International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 6 Issue 12, December - 2022, Pages: 151-159

Spline Truncated Nonparametric Regression Models for Longitudinal Data and Implementation in Modeling Indonesia's Family Planning Program

Dita Amelia*, Sa'idah Zahrotul Jannah
Statistics Study Program, Mathematics Department
FST Universitas Airlangga
Surabaya, Indonesia
dita.amelia@fst.unair.ac.id

Abstract: *The most widely used nonparametric and semiparametric regression models in the last decade are spline regressions. Spline is a model that has a very special and very good statistical interpretation and visual interpretation. In regression analysis, the data that is often used is cross section data. But regression analysis can also be applied to longitudinal data. In its application in the social field, longitudinal data studies are also used involving research subjects in the form of regions. Several social issues that have been discussed a lot lately are regarding Indonesia's population growth which continues to increase. Through various Family Planning programs called "Keluarga Berencana (KB)" program, it is hoped that it can reduce the rate of population growth so that population growth remains balanced. Therefore, further analysis was carried out to estimate the success of KB Program through the percentage indicator Contraceptive Prevalence Rate (CPR) with poverty depth index, the average length of schooling for the population aged ≥ 15, and the proportion of women aged 20-24 years who are married or living together before the age of 18 as predictor variables. The model estimator was obtained by the WLS method. At the model application stage, longitudinal data is used with an analysis unit of 34 provinces in Indonesia in 2017 to 2021 which is divided into groups of Java Bali, Outer Java Bali I, and Outer Java Bali II. From the three groups, it was found that the optimum knots used were three knots with a GCV value of* $1{,}11\text{x}10^{-26}$ *for the Java-Bali group,* $1{,}05\text{x}10^{-26}$ *for the Outer Java-Bali I group, and* $3{,}19\text{x}10^{-27}$ *for the Outer Java-Bali II group. Each variable has a different effect on each subject. The number of knots in the nonparametric spline regression model has an influence in relating the behavior of the data in each sub-interval or segment. The optimal knot value is given by the smallest GCV value and the application of longitudinal data will provide a more accurate model for each observed subject.*

Keywords—nonparametric; splines; longitudinal data; KB; GCV

1. INTRODUCTION

In some research cases, it is often desirable to know the pattern of functional relationships between one or more variables. The method that is often used is regression analysis. Regression analysis is a statistical analysis that is used to determine the relationship pattern between one or more variables. The variables contained in the regression analysis consist of response variables and predictor variables. In order to be able to model one or more variables, the first thing that should be done is to investigate whether these variables are rationally correlated or not. If rationally correlated, statistical modeling can be done using regression analysis [1].

There are three types of regression developed by the researchers, namely parametric regression, nonparametric regression, and semiparametric regression. Parametric regression is a regression where the shape of the regression curve pattern is known, while nonparametric regression the shape of the regression curve pattern is unknown. Meanwhile, if some of the regression curve patterns are known and some are unknown, the regression analysis used is semiparametric regression. Generally, in regression analysis, before modeling using one of the three types of regression, it is preceded by examining the scatter plot between each predictor variable and the response variable. If the scatter plot shows the tendency of the data to follow a linear, quadratic, cubic, or polynomial pattern, a parametric regression model is used. However, if the scatter plot between the predictor variable and the response variable does not show a tendency for the data to follow a certain pattern, the nonparametric regression model is used.

In the regression analysis, the data used is cross sectional data. However, regression analysis can also be applied to longitudinal data or what is commonly called panel data. Longitudinal data is a combination of cross section data and time series data [2].

Longitudinal studies with nonparametric regression approach have also been developed by several researchers including [3] with spline approach, [4] with gee smoothing spline. Reference [5] used penalized splines with AIDS case studies. While [6] used a semiparametric spline regression model approach for longitudinal data on cases of CD4 levels in HIV patients.

In its application in the social field, longitudinal data studies are also used involving area-based research subjects. The research uses longitudinal data with the research subjects being the areas carried out by [7] regarding the factors that influence the human development index in regencies/cities in East Java province with 38 regencies/cities in East Java province as the research subjects.

Several social issues that have been widely discussed lately are regarding the growth of Indonesia's population which continues to increase from year to year. Based on the results of the 2020 population census, it is known that the population growth rate is 1,25% and a population development policy is needed. In overcoming these problems, the National Family Planning Coordinating Board or Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN) has a very important role. Through various family planning programs it is expected to reduce the rate of population growth so that population growth remains balanced. One of the main programs of the BKKBN is related to the use of contraceptives or family planning programs. Therefore, it is necessary to carry out further analysis to estimate the success of family planning through existing indicators based on factors that are thought to influence where this approach uses longitudinal data with a truncated spline nonparametric regression approach.

2. Literature Review

2.1 Nonparametric Regression

Nonparametric regression is one approach that explains the functional relationship between the response variable and the predictor variable if the regression curve is not assumed to have a certain shape [8].

$$y_i = m(x_i) + \varepsilon_i \, , i = 1,2,3, \dots \dots , n \quad (1)$$

y_i is the response variable, $m(x_i)$ is a regression function assuming the shape of the curve does not form a certain pattern, and x_i is a predictor variable, whereas ε_i is a random error with mean 0 dan variance σ^2. The extension of the nonparametric regression equation (1) is a multipredictor nonparametric regression with the following general form:

$$y_i = \sum_{j=1}^{p} m(x_{ij}) + \varepsilon_i \, , i = 1,2,3, \dots , n \, ; j = 1,2,3, \dots . , p \quad (2)$$

Several smoothing points are used to estimate the regression function $m(x_i)$ in equation (2) include kernel, local linear, local polynomial, spline, and fourier series [9-12].

