Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2023, 2023:45
https://doi.org/10.28919/cmbn/7946
ISSN: 2052-2541

MODELING OF POVERTY LEVEL IN SOUTH SULAWESI PROVINCE THROUGH SPLINE NONPARAMETRIC REGRESSION APPROACH

WAHIDAH ALWI[1,*], IRWAN[1], ERMAWATI[1], TRY AZISAH NURMAN[1], HERNAWATI[2], RISNAWATI IBNAS[1]

[1]Department of Mathematics, Faculty of Science and Technology, Universitas Islam Negeri Alauddin Makassar, Indonesia

[2]Department of Physics, Faculty of Science and Technology, Universitas Islam Negeri Alauddin Makassar, Indonesia

Abstract: This study discussed modeling poverty levels in South Sulawesi through Spline Nonparametric Regression Approach. Since no specific pattern was formed from the relation between the poverty level in South Sulawesi and the factors that influence it, the researchers used nonparametric regression modelling with Spline approach. The Spline model is very good in modelling data that has changeable patterns at certain subintervals. The aimed of this study were to investigate the factors that have the most influence on the poverty level and modelling the poverty level in South Sulawesi through nonparametric regression. The scope of this study was the use of Generalized Cross-Validation (GCV) in selecting optimal knot points; in this case, it used 1, 2, and 3 knots. This study found that three factors mainly influence the poverty level in South Sulawesi in 2017: unemployment, population growth, and literacy rates. The results showed that the best spline model was the model with three-knot

*Corresponding author
E-mail address: wahidah.alwi@uin-alauddin.ac.id
Received March 07, 2023

points and the minimum GCV value is 11.1155. In addition, the R^2 value is 79.75%. It means that the variables of unemployment, population growth, and literacy rate can explain 79.75% of the variation of the poverty variable, while other variables explain the remaining 20.25%.

Keywords: nonparametric regression; spline; knot point.

2020 AMS Subject Classification: 92B10.

1. INTRODUCTION

Poverty is often known as a state of lack of money or goods to ensure life sustainability, and is a classic problem faced by most developing countries and is also one of the economic indicators to see the level of community welfare in one area. The number of poor people in Indonesia fluctuates from year to year. The poor population was around 54.2 million people in 1976. This number then reduced to 42.3 million people (about 32.8 million people in urban areas, and about 9.5 million people in rural areas) in 1980, or decreased by approximately 21.95% from 1976. In 1990 the number of poor people decreased to around 27.2 million people or decreased by about 35.69% from 1980. However, the number of poor people increased to around 38.4 million people in 2002 and dropped to approximately 37.17 million people in 2007. Fluctuations in the number of poor people in Indonesia are caused by various factors, such as the economic crisis, the increase in population every year, government policies, etc. According to the Susenas data collected by BPS (Central Bureau of Statistics) in 2017, the poor population in South Sulawesi was 825.97 people or 9.48%. The poor people in urban areas were 166.50 people or 4.76%, while in rural areas, it was 659.470 people or 12.65% [1].

A previous study using the Spline nonparametric regression approach for modeling the rate of economic growth in East Java found that districts/cities in East Java with the highest economic growth rates were Bojonegoro district at 9.2%. In contrast, the lowest economic growth rate was Sampang district which showed 6.14%. All variables are significant to the model, namely the level of labor participation (x1), regional income and expenditure budgets (x2), the number of medium/large industries (x3), and general flow funds (x4), with a coefficient of determination of 85.66% [2]. Another study on modeling the poverty indicator in Indonesia found that the higher the literacy rate, the lower the percentage of poor people in Indonesia. Still, it was not when the literacy rate was between 93.15% to 97.67% that the percentage of the poor increased. The higher the literacy rate was, the lower the tendency of the poverty index would be [3].

Based on the previous studies, the researchers were interested in modeling the average poverty rate in South Sulawesi by using the spline nonparametric regression approach. One of the methods that could be used to know factors that influence the poverty level in South Sulawesi was nonparametric regression analysis. The researchers used a nonparametric regression model with the Spline approach because the relation between the poverty level in South Sulawesi and the factors that influenced it did not form a certain pattern. The spline method is very good in modeling data that had changeable patterns at certain subintervals. Spline is a model that has statistical and visual interpretations and has a very good ability to be generalized to complex and complicated statistical modeling [4]–[6]. One method to select the optimal smoothing parameter in the spline estimator is Generalized Cross Validation (GCV) [7]–[9]. The GCV method has been widely used to select the glazing parameters and gives unbiased results [10]–[15].

The problems investigated in this study were the factors that mostly influence the poverty level and modeling the poverty level in Sulawesi Selatan through nonparametric regression. The scope of the problem of this study was the use of GCV in selecting optimal knot points. In this case, it used 1, 2, and 3 knots.

2. PRELIMINARIES

2.1. Dataset

This study used secondary data taken from the Central Bureau of Statistics of South Sulawesi in 2017. The research unit consisted of 24 districts in South Sulawesi. The response variable (Y) was the average poverty level per district in South Selatan. While the predictor variables were some factors that influence the average level of poverty in South Sulawesi in 2017, namely population growth rate (X1) and literacy rate (X2).

The steps of data analysis carried out in this study are as follows:

1. To create descriptive statistics from each variable and determine the poverty level characteristics in South Sulawesi.
2. To make a scatter plot between the poverty level (y) and each predictor variable.
3. To model the poverty rate in South Sulawesi using a spline with some knot points.
4. To Select the optimal knot point based on the minimum GCV.
5. To model the poverty rate with its predictor variables using a spline with optimal knots.
6. To test the significance of the parameters.
7. To calculate the R2.
8. To make interpretations and conclusions.

