

**MODELING EAST JAVA INDONESIA LIFE EXPECTANCY USING
SEMIPARAMETRIC REGRESSION MIXED SPLINE TRUNCATED AND
FOURIER SERIES**

Khaerun Nisa¹, I Nyoman Budiantara²

¹ Balai Penelitian dan Pengembangan Agama Makassar (BLAM)

² Institut Teknologi Sepuluh Nopember (ITS)

e-mail: kn.khaerunnisa@gmail.com

DOI: 10.14710/medstat.13.2.149-160

Article Info:

Received: 27 June 2019

Accepted: 23 November 2020

Available Online: 28 December
2020

Keywords:

*Life Expectancy,
Semiparametric Regression,
Mixed Estimator, Spline
Truncated, Fourier Series*

Abstract: Life expectancy is one of the indicators used to evaluate the government's performance in improving the well-being of the population. High life expectancy in an area indicates that people in the area have been assured of health and poverty has been well overcome, and vice versa. Based on national socioeconomic survey (SUSENAS) data, showing life expectancy in East Java Province from 2009 to 2013 increased by 69.15 years to 70,19 years. Although overall life expectancy in East Java province has increased, there are still some areas that have life expectancy below 65 years. This is not from the different characteristics of each religion. Therefore, the main purpose of this study is to model life expectancy in East Java using semiparametric regression with a mixed estimator of Spline Truncated and Fourier Series. Based on the research that has been done, the results that modeling the data of life expectancy using mixed estimator of Spline Truncated and Fourier Series produced a value of R^2 of 99,62% which means that the predictor variables are able to explain the response variabel life expectancy of 99.62%.

1. INTRODUCTION

According to the Indonesian Central Bureau of Statistics, life expectancy is the average estimate from the age of the newborn to the death. The life expectancy of East Java population based on the National Economic Survey (SUSENAS) continued to increase since 2009 by 69.15 years until 2013 which reached 70.19 years. Although overall East Java life expectancy has increased, there are 7 districts in East Java that have life expectancy below 65 years. The difference in life expectancy in districts/cities in East Java is certainly inseparable from the economic, social and cultural factors of each region that have own characteristics. Some of the factors that are suspected to affect life expectancy in East Java are infant Mortality Rate, percentage of infants aged 0-11 months who are breast milked 4-6 months, Labor Force Participation Rate, Literacy Rate, and average school length.

Research on life expectancy has been conducted by several researchers including Cleries et al (2009) who conducted research to find out the trends in the mortality rate of the trends in the mortality rate of the Spanish population from 1977 to 2001 and its impact on life expectancy using the Bayesian Age Period Cohort (APC). Halicioglu (2011) conducted research to find out the factors that affect life expectancy in Turkey with Autoregressive Distributed Lag (ADRL) approach. As well as Sugiantari (2013) analyzed the factors that affect life expectancy in East Java using Spline Semiparametric Regression.

Another method that can be used to model life expectancy is semiparametric regression with a mixed estimator of Spline Truncated and Fourier Series Nisa et al (2016). Semiparametric regression is a combination of parametric components and nonparametric components. This method plays an important role in solving the problem of regression modeling which has a relationship between response variables and their predictor variables some follow certain patterns, others have fickle patterns at certain sub-sub intervals, and others have repetitive patterns.

Research using mixed estimators was conducted by Budiantara et al (2015) using Spline and Kernel mixture estimators in nonparametric regression. Sudiarsa et al (2015) involved the estimators of Fourier and Spline Truncated in nonparametric regression. Research using semiparametric regression was conducted by Pane et al (2014) involving Fourier Series estimator, Wibowo et al, (2013) multiresponsive modeling on semiparametric regression. Then the Fourier Series estimator with a case study of rice production in Central Java (Asrini & Budiantara, 2014). Multivariable semiparametric regression model with combined estimator of Fourier Series and Kernel (Nisa et al, 2017).

The purpose of this study is to describe life expectancy, infant mortality rate, percentage of babies breastfed, average school length, labor force participation rate and literacy rate in East Java Province and model life expectancy in East Java Province using mixed semiparametric regression Spline Truncated dan Fourier Series.

2. LITERATURE REVIEW

2.1. Parametric Regression, Nonparametric Regression, Semiparametric Regression

Parametric regression is one of the widely used statistical methods that largely describes the pattern of relationship between response variables and predictor variables. Ellis et al (1968) stated that the linear form of parametric regression is

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + \varepsilon_i, i = 1, 2, \dots, n. \quad (1)$$

with y_i as a response variable, $x_{1i}, x_{2i}, \dots, x_{pi}$ as the predictor variable, is an unknown parameter and ε_i is an independent random error, distributed normally with zero mean and σ^2 variance. Estimator parameter model obtained based on various methods that have been known in statistics namely least square, maximum likelihood Wahba (1990) in (Rismal et al, 2016).

Nonparametric regression is a regression whose pattern of relationship between the response variable and the predictor variable is unknown in shape. Suppose given n independent observations, i.e. pairs $(y_i, z_i), i = 1, 2, \dots, n$. The pattern of the relationship between variables y_i and z_i is unknown and follows the regression model:

$$y_i = g(z_i) + \varepsilon_i \quad (2)$$

with y_i as a response variable, z_i is a predictor variable, ε_i is an independent random error, normal distribution with zero mean and σ^2 variance, while $g(z_i)$ is an unknown regression function of the pattern form.

Semiparametric regression is a combination of parametric components and nonparametric components. Supposing given data paired (x_i, y_i, z_i) or the relationship between x_i, z_i and y_i assumed to follow a semiparametric regression model:

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

with y_i is a response variable, x_i and z_i are predictor variables, and ε_i is an independent and normally distributed random error with zero mean and σ^2 variance. $f(x_i)$ is a regression function known to form a pattern while $g(z_i)$ is a regression function known to form the pattern.

