Sigma J Eng Nat Sci, Vol. 42, No. 1, pp. 1–10, February, 2024

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2024.00001

Research Article

An amalgamation of crisp and fuzzy quantile regression model

Saima MUSTAFA[1], Hina BASHARAT[1], Ali AKGÜL[2,3,*], Mohsin SHAHZAD[1], Abdelhamied Farrag SAYED[4]

[1]Department of Mathematics and Statistics, PMAS-Arid Agriculture University Rawalpindi, 46300, Pakistan
[2]Department of Mathematics, Faculty of Art and Science, Siirt University, Siirt, 56100, Türkiye
[3]Department of Electronics and Communication Engineering, Saveetha School of Engineering, SIMATS, Chennai, Tamilnadu, 602105, India
[4]Department of Mathematics, Al-Lith University College, Umm Al Qura University, P. O Box 112, Al-Lith 21961, Makkah Al Mukaramah, 21955, Kingdom of Saudi Arabia

ARTICLE INFO

Article history
Received: 16 October 2021
Revised: 09 December 2021
Accepted: 27 February 2022

Keywords:
Fuzzy Regression; Quantile Regression; Fuzzy Data; Linear Programming; Simplex Procedure

ABSTRACT

Fuzzy set theory is the most powerful tool to describe the process of uncertainty which exist in real world and fuzzy regression is an important research topic which can be used for prediction by establishing the functional relationship between fuzzy variables. Quantile regression is also a significant statistical method for estimating and drawing inferences about conditional quantile functions. This study introduced the idea of quantile regression with respect to fuzzy. The ordinary fuzzy regression is based on least square method but here we have introduced the idea of weighted least absolute deviation method in fuzzy regression. We have considered two different cases for the illustration of our proposed technique, firstly when the input and output are taken as fuzzy and secondly, the input and output are taken as fuzzy but the parameters are crisp. The algorithm for each case is based on linear programming problem (LPP). The LPP is constructed for individual case and solved it by the method of Simplex procedure. The proposed work is then compared with the conventional fuzzy regression by using AIC criterion. Empirical study shows that the proposed technique works best in every situation where the fuzzy regression fails and also provide the results in depth.

Cite this article as: Mustafa S, Basharat H, Akgül A, Shahzad M, Sayed AF. An amalgamation of crisp and fuzzy quantile regression model. Sigma J Eng Nat Sci 2024;42(1):1–10.

INTRODUCTION

Regression is one of the most important statistical measure for applied research and is applied to evaluate the relationship between response variable and explanatory variables. It also make some assumptions about the shape or distribution of residuals like heteroscedasticity, auto-correlation and multi-collinearity. If these assumptions are present in the model then the results of the estimates become misleading. An ordinary least squares (OLS) technique is used for estimating the standard regression parameters.

***Corresponding author.**
*E-mail address: aliakgul00727@gmail.com

This paper was recommended for publication in revised form by Regional Editor Hijaz Ahmad

Published by Yıldız Technical University Press, İstanbul, Turkey

The output of the typical linear regression methodology is the mean average between a set of explanatory variables, while response variable is called conditional mean whose function is stated as $E(Y/X)$. As one might be interested in conditional distribution of Y by observing the relationship at different points but the linear regression technique only offers a limited view of the relationship. Therefore another technique called quantile regression is used in this situation which has capability to describe the relationship in the conditional distribution of Y at different points. Quantile regression falls in a non parametric approach as non parametric is not based on any assumptions, similarly quantile regression also makes no assumption about the distribution of the residual or error term. Quantile regression permits us to study the effect of different predictors on various quantiles of the response variable, so that it provides a comprehensive picture of the relationship between response and explanatory variable. It is more vigorous to outliers than the casual least squares regression.

Measurement results and observational data are often not precise numbers but may be less or more precise, it is also called fuzzy. The idea of assembling sets as generalized sets was firstly published in [1] Menger, and later Zadeh [2] introduced the term 'fuzzy set' [3]. All types of data which cannot be shown as precise numbers or cannot be categorized precisely are said to be non-precise or fuzzy. The data in the form of linguistic or verbal descriptions like low temperature, high speed and high blood pressure etc are the examples of fuzzy. Fuzzy set theory is the most powerful tool to describe the process of uncertainty and imprecise information which exist in real world, and fuzzy regression is an important research topic which can be used to fulfill predicting by establishing the functional relationship between fuzzy variables. There are a many studies have been done to merge statistical techniques and fuzzy set theory. Li and Zeng [4] modified fuzzy regression based on least absolute deviation which is also known as median regression. They introduced a new distance measure between fuzzy number and merge with median regression or LAD and then parameters are considered to be a trapezoidal fuzzy number, Otadi [5] discussed the case fuzzy input and fuzzy coefficients in fuzzy polynomial regression, Choi [6] applied LAD (least Absolute Deviation) estimation technique to generate fuzzy regression analysis, Tian and Zhu [7] used EM-Algorithm to derive a new procedure for estimating linear quantile regression, Das and Ghosal [8] used Bayesian method to solve simultaneous quantile regression. In this approach, quantile function is obtained by inserting the prior distribution on monotone increasing function, Kim et al. [9] considered a single case fuzzy input-output with LAD (Least Absolute Deviation) regression, D'Urso [10] applied least square procedure for estimating the parameters of different fuzzy linear regression, where the model is based on fuzzy output with fuzzy input and Tanaka and Guo [11] recommended a technique to evaluate the efficiency of the model with fuzzy input and output. Similarly, Rasheed and Mustafa [12] done comparative study between simple linear regression, quantile regression and spline regression and Khalid and Mustafa [13] developed a quantile regression model using B-spline and penalties where she used method least absolute deviation (LAD) and include penalty term in the objective function.

