Modeling Stunting Prevalence in Indonesia : Mixed Spline Truncated and Fourier Series Nonparametric Regression

Hartina Husain[1*], Irmayani[2], and Andi Oxy Raihan Machikami Rahman[1]
[1]Data Science Department, Bacharuddin Jusuf Habibie Institute of Technology, Parepare, Indonesia
[2]Actuarial Science Department, Bacharuddin Jusuf Habibie Institute of Technology, Parepare, Indonesia
*Corresponding author: hartinahusain@ith.ac.id

Received: 31 May 2024
Revised: 2 September 2024
Accepted: 4 November 2024

ABSTRACT – **Stunting is a condition of failure to grow in children that occurs due to malnutrition chronic so that the child's height is shorter compared to his age. This research aims to model the factors that influence the prevalence of stunting in Indonesia based on a literature study using mixed spline truncated and fourier series nonparametric regression method. Data used is secondary data regarding the prevalence of stunting and several suspected factors influencing it, namely the percentage of the population with health insurance and the percentage of the population who smoked last month (Age ≥ 15 Years). Data was sourced from publications from the Ministry of Health and Badan Pusat Statistik (BPS) in 2022. The results show that the model combines a spline truncated component with one knot and a fourier series component with one oscillation , resulting in a minimum Generalized Cross Validation (GCV) Value of 34.46 and an Mean Square Error (MSE) of 4.89.**
Keywords – **Fourier Series, GCV, Spline Truncated, Stunting.**

I. INTRODUCTION

The world is dealing with a severe issue of stunting, particularly in poor and developing countries including Indonesia [1]. The Indonesian government is focusing on stunting as part of its national stunting prevention plan. Stunted children are those who lack height due to malnutrition. As a result, the child is shorter than average for their age. Indonesia is working to reduce the incidence of stunting by international goals which is the World Health Assembly's (WHA) commitment to reduce the prevalence of stunting by 40% by 2025. In keeping with that, the Sustainable Development Goals (SDGs) aim to eradicate any malnutrition by the year 2030. The prevalence of stunting in children under five must decrease to 19.4%. In 2024 to meet this target, which calls for measures to be undertaken to accelerate the reduction of stunting [2].

The Ministry of Health reports that in 2022, the stunting rate was 21.6%, down from 24.4% in 2021 [3]. The government conveyed through the Coordinating Ministry for Human Development and Cultural Affairs Indonesia targeting every year the stunting rate decreases by 3%. At this moment, the Indonesian government is currently concentrating on the issue of stunting. Stunting is a condition that is caused by many factors [4]. The impact of stunting on children can affect them in the short term as well as in the future. Short-term effects include disruptions or harm to the development of children's brains, a decrease in Intelligence Quotient, and a compromised immune system that makes a kid more vulnerable to illness or infection [5]. Therefore, a stunting treatment is needed based on an understanding of what factors can affect the prevalence of stunting.

Previous studies on stunting have frequently been conducted by other researchers using a variety of statistical techniques. One of them is research on stunting with Geographically Weighted Regression (GWR) was used to investigate spatial variation in stunting cases in NTT Province, identifying key factors such as the percentage of pregnant women at risk of Chronic Energy Deficiency (CED), the percentage of babies fully immunized, the percentage of women who graduated from Junior College, the percentage of poor people, the percentage of babies exclusively breastfed, and the percentage of underage women who have ever married [6]. While GWR is effective in capturing spatial heterogeneity, its limitation lies in its inability to handle complex nonlinear relationships between variables, especially when the form of these relationships is unknown or irregular. This creates the need for more flexible models when the exact nature of the data pattern is uncertain.

Other research on stunting has also been conducted with the spline truncated nonparametric regression method [7][8]. The results obtained from this study show that the best model obtained by the spline truncated nonparametric regression method with a combination of the number of knots 3,1,2,2,3 on each variable that affects it. The five independent variables include the percentage of children under five who are fully immunized, the percentage of pregnant women at risk of CED, the percentage of children under five who are breastfed for 6 months, the percentage of households with proper sanitation and the percentage of babies born receiving initiate early breastfeeding, which all have a significant effect on the model. Furthermore, the coefficient of determination obtained from the model is 80.77% indicating its effectiveness in explaining the variation in stunting prevalence [7].

