Modeling of poverty in Kalimantan with nonparametric spline regression approach

A Halida, N Salam, and A S Lestia
Universitas Lambung Mangkurat, South Kalimantan 70714, Indonesia
annisahalidaaa@gmail.com

Abstract. Poverty is a problem that is always present in any country, including Indonesia. Kalimantan is one of the islands in Indonesia that has not been free from poverty. Therefore, more effective efforts to further reduce poverty are needed. The purpose of this study was to develop a model that can explain significant factors to poverty in Kalimantan. To achieve the objective, first, factors that may have significant influence to the changes in the percentage of poor people should be identified, using regression analysis. In this study, nonparametric regression analysis was used with a spline approach since the relationship between poverty and the explanatory factors did not have a particular pattern. The results of this study showed that, the best was using three knot points, where Open Unemployment Rate (OUR), Human Development Index (HDI), and Economic Growth have a significant effect on poverty in Kalimantan.

1. Introduction
Poverty is a problem that is always present in any country, including Indonesia. Kalimantan is an island in Indonesia that has not been free yet from poverty. Based on the Official Statistics News or Berita Resmi Statistik No. 56/07/Th. XXIII by the Indonesian Central Bureau of Statistics or Badan Pusat Statistik Indonesia shows that the percentage of poor people in Kalimantan was 5.81 percent by March 2020 (it was increase up to 0.56% from September 2019) [1]. Considering that fact, more effective efforts are needed to reduce poverty in Kalimantan. The first step that can be done is to identify factors that may have significant contribution to changes in the percentage of poor people using regression analysis.

In this study, nonparametric regression analysis was used because the relationship pattern between poverty and the explanatory factors did not have a specific pattern such as a linear, cubic, quadratic, or other pattern [2]. One of several methods of approach in nonparametric regression that can be used to estimate the function is to use the spline approach. The spline is segmented polynomials (piecewise polynomials) with flexibility properties [3]. Because of this properties, the spline regression model can effectively adapt to data characteristics [4]. In analyzing the relationship between poverty and the factors that influence it used the analysis of nonparametric regression spline. The analysis of nonparametric...
regression spline was chosen because spline approach has advantages in overcoming the data patterns that
tend to rise or fall sharply, and the curve produced is relatively smooth [5].

Previously, nonparametric spline regression was used in a study by Ni Putu Dera Yanthi and I Nyoman Budiantara in 2016 [6] to review the human development index in Central Java. The results showed that using a combination of knot points $3,3,2,1,2$ was the best model. And further research by Tirta Purnaraga, Sifriyani, and Surya Prangga in 2020 [7] has reviewed the rate on economic growth in Kalimantan with a nonparametric spline regression approach. The result showed that the best model is was using three knot points. Based on this, this paper will examine the nonparametric spline regression model that was formed, and to find out the factors that have a significant effect on poverty in Kalimantan.

2. Materials and Methods

2.1 Data Sources
This research has used the data from the publications of Central Bureau of Statistics (BPS) of South Kalimantan, Central Kalimantan, East Kalimantan, West Kalimantan, and North Kalimantan in 2019, with the unit observations were 56 districts/cities.

2.2 Research Variable
The variables of study are given in Table 1, where the dependent variables are Poverty ($Y$) and the independent variables are Upon Unemployment Rate (OUR), Human Development Index (HDI), and Economic Growth.

| Variable                     | Symbol | Unit          | Measuring Scale |
|------------------------------|--------|---------------|-----------------|
| Poverty                      | $Y$    | Percent       | Ratio           |
| Open Unemployment Rate (OUR) | $X_1$  | Percent       | Ratio           |
| Human Development Index (HDI)| $X_2$  | Percent       | Ratio           |
| Economic Growth              | $X_3$  | Percent       | Ratio           |

2.3 Analysis Steps
1) Describing research data descriptively.
2) Created a scatterplot between the dependent and independent variable, to determine the shape of the data pattern formed.
3) Modeling the dependent and independent variables using knot points.
4) Selecting the optimal knot point based on the smallest GCV.
5) Modeling the percentage of poverty in Kalimantan Island using optimal knot points.
6) Performing significance testing on parameters simultaneously and individually.
7) Testing the residual assumptions.
8) Calculating the coefficient of determination ($R^2$).
9) Interpret models and take conclusions.

