BAREKENG: Journal of Mathematics and Its Applications
September 2026 Volume 20 Issue 3 Page 2179-2194
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol20iss3pp2179-2194

EVALUATION OF THE FLEXIBILITY OF NADARAYA-WATSON KERNEL AND PENALIZED SPLINE ESTIMATORS IN BIVARIATE RESPONSE NONPARAMETRIC REGRESSION MODELS

Cinta Rizki Oktarina[1], Sigit Nugroho[2*], Idhia Sriliana[3]

[1,2,3]*Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Bengkulu*
Jln. WR. Supratman, Kandang Limun, Bengkulu, 38371, Indonesia

*Corresponding author's e-mail: * **snugroho@unib.ac.id***

Article Info

Article History:

Received: 5th August 2025
Revised: 22nd November 2025
Accepted: 9th March 2026
Available online: 8th April 2026

Keywords:

Bandwidth;
Bivariate nonparametric regression;
Kernel Nadaraya-Watson;
Knot;
Penalized Spline.

ABSTRACT

Nonparametric regression is a flexible approach used when the functional relationship between predictors and responses is unknown. In the context of multiple responses, bivariate nonparametric regression allows modeling two correlated response variables, such as stunting and wasting prevalence, which remain critical issues in public health. This study aims to evaluate the flexibility and performance of two nonparametric estimators, the Nadaraya-Watson Kernel and the Penalized Spline, for modeling bivariate response data. The research was conducted in two stages: (1) simulation using variations in sample sizes (50, 100, 150, 200) and error variances based on exponential and trigonometric functions, and (2) application to real data on stunting and wasting prevalence in Indonesia (2024) obtained from Statistics Indonesia (BPS), with socioeconomic and health-related predictors. Model performance was assessed using RMSE, MSE, and R-squared, complemented by MANOVA, orthogonal polynomial contrasts, and Tukey's post-hoc test to examine significant differences across scenarios. Simulation results indicate that the Nadaraya-Watson Kernel estimator consistently outperformed the Penalized Spline, providing lower RMSE and MSE values and greater stability, particularly for larger sample sizes and smaller error variances. Orthogonal polynomial analysis revealed a quadratic relationship between sample size and RMSE, with occasional cubic patterns, while error variance consistently exhibited a quadratic trend. In the applied study, the Nadaraya-Watson Kernel with a Gaussian kernel achieved high accuracy, with an MSE of 0.00086 and an R-squared value indicating a strong model fit. However, this high R-squared value may reflect potential overfitting, which warrants further validation through cross-validation. These findings demonstrate that the Nadaraya-Watson Kernel offers an effective approach for bivariate nonparametric regression, supporting data-driven policy decisions in nutrition and public health.

How to cite this article:

C. R. Oktarina, S. Nugroho, and I. Sriliana, "EVALUATION OF THE FLEXIBILITY OF NADARAYA-WATSON KERNEL AND PENALIZED SPLINE ESTIMATORS IN BIVARIATE RESPONSE NONPARAMETRIC REGRESSION MODELS", *BAREKENG: J. Math. & App.*, vol. 20, no. 3, pp. 2179-2195, Sep, 2026.

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

Research Article · Open Access

1. INTRODUCTION

Regression analysis is a method used to explain how one or more response variables depend on one or more predictor variables [1]. There are three main approaches to estimating the regression curve: parametric, nonparametric, and semiparametric regression [2]. If the functional form of the relationship between predictors and responses is known, parametric regression can be applied [3]. However, in practice, data often do not follow any specific pattern. When the relationship between predictor and response variables is unknown, nonparametric regression is more appropriate for modeling such relationships [3], [4]. Nonparametric regression analysis has evolved beyond univariate response analysis to include bivariate and multivariate response modeling. Bivariate response regression involves two correlated response variables. Significant correlation between the response variables is a key requirement for implementing this approach [5]. Various functions have been employed in nonparametric regression modeling, including popular ones such as splines, kernels, local polynomials, Fourier series, wavelets, MARS, and others [6]. Nonparametric regression using the kernel approach, or local averaging regression, is particularly useful when data points are unevenly spaced or when predictors are stochastic. The kernel method relies on a kernel function and bandwidth for estimation. It is known for its flexibility, simple mathematical formulation, relatively fast convergence, and computational efficiency [7]. Despite these advantages, the kernel approach can be compared with the spline regression method, which is also highly flexible.

Spline regression offers both strong statistical and visual interpretability [3]. It involves piecewise polynomial functions that are continuous across their domain [8]. Splines use connecting points called knots, which allow for flexible modeling of complex data patterns [3]. Besides the selection of knot locations and the number of knots, determining the optimal smoothing parameter λ is crucial. This leads to the penalized spline regression model. [9] introduced spline functions for approximating univariate nonparametric regression curves, and developed M-type splines to handle outliers. Further, [10] applied Bayesian approaches to solve multivariate nonparametric spline regression problems. Research on kernel and spline methods has been widely applied. [11] studied regional forecasting of $PM_{2.5}$ concentrations using a novel model based on empirical orthogonal function analysis and the Nadaraya-Watson kernel estimator. Their model achieved an average prediction accuracy of 74.38% and over 92% cumulative variance explained with varying bandwidths by season. Another study by [12] focused on outlier detection using penalized spline regression to model the poverty depth index as the response variable. Their model achieved an R-square of 69.10% with optimal knot numbers of 1, 2, 4, 1, 5, 3, and 1 for each predictor.

This study employs two nonparametric regression estimators, the Nadaraya-Watson Kernel and Penalized Spline, to estimate a bivariate response nonparametric regression model to capture complex data patterns more effectively. It also evaluates the flexibility and performance of both estimators using both simulated and real datasets. Nonparametric regression has gained popularity due to its ability to model complex relationships without assuming specific functional forms. Bivariate approaches using kernel and spline methods offer greater flexibility, especially when analyzing two interrelated response variables, such as stunting and wasting in child nutrition studies. Based on the Indonesian Nutritional Status Survey (SSGI), stunting and wasting prevalence in Indonesia has declined since 2019, yet remains above WHO thresholds (20% for stunting and 5% for wasting). Nationally, the stunting rate dropped from 24.4% in 2021 to 21.6% in 2022, while wasting decreased from 7.1% to 6.2%. With the national target set to reduce stunting to 14% and wasting below 5% by 2024, this goal has now passed, and the findings of this study are applied to evaluate stunting and wasting data in Indonesia. The novelty of this study lies in the integration of bivariate nonparametric regression with comprehensive performance evaluation using MANOVA, orthogonal polynomial contrasts, and Tukey's post-hoc tests, applied to both simulated scenarios and up-to-date empirical data on stunting and wasting. To our knowledge, no prior research has combined these approaches for modeling child nutrition indicators in Indonesia.

2. RESEARCH METHODS

2.1 Nonparametric Regression

Nonparametric regression offers high flexibility in estimating curves from data without assuming a specific functional form as in parametric models. Generally, nonparametric regression models only assume

that the regression function has an infinite-dimensional form [3]. The general form of a nonparametric regression model is as follows [13]:

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

There are several functions used in nonparametric regression modeling, including spline and kernel [14].

2.2 Kernel Nadaraya-Watson

Kernel nonparametric regression, or local averaging regression, is a widely applied method in situations where data points are unevenly distributed or the predictor variable is random. This approach is based on a weighted average of the response variable, where the weights are determined by the distance between predictor observations, measured using a bandwidth parameter (h) [15]. Kernel nonparametric regression is derived from local polynomial regression, specifically regarded as a special case of zero-degree polynomial regression, known as the local constant approach. The Nadaraya-Watson kernel estimator relies on the smoothing parameter, bandwidth (h), which governs the smoothness of the curve, larger values of h result in smoother curve estimates [16]. In addition to bandwidth selection, the choice of kernel function also plays a crucial role in estimating the response function. An appropriate kernel function enhances the accuracy of the estimated function. The closer the estimate is to the true underlying function, the more efficient the estimation becomes [17]. This study employs three types of kernels, namely Epanechnikov, Gaussian, and Biweight [18]. Where $K(t_{\mathrm{I}})$ is the kernel function used as a weight in the kernel estimator involving the bandwidth parameter [19]. A kernel is a real-valued, continuous, bounded, and symmetric function [3]. Based on this definition, the kernel function possesses these two key properties [20]:

$$1. \int_{-\infty}^{\infty} K(t)dt = 1, \tag{2}$$

$$2. \int_{-\infty}^{\infty} tK(t)dt = 0. \tag{3}$$

By assuming that the response variable and the predictor variable are independent, the estimator $\hat{m}(t_i)$ and the addition of $\frac{1}{n}$ to the denominator is necessary to normalize the kernel weights, allowing the estimator $\hat{m}(t_i)$ to converge more quickly and yield more stable results, particularly for large sample sizes [3]. With this adjustment, the revised form of the estimator is expressed as:

$$\hat{m}(t_i) = \frac{K\left(\sum_{g=1}^{G} \frac{t_g - t_{gi}}{h_g}\right) y_i}{\sum_{i=1}^{n} K\left(\sum_{g=1}^{G} \frac{t_g - t_{gi}}{h_g}\right)} \tag{4}$$

$$= W_h(t_i) y_i. \tag{5}$$

The weight function $W_h(t_i)$ can be defined as follows:

$$W_h(t_i) = \begin{bmatrix} W_h(t_1) & 0 & \cdots & 0 \\ 0 & W_h(t_2) & \vdots & \vdots \\ 0 & \vdots & \ddots & 0 \\ 0 & \cdots & 0 & W_h(t_n) \end{bmatrix}. \tag{6}$$

