JTAM (Jurnal Teori dan Aplikasi Matematika)
http://journal.ummat.ac.id/index.php/jtam
p-ISSN 2597-7512 | e-ISSN 2614-1175
Vol. 10, No. 2, April 2026, pp. 647-660

Estimation of Stunting and Wasting Prevalence in Southern Part of Sumatra Using Nadaraya-Watson Kernel and Penalized Spline

Cinta Rizki Oktarina[1], Sigit Nugroho[2*], Idhia Sriliana[2]
[1]Department of Mathematics, Institut Teknologi Batam, Indonesia
[2]Department of Statistics, Universitas Bengkulu, Indonesia
snugroho@unib.ac.id

Article History:
Received : 03-11-2025
Revised : 05-01-2026
Accepted : 06-01-2026
Online : 01-04-2026

Keywords:
Bivariate Nonparametric Regression;
Kernel Nadaraya Watson;
Penalized Spline;
Stunting;
Wasting.

ABSTRACT

This study aims to estimate the prevalence of stunting and wasting in the southern region of Sumatra using a bivariate nonparametric regression framework based on the Nadaraya-Watson Kernel and Penalized Spline estimators with Penalized Weighted Least Squares (PWLS). The analysis utilizes data from the 2023 Indonesian Toddler Nutrition Survey, comprising 60 regencies and cities across five provinces, namely Bengkulu, South Sumatra, Lampung, Jambi, and Bangka Belitung. By jointly modeling stunting and wasting as correlated response variables, this study seeks not only to compare methodological performance, but also to provide empirical insights into the nonlinear patterns underlying child nutritional outcomes influenced by maternal-child health and socioeconomic conditions. Model performance was evaluated using the Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). The empirical results indicate that the Nadaraya-Watson Kernel estimator outperforms the Penalized Spline approach, yielding a substantially lower prediction error (MSE = 0.0008), high goodness-of-fit values (R^2 of 99.98% for stunting and 99.95% for wasting), and relatively small RMSE values of 0.038 and 0.017, respectively. These findings suggest that the kernel-based estimator provides stable and accurate predictions within the data structure considered, particularly in capturing complex nonlinear relationships between predictors and nutritional outcomes. Furthermore, the results reveal that the effects of health-related and socioeconomic factors vary across different prevalence levels, underscoring the importance of nonparametric methods in accommodating heterogeneous and nonlinear response patterns. In line with previous evidence emphasizing integrated, multisectoral approaches to child nutrition improvement, the findings highlight the relevance of combining health interventions with broader social protection strategies. Nevertheless, the interpretation of results is subject to methodological caution, given the limited sample size and the aggregated nature of the data. Overall, this study demonstrates the potential of bivariate nonparametric regression as a complementary analytical tool for health data analysis and evidence-based policy formulation related to stunting and wasting reduction.

https://doi.org/10.31764/jtam.v10i2.36177

A. INTRODUCTION

Stunting and wasting remain critical public health issues in Indonesia, reflecting chronic and acute forms of child undernutrition that continue to pose serious challenges to human capital development and long-term economic growth. These two nutritional indicators are closely interrelated and often coexist, particularly in regions with persistent socioeconomic disparities. In the southern part of Sumatra, variations in household welfare, access to health services, and maternal and child health practices contribute to complex nutritional patterns that are difficult to explain using simple linear assumptions. Consequently, robust and flexible statistical modelling is required to accurately describe the underlying relationships and support evidence-based nutritional interventions.

Regression is a fundamental statistical technique employed to explain how one or more dependent variables are influenced by one or more independent variables (Montgomery & Vining, 2021). According to the estimation strategy used, regression approaches are generally grouped into parametric, nonparametric, and semiparametric methods. Parametric regression is applied when the functional form describing the relationship between the dependent and independent variables is already specified or assumed beforehand (Budiantara, 2019). However, in practice, not all data follow a specific functional form. When the form of the relationship is unknown, nonparametric regression becomes a more suitable alternative because of its high flexibility in modelling complex relationships between variables without requiring a predetermined functional form (Mestiri, 2023). Along with the development of nonparametric regression methods, this analysis has been extended beyond a single response to accommodate multiple correlated response variables (Islamiyati et al., 2022).

Nonparametric regression has progressed from handling single response variables to accommodating bivariate and multivariate responses. In a bivariate setting, two related outcome variables are modelled simultaneously (Anisar et al., 2023). Various techniques are applied within nonparametric regression, such as spline methods, kernel approaches, local polynomial estimators, Fourier series, wavelet methods, and MARS (Oktarina, Sriliana, et al., 2025). Among these approaches, kernel estimation is particularly popular because it can represent intricate data patterns by utilizing kernel functions and bandwidth parameters (Sriliana et al., 2022). This method does not depend on strict linearity assumptions and is capable of yielding smooth function estimates (Hidayat et al., 2021; Oktarina, Nugroho, et al., 2025). A widely recognized kernel-based estimator is the Nadaraya-Watson estimator, known for its flexibility in capturing nonlinear relationships (Ali et al., 2021).

In addition to kernel estimators, spline estimators are also widely applied because of their clear statistical and visual interpretation (Kirkby et al., 2023). Spline regression utilizes piecewise polynomial functions that are continuous at the connecting points known as knots (Yang et al., 2023). However, most existing studies have focused only on univariate nonparametric regression or applied each estimator separately, without directly comparing their flexibility within a bivariate framework using real data (Ali et al., 2021; Sriliana et al., 2023). This gap highlights the need for a comprehensive evaluation of nonparametric estimators particularly Kernel Nadaraya-Watson and Penalized Spline in modelling correlated response variables such as Stunting and Wasting.

