ESKİŞEHİR TECHNICAL UNIVERSITY JOURNAL OF SCIENCE AND TECHNOLOGY
A- APPLIED SCIENCES AND ENGINEERING

2021, 22(4), pp. 353-365, DOI: 10.18038/estubtda.1033350

RESEARCH ARTICLE

COMPARATIVE ANALYSIS FOR FUZZY NONPARAMETRIC REGRESSION MODELS

Munevvere YILDIZ [1, *], Memmedaga MEMMEDLI [2]

[1] Department of Banking and Insurance, Bolvadin Faculty of Applied Sciences, Afyon Kocatepe University, Afyonkarahisar, Turkey
[2]mmammadov@anadolu.edu.tr, Eskisehir, Turkey

ABSTRACT

Statistical modeling is essential to revealing the relationships between variables. These statistical models can be classified as parametric and nonparametric methods in studies using crisp values. However, most of the collected data are inherently fuzzy. In this context, the fuzzy expression of methods using precise data is a matter of curiosity for researchers. The methods with fuzzy input and output variables have been developed for a long time. The study aims to describe nonparametric local polynomial regression models in fuzzy structure to examine the results for cases where the input variable is a crisp number, and the output variable is a symmetrical triangular and trapezoidal fuzzy number. According to the results, the bandwidth parameter was smaller in models where the degree of the polynomial was taken as one and larger in the case of three. In addition, the bandwidth parameter was found to be larger in models using the Epanechnikov kernel.

Keywords: Fuzzy local polynomial regression, Symmetric triangular fuzzy number, Trapezoidal fuzzy number, Generalized cross-validation, Mean square error

1. INTRODUCTION

Regression analysis is a method that is frequently used to reveal the functional form of the relationship between at least two variables and to produce predictions. However, some information that must be obtained before the analysis and some assumptions that must be provided divide regression analysis into two kinds as parametric and nonparametric. In cases where the assumptions are met and the shape of the functional relationship between the variables is known, parametric approaches are adopted; otherwise, nonparametric techniques are followed. It is an essential advantage of nonparametric methods. It provides a versatile method in investigating the general relationship between two variables and can produce estimates without reference to a particular parametric model. Both linear and nonlinear relationships can be modeled [1, 2]. There are many studies on parametric and nonparametric regression models in the literature. The vast majority of these studies were concerned with crisp data. It is common to use variables that do not contain fuzziness, where the information about the dependent and independent variables have crisp values. However, not all data in nature are precise. There are also data that contain uncertainty, in other words, fuzzy at its core. In order to derive and use such data, it is crucial to process and express classical statistical methods with fuzzy information.

There are relatively more studies on fuzzy parametric regression models in the literature [3-15]. The interest in parametric methods continues from the past to the present. However, there are few studies due to many parameters used in nonparametric regression methods and the relative difficulty of their expression and comparison.

Nonparametric regression models mostly focused on smoothing techniques. Kernel smoothing, k-nearest neighbor smoothing, local polynomial smoothing, and spline smoothing are the most used techniques. In the process of expressing nonparametric regression models in a fuzzy structure, different approaches have been put forward within the framework of smoothing techniques.

*Corresponding Author: munevvereyildiz@aku.edu.tr
Received: 10.10.2021 Published: 29.12.2021

In their study, Cheng and Lee [16] expressed the fuzzy nonparametric regression method by fuzzifying the k-nearest neighbor and kernel regression estimators. Cheng and Lee [17] used radial basis function networks in fuzzy regression expression. Wang et al. [18] used kernel smoothing and k-nearest neighbor techniques for the fuzzy expression of the local polynomial regression model. Yildiz [19], Memmedli and Yildiz [20] and Memmedli et al. [21] stated the formulation that can be used to calculate the cross-validation (CV), generalized cross-validation criteria (GCV), and the averaged squared error value, which is the model performance criterion, for the fuzzy structure in the fuzzy nonparametric local polynomial regression approach. Razzaghnia and Danesh [22] performed the selection of the best smoothing parameter in the local linear smoothing technique when the input variable is crisp, and the output variable is a trapezoidal fuzzy number. Danesh et al. [23] used the adaptive-neuro fuzzy system to estimate the fuzzy nonparametric regression function with precise input and fuzzy output. Hesamian and Akbari [24] extended the classical bandwidth selection for the population is normal and known or unknown fuzzy variance. In addition, different fuzzy distance approaches and cross-validation criteria were calculated and compared. Hesamian and Akbari [25] proposed a fuzzy spline method. They compared other fuzzy nonlinear regression methods in the literature with the proposed method. The results obtained with the proposed method produced better predictions. Hesamian and Akbari [26] expressed a fully fuzzy nonparametric regression model with a combination of both parametric and nonparametric methods. Naderkhani et al. [27] conducted research on selecting the best smoothing parameter for local linear smoothing, k-nearest neighbor smoothing, and kernel smoothing techniques from nonparametric methods based on adaptive neuro-fuzzy inference systems. Danesh et al. [28] presented a fuzzy inference system for fuzzy regression function prediction. In this study, they were made a comparative study for other fuzzy nonparametric regression technics. The results showed that the proposed method significantly decrease the boundary effect.

