Optimal Knots Point and Bandwidth Selection in Modeling Mixed Estimator Nonparametric Regression

A T R Dani, V Ratnasari*, and I N Budiantara

Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember, Surabaya, East Java, Indonesia

*Corresponding author: vita_ratna@statistika.its.ac.id

Abstract. The nonparametric regression model that is currently developing is limited to only one estimator form that is used. This happens, due to the assumption from the researcher that each of these predictors has the same data pattern. However, when we model the predictor variable with the response variable, what happens is that each predictor may have a different pattern, so it is necessary to develop a mixed estimator. The mixed estimator used in this study is the truncated spline and the Gaussian Kernel. Furthermore, it becomes a separate problem, when combining the truncated spline with the Gaussian Kernel, so it must determine the optimal knot point and bandwidth. Determine the optimal knot point and bandwidth is very important. The methods used are Cross-Validation (CV), Generalized Cross-Validation (GCV), and Unbiased Risk (UBR). In this study, the nonparametric regression model was applied to the data on the percentage of poor people in Districts/Cities on Borneo Island in 2019. Based on analysis, the modeling of the nonparametric regression approach with a mixed estimator of the truncated spline and the Gaussian Kernel uses the GCV method which gives the best results. This is supported by the coefficient of determination obtained by 90.52% and the MSE value of 1.10. This means that the nonparametric regression model built can explain the effect of the predictor variable on the response variable, namely the percentage of poor people by 90.52%.

1. Introduction

Regression analysis is a statistical method used to determine the pattern of relationships between one or more variables. The variables used in the regression analysis consist of response and predictor variables [1]. The purpose of regression analysis is to find the form of parameter estimation that matches the shape of the regression curve [2]. There are several approaches developed by researchers in statistical modeling using regression analysis, namely parametric regression, nonparametric regression, and semiparametric regression [3]. Recently, the approach using nonparametric regression has received a lot of attention from researchers [4]. The nonparametric regression approach does not depend on the assumption of a certain regression curve shape, so it will provide high flexibility, where it is expected that the data will adjust itself to the estimated form of the regression curve without being influenced by the researcher's subjectivity [5].

In nonparametric regression modeling, there are several estimators developed by researchers including Kernel [6, 7], Spline [8, 9], and Fourier series [10, 11]. Among the several methods, nonparametric regression with the spline and Kernel approach is an estimator method that is usually used in research. One of the advantages of the spline is that it is flexible, meaning that this function tends to find its own form of data estimation wherever the data pattern moves [12]. Apart from spline, the nonparametric regression approach with Kernel functions is also of interest. The kernel has a more flexible form, very good at modeling data that does not have a certain pattern [13].
The nonparametric regression model that is currently developing is limited to only one form of an estimator for all predictor variables used. But in fact, when we model predictor variables with response variables, what happens is that each predictor may have a different pattern [14]. Then a mixed estimator was developed by [14, 15, 16].

In nonparametric regression truncated spline, determine the optimal knot point is very important and crucial [17]. The same thing applies to Kernel regression, where the thing that must be considered is the choice of bandwidth [13]. It becomes a separate problem when using a mixed estimator model, especially combining the spline with the Kernel, so it must determine the optimal knot point and bandwidth. Determination of the optimal point of knots and bandwidth will greatly affect the regression curve that will be formed. To choose the optimal knot point and bandwidth parameters, you can use various methods that have been developed by researchers, for example Cross-Validation (CV) by [18], Generalized Cross-Validation (GCV) by [8], and Unbiased Risk (UBR) by [19].

In this study, a nonparametric regression model was applied to analyze problems related to poverty. Based on BPS data (2020), in March 2019 the percentage of poor people in Indonesia was 9.41% with the number of poor people in March 2019 of 25.14 million people. Poverty in brief is often understood as a low level of welfare [20]. The problem of poverty is a very serious problem faced by countries in the world, especially developing countries including Indonesia [21]. Poverty eradication is the main goal in the Sustainable Development Goals (SDG’s) for 2016 to 2030.