2.3 Penalized Spline

Spline regression utilizes piecewise and continuous polynomial functions [4]. Penalized spline regression offers high flexibility in estimating the function y [7], which is assumed to be smooth and to belong to the Sobolev space $V_2^q(a, b)$. One of the main advantages of spline regression is its ability to capture changes in data patterns. According to [4], nonparametric spline regression involves knots as crucial points where the behavior of the data or function changes. In addition to the location and number of knots, selecting an optimal smoothing parameter λ is essential. In general, a spline function of order m with the j knot for each response can be expressed as follows [13]:

$$g(t_i) = \delta_0 + \sum_{g=1}^{G} \sum_{m=1}^{M} \left(\delta_{mg} t_{gi}^m + \sum_{k=1}^{K} \phi_{gk} \left(t_{gi} - \xi_{gk}\right)_+^m \right). \tag{7}$$

With the truncated function defined as follows [4]:

$$\left(t_{gi}-\xi_{gk}\right)_{+}^{m}=\begin{cases}\left(t_{gi}-\xi_{gk}\right)^{m}, t_{gi}\geq\xi_{gk}\\ 0 \qquad\qquad , t_{gi}<\xi_{gk}\end{cases}. \tag{8}$$

2.4 Bivariate Response Regression

In bivariate response regression, the relationship between response variables can be measured using the Pearson correlation. The correlation coefficient $\hat{\rho}$ ranges from $-1\leq\hat{\rho}\leq 1$ and is calculated using Eq. (9):

$$\hat{\rho}=\frac{s_{y^{(1)}y^{(2)}}}{s_{y^{(1)}}s_{y^{(2)}}}. \tag{9}$$

If $\hat{\rho}$ approaches 1, it indicates a strong and positive relationship between $y^{(1)}$ and $y^{(2)}$. If it approaches -1, the relationship is strong and negative. When $\hat{\rho}$ is close to 0, it suggests a very weak or no relationship. The steps for testing the Pearson correlation hypothesis are as follows: First, the hypotheses are defined as $H_0\colon\rho=0$ and $H_1\colon\rho\neq 0$. The significance level is set at $\alpha=0.05$. The test statistic for detecting zero correlation is calculated using the following formula [8]:

$$t_{stat}=\left|\frac{\hat{\rho}\sqrt{n-2}}{\sqrt{1-\hat{\rho}^2}}\right|. \tag{10}$$

In model estimation, the weight matrix plays an important role because it can accommodate the correlation between errors $\left(\varepsilon_i^{(1)}\right)$ and $\left(\varepsilon_i^{(2)}\right)$ in one observation. This relationship is represented through the covariance matrix [21], [22].

$$\boldsymbol{W}=\begin{bmatrix}s_{y^{(1)}}^2\boldsymbol{I} & \left(s_{y^{(1)}y^{(2)}}\right)\boldsymbol{I}\\ \left(s_{y^{(2)}y^{(1)}}\right)\boldsymbol{I} & s_{y^{(2)}}^2\boldsymbol{I}\end{bmatrix}^{-1}, \tag{11}$$

$$\boldsymbol{W}=\begin{bmatrix}z_{y^{(1)}}^2\boldsymbol{I} & \left(z_{y^{(1)}y^{(2)}}\right)\boldsymbol{I}\\ \left(z_{y^{(2)}y^{(1)}}\right)\boldsymbol{I} & z_{y^{(2)}}^2\boldsymbol{I}\end{bmatrix}. \tag{12}$$

2.5 Bivariate Response Nonparametric Regression with Nadaraya-Watson Kernel

The bivariate nonparametric regression using the Kernel Nadaraya-Watson approach is employed to model the relationship between two response variables and one or more predictor variables without requiring specific distributional assumptions. This method incorporates two types of weights: one based on the distance between observations and the prediction point, and another to account for the correlation between response variables [23]. The combination of these weights results in a more accurate estimator. The combined weight matrix $\mathbf{V}$ in this regression is defined as the product of $\boldsymbol{W}$ and $\boldsymbol{W}_b$, with dimensions $2n\times 2n$, as stated in Eqs. (6) and (12).

$$\begin{aligned}\boldsymbol{V} &= \boldsymbol{W}\times\boldsymbol{W}_b\\ &= \begin{bmatrix}z_{y^{(1)}}^2\boldsymbol{I} & \left(z_{y^{(1)}y^{(2)}}\right)\boldsymbol{I}\\ \left(z_{y^{(2)}y^{(1)}}\right)\boldsymbol{I} & z_{y^{(2)}}^2\boldsymbol{I}\end{bmatrix}\begin{bmatrix}\boldsymbol{W}_h & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{W}_h\end{bmatrix}\\ &= \begin{bmatrix}\boldsymbol{z_{y^{(1)}}^2 W_h I} & \left(\boldsymbol{z_{y^{(1)}y^{(2)}}}\right)\boldsymbol{W_h I}\\ \left(\boldsymbol{z_{y^{(2)}y^{(1)}}}\right)\boldsymbol{W_h I} & \boldsymbol{z_{y^{(1)}}^2 W_h I}\end{bmatrix}.\end{aligned} \tag{13}$$

Thus, based on Eq. (13), the estimator of the bivariate nonparametric regression using the Kernel Nadaraya-Watson approach with weighting can be written as follows [15]:

$$\boldsymbol{y}^{*}=\boldsymbol{T}_{t_0}\boldsymbol{\beta}(t_0)+\boldsymbol{\varepsilon}^{*}, \tag{14}$$

where $\boldsymbol{y}^*$ is a response vector of size $2n \times 1$, $\boldsymbol{T}_{t_0}$ is a nonparametric kernel predictor matrix of size $2n \times 1$, $\boldsymbol{V} = \boldsymbol{W}\boldsymbol{W}_b$ is a combined weighting matrix of size $2n \times 2n$, and $\boldsymbol{\varepsilon}^*$ is a random error vector of size $2n \times 1$. According to Eq. (14), the parameter $\widehat{\boldsymbol{\beta}}(t_0)$ is obtained by minimizing the function $R(t_0)$ with respect to $\boldsymbol{\beta}(t_0)$. The parameter $\boldsymbol{\beta}(t_0)$ in Eq. (14) is obtained by minimizing the function $R(t_0)$, which is defined as:

$$R(t_0) = \left(\boldsymbol{y}^* - \boldsymbol{T}_{t_0}\boldsymbol{\beta}(t_0)\right)' \boldsymbol{V} \left(\boldsymbol{y}^* - \boldsymbol{T}_{t_0}\boldsymbol{\beta}(t_0)\right). \tag{15}$$

Thus, the parameter estimate is obtained as follows:

$$\widehat{\boldsymbol{\beta}}(t_0) = \left(\boldsymbol{T'}_{t_0}\boldsymbol{V}\boldsymbol{T}_{t_0}\right)^{-1}\boldsymbol{T'}_{t_0}\boldsymbol{V}\boldsymbol{y}^*. \tag{16}$$

Therefore, $\boldsymbol{\beta}(t_0)$ can be written as follows:

$$\hat{\beta}\,(t_0) = \sum_{i=1}^{n} \frac{\left[\left(z^2_{y^{(1)}} + 2z_{y^{(2)}y^{(1)}} + z^2_{y^{(2)}}\right)\left(\sum_{i=1}^{n} K\left(\sum_{g=1}^{G}\left(\frac{t_g - t_{gi}}{h_g}\right)\right)\right) y_i^{(r)}\right]}{\frac{1}{n}\sum_{i=1}^{n}\left[\left(z^2_{y^{(1)}} + 2z_{y^{(2)}y^{(1)}} + z^2_{y^{(2)}}\right)\left(\sum_{i=1}^{n} K\left(\sum_{g=1}^{G}\left(\frac{t_g - t_{gi}}{h_g}\right)\right)\right)\right]}. \tag{17}$$

The selection of the optimal kernel function and bandwidth is commonly performed by minimizing the Generalized Cross-Validation (GCV) criterion, which offers asymptotic optimality, making it highly effective under various data conditions [9]. The bandwidth parameter h plays a crucial role in controlling the smoothness of the kernel function estimate: a larger h produces smoother estimates with reduced fluctuation but increased bias, leading to underfitting, while a smaller h results in rougher estimates with higher variance and risk of overfitting. The GCV method for determining the optimal bandwidth and kernel function is defined as follows [24]:

$$GCV\left(h_{opt}\right) = \frac{MSE\left(h_{opt}\right)}{\left(1 - 2n^{-1}tr(\boldsymbol{B})\right)^2}, \tag{18}$$

where $\boldsymbol{MSE}(\boldsymbol{h_{opt}}) = \boldsymbol{2n^{-1}}(\boldsymbol{y}^* - \widehat{\boldsymbol{y}}^*)'(\boldsymbol{y}^* - \widehat{\boldsymbol{y}}^*)$ represents the mean squared error at the optimal bandwidth $\boldsymbol{h_{opt}}$, and $\boldsymbol{B} = \boldsymbol{T_{t_0}}(\boldsymbol{T'_{t_0}}\boldsymbol{V}\boldsymbol{T_{t_0}})^{-1}\boldsymbol{T'_{t_0}}\boldsymbol{V}$ denotes the smoothing matrix of dimension $\boldsymbol{2n \times 2n}$. Here, $\boldsymbol{y}^*$ and $\widehat{\boldsymbol{y}}^*$ are the observed and fitted response vectors, respectively, while $\boldsymbol{V}$ is the combined weighting matrix.

