ESTIMATION OF A BI-RESPONSE TRUNCATED SPLINE NONPARAMETRIC REGRESSION MODEL ON LIFE EXPECTANCY AND PREVALENCE OF UNDERWEIGHT CHILDREN IN INDONESIA

Anggi Putri Anisar[1*], Sifriyani[2], Andrea Tri Rian Dani[3]

[1,2,3]*Statistics Study Program, Faculty of Mathematics and Natural Sciences, Mulawarman University
St. Barong Tongkok No.04, Samarinda, 75123, Indonesia*

Corresponding author's e-mail: *anggianisar07@gmail.com

ABSTRACT

Researchers use the nonparametric regression method because it provides excellent flexibility in the modeling process. Nonparametric regression procedures can be used if the relationship pattern between the predictor and response variables is unknown. The truncated spline method is one of the most frequently used nonparametric regression methods. A truncated spline is a polynomial slice with continuous segmented properties, and the resulting curve is relatively smooth. The advantage of truncated splines is that they can be used on data that experience behavior changes at specific intervals. The nonparametric spline truncated bi-response regression approach is used when one or more predictor variables affect the two response variables with the assumption that there is a correlation between the response variables. This study aimed to obtain the best spline truncated bi-response nonparametric regression model on life expectancy data and the prevalence of underweight children in Indonesia in 2021. The data used comes from the Central Bureau of Statistics and the Indonesian Ministry of Health. The optimal knot point selection method uses the Generalized Cross Validation (GCV) method. The results showed that the best model formed was obtained using three-knot points based on a minimum GCV value of 22.77 and a coefficient of determination of 99.58%.

1. INTRODUCTION

Regression analysis is one of the statistical methods used to determine the effect of the predictor variable on the response variable. The main goal of a regression analysis is to find an estimate of the regression curve [1]. There are three approaches commonly used to estimate the shape of the regression curve, namely the parametric approach, the nonparametric approach, and the semiparametric approach. The parametric regression approach is used when it is assumed that the shape of the regression curve of data is known to be in condition [2]. The nonparametric regression approach is used when considering the form of the regression curve of an unknown data shape. A semiparametric method is used if the pattern of the regression curve is partly known and partly the way is strange [3].

Truncated splines have helpful statistical properties and should be considered a method for analyzing regression relationships [4]. The truncated spline approach is a regression model with an exceptional and reasonable interpretation of visual statistics because there are several advantages in estimating the regression curve [5]. A truncated spline is a function built on polynomial and truncated components, a slice of a polynomial with vertices, which can deal with changing data behavior patterns. The truncated spline approach is used to overcome the problem of data analysis modeling; namely, the relationship between the response variable and the predictor variable does not follow a specific pattern. There is a change in the way at certain subintervals. Certain sub-intervals can be used to overcome or reduce data patterns that experience a sharp increase with the help of knot points. The knot point is a fusion that shows the pattern of behavior of the truncated spline function at different intervals, with the presence of nodes, the resultingmodel will follow a relationship pattern the suits the behavior of the data [6] [7].

In real cases, we often face the problem that two or more dependent variables are observed in several independent variable values, and there is a correlation between the responses. As one of the methods to overcome this problem, the multiresponse nonparametric regression model is to model the function that represents the association of the variables used [8]. In this study, researchers used nonparametric bi-response spline truncated regression analysis. Nonparametric bi-response truncated spline regression is a regression model with two response variables. Regression analysis, especially in the bi-response truncated spline nonparametric regression, requires an appropriate parameter estimation method. Several techniques can be used to estimate parameters, including the Ordinary Least Square (OLS) method [9]. The principle of the least squares method is to minimize the sum of the squared errors. One of the methods that can be used in nonparametric regression using two response variables with the same principle as the OLS method is the Weighted Least Square (WLS) method with the help of a weighting matrix, namely the variance-covariance matrix [10] . Using the covariance variant matrix can produce accurate estimation results [11].

Health is an essential indicator in improving the quality of life, which supports human development [7][12]. In Indonesia, health problems are still a severe problem and a significant concern for the government. In this case, the government should consider combining policy interventions to reduce economic pressures, guarantee access to health care, and facilitate optimal health [13]. A high degree of public health can indicate the success of health and socio-economic development programs, which can indirectly increase life expectancy [3]. Life Expectancy is the average number of years of life that is still temporarily lived by someone who has reached a certain age in a specific year. There is an imbalance in the level of life expectancy. In general, rural areas have a higher life expectancy when compared to urban areas, which have a lower life expectancy[14]. Based on the Central Bureau of Statistics, life expectancy in Indonesia in 2019 was 71.34 and continued to increase until 2021 to 71.57. A critical factor in achieving optimal health status is maintaining a good nutritional state [10]. In developing countries, children under five (0-5 years) are most vulnerable to malnutrition. This nutritional status problem causes children to suffer from various infections and has low dietary status. One part of the prevalence of underweight toddlers that concerns researchers is malnourished. Being Underweight is a baby's failure to achieve ideal body weight, affecting height growth [15]. Several countries are concerned about improving and supporting health progress, so health-related policies are carried out optimally, such as free health checks, socializing the importance of health, and providing adequate health facilities [16][17].

