Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2025, 2025:90
https://doi.org/10.28919/cmbn/9342
ISSN: 2052-2541

ANALYSIS OF SPECIAL INDEX OF STUNTING TREATMENT THROUGH CONFIDENCE INTERVAL OF LONGITUDINAL SEMIPARAMETRIC SPLINE MODEL PARAMETERS

SITTI BALQIES GANDE, ANNA ISLAMIYATI*, GEORGINA MARIA TINUNGKI

Department of Statistics, Faculty of Mathematics and Natural Sciences, Hasanuddin University, Makassar 90245, Indonesia

Abstract: Stunting is a major concern in many nations, including Indonesia, and it has long-term implications for human resource quality. The Special Index of Stunting Treatment (SIST) is one of the social population problems that can be modeled with regression. The combination of parametric and nonparametric regression techniques is known as a semiparametric regression. This study used longitudinal data, obtained from repeated observations within a certain time span on the same individual in sequence. This study's estimation method makes use of a truncated spline, which has great flexibility and statistical interpretation. In the statistical inference of the parameter, the calculation of the confidence interval is crucial. Providing an estimated range for a population parameter based on two boundary points enables quantification of the level of confidence in the accuracy and reliability of the estimate. Six dimensions are formed in the acceleration of stunting reduction: health, education, nutrition, housing, food, and social protection. Results of the confidence interval estimation of truncated semiparametric spline regression parameters on longitudinal data. The optimal model was obtained using three knot points, resulting in a minimum GCV of 12.7066, an MSE of 2.0225, and R^2 93.66%. This study shows that this model is feasible and the predictor variables have a negative and positive influence on the Special Index of Stunting Treatment (SIST).

Keywords: confidence interval; semiparametric regression; truncated spline; longitudinal data; GCV; SIST.

2020 AMS Subject Classification: 62P10.

*Corresponding authors
E-mail addresses: annaislamiyati701@gmail.com
Received May 16, 2025

1. INTRODUCTION

Stunting is a nutritional problem that requires serious handling in Indonesia [1]. Stunting is a growth and developmental problem experienced by children as a result of poor nutrition, frequent illnesses, and lack of psychosocial stimulation [2]. According to the Indonesian Nutrition Status Survey in 2021, the prevalence of stunting in Indonesia is 24.4%, with a target of 14% by 2024. Stunting, or low height for age, is caused by long-term lack of dietary intake and/or repeated illnesses [3]. Some of the variables that impact the high incidence of stunting in Indonesia include inadequate parenting, poor access to health facility services, particularly ANC, especially for pregnant women, restricted availability of healthy food for families, and limited access to clean water [4]. There have been many studies on stunting, including sparse categorical principal component analysis [5], support vector machine [6], nonparametric spline truncated regression [7], spline binary logistic [8], and ordinal logistic regression models [9]. This study used data from the Special Index of Stunting Treatment (SIST), in which one predictor variable has a known association pattern with the response variable; however, the link between the other predictor factors and the response variable is unclear. This problem can be solved by using a semiparametric regression model.

Semiparametric regression combines parametric and nonparametric regression methods. Parametric regression is a type of regression in which the relationship pattern between the response variable and the predictor is known. Nonparametric regression is a type of regression in which the pattern of connection between the response variables and predictors is unknown [7]. Some estimators that have been performed in semiparametric regression include the kernel [10-11], Fourier series [12], local linear [13], local polynomial [14] and spline [15-18]. Among these estimators, spline is the most commonly used because it has the flexibility to estimate functions or data that vary on certain sub-intervals [18].

One of the most important parts of statistical inference is the confidence interval. With confidence interval estimation, it can be determined whether the value of a parameter is significant, especially the parameters in a regression model [19]. Confidence interval estimation approaches a parameter value using two points, which makes it possible to measure the degree of confidence in the accuracy of the estimate. Some researchers have addressed confidence interval estimation, namely, confidence interval estimation of semiparametric regression model parameters, confidence interval estimation with known variance using spline nonparametric regression [20], and confidence interval of truncated spline semiparametric regression [21]. Although these investigations remain confined to cross-sectional data, they may be advanced into longitudinal data. Longitudinal data refer to the information that is observed and measured regularly over specified time intervals. Numerous scholars have conducted studies on longitudinal data in semiparametric regression, focusing on the selection of optimal knot points [22], using longitudinal semiparametric regression with the least squares spline estimator [17].

