BAREKENG: Journal of Mathematics and Its Applications
June 2025 Volume 19 Issue 2 Page 0697–0708
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol19iss2pp0697-0708

NONPARAMETRIC REGRESSION MODELING USING THE SPLINE APPROACH TO STUNTING CASES IN INDONESIA

Mohamat Fatekurohman[1], Siti Nur Khasanah[2*], Yuliani Setia Dewi[3]

[1,2,3]*Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Jember*
Jln. Kalimantan No. 37, Jember, 68121, Indonesia

*Corresponding author's e-mail: * **nurkhas4427@gmail.com***

Article History:

Received: 1st June 2024
Revised: 12th November 2024
Accepted: 21st January 2025
Published: 1st April 2025

Keywords:

GCV;
Knots;
Nonparametric Spline Regression;
Stunting;

ABSTRACT

Indonesia is the fourth ranked country in the world and second in Southeast Asia with the highest stunting cases of 21.6%. According to the provisions of the World Health Organization (WHO), the maximum tolerance standard for stunted toddlers is 20 percent or one-fifth of the total number of toddlers, so the stunting rate in Indonesia is still relatively high. The high stunting rate in Indonesia can affect the quality of Indonesia's human resources, so early detection and immediate management of stunted toddlers are needed. Stunting is a condition of failure to grow due to chronic malnutrition which is caused by inadequate nutritional intake for a long time, resulting in being shorter than standard. This research aims to determine several factors that influence stunting in toddlers in Indonesia using the nonparametric spline regression method with one knot, two knots, three knots and the best model is found to be the one knot model. The results of regression nonparametric spline modeling with one knot are GCV of 14.32605 and R^2 of 81.1%. From the five variables, namely toddlers receiving complete basic immunization (x_1), babies receiving exclusive breast milk for 6 months (x_2), babies born receiving IMD (x_3), children aged 6-23 months consuming five of the eight food groups and drink throughout the day (x_4), households having access to proper sanitation (x_5), the following results were obtained: the variable that don't have a significant effect was toddlers receiving complete basic immunization (x_1), while the other four has a significant effect.

How to cite this article:

M. Fatekurohman, S. N. Khasanah and Y. S. Dewi., "NONPARAMETRIC REGRESSION MODELING USING THE SPLINE APPROACH TO STUNTING CASES IN INDONESIA," *BAREKENG: J. Math. & App*., vol. 19, iss. 2, pp. 0697-0708, June, 2025.

1. INTRODUCTION

Toddler stunting is a condition where children under five (babies under five years) fail to grow so they are shorter than standard due to long-term malnutrition [1]. Indonesia is ranked fourth in the world and second in Southeast Asia in 2020. According to the Indonesian Nutrition Status Survey (SSGI) [2], the stunting rate in Indonesia reached 24.4% or around 23 million children in 2021, while in 2022 the figure fell to 21.6%. In accordance with the provisions of the World Health Organization (WHO) [3], the maximum tolerance standard for stunted toddlers is 20 percent or one fifth of the total number of toddlers, so the stunting rate in Indonesia is still relatively high. The high level of stunting in Indonesia can affect the quality of Indonesia's human resources, so early detection and immediate management of stunted toddlers are needed.

Regression analysis is a statistical method used to understand and model the relationship between two or more variables [4]. Regression analysis is often used to explain how changes in one variable can affect other variables, identify functional relationships between independent variables (x) and dependent variables (y), regression analysis helps better predictions and decision making [5][6].

Nonparametric regression is a method of regression analysis employed to identify the relationship between independent and dependent variables when there is no predetermined pattern or when the regression curve is not defined [7][8]. Along with the development of computing and several limitations in parametric regression models, non-parametric regression models which do not require many assumptions are becoming more widely used to solve problems in various applied fields [9]. The ability of nonparametric regression is supported by the existence of parameters in each type of nonparametric regression method which makes the estimation of the regression curve more flexible. One nonparametric regression method is spline [10]. Spline is a polynomial model that has segmented or discontinuous properties that are formed at each knot point and can produce a regression function that fits the data. The spline method is very good at modeling data that has changing patterns in certain sub-intervals.

The case of stunting is a complex malnutrition problem and is influenced by various factors, so the relationship between the incidence of stunting and the factors that influence the occurrence of stunting may not follow a linear or parametric pattern. Previous research on poverty modeling in East Java province used nonparametric spline regression [11]. The analysis was carried out on 5 variables, namely unemployment (x_1), population (x_1), literacy aged 15+ (x_3), average age of 15+ (x_4), health community development index (x_5) showed that all variables can explain the model so as to produce a good model. In this case, nonparametric spline regression is a relevant method in identifying and modeling the relationship between various independent factors and dependent factors. Nonparametric spline regression can identify and model the relationship between various independent factors and dependent factors more flexibly, allowing adjustments to data patterns and identifying significant variables. The nonparametric spline regression method can help provide insight into the factors that contribute to stunting, thereby enabling more effective decision making in planning health programs and policies [12]. This research aims to model the factors that influence stunting in Indonesia so that it provides benefits for the Indonesian community and government in knowing the factors that influence stunting and being able to take policies to reduce the level of stunting in Indonesia.

