such as least squares regression, mixed effects models, and even traditional non-parametric methods, may give poor results in statistical analysis, so the researchers Bottai et al (2010) studied the use of a quantitative regression model based on the logistic transformation of quantities for the finite outcome variable.

The researchers Galarza & et al (2020). Studied a logistic quantum regression model using a logistic correlation function through EM algorithm with long tail distributions by relying on a packet R.

sometimes independent variables are not controlled within a certain distribution, so they are called nonparametric variables. In addition to that time-restricted dependent variable, it may be a fuzzy variable. As an extension of the researchers Bottai & et al (2010), the quantitative regression will be studied on the logistic transfer function according to certain conditions, including that the dependent variable represents the fuzzy number. Trigonometric and some independent variables are nonparametric, so the model is called a fuzzy semi-parametric logistic quantitative regression model.

2. Fuzzy set and triangular fuzzy number

The fuzzy set defines a set of numbers z belonging to the universal set Z with a certain membership function$\mu_{\bar{A}(z)}$, i.e., they are ordered in pairs as follows: **(Salmani & et al, 2017).**

$$\bar{A} = \{(z, \mu_{\bar{A}(z)}) \setminus z \in Z\} \tag{1}$$

Where $\mu_{\bar{A}(z)}$ is the grade of membership of z in $\bar{A}$, $\mu_{\bar{A}(z)}$belong to the interval [0, 1] $\{\mu_{\bar{A}(z)} \in [0,1]\}$.

Let $\tilde{Z} = (l_z, z, r_z)$ a triangular fuzzy number is defined to be a fuzzy set of the set of real numbers R, It has a membership function according to the following formula: **(Gani & Assarudeen, 2012)**

$$\mu_{\tilde{Z}(x)} = \begin{cases} 0 & for\ x<l_z \\ \frac{x-l_z}{z-l_z} & for\ l_z\le x\le z \\ \frac{r_z-x}{r_z-z} & for\ z\le x\le r_z \\ 0 & for\ x>r_z \end{cases} \quad (2)$$

The membership functions above hold the following conditions

1- l_z to z is increasing function.
2- z to r_z is decreasing function.
3- $l_z \le z \le r_z$

The triangular fuzzy number $\tilde{Z} = (l_z, z, r_z)$ is called positive triangular fuzzy number if $l_z, z, r_z > 0$ and called negative triangular fuzzy number if$l_z, z, r_z < 0$.

Let two triangular fuzzy numbers $\tilde{Z} = (l_z, z, r_z)$ and $\tilde{C} = (l_c, c, r_c)$ the arithmetic operations of triangular fuzzy numbers is as follows:

1- The triangular fuzzy numbers $\tilde{Z} = (l_z, z, r_z)$ and $\tilde{C} = (l_c, c, r_c)$ are equal if and only if $l_z = l_c, z = c$ and $r_z = r_c$.
2- The addition of triangular fuzzy numbers is as follows $\tilde{Z} + \tilde{C} = (l_z + l_c, z + c, r_z + r_c)$.
3- The subtraction of triangular fuzzy numbers is as follows $\tilde{Z} - \tilde{C} = (l_z - l_c, z - c, r_z - r_c)$.
4- The multiplication of triangular fuzzy numbers is as follows$\tilde{Z} * \tilde{C} = (\min(l_z l_c, l_z r_c, r_z l_c, r_z r_c), zc, max(l_z l_c, l_z r_c, r_z l_c, r_z r_c))$.
5- The division of triangular fuzzy numbers is as follows $\tilde{Z}/\tilde{C} = (\min\left(\frac{l_z}{l_c}, \frac{l_z}{r_c}, \frac{r_z}{l_c}, \frac{r_z}{r_c}\right), z/c, max\left(\frac{l_z}{l_c}, \frac{l_z}{r_c}, \frac{r_z}{l_c}, \frac{r_z}{r_c}\right))$