Several previous studies have demonstrated the effectiveness of kernel- and spline-based nonparametric estimators in modelling complex and nonlinear relationships. Ali (2019) proposed a modification of the Nadaraya Watson kernel estimator using an adaptive bandwidth approach based on robust mean, median, and harmonic mean, which was found to be more accurate than classical methods in terms of Mean Squared Error (MSE). Meanwhile, Oktarina, Sriliana, et al. (2025) developed a Penalized Spline Semiparametric Regression model for bivariate responses and demonstrated that the Penalized Weighted Least Square (PWLS) method performed better than the Penalized Least Square (PLS) method in modelling complex relationships among poverty indicators. These findings are consistent with established nonparametric regression literature, which emphasizes the flexibility and robustness of kernel- and spline-based estimators in capturing nonlinear relationship patterns (Budiantara, 2019; Fan & Gijbels, 1996; Ruppert et al., 2003; Sriliana et al., 2023). Both studies confirmed the effectiveness of the Kernel and Spline approaches in producing flexible and accurate estimations. Therefore, this study focuses on evaluating the flexibility of both estimators within the framework of bivariate nonparametric regression, particularly in modelling Stunting and Wasting cases in Sumatra. These findings indicate that both kernel and spline estimators have advantages in modelling complex and nonlinear relationships.

Despite these methodological advances, existing studies have largely applied these methods in contexts that are not directly linked to nutritional outcomes, or have focused on methodological development without explicitly addressing interrelated indicators such as stunting and wasting. Given that stunting and wasting represent chronic and acute forms of undernutrition that are biologically and statistically interconnected, a modelling framework capable of jointly capturing their nonlinear dependence structure is essential. This motivates the need for a bivariate nonparametric regression approach that can adequately reflect the complexity of child nutritional conditions.

Based on this background, the present study focuses on evaluating the flexibility of the Kernel Nadaraya-Watson and Penalized Spline estimators within a bivariate nonparametric regression framework, with specific application to stunting and wasting cases in Sumatra. The analysis is conducted using nutritional status data of children under five years old in the southern part of Sumatra for the year 2023, where the prevalence of stunting and wasting serves as two correlated response variables representing poor nutritional conditions. By emphasizing the joint modelling of these two critical nutritional indicators, this study provides an applied comparative assessment of estimator flexibility using real regional data and contributes empirical evidence that may support data-driven nutritional intervention strategies and regional nutrition policy formulation in southern Sumatra.

B. METHODS

1. Research Design, Data Source, and Variables

This study uses an applied quantitative design based on bivariate nonparametric regression to estimate the prevalence of Stunting and Wasting in the southern part of Sumatra. The model involves two correlated response variables Stunting and Wasting and four predictor variables representing maternal, child health, and socioeconomic factors. The data were obtained from the 2023 Indonesian Toddler Nutrition Survey published by the Ministry of Health, consisting of 60 observations from regencies and cities across five provinces: Bengkulu, South Sumatra, Lampung, Jambi, and Bangka Belitung. The independent variables include X_1 (proportion of children under five receiving complete basic immunization), X_2 (proportion of children under five with low birth weight), X_3 (proportion of children under five who are exclusively breastfed), and X_4 (proportion of individuals living below the poverty line). All variables are expressed as percentages and are assumed to influence the nutritional condition of children under five in southern part of Sumatra.

2. Analysis Procedure

The analysis in this study was conducted through two stages of estimation using the Nadaraya-Watson Kernel and Penalized Spline methods.

a. The first stage involved compiling data on stunting and wasting and their associated determinants, examining relationships between response variables using Pearson's correlation, and generating scatterplot visualizations. The regression curve was then estimated by choosing an appropriate kernel, bandwidth, and weighting matrix. Model accuracy was assessed using the Coefficient of Determination (R^2) and the Root Mean Squared Error (RMSE).
b. The second stage adopted similar preliminary procedures, followed by estimating the regression model through the Penalized Weighted Least Squares (PWLS) technique. The smoothing parameter (λ) and knot placement were selected using the Generalized Cross-Validation (GCV) criterion. Each response variable was then fitted separately, and model performance was again evaluated using R^2 and RMSE before conducting further analysis and interpretation.

3. Correlation in Response

The use of weighting matrices is grounded in bivariate regression theory, where it is essential to examine whether the dependent variables are statistically related. This association can be evaluated using correlation analysis (Al Barra & Saputro, 2025). The Pearson correlation coefficient is frequently applied for this purpose, as it measures the strength of the linear relationship between two response variables. Pearson's coefficient, denoted by $\hat{\rho}$, ranges from -1 to 1 and is calculated as described in (Deng et al., 2021):

$$\hat{\rho} = \frac{s_{y^{(1)}y^{(2)}}}{s_{y^{(1)}}s_{y^{(2)}}} \tag{1}$$

Based on hypothesis testing results, conclusions can be drawn about the presence and strength of relationships between response variables, which is essential in bivariate regression analysis. The weighting matrix plays a key role in estimating parameters within bivariate nonparametric regression models, as it accounts for correlations between responses within the same observation, improving estimation accuracy (Anisar et al., 2023; Oktarina, Sriliana, et al., 2025). In such models, correlated error terms between responses are addressed by representing their covariance structure for each observation.

$$\boldsymbol{W} = \begin{bmatrix} s^2_{y^{(1)}}\boldsymbol{I} & (s_{y^{(1)}y^{(2)}})\boldsymbol{I} \\ (s_{y^{(2)}y^{(1)}})\boldsymbol{I} & s^2_{y^{(2)}}\boldsymbol{I} \end{bmatrix}^{-\mathbf{1}} \tag{2}$$