2. RESEARCH METHODS

2.1 Stunting

Stunting is a condition of failure to grow due to chronic malnutrition caused by inadequate nutritional intake for a long time, so that it is shorter than the standard. Body length or height index according to age (PB/U or TB/U) with nutritional status categories and thresholds (Z-Score) including very short (severely stunted) < -3 SD, short (stunted) -3 SD up to < -2 SD, normal -2 SD to +3 SD, high > +3 SD (Regulation of the Minister of Health of the Republic of Indonesia No. 2 of 2020 concerning Child Anthropometry Standards) [13]. According to the Decree of the Minister of Health of the Republic of Indonesia No. HK.01.07/MENKES/19282022 concerning National Guidelines for Medical Services for Stunting Management, there are many factors that influence the occurrence of stunting [12]. The following are several factors that influence stunting, including: toddlers receiving complete basic immunization, babies receiving exclusive breast milk for 6 months, babies born receiving Early Breastfeeding Initiation (EBI), children aged

6-23 months consuming five of the eight food and drink groups throughout the day [14], households having access to proper sanitation [15].

2.2 Regression Analysis

Regression analysis is a method used to understand and model the relationship between two or more variables. The variables used to determine the relationship are the dependent variable (y) and the independent variable (x) [16]. The general form, simple regression analysis model is written as Equation (1).

$$y = \beta_0 + \beta_i x_i + \varepsilon_i \tag{1}$$

2.3 Nonparametric Regression

Nonparametric regression has high flexibility to form cuts according to the data at hand. It can be seen that the data follows an estimated regression curve based on a piecewise function on the interval [17][18]. The nonparametric regression model is as follows:

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

2.4 Nonparametric Spline Regression

Nonparametric spline regression is a segmented polynomial that has flexibility properties. The spline function is highly dependent on the knot points. Knot points are certain points that are used to divide the independent variable into different segments in the spline model. In general, the spline function is of order q with knot points $k_1, k_2, \dots, k_r$ is defined as a function f on Equation (3) [19][20].

$$f(x_i) = \sum_{j=0}^{q} \beta_j x_i^j + \sum_{m=1}^{r} \beta_{q+m} (x_i - k_m)^q \tag{3}$$

If Equation (3) substituted in Equation (2) then the nonparametric spline regression equation is obtained as follows Equation (4)

$$y_i = \sum_{j=0}^{q} \beta_j x_i^j + \sum_{m=1}^{r} \beta_{q+m} (x_i - k_m)^q + \varepsilon_i \tag{4}$$

with function $(x_i - k_m)^q$ which is a piecewise function, is obtained Equation (4)

$$(x_i - k_m)^q = \begin{cases} (x_i - k_m)^q & , \quad x_i \geq k_m \\ 0 & , \quad x_i < k_m \end{cases} \tag{5}$$

One method for selecting optimal knots is to use GCV (Generalized Cross Validation). GCV is a method used to find the optimal knot point obtained from the minimum GCV. GCV can be used as a parameter selection to minimize Mean Squared Error (MSE) [21]. An MSE value that is low or close to zero indicates a good model, the MSE value can be used Equation (6)

$$MSE(k) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{f}(x_i))^2 \tag{6}$$

Next, the GCV value is obtained from Equation (7)

$$GCV(k_1, k_2, \dots, k_r) = \frac{MSE(k_1, k_2, \dots, k_r)}{(n^{-1} Trace[I - H(k_1, k_2, \dots, k_r)])^2} \tag{7}$$

2.5 Coefficient of Determination (R^2)

R^2 is a tool for measuring the proportion of diversity or total variation around the mean y value that is explained by the regression model [11]. The larger the value, the better the model obtained because it is able to explain a lot of data R^2 [22]. The equation obtained R^2 is

$$R^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} \times 100\% \quad (8)$$

2.6 Parameters Test

Parameter tests are used to obtain a regression model with optimal knot points based on the minimum GCV [23]. Parameter tests are useful for finding out whether nonparametric variables influence the response variable or not [24]. Testing has two stages, namely:

1. Simultaneous Test

   The simultaneous test is a simultaneous test of regression curve parameters using the F test. The hypothesis in the simultaneous test is as follows

   a. $H_0 : \beta_1 = \beta_2 = \cdots = \beta_{q+m} = 0$ (no independent variable has a significant influence on the dependent variable).

   b. $H_1 : \exists\beta : \beta_h \neq 0; h = 1,2, \ldots, (q+m)$ (at least one independent variable has a significant effect on the dependent variable).