The aim of this study is to express the method in a fuzzy structure for the local polynomial approach, which is one of the nonparametric regression methods when an input variable is a crisp number, and an output variable is a fuzzy number. For the cases where the output variable is a symmetric triangular fuzzy number and a trapezoidal fuzzy number, the results are revealed by a simulation study. In the literature, there are many studies on the expression of parametric regression in the fuzzy structure. However, the existence of a small number of studies on fuzzy nonparametric methods and their openness to development have been the main motivation source. In cases where assumptions about parametric models are not provided, nonparametric methods are used. In this sense, the study will contribute to the literature by applying the nonparametric local polynomial regression method on different fuzzy number types.

This paper is organized as follows. Section two gives brief information about some definitions of various technical expressions. In section three, the methodology is expressed in a broader framework, and section four gives simulation study results. The last section gives the conclusions of the study.

2. DEFINITIONS

In this part of the study will be explained the concept of fuzzy numbers and the definitions of symmetric triangular fuzzy numbers, trapezoidal fuzzy numbers, and Diamond distance.

Definition 1 – Fuzzy Number:

X is a classical set (universe) and $\mu_A(x): X \to [0,1]$ represents the membership function, A is defined as the set of the following pairs as in Equation (1).

$$A = \{(x, \mu_A(x)), x \in X\} \tag{1}$$

If the maximum membership value of the fuzzy set A is equal to one, the normality condition is met (for $\exists x \in X,$ if $\mu_A(x) = 1$). For $\forall x_1 \in X,\ \forall x_2 \in X$ and $\lambda \in [0,1]$, $\mu_A(\lambda x_1 + (1-\lambda)x_2) \geq$

$\min(\mu_A(x_1), \mu_A(x_2))$ the fuzzy set is defined as convex if and only if each α-cut is a convex set. A normal and convex fuzzy set A of the set of real numbers $\Re$ is called a fuzzy number [29].

Definition 2- LR Fuzzy Number:

The LR representation of the membership function $\mu_M(x)$ for a fuzzy number M is defined by the Equation (2).

$$\mu_M(x) = \begin{cases} L[(m-x)/\alpha], x < m, \alpha > 0 \\ 1, \quad x = m \\ R[(x-m)/\beta], x > m, \beta > 0 \end{cases} \tag{2}$$

The $L(x)$ and $R(x)$ functions satify the following conditions for $\Re^+ \rightarrow [0,1]$.

i. $L(x) = L(-x)$, $R(x) = R(-x)$
ii. $L(0) = 1$, $R(0) = 1$
iii. L and R does not increase in the range $[0, \infty]$ (for $x \geq 0$, $L(x)$ and $R(x)$ are definitely decreases.)

The LR fuzzy number is written as $M = (m, \alpha, \beta)_{LR}$ [29].

Definition 3- Symmetric Triangular Fuzzy Number:

When $A = (a_1, a_2, a_3)$ is a fuzzy number, a_1 is the lower limit of fuzzy number, a_2 is the center of fuzzy number and a_3 is the upper limit of fuzzy number, and if it is defined as Equation (3), it is named triangular fuzzy number.

$$\mu_A(x) = \begin{cases} 0, & x < a_1 \\ \dfrac{x-a_1}{a_2-a_1}, & a_1 \leq x \leq a_2 \\ \dfrac{a_3-x}{a_3-a_2}, & a_2 \leq x \leq a_3 \\ 0, & x > a_3 \end{cases} \tag{3}$$

If $a_2 - a_1 = a_3 - a_2$, A is named as symmetric triangular fuzzy number [30].

Definition 4- Trapezoidal Fuzzy Number:

$A = (a_1, a_2, a_3, a_4)$ is a trapezoidal fuzzy number, and it contains many points in its structure whose membership degree is equal to 1. The trapezoidal fuzzy number is expressed with a membership function in Equation (4).

$$\mu_A(x) = \begin{cases} 0, & x < a_1 \\ \frac{x-a_1}{a_2-a_1}, & a_1 \leq x \leq a_2 \\ 1, & a_2 \leq x \leq a_3 \\ \frac{a_4-x}{a_4-a_3}, & a_3 \leq x \leq a_4 \\ 0, & x > a_4 \end{cases} \tag{4}$$

A trapezoidal fuzzy number turns into a triangular fuzzy number if $a_2 = a_3$ [30].

Definition 5- Diamond Distance:

The two LR fuzzy numbers like $A = (l_{A,} c_{A,} u_{A,})_{LR}$ and $B = (l_{B,} c_{B,} u_{B,})_{LR}$, the distance between these numbers is defined by Equation (5) using Diamond distance. It is a distance method that is frequently used to operate between two fuzzy numbers [13].

$$d^2(A,B) = (l_A - l_B)^2 + (c_A - c_B)^2 + (u_A - u_B)^2 \tag{5}$$

3. METHODOLOGY

The fuzzy expression of local polynomial regression, bandwidth selection criteria, and model performance criteria will be discussed in this part of the study.