2. RESEARCH METHODS

2.1 Regression Analysis

Regression analysis is a method that studies a relationship between response variables and predictor variables. There are three models of approaches based on the regression curve, namely the nonparametric regression approach, the semiparametric regression approach and the parametric regression approach [1]. The model for can be written in **Equation (1)**:

$$y_i = f(x_i) + \varepsilon_i ; i = 1, 2, ..., n \tag{1}$$

with:

y_i : response variable for the third observation to-i

x_i : predictor variable to-i

$f(x_i)$: regression curve

ε_i : *error* to-i

2.2 Multivariable Truncated Spline Bi-response Nonparametric Regression

Nonparametric spline truncated regression analysis if there is one response variable and one predictor variable, then it is called univariable spline truncated nonparametric regression. Nonparametric spline truncated regression analysis if there is one response variable with more than one predictor variable, then the regression is called multivariable spline truncated nonparametric regression [18]. Truncated spline nonparametric bi-response regression is defined as a nonparametric regression method that has more than one response variable and between these variables there is a strong correlation or relationship, both logically and mathematically [7]. The model for nonparametric spline truncated bi-response regression can be written in **Equation (2)**:

$$y_{1i} = \sum_{j=1}^{l} f(x_{ij}) + \varepsilon_{1i}$$
$$y_{2i} = \sum_{j=1}^{l} g(x_{ij}) + \varepsilon_{2i} \tag{2}$$

where is function f and g a regression curve of unknown shape and is approximated by a multivariable bi-response spline function with **Equation (3)**:

$$f(x_{ij}) = \alpha + \sum_{h=1}^{m} \theta_{hj} x_{ij}^h + \sum_{k=1}^{r} \tau_{m+k;j} (x_{ij} - K_{kj})_+^m$$
$$g(x_{ij}) = \lambda + \sum_{h=1}^{m} \psi_{hj} x_{ij}^h + \sum_{k=1}^{r} \varphi_{m+k;j} (x_{ij} - K_{kj})_+^m \tag{3}$$

dengan:

$f(x_{ij})$: first response variable regression function

$g(x_{ij})$: second response variable regression function

α : the first response variable constant

λ : second response variable constant

x_{ij} : predictor variable to observation to-i , $i = 1, 2, ..., n$ and j is the number of predictor variables
       $j = 1, 2, ..., l$

θ_{hj} : the coefficient of the regression parameter of the polynomial function, the first response variable
       where $h = 1, 2, ..., m$

$\tau_{m+k;j}$: the coefficient of the truncated spline regression parameter of the first response variable, where k
       is the knot point $k = 1, 2, ..., r$

ψ_{hj} : the coefficient of the regression parameter of the polynomial function, the second response
       variable where $h = 1, 2, ..., m$

$\varphi_{m+k;j}$: the coefficient of the second response variable spline truncated regression parameter, where k is

the knot point $k = 1, 2, ..., r$

K_{kj} : optimum knot point value to- k

The regression model in **Equation (2)** and **Equation (3)** can be written in the following form of matrix notation:

$$\tilde{y} = X(k)\tilde{\beta} + \tilde{\varepsilon} \tag{4}$$

which can be explained:

$$\tilde{y} = \begin{pmatrix} \tilde{y}_1 \\ --- \\ \tilde{y}_2 \end{pmatrix}_{2n \times 1} = \begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n} \\ --- \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2n} \end{pmatrix} \text{ dan } X(k) = \begin{pmatrix} C & \cdots & 0 \\ \cdots & \ddots & \cdots \\ 0 & \cdots & D \end{pmatrix} \tag{5}$$

with:

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

$$D = \begin{bmatrix} 1 & x_{11}^1 & \cdots & x_{11}^m & (x_{11} - \lambda_1^1)_+^m & \cdots & (x_{11} - \lambda_r^1)_+^m & \cdots & x_{l1}^1 & \cdots & x_{l1}^m & (x_{l1} - \lambda_1^1)_+^m & \cdots & (x_{l1} - \lambda_r^1)_+^m \\ 1 & x_{12}^1 & \cdots & x_{12}^m & (x_{11} - \lambda_1^1)_+^m & \cdots & (x_{12} - \lambda_r^1)_+^m & \cdots & x_{l2}^1 & \cdots & x_{l2}^m & (x_{l2} - \lambda_1^1)_+^m & \cdots & (x_{l2} - \lambda_r^1)_+^m \\ \vdots & \vdots & \ddots & & \vdots & & \ddots & & \ddots & \vdots & \ddots & & \vdots & & \ddots \\ 1 & x_{1n}^1 & \cdots & x_{1n}^m & (x_{11} - \lambda_1^1)_+^m & \cdots & (x_{1n} - \lambda_r^1)_+^m & \cdots & x_{ln}^1 & \cdots & x_{ln}^m & (x_{lj} - \lambda_1^1)_+^m & \cdots & (x_{1n} - \lambda_r^1)_+^m \end{bmatrix}$$

for value $\tilde{\beta}$ and $\tilde{\varepsilon}$ is a parameter vector and an error vector which has the following elements[12]:

$$\tilde{\beta} = \begin{pmatrix} \tilde{\beta}_1 \\ --- \\ \tilde{\beta}_2 \end{pmatrix} = \begin{pmatrix} \alpha \\ \theta_{11} \\ \theta_{12} \\ \vdots \\ \theta_{ml} \\ \tau_{(1+1)l} \\ \tau_{(1+2)l} \\ \vdots \\ \tau_{(m+r)l} \\ --- \\ \lambda \\ \psi_{11} \\ \psi_{12} \\ \vdots \\ \psi_{ml} \\ \varphi_{(1+1)l} \\ \varphi_{(1+2)l} \\ \vdots \\ \varphi_{(m+r)l} \end{pmatrix} , \tilde{\varepsilon} = \begin{pmatrix} \tilde{\varepsilon}_1 \\ --- \\ \tilde{\varepsilon}_2 \end{pmatrix} = \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{1n} \\ --- \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \vdots \\ \varepsilon_{2n} \end{pmatrix} \tag{6}$$

2.3 Estimation of Bi-response Spline Truncated Nonparametric Regression Models

The parameter estimation method in this study uses the Weighted Least Square (WLS) method, because this method is able to overcome correlations in the same observation subject [2]. In the bi-response spline nonparametric regression using the Weighted Least Square (WLS) optimization method, it is assumed that there is a correlation between errors in the first response (ε_{1i}) with an error in the second response (ε_{2i}). Value $\text{cov}(y_1, y_2)$ assumed to be homogeneous but the error variance value for each response is not the same, so the variance-covariance (W) matrix is used. Determination of the weighting matrix is by calculating the value of the covariance variance of the first response variable and the second response variable [19] [9].

$$W(y_1, y_2) = \begin{pmatrix} \text{var}(y_1) & \text{cov}(y_1, y_2) \\ \text{cov}(y_2, y_1) & \text{var}(y_2) \end{pmatrix} \tag{7}$$

where to find the values of the variances as follows:

$$\text{cov}(y_1, y_2) = \frac{1}{n}\sum_{i=1}^{n}(y_{1i} - \overline{y}_1)(y_{2i} - \overline{y}_2) \tag{8}$$

$$\text{cov}(y_2, y_1) = \frac{1}{n}\sum_{i=1}^{n}(y_{2i} - \overline{y}_2)(y_{1i} - \overline{y}_1) \tag{9}$$

$$\text{var}(y_1) = \frac{1}{n}\sum_{i=1}^{n}(y_{1i} - \overline{y}_1)^2 \tag{10}$$

$$\text{var}(y_2) = \frac{1}{n}\sum_{i=1}^{n}(y_{2i} - \overline{y}_2)^2 \tag{11}$$

The **W** weighting matrix can be written in **Equation (12)**:

$$\mathbf{W} = \left|\begin{array}{cccc|cccc} \sigma_1^2 & 0 & \cdots & 0 & \sigma_{11} & 0 & \cdots & 0 \\ 0 & \sigma_1^2 & & \vdots & 0 & \sigma_{22} & & \vdots \\ \vdots & & \ddots & 0 & \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_1^2 & 0 & \cdots & 0 & \sigma_{nn} \\ - & - & - & - & - & - & - & - \\ \sigma_{11} & 0 & \cdots & 0 & \sigma_2^2 & 0 & \cdots & 0 \\ 0 & \sigma_{22} & & \vdots & 0 & \sigma_2^2 & & \vdots \\ \vdots & & \ddots & 0 & \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_{nn} & 0 & \cdots & 0 & \sigma_2^2 \end{array}\right|_{2n \times 2n} \tag{12}$$

then to obtain the estimator, the completion of parameter optimization is carried out with WLS, for the sum of squared errors [3] is given by **Equation (13)**:

$$\left\{\tilde{\varepsilon}^T W \tilde{\varepsilon}\right\} = \left\{\left(\tilde{y} - X(k)\tilde{\beta}\right)^T W \left(\tilde{y} - X(k)\tilde{\beta}\right)\right\} \tag{13}$$

$$= \tilde{y}^T W \tilde{y} - \tilde{y}^T W X(k)\tilde{\beta} - X(k)^T \tilde{\beta}^T W \tilde{y} + X(k)^T \tilde{\beta}^T W X(k)\tilde{y} \tag{14}$$

$$= \tilde{y}^T W \tilde{y} - 2\tilde{\beta}^T X(k)^T W \tilde{y} + \tilde{\beta}^T X(k)^T W X(k)\tilde{\beta}$$

then **Equation (14)** obtained is derived from $\tilde{\beta}$ write the result in **Equation (15)**:

$$\frac{\partial\varphi(\tilde{\beta})}{\partial(\tilde{\beta})} = \frac{\partial(\tilde{y}^T W \tilde{y} - 2\tilde{\beta}^T X(k)^T W \tilde{y} + \tilde{\beta}^T X(k)^T W X(k)\tilde{\beta})}{\partial(\tilde{\beta})}$$

$$= -2X(k)^T W \tilde{y} + 2X(k)^T W X(k)\hat{\tilde{\beta}} \tag{15}$$

$$-2X(k)^T W \tilde{y} + 2X(k)^T W X(k)\hat{\tilde{\beta}} = 0$$

$$X(k)^T W X(k)\hat{\tilde{\beta}} = X(k)^T W \tilde{y}$$

$$\hat{\tilde{\beta}} = (X(k)^T W X(k))^{-1} X(k)^T W \tilde{y} \tag{16}$$

so that the estimated form of the spline model in bi-respon nonparametric regression [3] becomes:

$$\hat{f}(x) = X(k)\hat{\tilde{\beta}}$$
$$= X(k)(X(k)^T W X(k))^{-1} X(k)^T W \tilde{y} \tag{17}$$