The purpose of this research is to examine the estimator form of a mixed nonparametric regression truncated spline and Gaussian Kernel, which is then applied to the data on the percentage of poor people in Districts/Cities on Borneo Island in 2019. The method of selecting the optimal knot point and bandwidth uses CV, GCV, and UBR. The nonparametric regression approach, especially the estimator of the mixed of the truncated spline and the Gaussian Kernel, using the correct predictor variables, is expected to be a statistical model capable of analyzing poverty problems.

2. Methodology

2.1. Truncated Spline Nonparametric Regression

A truncated spline nonparametric regression approach is used if the regression curve between the response and the predictor variable does not form a pattern [22]. The data characteristic of the truncated spline estimator is a changing pattern at certain sub-intervals [2]. The study of nonparametric regression truncated multivariable spline on paired data \((x_i, x_{i1}, \ldots, x_{iq}, y_i)\) which assumed the relationship between the predictor variable and the response variable followed a nonparametric regression model.

\[ y_i = f(x_{i1}, x_{i2}, \ldots, x_{iq}) + \varepsilon_i, i = 1, 2, \ldots, n \]  

(1)

Furthermore, the regression curve is assumed to be additive:

\[ f(x_{i1}, x_{i2}, \ldots, x_{iq}) = f_1(x_{i1}) + f_2(x_{i2}) + \cdots + f_q(x_{iq}) = \sum_{p=1}^{q} f_p(x_{ip}), i = 1, 2, \ldots, n \]  

(2)

The component of degree 1 (linear) truncated spline nonparametric regression for each predictor variable can be written in general:

\[ f_p(x_{ip}) = \delta_{0p} + \sum_{j=1}^{m} \sum_{p=1}^{q} \delta_{jp} x_{ip} + \sum_{k=1}^{r} \sum_{p=1}^{q} \phi_{j,k,p} (x_{ip} - K_{ip})^m \]  

(3)

Equation (3) can be presented in the form of a matrix:
\[ \sum_{p=1}^{q} \tilde{f}_{p}(x_{pi}) = X_{q} \delta + X_{1}(K_{1}) \hat{\theta}_{1} + X_{2}(K_{2}) \hat{\theta}_{2} + \ldots + X_{q}(K_{q}) \hat{\theta}_{q} \]  

(4)

Based on Equation (4), so \( \sum_{p=1}^{q} \tilde{f}_{p}(x_{pi}) = X(\phi)\hat{\beta} \).

2.2. Gaussian Kernel Nonparametric Regression

The Kernel estimator is a linear estimator that is similar to other estimators, but the difference is that there is bandwidth in the Kernel [9]. Suppose given paired data \( t_{i} \) and \( y_{i} \), where \( i = 1, 2, \ldots, n \) follows a nonparametric regression model. Furthermore, the relationship \( t_{i} \) and \( y_{i} \) can be modeled functionally into:

\[ y_{i} = h(t_{i}) + \varepsilon_{i}; i = 1, 2, \ldots, n \]  

(5)

The regression curve of \( h(t_{i}) \) is an unknown shape. It to be approximated by the Kernel function, with the regression curve estimation is presented in Equation (6):

\[ \hat{h}_{a}(t) = n^{-1} \left[ \sum_{i=1}^{n} W_{ai}(t) \right] y_{i} \]  

(6)

where \( W_{ai}(t) = \frac{K_{a}(t-t_{i})}{n^{-1} \sum_{j=1}^{n} K_{a}(t-t_{j})} \) and \( K_{a}(t-t_{i}) = \frac{1}{\alpha} K\left( \frac{t-t_{i}}{\alpha} \right) \).

The Kernel function used is the Gaussian Kernel: \( K(z) = \frac{1}{\sqrt{2\pi}} \exp \left( \frac{1}{2} (-z^{2}) \right) ; I_{[-\alpha,\alpha]}(z) \).