Based on this description, the challenge in this work is how to model the Special Index of Stunting Treatment (SIST) with a confidence interval estimation method for the parameters of the spline semiparametric regression model on longitudinal data. The problem restriction in this study is the identification of ideal knot locations utilizing Generalized Cross Validation (GCV) with one, two, and three knots.

2. PREPARATION

A spline is a piecewise polynomial function that consists of several polynomials defined at certain sub-intervals. This feature is shown by a truncated function coupled with the estimator, and the pieces are referred to as knots. Knot points are joint combinations that illustrate the change in the behavioral pattern of the function at different intervals. If a spline function is programmed with knot points $k_1, k_2, \cdots, k_r$ is a function expressed in the form [23]

$$g(z_i) = \sum_{s=1}^{m} \alpha_s z_i^s + \sum_{k=1}^{r} \alpha_k (z_i - k_k)_+^m \tag{1}$$

With, $\sum_{s=1}^{m} \alpha_s z_i^s$ is the polynomial component and $\sum_{k=1}^{r} \alpha_k (z_i - k_k)_+^m$ is the truncated component with:

$$(z_i - k_k)_+^m = \begin{cases} (z_i - k_k)^m, & z_i \geq k_k \\ 0 & , z_i < k_k \end{cases} \tag{2}$$

The semiparametric regression integrates parametric and nonparametric regression methods. If given data (x_{ij}, z_{ij}, y_{ij}), where; $i = 1,2, \dots, n\,; j = 1,2, \dots, t$ the data are considered to be linked and follow longitudinal semiparametric regression. Thus, the equation model is as follows in (3).

$$y_{ij} = f(x_{ij}) + g(z_{ij}) + \varepsilon_{ij} \tag{3}$$
$$= f(x_{1ij}, x_{2ij}, \dots x_{pij}) + g(z_{1ij}, z_{2ij}, \dots, z_{qij}) + \varepsilon_{ij}$$

Where y_i is the response variable on the-i th data, $f(x_{1ij} + x_{2ij} + \cdots x_{pij})$ is the parametric component predictor variable approximated by a linear function while $g(z_{1ij}, z_{2ij}, \dots, z_{qij})$ is a nonparametric function of unknown shape.

If the semiparametric regression on the longitudinal data is additive, then the following equation can be formed:

$$y_{ij} = \sum_{u=1}^{p} f(x_{uij}) + \sum_{s=1}^{q} g(z_{sij}) \tag{4}$$

where, $u = 1,2, \dots p$; $s = 1,2, \dots q$. If the truncated polynomial function is assumed to be the value of $m = 1$, then based on the model in equation (4), the semiparametric regression model on longitudinal data can be derived in Equation (5) as follows:

$$y_{ij} = \beta_{0i} + \sum_{u=1}^{p} \beta_{iu} x_{iuj} + \sum_{s=1}^{q} \left(\alpha_{1i} z_{sij} + \sum_{k=1}^{r} \alpha_{s(1+k)i} (z_{sij} - K_{ski})_{+}^{1} \right)_{+} + \varepsilon_{ij} \tag{5}$$

where its truncated function$(z_{sij} - K_{ski})_{+}^{1}$ is defined as follows

$$(z_{sij} - K_{ski})_{+}^{1} = \begin{cases} (z_{sij} - K_{ski})^{1}, & z_{sij} \geq K_{ski} \\ 0 & , z_{sij} < K_{ski} \end{cases}$$

Based on Equation (5), the following matrix form can be presented:

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

Where,

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} S_1 & 0 & 0 & 0 \\ 0 & S_2 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & S_n \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{7}$$

Where, $\boldsymbol{y}$ is a vector of size $nt \times 1$, $\boldsymbol{S}$ is $[\boldsymbol{X} \vdots \boldsymbol{Z}]$ of size $nt \times n[1 + p + q(1 + r)]$ which contains parametric and nonparametric component variables, $\boldsymbol{\delta}$ is a consists of parameter vectors $\hat{\beta}$ and $\hat{\alpha}$ size $n[1 + p + q(r + 1)] \times 1$ that includes parametric and nonparametric component parameters and $\boldsymbol{\varepsilon}$ is random error vector size $nt \times 1$. Based on Equation (6), if it is assumed that the random error $\boldsymbol{\varepsilon}$ is independent and normally distributed with mean 0 and variance σ_i^2, then estimator $\boldsymbol{\delta}$ can be obtained by solving the Maximum Likelihood estimation (MLE).

