Journal of Theoretical and Applied Information Technology
15th November 2023. Vol.101. No 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DEVELOPING A MIXED NONPARAMETRIC REGRESSION MODLLING (SIMULATION STUDY)

AMJED MOHAMMED SADEK [1] , LEKAA MOHAMMED ALI [2]

[1] Department of Mathematics, Faculty of Basic Education, University of Telafer, Mosul, Iraq

[2] Department of Statistics, Faculty of Administration and Economics, University of Baghdad Baghdad, Iraq

E-mail: [1] amjed.m.sadek@uotelafer.edu.iq, [2] Lekaa.a@coadec.uobaghdad.edu.iq

ABSTRACT

The relationship between the response variable and the explanatory variable can show discernible patterns in certain cases, while in others, such patterns can remain unknown. Nonparametric regression techniques can potentially be employed to ascertain the unidentified pattern of relationships. The nonparametric regression technique provides a high degree of flexibility. In this work, we propose a new function for the kernel and use it with the mixed nonparametric regression between the kernel and truncated spline to compare it with the mixed Gaussian and the mixed biweight. Therefore, according to the study performed, we suppose that x_1, x_2, x_3 and x_4 have a specific pattern that was used with the spline method. On the other hand, x_5, x_6, x_7 and x_8 do not have a specific pattern that was used with the kernel method. Based on the GCV and MSE values, the best model was produced using the optimal bandwidth for each variable and one point of optimal knot for various sample sizes. The simulation study demonstrated that the mixed model with the proposed function has a suitable and superior performance when compared to the mixed Gaussian and mixed biweight models. The results of this investigation clearly demonstrated this model's superiority over its competitors. The highest obtained results have been confirmed by the coefficients of determination (R-square), which are 90.5%, 94.4%, and 97.2%, and the mixed nonparametric model with suggested function (AMS) provided the lowest mean square error (MSE) values of 4,074, 2,185, and 2,361 for various sample sizes. This indicates that the model will be able to produce accurate predictions and improve the performance of the data that we have been concentrating on.

Keywords: *Kernel Regression, Spline Truncated, Nonparametric Regression, Mixed Estimator, Mean Square Error.*

1. INTRODUCTION

Regression analysis is a method that statisticians use to determine the relationship between a response and one or more explanatory variables[1]. When the nature of the relationship patterns is uncertain or hard to figure out, such as their linearity, quadraticity, cubicity, or other specific form, the nonparametric regression model is used. Nonparametric models of regression demonstrate a high degree of flexibility because they possess the ability to mitigate the potential for misspecification. The shape of the regression function is determined solely by the data, independent of any subjective input from the researcher [2]. It is possible to explain the relationship by using a regression curve and applying either parametric or nonparametric regression estimation methods. The method of estimation known as parametric regression is used when it is possible to determine the form that the regression curve will take. Nonparametric regression is the method of estimation that is used when there are no shapes that can be ruled out for the regression function and preliminary information about the regression curve is limited. The regression model is very suitable for analyzing unknown data patterns due to its inherent flexibility [3, 4] .Regression models that are nonparametric compared to parametric ones require fewer conditions. Because of this, nonparametric regression models are a useful tool for researchers [5]. Recently, a variety of techniques, including kernels, local polynomials, truncated splines, Fourier, Wavelet, and others, have been developed for the nonparametric approach [6]. Specifically, truncated splines and kernel splines in fact, one of the most utilized estimation methods in nonparametric regression is the spline-truncated estimate. This estimator excels in its capacity to deal with data whose behavior varies at sub-specified intervals, and it provides exceptional statistical and visual interpretation as well [7,8]. In addition to the spline, the kernel is an excellent choice for modeling data that does not conform to any particular pattern [7]. There are many kernel functions, including triangular, quartic, epanechnikov, uniform, Gaussian, and others. In the context of kernel regression, the selection of bandwidth has more significance compared to the choice of kernel functions [9]. In the context of nonparametric regression, there are several fundamental premises that must be considered. The researcher only employs one version of the model estimator for each explanatory variable, and the pattern in each explanatory multivariable is expected to have the same pattern. The ensuing data patterns in their application to various situations frequently diverge from each explanatory variable. As a result, if just one estimator has been used to estimate the nonparametric regression curve, the resulting estimator does not fit the data pattern. As a result, the estimation of the regression model that results is less accurate and more likely to have high errors [4]. Many methods have been developed to tackle issues that can be resulted from dealing with multivariate models. Some researchers employed mixture model as a valid tool to cluster data to a number of subsets and use the appropriate model with each subset [10 , 11, 12]. In this study, we proposed a new kernel function and employed it in the mixed nonparametric estimator between the truncated spline and kernel regression. The objective of this study is to build a nonparametric mixture model using a new kernel function and then compare its performance to that of the Gaussian mixed model and the biweight mixed model. In addition, the simulation data had been submitted to a mixed estimator known as the (AMS), which compares the Gaussian and biweight models. The simulation data in this study is obtained from six different functions. It is assumed that certain exploratory variables are approximated using truncated splines, while others are estimated using kernel regression. Hence, the mixed model incorporating the suggested function exhibited superior performance in comparison to both the mixed Gaussian and mixed biweight models.

The rest of this paper is organized as follows: Sections 4.1 and 4.2 introduce the Methodology and methods used in this study. The form of the nonparametric regression model estimator using the proposed function (AMS), Gaussian, and biweight is presented in Section 4.3. Section 4.4 presents the selection of optimal knot points and bandwidth parameters to obtain the best model. Section 5 demonstrates the evaluation criteria used in this study. In Section 6, we illustrated the simulation study and computed the GCV, MSE, and R^2values. Finally, the last section illustrates the conclusions of the simulation study.

2. RESEARCH CONTRIBUTIONS

This paper makes significant contributions to the fields of statistics, mixed nonparametric regression modelling, applied mathematics, and computer science. by providing a comprehensive analysis of various statistical nonparametric regression techniques. The analysis includes a comparison of the performance and accuracy of these techniques on different sample sizes, highlighting their strengths and limitations. Additionally, the paper proposes a novel approach that combines nonparametric regression methods to further improve prediction accuracy and model flexibility, paving the way for future research and advancements in this area.

i- Introducing a new kernel function and incorporating it into the mixed model in order to improve the accuracy and flexibility of the mixed model for future predictions.

ii- The best optimal knot points and the optimal parameter value were selected by using the most accurate criteria for different sample sizes. These criteria were chosen based on their ability to minimize errors and maximize precision in the selection process. The selected knot points and parameter values will ensure the highest level of accuracy and efficiency in the given sample sizes.

iii- The mean square error (MSE) and the coefficient of determination (R-square) can be used to compare the proposed function (AMS) with the Gaussian and biweight functions and find the most accurate and reliable model for the data. These criteria provide quantitative measures of how well the function fits the data, allowing for a more objective evaluation. By considering both the mean square error (MSE) and coefficient of determination (R-square), we can gain a comprehensive understanding of the model's performance and make an informed decision on which function is best suited for the given data.

