International Journal of Multidisciplinary Research and Growth Evaluation.

Modeling of Infant and Child Mortality Rates Using Bi-response Nonparametric Regression Model Based on Penalized Spline Estimators

Firda Aulia Pratiwi [1], Mubadi'ul Fitriyani [2], Marisa Rifada [3], Nur Chamidah [4*]
[1-4] Department of Mathematics, Faculty of Science and Technology, Universitas Airlangga, Surabaya, East Java, Indonesia
[3, 4] Research Group of Statistical Modeling in Life Science, Faculty of Science and Technology, Universitas Airlangga, Surabaya, East Java, Indonesia

* Corresponding Author: **Nur Chamidah**

Article Info

ISSN (online): 2582-7138
Volume: 06
Issue: 01
January-February 2025
Received: 26-11-2024
Accepted: 29-12-2024
Page No: 941-950

Abstract
To achieve the Sustainable Development Goals (SDGs), it is essential to monitor the Infant Mortality Rate (IMR) and Child Mortality Rate (CMR), as these indicators are critical for assessing child health. West Java Province recorded the highest number of IMR cases, with 5,234 infants and 299 children. Compared to 2022, when there were 2,959 IMR-related fatalities and 173 CMR-related deaths, this represents a significant increase. The purpose of this study is to model the IMR and CMR using bi-response nonparametric regression model based on penalized spline estimators which has more flexibility to adapt pattern of data. The flexibility in penalized spline estimator is controlled by smoothing parameters, i.e. Lambda parameter, knot points and the number of knots. For determining the optimal smoothing parameter, we use criteria the minimum of Generalized Cross Validation (GCV) value. Based on the minimum GCV, we find that for an increase in the percentage of health facility deliveries, exclusive breastfeeding, and complete basic immunization tends to decrease IMR and CMR. Conversely, an increase in obstetric complications tends to increase IMR and CMR. The best model with Mean Squared Error (MSE) of 1.932198 and a coefficient of determination (R^2) of 0.8787218. It's mean that the model satisfy goodness of fits criteria.

Keywords: penalized spline, nonparametric regression, infant mortality rate, child mortality rate, weighted least squares

Introduction

The Under-Five Mortality Rate (U5MR) is the probability of a child dying between birth and their fifth birthday, expressed per thousand live births (Mutaqin *et al*., 2023) [1]. According to the Minister of Health Regulation No. 25 of 2014, infants are defined as children aged 0-11 months, while toddlers are children aged 12-59 months (Kementerian Kesehatan Republik Indonesia, 2024) [2]. Reducing infant, toddler, and neonatal mortality is a key objective of child health initiatives aimed at ensuring survival and well-being (Kementerian Kesehatan Republik Indonesia, 2023) [3]. By 2030, Indonesia aims to reduce under-five mortality to 25 deaths per 1,000 live births, aligning with the third goal of the Sustainable Development Goals (SDGs), which promotes healthy lives and well-being for all (United Nations) [4].

In 2023, Indonesia recorded a total of 34,087 under-five deaths, with 32,445 of these being infants (0-11 months), accounting for 94.8% of all under-five deaths. Meanwhile, 1,781 deaths occurred in children aged 12-59 months, representing 5.2% of the total. These figures reflect a significant increase from 2022, when the total number of under-five deaths was 21,447 (Kementerian Kesehatan Republik Indonesia, 2023) [3]. West Java province recorded the highest number of Infant Mortality Rate (IMR) cases, with 5,234 deaths, and ranked second in Child Mortality Rate (CMR) with 299 deaths. The IMR in West Java for 2023 was 6.40 per 1,000 live births (Dinkes Provinsi Jawa Barat, 2023) [5]. This figure has increased compared to 2022 where IMR cases occurred as many as 2959 deaths and CMR as many as 173 deaths (Dinkes Provinsi Jawa Barat, 2022) [6].

Previous research has explored the factors affecting IMR and CMR. For instance, Dwiargatra and Purhadi (2020) [7] used bivariate gamma regression and identified low birth weight and poverty as significant predictor variables, though they did not find strong correlations between the predictors and outcomes. In regression analysis, both parametric and nonparametric methods can be used to model relationships between predictor variables and outcomes (Nurhuda, Wasono, & Nohe, 2022) [8]. Parametric regression is applied when the form of the relationship is well-defined, whereas nonparametric regression is preferred when the shape of the relationship is uncertain, providing greater flexibility in estimating the regression curve (Suparti, Prahutama, & Santoso, 2018; Lestari *et al*., 2019) [9, 10]. Therefore, in this study, the factors influencing IMR and CMR will be modeled using a bi-response nonparametric regression approach.
Nonparametric regression models can be estimated using various estimators, including Kernel (Lestari *et al*., 2019; Astuti *et al*., 2018) [10, 11], Local Linear (Darnah *et al*., 2019) [12], Local Polynomial (Chamidah & Lestari, 2019)[13], Truncated Spline (Juniar *et al*., 2024) [14], Least Square Spline (Massaid *et al*., 2019) [15], and Penalized Spline (Chamidah, Rifada, & Amelia, 2022; Sulistianingsih, Kurniasari, & Kusnandar, 2019) [16, 17]. Penalized spline estimators, as noted by Zia *et al*. (2017) [18], offer an advantage in spline regression by focusing on key knot points in the data, which are critical for model accuracy. However, identifying the optimal knot points can be time-consuming and resource-intensive. Penalized splines address this challenge by placing knots at quantile points of the predictor variable, simplifying the process. In a study by Putri (2017) [19] using poverty data from Indonesia, the nonparametric model performed better than the parametric model in terms of R^2. Thus, this study aims to estimate the factors affecting IMR and CMR in West Java in 2023 using a bi-response nonparametric regression model with Penalized Spline estimators.