2.6 Bivariate Response Nonparametric Regression with Penalized Spline

Bivariate Response Nonparametric Regression with Penalized Spline explains the study of the dependence of one or more response variables on one or more predictor variables using the Penalized Spline estimator. Suppose that paired data $(t_1, t_2, \dots, t_G)$ are given, where the relationship pattern between the variable t_g and $y^{(r)}$ is unknown. The relationship between the variables t_g and $y^{(r)}$ is assumed to follow a nonparametric regression model. One approach that can be used to estimate the parameters in the bivariate response nonparametric Penalized Spline method is the Penalized Weighted Least Squares (PWLS) estimator. The PWLS estimator employs a smoothing parameter to control the roughness of the regression function and incorporates weights in the parameter estimation process. PWLS involves a weight in the form of the inverse of the variance-covariance matrix of the response variables, denoted by $\boldsymbol{W}^{-1}$, as shown in Eq. (12). The bivariate response nonparametric regression model using the PWLS estimator can be expressed as follows:

$$y_i^{(r)} = \delta_0^{(r)} + \sum_{g=1}^{G}\left(\delta_{mg}^{(r)} t_{gi}^{m(r)} + \sum_{k=1}^{K_g} \phi_{gk}^{(r)}\left(t_{gi} - \xi_{gk}\right)_+^m\right) + \varepsilon_i^{(r)};\ r = 1,2;\ i = 1,2,\dots,n. \tag{19}$$

Eq. (19), when expressed in matrix form, can be written as follows [4]:

$$\boldsymbol{y}^{**} = \boldsymbol{T}^{**}\boldsymbol{\delta}_{PWLS} + \boldsymbol{\varepsilon}^{**}.$$

The parameter $\boldsymbol{\delta}_{PWLS}$ in the Penalized Weighted Least Squares estimator is obtained by minimizing the function P, which is expressed as follows [25]:

$$P = \frac{1}{2n}\sum_{i=1}^{n} W_i(y - g(t_i))^2 + \lambda \int_0^1 (g''(t_i))^2 dt\,. \tag{20}$$

Eq. (20) can also be expressed in matrix form as presented by [26]:

$$\boldsymbol{P} = \boldsymbol{2n^{-1}}(\boldsymbol{y^{**}} - \boldsymbol{T^{**}\delta_{PWLS}})'\boldsymbol{W^{-1}}(\boldsymbol{y^{**}} - \boldsymbol{T^{**}\delta_{PWLS}}) + \boldsymbol{\lambda\delta'_{PWLS}D^{**}\delta_{PWLS}}, \quad (21)$$

where $\mathbf{y}^{**}$is a $2n \times 1$ response vector; $\mathbf{T}^{**}$ is a $2n \times 2\left(1 + GM + \sum_{g=1}^{G} K_g\right)$ nonparametric spline predictor matrix; $\boldsymbol{\delta}_{PWLS}$ is a $2\left(1 + GM + \sum_{g=1}^{G} K_g\right) \times 1$ parameter vector; and $\boldsymbol{\varepsilon}^{**}$ is a $2n \times 1$ random error vector. The spline basis matrix $\mathbf{D}^{**}$ is $2\left(1 + GM + \sum_{g=1}^{G} K_g\right) \times 2\left(1 + GM + \sum_{g=1}^{G} K_g\right)$. Based on Eq. (21), $\widehat{\boldsymbol{\delta}}_{\boldsymbol{PWLS}}$ is obtained by taking the derivative of the function P with respect to $\boldsymbol{\delta_{PWLS}}$. Thus, it is obtained:

$$\widehat{\boldsymbol{\delta}}_{\boldsymbol{PWLS}} = \left(\boldsymbol{T^{**\prime}W^{-1}T^{**}} + n\lambda\boldsymbol{D^{**}}\right)^{-1}\boldsymbol{T^{**\prime}W^{-1}y^{**}}. \quad (22)$$

After obtaining the estimator $\widehat{\boldsymbol{\delta}}_{PWLS}$, the next step is to explain the role and location of the knots as well as the smoothing parameter λ in the Penalized Spline model. A knot (ξ_k) is a point where the behavior of a function changes across different intervals. Penalized Spline regression applies knots located at the quantile points, which are the unique values of the predictor variable after the data are sorted. The method for determining the location of the knots in Penalized Spline regression can be expressed as follows [27]:

$$\xi_k = \frac{j}{K+1,}, j = 1,2,3,\ldots,K. \quad (23)$$

In determining the optimal smoothing parameter λ as well as the number and location of knots, a commonly used method is the minimization of Generalized Cross Validation (GCV) [27]. One advantage of the GCV method is its asymptotic optimality property [9]. The smoothing parameter λ plays a role in controlling the roughness penalty. As the value of λ increases, the estimated function becomes smoother, while a decrease in λ results in a rougher estimate. The GCV method can be defined as follows [13]:

$$GCV\left(\xi_{opt}, \lambda_{opt}\right) = \frac{MSE\left(\xi_{opt}, \lambda_{opt}\right)}{\left(1 - 2n^{-1}tr(\boldsymbol{A})\right)^2}. \quad (24)$$

The matrix $\boldsymbol{A}$ represents the projection matrix in the PWLS model, defined as $\boldsymbol{A} = \boldsymbol{T^{**}}(\boldsymbol{T^{**\prime}VT^{**}} + \boldsymbol{\lambda D^{**}})^{-1}\boldsymbol{T^{**\prime}V}$

2.7 Goodness of The Model

The estimated model benefits researchers, governments, and society in decision-making by minimizing estimation errors. R-Squared and RMSE are used to evaluate the model's performance. R-Squared shows how well the model explains data variation, ranging from $-\infty$ (worst) to $+1$ (best), with values near 1 indicating a strong model. It represents the proportion of variance explained by the model. The R-Squared formula [28] is defined as follows:

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}. \quad (25)$$

RMSE (Root Mean Squared Error) is a common model evaluation metric used to measure prediction error. It is calculated as the square root of the average squared differences between observed and predicted values. This method is optimal when the error distribution is normal, as it is more sensitive to large errors compared to other metrics like MAE (Mean Absolute Error). The RMSE formula is defined as follows [29]:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}. \quad (26)$$

A smaller RMSE value indicates a better model. The quality categories of the model are as follows: "Very good" if RMSE $\leq 0.31\times s$, "Good" if RMSE $\leq 0.45\times s$, "Acceptable" if RMSE $\leq 0.83\times s$, and "Unsatisfactory" if RMSE $> 0.83\times s$, where s denotes the standard deviation of the observed data (y_i) [30].

3. RESULTS AND DISCUSSION

3.1 Data and Analysis

The data used in this study are divided into two types, namely simulation data and applied study data. In the simulation data, the predictor variable x_{gi} follows a distribution of $N(5,2)$ and $N(7,3)$, and c_{gi} follows a distribution of $U(0,10)$ and $U(0,15)$, where x_{gi} represents the i-th simulated value of the g-th predictor variable and c_{gi} denotes the i-th simulated auxiliary variable. The errors follow a Multivariate Normal (MVN) distribution, $MVN(\mathbf{0},\boldsymbol{\Sigma})$, where $\mathbf{0}$ is a 2×1 zero vector and $\boldsymbol{\Sigma}$ is a 2×2 variance-covariance matrix with $\boldsymbol{\Sigma}_1 = \begin{bmatrix} 0.25 & 0.25 \\ 0.25 & 0.50 \end{bmatrix}, \boldsymbol{\Sigma}_2 = \begin{bmatrix} 0.50 & 0.25 \\ 0.25 & 1.00 \end{bmatrix}, \boldsymbol{\Sigma}_3 = \begin{bmatrix} 1.50 & 0.25 \\ 0.25 & 2.50 \end{bmatrix}$. The functions used in the simulation include the exponential function:

$$z(t_i) = 1.54(\exp(-5.12\mathrm{x}_{1\mathrm{i}}) + \exp(-9.12\mathrm{x}_{2\mathrm{i}})), \tag{27}$$

and the trigonometric function:

$$h(t_i) = sin(c_{1i}) + cos(c_{2i})\,. \tag{28}$$

For the applied study data, the variables used include stunting prevalence (Y_1), wasting prevalence (Y_2), Complete Basic Immunization (T_1), Low Birth Weight (T_2), Exclusive Breastfeeding (T_3), and Percentage of Poverty Population (T_4), which are measured in percentage units and obtained from the sources Profil Kesehatan and the Central Statistics Agency (BPS). The following are the stages of data analysis for the simulation study:

1. Take sample sizes of $n = 50, n = 100, n = 150, n = 200$.
2. Generate error, x_{gi} and c_{gi}, and define the regression curve.
3. Calculate the value of $y^{(r)}$ according to Eqs. (17) and (19).
4. Estimate the model by involving variations in order, location, the number of knots, and smoothing parameters based on minimum GCV using the formula given in Eq. (18) for the Kernel Nadaraya-Watson estimator and Eq. (24) for the Penalized Spline estimator.
5. Estimate the function using Kernel Nadaraya-Watson and Penalized Spline.
6. Perform 100 repetitions for each scenario.
7. Compare results using MSE, R^2, and RMSE.
8. Perform a MANOVA test to examine the differences between scenarios, followed by orthogonal polynomial contrasts to detect trend patterns, and a Tukey test to identify significant pairwise differences.
9. Conclude the performance of the Kernel Nadaraya-Watson and Penalized Spline estimators in estimating nonparametric regression curves.

