Mixture Model Nonparametric Regression and Its Application

Narita Yuri Adrianingsih\(^1\), I Nyoman Budiantara\(^2\)*, Jerry Dwi Trijoyo Purnomo\(^3\)

Department of Statistics, Institut Teknologi Sepuluh Nopember; Arif Rahman Hakim Street, Surabaya 60111, Indonesia

naritayuria98@gmail.com\(^1\), nyomanbudiantara65@gmail.com\(^2\)*, jerry@statistika.its.ac.id\(^3\)

Abstract. Regression analysis is a statistical analysis used to determine the pattern of relationships between predictor variables and response variables. There are two models of estimation approaches in regression analysis, namely parametric regression, and nonparametric regression. The parametric regression approach is used in the shape of the regression curve is known. In cases with unknown relationship patterns, the development is done using nonparametric regression. Nonparametric regression is a model estimation method which is based on an approach that is not bound by certain assumptions of the regression curve shape. Nonparametric regression varies greatly with variable curves that are different between one predictor variable with another predictor variable. In nonparametric regression, there are several types of the recommended kernel, spline, and Fourier series. In many cases, however, these conservative nonparametric regression methods cannot handle more complex problems. Mixture methods by combining several methods such as a mixture of spline and Fourier series, kernel and Fourier series, and so on, give a better result. This study aims to obtain estimates of a mixture of a truncated spline, kernel, and Fourier series by using the Ordinary Least Square (OLS) method and obtain methods for selecting knots, bandwidth, and optimal oscillation parameters with the smallest GCV. The results of this study are the formulation of a mixed estimation model of the truncated spline, kernel, and Fourier series and the smallest GCV formula to obtain the optimum location and number of points of knots, bandwidth, and oscillations.

1. Introduction
Regression analysis is a statistical method of a regression function or regression curve [1]. The main objective of regression analysis is to provide a means of exploring the relationship between response variables and predictors, and making predictions [2]. In the regression analysis, there are 3 types of models, namely the parametric regression model, the nonparametric model, and the semiparametric model. Early identification of the relationship pattern can be done using a scatter plot. The nonparametric regression model is very good for use for unknown data patterns because it has high flexibility, where the data is expected to find its own form of regression quotient estimation without being influenced by the researcher's subjectivity [1]. There are several types of estimators in nonparametric regression, namely kernel, spline truncated, Fourier series, local polynomial, and wavelet. There have been many studies for several types of estimators, such as those conducted by Prahatama [3], Tripena and Budiantara [4], Budiantara et al [5], Wisisono et al [6], Saputro et al [7] [8], Antomiadis et al. [9].
Truncated spline regression is of considerable interest to some researchers because truncated spline estimators are able to handle subtle characteristics of data or functions or data that has changed behavior at certain sub-intervals [10]. The kernel estimator is a development of the histogram estimator. This estimator is a linear estimator that is similar to other nonparametric regression estimators, the difference is only because the estimator kernel focuses more on the use of the bandwidth method [1]. The advantages of kernel estimators are its ability to model data that does not have a certain pattern [11], is more flexible, has an easy mathematical form, and can achieve a relatively fast convergence rate [12]. From a computational point of view, it is easier to perform and implement [13]. Fourier Series Meters are trigonometric polynomials that have flexibility so that they can adjust effectively to the nature of local data [14], there is a tendency to repeat itself [15]. The repetitive data pattern referred to is the repetition of the dependent variable values for different independent variables [3].

Based on the advantages possessed by the truncated spline estimator, kernel, and Fourier series, recently there have been many research developments regarding the mixture of the truncated spline estimator, kernel, and Fourier series that have been carried out by researchers such as Ratnasari et al [16], Sudiarsa et al [17], Afifah [18], Budiantara et al [19], Suparti et al [20], Budiantara et al [21].

Based on the results of previous studies, it will be carried out the development of a nonparametric regression estimator for a mixed additive model between the truncated spline, kernel, and Fourier series obtained through Ordinary Least Squares (OLS). This research is focused and expected to have better interpretations and predictions, namely handling data characters that follow repetitive patterns at certain intervals, have unknown patterns, and have changing patterns in certain sub-intervals, and determine the GCV formula from the model. the mixture so as to get the location and many points of optimal knots, bandwidth, and oscillations.