Materials

1. Bi-response Nonparametric Regression

According to Ampulembang *et al*. (2015) [20], The following equation represents the bi-response nonparametric regression model:

$$y_i^{(r)} = f^{(r)}(x_i) + \varepsilon_i^{(r)}, r = 1,2\,; i = 1,2,\dots,n \quad (1)$$

The variables y, x, f, and ε represent the response variable, predictor variable, regression function, and random error, respectively. An alternative expression for Equation (1), applied to the i-th observation of the r-th response variable, is as follows:

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

where $\boldsymbol{f}(\boldsymbol{x_i}) = (f^{(1)}(x_i), f^{(2)}(x_i))$ is the vector of the regression function, and $\boldsymbol{y_i} = \left(y_i^{(1)}, y_i^{(2)}\right)$ represents the bi-response variables. The error term $\boldsymbol{\varepsilon_i} = \left(\varepsilon_i{}^{(1)}, \varepsilon_i{}^{(2)}\right)$ is a vector of random errors with a mean of zero and variance matrix Σ for each index *i*. This model assumes that multiple observations on the two response variables, $y_i^{(1)}$ and $y_i^{(2)}$, exhibit correlation. The correlation between these response variables is quantified by the coefficient:

$$\rho_{y_i^{(1)}, y_i^{(2)}} = \frac{cov\left(y_i^{(1)}, y_i^{(2)}\right)}{\sqrt{Var\left(y_i^{(1)}\right) Var\left(y_i^{(2)}\right)}} \neq 0$$

where $\rho_{y_i^{(1)}, y_i^{(2)}}$ reflects the degree of association between the two correlated response variables $y_i^{(1)}$ and $y_i^{(2)}$.

2. Bi-response Nonparametric Regression based on Penalized Spline Estimator

According to Lestari *et al*. (2018) [21], the nonparametric regression model based on penalized spline estimator for bi-response response is represented by the following equation:

$$y_i{}^{(r)} = \theta_{oj}^{(r)} + \sum_{j=1}^{p}\left(\sum_{h=1}^{qjr} \theta_{jh}^{(r)} x_{ji}^{(h)} + \sum_{l=1}^{m_j} \phi_{jl}^{(r)} \left(x_{ji} - K_{jl}\right)_+^{qjr}\right) + \varepsilon_i^{(r)} \quad (3)$$

where $\theta_{oj}^{(r)}$ represents the intercept coefficient for the *j*-th predictor variable within the *r*-th response variable, $\theta_{jh}^{(r)}$ and $\phi_{jl}^{(r)}$ denote the coefficients for the predictor variables and the knots in the spline function for the *r*-th response variable, respectively. The truncated function $\left(x_{ji} - K_{jl}\right)_+^{qjr}$ is defined as:

$$\left(x_{ji} - K_{jl}\right)_+^{qjr} = \begin{cases} \left(x_{ji} - K_{jl}\right)_+^{qjr}, \text{if } x_{ji} \geq K_{jl} \\ 0, \text{ if } x_{ji} < K_{jl} \end{cases}$$

3. Penalized Spline Estimator

The function of Penalized Least Square (PLS), as described by Ruppert *et al*. (2003) [22], can be formulated as follows:

$$PLS(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i^{(r)} - f(x_{ji})\right)^2 + \lambda\left(\sum_{j=1}^{p}\sum_{l=1}^{m_j}\left(\phi_{jl}^{(r)}\right)^2\right), \lambda \geq 0 \quad (4)$$

where λ represents the smoothing parameter. According to Wood (2006) [23], the smoothing parameter can be estimated using the following expression:

$$\lambda = \frac{1.5^{h-1}}{10^8} \quad (5)$$

where λ is constrained, with $h = 1, 2, 3, \dots, H$ representing specific tuning values for optimization. Minimizing the PLS function can also be achieved by utilizing the PLS functional matrix, as shown in Equation (6):

$$Q = \frac{1}{n}(\boldsymbol{y} - \boldsymbol{T\beta})^T(\boldsymbol{y} - \boldsymbol{T\beta}) + \lambda\boldsymbol{\beta}^{\boldsymbol{T}}\boldsymbol{D\beta} \quad (6)$$

In this formulation, Q represents the penalized objective function, $\boldsymbol{T}$ is the matrix of predictor variables, $\boldsymbol{\beta}$ is the vector of coefficients, and $\boldsymbol{D}$ is a matrix associated with the penalty applied to the spline coefficients. This framework enables effective smoothing in the penalized spline estimation process, balancing the fit of the model with the complexity of the spline.

4. Weighted Least Square (WLS) Parameter Estimation and Penalized Spline Nonparametric Regression Model Parameter Estimation

In bi-response regression, the Weighted Least Squares (WLS) method is employed to minimize the weighted sum of squared errors, which is particularly useful when the assumption of constant error variance is violated (Greene, 2003)[24]. The WLS objective function can be formulated as follows:

$$\boldsymbol{\varepsilon}^{\boldsymbol{T}}\boldsymbol{W}_{\boldsymbol{\varepsilon}} = (\boldsymbol{y} - \boldsymbol{T\beta})^{\boldsymbol{T}}\boldsymbol{W}(\boldsymbol{y} - \boldsymbol{T\beta}) \quad (7)$$

In this context, the variation-covariance matrices of Respondents 1 and 2 are inverted, as the WLS estimator uses the weight matrix $\boldsymbol{W}$:

$$\boldsymbol{W} = [\Sigma]^{-1} = \begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}^{-1}$$

A penalized spline estimator model without weights is then applied in the nonparametric regression estimation for the bi-response. The model estimation can be expressed mathematically as:

$$\hat{\boldsymbol{y}} = \boldsymbol{T}(\boldsymbol{T}^{\boldsymbol{T}}\boldsymbol{T} + \boldsymbol{n}\lambda\boldsymbol{D})^{-1}\,\boldsymbol{T}^{\boldsymbol{T}}y = \boldsymbol{H}(\lambda)\boldsymbol{y} \quad (8)$$

where $\boldsymbol{H}(\lambda)$ represents the smoothing matrix that incorporates the penalty parameter λ.