3. RELATED WORK

Researchers that work on the development of nonparametric regression models almost always employ the same kind of estimate approach for some or all of the variables that are predicted. This is because the researchers only use a single type of regression function for all of the variables of prediction, which is based on the premise that all of the predictors are regarded as having the same data pattern. although in practice, there are numerous situations in which each predictor variable exhibits a unique pattern. Because of this, the estimate of the regression model becomes less accurate, resulting in significant inaccurate information. As a result, a number of researchers came up with nonparametric regression estimators that combined two distinct types of estimators. These estimators are known as mixed estimators.

Alan and Rukun [13] developed a mixed model using local polynomial and spline-truncated methods, using the square weighted least squares (WLS) criterion to estimate parameters. They also used kernel functions as weights and the Generalized Cross Validation (GCV) criterion to determine optimal band width and knot points. Despite the complexity, the mixed model is not guaranteed to be superior to the simple model. To figure out nonparametric regression curves for longitudinal data, Maulidia et al. [14] used spline-truncated and kernel estimators. They used cross-sectional data from the kernel estimator and spline-trimmed weighted least squares (WLS) optimization. The optimal bandwidth and knots were chosen using the Generalized cross-validation (GCV) approach to select the best model. In addition, Adrianingsih et al. [15] presented a method that incorporated three methods of the modelling process, which are kernel, truncated spline, and Fourier series. The goal was to obtain a mixed non-parametric regression model, which was then applied to the actual data. The mixed estimation model was chosen based on the smallest value of the generalized cross-validation (GCV). Three knot points and three oscillations were used in this study, and the smallest value was chosen for each: The generalized cross validation (GCV), mean squares of error (MSE) at each node point, and oscillations. They also used the simulation program to estimate the parameters of the estimated model and prove the efficiency of the estimated model based on the criterion of the coefficient of determination (R-square). Furthermore, according to Ratnasari et al. [16] conducted a study on mixed estimators using truncated spline and Gaussian kernel methods. They investigated three methods for selecting the optimal node point and smoothing parameter: cross-validation (CV), Generalized cross-validation (GCV), and unbiased risk (UBR). A simulation study was conducted to compare the performance of these methods on a non-parametric regression model. The results showed that the Generalized cross-validation (GCV) method provided better performance and accuracy. In the same year, Researchers Nurcahayani et al. [17] provided a method for estimating a multivariate semi-parametric regression curve by combining truncated spline and Fourier series. They used Penalized Least Square (PLS) optimization and minimum Generalized cross validation (GCV) to estimate model parameters. The optimal model was determined based on the GCV value and the R-square.

4. METHODOLOGY

4.1 Truncated spline Non-parametric Regression

Spline Truncated Regression has become a widely utilized estimate since it provides an excellent visual interpretation, is capable of handling smooth functions, and is elastic. One of the merits of this regression is that it tends to get its data estimates anywhere the data patterns move and is flexible. This is one of the reasons why this model is advantageous [18]. Spline regression is a type of polynomial regression in which several polynomial segments are combined at knots to create a continuous model [3]. Overall, the spline truncated regression model, characterized by its degrees and knots, is a mathematical function that can be expressed in the form of an equation.

$$s(v_i) = \sum_{s=0}^{p} \theta_s v_i^s + \sum_{r=1}^{k} \beta_r (v_i - \tau_k)_+^p \qquad (1)$$

Where θ_s and β_r are unknown parameters, and p is the spline truncated function degree with knot points. Therefore, the function of truncated can be described as following

$$(v_i - \tau_k)_+^p = \begin{cases} (v_i - \tau_k)^p, & v_i \geq \tau_k \\ 0, & v_i \leq \tau_k \end{cases}$$

The truncated spline regression model can generally be expressed as the following formula:

$$y_i = \sum_{s=0}^{p} \theta_s v_i^s + \sum_{r=1}^{k} \beta_r (v_i - \tau_k)_+^p + \varepsilon_i \qquad (2)$$

$$= \theta_0 + \theta_1 v_1 + \cdots + \theta_p v_i^p + \beta_1 (v_1 - \tau_1)_+^p + \cdots + \beta_r (v_1 - \tau_k)_+^p + \varepsilon_1$$

Journal of Theoretical and Applied Information Technology
15th November 2023. Vol.101. No 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The above equation may be expressed as a matrix, expressed by

$$S(v_i) = H(k)\varphi \tag{3}$$

Where

$$H(k)=\begin{pmatrix} 1 & v_1 & \cdots & v_1^p & (v_1-\tau_1)_+^p & \cdots & (v_1-\tau_k)_+^p \\ 1 & v_2 & \cdots & v_2^p & (v_2-\tau_2)_+^p & \cdots & (v_2-\tau_k)_+^p \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & v_n & \cdots & v_n^p & (v_n-\tau_n)_+^p & \cdots & (v_n-\tau_k)_+^p \end{pmatrix} \tag{4}$$

$S(v_i) = [s(v_1) \;\; s(v_2) \cdots s(v_n)]^T$, $\varphi = [\theta_1 \; \theta_2 \; \cdots \theta_P \; \beta_1 \; \beta_2 \; \cdots \beta_r]^T$ The size of the vector $S(v_i)$ is $n \times 1$, and the size of the matrix $H(k)$ is $(p+k+1) \times n$ while the size of the vector φ is $(p+k+1) \times 1$.

4.2 Kernel Non-parametric Regression

The main goal of non-parametric regression is to figure out the regression function. Several methods for estimating this function have been proposed; the most commonly used is the kernel estimator. The kernel estimator has a flexible property. Also, the math form is easy to understand, and convergence levels can be found quickly [9,15] One of the most significant nonparametric kernel estimators of a regression function is the Nadaraya [19] and Watson [20] kernel regression estimator. The regression curve can be approximated

$$m_h(u_i) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{K_h(u-u_i)}{n^{-1}\sum_{i=1}^{n}K_h(u-u_i)}\right]y_i = \frac{1}{n}\sum_{i=1}^{n}W_{h_i}(u_i)y_i \tag{5}$$

The equation shown above may also be written as follows:

$$m_h(u) = \frac{1}{n}W_{h_1}(u)y_1 + \frac{1}{n}W_{h_2}(u)y_2 + \ldots + \frac{1}{n}W_{h_n}(u)y_n \tag{6}$$

Furthermore, the equation (6) can be written in matrix form as follows:

$$m_h(u) = D(h)y \tag{7}$$

Where

$$D(h)=\begin{pmatrix} n^{-1}W_{h_1}(u_1) & n^{-1}W_{h_2}(u_1) & \cdots & n^{-1}W_{h_n}(u_1) \\ n^{-1}W_{h_1}(u_2) & n^{-1}W_{h_2}(u_2) & \cdots & n^{-1}W_{h_2}(u_2) \\ \vdots & \vdots & \ddots & \vdots \\ n^{-1}W_{h_1}(u_n) & n^{-1}W_{h_2}(u_2) & \cdots & n^{-1}W_{h_n}(u_n) \end{pmatrix}$$

$m_h(u) = [m_h(u_1) \quad m_h(u_2) \cdots m_h(u)]^T$, and $y = [y_1 \quad y_2 \cdots y_n]^T$, The vectors $m_h(u)$ and y are sized $n \times 1$,and the matrix $D(h)$ is sized $n \times n$.