The following are the steps for analyzing applied data using the Nadaraya-Watson kernel estimator and penalized spline:

1. Collect and describe the data on Stunting and Wasting, as well as the factors influencing them, based on previous theories and research.
2. Measure the strength of the relationship between the two response variables using Pearson correlation.
3. Visualize the data using scatterplots between response variables and predictors to determine the data relationship pattern.
4. Estimate the function using the Nadaraya-Watson kernel estimator:
   a. Determine the kernel function, upper and lower bounds, bandwidth value, and local data points.
   b. Construct the weight matrix for the Nadaraya-Watson kernel.
   c. Estimate the function using the optimal bandwidth value.
   d. Obtain the estimated function for response 1 and response 2.
5. Estimate the function using the Penalized Spline estimator:
   a. Determine the maximum order, bounds, lambda value, and the number of knots.
   b. Perform the PWLS estimation model using the order, knot locations, and optimal lambda.
   c. Obtain the estimated function for response 1 and response 2.
   d. Construct the nonparametric biresponse Penalized Spline model.
6. Create a plot of observed data and estimated response variables.
7. Evaluate the model performance using R-squared and RMSE.

3.2 Simulation Study

This section presents the application of Spline and Kernel on simulated data to assess the performance of the obtained estimator. The simulation is conducted with several variations in sample size and variance values of $\boldsymbol{\Sigma}$ as shown in Subsection 3.1, using two predictor variables for both response variables. The following is a scatter plot between each response variable and predictor variable from simulated data with the scenario $n = 100$ and $\boldsymbol{\Sigma}_2 = \begin{bmatrix} 0.50 & 0.25 \\ 0.25 & 0.10 \end{bmatrix}$.

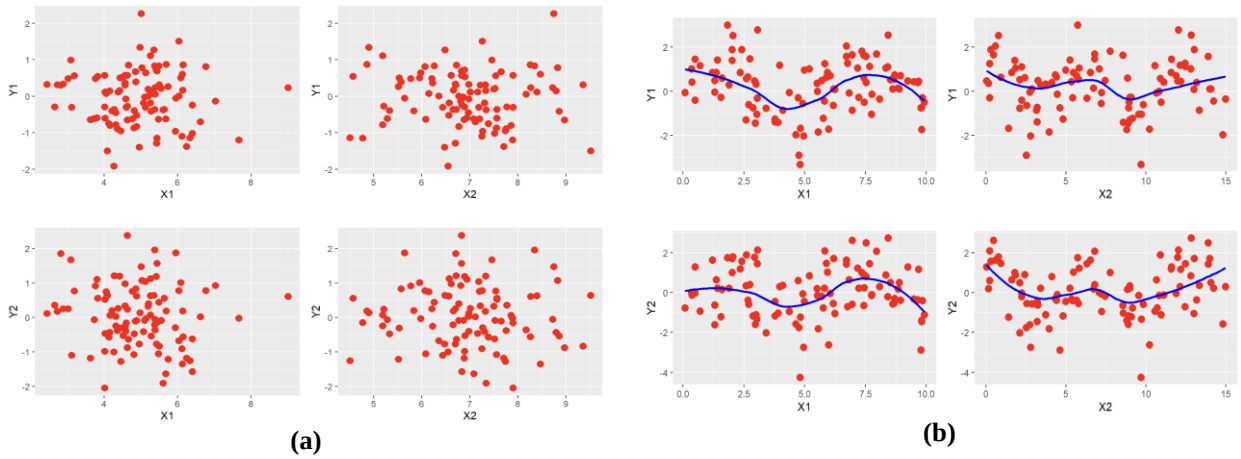


Figure 1. Relationship Between Predictor Variables and Response Variable for Simulated Data (a) Exponential Function (b) Trigonometric Function
Data source: (The visualization results were obtained from simulated data using RStudio with the ggplot2 package)

The results of the function estimation based on the optimal combination for exponential-kernel, exponential-spline, trigonometric-kernel, and trigonometric-spline are summarized in the table below. For both exponential-kernel and trigonometric-kernel estimators, the optimal bandwidth selected was 100 for all scenarios, determined by minimizing the GCV value. Meanwhile, for exponential-spline and trigonometric-spline estimators, the optimal combination consisted of 3 knots, polynomial order of 3-3 for $y^{(1)}$ and $y^{(2)}$, and a smoothing parameter $\lambda = 5$, also selected based on the minimum GCV:

Table 1. Simulated Optimal Combination

n	Exp-Kernel		Exp-Spline		Trigo-Kernel		Trigo-Spline		
	$\boldsymbol{\Sigma}$	GCV	$\boldsymbol{\Sigma}$	GCV	$\boldsymbol{\Sigma}$	GCV	$\boldsymbol{\Sigma}$	ξ	GCV
50	$\boldsymbol{\Sigma}_1$	8.878×10^{-10}	$\boldsymbol{\Sigma}_1$	1.531×10^{-5}	$\boldsymbol{\Sigma}_1$	2.882×10^{-7}	$\boldsymbol{\Sigma}_1$	3-3	3.328×10^{-1}
	$\boldsymbol{\Sigma}_2$	1.127×10^{-9}	$\boldsymbol{\Sigma}_2$	7.715×10^{-6}	$\boldsymbol{\Sigma}_2$	5.022×10^{-7}	$\boldsymbol{\Sigma}_2$	3-2	7.515×10^{-1}
	$\boldsymbol{\Sigma}_3$	1.673×10^{-9}	$\boldsymbol{\Sigma}_3$	3.465×10^{-6}	$\boldsymbol{\Sigma}_3$	3.951×10^{-7}	$\boldsymbol{\Sigma}_3$	3-3	2.402×100
100	$\boldsymbol{\Sigma}_1$	1.772×10^{-9}	$\boldsymbol{\Sigma}_1$	5.575×10^{-6}	$\boldsymbol{\Sigma}_1$	2.372×10^{-7}	$\boldsymbol{\Sigma}_1$	3-2	2.801×10^{-1}
	$\boldsymbol{\Sigma}_2$	2.011×10^{-9}	$\boldsymbol{\Sigma}_2$	2.696×10^{-6}	$\boldsymbol{\Sigma}_2$	2.816×10^{-7}	$\boldsymbol{\Sigma}_2$	3-3	1.003×100
	$\boldsymbol{\Sigma}_3$	2.685×10^{-9}	$\boldsymbol{\Sigma}_3$	9.158×10^{-7}	$\boldsymbol{\Sigma}_3$	5.235×10^{-7}	$\boldsymbol{\Sigma}_3$	3-3	2.129×100
150	$\boldsymbol{\Sigma}_1$	1.011×10^{-9}	$\boldsymbol{\Sigma}_1$	1.849×10^{-6}	$\boldsymbol{\Sigma}_1$	3.435×10^{-7}	$\boldsymbol{\Sigma}_1$	3-2	4.580×10^{-1}
	$\boldsymbol{\Sigma}_2$	9.429×10^{-10}	$\boldsymbol{\Sigma}_2$	1.289×10^{-6}	$\boldsymbol{\Sigma}_2$	3.916×10^{-7}	$\boldsymbol{\Sigma}_2$	3-3	7.996×10^{-1}
	$\boldsymbol{\Sigma}_3$	4.346×10^{-9}	$\boldsymbol{\Sigma}_3$	4.117×10^{-7}	$\boldsymbol{\Sigma}_3$	4.471×10^{-7}	$\boldsymbol{\Sigma}_3$	3-3	2.734×100
200	$\boldsymbol{\Sigma}_1$	7.216×10^{-10}	$\boldsymbol{\Sigma}_1$	1.296×10^{-6}	$\boldsymbol{\Sigma}_1$	3.781×10^{-7}	$\boldsymbol{\Sigma}_1$	3-2	3.623×10^{-1}
	$\boldsymbol{\Sigma}_2$	2.227×10^{-9}	$\boldsymbol{\Sigma}_2$	6.006×10^{-7}	$\boldsymbol{\Sigma}_2$	3.655×10^{-7}	$\boldsymbol{\Sigma}_2$	3-2	9.611×10^{-1}
	$\boldsymbol{\Sigma}_3$	3.282×10^{-9}	$\boldsymbol{\Sigma}_3$	2.367×10^{-7}	$\boldsymbol{\Sigma}_3$	7.342×10^{-7}	$\boldsymbol{\Sigma}_3$	3-3	2.806×100

The simulation process was repeated one hundred times for each sample variation and covariance matrix variation to ensure the stability of the results and to minimize the influence of random variations that may occur in a single simulation run. Through this repetition, the obtained and RMSE values are expected to better represent the overall performance of the methods used. The simulation results are summarized based on their mean values as follows:

Table 2. Average RMSE of Simulation Results

n	Σ	RMSE $y^{(1)}$ RMSE $y^{(2)}$			
		Exp-Kernel	**Exp-Spline**	**Trigo-Kernel**	**Trigo-Spline**
50	Σ_1	2.389×10^{-5} 3.411×10^{-5}	4.518×10^{-1} 6.456×10^{-1}	4.886×10^{-4} 5.414×10^{-4}	7.175×10^{-1} 8.672×10^{-1}
	Σ_2	3.221×10^{-5} 4.411×10^{-5}	6.524×10^{-1} 9.247×10^{-1}	5.568×10^{-4} 6.324×10^{-4}	8.503×10^{-1} 1.083×10^{-0}
	Σ_3	5.459×10^{-5} 7.021×10^{-5}	1.131×100 1.475×100	7.138×10^{-4} 8.611×10^{-4}	1.259×100 1.598×100
100	Σ_1	2.395×10^{-5} 3.429×10^{-5}	4.801×10^{-1} 6.784×10^{-1}	5.120×10^{-4} 5.586×10^{-4}	7.376×10^{-1} 9.013×10^{-1}
	Σ_2	3.279×10^{-5} 4.819×10^{-5}	6.751×10^{-1} 9.673×10^{-1}	5.567×10^{-4} 6.407×10^{-4}	8.856×10^{-1} 1.142×100
	Σ_3	5.976×10^{-5} 7.706×10^{-5}	1.173×100 1.522×100	7.249×10^{-4} 8.465×10^{-4}	1.318×100 1.650×100
150	Σ_1	2.497×10^{-5} 3.491×10^{-5}	4.881×10^{-1} 6.782×10^{-1}	5.100×10^{-4} 5.594×10^{-4}	7.449×10^{-1} 9.067×10^{-1}
	Σ_2	3.532×10^{-5} 4.945×10^{-5}	6.896×10^{-1} 9.868×10^{-1}	5.567×10^{-4} 6.448×10^{-4}	9.096×10^{-1} 1.171×100
	Σ_3	6.062×10^{-5} 7.901×10^{-5}	1.206×100 1.536×100	7.266×10^{-4} 8.516×10^{-4}	1.187×100 1.737×100
200	Σ_1	2.472×10^{-5} 3.485×10^{-5}	4.875×10^{-1} 6.956×10^{-1}	5.195×10^{-4} 5.709×10^{-4}	7.518×10^{-1} 9.169×10^{-1}
	Σ_2	3.462×10^{-5} 4.834×10^{-5}	6.938×10^{-1} 9.819×10^{-1}	5.632×10^{-4} 6.412×10^{-4}	9.046×10^{-1} 1.157×10^{-1}
	Σ_3	5.988×10^{-5} 7.718×10^{-5}	1.205×100 1.559×100	7.321×10^{-4} 8.572×10^{-4}	1.369×100 1.702×100

Based on Table 1, statistically, differences cannot be concluded solely by comparing the means of each scenario. Therefore, the next step is to test the results obtained. To determine whether there are significant differences based on sample size scenarios and error variance scenarios, a test for mean differences between pairs is conducted. This test is performed based on the following hypotheses:

H_0: There is no significant difference between the mean RMSE $y^{(1)}$ and RMSE $y^{(2)}$.
H_1: There is a significant difference between the mean RMSE $y^{(1)}$ and RMSE $y^{(2)}$.

Table 3. MANOVA Results

	***p*-value**			
	Exp-Kernel	**Exp-Spline**	**Trigo-Kernel**	**Trigo-Spline**
Sample	$<0.001^{**}$	$<0.001^{**}$	0.127	$<0.001^{**}$
Variance	$<0.001^{**}$	$<0.001^{**}$	$<0.001^{**}$	$<0.001^{**}$

Note: ** indicates statistical significance at $\alpha = 0.05$.

According to the MANOVA results in Table 3. There are statistically significant differences in RMSE values across sample size and error variance scenarios for all estimators except the trigonometric-kernel estimator with respect to sample size ($p = 0.127 > 0.05$). This suggests that the performance of the trigonometric kernel is less sensitive to changes in sample size, unlike the other three estimators.

Table 4. Orthogonal Polynomial Results

Category	Source	F_{stat} Contrast							
		Exp-Kernel		**Exp-Spline**		**Trigo-Kernel**		**Trigo-Spline**	
		$y^{(1)}$	$y^{(2)}$	$y^{(1)}$	$y^{(2)}$	$y^{(1)}$	$y^{(2)}$	$y^{(1)}$	$y^{(2)}$
Sample	Linear	✓	✓	✓	✓	-	-	✓	✓
	Quadratic	✓	✓	✓	✓	-	-	✓	✓
	Cubic	×	×	×	✓	-	-	✓	✓
Variance	Linear	✓	✓	✓	✓	✓	✓	✓	✓
	Quadratic	✓	✓	✓	✓	✓	✓	✓	✓

Note: ✓ indicates statistically significant contrast at α = 0.05; × indicates not significant; "-" indicates not applicable.

The orthogonal polynomial analysis indicates that, in general, the relationship between sample size and RMSE follows a quadratic pattern, as evidenced by the significance of quadratic contrasts in most methods. Although in certain cases, such as specific spline-based methods, a cubic trend also appears, it remains inconsistent. Conversely, all relationships between error variance and RMSE consistently exhibit a quadratic pattern across all estimation methods. To further illustrate the differences between groups, a boxplot visualization of the RMSE results is presented:

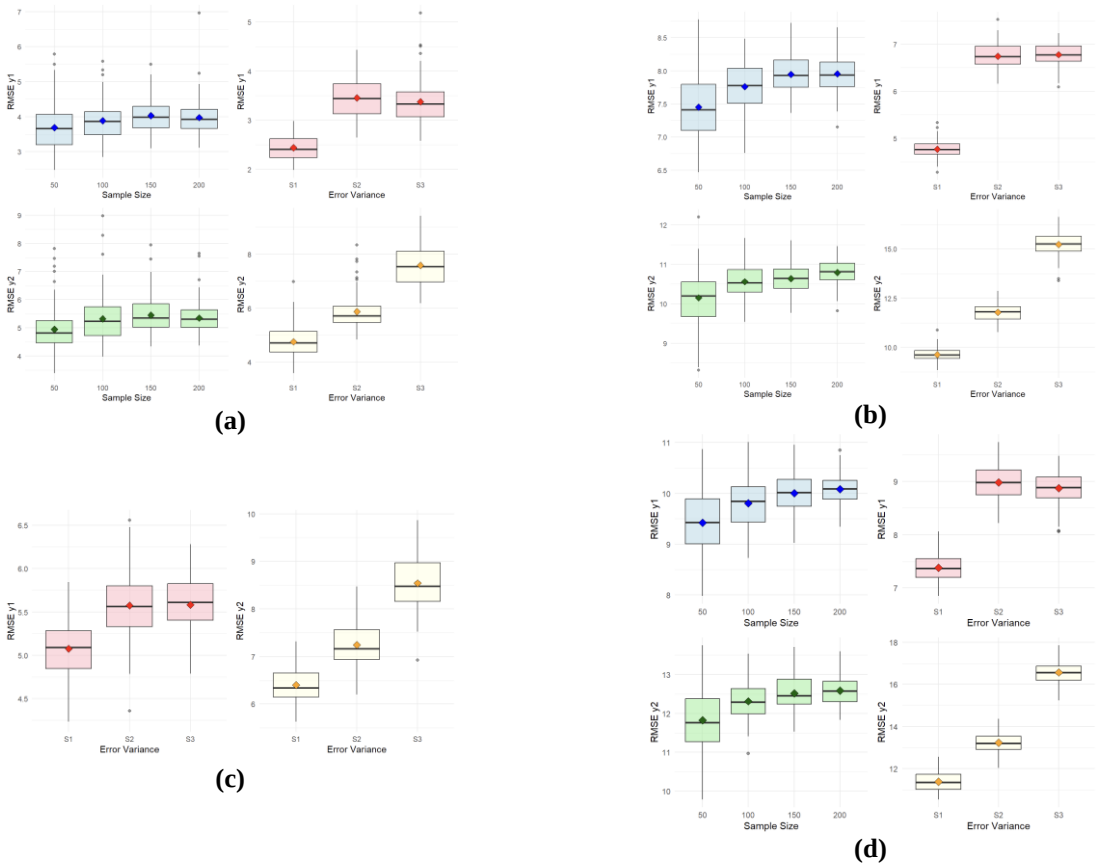


Figure 2. Visualization of RMSE Boxplot Results:
(a) Exponential-Kernel, (b) Exponential-Spline, (c) Trigonometric-Kernel, and (d) Trigonometric-Spline
Data source: (The visualization results were obtained from simulated data using RStudio with the ggplot2 package)

To further identify which pair exhibits significant differences, a follow-up Tukey test was conducted. This test aims to compare all pairs of means to identify statistically different pairs, as summarized in the following table:

Table 5. Tukey's Further Test Results

		Group							
		Exp-Kernel		Exp-Spline		Trigo-Kernel		Trigo-Spline	
		$y^{(1)}$	$y^{(2)}$	$y^{(1)}$	$y^{(2)}$	$y^{(1)}$	$y^{(2)}$	$y^{(1)}$	$y^{(2)}$
Sample	200	a	a	a	a	-	-	a	a
	150	a	a	a	b	-	-	a	a
	100	a	a	b	b	-	-	b	b
	50	b	b	c	c	-	-	c	c
Variance	$\boldsymbol{\Sigma}_3$	a	a	a	a	a	a	a	a
	$\boldsymbol{\Sigma}_2$	a	b	a	b	a	b	b	b
	$\boldsymbol{\Sigma}_1$	b	c	b	c	b	c	c	c

This table summarizes the results of Tukey's Post-Hoc test comparing methods based on sample size and variance. For sample size, groups with the same letter indicate no significant difference. The results show that across all methods, larger sample sizes (200, 150, 100) are mostly grouped under "a", indicating similar performance. While the smallest sample size (50) consistently forms a separate group ("b" or "c"), suggesting significantly higher RMSE for smaller samples. For variance, the highest variance level ($\boldsymbol{\Sigma}_3$) is grouped as "a" while the lowest ($\boldsymbol{\Sigma}_1$) consistently appears as "b" or "c", indicating that smaller error variances

significantly reduce RMSE. These findings reinforce that increasing sample size and reducing error variance both contribute to improved model accuracy.