Furthermore, parameter estimation in bi-response nonparametric spline regression is obtained by minimizing the PLS function using the WLS method. To achieve optimal estimation, the function Q must satisfy the condition ∂Q/∂β=0, ensuring that Q attains its minimum value.

$$\hat{\boldsymbol{y}} = \boldsymbol{T}(\boldsymbol{T}^{\boldsymbol{T}}\boldsymbol{W}\boldsymbol{T} + \boldsymbol{n}\lambda\boldsymbol{D})^{-1}\,\boldsymbol{T}^{\boldsymbol{T}}\boldsymbol{W}y = \boldsymbol{H}(\lambda)^{*}\boldsymbol{y} \quad (9)$$

The choice of the smoothing parameter (λ) and knot points determines the smoothness level of the penalized spline nonparametric regression. The smallest Generalized Cross Validation (GCV) value provides the ideal smoothing parameter. The formula for determining the GCV value in bi-response regression, as stated by Eubank, is:

$$GCV(\lambda) = \frac{MSE}{\left[\frac{1}{n}trace(\boldsymbol{I} - \boldsymbol{H}(\lambda))\right]^2} \quad (10)$$

where $\boldsymbol{H}(\lambda) = \boldsymbol{T}(\boldsymbol{T}^{\boldsymbol{T}}\boldsymbol{W}\boldsymbol{T} + \boldsymbol{n}\lambda\boldsymbol{D})^{-1}\,\boldsymbol{T}^{\boldsymbol{T}}\boldsymbol{W}$ and $MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$.

The bi-response regression model can be represented nonparametrically if the smoothing parameter and knot values are selected using the minimal GCV criterion. The fit of the resulting bi-response regression model can be evaluated using the coefficient of determination, as shown in the following equation:

$$R^2 = 1 - \frac{\sum_{j=1}^{2}\sum_{i=1}^{n}\left((y_i^{(j)} - \hat{y}_i^{(j)})\right)^2}{\sum_{j=1}^{2}\sum_{i=1}^{n}\left((y_i^{(j)} - \bar{y}^{(j)})\right)^2} \quad (11)$$

This is achieved by representing the actual value of the j-th response variable, $y_i^{(j)}$, in the i-th observation. Meanwhile, $\bar{y}^{(j)}$ denotes the mean of the j-th response variable across all observations, and $\hat{y}_i^{(j)}$ represents the estimated mean of the j-th response variable for the i-th observation.

Methods

The 2023 West Java Health Profile and data from the West Java Central Statistics Agency were used in this study. The research employed 27 cities/regency in West Java Province in 2023 as its observation units. The dataset comprised two response variables and four predictor variables. Table 1 provides a list of all variables utilized in this research.

Table 1: List of the Research Variables

Variable	Description
Y_1	Infant Mortality Rate (IMR)
Y_2	Child Mortality Rate (CMR)
X_1	Percentage of Health Facility Deliveries
X_2	Percentage of Exclusive Breastfeeding
X_3	Percentage of Complete Basic Immunization
X_4	Percentage of Obstetric Complications

The parameters for Infant Mortality Rate (IMR) and Child Mortality Rate (CMR) are estimated using the Weighted Least Squares (WLS) method within a penalized spline multipredictor nonparametric regression framework.

Results and Discussion

1. Descriptive Statistics

A descriptive statistical analysis was conducted to provide an overview of the research variables, with the results presented in Figure 1, Figure 2, and Table 2 below.