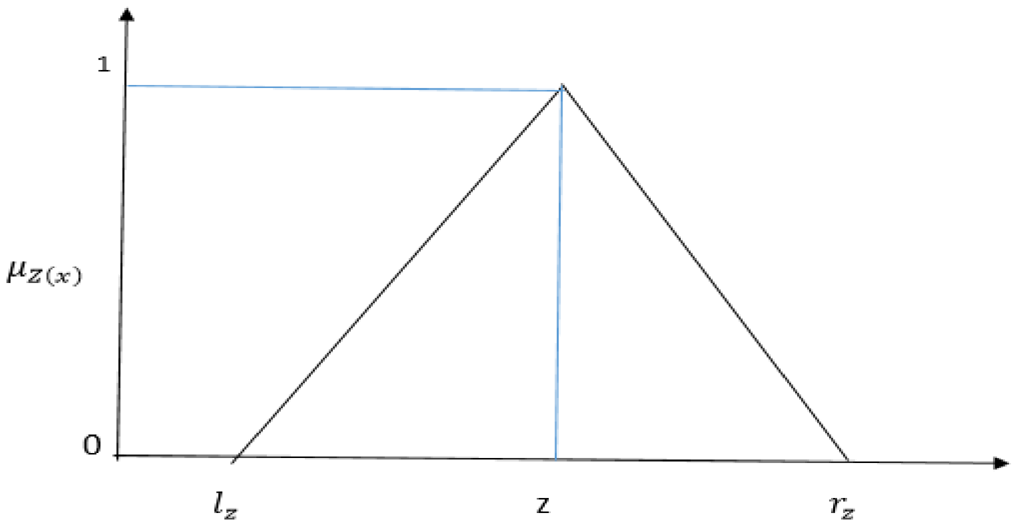

Figure 1: the triangular fuzzy number

3. Fuzzy Semi-Parametric Logistic Quantile Regression

Based on the same method presented by the researchers (Bottai & et al.) in 2010, we assume have n observation with some fuzzy continuous outcome $\tilde{y}_i = \left(l_{y_i}, y_i, r_{y_i}\right), i = 1,2, \dots \dots n$, where the limits of the fuzzy dependent variable are defined $\tilde{y}_{min}$ and $\tilde{y}_{max}$, a set of k independent parametric variables $x_i = \{x_{1i}, x_{2i}, \dots \dots x_{ki}\}^T$, and non-parametric variable t_i in smoothing function $f(t_i)$, The fuzzy semi-parametric quantile regression model is**. (Columbu & Bottai, 2017).**

$$\tilde{y}_i = x_i^T \tilde{\beta}(\tau) + f(t_i) + \varepsilon_i \quad (3)$$

Where $\tilde{\beta}(\tau) = \left(l_{\beta(\tau)}, \beta(\tau), r_{\beta(\tau)}\right) = \{\tilde{\beta}_1(\tau), \tilde{\beta}_2(\tau), \dots., \tilde{\beta}_P(\tau)\}^T$ the fuzzy regression parameters, for all $\tau \in (0,1)$, ε_i represent error term we assume that $P(\varepsilon_i < 0 \backslash x_i, t_i) = \tau$ or equivalently, that$P\left(\tilde{y}_i < x_i^T \tilde{\beta}_P + f(t_i) \backslash x_i, t_i\right) = \tau$, where τ represent the quantile of the fuzzy conditional distribution of $\tilde{y}_i$ given xi, t_i,the fuzzy semi-parametric logistic quantile regression function as follow: **(Orsini & Bottai, 2011).**

$$Q_{\tilde{y}}(\tau) = \frac{exp\left(x_i^T \tilde{\beta}(\tau) + f(t_i)\right)\tilde{y}_{max} + \tilde{y}_{min}}{1 + exp\left(x_i^T \tilde{\beta}(\tau) + f(t_i)\right)} \quad (4)$$

When$\tau = 0.5$, then $Q_{\tilde{y}}(0.5)$ represent the conditional median, it is the value that divides the conditional distribution of the fuzzy response variable into two parts with equal probability, Since there are no other assumptions regarding the error term.

The fuzzy semi-parametric logistic quantitative regression has many properties that are different from classical regression it is equivalent to monotonic transformations of the fuzzy outcome variable that is $Q_{h\tilde{y}}(\tau) = h\{Q_{\tilde{y}}(\tau)\}$ when any increasing function h, this property is exploited to define fuzzy semi-parametric logistic quantitative regression to model continuous outcomes that are constrained within a time interval as $\tilde{y} \in (\tilde{y}_{min}, \tilde{y}_{max})$.