$$\boldsymbol{\varepsilon} = \boldsymbol{y} - \boldsymbol{S\delta} \tag{8}$$

So that the likelihood function is obtained

$$L(\boldsymbol{\delta}) = \prod_{i=1}^{n}\left(\frac{1}{\sigma_i\sqrt{2\pi}}\right)\left[\exp\left\{-\frac{1}{2}\left(\frac{\boldsymbol{y}-\boldsymbol{S\delta}}{\sigma_i}\right)^{\boldsymbol{2}}\right\}\right] \tag{9}$$

$$= \prod_{i=1}^{n}\left(\frac{1}{\sigma_i\sqrt{2\pi}}\right)\exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma_i}(\boldsymbol{y}-\boldsymbol{S\delta})^T(\boldsymbol{y}-\boldsymbol{S\delta})\right)\right\}$$

If,

$$\boldsymbol{W} = \begin{pmatrix} 1/\sigma_1^2 & 0 & \cdots & 0 \\ 0 & 1/\sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/\sigma_n^2 \end{pmatrix},$$

then $\boldsymbol{W}$ is a weighted matrix of size $nt \times nt$ that addresses cases of heteroscedasticity**.** The optimization solution was then carried out with the following explanation,

$$L(\boldsymbol{\delta}) = \prod_{i=1}^{n}\left(\frac{1}{\sigma_i\sqrt{2\pi}}\right)\exp\left\{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{S\delta})^T\boldsymbol{W^{-1}}(\boldsymbol{y}-\boldsymbol{S\delta})\right\} \tag{10}$$

The Log-likehood function:

$$L(\boldsymbol{\delta}) = \log\big(L(\boldsymbol{\delta})\big) \tag{11}$$

$$= \log\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{S\delta})^T\boldsymbol{W^{-1}}(\boldsymbol{y}-\boldsymbol{S\delta})\right\}$$

$$= n.\log(1) - n\log(\sigma) - n\log\big(\sqrt{2\pi}\big) - \frac{1}{2}(\boldsymbol{y}^T\boldsymbol{W^{-1}y} - 2\boldsymbol{S}^T\boldsymbol{W^{-1}y\delta} +$$

$$\boldsymbol{S}^T\boldsymbol{W^{-1}S\delta}^T\boldsymbol{\delta})$$

To get the estimation of $\boldsymbol{\delta}$, it can be partially derivatized and then equated zero, to get the estimation of $\boldsymbol{\delta}$, as follows

$$\frac{\partial(l(\boldsymbol{\delta}))}{\partial\boldsymbol{\delta}} = \frac{\partial}{\partial\boldsymbol{\delta}}\Big[n.\log(1) - n\log(\sigma) - n\log\big(\sqrt{2\pi}\big) - \frac{1}{2}(\boldsymbol{y}^T\boldsymbol{W^{-1}y} -$$

$$2\boldsymbol{S}^T\boldsymbol{W^{-1}y\delta} + \boldsymbol{S}^T\boldsymbol{W^{-1}S\delta}^T\boldsymbol{\delta})\Big]$$

$$= -0 - 0 - 0 + \boldsymbol{S}^T\boldsymbol{W^{-1}y} - \frac{1}{2}.2\,\boldsymbol{S}^T\boldsymbol{W^{-1}S\hat{\delta}} = \boldsymbol{S}^T\boldsymbol{W^{-1}y} - \boldsymbol{S}^T\boldsymbol{S\hat{\delta}} \tag{12}$$

Then the estimation for $\hat{\boldsymbol{\delta}}$ is obtained as follows:

$$\hat{\boldsymbol{\delta}} = (\boldsymbol{S}^T\boldsymbol{W^{-1}S})^{-1}\boldsymbol{S}^T\boldsymbol{W^{-1}y} \tag{13}$$

By $\widehat{\boldsymbol{\delta}} = \begin{bmatrix} \hat{\beta} \\ \vdots \\ \hat{\alpha} \end{bmatrix}$

based on the estimation of $\widehat{\boldsymbol{\delta}}$ in equation (13), then obtained:

$$\begin{aligned} \widehat{\boldsymbol{y}} &= \boldsymbol{S}\widehat{\boldsymbol{\delta}} \\ &= \boldsymbol{S}(\boldsymbol{S}^T\boldsymbol{W}^{-\mathbf{1}}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{W}^{-\mathbf{1}}\boldsymbol{y} \\ &= \boldsymbol{A}\boldsymbol{y} \end{aligned}$$