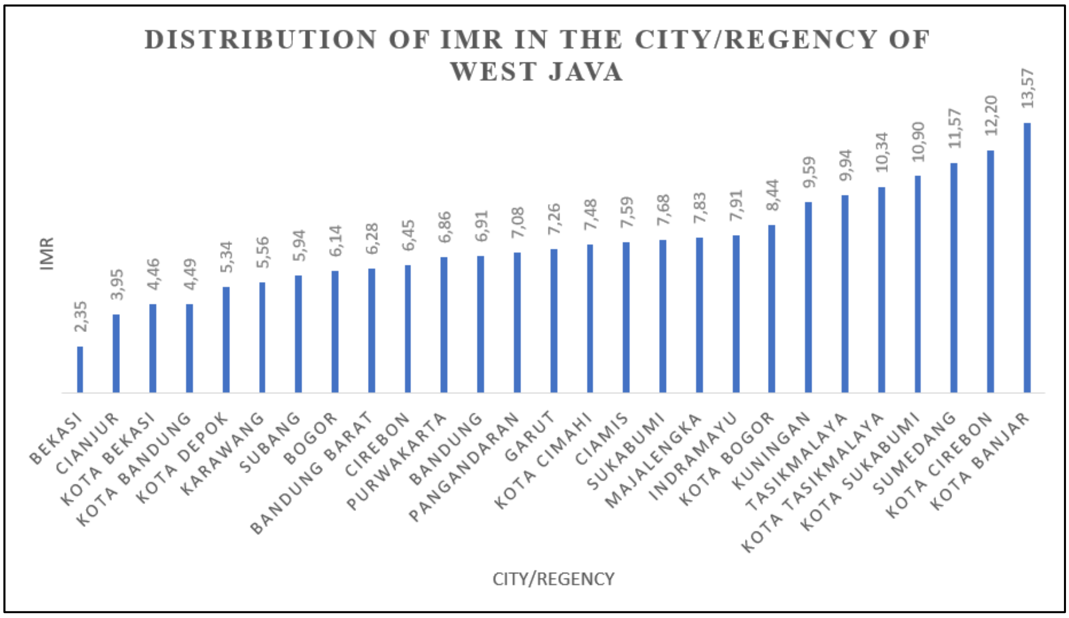

Fig 1: Infant Mortality Rate (IMR) Distribution across Cities in West Java

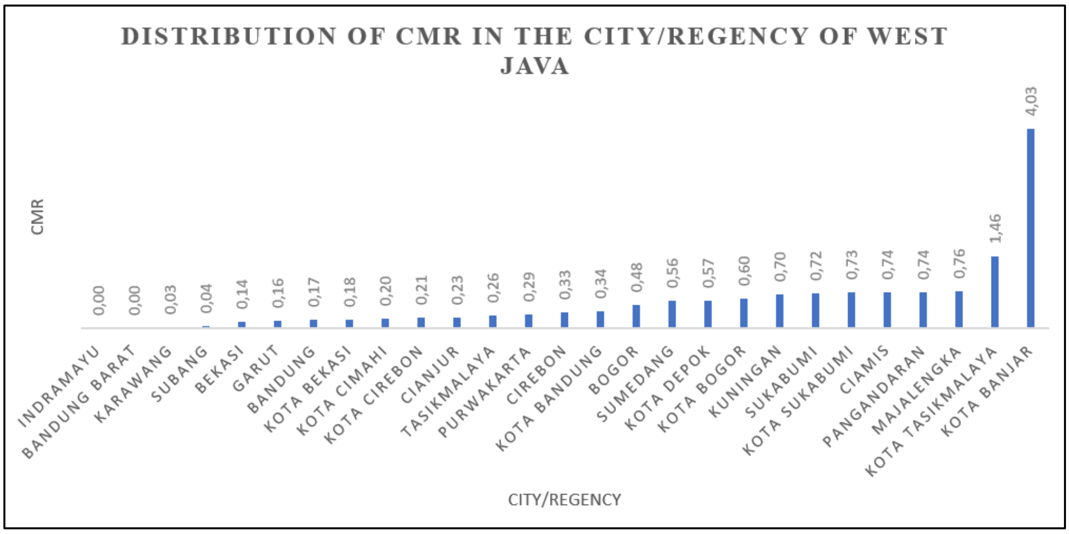

Fig 2: Child Mortality Rate (CMR) Distribution across Cities in West Java

Table 2: Descriptive Statistics of the Research Variables

Variable	Mean	St.Dev	Median	Min	Max
IMR (Y_1)	7.56	2.63	7.26	2.35	13.57
CMR (Y_2)	0.56	0.77	0.33	0.00	4.03
Percentage of Health Facility Deliveries (X_1)	95.96	6.9	96.8	84.7	121.4
Percentage of Exclusive Breastfeeding (X_2)	74.35	16.51	73.90	41.6	133.60
Percentage of Complete Basic Immunization (X_3)	98.07	7.53	98.40	82.7	123.8
Percentage of Obstetric Complications (X_4)	108	43.51	102	28	242

Based on Table 2, the Infant Mortality Rate (IMR) has an average of 7.56 with a standard deviation of 2.63, while the Child Mortality Rate (CMR) shows a lower mean of 0.56 with a standard deviation of 0.77, suggesting higher variability in IMR. Health facility deliveries are notably high, averaging 95.96%, reflecting substantial access to healthcare facilities. Exclusive breastfeeding rates are moderate, with an average of 74.35% and greater variability with standard deviation of 16.51, indicating differences in breastfeeding practices. Basic immunization coverage is also high, with a mean of 98.07%, signifying robust immunization efforts. However, the percentage of obstetric complications is relatively high, averaging 108, and showing significant variability, i.e standard deviation of 43.51, which might indicate healthcare challenges related to maternal health.

2. Correlation Test

The Pearson correlation test between the Infant Mortality Rate (IMR) and Child Mortality Rate (CMR) yielded a correlation coefficient of 0.5768, with a p-value of 0.0016, which is below the significance level of α = 5%. This result leads to the rejection of the null hypothesis (H_0), suggesting a significant positive association between IMR and CMR. This correlation supports the application of a biresponse regression analysis, as both response variables are related and may share predictive factors. To further explore these relationships, a scatterplot was generated to visualize the influence of the four predictor variables (health facility deliveries, exclusive breastfeeding, basic immunization, and obstetric complications) on IMR and CMR, providing a basis for examining potential predictive patterns.

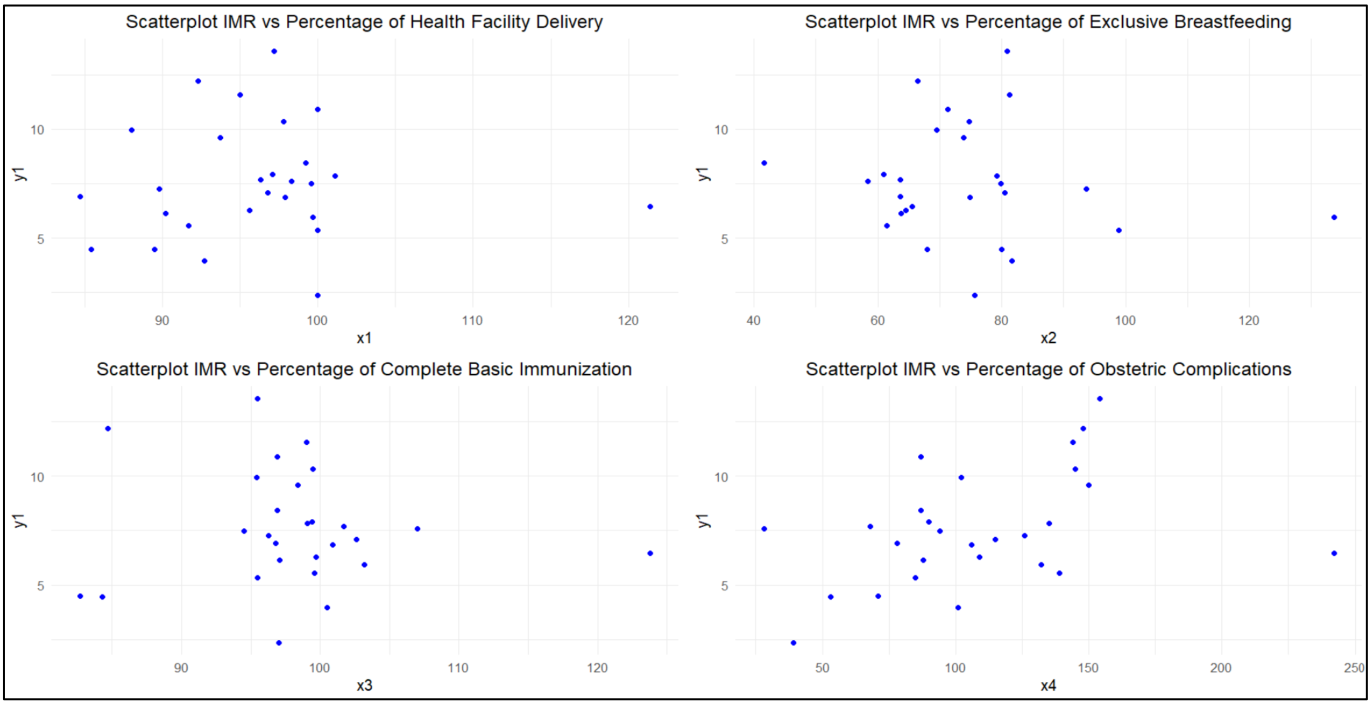

Fig 3: Scatterplot of response variables Y_1 against predictor variables X_1, X_2, X_3, X_4

