Proceeding of ICSA 2019, p: 69-82
ISBN 978-979-19256-3-1 (PDF)

Truncated Spline Estimation of Percentage Poverty Modeling in Papua Province

Ni Putu Ayu Mirah Mariati[1,2], I Nyoman Budiantara[1*], Vita Ratnasari[1]

[1]*Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia*

[2]*Universitas Mahasaraswati, Denpasar Bali, 80233, Indonesia*

[*)]Corresponding author: nyomanbudiantara65@gmail.com

Abstract

In estimating the regression curve there are three approaches, namely parametric regression, nonparametric regression and semiparametric regression. Nonparametric regression approach has high flexibility. Nonparametric regression approach that is quite popular is Truncated Spline. Truncated Spline is a polynomial pieces which have segmented and continuous. One of the advantages of Spline is that it can handle data that changes at certain sub intervals, so this model tends to search for data estimates wherever the data pattern moves and there are points of knots. In reality, data patterns often change at certain sub intervals, one of which is data on poverty in the Papua Province. Papua Province is ranked first in the percentage of poor people in Indonesia. The best of model Truncated Spline in nonparametric regression for the poverty model in Papua Province is using a combination of knot.

***Keywords*:** nonparametric regression, poverty, truncated spline.

Introduction

Modeling for one or more variables, the first thing that should be done is whether the variable is rationally correlated or not. If rationally happen correlation, then do modeling Statistics using regression analysis. Regression analysis is one of the Statistics analysis is used to investigate the pattern of functional relationships between one or more variables. Regression analysis to estimate the regression curve there are three approaches, namely parametric regression, nonparametric regression and semiparametric regression. In parametric regression approach there is an assumption that is very strong and rigid that the curve shape is known for example linear regression, quadratic, cubic, polynomial of degree p and others. The main objective in the regression analysis is to find forms of regression curve estimation. Nonparametric regression curves are only assumed to be smooth in the sense contained in a particular function space. The nonparametric regression approach that is quite popular is Spline. Spline is a form of estimator which is also often used in nonparametric regression because it has good visual interpretation, is flexible, and is able to handle character functions that are smooth (Eubank, 1988); (Budiantara, et al., 2015).

Spline as the data pattern approach introduced by Whittaker in 1923. Spline popularized by Schoenberg in 1942. While the Spline based on an optimization problem was developed by Reinsch in 1967 (Wahba, 1990). Spline are pieces of polynomial that are segmented and continuous. The advantage of Spline is that this model tends to look for data estimates wherever the data pattern moves, can describe changes in behavior patterns of functions at certain sub intervals (Liang, 2006). In addition, it also has advantages in overcoming data patterns that show sharp up/down with the help of knot and the resulting curve is relatively smooth. This advantage occurs because in Spline there are knots, which are joint fusion points that indicate changes in data behavior patterns (Eubank R., 1999). With these knot, Spline can provide flexibility that is better than polynomials, making it possible to adjust effectively to local characteristics of a function or data. Spline approach has a function basis. The basic functions commonly used include Truncated Spline. Truncated Spline is a function where there are changes in the behavior patterns of different curves at different intervals.

In daily life data is often found to change at certain sub-intervals, one of which is poverty data. Poverty is a situation where when a person or group of people cannot meet the needs of its economy to achieve prosperity and welfare. Poverty is multidimensional which has meaning because humans have many needs and various kinds so that poverty has a variety of aspects, namely the primary aspect which consists of assets, social political organization, knowledge and skills (Budiantara, et al., 2019). Poverty Alleviation is the goal of Sustainable development goals. One of the countries that have problems with poverty percentage is Indonesia. Indonesia is a developing country with a very large population, with the fourth largest population in the world. There is a close relationship between high levels of unemployment, the extent of poverty, education and income distribution is uneven. The open unemployment rate shows only aspects of the problem of employment opportunities in developing countries. If they do not work, the consequence is that they cannot fulfill their needs properly, this condition has an impact on the creation and swelling of the number of poverty. Human development is the goal of development itself. Human development plays a role in shaping a country's ability to absorb modern technology and to develop its capacity to create growth and sustainable development (Todaro & Stephen, 2014). Human development in Indonesia is synonymous with poverty reduction (Lanjouw, et al., 2001).

