BAREKENG: Journal of Mathematics and Its Applications
September 2023 Volume 17 Issue 3 Page 1565–1574
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol17iss3pp1565-1574

# TUBERCULOSIS CASE MODEL USING GCV AND UBR KNOT SELECTION METHODS IN TRUNCATED SPLINE NONPARAMETRIC REGRESSION

**Sitti Anggraeni [1], Sifriyani[2*], Qonita Qurrota A'yun[3]**

[1,2]*Statistical Study Program, Faculty of Mathematics and Natural Sciences, Mulawarman University*
[3]*Mathematics Study Program, Faculty of Mathematics and Natural Sciences, Mulawarman University*
*Barong Tongkok Street No.04, Samarinda, 75123, Indonesia*

*Corresponding author's e-mail:* ***sifriyani@fmipa.unmul.ac.id**

*ABSTRACT*

***Article History:***

*Received: 17th March 2023*
*Revised: 6th August 2023*
*Accepted: 16th August 2023*

***Keywords:***

*GCV;*
*Nonparametric regression;*
*Truncated spline;*
*Tuberculosis;*
*UBR.*

*The nonparametric regression approach is used when the shape of the regression curve is not known. The advantage of nonparametric regression is that it has a high degree of flexibility. The truncated spline is a method in the nonparametric regression approach, which can overcome changing data patterns at certain sub-intervals with the help of knot points. The purpose of this research is to obtain the best truncated spline nonparametric regression model estimates based on the GCV and UBR knot point selection methodsThe data used in this study came from the publications of the Indonesian Ministry of Health and BPS Indonesia. The response variable used is the percentage of successful treatment of tuberculosis patients in Indonesia with predictor variables namely the percentage of people who smoke over the age of 15 years, the percentage of households that have access to proper sanitation, the percentage of poor people, the percentage of food processing establishments that meet the standard requirements , national health insurance membership coverage and percentage of accredited hospitals. The results showed that the best model came from the GCV method using three knots. This model produces an MSE value of 3.65 with $R^2$ value of 97.04. The $R^2$ value indicates that the predictor variable used in this study affects the response variable by 97.04% while the other 2.96% is influenced by other variables that are not included in this study.*



***How to cite this article:***

S. Anggraeni, Sifriyani and Q. Q. A'yun., "TUBERCULOSIS CASE MODEL USING GCV AND UBR KNOT SELECTION METHODS IN TRUNCATED SPLINE NONPARAMETRIC REGRESSION," *BAREKENG: J. Math. & App.*, vol. 17, iss. 3, pp. 1565-1574, September, 2023.


Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

**Research Article • Open Access**

## 1. INTRODUCTION

Regression analysis approach is divided into 3, namely the parametric regression approach, the semiparametric regression approach and the nonparametric regression approach. The parametric regression approach is used when the shape of the regression curve is known, for example linear, cubic, quadratic polynomial and others. The semiparametric regression approach is a combination of the parametric regression approach and the nonparametric regression approach, where some of the regression curves are known and some are unknown. While the nonparametric regression approach is used if the shape of the regression curve is unknown. In real conditions the shape of the regression curve is not always known, so the recommended approach is a nonparametric regression approach [1], [2].

The nonparametric regression approach has a high degree of flexibility, so that the data are expected to be able to find their own estimated form of the regression curve without being influenced by the subjectivity of the researcher [3], [4]. One of the methods used in the nonparametric regression approach is spline.

Spline is a type of piecewise polynomial, which means a polynomial that has truncated or segmented properties [5], [6]. Spline has a function basis, one of which is a truncated spline. Truncated splines can overcome the nature of data patterns that change at certain intervals [7], [8]. Changes in data patterns that rise or fall sharply can be overcome with the help of knots. The knot point is a joint point that indicates a change in the data pattern . The number of knot points can be determined based on changes in data patterns that occur at certain sub-intervals. The goodness of the spline regression model is influenced by the selection of optimal knot points. Several methods that can be used to obtain optimal knot points include the Cross Validation (CV), Unbiased Risk (UBR), Generalized Maximum Likelihood (GML) and Generalized Cross Validation (GCV) [9]. The selection of knot points in this study used the GCV and UBR methods.