It has been observed that most of the the researchers relates the fuzzy set theory with the linear regression, logistic regression or narrates to the statistical models and named them fuzzy linear regression. While the work done on quantile regression, mostly includes the the quantile regression with splines and most of them with Bayesian approach. Therefore, this study has attempted to merge the crisp quantile regression with fuzzy set theory which motivates the combination of both theories in order to obtain more reliable models which could be successfully applied in many practical studies of quantile regression models in real world application when data is not precise or exact for example the proposed model can be used in bio medical research study to understand the relationships among different fuzzy variables outside of the conditional mean of the response, in the field of epidemiology for the risk prediction and to examine the effects of growth trajectories and similarly they can be used in weather forecasting, wind speed forecasting and estimation of food expenditures.

The rest of the paper is organized as follows, some basic concepts of fuzzy set theory and its related components and some techniques which are used for present work are discussed in Section 2. In Section 3, methodology of proposed work is discussed. In Section 4, we have discussed the numerical computations and some concluding remarks are given in Section 5.

PRELIMINARIES

In this section, some basic definitions related to present work are discussed.

Quantile Regression and Weighted Absolute Function

Quantile regression is a semi parametric approach because it avoids the assumptions regarding the parametric distribution of the error term. Quantile regression contributes much additional information for the response variable. In OLS, the estimates of linear regression is the conditional mean function, while the estimates of quantile regression is based on conditional quantile function $Q(Y/X)$. The special case of quantile regression is known as Least Absolute Deviation Regression and simply abbreviated as LAD. LAD studies the connection between the explanatory and response variable using the conditional of median function $Q_{0.5}(Y/X)$ where the median equals to the 50^{th} percentile of the observed or experimental distribution. The quantille ($\theta \rightarrow [0,1]$), splits the data into proportions θ for under estimation and $(1-\theta)$ for over estimation. If $\theta = 0.5$ then it becomes Median regression.

The estimation of parameter $\beta's$ in these three models are different. If ϵ_i is the model prediction error, then OLS minimizes as,

$$min \sum (Y - X^T\beta)^2$$

$$= min \sum_{i=1}^{n} (\epsilon_i)^2$$

The conditional mean function is represented as,

$$E(Y/X) = X^T\beta$$

Median regression is estimated by least absolute deviation (LAD) approach, which minimizes

$$min \sum |Y - X^T\beta|$$

$$= min \sum_{i=1}^{n} |\epsilon_i|$$

In quantile regression model (QRM), the model and its conditional mean function is given as,

$$Y_i = X_i^T\beta_\tau + \epsilon_t \qquad where \quad i = 1, \ldots, T$$

where the τ^{th} quantile of ϵ_t is zero, and ($\theta \in [0,1]$) quantile. Quantile regression minimizes the sum of the unsymmetrical penalties $\theta|\epsilon_t|$ for under prediction and $(1-\theta)|\epsilon_t|$ for over prediction. The estimator of quantile regression for any θ belongs to [0,1] minimizes the objective function;

$$min_\beta \sum_{Y \geqslant X'\beta} \theta|Y - X^T\beta| + \sum_{Y < X'\beta} (1-\theta)|Y - X^T\beta|$$

$$= min \sum_{\epsilon \geqslant 0} \theta|\epsilon_t| + \sum_{\epsilon < 0} (1-\theta)|\epsilon_t|$$

An asymmetric absolute loss function for $\theta \in [0,1]$ is,

$$\rho_\theta(\epsilon) = \theta I_{\{\epsilon \geqslant 0\}}|\epsilon_t| + (1-\theta) I_{\{\epsilon < 0\}}|\epsilon_t|$$

$$\rho_\theta(\epsilon) = \sum_{t=1}^{T} [\theta I_{\{\epsilon \geqslant 0\}} + (1-\theta) I_{\{\epsilon < 0\}}]|\epsilon_t|$$

where $I_{\{\epsilon \geqslant 0\}}$ and $I_{\{\epsilon < 0\}}$ are the check or indicator functions, defined by Yu et al. [14]. The equation

$$\rho_\theta = \sum_{\epsilon \geqslant 0} \theta|\epsilon_t| + \sum_{\epsilon < 0} (1-\theta)|\epsilon_t|$$

is a non-differential function and the loss function is a piece-wise linear. Therefore,it is a linear programming problem and its computation is quite difficult so it requires linear programming method for example simplex method to minimize it, which is an appropriate method to acquire a solution in a finite number of iterations (see for example [15]). An alternative simpler notation of $\rho_\theta(\epsilon)$ is

$$\rho_\theta(\epsilon) = \theta\epsilon^+ + (1-\theta)\epsilon^-$$

where ϵ^+ represents $I_{\{\epsilon \geqslant 0\}}|\epsilon_t|$ and ϵ^- represents $I_{\{\epsilon < 0\}}|\epsilon_t|$. The objective function is represented as,

$$min_\beta \sum_{t=1}^{T} \theta\epsilon_t^+ + (1-\theta)\epsilon_t^-$$

subject to the constraint,

$$Y = X^T\beta + \epsilon^+ - \epsilon^-$$

Fuzzy Logic and Sets

Fuzzy logic is a generalization of classical logic. It is a mathematical technique to handle with the uncertainty, where it model those problems which are utilizing the imprecise data. As fuzzy logic deals with linguistic terms by assigning the suitable membership degree. So it can say that fuzzy frequently reports with uncertainty, imprecision and vagueness. In this section we recall some notations and preliminary concepts about fuzzy sets and numbers (see [3]). The mathematical definition of fuzzy sets can be defined as;

Definition 1 A fuzzy set is a collection of ordered pairs, Y be universe of discourse and y is specific element of Y, so the fuzzy set A can be written as,

$$A = \{(y, \mu_{\tilde{A}}(y)), y \in Y\}$$

where $\mu_{\tilde{A}}(y)$ showing a degree of membership of element y with $\mu_{\tilde{A}}(y) \rightarrow [0,1]$ and each ordered pair (y, $\mu_{\tilde{A}}(y)$) is known as singleton.