2.4 Theoretical Review
2.4.1 Descriptive Statistics. Descriptive statistics is a method related to collecting or presenting data to obtain information. In descriptive statistics there are measures that are used as a summary value of the values of existing observations. These measures include the size of the data center and the size of the spread of the data [8].
2.4.2 Regression Analysis. Regression analysis can explain the relationship between two or more variables [9]. In regression, to estimate its function there are three kinds of approaches include regression parametric, nonparametric, and semiparametric [10]. The following is the equation of regression analysis in matrix form as follows.

\[ y = X\beta + \epsilon \]  

(1)

where,

- \( y \) = vector column size \( nx1 \) of the dependent variable
- \( X \) = matrix size \( nx(m + q) \) of the independent variables
- \( \beta \) = estimated value of the regression parameter with size \( (m + q)x1 \)
- \( \epsilon \) = column vector size \( nx1 \) from error

2.4.3 Nonparametric Regression. Nonparametric regression analysis is applied if the relationship between dependent and independent variables do not have a specific form of curve regression [3]. In general, the equation model can be seen as follows.

\[ y_i = f(x_i) + \varepsilon_i \]  

(2)

where,

- \( y_i \) = dependent variable
- \( x_i \) = independent variable
- \( f(x_i) \) = function regression that does not follow a specific pattern
- \( \varepsilon_i \sim \text{IIDN (Identical, Independent, Normally Distribution)} \)

2.4.4 Spline Nonparametric Regression. Spline nonparametric regression is a method of plotting data that takes the smoothness of the curve into account. Splines can adapt effectively to existing data, so that results are close to the truth. This occurs because the spline contains a knot point, which is a joint fusion point that indicates a change in data behavior pattern [3]. If the spline nonparametric function order \( m \) with knot points \( k_1, k_2, k_3, \ldots, k_r \), the equation model can be written as follows.

\[ y_i = \sum_{j=0}^{m} \beta_j x_i^j + \sum_{q=1}^{r} \beta_{m+q} (x_i - k_q)_+^m + \varepsilon_i, \quad i = 1, 2, \ldots, n \]  

(3)

with the function \((x_i - k_q)_+^m\) which is a truncated function, is presented in the following form.

\[ (x_i - k_q)_+^m = \begin{cases} (x_i - k_q)^m, & x_i \geq k_q \\ 0, & x_i < k_q \end{cases} \]  

(4)

2.4.5 Selection of Optimal Knot Point. The method to find the optimal knot points is Generalized Cross Validation (GCV). GCV method will be better used in normally distribution data [11]. The functions of the GCV can be written as follows:

\[ GCV(K) = \frac{\text{MSE}(K)}{\text{trace}(I - A(K))^2} \]  

(5)

with the function of Mean Square Error (MSE) as follows.
\[ MSE(K) = n^{-1} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]  

(6)

and \( \mathbf{A}(\mathbf{K}) \) is obtained from \( \hat{\mathbf{y}} = \mathbf{A}(\mathbf{K})\mathbf{y} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \).

2.4.6 Testing Model Parameters

2.4.6.1 Simultaneous Test

Hypotheses for the simultaneous test are given as follows.

\[ H_0: \beta_1 = \beta_2 = \ldots = \beta_{m+q} = 0 \]
\[ H_1: \text{at least one } \beta_j \neq 0, j = 1,2,\ldots,m+q \]

Mathematically, the formula for the simultaneous test using F test can be written as follows.

\[ F_{\text{count}} = \frac{MS_{\text{regression}}}{MS_{\text{error}}} \]  

(7)

The rejection area of \( H_0 \) is if \( F_{\text{count}} > F_{\alpha;(m+q,n-(m+q)-1)} \) or \( p_{\text{value}} < \alpha \). Then \( H_0 \) is rejected, it can be concluded at least one on the parameters that are significant to the model. The rejection area of \( H_0 \) is if \( F_{\text{count}} > F_{\alpha;(m+q,n-(m+q)-1)} \) or \( p_{\text{value}} < \alpha \) or it can be concluded at least one of the parameters is significant to the model [5].