2.4 Relationship Between Response Variables

The correlation coefficient is a value that measures the closeness of the relationship between two variables [19]. The correlation value can be used Pearson Product Moment correlation coefficient which is written in **Equation (18)**:

$$r = \frac{n \sum_{i=1}^{n} y_{1i} y_{2i} - \left(\sum_{i=1}^{n} y_{1i} \right) \left(\sum_{i=1}^{n} y_{2i} \right)}{\sqrt{\left[n \sum_{i=1}^{n} y_{1i}^2 - \left(\sum_{i=1}^{n} y_{1i} \right)^2 \right] \left[n \sum_{i=1}^{n} y_{2i}^2 - \left(\sum_{i=1}^{n} y_{2i} \right)^2 \right]}} \tag{18}$$

Based on calculations with the Pearson correlation formula, a correlation coefficient value will be obtained, which based on this value can be seen the closeness of the relationship between the response variables that will be used [20]. To test the hypothesis that there is a significant relationship between the first response variable and the second response variable, the correlation significance test is used with the formula:

$$H_0 : r_{y_1 y_2} = 0$$

$$H_1 : r_{y_1 y_2} \neq 0$$

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \tag{19}$$

The area of rejection if value $|t| < t_{\frac{\alpha}{2}(n-2)}$ or $p\text{-}value < \alpha$, so H_0 rejected, which means that it can be concluded that the two variables are not mutually exclusive (correlated).

2.5 Optimal Knot Point Selection

Theoretically, the GCV method has the asymptotic optimal property shown which is not shared by other methods. Another advantage of the GCV method is that it is invariant to transformation and in calculating the population variance does not need to be known [6]. Selection of optimal knot points using the GCV method selects the function that has the minimum GCV value. The GCV method is generally defined:

$$GCV = \frac{MSE(k)}{\left[2n^{-1} trace(\mathbf{I} - \mathbf{A}(k)) \right]^2} \tag{20}$$

with:

I : sized identity matrix $n \times n$
n : number of observations

$$MSE(k) = 2n^{-1} \left(\tilde{y} - X(k)\tilde{\beta} \right)^T \left(\tilde{y} - X(k)\tilde{\beta} \right) \tag{21}$$

$$A(k) = X(k) \left(X(k)^T W X(k) \right)^{-1} X(k)^T W \tag{22}$$

2.6 **Best Model** Selection **Criteria**

Determination coefficient value (R^2) is one of the quantities used to measure the feasibility of the regression model and a measure of the accuracy or accuracy of the regression model which shows the magnitude of the contribution of the predictor variable to changes in the response variable [11]. A good

model is a model that has value (R^2) high. Value (R^2) obtained is obtained from **Equation (23)**:

$$R^2 = \left(1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \right) \times 100\% \tag{23}$$

2.7 Research Variables

The method used in this study is a bi-response spline nonparametric regression with optimal knot point selection using the Generalized Cross Validation (GCV) method. The aim is to obtain the best estimate of the bi-response nonparametric regression model using a truncated spline approach to life expectancy data and the prevalence of underweight children in Indonesia. The data used is life expectancy data (y_1), majority of underweight toddlers (y_2), Percentage of poor people (x_1), Percentage of babies who are breastfed aged 0-5 months (x_2), Percentage of children aged 12-23 months who received complete primary immunization (x_3), and percentage rate of household members who have health insurance (x_4) in 2021 obtained from the website of the Central Statistics Agency and the Ministry of Health. The research data used covers all provinces in Indonesia, namely 34 provinces spread across Indonesia as of 2021

2.8 Stage of Analysis

The steps of analysis with nonparametric spline truncated bi-response regression are as follows:

1. Perform correlation testing between response variables.

2. Knowing the pattern of data between predictor variables on the response variables

3. Modeling the expected life expectancy and prevalence of underweight children using a linear truncated bi-response spline nonparametric regression approach for 1 knot, 2 knots, and 3 knots.

4. Selecting the optimal knot point using the Generalized Cross Validation (GCV) method and the coefficient of determination for the best model

5. Modeling the expected age and prevalence of underweight children with the best model formed

3. RESULT AND DISCUSSION

3.1 Correlation Testing

One of the requirements in conducting nonparametric spline truncated bi-response regression modeling is that the first response variable and the second response variable must have a strong relationship in theory and through statistical testing. The relationship between life expectancy and the prevalence of underweight children can be significantly shown by correlation testing. The correlation test hypothesis is written as follows:

Table 1. Correlation Test Between Response Variables

P-value	0.00
Correlation Value	-0.64

Based on **Table 1**, it can be seen that the p-value = 0.00 so it is concluded that H_0 is rejected or it is known that there is a significant relationship between life expectancy and the prevalence of underweight children, so that nonparametric spline truncated bi-response regression modeling is carried out. The correlation value of -0.64 means that the relationship between life expectancy and the prevalence of underweight children has a relatively strong level of cohesiveness, which means that the variable life expectancy and the majority of malnourished children have a close relationship or give each other a link.
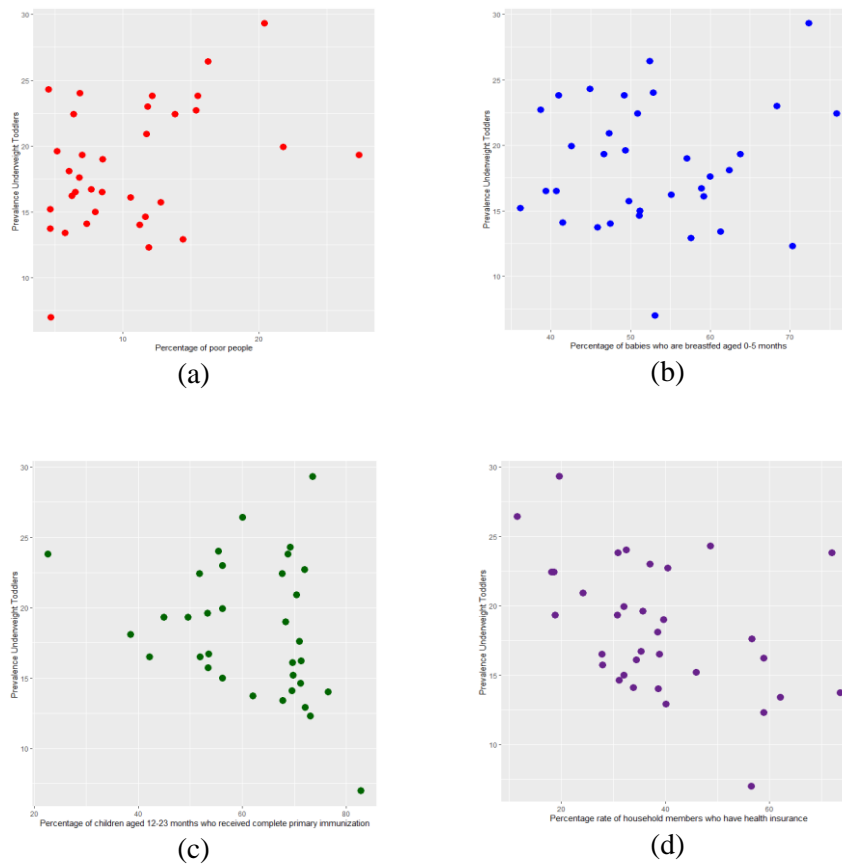
3.2 Data Patterns Between Response Variables and Predictor Variables

a. The pattern of Data Between Life Expectancy with Predictor Variables
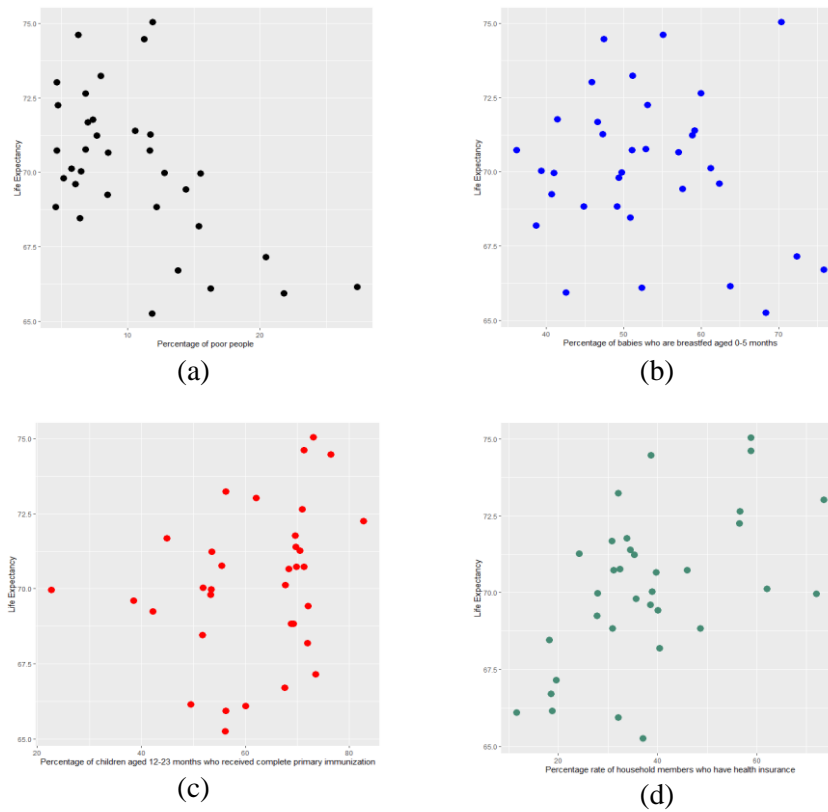
Figure 1. **Intermediate data pattern (a)** (y_1) **and** (x_1) **, (b)** (y_1) **and** (x_2) **(c)** (y_1) **and** (x_3) **, and (d)** (y_1) **and** (x_4)

b. The pattern of Data Between Prevalence Underweight Toddlers with Predictor Variables

Figure 2. **Intermediate data pattern (a)** (y_2) **and** (x_1) **, (b)** (y_2) **and** (x_2) **, (c)** (y_2) **and** (x_3) **, and (d)** (y_2) **and** (x_4)

3.3 Optimal Knot Point Selection

Selection of the optimal knot point with the GCV method by having a minimum GCV value. Minimum GCv values at 1-knot points, 2-knot points, and 3-knot points can be seen in Table 2 below. Intermediate data pattern.