3.1. Fuzzy Local Polynomial Regression

The purpose of fuzzy techniques is to include all the inherent uncertainty in the data into the model. Therefore, models based on fuzzy data contain more information than models in which the original uncertainty in the data is rejected or optionally omitted [31]. In this respect, it is essential to express all methods in fuzzy structure to be used in fuzzy data. When expressed with such fuzzy models, datasets with inherent fuzziness give better and more effective results. A fuzzy nonparametric local polynomial regression model will be expressed in this part of the study. Considering the case where the input variable X is a crisp number and the output variable Y is a fuzzy number, a fuzzy nonparametric regression model with a single explanatory variable is expressed as in Equation (6).

$$\tilde{y} = \mu(x)\ \{+\}\ \varepsilon \tag{6}$$

Here, $\mu(x) = (l(x), c(x), u(x))_{LR}$ is the fuzzy function defined in the crisp set $\mathcal{D}$ and whose values are LR fuzzy numbers, ε is the fuzzy error term. The $\{+\}$ symbol shows the operator that measures the difference between the observed and estimated fuzzy outputs whose definition depends on the fuzzy ranking methods used [16, 18]. Assuming that (x_i, y_i) $(i = 1,2,\dots,n)$ are observation data for the model (6) with crisp inputs and LR fuzzy outputs, for each output $\tilde{y}$, the c_y is the center, α_y and β_y are the left and right spreads components of LR fuzzy number as (c_y, α_y, β_y). Taking $l_y = c_y - \alpha_y$ and $u_y = c_y + \beta_y$, this number can be written as $(l_y, c_y, u_y)_{LR}$ with the help of left, center, and right limit points. The membership function for $A = (l_A, c_A, u_A)_{LR}$ that is a LR fuzzy number with l_A, c_A ve u_A the lower limit, center, and upper limit with real number as expressed as Equation (7).

$$\mu_A(t) = \begin{cases} L\left(\dfrac{c_A - t}{c_A - l_A}\right), & l_A \le t \le c_A \\ R\left(\dfrac{t - c_A}{u_A - c_A}\right), & c_A \le t \le u_A \\ 0, & otherwise \end{cases} \tag{7}$$

Here, the functions $L(.)$ and $R(.)$ are continuous, strictly decreasing in the interval $[0,1]$ and $L(0) = R(0) = 1$, $L(1) = R(1) = 0$. In this case, Equation (5) becomes the expression in Equation (8).

$$\tilde{y} = \mu(x)\ \{+\}\ \varepsilon = (l(x), c(x), u(x))_{LR}\{+\}\ \varepsilon \tag{8}$$

$l(x), c(x), u(x)$ are the left, center and right limit functions of a LR function, when they have a continuous derivative of order $(p+1)^{th}$ in domain $\mathcal{D}$, these functions can be written as Equation (9) as approximately p^{th} order Taylor polynomial with the neighborhood of a given point $x_0 \in \mathcal{D}$.

$$\begin{aligned} l(x) &\approx l(x_0) + l'(x_0)(x - x_0) + \cdots + \frac{l^{(p)}(x_0)}{p!}(x - x_0)^p \\ c(x) &\approx c(x_0) + c'(x_0)(x - x_0) + \cdots + \frac{c^{(p)}(x_0)}{p!}(x - x_0)^p \\ u(x) &\approx u(x_0) + u'(x_0)(x - x_0) + \cdots + \frac{u^{(p)}(x_0)}{p!}(x - x_0)^p \end{aligned} \tag{9}$$

In the method using the weighted least squares approach, the Diamond distance, which is a distance measurement for fuzzy sets, will be taken into account to calculate the difference between the actual and predicted values [13]. The application of the Diamond distance to the local polynomial fit estimation is expressed by Equation (10).

$$\begin{aligned}
&\sum_{i=1}^{n} d^2(\mu(x_i), y_i) K_h(|x_i - x_0|) = \\
&\sum_{i=1}^{n} d^2\Big(\big(\, l(x_i), c(x_i), u(x_i)\big)\,, (l_{y_i}, c_{y_i}, u_{y_i})\,\Big) K_h(|x_i - x_0|) = \\
&\sum_{i=1}^{n}\left(l_{y_i} - l(x_0) - l'(x_0)(x - x_0) - \cdots - \frac{l^{(p)}(x_0)}{p!}(x - x_0)^p\right)^2 K_h(|x_i - x_0|) + \\
&\sum_{i=1}^{n}\left(c_{y_i} - c(x_0) - c'(x_0)(x - x_0) - \cdots - \frac{c^{(p)}(x_0)}{p!}(x - x_0)^p\right)^2 K_h(|x_i - x_0|) + \\
&\sum_{i=1}^{n}\left(u_{y_i} - u(x_0) - u'(x_0)(x - x_0) - \cdots - \frac{u^{(p)}(x_0)}{p!}(x - x_0)^p\right)^2 K_h(|x_i - x_0|)
\end{aligned} \tag{10}$$

The primary purpose is to minimize Equation (10). Minimizing the value of each sum separately will also help to obtain the desired solution. For a given kernel function $K(\cdot)$ and smoothing parameter h, $K_h(|x_i - x_0|) = \frac{1}{h} K\left(\frac{|x_i - x_0|}{h}\right)$, $i = 1,2, \dots, n$, is a weight function whose role is to assign more weight to observations close to a given point x_0 and less to points farther away. Equation (11) expresses the Gauss kernel function, and Equation (12) represents the general structure of the Epanechnikov kernel function.

