PARAMETER
JURNAL MATEMATIKA, STATISTIKA DAN TERAPANNYA

e-ISSN. 2829-6303
Vol 04 No 2, August 2025
Pages 311-328

Modeling Poverty Rates in Indonesia Using Spline Nonparametric Regression

Ni Putu Nanik Hendayanti[1*], Maulida Nurhidayati[2], Luh Putu Ersamiya Ika Jayanthi[3], Luh Made Ira Angga Widyanti[4], I Kadek Dwi Cahyadinata[5], I Made Yoga Parwata[6]

[1,3,4,5,6]Information Systems Study Program, Faculty of Informatics and Computer Science, ITB STIKOM Bali
Raya Puputan St, No. 86 Renon, Denpasar, 80226, Bali, Indonesia
[2]Zakat and Waqf Management Study Program, Faculty of Economics and Islamic Business, Institut Agama Islam Negeri Ponorogo
Pramuka St, No. 156 Ponorogo, 63471, Jawa Timur, Indonesia

E-mail Correspondence Author: ***nanik@stikom-bali.ac.id***

Abstract
Poverty is a complex and multidimensional issue that remains a major focus in Indonesia as a developing country. Data show that the poverty rate in March 2024 reached 9.03%, exceeding the government's target of 6.5–7.5%. This study aims to analyze the factors that influence poverty levels in Indonesia by using secondary data obtained from the Central Statistics Agency (BPS) in 2023. The analytical method employed is Spline Nonparametric Regression, which is considered appropriate for processing social and economic data that tend to be sparse and non-stationary, such as Indonesia's poverty data. The results of this study indicate that among the four variables analyzed, only the Mean Years of Schooling has a significant effect on the percentage of the poor population in Indonesia. Other variables, namely GDP growth at constant prices, the open unemployment rate, and life expectancy, were not proven to have a significant effect on provincial poverty levels. The best model obtained uses Spline Nonparametric Regression with two knot points, as it has the smallest GCV value compared to the one-knot and three-knot models. Therefore, the findings of this study are expected to provide input for the government in formulating more targeted poverty alleviation policies, particularly through improvements in education.

Keywords: *Economics, Nonparametric Regression, Poverty Rate, Social, Spline*

: https://doi.org/10.30598/parameterv4i1pp311-328

Submitted: April 2025
Accepted: August 2025

1. INTRODUCTION

The problem of poverty is one of the economic challenges encountered in almost all countries throughout the world, particularly in developing countries, and Indonesia is no exception. Poverty is linked to different dimensions such as social, economic, and demographic factors. As a developing country with a growing population each year, Indonesia continues to face a high poverty rate. As of March 2024, Indonesia's poor population stood at 9.03 percent, showing a slight improvement from March 2023 and from 9.57 percent in September 2022. However, this figure remains above the national poverty reduction target stated in the 2020–2024 National Medium-Term Development Plan (RPJMN), which ranges between 6.5% and 7.5% **[1]**. This suggests that efforts to alleviate poverty remain a significant and ongoing challenge, particularly in the face of global economic uncertainties and the lingering effects of the COVID-19 pandemic.

According to **[2]**, poverty in developing nations tends to be structural, rooted in inequalities in access to education, healthcare, employment, and financial capital. **[3]** Further notes that Indonesia's poverty problem is intensified by the prevalence of informal labor and the vulnerability of both rural and urban poor populations. **[4]** argue that while poverty rates have gradually declined, external shocks such as the pandemic have pushed many near-poor households back into poverty. In addition, **[5]** emphasize that multidimensional poverty reflected through indicators such as education and health remains widespread and continues to hinder inclusive development. Therefore, a comprehensive and resilient poverty reduction strategy, incorporating both income-based and multidimensional approaches, is essential for achieving long-term social and economic progress **[6]**.

The factors that affect the poverty rate in Indonesia are complex, including economic aspects such as economic growth, job opportunities, and inflation, as well as social aspects such as education, health, and access to public services. To better understand the influence of these factors, a more sophisticated analytical approach is needed. Spline Nonparametric regression can be an effective tool for capturing nonlinear relationships between variables, which are often unexplained by linear regression methods. In this research the nonparametric regression method is adopted because the spline nonparametric regression is the method for estimating the model in accordance with the principle not constrained by the assumption of the shape of the regression curve **[7]**. Nonparametric regression offers greater flexibility in data analysis because it does not require assumptions about the shape of the data distribution or the linear relationships between variables. This approach is more appropriate to analyze heterogeneous and dynamic socio-economic data, such as poverty data in Indonesia.

Some studies that have relevance to this study are **[8]** Analysis of Multivariate Spline Nonparametric Regression to The Factor Affecting Poverty in South Sulawesi" The results of the study indicate that The best model produced by the nonparametric multivariate spline Regression in South Sulawesi is a model from tar 2 and the most influential Factor Between School Participation Rate and Per Capita GDP. Another one is **[9]** entitled "Binary Logistic Regression to Distinguish the Accuracy in Classifying Poverty Depth in Provinces in Indonesia". The findings of the study indicate that variables affecting the depth of poverty index were MYS and EYS, therefore, when the original measurement data were used, the classification results of the logistic regression model were different. This is shown by the existence of 3 provinces that should be included in the low Poverty Depth Index but are predicted to be high in the Poverty Depth Index, such as West Sumatra, West Kalimantan, and Maluku. Further, **[10]** explored the "Comparison of

Weighting Seemingly Unrelated Regression-Spatial Durbin Model for Poverty and Unemployment Factors". The result found that the SUR-SDM modeling using Customize weight gave R-Square smaller than Queen weight in two response variable since 80.6%, which compared to Queen by 80.64% for the variable poverty, and then the variable Unemployment the Customize weight gave the percentage of 92.51% smaller than Queen by 92.53% **[11]** studied on Nonparametric Linear Spline Regression Modeling of the Percentage of Poor Population in Kalimantan. Results of his research was obtained for R2 74.48% and the model formed is proper to use to model the pattern of the data and 5 Variables that have a significant effect on the Percentage of Poor Population that is the Population Growth Rate, Average age of school, Old School expectancy levels, Open Unemployment Rate, and Labor Force Participation Rate.