2.4.6.2 Individual Test (Partial)

The partial test is a test of the significance of each independent variable [12]. Hypotheses for the individual test (partial) are given as follows.

\[ H_0: \beta_j = 0; j = 1,2,\ldots,m+q \]
\[ H_1: \beta_j \neq 0; j = 1,2,\ldots,m+q \]

Mathematically, the formula for the individual test (partial) using the t test can be written as follows.

\[ t_{\text{count}} = \frac{\tilde{\beta}_j}{SE(\tilde{\beta}_j)}; j = 1,2,\ldots,m+q \]  

(8)

The rejection area of \( H_0 \) is if \( t_{\text{count}} > t_{\alpha/2(n-(m+q)-1)} \) or \( p_{\text{value}} < \alpha \) or can be concluded that the individual parameter has a significant effect to the model.

2.4.7 Residual Test

2.4.7.1 Identical Assumption Test of Residuals Variance.

The variance of residuals should be assumed to be homogeneous or no heteroscedasticity [13]. The Glejser test can be used to identify the assumption with the hypotheses as follows.

\[ H_0: \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_n^2 = \sigma^2 \]
\[ H_1: \text{at least one } \sigma_i^2 \neq \sigma^2; i = 1,2,\ldots,n \]

The Glejser test formula can be mathematically written as follows.

\[ F_{\text{count}} = \frac{MS_{\text{regression}}}{MS_{\text{residual}}} = \frac{\sum_{i=1}^{n}(|\hat{e}_i| - \hat{e}_i)^2/((m+q)-1)}{\sum_{i=1}^{n}(|e_i| - |\hat{e}_i|)^2/(n-(m+q))} \]  

(9)
The rejection area of $H_0$ is when $F_{\text{count}} < F_{\text{table}}$ and $p-value > \alpha$, so that it can be concluded that identical assumptions of residuals’ variance is fulfilled.

2.4.7.2 *Independent Assumption Test of Residuals*. Residuals should meet the assumption of independency or there is no a relationship between residuals [13]. This can be done by using plot the Autocorrelation Function (ACF).

2.4.7.3 *Normality Assumption Test of Residuals’ Distribution*. The residuals should be assumed to be normally distributed. This can be done by using Kolmogorov-Smirnov test [3]. The hypotheses are given as follows.

$H_0 : F(\varepsilon) = F_0(\varepsilon)$ (Residual is normally distributed)

$H_1 : F(\varepsilon) \neq F_0(\varepsilon)$ (Residual not normally distributed)

where,

$F(\varepsilon)$ = cumulative distribution function based on sample data

$F_0(\varepsilon)$ = cumulative frequency distribution function

The Kolmogorov-Smirnov test can be mathematically written as follows.

$$D = \max|S_N(\varepsilon) - F_0(\varepsilon)|$$

where,

$S_N(\varepsilon)$ = the observed cumulative probability function of a random sample

$H_0$ will be rejected if $|D| > D_\alpha$, where $D_\alpha$ is the critical value of the Kolmogorov-Smirnov table or $p-value < \alpha$.

2.4.8 *Model Goodness Criteria*. The coefficient of determination is the value of the proportion of a total variance that can be explained by the model [9]. The formula of $R^2$ as follows.

$$R^2 = \frac{SSR}{SST} \times 100\%$$

where,

$SSR$ = Sum of Square Regression

$SST$ = Sum of Square Total

2.4.9 *Poverty*. Poverty is often associated with low quality of life, which is defined as a condition of life characterized by a lack of basic needs. These basic needs are goods and services that are needed by everyone [14]. Poverty can be associated with material, social, and emotional deprivation. Poverty is a complex problem and is partly the result of political processes and policy development [15].