Where $\boldsymbol{A} = \boldsymbol{S}(\boldsymbol{S}^T\boldsymbol{W}^{-\mathbf{1}}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{W}^{-\mathbf{1}}\boldsymbol{y}$. Knot point selection with Generalized Cross Validation (GCV).

$$GCV(k) = \frac{MSE(k)}{[n^{-1}trace(\boldsymbol{I} - \boldsymbol{A}[\boldsymbol{k}])^2]} \tag{14}$$

where $GCV(k)$ is a vector containing the GCV values of the knot points, $MSE(k) = n^{-1}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$, $\boldsymbol{I}$ is the identity matrix, n is the number of observations, and the matrix $\boldsymbol{A}[\boldsymbol{k}]$ is obtained $\widehat{\boldsymbol{y}} = \boldsymbol{S}\widehat{\boldsymbol{\delta}}$. Furthermore, assuming $\varepsilon \sim N(0, \sigma^2\boldsymbol{W}^{-\mathbf{1}})$ then $y \sim N\left(\boldsymbol{S}\widehat{\boldsymbol{\delta}}, \sigma^2\boldsymbol{W}^{-\mathbf{1}}\right)$. The expectation and variance of the estimate $\widehat{\boldsymbol{\delta}}$ are respectively:

$$\begin{aligned} E\left(\widehat{\boldsymbol{\delta}}\right) &= E((\boldsymbol{S}^T\boldsymbol{W}^{-\mathbf{1}}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{W}^{-\mathbf{1}}\boldsymbol{y}) \\ &= (\boldsymbol{S}^T\boldsymbol{W}^{-\mathbf{1}}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{W}^{-\mathbf{1}}\boldsymbol{S}\boldsymbol{\delta} \\ &= \boldsymbol{\delta} \end{aligned} \tag{15}$$

$$\begin{aligned} Var\left(\widehat{\boldsymbol{\delta}}\right) &= Var((\boldsymbol{S}^T\boldsymbol{W}^{-\mathbf{1}}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{W}^{-\mathbf{1}}\boldsymbol{y}) \\ &= (\boldsymbol{S}^T\boldsymbol{W}^{-\mathbf{1}}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{W}^{-\mathbf{1}}).Var\,(\boldsymbol{y}).\boldsymbol{W}^{-\mathbf{1}}(\boldsymbol{S}^T\boldsymbol{W}^{-\mathbf{1}}\boldsymbol{S})^{-1} \\ &= \sigma^2(\boldsymbol{S}^T\boldsymbol{W}^{-\mathbf{1}}\boldsymbol{S})^{-\mathbf{1}} \end{aligned} \tag{16}$$

It is proven that the estimator. $\widehat{\boldsymbol{\delta}} \sim N\left(\widehat{\boldsymbol{\delta}}, \sigma^2(\boldsymbol{S}^T\boldsymbol{W}^{-\mathbf{1}}\boldsymbol{S})^{-\mathbf{1}}\right)$

A confidence interval is an interval that contains parameters with a certain probability. Confidence intervals $(1-\alpha)100\%$ for semiparametric regression model parameters using a truncated spline on longitudinal data, where $\delta_u, u = 1,2, \dots nt - n\big(p + q(1+r)\big)$, variance σ^2 is unknown.

$$T_u = \left(x_{i1}, x_{i2}, \dots, x_{ip,} z_{i1}, z_{i2}, \dots, z_{iq}, y\right) = \frac{\widehat{\boldsymbol{\delta}}_u - \boldsymbol{\delta}_u}{\sqrt{MSE(\boldsymbol{S}^T\boldsymbol{W}^{-\mathbf{1}}\boldsymbol{S})^{-\mathbf{1}}_{uu}}} \tag{17}$$

Furthermore, the confidence interval $(1-\alpha)$ can be obtained by solving the probability in Equation (18).

$$P\left(A_u \le T_u\left(x_{i1}, x_{i2}, \dots, x_{ip,} z_{i1}, z_{i2}, \dots, z_{iq}, y\right) \le B_u\right) = 1-\alpha \tag{18}$$