[12] studied on "Modeling of Poverty Level in South – South Selawesi Province Through Spline Nonparametric Regression Approach". Unemployment, population growth, and literacy rate were the key factors that contributed to the poverty level in the South Sulawesi in 2017, as it was revealed in this study. The best spline model was three-knotted. **[13]** researched on"Nonparametric Regression Estimator of Multivariable Truncated Spline for Categorical Data" the finding of the research reveal that the nonparametric regression model of Truncated Spline gives a better estimate than the logistic regression model. Some features of the proposed approach. **[14]** on "Smoothing Spline Estimator in Nonparametric Regression (Application: Poverty in Papua Province). From his research, the non-parametric Smoothing Spline is the best model for the poverty model in the Papua Province with GCV = 92.77 and R2 = 99.99%. **[15]** in a study "The Effect of Human Development Index on Poverty Model in Indonesia using Penalized Basis Spline Nonparametric Regression". His research shows that GCV value indicator and the MSE value, Of best model is an model 3 knots, smoothing parameter is 1000 and GCV value equal 11.26236 and MSE value is 11.08420.

In the context of Indonesia's development, this research is very relevant to support the government's efforts in formulating more targeted poverty alleviation policies. The data and analysis generated from nonparametric models can help governments understand the specific influence of economic and social factors in different regions, so that policies can be more appropriate to local conditions. With increasing pressure to reduce socio-economic inequality and improve equitable development across Indonesia, a more flexible and adaptive approach to analysis is essential to guide evidence-based policies. Previous studies have applied splines to regional poverty, but have not sufficiently explored the national scale with the latest post-pandemic data.

This article aims to employ nonparametric regression analysis to examine the influence of economic and social factors on poverty levels in Indonesia. Given the non-parametric approach used, this analysis is expected to depict the nuances derives from the dynamics between the multitude of factors and to provide in-depth understanding as a basis for more effective policies for poverty alleviation in most regions in Indonesia.

2. RESEARCH METHODS

2.1 Data

The data in this research was a secondary data which is obtained from the website of the Central Statistics Agency at https://www.bps.go.id/id. We use the data from the provinces in Indonesia in 2023.

2.2 Research Variables

The independent variable is a variable that influences or causes changes in, the dependent variable **[16]**. In this study, the independent variables consist of economic factors and social factors. The economic factors are represented by the GDP growth rate at constant prices (X1) and the open unemployment rate (X2). The social factors are represented by the mean years of schooling (X3) and life expectancy rate (X4).

Dependent variables are called output, criterion, or consequential variables. It is referred to as a bound variable in Indonesian. Bound variables are[sic]influenced or can be a result of independent variables **[16].** The independent variable is poverty level (X), represented as a percentage of the poor in the region. The data used was data on each province of Indonesia in 2023 for 38 good trial provinces. In the presentation of the data presented, there are 4 provinces where the data is incomplete, namely Southwest Papua, South Papua, Central Papua, and Mountainous Papua. The four provinces were subsequently removed from the research sample, and 34 provinces remained to be analyzed. The arrangement of research data is shown in **Table 1.**

Table 1. Research Data Arrangement

Province	Y	X1	X2	X3	X4
1	Y1	X1,1	X2,1	X3,1	X4,1
2	Y2	X1,2	X2,2	X3,2	X4,2
…	…	…	…	…	…
33	Y33	X1,33	X2,33	X3,33	X4,33
34	Y34	X1,34	X2,34	X3,34	X4,34

The model of independent and dependent variables appears in **Figure 1.**

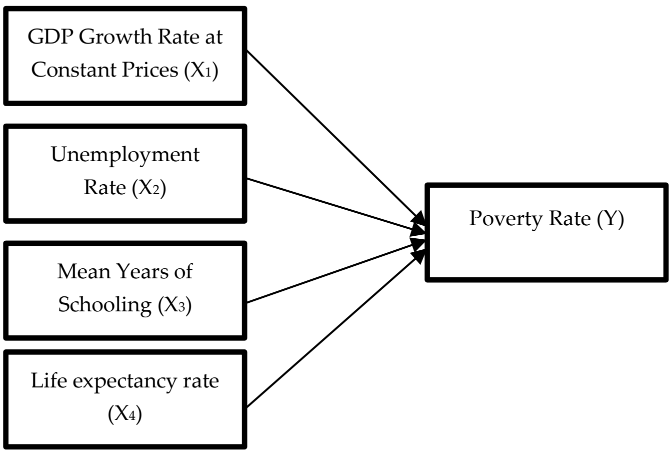

Figure 1. Conceptual Research Variables

2.3 Data Analysis Techniques

The steps of the analysis performed in the current study were:

a. Organize provincial data based on variables X1 to X4 and Y.

b. Analyze each variable descriptively to understand its distribution, mean, standard deviation, and range (minimum and maximum).
c. Describe the scatterplot between Y and each X to see the nonlinear relationship pattern.
d. Apply the Spline Nonparametric Regression technique to estimate the nonlinear relationship between variable Y and variable X.
e. Determine the optimal knot points by using the Generalized Cross Validation (GCV) criterion. The model with the smallest GCV value is considered to have the best fit.
f. Construct the Spline Nonparametric Regression model based on the selected optimal knot points.
g. Perform simultaneous and partial significance tests on the model parameters. If any predictor variables are found to be insignificant, they are excluded from the model. The analysis process is then repeated starting from step (d) to refine the model.
h. Check that residuals from the spline nonparametric regression are homoscedastic (identical variance), independent, and normally distributed. If normality is not met, transform the data and repeat the analysis from step (d).
i. Compute the coefficient of determination (R^2) to assess how well the model explains the variation in the poverty rate.
j. Interpret the model results, identify significant factors, draw conclusions, and provide recommendations for poverty alleviation policies.

2.4 Nonparametric regression

Nonparametric regression is a form of regression analysis in which the relationship between the response variables and the predictor is not specified in advance, but is instead allowed to be flexible, that is, only smooth functions in some function space are assumed. Nonparametric regression models can be generally expressed as [17]:

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

a. Estimator Spline

Spline is a segmented polynomial model. Segmented polynomials play an important role in the theory and application of statistics. The spinline is highly dependent on the knots point. The point of knots is a point of joint fusion where there is a pattern of behavior change of a function at different intervals [18].