The rate of economic growth is an increase in Gross Regional Domestic Product without regard to whether the increase is greater or smaller (Parkin, 2012). Growth and poverty have a very strong correlation, because in the early stages of the development process poverty levels tended to increase and at the end of the final stage of development the number of poor people gradually diminished (Mankiw, 2009). Research on Improving school enrollment rate elementary, junior high schools and universities as well as economic growth and a significant negative effect on poverty, which means that economic growth will reduce poverty. As well as in education and significant negative effect on poverty, meaning that the higher the level of education will reduce poverty (Teerawichitchainan & Knodel, 2018).

The unemployment rate significantly influence the level of poverty, the government needs to improve the quality of human resources through the community health status and increased access to education to remote areas (Chuang & Lai, 2007). There is a ranking table of percentages poor people by province in Indonesia. In the table, Papua Province is ranked first (Ratnasari, et al., 2015). In addition, Eddy conducted research on macro mapping in poverty. In this research based on the depth index of poverty because the depth index is a measure of the average expenditure gap of each poor population, in this study there is a ranking of Depth of Poverty Index in Indonesia, Papua Province is the highest depth index (Eddy & Agung, 2010). Papua Province is a province that has the highest poverty rate in Indonesia. Then poverty programs and more attention from the government are still very much needed in the Papua Province in overcoming problems of poverty. Based on this, the percentage of poverty will be model using Truncated Spline regression in one of the provinces in Indonesia, namely Papua Province.

Methods

Estimator of Truncated Spline using Least Square Method. After getting Truncated Spline estimator, applied to data on poverty in Papua Province. Below is given the data in 2017 with the following variables: the response variable is the Poverty Percentage (y) while successive predictor variables Open Unemployment Rate (x_1), Human Development Index (x_2), School Expectation Numbers (x_3), and Gross Regional Domestic Product (x_4).

a. To get Truncated Spline in nonparametric regression multivariable curve estimation the steps are as follows:

1. Form the model with Nonparametric Spline Multivariable

$$y_i = \sum_{p=1}^{q} f_p\left(x_{pi}\right) + \varepsilon_i ; i = 1, 2, ..., n$$

2. Approaching Truncated Spline Regression Curve level m witk r knot

$$f_p\left(x_{pi}\right) = \sum_{v=1}^{m} \beta_{vp} x_{pi}^{v} + \sum_{k=1}^{r} \beta_{p(k+m)} \left(x_{pi} - K_{pk}\right)_{+}^{m}$$

3. Forming optimization Least Square

$$\underset{\underset{\sim}{\beta} \in R^{p(m+r)}}{Min} \left\{ n^{-1} \sum_{i=1}^{n} \left(y_i - \sum_{p=1}^{q} \left(\sum_{v=1}^{m} \beta_{vp} x_{pi}^{v} + \sum_{k=1}^{r} \beta_{p(k+m)} \left(x_{pi} - K_{pk}\right)_{+}^{m} \right) \right)^2 \right\}$$

4. Present the Least Square equation in matrix form

$$\underset{\underset{\sim}{\beta} \varepsilon R^{p(m+r)}}{Min} \left\{ n^{-1} \left\| \underset{\sim}{y} - X \underset{\sim}{\beta} \right\|^2 \right\}$$

5. Complete optimization (4) with partial derivatives.

b. Application on Poverty Data in Papua Province

The steps for applying the poverty data model in Papua Province are as follows:

1. Make a Scatterplot $\left(x_1, y_1\right), \left(x_2, y_2\right), \left(x_3, y_3\right), \left(x_4, y_4\right)$.
2. Model data using Truncated Spline.
3. Select the optimal knot point with the GCV method.
4. Calculating the value of MSE and R^2 models Truncated Spline is the best.