2. Individual Test

   The individual test functions to determine individual parameters that have a significant influence on the response. The individual testing hypothesis is

   a. $H_0 : \beta_h = 0$ (the independent variable h does not have a significant effect on the dependent variable).

   b. $H_1 : \beta_h \neq 0; h = 1,2, \ldots, (q+m)$ (the h independent variable has a significant effect on the dependent variable).

3. RESULTS AND DISCUSSION

This chapter will discuss and analyze indicators that influence the number of stunting cases in Indonesia. The analysis in this chapter includes descriptive statistical analysis and stunting modeling. Modeling uses a nonparametric spline regression approach using one knot, two knots, three knots.

3.1 Descriptive Statistical Analysis

This descriptive statistical analysis presents the minimum, maximum, mean and variance values of each variable used. The variables used are the number of stunted toddlers (y), toddlers receiving complete basic immunization, (x_1), babies receiving exclusive breast milk for 6 months (x_2), babies born receiving Early Breastfeeding Initiation (EBI) (x_3), children aged 6-23 months consuming five of the eight food and drink groups throughout the day (x_4), households having access to proper sanitation(x_5). Descriptive statistical analysis in this research can be seen in Table 1.

Table 1. Descriptive Statistics

Variable	Minimum	Maximum	Mean	Variance
y	8.00	35.30	23.60	43.26514
x_1	22.52	83.89	62.18	182.9924
x_2	53.60	79.69	69.00	54.03829
x_3	48.87	73.59	62.88	40.01852
x_4	33.63	72.54	52.02	107.1303
x_5	40.34	96.21	81.00	95.7445

The descriptive statistical analysis examines key health indicators related to stunting in children under five, focusing on variables like immunization, breastfeeding, dietary diversity, and sanitation.

y (Stunted Toddlers) : Ranges from 8.00 to 35.30, with an average of 23.60 and high variance, indicating significant differences in stunting rates.

x_1 (Immunization) : Averages 62.18 with a wide range (22.52–83.89) and high variance, suggesting disparities in coverage.
x_2 (Exclusive Breastfeeding) : Averages 69.00 with moderate variance, showing relatively consistent practices.
x_3 (Early Breastfeeding/EBI) : Mean of 62.88 with low variance, indicating uniformity in early breastfeeding.
x_4 (Dietary Diversity) : Averages 52.02 with wide disparities in child nutrition, as shown by the high variance.
x_5 (Sanitation) : Mean of 81.00, with moderate differences in access to proper sanitation.
Overall, the data suggest areas for targeted intervention to reduce stunting.

3.2 Scatter Plot Pattern Analysis

Analysis of the relationship pattern between the number of stunted toddlers and each predictor variable can be done using a scatter plot. Scatter plot analysis was carried out before modeling. The scatter plots produced in this research have a random pattern or do not form a particular pattern. This identifies the method used is spline nonparametric regression for model estimation.

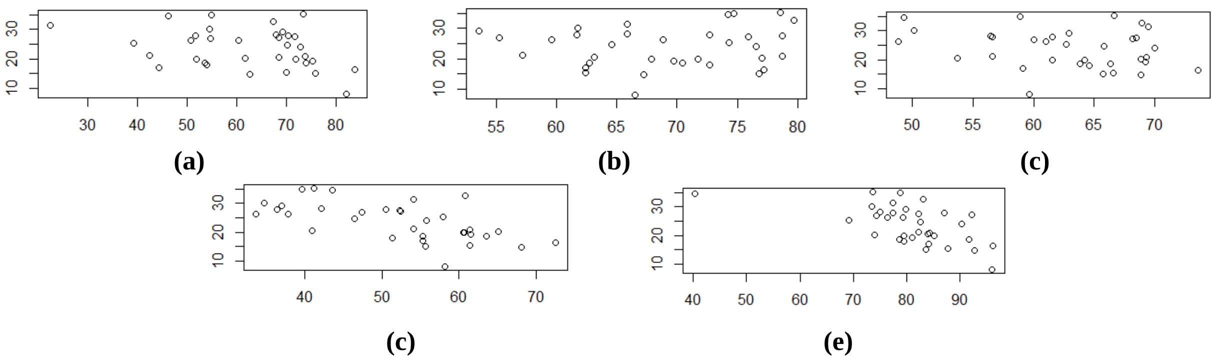

Figure 1. Scatter Plots between Dependent and Independent Variables, (a) Scatter Plot of y vs x_1, (b) Scatter Plot of y vs x_2, (c) Scatter Plot of y vs x_3, (d) Scatter Plot of y vs x_4, (e) Scatter Plot of y vs x_5

3.3 Modeling

The selection of optimum knot points in nonparametric spline regression was carried out in order to obtain the best model for this research. The knot points used in this research are one knot, two knots, three knots, and a combination of knots. The optimum knot point is obtained from the minimum GCV value. The nonparametric spline regression model with one knot on stunting data in Indonesia in 2022 is as follows.

$$\begin{aligned}\hat{y} = \ & \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 (x_1 - k_1) + \hat{\beta}_3 x_2 + \hat{\beta}_4 (x_2 - k_2) + \hat{\beta}_5 x_3 \\ & + \hat{\beta}_6 (x_3 - k_3) + \hat{\beta}_7 x_4 + \hat{\beta}_8 (x_4 - k_4) + \hat{\beta}_9 x_5 + \hat{\beta}_{10} (x_5 - k_5)\end{aligned} \quad (9)$$

The GCV value of the one-knot spline nonparametric regression model is presented in Table 2.