4.3- The Form of the Non-parametric Regression Mixed Model Estimator of Truncated Spline and Kernel Regression

Under this mixed estimator, we propose a new function to use with this mixed estimator, and then we compare the new mixed method with the Gaussian kernel and the biweight kernel. Furthermore, there are two components for explanatory variables: the first one will be estimated using the spline truncated method, while the kernel method will be used to estimate the second. So, Both the spline-truncated method and the kernel method will be used to estimate the regression curve or regression function. Consider the data (v_i, u_i, y_i) and the relationship between explanatory variables (v_i, u_i) and a response variable y_i are presumed to follow a nonparametric regression model. Generally, the non-parametric regression model for the mixture of the spline and kernel is represented by the equation below.

$$y_i = \mu(v_i, u_i) + \varepsilon_i \;\; , i = 1,2,\ldots,n \tag{8}$$

Moreover, Assume the regression curve $\mu(v_i, u_i)$ is to be additive, which means that it can be written as the following form:

$$\mu(u_i, v_i) = s_k(v_i) + m_h(u_i) \tag{9}$$

The function $s_k(v_i)$ and $m_h(u_i)$ is presumed to be unknown and smooth, which means that it is continuous and differentiable. A random error ε_i has a normal distribution with a zero mean and constant variance. Furthermore, the function $s_k(v_i)$ is approached using the method of the truncated spline with knots $(\tau_1 \; \tau_2 \cdots \tau_k)^T$, while the function $m_h(u_i)$ is approached using the Nadaraya-Watson kernel with a vector of the smoothing parameter (bandwidth). Consequently, the shape of the equation that appears in (9), can be represented in the following form:

$$\hat{\mu}_{k,h}(u_i, v_i) = \hat{s}_k(v_i) + \hat{m}_h(u_i) \tag{10}$$

Where $k = (\tau_1, \tau_2, \ldots, \tau_k)^T$is the knot points vector and $h = (h_1, h_2, \ldots, h_8)^T$is the bandwidth parameters. consider the basis of the truncated spline space:

$$\begin{pmatrix} 1, v, v^2, \ldots, v^p, (v-\tau_1)I(v \geq \tau_1), (v-\tau_2)I(v \geq \tau_2), \ldots, \\ (v-\tau_k)I(v \geq \tau_k) \end{pmatrix} \tag{11}$$

With I being the indicator function, the function $s_k(v_i)$ can be written as the following form:

$$s_k(v_i) = \theta_0 + \theta_1 v_i + \ldots + \theta_p v_i^p + \beta_1 (v_i - \tau_1)^p I(v_i \geq \tau_1) + \ldots + \beta_k (v_i - \tau_k)^p I(v_i \geq \tau_k) \quad (12)$$

Where $(\theta_o, \theta_1, \ldots, \theta_p, \beta_1, \beta_2, \ldots, \beta_k)$ being unknown parameters. Furthermore, the estimate of the function $m_h(u_i)$ can be provided by utilizing the Nadaraya-Watson kernel as follows:

$$\hat{m}_h(u_i) = n^{-1} \sum_{i=1}^{n} \left[\frac{K_h(u-u_i)}{n^{-1} \sum_{i=1}^{n} K_h(u-u_i)} \right] y_i = n^{-1} \sum_{i=1}^{n} W(u) y_i \quad (13)$$

Where $W(u)$ and $K_h(u-u_i)$ may be expressed as the following formula:

$$w(u) = \frac{k_h(u-u_i)}{n^{-1} \sum_{i=1}^{n} k_h(u-u_i)} \quad \text{and} \quad k_h(m-m_i) = \frac{1}{h} k\left(\frac{u-u_i}{h}\right)$$

With $k(u)$ is the kernel function. There are many types of kernel functions [21]. In this study, we propose a new function for the kernel and use the Gaussian kernel and the biweight kernel with the following formula:(see Table 1)

Table 1: *Kernel Function*

No.	Kernel functions	function formula
1	Proposed function (AMS)	$k(u) = \frac{122377}{156250} \exp\left[-\left(u^6 + u^4 + u^2\right)\right], u \in [-\infty, \infty]$
2	Gaussian	$k(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-u^2\right), u \in [-\infty, \infty]$
3	Biweight	$k(u) = \frac{15}{16}\left(1-u^2\right)^2, u \in (-1, 1)$

Therefore, the kernel estimator that is often used for estimating random instances is the Nadaraya-Watson kernel estimator. Hence, the multivariate Nadaraya-Watson kernel estimator can be defined in the following form: [19,20]

$$\hat{m}_h(u_i) = \frac{\frac{1}{n} \sum_{i=1}^{n} \left[\prod_j k_h\left(\frac{u_j - u_{ji}}{h_j}\right) \right]}{\frac{1}{n} \sum_{i=1}^{n} \left[\prod_j k_h\left(\frac{u_j - u_{ji}}{h_j}\right) \right]} y_i, j = 1, 2, \ldots, q \quad (14)$$

To get an estimate of the combined nonparametric regression model of truncated spline and kernel regression in the form of (9), we will be using the method of least squares. According to Ratnasari et al.[22] presented some theorems and lemmas. If the function $s_j(v_{ji}), j = 1, 2, \ldots, p$ is approached using the linear truncated spline function with knots $k_j(\tau_{j1}, \tau_{j2}, \ldots, \tau_{jk})^T$, Therefore, it can be expressed as:

$$\sum_{j=1}^{p} S_j(v_{ji}) = H(k)\varphi \quad (15)$$

With $\varphi = (\theta_{j0}, \theta_{j1}, \beta_{j1}, \beta_{j2}, \ldots, \beta_{jk})$,
$S_j(v_{ji}) = \left(s_1(v_{1i}), s_2(v_{2i}), \ldots, s_p(v_{pi})\right)^T$ and
$H(k) = \left(H(k_1), H(k_2), \ldots, H(k_p)\right)$

$$H(k_j) = \begin{bmatrix} v_{j1} & (v_{j1} - \tau_{j1})_+ & \cdots & (v_{ji} - \tau_{jk})_+ \\ \vdots & \vdots & \ddots & \vdots \\ v_{jn} & (v_{jn} - \tau_{j1})_+ & \cdots & (v_{jn} - \tau_{jk})_+ \end{bmatrix}$$

Moreover, if the function $m_r(u_{ri}), r = 1, 2, \ldots, q$ is approached using the kernel Nadaraya-Watson. Using formula (14), the following matrix shape gives the kernel estimator for a multivariate regression curve.

$$\sum_{r=1}^{q} m_r(u_{ri}) = F(h)y \quad (16)$$

Where $[y_1\ y_2\ \ y_n]^T$ is the vector of the response variable. Additionally, the formula (16) can be expressed in the next matrix representation.

$$\sum_{j=1}^{p} s_j(v_j, k; h) = M(k, h)y \quad (17)$$

With
$M(k,h) = H(k)[H(k)^T H(k)]^{-1} H(k)^T (I - F(h))$

Further, the estimate of the mixed nonparametric regression between the truncated spline, and the kernel estimators is provided by:

$$\hat{\mu}(v_1, v_2, \ldots, v_p, u_1, u_2, \ldots, u_q; k, h) = M(k,h) + F(h) = L(k,h)y$$

