BAREKENG: Journal of Mathematics and Its Applications
September 2024 Volume 18 Issue 3 Page 2015–2028
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol18iss3pp2015-2026

MODELING STUNTING PREVALENCE IN INDONESIA USING SPLINE TRUNCATED SEMIPARAMETRIC REGRESSION

Rizki Dwi Fadlirhohim[1*], **Sifriyani**[2], **Andrea Tri Rian Dani**[3]

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

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

ABSTRACT

Article History:

Received: 18th April 2024
Revised: 10th May 2024
Accepted: 21st July 2024
Published: 1st September 2024

Keywords:

Semiparametric Regression;
Stunting Prevalence;
Spline Truncated.

Semiparametric regression combines parametric and nonparametric regression approaches. It is employed when the relationship pattern of the response variable is known with some predictors, while for other predictors, the relationship pattern is uncertain. The parametric regression component in this study is linear regression, while the nonparametric component utilizes a spline truncated estimator, resulting in a semiparametric spline truncated regression model. The case study focuses on the prevalence of stunting across 34 provinces in Indonesia in 2022, revealing a relatively high prevalence of 21.60%. The research aims to determine the optimal number of knots, the best model, and factors influencing stunting prevalence in Indonesia. The findings indicate that the optimal three-knot model with a GCV of 9.30 yields an RMSE of 1.70 and R^2 *of 92.71%. Significance tests for simultaneous and partial parameters reveal that all predictor variables significantly influence stunting prevalence.*

How to cite this article:

Fadlirhohim, R. D., Sifriyani and A. T. R. Dani., "MODELING STUNTING PREVALENCE IN INDONESIA USING SPLINE TRUNCATED SEMIPARAMETRIC REGRESSION," *BAREKENG: J. Math. & App.*, vol. 18, iss. 3, pp. 2015-2028, September, 2024.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Regression analysis is a statistical technique that clarifies the relationship between a predictor variable and response variables [1]. A function can be approximated in regression modeling using parametric, nonparametric, and semiparametric approaches [2]. Parametric regression is suited for scenarios with a preconceived notion about the functional relationship between the response variable and the predictors [3]. Nonparametric regression is a method suited for situations where the underlying form of the association between the response variable and the predictors is uncertain [4].

Given the current advancements in technology and information, there is no guarantee that regression modeling only considers one component, such as utilizing either parametric or nonparametric regression alone. In cases of multivariable regression, if one predictor variable exhibits a known relationship pattern with the response variable while the other does not, then a semiparametric regression approach can be employed to address such issues.

Semiparametric regression combines features of both parametric and nonparametric regression techniques for a more flexible analysis. Semiparametric regression comes into play when the relationship between the response variable and some predictor variables is straightforward. However, the influence of other predictors needs to be clarified or defined [5]. The parametric component of this research is estimated using linear regression. In contrast, the nonparametric component is estimated using a spline truncated.

Linear regression is employed when there is an assumption that the relationship between the predictor and response variables follows a linear pattern, which can be identified through scatter plots and other diagnostic techniques [6]. A spline truncated comprises polynomial segments possessing both segmented and continuous properties, rendering it effective in capturing the local characteristics of the data function [7]. An advantage of spline truncated lies in its flexibility. It allows the model to autonomously adapt to the data pattern by estimating values wherever the pattern shifts, facilitated by knot points [8]. Several studies have investigated spline truncated semiparametric regression, including ([9], [10], [11])

Semiparametric spline truncated regression can be applied in various fields, including economics, education, and health. This study uses a case study related to stunting, which falls within the health domain. Stunting, a condition defined by stunted physical growth and impaired cognitive development due to chronic malnutrition in children, is a prevalent public health concern in Indonesia, particularly impacting toddlers [12]. The data shows that Indonesia has a relatively high prevalence of stunting, especially in areas with significant poverty levels. Health Minister Budi Gunadi Sadikin hopes that in this year of normalcy, the reduction in stunting cases can be even more pronounced, aiming to achieve the target of reducing stunting to 14 percent by 2024. Stunting in Indonesia is a serious public health problem believed to be influenced by various complex factors, including poor nutrition, inadequate sanitation, limited access to healthcare services, and socioeconomic factors [13]. Based on the previous explanations, the researchers are interested in modeling the prevalence of stunting in Indonesia using semiparametric spline truncated regression.

2. RESEARCH METHODS

The research design employed in this research is ex post facto, indicating that data collection occurred after all relevant events had transpired. The data utilized are secondary data obtained from the Indonesian Nutrition Status Survey and the official website of the Indonesian Central Bureau of Statistics. The sampling technique adopted in this research is purposive sampling or judgmental sampling. The software used in data processing is R Studio.

2.1 Data and Data Sources

The population under investigation in this research encompasses all toddlers with stunting prevalence across the 34 provinces of Indonesia. The sample utilized in this study comprises toddlers with stunting prevalence across the same 34 provinces of Indonesia in the year 2022. The research variables encompass predictor and response variables, as outlined in Table 1.

