



Available online at <http://scik.org>

Commun. Math. Biol. Neurosci. 2023, 2023:48

<https://doi.org/10.28919/cmbn/7903>

ISSN: 2052-2541

BI-RESPONSE TRUNCATED SPLINE NONPARAMETRIC REGRESSION WITH OPTIMAL KNOT POINT SELECTION USING GENERALIZED CROSS- VALIDATION IN DIABETES MELLITUS PATIENT'S BLOOD SUGAR LEVELS

SIFRIYANI^{1,2,*}, AR RUM MIA SARI^{1,2}, ANDREA TRI RIAN DANI¹, SYATIRAH JALALUDDIN³

¹Department of Statistics, Faculty of Mathematics and Natural Sciences, Mulawarman University, Samarinda, Indonesia

²Laboratory of Applied Statistics, Department of Mathematics, Mulawarman University, Samarinda, Indonesia

³Department of Statistics, Universitas Negeri Makassar, Makassar, Indonesia

³Department of Statistics, Faculty of Mathematics and Natural Sciences, Lambung Mangkurat University, Banjarmasin, Indonesia

⁵School of Medicine, Universitas Islam Negeri Alauddin, Makassar, Indonesia

Copyright © 2023 the author(s). This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract: This article discusses statistical modeling implemented in the health sector. This study used a bi-response nonparametric regression method with truncated spline estimation that used two response variables. The nonparametric regression method is used when the regression curve is not known for its shape and pattern. This study aims to model the blood sugar levels of people with diabetes mellitus. The data used are blood sugar levels of people with diabetes mellitus before fasting, blood sugar levels of people with diabetes mellitus two hours after fasting, cholesterol levels, and triglyceride levels. Determination of the optimal knot point using Generalized Cross-Validation. The parameter estimation method used is Weighted Least-Squares. The best model was obtained from the study results,

*Corresponding author

E-mail address: sifriyani@fmipa.unmul.ac.id

Received February 05, 2023

namely the bi-response truncated spline regression model with three-knot points where the minimum GCV value is 8.573 and has an R^2 value of 99.62%.

Keywords: bi-response; diabetes mellitus; generalized cross-validation; knots; truncated spline.

2020 AMS Subject Classification: 92C50.

1. INTRODUCTION

Regression analysis is a way that can be used to determine the relationship of a response variable with one or more predictor variables [1]. Regression analysis is an analytical method used to analyze data and draw meaningful conclusions about the relationship of variable dependence on other variables [2]. In the regression analysis, there are three approaches to estimating the regression curve: parametric, nonparametric, and semiparametric [3], [4]. If data in the form of a regression curve forms a known pattern, such as a linear, quadratic, cubic, or polynomial form of degree k , then the proper approach is parametric. The nonparametric approach is used when the shape of the regression curve of data is unknown. A semiparametric method is used when the shape of the regression curve is partly known and partly the pattern is unknown [5], [6].

The state of the data where the regression function is unknown often makes researchers decide to complete their research or analyze the data using nonparametric regression analysis [7]. Some authors, such as Hardle [8] and Wahba [9], suggest using nonparametric regression to model data for good flexibility. The nonparametric regression model is a regression method that is used when the regression curve between the response variable and the predictor variable is unknown in shape or pattern [10], [11]. This is because the nonparametric regression model has high flexibility in forming the regression curve and, in particular, does not require parametric assumptions [12]. Several nonparametric regression model approaches that are widely used include Kernel [13]–[16], truncated spline [17]–[20], Fourier series [14], [21], [22], and so on. A popular nonparametric regression approach is the truncated spline.

A truncated spline is a truncated polynomial function of order k , in which the part has connecting points called knot points [3], [23]. The knot point is the joint point where there is a

CROSS-VALIDATION IN DIABETES MELLITUS PATIENT'S BLOOD SUGAR LEVELS

change in the behavior pattern of the function or curve. The order in the function indicates the degree of the polynomial degree of the function. These knot points and orders will then be used to determine the truncated spline regression model [10]. Truncated splines can describe changes in behavior patterns and processes at sub-intervals, overcome data patterns that show sharp ups or downs with the help of knots, and the curves produced are relatively smooth [3].

Current nonparametric regression research has been studied and developed based on the type of response and the number of response variables involved in regression modeling. Regression research based on the number of responses is divided into single-response, bi-response, and multiresponse regression models [24], [25]. The truncated spline regression analysis used to investigate one predictor variable with one response variable is called a univariable truncated spline. If, in the regression analysis, there is one response variable with more than one predictor variable, then the regression is called a multivariable truncated spline. At the same time, the regression analysis in which there are two response variables is called bi-response truncated spline regression. Biresponse Truncated Spline Regression is included in multivariate regression because it consists of one or more correlated response variables and one or more predictor variables [26], [27]. The existence of the response correlation in the bi-response regression model can be determined using the Pearson correlation coefficient. Some studies that examine bi-response truncated spline regression include: [24], [25], [28]

The application of regression analysis is widely used in various problems in the health sector. One of the cases of health with a high risk of death is diabetes. WHO says diabetes mellitus can be defined as a disease or condition with a chronic metabolic disorder with multiple etiologies, which can be characterized by high blood sugar levels and also accompanied by metabolic disturbances in carbohydrates, lipids, and proteins as a result of functional insufficiency [29], [30].