2. Method

2.1. Spline Truncated Nonparametric Regression

The most widely used approach in nonparametric regression is Spline Truncated regression. Spline regression is a polynomial regression in which different polynomial segments are joined at knots and are continuous. One of the advantages of this regression is that this model tends to find its own data estimates wherever the data patterns move and are flexible.

In general, a spline function truncated with degrees $M$ and knots $\xi_1, \xi_2, \ldots, \xi_r$ is a function that can be written in the following form:

$$h(z_i) = \sum_{j=0}^{M} \alpha_j z_i^j + \sum_{k=1}^{r} \beta_k (z_i - \xi_k)_+^M$$

(1)

with truncated function:

$$(z_i - \xi_k)_+^M = \begin{cases} (z_i - \xi_k)^M, & z_i \geq \xi_k \\ 0, & z_i < \xi_k \end{cases}$$

In general, the truncated spline model can be written as follows:

$$y_i = \sum_{j=0}^{M} \alpha_j z_i^j + \sum_{k=0}^{r} \beta_k (z_i - \xi_k)_+^M + \varepsilon_i$$

$$= \alpha_0 + \alpha_1 z_i + \ldots + \alpha_M z_i^M + \beta_1 (z_i - \xi_1)_+^M + \ldots + \beta_r (z_i - \xi_r)_+^M + \varepsilon_i$$

(2)

Equation (1) can be denoted in the form of a matrix to be:

$$h(z_i) = G(\xi)\theta$$
where:

\[
G(\xi) = \begin{bmatrix}
1 & z_1 & \cdots & z_n^M & (z_1 - \xi_1)^M & \cdots & (z_n - \xi_n)^M \\
1 & z_2 & \cdots & z_n^M & (z_2 - \xi_1)^M & \cdots & (z_n - \xi_n)^M \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
1 & z_n & \cdots & z_n^M & (z_n - \xi_1)^M & \cdots & (z_n - \xi_n)^M
\end{bmatrix}
\]

\[
h(z_i) = (h(z_1) \ h(z_2) \ \ldots \ h(z_n))^T
\]

\[
\theta = (\alpha_1 \ \alpha_2 \ \ldots \ \alpha_M \ \beta_1 \ \beta_2 \ \ldots \ \beta_M)^T
\]

Vectors \(h(z_i)\) are sized \(n \times 1\), matrices \(G(\xi)\) are sized \((M + r + 1) \times n\) while vectors \(\theta\) are sized \((M + r + 1) \times 1\).

2.2. Kernel Nonparametric Regression

The advantage of kernel estimators is that they are flexible. The mathematical form is very easy and can reach convergence levels relatively quickly. The regression curve \(g(t_i)\) is approached by the kernel function, the estimation of the regression curve can be written in the equation:

\[
g_a(t) = n^{-1} \sum_{i=1}^{n} \frac{K_a(t-t_i)}{n^{-1} \sum_{i=1}^{n} K_a(t-t_i)} y_i = n^{-1} \sum_{i=1}^{n} W_{a_i}(t) y_i
\]

(3)

If the addition form in the equation is described completely then

\[
g_a(t) = n^{-1}W_{a_1}(t)y_1 + n^{-1}W_{a_2}(t)y_2 + \ldots + n^{-1}W_{a_n}(t)y_n
\]

(4)

If it is notated in matrix form, it will be

\[
g_a(t) = V(\alpha)y
\]

(5)

where

\[
V(\alpha) = \begin{bmatrix}
n^{-1}W_{a_1}(t_1) \ & n^{-1}W_{a_2}(t_1) \ & \cdots \ & n^{-1}W_{a_n}(t_1) \\
n^{-1}W_{a_1}(t_2) \ & n^{-1}W_{a_2}(t_2) \ & \cdots \ & n^{-1}W_{a_n}(t_2) \\
\vdots \ & \vdots \ & \ddots \ & \vdots \\
n^{-1}W_{a_1}(t_n) \ & n^{-1}W_{a_2}(t_n) \ & \cdots \ & n^{-1}W_{a_n}(t_n)
\end{bmatrix}
\]

Vector \(g_a(t)\) is sized \(n \times 1\), vector \(y\) are sized \(n \times 1\), and matrix \(V(\alpha)\) is sized \(n \times n\).