4. Nadaraya-Watson Kernel Biresponse Nonparametric Regression

The bivariate nonparametric regression using the Kernel Nadaraya-Watson estimator models relationships between two response variables and one or more predictors without assuming a specific data distribution (Sadek & Mohammed, 2024). The kernel smooths the functional relationship and applies weights reflecting each observation's influence based on its distance from the estimation point (Shi et al., 2022). This method simultaneously analyzes correlated responses common in health, economics, and social sciences by incorporating an additional weighting function to capture interdependence and improve estimation efficiency (Tosatto et al., 2021). Thus, the Nadaraya-Watson Kernel Biresponse Nonparametric Regression estimator can be formulated as (Selk & Gertheiss, 2022):

$$\widehat{\boldsymbol{\beta}}\left(\boldsymbol{t_0}\right)=\sum_{i=1}^{n}\frac{\left[\left(z_{y^{(1)}}^2+2z_{y^{(2)}y^{(1)}}+z_{y^{(2)}}^2\right)\left(K\left(\sum_{g=1}^{G}\left(\frac{t_g-t_{gi}}{h_g}\right)\right)y_i^{(r)}\right)\right]}{\frac{1}{n}\sum_{i=1}^{n}\left[\left(z_{y^{(1)}}^2+2z_{y^{(2)}y^{(1)}}+z_{y^{(2)}}^2\right)\left(K\left(\sum_{g=1}^{G}\left(\frac{t_g-t_{gi}}{h_g}\right)\right)\right)\right]} \tag{3}$$

To obtain an optimal estimation, it is essential to determine the most suitable kernel function and bandwidth. The commonly used criterion for this purpose is Generalized Cross-Validation (GCV). The advantage of the GCV method lies in its asymptotic optimality, which makes it robust and efficient across a variety of data structures (Patil & LeJeune, 2024). In this study, the selection of the kernel bandwidth is conducted by taking into account the measurement unit and scale of the response variable, in order to ensure consistency of estimation and interpretability of the results. The bandwidth parameter h regulates the smoothness of the estimated function larger bandwidths produce smoother estimates but may cause bias, while smaller bandwidths increase sensitivity to noise, potentially leading to overfitting (Misiakiewicz & Saeed, 2024). The optimal bandwidth is determined as:

$$GCV\left(h_{opt}\right)=\frac{MSE\left(h_{opt}\right)}{\left(1-2n^{-1}tr(\boldsymbol{B})\right)^2} \tag{4}$$

5. Spline-Penalized Biresponse Nonparametric Regression

Penalized splines (P-splines) offer a flexible framework for modelling relationships between one or more predictors and two response variables simultaneously. Using paired data $(t_1, t_2, \ldots, t_g)$, the Penalized Weighted Least Squares (PWLS) estimator incorporates smoothing parameters to control the roughness of the estimated function (Islamiyati et al., 2022). This approach accounts for the inverse covariance matrix of the responses, leading to the model:

$$y_i^{(r)}=\delta_0^{(r)}+\sum_{g=1}^{G}\left(\delta_{mg}^{(r)}t_{gi}^{m\,(r)}+\sum_{k=1}^{K_g}\phi_{gk}^{(r)}\left(t_{gi}-\xi_{gk}\right)_+^m\right)+\varepsilon_i^{(r)};\ r=1,2 \tag{5}$$

The estimator depends on selecting appropriate knots and a smoothing parameter λ. Knots (ξ_k) are placed at quantile points of the predictor variable, representing distinct data segments once sorted (L. Yang & Hong, 2017). In this study, the number of knots is restricted to three to maintain a parsimonious model structure and to mitigate the risk of overfitting, particularly given the limited sample size (Pahlepi et al., 2025). The location of these knots is defined as:

$$\xi_k = \frac{j}{K+1,}, k = 1,2,3,..,K \tag{6}$$

The performance of the penalized spline estimator is also controlled by the smoothing parameter λ, which regulates the trade-off between model flexibility and smoothness. Small values of λ may lead to overfitting, whereas large values may result in underfitting. The optimal smoothing parameter is selected using the GCV criterion, defined as:

$$GCV(\xi_{opt}, \lambda_{opt}) = \frac{MSE(\xi_{opt}, \lambda_{opt})}{\left(1 - 2n^{-1}tr(\boldsymbol{A})\right)^2} \tag{7}$$

The GCV criterion serves as an implicit validation mechanism that controls both overfitting and underfitting in the penalized spline regression model. By minimizing the GCV value, the selected smoothing parameter λ achieves an optimal balance between bias and variance, ensuring stable model generalization, particularly under limited sample size conditions.

6. Goodness of Evaluation Model

Assessing the performance of a model is crucial to determine its reliability and usefulness in research and policy analysis. Two widely applied indicators to measure model quality are the Coefficient of Determination (R^2) and the Root Mean Squared Error (RMSE) (Chicco et al., 2021). The R^2 statistic describes the proportion of variability in the observed data that can be explained by the model. Its value ranges from -1 (very poor fit) to +1 (perfect fit), where values closer to 1 reflect stronger explanatory capability. The formula for R^2 is defined as:

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{8}$$

RMSE is another frequently used measure to determine the predictive accuracy of a model by calculating the average magnitude of prediction errors. It is computed as the square root of the mean of the squared differences between observed and predicted outcomes. A lower RMSE suggests better model accuracy, particularly when the residuals approach a normal distribution (Pahlepi et al., 2025). The mathematical expression for RMSE is given by (Hodson, 2022):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{7}$$

Together, R^2 and RMSE provide a comprehensive evaluation of model quality, where R^2 indicates explanatory strength, and RMSE represents predictive precision.

C. RESULT AND DISCUSSION

1. Correlation between Response Variables

To examine the dependence structure between the two response variables, Pearson's correlation coefficient was employed. This analysis aims to identify whether a statistically significant linear association exists between stunting and wasting rates. The results of the correlation analysis are presented in Table 1.

Table 1. Correlation output of $\boldsymbol{Y^{(1)}}$ and $\boldsymbol{Y^{(2)}}$

$\boldsymbol{H_0: \rho = 0}$	
Statistics	**Value**
$\hat{\rho}$	0.677
t-test	7.022
t-crit	2.301
p-value	<0.001

Based on the correlation test, the estimated correlation coefficient between stunting and wasting is $\hat{\rho} = 0.677$. The corresponding t-value is 7.022, which exceeds the critical value ($t_{\text{crit}} = 2.301$), with a p-value less than 0.001. These results indicate a statistically significant positive linear relationship between the two response variables.