4.4-Selection knot points and optimal bandwidth parameter of the mix spline truncated and kernel regression

The Knots point and optimum bandwidth parameters are required for the estimation of nonparametric curve regression using a mixed truncated spline and kernel. Generalized Cross-Validation (GCV) is a technique that may be utilized in order to select the knot point and determine the appropriate parameter bandwidth [4]. The generalized cross-validation (GCV) method shows asymptotically optimal properties, making it

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

suitable for application in cases involving large sample sizes. One characteristic that distinguishes the asymptotically optimal trait from other approaches, such as cross-validation (CV), is its lack of shared attributes [23]. Thus, the next formula provides the expression for the criteria function used for generalized cross-validation (GCV), denoted by the formula (18).

$$GCV(k,h) = \frac{n^{-1}\left\| y_i - \hat{\mu}(v_1,v_2,\ldots,v_p,u_1,u_2,\ldots,u_q\,;k,h)\right\|^2}{(n^{-1}trace[I-M(k,h)-F(h)])^2} \quad (18)$$

Where I is an identity matrix, and the vector of the optimum knot point $k_{opt} = (\tau_{j1},\tau_{j2},\ldots,\tau_{j9})^T$.In this study, 1-knot point $(j=1)$,2-knots point $(j=2)$,3-knots point $(j=3)$ and four knots point $(j=4)$ were used. Furthermore, the k_{opt} and h_{opt} obtained by optimization:

$$GCV(k_{opt},h_{opt}) = \underset{k,h}{Min}\{G(k,h)\} \quad (19)$$

5. EVALUATION CRITERIA

Finding the optimal model to describe the relationship between the response and explanatory variables based on predetermined criteria is one of the purposes expected from performing regression analysis. The criteria that are employed to evaluate the appropriateness of the model of regression are Mean Square Error (MSE) and coefficient of determination. The lower the Mean Squared error (MSE) value obtained, the better the model obtained. In contrast, the model obtained will be best if the value of R^2 is substantial [24]. mean squared error Creritia is one of the most commonly used due to its clarity, interpretability, and adaptability for optimization, it continues to be a common option for assessing regression models [25]. In addition, R squared is a statistical metric that is used to evaluate the explanatory power of a model as well as its fit to the data. It may be interpreted, and it provides a standardized measure that can be used to compare the performance of the model with a variety of independent variables. It also helps in picking the best appropriate model when several regression models are available, with larger R-squared values suggesting a better fit. This is because R-squared values are proportional to the amount of overlap between the variables. However, one has to consider other aspects, such as the complexity of the model and the possibility of overfitting [26].

6. SIMULATION STDY

By using R software programming, we wish to test how well the proposed new function (AMS) with the mixture of the non-parametric regression of kernel and spline truncated performs, then compare the new proposed function with a Gaussian and biweight kernel. The explanatory variable is generated from $Norm(5,1)$ and the number of the explanatory variable is Considered $(p=8)$ with three different sample sizes $(n=50,130,200)$, random error generated (70%) from $Norm(5,1)$and (30%) from $Norm(10,4)$. Furthermore, the optimal bandwidth parameter and knot points are obtained by using generalized cross-validation (GCV). The generated data are repeated (100) times for each sample size to compute the Mean Square error (MSE) and coefficient of determination R^2. Moreover, we use the following regression functions to generate a new data.

$$Z_{i1} = sin(\pi x) \quad (1)$$
$$Z_{i2} = 5\,sin(2\pi(1-x)^3) \quad (2)$$
$$Z_{i3}, Z_{i7}, Z_{i8} = 100/5x^2 \quad (3)$$
$$Z_{i4} = 20/2 + x + x^2 \quad (4)$$
$$Z_{i5} = 5\,cos(2\pi(1-x)^3) \quad (5)$$
$$Z_{i6} = 4\,cos(2\pi(1-x)^4) \quad (6)$$

Therefore, the combined model of the Non-parametric Regression between the truncated spline and kernel that has been created will have the following form:

$$y_i = z_{i1} + z_{i2} + z_{i3} + z_{i4} + z_{i5} + z_{i6} + z_{i7} + z_{i8} + \varepsilon_i$$

Moreover, suppose that the explanatory variables x_1, x_2, x_3 and x_4 have specific patterns and are used with the truncated spline, which is named v_1, v_2, v_3 and v_4, and x_5, x_6, x_7 and x_8 do not form a particular pattern, then is used with kernel regression ,which is named u_5, u_6, u_7 and u_8.The bandwidth of the Nadaraya-Watson kernel method for nonparametric regression curve estimation is highly important. Hence, in order to achieve optimal outcomes, it is imperative to carefully choose the most suitable bandwidth parameter. Determining the optimal bandwidth h_j for each variable y is performed by evaluating the minimum generalized cross-validation (GCV) using Equation (17). As shown in Table 2, the values of the bandwidth parameter and GCV, MSE, and R^2 for the proposed function AMS, Gaussian and biweight kernels.

Table 2: GCV, MSE, and R-Squares Values To The Proposed Kernel Function (AMS), Gaussian, and Biweight for Different Sample Sizes

n	h_j	GCV (AMS)	MSE (AMS)	R2 (AMS)	GCV (GAUS)	MSE (GAUS)	R2 (GAUS)	GCV (BIW)	MSE (BIW)	R2 (BIW)
50	0.428	**0.00629**	**4.074**	**90.5%**	0.0381	26.408	43.1%	0.35184	142.164	35.18%
	0.368	0.03766	4.533	87.5%	0.1396	27.913	24.4%	0.98402	182.165	13.17%
	0.391	0.33918	5.642	91.8%	0.4394	26.951	22.3%	1.65214	142.676	15.28%
	0.495	0.12758	5.223	87.46%	0.1469	27.483	30.4%	0.97183	122.825	18.03%
130	0.374	**0.00545**	**2.185**	**94.4%**	0.0517	18.843	51.7%	0.65909	123.176	21.40%
	0.409	0.01517	3.457	92.7%	0.0809	21.557	38.5%	0.95840	129.907	20.90%
	0.319	0.22634	4.075	96.3%	0.1772	23.006	35.7%	1.49122	137.139	20.51%
	0.376	0.01925	4.277	91.5%	0.0824	24.047	48.7%	0.8888	119.824	23.04%
200	0.348	**0.00922**	**2.361**	**97.2%**	0.0639	16.291	52.3%	0.73022	111.786	39.23%
	0.429	0.00943	2.375	95.2%	0.0697	17.287	51.6%	0.78517	115.761	42.39%
	0.317	0.01604	3.468	98.7%	0.0815	17.563	54.0%	1.00117	122.824	53.59%
	0.357	0.01310	3.369	97.4%	0.0718	18.389	53.9%	0.73409	112.995	39.65%

According to the simulation results offered in Table 2, it is observed that the proposed function (AMS) of the mixed non-parametric regression model yields the lowest value for the GCV and MSE for different sample sizes (n=50,130, and 200), as highlighted in bold. indicating its superior performance compared to the Gaussian and Biweight functions. Hence, the optimal bandwidths for different sample sizes and for each explanatory variable x_5, x_6, x_7 and x_8, which is named u_5, u_6, u_7 and x_8, are present in Table 3.