In general, a p-ordered spline function is any function that can be written in the form of:

$$f(x) = \sum_{i=1}^{p} \alpha_i x^i + \sum_{i=1}^{h} \delta_j \left(x - k_j\right)_+^p \tag{2}$$

Suppose given is a vector with a function value at the node points. The estimated spline $\mathbf{f} = \left(f(x_1), f(x_2), \dots, f(x_n)\right) f k_1, k_2, \dots, k_n \hat{\mathbf{f}}_\lambda$ is the estimated value of the $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$ following:

$$\hat{\mathbf{f}}_\lambda = \begin{bmatrix} \hat{f}_\lambda(x_1) \\ \hat{f}_\lambda(x_2) \\ \vdots \\ \hat{f}_\lambda(x_n) \end{bmatrix} = (S_\lambda)_{n\times n} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \text{or} \quad \hat{\mathbf{f}}_\lambda = S_\lambda y \tag{3}$$

With $\hat{\mathbf{f}}_\lambda$ is a spline function with knot points for smoothing parameters and is a positively defined smoother matrix of size [19] $k_1, k_2, \dots, k_n \lambda > 0 S_\lambda n \times n$.

b. Spline regression in nonparametric

According to [17], the estimation of the is f(x) a smooth estimator. The general forms of $f_\lambda(x)$m-order spline regression are as follows:

$$y = \beta_0 + \sum_{j=1}^{m} \beta_j x^j + \sum_{k=1}^{N} \beta_{j+k} \left(x - K_k\right)_+^m + \varepsilon \tag{4}$$

Using observation data as many as n, the matrix form of equation (4) can be written as follows:

$$\mathbf{y} = \mathbf{X_1}\boldsymbol{\delta}_1 + (\mathbf{X} - \mathbf{K})\boldsymbol{\delta}_2 + \boldsymbol{\varepsilon} \tag{5}$$

With

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix};\ \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix};\ \boldsymbol{\delta}_1 = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix};\ \mathbf{X_1} = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & x_2^2 & \cdots & x_2^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m \end{bmatrix};\ \boldsymbol{\delta}_1 = \begin{bmatrix} \beta_{m+1} \\ \beta_{m+2} \\ \beta_{m+3} \\ \vdots \\ \beta_{m+n} \end{bmatrix} \tag{6}$$

$$(\mathbf{X} - \mathbf{K}) = \begin{bmatrix} (x_1 - k_1)^m & (x_1 - k_2)^m & \cdots & (x_1 - k_N)^m \\ (x_2 - k_1)^m & (x_2 - k_2)^m & \cdots & (x_2 - k_N)^m \\ \vdots & \vdots & \ddots & \vdots \\ (x_n - k_1)^m & (x_n - k_2)^m & \cdots & (x_n - k_N)^m \end{bmatrix} \tag{7}$$

c. Optimal Bandwidth Selection

Bandwidth (h), is the smoothing parameter and it controls the smoothness of the estimated curve. If the bandwidth is too narrow, the curve will be under-smoothing, that is, the curve will be quite rough and highly fluctuated; whereas if the bandwidth is too wide, the curve will be over-smoothing, which means, the curve matches the data pattern accurately, but it is quite smooth [20] Thus, the optimum bandwidth should be chosen. One way for obtaining the optimal h is through the GCV criteria which are defined by:

$$\mathrm{GCV} = \frac{\mathrm{MSE}}{\left(\frac{1}{n}\mathrm{tr}(\mathrm{I} - \mathrm{H(h)})\right)^2} \tag{8}$$

with $\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(y - m_h(x_i)\right)^2$.

The goodness of an estimator can be seen from the degree of its error. The smaller the error rate, the better the estimate.

d. Metose Least Square

One of the procedures used to estimate parameters is the least square method. In this method, an estimator will be obtained for the appropriate model for a dataset, as well as characterize the statistical properties of the obtained estimate. The least square method is generally divided into two, namely the ordinary least square (OLS) and the weighted least square (WLS) [21].

In the OLS method to define parameter estimation is as a value that minimizes the number of squares between the observation and the model called the number of squares of the error which is defined as follows:

$$\boldsymbol{\varepsilon}^{\mathbf{T}}\boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^{\mathbf{T}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (9)$$

The standard calculus technique is used to obtain the minimum number of squared errors, which is the property of a quadratic equation that reaches a minimum value when the derivative of the function is equal to zero. So, to obtain an estimator by decreasing $\boldsymbol{\varepsilon}^{\mathbf{T}}\boldsymbol{\varepsilon}$to $\boldsymbol{\beta}$and making a derivative of a function equal to zero.

3. RESULTS AND DISCUSSION

3.1 Descriptive Statistical Analysis

In this descriptive statistical analysis, the distribution of each variable consisting of mean, standard deviation, as well as minimum, and maximum, is explained. **Table 2** shows the descriptive statistics of the data.

The percentage of poor people (Y) in provinces in Indonesia varies from the lowest of 4.25% to the highest of 26.03%, with an average of 10.09%. This means that, in general, about 10% of the population in each province is in the poor category, but there are some provinces with much lower or much higher percentages. This variation can be seen from the standard deviation of 5.18%, which shows that the percentage of poor people in these provinces is quite diverse and not all of them are close to the average.

The GDP Growth Rate at Constant Prices (X1) in provinces in Indonesia shows considerable variation, with the lowest value being 1.80% and the highest value reaching 20.49%. The average GDP Growth Rate at Constant Prices is 5.40%, which shows that in general, provinces in Indonesia experience annual economic growth of 5.40%. However, the standard deviation of 3.03% shows a significant difference between provinces, where some provinces have much lower or much higher growth than the average.

Table 2. Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Y	34	4.25	26.03	10.0891	5.18351
X1	34	1.80	20.49	5.4024	3.02511
X2	34	2.27	7.52	4.6138	1.41907
X3	34	7.15	11.45	8.9276	0.91191
X4	34	63.21	73.62	69.4318	2.30980

The Unemployment rate (X2) in provinces in Indonesia varies from the lowest value of 2.27% to the highest value of 7.52%. The average Unemployment rate is 4.61%, which means that, in general, about 4.61% of the working-age population in these provinces is unemployed. The standard deviation of 1.42% indicates that the unemployment rate between provinces is quite diverse, with some provinces having lower or higher than average unemployment rates.

The Mean Years of Schooling (X3) in provinces in Indonesia varies, with the lowest score being 7.15 years and the highest score reaching 11.45 years. The Mean Years of Schooling overall was 8.93 years, which shows that, in general, the population aged 25 and above in these provinces has completed education up to almost 9th grade (junior high school) on average. The standard deviation of 0.91 years indicates that the Mean Years of Schooling in most provinces is not that different from the national average, although there are some provinces with lower or higher numbers.