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{11}$$

$$K(x) = \begin{cases} 0.75(1 - x^2), & |x| < 1 \\ 0, & otherwise \end{cases} \tag{12}$$

The kernel functions are closely related to the smoothing parameter in models. In local polynomial models, the degree of the polynomial and the selected smoothing parameter are significant to establish the bias-variance balance. Expressing the relationship between the degree of polynomial and kernel functions, Fan and Gijbels [32] stated that the variance between the model with the degree of two and the model with the degree of three is the same bias decreases as the degree of the polynomial increases. At this point, the result is that the fitting performed with a single degree performs better asymptotically. For this reason, it was preferred to work with local linear and local cubic models in the study.

3.2. Selection of Smoothing Parameter

The smoothing parameter selection in local polynomial fit is another critical consideration. Selecting the smoothing parameter large will cause over smoothing in the functional structure, and choosing it too small will cause under smoothing. Selecting this parameter at the optimal value is important to create a balanced model [1]. For the selection of bandwidth parameters, cross-validation criteria, GCV criteria, Akaike Information Criteria (AIC) are frequently used. For selecting the bandwidth value in the study, the GCV expressed in fuzzy structure in the study of Yildiz [19] will be used. A hat matrix is used to define the GCV criterion, which allows the observation values to be mapped to the predictive values. The GCV criterion is expressed by Equation (13) with H being hat matrix.

$$GCV(h) = \frac{1}{n}\sum_{i=1}^{n} \frac{\big(y_i - \hat{\mu}(x_i, h)\big)^2}{\left(1 - \frac{1}{n} tr(H)\right)^2} \tag{13}$$

Equation (13) can be expressed by using the Diamond distance as follows.

$$GCV(h) = \frac{1}{n}\sum_{i=1}^{n}\frac{d^2(y_i - \hat{\mu}(x_i,h))}{\left(1-\frac{1}{n}tr(H)\right)^2} = \frac{1}{n}\sum_{i=1}^{n}\frac{\left(l_{y_i} - \hat{l}(x_i,h)\right)^{\mathbf{2}} + \left(c_{y_i} - \hat{c}(x_i,h)\right)^{\mathbf{2}} + \left(u_{y_i} - \hat{u}(x_i,h)\right)^{\mathbf{2}}}{\left(1-\frac{1}{n}tr(H)\right)^2} \quad (14)$$

Since the elements of the hat matrix H are crisp numbers, it can be written as a sum of three of the appropriate expressions for the variables l, c, u.

$$GCV(h) = GCV(l,h) + GCV(c,h) + GCV(u,h) \quad (15)$$

Here,

$$GCV(l,h) = \sum_{i=1}^{n}\frac{\left(l_{y_i} - \hat{l}(x_i,h)\right)^{\mathbf{2}}}{\left(1-\frac{1}{n}tr(H)\right)^2}, \quad GCV(c,h) = \sum_{i=1}^{n}\frac{\left(c_{y_i} - \hat{c}(x_i,h)\right)^{\mathbf{2}}}{\left(1-\frac{1}{n}tr(H)\right)^2},$$

$$GCV(u,h) = \sum_{i=1}^{n}\frac{\left(u_{y_i} - \hat{u}(x_i,h)\right)^{\mathbf{2}}}{\left(1-\frac{1}{n}tr(H)\right)^2} \quad (16)$$

For the selection of optimum $h = h_0$ value of the smoothing parameter, it is necessary to solve the $GCV(h_*)=\min_h GCV(h)$ minimization problem.

3.3. Model Performance Criteria

After mentioning the degree of the polynomial and the choice of bandwidth in the presentation of each model, the criterion to be used can be expressed in various ways for precise data when it is desired to make a comparison between the models. In order to make a comparison between local linear and local cubic models expressed in fuzzy structure, the average squared error based on Diamond distance is expressed in Yildiz's [19] study. However, it will be represented as mean square error (MSE) in this study due to its more formal use. The MSE criterion is expressed by Equation (17) when comparing different models.

$$MSE(h) = \frac{1}{n}\sum_{i=1}^{n} d^2\left(\mu(x_i), \hat{\mu}(x_i)\right)$$
$$= \frac{1}{n}\sum_{i=1}^{n}\left[(l(x_i) - \hat{l}(x_i))^2 + (c(x_i) - \hat{c}(x_i))^2 + (u(x_i) - \hat{u}(x_i))^2\right] \quad (17)$$

The MSE criterion is obtained by finding the mean of the squares of the difference between the actual values of the function and the predicted values. The minimum value of this performance criterion indicates the desired state for the models. It is possible to define the model with the smaller MSE criterion as a better model.

4. SIMULATION STUDY AND RESULTS

The simulation study aims to analyze and compare different fuzzy number types by using some parameters derived for fuzzy local polynomial models in Yildiz [19], Memmedli and Yildiz [20], and Memmedli et al. [21] studies. In the analysis, two cases where the input variable is the exact number and the output variable is the symmetric triangular fuzzy number and the trapezoidal fuzzy number were considered. The analyzes in the study were carried out in the R program with the help of the locpol package [33].

For this purpose, a symmetrical triangular fuzzy number for the output variable will be produced to be used in the simulation study with the help of the function used in Cheng and Lee's [16] studies.