2.2 Truncated Spline Nonparametric Regression

The Spline Estimator has good ability to estimate data functions that have different behavior in sub-intervals different. Functions that have different behavior in each sub-interval are connected to each other by points called knot points [13]. In general, truncated spline functions are p-order with knot points $\tau_1, \tau_2, \dots, \tau_k$ is any function that can be represented in the following form [14]:

$$m(x) = \beta_0 + \beta_1 x + \dots + \beta_p x^p + \sum_{r=1}^{k} \beta_{p+r} (x - \tau_r)_+^p \quad (3)$$

β is regression coefficient, and

$$(x - \tau_r)_+^p = \begin{cases} (x - \tau_r)^p & , \; x \geq \tau_r \\ 0 & , \; x < \tau_r \end{cases}$$

2.3 Estimation of Truncated Spline Nonparametric Regression Parameters

Suppose there are n observations $\{x_i, y_i\}_{i=1}^n$ which satisfies equation (1) with $m(x_i)$ is a truncated spline function that has been described in equation (3). An estimate of the $m(x)$ curve function can be obtained by estimating the coefficients $\underset{\sim}{\beta} = \left(\beta_0, \beta_1, \dots, \beta_{p+1}, \beta_{p+2}, \dots, \beta_{(p+r)1}\right)^T$ with $\underset{\sim}{y} = (y_1, y_2 \, , \dots \, , \, y_n)^T$ and

$$X = \begin{bmatrix} 1 & x_1 & x_1^p & (x_1 - \tau_1)_+^p & \dots & (x_1 - \tau_k)_+^p \\ 1 & x_2 & x_2^p & (x_2 - \tau_1)_+^p & \dots & (x_2 - \tau_k)_+^p \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^p & (x_n - \tau_1)_+^p & \dots & (x_n - \tau_k)_+^p \end{bmatrix} \quad (4)$$

Then the estimation of $\underset{\sim}{\beta}$ can be obtained using the least squares method [14] so that the following parameter estimates are obtained:

$$\underset{\sim}{\hat{\beta}} = (\boldsymbol{X^T X})^{-1} \mathbf{X^T} \underset{\sim}{y}$$

$\underset{\sim}{y}$ can be obtained by using the following approach:

$$\underset{\sim}{\hat{y}} = X[X^T X]^{-1}[X^T \underset{\sim}{y}] = A(\lambda) \underset{\sim}{y} \quad (5)$$

with $A(\lambda) = X[X^T X]^{-1} X^T$; λ is a smoothing parameter with order (p), number of knots (k), and knot point $(\tau_1, \tau_2, \dots, \tau_k)$, can be written in notation $h = \big(p, k, (\tau_1, \tau_2, \dots, \tau_k)\big)$.

2.4 Generalized Cross Validation (GCV)

In nonparametric regression, one of the criteria for determining the optimum smoothing parameter is Generalized Cross Validation (GCV) ([14], [15]).

$$GCV(\lambda) = \frac{MSE(\lambda)}{(n^{-1} tr[I - A(\lambda)])^2} \quad (6)$$

$MSE(\lambda) = n^{-1} \sum_{i-1}^{n} (y_i - \bar{y})^2$, $A(\lambda)$ obtained from the relationship $\hat{y} = A(\lambda) y$, I is the identity matrix, and n is the number of observations. The optimal knot value is given by the smallest GCV value.

In nonparametric biresponse regression, the random error vector for each response $\underset{\sim}{\varepsilon}^{(1)}$, $\underset{\sim}{\varepsilon}^{(2)}$ correlated with each other, so to estimate the parameter $\underset{\sim}{\beta}$ the Weighted Least Square (WLS) method is used. WLS is used to minimize the number of weighted error squares [16] which is formulated as follows:

$$\underset{\sim}{\varepsilon}^T \underset{\sim}{W} \underset{\sim}{\varepsilon} = (\underset{\sim}{\mathrm{y}} - \underset{\sim}{\mathbf{X}\beta})^T \mathbf{W} (\underset{\sim}{y} - \underset{\sim}{\boldsymbol{X}\beta}) \quad (7)$$

$\underset{\sim}{\beta}$ is a WLS estimator and the weighting W is the inverse of the variance-covariance matrix of $\underset{\sim}{\varepsilon}$ or $\underset{\sim}{y}$ with the condition X, which is denoted by $Var(\underset{\sim}{\varepsilon} | \boldsymbol{X}) = Var(\underset{\sim}{y} | \boldsymbol{X}) = \Sigma$

Equation (7) is then derived from $\underset{\sim}{\beta}$ so that the WLS estimator is obtained as follows:

$$\underset{\sim}{\hat{\beta}} = (\boldsymbol{X^T W^{-1} X})^{-1} \mathbf{X^T} \boldsymbol{W^{-1}} \underset{\sim}{y} \quad (8)$$

2.5 Longitudinal Data

Longitudinal data assumes that the subjects measured are independent of each other so there is no correlation, but the observations within the subjects are mutually dependent so there is a correlation [17]. When viewed based on the regression approach, nonparametric regression is known to be more adaptive to the data and does not require certain assumptions as in the parametric regression approach, so nonparametric regression is a good alternative for handling longitudinal data [18] which cannot be analyzed accurately through a parametric approach [19]. In longitudinal data, if there is more than one response, then there must be a correlation of each response on the same subject.