UBR is a knot point selection method that requires the estimated value of the variance to be known. The UBR method can produce good values if it is used on data that is not normally. Whereas GCV would be better used on gaussian data or normally distributed. The advantages of the GCV method are that it is asymptotically optimal, simple and efficient in calculations, does not require information about the variance and invariance of the transformation [10].

Research using the GCV and UBR methods has been carried out by previous researchers including Comparison of Selection of Optimal Knot Points Using Unbiased Risk (UBR) and Generalized Cross Validation (GCV) Methods on Nonparametric Regression Splines with Case Studies of Total Infant Mortality Rate in Central Java Province in 2017 [10] and Spline Nonparametric Regression Modeling with Optimal Knot Selection Using Unbiased Risk (UBR) and Generalized Cross Validation (GCV) Methods in the case of the Human Development Index in Central Java [9]. The results of the two studies concluded that the GCV method is more appropriate to use in selecting optimal knot points compared to the UBR method.

This study will use the GCV and UBR methods to select optimal knot points in tuberculosis cases in Indonesia. Tuberculosis is an infectious disease caused by the bacterium mycobacterium tuberculosis and is one of the 10 biggest causes of death in the world. Indonesia is ranked 2nd with the highest number of tuberculosis sufferers in the world after India. In 2020 the number of tuberculosis cases in Indonesia was 351.936 cases, which has decreased compared to all tuberculosis cases that occurred in 2019. The treatment success rate is an indicator used to evaluate the progress of tuberculosis treatment. The treatment success rate can also be interpreted as the number of all tuberculosis cases that are cured and fully treated among all tuberculosis cases that are treated and reported [11]. Research on tuberculosis cases has been carried out by previous researchers including Multivariable Nonparametric Regression with a Truncated Spline Approach in Tuberculosis Cases [12] and Modeling of Factors Affecting Tuberculosis Recovery in East Java Province with Spline Truncated Regression [13]. In this study, one method of selecting knots was used, namely GCV, while in this study two methods were used to select points of knots, namely GCV and UBR. In addition, tuberculosis cases in this study used the percentage success rate of treatment of tuberculosis patients.

## 2. RESEARCH METHODS

This study uses a truncated spline nonparametric regression approach with the GCV and UBR knot point selection method on data on the success rate of treatment of tuberculosis patients in Indonesia in 2020. The data collection technique used was a purposive sampling technique.

### 2.1 Regression Analysis

Regression analysis is a statistical method used in modeling the relationship between two or more observational variables. This method can be used to identify a relationship between two variables, namely the predictor variable (independent) and the response variable (dependent) [14].

### 2.2 Truncated Spline Nonparametric Regression

Nonparametric regression is a method used to identify patterns of relationship between response variables and predictor variables when the shape of the regression curve is unknown. In nonparametric regression, the data is expected to be able to obtain the shape of the regression curve on its own without considering the subjectivity of the researcher [15]. One model that is often used in nonparametric regression is the spline.

The spline function in general can be written in the following equation.

$$y = \sum_{j=0}^{m} \beta_j x_i^j + \sum_{h=1}^{q} \beta_{m+h} \left(x_i - K_h\right)_+^m + \varepsilon \tag{1}$$

With:

| | | |
|---|---|---|
| $y$ | = | Response variable |
| $\beta_j$ | = | Regression parameter coefficients for polynomial functions |
| $x_i$ | = | Predictor variables |
| $\beta_{m+h}$ | = | Regression parameter coefficients for truncated functions |
| $K_h$ | = | Point knots |
| $\varepsilon$ | = | Errors that are assumed to be identical, independent and normally distributed are stated in $\varepsilon \sim \mathrm{IIDN}(0, \sigma^2)$ |

### 2.3 Generalized Cross Validation (GCV)

GCV is one of the methods used in selecting optimal knot points. The resulting optimal knot point comes from the smallest GCV value [16]. The GCV function can be seen in the following equation.