- Symmetric Triangular Fuzzy Number

$g(x)$ is a function defined in the range [0,10].

$$g(x) = 10 + 5\sin(0.025\pi(1-x)^2)$$

Let the following definitions be made for $x_i = 0.1i$ $(i = 1,2,\dots,100)$ values in the range [0,10].

$$y_i = g(x_i) + rand[-0.5, 0.5]$$

$$\sigma_i = \frac{1}{4} g(x_i) + rand[-0.25, 0.25]$$

In the equations $rand[a_1, a_2]$ for each i. term in the interval $[a_1, a_2]$ represents a random number generated independently from the uniform distribution. Based on these equations, symmetric triangular fuzzy numbers are derived by expressing them as follows.

$$Y_i = (l_{y_i}, c_{y_i}, u_{y_i})_T = (y_i - \sigma_i, y_i, y_i + \sigma_i)_T \qquad i = 1,2,\dots,100$$

- Trapezoidal Fuzzy Number

With the help of a function used in Naderkhani et al.'s [27] study, a trapezoidal fuzzy number will be produced for the output variable, and a simulations study will be carried out. The x_i's are defined in the range [0,1] with a uniform distribution, and the $f(x)$ function is expressed as follows with $i = 1,2,\dots,100$.

$$f(x) = \frac{x^2}{5} + 2e^{x/10}$$

Using the $f(x)$ function in the following functional structures, the trapezoidal fuzzy number will be determined as Equation (18).

$$y_i = f(x_i) + rand[-0.5, 0.5]$$

$$e_i = \frac{1}{4f(x_i)} + rand[0,1]$$

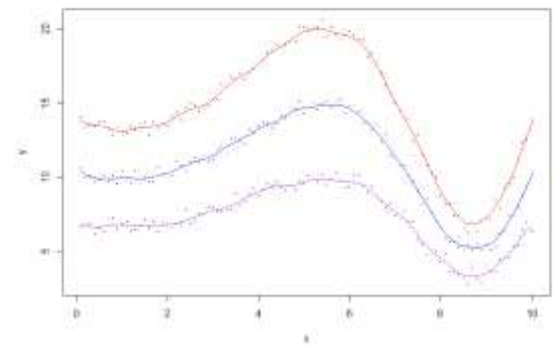
$$\tilde{Y}_i = (Y_i^1, Y_i^2, Y_i^3, Y_i^4) = (y_i - e_i, y_i + \frac{1}{3e_i}, y_i + \frac{2}{3e_i}, y_i + e_i) \tag{18}$$

After producing the symmetric triangular fuzzy number and trapezoidal fuzzy number type outputs required for the simulation study, the bandwidth values that make the smallest of the GCV criteria will be determined using Gauss and Epanechnikov kernels for fuzzy local linear and fuzzy local cubic nonparametric models. Table 1 shows the results obtained when the output variable is a symmetric triangular fuzzy number.

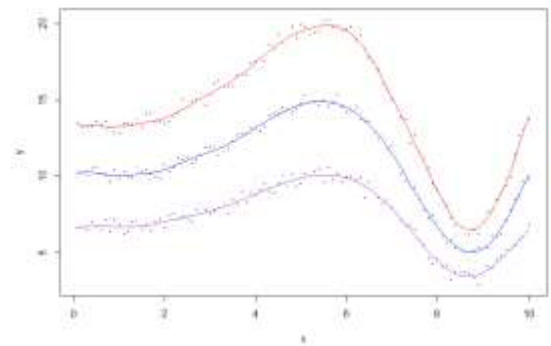
Table 1. Bandwidth and GCV values for symmetric triangular fuzzy number output

	Gauss Kernel			Epanechnikov Kernel		
	h	**GCV**	**MSE**	**h**	**GCV**	**MSE**
Local linear smoothing (LLS)	0.20	0.3590	0.226404	0.40	0.3600	0.229386
Local cubic smoothing (LCS)	0.50	0.3324	0.245779	1.70	0.3306	0.262345

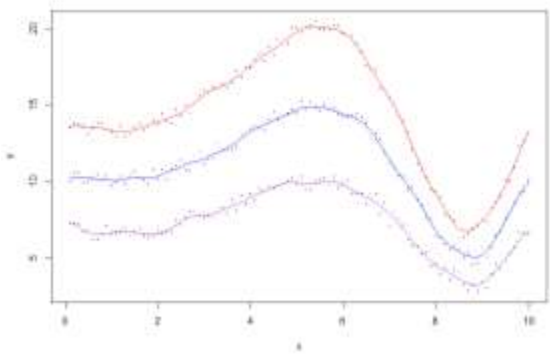
According to the information in Table 1, if a fuzzy nonparametric local linear model is used, a bandwidth value of 0.20 will be reached using the Gaussian kernel with the GCV formulation expressed in fuzzy structure. On the other hand, when the Epanechnikov kernel is used, the bandwidth parameter is determined as 0.40. In the fuzzy expression of the local cubic model, the bandwidth value of the model created with Gaussian kernel is 0.50, and when the Epanechnikov kernel is used, it has the smallest GCV value of 1.70 bandwidth. In addition, the performance criteria MSE values of the four models were found to be close to each other. In Figure 1, the fit curves formed within the framework of the optimal bandwidth parameter determined in the cases where the degree of the polynomial is one and three and in the models using Gaussian and Epanechnikov kernels are given.