2.6 Truncated Spline Nonparametric Regression on Longitudinal Data

The nonparametric regression equation on longitudinal data assumes paired data x_{ij}, y_{ij} and fulfill the following equation:

$$\underset{\sim}{y_{ij}} = \underset{\sim}{m}(x_{ij}) + \underset{\sim}{\varepsilon_{ij}} \tag{9}$$

$\underset{\sim}{m}(x_{ij}) = \left(m^{(1)}(x_{ij}) \;\; m^{(2)}(x_{ij})\right)^T$ and $\varepsilon_{ij} = \left({\varepsilon_{ij}}^{(1)} \;\; {\varepsilon_{ij}}^{(2)}\right)$ is a random error with mean 0 and variance $\sum_i$, $i = 1,2,\ldots,n$ represents the index for the observed subjects and $j = 1,2,\ldots,t_i$ represents the index for observations in each subject. Nonparametric regression on longitudinal data based on the truncated weighted spline estimator as follows:

$$\begin{aligned} \underset{\sim}{\hat{y}} &= \boldsymbol{X}\underset{\sim}{\hat{\beta}} \\ &= (\boldsymbol{X}^T\mathbf{W}\boldsymbol{X})^{-1}\mathbf{X}^{-\mathbf{T}}\boldsymbol{W}\underset{\sim}{y} \\ &= \boldsymbol{A}\underset{\sim}{y} \end{aligned} \tag{10}$$

$\boldsymbol{A} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}$ and Mean Square Error (MSE):

$$MSE = (2T)^{-1}[(\underset{\sim}{y} - \underset{\sim}{\hat{y}})^T(\underset{\sim}{y} - \underset{\sim}{\hat{y}})] \tag{11}$$

Furthermore, the MSE formula can be derived in nonparametric regression on longitudinal data and the MSE formula is obtained as follows:

$$MSE = \frac{1}{2T}\underset{\sim}{y}^T(\boldsymbol{I} - \boldsymbol{A})^T(\boldsymbol{I} - \boldsymbol{A})\underset{\sim}{y} \tag{12}$$

the MSE value is then used to calculate the GCV value. The GCV value in nonparametric regression on longitudinal data is formulated in the following equation:

$$GCV(\lambda) = \frac{2T^{-1}\underset{\sim}{y}^T(\boldsymbol{I}-\boldsymbol{A})^T(\boldsymbol{I}-\boldsymbol{A})\underset{\sim}{y}}{[2T^{-1}(\boldsymbol{I}-\boldsymbol{A})]^2} \tag{13}$$

Furthermore, the model suitability test is carried out by calculating the Goodness of Fit criteria, namely MSE and R^2 with

$$R^2 = 1 - \frac{SSE}{SST} \tag{14}$$

$SSE = (\underset{\sim}{y} - \underset{\sim}{\hat{y}})^T(\underset{\sim}{y} - \underset{\sim}{\hat{y}})$, and

$SST = (\underset{\sim}{y} - \underset{\sim}{\bar{y}})^T(\underset{\sim}{y} - \underset{\sim}{\bar{y}})$.

3. Methodology

The data used in this study is secondary data from Central Bureau of Statistics or Badan Pusat Statistik (BPS) relating to the success of family planning and indicators that are thought to have an influence from various aspects. The unit of observation used in the study were 34 provinces in Indonesia for five years from 2017 to 2021. The research variables used consisted of one response variable with three predictor variables in longitudinal data. The research variables are presented in Table 1.

Table 1. Research Variable

Variable	Variable Description
Y_1	Percentage of Women Aged 15-49 Years and Married Status Using Contraceptives / Contraceptive Prevalence Rate (CPR)
X_1	Poverty Depth Index (P1) by Province and Region
X_2	Average Years of Schooling of Population Age ≥ 15 Years by Province
X_3	Proportion of women aged 20-24 years who are married or living together before the age of 18 by province

The scope of research is limited to 34 provinces in Indonesia in 2017-2021, where the analysis will be divided into three groups based on the division of the BKKBN as in Table 2.

Table 2. Provincial Groups based on BKKBN

No.	Java-Bali	Outer Java-Bali I	Outer Java-Bali II
1	DKI Jakarta	Aceh	Riau
2	Jawa Barat	Sumatera Utara	Jambi
3	Jawa Tengah	Sumatera Barat	Bengkulu
4	DI Yogyakarta	Sumatera Selatan	NTT
5	Jawa Timur	Lampung	Kalimantan Tengah
6	Banten	NTB	Kalimantan Timur
7	Bali	Kalimantan Barat	Sulawesi Tengah
8		Kalimantan Selatan	Sulawesi Tenggara
9		Sulawesi Utara	Maluku
10		Sulawesi Selatan	Maluku Utara
11		Bangka Belitung	Papua Barat
12		Gorontalo	Papua
13		Sulawesi Barat	Kepualuan Riau
14			Kalimantan Utara

The research flow chart is depicted in Figure 1 as follows:

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 6 Issue 12, December - 2022, Pages: 151-159

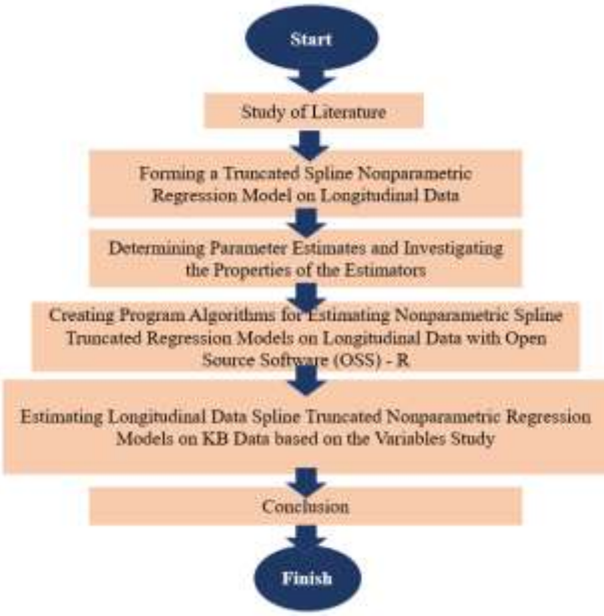

Fig. 1. Research Flowchart

4. Results and Discussion

To provide an overview or initial information related to the relationship between the response variable and each predictor variable, a scatter plot is made. This scatter plot will provide information about the pattern of the regression curve shape that will be used in modeling.