2. Visualization of Relationship Pattern of Response Variable and Each Predictor Variable

Prior to model estimation, scatterplots were constructed to explore the association patterns between the response variables and each predictor variable. The use of scatterplots enables preliminary identification of possible linear or nonlinear relationships, data clustering, and the presence of extreme observations. This exploratory visualization phase provides an initial description of the distributional patterns between predictors and the response variables.

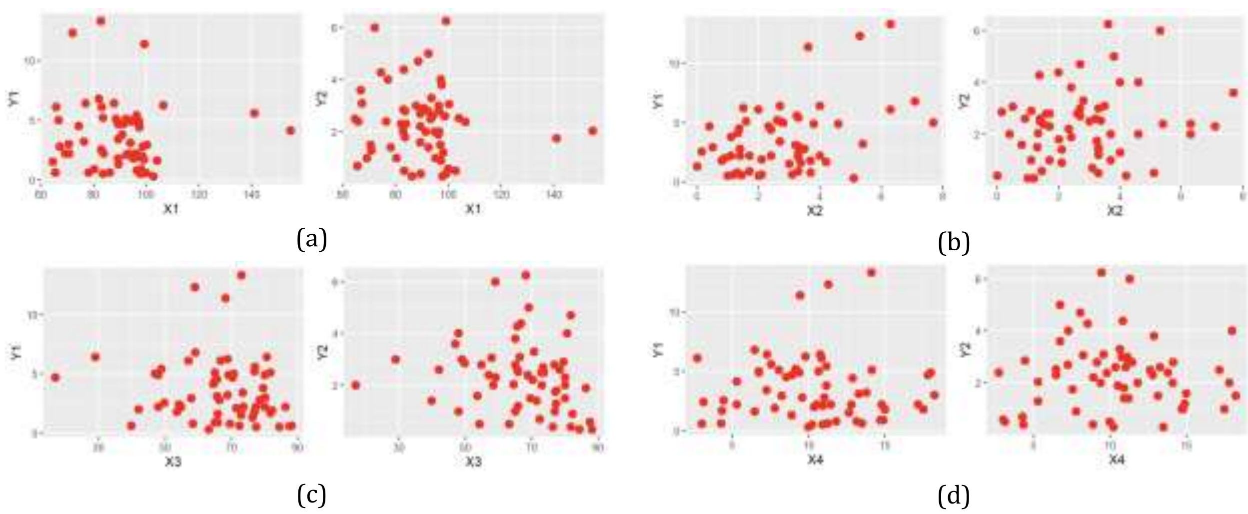


Figure 1. Scatter Pattern of The Relationship Between Response Variables and Predictor Variables (a) $\boldsymbol{X_1}$, (b) $\boldsymbol{X_2}$, (c) $\boldsymbol{X_3}$, (d) $\boldsymbol{X_4}$

Figure 1 shows that the data points are widely scattered across the range of each predictor variable and do not exhibit a clear linear pattern between the predictors and the response variables. The absence of a clear linear pattern in the scatterplots suggests potential nonlinear relationships between predictors and stunting and wasting outcomes. This empirical pattern supports the application of nonparametric regression methods to capture complex associations in child nutritional data.

3. Nadaraya-Watson Kernel Biresponse Nonparametric Regression

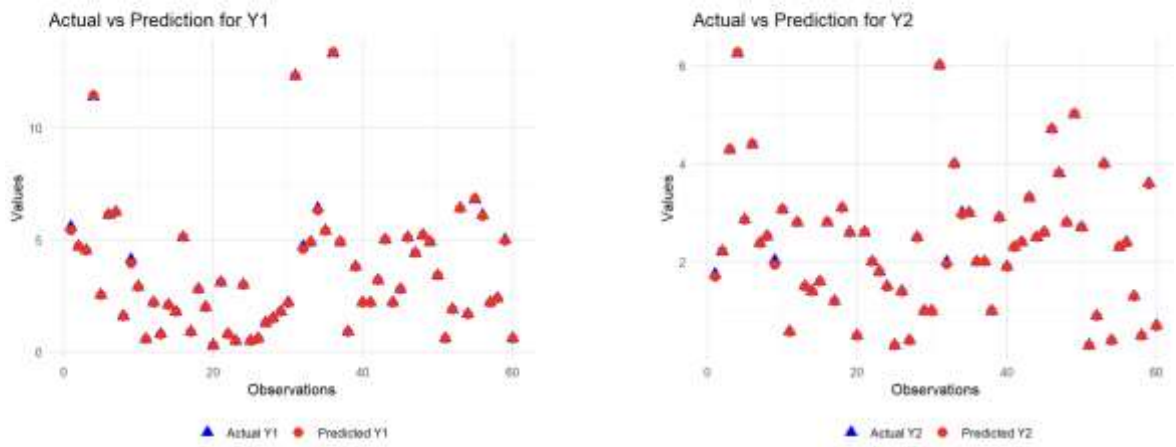
The bivariate Nadaraya-Watson kernel regression models the relationship between two responses and multiple predictors without assuming a parametric form. Three kernel functions Epanechnikov, Gaussian, and Biweight were used, with optimal bandwidth selected using the Generalized Cross-Validation (GCV) method. Each observation contributes to the weighted average of responses based on proximity, producing smooth local estimates. The estimated results based on three kernel functions are shown in Table 2.