Based on the Kernel function that applies to each \( t = t_{1}, t = t_{2}, \ldots, t = t_{n} \), then Equation (6) can be written in form of a matrix to be \( D(\alpha)\tilde{Y} \).

3. Result and Discussion

3.1. Mixed Estimator Nonparametric Regression Truncated Spline and Gaussian Kernel

Given a pair of data with a predictor variable of \( q \) for the truncated spline component and only 1 for the Kernel component such that it can be written \( (x_{q}, x_{2}, \ldots, x_{q}, t_{i}, y_{i}) \), with \( i = 1, 2, \ldots, n \). A nonparametric regression model of truncated spline and Gaussian Kernel can be written:

\[ y_{i} = \sum_{p=1}^{q} f_{p}(x_{pi}) + h(t_{i}) + \varepsilon_{i} \]  

(7)

Based on Equation (7), it can be written briefly \( y_{i} = \mu(x_{i}, t_{i}) + \varepsilon_{i} \). Furthermore, it can be presented in the form of a matrix to be:

\[ \tilde{Y} = X(\phi)\tilde{\beta} + D(\alpha)\tilde{Y} + \tilde{\varepsilon} \]  

(8)

The method of parameter estimation used in this study is Ordinary Least Squares (OLS). The sum squared of error can be found with:
$$Q(\beta | \phi, \alpha) = \sum_{i=1}^{n} \hat{e}_i^2 = \hat{e}^T \hat{e}$$
$$= \left\| \left[ \mathbf{I} - \mathbf{D}(\alpha) \right] \bar{Y} \right\|^2 - 2 \hat{\beta}^T \mathbf{X}(\phi)^T \left[ \mathbf{I} - \mathbf{D}(\alpha) \right] \bar{Y} + \hat{\beta}^T \mathbf{X}(\phi)^T \mathbf{X}(\phi) \hat{\beta}$$

(9)

To get an estimator of $\hat{\beta}$, you can do a partial derivative of Equation (9) on $\hat{\beta}$ as follows:

$$\frac{\partial Q(\hat{\beta} | \phi, \alpha)}{\partial \hat{\beta}} = \frac{\partial}{\partial \hat{\beta}} \left( \left\| \left[ \mathbf{I} - \mathbf{D}(\alpha) \right] \bar{Y} \right\|^2 - 2 \hat{\beta}^T \mathbf{X}(\phi)^T \left[ \mathbf{I} - \mathbf{D}(\alpha) \right] \bar{Y} + \hat{\beta}^T \mathbf{X}(\phi)^T \mathbf{X}(\phi) \hat{\beta} \right)$$
$$= -2 \mathbf{X}(\phi)^T \left[ \mathbf{I} - \mathbf{D}(\alpha) \right] \bar{Y} + 2 \mathbf{X}(\phi)^T \mathbf{X}(\phi) \hat{\beta}$$

(10)

Equation (10) is equal to zero:

$$\frac{\partial Q(\hat{\beta} | \phi, \alpha)}{\partial \hat{\beta}} = 0$$

(11)

Then it will be obtained:

$$\mathbf{X}(\phi)^T \mathbf{X}(\phi) \hat{\beta} = \mathbf{X}(\phi)^T \left[ \mathbf{I} - \mathbf{D}(\alpha) \right] \bar{Y}$$

(12)

So that

$$\mathbf{X}(\phi)^T \mathbf{X}(\phi) \hat{\beta} = \mathbf{X}(\phi)^T \left[ \mathbf{I} - \mathbf{D}(\alpha) \right] \bar{Y}$$

(13)

Then we will get an estimator for the following:

$$\hat{\beta} = \left[ \mathbf{X}(\phi)^T \mathbf{X}(\phi) \right]^{-1} \mathbf{X}(\phi)^T \left[ \mathbf{I} - \mathbf{D}(\alpha) \right] \bar{Y}$$

(14)

Equation (14) can be written as $\hat{\beta} = \mathbf{C}(\phi, \alpha) \bar{Y}$, with $\mathbf{C} = \left[ \mathbf{X}(\phi)^T \mathbf{X}(\phi) \right]^{-1} \mathbf{X}(\phi)^T \left[ \mathbf{I} - \mathbf{D}(\alpha) \right]$.