If the significance level is α then the confidence interval $(1-\alpha)100\%$ for the parameter $\widehat{\boldsymbol{\delta}}$ in Equation (6) is as follows

$$P\left(\begin{array}{c} \hat{\delta}_u - t_{\frac{\alpha}{2}, nt-n(p+q(r+1))} \le \sqrt{\frac{y^T \boldsymbol{A} y}{nt-(p+q(r+1))}}\boldsymbol{k} \le \delta_u \le \\ \hat{\delta}_u + t_{\frac{\alpha}{2}, nt-n(p+q(r+1))}\sqrt{\frac{y^T \boldsymbol{A} y}{nt-(p+q(r+1))}}\boldsymbol{k} \end{array}\right) = 1-\alpha \tag{19}$$

$\boldsymbol{k} = (\boldsymbol{S}^T \boldsymbol{W}^{-\mathbf{1}} \boldsymbol{S})_{uu}^{-\mathbf{1}}$ denotes the element of the u th row u th column of the $\sqrt{(\boldsymbol{S}^T \boldsymbol{W}^{-\mathbf{1}} \boldsymbol{S})_{uu}^{-\mathbf{1}}}$ matrix.

3. RESULTS

This study uses data from the Special Index of Stunting Treatment (SIST) in 2018-2023 data in Indonesia as well as several influencing factors, namely the health, education, nutrition, housing, food, and social protection dimensions.

Table 1. Descriptive Statistics

Variables	N	Min	Max	Mean
y	204	45.5	85.8	68.145
x_1	204	36.1	97.4	70.979
x_2	204	11.7	77.6	38.558
x_3	204	52.2	100.0	80.473
x_4	204	49.6	97.5	82.175
x_5	204	25.1	97.6	76.330
x_6	204	35.0	95.7	57.345

Table 1. shows that the average SIST is 68.145 percent with a maximum value of 85.8 percent obtained by East Java Province in 2023, while the minimum value is in Papua Province at 45.5 percent. The average Health Dimension in Indonesia in 2018-2023 is 70.979, for the Education Dimension in Indonesia in 2018-2023 is 38.558, the Nutrition Dimension in Indonesia in 2018-2023 is 80.437, and the Housing Dimension in Indonesia in 2018-2023 is 82.175, the Food

Protection Dimension in Indonesia in 2018-2023 is 76.330 and the Social Protection Dimension in Indonesia in 2018-2023 is 57.345.

The relationship pattern between the response and predictor variables was analyzed using a scatter plot before beginning any modeling. The initial form of the relationship between the response and predictor variables was determined using at scatter plot. The figure below shows at scatter plot between the response variable and predictor variables.

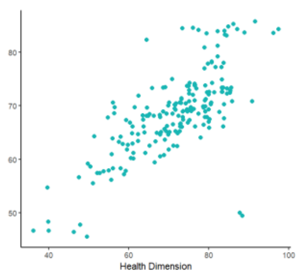

Figure 1: Health Dimension and SIST

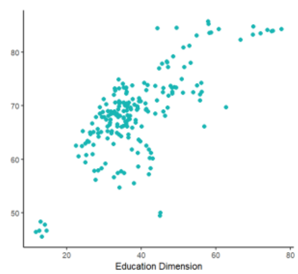

Figure 2: Education dimension and SIST

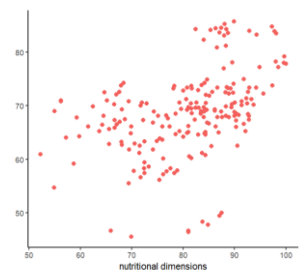

Figure 3. Nutrition Dimension and SIST

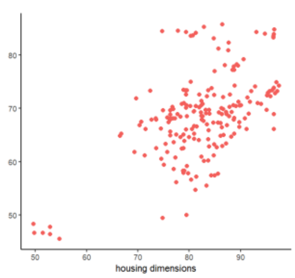

Figure 4. Housing Dimension and SIST

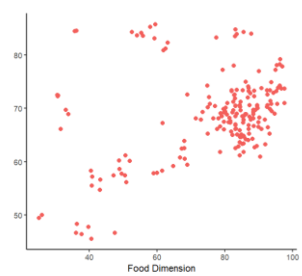

Figure 5. Food Dimension and SIST

Figure 6. Social Protection Dimension and SIST

Figure 1. Scatter plot of predictor variables and Special Index of Stunting Treatment (SIST)

Based on Figure 1, the relationship pattern between SIST and several predictor variables forms two relationship patterns, where the green scatter plot forms a regular pattern that includes the health and education dimensions. In addition, for the red-colored scatter plot, it makes a pattern that is not known to its functional form, where the variables included are the nutrition, housing, food, and social protection dimensions. Therefore, special index of stunting treatment data with its predictor variables will be modeled using semiparametric spline regression on longitudinal data.