Where $\tilde{y}_{min}, \tilde{y}_{max}$ the limits of the feasible interval of the fuzzy outcome variable do not represent the smallest and largest observed sample values.
We assume there exists a set of fuzzy parameters $\tilde{\beta}(\tau) = \left(l_{\beta(\tau)}, \beta(\tau), r_{\beta(\tau)}\right)$for any quantile *p*, and a known increasing function *h* from the interval $(\tilde{y}_{min}, \tilde{y}_{max})$ to the real line referred to as link function, such that

$$h^*\{Q_{\tilde{y}}(\tau)\} = x_i^T \tilde{\beta}(\tau) + f(t_i) \quad (5)$$

Since the fuzzy dependent variable is continuous, from among a set of options to the link function h^*, the logistic transformation was chosen as follows: **(Orsini & Bottai, 2011).**

$$h^*(\tilde{y}_i) = log\left(\frac{\tilde{y}_i - \tilde{y}_{min}}{\tilde{y}_{max} - \tilde{y}_i}\right) = logit(\tilde{y}_i) \quad (6)$$

Regression coefficients can be estimated using fuzzy semi-parametric logistic quantitative regression by regressing the transformed outcome $h^*(\tilde{y}_i)$ as follows:

$$Q_{h^*(\widetilde{y})}(\tau) = Q_{logit(\widetilde{y})}(\tau) = x_i^T \tilde{\beta}(\tau) + f(t_i) \quad (7)$$

The above transformation that depends on the linkage function h is similar to the transformation used in classical logistic regression.

4. Estimated Model

In this section, the fuzzy semi-parametric logistic quantile regression model will be estimated after it has been transformed into a linear form, through two steps; first, the parametric part (the fuzzy parameters) is estimated relying on the following equation: (Koenker, 2005).

$$min_{\tilde{\beta}} \sum_{i=1}^{n} \rho_\tau(\tilde{y}_i - x_i^T \tilde{\beta}(\tau)) \quad (8)$$

Where

$$\rho_{\theta\tau} = \begin{cases} \tau z & if\ z \geq 0 \\ -(1 - \tau z) & if z < 0 \end{cases}$$

In the second step; the non-parametric part represents the fuzzy smooth function $f(t_i)$ is estimated to the model as follows**: (Shemail & Mohammed, 2022).**

$$\hat{f}(t_i) = \sum_{i=1}^{n} W_j(t_i) \otimes (\tilde{y}_i \ominus (\oplus_{j=1}^{p} (\tilde{\beta}(\tau) \oplus x_{ij}))) \quad (9)$$

$$\hat{f}(t_i) = (\sum_{i=1}^{n} W_j(t_i) \otimes \tilde{y}_i) \ominus \sum_{i=1}^{n} W_j(t_i) \otimes (\oplus_{j=1}^{p} (\tilde{\beta}(\tau) \oplus x_{ij}))$$

Where

$$W_j(t_i) = \frac{k(\frac{t_i - t_j}{h})}{\sum_{i=1}^{n} k(\frac{t_i - t_j}{h})} \qquad (10)$$

$W_j(t_i)$ Represent the smoothing weights and sum of the weights equal to one$\sum_{i=1}^{n} w_{hi}(t_j) = 1$, $k(.)$ the kernel function, h represent bandwidth. We depending on Nadaraya-Watson to estimate$W_j(t_i)$.

By referring to the equation 9, we get the estimation of the fuzzy semi-parametric logistic quantile regression model as follow: **(Hesamian& Akbari, 2017)**

$$\hat{\tilde{y}}_i = \oplus_{j=1}^{p}(\hat{\tilde{\beta}}(\tau) \oplus x_{ij}) + (\sum_{i=1}^{n} W_j(t_i) \otimes \tilde{y}_i) \ominus \sum_{i=1}^{n} W_j(t_i) \otimes (\oplus_{j=1}^{p}(\hat{\tilde{\beta}}(\tau) \oplus x_{ij})) \qquad (11)$$

$$\hat{\tilde{y}}_i = (\sum_{i=1}^{n} W_j(t_i) \otimes \tilde{y}_i + \oplus_{j=1}^{p}(\hat{\tilde{\beta}}(\tau) \oplus x_{ij}^*) \qquad (12)$$