Truncated spline curve estimator form is:

$$\sum_{p=1}^{q} \hat{f}_p(x_p) = \mathbf{X}(\phi) \hat{\beta}$$

(15)

And we can write it:

$$\sum_{p=1}^{q} \hat{f}_p(x_p) = \mathbf{X}(\phi) \hat{\beta} \rightarrow \mathbf{X}(\phi) \left[ \mathbf{X}(\phi)^T \mathbf{X}(\phi) \right]^{-1} \mathbf{X}(\phi)^T \left[ \mathbf{I} - \mathbf{D}(\alpha) \right] \bar{Y}$$

(16)

Equation (16) can be written as $\sum_{p=1}^{q} \hat{f}_p(x_p) = \mathbf{A}(\phi, \alpha) \bar{Y}$. Then we can write the Gaussian Kernel estimator form $\hat{h}_\alpha(t) = \mathbf{D}(\alpha) \bar{Y}$.

Based on Equation (14) and the estimator form of each component, we will get a mixed estimator of the nonparametric regression truncated spline and Gaussian Kernel as follows:

$$\hat{\mu}_{p, \alpha}(x, t) = \sum_{p=1}^{q} \hat{f}_p(x_p) + \hat{h}_\alpha(t)$$
$$= \left[ \mathbf{A}(\phi, \alpha) + \mathbf{D}(\alpha) \right] \bar{Y}$$
$$= \mathbf{B}(\phi, \alpha) \bar{Y}$$

(17)
Matrix $B(\phi, \alpha)$ is highly dependent on $A(\phi, \alpha)$ which is a truncated spline component with knots $\phi = (K_1, K_2, \ldots, K_T)^T$ and $D(\alpha)$ which is a component of the Gaussian Kernel with bandwidth.

### 3.2. Methods for Selecting Knot Points and Bandwidth

The method used in the process of selecting optimal knots and bandwidth points is Cross-Validation (CV), Generalized Cross-Validation (GCV), and Unbiased Risk (UBR). The CV, GCV, and UBR equations have been modified based on the nonparametric regression model used, namely the mixed estimator.

1. **Cross-Validation (CV)**

   $$CV(\phi_{opt}, \alpha_{opt}) = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{(y_i - \hat{y}_i)}{1 - \left[ A_u(\phi, \alpha) + D_u(\alpha) \right]^T} \right)^2$$

2. **Generalized Cross-Validation (GCV)**

   $$GCV(\phi_{opt}, \alpha_{opt}) = \frac{\text{MSE}(\phi, \alpha)}{\left( \frac{1}{n}\text{trace}\left[ I - \left[ A(\phi, \alpha) + D(\alpha) \right]\right] \right)^2}$$

3. **Unbiased Risk (UBR)**

   $$UBR(\phi_{opt}, \alpha_{opt}) = n^{-1} \left\{ \frac{\| (I - [A(\phi, \alpha) + D(\alpha)])Y \|^2}{\text{trace}\left[ (I - [A(\phi, \alpha) + D(\alpha)])^T (I - [A(\phi, \alpha) + D(\alpha)]) \right]} + \frac{\hat{\sigma}^2}{n} \text{trace}\left[ A(\phi, \alpha) \right] \right\}$$

   with $\hat{\sigma}^2 = \frac{\| (I - [A(\phi, \alpha) + D(\alpha)])Y \|^2}{\text{trace}\left[ (I - [A(\phi, \alpha) + D(\alpha)])^T (I - [A(\phi, \alpha) + D(\alpha)]) \right]}$

### 3.3. Data Application

The data used in this study are secondary data for 2019 published by the Badan Pusat Statistik (BPS). The observation unit is the Districts/Cities on the Borneo Island, which currently of 5 Provinces. The variables used in this study are shown in Table 1.