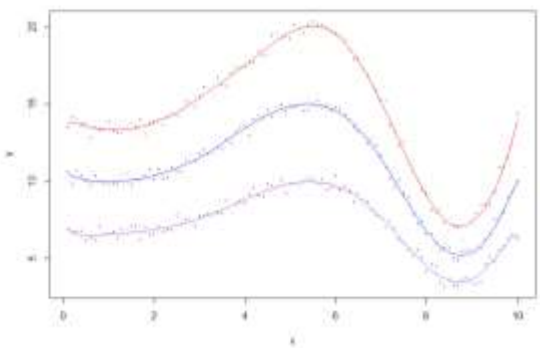
(a)



(b)



(c)



(d)

Figure 1. Model fit curves for the symmetric triangular fuzzy number output variable: (a) deg=1, h=0.20, Gauss Kernel; (b) deg=3, h=0.50, Gauss Kernel; (c) deg=1, h=0.40, Epanechnikov Kernel; (d) deg=3, h=1.70, Epanechnikov Kernel

Straight lines in the graphs are the estimation curves for the output variable's right scatter, center, and left scatter value in the form of symmetric triangular fuzzy numbers for the fitted model. The points on the graphs represent the actual observed values. In the examinations made when the output variable is a symmetric triangular fuzzy number, it is seen that the fit curves with a polynomial degree of one exhibit a more wavy structure. It can be stated that this situation can be explained by the increase in the bandwidth parameter when the polynomial degree is increased to three. On the other hand, in cases where the Epanechnikov kernel is used in the modeling, it is seen that the fitting occurs with a wider bandwidth compared to the model using the Gaussian kernel. Bandwidth parameter has an important place in nonparametric models in terms of obtaining correct model results.

In the study, the MSE criterion was used to evaluate the performance of the linear and cubic models created with the optimal bandwidth parameter. The bandwidth value, GCV value, and MSE values calculated according to the kernel function used in the models where the output variable is trapezoidal fuzzy number are presented in Table 2. As a numerical measure based on the difference between actual values and predicted values, MSE values are close to each other in all models.

Table 2. Bandwidth and GCV values for trapezoidal fuzzy number output

	Gauss Kernel			Epanechnikov Kernel		
	h	**GCV**	**MSE**	**h**	**GCV**	**MSE**
Local linear smoothing (LLS)	0.45	0.6589	0.531616	0.85	0.6589	0.529351
Local cubic smoothing (LCS)	0.95	0.6425	0.545452	2.30	0.6511	0.548633

For the fuzzy nonparametric local linear model in which the output variable is derived as a trapezoidal fuzzy number, the optimal bandwidth parameter was 0.45 with Gaussian kernel and 0.85 with Epanechnikov kernel according to the GCV criterion. On the other hand, in the cubic model, the bandwidth for the Gaussian kernel is 0.95, while it is 2.30 for the Epanechnikov kernel. The model fit curves for the selected bandwidth parameters for the trapezoidal fuzzy output variable and the use of Gaussian and Epanechnikov kernels in cases where the degree is one and three are shown in Figure 2.