Life expectancy rate (X4) in provinces in Indonesia varies, with the lowest value being 63.21 years and the highest value reaching 73.62 years. The average overall Life expectancy rate was 69.43 years, which suggests that, in general, the population in these provinces has an estimated life expectancy of up to about 69 years. The standard deviation of 2.31 years shows that the difference in Life expectancy rate between provinces is not too large, although there are some provinces with lower or higher than average figures.

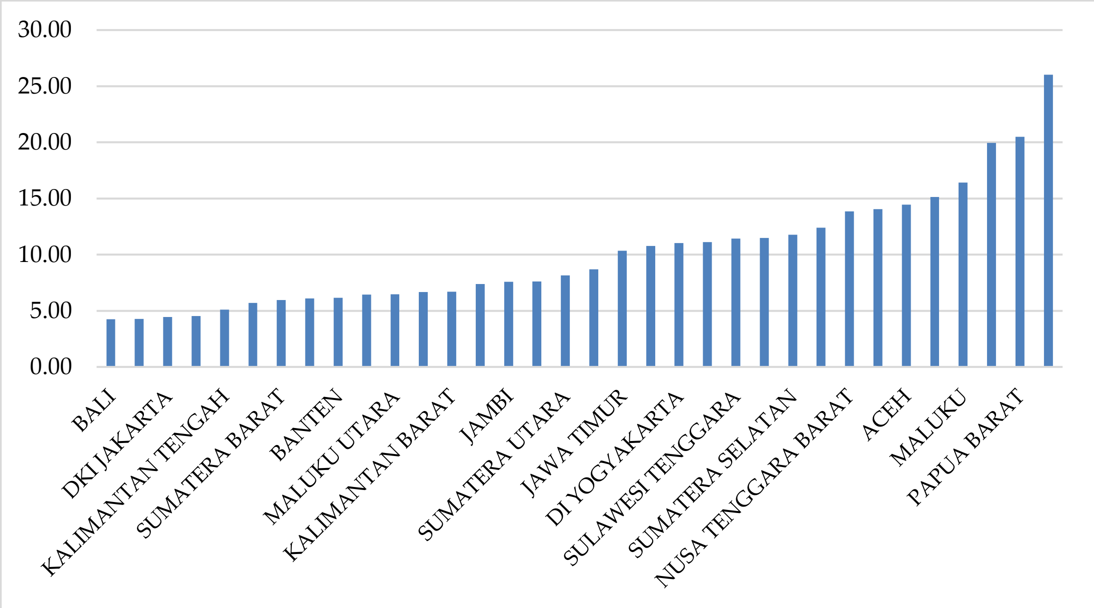

Figure 2. Percentage of Poor Population in Provinces in Indonesia

Figure 2 shows the distribution of the percentage of poor people in Indonesian provinces, which shows considerable variation. The province with the lowest percentage of poor people is Bali, followed by DKI Jakarta, Central Kalimantan, and West Sumatra. Most provinces, such as East Java, DI Yogyakarta, and Southeast Sulawesi, are in the medium value range. Meanwhile, West Papua has the highest percentage of poor people, followed by Maluku and Aceh, which shows that these provinces have numbers that are well above average. Overall, this graph shows inequality in the variable distribution of the percentage of poor people among provinces in Indonesia.

3.1. Patterns of Relationship of the Percentage of Poor Population with Factors Suspected of Influencing

The first step taken in using regression analysis is to create a scatterplot. This is done to determine the pattern of relationship between the response variable and each predictor variable. The relationship pattern is used to determine the regression method to be used.

The scatterplot displayed in Figure 3 (a) is a type of relationship pattern that can exist between the percentage of the poor population and the GDP Growth Rate at Constant Prices. The patterns created on the percentage of poor population and the GDP Growth Rate at Constant Prices do not generate pattern clean and they were added as a nonparametric element.

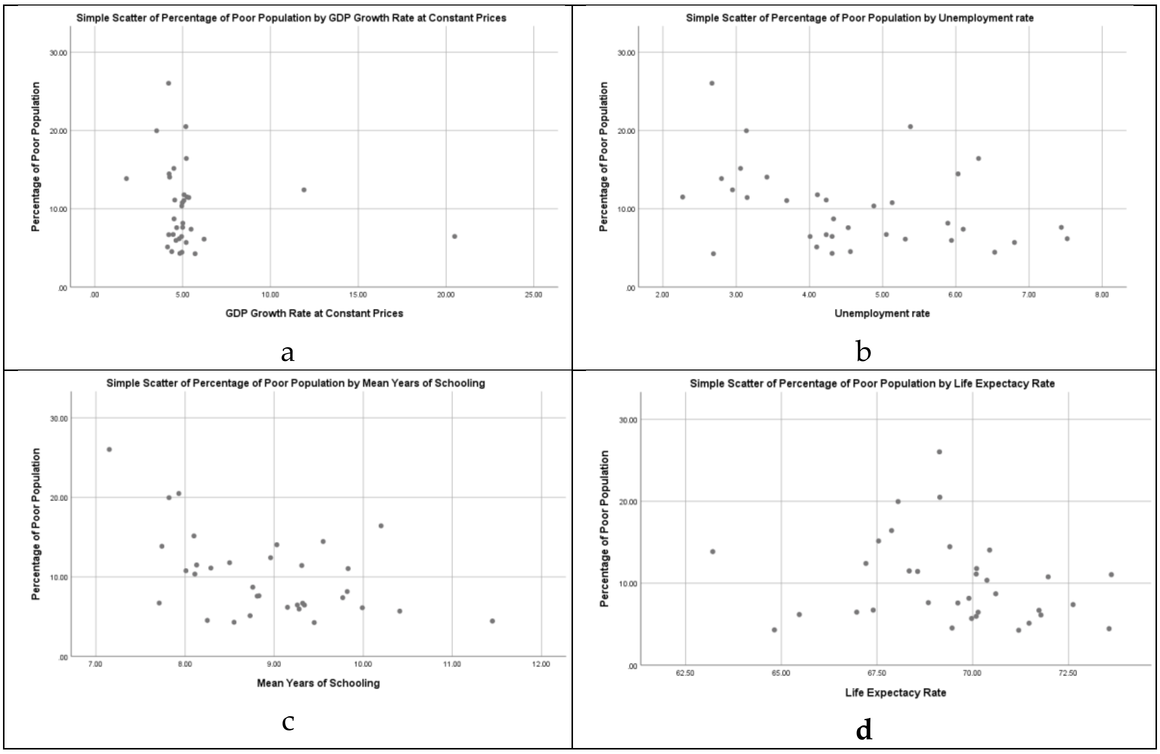

Figure 3 (a, b, c, d). Patterns of Relationship of the Percentage of Poor Population with Factors Suspected of Influencing

The scatterplot of Figure 3 (b) corresponds to a type of relationship between the variable percentage of the poor population versus the unemployment rate. The patterns that occur in the poorpercent and unemp are not patterned, so they constitute a non-parametric component.

