p-ISSN 2338 – 3216
e-ISSN 2528 – 1070

JURNAL STATISTIKA
UNIVERSITAS MUHAMMADIYAH SEMARANG
12(2) 2024: 29-41

MODELING OF FACTORS INFLUENCING GENDER DEVELOPMENT INDEX (GDI) IN PAPUA PROVINCE USING SPLINE NONPARAMETRIC REGRESSION

Sintah[1], Ferry Kondo Lembang[2], Norisca Lewaherilla[3*]
[1,2,3]Statistics Study Program, Department of Mathematics,
Faculty of Mathematics and Natural Science, Pattimura University, Indonesia
***e-mail**: *lewaherillanorisca@gmail.com*

Article Info:
Received: July 7th, 2024
Accepted: November 27th, 2024
Available Online: December 31th, 2024

Keywords:
GCV, Gender Development Index (GDI), Knots, Spline Nonparametric Regression

Abstract: *The Gender Development Index (GDI) is an index of achieving basic human development capabilities to measure success in efforts to develop the quality of human life by considering gender inequality. Papua Province is the province with the lowest GDI score when compared to the 34 provinces in Indonesia. This condition shows that there is still a development gap between the male and female genders. For this reason, it is necessary to research the factors that are suspected to influence GDI in Papua Province. In this study, the pattern of GDI data and the factors that are thought to influence it do not form a specific pattern, so we used spline nonparametric regression. Based on this study, the best model is obtained by the optimal knot point based on the smallest Generalized Cross Validation (GCV) value, which are 3-knots and six significant variables, namely, Life Expectancy* (X_1)*, Expected Years of Schooling* (X_2)*, Female Income Contribution* (X_3)*, Sex Ratio* (X_4)*, Female Labor Force Participation Rate* (X_5)*, High School Enrollment Rate*(X_6)*. This model has an* R^2 *of 99.95%. The predictor variable used has an effect of 99.95%, and other variables influence the rest.*

1. INTRODUCTION

Gender issues are a concern in human development that has long been a concern of the world. So, the global development agenda entitled Sustainable Development Goals (SDGs) was a commitment to encourage gender equality and empowerment of women [1]. In general, gender is defined as differences in roles, positions, responsibilities, and division of labor between men and women determined by society based on the nature of women and men who are considered appropriate according to the norms, customs, beliefs, or habits of society [2]. Gender-oriented development prioritizes the concept of justice without any gender differences, which aims to realize a harmonious partnership between men and women and to realize gender equality and justice in various fields of life and development [3].

The level of development success that brings gender issues can be measured by the Gender Development Index (GDI). That index that measures the achievement of human development in health, education, and a region's economy, considering gender equality between men and women. The gap between men and women in Indonesia still occurs frequently, as evidenced by the many cases of violence against women.

Based on the data from Komisi Nasional Perempuan, the number of cases of violence, especially against the female gender, is increasing every year. This data was evidenced by the

[DOI: 10.14710/JSUNIMUS.12.2.2024.29-41]

number of cases recorded reaching 338,496 in 2021, which has increased compared to 226,062 cases in 2020 [4]. It shows the conditions of gender inequality and discrimination against women are still terrible, so it is one of the severe problems that occur in Indonesia. Therefore, to realize gender equality, the government established regulations concerning the National Development Program in UU No. 25 of 2000. It is a strategy to integrate gender to reduce the gap between Indonesian women and men in accessing and obtaining development benefits and increasing participation and control over the development process [5].

The previous research about the GDI for Central Java Province in 2021 used the Geographically Weighted Regression model [6]. Central Java Province ranked 10th nationally with a GDI value of 92.48%, or the gender gap was relatively low. They found that the significant variables are the average length of schooling, school participation rate, and gender ratio. The regression method that can be used to analyze gender equality issues. It has three approaches: parametric regression, nonparametric regression, and semiparametric regression [7]. In 2023, spline nonparametric regression was used to analyze the Provincial Happiness Index [8]. The results showed that the best Spline Nonparametric Regression model for the Provincial Happiness Index in Indonesia was a model with a combination of 2,1,3,2,3 knot points, which had a minimum Generalized Cross Validation (GCV) value of 2.38, with four predictor variables that had a significant effect between others are Literacy Rate, Open Unemployment Rate, Clean and Healthy Life Behavior, and Percentage of Poor Population. The other research can be seen in [9].

Based on data published by the Central Statistics Agency (Badan Pusat Statistik/BPS), the lowest GDI value in Indonesia in 2021 is located in Asmat Regency, Papua Province. Papua Province is one of the largest provinces in Indonesia, with an area of 319,036.05 km2, equivalent to 16.64% of the total area of the state of Indonesia. The projected population as of 2020 is 3,435,430 people, and the male population in Papua Province is higher than the female population. The male population is 1,802,213, while the female population is 1,633,217. On the other hand, gender equality can be realized if there is no gap between the Human Development Index (HDI) of men and women.