Where $x_{ij}^* = x_{ij} - (\sum_{i=1}^{n} W_j(t_i) x_{ij}$ Relying on the triangular fuzzy number, the fuzzy model estimate is as follows:

$$\hat{y}_i = \sum_{i=1}^{n} W_j(t_i) \otimes y_i + \oplus_{j=1}^{p}(\hat{\beta}(\tau) \oplus x_{ij}^*)$$
$$l_{\hat{y}_i} = \sum_{i=1}^{n} W_j(t_i) \otimes l_{y_i} + \oplus_{j=1}^{p}(l_{\hat{\beta}(\tau)} S_{ji} x_{ij}^*) - \oplus_{j=1}^{p}(r_{\hat{\beta}(\tau)}(1 - S_{ji}) x_{ij}^*$$
$$r_{\hat{y}_i} = \sum_{i=1}^{n} W_j(t_i) \otimes r_{y_i} + \oplus_{j=1}^{p}(r_{\hat{\beta}(\tau)} S_{ji} x_{ij}^*) - \oplus_{j=1}^{p}(l_{\hat{\beta}(\tau)}(1 - S_{ji}) x_{ij}^*$$

And

$$S_{ji} = \begin{cases} 1 & x_{ij}^* \geq 0 \\ 0 & x_{ij}^* < 0 \end{cases}$$

Based on a matrix form the estimation of the fuzzy semi-parametric logistic quantile regression model as follow:

$$\hat{\tilde{y}} = [\mathrm{Wy} + X^* \hat{\beta}, W L_y + X_S^* L_{\hat{\beta}} - X_{1-S}^* R_{\hat{\beta}}, W R_y + X_S^* R_{\hat{\beta}} - X_{1-S}^* L_{\hat{\beta}}] \quad (13)$$

Where

$$W = \begin{bmatrix} w_{11} & w_{12} & . & . & . & w_{1n} \\ w_{21} & w_{22} & . & . & . & w_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ . & . & . & . & . & . \\ w_{n1} & w_{n2} & . & . & . & w_{nn} \end{bmatrix} \quad X = \begin{bmatrix} x_{11} & x_{12} & . & . & . & x_{1p} \\ x_{21} & x_{22} & . & . & . & x_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ . & . & . & . & . & . \\ x_{n1} & x_{n2} & . & . & . & x_{np} \end{bmatrix}$$

$$X^* = \begin{bmatrix} x_{11}^* & x_{12}^* & . & . & . & x_{1p}^* \\ x_{21}^* & x_{22}^* & . & . & . & x_{2p}^* \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ . & . & . & . & . & . \\ x_{n1}^* & x_{n2}^* & . & . & . & x_{np}^* \end{bmatrix} X_s^* = \begin{bmatrix} s_{11}x_{11}^* & s_{12}x_{12}^* & . & . & . & s_{1p}x_{1p}^* \\ s_{21}x_{21}^* & s_{22}x_{22}^* & . & . & . & s_{2p}x_{2p}^* \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ . & . & . & . & . & . \\ s_{n1}x_{n1}^* & s_{n2}x_{n2}^* & . & . & . & s_{np}x_{np}^* \end{bmatrix}$$

$$X_{1-S}^* = \begin{bmatrix} (1-s_{11})x_{11}^* & (1-s_{12})x_{12}^* & . & . & . & (1-s_{1p})x_{1p}^* \\ (1-s_{21})x_{21}^* & (1-s_{22})x_{22}^* & . & . & . & (1-s_{2p})x_{2p}^* \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ . & . & . & . & . & . \\ (1-s_{n1})x_{n1}^* & (1-s_{n2})x_{n2}^* & . & . & . & (1-s_{np})x_{np}^* \end{bmatrix}$$

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ . \\ . \\ . \\ \hat{\beta}_p \end{bmatrix}, L_{\hat{\beta}} = \begin{bmatrix} l_{\hat{\beta}_1} \\ l_{\hat{\beta}_2} \\ . \\ . \\ . \\ l_{\hat{\beta}_p} \end{bmatrix}, R_{\hat{\beta}} = \begin{bmatrix} r_{\hat{\beta}_1} \\ r_{\hat{\beta}_2} \\ . \\ . \\ . \\ r_{\hat{\beta}_p} \end{bmatrix} y = \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix}, L_y = \begin{bmatrix} l_{y_1} \\ l_{y_2} \\ . \\ . \\ . \\ l_{y_n} \end{bmatrix}, R_y = \begin{bmatrix} r_{y_1} \\ r_{y_2} \\ . \\ . \\ . \\ r_{y_n} \end{bmatrix}$$