The plot presented in the Figure 3 (c) is a type of the relationship pattern tithe variables percentage of poor population and the Mean years of schooling. The patterns shown in both percentage of poor population and MYS did not follow a specific pattern so they were considered a nonparametric component.

Figure 3 (d)is a scatter plot, which is a relation pattern between percentage variable of the poor population and Life expectancy rate. The patterns obtained on the both variables such as percentage of poor population and Life expectancy rate do not have a specific pattern therefore are nonparamertic components.

3.2. Selection of Optimal Knot Points

In Spline Nonparametric Regression model, the best model is chosen and the node point of the model is the term of common merging point where the value changes the form of the equation whose behavior changes over the data. Optimal knot point by Generalized Cross Validation (GCV) procedure on variables suspected to have influence on the percentage of the poor population. The minimum GCV tells us the best knot point. In this study, optimal knot point selection was one knot point, two knot points, and three knot points.

a. Modeling with a Single Point Knot

The GCV values for the nonparametric spline regression model with one node point are shown in Table 3 below:

Table 3. GCV Value of One Point Knot

No	GCV	X1	X2	X3	X4
1	22.97	7.14	3.77	8.38	66.18
2	23.10	7.52	3.88	8.47	66.40
3	23.13	6.76	3.66	8.29	65.97
4	23.15	7.90	3.98	8.55	66.61
5	23.30	8.28	4.09	8.64	66.82

Nonparametric regression model of spline with at one point knot on the percentage of the provincial poor population in Indonesia is as follows:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_{11}x_1 + \hat{\beta}_{12}(x_1 - k_{11})_+^1 + \hat{\beta}_{21}x_2 + \hat{\beta}_{22}(x_2 - k_{21})_+^1 + \hat{\beta}_{31}x_3 + \hat{\beta}_{32}(x_3 - k_{31})_+^1 + \hat{\beta}_{41}x_4 + \hat{\beta}_{42}(x_4 - k_{41})_+^1$$

A summary of the ten GCV values for the spline nonparametric regression model with one point of node can be found as Table below This point will be used afterwards as the minimum GCV point for the best model.

The results of the GCV value for the nonparametric regression model of the spline with one node point in **Table 3** show that the minimum GCV value is 22.97. The value of the node points for each predictor variable is as follows: X1 = 7.14; X2 = 3.77; X3= 8.38; and X4=66.18.

b. Modeling with two Knot Points

The nonparametric regression (spline with two nodes points) model with the percentage provincial poor population in Indonesia as a dependent variable:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_{11}x_1 + \hat{\beta}_{12}(x_1 - k_{11})_+^1 + \hat{\beta}_{13}(x_1 - k_{12})_+^1 + \hat{\beta}_{21}x_2 + \hat{\beta}_{22}(x_2 - k_{21})_+^1 + \widehat{\beta_{23}}(x_2 - k_{22})_+^1 + \hat{\beta}_{31}x_3 + \hat{\beta}_{32}(x_3 - k_{31})_+^1 + \hat{\beta}_{33}(x_3 - k_{32})_+^1 + \hat{\beta}_{41}x_4 + \hat{\beta}_{42}(x_4 - k_{41})_+^1 + \hat{\beta}_{43}(x_4 - k_{42})_+^1$$

Below is a table of the ten GCV values for the spline nonparametric regresssion model with two knot points whose value would ultimately be chosen as the minimum GCV for the best model. Results of GCV value to spline non-parametric regression model-2 nodes in **Table 4** indicate that the minimum GCV value is 16.40. Node points values for each predictor variable are: X1 = 9.43; 12.86; X2 = 4.1; 5.38; X3= 8.91; 9.69 and X4=67.46; 69.37.

Table 4. GCV Value of Two Point Knot

No	GCV	X1	X2	X3	X4
1	16.40	9.43	4.41	8.91	67.46
		12.86	5.38	9.69	69.37
2	16.41	9.43	4.41	8.91	67.46
		12.48	5.27	9.61	69.16
3	16.49	12.10	5.16	9.52	68.95
		12.48	5.27	9.61	69.16
4	16.71	5.23	3.23	7.94	65.12
		12.10	5.16	9.52	68.95
5	17.13	12.10	5.16	9.52	68.95
		12.86	5.38	9.69	69.37

c. Modeling with three Knot Points

The nonparametric spline regression model with three node points on Percent provincial poor population of Indonesia is specified as:

$$\begin{aligned}\hat{y} = \hat{\beta}_0 + \hat{\beta}_{11}x_1 + \hat{\beta}_{12}(x_1 - k_{11})_+^1 + \hat{\beta}_{13}(x_1 - k_{12})_+^1 + \hat{\beta}_{14}(x_1 - k_{13})_+^1 + \hat{\beta}_{21}x_2 \\ + \hat{\beta}_{22}(x_2 - k_{21})_+^1 + \hat{\beta}_{23}(x_2 - k_{22})_+^1 + \hat{\beta}_{24}(x_2 - k_{23})_+^1 + \hat{\beta}_{31}x_3 \\ + \hat{\beta}_{32}(x_3 - k_{31})_+^1 + \hat{\beta}_{33}(x_3 - k_{32})_+^1 + \hat{\beta}_{34}(x_3 - k_{33})_+^1 + \hat{\beta}_{41}x_4 \\ + \hat{\beta}_{42}(x_4 - k_{41})_+^1 + \hat{\beta}_{43}(x_4 - k_{42})_+^1 + \hat{\beta}_{44}(x_4 - k_{43})_+^1\end{aligned}$$

The following is a table showing the ten GCV values for the spline nonparametric regression model with three knot points that will later be selected as the minimum GCV for the best model.