To identify the optimum model, the Generalized Cross Validation (GCV) approach uses the minimal value to find the optimal knot point. One to three ideal knot points were employed in this investigation. Table 2 indicates that the minimum GCV value with three knot points was 12.7066.

These findings are used to model the Special Index of Stunting Treatment. Table 3 displays the outcomes of utilizing three knot points to estimate the parameters of the spline semiparametric regression model on longitudinal data.

Table 2. Optimal Knot Point Selection

Knot Point	1	2	**3**
GCV	14.9384	14.0566	12.7066

Table 3. Parameter Estimation

Parameters	Estimation	Parameters	Estimation
$\boldsymbol{\beta_0}$	-1.0164	α_{24}	0.2975
$\boldsymbol{\beta_1}$	0.1757	α_{31}	0.3578
$\boldsymbol{\beta_2}$	0.1771	α_{32}	9.2684
$\boldsymbol{\alpha_{11}}$	0.1320	α_{33}	-10.2171
$\boldsymbol{\alpha_{12}}$	0.5826	α_{34}	0.7373
$\boldsymbol{\alpha_{13}}$	-0.6273	α_{41}	0.0435
$\boldsymbol{\alpha_{14}}$	0.1120	α_{42}	0.1308
$\boldsymbol{\alpha_{21}}$	0.1736	α_{43}	-0.0203
$\boldsymbol{\alpha_{22}}$	-0.8676	α_{44}	0.0566
$\boldsymbol{\alpha_{23}}$	0.5977		

The longitudinal model created by the spline semiparametric regression model is:

$$\hat{\boldsymbol{y}} = -1.0164 + 0.1757x_{i1} + 0.1771x_{i2} + 0.1320z_{i1} + 0.5826\,(z_{i1} - 68.4) - 0.6273(z_{i1} - 68.7)$$
$$+0.1120(z_{i1} - 71.6) + 0.1736z_{i2} - 0.8676(z_{i2} - 75.1) + 0.5977(z_{i2} - 75.4\,) + 0.2975$$
$$(z_{i2} - 77) + 0.3578z_{i3} + 9.2684(z_{i3} - 51.9) - 10.2171(z_{i3} - 52.6) + 0.7373$$
$$(z_{i3} - 64.8) + 0.0435z_{i4} + 0.1308(z_{i4} - 45\,) - 0.0237(z_{i4} - 45.2\,) + 0.0566(z_{i4} - 47.1)$$

Confidence interval estimation follows the acquisition of the parameter estimation model for the longitudinal spline semiparametric regression model to improve the accuracy of the estimated results in describing population circumstances. First, a test for the standard distribution was conducted. The acquired value of 0.09 > 0.05, which indicates a normal distribution, was subjected to the normality assumption test. Confidence intervals were constructed using the best spline semiparametric regression model parameter estimation findings.

Table 4. Confidence Interval

Parameters	Estimate	lower limit	Upper limit	Parameters	Estimate	lower limit	Upper limit
β_0	-1.0164	-8.6971	6.6643	α_{24}	0.2975	-2.5834	3.1785
β_1	0.1757	0.0855	0.2659	α_{31}	0.3578	0.1195	0.5962
β_2	0.1771	0.0881	0.2660	α_{32}	9.2684	-0.0809	18.6177
α_{11}	0.1320	-0.1214	0.3855	α_{32}	10.2171	-19.8128	-0.6214
α_{12}	0.5826	13.8103	14.9756	α_{34}	0.7373	0.1949	1.2797
α_{13}	-0.6273	-16.3543	15.0995	α_{41}	0.0435	-0.3794	0.4666
α_{14}	0.1120	-1.5550	1.7790	α_{42}	0.1308	-1.6525	1.9142
α_{21}	0.1736	-0.0394	0.3868	α_{43}	-0.0203	-1.9085	1.8677
α_{22}	-0.8676	-16.3103	14.5750	α_{44}	0.0566	-1.1318	1.2452
α_{23}	0.5977	-17.2100	18.4055				