3. Results and Discussion

| Variable | Mean | Variance | Minimum | Maximum |
|----------|------|----------|---------|---------|
| $Y$      | 17.396 | 161.9571 | 1.340   | 56.340  |
| $X_1$    | 4.355  | 2.717716 | 1.740   | 9.190   |
| $X_2$    | 70.41  | 19.41239 | 62.66   | 80.77   |
| $X_3$    | 5.196  | 2.368759 | 2.180   | 7.970   |
Based on Table 2, it is known that the mean percentage of Poverty in Kalimantan Island in 2019 was 17.40% with a variance of 161.96%. The highest percentage of poor people was in Kutai Kartanegara (56.34%), and the lowest was in Tana Tidung (1.34%). The mean of Open Unemployment Rate (OUR) in 2019 was 4.35% with highest is in Bontang (9.19%), while the lowest was in Pulang Pisau (1.74%). The mean percentage of Human Development Index (HDI) in 2019 was 70.41% with the highest HDI was in Palangkaraya, Central Kalimantan Province (80.77%), while the lowest was in North Kayong Regency, West Kalimantan Province (62.66%). Finally, the mean percentage of Economic Growth in 2019 was 5.20% with the highest rate was in East Kutai Regency (7.97%) and the lowest was in Bontang (2.18%).

3.1 Analysis of Relationship Patterns
Figure 1 shows the pattern that exists between Poverty and the factors (OUR, HDI, and Economic Growth) in Kalimantan.

![Figure 1. Scatterplot Poverty and the three explanatory variables](image-url)
Figure 1 has shown scatterplots to see the pattern of the relationship between Poverty and OUR (Figure 1a), Poverty and HDI (Figure 1b), and Poverty and Economic Growth (Figure 1c), it is found that the three patterns of the relationship do not follow a certain pattern, such as a linear, cubic, quadratic, or other patterns so that, nonparametric regression analysis was used to analyse the data.

3.2 Selection of Optimal Knot Points

3.2.1 Selection of Optimal Knot Point with 1 Knot Point

| Table 2. GCV Value Using 1 Knot Point |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| $X_1$           | $X_2$           | $X_3$           | GCV             |
| 2.500204        | 64.50796        | 2.770816        | 185.8758        |
| 2.652245        | 64.87755        | 2.888980        | 185.6649        |
| 2.804286        | 65.24714        | 3.007143        | 185.8850        |
| 2.956327        | 65.61673        | 3.125306        | 186.2633        |
| 3.108367        | 65.98633        | 3.243469        | 186.6881        |
| 3.564490        | 67.09510        | 3.597959        | 186.5098        |
| 3.716531        | 67.46469        | 3.716122        | 183.8462        |
| **3.868571**    | **67.83429**    | **3.834286**    | **182.2519**    |
| 4.020612        | 68.20388        | 3.952449        | 183.4639        |
| 4.172653        | 68.57347        | 4.070612        | 185.4871        |

Based on Table 3, it is known that the optimal knot point of the OUR ($X_1$) is at 3.87, the HDI ($X_2$) is at 67.84, and Economic Growth ($X_3$) is at 3.83.