Linguistic Terms

Linguistic variables are essential for fuzzy sets. Basically, these are the words that are converted into numeric values by using fuzzy set theory. It is very helpful for dealing with vague and inadequate information as a first step to overcome ambiguity.

Membership Function

Membership is a degree of truthfulness of an input. The degree always lies between 0 to 1. It is denoted by $\mu_{\tilde{A}}(y)$, where $\tilde{A}$ is a fuzzy set and $y \in Y$. This functions permits to draw fuzzy sets graphically, where x-axis has a values of universal set and y-axis has values of membership on an interval [0,1].There are different types of membership functions and some of them that at are most commonly used are triangular, trapezoidal and gaussian membership function.

Definition 2 Suppose x be a subset of X, a membership function of a fuzzy set $\tilde{A}$ is symbolized as $\mu_{\tilde{A}}(\mathrm{x})$, where $0 \leq \mu_{\tilde{A}}(\mathrm{x}) \leq 1$. The value of $\mu_{\tilde{A}}(\mathrm{x})$ is called the degree of membership x ∈ X which is defined on a fuzzy set $\tilde{A}$ If $\mu_{\tilde{A}}(\mathrm{x})$ is zero, it will shows that it is not a member of fuzzy set $\tilde{A}$ and if $\mu_{\tilde{A}}(\mathrm{x}) = 1$ then is shows that it a fully member of given fuzzy set.

Fuzzy Number

Definition 3 A fuzzy number B is called a fuzzy subset of R, then the membership function $\mu_B(\mathrm{x})$ must obey the following standards,

1. $\alpha - cut: \mu_B(x) \rightarrow [0,1]$
2. The set of $\alpha - cut$ must have such elements of x whose membership magnitude is $\mu_B(x) \geq \alpha$.

3. There exist x, i.e. $\mu_B(x) = 1$.
4. $\mu_B(x)$ is a piece-wise continuous membership function.

Definition 4 If x belongs to the set X, and all the elements of X has non-zero membership magnitude in B, i.e,

$$B = \{x \in X \ |\mu_B(x) > 0\}$$

$\mu_B(x)$ is membership function defined on a closed interval [0,1].

Theorem 1 Let A = (a_l, a_m, a_r) and B = (b_l, b_m, b_r) are two triangular fuzzy numbers in a universe of discourse X, then

1. $A \oplus B = (a_l, a_m, a_r) + (b_l, b_m, b_r) = (a_l \oplus b_l, a_m \oplus b_m, a_r \oplus b_r)$
2. $-(a_l, a_m, a_r) = (a_r, -a_m, a_l)$

Fuzzy Regression

The term fuzzy regression analysis was introduced by Tanaka and Guo [11]. Possibilistic and least-squares are two main approaches of fuzzy regression models. The idea of best fit in both approaches includes the problem which is associated with the funtion optimization. Particularly, in possibilistic method, this functional takes the form of a measure of the spreads of the estimated output, either as a weighted linear sum involving the estimated coefficients in linear regression, or as quadratic form in the case of exponential possibilistic regression. Secondly in the standard least-squares approach, the functional to be minimized the distance between the observed and estimated outputs. This reduces to a class of quadratic optimization problems and constrained quadratic optimization D′Urso [10].

Linear Programming Problem

The problem of optimizing the linear function subject to its linear restriction or constraints, where the motive is to find the vector $\in R$. The constraints may be in the form of equalities or inequalities such as (=, ≤, ≥) respectively. Let z be the expression which is being minimized or maximized is called an objective function, and let the variable x_i, $(i = 1,2,\ldots,n)$ involved in the problem are called decision variables. Each line of constraints ended with the constant d_i. If the set of variables x_i fulfills all the restrictions then it is called feasible state and the set of these points in this state is known as feasible region.

The standard linear programming problem can be expressed as,

$$min \ z = \sum_{i=1}^{n} c_i x_i$$

Subject to,

$$\sum_{i=1}^{n} a_i x_i = d$$

where $x_i \geq 0 \ \ and \ \ i = 1,2,\ldots,n$. The above linear programming problem can be solved by using the simplex method which is one of the most useful and efficient algorithms ever invented, and it is still the standard method employed on computers to solve optimization problems.

The given LPP is easy to solve when the variables involve in the problem and their respective constraints are lesser in number but when the number of variables and constraints become larger, it is not possible to solve it manually, so one can overcome this complexity by using appropriate statistical software.

Fuzzy Quantile Regression

In this study we have introduced quantile regression in fuzzy logic which gives more detailed and precise results whenever the information is inexact or imprecise.