Table 2. Comparison of GCV Values of Kernel Functions

Kernel Function	GCV	Combination Number	Optimal Bandwidth t_1, t_2, t_3, t_4
Epanechnikov	0.00425	159900	100, 75, 100, 100
Gaussian	0.00090	159920	100, 80, 100, 100
Biweight	0.12373	160000	100, 100, 100, 100

Based on Table 2, the GCV values differ across the three kernel functions. The Gaussian kernel yields the lowest GCV value (0.00090), followed by the Epanechnikov kernel (0.00425) and the Biweight kernel (0.12373). The corresponding optimal bandwidth combination for the Gaussian kernel is (100, 80, 100, 100). The lower GCV value obtained by the Gaussian kernel indicates superior estimation performance compared to the Epanechnikov and Biweight kernels in the bivariate Nadaraya-Watson framework. This result suggests that the Gaussian kernel provides a more effective balance between bias and variance, which is essential for capturing complex and nonlinear relationships between predictors and response variables.

Accordingly, the Gaussian kernel is identified as the most appropriate kernel function for this study. The smooth nature of the Gaussian function allows more flexible adaptation to data patterns, leading to stable and consistent predictions for both response variables, namely stunting prevalence ($y^{(1)}$) and wasting prevalence ($y^{(2)}$), across provinces in the southern part of Sumatra. Therefore, subsequent analyses and visualizations of prediction results are conducted using the Gaussian kernel, as shown in Figure 2.



(a) (b)

Figure 2. Plot of Kernel Prediction Results (a) Response 1, (b) Response 2

4. Spline-Penalized Biresponse Nonparametric Regression

The Penalized Spline estimation, following the Kernel Nadaraya-Watson model, uses the Penalized Weighted Least Squares (PWLS) method to improve precision and reduce bias. The model's flexibility depends on the polynomial order, smoothing parameter (λ), and number of knots. In this study, three polynomial orders (linear, quadratic, cubic) and λ values ranging from 5 to 100 were tested, while the number and placement of knots were optimized using the minimum Generalized Cross-Validation (GCV) value to balance smoothness and flexibility. The optimal combination of order, knot location, number of knots, and λ value is summarized in Table 3.

Table 3. Optimal Knot Location

ξ_1	ξ_2	ξ_3	ξ_4
82.00	1.36	58.85	8.24
90.80	2.50	69.00	10.76
97.20	2.63	77.13	13.15
$Orde$ (3,3)			
$\lambda = 10$			
$GCV = 8.639 \times 10^{-8}$			

The selected model corresponds to polynomial order (3,3) with λ = 10, yielding a minimum GCV value of 8.639×10^{-8}. The relationship between λ and the GCV value is illustrated in Figure 3.

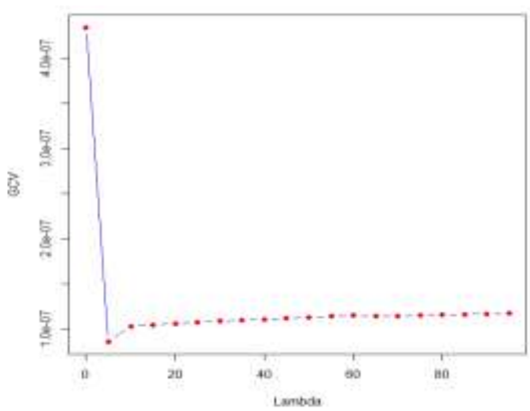


Figure 3. Increase in Lambda Value Against GCV

Based on the optimal combination of order, knot locations, number of knots, and lambda value, the bivariate nonparametric Penalized Spline regression models for both response variables are obtained as follows:

$$\begin{aligned}\hat{y}_i^{(1)} = {} & -4.431 + 1.224t_{1i} - 0.0003t_{1i}^2 - 0.000058t_{1i}^3 - 0.00059(t_{1i} - 82.000)_+^3 \\ & - 0.000038(t_{1i} - 90.800)_+^3 - 0.00059(t_{1i} - 97.200)_+^3 - 1.127t_{2i} + 0.386t_{2i}^2 \\ & - 0.016t_{2i}^3 + 0.00011(t_{2i} - 1.360)_+^3 + 0.003(t_{2i} - 2.500)_+^3 + 0.005(t_{2i} \\ & - 3.630)_+^3 + 0.754t_{3i} - 0.0251t_{3i}^2 + 0.00023t_{3i}^3 - 0.0021(t_{3i} - 58.850)_+^3 \\ & + 0.0051(t_{3i} - 69.000)_+^3 - 0.0099(t_{3i} - 77.130)_+^3 - 29.950t_{4i} + 4.260t_{4i}^2 \\ & - 0.188t_{4i}^3 + 0.185(t_{4i} - 8.240)_+^3 + 0.058(t_{4i} - 10.760)_+^3 - 0.0028(t_{4i} \\ & - 13.150)_+^3\end{aligned}$$

$$\begin{aligned}\hat{y}_i^{(2)} = {} & -0.0095 + 0.5142t_{1i} - 0.0054t_{1i}^2 + 1.9742 \times 10^{-5}t_{1i}^3 - 0.0003(t_{1i} - 82.000)_+^3 \\ & + 0.0009(t_{1i} - 90.800)_+^3 - 0.0006(t_{1i} - 97.200)_+^3 + 0.3607t_{2i} - 0.1018t_{2i}^2 \\ & + 0.0104t_{2i}^3 - 0.0006(t_{2i} - 1.360)_+^3 - 0.0005(t_{2i} - 2.500)_+^3 + 0.0059(t_{2i} \\ & - 3.630)_+^3 + 0.2216t_{3i} - 0.0059t_{3i}^2 + 0.00005t_{3i}^3 - 0.0009(t_{3i} - 58.850)_+^3 \\ & + 0.0031(t_{3i} - 69.000)_+^3 - 0.0064(t_{3i} - 77.130)_+^3 - 8.8032t_{4i} + 1.3096t_{4i}^2 \\ & - 0.0596t_{4i}^3 + 0.0596(t_{4i} - 8.240)_+^3 + 0.0205(t_{4i} - 10.760)_+^3 + 0.0180(t_{4i} \\ & - 13.150)_+^3\end{aligned}$$