2.2. Spline Nonparametric Regression

Nonparametric regression is one statistic method used to know the relationship between the response and predictor variables. The regression curve shape is unknown, or there is no complete past information about the data pattern [16]. Nonparametric regression is so flexible in modeling the pattern of the data. Generally, the Spline nonparametric regression model is:

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

Therefore, $\boldsymbol{y_i}$ is the response variable, $\boldsymbol{t_i}$ is the predictor variable, and $\boldsymbol{f(t_i)}$ is a regression function that the pattern is unknown, ε_i~IIDN (0, $\sigma^2 I$). In the spline nonparametric regression model, the regression curve is approximated by a spline function of order p with the k1, k2, ..., kr knot points as in the following form:

$$\boldsymbol{f(t_i) = \sum_{j=0}^{p} \beta_j t_i^j + \sum_{l=1}^{r} \beta_{p+l} (t_i - k_l)_+^p} \tag{2}$$

β is the model variable, p is the spline orde, and k is the number of spline knots. Based on the equations (1) and (2), the regression equation as follows:

$$\boldsymbol{y_i = \sum_{j=0}^{p} \beta_j t_i^j + \sum_{l=1}^{r} \beta_{p+l} (t_i - k_l)_+^p + \varepsilon_i,\ i = 1, 2, \ldots, n} \qquad ((3)$$

2.3. Selecting the optimum knot point

The best spline nonparametric regression model is obtained from the optimal knot point. The optimal knot point is obtained from the minimum GCV value as the GCV value modification of the CV (Cross Validation) method. The GCV method is the most widely used because it has asymptotic optimal properties. The Generalized Cross-Validation (GCV) method can be written as follows:

$$\boldsymbol{GCV(k_1, k_2, \dots, k_l) = \frac{MSE(k_1 k_2 \dots k_l)}{[n^{-1} trace(I - H(k_1 k_2 \dots k_l)]^2}} \quad (4)$$

Thus, I is the identity matrix, and n is the number of observations.

$$\boldsymbol{GMSE(k_1, k_2, \dots, k_l) = n^{-1} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \quad (5)$$

2.4. Model Parameter Testing

a. Simultaneous Testing of Model Parameters

A simultaneous test needs to be carried out to know the significance of the model variables simultaneously. The hypotheses for the concurrent test are:

$H_0 = \beta_1 = \beta_2 = \cdots \beta_m = \gamma_1 = \gamma_2 = \cdots \gamma_{p+r} = 0$

$H_1 =$ at least one $\gamma_j \neq 0$ or $\beta_k \neq 0$

With $P + r$ is nonparametric regression parameter.

Statistic test:

$$\boldsymbol{F_{test} = \frac{MSR}{MSE}} \quad (6)$$

With

$$\boldsymbol{MSE = \frac{y'y - b'X(k)'y}{n - (m + p + r + 1)}}$$

$$\boldsymbol{MSR = \frac{b'X(k)'y - n\bar{y}^2}{m + p + r}}$$

The conclusion is H_0 is rejected if F_{test} is greater than $F_{\alpha;(m+(p+r),n-(m+(p+r))-1)}$ or p-value $< \alpha$, which causes at least there is one parameter that is significant to the model.

b. Individual Testing of Model Parameters

The individual testing parameter is used to obtain a conclusion from simultaneous model parameters with at least one significant model parameter. The hypothesis of individual testing as follows:

$H_0 = \beta_k = 0$

$H_1 = \beta_k \neq 0, k = 1,2, \dots, m$

The following are test hypotheses of nonparametric regression:

$H_0 = \gamma_j = 0$

$H_1 = \gamma_j \neq 0, j = 1,2, \dots, p + r$

Statistic test:

$$\boldsymbol{T_{test}} = \frac{\boldsymbol{\hat{\beta}_k}}{\boldsymbol{se\ (\hat{\beta}_k)}}\text{, for parametric parameters} \quad (7)$$

$$\boldsymbol{T_{test}} = \frac{\boldsymbol{\hat{\gamma}_k}}{\boldsymbol{se\ (\hat{\gamma}_k)}}\text{, for non parametric parameters} \quad (8)$$

3. Main Results

3.1. The characteristics of poverty and its factors in districts/cities of South Sulawesi Province in 2017

Based on data collected, some factors influenced the poverty in South Sulawesi, including the unemployment rate, population growth rate, and literacy rate. The characteristics of the three variables are presented in the following table 1 and figure 1.

Table 1. Descriptive statistic of response and predictor variables

Variables	Minimum	Maximum	Mean	Variance
Y	4,62	16,2	10,22	10,69
X_1	1,87	10,96	4,917	5,316
X_2	1,53	17,13	4,162	10,874
X_3	3,67	4,49	4,162	0,047

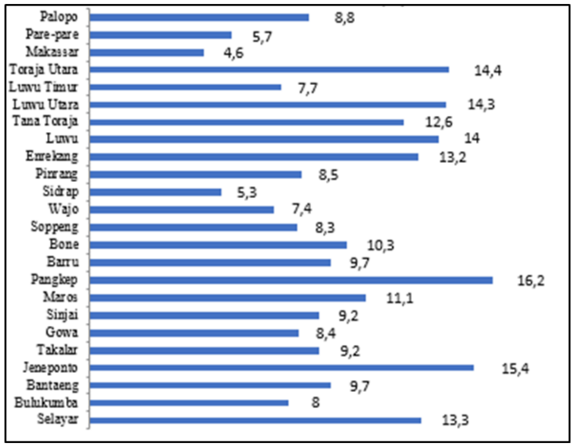

Figure 1. The poverty level in districts/cities of South Sulawesi Province in 2017

Table 1 and figure 1 show that the response variable (Y) is the poverty in South Sulawesi Province. The average poverty was 10.221 in 2017, which meant that 10 out of 100 residents of South Sulawesi are still categorized as poor. The variance value was 10.690, which indicated that the small variance value showed the data of the poor population in each district and city tend not to vary. The highest number of poor people was Pangkep Regency at 16.2%, while the lowest number was Makassar at 8.1%. The following figure 2 presents data on economic growth in South Sulawesi Province in 2017.

Variable X_1 was the unemployment rate to affect poverty. Based on table 1 and figure 1, the unemployment rate variable had an average of 4.9171, indicating that 4 out of 100 residents were still unemployed.