$$GCV(K) = \frac{MSE(\boldsymbol{K})}{\left[n^{-1} trace(\boldsymbol{I} - \boldsymbol{A}(\boldsymbol{K}))\right]^2} \tag{2}$$

$$= n^{-1} \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} (y_i - \hat{f}(x_i))^2}{\left[n^{-1} trace(\boldsymbol{I} - \boldsymbol{A}(\boldsymbol{K})\right]^2}$$

$$= n^{-1} \frac{\underset{\sim}{y}'(\boldsymbol{I} - \boldsymbol{A}(\boldsymbol{K}))'(\boldsymbol{I} - \boldsymbol{A}(\boldsymbol{K}))\underset{\sim}{y}}{\left[n^{-1} trace(\boldsymbol{I} - \boldsymbol{A}(\boldsymbol{K})\right]^2}$$

With:

$$\hat{f}\left(x_i\right) = \boldsymbol{A}(\boldsymbol{K})\underset{\sim}{y} = \left[\boldsymbol{X}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}\right]\underset{\sim}{y} \tag{3}$$

$$\boldsymbol{A(K)} = \left[\boldsymbol{X}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}\right]$$

### 2.4 Unbiased Risk (UBR)

UBR is one method that is also used in selecting optimal knot points. The resulting optimal knot point comes from the minimum UBR value. The formula used in the UBR method is given in the following equation.

$$\hat{R}(\boldsymbol{K}) = \frac{1}{n}\left\|(\mathrm{I} - \boldsymbol{A}(\boldsymbol{K}))\underset{\sim}{y}\right\|^2 - \frac{\sigma^2}{\mathrm{n}} trace\left[\boldsymbol{I} - \boldsymbol{A}(\boldsymbol{K})\right]^2 + \frac{\sigma^2}{n} trace\left[\boldsymbol{A}^2\right](\boldsymbol{K}) \tag{9}$$

With:

$$A(K) = X\left(X^t X\right)^{-1} X^t \quad (10)$$

The estimated value of obtained using the following formula:

$$\hat{\sigma}^2 = \frac{\left\| \left[ I - A(K) \right] \underset{\sim}{y} \right\|}{trace\left[ I - A(K) \right]} \quad (11)$$

## 2.5 Best Model Selection Criteria

The method that is often used in selecting the best model is the Mean Square Error (MSE). MSE is a good condition of the model that has a non-negative value. A good model comes from the smallest MSE value produced by the model [17]. The following is the equation for calculating the MSE value in the truncated spline regression model.

$$MSE = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)}{n} \quad (12)$$

The coefficient of determination can be used to measure the strength of a model in explaining the response variable. The value of the coefficient of determination ranges from zero to one [18]. The following is the equation used to calculate the coefficient of determination.

$$R^2 = \frac{SSR}{SST} \quad (13)$$

## 2.6 Multicollinearity Test

The multicollinearity test is a test used to determine whether there is a high correlation between predictor variables in the regression model. This test can use the VIF method, where if the VIF value is more than 10 then it indicates the occurrence of multicollinearity [19]. The following is the equation used to find the VIF value.

$$\text{VIF} = \frac{1}{1 - R^2} \quad (14)$$

## 2.7 Research Variables

The variables used in this study consisted of response variables and predictor variables. The response variable is the success rate of treatment of tuberculosis patients in Indonesia. Predictor variables are the percentage of the population smoking over the age of 15 years , the percentage of households having access to proper sanitation, the percentage of national health insurance, the percentage of poor people, the percentage of food processing establishments that meet standard requirements and the percentage of accredited hospitals. Variable notations, variables description and data sources are presented in Table 1.

**Table 1. Research Variables**

| Notation | Operational Definition | Data Source |
|---|---|---|
| $y$ | Percentage of success rate of treatment of tuberculosis patients | Ministry of Health of the Republic of Indonesia |
| $x_1$ | Percentage of the population smoking over the age of 15 years | bps.go.id |
| $x_2$ | The percentage of households having access to proper sanitation | Ministry of Health of the Republic of Indonesia |
| $x_3$ | The percentage of national health insurance | Ministry of Health of the Republic of Indonesia |
| $x_4$ | The percentage of poor people | Ministry of Health of the Republic of Indonesia |
| $x_5$ | The percentage of food processing establishments that meet standard requirements | Ministry of Health of the Republic of Indonesia |
| $x_6$ | Percentage of accredited hospitals | Ministry of Health of the Republic of Indonesia |

### 2.8 Stage of Analysis

The stages of analysis in this study are as follows.