The results of the GCV value for the spline nonparametric regression model with three knot points in **Table 5** show that the minimum GCV value is 17.27. The value of the knot points for each predictor variable is as follows: X1 = 6.76; 11.72; 12.10; X2 = 3.66; 5.06; 5.16; X3 = 8.29; 9.43; 9.52; and X4 = 65.97; 68.73; 68.95.

Table 5. GCV Value of Three Knot Points

N0	GCV	X1	X2	X3	X4
		6.76	3.66	8.29	65.97
1	17.27	11.72	5.06	9.43	68.73
		12.1	5.16	9.52	68.95
		6.38	3.56	8.2	65.76
2	17.36	11.72	5.06	9.43	68.73
		12.1	5.16	9.52	68.95
		7.14	3.77	8.38	66.18
3	17.97	11.72	5.06	9.43	68.73
		12.1	5.16	9.52	68.95
		7.14	3.77	8.38	66.18
4	18.02	7.52	3.88	8.47	66.4
		9.43	4.41	8.91	67.46
		6.76	3.66	8.29	65.97
5	18.24	7.52	3.88	8.47	66.4
		9.43	4.41	8.91	67.46

3.3. Best Model Selection

The best model is determined from the minimum GCV value between the minimum GCV value of one node point, two node points, and three node points. Here is a table with a few of the knot points for comparison.

Table 6. GCV Value Comparison

No	GCV	X1	X2	X3	X4
1 point Knot	22.97	7.14	3.77	8.38	66.18
2 knot points	16.40	9.43	4.41	8.91	67.46
		12.86	5.38	9.69	69.37
		6.76	3.66	8.29	65.97
3 knot points	17.27	11.72	5.06	9.43	68.73
		12.1	5.16	9.52	68.95

As can be observed from **Table 6**, the minimum GCV value is obtained when using a spline nonparametric regression model with 2 knot points.

3.4. Parameter Significance Testing of Spline Nonparametric Regression Model

Spline Nonparamentric Testing of the predictors of the model Spline Nonparametic regression is done together and then separately to see if the predictors are significant to the model.

a. Simultaneous Parameter Testing

Simultaneous parameter testing was used to find out whether the predictor variables had a simultaneous effect on the model with the hypothesis used.

H0 : $\beta 1 = \beta 2 = \ldots = \beta 12 = 0$

H1: there is at least one $\beta k \neq 0$; k = 1, 2, ..., 12

The following is an ANOVA table of the Spline Nonparametric Regression model

Table 7. Simultaneous Test Results

Source	df	SS	MS	F	p-value
Regression	12	512.066	42.67217	2.392173	0.03866
Error	21	374.6032	17.83825		
Total	33	886.6693			

Based on Table 7, the Fvalue of 2.392 was obtained which was then compared with the value of F(0.05; 12; 21) by 2.25. From these two values, a decision can be made to reject H0 because the value of Fcal is greater than the value of F table. Then seen from the p-value of 0.039 which is less than the value of α (0.05), it means that there is at least one significant parameter to the model.

b. Individual Parameter Testing

In simultaneous testing results, a significant parameter was concluded for the model. These parameters were determined by individual testing of the predictor variables using t-test with the hypothesis stated as follows.

H0 : $\beta k = 0$

H1 : $\beta k \neq 0$; k = 1, 2, …, 12

The following are the results of individual parameter testing of the spline nonparametric regression model.

Table 8. Individual Parameter Test Results

Variable	Parameter	Estimator	t statistic	P-Value	Results
Constant	b0	-36.294	-0.430	0.672	Insignificant
	b11	-1.779	-1.218	0.237	Insignificant
X1	b12	6.522	1.437	0.165	Insignificant
	b13	-5.831	-1.393	0.178	Insignificant
	b21	-2.392	-1.294	0.210	Insignificant
X2	b22	4.054	0.770	0.450	Insignificant
	b23	-1.579	-0.283	0.780	Insignificant
	b31	-6.320	-2.095	0.048	Signifikan
X3	b32	11.372	1.618	0.121	Insignificant
	b33	-6.643	-1.045	0.308	Insignificant
	b41	1.738	1.407	0.174	Insignificant
X4	b42	-1.393	-0.556	0.584	Insignificant
	b43	-1.285	-0.616	0.545	Insignificant

Based on the results of individual parameter testing in Table 8, it is known that the variable of Mean Years of Schooling has a significant effect on the percentage of the provincial poor population in Indonesia.

c. Residual Assumption Examination

Residual assumption tests are performed to check whether the generated residual satisfy the assumptions which are identical, independent, and normally distributed.

1. Identical Residual Assumption Checks

Examination of identical residual assumptions was carried out with the aim of finding out whether the residual variance was homogeneous and the absence of heteroscedasticity. If both of these things occur, then the residual in this study does not meet the assumption of identical residuals. The examination of these identical assumptions was carried out using the Glejser test with the following hypothesis.

H0: Residual is identical

H1: Residual tidak identic

The results of the residual assumption examination are identical to those using the Glejser test are shown in **Table 9.**

Table 9. Identical Residual Examination Results

Source	Df	SS	MS	Fhit	p-value
Regression	12	53.070	4.422	0.776	0.668
Error	21	119.69	5.700		
Total	33	172.76			

Based on **Table 9**, the Fcal value was obtained as 0.776. When compared to the value of F0.05; 12; 21 of 2.25 results in a decision to fail to reject H0 because the value of F is smaller than the value of Ftable. The p-value obtained is 0.668 where the value is greater than the value of α (0.05) so that it fails to subtract H0. This means that in this study, heteroscedasticity does not occur or has met identical assumptions.

b. Normal Distribution Residual Assumption Check

A review of the assumption of normality of residual was performed for the residual model. The test technique that was applied for examination of the remaining normality assumption is Kolmogorov-Smirnov test.

Based on the results shown in **Table 10**, it is known that the p-value=0.176 is more than 0.05, then it fails to subtract H_0 and it is concluded that the residual is normally distributed.