The spline truncated approach has inherent limitations, particularly in modeling data that exhibit underlying periodic patterns. The relationships between predictor variables in stunting prevalence may encompass diverse and recurring patterns. While the spline truncated method is adept at capturing localized variations and non-linear fluctuations, it may not adequately address these cyclical trends. To overcome this limitation, a mixed model approach that integrates both spline truncated and Fourier series estimators provides a more comprehensive solution. The spline truncated estimator effectively captures non-linear, irregular variations within the data, whereas the Fourier series estimator is specifically designed for analyzing periodic, cyclical patterns. By combining these two methodologies, the mixed model enhances the ability to capture both short-term variations and long-term cycles, leading to a more precise understanding of the factors influencing stunting prevalence. Furthermore, the mixed model approach is inherently superior to using a single estimator, as it leverages the strengths of both methods, thereby offering a more robust analysis that accounts for the complexities and nuances present in the data.

II. LITERATURE REVIEW

A. Nonparametric Regression Analysis

Regression analysis is a statistical method used to investigate the relationship between a dependent variable and one or more independent variables. There are several types of regression, including parametric, nonparametric, and semiparametric. If the data pattern is clear (e.g., linear, quadratic, or cubic), then parametric regression can be used. However, not all data have a clear pattern. In such cases, nonparametric regression models the relationship between the response and predictor variables [9]. The general formula for a nonparametric regression model is:

$$y_i = f(x_i) + \varepsilon_i \, , i = 1,2,3 \cdots n \, , \varepsilon_i \sim iidN(0, \sigma^2) \tag{1}$$

Where

y_i : Dependent variable

$f(x_i)$: Nonparametric regression function

In nonparametric regression, researchers have the freedom to choose the analysis method that best suits the pattern of the data they have. This is in contrast to parametric regression, which is constrained to a specific functional form. Several commonly used estimator approaches include Kernel, Local Polynomial, Spline, Histogram, Fourier Series, Orthogonal Series, k-NN, MARS, Neural Network (NN), Wavelets, and other approaches.

Spline nonparametric regression is a suitable estimation technique for dealing with data that exhibits varying behavioral patterns over specific sub-intervals. The spline nonparametric regression model can be expressed as in Equation (1), where $f(x_i)$ represents the regression curve approximated by a spline function of order/degree p Spline functions are highly dependent on knot points, which represent the locations of pattern changes in the data [10].Therefore, knot points are specified in the spline regression equation $K_1, K_2, \dots, K_M$. The expression for is given by the following equation:

$$f(x_i) = \sum_{j=0}^{p} \xi_j x_i^j + \sum_{k=0}^{M} \Phi_k (x_i - K_k)_+^p \tag{2}$$

Where is a truncated function given by:

$$(x_i - K_k)_+^p = \begin{cases} (x_i - K_k)_+^p & ; \ x_i \geq K_k \\ 0 & ; \ x_i < K_k \end{cases}$$

where $\xi_0, \xi_1, \cdots, \xi_p, \Phi_1, \cdots, \Phi_m$ are parameter of spline truncated function, $i = 1,2, \cdots, n$ denotes the amount of data and K_k with $k = 1,2, \cdots, M$ in denotes knots.

The fourier series is a trigonometric polynomial function that has levels of flexibility to deal with data that has repetitive patterns. The Fourier series is used to estimate the curve regression showing sinus and cosinus waves [11]. The nonparametric approximation function using the fourier series can be written as follows:

$$g(t_i) = bt_i + \frac{\alpha_0}{2} + \sum_{l=1}^{L} \alpha_l \cos lt_i \tag{3}$$

Where b, α_0, α_l is parameter of fourier series model and $l = 1,2, \cdots, L$ is oscillation parameters.

B. Generalized Cross Validation

Getting the optimal spline function depends on selection knot points K which are a combination of changes function at different intervals. The fourier series depends on the parameters oscillation l which shows the number of oscillations of the wave cosinus in the model. Values of K and l that are too small will produce a curve that is under smoothing is very rough and very volatile. If a value that is too large is obtained, it will result in an over-curve smoothing that is very smooth but does not match the data pattern [12].