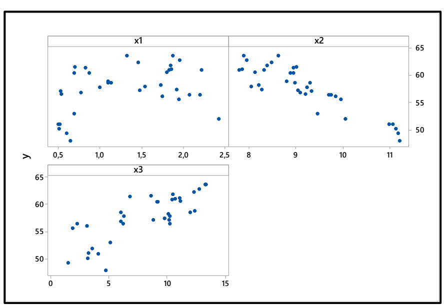

Fig. 2. Scatter Plot between Response Variables and Each Predictor Variable in the Java Bali Group

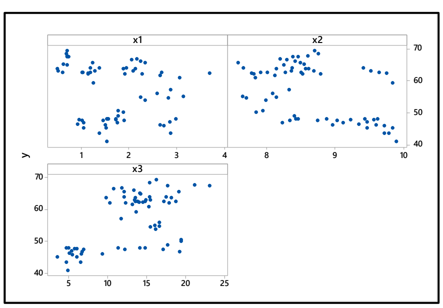

Fig. 3. Scatter Plot between Response Variables and Each Predictor Variable in the Outer Java Bali I Group

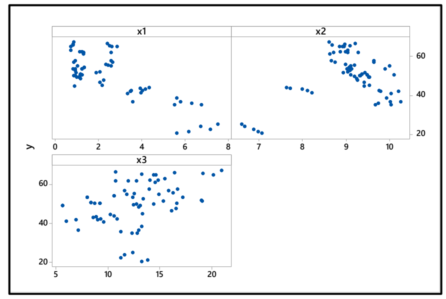

Fig. 4. Scatter Plot between Response Variables and Each Predictor Variable in the Outer Java Bali II Group

Based on Figure 2, Figure 3, and Figure 4 it is known that the data plots are spread out and do not form a certain pattern so that a nonparametric regression approach is an approach that can be offered. In this study, nonparametric regression was used because it involved response variables with a truncated spline approach for longitudinal data. Determination of optimal knots using the GCV method using one, two, and three knot points. The criteria for the goodness of the model used is MSE. In this application, the weighting matrix model used is the number of observations in each subject. The weight matrix **V** is defined as $\mathbf{V} = \left[diag\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right)\right]$. Analysis was performed for each group with the aim of modeling each subject in the group. The following are the results of modeling in each group.

4.1 Model in the Java-Bali group

In the truncated spline nonparametric regression model approach, it is known that there is a point called the knot point, which is a joint point where there is a change in the pattern of behavior in the function. In a plot between the response variable and the predictor variable, several cuts or segments can be made based on the knot points. The location of the knot point and the number of knots is very important. The GCV method is used to determine the location of the optimal knot point in each variable. The number of knots used varies, namely one knot point, two knots, and three knots for each variable. Following are the knot points, GCV values for each variable in each subject using **V** weights.

Table 3. Summary of Smallest GCV for 1 Knot of the Java-Bali Group

No.	Province	Knot			GCV
		X_1	X_2	X_3	
1	DKI Jakarta	0,58	11,10	3,30	
2	Jawa Barat	1,31	8,79	11,90	
3	Jawa Tengah	1,92	8,05	10,49	
4	DIY	2,13	9,89	4,33	$3{,}42\text{x}10^{-24}$
5	Jawa Timur	1,89	8,16	12,09	
6	Banten	0,95	9,11	7,75	
7	Bali	0,62	9,23	7,99	

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 6 Issue 12, December - 2022, Pages: 151-159

Table 4. Summary of Smallest GCV for 2 Knot of the Java-Bali Group

No.	Province	Knot			GCV
		X_1	X_2	X_3	
1	DKI Jakarta	0,55	11,06	2,77	$1{,}24\text{x}10^{-25}$
		0,62	11,17	4,22	
2	Jawa Barat	1,25	8,69	11,38	
		1,42	8,95	12,81	
3	Jawa Tengah	1,81	7,97	10,28	
		2,11	8,19	10,86	
4	DIY	2,02	9,83	3,61	
		2,32	9,99	5,58	
5	Jawa Timur	1,86	8,07	11,62	
		1,93	8,30	12,91	
6	Banten	0,89	9,04	7,25	
		1,04	9,23	8,62	
7	Bali	0,59	9,14	7,15	
		0,67	9,38	9,45	

Table 5. Summary of Smallest GCV for 3 Knot of the Java-Bali Group

No.	Province	Knot			GCV
		X_1	X_2	X_3	
1	DKI Jakarta	0,49	10,97	1,52	$1{,}11\text{x}10^{-26}$
		0,58	11,11	3,43	
		0,62	11,17	4,22	
2	Jawa Barat	1,10	8,47	10,15	
		1,32	8,81	12,03	
		1,42	8,95	12,81	
3	Jawa Tengah	1,54	7,78	9,78	
		1,95	8,07	10,54	
		2,11	8,19	10,86	
4	DI Yogyakarta	1,75	9,69	1,92	
		2,16	9,90	4,51	
		2,32	9,99	5,58	
5	Jawa Timur	1,80	7,88	10,50	
		1,89	8,18	12,20	
		1,93	8,30	12,91	
6	Banten	0,77	8,88	6,06	
		0,96	9,13	7,87	
		1,04	9,23	8,62	
7	Bali	0,52	8,94	5,16	
		0,62	9,25	8,19	
		0,67	9,38	9,45	

Table 3, Table 4, and Table 5 show the knot points formed for each smallest GCV value. These results are obtained from the running program. Each running program will generate many alternative knot points with various GCV values. After knowing the GCV value of each knot point, the next step is to choose the smallest GCV value between one, two and three knots. It is known that the smallest GCV is $1{,}11\text{x}10^{-26}$. It is this smallest GCV value that produces the optimum knot points as in Table 5, in general the running program produces 17.296 alternative combinations of knot points with each variable having three knot points. Because the smallest GCV is indicated by GCV with 3 knots, the knot points will be used in modeling. The next step is to use optimal knot points to be modeled to obtain model parameter estimates. The results of the program are all parameter estimates to form a model for each subject. The Java-Bali group consists of seven provinces or seven subjects so that all parameter estimates for the seven provinces will be produced. As an example, the following parameter estimates are generated for the DKI Jakarta Province model as shown in Table 6.