Diabetes doesn't just cause premature death worldwide. This disease is also a significant cause of blindness, heart disease, and kidney failure. The International Diabetes Federation (IDF) organization estimates that at least 463 million people aged 20-79 years in the world have diabetes in 2019, or the equivalent of a prevalence rate of 9.3% of the total population at the same age. The

prevalence of diabetes is estimated to increase with the increasing age of the population to 19.9% or 111.2 million people aged 65-79 years. The number is predicted to grow until it reaches 578 million in 2030 and 700 million in 2045 [31]. Based on this background, researchers used bi-response truncated spline regression and applied a model of blood sugar levels in patients with diabetes mellitus in inpatients at Abdul Wahab Sjahranie Hospital (AWS) Samarinda, East Kalimantan.

2. PRELIMINARIES

A. Nonparametric Regression

In general, the general nonparametric regression model with data pairs (x_i, y_i) is written in Equation (1).

$$y_i = f(x_i) + \varepsilon_i \quad (1)$$

With

y_i : response variable

x_i : predictor variable

$f(x_i)$: regression curve with no known relationship pattern

ε_i : random errors $\varepsilon_i \sim IIDN(0, \sigma^2)$.

The general purpose of regression analysis is to estimate or predict a value of the response variable when the predictor variable is assigned a value [32] – [34]. In other words, it is to find the form of parameter estimation that matches the shape of the regression curve.

B. Truncated Spline Nonparametric Regression

A truncated spline is a regression model that can adapt to changes in data patterns, so it is often said to have high flexibility. The truncated spline function is written in Equation (2).

$$f(x_i) = \beta_0 + \sum_{h=1}^m \beta_h x_i^h + \sum_{k=1}^r \beta_{m+k} (x_i - K_k)_+^m \quad (2)$$

If Equation (2) is substituted in Equation (1), a truncated spline nonparametric regression model

will be obtained.

$$y_i = \beta_0 + \sum_{h=1}^m \beta_h x_i^h + \sum_{k=1}^r \beta_{m+k} (x_i - K_k)_+^m + \varepsilon_i \quad (3)$$

The truncated function in Equation (3) is described in Equation (4).

$$(x_i - K_k)_+^m = \begin{cases} (x_i - K_k)_+^m, & x_i \geq K_k \\ 0, & x_i < K_k \end{cases} \quad (4)$$

The regression model in Equation (3) can be written in matrix form as in Equation (5).

$$\mathbf{y} = \mathbf{X}(k)\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (5)$$

Using the Maximum Likelihood Estimation (MLE) in the parameter estimation process. Obtained parameter estimation of $\hat{\boldsymbol{\beta}}$ in Equation (6).

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}(k)^T \mathbf{X}(k))^{-1} \mathbf{X}(k)^T \mathbf{y} \quad (6)$$

C. Bi-response Truncated Spline Nonparametric Regression

Bi-response truncated spline is defined as one of the nonparametric regression models with more than one response variable. Between these response variables, there is a strong correlation or relationship. The model for bi-response truncated spline is written in Equation (7).

$$\begin{aligned} y_{1i} &= \sum_{j=1}^l m(x_{ij}) + \varepsilon_{1i} \\ y_{2i} &= \sum_{j=1}^l z(x_{ij}) + \varepsilon_{2i} \end{aligned} \quad (7)$$

The functions $m(x_{ij})$ and $z(x_{ij})$ are regression curves of unknown shape and will be approximated by a bi-respon truncated spline in Equation (8).

$$\begin{aligned} m(x_{ij}) &= \delta + \sum_{j=1}^l \sum_{h=1}^m \xi_{hj} x_{ij}^h + \sum_{j=1}^l \sum_{k=1}^r \varphi_{m+k,j} (x_{ij} - K_{kj})_+^m \\ z(x_{ij}) &= \phi + \sum_{j=1}^l \sum_{h=1}^m \psi_{hj} x_{ij}^h + \sum_{j=1}^l \sum_{k=1}^r \gamma_{m+k,j} (x_{ij} - K_{kj})_+^m \end{aligned} \quad (8)$$

The regression model in Equation (8) can be written in matrix notation as:

$$\mathbf{y} = \mathbf{X}(k)\boldsymbol{\eta} + \boldsymbol{\varepsilon} \quad (9)$$

With:

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \text{---} \\ \mathbf{y}_2 \end{pmatrix} = \begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n} \\ \text{---} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2n} \end{pmatrix}; \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \text{---} \\ \boldsymbol{\varepsilon}_2 \end{pmatrix} = \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{1n} \\ \text{---} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \vdots \\ \varepsilon_{2n} \end{pmatrix}; \quad \mathbf{X}(\Phi) = \begin{pmatrix} \mathbf{C} & \cdots & \mathbf{0} \\ \cdots & \ddots & \cdots \\ \mathbf{0} & \cdots & \mathbf{D} \end{pmatrix}$$

Where,

$$\mathbf{C} = \begin{pmatrix} 1 & x_{11}^1 & \cdots & x_{11}^m & (x_{11} - K_1^1)_+^m & \cdots & (x_{11} - K_r^1)_+^m & \cdots & x_{l1}^1 & \cdots & x_{l1}^m & (x_{l1} - K_1^1)_+^m & \cdots & (x_{l1} - K_r^1)_+^m \\ 1 & x_{12}^1 & \cdots & x_{12}^m & (x_{12} - K_1^1)_+^m & \cdots & (x_{12} - K_r^1)_+^m & \cdots & x_{l2}^1 & \cdots & x_{l2}^m & (x_{l2} - K_1^1)_+^m & \cdots & (x_{l2} - K_r^1)_+^m \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n}^1 & \cdots & x_{1n}^m & (x_{1n} - K_1^1)_+^m & \cdots & (x_{1n} - K_r^1)_+^m & \cdots & x_{ln}^1 & \cdots & x_{ln}^m & (x_{ln} - K_1^1)_+^m & \cdots & (x_{ln} - K_r^1)_+^m \end{pmatrix}$$