The initial observations in this study were 56 Districts/Cities on the Borneo Island. Furthermore, the cities of Bontang, Banjarmasin, and Pontianak were not used in research observations because they were outlier data. The observations in this study were left 53 Districts/Cities.

The pattern of relationship between the response variable and each predictor variable can be seen from the scatter plot in Figure 1.

Based on Figure 1, a nonparametric regression model will be applied using a mixed estimator of the truncated spline and Gaussian Kernel. This is because there is a pattern of relationship between predictor and response variables that tend to fluctuate in certain sub-intervals $(x_i - x_k)$ and there are also undefined pattern $(t_i)$. 

---

5


| Variable   | Notation | Description                                           |
|------------|----------|-------------------------------------------------------|
| Response   | $y$      | Percentage of Poor Population                         |
| $x_1$      |          | Life Expectancy                                       |
| $x_2$      |          | Open Unemployment Rate                                |
| $x_3$      |          | Average Length of Schooling                           |
| Predictor  | $x_4$    | Economic Growth Rate                                  |
| $x_5$      |          | Percentage of Households with Adequate Access to Sanitation |
| $x_6$      |          | Percentage of Households with Access to Adequate Drinking Water |
| $t_1$      |          | Population Density                                    |

Figure 1. Scatter plot between predictor and response variables
The best estimator model for the mixed of truncated spline and Gaussian Kernel, will be selected by looking at the smallest CV, GCV and UBR values from several models based on the number of knot points and bandwidth. This study uses the same number of points of knots for each predictor variable used, namely 1 to 3 points of knots.

**Table 2 Cross-Validation (CV) method**

| Knot Point Locations | Bandwidth | CV   | MSE  | R²   |
|----------------------|-----------|------|------|------|
|                      |           |      |      |      |
| x₁ x₂ x₃ x₄ x₅ x₆  | t₁        |      |      |      |
| 1 knot               | 69.65 5.16 9.23 5.73 65.71 62.46 | 2.55 | 4.27 | 2.53 | 66.00% |
| 2 knots              | 67.01 3.92 7.90 4.40 49.85 40.97 | 2.18 | 4.27 | 2.31 | 73.35% |
| 3 knots              | 66.63 3.74 7.71 4.22 47.59 37.89 | 1.41 | 4.18 | 1.82 | 83.80% |

**Table 3 Generalized Cross-Validation (GCV) method**

| Knot Point Locations | Bandwidth | GCV | MSE  | R²   |
|----------------------|-----------|-----|------|------|
|                      |           |     |      |      |
| x₁ x₂ x₃ x₄ x₅ x₆  | t₁        |     |      |      |
| 1 knot               | 69.65 5.16 9.23 5.72 65.71 62.47 | 2.55 | 5.88 | 2.52 | 66.00% |
| 2 knots              | 71.16 5.87 9.99 6.48 74.77 74.75 | 1.82 | 3.09 | 1.29 | 85.10% |
| 3 knots              | 64.37 2.68 6.57 3.08 34.00 19.47 | 1.09 | 1.51 | 1.10 | 90.52% |

**Table 4 Unbiased Risk (UBR) method**

| Knot Point Locations | Bandwidth | UBR  | MSE  | R²   |
|----------------------|-----------|------|------|------|
|                      |           |      |      |      |
| x₁ x₂ x₃ x₄ x₅ x₆  | t₁        |      |      |      |
| 1 knot               | 73.80 7.11 11.32 7.80 90.62 96.24 | 3.09 | 2.68×10⁻⁴ | 2.79 | 60.13% |
| 2 knots              | 71.92 6.22 10.37 6.86 79.30 80.89 | 2.58 | 1.33×10⁻⁵ | 2.48 | 70.12% |
| 3 knots              | 64.37 2.68 6.57 3.08 34.00 19.47 | 1.92 | 1.79×10⁻⁶ | 2.18 | 79.23% |

Based on Tables 2, 3, and 4, it is known that the CV, GCV, and UBR methods produce high R² values and low MSE values when using 3 knots.