1. Perform descriptive statistical analysis on each variable.
2. Make a scatter plot between the Tuberculosis Treatment Success Rate data and each predictor variable.
3. Perform multicollinearity detection on each predictor variable.
4. Find estimation of a truncated nonparametric spline regression model using one knot, two knots and three knots using the GCV method.
5. Find Estimation of a truncated nonparametric spline regression model using one knot, two knots and three knots using the UBR method.
6. Select the best nonparametric spline regression model based on the coefficient of determination $(R^2)$ and Mean Square Error (MSE) based on the optimal knot point on the GCV and UBR methods.
7. Interpret the results of the analysis obtained.
8. Make conclusions.

## 3. RESULTS AND DISCUSSION

The first step is to look at data patterns using a scatterplot and then perform multicollinearity detection. After that, select the optimal knot point using the GCV and UBR methods. Then compare the optimal knot point values obtained from the two methods. Then choose the optimal knot point based on the smallest Mean Square Error (MSE) value and the highest $R^2$ value.

### 3.1 Data Pattern

One of the stages of analysis in nonparametric spline truncated regression is knowing the relationship pattern between predictor variables and response variables. The pattern of this relationship can be known by using a scatterplot. The following is a scatterplot between predictor variables and response variables.

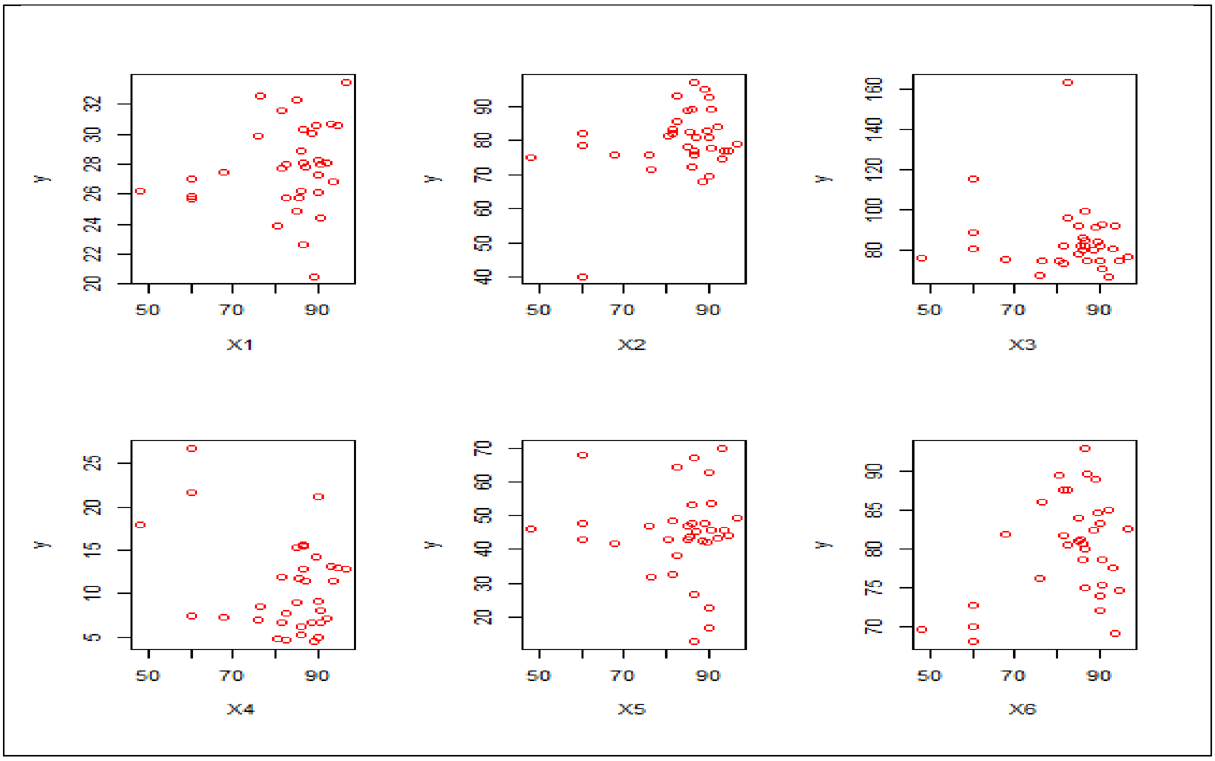


**Figure 1. Data Pattern**

Based on **Figure 1** above, it can be seen that the data pattern between each predictor variable and the response variable does not form a specific pattern such as linear, quadratic and cubic, where the data spreads from other data. This condition is not appropriate if forced to use a parametric regression approach, so that a nonparametric regression approach will be appropriate for this data.