Table 2. The Smallest GCV Value and R^2 at Optimal Knot Points

Point Knots	GCV Value	R^2 Value	MAPE Value
1	27.03	98.93 %	8.89 %
2	25.50	99.29 %	6.70 %
3	22.77	99.58 %	5.63 %

Based on **Table 2**. it is found that the selection of the best model is the one with the minimum GCV value obtained is 2.77 namely in the model with 3-knot points with a determination coefficient value of 99.58% and a MAPE value of 5.63% which is the minimum. the optimal knot point can be used as the best model is a nonparametric bi-response spline truncated regression model with 3-knot points. The coefficient of determination at the three-knot points in **Table 2** means that as much as 99.58% of the variation that occurs in Life Expectancy and prevalence of underweight toddlers is influenced by the percentage of poor people. the rate of babies who are breastfed aged 0-5 months. the percentage of children aged 12-23 month of receiving complete primary immunization. the percentage of household members having health care insurance. At the same time. for the remaining 0.42%. the variation that occurs in life expectancy and prevalence of under-fives is influenced by other factors that have not been studied. The MAPE value of 5.63%. which is less than 10%. indicates that the prediction results from the regression model are very accurate and reasonable for predicting future discounts.

3.4 Best Model of Bi-response Spline Truncated Nonparametric Regression

The location of the optimal knot point in the nonparametric spline truncated bi-response regression model of 3-knot issues based on the minimum GCV value can be seen in **Table 3**.

Table 3. Optimal Knot Locations at 3-Knot Points

Predictor Variables	Optimal Knot Point	GCV Value
x_1	12.43 13.22 15.58	
x_2	49.89 51.26 55.37	22.77
x_3	43.45 45.52 51.74	
x_4	32.98 35.12 41.53	

Based on the results obtained in **Table 3**. it can be concluded that the optimal knot point at 3-knot points for the predictor variable used with parameter estimation obtained based on the optimal knot point for the bi-response spline truncated nonparametric regression model is accepted in **Table 4**.

Table 4. Parameter Estimation at 3 Knot Points

Response Variable	Predictor Variables	Parameter Estimation	
	Constant	$\hat{\alpha}$	67.83
		$\hat{\theta}_1$	0.15
	x_1	$\hat{\tau}_{11}$	36.46
y_1		$\hat{\tau}_{12}$	-0.71
		$\hat{\tau}_{13}$	0.57
	x_2	$\hat{\theta}_2$	-0.04
		$\hat{\tau}_{21}$	-0.83

Response Variable	Predictor Variables	Parameter Estimation	
		$\hat{\tau}_{22}$	-0.11
		$\hat{\tau}_{23}$	-0.43
		$\hat{\theta}_3$	0.01
	x_3	$\hat{\tau}_{31}$	-1.26
		$\hat{\tau}_{32}$	1.60
		$\hat{\tau}_{33}$	1.18
		$\hat{\theta}_3$	0.01
	x_3	$\hat{\tau}_{31}$	-1.26
		$\hat{\tau}_{32}$	1.60
		$\hat{\tau}_{33}$	1.18
		$\hat{\theta}_4$	0.06
	x_4	$\hat{\tau}_{41}$	-1.37
		$\hat{\tau}_{42}$	0.12
		$\hat{\tau}_{43}$	1.48
	Constant	$\hat{\lambda}$	-2.11
		$\hat{\psi}_1$	0.71
	x_1	$\hat{\varphi}_{11}$	-20.41
		$\hat{\varphi}_{12}$	32.73
		$\hat{\varphi}_{23}$	-13.19
		$\hat{\psi}_2$	0.44
	x_2	$\hat{\varphi}_{21}$	-9.12
		$\hat{\varphi}_{22}$	10.88
y_2		$\hat{\varphi}_{23}$	-1.95
		$\hat{\psi}_3$	-0.73
	x_3	$\hat{\varphi}_{41}$	1.79
		$\hat{\varphi}_{42}$	0.25
		$\hat{\varphi}_{33}$	-1.41
		$\hat{\psi}_4$	0.36
	x_4	$\hat{\varphi}_{41}$	5.09
		$\hat{\varphi}_{42}$	6.63
		$\hat{\varphi}_{43}$	-1.86

Based on the selection of optimal knot points. a truncated bi-response spline nonparametric regression model with 3-knot points is the best model that can be used to model life expectancy and the prevalence of underweight children in Indonesia. The best model for the life expectancy variable with nonparametric bi-response spline truncated regression with three-knot points can be written in **Equation (1)**:

$$
\begin{aligned}
\hat{y}_1 = {} & 67.83 + 0.15x_{i1} + 36.46(x_{i1}-12.43)_+ - 0.71(x_{i1}-13.22)_+ + 0.57(x_{i1}-15.58)_+ - 0.04x_{i2} \\
& -0.83(x_{i2}-49.89)_+ - 0{,}11(x_{i2}-51.26)_+ - 0.43(x_{i2}-55.37)_+ + 0.01x_{i3} - \\
& 1.26(x_{i3}-43.45)_+ + 1.60(x_{i3}-45.52)_+ + 1.18(x_{i3}-51.74)_+ + 0.06x_{i4} \\
& -1.37(x_{i4}-32.98)_+ + 0.12(x_{i4}-35.12)_+ + 1.48(x_{i4}-41.53)_+
\end{aligned}
\tag{1}
$$