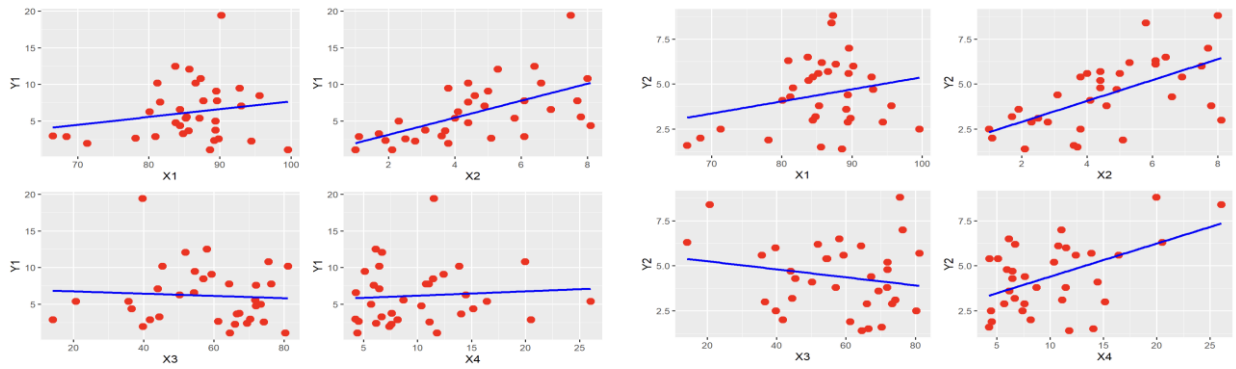
3.3 Data application

After conducting the modeling using simulated data, both methods were applied to the prevalence data of Stunting and Wasting. These two indicators often occur simultaneously and reflect poor nutritional status, making a simultaneous analysis of both variables highly important. As an initial step, a correlation test between the two response variables was performed, and the results are summarized in the following table:

Table 6. Correlation Test

$H_0: \rho = 0$	
Statistics	Value
$\hat{\rho}$	0.216
t_{value}	2.735
t_{table}	2.264
$p - value$	0.007

Based on Table 6, the calculated t-value (2.735) is greater than the critical t-value (2.264). Thus, the null hypothesis (H_0) is rejected, indicating that at the 0.05 significance level, there is a significant positive correlation of 0.216 between the response variables, namely the prevalence of Stunting and Wasting in Indonesia in 2024. Therefore, the correlation assumption is satisfied, suggesting that the prevalence of Stunting and Wasting can be modeled simultaneously using a bivariate approach. Furthermore, the relationship patterns between the predictor and response variables were examined through scatter plot visualization.



(a) (b)

Figure 3. Relationship Between Predictor Variables and Response Variable: (a) Response 1 and (b) Response 2

Data source: (The visualization results were obtained from simulated data using RStudio with the ggplot2 package)

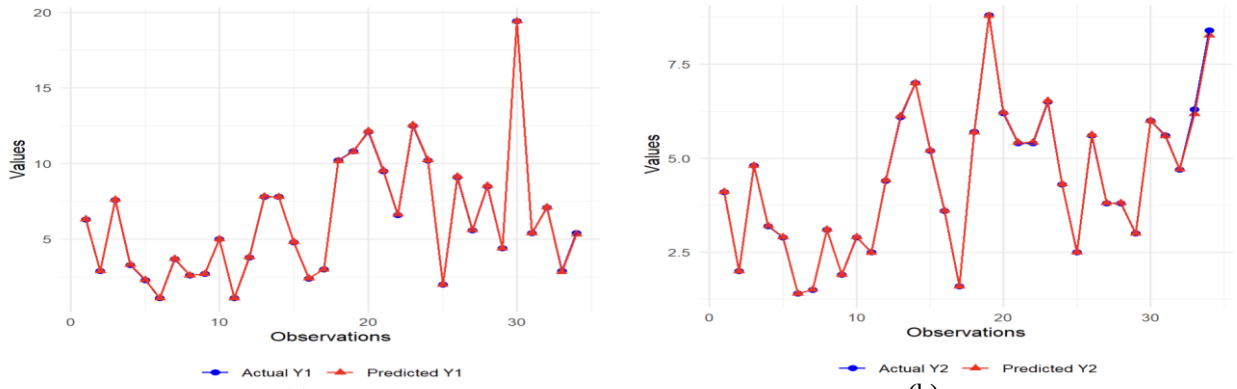
Based on the scatter plot results, there is no clear relationship pattern between the response variables and the predictor variables, indicating that the prevalence data of Stunting and Wasting are suitable for modeling using both methods. The bivariate nonparametric regression using the Nadaraya-Watson Kernel is employed to model two response variables without assuming any specific distribution, utilizing distance-based weighting of predictors. This study uses the Epanechnikov, Gaussian, and Biweight kernels and determines the optimal bandwidth within the range of 5 to 100 with an increment of 5 based on the minimum GCV criterion. The local estimation points are taken from the latest observation data to calculate the weighted average of the response values based on the proximity of the predictors. The next step is to estimate the regression function using the kernel functions determined in the previous stage. The estimation is performed by applying the selected kernel functions, namely, Epanechnikov, Gaussian, or Biweight, to the predictor variables. The estimation results based on the three kernel functions are obtained as follows:

Table 7. Comparison of GCV Values of Kernel Functions

Kernel Function	GCV	Combination Number	Optimal Bandwidth t_1, t_2, t_3, t_4
Epanechnikov	0.00383	160000	100,100,100,100

Kernel Function	GCV	Combination Number	Optimal Bandwidth t_1, t_2, t_3, t_4
Gaussian	0.00088	159940	100,85,100,100
Biweight	0.01352	159959	95,90,100,100

Based on Table 7, the Gaussian kernel produces the smallest GCV value (0.00088) compared to Epanechnikov (0.00383) and Biweight (0.01352). Therefore, it is chosen as the best kernel function. The estimated function $\hat{y}_i^{(1)}$ and $\hat{y}_i^{(2)}$ is then plotted to compare the original and predicted data. As shown in Fig. 4.



(a) (b)

Figure 4. Plot of Kernel Function Estimation Results
(a) Response 1 (b) Response 2
Data source: (The visualization results were obtained from simulated data using RStudio with the ggplot2 package)

After applying the Kernel Nadaraya-Watson estimator. The analysis is continued using the Penalized Spline estimator to model the relationship between the response variables and predictors without assuming a specific functional form. Parameter estimation is performed using Penalized Weighted Least Squares to improve model accuracy. In bivariate nonparametric regression using Penalized Spline, the selection of polynomial order, Lambda value, and number of knots greatly determines the model's flexibility. The maximum order is limited to 3, with combinations of order 1 (linear), order 2 (quadratic), and order 3 (cubic), where the optimal order is selected based on the minimum GCV value. Lambda values are tested within the range of 5 to 100 with increments of 5 to achieve the best smoothing level based on GCV. The number of knots was restricted to a maximum of three following previous studies on limited-sample spatial or health-related data, where excessive knots tended to increase variance without significantly improving fit [19]. This restriction also reduces computational complexity and ensures model stability, with the optimal number also determined based on the minimum GCV value. The next step is to determine the optimal combination of order, knot locations, number of knots, and lambda values, which is summarized in the following table.