Table 6. Parameter Estimation for DKI Jakarta Province

Par	Value	Par	Value	Par	Value
β_0	1,432	β_5	-10,984	β_9	0,662
β_1	-0,631	β_6	-0,795	β_{10}	-0,232
β_2	4,405	β_7	-0,572	β_{11}	-0,336
β_3	0,901	β_8	-0,541	β_{12}	-5,006
β_4	-1,122				

The model formed for DKI Jakarta Province is:

$$\begin{aligned}\hat{Y}_{\mathrm{i}} = {} & 1{,}432 - 0{,}631X_{1i} - 1{,}122(X_{1i} - 0{,}49)_+ \\ & -0{,}572(X_{1i} - 0{,}58)_+ - 0{,}232(X_{1i} - 0{,}62)_+ + 4{,}405X_{2i} \\ & -10{,}984(X_{2i} - 10{,}97)_+ - 0{,}541(X_{2i} - 11{,}11)_+ \\ & -0{,}336(X_{2i} - 11{,}17)_+ + 0{,}901X_{3i} - 0{,}795(X_{3i} - 1{,}52)_+ \\ & +0{,}662(X_{3i} - 3{,}43)_+ - 5{,}006(X_{3i} - 4{,}22)_+\end{aligned}$$

This model is the best model with a GCV value of $1{,}11\text{x}10^{-26}$ and an MSE value of $4{,}265\text{x}10^{-22}$. Models for other provinces in the Java-Bali group can also be written by considering estimated parameter values and knot points for each of these provinces.

4.2 Model in the Outer Java-Bali I Group

Following are the knot points, GCV values for each variable in each subject using **V** weights in the Outer Java Bali I group as in Table 7, Table 8, and Table 9.

Table 7. Summary of Smallest GCV for 1 Knot Group Outer Java Bali I

No.	Province	Knot			GCV
		X_1	X_2	X_3	
1	Aceh	2,88	9,66	5,98	$7{,}38\text{x}10^{-25}$
2	Sumatera Utara	1,61	9,78	5,99	
3	Sumatera Barat	1,00	9,33	5,80	
4	Sumatera Selatan	2,25	8,67	13,08	
5	Lampung	2,07	8,45	11,39	
6	NTB	2,86	7,98	16,26	
7	Kalimantan Barat	1,16	7,87	17,47	
8	Kalimantan Selatan	0,70	8,63	20,73	
9	Sulawesi Utara	1,30	9,70	14,79	
10	Sulawesi Selatan	1,64	8,79	13,07	
11	Bangka Belitung	0,63	8,41	17,32	

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 6 Issue 12, December - 2022, Pages: 151-159

12	Gorontalo	3,36	8,15	14,17	
13	Sulawesi Barat	1,85	8,22	18,72	

Table 8. Summary of Smallest GCV for 2 Knot Group Outer Java Bali I

No.	Province	Knot			GCV
		X_1	X_2	X_3	
1	Aceh	2,88	9,67	6,02	4,91x10^{-26}
		2,94	9,73	6,35	
2	Sumatera Utara	1,61	9,79	6,02	
		1,67	9,84	6,29	
3	Sumatera Barat	1,01	9,33	5,87	
		1,03	9,41	6,42	
4	Sumatera Selatan	2,26	8,67	13,11	
		2,30	8,73	13,35	
5	Lampung	2,08	8,45	11,43	
		2,13	8,51	11,81	
6	NTB	2,88	7,99	16,29	
		3,02	8,07	16,47	
7	Kalimantan Barat	1,17	7,88	17,58	
		1,20	7,95	18,43	
8	Kalimantan Selatan	0,71	8,63	20,89	
		0,71	8,69	22,16	
9	Sulawesi Utara	1,30	9,71	14,83	
		1,34	9,78	15,12	
10	Sulawesi Selatan	1,64	8,80	13,19	
		1,69	8,89	14,09	
11	Bangka Belitung	0,63	8,42	17,41	
		0,67	8,49	18,18	
12	Gorontalo	3,38	8,16	14,25	
		3,55	8,25	14,84	
13	Sulawesi Barat	1,85	8,23	18,77	
		1,87	8,32	19,15	

Table 9. Summary of Smallest GCV for 3 Knot Group Outer Java Bali I

No.	Province	Knot			GCV
		X_1	X_2	X_3	
1	Aceh	2,65	9,43	4,64	1,05X10^{-26}
		2,83	9,62	5,74	
		2,90	9,69	6,14	
2	Sumatera Utara	1,38	9,56	4,85	
		1,56	9,74	5,78	
		1,63	9,81	6,12	
3	Sumatera Barat	0,92	9,03	3,55	
		0,99	9,27	5,39	
		1,01	9,36	6,08	
4	Sumatera Selatan	2,09	8,42	12,10	
		2,22	8,62	12,90	
		2,27	8,70	13,20	
5	Lampung	1,89	8,20	9,82	
		2,04	8,40	11,10	
		2,10	8,48	11,58	
6	NTB	2,26	7,65	15,50	
		2,75	7,92	16,13	
		2,93	8,02	16,36	
7	Kalimantan Barat	1,01	7,58	13,95	
		1,14	7,82	16,83	
		1,18	7,90	17,90	
8	Kalimantan Selatan	0,67	8,38	15,46	
		0,70	8,58	19,77	
		0,71	8,66	21,36	
9	Sulawesi Utara	1,14	9,41	13,58	
		1,27	9,65	14,57	
		1,32	9,73	14,94	
10	Sulawesi Selatan	1,46	8,43	9,36	
		1,60	8,72	12,40	
		1,66	8,83	13,52	
11	Bangka Belitung	0,49	8,14	14,15	
		0,60	8,36	16,74	
		0,65	8,45	17,70	
12	Gorontalo	2,66	7,78	11,71	
		3,23	8,08	13,73	
		3,45	8,20	14,47	
13	Sulawesi Barat	1,75	7,85	17,17	
		1,83	8,15	18,44	
		1,86	8,27	18,91	

The smallest GCV is shown by the GCV with 3 knots so that the knot points will be used in modeling. Table 10 contains the resulting parameter estimates for the Aceh Province model.