Fuzzy Linear Regression

We have considered the model which is represented as,

$$\tilde{Y}_i = \tilde{A}_o \oplus \tilde{A}_1\tilde{X}_1 \oplus \ldots \oplus \tilde{A}_k\tilde{X}_k \oplus E \qquad (1)$$

where $\tilde{Y}_i$, $\tilde{A}$ and X_k are triangular fuzzy numbers. The above model in matrix form is given as,

$$\begin{bmatrix}\tilde{Y}_1\\ \tilde{Y}_2\\ \vdots\\ \tilde{Y}_n\end{bmatrix} = \begin{bmatrix}1 & \tilde{X}_{11} & \tilde{X}_{12} & \cdots & \tilde{X}_{1k}\\ 1 & \tilde{X}_{21} & \tilde{X}_{22} & \cdots & \tilde{X}_{2k}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 1 & \tilde{X}_{n1} & \tilde{X}_{n2} & \cdots & \tilde{X}_{nk}\end{bmatrix}\begin{bmatrix}\tilde{A}_o\\ \tilde{A}_1\\ \vdots\\ \tilde{A}_n\end{bmatrix} + \begin{bmatrix}E_1\\ E_2\\ \vdots\\ E_n\end{bmatrix}$$

In simpler form the above model can be expressed as,

$$\tilde{Y} = \tilde{X}\tilde{A} \oplus \epsilon \qquad (2)$$

By applying the method of least square separately we can get the left, middle and right of the parameter estimates which are given below;

$$\tilde{A}_l = (X_l^T X_l)^{-1}(X_l^T Y_l) \qquad (3)$$

$$\tilde{A}_m = (X_m^T X_m)^{-1}(X_m^T Y_m) \qquad (4)$$

$$\tilde{A}_r = (X_r^T X_r)^{-1}(X_r^T Y_r) \qquad (5)$$

Proposed Models

We have considered two different cases for the illustration of our proposed technique, firstly when the input and output are taken as fuzzy and secondly, when the input and output are taken as fuzzy but the parameters are crisp.

Case-I: Fuzzy output, fuzzy input and fuzzy parameter (FFF)

For this case we considered the following model,

$$\tilde{Y}_{i\theta} = \tilde{A}_{0\theta} \oplus \tilde{A}_{1\theta}\tilde{X}_1 \oplus \ldots \oplus \tilde{A}_{k\theta}\tilde{X}_k \oplus E \qquad (6)$$

where θ=Specified Quantile and $\tilde{Y}_i\,\theta$, $\tilde{A}\theta$ and $\tilde{X}_k$ are triangular fuzzy numbers.These numbers are represented as,

$$\tilde{Y}_{i\theta} = (Y_{l\theta}, Y_{m\theta}, Y_{r\theta})$$

$$\tilde{A}_{i\theta} = (a_{il}, a_{im}, a_{ir})_\theta$$

$$\tilde{X}_{ij} = (X_{il}, X_{im}, X_{ir})$$

By using definition (4), equation (6) can be written as,

$$\tilde{Y}_{i\theta} = [(y - L^{-1}(\alpha) \times e)_\theta, (y_\theta), (y + L^{-1}(\alpha) \times e)_\theta] \quad (7)$$

where $L^{-1}(\alpha)$ = Inverse Shape function, e = spread of fuzzy number and for instance α= 0 is used (see for example [11]). So equation (6) can be written as,

$$\tilde{Y}_\theta = (a_{0l}, a_{0m}, a_{0r})_\theta \oplus (a_{1l}, a_{1m}, a_{1r})_\theta \tilde{X}_1 \oplus \ldots \oplus (a_{kl}, a_{km}, a_{kr})_\theta \tilde{X}_k \quad (8)$$

By using theorem (1) we have,

$$\tilde{Y}_{i\theta} = [(y - L^{-1}(\alpha) \times e)_\theta, (y_\theta), (y + L^{-1}(\alpha) \times e)_\theta] = (\sum_{i=1}^{n} a_{il}\tilde{X}_{i\theta}, \sum_{i=1}^{n} a_{im}\tilde{X}_{i\theta}, \sum_{i=1}^{n} a_{ir}\tilde{X}_{i\theta}) \quad (9)$$

To minimize the difference between observe and estimated fuzzy numbers we can decompose the above equation into three equations(see for example [16]) and are given as,

$$Y_{l\theta} = (y - L^{-1}(\alpha) \times e)_\theta = \sum_{i=1}^{k} a_{il}\tilde{X}_{i\theta} \quad (10)$$

$$Y_{m\theta} = (y_\theta) = \sum_{i=1}^{k} a_{im}\tilde{X}_{i\theta} \quad (11)$$

$$Y_{r\theta} = (y + L^{-1}(\alpha) \times e)_\theta = \sum_{i=1}^{k} a_{ir}\tilde{X}_{i\theta} \quad (12)$$

To find the parameter of the left triangular fuzzy number we have to minimize the equation (10), for this purpose linear programming problem is given as follows,

$$min \sum_{i=1}^{n} \rho_\theta |Y_{l\theta} - \sum_{i=1}^{k} a_{il}\tilde{X}_{i\theta}| = min \sum_{i=1}^{n} \rho_\theta |\delta_{il}| \quad (13)$$

where,

$$\delta_{il} = \mu_{1l} - \mu_{2l}$$

$$\mu_{1l} = I(\delta_{il} > 0)$$

$$\mu_{2l} = I(\delta_{il} < 0)$$

where I represents indicator function.

So the equation (10) becomes,

$$min \sum_{i=1}^{n} [\theta\mu_{1l} - (\theta - 1)\mu_{2l}] \quad (14)$$

And the objective function is,

$$min \sum_{i=1}^{n} [\theta\mu_{1l} + (1 - \theta)\mu_{2l}] \quad (15)$$

Subject to the constraints,

$$Y_{l\theta} - \sum_{i=1}^{k} a_{il}\tilde{X}_{i\theta} = \mu_{1l} - \mu_{2l} \quad (16)$$