Table 1. Research Variables

	Variable	Unit
Response	Prevalence of Stunting	Percent (%)
Predictor	The percentage of infants under 6 months old who receive exclusive breastfeeding.	Percent (%)
	The percentage of infants who receive early initiation of breastfeeding.	Percent (%)
	The percentage of children aged 12-23 months who receive complete basic immunization.	Percent (%)
	The percentage of households with access to adequate sanitation.	Percent (%)
	Human Development Index	

2.2 Parametric Regression Analysis

Parametric regression leverages statistical techniques where the relationship between variables takes a predetermined form, like linear, quadratic, or exponential. This method relies on specific assumptions to create distributions with specific parameter values. Some basic assumptions for using parametric regression include that the samples obtained are identically independent and normally distributed (normality of errors), there is no association among predictor variables (multicollinearity), there is no influence from variables in the model through time lag (autocorrelation), and the error properties have unequal variances (heteroskedasticity) [14].

The multiple linear regression model given i observations with p predictor variables can be expressed as **Equation (1)**.

$$y_i = \beta_0 + \sum_{p=1}^{l} \beta_p x_{pi} + \varepsilon_i,\ i = 1,2,\ldots,n;\ p = 1,2,\ldots,l \tag{1}$$

Based on **Equation (1)**, it can be written in matrix form as **Equation (2)**.

$$y = \boldsymbol{X\beta} + \varepsilon \tag{2}$$

where,

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{l1} \\ 1 & x_{12} & \cdots & x_{l2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{ln} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_l \end{bmatrix}, \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{3}$$

2.3 Nonparametric Spline Truncated Regression Analysis

Nonparametric regression is a technique employed to determine the influence of predictor variables on the response variable where the data's pattern (regression curve) is unknown or does not conform to a specific pattern [15]. Spline truncated nonparametric regression is a smooth segmented polynomial model. The spline truncated function has connecting points called knot points [16]. Spline truncated has the advantage of handling data patterns that exhibit sharp increases or decreases with the help of knot points, and the resulting curve appears relatively smooth [17]. Below is the nonparametric spline truncated regression model written as **Equation (3)**.

$$y_i = \delta_0 + \sum_{j=1}^{s}\sum_{h=1}^{m} \delta_{jh} z_{ji}^h + \sum_{j=1}^{s}\sum_{k=1}^{r} \delta_{j_{m+k}} (z_{ji} - K_{kj})_+^m + \varepsilon_i,\ i = 1,2,\ldots,n \tag{4}$$

where the truncated (piecewise) function is described in **Equation (4)** [17], [18], [19].

$$(z_{ji} - K_{kj})_+^m = \begin{cases} (z_{ji} - K_{kj})_+^m & , z_{ji} \geq K_{kj} \\ 0 & , z_{ji} < K_{kj} \end{cases} \tag{5}$$

Based on **Equation (3)**, it can be written in matrix form as **Equation (5)**.

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\delta} + \boldsymbol{\varepsilon} \tag{6}$$

where,

$$\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}^{\mathrm{T}} \tag{7}$$

$$\mathbf{Z} = \begin{bmatrix} 1\, z_{11}\, z_{11}^2 \ldots z_{s1}^2\; z_{s1}^m\; (z_{11}-K_{11})_+^m \ldots (z_{11}-K_{s1})_+^m \ldots (z_{s1}-K_{1r})_+^m \ldots (z_{s1}-K_{sr})_+^m \\ 1\, z_{12}\, z_{12}^2 \ldots z_{s2}^2\; z_{s2}^m\; (z_{12}-K_{11})_+^m \ldots (z_{12}-K_{s1})_+^m \ldots (z_{s2}-K_{1r})_+^m \ldots (z_{s2}-K_{sr})_+^m \\ \vdots\;\vdots\;\vdots\;\ddots\;\vdots\;\vdots \quad \vdots \quad \ddots \quad \vdots \quad \ddots \quad \vdots \quad \ddots \quad \vdots \\ 1\, z_{1n}\, z_{1n}^2 \ldots z_{sn}^2\; z_{sn}^m\; (z_{1n}-K_{11})_+^m \ldots (z_{1n}-K_{s1})_+^m \ldots (z_{sn}-K_{1r})_+^m \ldots (z_{sn}-K_{sr})_+^m \end{bmatrix} \tag{8}$$

$$\boldsymbol{\delta} = \begin{bmatrix} \delta_0 & \delta_{11} & \delta_{12} & \ldots & \delta_{1m} & \ldots & \delta_{sm} & \delta_{s(m+1)} & \ldots & \delta_{s(m+r)} \end{bmatrix}^{\mathrm{T}} \tag{9}$$

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 & \varepsilon_2 & \cdots & \varepsilon_n \end{bmatrix}^{\mathrm{T}} \tag{10}$$

2.4 Semiparametric Spline Truncated Regression Analysis

Semiparametric regression is a statistical analysis technique that integrates elements of both parametric regression and nonparametric regression methodologies [20]. Assuming a regression analysis involving paired data (x_i, y_i, z_i), where the relationship among these variables is hypothesized to adhere to a semiparametric regression model, it can be represented as **Equation (6)**.

$$y_i = f(x_i) + h(z_i) + \varepsilon_i \tag{11}$$

In semiparametric spline truncated regression, given paired data $(x_{1i}, x_{2i}, \ldots, x_{li}, z_{1i}, z_{2i}, \ldots, z_{si}, y_i)$ that exhibit a relationship, it is further assumed that the pattern of this relationship follows a semiparametric regression model. The semiparametric spline truncated regression curve can be expressed as in **Equation 7** [21].

$$\mu(x_i, z_i) = \sum_{p=1}^{l} f(x_{pi}) + \sum_{j=1}^{s} h(z_{ji}) \tag{12}$$

Based on **Equation (7)**, it can be written in matrix form as **Equation (8)** [22].