5. Goodness of Fit and Mean Square Error

The comparison was made between the estimated fuzzy semi-parametric logistic quantile regression model when the dependent variable represents the fuzzy triangular number, we use the goodness of fit $S(\tilde{y}_i, \hat{\tilde{y}}_i)$ and mean square error $MSE(\tilde{y}_i)$ is as follow: **(Hesamian& Akbari, 2017)**

$$S(\tilde{y}_i, \hat{\tilde{y}}_i) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{1+DK(\tilde{y}_i, \hat{\tilde{y}}_i)^2} \quad (14)$$

$$S(\tilde{y}_i, \hat{\tilde{y}}_i) = \frac{1}{n}\sum_{i=1}^{n} DK(\tilde{y}_i, \hat{\tilde{y}}_i)^2 \quad (15)$$

Where $DK(\tilde{y}_i, \hat{\tilde{y}}_i)^2$ The squared distance between the fuzzy dependent variable$\tilde{y}_i$ and the model estimate $\hat{\tilde{y}}_i$, there are several ways to calculate the distance as follows: **(Hesamian & Akbari,2019).**

$$DK_1(\tilde{y}_i, \hat{\tilde{y}}_i) = y - \hat{y} - 0.25(L_y - L_{\hat{y}}) + 0.25(R_y - R_{\hat{y}}) \quad (16)$$

$$DK_2(\tilde{y}_i, \hat{\tilde{y}}_i) = y - \hat{y} - 0.167(L_y - L_{\hat{y}}) + 0.167(R_y - R_{\hat{y}}) \quad (17)$$

$$DK_3(\tilde{y}_i, \hat{\tilde{y}}_i) = y - \hat{y} - 0.221(L_y - L_{\hat{y}}) + 0.221(R_y - R_{\hat{y}}) \quad (18)$$

In which

$K_1(x) = 1 \quad for\ |x| \le 1$

$K_2(x) = (1 - |x|) \quad for\ |x| \le 1$

$K_3(x) = 35/32(1 - |x|^2)^3 \quad for\ |x| \le 1$

6. Simulation

In this section, the simulation experiment will be conducted using the R program by using four different sample sizes, 25, 50, 75, and 100, with a repetition of 100. The simulation experiment will be presented through the following steps:

1- Three independent variables are generated according to the normal distribution with different means and variances as follows:

$X_1 \sim (0,1)$
$X_2 \sim (1.3,0.8)$
$X_3 \sim (2.4,1.4)$

2- The non-parametric variable is generating as follow: **(Hesamian & Akbari,2019).**

$t_i = \frac{i}{n}$, $f(t_i) = 1 + \frac{t_i(t_i+0.5)}{t_i^3+1} * \sin(t_i)$

3- The initial values of the parameters represent $\beta_0 = 0.5, \beta_1 = 0.02, \beta_2 = -0.3, \beta_3 = 0.4$ when the τ quantile of the fuzzy conditional distribution of $\tilde{y}_i$ given xi, t_i is $\tau = 0.2,0.4,0.6,0.8,1$.

4- The fuzzy error term is generated by exponential distribution and uniform distribution as follows:
$e_i \sim \exp(8.2)$ $l_{e_i} \sim unif(0.6,0.8)$ $r_{e_i} \sim (0.9,1.2)$

5- The fuzzy dependent variable $\tilde{y}_i = (l_{y_i}, y_i, r_{y_i})$ is generated by relying on the independent variables, the nonparametric variable, the parameters, and the fuzzy error term as follows:

$l_{y_i} = x_i^T \tilde{\beta}(\tau) + f(t_i) + l_{e_i}$
$y_i = x_i^T \tilde{\beta}(\tau) + f(t_i) + \varepsilon_i$
$y_i = x_i^T \tilde{\beta}(\tau) + f(t_i) + \varepsilon_i$