### 3.2 Multicollinearity

Multicollinearity detection aims to determine whether there is a correlation between predictor variables in the regression model. A good model is a model that has no correlation between the predictor variables. The results of the multicollinearity test can be seen in **Table 2**.

**Table 2. Multicollinearity detection using the VIF method**

| Variabel | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ |
|---|---|---|---|---|---|---|
| VIF | 1.29 | 2.04 | 1.22 | 1.77 | 1.30 | 1.34 |

Based on **Table 2** above, it can be seen that the VIF values for all predictor variables are less than 10 (VIF <10), so it can be concluded that there is no correlation between predictor variables and there is no multicollinearity in the regression model. Thus all predictor variables can be used in nonparametric spline truncated regression testing.

### 3.3 Selection of Optimal Knot Points Using the GCV Method

In the GCV method, the optimal knot point value is based on the minimum GCV value. Selection of optimal knot points using the GCV method is divided into one knot point, two knot points and three knot points. The following is the result of a comparison of the knot points on the GCV method.

**Table 3. Comparison of GCV Knot Points**

| Knot Point | GCV Value | MSE | $R^2$ |
|---|---|---|---|
| 1 Point Knot | 165.93 | 63.30 | 48.64 |
| 2 Point Knot | 118.03 | 26.14 | 78.79 |
| **3 Point Knot** | **34.90** | **3.65** | **97.04** |

Based on **Table 3** above, it can be seen that the minimum GCV value is 34.90. So that the optimal knot point is obtained from three knot points. The results of parameter estimation from the truncated spline nonparametric regression model using three knot points can be seen in **Table 4** below.

**Table 4. Optimal Knot Point Parameter Estimation GCV Method**

| Variable | Parameter | Estimation |
|---|---|---|
| *Constant* | $\beta_0$ | 19.69 |
| $x_1$ | $\beta_{11}$ | 0.88 |
| | $\beta_{12}$ | 0.45 |
| | $\beta_{13}$ | 0.01 |
| | $\beta_{14}$ | -0.60 |
| $x_2$ | $\beta_{21}$ | -0.20 |
| | $\beta_{22}$ | 0.21 |
| | $\beta_{23}$ | 9.27 |
| | $\beta_{24}$ | 6.18 |
| $x_3$ | $\beta_{31}$ | 3.09 |
| | $\beta_{32}$ | -0.68 |
| | $\beta_{33}$ | 6.11 |
| | $\beta_{34}$ | 4.88 |
| $x_4$ | $\beta_{41}$ | -0.70 |
| | $\beta_{42}$ | -0.47 |
| | $\beta_{43}$ | -0.23 |

| | | |
|---|---|---|
| | $\beta_{44}$ | 4.23 |
| $x_5$ | $\beta_{51}$ | 2.82 |
| | $\beta_{52}$ | 1.41 |
| | $\beta_{53}$ | -15.70 |
| | $\beta_{54}$ | 23.44 |
| $x_6$ | $\beta_{61}$ | 15.12 |
| | $\beta_{62}$ | -0.53 |
| | $\beta_{63}$ | -0.35 |
| | $\beta_{64}$ | -0.17 |

Based on **Table 4** bove, a truncated spline nonparametric regression model is obtained using three knot points as follows.

$$\hat{y}=19,69+0,88x_1+0,45(x_1-27,62)_+ +0,01(x_1-28,15)_+ -0,60(x_1-28,42)_+ -0,20x_2+0,21(x_2-71,53)_+$$
$$+9,27(x_2-73,84)_+ +6,18(x_2-74,99)_+ +3,09x_3-0,68(x_3-119,93)_+ +6,11(x_3-123,84)_+$$
$$+4,88(x_3-125,79)_+ -0,70x_4-0,47(x_4-16,77)_+ -0,23(x_4-17,68)_+ +4,23(x_4-18,13)_+$$
$$+2,82x_5+1,41(x_5-44,27)_+ -15,70(x_5-46,61)_+ +23,44(x_5-47,78)_+ +15,12x_6$$
$$-0,53(x_6-81,77)_+ -0,35(x_6-82,78)_+ -0,17(x_6-83,28)_+$$

The model above produces an $R^2$ value of 97.04 which indicates that the variable percentage of the population smokes over 15 years, the percentage of households having access to proper sanitation, the coverage of national health insurance, the percentage of poor people, the percentage of food processing establishments that meet the requirements standards and the percentage of accredited hospitals have an effect of 97.04% on the success rate of treatment of tuberculosis patients in Indonesia while 2.96% are influenced by other variables.