In various nonparametric regression studies, a common method for selecting optimal knot points and oscillation is Generalized Cross Validation (GCV). The general formula for GCV in nonparametric regression is given below [10]:

$$GCV(K,l) = \frac{MSE(k,l\,)}{(n^{-1}trace[\boldsymbol{I} - \boldsymbol{A}(k,l)])^2} \quad (4)$$

and

$$MSE\ (k,l) = \frac{(\tilde{y} - \hat{\tilde{y}})^T(\tilde{y} - \hat{\tilde{y}})}{n}$$

C. Stunting Prevalence

Stunted growth in toddlers refers to their inability to grow optimally due to prolonged malnutrition. This condition leads to children being shorter than the standards set by the World Health Organization (WHO) (Kemenkes RI, 2013). Short stature in toddlers indicates chronic malnutrition experienced by mothers or mothers-to-be during pregnancy, the fetal period, or infancy, as well as other factors that indirectly affect health [13].

Toddlers require special attention in various aspects to prevent stunting, which can hinder their mental development, physical growth, and overall health [14]. Their growth and development can be assessed through various nutritional status indicators. A common anthropometric standard for evaluating children's nutritional status is based on their height-for-age or length-for-age z-scores. Children with a z-score between -3 and <-2 are considered stunted, while those with a z-score below -3 are considered severely stunted [3].

III. METHODOLOGY

This research method involves a literature review on mixed truncated spline nonparametric regression estimation and Fourier series. The estimation results are then applied to stunting prevalence data in Indonesia. The data used is secondary data for 2022, obtained from publications of the Ministry of Health and Badan Pusat Statistik (BPS). The data type is cross-sectional, consisting of 34 provinces in Indonesia. This study uses one response variable and two predictor variables as shown in Table 1

Table 1 **Research Variables**

Variables	Description
Y	Stunting Prevalence
X_1	Percentage of the population who has health insurance
X_2	Percentage of the population who smoked last month (Age ≥ 15 Years).

The analysis steps involved in this study are as follows:

1. Obtaining the estimation of the mixed nonparametric regression model Spline Truncated and Fourier Series with the following steps:
   a. Given the response variable Y, the nonparametric spline truncated $x_1, x_2, \cdots, x_p$ component f(x), and the Fourier series component $t_1, t_2, \cdots, t_q$
   b. Approximate the function $f(x_{pi})$ with a truncated linear spline function with M knots.
   c. Approximate the function $g(t_{qi})$ and the Fourier series.
   d. Express the curve $\tilde{f}(x), and\ \tilde{g}(t)$ in matrix form where $\tilde{\xi}$ truncated spline regression parameters and $\tilde{\alpha}$ Fourier series regression parameters are present.
   e. Expressing the mixed nonparametric regression of truncated spline and Fourier series in matrix form.
   f. Solving the optimization problem using the Ordinary Least Squares (OLS) method to find $\hat{\tilde{\xi}}$ and $\hat{\tilde{\alpha}}$.
2. Applying the obtained mixed estimation to model stunting prevalence data in Indonesia using the following steps:
   a. Conducting data exploration and descriptive analysis of each research variable
   b. Creating scatter plots of the data to understand the shape of the relationship pattern between the response and predictors
   c. Determining predictor variables using truncated spline regression curves and Fourier series regression curves with the minimum GCV criterion.
   d. Modeling stunting prevalence data with mixed nonparametric regression of truncated spline and Fourier series
   e. Selecting optimal knot points, oscillation parameters, and bandwidth using the minimum GCV criterion.
   f. Calculating the coefficient of determination (R-squared) for the obtained model
   g. Create conclusions

IV. RESULTS AND DISCUSSIONS

A. *Mixed of Spline Truncated and Fourier Series Nonparametric Regression Model*

Given data with response variable Y and predictor variable consisting of nonparametric components of a spline truncated x_1, $x_2, \cdots, x_P$ and fourier series $t_1, t_2, \cdots, t_Q$. The form of a nonparametric regression model that contains these variables can be expressed as follows :

$$y_i = \sum_{p=1}^{P} f(x_{pi}) + \sum_{q=1}^{Q} g(t_{qi}) + \varepsilon_i \tag{5}$$

Where $i = 1, 2, \cdots, n$

ε_i assumed to be independent, identical, and normally distributed with zero mean and variance σ_i^2. The function $f(x_{pi})$ with p variables is approximated by a spline truncated function, namely :

$$f(x_{pi}) = \xi_0 + \xi_p x_{pi} + \sum_{k=1}^{M} \Phi_{pk}(x_{pi} - K_{pk})_+ \tag{6}$$

Where M represents the number of knots with a spline truncated function given by :

$$(x_{pi} - K_{pk})_+ = \begin{cases} (x_{pi} - K_{pk})_+ & ; x_{pi} \geq K_{pk} \\ 0 & ; x_{pi} < K_{pk} \end{cases}$$