Table 2. Knot Points and GCV for Single Knot Point Splines

GCV	x_1	x_2	x_3	x_4	x_5
19.47740	23.77245	54.13245	49.37449	34.42408	41.48020
22.35994	25.02490	54.66490	49.87898	35.21816	42.62041
⋮	⋮	⋮	⋮	⋮	⋮
14.32605	**80.13265**	**78.09265**	**72.07653**	**70.15776**	**92.78939**
14.53342	81.38510	78.62510	72.58102	70.9514	93.92959
14.68320	82.63755	79.15755	73.08551	71.4592	95.06980

Estimation of nonparametric regression models splines by using two knot points in modeling the percentage of stunted toddlers in Indonesia is modeled in Equation (10).

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2(x_1 - k_1) + \hat{\beta}_3(x_1 - k_2) + \hat{\beta}_4 x_2 + \hat{\beta}_5(x_2 - k_3) + \hat{\beta}_6(x_2 - k_4) + \hat{\beta}_7 x_3 + \hat{\beta}_8(x_3 - k_5) + \hat{\beta}_9(x_3 - k_6) + \hat{\beta}_{10} x_4 + \hat{\beta}_{11}(x_4 - k_7) + \hat{\beta}_{12}(x_4 - k_8) + \hat{\beta}_{13} x_5 + \hat{\beta}_{14}(x_5 - k_9) + \hat{\beta}_{15}(x_5 - k_{10}) \quad (10)$$

The GCV value of the two-knot spline nonparametric regression model is presented in **Table 3**.

Table 3. Knot Points and GCV for Two Knot Point Splines

GCV	x_1	x_2	x_3	x_4	x_5
22.79799	23.77245 25.0249	54.13245 54.6649	49.37449 49.87898	34.42408 35.21816	41.4802 42.62041
⋮	⋮	⋮	⋮	⋮	⋮
14.66208	**23.77245** **81.3851**	**54.13245** **78.6251**	**49.37449** **72.58102**	**34.42408** **70.95184**	**41.4802** **93.92959**
⋮	⋮	⋮	⋮	⋮	⋮
15.73442	81.3851 82.63755	78.6251 79.15755	72.58102 73.08551	70.95184 71.74592	93.92959 95.0698

Estimation of a spline nonparametric regression model using three knot points in modeling the percentage of stunted toddlers in Indonesia is modeled in **Equation (11)**.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2(x_1 - k_1) + \hat{\beta}_3(x_1 - k_2) + \hat{\beta}_4(x_1 - k_3) + \hat{\beta}_5 x_2 + \hat{\beta}_6(x_2 - k_4) + \hat{\beta}_7(x_2 - k_5) + \hat{\beta}_8(x_2 - k_6) + \hat{\beta}_9 x_3 + \hat{\beta}_{10}(x_3 - k_7) + \hat{\beta}_{11}(x_3 - k_8) + \hat{\beta}_{12}(x_3 - k_9) + \hat{\beta}_{13} x_4 + \hat{\beta}_{14}(x_4 - k_{10}) + \hat{\beta}_{15}(x_4 - k_{11}) + \hat{\beta}_{16}(x_4 - k_{12}) + \hat{\beta}_{17} x_5 + \hat{\beta}_{18}(x_5 - k_{13}) + \hat{\beta}_{19}(x_5 - k_{14}) + \hat{\beta}_{20}(x_5 - k_{15}) \quad (11)$$

The GCV value of the three-knot spline nonparametric regression model is presented in **Table 4**.

Table 4. Knot Points and GCV for Three Knot Point Splines

GCV	x_1	x_2	x_3	x_4	x_5
22.79799	23.77245 25.0249 26.27735	54.13245 54.6649 55.19735	49.37449 49.87898 50.38347	34.42408 35.21816 36.01225	41.4802 42.62041 43.76061
⋮	⋮	⋮	⋮	⋮	⋮
15.03168	**45.06408** **75.12286** **82.63755**	**63.18408** **75.96286** **79.15755**	**57.95082** **70.05857** **73.08551**	**47.92347** **66.98143** **71.74592**	**60.86367** **88.22857** **95.0698**
⋮	⋮	⋮	⋮	⋮	⋮
16.26281	80.13265 81.3851 82.63755	78.09265 78.6251 79.15755	72.07653 72.58102 73.08551	70.15776 70.95184 71.74592	92.78939 93.92959 95.0698

Based on **Table 2** until **Table 4**, the optimum knot point will be selected based on the minimum GCV value and a comparison of GCV values will be obtained as follows **Table 5**.