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\delta} + \boldsymbol{\varepsilon} \tag{13}$$

2.5 Generalized Cross-Validation (GCV)

The utilization of the GCV method in selecting knot points is considered the best because it possesses asymptotically optimal properties. With a large sample size, its optimality remains, making it asymptotically optimal and closely related to the sample size [23]. GCV is also invariant to transformations, meaning that even if data is transformed, its GCV generally remains the same [24].

The basis of the GCV method is to select the value of k by minimizing the value of GCV. **Equation (9)** for the GCV value [25].

$$GCV = \frac{MSE(\mathbf{K})}{\left[n^{-1} trace(\mathbf{I} - \mathbf{A}(\mathbf{F})\right]^2} \tag{14}$$

where,

$$MSE(\mathbf{K}) = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n} \tag{15}$$

I : The identity matrix is of size $n \times n$
A(F) : Matrix of size $n \times n$

2.6 Significance Testing of Model Parameters

Significance testing of parameters in regression models aims to determine whether those parameters indicate a relationship between the response variable and predictor variables and assess the parameters' suitability in explaining the model [26].

Simultaneous significance testing of parameters determines whether the parameters of the semiparametric spline truncated regression model collectively exhibit significance. Significance testing of parameters using simultaneous testing with hypotheses, critical regions, and the test statistic distributed as F is as follows.

$H_0 : \beta_1 = \beta_2 = \beta_3 = ... = \beta_l = \delta_{11} = \delta_{sm} = ... = \delta_{s,m+1} = ... = \delta_{s(m+r)} = 0$

H_1 :There is at least one $(\beta_p, \delta_{j(h+k)}) \neq 0$; where $p = 1,2,...,l; j = 1,2,...,s;\ k = 1,2,...,r$

The test statistic used in simultaneous parameter testing can be expressed in the following **Equation (10)**:

$$F = \frac{MSR}{MSE} \tag{16}$$

where,

$$MSR = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{l + (sm + sr)} \tag{17}$$

$$MSE = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n - (l + (sm + sr) + 1} \tag{18}$$

H_0 is rejected when $F > F_{\alpha;(l+(sm+sr),n-(l+(sm+sr)+1)}$ or $p - value < \alpha$, which means that at least one parameter in the semiparametric spline truncated regression model is significant. [27].

Partial parameter testing is conducted when the simultaneous parameter testing concludes that at least one parameter is significant. The hypotheses for partial testing are as follows.

$H_0 : \beta_p = 0;\ \delta_j = 0$

$H_1 : \beta_p \neq 0;\ \delta_j \neq 0,\ p = 1,2,...,l,\ j = 1,2,...,sm + sr$

Individual testing is performed using the t-test. The test statistic used is given by **Equation (11)** and **Equation (12)**.

$$t = \frac{\hat{\beta}_p}{SE(\hat{\beta}_p)} \tag{19}$$

$$t = \frac{\hat{\delta}_j}{SE(\hat{\delta}_j)} \tag{20}$$

H_0 is rejected when $|t| > t_{\frac{\alpha}{2},(n-(l+(sm+sr)-1)}$ or $p - value < \alpha$ [28].

3. RESULTS AND DISCUSSION

3.1 Descriptive Statistics

Descriptive statistics can provide a general explanation of the variable under investigation in a more understandable form [29]. Before conducting analysis and discussion, the distribution of stunting prevalence in Indonesia in 2022 can be visualized in **Figure 1**. According to the World Health Organization (WHO), the prevalence of stunting is categorized into four levels: low, medium, high, and very high [30].

Figure 1. Map of stunting prevalence distribution in Indonesia in 2022
***Source**: Quantum Geographic Information System (QGIS)*

Based on **Figure 1**, it is known that in the first group, there are 12 provinces with low stunting prevalence in the range of 0%-20%. Furthermore, in the second group, there are 17 provinces with moderate stunting prevalence in the range of 20.01%-30%. The third group consists of 5 provinces with high stunting prevalence in the range of 30.01%-40%. In this study, descriptive statistics used are mean, minimum, and maximum, displayed in **Table 2**.

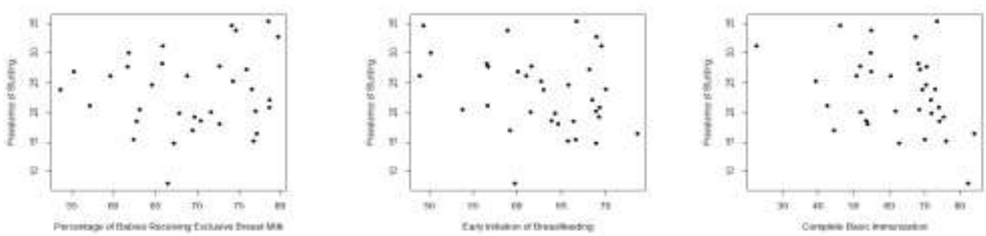
Table 2. Descriptive statistics of response and predictor variables