It can be expressed in matrix form as:

$$\begin{pmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & (x_{11} - K_{11})_+ & \cdots & (x_{P1} - K_{PM})_+ \\ 1 & x_{12} & (x_{12} - K_{11})_+ & \cdots & (x_{P2} - K_{PM})_+ \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & (x_{1n} - K_{11})_+ & \cdots & (x_{Pn} - K_{PM})_+ \end{pmatrix} \begin{pmatrix} \xi_0 \\ \xi_1 \\ \Phi_{11} \\ \vdots \\ \Phi_{PM} \end{pmatrix}$$

Or in matrix notation it can be written as :

$$\tilde{f}(x) = \boldsymbol{G}(\boldsymbol{x})\tilde{\xi} \tag{7}$$

With $\tilde{f}(x)$ of size $n \times 1$, Matrix $\boldsymbol{G}(\boldsymbol{x})$ of size $n \times (2P + PM)$ and $\tilde{\xi}$ of size $(2P + PM) \times 1$.
Next, $g(t_{qi})$ with variable q is approximated by a Fourier Series, namely :

$$q(t_{qi}) = b_q t_{qi} + \frac{\alpha_{0q}}{2} + \sum_{l=1}^{L} \alpha_{lq} \cos l t_{qi} \tag{8}$$

Or in matrix notation it can be written as :

$$\begin{pmatrix} g(t_1) \\ g(t_2) \\ \vdots \\ g(t_n) \end{pmatrix} = \begin{pmatrix} t_{11} & \frac{1}{2} & \cos t_{11} & \cdots & \cos L t_{Q1} \\ t_{12} & \frac{1}{2} & \cos t_{12} & \cdots & \cos L t_{Q2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_{1n} & \frac{1}{2} & \cos t_{1n} & \cdots & \cos L t_n \end{pmatrix} \begin{pmatrix} b_0 \\ \alpha_{01} \\ \alpha_{11} \\ \vdots \\ \alpha_{LQ} \end{pmatrix}$$

Or in matrix notation it can be written as

$$\tilde{g}(t) = \boldsymbol{D}(\boldsymbol{t})\tilde{\alpha} \tag{9}$$

With $\tilde{g}(x)$ of size $n \times 1$, Matrix $\boldsymbol{D}(\boldsymbol{t})$ of size $n \times (Q + (L + 2))$ and $\tilde{\alpha}$ of size $(Q + (L + 2)) \times 1$. Meanwhile, L represents the number of oscillation parameters.

Furthermore, Mixed of spline truncated and fourier series nonparametric regression model can be written as follows :

$$\tilde{y} = \boldsymbol{G}(\boldsymbol{x})\tilde{\xi} + \boldsymbol{D}(\boldsymbol{t})\tilde{\alpha} + \tilde{\varepsilon} \tag{10}$$

To obtain parameter estimation results in model (10), it can be done using Ordinary Least Square (OLS) optimization method:

$$\underset{\tilde{\xi},\tilde{\alpha}}{\text{Min}} Q(\tilde{\xi}, \tilde{\alpha}) = \underset{\tilde{\xi},\tilde{\alpha}}{\text{Min}}\ \tilde{\varepsilon}^T \tilde{\varepsilon}$$
$$= \underset{\tilde{\xi},\tilde{\alpha}}{\text{Min}} \left(\tilde{y} - \boldsymbol{G}(\boldsymbol{x})\tilde{\xi} - \boldsymbol{D}(\boldsymbol{t})\tilde{\alpha}\right)^T \left(\tilde{y} - \boldsymbol{G}(\boldsymbol{x})\tilde{\xi} - \boldsymbol{D}(\boldsymbol{t})\tilde{\alpha}\right)$$

Estimated results are obtained

$$\hat{\tilde{\xi}} = \left(\boldsymbol{G}(\boldsymbol{x})^T \boldsymbol{G}(\boldsymbol{x})\right)^{-1} \boldsymbol{G}(\boldsymbol{x})^T \tilde{y} - \left(\boldsymbol{G}(\boldsymbol{x})^T \boldsymbol{G}(\boldsymbol{x})\right)^{-1} \boldsymbol{G}(\boldsymbol{x})^T\ \boldsymbol{D}(\boldsymbol{t})\hat{\tilde{\alpha}} \tag{11}$$

$$\hat{\tilde{\alpha}} = \left(\boldsymbol{D}(\boldsymbol{t})^T\boldsymbol{D}(\boldsymbol{t})\right)^{-1}\boldsymbol{D}(\boldsymbol{t})^T\tilde{y} - \left(\boldsymbol{D}(\boldsymbol{t})^T\boldsymbol{D}(\boldsymbol{t})\right)^{-1}\boldsymbol{D}(\boldsymbol{t})^T\,\boldsymbol{G}(\boldsymbol{x})\hat{\tilde{\xi}} \quad (12)$$