$$\mathbf{D} = \begin{pmatrix} 1 & x_{11}^1 & \cdots & x_{11}^m & (x_{11} - K_1^1)_+^m & \cdots & (x_{11} - K_r^1)_+^m & \cdots & x_{l1}^1 & \cdots & x_{l1}^m & (x_{l1} - K_1^1)_+^m & \cdots & (x_{l1} - K_r^1)_+^m \\ 1 & x_{12}^1 & \cdots & x_{12}^m & (x_{12} - K_1^1)_+^m & \cdots & (x_{12} - K_r^1)_+^m & \cdots & x_{l2}^1 & \cdots & x_{l2}^m & (x_{l2} - K_1^1)_+^m & \cdots & (x_{l2} - K_r^1)_+^m \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n}^1 & \cdots & x_{1n}^m & (x_{1n} - K_1^1)_+^m & \cdots & (x_{1n} - K_r^1)_+^m & \cdots & x_{ln}^1 & \cdots & x_{ln}^m & (x_{ln} - K_1^1)_+^m & \cdots & (x_{ln} - K_r^1)_+^m \end{pmatrix}$$

Matrix $\mathbf{0}$ is a null matrix of size $n \times (l(m+r))$.

$$\boldsymbol{\eta} = \begin{pmatrix} \boldsymbol{\eta}_1 \\ \text{---} \\ \boldsymbol{\eta}_2 \end{pmatrix}$$

With:

$$\boldsymbol{\eta}_1 = (\delta, \xi_{11}, \xi_{12}, \dots, \xi_{ml}, \varphi_{(m+1)l}, \varphi_{(m+2)l}, \dots, \varphi_{(m+r)l})^T$$

$$\boldsymbol{\eta}_2 = (\phi, \psi_{11}, \psi_{12}, \dots, \psi_{ml}, \gamma_{(m+1)l}, \gamma_{(m+2)l}, \dots, \gamma_{(m+r)l})^T$$

The parameter estimation method used is Weighted Least Squares (WLS). The advantage of WLS is that it can overcome correlations in the same observation subject. In nonparametric regression, bi-response truncated spline, there is a correlation between the error in the first response ε_{1i} and the error in the second response ε_{2i} . The \mathbf{W} weighting matrix is presented in Equation (10).

$$\mathbf{W} = \left(\begin{array}{cccc|cccc} \sigma_1^2 & 0 & \cdots & 0 & \sigma_{11} & 0 & \cdots & 0 \\ 0 & \sigma_1^2 & \cdots & 0 & 0 & \sigma_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_1^2 & 0 & 0 & \cdots & \sigma_{nn} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \sigma_{11} & 0 & \cdots & 0 & \sigma_2^2 & 0 & \cdots & 0 \\ 0 & \sigma_{22} & \cdots & 0 & 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{nn} & 0 & 0 & \cdots & \sigma_2^2 \end{array} \right) \quad (10)$$

Parameter estimation of $\hat{\boldsymbol{\eta}}$ is written in Equation (11).

$$\hat{\boldsymbol{\eta}} = (\mathbf{X}(\Phi)^T \mathbf{W} \mathbf{X}(\Phi))^{-1} \mathbf{X}(\Phi)^T \mathbf{W} \mathbf{y} \quad (11)$$

D. Generalized Cross-Validation (GCV)

Theoretically, the GCV method has asymptotically optimal properties, is invariant to transformation, and does not require variance information in its calculations. The formula for selecting the optimal knot point in the bi-response truncated spline regression using GCV is written in Equation (12).

$$GCV(\Phi) = \frac{MSE(\Phi)}{\left(\frac{1}{2n} \text{trace}(\mathbf{I} - \mathbf{A}(\Phi)) \right)^2} \quad (12)$$

With

$$MSE(\Phi) = n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$\mathbf{A}(\Phi) = \mathbf{X}(\Phi) (\mathbf{X}(\Phi)^T \mathbf{W} \mathbf{X}(\Phi))^{-1} \mathbf{X}(\Phi)^T \mathbf{W}$$

3. RESEARCH METHODOLOGY

A. Data Sources

The data used in this study is secondary data obtained from Abdul Wahab Syahrani Hospital (AWS) in 2022. The research variables used in this study are presented in Table 1.

| Variable | Notation | Description |
|---|----------|--|
| Fasting blood sugar levels | y_1 | Blood sugar levels of patients with diabetes mellitus were obtained before eating. This test is done after fasting or after not consuming food/drink for eight hours |
| Blood sugar level two hours after fasting | y_2 | Blood sugar levels of patients with diabetes mellitus were obtained after consuming food/drinks for two hours. |
| Total cholesterol levels | x_1 | Total cholesterol levels are taken in patients with diabetes mellitus. |
| Triglyceride levels | x_2 | Triglycerides are a type of fat that is found in the blood. Triglyceride levels were taken in patients with diabetes mellitus. |

TABLE 1. Description of Study Variables

B. Data Analysis Technique

Data analysis techniques in this study consisted of descriptive statistical analysis and bi-response truncated spline nonparametric regression analysis using the help of software R.

1. Define response variables and predictor variables.
2. Make descriptive statistics for response variables and predictor variables.
3. Perform correlation testing between response variables.
4. Make a scatter plot between the predictor and response variables to find the relationship pattern.
5. Modeling blood sugar levels in patients with diabetes mellitus with a linear spline of one knot, two knots, and three knots with the following stages.
 - a. Create a nonparametric bi-response regression model.
 - b. Determine the W matrix used as a weighting.
 - c. Weighted Least Square (WLS) optimization.