After establishing the PWLS model for the Stunting and Wasting case study in the Part of Southern Sumatra in 2023, the next step is to segment the population based on predictor variables and interpret the regression coefficients to understand the relative influence of each variable on children's nutritional status. Through this analysis, it is expected to gain a deeper understanding of the factors affecting nutritional status, serving as a basis for more effective decision-making and policy planning. The segmentation results for each variable are presented as follows.

$$\begin{aligned}f^{(1)}(t_{1i}) = {} & 1.224t_{1i} - 0.0003t_{1i}^2 - 0.000058t_{1i}^3 - 0.00059(t_{1i} - 82.000)_+^3 - 0.000038(t_{1i} \\ & - 90.800)_+^3 - 0.00059(t_{1i} - 97.200)_+^3\end{aligned}$$

$$f^{(1)}(t_{1i}) = \begin{cases} 1.224t_{1i} - 0.0003t_{1i}^2 - 0.000058t_{1i}^3. & 0 < t_{1i} \leq 82.00 \\ -10.679t_{1i} + 0.1448t_{1i}^2 - 0.000648t_{1i}^3 325.307. & 82.00 < t_{1i} \leq 90.80 \\ -11.618t_{1i} + 0.1551t_{1i}^2 - 0.00068t_{1i}^3 - 353.74. & 90.80 < t_{1i} \leq 97.20 \\ -28.3406t_{1i} + 0.3274t_{1i}^2 - 0.00127t_{1i}^3 - 895.568. & t_{1i} > 97.20 \end{cases}$$

Based on the segmentation results of the Low Birth Weight (LBW) Percentage variable, the effect of t_{1i} on stunting prevalence ($y^{(1)}$) varies across four segments. In Segment I, an increase in t_{1i} raises stunting prevalence, while in Segments II and III, further increases reduce it, suggesting improved adaptation or interventions. In Segment IV, the decline continues but becomes more gradual, indicating that at very high LBW levels, additional increases have a diminishing effect on stunting reduction.

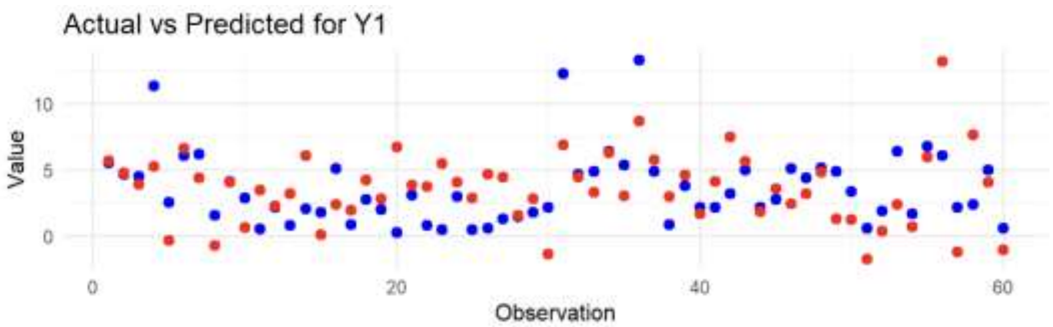
$$f^{(2)}(t_{1i}) = 0.5142t_{1i} - 0.0054t_{1i}^2 + 1.9742 \times 10^{-5}t_{1i}^3 - 0.0003(t_{1i} - 82.000)_+^3 + 0.0009(t_{1i} - 90.800)_+^3 - 0.0006(t_{1i} - 97.200)_+^3$$

$$f^{(2)}(t_{1i}) = \begin{cases} 0.5142t_{1i} - 0.0054t_{1i}^2 + 1.9742 \times 10^{-5}t_{1i}^3. & 0 < t_{1i} \leq 82.00 \\ -5.7393t_{1i} + 0.0708t_{1i}^2 - 0.00029t_{1i}^3 + 170.924. & 82.00 < t_{1i} \leq 90.80 \\ 16.2738t_{1i} - 0.0708t_{1i}^2 + 0.0006t_{1i}^3 - 495.353. & 90.80 < t_{1i} \leq 97.20 \\ -1.582t_{1i} + 0.0121t_{1i}^2 - 0.00003t_{1i}^3 + 83.1949. & t_{1i} > 97.20 \end{cases}$$

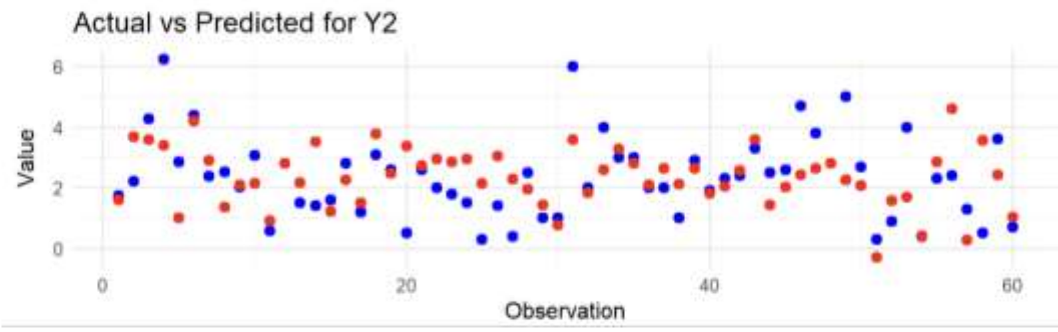
The segmentation results of the Low Birth Weight (LBW) Percentage variable show a nonlinear relationship with wasting prevalence ($y^{(2)}$). In Segment I, an increase in t_{1i}raises wasting prevalence, while in Segment II it decreases, suggesting compensatory health responses. In Segment III, the relationship turns positive again, and in Segment IV, wasting slightly declines. Overall, the LBW-wasting relationship alternates between positive and negative effects, indicating that its influence varies across prevalence levels and requires region-specific intervention strategies.