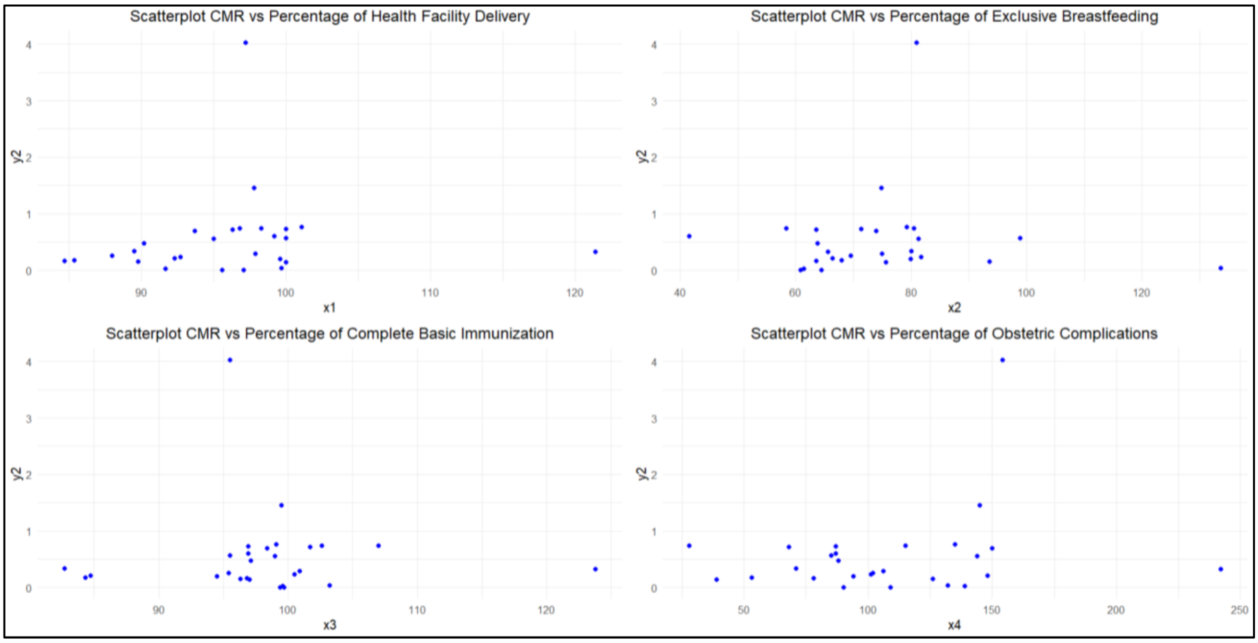

Fig 4: Scatterplot of response variables Y_2 against predictor variables X_1, X_2, X_3, X_4

The scatterplots reveal that there is no strong linear association between the Child Mortality Rate (CMR) and the four predictor variables. In each plot, data points are widely dispersed, particularly around lower CMR values, without clear trends indicating that higher levels of these health indicators correspond to lower or higher CMR. This lack of observable patterns suggests that the relationship between CMR and these predictors cannot be identified. The next step is to examine the linear relationships between the response variables and predictors to determine if significant associations exist. The results of this analysis are presented in Table 3.

Table 3: Results of the Correlation Analysis between Response Variables and Predictors

Variable		*P-value*	Alpha	Decision	Conclusion
Response	**Predictor**				
$\boldsymbol{Y_1}$	X_1	0.9104	0,05	Accept H_0	Not Linear
	X_2	0.5219	0,05	Accept H_0	Not Linear
	X_3	0.7751	0,05	Accept H_0	Not Linear
	X_4	0.0378	0,05	Reject H_0	Linear
$\boldsymbol{Y_2}$	X_1	0.5625	0,05	Accept H_0	Not Linear
	X_2	0.9079	0,05	Accept H_0	Not Linear
	X_3	0.9960	0,05	Accept H_0	Not Linear
	X_4	0.2762	0,05	Accept H_0	Not Linear

Based on the results in Table 3, it was implied that only obstetric complications may have a direct linear effect on the Infant Mortality Rate (Y_1), while no linear associations were found for the other predictors with either Y_1 or Y_2. Given the lack of significant linear relationships for most predictors, a nonparametric approach would be more appropriate. Nonparametric methods can capture complex, nonlinear associations that may exist between the response variables and predictors, providing a more flexible framework that does not rely on the strict assumptions of linearity. This approach could yield more accurate insights into the underlying relationships in the data.

3. Multipredictor Biresponse Nonparametric Regression Modelling Based on Penalized Spline Estimator

Based on the analysis of IMR and CMR in relation to each influential predictor variable, the optimal smoothing parameters were identified using the criterion of minimal Generalized Cross-Validation (GCV) value, as shown in Table 4. These parameters include the order of the smoothing function, the number of knots, and the specific knot points for each predictor variable. By minimizing the GCV value, the chosen parameters provide the best balance between model fit and complexity, ensuring a smooth yet flexible representation of the relationship between IMR, CMR, and their predictors.

Table 4: Optimal Order and Number of Knots for Each Predictor Variable Based on GCV value

Predictor	Orde		Total of Knot	Knot Point	Minimum GCV Value
	Y_1	Y_2			
X_1	1	1	1	96.8	4.173423
X_2	1	1	1	73.9	4.349335
X_3	1	1	1	98.4	4.31414
X_4	1	1	1	102	3.758504

The analysis results in Table 4 indicate that the spline model for each predictor yields distinct knot points and minimum GCV values. For predictor X_1, a spline model with a single knot at 96.8 achieves a minimum GCV value of 4.1734. For predictor X_2, the spline model with one knot at 73.9 yields a minimum GCV value of 4.3493. Similarly, the spline model for predictor X_3 with a knot at 98.4 results in a minimum GCV value of 4.3141. Lastly, the model for predictor X_4 with a knot at 102 produces the lowest GCV value at 3.7585. These results highlight that each predictor has a unique optimal knot point.