Table 5. Comparison of GCV Values

Model		GCV	R^2
1 knot	**:**	**14.32605**	**81.55%**
2 knots	:	14.66208	86.67%
3 knots	:	15.03168	92.07%

3.4 Significance Test of Spline Nonparametric Regression Model Parameters

The best nonparametric spline regression model is known from the nonparametric spline regression model with 1 knot, then testing the significance of the nonparametric spline regression model parameters is carried out. The significance test of the regression model parameters was carried out to determine whether the predictor variables had a significant influence on the percentage of stunted toddlers in Indonesia.

1. Simultaneous Testing

Simultaneous or simultaneous testing is testing the significance of all parameters contained in the model as a whole. Testing the significance of model parameters simultaneously for a model with one knot has the following hypothesis:

$H_0 : \beta_0 = \beta_1 = \beta_2 = \cdots = \beta_{10} = 0$

$H_1 : \exists\beta : \beta_h \neq 0; h = 1,2, \dots, 10$

The following are the results of simultaneous parameter significance tests presented in **Table 6**.

Table 6. ANOVA Results of Simultaneous Testing

Sources of Variation	Df	SS	M.S	$F_{statistic}$	p-value
Regression	10	1157.838	115.7838		
Error	23	269.9117	7.938579	4.58495	1.023534×10^{-7}
Total	33	1427.75			

Based on the values in **Table 6**, it can be seen that $F_{statistic} = 4.58495$ and the p-value = 1.023534×10^{-7} with a $F_{(0.05;10;23)} = 2.275$ and significance level of 0.05 indicates that $F_{statistic} = 4.58495$ is greater than $F_{table} = 2.275$ while the p-value is less than the significance level so it was decided to reject H_0, which means that simultaneously the variable toddlers receiving complete basic immunization, babies receiving exclusive breast milk for 6 months, babies born receiving Early Breastfeeding Initiation (EBI), children aged 6-23 months consuming five of the eight food and drink groups throughout the day, and the households having access to proper sanitation, which influences the percentage of stunting in Indonesia in 2022. The model in this research obtained R^2 or a coefficient of determination of 81.1%. This value shows that the five variables in the research have an influence of 81.1% on the percentage of stunted toddlers in Indonesia, so the model is quite good.

2. Individual Testing

Individual testing aims to determine variables that have a significant influence on the percentage of stunting in Indonesia in 2022.

$H_0 : \beta_h = 0$

$H_1 : \beta_h \neq 0; h = 1,2, \dots, 10$

The results of individual testing show that the expected variables influence the percentage of stunted toddlers in Indonesia. **Table 7** shows partial test results for each parameter.

Table 7. Partial Test Results

Variable	Parameter	Estimate	$t_{statistic}$	p-value	Decision
Constant	β_0	41.50137383	5.892513	5.252×10^{-6}	Significant
x_1	β_1	-0.08860576	-1.943192	6.433×10^{-2}	No
	β_2	0.14025499	1.568889	1.303×10^{-1}	No
x_2	β_3	0.39107847	2.824020	9.623×10^{-3}	Significant
	β_4	-0.53491455	-7.615098	9.889×10^{-8}	Significant
x_3	β_5	-0.23628512	-3.144718	4.539×10^{-3}	Significant
	β_6	5.57883418	2.216802	3.680×10^{-2}	Significant
x_4	β_7	20.94161693	3.730068	1.097×10^{-3}	Significant
	β_8	2.24716931	2.216802	3.680×10^{-2}	Significant
x_5	β_9	3.53710995	2.216802	3.680×10^{-2}	Significant
	β_{10}	-5.84431945	-1.541874	1.368×10^{-1}	No

Based on **Table 7** There are 11 parameters in the nonparametric spline regression model with a confidence level of 95%, if the p-value is less than α, namely 0.05, then the parameter estimate is significant. In total, there were 8 significant parameters and 3 parameters that were not significant out of 11 parameters, so that 4 variables were thought to be influential, namely the variables of babies receiving exclusive breast milk for 6 months, babies born receiving Early Breastfeeding Initiation (EBI), children aged 6-23 months consuming five of the eight food and drink groups throughout the day, and the households having access to proper sanitation has an effect on the percentage of stunting in Indonesia and 1 variable is thought to have no significant effect, namely toddlers receiving complete basic immunization variable.