### 3.4 Selection of Optimal Knot Points Using the UBR Method

In the UBR method, the optimal knot point value is based on the minimum UBR value. Selection of optimal knot points using the UBR method is divided into one knot point, two knot points and three knot points. The following is the result of a comparison of the knot points on the UBR method.

**Table 5. Comparison of Knot Points of the UBR Method**

| Knot Point | Nilai UBR | MSE | $R^2$ |
|---|---|---|---|
| 1 Point Knot | $3.25\times10^{-12}$ | 81.64 | 33.75 |
| **2 Point Knot** | $\mathbf{2.64\times10^{-12}}$ | **57.48** | **53.36** |
| 3 Point Knot | $3.40\times10^{-12}$ | 55.88 | 54.66 |

Based on **Table 5** above, it can be seen that the minimum UBR value is 53.36. So that the optimal knot point is obtained from two knot points. The parameter estimation results from the truncated spline nonparametric regression model with two knot points can be seen in **Table 6** below.

**Tabel 6. Optimal Knot Point Parameter Estimation UBR Method**

| Variable | Parameter | Estimation |
|---|---|---|
| *Constant* | $\beta_0$ | 19.69 |
| $x_1$ | $\beta_{11}$ | 0.88 |
| | $\beta_{12}$ | 0.45 |
| | $\beta_{13}$ | 0.01 |
| $x_2$ | $\beta_{21}$ | -0.60 |
| | $\beta_{22}$ | -0.20 |

| | | |
|---|---|---|
| | $\beta_{23}$ | 0.20 |
| $x_3$ | $\beta_{31}$ | 17.31 |
| | $\beta_{32}$ | 8.66 |
| | $\beta_{33}$ | 3.26 |
| $x_4$ | $\beta_{41}$ | -1.08 |
| | $\beta_{42}$ | -1.30 |
| | $\beta_{43}$ | -0.65 |
| $x_5$ | $\beta_{51}$ | 7.90 |
| | $\beta_{52}$ | 3.95 |
| | $\beta_{53}$ | -124.54 |
| $x_6$ | $\beta_{61}$ | 263.98 |
| | $\beta_{62}$ | -0.99 |
| | $\beta_{63}$ | -0.50 |

Based on **Table 6** truncated spline nonparametric regression model is obtained using two knot points as follows.

$$\hat{y}=19,69+0,88x_1+0,45(x_1-27,62)_+ +0,01(x_1-28,15)_+ -0,60x_2-0,20\left(x_2-71,53\right)_+ +0,21\left(x_2-73,84\right)_+$$
$$+17,31x_3-8,66\left(x_3-119,93\right)_+ +3,26\left(x_3-123,84\right)_+ -1,08x_4-1,31(x_4-16,77)_+ -0,65\left(x_4-17,68\right)_+$$
$$+7,90x_5+13,95(x_5-44,27)_+ -124,54\left(x_5-46,61\right)_+ +263,98x_6-0,99(x_6-81,77)_+ -0,50\left(x_6-82,78\right)_+$$

The model above produces an $R^2$ value of 53.36 which indicates that the variable percentage of the population smokes over 15 years, the percentage of households having access to proper sanitation, the coverage of national health insurance, the percentage of poor people, the percentage of food processing establishments that meet the requirements standards and the percentage of accredited hospitals have an effect of 53.36% on the success rate of treatment of tuberculosis patients in Indonesia while 46.64% are influenced by other variables.

### 3.5 Comparison of GCV and UBR Methods in Selection of Optimal Knot Points

The minimum GCV and UBR values for each knot point have been obtained in the previous discussion. Then choose the optimal knot point by comparing the GCV and UBR values based on the smallest Mean Square Error (MSE) value and the largest value.