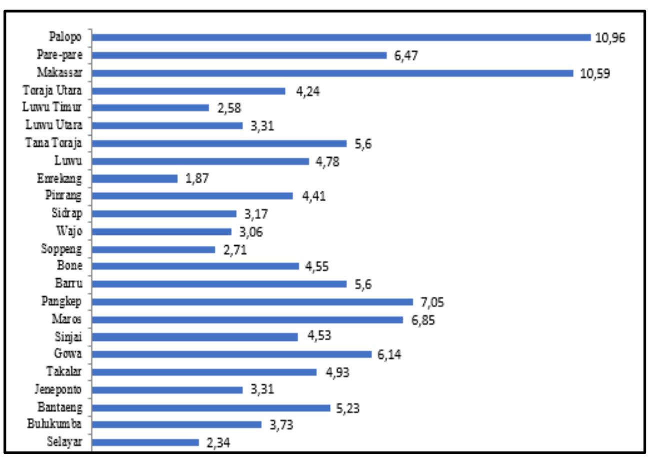

Figure 2. Unemployment rate in districts/cities of South Sulawesi in 2017

The variance value was 5.316, indicating it was quite small, which showed that unemployment data for each district and city was not too varied. The highest percentage was found in Palopo City, which was 10.96% unemployment, while the lowest was in Enrekeng Regency, 1.87%.

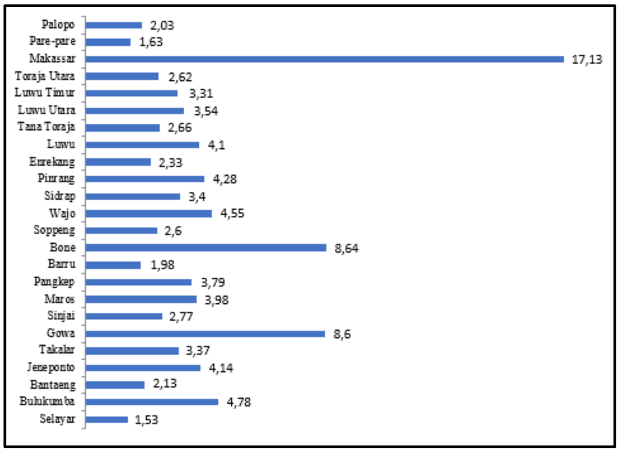

Figure 3. The population growth rate in districts/cities in South Sulawesi in 2017

Variable X_2 was the rate of population growth that affecting poverty. Table 1 and Figure 3 showed that the population growth rate variable had an average of 4.1621, and the variance value was 10.874. The variance value was quite small, indicating that the population growth rate for each district and city was not too varied. The highest percentage of population growth was in Makassar City, 17.13% of the population, while the lowest was in Selayar Regency at 1.53%.

Variable X_3 was the literacy rate. Table 1 and Figure 4 showed that the average value of the literacy rate was 4.1617%, with a variance of 0.047. The fairly small variance value indicated that the literacy rate data for each district and city was not too varied. The highest value of literacy rate in South Sulawesi was in Makassar City at 4.49%, while the lowest was in Wajo Regency at 3.67%.

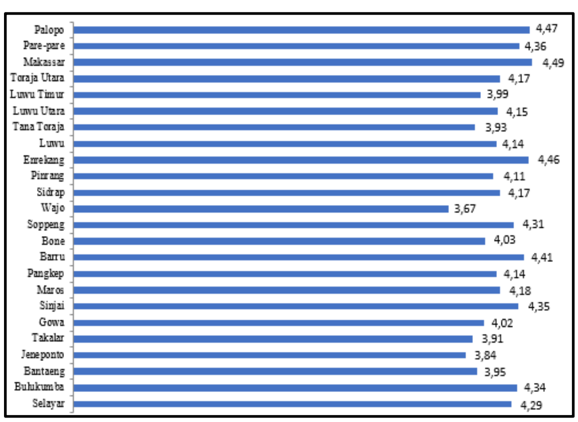

Figure 4. Literacy Rate in Districts/Cities of South Sulawesi Province in 2017

3.2. Analysis of the relationship pattern of factors influencing poverty in districts/cities in South Sulawesi

The first step in conducting regression analysis is to make a scatter plot to determine the relationship pattern of the predictor variable to the response variable, which is the number of poor people. The pattern of this relationship is used to determine the regression method used. The following is a scatter plot among the factors influencing poverty in South Sulawesi.

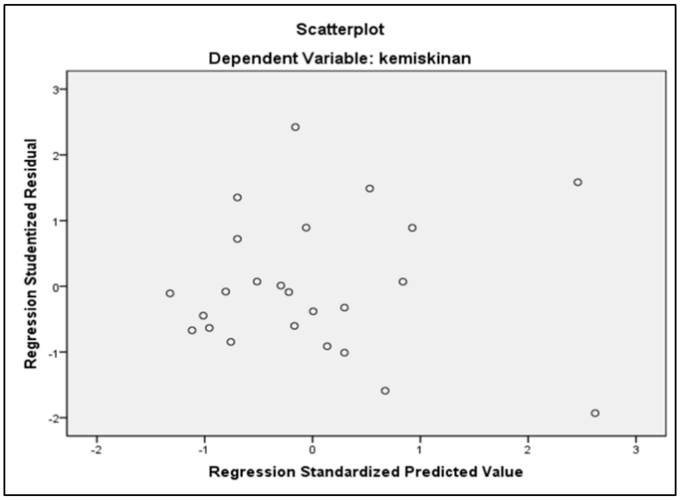

Figure 5. The Pattern of the Relationship between Unemployment Rate and Poverty in Regencies/Cities of South Sulawesi in 2017

Based on Figure 5, the Unemployment Rate (X_1) variable and poverty show an uncertain pattern of relationship. Therefore, the estimation of the model uses in nonparametric regression. Figure 6 shows the relationship between Population (X_2) and poverty. It indicates an uncertain relationship pattern so that the estimated model used is nonparametric regression. Figure 7 shows the relationship pattern formed between the Literacy Rate Variable (X_3) and poverty. It indicates an uncertain pattern of relationship so that the estimation model used is nonparametric regression.