3.5 Model Interpretation

The best model with minimum GCV is met, then model interpretation is carried out. The spline nonparametric regression model is formed using the optimal knot point, namely one knot point with **Equation (10)** and the resulting model interpretation:

$$\hat{y} = 41.50137383 - 0.08860576x_1 + 0.14025499(x_1 - 80.13265) + 0.39107847x_2 - 0.53491455(x_2 - 78.09265) - 0.23628512x_3 + 5.57883418(x_3 - 72.07653) + 20.94161693x_4 + 2.24716931(x_4 - 70.15776) + 3.53710995x_5 - 5.84431945(x_5 - 92.78939)$$

Model interpretation for significant variables was carried out to determine their influence on the percentage of stunted toddlers in Indonesia with significant variables. Based on this model, it can be interpreted as follows:

1. If the variables x_1, x_3, x_4, and x_5 are considered constant, then the significance of the effect of babies receiving exclusive breast milk for 6 months (x_2) on the percentage of stunted toddlers is as follows:

$$\hat{y} = 0.39107847x_2 - 0.53491455(x_2 - 78.09265)$$
$$= \begin{cases} 0.39107847x_2 & ; \quad x_2 < 78.09265 \\ 41.772895 - 0.53491455\, x_2 & ; \quad x_2 \geq 78.09265 \end{cases} \quad (12)$$

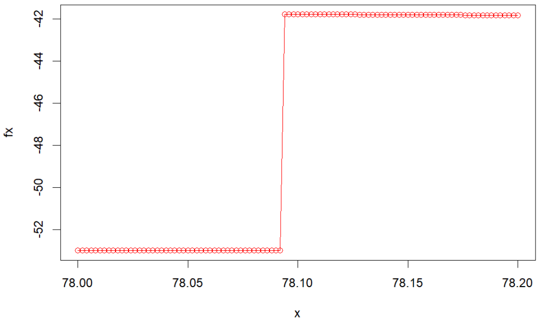

Figure 2. **Graphic of Babies Getting Exclusive Breast Milk**

2. If the variables x_1, x_2, x_4, and x_5 are considered constant, then the significance of the effect of babies born receiving Early Breastfeeding Initiation (EBI) (x_3) on the percentage of children with stunting is as follows:

$$\hat{y} = -0.23628512x_3 + 5.57883418(x_3 - 72.07653)$$
$$= \begin{cases} -0.23628512x_3 & ; \quad x_3 < 72.07653 \\ -402.103009 + 5.57883418\, x_3 & ; \quad x_3 \geq 72.07653 \end{cases} \quad (13)$$

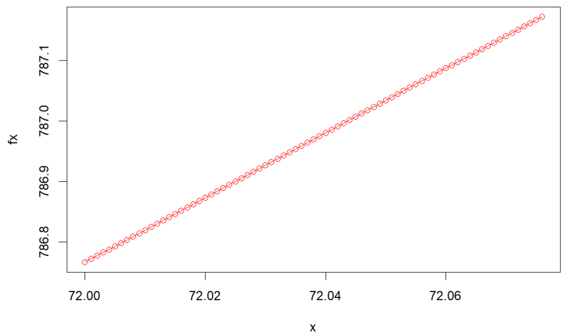

Figure 3. **Early Initiation of Breastfeeding**

3. If the variables x_1, x_2, x_3, dan x_5 are considered constant, then the significance of the effect of children aged 6-23 months consuming five of the eight food and drink groups throughout the day (x_4) on the percentage of stunted toddlers is as below:

$$\hat{y} = 20.94161693x_4 + 2.24716931(x_4 - 70.15776) \\ = \begin{cases} 20.94161693x_4 & ; \quad x_4 < 70.15776 \\ -157.65636 + 2.24716931\, x_4 & ; \quad x_4 \geq 70.15776 \end{cases} \tag{14}$$

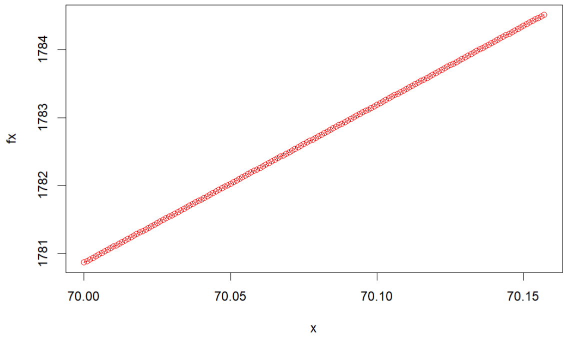

Figure 4. Chart Consume 5 of 8 Food and Drink Groups

4. If the variable x_1, x_2, x_3, dan x_4 are considered constant, then the significance of the effect of households having access to proper sanitation (x_5) on the percentage of stunted toddlers as below:

$$\hat{y} = 3.53710995x_5 - 5.84431945(x_5 - 92.78939) \\ = \begin{cases} 3.53710995x_5 & ; \quad x_5 < 92.78939 \\ 542.2908 - 5.84431945\, x_5 & ; \quad x_5 \geq 92.78939 \end{cases} \tag{15}$$