Results and Discussion

1. Model of Truncated Spline in Nonparametric Regression Multivariable

In this section we discuss multivariable nonparametric regression Truncated Spline estimation. Spline is the sum of the polynomial functions with a function (truncated). Given a multivariable nonparametric regression model:

$$y_i = \sum_{p=1}^{q} f_p\left(x_{pi}\right) + \varepsilon_i; i = 1, 2, ..., n$$

Next, the regression curve is approached with the multivariable Spline function as follows:

$$f_p\left(x_{pi}\right) = \sum_{v=1}^{m} \beta_{vp} x_{pi}^{v} + \sum_{k=1}^{r} \beta_{p(k+m)} \left(x_{pi} - K_{pk}\right)_+^m$$

Knots that show changes in the behavior patterns of these functions at different sub-intervals. Thus the regression model can be written as:

$$\begin{aligned}\sum_{p=1}^{q} f_p\left(x_{pi}\right) &= \sum_{p=1}^{q}\left(\beta_{1p}x_{pi}^{1} + ... + \beta_{mp}x_{pi}^{m} + \beta_{p(1+m)}\left(x_{pi} - K_{p1}\right)_+^m + ... + \beta_{p(r+m)}\left(x_{pi} - K_{pr}\right)_+^m\right)\\
&= \left(\beta_{11}x_{1i}^{1} + ... + \beta_{m1}x_{1i}^{m} + \beta_{1(1+m)}\left(x_{1i} - K_{11}\right)_+^m + ... + \beta_{1(r+m)}\left(x_{1i} - K_{1r}\right)_+^m\right) + ... +\\
&\quad \left(\beta_{1q}x_{qi}^{1} + ... + \beta_{mq}x_{qi}^{m} + \beta_{q(1+m)}\left(x_{qi} - K_{q1}\right)_+^m + ... + \beta_{q(r+m)}\left(x_{qi} - K_{qr}\right)_+^m\right) \qquad (1)\end{aligned}$$

If equation (1) is presented in the form of a matrix, it is obtained:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11}^1 & \cdots & x_{11}^m & \left(x_{11} - K_{11}\right)_+^m & \cdots & \left(x_{11} - K_{1r}\right)_+^m \\ x_{12}^1 & \cdots & x_{12}^m & \left(x_{12} - K_{11}\right)_+^m & \cdots & \left(x_{12} - K_{1r}\right)_+^m \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_{1n}^1 & \cdots & x_{1n}^m & \left(x_{1n} - K_{11}\right)_+^m & \cdots & \left(x_{1n} - K_{1r}\right)_+^m \end{bmatrix} \begin{bmatrix} \beta_{11} \\ \vdots \\ \beta_{m1} \\ \beta_{1(1+m)} \\ \vdots \\ \beta_{1(r+m)} \end{bmatrix} + ... +$$

$$\begin{bmatrix} x_{q1}^{1} & \cdots & x_{q1}^{m} & \left(x_{q1}-K_{q1}\right)_{+}^{m} & \cdots & \left(x_{q1}-K_{qr}\right)_{+}^{m} \\ x_{q2}^{1} & \cdots & x_{q2}^{m} & \left(x_{q2}-K_{q1}\right)_{+}^{m} & \cdots & \left(x_{q2}-K_{qr}\right)_{+}^{m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_{qn}^{1} & \cdots & x_{qn}^{m} & \left(x_{qn}-K_{q1}\right)_{+}^{m} & \cdots & \left(x_{qn}-K_{qr}\right)_{+}^{m} \end{bmatrix} \begin{bmatrix} \beta_{1q} \\ \vdots \\ \beta_{mq} \\ \beta_{q(1+m)} \\ \vdots \\ \beta_{q(r+m)} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{2}$$

Equation (2) can be written as:

$$\underset{\sim}{y} = X\left(K_{11},...,K_{1r} \mid \cdots \mid K_{q1},...,K_{qr}\right)\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