**Table 7. Comparison of the GCV Method and the UBR Method**

| Method | Number of Knot Points | Minimum Value | MSE | $R^2$ |
|---|---|---|---|---|
| **GCV** | **3 Point Knot** | **34.90** | **3.65** | **97.04** |
| UBR | 2 Point Knot | $2,64\times10^{-12}$ | 57.48 | 53.36 |

Based on **Table 7** it can be seen that the GCV method with three knots produces an MSE value of 3.65 with a $R^2$ value of 97.04 while the UBR method produces an MSE value of 57.48 with a $R^2$ value of 53.36. So it can be concluded that the GCV method using three knot points is more appropriate to use to select the optimal knot point in the case of the success rate of treatment of tuberculosis patients in Indonesia in 2020.

## 4. CONCLUSIONS

Based on the results of the analysis and discussion that has been carried out in the previous section, the following conclusions are obtained.

1. Estimation of the best truncated spline nonparametric regression model based on the GCV knot point selection method in data on the success rate of treatment of tuberculosis patients in Indonesia in 2020 comes from three knot points. The estimation results of the model are as follows.

$$\hat{y}=19,69+0,88x_1+0,45(x_1-27,62)_+ +0,01(x_1-28,15)_+ -0,60x_2-0,20(x_2-71,53)_+ + 0,21(x_2-73,84)_+ +17,31x_3-8,66(x_3-119,93)_+ +3,26(x_3-123,84)_+ -1,08x_4 - 1,31(x_4-16,77)_+ -0,65(x_4-17,68)_+ +7,90x_5+13,95(x_5-44,27)_+ -124,54(x_5-46,61)_+ + 263,98x_6-0,99(x_6-81,77)_+ -0,50(x_6-82,78)_+$$

2. Estimation of the best truncated spline nonparametric regression model based on the UBR knot point selection method in data on the success rate of treatment of tuberculosis patients in Indonesia in 2020 comes from two knot points. The estimation results of the model are as follows.

$$\hat{y}=19,69+0,88x_1+0,45(x_1-27,62)_+ +0,01(x_1-28,15)_+ -0,60x_2-0,20(x_2-71,53)_+ + 0,21(x_2-73,84)_+ +17,31x_3-8,66(x_3-119,93)_+ +3,26(x_3-123,84)_+ -1,08x_4 - 1,31(x_4-16,77)_+ -0,65(x_4-17,68)_+ +7,90x_5+13,95(x_5-44,27)_+ -124,54(x_5-46,61)_+ + 263,98x_6-0,99(x_6-81,77)_+ -0,50(x_6-82,78)_+$$

3. The best model with the GCV method produces an MSE value of 3.65 with an $R^2$ value of 97.64. While the UBR method produces an MSE value of 57.48 with an $R^2$ value of 53.36. So it can be concluded that the GCV method is more appropriate to use to obtain optimal knot point values in this study.

## REFERENCES


[1] S. Sifriyani, S. H. Kartiko, and G. Gunadi, "Development of Nonparametric Geographically Weighted Regression Using Truncated Spline Approach," *Songklanakarin J. Sci. Technol.*, vol. 40, no. 2, 2018.

[2] S. Sifriyani, H. Ilmi, and Z. Mar'ah, "Application of Nonparametric Geographically Weighted Spline Regression Model for Spatial Mapping of Open Unemployment Rate in Kalimantan Application of Nonparametric Geographically Weighted Spline Regression Model for Spatial Mapping of Open Unemployment," *J. Phys. Conf. Ser.*, vol. 2123, no. 1, p. 012038, 2021.

[3] F. R. Sihotang, S. Sifriyani, and S. Prangga, "Aplikasi Regresi Nonparametrik Spline Birespon Pada Data Kualitas Air Di Das Mahakam," *Media Bina Ilm.*, vol. 16, no. 3, pp. 6515–6526, 2021.

[4] S. Sifriyani, I. N. Budiantara, S. H. Kartiko, and G. Gunadi, "A New Method of Hypothesis Test for Truncated Spline Nonparametric Regression Influenced by Spatial Heterogeneity and Application, Abstract and Applied Analysis," vol. 2018.