3.2.2 Selection of Optimal Knot Point with 2 Knot Points

| Table 3. GCV Value Using 2 Knot Point |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| $X_1$           | $X_2$           | $X_3$           | GCV             |
| 3.412449        | 66.72551        | 3.479796        | 153.7702        |
| 3.868571        | 67.83429        | 3.834286        | 155.3410        |
| 3.412449        | 66.72551        | 3.479796        | 155.3410        |
| 4.020612        | 68.20388        | 3.952449        | 183.4639        |
| **3.564490**    | **67.0951**     | **3.597959**    | **138.8858**    |
| **3.716531**    | **67.46469**    | **3.716122**    | **138.8858**    |
| 3.564490        | 67.0951        | 3.597959        | 139.1843        |
| 3.868571        | 67.83429        | 3.834286        | 149.0474        |
| 3.564490        | 67.0951        | 3.597959        | 149.0474        |
| 4.020612        | 68.20388        | 3.952449        | 183.4639        |
| 3.716531        | 67.46469        | 3.716122        | 149.4456        |
| 3.868571        | 67.83429        | 3.834286        | 158.0423        |
| 5.388980        | 71.5302         | 5.015918        | 158.0423        |
| 5.997143        | 73.00857        | 5.488571        | 158.0715        |
| 5.541020        | 71.8998         | 5.134082        | 155.7363        |
| 5.693061        | 72.26939        | 5.252245        | 157.996         |
| 5.541020        | 71.8998         | 5.134082        | 155.7363        |
| 5.845102        | 72.63898        | 5.370408        | 157.996         |
| 5.541020        | 71.8998         | 5.134082        | 155.7363        |
| 5.997143        | 73.00857        | 5.488571        | 157.996         |

Based on Table 4, it is known that the two optimal knot points for UOR ($X_1$) are at 3.56 and 3.72, the HDI ($X_2$) are at 67.10 and 67.46, and the Economic Growth ($X_3$) are at 3.60 and 3.72.
3.2.3 Selection of Optimal Knot Point with 3 Knot Points

Table 4. GCV Value Using 3 Knot Point

|        | X_1    | X_2    | X_3    | GCV       |
|--------|--------|--------|--------|-----------|
| 3.564490 | 67.09510 | 3.597959 |
| 3.868571 | 67.83429 | 3.834286 | 122.1991 |
| 4.020612 | 68.20388 | 3.952449 |
| 3.564490 | 67.09510 | 3.597959 |
| 3.868571 | 67.83429 | 3.834286 | 126.3655 |
| 4.172653 | 68.57347 | 4.070612 |
| 3.564490 | 67.09510 | 3.597959 |
| 3.868571 | 67.83429 | 3.834286 | 127.5801 |
| 4.324694 | 68.94306 | 4.188776 |
| 3.564490 | 67.09510 | 3.597959 |
| 3.868571 | 67.83429 | 3.834286 | 130.0301 |
| 4.476735 | 69.31265 | 4.306939 |
| 3.564490 | 67.09510 | 3.597959 |
| 3.868571 | 67.83429 | 3.834286 | 130.2926 |
| 5.845102 | 72.63898 | 5.370408 |
| 3.564490 | 67.09510 | 3.597959 |
| 3.868571 | 67.83429 | 3.834286 | 127.5619 |
| 5.997143 | 73.00857 | 5.488571 |
| 3.564490 | 67.09510 | 3.597959 |
| 3.868571 | 67.83429 | 3.834286 | 127.3463 |
| 6.149184 | 73.37816 | 5.606735 |
| 3.564490 | 67.09510 | 3.597959 |
| 3.868571 | 67.83429 | 3.834286 | 127.4261 |
| 6.301224 | 73.74776 | 5.724898 |
| 3.564490 | 67.09510 | 3.597959 |
| 3.868571 | 67.83429 | 3.834286 | 129.1877 |
| 6.453265 | 74.11735 | 5.843061 |
| 3.564490 | 67.09510 | 3.597959 |
| 3.868571 | 67.83429 | 3.834286 | 130.2367 |
| 6.605306 | 74.48694 | 5.961224 |

Based on Table 5, it is known that the three optimal knot points for OUR (X_1) are at 3.56, 3.87 and 4.02, the HDI (X_2) are at 67.10, 67.83 and 68.20, and the Economic Growth (X_3) are at 3.60, 3.83 and 3.95.

3.3 Selection of the Best Knot Point

The next step is to compare the GCV values at each knot point in order to get the best model.

Table 5. GCV Value on Each Knot Point

| Number of Knot Point | GCV Minimum |
|----------------------|-------------|
| 1 Knot Point         | 182.2519    |
| 2 Knot Points        | 138.8858    |
| 3 Knot Points        | 122.1991    |
Based on Table 6, minimum GCV value is at three knot points. It can be concluded that the best model is formed using the three knot points.