Table 8. Optimal Knot Location

ξ_1	ξ_2	ξ_3	ξ_4
84.533	2.875	51.333	6.680
89.000	4.500	66.733	11.430
	6.325		
$Order\ (3,3)$			
$\lambda = 5$			
$GCV = 7.305 \times 10^{-8}$			

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:

$$\hat{y}_i^{(1)} = 50.518 - 3.450t_{1i} - 0.031t_{1i}^2 + 0.0004t_{1i}^3 - 0.036(t_{1i} - 84.533)_+^3 + 0.091(t_{1i} - 89.000)_+^3$$
$$+28.277t_{2i} - 6.012t_{2i}^2 + 0.415t_{2i}^3 - 0.032(t_{2i} - 2.875)_+^3 - 0.033(t_{2i} - 4.500)_+^3$$
$$-0.010(t_{2i} - 6.325)_+^3 - 4.635t_{3i} + 0.142t_{3i}^2 - 0.001t_{3i}^3 + 0.002(t_{3i} - 51.333)_+^3$$

$$+0.003(t_{3i}-66.733)_+^3+80.019t_{4i}-10.839t_{4i}^2+0.493t_{4i}^3-0.398(t_{4i}-6.680)_+^3$$
$$-0.110(t_{4i}-11.430)_+^3 \quad (29)$$

$$\hat{y}_i{}^{(2)}=-0.339-0.214t_{1i}-0.025t_{1i}^2+0.0002t_{1i}^3-0.015(t_{1i}-84.533)_+^3+0.037(t_{1i}-89.000)_+^3$$
$$+8.136t_{2i}-1.562t_{2i}^2+0.099t_{2i}^3-0.007(t_{2i}-2.875)_+^3-0.004(t_{2i}-4.500)_+^3$$
$$-0.0005\ (t_{2i}-6.325)_+^3-1.866t_{3i}+0.056t_{3i}^2-0.0004t_{3i}^3+0.001(t_{3i}-51.333)_+^3$$
$$+0.00006(t_{3i}-66.733)_+^3+26.907t_{4i}-3.510t_{4i}^2+0.153t_{4i}^3-0.107(t_{4i}-6.680)_+^3$$
$$-0.055(t_{4i}-11.430)_+^3 \quad (30)$$

After establishing the PWLS model for the Stunting and Wasting case study by 2024 in Indonesia, 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 Variable T_2 are presented as follows:

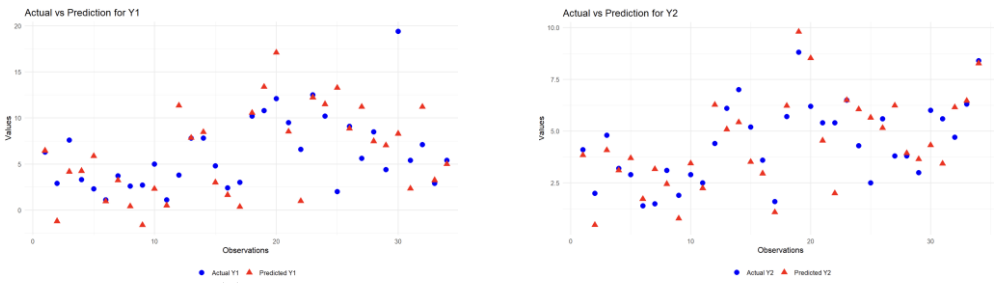
$$f^{(1)}(t_{2i})=28.277t_{2i}-6.012t_{2i}^2+0.415t_{2i}^3-0.032(t_{2i}-2.875)_+^3-0.033(t_{2i}-4.500)_+^3$$
$$-0.010(t_{2i}-6.325)_+^3 \quad (31)$$

$$f^{(1)}(t_{2i})=\begin{cases}8.277t_{2i}-6.012t_{2i}^2+0.415t_{2i}^3, & 0<t_{2i}\le 2.875\\ 7.483t_{2i}-0.537t_{2i}^2+0.383+0.760, & 2.875<t_{2i}\le 4.500\\ 5.479t_{2i}-5.291t_{2i}^2-0.350t_{2i}^3+3.767, & 4.500<t_{2i}\le 6.325\\ 4.278t_{2i}-5.101t_{2i}^2+0.340t_{2i}^3+6.298, & t_{2i}>6.325\end{cases}$$

Table 9. Summary of t_2 Segmentation in Response 1

Knot	**Segment**			
t_2	I	II	III	IV
2.875				
4.500	+	+	+	+
6.325				

Based on the segmentation results for Low Birth Weight, the effect of t_{2i} on prevalence of stunting shows a consistent pattern across all segments. In Segments I, II, III, and IV, an increase in t_{2i} by one unit leads to an increase in $y^{(1)}$. This indicates that higher values of t_{2i} are associated with a higher prevalence of Stunting with no evidence of reduction in any segment. Therefore, this factor demonstrates a strong positive relationship with Stunting prevalence, emphasizing the need for targeted interventions to mitigate the impact of increasing t_{2i} on child health outcomes. The estimated function $\hat{y}_i{}^{(1)}$and $\hat{y}_i{}^{(2)}$ is then plotted to compare the original and predicted data as shown in Fig. 5.



(a) (b)

Figure 5. Plot of spline function estimation results
(a) Response 1 (b) Response 2

Data source: (The visualization results were obtained from simulated data using RStudio with the ggplot2 package)

The selection of the best model between the Bivariate Nonparametric Regression Kernel Nadaraya-Watson and the Penalized Spline was based primarily on the Mean Squared Error (MSE) as the main performance criterion, rather than on the GCV value. The GCV was employed solely for determining the optimal smoothing parameters within each model. The Kernel model achieved an MSE of 0.00086, whereas the Penalized Spline model produced a substantially higher MSE of 8.708. Therefore, the Kernel model was selected as the superior estimator due to its lower MSE and better generalization performance. Thus, the Kernel model was chosen for its superior ability to generalize the data. Further evaluation showed that the Kernel Nadaraya-Watson model had very high R^2 values, namely, 0.9999 for $y^{(1)}$ and 0.9996 for $y^{(2)}$, along with low RMSE values, indicating the model's ability to explain nearly all data variation with minimal prediction error. Therefore, this model can be relied upon for data-based analysis and decision-making. These findings are consistent with previous studies. Study [31] reported superior performance of the Nadaraya Watson Kernel, particularly with Gaussian kernels, in achieving high predictive accuracy. Meanwhile, study [32] demonstrated the application of spline-based methods, including Penalized Spline, for modeling complex bivariate relationships. Together, these studies support the relevance of the methods applied in this research.

4. CONCLUSION

This study compared the flexibility of the Kernel Nadaraya-Watson and Penalized Spline estimators in bivariate nonparametric regression using simulated and empirical data. Based on the Mean Squared Error (MSE), the Kernel estimator showed lower prediction errors and greater stability across sample sizes, while the extremely high R^2 values suggest potential overfitting that requires further validation through cross-validation. Although the Penalized Spline model produced smaller GCV values, its higher MSE may result from parameter constraints, particularly the limited number of knots, which reduced model flexibility rather than indicating methodological weakness. Overall, the Kernel estimator shows promising potential for health-related applications, but future research should include parameter tuning and validation to ensure more reliable and generalizable results.

Author Contributions

Cinta Rizki Oktarina: Conceptualization, Methodology, Data Curation, Formal Analysis, Software, Visualization, Writing-Original Draft, Validation. Sigit Nugroho: Supervision, Methodology, Validation, and Writing–Review and Editing. Idhia Sriliana: Supervision, Resources, and Writing–Review and Editing. All authors discussed the results and contributed to the final manuscript.

Funding Statement

The author would like to express gratitude to the Research and Community Service Institute (LPPM) of Universitas Bengkulu and the Directorate General of Higher Education. Science. and Technology (Kemdiktisaintek) for the financial support provided through the Master's Thesis Research Grant under contract number 2876/UN30.15/PT/2025

Acknowledgment

The authors would like to express their sincere gratitude to Statistics Indonesia (BPS) for providing access to the 2024 stunting and wasting data, and to all colleagues and reviewers who provided valuable feedback, suggestions, and constructive comments that helped improve the quality of this manuscript.

Declarations

The authors declare that they have no conflicts of interest to report for this study.

Declaration of Generative AI and AI-assisted Technologies

AI-assisted technology (ChatGPT) was used to support light paraphrasing and sentence restructuring for clarity. The authors confirm that the underlying ideas, arguments, data analyses, and conclusions are original and were not generated by AI. All AI-assisted edits were critically reviewed and validated by the authors.

REFERENCES

[1] D. C. Montgomery, E. A. Peck, and G. VINNING, *LINEAR REGRESSION ANALYSIS*, 6th ed. John Wiley & Sons, Inc, 2021. doi: https://doi.org/10.2307/1268395

[2] I. Sriliana, I. N. Budiantara, and V. Ratnasari, "THE MIXED ESTIMATOR OF TRUNCATED SPLINE AND LOCAL LINEAR IN MULTIVARIABLE NONPARAMETRIC REGRESSION," *AIP Conf. Proc.*, vol. 2554, no. 1, 2023. doi: https://doi.org/10.1063/5.0104167

[3] R. L. Eubank, *NONPARAMETRIC REGRESSION AND SPLINE SMOOTHING*, 2nd ed. New York: Marcel Dekker, 1999. doi: https://doi.org/10.1201/9781482273144

[4] I. N. Budiantara, *REGRESI NONPARAMETRIK SPLINE TRUNCATED*. Surabaya: ITS Press, 2019.

[5] P. P. Gabrela, J. D. T. Purnomo, and I. N. Budiantara, "THE ESTIMATION OF MIXED TRUNCATED SPLINE AND FOURIER SERIES ESTIMATOR IN BI-RESPONSE NONPARAMETRIC REGRESSION," *AIP Conf. Proc.*, vol. 2903, no. 1, 2023. doi: https://doi.org/10.1063/5.0177224.