The findings of this study indicate that stunting and wasting are influenced by a complex interaction between maternal child health aspects and socioeconomic conditions. This result is consistent with Bhutta et al. (2025), who emphasize that improvements in child growth require integrated, multisectoral approaches rather than nutrition-specific interventions alone. The observed nonlinear relationships suggest that the effects of these factors vary across different prevalence levels, highlighting the relevance of nonparametric methods in capturing such complexity. These findings underscore the importance of comprehensive health and social protection policies in efforts to reduce stunting and wasting.

Overall, the relationship between low birth weight and wasting prevalence is nonlinear, alternating between positive and negative effects across intervals. This indicates that LBW's impact on child nutrition varies by prevalence level, requiring region-specific interventions. The same segmentation and interpretation apply to all predictors, reflecting their distinct nonlinear effects on both response variables. Based on the estimated functions $\hat{y}^{(1)}$and $\hat{y}^{(2)}$, prediction results are plotted against actual data in Figure 4.



(a)



(b)

Figure 4. Plot of Penalized Spline Prediction Results (a) Response 1, (b) Response 2

5. Model Selection and Accuracy

The best model was selected based on the Mean Squared Error (MSE), where the Kernel Nadaraya-Watson estimator achieved the lowest value (0.0008) compared to the Penalized Spline model (4.621), indicating higher predictive accuracy. The Kernel model also showed excellent performance with R^2 values of 99.98% (Response 1) and 99.95% (Response 2), and low RMSE values of 0.038 and 0.017, respectively. These results confirm that the Kernel Nadaraya-Watson model provides superior accuracy and reliability, making it the most suitable for further analysis.

Previous studies have shown that kernel- and spline-based nonparametric estimators are effective in modeling complex and nonlinear relationships. Adaptive Kernel Nadaraya–Watson approaches have been reported to improve estimation accuracy in terms of MSE, while Penalized Spline models using Penalized Weighted Least Squares (PWLS) demonstrate superior performance in capturing nonlinear patterns in bivariate settings. Overall, the literature consistently highlights the flexibility and robustness of kernel and spline estimators, supporting their suitability for modeling correlated responses such as stunting and wasting.

D. CONCLUSION

The objectives stated in the Introduction have been achieved through the Results and Discussion. This study analyzed the relationship between stunting and wasting prevalence in the southern part of Sumatra using bivariate nonparametric regression models. Based on the GCV criterion, the Kernel Nadaraya-Watson model with a Gaussian kernel yielded lower GCV values than the Penalized Spline model, indicating more stable estimation performance under the given data structure. The results also confirm a significant positive correlation between stunting and wasting, while the estimated models reveal nonlinear and segment-specific relationships between nutritional outcomes and their associated factors. These findings demonstrate that flexible nonparametric approaches are appropriate for capturing the complexity of child nutrition patterns, particularly in settings with limited sample sizes.

Although nonparametric regression does not rely on strict distributional assumptions, future studies are encouraged to incorporate explicit outlier detection and control procedures. Given the flexibility of nonparametric models and the limited sample size, influential or extreme observations may disproportionately affect local estimation and smoothing behavior. Implementing robust preprocessing or outlier-handling strategies can improve model stability and enhance the reliability of empirical interpretations.

ACKNOWLEDGEMENT

The author sincerely acknowledges the support provided by the Research and Community Service Institute (LPPM) of the University of Bengkulu, as well as the Directorate General of Higher Education, Science, and Technology (Kemdiktisaintek), whose assistance and facilitation greatly contributed to the completion of this research. The author also extends sincere appreciation to the Institut Teknologi Batam (ITEBA) for the academic guidance and institutional encouragement provided throughout this work. Furthermore, the organizing committee of the Brawijaya International Conference on Statistical Modeling and Data Science (BICOSDA 2025) is gratefully acknowledged for offering the opportunity to present and disseminate this study in an international academic forum.

REFERENCES

Al Barra, A. F., & Saputro, D. R. S. (2025). Knot Optimization for Bi-Response Spline Nonparametric Regression With Generalized Cross-Validation (Gcv). *Barekeng*, *19*(1), 271–280. https://doi.org/10.30598/barekengvol19iss1pp0271-0280

Ali, T. H. (2019). Modification of The Adaptive Nadaraya-Watson Kernel Method for Nonparametric Regression (Simulation Study). *Communications in Statistics: Simulation and Computation*, *51*(2), 391–403. https://doi.org/10.1080/03610918.2019.1652319

Ali, T. H., Hayawi, H. A. A. M., & Botani, D. S. I. (2021). Estimation of the bandwidth parameter in Nadaraya-Watson kernel Non-Parametric Regression Based on Universal Threshold Level. *Communications in Statistics: Simulation and Computation*, *52*(4), 1476–1489. https://doi.org/10.1080/03610918.2021.1884719

Anisar, A. P., Sifriyani, S., & Dani, A. T. R. (2023). Estimation of a Bi-Response Truncated Spline Nonparametric Regression Model on Life Expectancy and Prevalence of Underweight Children in Indonesia. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, *17*(4), 2011–2022. https://doi.org/10.30598/barekengvol17iss4pp2011-2022

Bhutta, A., Islam, M., Gaffey, M. F., Victora, C. G., Menon, P., Katz, J., Horton, S. E., Yearwood, J., & Black, R. E. (2025). What works for reducing stunting in low-income and middle-income countries ? Cumulative learnings from the Global Stunting Exemplars Project. *The American Journal of Clinical Nutrition*, *121*(March), 113–128. https://doi.org/10.1016/j.ajcnut.2025.03.004

Budiantara, I. N. (2019). *Regresi Nonparametrik Spline Truncated*. ITS Press.