- d. Get the value of $\hat{\eta}$.
 - e. Get estimates for the regression curve $m(x_{ij})$ and $z(x_{ij})$
 - f. Determine the optimal knot point using the Generalized Cross Validation (GCV) method, where the smallest GCV value is the optimal knot point.
6. Calculates the R^2 value of the model.

4. MAIN RESULTS

In this section, we will explain the results of a bi-response truncated spline nonparametric regression study applied to blood sugar levels of people with diabetes mellitus in inpatients at the Abdul Wahab Sjahranie Hospital (AWS) Samarinda, East Kalimantan.

A. Descriptive Statistics

Description of the data using descriptive statistical analysis is used to see an overview of the data used. The descriptive statistical analysis used in this study is the average, minimum, maximum, and standard deviation. The results of the descriptive statistical analysis are presented in Table 2.

| Variable | n | Mean | Minimum | Maximum | Standard Deviation |
|----------|-----|---------|---------|---------|--------------------|
| y_1 | 115 | 213.435 | 60 | 427 | 94.990 |
| y_2 | 115 | 233.956 | 86 | 543 | 100.836 |
| x_1 | 115 | 154.009 | 54 | 255 | 45.478 |
| x_2 | 115 | 153.045 | 63 | 274 | 46.939 |

TABLE 2. Descriptive Statistics

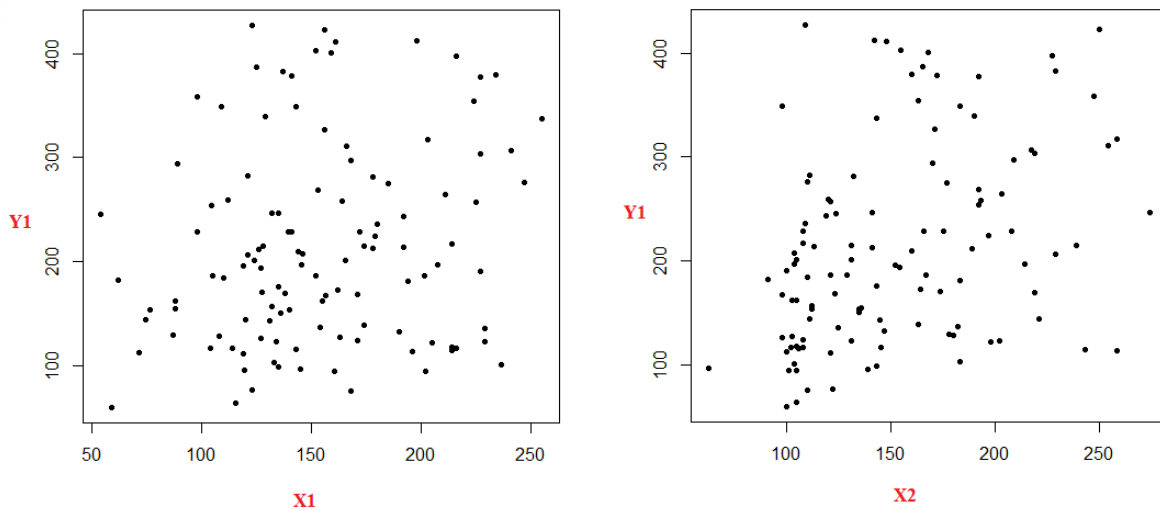
Based on Table 2, it can be seen that the lowest blood sugar level in diabetes mellitus patients before fasting (y_1) is 60 mg/dl, and the highest is 427 mg/dl. The average blood sugar level of diabetes mellitus patients before fasting was 213.435 mg/dl, with a standard deviation of 94.990 mg/dl. The blood sugar level of diabetes mellitus patients two hours after fasting (y_2) was the lowest was 86mg/dl, and the highest was 543mg/dl. The average blood sugar level of patients with

diabetes mellitus two hours after fasting was 233.956 mg/dl, with a standard deviation of 100.836 mg/dl.

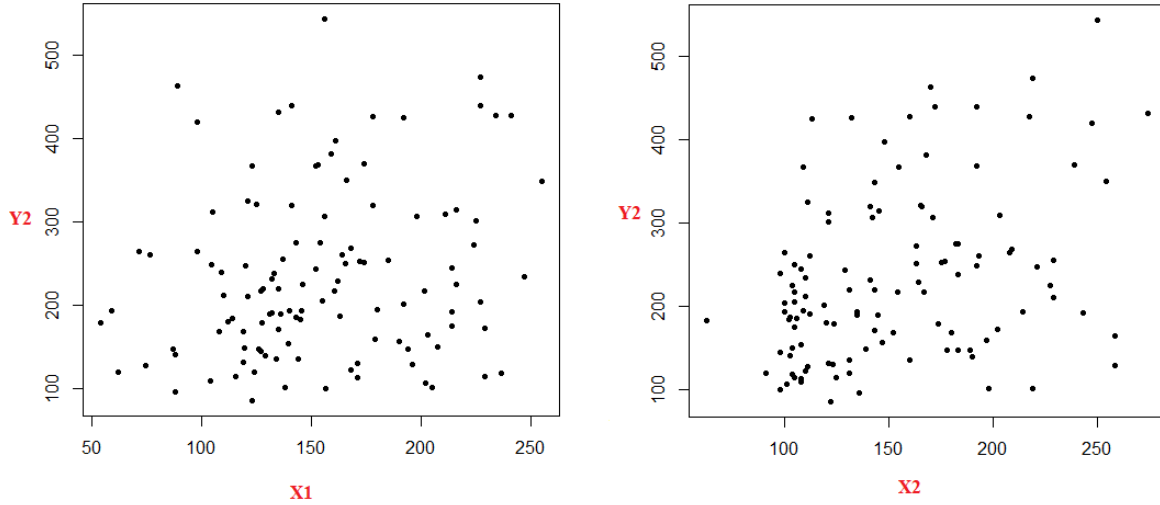
B. Scatter Plot

The initial step in the bi-response truncated spline regression is to determine the relationship pattern between each response variable and the predictor variable using a scatter plot. The goal is to see the data patterns that are formed. Based on the results of the scatter plot that has been carried out on the response variable, namely data on blood sugar levels in patients with diabetes mellitus before fasting (y_1) with predictor variables, namely data on total cholesterol levels in patients with diabetes mellitus (x_1) and data on triglyceride levels in patients with diabetes mellitus (x_2) can be seen in Figure 1.