By example $\boldsymbol{R} = \left(\boldsymbol{G}(\boldsymbol{x})^T\boldsymbol{G}(\boldsymbol{x})\right)^{-1}\boldsymbol{G}(\boldsymbol{x})^T$ and $\boldsymbol{S} = \left(\boldsymbol{D}(\boldsymbol{t})^T\boldsymbol{D}(\boldsymbol{t})\right)^{-1}\boldsymbol{D}(\boldsymbol{t})^T$ then the form $\hat{\tilde{\xi}}$ and $\hat{\tilde{\alpha}}$ can be simplified to:

$$\hat{\tilde{\xi}} = \boldsymbol{R}\tilde{y} - \boldsymbol{R}\boldsymbol{D}(\boldsymbol{t})\hat{\tilde{\alpha}} \quad (13)$$

$$\hat{\tilde{\alpha}} = \boldsymbol{S}\tilde{y} - \boldsymbol{S}\boldsymbol{G}(\boldsymbol{x})\hat{\tilde{\xi}} \quad (14)$$

Then eliminate the equations (13) and (14) to get parameter values $\hat{\tilde{\xi}}$ that do not contain other parameters

$$\hat{\tilde{\xi}} = (\boldsymbol{I} - \boldsymbol{R}\boldsymbol{D}(\boldsymbol{t})\boldsymbol{S}\boldsymbol{G}(\boldsymbol{x}))^{-1}(\boldsymbol{R} - \boldsymbol{R}\boldsymbol{D}(\boldsymbol{t})\boldsymbol{S})\tilde{y}$$

$$\hat{\tilde{\xi}} = \boldsymbol{B}(k,l)\tilde{y} \quad (15)$$

Where $\boldsymbol{B}(k,l) = (\boldsymbol{I} - \boldsymbol{R}\boldsymbol{D}(\boldsymbol{t})\boldsymbol{S}\boldsymbol{G}(\boldsymbol{x}))^{-1}(\boldsymbol{R} - \boldsymbol{R}\boldsymbol{D}(\boldsymbol{t})\boldsymbol{S})$

The same thing is done to get the value of parameter $\hat{\tilde{\alpha}}$

$$\hat{\tilde{\alpha}} = (\boldsymbol{I} - \boldsymbol{S}\boldsymbol{G}(\boldsymbol{x})\boldsymbol{R}\boldsymbol{D}(\boldsymbol{t}))^{-1}(\boldsymbol{S} - \boldsymbol{S}\boldsymbol{G}(\boldsymbol{x})\boldsymbol{R})\tilde{y}$$

$$\hat{\tilde{\alpha}} = \boldsymbol{C}(k,l)\tilde{y} \quad (16)$$

Where $\boldsymbol{C}(k,l) = (\boldsymbol{I} - \boldsymbol{S}\boldsymbol{G}(\boldsymbol{x})\boldsymbol{R}\boldsymbol{D}(\boldsymbol{t}))^{-1}(\boldsymbol{S} - \boldsymbol{S}\boldsymbol{G}(\boldsymbol{x})\boldsymbol{R})$

Based on equations (15) and (16), the mixed estimator of spline truncated and fourier series in nonparametric regression model can be written as follows:

$$\begin{aligned}\hat{\tilde{y}} &= \boldsymbol{G}(\boldsymbol{x})\hat{\tilde{\xi}} + \boldsymbol{D}(\boldsymbol{t})\hat{\tilde{\alpha}} \\ &= \boldsymbol{G}(\boldsymbol{x})\boldsymbol{B}(k,l)\tilde{y} + \boldsymbol{D}(\boldsymbol{t})\boldsymbol{C}(k,l)\tilde{y} \\ &= (\boldsymbol{G}(\boldsymbol{x})\boldsymbol{B}(k,l) + \boldsymbol{D}(\boldsymbol{t})\boldsymbol{C}(k,l))\tilde{y} \\ &= \boldsymbol{A}(k,l)\tilde{y}\end{aligned} \quad (17)$$