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The Coefficient of Determination R-squared is More Informative Than SMAPE, MAE, MAPE, MSE and RMSE in Regression Analysis Evaluation. *PeerJ Computer Science*, *7*(e623), 1–24. https://doi.org/10.7717/PEERJ-CS.623

Deng, J., Deng, Y., & Cheong, K. H. (2021). Combining conflicting evidence based on Pearson correlation coefficient and weighted graph. *International Journal of Intelligent Systems*, *36*(12), 7443–7460. https://doi.org/10.1002/int.22593

Fan, J., & Gijbels, I. (1996). *Lokal Polynomial Modelling and Its Applications* (1st ed., Issue 1985). Springer Science Business Media.

Hidayat, R., Budiantara, I. N., Otok, B. W., & Ratnasari, V. (2021). The regression curve estimation by using mixed smoothing spline and kernel (MsS-K) model. *Communications in Statistics - Theory and Methods*, *50*(17), 3942–3953. https://doi.org/10.1080/03610926.2019.1710201

Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. *Geoscientific Model Development*, *15*(14), 5481–5487. https://doi.org/10.5194/gmd-15-5481-2022

Islamiyati, A., Anisa, A., Zakir, M., Sirajang, N., Sari, U., Affan, F., & Usrah, M. J. (2022). The Use of Penalized Weighted Least Square To Overcome Correlations Between Two Responses. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, *16*(4), 1497–1504. https://doi.org/10.30598/barekengvol16iss4pp1497-1504

Kirkby, J. L., Leitao, Á., & Nguyen, D. (2023). Spline local basis methods for nonparametric density estimation. *Statistics Surveys*, *17*(2023), 75–118. https://doi.org/10.1214/23-SS142

Mestiri, S. (2023). *Nonparametric regression models : theories and Sami Mestiri To cite this version : HAL Id : hal-04004433 Nonparametric regression models : theories and applications in R*. https://hal.science/hal-04004433v1

Misiakiewicz, T., & Saeed, B. (2024). *A non-Asymptotic Theory of Kernel Ridge Regression: Deterministic Equivalents, Test Error, and GCV Estimator* (pp. 1–131). arXiv. http://arxiv.org/abs/2403.08938

Montgomery, D. C., & Vining, E. A. P. G. G. (2021). Linear Regression Analysis. In *Mechanical Reliability Improvement* (6th ed.). JOHN WILEY & SONS, INC. https://doi.org/10.1201/9780203910993.ch7

Oktarina, C. R., Nugroho, S., Sriliana, I., & Novianti, P. (2025). *Estimation of Stunting and Wasting in Sumatra 2022 with Nadaraya-Watson Kernel and Penalized Spline*. *8*(3), 197–208.

Oktarina, C. R., Sriliana, I., & Nugroho, S. (2025). Penalized Spline Semiparametric Regression for Bivariate Response in Modeling Macro Poverty Indicators. *Indonesian Journal of Applied Statistics*, *8*(1), 13–23. https://doi.org/https://doi.org/10.13057/ijas.v8i1.94370

Pahlepi, R., Sriliana, I., Agwil, W., & Oktarina, C. R. (2025). Biresponse Spline Truncated Nonparametric Regression Modeling for Longitudinal Data on Monthly Stock Prices of Three Private Banks in Indonesia. *Barekeng*, *19*(4), 2467–2480. https://doi.org/10.30598/barekengvol19iss4pp2467-2480

Patil, P., & LeJeune, D. (2024). *Asymptotically free sketched ridge ensembles: Risks, cross-validation, and tuning* (pp. 1–45). ArXiv. http://arxiv.org/abs/2310.04357

Ruppert, D., Wand, M. P., & Carroll, R. . (2003). Semiparametric Regression. In *Regression Modeling*. Cambridge University Press. https://doi.org/10.1201/9781420091984-c17

Sadek, A. M., & Mohammed, L. A. (2024). Evaluation of the Performance of Kernel Non-parametric Regression and Ordinary Least Squares Regression. *International Journal on Informatics Visualization*, *8*(3), 1352–1360. https://doi.org/10.62527/joiv.8.3.2430

Selk, L., & Gertheiss, J. (2022). Nonparametric Regression and Classification with Functional, Categorical, and Mixed Covariates. *Advances in Data Analysis and Classification*, *17*(2), 519–543. https://doi.org/10.1007/s11634-022-00513-7

Shi, X., Jiang, D., Qian, W., & Liang, Y. (2022). Application of the Gaussian Process Regression Method Based on a Combined Kernel Function in Engine Performance Prediction. *ACS Omega*, *7*(45), 41732–41743. https://doi.org/10.1021/acsomega.2c05952

Sriliana, I., Budiantara, I. N., & Ratnasari, V. (2022). A Truncated Spline and Local Linear Mixed Estimator in Nonparametric Regression for Longitudinal Data and Its Application. *Symmetry*, *14*(12), 1–19. https://doi.org/10.3390/sym14122687

Sriliana, I., Budiantara, I. N., & Ratnasari, V. (2023). Poverty gap index modeling in Bengkulu Province using truncated spline regression for longitudinal data. *AIP Conference Proceedings*, *2975*(1). https://doi.org/https://doi.org/10.1063/5.0181178

Tosatto, S., Akrour, R., & Peters, J. (2021). An Upper Bound of the Bias of Nadaraya-Watson Kernel Regression Under Lipschitz Assumptions. *Stats*, *4*(1), 1–17. https://doi.org/10.3390/stats4010001

Yang, G., Zhang, B., & Zhang, M. (2023). Estimation of Knots in Linear Spline Models. *Journal of the American Statistical Association*, *118*(541), 639–650. https://doi.org/10.1080/01621459.2021.1947307

Yang, L., & Hong, Y. (2017). Adaptive Penalized Splines for Data Smoothing. *Computational Statistics and Data Analysis*, *108*(2017), 70–83. https://doi.org/10.1016/j.csda.2016.10.022