Based on Table 4, 95% confidence interval of the model parameters. The predictor variables that do not have a significant effect on the Special Index for Handling Stunting are Nutrition Dimension, Housing Dimension, and Social Protection Dimension because the parameter confidence interval is obtained in the range from negative values to positive values that contain zero values. However, the Food, Education, and Health Dimensions significantly improved the Special Index of Stunting Treatment (SIST). Therefore, every one unit increase in the Special Index of Stunting Treatment (SIST) will increase the value of the health, nutrition, and food dimensions by respectively, 0.1757, 0.1771, and 0.3578. The semiparametric regression model with a linear truncated spline approach on this longitudinal data has a value MSE of 2.0225 and a value R^2 of 93.66%.

4. CONCLUSION

Based on the findings of the 95% confidence interval estimation for the parameters of the semiparametric spline regression model using longitudinal data of the Special Index of Stunting Treatment (SIST) in Indonesia between 2018 and 2023, it was shown that among the six predictor variables is health, education, food, nutrition, housing, and social protection. Three had a statistically significant effect: health, education, and food. By contrast, the nutrition, housing, and social protection dimensions exhibit confidence intervals that range from negative to positive values and include zero, indicating that these dimensions do not have a sufficiently strong or consistent influence on the stunting reduction index.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interest.

REFERENCES

[1] M.S. Noor, M.D. Andrestian, R.A. Dina, et al. Analysis of Socioeconomic, Utilization of Maternal Health Services, and Toddler's Characteristics as Stunting Risk Factors, Nutrients 14 (2022), 4373. https://doi.org/10.3390/nu14204373.

[2] A.F. Suryono, A. Kurniawan, P.A.P. Widyangga, M.S. Dewanti, Modeling the Stunting Prevalence Rate in Indonesia Using Multi-Predictor Truncated Spline Nonparametric Regression, J. Apl. Stat. Komput. Stat. 16 (2024), 1-14. https://doi.org/10.34123/jurnalasks.v16i1.719.

[3] H. Torlesse, A.A. Cronin, S.K. Sebayang, et al. Determinants of Stunting in Indonesian Children: Evidence From a Cross-Sectional Survey Indicate a Prominent Role for the Water, Sanitation and Hygiene Sector in Stunting Reduction, BMC Public BMC Public Health 16 (2016), 669. https://doi.org/10.1186/s12889-016-3339-8.

[4] T. Huriah, N. Nurjannah, Risk Factors of Stunting in Developing Countries: a Scoping Review, Open Access Maced. J. Med. Sci. 8 (2020), 155-160. https://doi.org/10.3889/oamjms.2020.4466.

[5] A. Islamiyati, M. Nur, A. Salam, et al. Risk Factor Analysis for Stunting Incidence Using Sparse Categorical Principal Component Logistic Regression, MethodsX 14 (2025), 103186. https://doi.org/10.1016/j.mex.2025.103186.

[6] M. Fadil, A. Islamiyati, S.A. Thamrin, Classification of Nutritional Status in Toddlers Using the Support Vector Machine Method, Commun. Math. Biol. Neurosci. 2025 (2025), 57. https://doi.org/10.28919/cmbn/9126.

[7] A. Islamiyati, A. Kalondeng, M. Zakir, S. Djibe, U. Sari, Detecting Age Prone to Growth Retardation in Children Through a Bi-Response Nonparametric Regression Model with a Penalized Spline Estimator, Iran. J. Nurs. Midwifery Res. 29 (2024), 549-554. https://doi.org/10.4103/ijnmr.ijnmr_342_22.

[8] A. Islamiyati, Anisa, M. Zakir, et al. The Use of the Binary Spline Logistic Regression Model on the Nutritional Status Data of Children, Commun. Math. Biol. Neurosci. 2023 (2023), 37. https://doi.org/10.28919/cmbn/7935.

[9] S. Arifin, A. Islamiyati, E.T. Herdiani, Ability of Ordinal Spline Logistic Regression Model in the Classification of Nutritional Status Data, Commun. Math. Biol. Neurosci. 2023 (2023), 83. https://doi.org/10.28919/cmbn/8072.

[10] A.T. Ampa, I. Nyoman Budiantara, I. Zain, Kernel Multivariable Semiparametric Regression Model in

Estimating the Level of Open Unemption in East Java Province, J. Phys.: Conf. Ser. 1899 (2021), 012127. https://doi.org/10.1088/1742-6596/1899/1/012127.