Table 3: *Optimal Bandwidth For Different Sample Sizes*

Optimal band widths	Sample size		
	50	130	200
h_5	0.4278	0.3732	0.3475
h_6	0.3674	0.4083	0.4286
h_7	0.3906	0.3192	0.3163
h_8	0.4943	0.3754	0.3561

Additionally, the following step is the choosing of the knot points with the associated explanatory variables x_1, x_2, x_3 and x_4 with a truncated spline method. The function $s_k(v_i)$ is approached by a truncated spline linear function with knots τ_{kj}. In this study, a 1-knot point $(j = 1)$, a 2-knots point $(j = 2)$, a 3-knots point $(j = 3)$ and 4-knots point $(j = 4)$ were used.The number and location of optimal knot points have been determined using the Generalized Cross-Validation (GCV) requirements and MSE, considering the previously acquired optimal bandwidths. Tables (4, 5, 6) shows the number of knot points for different sample sizes (n=50,130, and 200). with associated explanatory variables x_1, x_2, x_3 and x_4,which are named v_1, v_2, v_3 *and* v_4.

Table 4: *GCV and MSE Values With All Knots Point For Different Functions (AMS, Gaus, Biw) With n=50*

n	knots	V_1	V_2	V_3	V_4	GCV (AMS)	MSE (AMS)	R^2 (AMS)	GCV (GAUS)	MSE (GAUS)	R^2 (GAUS)	GCV (BIW)	MSE (BIW)	R2 (BIW)
50	1	**5.15**	**5.14**	**5.35**	**4.64**	**0.00629**	**4.074**	90.5%	0.0381	26.408	43.1%	0.35184	142.164	35.18 %
	2	4.92 5.96	4.98 5.89	5.09 5.88	4.48 5.58	0.03766	4.533	87.5%	0.1396	27.913	24.4%	0.98402	182.165	13.17 %
	3	4.69 5.32 6.33	4.79 5.29 6.32	4.76 5.59 6.04	4.24 4.88 6.38	0.33918	5.642	91.8%	0.4394	26.951	22.3%	1.65214	142.676	15.28 %
	4	4.43 4.37 4.55 3.86	4.92 4.98 5.09 4.48	5.32 5.29 5.59 4.88	5.96 5.89 5.88 5.58	0.12758	5.223	87.46%	0.1469	27.483	30.4%	0.97183	122.825	18.03 %

Table 5: *GCV and MSE Values With all Knots Point For Different Functions (AMS, GAUS, BIW) With n=130*

n	knots	V_1	V_2	V_3	V_4	GCV (AMS)	MSE (AMS)	R^2 (AMS)	GCV (GAUS)	MSE (GAUS)	R^2 (GAUS)	GCV (BIW)	R^2 (BIW)	MSE (BIW)
130	1	**5.17**	**4.99**	**5.09**	**5.14**	**0.00545**	**2.185**	94.4%	0.0517	18.843	51.7%	0.65909	21.40 %	123.176
	2	4.75 5.87	4.78 5.75	4.84 5.89	4.88 5.82	0.01517	3.457	92.7%	0.0809	21.557	38.5%	0.95840	20.90 %	129.907
	3	4.55 5.39 6.19	4.51 5.16 6.27	4.46 5.29 6.13	4.59 5.33 6.28	0.22634	4.075	96.3%	0.1772	23.006	35.7%	1.49122	20.51 %	137.139
	4	4.24 4.75 5.39 5.87	4.31 4.78 5.16 5.75	4.17 4.84 5.29 5.89	4.27 4.88 5.33 5.82	0.01925	4.277	91.5%	0.0824	24.047	48.7%	0.8888	23.04 %	119.824

Table 6: *GCV and MSE Values With all Knots Point For Different Functions (AMS, Gaus, Biw) With n=200*

n	knots	V_1	V_2	V_3	V_4	GCV (AMS)	MSE (AMS)	R^2 (AMS)	GCV (GAUS)	MSE (GAUS)	R^2 (GAUS)	GCV (BIW)	MSE (BIW)	R^2 (BIW)
200	1	**4.96**	**5.09**	**5.01**	**5.05**	**0.00922**	**2.361**	97.2%	0.0639	16.291	52.3%	0.73022	111.79	39.23%
	2	4.69 5.69	4.85 5.89	4.82 5.81	4.77 5.98	0.00943	2.375	95.2%	0.0697	17.287	51.6%	0.78517	115.77	42.39%
	3	4.24 4.96 5.69	4.18 5.09 5.89	4.23 5.01 5.81	4.06 5.05 5.98	0.01604	3.468	98.7%	0.0815	17.563	54.0%	1.00117	122.83	53.59%
	4	4.24 4.69 5.19 5.69	4.18 4.85 5.29 5.89	4.23 4.82 5.31 5.81	4.06 4.77 5.27 5.98	0.01310	3.369	97.4%	0.0718	18.389	53.9%	0.73409	112.99	39.65%

The GCV, MSE, and R^2 values are computed for one knot, two knots, three knots, and four knots for different sample sizes (n=50,130, and 200) in Tables 4, 5, and 6. It shows that the GCV and MSE exhibit their lowest values at one knot for all sample sizes, and the R^2 values for each sample at one knot point are 95.4%, 94.4%, and 97.2% respectively. When compared, the proposed function AMS exhibits the lowest value among GUAS and BIW. Hence, considering the minimum values of GCV and MSE achieved through the utilization of one knot, the optimal knot points can be determined and are indicated in Table 7.

Table 7: *Optimal Knot Points For Each Sample Size*

Independent variables	Optimal knot points		
	n		
	50	130	200
V_1	5.15	5.17	4.96
V_2	5.14	4.99	5.09
V_3	5.34	5.09	5.01
V_4	4.64	5.14	5.05

Based on the optimal bandwidth parameter values we showed in Table 3 and the proposed function (AMS), here is the kernel estimator formula for the explanatory variables that were used with the kernel for different sample sizes (n = 50, 130, and 200). Respectively

$$\hat{m}_h\left(u_5,u_6,u_7,u_8,h\right)=\frac{\sum^{50}\left[k\left(\frac{u_5-u_{i5}}{0.4278}\right).k\left(\frac{u_6-u_{i6}}{0.3674}\right).k\left(\frac{u_7-u_{i7}}{0.3906}\right).k\left(\frac{u_8-u_{i8}}{0.4943}\right)\right]y_i}{\sum^{50}\left[k\left(\frac{u_5-u_{i5}}{0.4278}\right).k\left(\frac{u_6-u_{i6}}{0.3674}\right).k\left(\frac{u_7-u_{i7}}{0.3906}\right).k\left(\frac{u_8-u_{i8}}{0.4943}\right)\right]} \quad (1)$$

$$\hat{m}_h\left(u_5,u_6,u_7,u_8,h\right)=\frac{\sum^{130}\left[k\left(\frac{u_5-u_{i5}}{0.3732}\right).k\left(\frac{u_6-u_{i6}}{0.4083}\right).k\left(\frac{u_7-u_{i7}}{0.3192}\right).k\left(\frac{u_8-u_{i8}}{0.3754}\right)\right]y_i}{\sum^{130}\left[k\left(\frac{u_5-u_{i5}}{0.3732}\right).k\left(\frac{u_6-u_{i6}}{0.4083}\right).k\left(\frac{u_7-u_{i7}}{0.3192}\right).k\left(\frac{u_8-u_{i8}}{0.3754}\right)\right]} \quad (2)$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$\hat{m}_h(u_5,u_6,u_7,u_8,h)=\frac{\sum^{200}\left[k\left(\frac{u_5-u_{i5}}{0.3475}\right).k\left(\frac{u_6-u_{i6}}{0.4286}\right).k\left(\frac{u_7-u_{i7}}{0.3163}\right).k\left(\frac{u_8-u_{i8}}{0.3561}\right)\right].y_i}{\sum^{200}\left[k\left(\frac{u_5-u_{i5}}{0.3475}\right).k\left(\frac{u_6-u_{i6}}{0.4286}\right).k\left(\frac{u_7-u_{i7}}{0.3163}\right).k\left(\frac{u_8-u_{i8}}{0.3561}\right)\right]} \quad (3)$$

With $h=[h_5,h_6,h_7,h_8]^T$.