Table 10. Parameter Estimation for Aceh Province

Par	Value	Par	Value	Par	Value
β_0	0,816	β_5	-4,587	β_9	-3,143
β_1	1,792	β_6	-1,379	β_{10}	9,886
β_2	2,178	β_7	13,787	β_{11}	-5,858
β_3	4,027	β_8	-4,013	β_{12}	-1,646
β_4	-0,393				

Model formed for Aceh Province

$$\begin{aligned}\hat{Y}_i = {} & 0{,}816 + 1{,}792X_{1i} - 0{,}393(X_{1i} - 2{,}65)_+ \\ & + 13{,}787(X_{1i} - 2{,}83)_+ + 9{,}886(X_{1i} - 2{,}90)_+ + 2{,}178X_{2i} \\ & - 4{,}587(X_{2i} - 9{,}43)_+ - 4{,}013(X_{2i} - 9{,}62)_+ \\ & - 5{,}858(X_{2i} - 9{,}69)_+ + 4{,}027X_{3i} - 1{,}379(X_{3i} - 4{,}64)_+ \\ & - 3{,}143(X_{3i} - 5{,}74)_+ - 1{,}646(X_{3i} - 6{,}14)_+\end{aligned}$$

This model is the best model with a GCV value of 1,05x10^{-26} and an MSE value of 4,783x10^{-21}. Models for other provinces in the Java-Bali group can also be written by considering estimated parameter values and knot points for each of these provinces.

4.3 Model in the Outer Java-Bali II Group

Following are the knot points, GCV values for each variable in each subject using **V** weights in the Outer Java Bali II group as shown in Table 11, Table 12, and Table 13.

Table 11. Summary of Smallest GCV for 1 Knot Group Outer Java Bali II

No.	Province	Knot			GCV
		X_1	X_2	X_3	
1	Riau	1,24	9,38	9,00	$1{,}33 \times 10^{-25}$
2	Jambi	1,24	8,90	13,52	
3	Bengkulu	2,71	9,15	13,41	
4	NTT	4,21	8,02	9,13	
5	Kalimantan Tengah	0,80	8,90	19,27	
6	Kalimantan Timur	1,11	9,95	12,29	
7	Sulawesi Tengah	2,55	9,01	15,38	
8	Sulawesi Tenggara	2,12	9,34	17,30	
9	Maluku	3,50	10,09	9,59	
10	Maluku Utara	0,92	9,35	15,95	
11	Papua Barat	6,36	9,92	12,58	
12	Papua	6,92	6,91	12,99	
13	Kepualuan Riau	0,98	10,26	5,96	
14	Kalimantan Utara	1,10	9,31	14,61	

Table 12. Summary of Smallest GCV for 2 Knot Group Outer Java Bali II

No.	Province	Knot			GCV
		X_1	X_2	X_3	
1	Riau	1,16	9,23	7,38	$1{,}699 \times 10^{-26}$
		1,26	9,42	9,40	
2	Jambi	1,17	8,76	12,18	
		1,26	8,94	13,86	
3	Bengkulu	2,57	9,04	12,12	
		2,75	9,18	13,73	
4	NTT	4,07	7,83	7,63	
		4,24	8,07	9,50	
5	Kalimantan Tengah	0,75	8,75	17,48	
		0,81	8,93	19,71	
6	Kalimantan Timur	0,99	9,79	10,57	
		1,14	9,98	12,72	
7	Sulawesi Tengah	2,44	8,84	14,03	
		2,57	9,06	15,72	
8	Sulawesi Tenggara	2,00	9,15	15,40	
		2,15	9,39	17,77	
9	Maluku	3,42	9,93	8,30	
		3,52	10,14	9,92	
10	Maluku Utara	0,87	9,19	14,60	
		0,93	9,40	16,29	
11	Papua Barat	5,95	9,80	11,91	
		6,46	9,95	12,74	
12	Papua	6,30	6,75	12,15	
		7,07	6,94	13,20	
13	Kepualuan Riau	0,84	10,14	4,51	
		1,02	10,29	6,32	
14	Kalimantan Utara	0,98	9,21	12,51	
		1,13	9,33	15,13	

Table 13. Summary of Smallest GCV for 3 Knot Group Outer Java Bali II

No.	Province	Knot			GCV
		X_1	X_2	X_3	
1	Riau	1,07	9,08	5,75	$3{,}199 \times 10^{-27}$
		1,14	9,21	7,17	
		1,29	9,46	9,91	
2	Jambi	1,11	8,63	10,84	
		1,17	8,75	12,01	
		1,28	8,98	14,28	
3	Bengkulu	2,42	8,92	10,84	
		2,55	9,02	11,96	
		2,79	9,22	14,13	
4	NTT	3,93	7,64	6,14	
		4,05	7,81	7,45	
		4,29	8,13	9,97	
5	Kalimantan Tengah	0,71	8,61	15,69	
		0,75	8,73	17,26	
		0,82	8,98	20,27	
6	Kalimantan Timur	0,87	9,64	8,85	
		0,97	9,77	10,36	
		1,17	10,03	13,26	
7	Sulawesi Tengah	2,34	8,66	12,68	
		2,43	8,82	13,86	
		2,60	9,11	16,14	
8	Sulawesi Tenggara	1,88	8,95	13,50	
		1,99	9,12	15,16	
		2,19	9,45	18,37	
9	Maluku	3,33	9,76	7,00	
		3,40	9,91	8,14	
		3,55	10,19	10,32	
10	Maluku Utara	0,82	9,02	13,26	
		0,86	9,17	14,44	
		0,95	9,45	16,71	
11	Papua Barat	5,54	9,68	11,24	
		5,90	9,79	11,83	
		6,59	9,99	12,95	
12	Papua	5,68	6,60	11,31	
		6,22	6,73	12,05	
		7,27	6,99	13,47	
13	Kepualuan Riau	0,71	10,02	3,07	
		0,83	10,12	4,33	
		1,06	10,33	6,77	
14	Kalimantan Utara	0,86	9,11	10,42	
		0,96	9,20	12,25	
		1,16	9,36	15,79	

The smallest GCV is shown by the GCV with 3 knots so that the knot points will be used in modeling. Table 14 contains the resulting parameter estimates for the Riau Province model.