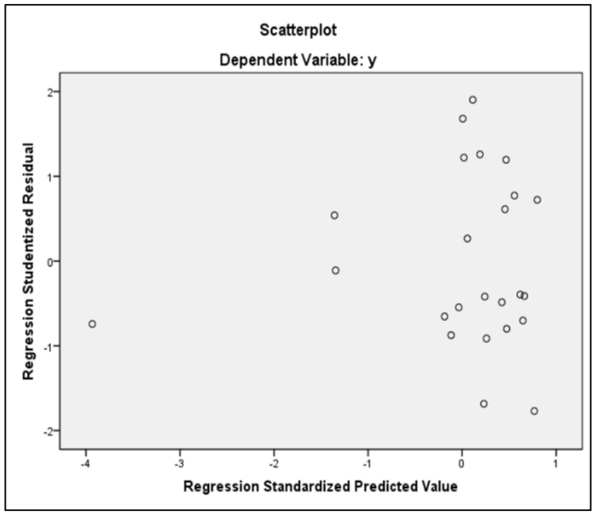

Figure 6. The pattern of the Relationship between Population and Poverty in Regency/City of South Sulawesi in 2017

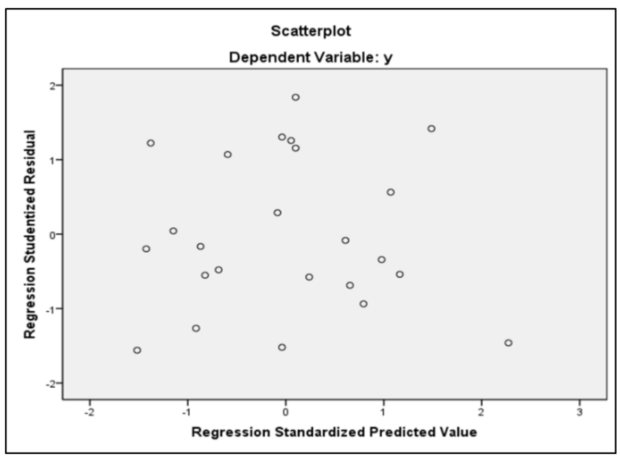

Figure 7. The pattern of Relationship between Literacy Rate and Poverty in Regencies/Cities of South Sulawesi in 2017

3.3. Selecting the optimum knot points

The knot point is the point of change in data behavior at certain sub-intervals. The best spline nonparametric regression model can be obtained from the optimal knot point using the Generalized Cross Validation (GCV) method. The minimum GCV value is the optimal knot point. The selection of optimal knot points with one, two, and three-knot points is explained as follows.

3.3.1. Selecting knot points with one-knot point

The spline nonparametric regression estimation model with the one-knot point on the poverty rate in South Sulawesi is as follows.

$$\hat{\boldsymbol{y}} = \hat{\boldsymbol{\beta}}_0 + \hat{\boldsymbol{\beta}}_1 \boldsymbol{x}_1 + \hat{\boldsymbol{\beta}}_2 (\boldsymbol{x}_1 - \boldsymbol{k}_1) + \hat{\boldsymbol{\beta}}_3 \boldsymbol{x}_2 + \hat{\boldsymbol{\beta}}_4 (\boldsymbol{x}_2 - \boldsymbol{k}_2) + \hat{\boldsymbol{\beta}}_5 \boldsymbol{x}_3 + \hat{\boldsymbol{\beta}}_6 (\boldsymbol{x}_3 - \boldsymbol{k}_3)$$

Table 2 shows ten GCV values around the minimum GCV for the one-knot spline nonparametric regression model.

Table 2. Value of GCV One Knot-Point

X_1	X_2	X_3	GCV
2.05551	1.84837	3.68674	14.6526
2.24102	2.16674	3.70347	14.5692
2.4263	2.4851	3.7202	13.9515
2.61204	**2.80347**	**3.73694**	**13.2867**
2.79755	3,12184	3.75367	13.9679
2.98306	3.4402	3.77041	14.421
3.16857	3.75857	3.78714	14.2315
3.35408	4.07694	3.80388	14.6443
3.53959	4.39531	3.82061	14.8992
3.7251	4.71367	3.83735	15.085

Based on Table 2, it is known that the minimum GCV value for the spline nonparametric regression model with a one-knot point is 13.2867. The value is obtained from one point of optimal knots on each predictor variable. The optimal knot point for the Unemployment Rate Variable (X_1) is at the 2.612041-knot point, the Population Growth Variable (X_2) is at the 2.80347-knot point, and the Literacy Rate Variable (X_3) is at the 3.73694-knot point.

3.3.2. Selecting knot points with two-knot points

After selecting knot points with the one-knot point, the next step is selecting the optimal knot points using two-knot points for each variable. The following is the spline nonparametric regression model of poverty in South Sulawesi with two-knot points.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 (x_1 - k_1) + \hat{\beta}_3 (x_1 - k_2) + \hat{\beta}_4 x_2 + \hat{\beta}_5 (x_2 - k_3) + \hat{\beta}_6 (x_2 - k_4) + \hat{\beta}_7 x_3 + \hat{\beta}_8 (x_3 - k_5) + \hat{\beta}_9 (x_3 - k_6)$$

Table 3 below shows ten GCV values around the minimum GCV value for a two-point knot spline nonparametric regression model.

Table 3. Value of GCV Two Knot-Points

X_1	X_2	X_3	GCV
1.87 2.05551	1.53 1.84837	3.67 3.68673	14.5604
1.87 2.24102	1.53 2.16674	3.67 3.70347	14.5692
1.87 2.42653	1.53 2.4851	3.67 3.7202	13.9515
1.87 2.61204	1.53 2.80347	3.67 3.73694	13.2867
1.87 2.79755	1.53 3.12184	3.67 3.75367	13.9679
1.87 2.98306	1.53 3.4402	3.67 3.77041	14.421
1.87 3.16857	1.53 3.75857	3.67 3.78714	14.2315
1.87 3.35408	1.53 4.07694	3.67 3.80388	14.6443
.	.	.	.
.	.	.	.
1,87 10.7745	1.53 16.8116	3.67 4.47327	15.3351
1.87 **10.96**	**1.53** **17.13**	**3.67** **4.49**	**12.6553**

Table 3 above shows the ten minimum GCV values around the GCV value for the spline nonparametric regression model with two-knot points is 12.6553. The value is obtained from two optimal knot points on each predictor variable. The optimal knot point for the Unemployment Variable (X_1) is at 1.87 and 10.96 knots. The Population Growth Variable (X_2) is at 1.53 and 17.13 knots. Furthermore, the Literacy Rate Variable (X_3) is at 3.67 and 4.49 knots.