Also, the estimation of the unknown truncated spline parameters (θ, v, and k) for the mixture nonparametric regression model with truncated spline and kernel is done at a one-knot point for each relevant explanatory variable and the optimal bandwidth. The results of this estimation are as follows: (see Table 8)

Table 8: *Estimates Of The Parameters For The Truncated Spline To One Knot Point For Different Sample Size.*

	50		130		200	
Independent variables	**Parameters**	**Estimates**	**Parameters**	**Estimates**	**Parameters**	**Estimates**
Constant	θ_0	0.5144	θ_0	-0.1969	θ_0	0.4418
v_1	θ_1	-0.0025	θ_1	-0.2075	θ_1	-0.3565
	β_{11}	0.2328	β_{11}	0.04543	β_{11}	0.8227
v_2	θ_2	-0.0622	θ_2	0.2251	θ_2	-0.06903
	β_{21}	-0.1735	β_{21}	0.07107	β_{21}	-0.4608
v_3	θ_3	-0.1504	θ_3	0.07296	θ_3	0.3973
	β_{31}	-0.3078	β_{31}	-0.1921	β_{31}	-0.5987
v_4	θ_4	-1.2247	θ_4	-0.9409	θ_4	1.1951
	β_{41}	0.7054	β_{41}	0.7251	β_{41}	0.01236

Therefore, the mixed non-parametric model between the kernel and the truncated spline with the use of the proposed kernel function (AMS) for each sample size (n = 50, 130, and 200) is given as follows:

$$\hat{y}_i=0.5144-0.0025\,v_{1i}-0.0622\,v_{2i}-0.1504\,v_{3i}-1.2247\,v_{4i}+0.2328(v_{1i}-5.15)_+^1$$
$$-0.1735(v_{2i}-5.14)_+^1-0.3078(v_{3i}-5.34)_+^1+0.7054(v_{4i}-4.64)_+^1$$
$$+\frac{\sum^{50}\left[k\left(\frac{u_5-u_{i5}}{0.4278}\right).k\left(\frac{u_6-u_{i6}}{0.3674}\right).k\left(\frac{u_7-u_{i7}}{0.3906}\right).k\left(\frac{u_8-u_{i8}}{0.4943}\right)\right].y_i}{\sum^{50}\left[k\left(\frac{u_5-u_{i5}}{0.4278}\right).k\left(\frac{u_6-u_{i6}}{0.3674}\right).k\left(\frac{u_7-u_{i7}}{0.3906}\right).k\left(\frac{u_8-u_{i8}}{0.4943}\right)\right]} \quad (1)$$

$$\hat{y}_i=-0.1969-0.2075\,v_{1i}+0.2251\,v_{2i}+0.0296\,v_{3i}-0.9409\,v_{4i}+0.4543(v_{1i}-5.17)_+^1$$
$$-0.01707(v_{2i}-4.99)_+^1-0.1921(v_{3i}-5.09)_+^1+0.7251(v_{4i}-5.14)_+^1$$
$$+\frac{\sum^{130}\left[k\left(\frac{u_5-u_{i5}}{0.3732}\right).k\left(\frac{u_6-u_{i6}}{0.4083}\right).k\left(\frac{u_7-u_{i7}}{0.3192}\right).k\left(\frac{u_8-u_{i8}}{0.3754}\right)\right].y_i}{\sum^{130}\left[k\left(\frac{u_5-u_{i5}}{0.3732}\right).k\left(\frac{u_6-u_{i6}}{0.4083}\right).k\left(\frac{u_7-u_{i7}}{0.3192}\right).k\left(\frac{u_8-u_{i8}}{0.3754}\right)\right]} \quad (2)$$

$$\hat{y}_i=0.4418-0.3565\,v_{1i}+0.06903\,v_{2i}+0.3973\,v_{3i}+1.1951\,v_{4i}+0.8227(v_{1i}-4.96)_+^1$$
$$-0.4608(v_{2i}-5.09)_+^1-0.5987(v_{3i}-5.01)_+^1+0.01236(v_{4i}-5.05)_+^1$$
$$+\frac{\sum^{200}\left[k\left(\frac{u_5-u_{i5}}{0.3475}\right).k\left(\frac{u_6-u_{i6}}{0.4286}\right).k\left(\frac{u_7-u_{i7}}{0.3163}\right).k\left(\frac{u_8-u_{i8}}{0.3561}\right)\right].y_i}{\sum^{200}\left[k\left(\frac{u_5-u_{i5}}{0.3475}\right).k\left(\frac{u_6-u_{i6}}{0.4286}\right).k\left(\frac{u_7-u_{i7}}{0.3163}\right).k\left(\frac{u_8-u_{i8}}{0.3561}\right)\right]} \quad (3)$$

Where

$k(u)=\frac{122377}{156250}exp[-(u^6+u^4+u^2)], u\in[-\infty,\infty]$

15th November 2023. Vol.101. No 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

is the proposed new function (AMS) for the kernel. The mixed model with the proposed function (AMS) achieves the highest coefficient of determination R^2, for each sample size (n=50,130 and 200), 90.5%,94.4% and 97.2% respectively, compared to Gaussian mixed and Biweight mixed.

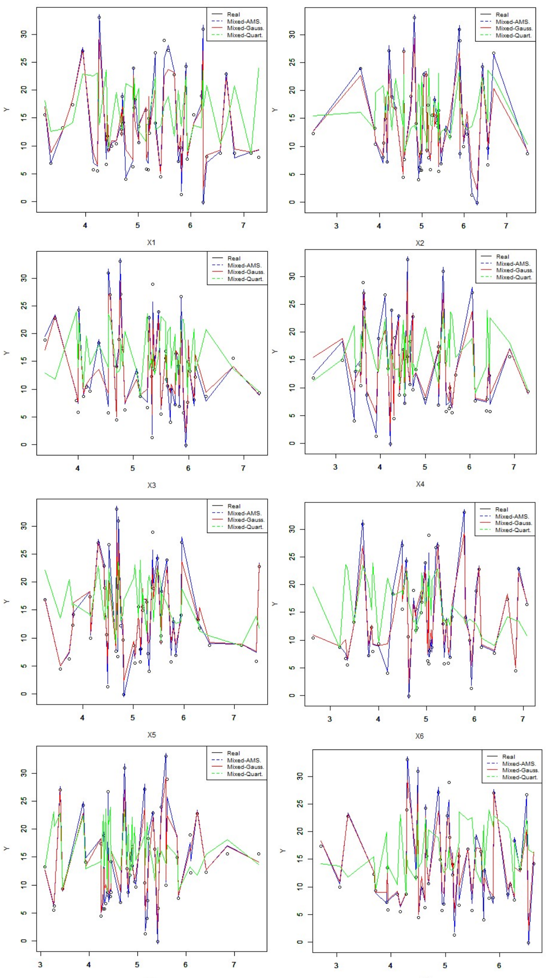

Figure 1. Plot Of y and The Original Simulation Data Is Generated With The Mixed Model Using Proposed Function (AMS), Mixed Gaussian, and Mixed Biweight Using Sample Size 50.