[6] I. Sriliana, I. N. Budiantara, and V. Ratnasari, "A TRUNCATED SPLINE AND LOCAL LINEAR MIXED ESTIMATOR IN NONPARAMETRIC REGRESSION FOR LONGITUDINAL DATA AND ITS APPLICATION," *Symmetry (Basel).*, vol. 14, no. 12, 2022. doi: https://doi.org/10.3390/sym14122687.

[7] R. Hidayat, I. N. Budiantara, B. W. Otok, and V. Ratnasari, "THE REGRESSION CURVE ESTIMATION BY USING MIXED SMOOTHING SPLINE AND KERNEL (MSS-K) MODEL," *Commun. Stat. - Theory Methods*, vol. 50, no. 17, pp. 3942–3953, 2021. doi: https://doi.org/10.1080/03610926.2019.1710201

[8] R. Pahlepi, I. Sriliana, W. Agwil, and C. R. Oktarina, "BIRESPONSE SPLINE TRUNCATED NONPARAMETRIC REGRESSION MODELING FOR LONGITUDINAL DATA ON MONTHLY STOCK PRICES OF THREE PRIVATE BANKS IN INDONESIA," *Barekeng*, vol. 19, no. 4, pp. 2467–2480, 2025. doi: https://doi.org/10.30598/barekengvol19iss4pp2467-2480

[9] G. Wahba, *SPLINE MODELS FOR OBSERVATIONAL DATA*. PENNSYLVANIA: SOCIETY FOR INDUSTRIAL AND APPLIED MATHEMATICS, 1990. doi: https://doi.org/10.1137/1.9781611970128

[10] Y. R. Yue, D. Simpson, F. Lindgren, and H. Rue, "BAYESIAN ADAPTIVE SMOOTHING SPLINES USING STOCHASTIC DIFFERENTIAL EQUATIONS," *Bayesian Anal.*, vol. 9, no. 2, pp. 397–424, 2014. doi: https://doi.org/10.1214/13-BA866

[11] K. Xie, F. Meng, and D. Zhang, "REGIONAL FORECASTING OF PM2.5 CONCENTRATIONS: A NOVEL MODEL BASED ON THE EMPIRICAL ORTHOGONAL FUNCTION ANALYSIS AND NADARAYA–WATSON KERNEL REGRESSION ESTIMATOR," *Environ. Model. Softw.*, vol. 170, no. January, p. 105857, 2023. doi: https://doi.org/10.1016/j.envsoft.2023.105857

[12] A. R. Fadilah, A. Fitrianto, and I. M. Sumertajaya, "OUTLIER IDENTIFICATION ON PENALIZED SPLINE REGRESSION MODELING FOR POVERTY GAP INDEX IN JAVA," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 16, no. 4, pp. 1231–1240, 2022. doi: https://doi.org/10.30598/barekengvol16iss4pp1231-1240

[13] D. Ruppert, M. P. Wand, and R. Carroll, *SEMIPARAMETRIC REGRESSION*. Cambridge University Press, 2003. doi: https://doi.org/10.1201/9781420091984-c17

[14] N. Y. Adrianingsih, I. N. Budiantara, and J. D. T. Purnomo, "MIXTURE MODEL NONPARAMETRIC REGRESSION AND ITS APPLICATION," *J. Phys. Conf. Ser.*, vol. 1842, no. 1, 2021. doi: https://doi.org/10.1088/1742-6596/1842/1/012044

[15] J. Fan and I. Gijbels, *LOKAL POLYNOMIAL MODELLING AND ITS APPLICATIONS*, 1st ed., no. 1985. Springer Science Business Media, 1996.

[16] T. H. Ali, "MODIFICATION OF THE ADAPTIVE NADARAYA-WATSON KERNEL METHOD FOR NONPARAMETRIC REGRESSION (SIMULATION STUDY)," *Commun. Stat. Simul. Comput.*, vol. 51, no. 2, pp. 391–403, 2019. doi: https://doi.org/10.1080/03610918.2019.1652319

[17] C. P. A. Moraes, D. G. Fantinato, and A. Neves, "EPANECHNIKOV KERNEL FOR PDF ESTIMATION APPLIED TO EQUALIZATION AND BLIND SOURCE SEPARATION," *Signal Processing*, vol. 189, p. 108251, 2021. doi: https://doi.org/10.1016/j.sigpro.2021.108251.

[18] A. C. Guidoum, "KERNEL ESTIMATOR AND BANDWIDTH SELECTION FOR DENSITY AND ITS DERIVATIVES: THE KEDD PACKAGE," vol. 3, no. 1, pp. 1–22, 2020. doi: https://doi.org/10.48550/arXiv.2012.06102.

[19] I. Sriliana, I. N. Budiantara, and V. Ratnasari, "THE PERFORMANCE OF MIXED TRUNCATED SPLINE-LOCAL LINEAR NONPARAMETRIC REGRESSION MODEL FOR LONGITUDINAL DATA," *MethodsX*, vol. 12, no. July 2023, 2024. doi: https://doi.org/10.1016/j.mex.2024.102652

[20] Suparti, R. Santoso, A. Prahutama, and A. R. Devi, *REGRESI NONPARAMETRIK*, 1st ed. Ponorogo: Wade Group, 2017.

[21] Sifriyani, A. R. M. Sari, A. T. R. Dani, and S. Jalaluddin, "BI-RESPONSE TRUNCATED SPLINE NONPARAMETRIC REGRESSION WITH OPTIMAL KNOT POINT SELECTION USING GENERALIZED CROSS-VALIDATION IN DIABETES MELLITUS PATIENT'S BLOOD SUGAR LEVELS," *Commun. Math. Biol. Neurosci.*, vol. 2023, pp. 1–18, 2023. doi: https://doi.org/10.28919/cmbn/7903.

[22] C. R. Oktarina, S. Nugroho, and I. Sriliana, "Estimation of Stunting and Wasting Prevalence in Southern Part of Sumatra Using Nadaraya-Watson Kernel and Penalized Spline," vol. 10, no. 2, pp. 647–660, 2026.

[23] S. Tosatto, R. Akrour, and J. Peters, "AN UPPER BOUND OF THE BIAS OF NADARAYA-WATSON KERNEL REGRESSION UNDER LIPSCHITZ ASSUMPTIONS," *Stats*, vol. 4, no. 1, pp. 1–17, 2021. doi: https://doi.org/10.3390/stats4010001

[24] T. Misiakiewicz and B. Saeed, "A NON-ASYMPTOTIC THEORY OF KERNEL RIDGE REGRESSION: DETERMINISTIC EQUIVALENTS, TEST ERROR, AND GCV ESTIMATOR," pp. 1–131, 2024, [Online]. Available:

[25] P. Green and B. Silverman, *NONPARAMETRIC REGRESSION AND GENERALIZED LINEAR MODELS*. Springer Science Business Media, 1994. doi: https://doi.org/10.1007/978-1-4899-4473-3

[26] A. Islamiyati *et al.*, "THE USE OF PENALIZED WEIGHTED LEAST SQUARE TO OVERCOME CORRELATIONS BETWEEN TWO RESPONSES," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 16, no. 4, pp. 1497–1504, 2022. doi: https://doi.org/10.30598/barekengvol16iss4pp1497-1504

[27] L. Yang and Y. Hong, "ADAPTIVE PENALIZED SPLINES FOR DATA SMOOTHING," *Comput. Stat. Data Anal.*, vol. 108, pp. 70–83, 2017. doi: https://doi.org/10.1016/j.csda.2016.10.022

[28] D. Chicco, M. J. Warrens, and G. Jurman, "THE COEFFICIENT OF DETERMINATION R-SQUARED IS MORE INFORMATIVE THAN SMAPE, MAE, MAPE, MSE AND RMSE IN REGRESSION ANALYSIS EVALUATION," *PeerJ Comput. Sci.*, vol. 7, no. e623, pp. 1–24, 2021. doi: https://doi.org/10.7717/peerj-cs.623

[29] T. O. Hodson, "ROOT-MEAN-SQUARE ERROR (RMSE) OR MEAN ABSOLUTE ERROR (MAE): WHEN TO USE THEM OR NOT," *Geosci. Model Dev.*, vol. 15, no. 14, pp. 5481–5487, 2022. doi: https://doi.org/10.5194/gmd-15-5481-2022

[30] S. Dazzi, R. Vacondio, and P. Mignosa, "FLOOD STAGE FORECASTING USING MACHINE-LEARNING METHODS: A CASE STUDY ON THE PARMA RIVER (ITALY)," *Water (Switzerland)*, vol. 13, no. 12, pp. 1–22, 2021. doi: https://doi.org/10.3390/w13121612

[31] A. M. Sadek and L. A. Mohammed, "EVALUATION OF THE PERFORMANCE OF KERNEL NON-PARAMETRIC REGRESSION AND ORDINARY LEAST SQUARES REGRESSION," *Int. J. Informatics Vis.*, vol. 8, no. 3, pp. 1352–1360, 2024. doi: https://doi.org/10.62527/joiv.8.3.2430

[32] C. R. Oktarina, I. Sriliana, and S. Nugroho, "PENALIZED SPLINE SEMIPARAMETRIC REGRESSION FOR BIVARIATE RESPONSE IN MODELING MACRO POVERTY INDICATORS," *Indones. J. Appl. Stat.*, vol. 8, no. 1, pp. 13–23, 2025. doi: https://doi.org/10.12962/j27213862.v8i3.23330.