The HDI of males and females in Papua Province always show differences every year. The HDI for men in Papua Province in 2021 was 66.07, while for women, it was only 52.96, or there was a difference of 13.11. Based on the data on the HDI of Papua Province in 2021 and the factors that are thought to influence it, and the scatterplot does not follow a specific pattern. This HDI shows that there is still a gap between men and women in Papua Province. So, spline nonparametric regression can be used since it has the advantage of being a polynomial cut with segmented properties. This segmented nature has more flexibility than ordinary polynomials, making it possible to adjust more effectively to the local characteristics of a data function [10]. Therefore, it is necessary to research the factors that influence GDI to determine the factors that cause the low value of GDI in Papua Province by using spline nonparametric regression.

2. LITERATURE REVIEW

2.1 Gender Development Index (GDI)

The measurement of inter-gender inequality in HDI achievement uses the ratio of female HDI to male HDI to show the achievement of human development between women and men. The following formula is used to measure the value of GDI.

[DOI: 10.14710/JSUNIMUS.12.2.2024.29-41]

$$GDI = \frac{HDI_{women}}{HDI_{men}} \tag{1}$$

2.2. Spline Nonparametric Regression

Splines are polynomial cuts with segmented properties, so spline models have high flexibility and an excellent ability to handle data whose behavior changes at certain sub-intervals, making it possible to adjust more effectively to the local characteristics of a function or data [11]

Suppose there are data $(x_{1i}, x_{2i}, \dots, x_{pi})$and the relationship between $(x_{1i}, x_{2i}, \dots, x_{pi})$ and y_i is approximated by a nonparametric regression model, $y_i = f(x_{1i}, x_{2i}, \dots, x_{pi}) + \varepsilon_i$, with $i = 1, 2, 3, \dots, n$; where, y_i is the response variable, and f is a regression curve of unknown shape. Suppose the regression curve f is an additive model and is approximated by a spline function. In that case, the following regression model is obtained:

$$\begin{aligned} y_i &= \sum_{j=1}^{p} f(x_{ji}) + \varepsilon_i; \;\; i = 1, 2, 3, \dots, n \\ &= f(x_{1i}) + f(x_{2i}) + \dots + f(x_{pi}) + \varepsilon_i \end{aligned} \tag{2}$$

where:

$$f(x_{1i}) = \sum_{h=0}^{q} \beta_{hj} x_{ji}^{h} + \sum_{l=1}^{m} \beta_{(q+1)j} (x_{ji} - k_{lj})_{+}^{q} \tag{3}$$

with

$$(x_{ji} - k_{lj})_{+}^{q} = \begin{cases} (x - k_{lj})^q & , x_{ji} \geq k_{lj} \\ 0 & , x_{ji} < k_{lj} \end{cases}$$

and $k_{1j}, k_{2j}, \cdots, k_{mj}$ are knot points showing the function's behavioral change pattern at different sub-intervals. While the value of q is the degree of the polynomial. Equation (2) can be written as follow.

$$\begin{aligned} y_i = {} & \beta_{01} + \beta_{11} x_{1i} + \dots + \beta_{q1} x_{1i}^{q} + \alpha_{11}(x_{1i} - k_{11})_{+}^{q} + \dots + \alpha_{m1}(x_{1i} - \\ & k_{m1})_{+}^{q} + \beta_{02} + \beta_{12} x_{2i} + \dots + \beta_{q2} x_{2i}^{q} + \alpha_{12}(x_{2i} - k_{12})_{+}^{q} + \dots + \alpha_{m2}(x_{2i} - \\ & k_{m2})_{+}^{q} + \dots + \beta_{0p} + \beta_{1p^x} x_{pi} + \dots + \beta_{qp} x_{pi}^{q} + \alpha_{1p}(x_{pi} - k_{1p})_{+}^{q} + \dots + \\ & \alpha_{mp}(x_{pi} - k_{mp})_{+}^{q} + \end{aligned} \tag{4}$$

One of the methods used for optimal knot point selection is Generalized Cross Validation (GCV). The GCV method has asymptotic optimal properties compared to other methods, such as Cross Validation (CV) [12]. The best spline model with optimal knot points is obtained from the smallest GCV value [13]. According to [11], the GCV function can be written as follows:

$$GCV(k_1, k_2, \dots, k_j) = \frac{MSE(k_1, k_2, \dots, k_j)}{\left(n^{-1} trace[I - A](k_1, k_2, \dots, k_j)\right)^2} \tag{5}$$

where:

I : identity matrix

[DOI: 10.14710/JSUNIMUS.12.2.2024.29-41]

n : number of observations

k : point knot

then,

$A(k_1, k_2, \dots, k_j) = X(X^T X)^{-1} X^T$

and

$$MSE(k_1, k_2, \dots, k_j) = n^{-1} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \quad (6)$$

with $\hat{y} = A(k_1, k_2, \dots, k_j)y$.

By using the least square method to estimate the parameter β in the spline nonparametric regression model by solving optimization.

$$Min(\vec{\varepsilon}^T \vec{\varepsilon}) = Min\left\{\left(\vec{Y} - X[K_1, K_2, \dots, K_k]\vec{\beta}\right)^T \left(\vec{Y} - X[K_1, K_2, \dots, K_k]\vec{\beta}\right)\right\} \quad (7)$$

where:

$\vec{Y} = (y_1, y_2, \dots, y_n)^T$; $\vec{\beta} = (\beta_0, \beta_1, \beta_2, \dots, \beta_p, \beta_1, \beta_2, \dots, \beta_r)$; $\vec{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$;

and $X[K_1, K_2, \dots, K_k]$ is the matrix X that depends on the knot points.