Estimation of $\underset{\sim}{\beta} = \left(\underset{\sim}{\beta}_1,...,\underset{\sim}{\beta}_q\right)'$ obtained using the *Least Square* method. Parameter estimator $\underset{\sim}{\beta}$ obtained from completing optimization:

$$\underset{\underset{\sim}{\beta}\varepsilon R^{p(m+r)}}{Min}\left\{n^{-1}\left\|\underset{\sim}{y} - X\underset{\sim}{\beta}\right\|^2\right\}$$

$$= \left\{\left(\underset{\sim}{y} - X\left(K_{11},...,K_{1r} \mid \cdots \mid K_{q1},...,K_{qr}\right)\underset{\sim}{\beta}\right)'\left(\underset{\sim}{y} - X\left(K_{11},...,K_{1r} \mid \cdots \mid K_{q1},...,K_{qr}\right)\underset{\sim}{\beta}\right)\right\}$$

To complete the optimization using partial derivatives, for example:

$$\square\left(\underset{\sim}{\beta}\right) = \left\{\left(\underset{\sim}{y} - X\left(K_{11},...,K_{1r} \mid \cdots \mid K_{q1},...,K_{qr}\right)\underset{\sim}{\beta}\right)'\left(\underset{\sim}{y} - X\left(K_{11},...,K_{1r} \mid \cdots \mid K_{q1},...,K_{qr}\right)\underset{\sim}{\beta}\right)\right\} \tag{3}$$

Equation (3) to be:

$$= \underset{\sim}{y}'\underset{\sim}{y} - 2\underset{\sim}{\beta}X'\left(K_{11},...,K_{1r} \mid \cdots \mid K_{q1},...,K_{qr}\right)\underset{\sim}{y}$$

$$+\underset{\sim}{\beta}'X'\left(K_{11},...,K_{1r} \mid \cdots \mid K_{q1},...,K_{qr}\right)X\left(K_{11},...,K_{1r} \mid \cdots \mid K_{q1},...,K_{qr}\right)\underset{\sim}{\beta} \tag{4}$$

The next step in equation (4) is derived from $\underset{\sim}{\beta}$:

$$\frac{\partial\square\left(\underset{\sim}{\beta}\right)}{\partial\underset{\sim}{\beta}} = -2X'\left(K_{11},...,K_{1r} \mid \cdots \mid K_{q1},...,K_{qr}\right)\underset{\sim}{y}$$

$$+2X'\left(K_{11},...,K_{1r} \mid \cdots \mid K_{q1},...,K_{qr}\right)X\left(K_{11},...,K_{1r} \mid \cdots \mid K_{q1},...,K_{qr}\right)\underset{\sim}{\beta} \tag{5}$$

After equation (5) is derived the results are equated with 0, then the equation is obtained:

$$0=-2X'\left(K_{11},...,K_{1r}\,|\cdots|\,K_{q1},...,K_{qr}\right)\underset{\sim}{y}+$$

$$2X'\left(K_{11},...,K_{1r}\,|\cdots|\,K_{q1},...,K_{qr}\right)X\left(K_{11},...,K_{1r}\,|\cdots|\,K_{q1},...,K_{qr}\right)\underset{\sim}{\hat{\beta}} \qquad (6)$$

So that the estimator $\underset{\sim}{\hat{\beta}}$ is given by:

$$\underset{\sim}{\hat{\beta}}=\left(X'\left(K_{11},...,K_{1r}\,|\cdots|\,K_{q1},...,K_{qr}\right)X\left(K_{11},...,K_{1r}\vdots\cdots\vdots K_{q1},...,K_{qr}\right)\right)^{-1}X'\left(K_{11},...,K_{1r}\,|\cdots|\,K_{q1},...,K_{qr}\right)\underset{\sim}{y}$$

with $\underset{\sim}{\hat{\beta}}=\left(\underset{\sim}{\hat{\beta}}_1',...,\underset{\sim}{\hat{\beta}}_q'\right)'$.