The best model for the variable prevalence of underweight toddlers with nonparametric bi-response spline truncated regression with 3-knot points can be written in **Equation (2)**:

$$\hat{y}_2 = -2.11 + 0.714x_{i1} - 20.41(x_{i1} - 12.43)_+ + 32.73(x_{i1} - 13.22)_+ - 13.19(x_{i1} - 15.58)_+$$
$$+ 0.44x_{i2} - 9.12(x_{i2} - 49.89)_+ + 10.88(x_{i2} - 51.26)_+ - 1.95(x_{i2} - 55.37)_+ - 0.73x_{i3} -$$
$$1.79(x_{i3} - 43.45)_+ + 0.25(x_{i3} - 45.52)_+ - 1.41(x_{i3} - 51.74)_+ + 0.36x_{i4}$$
$$-5.09(x_{i4} - 32.98)_+ + 6.63(x_{i4} - 35.12)_+ - 1.86(x_{i4} - 41.53)_+$$

This model can be used to determine the prediction results on the first response variable and second response variable based on the best model in this case life expectancy and prevalence of underweight children using a nonparametric spline regression approach. This can be used as a consideration in making decisions in the future.

4. CONCLUSIONS

The conclusion that can be drawn from the research results is that the best nonparametric truncated bi-response spline regression model is obtained with three-knot points. The resulting GCV value is 22.77 and the coefficient of determination is 99.58%. The bi-response truncated spline nonparametric regression model is as follows:

1. Model for the life expectancy variable

$$\hat{y}_1 = 67.83 + 0.15x_{i1} + 36.46(x_{i1} - 12.43)_+ - 0.71(x_{i1} - 13.22)_+ + 0.57(x_{i1} - 15.58)_+ - 0.04x_{i2}$$
$$-0.83(x_{i2} - 49.89)_+ - 0,11(x_{i2} - 51.26)_+ - 0.43(x_{i2} - 55.37)_+ + 0.01x_{i3} -$$
$$1.26(x_{i3} - 43.45)_+ + 1.60(x_{i3} - 45.52)_+ + 1.18(x_{i3} - 51.74)_+ + 0.06x_{i4}$$
$$-1.37(x_{i4} - 32.98)_+ + 0.12(x_{i4} - 35.12)_+ + 1.48(x_{i4} - 41.53)_+$$

2. Model for the prevalence of underweight toddlers variable

$$\hat{y}_2 = -2.11 + 0.714x_{i1} - 20.41(x_{i1} - 12.43)_+ + 32.73(x_{i1} - 13.22)_+ - 13.19(x_{i1} - 15.58)_+$$
$$+ 0.44x_{i2} - 9.12(x_{i2} - 49.89)_+ + 10.88(x_{i2} - 51.26)_+ - 1.95(x_{i2} - 55.37)_+ - 0.73x_{i3} -$$
$$1.79(x_{i3} - 43.45)_+ + 0.25(x_{i3} - 45.52)_+ - 1.41(x_{i3} - 51.74)_+ + 0.36x_{i4}$$
$$-5.09(x_{i4} - 32.98)_+ + 6.63(x_{i4} - 35.12)_+ - 1.86(x_{i4} - 41.53)_+$$

The suggestions that the authors can convey from this study are that future research can use other research data. such as longitudinal data. and future research can also use nonparametric spline regression with different basic approaches. such as b-spline. p-spline. and so on.

REFERENCES

[1]   A. A. R. Fernandes, I. Nyoman Budiantara, B. W. Otok, and Suhartono, "Spline estimator for bi-responses nonparametric regression model for longitudinal data," *Appl. Math. Sci.*, vol. 8, no. 113–116, pp. 5653–5665, 2014, doi: 10.12988/ams.2014.47566.

[2]   T. Purnaraga, S. Sifriyani, and S. Prangga, "Regresi Nonparamaetrik Spline Pada Data Laju Pertumbuhan Ekonomi Di Kalimantan," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 14, no. 3, pp. 343–356, 2020, doi: 10.30598/barekengvol14iss3pp343-356.

[3]   A. B. Padatuan, S. Sifriyani, and S. Prangga, "Pemodelan Angka Harapan Hidup Dan Angka Kematian Bayi Di Kalimantan Dengan Regresi Nonparametrik Spline Birespon," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 15, no. 2, pp. 283–296, 2021, doi: 10.30598/barekengvol15iss2pp283-296.

[4]   Sifriyani, H. Ilmi, and Z. Mar'Ah, "Application of Nonparametric Geographically Weighted Spline Regression Model for Spatial Mapping of Open Unemployment Rate in Kalimantan," *J. Phys. Conf. Ser.*, vol. 2123, no. 1, 2021, doi: 10.1088/1742-6596/2123/1/012038.