3. METHODOLOGY

In this research, the data was obtained from the publication data of the BPS of Papua Province in 2021. Meanwhile, the observation units used in this study are 29 districts in Papua Province. The variables used in this study are secondary data obtained from previous research and theories related to gender development. The variables used can be seen in Table 1.

Table 1. Research Variable

Symbol	Respond Variable	Unit
Y	GDI	Percent
Symbol	Predictor Variable	Unit
X_1	Life Expectancy Year	Year
X_2	Expected Years of Schooling	Year
X_3	Female Income Contribution	Percent
X_4	Sex Ratio	Percent
X_5	Female Labor Force Participation Rate	Percent
X_6	High School Participation Rate	Percent

The analysis steps carried out in this study refer to the research.

1) Prepare data on HDI and variables that are thought to affect GDI.
2) Make the descriptive statistics of each variable to determine the characteristics of each district in Papua Province.
3) Make a scatter plot between the GDI (Y) of Papua Province and the independent variables (X) that are thought to have an effect to determine the shape of the data pattern.
4) Modeling the data with a spline approach of 1, 2, and 3 knots.
5) Selecting the optimal knot point based on the minimum GCV value.
6) Obtain a spline regression model with optimal knot points.

[DOI: 10.14710/JSUNIMUS.12.2.2024.29-41]

7) Testing the significance of spline regression parameters simultaneously.
8) Testing the spline regression parameters partially.
9) Testing the residual assumption.
10) Calculating the coefficient of determination (R^2).
11) Interpreting the model and drawing conclusions.

4. RESULTS AND DISCUSSION

4.1. Characteristics of the GDI in Papua Province

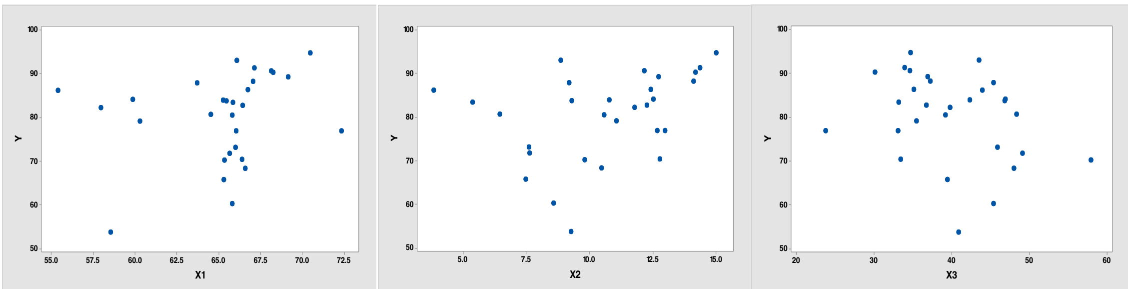
The descriptive statistics of the respond and preditor variables are shown in Table 2.

Table 2. Descriptive Statistics

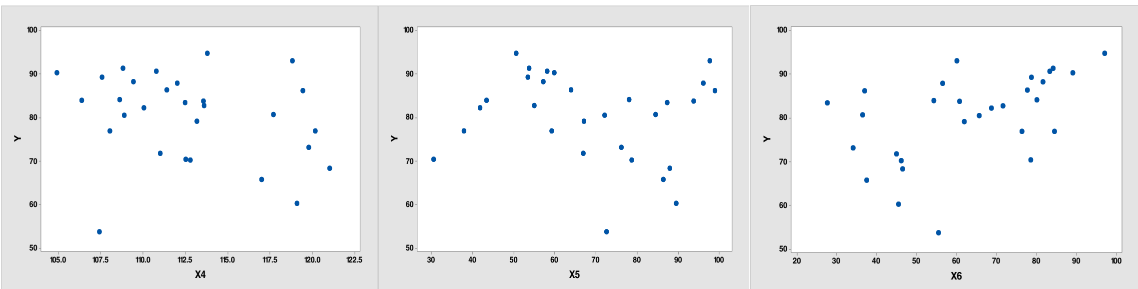
Variable	Mean	Deviation Standard	Minimum	Maximum
Y	80.10	10.05	53.72	94.65
X_1	65.26	3.67	55.43	72.36
X_2	10.58	2.82	3.87	15.02
X_3	40.09	7.12	23.84	57.93
X_4	112.80	4.61	104.95	121.06
X_5	69.03	19.14	30.71	99.00
X_6	62.93	19.17	27.71	97.12

4.2. Analysis of Factors Suspected of Affecting HDI in Papua Province

In this study, the relationship pattern between GPA as the response variable and each predictor variable to visualized by using a scatterplot.



(a) (b) (c)



(d) (e) (f)

Fig. 1 Scatterplot of (a) Life Expectancy Year; (b) Expected Years of Schooling; (c) Female Income Contribution; (d) Sex Ratio; (e) Female Labor Force Participation Rate; and (f) High School Enrollment Rate

[DOI: 10.14710/JSUNIMUS.12.2.2024.29-41]

Based on Figures 1, the scatterplot pattern between HDI and the six variables used does not form a certain pattern, so the model estimation can use a spline nonparametric regression approach.