2.3. Nonparametric Regression Fourier Series

Fourier series are trigonometric polynomials that have high flexibility. This Fourier series is used to estimate the regression curve that shows sine and cosine waves. The Fourier series estimator is used if the data investigated is unknown and there is a seasonal trend [4] [15]. The Fourier series equation is as follows:
\[ f(x_i) = bx_i + \frac{1}{2}\alpha_0 + \sum_{k=1}^{K}\alpha_k \cos kx_i \] (6)

where \( b, \alpha_0, \alpha_k, k = 1, 2, \ldots, K \) are model parameters.

The equation can be written as:

\[ f(x_i) = bx_i + \frac{1}{2}\alpha_0 + \alpha_1 \cos x_i + \alpha_2 \cos 2x_i + \ldots + \alpha_K \cos Kx_i \] (7)

If the equation is written in matrix form, the equation becomes:

\[ \mathbf{f}(x) = \mathbf{N}(K)\mathbf{\rho} \] (8)

where

\[ \mathbf{N}(K) = \begin{bmatrix} x_1 & 1 & \cos x_1 & \cos 2x_1 & \cdots & \cos Kx_1 \\ x_2 & 1 & \cos x_2 & \cos 2x_2 & \cdots & \cos Kx_2 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_n & 1 & \cos x_n & \cos 2x_n & \cdots & \cos Kx_n \end{bmatrix} \]

\[ \mathbf{f}(x) = \begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix}, \quad \mathbf{\rho} = \begin{bmatrix} b \\ \frac{1}{2}\alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_K \end{bmatrix} \]

Vector \( \mathbf{f}(x) \) are sized \( n \times 1 \), vector \( \mathbf{\rho} \) are sized \( n \times 1 \), and matrix \( \mathbf{N}(K) \) is sized \( n \times n \).

3. Result and Discussion

Paired data \((z_{i1}, z_{i2}, \ldots, z_{ip}, t_{i1}, t_{i2}, \ldots, t_{ip}, x_{i1}, x_{i2}, \ldots, x_{in}, y_i)\), \( i = 1, 2, \ldots, n \) which have a relationship are assumed to follow the nonparametric regression model. Variables \( z_{i1}, z_{i2}, \ldots, z_{ip}, t_{i1}, t_{i2}, \ldots, t_{ip}, x_{i1}, x_{i2}, \ldots, x_{in} \) are predictor variables and \( y_i \) is response variables. The form of the nonparametric regression model is:

\[ y_i = \mu(z_{i1}, z_{i2}, \ldots, z_{ip}, t_{i1}, t_{i2}, \ldots, t_{ip}, x_{i1}, x_{i2}, \ldots, x_{in}) + \epsilon_i \]

\[ = \mu(z_i, t_i, x_i) + \epsilon_i \] (9)

where \( z_i = (z_{i1}, z_{i2}, \ldots, z_{ip})', t_i = (t_{i1}, t_{i2}, \ldots, t_{ip})', \) and \( x_i = (x_{i1}, x_{i2}, \ldots, x_{in})' \). Random error \( \epsilon_i \) is normally distributed with \( E(\epsilon_i) = 0 \) and \( Var(\epsilon_i) = \sigma^2 \). Regression curve \( \mu(z_i, t_i, x_i) \) is assumed to be additive, so it can be written as

\[ \mu(z_i, t_i, x_i) = h(z_{i0}) + h(z_{i1}) + \ldots + h(z_{ip}) + g(t_{i0}) + g(t_{i1}) + \ldots + g(t_{ip}) + f(x_{i0}) + f(x_{i1}) + \ldots + f(x_{in}) \] (10)