The comparison of estimation methods is made in the fuzzy semi-parametric logistic quantile regression model when using triangular fuzzy number represent an independent variable, kernel function represent (Gaussian, Epanechnikov, Triangle) and τ equal to (0.2,0.4,0.6,0.8,1) based on the mean square error MSE(V) and the criterion goodness of fit $S(\widehat{V}, V)$ according to the following tables

Table 1: represent $S(\tilde{y}_i, \hat{\tilde{y}}_i)$ & $S(\tilde{y}_i, \hat{\tilde{y}}_i)$ to model estimator when sample size 25

Kernel function	Performance	τ=0.2	τ=0.4	τ=0.6	τ=0.8	τ=1
Gaussian	$S(\tilde{y}_i, \hat{\tilde{y}}_i)$	1.131384	1.159876	1.171052	0.9400513	0.326302
	$MSE(\tilde{y}_i)$	0.0105017	0.0271406	0.0484818	0.2244429	2.109086
Epanechnikov	$S(\tilde{y}_i, \hat{\tilde{y}}_i)$	1.151352	1.186768	1.208737	0.9653818	0.327589
	$MSE(\tilde{y}_i)$	0.0117612	0.0292926	0.0513755	0.2290452	2.113122
Triangle	$S(\tilde{y}_i, \hat{\tilde{y}}_i)$	1.15128	1.188286	1.207518	0.9658705	0.327604
	$MSE(\tilde{y}_i)$	0.0124797	0.0297677	0.0515621	0.2294883	2.11359

Table 2: represent $S(\tilde{y}_i, \hat{\tilde{y}}_i)$ & $S(\tilde{y}_i, \hat{\tilde{y}}_i)$ to model estimator when sample size 50

Kernel function	Performance	τ=0.2	τ=0.4	τ=0.6	τ=0.8	τ=1
Gaussian	$S(\tilde{y}_i, \hat{\tilde{y}}_i)$	1.248435	1.173173	1.095539	0.9479852	0.3178901
	$MSE(\tilde{y}_i)$	0.01362595	0.03535306	0.0794502	0.2267212	2.198936
Epanechnikov	$S(\tilde{y}_i, \hat{\tilde{y}}_i)$	1.27292	1.19108	1.1097	0.9599027	0.3184933
	$MSE(\tilde{y}_i)$	0.01560579	0.03708437	0.0807982	0.2286005	2.200847
Triangle	$S(\tilde{y}_i, \hat{\tilde{y}}_i)$	1.271866	1.190071	1.108947	0.9594122	0.3184589
	$MSE(\tilde{y}_i)$	0.01566583	0.03718067	0.0810207	0.2287428	2.200968

Table 3: represent $S(\tilde{y}_i, \hat{\tilde{y}}_i)$ & $S(\tilde{y}_i, \hat{\tilde{y}}_i)$ to model estimator when sample size 75

Kernel function	Performance	τ=0.2	τ=0.4	τ=0.6	τ=0.8	τ=1
Gaussian	$S(\tilde{y}_i, \hat{\tilde{y}}_i)$	1.200921	1.154948	1.083228	0.9378037	0.3668881
	$MSE(\tilde{y}_i)$	0.012563	0.0357774	0.07556379	0.2190008	1.77142
Epanechnikov	$S(\tilde{y}_i, \hat{\tilde{y}}_i)$	1.211991	1.166094	1.090782	0.9445642	0.3672586
	$MSE(\tilde{y}_i)$	0.013582	0.0367766	0.07657907	0.220187	1.772229
Triangle	$S(\tilde{y}_i, \hat{\tilde{y}}_i)$	1.211537	1.165544	1.090542	0.9441865	0.3672381
	$MSE(\tilde{y}_i)$	0.013652	0.0368465	0.07665346	0.2202382	1.772342

Table 4: represent $S(\tilde{y}_i, \hat{\tilde{y}}_i)$ & $S(\tilde{y}_i, \hat{\tilde{y}}_i)$ to model estimator when sample size 100