4.3. Selection of Optimal Knot Point and Minimum GCV

Table 3 shows that the modeling that produces the minimum GCV value is spline nonparametric regression modeling using 3-knot points.

Table 3 Comparisson of the Minimum GCV

Model	Minimum GCV
1-Knot point	99.409
2-Knot point	62.955
3-Knot point	**2.462**

4.4. Parameter Significance Testing

a) Simultaneous Test

This test is carried out on the regression model parameters on the response variable simultaneously to see the significance of the model's parameters as a whole by involving all predictor variables. Simultaneous parameter testing is done using the F test. The hypothesis used is as follows.

$H_0: \beta_1 = \beta_2 = \cdots = \beta_{24} = 0$

H_0: at least one significant parameter $\beta_1 \neq 0, q = 1,2, \dots, 24$

Table 4. Result of the Simultaneous Test (ANOVA)

	Degree of Freedom (df)	Sum Square (SS)	Mean Square (MS)	F_{count}	F_{table}
Regression	24	2828.017	117.834	347.063	5.774
Error	4	1.358	0.340		
Total	28	2829.376			

Based on Table 4, the decision to reject H0 is obtained because $F_{count} = 347.063 > F_{(0,05;24;4)} = 5.774$. It can be concluded that simultaneously, all predictor variables affect the HDI, or there is at least one significant parameter in the spline regression model.

b) Individual test

To determine which parameters significantly affect the regression model, partial testing or individual tests are carried out. The individual test is carried out after the simultaneous test results in a decision to reject H_0. Individual parameter testing is carried out using the t-test. Meanwhile, the hypothesis used is as follows.

$H_0: \beta_q = 0$

$H_1: \beta_q \neq 0, q = 1,2, \dots ,24$

Table 5 Result of Individual Test

Variable	Parameter	Estimator	$\lvert t_{count} \rvert$	*p-value*	Decision
X_1	β_0	1191.433	15.58	0.000	Significant
	β_1	-27.498	-15.67	0.000	Significant
	β_2	27.353	12.876	0.000	Significant

[DOI: 10.14710/JSUNIMUS.12.2.2024.29-41]

Variable	Parameter	Estimator	$\lvert t_{count}\rvert$	*p-value*	Decision
	β_3	39.41	9.358	0.000	Significant
	β_4	-6.444	-8.858	0.000	Significant
X2	β_5	0.021	0.228	0.824	Not Significant
	β_6	-7.264	-32.963	0.000	Significant
	β_7	30.047	16.125	0.000	Significant
	β_8	1.268	3.396	0.006	Significant
X3	β_9	139.299	9.661	0.000	Significant
	β_{10}	-4.408	-2.755	0.019	Significant
	β_{11}	-27.093	-30.053	0.000	Significant
	β_{12}	24.579	12.944	0.000	Significant
X4	β_{13}	-39.031	-9.21	0.000	Significant
	β_{14}	18.42	25.796	0.000	Significant
	β_{15}	-53.667	-27.183	0.000	Significant
	β_{16}	7.276	9.187	0.000	Significant
X5	β_{17}	-1.116	-3.612	0.004	Significant
	β_{18}	-31.026	-23.205	0.000	Significant
	β_{19}	0.714	4.855	0.001	Significant
	β_{20}	-0.569	-7.396	0.000	Significant
X6	β_{21}	4.762	27.628	0.000	Significant
	β_{22}	6.572	37.474	0.000	Significant
	β_{23}	1.134	8.623	0.000	Significant
	β_{24}	-7.305	-14.427	0.000	Significant

Based on Table 5, testing parameters individually by comparing the $|t_{count}|$ value and the t_{table} value, we got $|t_{count}| = 0.028 < t_{0.05;23} = 2.070$, or the $|t_{count}|$ value is more significant than the $t_{0.05;23}$. Then, the p-value 0.824 is greater than $\alpha = 0.05$. So, the decision is to reject H_0. From 24 parameters, only β_5 is not significant. Although the X_2 variable still has a significant effect on the GDI, because the β_6, β_7, and β_8 parameters are significant. So, it can be concluded that the variable of Life Expectancy (X_1), Expected Years of Schooling (X_2), Female Income Contribution (X_3), Sex Ratio (X_4), Female Labor Force Participation Rate (X_5), and the High School Enrollment Rate (X_6) have a significant effect.

4.5. Residual Assumption Testing

a) Identical Residual Assumption

To see the homogeneity of the residual variance, the identical assumption is checked. The identical residual assumption is fulfilled if there is no indication of heteroscedasticity. Checking the assumption of identical residuals can be done with the Glejser test, which performed by regressing the absolute of the residuals with the predictor variable [14].