FIGURE 1. Scatter Plot y_1 with x_1 and x_2



The scatter plot from data on blood sugar levels in patients with diabetes mellitus two hours after fasting (y_2) with predictor variables, namely data on total cholesterol levels in patients with diabetes mellitus (x_1) and data on triglyceride levels in patients with diabetes mellitus (x_2) can be seen in Figure 2.

FIGURE 2. Scatter Plot y_2 with x_1 and x_2 

Figures 1 and 2 show that the relationship pattern between the response variable and the predictor variable is unknown, so the appropriate solution is to use a nonparametric regression approach.

C. Bi-response Truncated Spline Nonparametric Regression

After identifying the relationship pattern between the response variable and the predictor variable using a scatter plot, it is known that all relationship patterns do not form a pattern, so the approach can be used in nonparametric regression. The following is a bi-response truncated spline regression model with two predictor variables and one-knot point written in Equations (13).

$$\begin{aligned}\widehat{y}_1 &= \widehat{\delta} + \widehat{\xi}_{11}x_{i1} + \widehat{\varphi}_{21}(x_{i1} - K_{11})_+ + \widehat{\xi}_{12}x_{i2} + \widehat{\varphi}_{22}(x_{i2} - K_{12})_+ \\ \widehat{y}_2 &= \widehat{\phi} + \widehat{\psi}_{11}x_{i1} + \widehat{\gamma}_{21}(x_{i1} - K_{11})_+ + \widehat{\psi}_{12}x_{i2} + \widehat{\gamma}_{22}(x_{i2} - K_{12})_+\end{aligned}\quad (13)$$

The bi-response truncated spline regression model with two predictor variables and two-knot points is presented in Equations (14).

$$\begin{aligned}\widehat{y}_1 &= \widehat{\delta} + \widehat{\xi}_{11}x_{i1} + \widehat{\varphi}_{21}(x_{i1} - K_{11})_+ + \widehat{\varphi}_{31}(x_{i1} - K_{21})_+ + \widehat{\xi}_{12}x_{i2} + \widehat{\varphi}_{22}(x_{i2} - K_{12})_+ \\ &\quad + \widehat{\varphi}_{32}(x_{i2} - K_{22})_+ \\ \widehat{y}_2 &= \widehat{\phi} + \widehat{\psi}_{11}x_{i1} + \widehat{\gamma}_{21}(x_{i1} - K_{11})_+ + \widehat{\gamma}_{31}(x_{i1} - K_{21})_+ + \widehat{\psi}_{12}x_{i2} + \widehat{\gamma}_{22}(x_{i2} - K_{12})_+ \\ &\quad + \widehat{\gamma}_{32}(x_{i2} - K_{22})_+\end{aligned}\quad (14)$$

The bi-response truncated spline regression model with two predictor variables and three-knot points is presented in Equations (15).

$$\begin{aligned}
\widehat{y}_1 &= \widehat{\delta} + \widehat{\xi}_{11}x_{i1} + \widehat{\phi}_{21}(x_{i1} - K_{11})_+ + \widehat{\phi}_{31}(x_{i1} - K_{21})_+ + \widehat{\phi}_{41}(x_{i1} - K_{31})_+ + \widehat{\xi}_{12}x_{i2} \\
&\quad + \widehat{\phi}_{22}(x_{i2} - K_{12})_+ + \widehat{\phi}_{32}(x_{i2} - K_{22})_+ + \widehat{\phi}_{42}(x_{i2} - K_{32})_+ \\
\widehat{y}_2 &= \widehat{\phi} + \widehat{\psi}_{11}x_{i1} + \widehat{\gamma}_{21}(x_{i1} - K_{11})_+ + \widehat{\gamma}_{31}(x_{i1} - K_{21})_+ + \widehat{\gamma}_{41}(x_{i1} - K_{31})_+ + \widehat{\psi}_{12}x_{i2} \\
&\quad + \widehat{\gamma}_{22}(x_{i2} - K_{12})_+ + \widehat{\gamma}_{32}(x_{i2} - K_{22})_+ + \widehat{\gamma}_{42}(x_{i2} - K_{32})_+
\end{aligned} \tag{15}$$

The bi-response truncated spline regression model, in general, that has been described previously is the first step taken to estimate parameters. Next, the model selection with the optimal knot point is carried out.

D. Selection of Optimal Knot Points

The best model selection from the bi-response truncated spline regression can be seen from the minimum GCV value and the maximum R^2 value. Table 3 shows the GCV value and R^2 value of each knot point.

| Knot Points | Minimum GCV | R^2 (%) |
|--------------------|--------------------|-----------------------------|
| 1 | 9.126 | 89.580 |
| 2 | 8.858 | 97.725 |
| 3 | 8.573 | 99.625 |

TABLE 3. Minimum GCV Value and R^2 Value of Each Knot Point

Selection of the best model is based on the model criteria that have the smallest GCV value and the maximum value of the coefficient of determination (R^2). This study's bi-response truncated spline regression model has the smallest GCV value, namely at three-knot points of 8.573. The model with three knots is the best bi-response truncated spline regression model. The coefficient of determination (R^2) for the three-point knot spline nonparametric bi-response regression model is 99.625, which states that 99.625% of the variation occurs in the total cholesterol level data variable in diabetes mellitus patients (x_1), the triglyceride level data variable in diabetes mellitus patients (x_2) and the remaining 0.370% is influenced by other factors that researchers do not yet know.