Regression curve estimator $\underset{\sim}{\hat{f}}(x)$ obtained from:

$$\underset{\sim}{\hat{f}}(x)=\left[\underset{\sim}{\hat{f}}(x_1)',\underset{\sim}{\hat{f}}(x_2)',...,\underset{\sim}{\hat{f}}(x_q)'\right]'=X\left(K_{11},...,K_{1r}\,|\cdots|\,K_{q1},...,K_{qr}\right)\underset{\sim}{\hat{\beta}}$$

$$=\left(X'\left(K_{11},...,K_{1r}\,|\cdots|\,K_{q1},...,K_{qr}\right)X\left(K_{11},...,K_{1r}\,|\cdots|\,K_{q1},...,K_{qr}\right)\right)^{-1}$$

$$X'\left(K_{11},...,K_{1r}\,|\cdots|\,K_{q1},...,K_{qr}\right)\underset{\sim}{y}$$

Estimation f can be written to be:

$$\hat{f}_p(x_{pi})=\sum_{v=1}^{m}\hat{\beta}_{vp}x_{pi}^{v}+\sum_{k=1}^{r}\hat{\beta}_{p(k+m)}\left(x_{pi}-K_{pk}\right)_+^m$$

$$\sum_{p=1}^{q}\hat{f}_p(x_{pi})=\sum_{p=1}^{q}\left(\sum_{v=1}^{m}\hat{\beta}_{vp}x_{pi}^{v}+\sum_{k=1}^{r}\hat{\beta}_{p(k+m)}\left(x_{pi}-K_{pk}\right)_+^m\right)$$

As a result the estimate for the regression curve is as follows:

$$\hat{y}=\sum_{p=1}^{q}\hat{f}_p(x_{pi})=\sum_{p=1}^{q}\sum_{v=1}^{m}\hat{\beta}_{vp}x_{pi}^{v}+\sum_{p=1}^{q}\sum_{k=1}^{r}\hat{\beta}_{p(k+m)}\left(x_{pi}-K_{pk}\right)_+^m$$

where $\hat{\beta}_{vp}$ and $\hat{\beta}_{p(k+m)}$, $v=1,2,...,m;k=1,2,...,r,;\ p=1,2,...,q$

2. Modeling Poverty in Papua Province

Will be shown in the scatterplot between the response variable and the predictor variable. With the following variables: the response variable is the Poverty Percentage (y) while successive predictor variables Open Unemployment Rate (x_1), Human Development Index (x_2), School Expectation Numbers (x_3), and Gross Regional Domestic Product (x_4).