Kernel function	Performance	τ=0.2	τ=0.4	τ=0.6	τ=0.8	τ=1
Gaussian	$S(\tilde{y}_i, \hat{\tilde{y}}_i)$	1.188296	1.175577	1.080426	0.9390991	0.328061
	$MSE(\tilde{y}_i)$	0.01405338	0.03793128	0.0838061	0.2172008	2.096226
Epanechnikov	$S(\tilde{y}_i, \hat{\tilde{y}}_i)$	1.196553	1.18481	1.086752	0.9441733	0.328334
	$MSE(\tilde{y}_i)$	0.01496802	0.03893817	0.0847230	0.2181768	2.09711
Triangle	$S(\tilde{y}_i, \hat{\tilde{y}}_i)$	1.196102	1.184345	1.086391	0.9438766	0.328315
	$MSE(\tilde{y}_i)$	0.01498265	0.03892153	0.0847374	0.2181734	2.097154

The above tables show the goodness of fit $S(\tilde{y}_i, \hat{\tilde{y}}_i)$ and the mean squares error $MSE(\tilde{y}_i)$ for the estimated model, when using the squared distance $DK_1(\tilde{y}_i, \hat{\tilde{y}}_i)$ in equation 16 for the estimated model. When the model is estimated by equation 13. the parametric part of the model will be estimated based on equation 8, and the non-parametric part depending on the Nadaraya-Watson estimator when the Kernel functions are represented (Gaussian, Epanechnikov, Triangle) and with different sample sizes.

7. Conclusions

- We notice the mean square error $\mathrm{MSE}(\tilde{y}_i)$ and the criterion of the goodness of fit $S(\tilde{y}_i, \hat{\tilde{y}}_i)$ have an inverse relationship.
- We notice the best kernel function used in the Nadaraya Watson estimators is Gaussian function in all samples size.
- We notice the best quantile of the fuzzy conditional distribution of $\tilde{y}_i$ given xi, t_i is equal τ=0.2 in all samples size.
- In all sample sizes, we notice that there is a convergence in the mean squares error and the criterion of the goodness of fit, because the dependent variable represents a fuzzy triangular number.

Reference

1- Koenker, R., & Hallock, K. F. (2001). Quantile regression. Journal of economic perspectives, 15(4), 143-156.
2- Koenker, R., & Bassett Jr, G. (1978). Regression quantiles. Econometrica: journal of the Econometric Society, 33-50.
3- Koenker, R., & Ng, P. (2005). Inequality constrained quantile regression. Sankhyā: The Indian Journal of Statistics, 418-440.
4- Columbu, S., & Bottai, M. (2017). Cognitive deterioration in cancer patients using logistic quantile regression.

5- Bottai, M., Cai, B., & McKeown, R. E. (2010). Logistic quantile regression for bounded outcomes. Statistics in medicine, 29(2), 309-317.
6- Galarza, C. E., Zhang, P., & Lachos, V. H. (2021). Logistic quantile regression for bounded outcomes using a family of heavy-tailed distributions. Sankhya B, 83(2), 325-349.
7- Salmani, F., Taheri, S. M., Yoon, J. H., Abadi, A., Alavi Majd, H., & Abbaszadeh, A. (2017). Logistic regression for fuzzy covariates: modeling, inference, and applications. International Journal of Fuzzy Systems, 19(5), 1635-1644.
8- Gani, A. N., & Assarudeen, S. M. (2012). A new operation on triangular fuzzy number for solving fuzzy linear programming problem. Applied Mathematical Sciences, 6(11), 525-532.
9- Orsini, N., & Bottai, M. (2011). Logistic quantile regression in Stata. The Stata Journal, 11(3), 327-344.
10- Koenker, R. (2005). Quantile Regression. Book, Department of Economics, University of Illinois, Urbana-Champaign Champaign, IL 61820 USA.
11- Shemail, A. H., & Mohammed, M. J. (2022). Comparison between some estimation methods for an intuitionistic fuzzy semi-parametric logistic regression model with practical application about covid-19. International Journal of Nonlinear Analysis and Applications, 13(1), 3723-3732.
12- Hesamian, G., & Akbari, M. G. (2017). Semi-parametric partially logistic regression model with exact inputs and intuitionistic fuzzy outputs. Applied Soft Computing, 58, 517-526.
13- Hesamian, G., & Akbari, M. G. (2019). Fuzzy quantile linear regression model adopted with a semi-parametric technique based on fuzzy predictors and fuzzy responses. Expert Systems with Applications, 118, 585-597.

Article submitted 1 February 2023. Published as resubmitted by the authors 21 February 2023.