CROSS-VALIDATION IN DIABETES MELLITUS PATIENT'S BLOOD SUGAR LEVELS

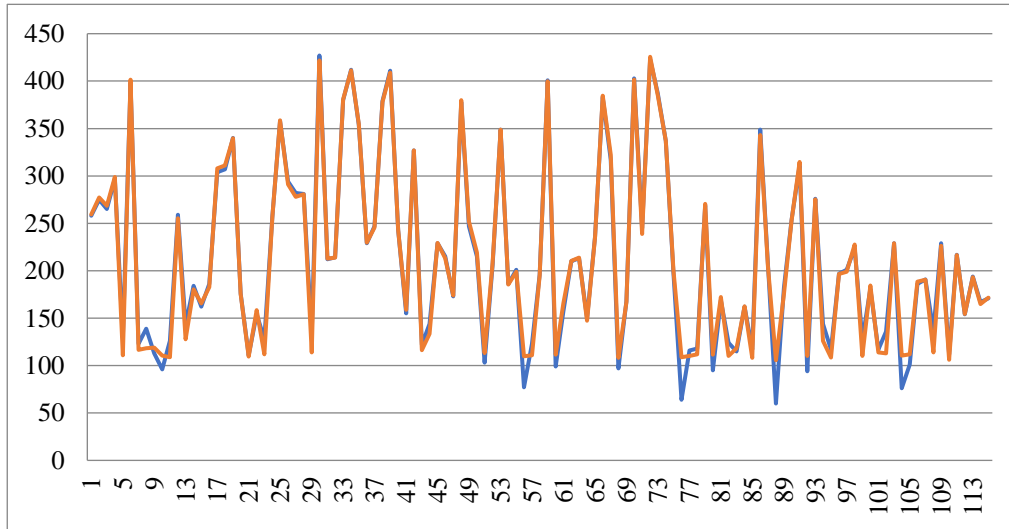
The formula of the bi-response truncated spline nonparametric regression model for three optimal knot points is in Equation (15). The estimation of model parameters with three-knot points has been described previously so that the best spline bi-response truncated spline regression model is written in Equation (16).

$$\begin{aligned}\widehat{y}_1 &= 98.002 + 0.049x_{i1} + 2.808(x_{i1} - 137.172)_+ + 3.595(x_{i1} - 157.965)_+ \\ &\quad + 1.773(x_{i1} - 164.896)_+ + 0.047x_{i2} + 0.067(x_{i2} - 150.310)_+ \\ &\quad + 0.260(x_{i2} - 172.138)_+ + 0.240(x_{i2} - 179.413)_+ \\ \widehat{y}_2 &= 117.441 + 0.007x_{i1} + 2.674(x_{i1} - 137.172)_+ + 1.749(x_{i1} - 151.034)_+ \\ &\quad + 0.075(x_{i1} - 199.551)_+ + 0.006x_{i2} + 0.047(x_{i2} - 150.310)_+ \\ &\quad + 0.081(x_{i2} - 164.862)_+ + 0.103(x_{i2} - 215.793)_+\end{aligned}\quad (16)$$

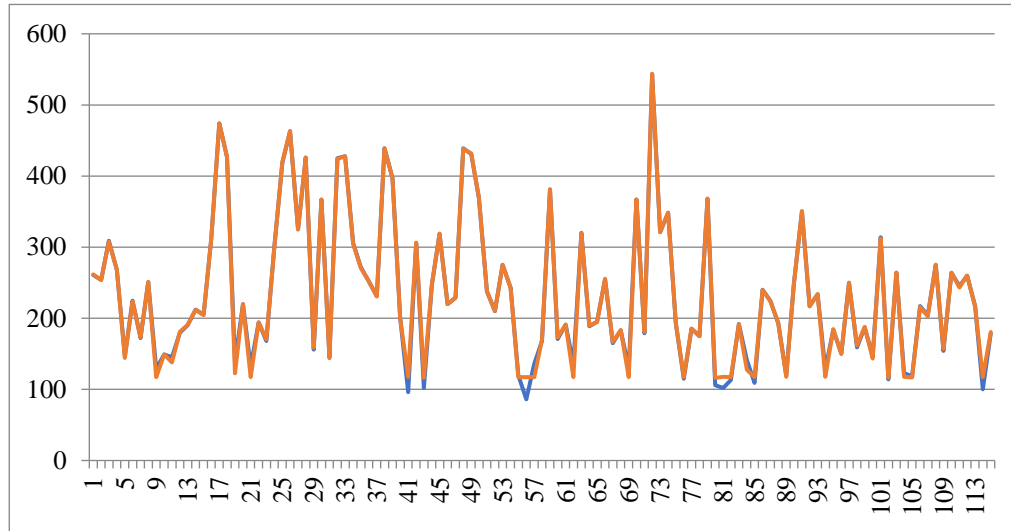
E. Model Fit Test

A comparison between the estimation results and data on blood sugar levels in diabetic patients is presented in the graph in Figure 3.

FIGURE 3. Actual vs. Prediction from Best Model



(a) first response



(b) second response

Figure 3 shows whether the resulting estimated or predictive data tend to approach the graph of the actual data or the original value data.