Where $\boldsymbol{A}(k,l) = (\boldsymbol{G}(\boldsymbol{x})\boldsymbol{B}(k,l) + \boldsymbol{D}(\boldsymbol{t})\boldsymbol{C}(k,l)$

To determine the optimal knot point k and oscillation parameters l with the smallest Generalized Cross Validation (GCV) value criteria, namely:

$$GCV\ (k,l) = \frac{MSE\ (k,l)}{\left(n^{-1}tr\left(\boldsymbol{I} - \boldsymbol{A}(k,l)\right)\right)^2} \quad (18)$$

The Mean Square Error (MSE) is

$$MSE\ (k,l) = \frac{\left(\tilde{y} - \hat{\tilde{y}}\right)^T\left(\tilde{y} - \hat{\tilde{y}}\right)}{n}$$

B. *Stunting Prevalence Model*

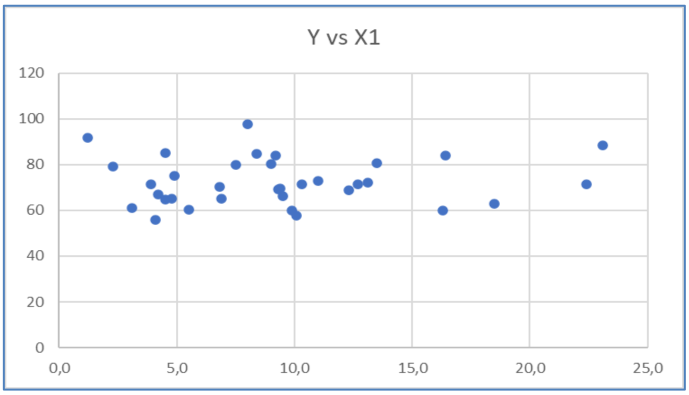
As an initial stage, descriptive analysis was carried out on each variable used in the research. This aims to describe the characteristics of the data concisely to gain a better understanding of the data.

Table 2 Descriptive Statistics

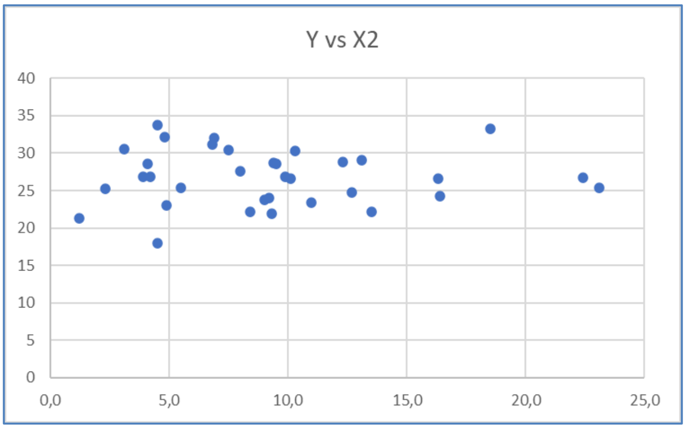
	N	Minimum	Maximum	Mean	Std. Deviation
Y	34	1.20	23.10	9.31	5.38
X1	34	55.91	97.50	72.49	10.35
X2	34	17.91	33.81	26.76	3.73

Table 2 presents descriptive statistics for variables analyzed in this study. The stunting prevalence is lowest in DKI Jakarta at 1.2, while West Sulawesi exhibits the highest prevalence at 23.1, indicating significant regional disparities in child growth and development. In terms of the percentage of the population who has health insuranceJambi has the lowest percentage at 55.91, whereas Aceh has the highest at 97.5. The standard deviation of 10.35 reflects quite high differences in access to healthcare services across provinces. Regarding the percentage of the population who smoked last month (Age ≥ 15 Years), Bali has the lowest percentage of smokers at 17.91, while Lampung has the highest at 33.81. These figures provide insight into the health behaviors of the population in different regions.

The scatter plots illustrate the initial relationship between the response variable and predictor variables, informing the decision to employ a non-parametric approach—specifically, truncated spline or Fourier series—based on the theoretical shape of the data. If the data pattern is cut into certain sub-intervals, it can be followed by spline truncated. Meanwhile, if the data has a repeating pattern or forms a sinus cosinus wave, it can be followed by fourier series. The comparison between the scatter plot of the stunting prevalence variable and each predictor variable, namely the percentage of the population who have health insurance and the percentage of the population who smoked in the last month (aged ≥ 15 years), is shown in Figure 1.