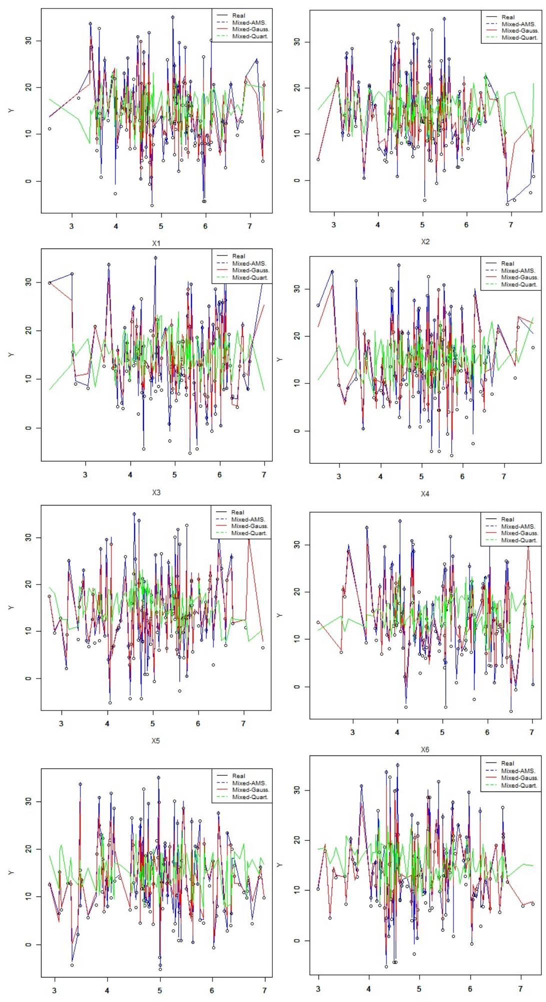

Figure 2. Plot Of y and The Original Simulation Data Is Generated With The Mixed Model Using Proposed Function (AMS), Mixed Gaussian, and Mixed Biweight Using Sample Size 130.

15th November 2023. Vol.101. No 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

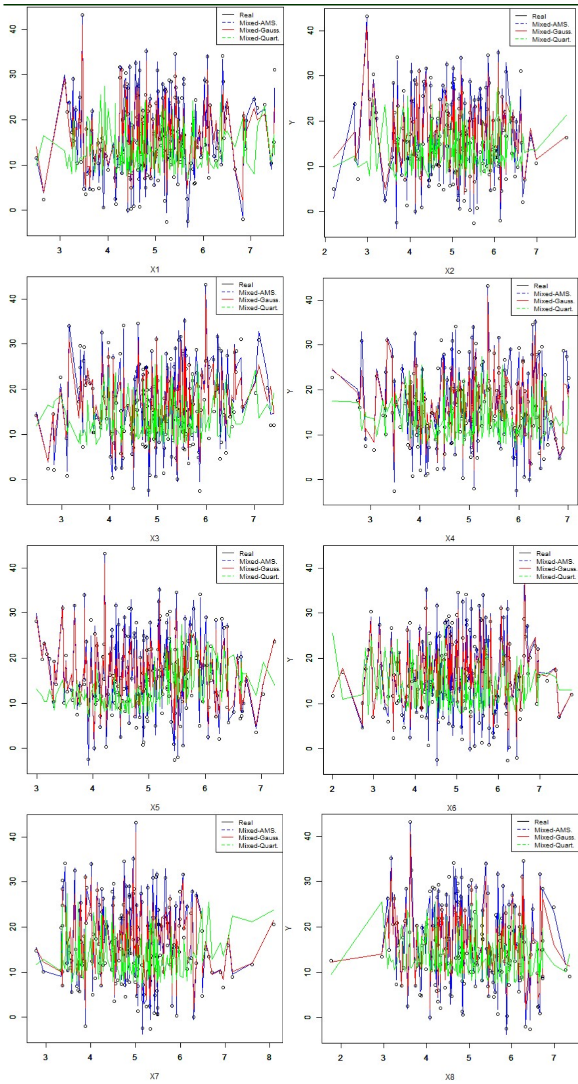

Figure 3. Plot Of y and The Original Simulation Data is generated with the mixed model using Proposed Function (AMS), Mixed Gaussian, and Mixed Biweight Using Sample Size 200.

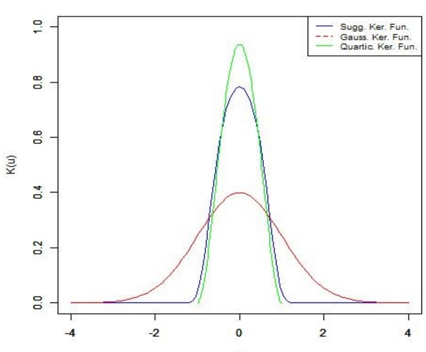

Figure 4. Shows That The Curve Of The Proposed Function (AMS) With Gaussian and Biweight.

Gaussian, and biweight were computed based on the sample sizes 50, 130, and 200 and are presented in Figures 1, 2, and 3. While the performance of the mixed nonparametric regression model using the suggested function (AMS) is superior and comparable with the mixed Gaussian, and mixed biweight models for different sample sizes, our comparable study is based on GCV and MSE; for each sample size, the GCV, MSE, and R^2values of the mixed (AMS), mixed Gaussian and mixed biweight are shown in Tables 4, 5, and 6 previously. Furthermore, the estimated mixed (AMS) model is extremely close to the simulation of the original data. Consequently, one may use this model to make highly accurate predictions. Compared to the Gaussian and the biweight mixed models, the mixed model (AMS) has the lowest values of GCV, MSE, and value of R^2, which are 90.5%, 95.4%, and 97.2%, respectively.