Variable	Mean	Minimum	Maximum
Prevalence of Stunting	23.25%	8.00%	35.30%
The percentage of infants under 6 months old who receive exclusive breastfeeding.	69.21%	53.60%	79.69%
The percentage of infants who receive early initiation of breastfeeding.	62.88%	48.87%	73.59%
The percentage of children aged 12-23 months who receive complete basic immunization.	62.18%	22.52%	83.89%
The percentage of households with access to adequate sanitation.	80.32%	68.88%	96.21%
Human Development Index	71.96	61.39	81.65

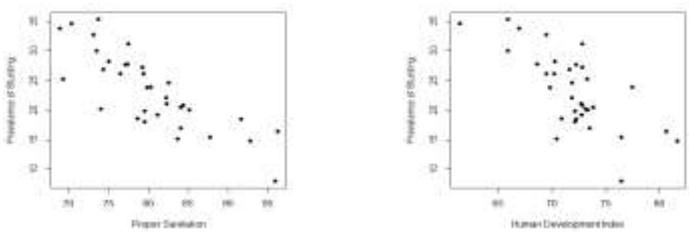
3.2 Scatter Plot

The initial step in modeling factors influencing the prevalence of stunting in Indonesia is to examine the relationship between the prevalence of stunting in Indonesia in 2022 and variables suspected to influence it using scatterplots.

Through a scatter plot, we can observe the relationship between stunting prevalence and suspected influencing factors, where the determination of variables as components of parametric and nonparametric can be made by observing the pattern obtained **[31]**. The scatter plot between response variables and predictor variables in this study can be seen in **Figure 2**.



(a) **(b)** **(c)**



(e) (d)

Figure 2. Scatter plot between predictor variables and response variable
***Source**: Rstudio*

Based on **Figure 2**, it can be observed that the data pattern between stunting prevalence and the percentage of infants under 6 months who receive exclusive breastfeeding (a), the percentage of infants who receive early initiation of breastfeeding (b), and the percentage of children aged 12-23 months who receive complete basic immunization (c) do not exhibit a specific pattern; the data are scattered, and there are some data points far from the distribution of other data points. The data pattern between stunting prevalence and the percentage of households with access to adequate sanitation (d) and the Human Development Index (e) tends to form a specific pattern, which is linear. The results of the identification regarding predictor variables are presented in **Table 3**.

Table 3. Identification of Predictor Variables

Variable	Estimator	Description
x_1	Linear Regression	The percentage of infants under 6 months old who receive exclusive breastfeeding.
x_2		The percentage of infants who receive early initiation of breastfeeding.
z_1	*Spline Truncated*	The percentage of children aged 12-23 months who receive complete basic immunization.
z_2		The percentage of households with access to adequate sanitation.
z_3		Human Development Index

3.3 Modeling Semiparametric Spline Truncated Regression

The mathematical model of semiparametric spline truncated regression on the prevalence of stunting in Indonesia in 2022 is shown in **Equation (13)**:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \delta_{11} z_{1i} + \delta_{21} z_{2i} + \delta_{31} z_{3i} + \sum_{j=1}^{3}\sum_{k=1}^{3} \delta_{j_{m+k}} (z_{ji} - K_{kj})_+^m + \varepsilon_i \tag{21}$$

3.4 Parameter Estimation and Optimal Knot Selection

Parameter estimation is conducted utilizing the Ordinary Least Squares (OLS) method. Additionally, **Table 4** showcases the positions of knot points along with the optimal count of knot points corresponding to the minimum GCV value.

Table 4. Minimum GCV Value

No	Number of Knot Points	Location of Knot Points	GCV	RMSE	R^2
1	1 Knot Point	$K_{11} = 73.83$, $K_{12} = 68.04$, $K_{13} = 70.11$	14.72	2.82	80.03%
2	2 Knot Point	$K_{11} = 67.44; K_{21} = 69.57$, $K_{12} = 61.99; K_{22} = 64.01$, $K_{13} = 55.08; K_{23} = 60.09$	12.58	2.29	86.79%

No	Number of Knot Points	Location of Knot Points	GCV	RMSE	R^2
3	3 Knot Point	$K_{11}=67.27; K_{21}=69.75; K_{31}=74.72$, $K_{12}=61.82; K_{22}=64.17; K_{32}=68.88$, $K_{13}=54.67; K_{23}=60.51; K_{33}=72.20$	9.30	1.70	92.71%

As indicated in **Table 4**, the lowest GCV value is attained with three knot points, yielding a GCV value of 9.30. The predicted values of y derived from the semiparametric spline truncated regression model employing three knot points are depicted in **Figure 3**.

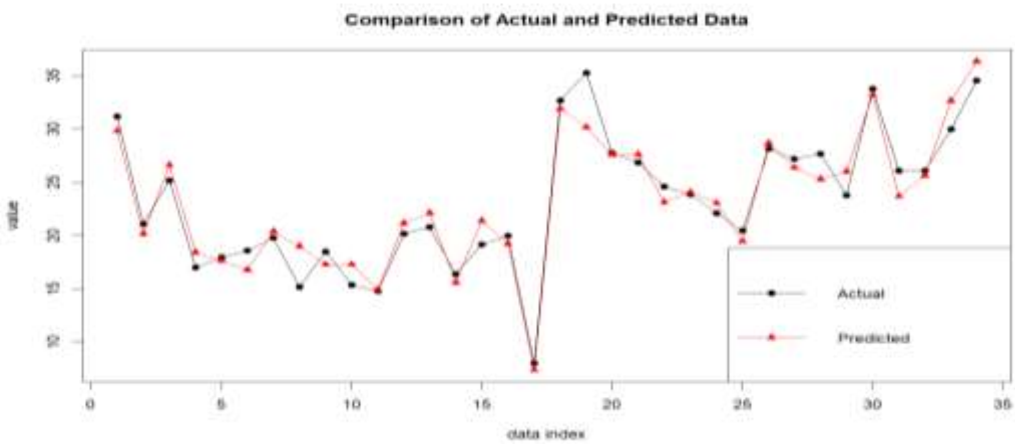


Figure 3. Comparison Plot of Actual Data with Predicted Data
***Source**: Rstudio*

The parameter estimation results of the semiparametric spline truncated regression model using the OLS method with three knot points are displayed in **Table 5**.