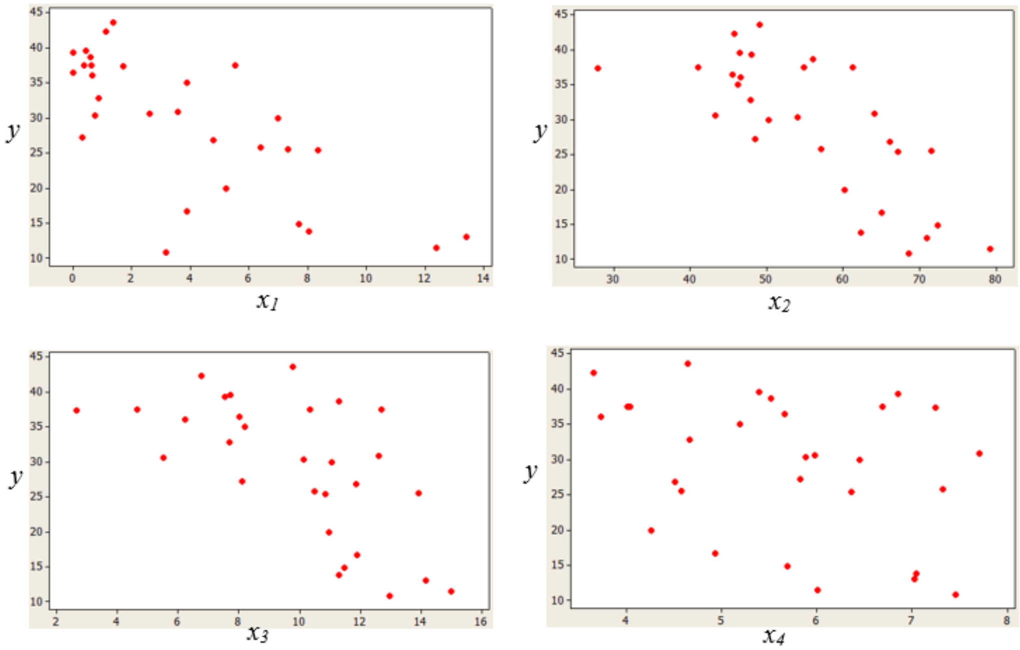

Figure 1. *Scatterplot between response variables with each predictor variable*

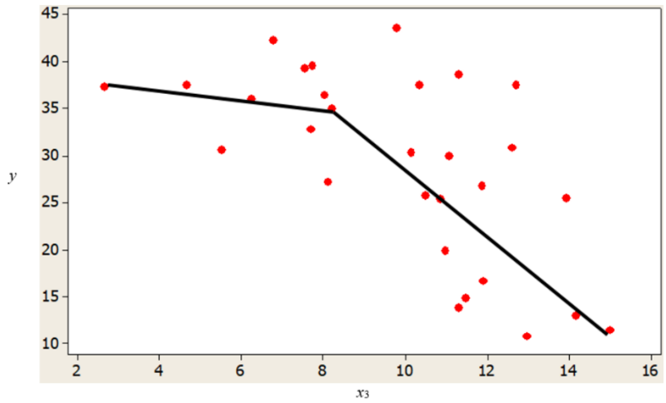

Figure 2. *Knot in Truncated Spline*

In Figure 1, there is a scatterplot between response variables with each predictor variable that changes in sub-intervals. Figure 2 shows that in the Truncated Spline there is a change in a certain sub interval and contains a point of knots. So, modeling poverty cases using nonparametric regression Truncated Spline. In this discussion, modeling will be conducted on variables that are thought to affect poverty. The method used to model this data is nonparametric regression Truncated Spline. The following is a mathematical model of the Truncated Spline in nonparametric regression.

$$\begin{aligned}\hat{y}_i \;=\; & \beta_0+\beta_1\, x_{1i}-\beta_{12}\,(x_{1i}-K_{11})_+ +\beta_{13}\,(x_{1i}-K_{12})_+ +\beta_{14}\,(x_{1i}-K_{13})_+ +\cdots+ \\ & \beta_4\, x_{4i}-\beta_{41}\,(x_{4i}-K_{41})_+ +\beta_{42}\,(x_{4i}-K_{42})_+ +\beta_{44}\,(x_{4i}-K_{43})_+ .\end{aligned}$$

Knots is the point where the pattern of data changes. To get the optimal knot point, use the GCV method. Choosing the optimal value of knots used the minimum GCV value. The point of knots used in this study are one knot, two knots, three knots and a combination of knots. The following will show the optimal of knots.

Table 1. Comparison of GCV Value of Various Knots

Knot	*GCV*	*MSE*	R^2
1 Knot	42,48	22,28	76,49
2 Knot	41,55	19,76	79,14
3 Knot	29,98	6,02	93,64
Combination of Knot	**26,92**	**8,19**	**91,35**

Based on **Table 1** obtained optimal knots are knots combinations. The combination of knots is a combination of one knots, two knots, and three knots. This combination is used to select the optimal knot. In selecting the optimal knot in the nonparametric regression Truncated Spline model with a combination of knots, the minimum GCV value was chosen. The results of the GCV will be compared with the results of the GCV in the previous experiment. The following is the GCV value of the Truncated Spline regression model with a combination of knots.