For more simplicity, the problem of optimization for the above equations can be converted into matrix form, such as

$$minZ_l = C_l^T \Psi_l \quad (17)$$

Subject to the constraints,

$$D_l \Psi_l = Y_l \quad (18)$$

where,

$$C_l^T = [\theta 1_n^T,\ (1 - \theta)1_n^T\ ,0,\ 0]$$

$$\Psi_l = [\mu_{1l},\ \mu_{2l}\ , a_{0l},\ a_{il}]$$

$$D_l = [I,\ -I\ , X_{il},\ -X_{il}]$$

$$Y_l = [y_1, y_2, \ldots, y_n]_l$$

where

1_n^T = Vectors of ones,

0= Vectors of zeros,

I= Identity matrix,

X_{ij}=Triangular fuzzy set of independent variable,

a_{ij}= left, middle and right fuzzy number parameter,

Similarly for equation (11), linear programming problem to obtain the parameter of the triangular fuzzy number is given as,

$$minZ_m = C_m^T \Psi_m \quad (19)$$

Subject to the constraints,

$$D_m \Psi_m = Y_m \quad (20)$$

where,

$$C_m^T = [\theta 1_n^T,\ (1 - \theta)1_n^T\ ,0,\ 0]$$

$$\Psi_m = [\mu_{1m},\ \mu_{2m}\ , a_{0m},\ a_{im}]$$

$$D_m = [I,\ -I\ , X_{im},\ -X_{im}]$$

$$Y_m = [y_1, y_2, \ldots, y_n]_m$$

And for equation (12), linear programming problem is represented as,

$$minZ_r = C_r^T \Psi_r \quad (21)$$

Subject to the constraints,

$$D_r \Psi_r = Y_r \quad (22)$$

where,

$$C_r^T = [\theta 1_n^T,\ (1-\theta)1_n^T\ ,0,\ 0]$$

$$\Psi_r = [\mu_{1r},\ \mu_{2r}\ ,a_{0r},\ a_{ir}]$$

$$D_r = [I,\ -I\ ,X_{ir},\ -X_{ir}]$$

$$Y_r = [y_1, y_2, \ldots, y_n]_r$$

Case-II: fuzzy output, fuzzy input and crisp parameter (FFC)

The model of quantile regression for fuzzy output, fuzzy Input and crisp parameter (FFC)is represented as,

$$\tilde{Y}_{i\theta} = a_{0\theta} \oplus a_{1\theta}\tilde{X}_{i1} \oplus \ldots \oplus a_{k\theta}\tilde{X}_{ik} \oplus E \quad (23)$$

where θ=Specified Quantile, $\tilde{Y}_{i\theta}$, and $\tilde{X}_k$ are triangular fuzzy numbers and $a_{i\theta}$ is a crisp parameter, such as

$$\tilde{Y}_{i\theta} = (Y_{l\theta}, Y_{m\theta}, Y_{r\theta})$$

$$\tilde{X}_{ij} = (X_{ij_l}, X_{ij_m}, X_{ij_r})$$

From the definition (4), membership functions of $\tilde{Y}_{i\theta}$ and $\tilde{X}_{ij}$ are represented as,

$$\mu_{\tilde{X}_{ij}}(x) = \begin{cases} 0, & x < l \\ \frac{x-l}{m-l}, & l \le x \le m \\ \frac{u-x}{u-m}, & m \le x \le u \\ 0 & x > u\ . \end{cases} \quad (24)$$

$$\mu_{\tilde{Y}_i}(y) = \begin{cases} 0, & y < l \\ \frac{y-l}{m-l}, & l \le y \le m \\ \frac{u-y}{u-m}, & m \le y \le u \\ 0 & y > u\ . \end{cases} \quad (25)$$

To find the crisp parameters $a_{0\theta}, a_{1\theta}, \ldots, a_{k\theta}$ we converted the fuzzy values $\tilde{Y}_i = (Y_l, Y_m, Y_r)$, and $\tilde{X}_{ij} = (X_{ijl}, X_{ijm}, X_{ijr})$ into crisp values by using the method of defuzzification.

Let X_{iD} and Y_{iD} be the defuzzified values of the fuzzy observation $\tilde{X}_{ij}$ and $\tilde{Y}_i$ respectively. There are numerous methods have been suggested in literature for defuzzification, here we are using the centroid method to defuzzify the values of $\tilde{X}_{ij}$ and $\tilde{Y}_i$. The formula for defuzzification are represented as,

$$X_{iD} = \frac{\int_{-\infty}^{\infty} x\mu_{\tilde{X}_{ij}}(x)dx}{\int_{-\infty}^{\infty} \mu_{\tilde{X}_{ij}}(x)dx} \quad (26)$$

$$Y_{iD} = \frac{\int_{-\infty}^{\infty} y\mu_{\tilde{Y}_{ij}}(y)dy}{\int_{-\infty}^{\infty} \mu_{\tilde{Y}_{ij}}(y)dy} \quad (27)$$

So the FFC model in equation (23) becomes,

$$Y_{iD(\theta)} = a_{0\theta} + a_{1\theta}X_{i1_D} + \ldots + a_{k\theta}X_{ik_D} + \varepsilon \quad (28)$$

$$Y_{iD(\theta)} = a_{0\theta} + \sum a_{i\theta}X_{ij_D} \quad (29)$$

The problem of minimization in quantile regression algorithm is given as,

$$Q^{\theta}(Y_D/X_D) = \theta\left(\sum_{Y_{iD} \ge a_{0\theta} + \sum a_{i\theta}X_{ij_D}} |Y_{iD} - a_{0\theta} - \sum a_{i\theta}X_{ij_D}|\right) + (1-\theta)\left(\sum_{Y_{iD} < a_{0\theta} + \sum a_{i\theta}X_{ij_D}} |Y_{iD} - a_{0\theta} - \sum a_{i\theta}X_{ij_D}|\right) \quad (30)$$