Table 5. Parameter Estimation

Variable	Parameter	Parameter Estimation Results	Variable	Parameter	Parameter Estimation Results
-	$\hat{\beta}_0$	132.52		$\hat{\delta}_{21}$	0.03
x_1	$\hat{\beta}_1$	-0.67	z_2	$\hat{\delta}_{22}$	109.33
x_2	$\hat{\beta}_2$	-0.86		$\hat{\delta}_{23}$	26.84
	$\hat{\delta}_{11}$	-0.01		$\hat{\delta}_{24}$	21.14
z_1	$\hat{\delta}_{12}$	-53.23		$\hat{\delta}_{31}$	0.09
	$\hat{\delta}_{13}$	-14.04	z_3	$\hat{\delta}_{32}$	-56.93
	$\hat{\delta}_{14}$	-10.29		$\hat{\delta}_{33}$	-10.89
				$\hat{\delta}_{34}$	-11.08

According to **Table 5** and **Table 4**, the parameter estimation outcomes of the semiparametric spline truncated regression on stunting prevalence in Indonesia in 2022, considering suspected influencing factors, have been derived. The semiparametric spline truncated regression model with three knot points can be expressed as **Equation (15)**.

$$\begin{aligned}\hat{y} = {} & 132.52 - 0.67x_{1i} - 0.86x_{2i} - 0.01z_{1i} + 0.03z_{2i} + 0.09z_{3i} - \\ & 53.23(z_{1i}-67.27)_+ + 109.33(z_{2i}-61.82)_+ - 56.93(z_{3i}-54.67)_+ - \\ & 14.04(z_{1i}-69.75)_+ + 26.84(z_{2i}-64.17)_+ - 10.89(z_{3i}-60.51)_+ - \\ & 10.29(z_{1i}-74.72)_+ + 21.14(z_{2i}-68.88)_+ - 11.08(z_{3i}-72.20)_+\end{aligned} \tag{22}$$

3.5 Parameter Significance Testing

This test is conducted to determine each predictor variable that significantly influences the response variable. The results of the simultaneous testing are presented in **Table 6**.

Table 6. Analysis of Variance

Source of Variance	Degrees of Freedom	Sum of Squares	Mean Square	F_{value}	$F_{\alpha,df1,df2}$
Regression	14	1256.96	89.78	17.27	2.25
Error	19	98.76	5.12		
Total	33	1355.72			

Based on **Table 6**, the obtained value of F_{value} is 17.27 and $F_{0,05;(14,19)}$ is 2.25. Since $F_{value} > F_{0,05;(14,19)}$, the decision is to reject H_0. Hence, it can be deduced that collectively, at least one predictor variable significantly impacts the prevalence of stunting in Indonesia in 2022. Moreover, the significance testing of individual model parameters can be carried out. Partial significance testing of model parameters uses the t-test statistic with degrees of freedom $t_{\alpha/2,(n-(l+(sm+sr)-1)}$. The results of the partial significance testing of model parameters are displayed in **Table 7**.

Table 7. Partial test

Variable	Parameter	Parameter Estimation Results	t_{value}	$t_{\frac{\alpha}{2},(n-(l+(sm+sr)-1)}$	Decision
-	$\hat{\beta}_0$	132.52	9.06*		H_0 rejected
x_1	$\hat{\beta}_1$	-0.67	-5.33*		H_0 rejected
x_2	$\hat{\beta}_2$	-0.86	-3.90*		H_0 rejected
z_1	$\hat{\delta}_{11}$	-0.01	-0.07		H_0 is not rejected
	$\hat{\delta}_{12}$	-53.23	-3.88*		H_0 rejected
	$\hat{\delta}_{13}$	-14.04	-2.39*		H_0 rejected
	$\hat{\delta}_{14}$	-10.29	-3.24*		H_0 rejected
z_2	$\hat{\delta}_{21}$	0.03	0.23	2.09	H_0 is not rejected
	$\hat{\delta}_{22}$	109.33	3.89*		H_0 rejected
	$\hat{\delta}_{23}$	26.84	2.26*		H_0 rejected
	$\hat{\delta}_{24}$	21.14	3.72*		H_0 rejected
z_3	$\hat{\delta}_{31}$	0.09	0.96		H_0 is not rejected
	$\hat{\delta}_{32}$	-56.93	-3.79*		H_0 rejected
	$\hat{\delta}_{33}$	-10.89	-1.61		H_0 is not rejected
	$\hat{\delta}_{34}$	-11.08	-3.85*		H_0 rejected

Note: (*) denotes parameters on predictor variables that are significant

Based on **Table 7**, 11 significant out of 15 parameters in the semiparametric spline truncated regression model are found. While four parameters are not significant, overall, we obtain 5 variables that significantly influence the model because in nonparametric regression, if at least one parameter is significant, it can be concluded that the variable significantly influences the model. These variables are the percentage of infants receiving exclusive breastfeeding for 6 months, the percentage of infants receiving early initiation of breastfeeding, the percentage of children aged 12-23 months receiving complete basic immunization, the percentage of households with access to adequate sanitation, and the Human Development Index.