3.3.3. Selecting knot points with three-knot points

The following is a spline nonparametric regression model of the poverty rate in South Sulawesi in 2017 with three-knot points.

$$\begin{aligned}\hat{\boldsymbol{y}} = {} & \hat{\boldsymbol{\beta}}_{\mathbf{0}} + \hat{\boldsymbol{\beta}}_{\mathbf{1}}\boldsymbol{x}_{\mathbf{1}} + \hat{\boldsymbol{\beta}}_{\mathbf{2}}(\boldsymbol{x}_{\mathbf{1}} - \boldsymbol{k}_{\mathbf{1}}) + \hat{\boldsymbol{\beta}}_{\mathbf{3}}(\boldsymbol{x}_{\mathbf{1}} - \boldsymbol{k}_{\mathbf{2}}) + \hat{\boldsymbol{\beta}}_{\mathbf{4}}(\boldsymbol{x}_{\mathbf{1}} - \boldsymbol{k}_{\mathbf{3}}) + \hat{\boldsymbol{\beta}}_{\mathbf{5}}\boldsymbol{x}_{\mathbf{2}} + \hat{\boldsymbol{\beta}}_{\mathbf{6}}(\boldsymbol{x}_{\mathbf{2}} - \boldsymbol{k}_{\mathbf{4}}) \\ & + \hat{\boldsymbol{\beta}}_{\mathbf{7}}(\boldsymbol{x}_{\mathbf{2}} - \boldsymbol{k}_{\mathbf{5}}) + \hat{\boldsymbol{\beta}}_{\mathbf{8}}(\boldsymbol{x}_{\mathbf{2}} - \boldsymbol{k}_{\mathbf{6}}) + \hat{\boldsymbol{\beta}}_{\mathbf{9}}\boldsymbol{x}_{\mathbf{3}} + \hat{\boldsymbol{\beta}}_{\mathbf{10}}(\boldsymbol{x}_{\mathbf{3}} - \boldsymbol{k}_{\mathbf{7}}) + \hat{\boldsymbol{\beta}}_{\mathbf{11}}(\boldsymbol{x}_{\mathbf{3}} - \boldsymbol{k}_{\mathbf{8}}) \\ & + \hat{\boldsymbol{\beta}}_{\mathbf{12}}(\boldsymbol{x}_{\mathbf{3}} - \boldsymbol{k}_{\mathbf{9}})\end{aligned}$$

Table 4 above shows seven GCV values, and one of them is the minimum for the three-knot spline nonparametric regression model.

Table 4. Value of GCV Three Knot-Points

X_1	X_2	X_3	GCV
2.05551	1.84837	3.68674	
2.24102	2.16674	3.70347	15.6858
2.42653	2.4851	3.7202	
2.05551	1.84837	3.68674	
2.24102	2.16674	3.70347	17.612
2.61204	2.80347	3.73694	
2.05551	1.84837	3.68674	
2.24102	2.16674	3.70347	20.0675
2.79755	3.12184	3.75367	
2.05551	1.84837	3.68674	
2.24102	2.16674	3.70347	20.1366
2.98306	3.4402	3.77041	

.	.	.	.
.	.	.	.
2.42653	2.4851	3.7202	
10.589	16,4833	4.45653	17.7716
10.7745	16.8116	4.47327	
2.61204	**2.80347**	**3.73694**	
2.79755	**3.12184**	**3.75367**	**11.1155**
2.98306	**3.4402**	**3.77041**	
2.61204	2.80374	3.73694	
2.79755	3.12184	3.75367	15.0393
3.16857	3.75857	3.78714	

Table 4 is known that the minimum GCV value for the spline nonparametric regression model with three-knot points is 11.1155. The value is obtained from three optimal knot points on each predictor variable. The optimal knot point for the Unemployment Variable (X_1) is at the knot point 2.61204; 2.79755; and 2.98306. The Population Growth Variable (X_2) is at the knot point 2.80347; 3.12184; and 3.4402. Then, the Literacy Rate Variable (X_3) is at the knot point 3.73694; 3.75367; and 3.77041.

3.3.4. Selecting the best knot point

The best knot points are the knot points that have a minimum GCV value. The following is a comparison of the minimum GCV values obtained at one-knot, two-knot, and three-knot points, as shown in Table 5. Based on that table, it is known that the minimum GCV value is spline nonparametric regression using three-knot points, which is 11.11546.

Table 5. The Comparison of CGV Value

Model	**GCV**
1 knot	13.2867
2 knots	12.6553
3 knots	11.1155

Based on the criteria of selecting the best model, it is known that the minimum GCV value is generated by a nonparametric spline regression model with three-knot points, which is 11.1155.

3.4. Testing the significant variables of spline nonparametric regression model

Testing the significant variables of the spline nonparametric regression is carried out after obtaining the best spline nonparametric regression model. This test is conducted to determine the factors that significantly affect poverty in South Sulawesi.

3.4.1. Simultaneous Testing

Simultaneous testing aims to determine the significance of the variables in the overall model. Testing Hypothesis is carried out to test the significance of variables simultaneously using the following hypotheses:

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_{12} = 0$

H_1 : is the least, $\beta_i \neq 0$; $where\ i = 1,2, \ldots, 12$

The analysis of variance from the nonparametric regression model is presented as the following table:

Table 6. The Comparison of CGV Value

Variation Source	Degree of Freedom	Sum Of Square (SS)	Mean Sum Of Square (MSS)	F_{table}	P-Value
Regression	12	99.07065	8.255888	0.61868	0.78918
Error	11	146.7889	13.34445		

Based on Table 6, it is known that the test statistic using F_{test} is 0.6186759 with a *p-value* of 0.7891845. At a significant level (α) of 5%, the p-value is greater than α, which meant fail to reject H_0. This result indicated that all predictor variables significantly affect the value of Poverty in South Sulawesi.

3.4.2. Individual Testing of Three Predictor Variables

Simultaneous test results show that there is at least one significant parameter of the spline nonparametric regression model. Hypothesis testing to test the significance of variables partially using the following hypothesis:

To find out the significant parameters, individual tests were carried out with the results as presented in Table 7.