The estimation of the multi-predictor bi-response nonparametric regression model will be conducted using weight **W** based on the optimal smoothing parameters identified in Table 4. After gathering the smoothing parameters (order, number of knots, and knot points) for each predictor, these weights will guide the model fitting process to capture the complex relationships between the predictors and response variables effectively. The results of this model estimation are presented in Table 5, providing the optimum lamda for each predictor within the bi-response framework.

Table 5: Model Estimation Results using **W** Weighting

Predictor	Orde		Total of Knot	Optimum of Lamda	Minimum of GCV
	Y_1	Y_2			
X_1	1	1	1	0.000001	4.250166
				0.050001	4.192245
				0.100001	4.164397
				0.150001	4.151618
				0.200001	**4.147375**
				0.250001	4.148340
				0.300001	4.152615
				0.350001	4.159034
X_2	1	1	1	0.400001	4.166839
				0.450001	4.175519
				0.500001	4.184720
				0.550001	4.194193
X_3	1	1	1	0.600001	4.203761
				0.650001	4.213297
				0.700001	4.222710
				0.750001	4.231937
X_4	1	1	1	0.800001	4.240932
				0.850001	4.249665
				0.900001	4.258117
				0.950001	4.266276

Based on Table 5, the minimum GCV value achieved is 4.1474, with the following optimal smoothing parameters for each predictor: (a) For predictor 1, both the first and second response orders are 1, with 1 knot. (b) For predictor 2, the first and second response orders are also 1, with a single knot. (c) For predictor 3, both response orders are set to 1, with 1 knot as well. (d) Predictor 4 has similar parameters, with both response orders at 1 and 1 knot. Additionally, the optimal smoothing parameter lambda is set at 0.20001. These parameters collectively contribute to minimizing the GCV value, optimizing the model fit for the multi-predictor bi-response framework.

The estimation of the nonparametric bi-response multi-predictor regression model based on a penalized spline estimator, applied to IMR data in West Java, is as follows:

$$\hat{Y}^{(1)} = -0.642 + 0.167X_1 - 0.354(X_1 - 96.8)_+ - 0.642 - 0.05X_2 + 0.007(X_2 - 73.9)_+ - 0.642 - 0.051X_3 - 0.124(X_3 - 98.4)_+ - 0.642 + 0.03X_4 + 0.027(X_4 - 102)_+ \quad (12)$$

The estimation of the nonparametric bi-response multi-predictor regression model based on a penalized spline estimator, applied to CMR data in West Java, is as follows:

$$\hat{Y}^{(2)} = -1.446 + 0.063X_1 - 0.111(X_1 - 96.8)_+ - 1.446 + 0.009X_2 - 0.017(X_2 - 73.9)_+ - 1.446 + 0.004X_3 - 0.063(X_3 - 98.4)_+ \quad (13)$$

$$-1.446 - 0.008X_4 + 0.027(X_4 - 102)_+$$

Based on equations (12) and (13), the variable Percentage of Health Facility Deliveries (X_1), assuming other variables are held constant, can be interpreted as follows:

$$\hat{f}^{(1)}(X_1) = \begin{cases} -0.642 + 0.167X_1, if\ X_1 \le 96.8 \\ 33.625 - 0.187X_1, if\ X_1 > 96.8 \end{cases} \quad (14)$$

$$\hat{f}^{(2)}(X_1) = \begin{cases} -1.446 + 0.063X_1, if\ X_1 \le 96.8 \\ 9.299 - 0.048X_1, if\ X_1 > 96.8 \end{cases} \quad (15)$$

From the model above, it can be observed that when the percentage of health facility deliveries is below 96.8%, each one percent increase in health facility deliveries tends to increase IMR by 0.167 and CMR by 0.063. Conversely, when the percentage of health facility deliveries exceeds 96.8%, each additional one percent increase tends to decrease IMR by 0.187 and CMR by 0.048.

Based on equations (12) and (13), the variable Percentage of Exclusive Breastfeeding (X_2), assuming other variables are held constant, can be interpreted as follows:

$$\hat{f}^{(1)}(X_2) = \begin{cases} -0.642 - 0.05X_2, if\ X_2 \leq 73.9 \\ -1.159 - 0.043X_2, if\ X_2 > 73.9 \end{cases} \quad (16)$$

$$\hat{f}^{(2)}(X_2) = \begin{cases} -1.446 + 0.009X_2, if\ X_2 \leq 73.9 \\ -0.1897 - 0.008X_2, if\ X_2 > 73.9 \end{cases} \quad (17)$$

From the model above, it can be observed that when the percentage of exclusive breastfeeding is below 73.9%, each one percent increase in exclusive breastfeeding tends to decrease IMR by 0.05 and increase CMR by 0.009. Conversely, when the percentage of exclusive breastfeeding exceeds 73.9%, each additional one percent increase tends to decrease IMR by 0.043 and CMR by 0.008.

Based on equations (12) and (13), the variable Percentage of Complete Basic Immunization (X_3), assuming other variables are held constant, can be interpreted as follows:

$$\hat{f}^{(1)}(X_3) = \begin{cases} -0.642 - 0.051X_3, if\ X_3 \leq 98.4 \\ 11.56 - 0.175X_3, if\ X_3 > 98.4 \end{cases} \quad (18)$$

$$\hat{f}^{(2)}(X_3) = \begin{cases} -1.446 + 0.004X_3, if\ X_3 \leq 98.4 \\ -4.753 - 0.059X_3, if\ X_3 > 98.4 \end{cases} \quad (19)$$

From the model above, it can be observed that when the percentage of complete basic immunization is below 98.4%, each one percent increase in immunization coverage tends to decrease IMR by 0.051 and increase CMR by 0.004. Conversely, when immunization coverage exceeds 98.4%, each additional one percent increase tends to decrease IMR by 0.175 and CMR by 0.059.