3.6 Interpretation of Spline Truncated Semiparametric Regression Models

The following is the model interpretation of each significant variable influencing the prevalence of stunting.

1. The influence of the percentage of households with access to adequate sanitation on the prevalence of stunting is written in **Equation (15)**.

$$\hat{y}_i = 132,52 - 0,67x_{1i} \tag{23}$$

The regression coefficient of 0.67 indicates that for every 1% increase in the percentage of households with access to adequate sanitation, the prevalence of stunting will decrease by 0.67%.

2. The influence of the Human Development Index on the prevalence of stunting is written in **Equation (16)**.

$$\hat{y}_i = 132,52 - 0,86x_{2i} \tag{24}$$

The regression coefficient of 0.86 indicates that for every 1 unit increase in the Human Development Index, the prevalence of stunting will decrease by 0.86%.

3. The influence of the percentage of infants receiving exclusive breastfeeding for 6 months on the prevalence of stunting is written in **Equation (17)**.

$$\hat{y}_i = -0,01z_{1i} - 53,23(z_{1i} - 67,27)_+ - 14,04(z_{1i} - 69,75)_+ - 10,29(z_{1i} - 74,72)_+ \tag{25}$$

The truncated function for **Equation (17)** can be presented in **Equation (18)**.

$$\begin{cases} -0,01z_{1i} & z_{1i} < 67,27 \\ -53,24z_{1i} + 3580,78 & 67,27 \le z_{1i} < 69,75 \\ -67,28z_{1i} + 4560,07 & 69,75 \le z_{1i} < 74,72 \\ -77,57z_{1i} + 5359,81 & z_{1i} \ge 74,72 \end{cases} \tag{26}$$

The interpretation of **Equation (18)** can be visualized through a map presentation in **Figure 4**.

Figure 4. Map depicting the distribution of the percentage of infants receiving exclusive breastfeeding for 6 months
***Source**: Quantum Geographic Information System (QGIS)*

Figure 4 shows that in the first group, there are 14 provinces with a low percentage of infants receiving exclusive breastfeeding for 6 months, ranging from less than 67.27%. Furthermore, in the second group, there are 4 provinces with a moderate percentage of infants receiving exclusive breastfeeding for 6 months, ranging from 67.27% to 69.75%. The third group comprises 7 provinces with a high percentage of infants receiving exclusive breastfeeding for 6 months, ranging from 69.75% to 74.72%. Lastly, the fourth group includes 9 provinces with a very high percentage of infants receiving exclusive breastfeeding for 6 months, ranging from equal to or more than 74.72%.

4. The influence of the percentage of infants receiving early initiation of breastfeeding on the prevalence of stunting is written in **Equation (19)**.

$$\hat{y}_i = 0,03z_{2i} + 109,33(z_{2i} - 61,82)_+ + 26,84(z_{2i} - 64,17)_+ + 21,14(z_{2i} - 68,88)_+ \tag{27}$$

The truncated function for **Equation (19)** can be presented in **Equation (20)**.

$$\begin{cases} 0,03z_{2i} & z_{2i} < 61,82 \\ 109,36z_{2i} - 6758,78 & 61,82 \le z_{2i} < 64,17 \\ 136,20z_{2i} - 8481,10 & 64,17 \le z_{2i} < 68,88 \\ 157,34z_{2i} - 9937,22 & z_{2i} \ge 68,88 \end{cases} \tag{28}$$

In **Equation (20)**, it can be interpreted through a map visualization in **Figure 5**.

Figure 5. Map depicting the distribution of the percentage of infants receiving early initiation of breastfeeding

Source: *Quantum Geographic Information System (QGIS)*

Based on **Figure 5**, it is known that in the first group, there are 14 provinces with a low percentage of infants receiving early initiation of breastfeeding, ranging from less than 61.82%. Furthermore, in the second group, there are 3 provinces with a moderate percentage of infants receiving early initiation of breastfeeding, ranging from 61.82% to 64.17%. The third group consists of 10 provinces with a high percentage of infants receiving early initiation of breastfeeding, ranging from 64.17% to 68.88%. Lastly, the fourth group includes 7 provinces with a very high percentage of infants receiving early initiation of breastfeeding, ranging from equal to or more than 68.88%.

5. The influence of the percentage of children aged 12-23 months receiving complete basic immunization on the prevalence of stunting is written in **Equation (21)**.

$$\hat{y}_i = 0,09\,z_{3i} - 56,93(z_{3i} - 54,67)_+ - 10,89(z_{3i} - 60,51)_+ - 11,08(z_{3i} - 72,20)_+ \tag{29}$$

The truncated function for **Equation (21)** can be presented in **Equation (22)**.

$$\begin{cases} 0,09z_{3i} & z_{3i} < 54,67 \\ -56,93z_{3i} + 3112,36 & 54,67 \le z_{3i} < 60,51 \\ -67,82z_{3i} + 3771,31 & 60,51 \le z_{3i} < 72,20 \\ -78,90z_{3i} + 4571,29 & z_{3i} \ge 72,20 \end{cases} \tag{30}$$

In **Equation (22)**, it can be interpreted through a map visualization in **Figure 6**.