$H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2 = \sigma^2$

H_1: at least exist $\sigma_1^2 \neq \sigma_2^2; i = 1,2,\ldots,n$

[DOI: 10.14710/JSUNIMUS.12.2.2024.29-41]

Table 6. Result of Glejser Test

	Degree of Freedom (df)	Sum Square (SS)	Mean Square (MS)	F_{count}	F_{table}
Regression	6	22.657	3.776	0.133	0.991
Error	22	623.081	28.322		
Total	28	645.738			

Based on Table 6, it is known that the p-value is 0.991. With a significance level of 0.05, the decision fails to reject H_0 because the p-value is more significant than $\alpha = 0.05$, or the value of $F_{count} = 0.133$ is smaller than $F_{table} = 2.55$. So, it can be concluded that there is no heteroscedasticity or in other words, the variation between residuals is the same. We said that the identical assumption on the residuals is fulfilled.

b) Independent Residual Assumption

The Run Test is used to test whether there is a high correlation between residuals. Where there is no correlation between residuals, it is said that the residuals are random [15]. The hypothesis used is as follows.

$H_0: \rho = 0$ (Independent Residual)

$H_0: \rho \neq 0$ (Dependent Residual)

Table 7. Result of Run *Test*

	Residual
Test Value[a]	1.27
Cases < Test Value	14
Cases >= Test Value	15
Total Cases	29
Number of Runs	17
Z	0.385
Asymp. Sig. (2-tailed)	0.700

Table 7 as an output with SPSS, shows that the probability value = 0.700 is more significant than α, which means accept H_0. So, it can be concluded that the residuals are random or there is no autocorrelation between residual values.

[DOI: 10.14710/JSUNIMUS.12.2.2024.29-41]

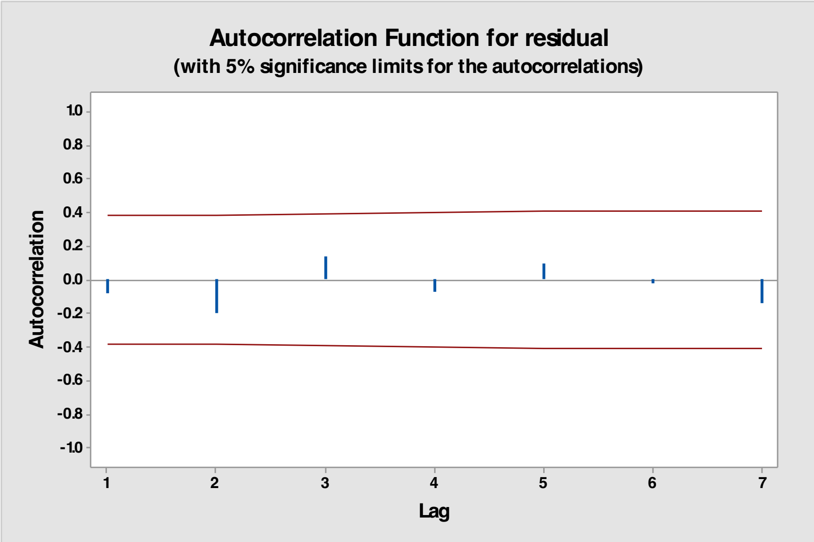


Fig. 2 Plot ACF Residual

Based on Figure 2, the autocorrelation value of the residuals is at the significant limit, or in other words, there are no lags that go out of bounds. So, it can be concluded that there is no correlation between residuals. So, the assumption of independent residuals is fulfilled.

c) Residual Normality Assumption Testing

The residual normality assumption test can be carried out to check whether the residuals follow a normal distribution. Testing can be done with the Kolmogorov-Smirnov (KS) test, with the hypothesis used as follows:

H_0: residuals are normally distributed

H_1: residuals are not normally distributed

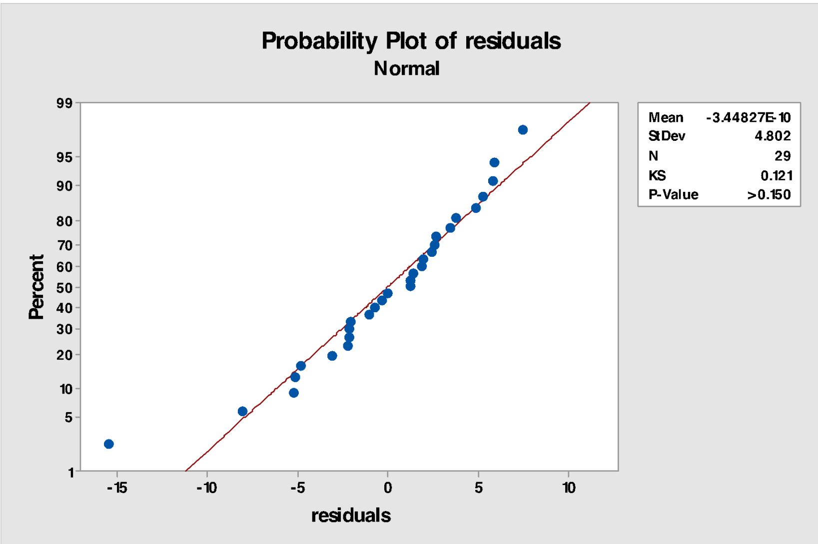


Fig. 2 Result of Kolmogov Smirnov Test

Based on Figure 2, the value of $KS_{count} = 0.121$ is smaller than $KS_{table} = 0.246$, and the p-value > 0.150. So, it fails to reject H_0. Thus, the residuals of the spline nonparametric regression model have met the assumption of normal distribution.