5. CONCLUSION

The best model from bi-response truncated spline nonparametric regression model with three-knot points. The resulting GCV value is 8.573, with an R^2 of 99.625%. The bi-response truncated spline regression model for the first response variable, namely data on blood glucose levels in patients with diabetes mellitus during fasting, is as follows.

$$\begin{aligned}\widehat{y}_1 = & 98.002 + 0.049x_{i1} + 2.808(x_{i1} - 137.172)_+ + 3.595(x_{i1} - 157.965)_+ \\ & + 1.773(x_{i1} - 164.896)_+ + 0.047x_{i2} + 0.067(x_{i2} - 150.310)_+ \\ & + 0.260(x_{i2} - 172.138)_+ + 0.240(x_{i2} - 179.413)_+\end{aligned}$$

The second response variable:

$$\begin{aligned}\widehat{y}_2 = & 117.441 + 0.007x_{i1} + 2.674(x_{i1} - 137.172)_+ + 1.749(x_{i1} - 151.034)_+ \\ & + 0.075(x_{i1} - 199.551)_+ + 0.006x_{i2} + 0.047(x_{i2} - 150.310)_+ \\ & + 0.081(x_{i2} - 164.862)_+ + 0.103(x_{i2} - 215.793)_+\end{aligned}$$

The application of nonparametric regression, especially the bi-response truncated spline model, has been successfully carried out in modeling blood sugar levels of people with diabetes mellitus

in inpatients at the Abdul Wahab Sjahranie Hospital (AWS) Samarinda, East Kalimantan.

CONFLICT OF INTERESTS

The author(s) declare that there is no conflict of interest.

ACKNOWLEDGEMENTS

The author(s) gratefully acknowledge the funding of Ministry of Education, Culture, Research and Technology (KEMENDIKBUD RISTEK) and Mulawarman University.

REFERENCES

- [1] H. Nurcahayani, I.N. Budiantara, I. Zain, The curve estimation of combined truncated spline and Fourier series estimators for multiresponse nonparametric regression, *Mathematics*. 9 (2021), 1141.
<https://doi.org/10.3390/math9101141>.
- [2] Sifriyani, I.N. Budiantara, S.H. Kartiko, et al. A new method of hypothesis test for truncated spline nonparametric regression influenced by spatial heterogeneity and application, *Abstr. Appl. Anal.* 2018 (2018), 9769150.
<https://doi.org/10.1155/2018/9769150>.
- [3] G. Wahba, Improper priors, spline smoothing and the problem of guarding against model errors in regression, *J. R. Stat. Soc.: Ser. B (Methodol.)* 40 (1978), 364–372. <https://doi.org/10.1111/j.2517-6161.1978.tb01050.x>.
- [4] G. Wahba, *Spline models for observational data*, SIAM, Philadelphia, 1990.
<https://doi.org/10.1137/1.9781611970128>.
- [5] I. Sriliana, I.N. Budiantara, V. Ratnasari, A truncated spline and local linear mixed estimator in nonparametric regression for longitudinal data and its application, *Symmetry*. 14 (2022) 2687.
<https://doi.org/10.3390/sym14122687>.
- [6] D.P. Rahmawati, I.N. Budiantara, D.D. Prastyo, et al. Mixture spline smoothing and kernel estimator in multi-response nonparametric regression, *IAENG Int. J. Appl. Math.* 51 (2021), 28.
- [7] Sifriyani, H. Ilmi, Z. Mar'ah, Application of nonparametric geographically weighted spline regression model for spatial mapping of open unemployment rate in kalimantan, *J. Phys.: Conf. Ser.* 2123 (2021), 012038.
<https://doi.org/10.1088/1742-6596/2123/1/012038>.

- [8] W. Hardle, *Applied nonparametric regression*, Humboldt University Berlin, 1994.
http://sfb649.wiwi.hu-berlin.de/fedc_homepage/xplore/ebooks/html/anr/index.html
- [9] G. Wahba, Y. Wang, Spline function, in: J.A. Matthews, *Encyclopedia of Environmental Change*, SAGE Publications, Thousand Oaks, California 91320, 2014. <https://doi.org/10.4135/9781446247501>.
- [10] V. Ratnasari, I.N. Budiantara, A.T.R. Dani, Nonparametric regression mixed estimators of truncated spline and gaussian kernel based on cross-validation (CV), generalized cross-validation (GCV), and unbiased risk (UBR) methods, *Int. J. Adv. Sci. Eng. Inform. Technol.* 11 (2021) 2400-2406. <https://doi.org/10.18517/ijaseit.11.6.14464>.
- [11] Sifriyani, S.H. Kartiko, I.N. Budiantara, et al. Development of nonparametric geographically weighted regression using truncated spline approach, *Songklanakarin J. Sci. Technol.* 40 (2018), 909-920.
<https://doi.org/10.14456/SJST-PSU.2018.98>.
- [12] R.L. Eubank, *Nonparametric regression and spline smoothing*, Marcel Dekker, New York, 1999.
- [13] R. Putra, M.G. Fadhlurrahman, Gunardi, Determination of the best knot and bandwidth in geographically weighted truncated spline nonparametric regression using generalized cross validation, *MethodsX.* 10 (2023), 01994. <https://doi.org/10.1016/j.mex.2022.101994>.
- [14] N.Y. Adrianingsih, I.N. Budiantara, J.D.T. Purnomo, Mixture model nonparametric regression and its application, *J. Phys.: Conf. Ser.* 1842 (2021), 012044. <https://doi.org/10.1088/1742-6596/1842/1/012044>.
- [15] S. Sifriyani, M. Rasjid, D. Rosadi, et al. Spatial-temporal epidemiology of COVID-19 using a geographically and temporally weighted regression model, *Symmetry.* 14 (2022), 742. <https://doi.org/10.3390/sym14040742>.
- [16] A.A.R. Fernandes, I.N. Budiantara, B.W. Otok, et al. Reproducing kernel Hilbert space for penalized regression multi-predictors: case in longitudinal data, *Int. J. Math. Anal.* 8 (2014), 1951-1961.
<https://doi.org/10.12988/ijma.2014.47212>.
- [17] A.T.R. Dani, V. Ratnasari, I.N. Budiantara, Optimal knots point and bandwidth selection in modeling mixed estimator nonparametric regression, *IOP Conf. Ser.: Mater. Sci. Eng.* 1115 (2021), 012020.
<https://doi.org/10.1088/1757-899x/1115/1/012020>.
- [18] R. Hidayat, I.N. Budiantara, B.W. Otok, et al. The regression curve estimation by using mixed smoothing spline and kernel (MsS-K) model, *Commun. Stat. - Theory Methods.* 50 (2020) 3942-3953.
<https://doi.org/10.1080/03610926.2019.1710201>.