Figure 6. Map depicting the distribution of the percentage of children aged 12-23 months receiving complete basic immunization

Source: *Quantum Geographic Information System (QGIS)*

Based on **Figure 6**, it is known that in the first group, there are 11 provinces with a low percentage of children aged 12-23 months receiving complete basic immunization, ranging from less than 54.67%. Furthermore, in the second group, there are 3 provinces with a moderate percentage of children aged 12-23 months receiving complete basic immunization, ranging from 54.67% to 60.51%. The third group consists of 12 provinces with a high percentage of children aged 12-23 months receiving complete basic immunization, ranging from 60.51% to 72.20%. Lastly, the fourth group includes 8 provinces with a very high percentage of children aged 12-23 months receiving complete basic immunization, ranging from equal to or more than 72.20%.

4. CONCLUSIONS

Based on the analysis and discussion results, the following conclusions are drawn:

1. The semiparametric regression model with the spline truncated approach on the prevalence of stunting in Indonesia in 2022, based on optimal knot points using the GCV method, is obtained from three knot points with a GCV value of 9.30, resulting in an RMSE of 1.70 and an R-squared value of 92.71%.
2. The best model for semiparametric regression with the spline truncated approach using three knot points is described as follows:

$$\hat{y} = 132.52 - 0.67x_{1i} - 0.86x_{2i} - 0.01z_{1i} + 0.03z_{2i} + 0.09z_{3i} - 53.23(z_{1i} - 67.27)_+ + 109.33(z_{2i} - 61.82)_+ - 56.93(z_{3i} - 54.67)_+ - 14.04(z_{1i} - 69.75)_+ + 26.84(z_{2i} - 64.17)_+ - 10.89(z_{3i} - 60.51)_+ - 10.29(z_{1i} - 74.72)_+ + 21.14(z_{2i} - 68.88)_+ - 11.08(z_{3i} - 72.20)_+$$

3. Based on the significance testing of parameters, both simultaneously and partially, it is known that the variables percentage of households with access to adequate sanitation, Human Development Index, percentage of infants under 6 months receiving exclusive breastfeeding, percentage of infants receiving early initiation of breastfeeding, and percentage of children aged 12-23 months receiving complete basic immunization significantly influence the prevalence of stunting.

REFERENCES

[1] W. Green, and B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. Arrowsmith Ltd., Bristol. 2000.

[2] I. N. Budiantara, *Regresi Nonparametrik Spline Truncated*. Surabaya: ITS Press. 2019.

[3] A. T. R. Dani, and L. Ni'matuzzahroh, "Penerapan Keluarga Model Spline Truncated Polinomial pada Regresi Nonparametrik. Inferensi," *Department of Statistics, ITS*, vol. 5, no. 1. pp. 37-44, 2022.

[4] Sifriyani., A. T. R. Dani, M. Fauziyah., M. N. Hayati, S. Wahyuningsih, and S. Prangga, "Spline And Kernel Mixed Estimators In Multivariable Nonparametric Regression For Dengue Hemorrhagic Fever Model," *Commun. Math. Biol. Neurosci*, pp. 1–15, 2023.

[5] A. T. R . Dani, and N. Y. Adrianingsih, "Estimasi Model Regresi Semiparametrik Spline Truncated menggunakan metode Maximum Likelihood Estimation (MLE)," *Jambura Journal Of Probability and Statistics*. vol. 2, no. 2. pp. 56-63, 2021.

[6] Litawati and I. N. Budiantara, "Pendekatan Regresi Nonparametrik Spline Untuk Pemodelan Laju Pertumbuhan Ekonomi (LPE) di Jawa Timur," *Jurnal Sains dan Seni Pomits*, vol. 2, no. 2, 2013.

[7] R. L. Eubank, *Nonparametric Regression and Spline Smoothing*. New York: Marcel Dekker Inc. Ebook. 1988.

[8] R. L. Eubank, *Nonparametric Regression and Spline Smoothing 2nd*. New York: Marcel Dekker Inc. Ebook. 1999.

[9] I. N. Budiantara, and A.P. Sugiantari, "Analisis Faktor-faktor yang Mempengaruhi Angka Harapan Hidup di Jawa Timur Menggunakan Regresi Semiparametrik Spline," *Jurnal Sains dan Seni Pomits*, vol. 2, no. 1, pp. 37-41, 2013.

[10] E. O. Permatasari, T.A. Rumiati, and M. Ishaq, "Analisis Faktor-Faktor yang Mempengaruhi Produksi Padi di Provinsi Jawa Timur Menggunakan Regresi Semiparametrik Spline," *Jurnal Sains dan Seni ITS*, vol. 5, no. 1, pp. 420-425, 2017.

[11] S. Side, W. Sanusi, and M. W. Maksum, "Model Regresi Semiparametrik Spline untuk Data Longitudinal pada Kasus Demam Berdarag Dengue di Kota Makassar," *Journal of Mathematics, Computations, and Statistics*. vol. 3, no. 1, pp. 20-31, 2020.