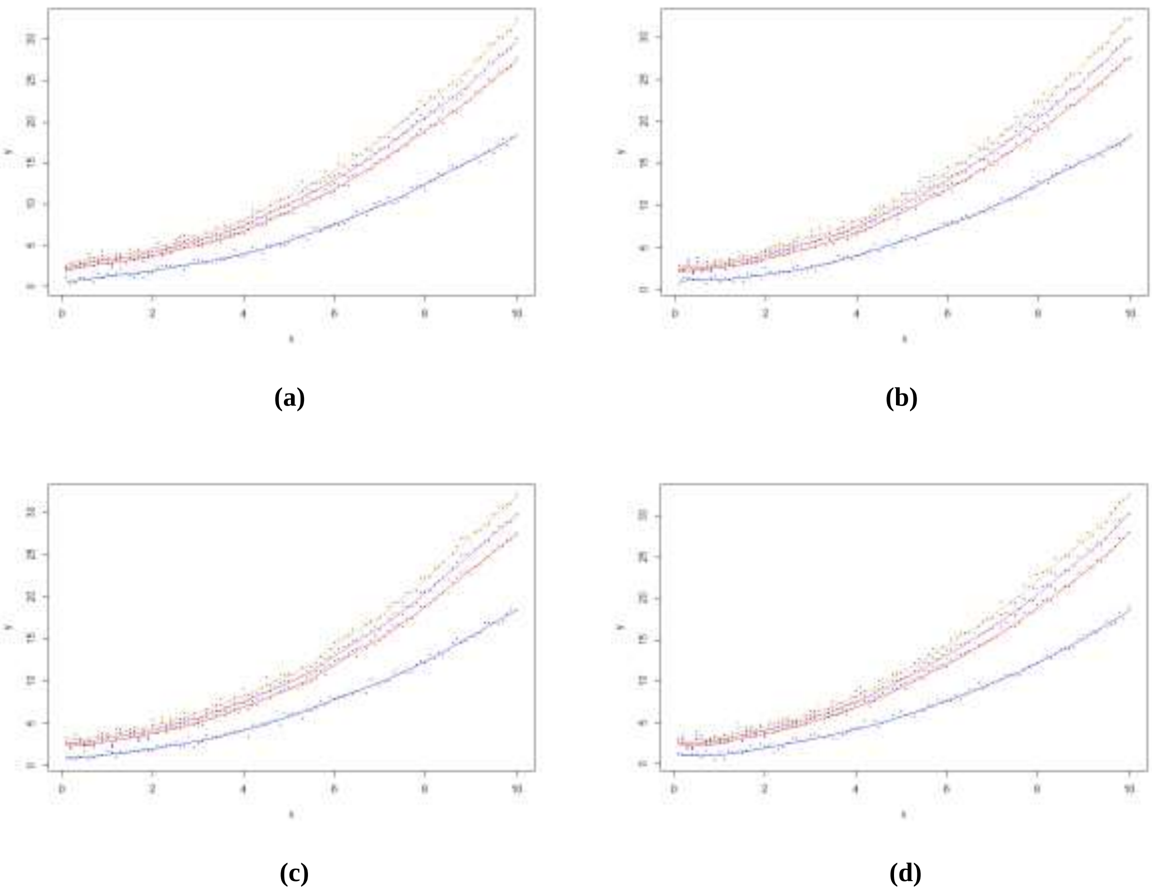


Figure 2. Model fit curves for the trapezoidal fuzzy number output variable: (a) deg=1, h=0.45, Gauss Kernel; (b) deg=3, h=0.95, Gauss Kernel; (c) deg=1, h=0.85, Epanechnikov Kernel; (d) deg=3, h=2.30, Epanechnikov Kernel

The structures shown with straight lines in the graphs are the estimation curves of the four parameters $\left(Y_i^1, Y_i^2, Y_i^3, Y_i^4\right)$ in the fuzzy number expression of the output variable in the form of trapezoidal fuzzy numbers for the fitted model. The points on the graphs represent the actual observation values. In the case where the output variable is a trapezoidal fuzzy number, similar results were observed to the results obtained with the symmetrical triangular fuzzy number. It is seen that the fit curves with a polynomial degree of one exhibit a more wavy structure. Also, in cases where the Epanechnikov kernel is used in the modeling, it is seen that the fitting occurs with a wider bandwidth compared to the model using the Gaussian kernel. Although the functional structure used to generate the trapezoidal fuzzy number is more linear, it is seen that the smoothness of the fit curves increases when the degree is taken as three. Again, in the case of using the Epanechnikov kernel, it can be stated that the wider bandwidth parameter is selected. When we look at the difference between the performances of the models created using the trapezoidal fuzzy number type output, it is observed that there is no high difference between the MSE values.

5. CONCLUSION

In today's conditions, although the collection of data in the exact number type seems to represent easy and clear solutions for people, technological developments and the increase in the use of artificial intelligence algorithms are the harbingers of future innovations in data collection. An important part of the studies in both technology and software is an indication that the way of data collection will be far from certain, based on probability and in a fuzzy structure. Although it is not very possible to collect fuzzy data in current conditions, no matter what discipline it is, it is essential to prepare scientifically for the location and status of the data in the future. In this framework, statistical methods have been expressed in fuzzy structure for data that can be collected in fuzzy structure for many years. However, at this point, since there is no direct fuzzy data collection process, researchers have focused on simulation studies.

The aim of this study is to contribute to the fuzzy nonparametric local polynomial regression approach and to observe and discuss the results with two different fuzzy number types. With the help of two separate functions are used to create the output variable of the symmetrical triangular fuzzy number and then the trapezoidal fuzzy number type. Common observations in studies using both number types are as follows;

- Bandwidth is chosen smaller in the fuzzy local linear model. This situation is the theoretical expectation.
- If the Epanechnikov kernel is preferred in creating the model, the selected bandwidth is wider than if the Gaussian kernel is selected.
- If the polynomial degree is three, the fit curves have a smoother structure than if the polynomial degree is one.
- When the output variable is a trapezoidal fuzzy number, the bandwidth value selected in all models is determined wider than the models whose output variable is a symmetrical triangular fuzzy number.
- It has been determined that the smallest GCV value obtained at the point of bandwidth selection and the MSE criterion values, which help us have information about the performance of the models, have higher values when the output variable is a trapezoidal fuzzy number. Due to the nature of the numbers, three parameters in the symmetric triangular fuzzy number type and four different parameters in the trapezoidal fuzzy number type affect these values.

There are many new fuzzy number types and fuzzy distance methods in the literature. Experimenting with different fuzzy numbers and expressing the formulations used in the calculation of criteria such as GCV and MSE, both according to the number type and with other fuzzy distance methods, will make important contributions in scientific terms.

ACKNOWLEDGMENTS

We would like to thank the editor and anonymous referees for their invaluable comments and support.

CONFLICT OF INTEREST

The authors stated that there are no conflicts of interest regarding the publication of this article.

REFERENCES

[1] Hardle W. Applied Nonparametric Regression. Cambridge University Press, New York, 1994.

[2] Fox J. Nonparametric Simple Regression: Smoothing Scatterplots. Sage Publications, California, 83 s., 2000.

[3] Tanaka H, Uejima S, Asai K. Linear Regression Analysis with Fuzzy Model. IEEE Transactions on Systems, Man., and Cybernetics 1982, 12, 903-907.

[4] Tanaka H. Fuzzy Data Analysis by Possibilistic Linear Models. Fuzzy Sets and Systems 1987, 24, 363-375.

[5] Tanaka H, Watada J. Possibilistic Linear Systems and Their Application to the Linear Regression Model. Fuzzy Sets and Systems 1988, 27, 275-289.