[DOI: 10.14710/JSUNIMUS.12.2.2024.29-41]

4.6. Criteria for Selecting the Best Model

One way to determine the best regression model is to use the coefficient of determination (R^2) value, which shows how good the regression model is in explaining the variability of HDI in Papua Province. The R^2 value is 99.95%. Thus, the spline nonparametric regression model obtained can explain the variability of HDI in Papua Province by 99.95%. In contrast, the residual is explained by other variables.

4.7. Spline Regression Model Interpretation

Based on the analysis results carried out previously, the minimum GCV value of the best spline nonparametric regression model is obtained, and the residual assumptions are met. Then, the interpretation of the model using three-knot points is carried out. The following is a model of the factors affecting HDI using spline nonparametric regression.

$$\begin{aligned}\hat{y} = 1191.433 &- 27.498X_1 + 27.353(X_1 - 58.885)_+^1 + 39.410(X_1 - 66.832)_+^1 \\ &- 6.444(X_1 - 70.287)_+^1 + 0.021X_2 - 7.264(X_2 - 6.146)_+^1 \\ &+ 30.047(X_2 - 11.379)_+^1 + 1.268(X_2 - 13.655)_+^1 + 139.299 \\ &- 4.408(X_3 - 30.797)_+^1 - 27.093(X_3 - 46.799)_+^1 + 24.597(X_3 - 53.756)_+^1 \\ &- 39.031X_4 + 18.420(X_4 - 108.238)_+^1 - 53.667(X_4 - 115.800)_+^1 \\ &+ 7.276(X_4 - 119.087)_+^1 - 1.116X_5 - 31.026(X_5 - 44.647)_+^1 \\ &+ 0.714(X_5 - 76.701)_+^1 - 0.569(X_5 - 90.638)_+^1 + 4.762X_6 \\ &+ 6.572(X_6 - 41.875)_+^1 + 1.134(X_6 - 74.56)_+^1 - 7.305(X_6 - 88.621)_+^1\end{aligned}$$

1) If X_2, X_3, X_4, X_5 and X_6 are constant, then the effect of Life Expectancy (X_1) on HDI is:

$$\begin{aligned}\hat{y} = 1191.433 &- 27.498X_1 + 27.353(X_1 - 58.885)_+^1 + 39.410(X_1 - 66.832)_+^1 \\ &- 6.444(X_1 - 70.287)_+^1\end{aligned}$$

$$= \begin{cases} 1191.433 - 27.498X_1 : X_1 < 58.885 \\ -419.248 - 0.145X_1 : 58.885 \le X_1 < 66.832 \\ -3053.097 + 39.265X_1 : 66.832 \le X_1 < 70.287 \\ -2600.168 + 32.821X_1 : X_1 \ge 70.287 \end{cases}$$

Based on the model obtained, it is known that if a Life Expectancy less than 58.885 years increases by one year, then the GDI will decrease by 27.498%. If life expectancy between 58.885 years and 66.832 years increases by one year, the HDI will decrease by 0.145%. If life expectancy between 66.382 years and 70.287 years increases by one year, the GDI will increase by 39.265%. Meanwhile, if the life expectancy of more than or equal to 70.287 years increases by one year, the GDI will increase by 32.821%.

2) If X_1, X_3, X_4, X_5 and X_6 are constant are held constant, the effect of Expected Years of Schooling (X_2) on HDI is as follows.

$$\begin{aligned}\hat{y} = 1191.433 + 0.021X_2 - 7.264(X_2 - 6.146)_+^1 + 30.047(X_2 - 11.379)_+^1 \\ +1.268(X_2 - 13.655)_+^1\end{aligned}$$

$$= \begin{cases} 1191.433 + 0.021X_2 : X_2 < 6.146 \\ 1236.078 - 7.243X_2 : 6.146 \le X_2 < 11.379 \\ 894.173 + 22.804X_2 : 11.379 \le X_2 < 13.655 \\ 876.858 + 24.072X_2 : X_2 \ge 13.655 \end{cases}$$

Based on the model obtained, it is known that if the Expected Years of Schooling of less than 6.146 years increases by one year, then the GDI will increase by 0.021%. If the Expected Years of Schooling between 6.146 years and 11.379 years increases by one year,

[DOI: 10.14710/JSUNIMUS.12.2.2024.29-41]

then the GDI will decrease by 7.243%. If the Expected Years of Schooling between 11.379 years and 13.655 years increase by one year, the GDI will increase by 22.804%. Meanwhile, if a Life Expectancy of more than or equal to 13.655 years increases by one year, the GDI will increase by 24.072%.