Table 14. Parameter Estimation for Riau Province

Par	Value	Par	Value	Par	Value
β_0	1,406	β_5	-9,764	β_9	1,790
β_1	6,379	β_6	-4,085	β_{10}	-0,244

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 6 Issue 12, December - 2022, Pages: 151-159

β_2	3,143	β_7	3,391	β_{11}	0,705
β_3	2,878	β_8	-2,346	β_{12}	-4,671
β_4	4,932				

The model formed for Riau Province:

$$\begin{aligned}\hat{Y}_{\mathrm{i}} = {} & 1{,}406 + 6{,}379X_{1i} + 4{,}932(X_{1i} - 1{,}07)_+ \\ & +3{,}391(X_{1i} - 1{,}14)_+ - 0{,}244(X_{1i} - 1{,}29)_+ + 3{,}143X_{2i} \\ & -9{,}764(X_{2i} - 9{,}08)_+ - 2{,}346(X_{2i} - 9{,}21)_+ \\ & +0{,}705(X_{2i} - 9{,}46)_+ + 2{,}878X_{3i} - 4{,}085(X_{3i} - 5{,}75)_+ \\ & +1{,}790(X_{3i} - 7{,}17)_+ - 4{,}671(X_{3i} - 9{,}91)_+\end{aligned}$$

This model is the best model with a GCV value of $3{,}199 \times 10^{-27}$ and an MSE value of $1{,}966 \times 10^{-21}$. Models for other provinces in the Java-Bali group can also be written by considering estimated parameter values and knot points for each of these provinces.

4.4 Discussion of Spline Truncated Nonparametric Regression Model for Longitudinal Data on KB Data

The benefit of the resulting model is to predict the percentage of CPR from each province in Indonesia. Examples of the model formed for DKI Jakarta Province are,

$$\begin{aligned}\hat{Y}_{\mathrm{i}} = {} & 1{,}432 - 0{,}631X_{1i} - 1{,}122(X_{1i} - 0{,}49)_+ \\ & -0{,}572(X_{1i} - 0{,}58)_+ - 0{,}232(X_{1i} - 0{,}62)_+ + 4{,}405X_{2i} \\ & -10{,}984(X_{2i} - 10{,}97)_+ - 0{,}541(X_{2i} - 11{,}11)_+ \\ & -0{,}336(X_{2i} - 11{,}17)_+ + 0{,}901X_{3i} - 0{,}795(X_{3i} - 1{,}52)_+ \\ & +0{,}662(X_{3i} - 3{,}43)_+ - 5{,}006(X_{3i} - 4{,}22)_+\end{aligned}$$

With segmented formed for each predictor variable, namely the poverty depth index variable (X_1)

$$\hat{Y} = \begin{cases} 1{,}432 - 0{,}631X_1 \ ; & X_1 < 0{,}49 \\ 1{,}982 - 1{,}753X_1 \ ; & 0{,}49 \le X_1 < 0{,}58 \\ 2{,}314 - 2{,}325X_1 \ ; & 0{,}58 \le X_1 < 0{,}62 \\ 2{,}458 - 2{,}577X_1 \ ; & X_1 \ge 0{,}62 \end{cases}$$

Variable average length of school (X_2)

$$\hat{Y} = \begin{cases} 1{,}432 + 4{,}405X_2 \ ; & X_2 < 10{,}97 \\ 121{,}917 - 6{,}579X_2 \ ; & 10{,}97 \le X_1 < 11{,}11 \\ 127{,}928 - 7{,}12X_2 \ ; & 11{,}11 \le X_1 < 11{,}17 \\ 131{,}681 - 7{,}456X_2 \ ; & X_1 \ge 11{,}17 \end{cases}$$

Variable proportion of married women (X_3)

$$\hat{Y} = \begin{cases} 1{,}432 + 0{,}901X_3 \ ; & X_1 < 1{,}52 \\ 2{,}640 + 0{,}106X_3 \ ; & 1{,}52 \le X_1 < 3{,}43 \\ 4{,}911 + 0{,}768X_3 \ ; & 3{,}43 \le X_1 < 4{,}22 \\ 7{,}024 + 0{,}267X_3 \ ; & X_1 \ge 4{,}22 \end{cases}$$

To make it easier to interpret the model, a description will be made for each segment or piece that is formed. The following is an example of an interpretation of the model formed in DKI Jakarta Province. The model interpretation of the poverty depth index variable (X_1) assuming other variables are constant is as follows.

From the model above, it is known that when the poverty depth index is less than 0,49, every 1-point increase in the poverty depth index tends to decrease the CPR percentage by 0,63%. When the poverty depth index is between 0,49 and 0,58, every 1-point increase in the poverty depth index tends to reduce the CPR percentage by 1,75%. When the poverty depth index is between 0,58 and 0,62, every 1 increase in the poverty depth index tends to reduce the CPR percentage by 2,32%. Meanwhile, when the poverty depth index is more than 0,62, every 1-point increase in the poverty depth index tends to reduce the CPR percentage by 2,58%. In this case it is known that poverty has an impact on the success of family planning. The greater the poverty index, the less likely it is to participate in the family planning program so that it can reduce the percentage of CPR, which is in accordance with the opinion of previous studies. Interpretation of the influence of other predictor variables can also be explained for each segment.