Based on Table 3, it can be seen that by using the GCV method in the process of selecting optimal knots and bandwidth points in the mixed estimator nonparametric regression model, better results are obtained than the CV and UBR methods. This is because, with the GCV method, high performance and accuracy are obtained, the R² value is 90.52% and the MSE value is 1.10. Visualization of the value with the best mixed estimator nonparametric regression model is shown in Figure 2.
Based on Figure 2, it can be seen that the value obtained from nonparametric regression modeling used a mixed estimator of the truncated spline and the Gaussian Kernel with the GCV method tends to follow the actual data pattern. Furthermore, the nonparametric regression model using mixed estimator with three knot points obtained from the GCV method is written in Equation (21).

\[
\hat{y} = \hat{\delta}_1 y_1 + \hat{\delta}_2 y_2 + \hat{\delta}_3 y_3 + \hat{\delta}_4 y_4 + \hat{\delta}_5 y_5 + \hat{\phi}_1 (x_1 - 64,37) + \hat{\phi}_2 (x_2 - 71,54) + \\
\hat{\phi}_3 (x_3 - 72,30) + \hat{\phi}_4 (x_4 - 2,68) + \hat{\phi}_5 (x_5 - 6,05) + \hat{\phi}_6 (x_6 - 6,40) + \hat{\phi}_7 (x_7 - 6,57) + \\
\hat{\phi}_8 (x_8 - 10,18) + \hat{\phi}_9 (x_9 - 10,56) + \hat{\phi}_{10} (x_{10} - 3,08) + \hat{\phi}_{11} (x_{11} - 6,66) + \hat{\phi}_{12} (x_{12} - 7,05) + \\
\hat{\phi}_{13} (x_{13} - 34,00) + \hat{\phi}_{14} (x_{14} - 77,03) + \hat{\phi}_{15} (x_{15} - 81,56) + \hat{\phi}_{16} (x_{16} - 19,47) + \hat{\phi}_{17} (x_{17} - 77,82) + \\
\hat{\phi}_{18} (x_{18} - 83,96) + \frac{1}{55} \sum_{i=1}^{55} \frac{1}{1,09} \left( \frac{t - t_i}{1,09} \right)^2 X_{18}^T X_{18} \right] \hat{y}.
\]

(21)

This means that the mixed estimator nonparametric regression model built is able to explain the effect of the predictor variable on the response variable, namely the percentage of poor people based on the $R^2$ value of 90.52%.

4. Conclusion

Based on the analysis, we get a nonparametric regression model using mixed estimator truncated spline and Gaussian Kernel as follows:

\[
\hat{Y} = X(\phi) \hat{\beta} + D(\alpha) \hat{\mu} + \hat{\epsilon}
\]

where estimate results from $\hat{\beta}$ is $X(\phi)^T X(\phi)^{-1} X(\phi)^T [I - D(\alpha)] \hat{y}$. In general, the nonparametric regression model using mixed estimator can be written as:
The results of the nonparametric regression model using mixed estimator truncated spline and Gaussian Kernel using 3 knot points with the GCV method, obtained a $R^2$ value of 90.52% and MSE value of 1.10. This means that the nonparametric regression model built is able to explain the effect of the predictor variable on the response variable, namely the percentage of poor people by 90.52%.

Acknowledgments

Thanks to the Education and Culture Office of East Borneo Province for providing “Kaltim Tuntas Scholarship” assistance to complete Postgraduate Education at the Sepuluh Nopember Institute of Technology Surabaya.