(a). stunting prevalence vs percentage of the population who has health insurance



(b). stunting prevalence vs the percentage of the population who smoked last month (Age ≥ 15 Years).

Figure 1 Scatter Plot Data

Based on the scatter plots displayed in Figure 1, both Figure 1a and Figure 1b reveal analogous patterns, indicating that they can be effectively modeled using either the spline truncated method or Fourier series. This observation stems from the identification of both repeating patterns and truncated segments at specific sub-intervals, highlighting the complexity and variability in the data. To enhance the modeling process, the criterion of the smallest Generalized Cross-Validation (GCV) value is employed. This approach allows for a systematic evaluation of the predictor variables, facilitating the determination of which can be more accurately approximated by either the spline truncated or Fourier series methods.

Table 3 Comparison of GCV Value

No	Variable Spline Truncated	Variable Fourier Series	GCV
1	X_1	X_2	35.75
2	X_2	X_1	34.46*

Based on Table 3, the smallest GCV value is 34.46, where variable X_2 is approximated using the spline truncated estimator, while variable X_1 is approximated using fourier series. The variable approximated by the spline truncated estimator is denoted as t and the variable approximated by the fourier series is represented as x .

Estimating results using mixed spline truncated and fourier series nonparametric regression model with one knot and one oscillation parameter are shown:

$$\hat{y} = \hat{\xi}_0 + \hat{\xi}_1 x_1 + \hat{\xi}_2 (x_1 - K_1)_+ + \hat{b}t_1 + \frac{\hat{\alpha}_0}{2} + \hat{\alpha}_1 \cos t_1$$

$$= -5.27 \times 10^{-10} + 0.165x_1 - 1.06(x_1 - 30.14)_+ + 0.062t_1 - \frac{5.30 \times 10^{-15}}{2} + 1.26 \cos t_1$$

The obtained smallest Generalized Cross Validation (GCV) value is 34.46 and the Mean Square Error (MSE) is 4.89.

Next, the model $\hat{y}$ obtained can be interpreted partially. The equation model that can be directly interpreted applies only to the variable approximated by the spline truncated estimator, which is the predictor variable representing the percentage of the population who smoked last month (age ≥ 15 years). In contrast, the regression curve approximated using the Fourier series estimator, where the predictor variable represents the percentage of the population with health insurance, cannot be directly interpreted. However, the analysis of the response variable still involves the percentage of the population with health insurance.

If variable t_1 is constant, then the spline truncated function representing the influence of the percentage of the population who smoked last month (age ≥ 15 years) on stunting prevalence is given by the following expression:

$$\hat{y} = -5.27 \times 10^{-11} + 0.165x_1 - 1.06(x_1 - 30.14)_+ + c$$

$$= \begin{cases} -5.27 \times 10^{-11} + 0.165x_1 & ; x_1 < 30.14 \\ 31.95 - 0{,}895x_1 & ; x_1 \geq 30.14 \end{cases}$$

The interpretation of the truncated spline model above is for provinces with a stunting prevalence of less than 30.14, an increase of one unit in the percentage of the population who smoked last month (age ≥ 15 years) will increase the stunting prevalence by 0.165. Meanwhile, for provinces with a stunting prevalence of 30.14 or higher, an increase of one unit in the percentage of the population who smoked last month (age ≥ 15 years) will decrease the stunting prevalence by 0.895.

V. CONCLUSIONS

The mixed estimator of spline truncated and fourier series in a nonparametric regression model can be expressed as

$$\hat{\tilde{y}} = \boldsymbol{G}(\boldsymbol{x})\boldsymbol{B}(k,l)\tilde{y} + \boldsymbol{D}(\boldsymbol{t})\boldsymbol{C}(k,l)\tilde{y},$$

where $\boldsymbol{B}(k,l) = (\boldsymbol{I} - \boldsymbol{RD}(\boldsymbol{t})\boldsymbol{SG}(\boldsymbol{x}))^{-1}(\boldsymbol{R} - \boldsymbol{RD}(\boldsymbol{t})\boldsymbol{S})$ and $\boldsymbol{C}(k,l) = (\boldsymbol{I} - \boldsymbol{SG}(\boldsymbol{x})\boldsymbol{RD}(\boldsymbol{t}))^{-1}(\boldsymbol{S} - \boldsymbol{SG}(\boldsymbol{x})\boldsymbol{R})$. The finding from the modeling of stunting prevalence data using mixed spline truncated and fourier series nonparametric regression yield the equation

$$\hat{y} = -5.27 \times 10^{-10} + 0.165x_1 - 1.06(x_1 - 30.14)_+ + 0.062t_1 - \frac{5.30 \times 10^{-15}}{2} + 1.26 \cos t_1$$

With a Generalized Cross Validation (GCV) of 34.46 and a Mean Square Error (MSE) of 4.89.