[12] V. E. Rahmawati, E. P. Pamungkasari, and B. Murti, "Determinants of Stunting and Child Development in Jombang District," *Journal of Maternal and Child Health*, vol. 3, no. 1, pp. 68–80, 2018.

[13] Kementerian Kesehatan RI, *Hasil Survei Status Gizi Indonesia (SSGI) 2022*. Jakarta: Kemenkes. 2022.

[14] Abdy, "Tinjauan Singkat tentang Regresi Parametrik dan Regresi Nonparametrik," *SAINTIFIK: Jurnal Matematika, Sains dan Pembelajarannya*. vol. 5, no. 1, 2019.

[15] W. Noviani, I. Purnamasari, and Sifriyani, "Pemodelan Regresi Nonparametrik Spline Linear Persentase Penduduk Miskin di Kalimantan," *Jurnal Siger Matematika*. vol. 1, no. 2. pp. 34-41, 2020.

[16] R. Mubarak, and I. N. Budiantara, “Analisis Regresi Spline Multivariabel untuk Pemodelan Kematian Penderita Demam Berdarah Dengue (DBD) di Jawa Timur,” *Jurnal Sains dan Seni ITS*, vol 1, no. 1, 2013.

[17] R. D. Ente, A. Islamiyati, and Raupong., “Pengaruh Indeks Massa Tubuh dan Trigliserida terhadap Guula Darah dengan Model Regresi Nonparametrik Spline Biprediktor,” *Jurnal Estimasi*, vol. 2, no. 2. pp. 71-79, 2021.

[18] Sifriyani, A. R. M. Sari, A. T. R. Dani, S. Jalaluddin. “Bi-response truncated spline nonparametric regression with optimal knot point selection using generalized cross-validation in diabetes mellitus patient's blood sugar levels,” *Communications in Mathematical Biology and Neuroscience*. vol. 48, 2023.

[19] Sifriyani, Diu, M.Y. Mar’ah, Z. Anggraini, D. and Jalaluddin, S., “Modeling of Dengue Hemorrhagic Fever Cases In Aws Hospital Samarinda Using Bi-Responses Nonparametric Regression With Estimator Spline Truncated,” *Communications in Mathematical Biology and Neuroscience*, vol. 27, 2023.

[20] S. B. Loklomin, “Pemodelan Indeks Pembangunan Manusia Di Kepulauan Maluku Dengan Pendekatan Estimasi Interval Parameter Model Regresi Semiparametrik Spline Truncated,” *BAREKENG: Journal of Mathematics and Itrts Applications.* vol. 13, no. 2, pp. 125-134, 2019.

[21] Y. M. Pratama, A. A. R. Fernandes, N. W. S. Wardhani, Nurjannah, and Solimun, “Statistical Modeling Of Tourism Investment Decisions In Indonesia Using Semiparametric Approach,” *BAREKENG: Journal of Mathematics and Itrts Applications.* vol. 18, no. 1, pp. 529-536, 2024.

[22] C. A. Putri, Indahwati and A. Kurnia, “Small Area Estimation Of Mean Years School In Bogor District Using Semiparametric P-Spline,” *BAREKENG: Journal of Mathematics and Itrts Applications.* vol. 16, no. 4, pp. 1541-1550, 2022.

[23] G. Wahba, *Spline Models for Observation Data Society for Industrial and Applied Mathematics*. Philadelphia: Pennsylvania. 1990.

[24] A. P. Anisar, Sifriyani, and A. T. R. Dani, “Estimation Of A Bi-Response Truncated Spline Nonparametric Regression Model On Life Expectancy And Prevalence Of Underweight Children In Indonesia,” *BAREKENG: Journal of Mathematics and Itrts Applications.* vol. 17, no. 4, pp, 2011-2022, 2023.

[25] S. Anggraeni, Sifriyani., and Q, Q. A’yun, “Tuberculosis Case Model Using Gcv And Ubr Knot Selection Methods In Truncated Spline Nonparametric Regression,” *BAREKENG: Journal of Mathematics and Itrts Applications*. vol. 17, no. 3, pp. 1565-1574.

[26] Sudjana., *Metoda Statistika*. Bandung: Tarsito. 2002.

[27] M. A. D. Octovanny, “Pemodelan Indeks Pembangunan Kesehatan Masyarakat Provinsi Papua Dengan Regresi Semiparametrik Truncated Spline,” *Prosiding Seminar Nasional Matematika, Statistika, dan Aplikasinya*, pp. 29-41, 2023.

[28] Utami, “Pendekatan Regresi Semiparametrik Spline Truncated Untuk Pemodelan Tingkat Pengangguran Terbuka Di Jawa Tengah,” *Jurnal Statistika*, vol. 6, no. 2, pp. 172-176.

[29] R. E. Walpole, *Pengantar Statistika edisi ke-3*. Jakarta: PT. Gramedia Pustaka Utama. 1995.

[30] World Health Organization. *Levels and Trends in Child Malnutrition*. UNICEF: WHO. 2012

[31] A. T. R. Dani, F. B. Putra, M. A. Zen, Sifriyani, M. Fauziyah, V. Ratnasari, and N. Y. Adrianingsih, “Pemodelan Kadar Hemoglobin Pada Pasien Demam Berdarah di Kota Samarinda Menggunakan Regresi Semiparametrik Spline Truncated,” *Jambura Journal of Probability and Statistics*, vol. 4, no. 2, pp. 56-64, 2023.