To reach the linear programming problem (LPP), we have solved the equation (30) and simplified by considering the model up to two parameters$a_{0\theta}$ and $a_{1\theta}$, so the obtained form is given as

$$Q^{\theta}(Y_D/X_D) = \theta\sum_{\varepsilon \ge 0}|\varepsilon_i| + (1-\theta)\sum_{\varepsilon < 0}|\varepsilon_i| \quad (31)$$

where $\varepsilon_i = Y_{iD} - a_{0\theta} - \sum a_{1\theta}X_{ijD}$ As the above equation is non differential and it is a problem of linear programming, so we can write the above equation into a simpler form, such as

$$\rho_\theta(\varepsilon) = \theta\varepsilon^+ + (1-\theta)\varepsilon^- \quad (32)$$

where ε^+ represents the indicator function $I_{\{\varepsilon \ge 0\}}|\varepsilon_i|$ and ε^- represents the indicator function $I_{\{\varepsilon < 0\}}|\varepsilon_i|$. The objective function for LPP is represented as

$$min_{a_i}\sum_{i=1}^{n}[\theta\varepsilon^+ + (1-\theta)\varepsilon^-] \quad (33)$$

Subject to the constraint,

$$Y = X^T A + \varepsilon^+ - \varepsilon^- \quad (34)$$

where $\varepsilon^+ \ge 0$ and $\varepsilon^- \ge 0$.

ESTIMATION OF ERROR AND EVALUATION OF MODELS

The errors of both models given in equation (6) and (23) are computed by taking the difference between observed and expected values of fuzzy numbers $\tilde{Y}_i$ and $\tilde{Y}_i$ and are represented as,

$$E = \int |\tilde{Y}_i - \tilde{Y}_i|dy \quad (35)$$

The evaluation of the proposed models are done by taking the advantage of Akaike information criterion (AIC). AIC is a tool which is used to check and evaluate the quality of the model. Minimum the value of AIC shows the correctness of the model. The formula to compute the AIC is given as,

$$AIC = 2(p - \varphi) \quad (36)$$

where p= number of parameters to be estimated
$\varphi = ln(\hat{L})$
where $ln(\hat{L})$ is the maximum of likelihood function of the model.

Numerical Computation

To illustrate the theoretical results and obtain efficiency of our proposed regression models we have considered the data set of a simulation study which is originally used by Chung [17]. The proposed linear programming problem (LPP) is consist of complex algorithm and it cannot be solved manually, therefore R-software is used to compute the model. The R-package of the proposed study is created under the latest version of R Software. The name of the proposed R-package is "FuzzReg", where the title of package is Fuzzy Regression, the version of the package is 0.1.1. The package "FuzzReg" give the estimates of Fuzzy Linear regression (FLR) and fuzzy quantile regression (FQR). The command for Fuzzy linear regression is created on the least square method while the command for fuzzy quantile regression is made as per the proposed technique. The details of programming with codes are available in appendix section.

Case-I

In this case we considered the model of quantile regression when the input and output are taken as fuzzy and is given in equation (6). The coefficients of the this model are estimated using the linear programming problem by the equation (10), (11) and (12) at 0.25, 0.5, 0.75 quantile such as $\theta = 0.25,0.5,0.75$). The data set is taken from the research paper of Chung [17] to explain and compare theoretical results of proposed models. The data contains one response triangular fuzzy number $\tilde{Y}_i$ and two explanatory triangular fuzzy numbers $\tilde{X}_{i1}$ and $\tilde{X}_{i2}$.

ESTIMATION OF MODELS AT DIFFERENT QUANTILES

For one response and two explanatory fuzzy numbers the model at $\theta = 0.25$ (at 25^{th} quantile) using equation (6) is given as,

$$\tilde{Y}_i = \tilde{A}_{0(0.25)} \oplus \tilde{A}_{1(0.25)} \otimes \tilde{X}_{i1} \oplus \tilde{A}_{1(0.25)} \otimes \tilde{X}_{i2} \quad (37)$$

where

$$\tilde{A}_{0(0.25)} = (a_{0l}, a_{0m}, a_{0r})_{0.25}$$

$$\tilde{A}_{1(0.25)} = (a_{1l}, a_{1m}, a_{1r})_{0.25}$$

$$\tilde{A}_{2(0.25)} = (a_{2l}, a_{2m}, a_{2r})_{0.25}$$

By inserting the numerical values of coefficients, the estimated model at 25^{th} quantile is represented as,

$$\hat{\tilde{Y}}_i = (8.53613619,\ 2.925556029,\ -9.930407981) \oplus (0.46700327,\ 0.489126288,\ 0.460826026) \otimes \tilde{X}_{i1} \oplus (0.01319303,\ 0.009639338,\ 0.009684359) \otimes \tilde{X}_{i2} \quad (38)$$

By following the same steps we also estimated the model at 50^{th} quantile and is given as

$$\hat{\tilde{Y}}_i = (5.35834965,\ 6.624282536,\ -5.25445763) \oplus (0.54513813,\ 0.488534315,\ 0.43391915) \otimes \tilde{X}_{i1} \oplus (0.01156551,\ 0.008782578,\ 0.01201645) \otimes \tilde{X}_{i2} \quad (39)$$

Similarly estimated model at 75^{th} is represented as,

$$\hat{\tilde{Y}}_i = ((15.013455958,\ 4.83179004,\ -34.71201711) \oplus (0.693813123,\ 0.48638943,\ 0.51345892) \otimes \tilde{X}_{i1} \oplus (0.001189305,\ 0.01044304,\ 0.01831212)) \otimes \tilde{X}_{i2} \quad (40)$$

Estimated Model Of Fuzzy Linear Regression

Case-I

The model of fuzzy linear regression in the case of fuzzy input, fuzzy Output and fuzzy parameter is given as,

$$\tilde{Y}_i = \tilde{A}_0 \oplus \tilde{A}_1 \otimes \tilde{X}_{i1} \oplus \tilde{A}_2 \otimes \tilde{X}_{i2}$$