From the segmentation model that has been formed, it is known that each predictor variable has a different effect on each response variable. In this analysis using longitudinal data, it can explain the more specific effects in each subject. The effects may be the same but may also be different because the characteristics of the subjects, in this case, are different provinces. Table 15 shows an overview of the influence of each variable in each province in the Java-Bali group in each segment.

Table 15. Effect of Predictor Variables in Each Segment

X_1		X_2		X_3	
----	DKI Jakarta	+---	DKI Jakarta, DI Yogyakarta	+-+-	DKI Jakarta
++++	Jawa Barat, Jawa Tengah, Banten, Bali	+-++	Jawa Barat, Jawa Timur, Banten	+--+	Jawa Barat, Bali
+---	DI Yogyakarta	+++-	Jawa Tengah	++--	Jawa Tengah, Banten
++--	Jawa Timur	+--+	Bali	+-++	DI Yogyakarta
				++++	Jawa Timur

5. Conclusion

The conclusion obtained based on the discussion that has been carried out previously, the application of the model to the success data for family planning was carried out in three discussion groups, namely the Java Bali group, Outer Java Bali I, and Outer Java Bali II with the optimum knot used of three knots. The following is a summary of the GCV and MSE values from the best modeling results for each group

Group	GCV	MSE
Java Bali	$1{,}11 \times 10^{-26}$	$4{,}265 \times 10^{-22}$
Outer Java Bali I	$1{,}05 \times 10^{-26}$	$4{,}783 \times 10^{-21}$
Outer Jawa Bali II	$3{,}199 \times 10^{-27}$	$1{,}966 \times 10^{-21}$

Using longitudinal data, it can explain the more specific effects in each subject. The effects may be the same but may also be different because the characteristics of the subjects, in this case, are different provinces.

6. REFERENCES

[1] Budiantara, I.N. (2009). Spline dalam Regresi Nonparametrik dan Semiparametrik, Sebuah Pemodelan Statistika Masa Kini dan Masa mendatang, Pidato Pengukuhan Guru Besar, Institut Teknologi Sepuluh Nopember, Surabaya.

[2] Gujarati, D. N. (2003). Basic Econometric, 4 edition, McGraw-Hill Companies, Inc., New York.

[3] Durban, M., Harezlak, J., Wand, M.P., & Carrol, R.J. (2004). Simple Fitting of Subject-Specific Curves for Longitudinal Data, Statistics in Medicine, hal. 1-24.

[4] Ibrahim, N.A. & Suliadi. (2008). Analyzing Longitudinal Data using GeeSmoothing Spline, Proceedings of The 8 WSEAS International Conference on Applied Computer and Applied Computational Science.

[5] Liang, H. dan Xiao, Y. (2006). Penalized Splines for Longitudinal Data With An Application in AIDS Studies, Journal of Modern Applied Statistical Methods, Vol. 5, No. 1, hal 130-139.

[6] Laome, L. (2009). Model Regresi Semiparametrik Spline untuk Data Longitudinal pada Kasus Kadar CD4 Penderita HIV, Paradigma, Vol. 13, No. 2, hal 189-194.

[7] Melliana, A. (2013). Analisis Statistika Faktor yang Mempengaruhi Indeks Pembangunan Manusia di Kabupaten/Kota Provinsi Jawa Timur dengan Menggunakan Regresi Panel, Tugas Akhir Program Sarjana, Institut Teknologi Sepuluh Nopember, Surabaya.

[8] Hardle, W. (1992). Applied Nonparametric Regression, Cambridge University Press, New York.

[9] Chamidah, N., Rifada, M., & Amelia, D. (2022). A theoretical discussion on modeling the number of COVID-19 death cases using penalized spline negative binomial regression approach, Commun. Math. Biol. Neurosci., Article ID 82.

[10] Chamidah, N., Widyanti, A., Trapsilawati, F., & Syafitri, U.D. (2021). Local Linear Negative Binomial Nonparametric Regression For Predicting The Number Of Speed Violations On Toll Road: A Theoretical Discussion, Commun. Math. Biol. Neurosci, 10, pp.1-11.

[11] Chamidah, N., & Lestari, B. (2019). Estimation of covariance matrix using multi-response local polynomial estimator for designing children growth charts: A theoretically discussion, Journal of Physics: Conference Series, Vol. 1397 012072.

[12] Chamidah, N., Selvira, D., Febriana, R.A., Ariyanto, & Reiza, S. (2021). Fourier Series Estimator for Predicting International Market Price of White Sugar, AIP Conference Proceedings 2329, 060035.

[13] Chamidah, N., Lestari, B., Massaid, A., & Saifudin, T. (2020). Estimating Mean Arterial Pressure Affected By Stress Scores Using Spline Nonparametric Regression Model Approach, Commun. Math. Biol. Neurosci. 2020(72), pp.1-12.

[14] Eubank, R.L. (1999). Nonparametric Regression and Spline Smoothing2nd Edition, Marcel Dekker, Inc., New York.

[15] Wahba, G. (1990). Spline Models for Observation Data, University of Winsconsin at Madison, Pensylvania.

[16] Hidayati, L., Chamidah, N., & Budiantara, I.N. (2020). Confidence Interval of Multiresponse Semiparametric Regression Model Parameters Using Truncated Spline, International Journal of Academic Applied Research (IJAAR) ISSN : 2643-9603, Vol. 4(1) pp.14 -18.

[17] Wu, H., & Zhang, J.T. (2006). Nonparametric Regression Method for Longitudinal Data Analysis, John Wiley & Sons, New Jersey.

[18] Wang, J.L. (2003). Nonparametric Regression Analysis of Longitudinal Data, Departmen of Statistics, University of California, Davis.

[19] Islamiyati, A., Fatmawati, Chamidah, N. (2020). Penalized Spline Estimator with Multi-Smoothing Parameters in Bi-Response Multi-Predictor Nonparametric Regression Model For Longitudinal Data, Songklanakarin J. Sci. Technol. 42 (4), 897-909.