3.4 Estimation of Model Parameters

The following is the estimation result of the nonparametric spline regression model parameters using three knot points.

\[
\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2(x_1 - 3.564490) + \hat{\beta}_3(x_1 - 3.868571) + \hat{\beta}_4(x_1 - 4.020612) + \\
\hat{\beta}_5x_2 + \hat{\beta}_6(x_2 - 67.09510) + \hat{\beta}_7(x_2 - 67.83429) + \hat{\beta}_8(x_2 - 68.20388) + \\
\hat{\beta}_9x_3 + \hat{\beta}_{10}(x_3 - 3.597959) + \hat{\beta}_{11}(x_3 - 3.834286) + \hat{\beta}_{12}(x_3 - 3.952449)
\]

3.5 Testing Model Parameters

3.5.1 Simultaneous Test

Table 6. Analysis of Variance

| Source         | Degrees Of Freedom (df) | Sum of Square (SS) | Mean Square (MS) | F_count | P – value |
|----------------|-------------------------|--------------------|------------------|---------|-----------|
| Regression     | 13                      | 3657.463           | 304.7886        | 2.496278 | 0.0139273 |
| Error          | 42                      | 5250.18            | 122.0972        |         |           |
| Total          | 55                      | 8907.643           |                  |         |           |

Using a significance level of \( \alpha = 0.05 \), it is obtained that \( F_{table} = 1.961218 \). Since the value of \( F_{count} > F_{table} \) and \( p – value < \alpha = 0.05 \), \( H_0 \) is rejected. It is concluded that there is at least one of the model parameters that has a significant effect to the model.

3.5.2 Individual Test (Partial)

Table 8 shows the results of individual (partial) parameters test.

Table 7. Partial Parameter Test

| Variable | Parameter | Estimasion Parameter | t_count | p – value | Description |
|----------|-----------|----------------------|---------|-----------|-------------|
| \( X_1 \) | \( \beta_0 \) | -55.84158            | -3.1412775 | 0.003041337 | Significant |
|          | \( \beta_1 \) | -26.52709            | -2.7863693 | 0.007900133 | Significant |
|          | \( \beta_2 \) | 29.31449             | 3.1923262  | 0.002638826 | Significant |
|          | \( \beta_3 \) | -20.30880            | -0.4356061 | 0.6653015  | Not Significant |
|          | \( \beta_4 \) | 21.05094             | 0.4900334  | 0.6266013  | Not Significant |
| \( X_2 \) | \( \beta_5 \) | -131.15276           | -3.1446814 | 0.003012788 | Significant |
|          | \( \beta_6 \) | -77.81049            | -4.2410945 | 0.000115902 | Significant |
|          | \( \beta_7 \) | 34.91459             | 4.2577594  | 0.000110032 | Significant |
|          | \( \beta_8 \) | 69.76296             | 3.1348015  | 0.003096359 | Significant |
| \( X_3 \) | \( \beta_9 \) | 15.13617             | 2.2483227  | 0.02973567 | Significant |
|          | \( \beta_{10} \) | 34.91459            | 4.2577594  | 0.000110032 | Significant |
|          | \( \beta_{11} \) | 69.76296            | 3.1348015  | 0.003096359 | Significant |
|          | \( \beta_{12} \) | -16.62175            | -2.2641375 | 0.02866542 | Significant |
With $\alpha=0.05$ obtained results from 12 parameters, there are 2 parameters that are not significant, the parameters $\hat{\beta}_3$ and $\hat{\beta}_4$ on variable $X_1$. In individual parameter testing, if there is only one parameter in the independent variable that is significant, then the independent variable can be said to be significant even though the other parameters contained in the variable are not significant.