Table 2. Optimal Point Selection with Combination Knot

No	Variable	Variations Knot	Knot	GCV
1	x_1 x_2 x_3 x_4	1 1 1 2	K_1=12,88; K_2=77,28; K_3=14,51; K_4=7,39; K_5=7,55;	41,55
2	x_1 x_2 x_3 x_4	2 3 1 2	K_1=12,36;K_2=12,88; K_3=65,40; K_4=67,38; K_5=69,35; K_6=14,51; K_7=7,39; K_8=7,55;	38,82
3	x_1 x_2 x_3 x_4	**3** **3** **1** **3**	**K_1=9,79; K_2=10,30; K_3=10,82;** **K_4=65,40; K_5=67,37; K_6=69,35;** **K_7=14,51;** **K_8=6,62; K_9=6,78; K_{10}=6,93;**	**26,92**
4	x_1 x_2 x_3 x_4	3 3 3 1	K_1=9,79;K_2=10,30;K_3=10,82; K_4=65,40; K_5=67,37; K_6=69,35; K_7=11,66;K_8=12,13;K_9=12,61; K_{10}=7,55;	36,77
5	x_1 x_2 x_3 x_4	3 3 1 1	K_1=9,79;K_2=10,30;K_3=10,82; K_4=65,40; K_5=67,37; K_6=69,35; K_7=14,51; K_8=7,55;	37,29

In **Table 1** and **Table 2** it can be seen that the minimum GCV value with a combination of knots is 26.92 with a combination of 3,3,1,3. The optimal knot of a combination knots are

(K_1=9,79; K_2=10,30; K_3=10,82); (K_7=14,51);

(K_4=65,40; K_5=67,37; K_6=69,35); (K_8=6,62; K_9=6,78; K_{10}=6,93).

3. Parameter Estimation of the Model Truncated Spline in Nonparametric Regression

In getting the model it is best to use optimal knots. From the results of the optimal knot point selection, the regression model using the knot point combination is the best. The results of parameter estimation using point knot combinations are as follows.

$$\begin{aligned}\hat{y} = \ & 106.56\text{-}1{,}01\ x_1 - 58{,}72\ (x_1 - 9{,}79)_+ + 8{,}77\ (x_1 - 10{,}30)_+ + 76{,}26\ (x_1 - 10{,}82)_+ + \\ & -1{,}79\ x_2 + 13{,}73\ (x_2 - 65{,}40)_+ - 41{,}52\ (x_2 - 67{,}37)_+ + 43{,}03\ (x_2 - 69{,}35)_+ + \\ & 4{,}82\ x_3\text{-}196{,}34\ (x_3 - 14{,}51)_+ - 4{,}64\ x_4 + 180{,}44(x_4 - 6{,}62)_+ \\ & -355{,}79(x_4 - 6{,}78)_+ + 195{,}97(x_4 - 6{,}93)_+ \ . \end{aligned} \tag{7}$$

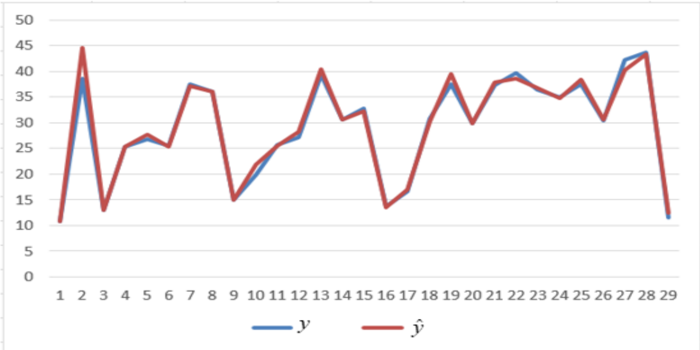

Figure 3. y with $\hat{y}$

In **Figure 3**, it can be seen that the Truncated Spline in nonparametric regression model with this knot combination is good because the estimated value of y with $\hat{y}$ is relatively small. The Truncated Spline regression model based on equation (7) has *GCV* of 26.92, *MSE* is obtained 8.19 and R^2=91.35%. This means that this model can explain percentage of poverty by 91.35%.