Based on equations (12) and (13), the variable Percentage of Obstetric Complications (X_4) with assuming other variables are held constant can be interpreted as follows:

$$\hat{f}^{(1)}(X_4) = \begin{cases} -0.642 + 0.03X_4, if\ X_4 \leq 102 \\ -3.396 + 0.057X_4, if\ X_4 > 102 \end{cases} \quad (20)$$

$$\hat{f}^{(2)}(X_4) = \begin{cases} -1.446 - 0.008X_4, if\ X_4 \leq 102 \\ -4.2 + 0.019X_4, if\ X_4 > 102 \end{cases} \quad (21)$$

From the model above, it can be observed that when the percentage of obstetric complications is below 102%, each one percent increase in obstetric complications tends to increase IMR by 0.01 and decrease CMR by 0.004. Conversely, when the percentage of obstetric complications exceeds 102%, each additional one percent increase tends to raise both IMR by 0.07 and CMR by 0.021.

Furthermore, the observed and estimated values for IMR and CMR across various cities or regencies are presented in Figure 5 and Figure 6. The red dots represent the actual observed values, while the blue line indicates the estimated values based on the model. Both Figure 5 and Figure 6 indicate minimal to no discrepancies between the observed data and the model estimates, suggesting a strong agreement between the two data sets across all response variables.

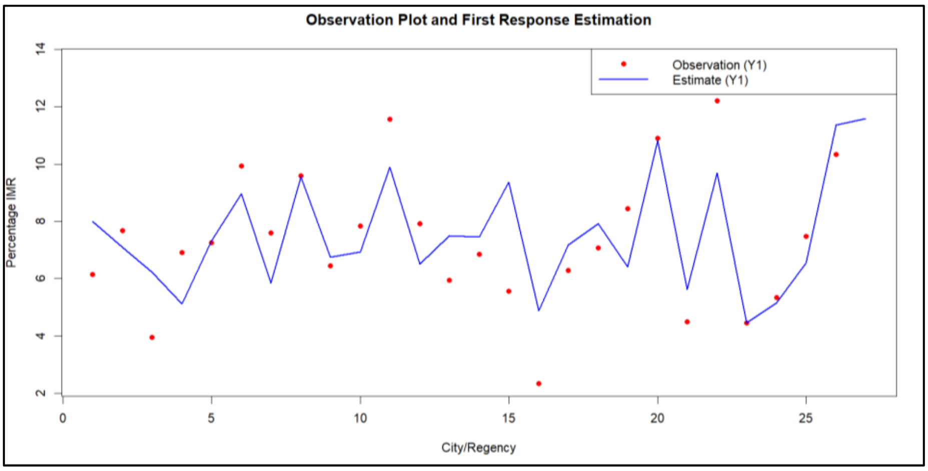

Fig 5: The Observed and Estimated Values for IMR across various cities/regencies

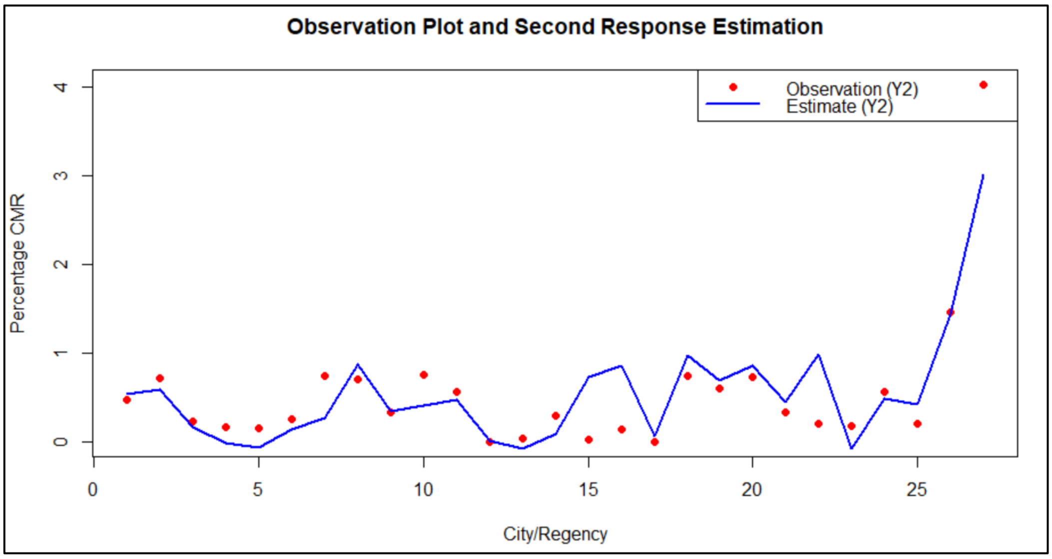

Fig 6: The Observed and Estimated Values for CMR across various cities/regencies

A comparison was conducted between parametric linear regression methods and bi-response nonparametric regression with multiple predictors to assess the effectiveness of each model. This analysis aims to evaluate the Mean Squared Error (MSE) and R^2 values, which indicate how well the four predictor variables explain the IMR and CMR response variables. The results of this comparison are presented in Table 6.

Table 6: Comparison of Parametric and Nonparametric Regression Model Results

Method	MSE	R^2
Bi-response Nonparametric Regression	1.932198	0.8787218
Bi-response Linear Parametric Regression	5.304398	0.2683264

Based on the analysis of both methods, the parametric regression model has an MSE of 5.3044 and an R^2 of 0.2683, indicating that it explains only 26.8% of the variation. In contrast, the nonparametric model achieves an MSE of 1.9322 and an R^2 of 0.8787, explaining approximately 87.8% of the variation. These results suggest that nonparametric regression is significantly more effective in capturing the complex relationships between the predictor and response variables.

Conclusion

The analysis demonstrated that the nonparametric regression model with a penalized spline estimator for modeling IMR and CMR in West Java in 2023 yielded a GCV of 3.5227, an MSE of 1.9322, and an R^2 of 0.8787. This model accounted for 87.87% of the variability in IMR and CMR, with the remaining 12.13% attributable to factors outside the model. According to the nonparametric model, an increase in the percentage of health facility deliveries, exclusive breastfeeding, and complete basic immunization is associated with reductions in IMR and CMR, while a rise in obstetric complications is associated with increases in these rates.

In comparison, a parametric regression model applied to the same dataset resulted in an MSE of 5.304398 and an R^2 of 0.2683264. These findings suggest that the nonparametric model performs more effectively in minimizing MSE and maximizing R^2, thereby providing a better fit for the nonlinear relationships between the predictor variables and IMR and CMR than the parametric model.