[6] Tanaka H, Hayashi I, Watada J. Possibilistic Linear Regression Analysis for Fuzzy Data. European Journal of Operational Research 1989, 40, 389-396.

[7] Bardossy A. Note on Fuzzy Regression. Fuzzy Sets and Systems 1990, 37, 65-75.

[8] Savic DA, Pedrycz W. Evaluation of Fuzzy Linear Regression Model. Fuzzy Sets and Systems 1991, 39, 51-63.

[9] Xizhao W, Minghu H. Fuzzy Linear Regression Analysis. Fuzzy Sets and Systems 1992, 51, 179-188.

[10] Chang PT, Lee ES. Fuzzy Linear Regression with Spreads Unrestricted in Sign. Computers and Mathematics with Applications 1994, Vol. 28, pp. 61-70.

[11] Ishibuchi H, Tanaka H. Fuzzy Regression Analysis Using Neural Networks. Fuzzy Sets and Systems 1992, 50, 257-265.

[12] Nasrabadi M, Nasrabadi E. A Mathematical-Programming Approach to Fuzzy Linear Regression Analysis. Applied Mathem. and Computation 2004, 155, 673-688.

[13] Diamond P. Fuzzy Least Squares. Information Sciences 1988, 46, 141-157.

[14] Hong H, Song JK, Do H. Fuzzy Least Squares Linear Regression Analysis Using Shape Preserving Operations. Information Sciences 2001, 138, 185-193.

[15] Kahraman C, Beskese A, Bozbura TF. Fuzzy Regression Approaches and Applications. StudFuzz 2006 Springer-Verlag Berlin Heidelberg, 201, 589–615.

[16] Cheng CB, Lee ES. Nonparametric Fuzzy Regression – k-NN and Kernel Smoothing Techniques. Computers and Mathematical with Applications 1999, 38, 239-251.

[17] Cheng CB, Lee ES. Fuzzy Regression with Radial Basis Function Networks. Fuzzy Sets and Systems 2001, 119, 291-301.

[18] Wang N, Zhang WX, Mei CL. Fuzzy Nonparametric Regression Based on Local Linear Smoothing Technique. An International Journal Information Sciences 2007, 177, 3882-3900.

[19] Yildiz M. Analysis of Nonparametric Fuzzy Regression Models. Ph.D. Dissertation, Anadolu University Graduate School of Sciences Statistical Department, 2013.

[20] Memmedli M, Yildiz M. Comparison study on smoothing parameter and sample size in nonparametric fuzzy local polynomial regression models. IV International Conference "Problems of Cybernetics and Informatics" (PCI'2012), September 12-14 2012.

[21] Memmedli M, Yildiz M, Ozdemir O. Parameter Selection of Fuzzy Nonparametric Local Polynomial Regression. 2nd International Symposium on Computing in Science&Engineering, June 1-4 2011, Kusadasi, Aydin, Turkey.

[22] Razzaghnia T, Danesh S. Nonparametric Regression with Trapezoidal Fuzzy Data. International Journal on Recent and Innovation Trends in Computing and Communication 2015, Volume:3 Issue: 6, 3826-3831.

[23] Danes S, Farnoosh R, Razzagnia T. Fuzzy nonparametric regression based on an adaptive neuro-fuzzy inference system. Neurocomputing 2016, 173, 1450-1460.

[24] Hesamian G, Akbari MG. Nonparametric Kernel Estimation Based on Fuzzy Random Variables. IEEE Transactions on Fuzzy Systems February 2017, Vol. 25, No.1.

[25] Hesamian G, Akbari MG. Fuzzy spline univariate regression with exact predictors and fuzzy responses. Journal of Computational and Applied Mathematics 2020, 375, 112803.

[26] Hesamian G, Akbari MG. A fuzzy nonlinear univariate regression model with exact predictors and fuzzy responses. Soft Computing 2021, 25:3247–3262. https://doi.org/10.1007/s00500-020-05375-9

[27] Naderkhani R, Behzad MH, Razzaghnia T, Farnoosh R. Fuzzy Regression Analysis Based on Fuzzy Neural Networks Using Trapezoidal Data. International Journal of Fuzzy Systems 2021, 23(5):1267–1280.

[28] Danesh M, Danesh S, Razzaghnia T, Maleki A. Prediction of Fuzzy Nonparametric Regression Function: A Comparative Study of a New Hybrid Method and Smoothing Methods. Global Analysis and Discrete Mathematics 2021, Volume 6, Issue 1, pp. 143-177.

[29] Lin CT, Lee GCS. Neural Fuzzy Systems A Neuro-Fuzzy Synergism to Intelligent Systems. Prentice-Hall, Inc., 1996.

[30] Lee KH. First Course on Fuzzy Theory and Applications. Springer-Verlag, Berlin Heidelberg New York, 2005.

[31] D'Urso P, Gastaldi T. An "orderwise" Polynomial Regression Procedure for Fuzzy Data. Fuzzy Sets and Systems 2002, 130, 1-19.

[32] Fan J, Gijbels I. Local Polynomial Modeling and Its Applications. Chapman & Hall/CRC, 1996.

[33] Cabrera JLO. locpol: Kernel Local Polynomial Regression, R package version 0.7-0. https://CRAN.R-project.org/package=locpol, 2018.