The estimated fuzzy regression model is given as,

$$\hat{\tilde{Y}}_i = (24.201139688,\ 0.82708141,\ -40.62302555) \oplus (0.500133049,\ 0.45968896, .47076543) \otimes \tilde{X}_{i1} \oplus (0.006719898,\ 0.01379444, .01958577) \otimes \tilde{X}_{i2} \quad (41)$$

Case-II

We considered the quantile regression model in this case when we have fuzzy input, fuzzy out put and crisp parameters.The estimation of the model given in equation (23) is done by using linear programming problem in equation (33) and (34). The data contains triangular fuzzy number i.e. $\tilde{Y}_{i\theta} = (Y_{l\theta}, Y_{m\theta}, Y_{r\theta})$ and $\tilde{X}_{ij} = (X_{ijl}, X_{ijm}, X_{ijr})$. In order to find the crisp parameters (a_0, a_1, a_2) we have converted the fuzzy observations into crisp values. Centeroid method given by equations (26) and (27) is used to defuzzify the fuzzy observations.

ESTIMATION OF MODELS AT DIFFERENT QUANTILES

The model for one response and two explanatory fuzzy numbers at 25^{th} quantile is given as,

$$Y_{iD} = a_{0(0.25)} \oplus a_{1(0.25)} \otimes X_{i1} \oplus a_{1(0.25)} \otimes X_{i2} \quad (42)$$

The intercept and the slopes of X_{ij} are computed using the equation (33). The estimated model at 25^{th} quantile is represented as,

$$\hat{\tilde{Y}}_i = (3.230283681) \oplus (0.513937886) \otimes \tilde{X}_{i1} \oplus (0.008082569) \otimes \tilde{X}_{i2} \quad (43)$$

Similarly the estimated mode at 50^{th} quantile is represented as,

$$\hat{\tilde{Y}}_i = (-2.00406574) \oplus (0.49955357) \otimes \tilde{X}_{i1} \oplus (0.01203797) \otimes \tilde{X}_{i2} \quad (44)$$

and estimated model at 75^{th} quantile is,

$$\hat{\tilde{Y}}_i = (-0.6244067) \oplus (0.5137353) \otimes \tilde{X}_{i1} \oplus (0.5137353) \otimes \tilde{X}_{i2} \quad (45)$$

Estimated Model of Fuzzy Linear Regression

Case-II

The model of fuzzy linear regression in case of fuzzy input, fuzzy output and crisp parameter, in terms of two input and one output fuzzy number is shown as,

$$\tilde{Y}_i = a_0 \oplus a_1\tilde{X}_{i1} \oplus a_2\tilde{X}_{i2}$$

and after estimating the parameters, estimated fuzzy linear regression model is given as,

$$\hat{\tilde{Y}}_i = (-1.57537822) \oplus (0.50189464)\tilde{X}_{i1} \oplus (0.01156812)\tilde{X}_{i2} \quad (46)$$

COMPARISON OF FUZZY QUANTILE AND FUZZY LINEAR REGRESSION

The evaluation and comparison between existing and proposed models are made by error estimation. Error estimation is done by deviating the observed values by estimated values as given in the equation (35). As fuzzy linear regression estimates a single line over the data while fuzzy quantile regression estimate multiple lines at specified quantile θ. For the sack of comparison, we have compared error of fuzzy quantile regression at 75^{th} quantile with fuzzy linear regression.

The error of fuzzy linear regression for the model given in Case-I is obtained as 3135.449 while the error of fuzzy quantile regression for the same model is obtained as 145.9895. In case-II, the errors are attained as 297.1842 and 51.84444 for fuzzy linear regression and fuzzy quantile regression respectively. It can be clearly seen that in both cases, proposed fuzzy quantile regression gives less error as compared to fuzzy linear regression and it can be concluded that fuzzy quantile regression performs better then the conventional fuzzy linear regression.

We also extended our comparison by using Akaike information criterion (AIC) given in section (Estimation of Error and Evaluation of Models). The more detail of errors and AIC for both cases are represented in table 1 and table 2. The value of AIC for proposed fuzzy quantile regression model is less than AIC value for fuzzy linear regression in both cases, which implies the better performance of our proposed models.

Table 1. Error and AIC of Fuzzy Linear Regression and Fuzzy Quantile Regression for Case-I

Models	Quantiles	Error	AIC
Fuzzy Quantile Regression	0.25	157.3862	114.4911
	0.5	145.9895	120.1598
	0.75	192.7083	127.6773
Fuzzy Linear Regression		3135.449	128.0495

Table 2. Error and AIC of Fuzzy Linear Regression and Fuzzy Quantile Regression for Case-II

Models	Quantiles	Error	AIC
Fuzzy Quantile Regression	0.25	60.2454	94.50432
	0.5	51.84444	94.00034
	0.75	61.1996	95.30643
Fuzzy Linear Regression		297.1842	95.36269