Table 10. Tests of Normality

	Kolmogorov-Smirnova			Shapiro-Wilk		
	Statistic	**Df**	**p-value.**	**Statistic**	**df**	**p-value.**
Residual	0.128	34	0.176	0.976	34	0.659

2. Coefficient of Determination

The value of the determination coefficient (R2) shows how good the regression model explains the variability of the percentage of poor people in the provinces in Indonesia. The following are the results of the calculation of the determination coefficient.

$$R^2 = \frac{\text{SS Regresi}}{\text{SS Total}} = \frac{512.066}{886.6693} = 0,5775 = 57,75\%$$

From the results of the calculation of the determination coefficient, the result was 57.75%, which means that the nonparametric regression spline model was able to explain the variability of the percentage of poor people in the provinces in Indonesia of 57.75% while the rest was explained by other variables that were not included in the model.

The R^2 value of 57.75% is considered small because the model is only able to explain part of the variability in poverty levels, while approximately 42.25% of the remaining variation is influenced by factors not yet included in the model. This can occur because the independent variables used are still limited, the presence of other important factors outside the variables analyzed (such as cultural aspects, local policies, or geographic conditions), the possibility of noise or errors in the data, and the complexity of the relationships between variables that cannot be fully captured by the model used.

3. Interpretation of Spline Nonparametric Regression Models

Modeling of spline nonparametric regression on the percentage of poor population in provinces in Indonesia in 2023 with the best model, namely using 2 knot points. Here is the best form of the spline nonparametric regression model with 2 node points and their parameter values.

$$\begin{aligned}\hat{y} = -36,294 - 1,779 * x_1 + 6,522 * (x_1 - 9,43)_+^1 - 5,831 * (x_1 - 12,86)_+^1 - 2,392 * x_2 \\ + 4,054 * (x_2 - 4,41)_+^1 - 1,579 * (x_2 - 5,38)_+^1 - 6,320 * x_3 + 11,372 \\ * (x_3 - 8,91)_+^1 - 6,643 * (x_3 - 9,69)_+^1 + 1,738 * x_4 - 1,393 * (x_4 - 67,46)_+^1 \\ - 1,285 * (x_4 - 69,37)_+^1\end{aligned}$$

Based on the results of the nonparametric regression analysis of the spline with the number of two knots, significant information was obtained about the influence of economic and social variables on the percentage of the poor population in Indonesia. This result is in contrast to the results of the study **[12],** which shows that the best model for modeling nonparametric spline regression is three-point knots. This result indicates that modeling with nonparametric spline regression must still test all the knots starting from 1 knot point, 2 knot points, and 3 knots, because each different data will give different results.

The results of this study provide relatively low R2 results when compared to the estimation method using Smoothing Spline conducted by **[13]**. However, this difference in results does not guarantee which method is the best, this is because the selection of variables will also give different results. So, to compare the best model of the 2 estimates, it can be done by choosing the same variable to compare so that the model that gives the best results is known. The results of the study **[14]** discuss the same nonparametric regression but with different estimation methods and data types. The results of this study show that nonparametric regression is very broad and possible to be developed.

The results of the estimates show that the GDP Growth Rate at Constant Prices (X1) has a negative relationship with the percentage of the poor population (Y), especially in the low to medium growth value range. This shows that an increase in national economic activity tends to be followed by a decrease in poverty rates. However, this relationship is not linear because at higher levels of economic growth, the decrease in poverty is no longer significant, indicating a possible saturation effect.

The unemployment rate (X2) shows a positive relationship with poverty, especially in the range of medium to high unemployment rates. When the unemployment rate increases, the percentage of the poor also tends to increase significantly. This pattern shows that unemployment is one of the main factors that worsen poverty conditions, especially if available jobs are not able to absorb the labor force optimally.

The Mean Years of Schooling (X3) shows a negative relationship with the percentage of the poor population. This means that the higher the Mean Years of Schooling in a region, the lower the poverty rate. However, the decrease in poverty tends to be sharper in the increase in Mean Years of Schooling from low to middle levels. Once it reaches a certain point, the increase in school length no longer has a significant impact on poverty reduction. This shows the importance of equal access to primary and secondary education in overcoming poverty.

Meanwhile, Life expectancy rate (X4) also shows a negative relationship with poverty rates. Areas with higher Life expectancy rates tend to have a lower percentage of poor populations. This indicates that areas with better access to health services and a high quality of life are able to reduce poverty rates. Just like the education variable, the effect of Life expectancy rate on poverty reduction is also more significant in the low to medium value range.

Overall, the nonparametric regression approach of the spline succeeded in uncovering a pattern of nonlinear relationships between economic and social variables to poverty. This model is more flexible than parametric regression because it is able to adjust the shape of the curve to the actual data, resulting in more accurate and informative estimates. Therefore, poverty alleviation policies should consider the nonlinear characteristics of this relationship to be more targeted and effective. Thus, this study not only strengthens the existing literature but also provides new empirical evidence that nonparametric spline regression is an adaptive and accurate method for modeling the complex relationship between socioeconomic indicators and poverty rates in Indonesia, especially in the context of regional heterogeneity and high social dynamics. From a policy perspective, the findings highlight the importance of prioritizing education specifically improving the mean years of schooling as a key lever in reducing poverty. Interventions such as expanding access to basic and secondary education, improving educational quality in underdeveloped regions, and aligning educational programs with local economic needs could significantly enhance poverty reduction efforts.

4. CONCLUSION

Based on the results of the research that has been carried out, the following conclusions were obtained:

1. Based on the results of individual parameter testing, it is shown that the variable of average length of school has a significant effect on the percentage of the provincial poor population in Indonesia.
2. The best model produced in the Spline Nonparametric Regression is with two knot points because it has the smallest GCV value compared to other models.

$$\begin{aligned}\hat{y} = {} & -36{,}294 - 1{,}779 * x_1 + 6{,}522 * (x_1 - 9{,}43)_+^1 - 5{,}831 * (x_1 - 12{,}86)_+^1 - 2{,}392 * x_2 \\ & + 4{,}054 * (x_2 - 4{,}41)_+^1 - 1{,}579 * (x_2 - 5{,}38)_+^1 - 6{,}320 * x_3 + 11{,}372 \\ & * (x_3 - 8{,}91)_+^1 - 6{,}643 * (x_3 - 9{,}69)_+^1 + 1{,}738 * x_4 - 1{,}393 \\ & * (x_4 - 67{,}46)_+^1 - 1{,}285 * (x_4 - 69{,}37)_+^1\end{aligned}$$

3. GDP and open unemployment rate as economic factors do not affect the percentage of poor people

4. The average length of schooling as a social factor has an influence on the percentage of the poor population, but another social factor, namely life expectancy, has no effect on the percentage of the poor population.