References

[1] Draper N R and Smith H 1992 *Applied Regression Analysis: Third Edition* USA: John Wiley and Sons, Inc
[2] Budiantara I N 2019 *Regresi Nonparametrik Spline Truncated* Surabaya: ITS Press
[3] Dani A T R, Adrianingsih N Y and Ainurrochmah A 2020 Pengujian Hipotesis Simultan Model Regresi Nonparametrik Spline Truncated dalam Pemodelan Kasus Ekonomi *JAMBURA Journal of Probability and Statistics* 1 98-106
[4] Prahutama A, Suparti and Utami T W 2018 Modelling Fourier Regression for Time Series Data with Case Study Modelling Inflation in Foods Sector in Indonesia *IOP Conf. Series: Journal of Physics* 974 1-9
[5] Devi A R, Budiantara I N and Ratnasari V 2018 Unbiased Risk and Cross-Validation Method for Selecting Optimal Knots in Multivariable Nonparametric Regression Spline Truncated AIP Conf. Proc 2021 060003
[6] Nadaraya E A 1989 *Nonparametric Estimation of Probability Densities and Regression Curves* London: Kluwer Academic Publisher
[7] Hardle W 1994 *Applied Nonparametric Regression* Berlin: Humboldt-Universitat zu Berlin
[8] Wahba G 1990 *Spline Models for Observational Data* Pennsylvania: SIAM
[9] Eubank R L 1999 *Nonparametric Regression and Spline Smoothing* New York: Marcel Dekker
[10] Bilodeau M 1992 Fourier Smoother and Additive Models *Canadian Journal of Statistics* 20 257-269
[11] Morton J and Silverberg L 1999 Fourier Series of Half-range Functions by Smooth Extension *Applied Mathematical Modelling* 33 812-821
[12] Saputro D R S, Demu K R and Widyantingsih P 2018 Nonparametric Truncated Spline Regression Model on the Data of Human Development Index (HDI) in Indonesia *IOP Conf. Series: Journal of Physics* 1028 1-4
[13] Aydin D, Gneri O I and Fit A 2016 Choice of Bandwidth for Nonparametric Regression Models using Kernel Smoothing: A Simulation Study *International Journal of Sciences Basic and Applied Research* 26 47-61
[14] Budiantara I N, Ratnasari V, Madu R and Zain I 2015 The Combination of Spline and Kernel Estimator for Nonparametric Regression and Its Properties *Applied Mathematical Sciences* 9 6083-6094
[15] Ratnasari V, Budiantara I N, Ratna M and Zain I 2016 Estimation of Nonparametric Regression Curve Using Mixed Estimator of Multivariable Truncated Spline and Multivariable Kernel *Global Journal of Pure and Applied Mathematics* 12 5047-5057
[16] Hidayat R, Budiantara I N, Otok B W and Ratnasari V 2019 Kernel-Spline Estimation of Additive Nonparametric Regression Model IOP Conf. Series: Materials Science and Engineering 546 1-9

[17] Setiawan R N S, Budiantara I N and Ratnasari V 2017 Application of Confidence Intervals for Parameters of Nonparametric Spline Truncated Regression on Index Development Gender in East Java IPTEK Journal of Sciences 2 49-55

[18] Craven P and Wahba G 1979 Smoothing Noisy Data with Spline Functions Numerische Mathematics 31 377-403

[19] Wang Y 1997 GRKPACK: Fitting Smoothing Spline Analysis of Variances Models to Data from Exponential Families Communication in Statistics: Simulation and Computation 26 756-782

[20] Badan Pusat Statistik 2020 Provinci Kalimantan Timur dalam Angka 2020 Kaltim: BPS

[21] Ferezagia D V 2018 Analisis Tingkat Kemiskinan di Indonesia Jurnal Sosial Humaniora 1 1-6

[22] Montoya E L, Ulloa N and Miler V 2014 A Simulation Study Comparing Knot Selection Methods with Equality Spaced Knots in a Penalized Regression Spline International Journal of Statistics and Probability 3 96-110