CROSS-VALIDATION IN DIABETES MELLITUS PATIENT'S BLOOD SUGAR LEVELS

- [19] I.N. Budiantara, M. Ratna, I. Zain, et al. Modeling the percentage of poor people in Indonesia using spline nonparametric regression approach, *Int. J. Basic Appl. Sci.* 12 (2012), 119-124.
- [20] N.P.A.M. Mariati, I.N. Budiantara, V. Ratnasari, Combination estimation of smoothing spline and Fourier series in nonparametric regression, *J. Math.* 2020 (2020), 4712531. <https://doi.org/10.1155/2020/4712531>.
- [21] I.W. Sudiarsa, I.N. Budiantara, Suhartono, et al. Combined estimator Fourier series and spline truncated in multivariable nonparametric regression, *Appl. Math. Sci.* 9 (2015), 4997-5010.
<https://doi.org/10.12988/ams.2015.55394>.
- [22] M.A.D. Octavanny, I.N. Budiantara, H. Kuswanto, et al. A new mixed estimator in nonparametric regression for longitudinal data, *J. Math.* 2021 (2021), 3909401. <https://doi.org/10.1155/2021/3909401>.
- [23] R. Hidayat, I.N. Budiantara, B.W. Otok, et al. An extended model of penalized spline with the addition of kernel functions in nonparametric regression model, *Appl. Math. Inform. Sci.* 13 (2019), 453–460.
<https://doi.org/10.18576/amis/130318>.
- [24] D.P. Rahmawati, I.N. Budiantara, D.D. Prastyo, et al. Mixed spline smoothing and kernel estimator in biresponse nonparametric regression, *Int. J. Math. Math. Sci.* 2021 (2021), 6611084. <https://doi.org/10.1155/2021/6611084>.
- [25] Nidhomuddin, N. Chamidah, A. Kurniawan, Confidence interval of the parameter on multipredictor biresponse longitudinal data analysis using local linear estimator for modeling of case increase and case fatality rates COVID-19 in Indonesia: A theoretical discussion, *Commun. Math. Biol. Neurosci.* 2022 (2022), 23.
<https://doi.org/10.28919/cmbn/6900>.
- [26] A. Tohari, N. Chamidah, Fatmawati, et al. Modelling the number of HIV and aids cases in East Java using biresponse multipredictor negative binomial regression based on local linear estimator, *Commun. Math. Biol. Neurosci.* 2021 (2021), 73. <https://doi.org/10.28919/cmbn/5652>.
- [27] S. Suparti, B. Warsito, R. Santoso, et al. Biresponses kernel nonparametric regression: inflation and economic growth, *Int. J. Criminol. Sociol.* 10 (2020), 465-471. <https://doi.org/10.6000/1929-4409.2021.10.54>.
- [28] H. Husain, I.N. Budiantara, I. Zain, Mixed estimator of spline truncated, Fourier series, and kernel in biresponse semiparametric regression model, *IOP Conf. Ser.: Earth Environ. Sci.* 880 (2021), 012046.
<https://doi.org/10.1088/1755-1315/880/1/012046>.
- [29] S. Mansour, A. Al Kindi, A. Al-Said, et al. Sociodemographic determinants of COVID-19 incidence rates in

- Oman: Geospatial modelling using multiscale geographically weighted regression (MGWR), *Sustain. Cities Soc.* 65 (2021), 102627. <https://doi.org/10.1016/j.scs.2020.102627>.
- [30] A.D. Association, Diagnosis and classification of diabetes mellitus, *Diabetes Care.* 33 (2010), S62–S69. <https://doi.org/10.2337/dc10-s062>.
- [31] Kementerian Kesehatan RI., InfoDATIN, Pusat Data dan Informasi Kementerian Kesehatan RI. Kementerian Kesehatan RI., Jakarta, pp. 1–10, 2020.
- [32] F. R. Sihotang, S. Sifriyani, S. Prangga, Aplikasi regresi nonparametrik spline birespons pada data kualitas air di das mahakam, *Media Bina Ilmiah*, 16 (2021), 6515-6526.
- [33] A. B. Padatuan, S. Sifriyani, S. Prangga, Pemodelan angka harapan hidup dan angka kematian bayi di kalimantan dengan regresi nonparametrik spline birespons, *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, 15 (2021), 283-296.
- [34] T. Purnaraga, S. Sifriyani, S. Prangga, Regresi nonparametrik spline pada data laju pertumbuhan ekonomi di kalimantan, *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, 14 (2020), 343-356.