CONCLUSION

This research study mainly concerns with combination of quantile regression and fuzzy set theory to obtain fuzzy quantile regression. Fuzzy regression analysis is considered as extensively used technique in the atmosphere of fuzzy. It is an emerging zone for research now a days. Quantile regression is also an important methodology for response variables. The situation in which the assumption of linear regression becomes fail, the quantile regression assists to produce required results. We have proposed two models of quantile regression with various combinations of input-output variables by using triangular fuzzy number.The proposed models are then converted into a linear programming problem (LPP) and are solved by simplex method, taking triangular fuzzy data set. Empirical work shows evidence in the favor of proposed work. The error and AIC of each proposed model are less than the error and AIC of fuzzy regression which demonstrates that the proposed models are best fitted models and provides excellent assistance to elegantly solve structural equation models in fuzzy environment. Empirical researchers can take advantage of proposed fuzzy quantile regression models ability to examine the impact of predictor variables on response distribution in uncertainty modeling. Other research work can be introduced in the framework of present approach to fuzzy quantile regression where any combination of the inputs, outputs and parameters could be fuzzy or type-2 fuzzy.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Menger K. Probabilistic theories of relations. Proc Natl Acad Sci U S A 1951;37:178−185. [CrossRef]
[2] Zadeh LA. Fuzzy sets. Inf Control 1965;8:338−353. [CrossRef]
[3] Viertl R. Statistical methods for fuzzy data. Hoboken, New Jersey: John Wiley and Sons; 2011. p. 1−5. [CrossRef]
[4] Li J, Zeng W. A new fuzzy regression model based on least absolute deviation. Eng Appl Artif Intell 2016;52:54−64. [CrossRef]
[5] Otadi M. Fully fuzzy polynomial regression with fuzzy neural networks. Neurocomputing 2014;142:486−493. [CrossRef]
[6] Choi SH. Fuzzy regression using least absolute deviation estimators. Soft Comput 2008;12:257−263. [CrossRef]
[7] Tian Y, Zhu Q. Estimation of linear composite quantile regression using EM algorithm. Stat Probab Lett 2016;117:183−191. [CrossRef]
[8] Das P, Ghosal S. Bayesian quantile regression using random B-spline series prior. arXiv 2016:90−101.
[9] Kim J, Kim DH, Choi SH. Least absolute deviation estimator in fuzzy regression. J Appl Math Comput 2005;18:649−656.
[10] D'Urso P. Linear regression analysis for fuzzy/crisp input and fuzzy/crisp output data. Comput Stat Data Anal 2003;42:47−72. [CrossRef]
[11] Tanaka H, Guo P. Fuzzy DEA: a perceptual evaluation method. Fuzzy Sets Syst 2000;119:149−160. [CrossRef]
[12] Rasheed S, Mustafa S. Regression analysis by using spline techniques (Dissertation Thesis). Islamabad, Pakistan: PMAS Arid Agriculture University; 2013.
[13] Khalid A, Mustafa S. Flexible smoothing of quantile regression models with B spline and penalties (Dissertation Thesis). Islamabad, Pakistan PMAS Arid Agriculture University Rawalpindi; 2015.
[14] Yu L, Stander J. Quantile regression: applications and current research areas. J R Stat Soc 2003;52:331−350. [CrossRef]
[15] Baum CF. Quantile regression (presentation). Boston College; 2013. p. 1−20.
[16] Kim B, Bishu RR. Evaluation of fuzzy linear regression models by comparing membership functions. Fuzzy Sets Syst 1998;100:342−352. [CrossRef]
[17] Chung HW. Linear regression analysis for fuzzy input and output data using the extension principle. Comput Math Appl 2003;45:1849−1859. [CrossRef]

APPENDIX

FUZZY LINEAR REGRESSION

Description

It gives the fuzzy estimates of fuzzy regression by using the method of least squares.

Usage

flr(Y, X, type)

Arguments

"type" is model to be used (1)
x is an input fuzzy number
y is an output fuzzy number
1 "Fuzzy output, Fuzzy input and Fuzzy Parameters"

Examples

If given Triangular Fuzzy Number
x_left<- c (1.5,3.0,4.5,6.5,8.0,9.5,10.5,12.0)
x_centre< -c (2.0,3.5,5.5,7.0,8.5,10.5,11.0,12.0)
x_right< -c (2.5,4.0,6.5,7.5,9.0,11.5,11.5,13.0)
y_left< -c (3.5,5.0,6.5,6.0,8.0,7.0,10.0,9.0)
y_centre< -c (4.0,5.5,7.5,6.5,8.5,8.0,10.5,9.5)
y_right< -c (4.5,6.0,8.5,7.0,9.0,9.0,11.0,10.0)
X< -cbind(x_left, x_centre, x_right) Y< -cbind(y_left, y_centre, y_right) flr(Y, X, type = 1)

FUZZY QUANTILE REGRESSION

Description

It gives the estimates of fuzzy quantile regression using the method of Weighted Least Absolute Deviation (WLAD). It converts the input variables into Linear Programming Problem (LPP) and uses the Simplex Algorithm to solve the LPP.

Usage

fqr(X, y_left, y_centre, y_right, t, type)

Arguments

t is a specified quantile ranges from 0 to 1 i.e t=[0,1] "type" specifies the model (Here 1 in our case)
x is an input fuzzy number
y_left is left output fuzzy number y_centre is center of output fuzzy number y right is a right output fuzzy number
1 "Fuzzy output, Fuzzy input and Fuzzy Parameters"

Examples

If given Triangular Fuzzy Number library ("lpsolve")
x_left< -c (1.5,3.0,4.5,6.5,8.0,9.5,10.5,12.0)
x_centre< -c (2.0,3.5,5.5,7.0,8.5,10.5,11.0,12.5)
x_right< -c (2.5,4.0,6.5,7.5,9.0,11.5,11.5,13.0)
y_left<-c (3.5,5.0,6.5,6.0,8.0,7.0,10.0,9.0)
y_centre< -c (4.0,5.5,7.5,6.5,8.5,8.0,10.5,9.5)
y_right< -c (4.5,6.0,8.5,7.0,9.0,9.0,11.0,10.0)
X< -cbind(x_left, x_centre, x_right) t<-0.5
fqr(X, y_left, y_centre, y_right, t, type = 1)