ACKNOWLEDGMENTS

The author extends deepest gratitude to STIKOM Bali for the financial support and facilities provided, which greatly contributed to the successful completion of this research.

FUNDING INFORMATION

This research is the output of an internal research grant funded by ITB STIKOM Bali.

AUTHOR CONTRIBUTIONS STATEMENT

First Author: Conceptualization, methodology, writing-original draft, software, validation. Second Author: Conceptualization, methodology, writing-original draft, software, validation. Third Author: Data curation, resources, draft preparation. Fourth Author: Formal analysis, validation. Fifth Author: Validation, writing-review & editing. Sixth Author: Validation, writing-review & editing. All authors discussed the results and contributed to the final manuscript.

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest

DATA AVAILABILITY

The data that support the findings of this study are available from a public source (https://www.bps.go.id/id) and were processed by the author using R software.

REFERENCES

[1] B. P. S. Indonesia, "Persentase Penduduk Miskin Maret 2024 turun menjadi 9,03 persen." Accessed: Dec. 26, 2024. [Online]. Available: https://www.bps.go.id/id/pressrelease/2024/07/01/2370/persentase-penduduk-miskin-maret-2024-turun-menjadi-9-03-persen-.html

[2] M. P. Todaro and S. C. Smith, *Economic Development*, 13th ed. Pearson, 2020.

[3] World Bank, "Indonesia Economic Prospects (IEP): Boosting the Recovery," World Bank. Accessed: Aug. 09, 2025. [Online]. Available: https://www.worldbank.org/en/country/indonesia/publication/indonesia-economic-prospect

[4] A. Suryahadi, V. Febriany, A. Yumna, and F. Azzahro, *The Dynamics of Poverty and Inequality in Indonesia: Trends, Policies, and Future Challenges*. SMERU Research Institute, 2022.

[5] Alkire, S., Kanagaratnam, U., & Nogales, R. (2023). *Multidimensional Poverty Index 2023: Unstacking global poverty – Data for high-impact action*. United Nations Development Programme (UNDP) and OPHI.

[6] OECD, "Economic Outlook for Southeast Asia, China and India 2022: Financing Sustainable Recovery from COVID-19."

[7] W. Sanusi, R. Syam, and R. Adawiyah, "Model regresi nonparametrik dengan pendekatan spline (studi kasus: Berat badan lahir rendah di rumah sakit ibu dan anak siti fatimah makassar)," *J. Math. Comput. Stat.*, vol. 2, no. 1, pp. 70–81, 2019.

[8] M. Q. Alfairus, M. A. Tiro, and M. K. Aidid, "Regresi Nonparametrik Spline Multivariat Untuk Menentukan Faktor Yang Mempengaruhi Kemiskinan di Sulawesi Selatan," *VARIANSI J. Stat. Its Appl. Teach. Res.*, vol. 3, no. 2, Art. no. 2, Oct. 2021, doi: 10.35580/variansiunm23857.

[9] N. P. N. Hendayanti and M. Nurhidayati, "Regresi Logistik Biner dalam Penentuan Ketepatan Klasifikasi Tingkat Kedalaman Kemiskinan Provinsi-Provinsi di Indonesia," *Sainstek J. Sains Dan Teknol.*, vol. 12, no. 2, pp. 63–70, 2020.
[10] L. P. S. Pratiwi, N. P. N. Hendayanti, and I. K. P. Suniantara, "Perbandingan pembobotan seemingly unrelated regression–spatial durbin model untuk faktor kemiskinan dan pengangguran," *J. Varian*, vol. 3, no. 2, pp. 51–64, 2020.
[11] W. Noviani, S. Sifriyani, and I. Purnamasari, "Pemodelan regresi nonparametrik spline linear persentase penduduk miskin di Kalimantan," *J. Siger Mat.*, vol. 1, no. 2, pp. 35–41, 2020.
[12] W. Alwi, T. A. Nurman, and R. Ibnas, "Modeling of poverty level in South Sulawesi Province through spline nonparametric regression approach," *Commun Math Biol Neurosci*, vol. 2023, p. Article-ID, 2023.
[13] A. S. Suriaslan, V. Ratnasari, and I. N. Budiantara, "Nonparametric Regression Estimator of Multivariable Truncated Spline for Categorical Data," *Available SSRN 4902127*, Accessed: Apr. 07, 2025. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4902127.
[14] N. P. Mariati, I. N. Budiantara, and V. Ratnasari, "Smoothing Spline Estimator in Nonparametric Regression (Application: Poverty in Papua Province)," in *7th International Conference on Research, Implementation, and Education of Mathematics and Sciences (ICRIEMS 2020)*, Atlantis Press, 2021, pp. 309–314. Accessed: Apr. 07, 2025. [Online]. Available: https://www.atlantis-press.com/proceedings/icriems-20/125953634.
[15] N. Hasanah, S. Bahri, and N. Fitriyani, "The Effect of Human Development Index on Poverty Model in Indonesia using Penalized Basis Spline Nonparametric Regression.," in *IOP Conference Series: Materials Science and Engineering*, IOP Publishing, 2021, p. 012055. Accessed: Apr. 07, 2025. [Online]. Available: https://iopscience.iop.org/article/10.1088/1757-899X/1115/1/012055/meta.
[16] Sugiono, *Metode Penelitian Kuantitatif*. Bandung: Alfabeta, 2018.
[17] Wang, X., & Yang, L. (2017). Spline estimation of partially linear models with measurement error. *Journal of Multivariate Analysis*, 156, 128–140.
[18] Eilers, P. H. C., Marx, B. D., & Durbán, M. (2015). Twenty years of P-splines. *Statistical Modelling*, 15(4), 279–299.
[19] Wagner, J., Kauermann, G., & Münnich, R. (2021). Matrix-free penalized spline smoothing with multiple covariates. *arXiv preprint arXiv:2101.06034*.
[20] Herrmann, E., Gijbels, I., & Mammen, E. (2018). Local polynomial fitting and bandwidth selection. *Journal of Statistical Planning and Inference, 195*, 61–77.
[21] E. Sinaga, "Penerapan Metode Least Squares Method Dalam Estimasi Penjualan Produk Elektronik," *J. Comput. Inform. Res.*, vol. 2, no. 2, pp. 44–48, 2023.