[5] A. B. Padatuan, S. Sifriyani, and S. Prangga, "Bayi Di Kalimantan Dengan Regresi Nonparametrik Spline Birespon," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 15, no. 2, pp. 283–296, 2021.

[6] S. Sifriyani, "Simultaneous Hypothesis Testing of Multivariable Nonparametric Spline Regression in the GWR Model," *Int. J. Stat. Probab.*, vol. 8, no. 4, pp. 32–46, 2019.

[7] T. Purnaraga and S. Prangga, "Regresi Nonparamaetrik Spline Pada Data Laju Pertumbuhan Ekonomi Di Kalimantan," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 14, no. 3, pp. 343–356, 2020.

[8] R. N. Zarkasi, S. Sifriyani, and S. Prangga, "Identifikasi Faktor-Faktor Yang Mempengaruhi Menggunakan Regresi Panel," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 15, no. 2, pp. 277–282, 2021.

[9] E. M. Purbasari, T. W. Utami, and M. Al Haris, "Perbandingan Pemilihan Titik Knot Optimal Menggunakan Metode Unbiased Risk (UBR) Dan Generalized Cross Validation (GCV) Dalam Regresi Nonparametrik Spline (Studi Kasus: Jumlah Kematian Bayi di Provinsi Jawa Tengah 2017)," 2019.

[10] W. Nurmalisa and I. M. Nur, "Pemodelan Regresi Nonparametrik Spline dengan Pemilihan Titik Knot Optimal Menggunakan Metode Unbiased Risk dan Generalized Cross Validation pada Kasus Indeks Pembangunan Manusia di Jawa Tengah," *Pros. Semin. Nasioanl Mhs. Unimus*, vol. 2, pp. 344–355, 2019.

[11] Kementerian Kesehatan Republik Indonesia, *Profil Kesehatan Indonesia 2020*, no. 1. Jakarta, 2021.

[12] M. H. Pratama, Sifriyani, and D. A. Nohe, "Regresi Nonparametrik Multivariabel dengan Pendekatan Spline Truncated Pada Kasus Tuberculosis," vol. 22, no. 1, pp. 87–93, 2022.

[13] N. Evitasari, S. S. Handajani, and H. Pratiwi, "Pemodelan Faktor-Faktor Yang Memengaruhi Kesembuhan Tuberkulosis Di Provinsi Jawa Timur Dengan Regresi Nonparametrik *Spline Truncated*," *Semin. Nas. Pendidik. Mat.*, vol. 3, no. 1, pp. 107–118, 2020.

[14] H. Yozza and A. Afrimayani, "Konstruksi Model Hubungan Dua Variabel Dengan Analisis Regresi *Spline*," *Sainstek J. Sains dan Teknol.*, vol. 11, no. 1, p. 16, 2019.

[15] E. K. Litawati and I. N. Budiantara, "Pendekatan Regresi Nonparametrik Spline Untuk Pemodelan Laju Pertumbuhan Ekonomi

( LPE ) di Jawa Timur,” *J. Sains Dan Seni POMITS*, vol. 2, no. 2, pp. 2337–3520, 2013.
[16] M. D. Pasarella, S. Sifriyani, and F. D. T. Amijaya, “Nonparametric Regression Model Estimation With The Fourier Series The Fourier Series Approach And Its Application To The Accumulative Covid-19 Data In Indonesia,” *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 16, no. 4, pp. 1167–1174, 2022.
[17] S. Sifriyani, M. Rasjid, D. Rosadi, S. Anwar, R. D. Wahyuni, and S. Jalaluddin, “Spatial-Temporal Epidemiology of COVID-19 Using a Geographically and Temporally Weighted Regression Model,” *Symmetry (Basel)*, vol. 14, no. 4,2022.
[18] Z. Tupti, N. I. Purnama, and D. Novita, “Analisis Faktor Determinasi Kinerja Karyawan,” *Proceding Semin. Nas. Kewirausahaan*, vol. 2, no. 1, pp. 512–524, 2021.
[19] M. Sriningsih, D. Hatidja, and J. D. Prang, “Penanganan Multikolinearitas Dengan Menggunakan Analisis Regresi Komponen Utama Pada Kasus Impor Beras Di Provinsi Sulut,” *J. Ilm. Sains*, vol. 18, no. 1, 2018.