References

1. Mutaqin JS, Makful MR, Hastono SP, Sipahutar T. Determinan Kematian Balita di Provinsi Bangka Belitung, Gorontalo, dan Papua Barat: Analisis Data SDKI 2017. J Biostat Kependudukan, dan Inform Kesehat. 2023;3(3):177. doi:10.51181/bikfokes.v3i3.7108
2. Kementerian Kesehatan Republik Indonesia. Peraturan Kesehatan Republik Indonesia Nomor 25 Tahun 2014 Tentang Upaya Kesehatan Anak.; c2024. https://peraturan.bpk.go.id/Details/117562/permenkes-no-25-tahun-2014
3. Kemenkes RI. Profil Kesehatan Indonesia 2023; 2023. https://kemkes.go.id/id/profil-kesehatan-indonesia-2023
4. United Nations. Transforming Our World: The 2030 Agenda for Sustainable Development. Vol Part F2740.; 2015. doi:10.1007/978-3-031-07461-5_1
5. Dinkes Provinsi Jawa Barat. Profil Kesehatan Jawa Barat 2023; 2023. http://repo.iain-tulungagung.ac.id/5510/5/BAB 2
6. Dinas Kesehatan Provinsi Jawa Barat. Profil Kesehatan Provinsi Jawa Barat 2022.; 2022.
7. Dwiargatra A, Purhadi. Pemodelan Faktor- Faktor yang Mempengaruhi Angka Kematian Bayi dan Angka Kematian Anak di Provinsi Jawa Timur Tahun 2017 Menggunakan Bivariat Gamma Regression. J Sains dan Seni ITS. 2019;8(2):2337-3520. doi:10.12962/j23373520.v8i2.44406
8. Nurhuda GN, Wasono W, Nohe DA. Nonparametric Regression Modeling Based on Spline Truncated Estimator on Simulation Data. J Mat Stat dan Komputasi. 2022;19(1):172-182. doi:10.20956/j.v19i1.21534

9. Suparti, Prahutama A, Santoso R. Mix local polynomial and spline truncated: The development of nonparametric regression model. In: Journal of Physics: Conference Series. Vol 1025. ; 2018:1-7. doi:10.1088/1742-6596/1025/1/012102
10. Lestari B, Fatmawati, Budiantara IN, Chamidah N. Smoothing parameter selection method for multiresponse nonparametric regression model using smoothing spline and Kernel estimators approaches. In: Journal of Physics: Conference Series. Vol 1397. ; 2019:1-8. doi:10.1088/1742-6596/1397/1/012064
11. Astuti DAD, Srinadi IGAM, Susilawati M. Pendekatan Regresi Nonparametrik Dengan Menggunakan Estimator Kernel Pada Data Kurs Rupiah Terhadap Dolar Amerika Serikat. E-Jurnal Mat. 2018;7(4):305-310. doi:10.24843/mtk.2018.v07.i04.p218
12. Darnah, Utoyo MI, Chamidah N. Modeling of Maternal Mortality and Infant Mortality Cases in East Kalimantan using Poisson Regression Approach Based on Local Linear Estimator. In: IOP Conference Series: Earth and Environmental Science. Vol 243. ; 2019:1-7. doi:10.1088/1755-1315/243/1/012023
13. Chamidah N, Lestari B. Estimation of covariance matrix using multi-response local polynomial estimator for designing children growth charts: A theoretically discussion. In: Journal of Physics: Conference Series. Vol 1397. ; 2019:1-8. doi:10.1088/1742-6596/1397/1/012072
14. Juniar MA, Fania A, Ulya D, Ramadhan R, Chamidah N. Modelling Crime Rates in Indonesia Using Truncated Spline Estimator. BAREKENG J Ilmu Mat dan Terap. 2024;18(2):1201-1216. doi:10.30598/barekengvol18iss2pp1201-1216
15. Massaid A, Hanif M, Febrianti D, Chamidah N. Modelling of Poverty Percentage of Non-Food per Capita Expenditures in Indonesia Using Least Square Spline Estimator. In: IOP Conference Series: Materials Science and Engineering. Vol 546. ; 2019:1-5. doi:10.1088/1757-899X/546/5/052044
16. Chamidah N, Rifada M, Amelia D. A Theoretical Discussion On Modeling The Number Of Covid-19 Death Cases Using Penalized Spline Negative Binomial Regression Approach. Commun Math Biol Neurosci. Published online 2022. doi:https://doi.org/10.28919/cmbn/7518
17. Sulistianingsih E, Kurniasari W, Kusnandar D. Estimasi Parameter Regresi Spline Dengan Metode Penalized Spline. Bimaster Bul Ilm Mat Stat dan Ter. 2019;8(2):175-184. doi:10.26418/bbimst.v8i2.31532
18. Zia NG, Suparti, Safitri D. Pemodelan Regresi Spline Menggunakan Metode Penalized Spline pada Data Longitudinal (Studi Kasus: Harga Penutupan Saham LQ45 Sektor Keuangan dengan Kurs USD terhadap Rupiah Periode Januari 2011-Januari 2016). J Gaussian. 2017;6(2):221-230. http://ejournal-s1.undip.ac.id/index.php/gaussian
19. Putri FPF. Estimasi Model Regresi Nonparametrik Birespon Multiprediktor Berdasarkan Estimator Penalized Spline. Universitas Airlangga; 2017.
20. Ampulembang AP, Otok BW, Rumiati AT, Budiasih. Bi-responses nonparametric regression model using MARS and its properties. Appl Math Sci. 2015;9(29-32):1417-1427. doi:https://doi.org/10.12988/ams.2015.5127
21. Lestari B, Chamidah N, Saifudin T. Estimasi Fungsi Regresi Dalam Model Regresi Nonparametrik Birespon Menggunakan Estimator Smoothing Spline dan Estimator Kernel. J Mat Stat dan Komputasi. 2018;15(2):20-25. doi:10.20956/jmsk.v15i2.5710
22. Ruppert D, Wand MP, Carroll RJ. Semiparametric Regression. Cambridge University Press; 2003.
23. Wood SN. Generalized Additive Models: An Introduction with R. CRC Press; 2006.
24. Greene WH. Econometric Analysis. 5th ed. Pearson Education; 2003.