[11] I. Selingerova, S. Katina, I. Horova, Comparison of Parametric and Semiparametric Survival Regression Models with Kernel Estimation, J. Stat. Comput. Simul. 91 (2021), 2717-2739. https://doi.org/10.1080/00949655.2021.1906875.

[12] R. Kustianingsih, M.F.F. Mardianto, B.A. Ardhani, undefined. Kuzairi, A. Thohari, R. Andriawan, T. Yulianto, Fourier Series Estimator in Semiparametric Regression to Predict Criminal Rate in Indonesia, AIP Conf. Proc. 2329 (2021), 060023. https://doi.org/10.1063/5.0042123.

[13] T.W. Utami, N. Chamidah, T. Saifudin, Hematocrit Modeling in Dhf Patients Based on Longitudinal Data Using Local Linear Estimator in Semiparametric Regression, AIP Conf. Proc. 3302 (2025), 020006. https://doi.org/10.1063/5.0262183.

[14] V. Fibriyani, N. Chamidah, T. Saifudin, Estimating Time Series Semiparametric Regression Model Using Local Polynomial Estimator for Predicting Inflation Rate in Indonesia, J. King Saud Univ. - Sci. 36 (2024), 103549. https://doi.org/10.1016/j.jksus.2024.103549.

[15] R.D. Fadlirhohim, S. Sifriyani, A.T.R. Dani, Modeling Stunting Prevalence in Indonesia Using Spline Truncated Semiparametric Regression, BAREKENG 18 (2024), 2015-2028. https://doi.org/10.30598/barekengvol18iss3pp2015-2028.

[16] W. Ramadan, N. Chamidah, B. Zaman, L. Muniroh, B. Lestari, Standard Growth Chart of Weight for Height to Determine Wasting Nutritional Status in East Java Based on Semiparametric Least Square Spline Estimator, IOP Conf. Ser.: Mater. Sci. Eng. 546 (2019), 052063. https://doi.org/10.1088/1757-899x/546/5/052063.

[17] M. Setyawati, N. Chamidah, A. Kurniawan, Modelling Scholastic Aptitude Test of State Islamic Colleges in Indonesia Using Least Square Spline Estimator in Longitudinal Semiparametric Regression, J. Phys.: Conf. Ser. 1764 (2021), 012077. https://doi.org/10.1088/1742-6596/1764/1/012077.

[18] N. Chamidah, B. Zaman, L. Muniroh, et al. Multiresponse Semiparametric Regression Model Approach to Standard Growth Charts Design for Assessing Nutritional Status of East Java Toddlers, Commun. Math. Biol. Neurosci. 2023 (2023), 30. https://doi.org/10.28919/cmbn/7814.

[19] B. Lestari, N. Chamidah, I.N. Budiantara, et al. Determining Confidence Interval and Asymptotic Distribution for Parameters of Multiresponse Semiparametric Regression Model Using Smoothing Spline Estimator, J. King Saud Univ. - Sci. 35 (2023), 102664. https://doi.org/10.1016/j.jksus.2023.102664.

[20] A. Islamiyati, undefined. Raupong, A. Kalondeng, U. Sari, Estimating the Confidence Interval of the Regression

Coefficient of the Blood Sugar Model Through a Multivariable Linear Spline with Known Variance, Stat. Transit. New Ser. 23 (2022), 201-212. https://doi.org/10.2478/stattrans-2022-0012.

[21] M. Setyawati, N. Chamidah, A. Kurniawan, et al. Confidence Interval for Semiparametric Regression Model Parameters Based on Truncated Spline with Application to Covid-19 Dataset in Indonesia, Data Metadata 3 (2024), 609. https://doi.org/10.56294/dm2024.609.

[22] A.T. Damaliana, I. Nyoman Budiantara, et al. Comparing Between Mgcv and Agcv Methods to Choose the Optimal Knot Points in Semiparametric Regression with Spline Truncated Using Longitudinal Data, IOP Conf. Ser.: Mater. Sci. Eng. 546 (2019), 032003. https://doi.org/10.1088/1757-899x/546/3/032003.

[23] N. Chamidah, B. Lestari, A. Massaid, et al. Estimating Mean Arterial Pressure Affected by Stress Scores Using Spline Nonparametric Regression Model Approach, Commun. Math. Biol. Neurosci. 2020 (2020), 72. https://doi.org/10.28919/cmbn/4963.