Consclusion

From the results of the study it can be concluded that the estimation of the Truncated Spline in nonparametric regression curve is as follows:

$$\hat{y} = \sum_{p=1}^{q}\sum_{v=1}^{m}\hat{\beta}_{vp}x_{pi}^{v} + \sum_{p=1}^{q}\sum_{k=1}^{r}\hat{\beta}_{p(k+m)}\left(x_{pi}-K_{pk}\right)_{+}^{m}$$

After getting an estimate Truncated Spline will then be applied model of Truncated Spline regression in the case of poverty in Papua province is as follows

$$\hat{y} = 106.56\text{-}1,01\ x_1 - 58,72\ (x_1 - 9,79)_+ + 8,77\ (x_1 - 10,30)_+ + 76,26\ (x_1 - 10,82)_+ +$$

$$-1,79\ x_2 + 13,73\ (x_2 - 65,40)_+ - 41,52\ (x_2 - 67,37)_+ + 43,03\ (x_2 - 69,35)_+ +$$

$$4,82\ x_3\text{-}196,34\ (x_3 - 14,51)_+ - 4,64\ x_4 + 180,44(x_4 - 6,62)_+$$

$$-355,79(x_4 - 6,78)_+ + 195,97(x_4 - 6,93)_+ .$$

The best of model Truncated Spline regression is obtained using a knot combination with GCV=26,92; MSE=8,19 and has R^2 of 91,35%. y estimates with $\hat{y}$ relatively small.

References

Budiantara, I. N., Ratnasari, V., Ratna, M., & Zain, I. (2015). The Combination of Spline and Kernel Estimator for Nonparametric Regression and its Properties. *Apllied Mathematical Sciences, 9*(122), 6083-6094.

Budiantara, I. N., Ratnasari, V., Ratna, M., Wibowo, W., Afifah, N., Rahmawati, D. P., & Octavanny, M. A. (2019). Modeling Percentage Of Poor People In Indonesia Using Kernel And Fourier Series Mixed Estimator In Nonparametric Regression. *Investigación Operacional, 40*(4), 538-551.

Chuang, Y. C., & Lai, W. W. (2007). The Sources of Taiwan’s Regional Unemployment: A Cross-Region Panel Analysis. *Hitotsubashi Journal of Economics, 49*, 47-65.

Eddy, S. S., & Agung. (2010). Faktor-faktor yang Mempengaruhi Kemiskinan secara Makro di Lima Belas Provinsi Tahun 2007. *Jurnal dan Organisasi Manajemen, 6*(2), 89-100.

Eubank, R. (1999). *Nonparametric Regression and Spline Smoothing.* New York: Marcel Dekker.

Eubank, R. L. (1988). *Spline Smoothing and Nonparametric Regression.* New York: Marcel Dekker.

Lanjouw, P., Pradhan, M., Saadah, F., Sayed, H., & Sparrow, R. (2001). *Poverty, Education and Health in Indonesia: Who Benefits from Public Spending.* Washington D C: World Bank.

Liang, H. (2006). Estimations in Partially Linear Models and Numerical Comparison. *Computational Statistics and Data Analysis, 50*, 675-687.

Mankiw, M. (2009). *Macroeconomics.* New York: Worth Publishers.

Parkin, M. (2012). *Macroeconomics.* New York: Addison Wesley.

Ratnasari, V., Budiantara, I. N., Zain, I., Ratna, M., & Mariati, N. M. (2015). Comparison Truncated Spline and Fourier Series in Multivariable Nonparametric Regression Models

(Applications: Data Of Poverty in Papua, Indonesia). *International Journal of Basic and Applied Sciences, 15*(4), 9-12.

Teerawichitchainan, B., & Knodel, J. (2018). Long-Term Care Needs in the Context of Poverty and Population Aging: the Case of Older Persons in Myanmar. *Journal of Cross-Cultural Gerontolog*, *33*(2), 143-162.

Todaro, M. P., & Stephen, C. S. (2014). *Economic Development.* New York: Addition Wesley Longman, Inc.

Wahba, G. (1990). *Spline Models For Observasion Data.* Pensylvania: SIAM.