A regression curve \( \mu(z_i, t_i, x_i) \) is called a nonparametric mixed regression curve which is grouped into three components of the regression curve, namely the components of the spline truncated regression curve, the kernel, and the Fourier series. Equation (10) can be written as:

\[ \mu(z_i, t_i, x_i) = \sum_{i=1}^{p} h(z_{io}) + \sum_{a=1}^{p} g(t_{ia}) + \sum_{i=1}^{n} f(x_{ii}) \] (11)
Component of regression curve $\sum_{x=1}^{p} h(z_{mx})$ is component of spline truncated regression curve, component of regression curve $\sum_{u=1}^{q} g(t_{mu})$ is component of kernel regression curve, while component of regression curve $\sum_{v=1}^{r} f(x_{mv})$ is component of Fourier series regression curve. The estimator for the component of the spline truncated regression curve $\hat{h}(z_{mx})$ in equation (11) will be:

$$
\hat{h}(z_{mx}) = \sum_{x=1}^{p} G_{s}(\xi_{x})\theta \\
= G_{1}(\xi_{1})\theta + G_{2}(\xi_{2})\theta + ... + G_{p}(\xi_{p})\theta \\
= \{G_{1}(\xi_{1}) + G_{2}(\xi_{2}) + ... + G_{p}(\xi_{p})\}\theta \\
= G(\Omega)\theta
$$

The estimator for the component of the kernel $\hat{g}(t_{mu})$ regression curve in equation (11) will be as follows:

$$
\hat{g}(t_{mu}) = \sum_{u=1}^{q} V_{u}(\alpha_{u})\gamma \\
= V_{1}(\alpha_{1})\gamma + V_{2}(\alpha_{2})\gamma + ... + V_{q}(\alpha_{q})\gamma \\
= \{V_{1}(\alpha_{1}) + V_{2}(\alpha_{2}) + ... + V_{q}(\alpha_{q})\}\gamma \\
= V(A)\gamma
$$

The estimator for the component of the Fourier $\hat{f}(x_{mv})$ curve series in equation (11) will be:
\[
\sum_{i=1}^{r} f_i(x_i) = \sum_{i=1}^{r} \left( b_i x_i + \frac{1}{2} \alpha_{0v} + \sum_{k=1}^{K} \alpha_{k} \cos kx_i \right)
\]

From equations (12), (13), (14) then equation (11) can be presented in the form of vectors and matrices as follows:

\[
y = \sum_{s=1}^{p} h(y_{s}) + \sum_{u=1}^{q} g(t_{u}) + \sum_{v=1}^{r} f(x_{v}) + \varepsilon
\]

\[
y = G(\Omega)\theta + V(A)y + Xa + \varepsilon
\]

\[
y = \chi + \varepsilon = h + g + f + \varepsilon
\]

From equation (15) we get the following equation: \( \varepsilon = y - G(\Omega)\theta - V(A)y - Xa \)

becomes

\[
\varepsilon = Y^* - G(\Omega)\theta - Xa; \quad \text{with } Y^* = y - V(A)y = (I - V(A))y
\]

A settlement will be carried out to get the estimated parameters in equation (16) using Ordinary Least Square (OLS) optimization. \( \min_{\theta, a} \left\{ (Y^* - G(\Omega)\theta - Xa)^T (Y^* - G(\Omega)\theta - Xa) \right\} \) First, the equation is simplified to get the estimated parameter \( c \), as follows:
\( Q = (Y^* - G(\Omega)\theta - Xa)^T (Y^* - G(\Omega)\theta - Xa) \)
\[ = Y^*T Y^* - Y^*T G(\Omega)\theta - Y^*T Xa - \theta^T G(\Omega)^T Y^* + \theta^T G(\Omega)^T G(\Omega)\theta \]
\[ + \theta^T G(\Omega)^T Xa - a^T X^T Y^* + a^T X^T G(\Omega)\theta + a^T X^T Xa \]
\[ = Y^*T Y^* - 2\theta^T G(\Omega)^T Y^* - 2a^T X^T Y^* \]
\[ + 2\theta^T G(\Omega)^T Xa + \theta^T G(\Omega)^T G(\Omega)\theta + a^T X^T Xa \]