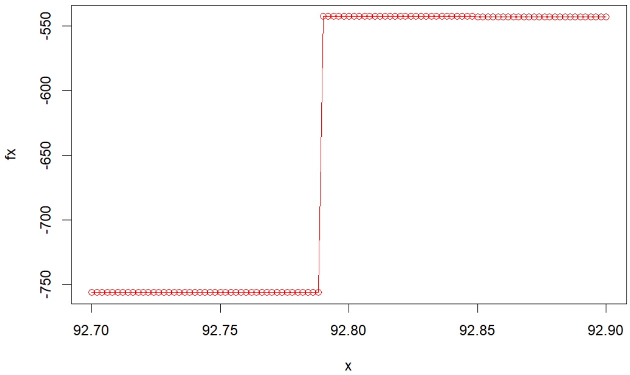

Figure 5. Chart of Houses with Proper Sanitation

3.5 Graph of the Percentage of Stunting Toddlers in Indonesia

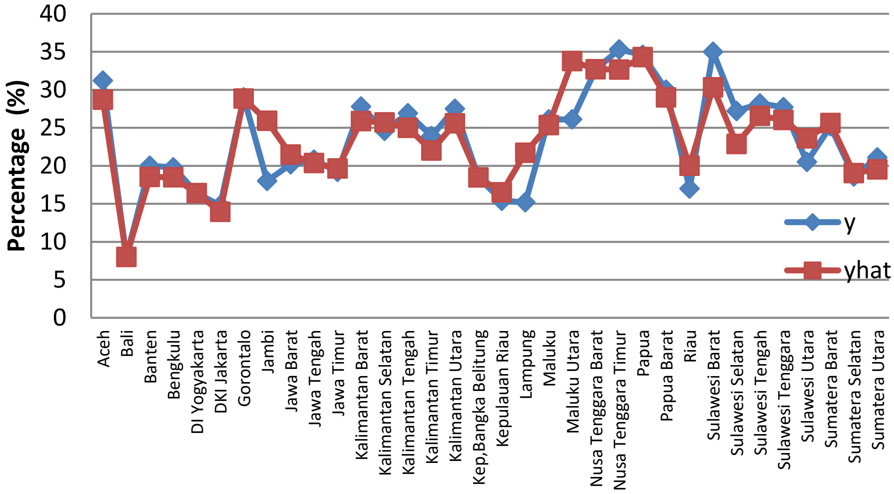

Figure 6. Graph of the Percentage of Stunting Toddlers in Indonesia

The graph above shows the difference between the actual percentage and the estimated percentage. There are 29 provinces where the actual percentage and the estimated percentage are the same or have a small difference, while several provinces that have different percentages include: Jambi, Lampung, North Maluku, West Sulawesi and South Sulawesi. From the explanation above, the nonparametric spline regression method is the appropriate method to use in this research because this method is able to model nonlinear relationships between response variables and predictor variables.

4. CONCLUSIONS

Based on the results of the analysis and discussion, it can be concluded that nonparametric spline regression modeling with the best model of one knot obtained a GCV of 14.32605 and value of R^2 is 81.1%. Of five independent variables, namely toddlers receiving complete basic immunization (x_1), babies receiving exclusive breast milk for 6 months (x_2), babies born receiving Early Breastfeeding Initiation (EBI) (x_3), children aged 6-23 months consuming five of the eight food and drink groups throughout the day (x_4), households having access to proper sanitation (x_5). The variable that does not have a significant effect is toddlers receiving complete basic immunization, (x_1). The lowest percentage of cases of stunting under five in Indonesia in 2022 using nonparametric spline regression is 8.00% in Bali Province, while the highest percentage is 34.29% in Papua Province. The nonparametric spline regression model obtained is

$$\begin{aligned}\hat{y} = \ & 41.50137383 + 0.39107847x_2 - 0.53491455(x_2 - 78.09265) - 0.23628512x_3 \\ & +5.57883418(x_3 - 72.07653) + 20.94161693x_4 + 2.24716931(x_4 - 70.15776) \\ & +3.53710995x_5 - 5.84431945(x_5 - 92.78939)\end{aligned}$$

REFERENCES

[1] M. R. Nugroho, R. N. Sasongko, and M. Kristiawan, "FAKTOR-FAKTOR YANG MEMPENGARUHI KEJADIAN STUNTING PADA ANAK USIA DINI DI INDONESIA," *J. Obs. J. Pendidik. Anak Usia Dini*, vol. 5, no. 2, pp. 2269–2276, 2021, doi: 10.31004/obsesi.v5i2.1169.

[2] Kementerian Kesehatan Republik Indonesia, "BUKU SAKU : HASIL SURVEI STATUS GIZI INDONESIA (SSGI) 2022," *Kementeri. Kesehat. Republik Indones.*, pp. 1–7, 2023.

[3] World Health Organization, *World Health Organization. Reducing stunting in children: equity considerations for achieving the Global Nutrition Targets 2025. World Health Organization; 2018.* 2018.