Table 7. The estimation of Regression Variables

Variable	**Parameter**	**Coefisient**	t_{value}	**P-Value**	**Decision**
X_1	$\hat{\beta}_1$	40.92965	1.898511	0.084162	Fail to reject the H_0
	$\hat{\beta}_2$	48.19676	0.57415	0.577418	Fail to reject the H_0
	$\hat{\beta}_4$	-34.11079	1.701529	0.116903	Fail to reject the H_0
	$\hat{\beta}_4$	-122.877	-6.13307	7.38×10^{-5}	Reject the H_0
X_2	$\hat{\beta}_5$	-37.061	-2.4638	0.031466	Reject the H_0
	$\hat{\beta}_6$	83.86478	4.313273	0.001228	Reject the H_0
	$\hat{\beta}_7$	-5.30923	-1.366627	0.199132	Fail to reject the H_0
	$\hat{\beta}_8$	-41.909	-5.18804	0.0003	Reject the H_0
X_3	$\hat{\beta}_9$	-29.1382	-0.45319	0.65922	Fail to reject the H_0
	$\hat{\beta}_{10}$	21.73152	2.630239	0.023397	Reject the H_0
	$\hat{\beta}_{11}$	-0.88841	-0.03902	0.969572	Fail to reject the H_0
	$\hat{\beta}_{12}$	-0.52363	-0.50201	0.625559	Fail to reject the H_0

Table 7 shows the significant and insignificant variables in each variable accompanied by t_{test} and *p-value*. When comparing the *p-value* with a significance level of 0.05, we get five variables that significantly influence the model. The significant variables are the unemployment rate variable (X_1), Population Growth (X_2), and Literacy Rate (X_3). The three variables are considered as variables that have a significant effect on poverty in South Sulawesi.

3.4.3. Coefficient of Determination

The value of the coefficient of determination (R^2) shows how good the regression model is in explaining the variability of the poverty rate in South Sulawesi.

$$R^2 = \frac{SS_{regression}}{SS_{total}} \times 100\% = \frac{2037.537}{2554.763} \times 100\% = 79.75\%$$

Based on the analysis, the R^2 value is 79.75%. It means that the variables of unemployment, population growth and literacy rate can explain 79.75% of the variation of the poverty variable, while other variables explain the remaining 20.25%.

4. DISCUSSION

4.1. Research Characteristics

Poverty is literally defined as the limitations experienced by a person, a family, a community, or countries that result in an uncomfortable life which threatens rights, justice, world relations, generations, and the future of a nation and state. Poverty is the inability of a person or group of people who live under the government's poverty line in terms of economy. The number of poor people in South Sulawesi continues to fluctuate every year. In 2017 the number of poor people in South Sulawesi was 813.070 people, and it increases compared to the previous year, which was 807.030 people.

The Central Bureau of Statistics uses the concept of the basic needs approach in measuring poverty. With this approach, poverty is an economic inability to meet basic food and non-food needs as measured from the expenditure side.

Based on the category of poor people following the Central Bureau of Statistics, it was noted that the poverty line of the population of South Sulawesi in 2017 was 283.461/capita/month. It indicated the minimum income required to obtain the living standard for food and non-food needs in one area. If it is lower than this number, then it is categorized as poor.

Based on the analyses of the scatter plot, it can be seen that the three variables, unemployment, economic growth, and literacy rate on poverty, have an uncertain pattern of relationship, so the estimation model used to analyze is nonparametric spline regression.

4.2. Model Interpretation of Poverty Levels using Spline Nonparametric Regression

After testing the spline nonparametric regression model and all residual assumptions are met, the researchers can interpret the obtained regression model. Based on the findings, it is known that the coefficient of determination or R^2 of the spline nonparametric regression model is

79.75% with three significant variables; unemployment, population growth, and literacy rates. Spline nonparametric regression model that is formed using optimal knot points, i.e., three-knot points, can be seen in the following equation:

$$\begin{aligned}\hat{y} = -7,267 + 40,929X_1 + 48,196(X_1 - 2,612) - 34,110(X_1 - 2,797) \\ - 122,877(X_1 - 2,983) - 37,061X_2 + 83,864(X_2 - 2,803) \\ - 5,309(X_2 - 3,121) - 41,909(X_2 - 3,440) - 29,138X_3 \\ + 21,731(X_3 - 3,736) - 0,888(X_3 - 3,753) - 0,523(X_3 - 3,770)\end{aligned}$$

The interpretation of the significant variable models is conducted to determine its effect on poverty. The three significant variables are unemployment, population growth and literacy rates. Based on this model, each influential variable can be interpreted as follows:

1. If X_2 and $X_3$are constant, the effect of the unemployment rate (X_1) on poverty is:

Therefore,

$$\hat{y} = \begin{cases} -7,267 + 40,929X_1 \ ; & x_1 < 2,612 \\ -133,154+89,125X_1 \ ; & 2,612 \le x_1 < 2,797 \\ -228,559 + 55,015X_1 ; & 2,797 \le x_1 < 2,983 \\ -595,101 - 57,862X_1 ; & x_1 \ge 2,983 \end{cases}$$

The spline estimation model in this equation uses a p-order spline model. The equation represents the boundaries of the region of the satisfied variable x. In this Equation, there are four intervals for each spline section. The Equation above can be presented in the figure as follows.

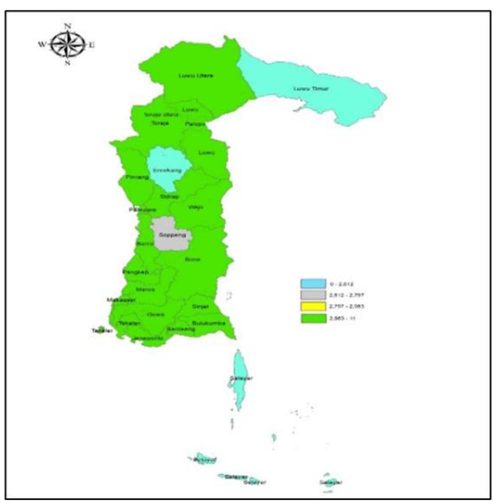

Figure 8. The Map of South Sulawesi based on the equation X1

Based on this Model, if the area with unemployment is less than 2.612, and it increases by 1%, then the poverty rate tends to increase by 40.929 percent. The districts/cities belonging to the first interval is as follows:

Table 8. Regencies/Cities in the First Interval of Unemployment

No	District/City
1	Luwu Timur
2	Enrekang
3	Selayar

Furthermore, in the second interval, if the area with unemployment ranges from 2.612 to 2.797, and it increases by one percent, then poverty tends to increase by 89.125 percent. The areas included in this interval are as follows:

Table 9. Regencies/Cities in the Second Interval of Unemployment

No	District/City
1	Soppeng

Moreover, in the third interval, if unemployment ranges from 2.797 to 2.983, and it increases by 1%, poverty tends to increase by 55.015%, and there are no areas that include this category. In the last interval, if unemployment is greater than 2.983, and it increases by 1%, then poverty tends to decrease by 57.862%. The areas included in this category are as follows:

Tabel 10. Districts/Cities in the Fourth Interval of Unemployment

No	District/City
1	Palopo
2	Pare-Pare
3	Makassar
4	Toraja Utara
5	Luwu Utara
6	Tanah Toraja
7	Luwu

8	Pinrang
9	Sidrap
10	Wajo
11	Bone
12	Barru
13	Pangkep
14	Maros
15	Sinjai
16	Gowa
17	Takalar
18	Jeneponto
19	Bantaeng
20	Bulukumba

2. If X_1 and X_3 are constant, then the effect of population growth (X_2) on poverty is:

$$\hat{y} = -7,267 - 37,061X_2 + 83,864(X_2 - 2,803) - 5,309(X_2 - 3,121) - 41,909(X_2 - 3,440)$$

Therefore,

$$\hat{y} = \begin{cases} -7,267 - 37,061X_2\ ; & x_2 < 2,803 \\ -242,337 + 46,803X_2; & 2,803 \le x_2 < 3,121 \\ -258,906 + 41,494X_2; & 3,121 \le x_2 < 3,440 \\ -403,072 - 0,415X_2\ ; & x_2 \ge 3,440 \end{cases}$$

The equations above can be presented in a figure 4.9. Based on this model, when the value of population growth is less than 2.803 and increases by 1%, then the poverty rate decreased by 37.061 percent. The areas included in this category are as presented in Table 4.11. Furthermore, in the second interval, if population growth ranges from 2.803 to 3.121, and it increases by 1%, poverty tends to increase by 46.803 percent. Moreover, in the last interval, if population growth is greater than 3.440 and increases by 1%, poverty decreases by 0.415%.

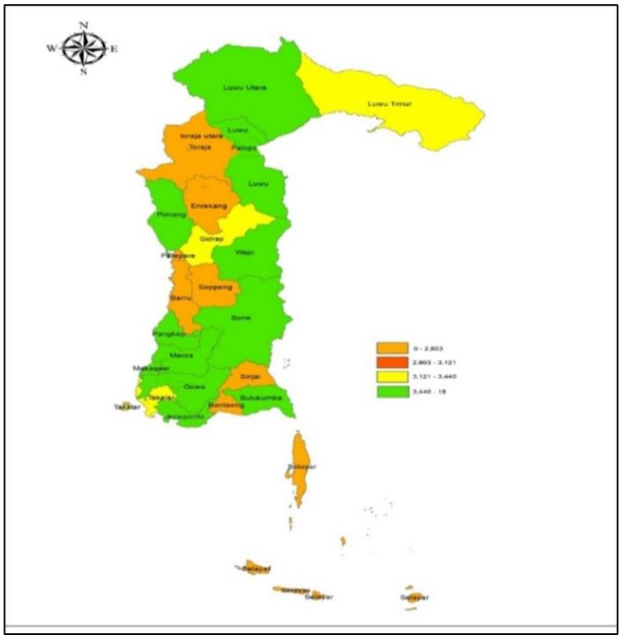

Figure 9. The Map of South Sulawesi based on the equation X_2

There are no areas included in this category. In the third interval, if population growth ranges from 3.121 to 3.440 and increases by 1%, then poverty increases by 41.494%.

Table 11. Regencies/Cities in the First Interval of Population Growth

No	District/City
1	Palopo
2	Pare-Pare
3	Toraja Utara
4	Tanah Toraja
5	Enrekang
6	Soppeng
7	Barru
8	Sinjai
9	Bantaeng
10	Selayar

The districts/cities belonging to the third interval are as follows:

Table 12. Regencies/Cities in the Third Interval of Population Growth

No	District/City
1	Luwu Timur
2	Sidrap
3	Takalar

The areas included in this category are as follows:

Tabel 13. Regencies/Cities in the Fourth Interval of Population Growth

No	District/City
1	Makassar
2	Luwu Utara
3	Luwu
4	Pinrang
5	Wajo
6	Bone
7	Pangkep
8	Maros
9	Gowa
10	Jeneponto
11	Bulukumba

3. If X_1 and X_2 are considered constant, then the influence of the Literacy Rate (X_3) on poverty is:

$$\boldsymbol{\hat{y} = -7,267 - 29,138X_3 + 21,731(X_3 - 3,736) - 0,888(X_3 - 3,753) - 0,523(X_3 - 3,770)}$$

Therefore,

$$\boldsymbol{\hat{y} = \begin{cases} -7,267 - 29,138X_3; & x_3 < 3,736 \\ -88,454 - 7,407X_3; & 3,736 \le x_3 < 3,753 \\ -91,786 - 8,295X_3; & 3,753 \le x_3 < 3,770 \\ -93,757 - 8,818X_3; & x_3 \ge 3,770 \end{cases}}$$

The Equation above can be presented in the figure as follows:

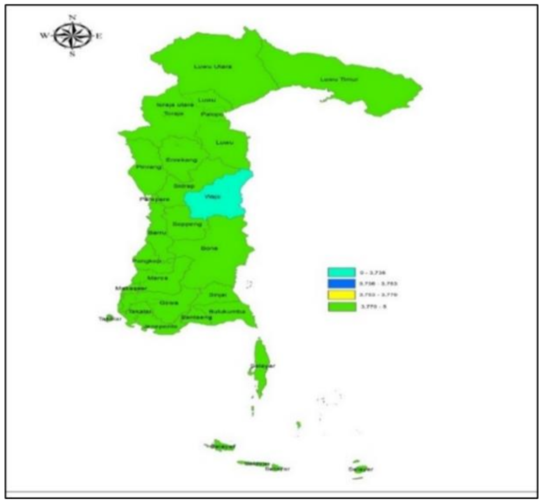

Figure 10. The Map of South Sulawesi based on the Equation of X_3

Based on this model, when the literacy rate is less than 3.736 and increases by 1%, the poverty rate decreased by 29.138%. The areas included in this category are as follows.