Then proceed to the first decrease in equation (17) and be accompanied by zero.

\[ \frac{\partial Q}{\partial \theta} = 0 \]
\[ -2G(\Omega)^T Y^* + 2G(\Omega)^T Xa + 2G(\Omega)^T G(\Omega)\theta = 0 \]
\[ \text{(18)} \]

Then simplify the equation again to get the estimated parameter \( a \).

\[ Q = (Y^* - G(\Omega)\theta - Xa)^T (Y^* - G(\Omega)\theta - Xa) \]
\[ = Y^*T Y^* - Y^*T G(\Omega)\theta - Y^*T Xa - \theta^T G(\Omega)^T Y^* + \theta^T G(\Omega)^T G(\Omega)\theta \]
\[ + \theta^T G(\Omega)^T Xa - a^T X^T Y^* + a^T X^T G(\Omega)\theta + a^T X^T Xa \]
\[ = Y^*T Y^* - 2\theta^T G(\Omega)^T Y^* - 2a^T X^T Y^* \]
\[ + 2a^T X^T G(\Omega)\theta + \theta^T G(\Omega)^T G(\Omega)\theta + a^T X^T Xa \]

Then the first decrease in equation (19) is continued and zero is added.

\[ \frac{\partial Q}{\partial a} = 0 \]
\[ -2X^T Y^* + 2X^T G(\Omega)\theta + 2X^T Xa = 0 \]
\[ \text{(20)} \]

From the results obtained in equations (18) and (20) it appears that it still contains parameters, so to solve the equation the substitution method is used. To facilitate the calculation, the equation can be written as follows:

\[ \hat{\theta} = MY^* - MXa \]
\[ \text{(21)} \]

with

\[ M = \left( G(\Omega)^T G(\Omega) \right)^{-1} G(\Omega)^T \]

\[ \hat{a} = NY^* - NG(\Omega)\theta \]
\[ \text{(22)} \]

with

\[ N = (X^T X)^{-1} X^T \]

The first step is to substitute equation (21) for equation (22) so that it is obtained:
\[
\hat{a} = NY * - NG(\Omega)(MY * - MX\hat{a}) \\
= NY * - NG(\Omega)MY * + NG(\Omega)MX\hat{a} \\
\hat{a} - NG(\Omega)MX\hat{a} = NY * - NG(\Omega)MY * \\
(I - NG(\Omega)MX)\hat{a} = (N - NG(\Omega)M)Y * \\
\hat{a} = (I - NG(\Omega)MX)^{-1}(N - NG(\Omega)M)y \\
\hat{a} = (I - NG(\Omega)MX)^{-1}(N - NG(\Omega)M)(I - V(A))y \\
\hat{a} = Cy 
\]
where \( C = (I - NG(\Omega)MX)^{-1}(N - NG(\Omega)M)(I - V(A)) \)

Then substitute equation (23) for equation (21) so that it is obtained as follows:

\[
\hat{\theta} = MY * - MX((I - NG(\Omega)MX)^{-1}(N - NG(\Omega)M)Y*) \\
= (M - MX((I - NG(\Omega)MX)^{-1}(N - NG(\Omega)M)y*) \\
= (M - MX((I - NG(\Omega)MX)^{-1}(N - NG(\Omega)M))(I - V(A))y \\
= Dy 
\]
where \( D = (M - MX((I - NG(\Omega)MX)^{-1}(N - NG(\Omega)M))(I - V(A)) \)

Based on equations (23) and (24), to obtain the nonparametric function estimation in equation (15) is:

\[
\hat{\chi} = \hat{h} + \hat{g} + \hat{f} \\
= G(\Omega)\hat{\theta} + V(A)y + X\hat{a} \\
= G(\Omega)Dy + V(A)y + XCy \\
= Ey 
\]
where \( E = G(\Omega)D + V(A) + XC \)