ACKNOWLEDGEMENT

We would like to thank the Bacharuddin Jusuf Habibie Institute of Technology for the financial support through Internal Grants provided for this research. We also thank all parties who participated in this research. Without their support and cooperation, this research would not have been successful.

REFERENCES

[1] T. Beal, A. Tumilowicz, A. Sutrisna, D. Izwardy, and L. M. Neufeld, "A review of child stunting determinants in Indonesia," *Matern. Child Nutr.*, vol. 14, no. 4, pp. 1–10, 2018, doi: 10.1111/mcn.12617.

[2] K. Rahmadhita, "Permasalahan Stunting dan Pencegahannya," *J. Ilm. Kesehat. Sandi Husada*, vol. 11, no. 1, pp. 225–229, 2020, doi: 10.35816/jiskh.v11i1.253.

[3] Kemenkes RI, *Profil Kesehatan Indonesia 2021*. 2022.

[4] C. R. Titaley, I. Ariawan, D. Hapsari, and A. Muasyaroh, "Determinants of the Stunting of Children in Indonesia : A Multilevel Analysis of the 2013 Indonesia Basic Health Survey," *Nutrients*, vol. 11, p. 1160, 2013.

[5] V. E. Togatorop, L. Rahayuwati, and R. D. Susanti, "Predictor of Stunting Among Children 0-24 Months Old in Indonesia: A Scoping Review," *J. Obs. J. Pendidik. Anak Usia Dini*, vol. 7, no. 5, pp. 5654–5674, 2023, doi: 10.31004/obsesi.v7i5.5222.

[6] A. Murad, F. Faruque, A. Naji, A. Tiwari, M. Helmi, and A. Dahlan, "Modelling geographical heterogeneity of diabetes prevalence and socio-economic and built environment determinants in Saudi City - Jeddah," *Geospat. Health*, vol. 17, no. 1, 2022, doi: 10.4081/gh.2022.1055.

[7] S. D. Anindita, "Pemodelan Persentase Balita Stunting di Indonesia Menggunakan Regresi Nonparametrik Spline Truncated," 2018.

[8] H. Husain, A. F. Dewi, and A. E. Wardani, "Permodelan Prevalensi Stunting Indonesia Menggunakan Regresi Nonparametrik Spline Truncated," *J. Anal. Res. Stat. Comput.*, vol. 3, no. 1, pp. 1–13, 2024, [Online]. Available: https://jarsic.org/main/article/view/26/19

[9] H. Husain, I. N. Budiantara, and I. Zain, "Mixed estimator of spline truncated, Fourier series, and kernel in biresponse semiparametric regression model," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 880, no. 1, 2021, doi: 10.1088/1755-1315/880/1/012046.

[10] Eubank R L, *Spline Smoothing and Nonparametric Regression*. New York, 1999.

[11] M. Fariz, F. Mardianto, P. Supervisor, N. Budiantara, and M. Si, "Semiparametric Regression Model of Biresponse Using Fourier Series," 1312.

[12] L. Ni'matuzzahroh, "Estimator Campuran Spline Truncated, Kernel, Dan Deret Fourier Dalam Regresi Nonparametrik Pada Data Longitudinal," 2021.

[13] A. J. Prendergast and J. H. Humphrey, "The stunting syndrome in developing countries," *Paediatr. Int. Child Health*, vol. 34, no. 4, pp. 250–265, 2014, doi: 10.1179/2046905514Y.0000000158.

[14] Kementerian Kesehatan Republik Indonesia, "Buku Saku Hasil Studi Status Gizi Indonesia (SSGI) Tahun 2022," *Kemenkes RI*, pp. 1–14, 2022, [Online]. Available: https://www.litbang.kemkes.go.id/buku-saku-hasil-studi-status-gizi-indonesia-ssgi-tahun-2021/