Table 14. Regencies/Cities in the First Interval of Literacy Rate

No	District/City
1	Wajo

Furthermore, in the second interval, if the literacy rate ranges from 3.736 to 3.753 and increases by 1%, poverty decreases by 7.407%. There are no areas included in this category. In the third interval, if the literacy rate ranges from 3.753 to 3.770 and increases by 1%, poverty decreases by 8.295 percent and no area is included in this interval.

In addition, in the fourth interval, if the literacy rate is greater than 3.770 and increases by 1%, poverty decreases by 8.818%. The areas that fall into this category are as follows.

Table 15. Regencies/Cities in the Fourth Interval of Literacy Rates

No	District/City
1	Selayar
2	Bulukumba
3	Jeneponto
4	Bantaeng

5	Takalar
6	Gowa
7	Sinjai
8	Maros
9	Pangkep
10	Barru
11	Sidrap
12	Soppeng
13	Pinrang
14	Enrekang
15	Luwu
16	Toraja Utara
17	Luwu Utara
18	Tanah Toraja
19	Luwu Timur
20	Bone
21	Palopo
22	Pare-Pare
23	Makassar

5. CONCLUSIONS

Three factors that influenced the poverty rate in South Sulawesi in 2017 were unemployment, population growth, and literacy rates. It could be formulated by using a spline nonparametric regression model as follows:

$$\begin{aligned}\boldsymbol{\hat{y}} = &-7,267 + 40,929X_1 + 48,196(X_1 - 2,612) - 34,110(X_1 - 2,797)\\ &- 122,877(X_1 - 2,983) - 37,061X_2 + 83,864(X_2 - 2,803)\\ &- 5,309(X_2 - 3,121) - 41,909(X_2 - 3,440) - 29,138X_3\\ &+ 21,731(X_3 - 3,736) - 0,888(X_3 - 3,753) - 0,523(X_3 - 3,770)\end{aligned}$$

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] Badan Pusat Statistik, Data dan Informasi kemiskinan kabupaten/kota tahun 2017, Jakarta, 2017.

[2] E.K. Litawati, I.N. Budiantara, Pendekatan regresi nonparametrik spline untuk pemodelan laju pertumbuhan ekonomi (LPE) di Jawa Timur, J. Sains Seni ITS, 2 (2013), D123–D128.

[3] D.A.W. Astiti, I.W. Sumarjaya, M. Susilawati, Analisis regresi nonparametrik spline multivariat untuk pemodelan indikator kemiskinan di Indonesia, E-Jurnal Mat. 5 (2016), 111-116. https://doi.org/10.24843/MTK.2016.v05.i03.p129

[4] I.N. Budiantara, Spline dalam regresi nonparametrik dan semiparametrik: sebuah pemodelan statistika masa kini dan masa mendatang, Pidato Pengukuhan untuk Jab. Guru Besar dalam Bid. Ilmu Mat. Stat. dan Probab. pada Jur. Stat. Fak. MIPA, 2009.

[5] I.N. Budiantara, Penelitian Bidang Regresi Spline Menuju Terwujudnya Penelitian Statistika yang Mandiri dan Berkarakter, in: Prosiding Seminar Nasional MIPA, 2011.

[6] H.I. Zebua, Pemodelan Kemiskinan di Sumatera Utara Menggunakan Regresi Nonparametrik Kernel dan Splines, in: Seminar Nasional Official Statistics, 2021, vol. 2021, no. 1, pp. 899–907. https://doi.org/10.34123/semnasoffstat.v2021i1.1087.

[7] J.D. Carew, G. Wahba, X. Xie, E.V. Nordheim, M.E. Meyerand, Optimal spline smoothing of fMRI time series by generalized cross-validation, NeuroImage. 18 (2003), 950-961. https://doi.org/10.1016/s1053-8119(03)00013-2.

[8] M. Maharani, D.R.S. Saputro, Generalized cross validation (GCV) in smoothing spline nonparametric regression models, J. Phys.: Conf. Ser. 1808 (2021) 012053. https://doi.org/10.1088/1742-6596/1808/1/012053.

[9] G. Wahba, Spline models for observational data, SIAM, 1990. https://doi.org/10.1137/1.9781611970128.

[10] A.R. Devi, R.F.W. Pratama, Suparti, Comparison of generalized cross validation and unbiased risk method for selecting optimal knot in spline truncated, J. Phys.: Conf. Ser. 1217 (2019) 012094. https://doi.org/10.1088/1742-6596/1217/1/012094.

[11] D.A. Girard, Asymptotic comparison of (partial) cross-validation, GCV and randomized GCV in nonparametric regression, Ann. Statist. 26 (1998), 315-334. https://doi.org/10.1214/aos/1030563988.

[12] D.A. Girard, Estimating the accuracy of (local) cross-validation via randomised GCV choices in kernel or smoothing spline regression, J. Nonparametric Stat. 22 (2010), 41–64. https://doi.org/10.1080/10485250903095820.

[13] K.C. Li, “Asymptotic optimality of C_L and generalized cross-validation in ridge regression with application to spline smoothing, Ann. Stat. 14 (1986), 1101-1112.

[14] B. Lestari, N. Budiantara, S. Sunaryo, et al. Spline estimator in multi-response nonparametric regression model, J. Ilmu Dasar, 11 (2010), 17-22.

[15] T.W. Utami, M.A. Haris, A. Prahutama, et al. Optimal knot selection in spline regression using unbiased risk and generalized cross validation methods, J. Phys.: Conf. Ser. 1446 (2020), 012049. https://doi.org/10.1088/1742-6596/1446/1/012049.

[16] R.L. Eubank, Nonparametric regression and spline smoothing, CRC press, 1999.