Spline, kernel, and Fourier series nonparametric mixed regression curve estimators are very dependent on knots, bandwidth, and oscillation parameters. To obtain optimum knots, bandwidth, and oscillation parameters, this study uses the Generalized Cross Validation (GCV) method as follows:

\[
GCV(\Omega, A, M) = \frac{MSE(\Omega, A, M)}{(n^{-1}\text{trace}(I - E(\Omega, A, M)))^2} 
\]
where \( MSE(\Omega, A, M) = n^{-1}y'(I - E(\Omega, A, M))(I - E(\Omega, A, M))y \)

Examples of applications that use this mixed estimator are as follows:

In Indonesia, the Human Development Index is an index used to measure the success of human development in a region. There are several factors that influence HDI (variable response), for example morbidity rate, high school net enrollment rate (APM SMA), GRDP per capita, these factors are variable predictors previously conducted by Dewanti’s research [22]. In this case, it will be investigated whether the predictor variable affects the response variable. The data taken is East Java BPS data. The scatterplot for each variable is as follows:
Judging from the three figures, it is assumed that HDI vs Morbidity follows the Fourier series pattern, HDI vs APM SMA follows the spline truncated pattern, and IPM vs GRDP per capita follows the kernel pattern. From here there are differences in patterns between one variable and another, so you can use a mixture estimator of the truncated spline, kernel, and Fourier series to solve the problem. Meanwhile, in Dewanti’s research [22], HDI vs Morbidity followed the spline truncated pattern.

4. Conclusion

Based on the analysis described above, then it can be concluded that:

Estimates for the Model of Spline Truncated, Kernel, and Fourier Series Nonparametric Regression are as follows:

\[ \hat{Y} = G(\Omega)Dy + V(A)y + XCy \]

where

\[ D = (M - MX((I - NG(\Omega)MX)^{-1}(N - NG(\Omega)M))(I - V(A))) \]
\[ C = (I - NG(\Omega)MX)^{-1}(N - NG(\Omega)M)(I - V(A)) \]

and the smallest GCV formula to obtain the optimum location and number of points of knots, bandwidth, and oscillations, as follows

\[ GCV(\Omega, A, M) = \frac{MSE(\Omega, A, M)}{(\text{trace}(I - E(\Omega, A, M)))^2} \]

where

\[ MSE(\Omega, A, M) = n^{-1}(I - E(\Omega, A, M))(I - E(\Omega, A, M))y \]

Acknowledgement

This paper is part of my research at Masters Statistics program in Institut Teknologi Sepuluh Nopember which has been supervised by the second and third authors. I would like to give a special thanks to the Institut Teknologi Sepuluh Nopember, second and third authors, and also want to thank my family and Tribuana Kalabahi University for providing all the facilities for me.

References

[1] Eubank, R.L. 1999, Nonparametric Regression and Spline Smoothing Second Edition, Marcel Dekker, New York.

[2] Silverman, B. W. 1985. Some Aspects of The Spline Smoothing Approach to Non-parametric Regression Curve Fitting. Journal of the Royal Statistical Society. Series B (Methodological), 47, 1-52.

[3] Prahutama, A. 2013. Model Regresi Nonparametrik dengan Pendekatan deret Fourier pada Kasus Tingkat Pengangguran Terbuka di Jawa Timur. Prosiding Seminar Nasional Statistika Universitas Diponegoro 2013.
[4] Tripena, A. and Budiantara, I. N. 2007. Fourier Estimator in Nonparametric Regression. *International Conference On Natural Sciences and Applied Natural Sciences*. Ahmad Dahlan University, Yogyakarta.

[5] Budiantara I.N., Mariati N.P.A.M., Ratnasari V., Zain I., Ratna M., Sudiarsa I.W., Mardianto M.F.F. and Hendayanti N.P.N. 2015. Comparison Truncated Spline and Fourier Series in Multivariable Nonparametric Regression Models (Application: Data of Poverty in Papua, Indonesia). *International Journal of Basic and Applied Sciences IJBAS-IJENS*, 15, 9-12.

[6] Wisisono, I. N. R., Nurwahidah, A. I., and Andriyana, Y. 2018. Regresi Nonparametrik dengan Pendekatan Deret Fourier pada Data Debit Air Sungai Citarum. *Jurnal Matematika*. Vol. 04, 2, pp. 75-82.