[4] A. Efendi, *Microsoff Excel : PENGOLAHAN DAN ANALISIS DATA + CD*. Penerbit Salemba.

[5] R. Kurniawan and B. Yuniarto, *ANALISIS REGRESI DASAR DAN TERAPANNYA DENGAN R*, 1st ed. Jakarta: Kencana, 2016.

[6] A. Efendi, N. Wardhani, R. Fitriani, and E. Sumarminingsih, *ANALISIS REGRESI : TEORI DAN APLIKASI DENGAN R*.

Malang: Universitas Brawijay Press, 2020.

[7] A. Fernandes and Solimun, *METODE ANALISIS DATA PENELITIAN*. Malang: Universitas Brawijay Press, 2022.

[8] N. Chamidah and B. Lestari, *ANALISIS REGRESI NONPARAMETRIK DENGAN PERANGKAT LUNAK R*. Surabaya: Airlangga University Press, 2022.

[9] Zulfikar and I. N. Budiantara, *MANAJEMEN RISET DENGAN PENDEKATAN KOMPUTASI STATISTIKA*. YOGYAKARTA: DEEPUBLISH, 2015.

[10] R. EUBANK, *SPLINE SMOOTHING AND NONPARAMETRIC REGRESSION*. New York: Camridge University Press, 1988.

[11] M. K. Dewi, “PEMODELAN KEMISKINAN DI PROVINSI JAWA TIMUR MENGGUNAKAN REGRESI NONPARAMETRIK SPLINE,” Jember, 2022.

[12] Kemenkes RI, “KEMENKES RI NO HK.01.07/MENKES/1928/2022 TENTANG PEDOMAN NASIONAL PELAYANAN KEDOKTERAN TATA LAKSANA STUNTING,” 2022.

[13] Kemenkes RI, “KEMENKES RI NO 2 TAHUN 2020 TENTANG STANDART ANTROPOMETRI ANAK,” 2020, doi: 10.7476/9788575415894.0004.

[14] Badan Pusat Statistik, “PROFIL KESEHATAN IBU DAN ANAK TAHUN 2022,” Jakarta: BPS, 2022.

[15] Badan Pusat Statistik, “SANITASI LAYAK,” *BPS*, 2022. https://bps.go.id/subject/29/perumahan.html#subjekViewTab1

[16] R. E. Walpole, *PENGANTAR STATISTIKA*, Edisi ke 3. Jakarta: PT. Gramedia Pustaka Utama, 1995.

[17] R. Eubank, *NONPARAMETRIC REGRESSION AND SPLINE SMOOTING*, 2nd ed. New York: Marcel Dekker, 1999.

[18] H. Wu and J.-T. Zhang, *NONPARAMETRIC REGRESSION METHODS FOR LONGITUDINAL DATA ANALYSIS*. United States: Wiley Interscience, 2006.

[19] I. Sriliana, I. N. Budiantara, and V. Ratnasari, “THE PERFORMANCE OF MIXED TRUNCATED SPLINE-LOCAL LINEAR NONPARAMETRIC REGRESSION MODEL FOR LONGITUDINAL DATA,” *MethodsX*, vol. 12, no. March, 2024, doi: 10.1016/j.mex.2024.102652.

[20] R. Putra, M. G. Fadhlurrahman, and Gunardi, “DETERMINATION OF THE BEST KNOT AND BANDWIDTH IN GEOGRAPHICALLY WEIGHTED TRUNCATED SPLINE NONPARAMETRIC REGRESSION USING GENERALIZED CROSS VALIDATION,” *MethodsX*, vol. 10, no. December 2022, p. 101994, 2023, doi: 10.1016/j.mex.2022.101994.

[21] A. Sayuti, D. Kusnandar, and M. N. Mara, “GENERALIZED CROSS VALIDATION DALAM REGRESI SMOOTHING SPLINE,” *Bul. Ilm. Mat. Stat. dan Ter.*, vol. 02, no. 3, pp. 191–196, 2013.

[22] S. Melinda, Sudarmin, and M. Nusrang, “PENDEKATAN REGRESI NONPARAMETRIK SPLINE TRUNCATED PADA INDEKS PEMBANGUNAN MANUSIA DI SULAWESI SELATAN,” *VARIANSI J. Stat. Its Appl. Teach. Res.*, vol. 3, no. 2, pp. 83–89, 2021, doi: 10.35580/variansiunm23860.

[23] W. Sanusi, R. Syam, and R. Adawiyah, “MODEL REGRESI NONPARAMETRIK DENGAN PENDEKATAN SPLINE (STUDI KASUS: BERAT BADAN LAHIR RENDAH DI RUMAH SAKIT IBU DAN ANAK SITI FATIMAH MAKASSAR),” *J. Math. Comput. Stat.*, vol. 2, no. 1, p. 70, 2020, doi: 10.35580/jmathcos.v2i1.12460.

[24] 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.