3.6 Residual Test

3.6.1 Identical Assumption Test of Residuals’ Variance

Table 8. Glejser Test Results

| Source       | df | Sum of Square | Mean Square | F count | $p – value$ |
|--------------|----|---------------|-------------|---------|-------------|
| Regression   | 13 | 895.311       | 74.60925    | 1.128175| 0.3637035   |
| Error        | 42 | 2843.706      | 66.1327     |         |             |
| Total        | 55 | 3739.017      |             |         |             |

With $\alpha = 0.05$ obtained $F_{0.005(3.52)}$ or $F_{table}$ of 1.961218. Since $F_{count} < F_{table}$ and $p – value > \alpha = 0.05$, $H_0$ is accepted, it can be concluded that the assumption of identical residuals’ variance has been met.

3.6.2 Independent Assumption Test of Residuals

Figure 2 shows that there is no autocorrelation since there is no out of the limit of significance. Therefore, it can be concluded that the assumption of independent residuals has been satisfied.

3.6.3 Normality Assumption Test of Residuals’ Distribution

Table 9. Kolmogorov-Smirnov Test

| Kolmogorov-Smirnov | $p - value$ |
|--------------------|-------------|
| 1.182              | 0.122       |

With $\alpha = 0.05$, it is known that $p – value > \alpha$, so that $H_0$ can be accepted. Thus, the assumption of normally distributed residuals has been satisfied.

3.7 Coefficient of Determination

$$R^2 = \frac{SS_{regression}}{SS_{total}} \times 100\%$$
Based on the calculation, it is obtained that $R^2$ value is 41.06%. This means that the model obtained is able to explain the variability of the Poverty percentage of 41.06%. Meanwhile, the rest is explained by other variables that are not included in the model.

3.8 Interpretation Model

The following is the model obtained.

$$\hat{y} = -55.84158 - 26.52709x_1 + 29.31449(x_1 - 3.564490) \cdot 10.30880(x_1 - 3.868571) \cdot + 34.91459(x_2 - 67.83429) \cdot + 69.76296(x_2 - 68.20388) \cdot + 15.13617x_3 \cdot + 34.91459(x_3 - 3.597959) \cdot + 69.76296(x_3 - 3.952449) \cdot$$

From the above model, the following interpretation is obtained.

1. If the variables $X_2$ and $X_3$ are constant, then the effect of the Open Unemployment Rate (OUR) on the percentage of Poverty in Kalimantan is as follows.

$$\hat{y} = -26.52709x_1 + 29.31449(x_1 - 3.564490) \cdot - 20.30880(x_1 - 3.868571) \cdot + 21.05094(x_1 - 4.020612) \cdot$$

$$\begin{align*}
   &-26.52709x_1, & &x_1 < 3.564490 \\
   &2.78740x_1 + 104.49120, & &3.564490 \leq x_1 \leq 3.868571 \\
   &-17.5214x_1 + 25.92518, & &3.868571 \leq x_1 \leq 4.020612 \\
   &3.52954x_1 - 110.56279, & &x_1 \geq 4.020612
\end{align*}$$

Regional groupings for the OUR in Kalimantan are also presented visually in Figure 3.

![Figure 3. Open Unemployment Rate (OUR) by Interval from Knot Point](image)

2. If the variables $X_1$ and $X_3$ are constant, then the influence of the Human Development Index (HDI) on the percentage of Poverty in Kalimantan is as follows.

$$\hat{y} = -131.15276x_2 - 77.81049(x_2 - 67.09510) \cdot + 34.91459(x_2 - 67.83429) \cdot + 69.76296(x_2 - 68.20388) \cdot$$
The regional groupings for the HDI in Kalimantan are also presented visually in Figure 4.

\[
\begin{align*}
\text{If } x_3 < 3.597959 & : & -131.15276x_2^2 + 208.96325x_2 + 5220.70260 ; \\
67.09510 \leq x_3 < 67.83429 & : & -208.96325x_2^2 + 5220.70260 ; \\
67.83429 \leq x_3 < 68.20388 & : & -174.04866x_2^2 + 2852.29618 ; \\
x_3 \geq 68.20388 & : & -104.2857x_2^2 + 1905.80836 ;
\end{align*}
\]

The regional groupings for Economic Growth in Kalimantan is also presented visually Figure 5.