[7] Saputro D. R. S., Sukmayanti A., Widyaningsih P. 2019. The nonparametric regression model using Fourier series approximation and penalized least squares (PLS) (case on data poverty in East Java). *Journal of Physics: Conf. Series*, 1188 012019, doi:10.1088/1742-6596/1188/1/012019.

[8] Saputro, D. R. S., Demu, K. R., and Widyaningsih, P. 2018. Nonparametric Truncated Spline Regression on the Data of Human Development Index (HDI) in Indonesia. *2nd International Conference on Statistics, Mathematics, Teaching, and Research*. IOP Publishing.

[9] Antomiadis, A., Bigot, and Spatinas, T. 2001. Wavelet Estimator in Nonparametric Regression: A Comparative Simulation Study. *Journal of Statistical Software*. Vol. 06, pp. 1-83.

[10] Budiantara, I.N. 2009. *Spline Dalam Regresi Nonparametrik Dan Semiparametrik: Sebuah Pemodelan Statistika Masa Kini dan Masa Mendatang. Pidato Pengukuhan Untuk Jabatan Guru Besar Dalam Bidang Ilmu Matematika Statistika dan Probabilitas, Pada Jurusan Statistika, Fakultas MIPA, Institut Teknologi Sepuluh Nopember*. ITS Press. Surabaya.

[11] Hardle, W. 1994. *Applied Nonparametric Regression*. Springer-Verlag. Berlin.

[12] Budiantara, I. N., & Mulianah. 2007. Pemilihan Bandwidth optimum dalam Regresi Semiparametrik Kernel dan Aplikasinya. *SIGMA: Jurnal Sains dan Teknologi*, Vol. 10, 2, 159-166.

[13] Klemela, J. 2014. *Multivariate Nonparametric Regression and Visualization: with R and Applications to Finance*. New Jersey: John Wiley and Sons, I.

[14] Asrini, L.I., and Budiantara, I.N. 2014. Fourier Series Semiparametric Regression Models (Case Study: The Production of Law Land Rice Irrigation in Central Java). *ARPN Journal of Engineering and Applied Sciences*, 9, 1501-1506.

[15] Bilodeau, M. 1992. Fourier Smoother and Additive Models. *The Canadian Journal of Statistics*, 3, 257-259.

[16] Ratnasari, V., Budiantara, I.N., Ratna, M., and Zain, I. 2015. Estimation of Nonparametric Regression Curve using Mixed Estimator of Multivariable Truncated Spline and Multivariable Kernel. *Global Journal of Pure and Applied Mathematics*. Vol. 10, 6, 5047-5057.

[17] Sudiarsa I.W., Budiantara I.N., Suhartono and Purnami S.W. 2015. Combined Estimator Fourier Series and Spline Truncated in Multivariable Nonparametric Regression. *Applied Mathematical Sciences*, 9, 4997-5010.

[18] Affifah, N. 2017. Mixed Estimator of Kernel and Fourier Series in Semiparametric Regression. *Journal of Physics Conference Series*. IOP Publishing.

[19] Budiantara I. N., Ratnasari V., Ratna M., and Zain I. 2015. The Combination of Spline and Kernel estimator for Nonparametrik Regression and Its Properties. *Applied Mathematical Science 9*, 122, 6083-6094.

[20] Suparti, Prahutama A., and Santoso R. 2018. Mix Local Polynomial and Spline Truncated: The Development of Nonparametric Regression Model. *IOP Conf Series: Journal of Physics: Conf Series* 1025 012102.

[21] Budiantara I. N., Ratnasari V., Ratna M., Wibowo W., Affifah N., Rahmawati D. P. and Octavanny M. A. D. 2019. Modelling Percentage of Poor People in Indonesia Using Kernel and Fourier
Series Mixed Estimator in Nonparametric Regression. *Revista Investigation Operacional*. Vol 40, 4, 538-550.

[22] Dewanti et al. 2020. Modelling of SDG's Achievement in East Java Using Biresponse Nonparametric Regression with Mixed Estimator Spline Truncated and Kernel. *IOP Conf Series: Journal of Physics: Conf Series* 1562 012016.