[5]   S. Sifriyani, M. Rasjid, D. Rosadi, S. Anwar, R. D. Wahyuni, and S. Jalaluddin, "Spatial-Temporal Epidemiology of COVID-19 Using a Geographically and Temporally Weighted Regression Model," *Symmetry (Basel).*, vol. 14, no. 4, 2022, doi: 10.3390/sym14040742.

[6]   M. D. Pasarella, S. Sifriyani, and F. D. T. Amijaya, "Nonparametrik Regression Model Estimation With the Fourier Series the Fourier Series Approach and Its Application To the Accumulative Covid-19 Data in Indonesia," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 16, no. 4, pp. 1167–1174, 2022, doi: 10.30598/barekengvol16iss4pp1167-1174.

[7]   S. P. Rifka Nurfiaza Zarkasi, Sifriyani, "IDENTIFIKASI FAKTOR-FAKTOR YANG MEMPENGARUHI MENGGUNAKAN REGRESI PANEL Identification of The Factors That Influence The Human Development Index In Kalimantan Using Panel Regression," vol. 15, no. 2, pp. 277–282, 2021.

[8]   N. Chamidah, B. Lestari, and T. Saifudin, "Modeling of blood pressures based on stress score using least square spline

estimator in bi-response non-parametric regression," *Int. J. Innov. Creat. Chang.*, vol. 5, no. 3, pp. 1200–1216, 2019.

[9]   K. Nakamura, S. Matsumura, and T. Mizutani, "Particle-to-surface frictional contact algorithm for material point method using weighted least squares," *Comput. Geotech.*, vol. 134, no. October 2020, p. 104069, 2021, doi: 10.1016/j.compgeo.2021.104069.

[10]   P. Dewanti, I. Nyoman Budiantara, and A. T. Rumiati, "Modelling of SDG's Achievement in East Java Using Bi-responses Nonparametric Regression with Mixed Estimator Spline Truncated and Kernel," *J. Phys. Conf. Ser.*, vol. 1562, no. 1, pp. 0–11, 2020, doi: 10.1088/1742-6596/1562/1/012016.

[11]   A. Islamiyati, Fatmawati, and N. Chamidah, "Estimation of Covariance Matrix on Bi-Response Longitudinal Data Analysis with Penalized Spline Regression," *J. Phys. Conf. Ser.*, vol. 979, no. 1, 2018, doi: 10.1088/1742-6596/979/1/012093.

[12]   A. Islamiyati, A. Kalondeng, N. Sunusi, M. Zakir, and A. K. Amir, "Biresponse nonparametric regression model in principal component analysis with truncated spline estimator," *J. King Saud Univ. - Sci.*, vol. 34, no. 3, p. 101892, 2022, doi: 10.1016/j.jksus.2022.101892.

[13]   J. I. Martínez, E. M. Román, E. L. Alfaro, C. Grandi, and J. E. Dipierri, "Geographic altitude and prevalence of underweight, stunting and wasting in newborns with the INTERGROWTH-21st standard," *J. Pediatr. (Rio. J).*, vol. 95, no. 3, pp. 366–373, 2019, doi: 10.1016/j.jped.2018.03.007.

[14]   I. P. De Ramos, A. H. Auchincloss, and U. Bilal, "Exploring inequalities in life expectancy and lifespan variation by race/ethnicity and urbanicity in the United States: 1990 to 2019," *SSM - Popul. Heal.*, vol. 19, no. September, p. 101230, 2022, doi: 10.1016/j.ssmph.2022.101230.

[15]   H. Y. Shin and H. T. Kang, "Recent trends in the prevalence of underweight, overweight, and obesity in Korean adults: The Korean National Health and Nutrition Examination Survey from 1998 to 2014," *J. Epidemiol.*, vol. 27, no. 9, pp. 413–419, 2017, doi: 10.1016/j.je.2016.08.014.

[16]   S. Syahrul, R. Kimura, A. Tsuda, T. Susanto, R. Saito, and F. Ahmad, "Prevalence of underweight and overweight among school-aged children and it's association with children's sociodemographic and lifestyle in Indonesia," *Int. J. Nurs. Sci.*, vol. 3, no. 2, pp. 169–177, 2016, doi: 10.1016/j.ijnss.2016.04.004.

[17]   R. K. Behl, B. D. Chaudhry, and K. Ramlila, "62201_0715078102_Artikel Journal annals of biology."

[18]   F. Bianchi, G. Bianchi, and D. Song, "The long-term impact of the COVID-19 unemployment shock on life expectancy and mortality rates," *J. Econ. Dyn. Control*, vol. 146, p. 104581, 2023, doi: 10.1016/j.jedc.2022.104581.

[19]   S. Sifriyani, "Simultaneous Hypothesis Testing of Multivariable Nonparametric Spline Regression in the GWR Model," *Int. J. Stat. Probab.*, vol. 8, no. 4, p. 32, 2019, doi: 10.5539/ijsp.v8n4p32.

[20]   A. I. R. Minum, P. A. M. Tirta, and M. Kota, "http://ejurnal.binawakya.or.id/index.php/MBI Vol.16 No.2 September 2021 Open Journal Systems," vol. 16, no. 2, pp. 6427–6438, 2021.