\[
\begin{align*}
\text{If } x_3 < 3.597959 & : & -131.15276x_2^2 + 208.96325x_2 + 5220.70260 ; \\
67.09510 \leq x_3 < 67.83429 & : & -208.96325x_2^2 + 5220.70260 ; \\
67.83429 \leq x_3 < 68.20388 & : & -174.04866x_2^2 + 2852.29618 ; \\
x_3 \geq 68.20388 & : & -104.2857x_2^2 + 1905.80836 ;
\end{align*}
\]

3. If the variables \( X_1 \) and \( X_2 \) are constant, then the effect of Economic Growth on the percentage of Poverty in Kalimantan is as follows.

\[
\hat{y} = \begin{cases} 
15.13617x_3 + 34.91459(x_3 - 3.597959) & ; \\
15.13617x_3 & ; \\
50.05076x_3 - 125.62126 & ; \\
119.81327x_3 - 393.11240 & ; \\
103.191977x_3 - 327.41578 & ; \\
\end{cases} \]

\[
\begin{align*}
\text{If } x_3 < 3.597959 & : & 69.76296(x_3 - 3.834286) + 16.62175(x_3 - 3.952449) ; \\
x_3 \geq 3.952449 & : & 69.76296(x_3 - 3.834286) + 16.62175(x_3 - 3.952449) ; \\
x_3 \geq 3.952449 & : & 69.76296(x_3 - 3.834286) + 16.62175(x_3 - 3.952449) ;
\end{align*}
\]

The regional grouping for Economic Growth in Kalimantan is also presented visually Figure 5.
4. Conclusion

The highest poverty percentage in Kalimantan in 2019 was in Kutai Kartanegara, while the lowest was in Tana Tidung. Based on the results of spline regression analysis, the best model was obtained using three knot points. Open Unemployment Rate (OUR), Human Development Index (HDI), and Economic Growth were the most significant that contribute to or can explain 41% the variation in changes of poverty’s percentage in Kalimantan. And for the government, both within the regency/city and provincial level in Kalimantan, to reduce the percentage of the poor population is to pay attention to the unemployment rate, and to increase the human development index and economic growth.

References

[1] Statistik B P 2020 Berita Resmi Statistik Profil Kemiskinan di Indonesia Maret 2020 No. 56/07/Th. XXIII

[2] Lestari B, Fatmawati, Budiantara I N and Chamidah N 2018 J. Phys. Conf. Ser. 1097

[3] Eubank R L 1999 Nonparametric Regression and Spline Smoothing (New York: Marcel Dekker Inc)

[4] Massaid A, Hanif M, Febrianty D and Chamidah N 2019 IOP Conf. Ser. Mater. Sci. Eng. 546

[5] Rinaldo Fernandes A A, Darmanto, Astuti A B, Solimun, Amalia L, Nurjannah, Yanti I, Arisoesilaningsih E and Isaskar R 2019 IOP Conf. Ser. Earth Environ. Sci. 239

[6] Ni Putu Dera Yanthi I N B 2016 J. Sains dan Seni ITS 5 271–6

[7] Purnaraga T, Sifriyani S and Prangga S 2020 Barekeng J. Ilmu Mat. dan Terap. 14 343–56

[8] Walpole R E 1982 Introduction to Statistics (USA: Macmillan Pub Co)

[9] Norman R. Draper H S 1992 Applied Regression Analysis Third Edition (New York: John Wiley & Sons, Inc)

[10] Rumlawang F Y, Aulele S N and Kasim N 2018 Barekeng J. Ilmu Mat. Dan Terap. 12 27

[11] Wahba G 1990 Spline Models for Observational Data (Pennsylvania: The Society for Industrial and Applied Mathematics)

[12] Mubarokah L 2016 J. Sains dan Seni ITS 5 96

[13] Gujarati D N 2004 Basic Econometric 4th Edition (New York: McGraw Hill Companies. Inc)

[14] Cica Sartika, M.Yani Balaka W A R 2016 J. Ekon. 1 106–18

[15] Alcock P 1997 Understanding Poverty Second Edition (London: Palgrave)