7.CONCLUSION AND FUTURE WORKS

This work presents a new function for the kernel and applies it to the mixed nonparametric regression model. The optimum bandwidth for each variable and the optimal one-knot point were determined for the best model for different sample sizes, considering the value of generalized Cross-Validation (GCV) and Mean square error (MSE) within the range of options explored. The simulation study conducted in Section 4 demonstrated that the mixed nonparametric model, which combines the kernel and spline truncated if implemented with the proposed function (AMS), demonstrates favourable suitability and superior predictive performance very well compared to the mixed Gaussian and mixed biweight models. This conclusion is confirmed by the coefficients of determination R^2, which are 90.5%, 94.4%, and 97.2%, respectively. The mixed nonparametric model (AMS) obtained the lowest mean square error values of 4.074, 2.185, and 2.361. The results show that the proposed function (AMS) for the kernel significantly improves the accuracy of the mixed nonparametric regression model. Additionally, it suggests that the mixed nonparametric model (AMS) not only outperforms the mixed Gaussian and mixed biweight models in terms of predictive accuracy but also provides a more precise estimation of the data. Furthermore, the low mean square error values indicate that the mixed nonparametric model is able to capture the underlying patterns in the data more effectively. On the other hand, one knot point was chosen based on the value (GCV). The accuracy of the mixed model may increase as the number of optimal knot points increases. To improve the accuracy of the mixed non-parametric regression model, we should investigate and compare diverse kernel functions. This will provide insight into their applicability to specific datasets and improve customization. The most efficient function for enhancing the model's precision can be identified through comparative analysis.

In the future, more research could be done to make the proposed field of study wider. This could include adding more explanatory variables and different kernel functions, as well as using different criteria to find the best bandwidth parameters and knot points. The model could also be used on real data from other fields to come up with new methods to make it more accurate and complete.

REFERENCE:

[1] L. A. Mohammed and M. Abd Hassan, "Multiple RegressionAnalysis with Epanechnikov & Gaussian Functions," *J. Econ. Adm. Sci.*, no. 59, pp. 1–14, 2018.
[2] J. S. Klemelä, *Multivariate nonparametric regression and visualization: With R and applications to finance*. John Wiley & Sons, 2014.
[3] R. L. Eubank, *Nonparametric regression and spline smoothing*. CRC press, 1999.
[4] I. N. Budiantara, V. Ratnasari, M. Ratna, and I. Zain, "The combination of spline and kernel estimator for nonparametric regression and its properties," *Appl. Math. Sci.*, vol. 9, no. 122, pp. 6083–6094, 2015.
[5] M. L. .Ali and M. Abd Hassan, "Comparison of nonparametric estimators in multiple regression analysis for Gamma and Beta functions," *J. Econ. Adm. Sci.*, vol. 24, no. 108, pp. 488–497, Jul. 2018.
[6] N. Chamidah and T. Saifudin, "Estimation of children growth curve based on kernel smoothing in multi-response nonparametric regression," *Appl. Math. Sci.*, vol. 7, no. 37, pp. 1839–1847, 2013.
[7] W. Härdle, *Applied nonparametric regression*, no. 19. Cambridge university press, 1990.
[8] I. N. Budiantara, M. Ratna, I. Zain, and W. Wibowo, "Modeling The Percentage of Poor People in Indonesia Using Spline Nonparametric Regression Approach," *Int. J. Basic Appl. Sci.*, vol. 12, no. 06, pp. 119–124, 2012.
[9] D. Aydin, "A Comparison of the Nonparametric Regression Models using Smoothing Spline and Kernel regression," *Int. J. Math. Comput. Sci. Eng.*, vol. 1, no. 12, pp. 253–257, 2007.
[10] F. Ali, "Bayesian Methods for Estimation the Parameters of Finite Mixture of Inverse Rayleigh Distribution," *Math. Probl. Eng.*, vol. 2023, 2023.
[11] N. Mohammed and F. Ali, "Estimation of parameters of finite mixture of Rayleigh distribution by the expectation-maximization algorithm," *J. Math.*, vol. 2022, 2022.
[12] F. Ali and J. Zhang, "Mixture model-based association analysis with case-control data in genome wide association studies," *Stat. Appl. Genet. Mol. Biol.*, vol. 16, no. 3, pp. 173–187, 2017, doi: 10.1515/sagmb-2016-0022.
[13] A. Prahutama and R. Santoso, "Mix local polynomial and spline truncated: the development of nonparametric regression model," in *Journal of Physics: Conference Series*, 2018, vol. 1025, no. 1, p. 12102.
[14] M. J. Maulidia, I. N. Budiantara, and J. D. T. Purnomo, "Nonparametric regression curve estimation using mixed spline truncated and kernel estimator for longitudinal data," in *AIP Conference Proceedings*, 2019, vol. 2194, no. 1, p. 20063.
[15] N. Y. Adrianingsih, I. N. Budiantara, and J. D. T. Purnomo, "Modeling with Mixed Kernel, Spline Truncated and Fourier Series on Human Development Index in East Java," in *IOP Conference Series: Materials Science and Engineering*, 2021, vol. 1115, no. 1, p. 12024.
[16] V. Ratnasari, I. N. Budiantara, and A. T. R. Dani, "Nonparametric Regression Mixed Estimators of Truncated Spline and Gaussian Kernel based on Cross-Validation (CV), Generalized CrossValidation (GCV), and Unbiased Risk (UBR) Methods," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 11, no. 6, pp. 2400–2406, 2021, doi: 10.18517/ijaseit.11.6.14464.
[17] H. Nurcahayani, I. N. Budiantara, and I. Zain, "The semiparametric regression curve estimation by using mixed truncated spline and fourier series model," in *AIP Conference Proceedings*, 2021, vol. 2329, no. 1. doi: 10.1063/5.0042870.
[18] I. N. Budiantara, "Inferensi statistik untuk model spline," *J. Mat-Stat*, vol. 7, no. 1, pp. 1–14, 2007.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[19] E. A. Nadaraya, "On estimating regression," *Theory Probab. Its Appl.*, vol. 9, no. 1, pp. 141–142, 1964.

[20] G. S. Watson, "Smooth regression analysis," *Sankhyā Indian J. Stat. Ser. A*, pp. 359–372, 1964.

[21] W. Härdle, M. Müller, S. Sperlich, and A. Werwatz, *Nonparametric and semiparametric models*, vol. 1. Springer, 2004.

[22] V. Ratnasari, I. N. Budiantara, M. Ratna, and I. Zain, "Estimation of nonparametric regression curve using mixed estimator of multivariable truncated spline and multivariable kernel," *Glob. J. Pure Appl. Math.*, vol. 12, no. 6, pp. 5047–5057, 2016.

[23] A. T. R. Dani, V. Ratnasari, and I. N. Budiantara, "Optimal Knots Point and Bandwidth Selection in Modeling Mixed Estimator Nonparametric Regression," in *IOP Conference Series: Materials Science and Engineering*, 2021, vol. 1115, no. 1, p. 12020.

[24] C. L. Cheng, Shalabh, and G. Garg, "Coefficient of determination for multiple measurement error models," *J. Multivar. Anal.*, vol. 126, pp. 137–152, 2014, doi: 10.1016/j.jmva.2014.01.006.

[25] D. M. Allen, "Mean square error of prediction as a criterion for selecting variables," *Technometrics*, vol. 13, no. 3, pp. 469–475, 1971, doi: 10.1080/00401706.1971.10488811.

[26] E. R. Dougherty, S. Kim, and Y. Chen, "Coefficient of determination in nonlinear signal processing," *Signal Processing*, vol. 80, no. 10, pp. 2219–2235, 2000, doi: 10.1016/S0165-1684(00)00079-7.