3) If X_1, X_2, X_4, X_5 and $X_6$are held constant, the effect of Female Income Contribution (X_3) on GDI is as follows.

$$\hat{y} = 1191.433 + 139.299X_3 - 4.408(X_3 - 30.797)_+^1 + 27.093(X_3 - 46.799)_+^1 + 24.579(X_3 - 53.756)_+^1$$

$$= \begin{cases} 1191.433 + 139.299X_3 : X_3 < 30.797 \\ 139.299 - 134.891X_3 : 30.797 \leq X_3 < 46.799 \\ 2595.111 + 107.798X_3 : 46.799 \leq X_3 < 53.756 \\ 1273.842 + 132.377X_3 : X_3 \geq 53.756 \end{cases}$$

Based on the model obtained, it is known that if women's income contribution of less than 30.797% increases by 1%, then the GDI will increase by 139.299%. If the contribution of women's income between 30.797% and 46.799% increases by 1%, the GDI will increase by 134.891%. If the contribution of women's income between 46.799% and 53.756% increases by 1%, then the GDI will increase by 107.798%. Meanwhile, if the contribution of women's income of more than 53.756% increases by 1%, the GDI will increase by 132.377%.

4) If X_1, X_2, X_3, X_5 and X_6 are held constant, the effect of Sex Ratio (X_4) on GDI is as follows.

$$\hat{y} = 1191.433 + 39.031X_4 - 18.420(X_4 - 108.238)_+^1 + 53.667(X_4 - 115.800)_+^1 + 7.276(X_4 - 119.087)_+^1$$

$$= \begin{cases} 1191.433 + 39.031X_4 : X_4 < 108.238 \\ -802.311 + 20.611X_4 : 108.238 \leq X_4 < 115.800 \\ 5412.274 + 33.056X_4 : 115.800 \leq X_4 < 119.087 \\ 4545.797 + 25.78X_4 : X_4 \geq 119.087 \end{cases}$$

Based on the model obtained, it is known that if the value of the sex ratio less than 108.238% increases by 1%, the GDI will decrease by 39.031%. If the sex ratio value between 108.238% and 115.800% increases by 1%, the GDI will increase by 20.611%. If the sex ratio value between 115.800% and 119.087% increases by 1%, the GDI will decrease by 33.056%. Meanwhile, if the value of the sex ratio greater than or equal to 119.087% increases by 1%, the GDI will decrease by 25.78%.

5) If X_1, X_2, X_3, X_4 and X_6 are considered constant, the effect of the Female Labor Force Participation Rate (X_5) on GDI is as follows.

$$\hat{y} = 1191.433 - 1.116X_5 - 31.026(X_5 - 44.647)_+^1 + 0.714(X_5 - 76.701)_+^1 - 0.569(X_5 - 90.638)_+^1$$

$$= \begin{cases} 1191.433 - 1.116X_5 : X_5 < 44.647 \\ 2576.651 - 32.142X_5 : 44.647 \leq X_5 < 76.701 \\ 2521.886 - 31.428X_5 : 76.701 \leq X_5 < 90.638 \\ 2573.459 - 31.997X_5 : X_5 \geq 90.638 \end{cases}$$

Based on the model obtained, it is known that if the value of Female Labor Force Participation Rate less than 44.647% increases by 1%, the GPA will decrease by 1.116%. If the value of Female Labor Force Participation Rate between 44.647% and 76.701%

[DOI: 10.14710/JSUNIMUS.12.2.2024.29-41]

increases by 1%, the GDI will decrease by 32.142%. If the value of Female Labor Force Participation Rate between 76.701% and 90.638% increases by 1%, the GDI will decrease by 31.428%. Meanwhile, if the value of Female Labor Force Participation Rate is greater than or equal to 90.638% and increases by 1%, the GDI will decrease by 31.997%.

6) If X_1, X_2, X_3, X_4 and X_5 are held constant, then the effect of High School Enrollment Rate (X_6) on GDI is as follows.

$$\hat{y} = 1191.433 + 4.7626X_6 + 6.572(X_6 - 41.875)_+^1 + 1.134(X_6 - 74.456)_+^1 - 7.305(X_6 - 88.621)_+^1$$

$$= \begin{cases} 1191.433 + 4.7626X_6 : X_6 < 41.875 \\ 916.231 - 11.334X_6 : 41.875 \le X_6 < 74.456 \\ 831.789 - 12.468X_6 : 74.456 \le X_6 < 88.621 \\ 1479.174 + 5.163X_6 : X_6 \ge 88.621 \end{cases}$$

Based on the model obtained, it is known that if the value of High School Enrollment Rate less than 41.875% increases by 1%, the GDI will increase by 4.762%. If the High School Enrollment Rate value between 41.875% and 74.456% increases by 1%, the GDI will increase by 11.334%. If the value between 74.456% and 88.621% increases by 1%, the GDI will increase by 12.468%. Meanwhile, if the value is more than 88.621% and increases by 1%, the GDI will increase by 5.163%.

5. CONCLUSION

The Gender Development Index (GDI) of Papua Province in 2021 has an average value of 80.10%. The Regency/City of Papua Province still has a condition of gender inequality, meaning that women's development achievements are lower than men's. The best spline regression model to model the factors affecting the GDI in Papua Province in 2021 based on the smallest GCV is 3-knot points. Based on the results of the analysis, it is obtained that the factors that significantly affect the GDI of the Regency/City in Papua Province are Life Expectancy, Expected Years of Schooling, Female Income Contribution, Sex Ratio, Female Labor Force Participation Rate, High School Enrollment Rate. Based on this optimal spline regression model, the R^2 value is 99.95%, which means the obtained model can explain the diversity of GDI values in Papua Province by 99.95%. In contrast, other variables explain the rest and residual assumptions that have been met and are good at modeling. The best spline regression model obtained is as follows:

$$\begin{aligned} \hat{y} = 1191.433 &- 27.498X_1 + 27.353(X_1 - 58.885)_+^1 + 39.410(X_1 - 66.832)_+^1 \\ &- 6.444(X_1 - 70.287)_+^1 + 0.021X_2 - 7.264(X_2 - 6.146)_+^1 \\ &+ 30.047(X_2 - 11.379)_+^1 + 1.268(X_2 - 13.655)_+^1 + 139.299 \\ &- 4.408(X_3 - 30.797)_+^1 - 27.093(X_3 - 46.799)_+^1 \\ &+ 24.597(X_3 - 53.756)_+^1 - 39.031X_4 + 18.420(X_4 - 108.238)_+^1 \\ &- 53.667(X_4 - 115.800)_+^1 + 7.276(X_4 - 119.087)_+^1 - 1.116X_5 \\ &- 31.026(X_5 - 44.647)_+^1 + 0.714(X_5 - 76.701)_+^1 \\ &- 0.569(X_5 - 90.638)_+^1 + 4.762X_6 + 6.572(X_6 - 41.875)_+^1 \\ &+ 1.134(X_6 - 74.56)_+^1 - 7.305(X_6 - 88.621)_+^1 \end{aligned}$$

[DOI: 10.14710/JSUNIMUS.12.2.2024.29-41]

REFERENCES

[1] S. Aprilianti and Y. Setiadi, "Faktor-faktor Yang Memengaruhi Indeks Pembangunan Gender di Indonesia Tahun 2020," in *Seminar Nasional Official Statistics*, 2022, pp. 245–254.

[2] S. U. R. Sari, "Aplikasi Metode Regresi Nonparametrik Spline Multivariabel untuk Pemodelan Faktor-Faktor yang Mempengaruhi Indeks Pembangunan Gender di Provinsi Jawa Barat," *Jurnal Statistika Universitas Muhammadiyah Semarang*, vol. 6, no. 2, 2018.

[3] A. Wahyu, L. Firdauzi, T. Waskita, M. Alfana, A. Pitoyo, and Y. Kusmanto, *Analisis Indeks Pembangunan Gender Provinsi Maluku Utara*. 2020. doi: 10.13140/RG.2.2.11702.52800.

[4] S. Madrim, "Kekerasan Terhadap Perempuan pada Tahun 2021 Merupakan Tertinggi dalam 10 Tahun Terakhir." Accessed: May 29, 2023. [Online]. Available: Https://Www.Voaindonesia.Com/A/Kekerasan-Terhadap-Perempuan-Pada-2021-Merupakan-Tertinggi-Dalam-10-Tahun-Terakhir-/6473578.Html

[5] Kementerian Pemberdayaan Perempuan dan Perlindungan Anak, "Pembangunan Manusia Berbasis Gender 2021," 2021.

[6] Z. D. Rizqiana, "Modelling Gender Development Index in Central Java In 2021 with Geographically Weighted Regression," *Journal of Statistical Methods and Data Science*, vol. 1, no. 1, pp. 25–36, 2023.

[7] Suparti, R. Santoso, A. Prahutama, and A. R. Devi, *Regresi Nonparametrik*. Wade Group, 2018.

[8] F. K. Lembang, V. Y. I. Ilwaru, and N. Lewaherilla, "Pemodelan Regresi Nonparametrik Spline dan Aplikasinya Pada Indeks Kebahagiaan Provinsi Di Indonesia," *Jurnal Riset dan Aplikasi Matematika (JRAM)*, vol. 7, no. 1, pp. 56–65, 2023.

[9] A. Indrasetianingsih, F. Fitriani, and P. J. Kusuma, "Klasifikasi Indeks Pembangunan Gender Di Indonesia Tahun 2020 Menggunakan Supervised Machine Learning Algorithms," *Inferensi*, vol. 4, no. 2, pp. 129–136, 2021.

[10] I. Aryantari, "Pemodelan Indeks Pembangunan Gender (IPG) di Provinsi Jawa Timur Menggunakan Regresi Nonparametrik Spline." Tugas Akhir. Surabaya: Institut Teknologi Sepuluh Nopember, 2017.

[11] R. L. Eubank, *Spline Smoothing and Nonparametric Regression*. 1988.

[12] G. Wahba, *Spline models for observational data*. SIAM, 1990.

[13] I. N. Budiantara, "Model Keluarga Spline Polinomial Truncated dalam Regresi Semiparametrik," *Berkala MIPA*, vol. 15, no. 3, pp. 55–61, 2005.

[14] D., N. Gurajati, *Basic Econometrics*, Fourth. Singapore: McGraw-Hill Inc., 2004.

[15] Samin. Radjid, "Pemodelan Prevalensi Balita Stunting Dengan Mengunakan Metode Regresi Nonparametrik Spline (Studi Kasus: Provinsi Maluku, Maluku Utara Dan Papua Tahun 2021)," Skripsi, Universitas Pattimmura, Ambon, 2022.

[DOI: 10.14710/JSUNIMUS.12.2.2024.29-41]