2.2. Semiparametric Regression Model with Mixed Estimator of Spline Truncated and Fourier Series

Given paired data $(x_{1i}, \dots, x_{pi}, t_{1i}, \dots, t_{qi}, z_{1i}, \dots, z_{ri}, y_i)$, $i = 1, 2, \dots, n$ assumed to follow a semiparametric regression model, as presented in the equation (4)

$$y_i = \mu(x_{1i}, x_{pi}, \dots, t_{1i}, \dots, t_{qi}, z_{1i}, \dots, z_{ri}, y_i) + \varepsilon_i \quad (4)$$

Kurva regresi μ regression curve in equation (4) is assumed to be additive, so it can be written as in equation (5)

$$y_i = \sum_{j=1}^p f_j(x_{ji}) + \sum_{s=1}^q g_s(t_{si}) + \sum_{l=1}^r h_l(z_{li}) + \varepsilon_i, i = 1, 2, \dots, n \quad (5)$$

$f_j(x_{ji})$ regression curve can be approached with linear function:

$$f(x) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} \quad (6)$$

$g_s(t_{si})$ regression curve can be approached with linear Spline Truncated with knot K_1, K_2, \dots, K_m . In general, linear Spline Truncated can be presented as follows:

$$G_i(t_i) = \sum_{j=1}^q \phi_{ji} t_i + \sum_{k=1}^m \alpha_{ki} (t_i - K_k)_+ \quad (7)$$

$$\text{with, } (t_i - K_k)_+ = \begin{cases} (t_i - K_k)_+ & , t_i \geq K_k \\ 0 & , t_i < K_k \end{cases}$$

Where $\phi_1, \alpha_1, \dots, \alpha_m$ are unknown parameters, K_1, K_2, \dots, K_m are knot point. Meanwhile, the $h_i(z_{li})$ regression curve can be approached with the Fourier series function.

$$h(z) = \sum_{i=1}^n b z_i + \frac{1}{2} a_0 + \sum_{k=1}^K a_k \cos k z_i \quad (8)$$

Where b, a_0, a_k are model parameters (8), $k = 1, 2, \dots, K$ are the number of cosine waves in a Fourier series function called oscillation.

The estimated result of the Spline Truncated and Fourier Series Semiparametric regression models given in the equation (4) can be obtained using the Penalized Least Square

method, i.e. by combining the goodness of fit and penalty containing the smoothing parameters (λ), presented in the equation (9).

$$\text{Min}_{f,g,h} \left\{ n^{-1} \sum_{i=1}^n \left(y_i - \sum_{j=1}^p f_j(x_{ji}) - \sum_{s=1}^q g_s(t_{si}) - \sum_{l=1}^r h_l(z_{li}) \right)^2 + \sum_{l=1}^r \lambda_l \int_0^{\pi} \frac{2}{\pi} (h_l''(z_l))^2 dz_l \right\} \quad (9)$$

Equation (9) can be written in the form of a matrix, indicated by the equation (10)

$$\text{Min}_{f,g,h} \left\{ n^{-1} (\tilde{Y} - \mathbf{X}\tilde{\beta} - \mathbf{G}\tilde{\phi} - \mathbf{D}\tilde{a})^T (\tilde{Y} - \mathbf{X}\tilde{\beta} - \mathbf{G}\tilde{\phi} - \mathbf{D}\tilde{a}) + \tilde{a}^T \mathbf{U}\tilde{a} \right\} \quad (10)$$

The description of the estimated results of the mixed semiparametric regression model can be seen in (Nisa et al., 2016). Obtained estimator linear parametric regression curve:

$$\hat{f}(x) = \mathbf{X}\hat{\beta} = \mathbf{A}(K, \lambda, k)\tilde{Y}, \quad (11)$$

with $\mathbf{A}(K, \lambda, k)$ is a matrix of coefficients in the estimator of parametric components. The full form can be seen in (Nisa et al., 2016).

Spline Truncated regression curve estimator:

$$\hat{g}(t) = \mathbf{G}\hat{\phi} = \mathbf{B}(K, \lambda, k)\tilde{Y}, \quad (12)$$

with $\mathbf{B}(K, \lambda, k)$ is a matrix of coefficients in the estimator of Spline Truncated estimator. The full form can be seen in (Nisa et al., 2016).

Thus, obtained estimation of semiparametric regression models of Spline Truncated and Fourier Series, as follows:

$$\begin{aligned} \hat{y} &= \hat{f}(x) + \hat{g}(t) + \hat{h}(z) \\ &= [\mathbf{A}(k, \lambda, K) + \mathbf{B}(k, \lambda, K) + \mathbf{C}(k, \lambda, K)]\tilde{Y} \\ &= \mathbf{F}(K, \lambda, k)\tilde{Y}, \end{aligned}$$

with $\mathbf{F}(K, \lambda, k) = \mathbf{A}(K, \lambda, k) + \mathbf{B}(K, \lambda, k) + \mathbf{C}(K, \lambda, k)$.

3. RESEARCH METHODOLOGY

3.1. Data Sources and Research Variables

This study uses secondary data obtained from the publication of Central Bureau of Statistics in 2013, namely the East Java Health Executive Report 2013, the East Java Education Executive Report 2013 and the National Economic Survey of East Java (SUSENAS) 2013. The observation unit used in this study are 38 districts/cities in East Java Province.

Where response variable (y) is life expectancy, and its predictor variables; (x_1) infant mortality rate, (x_2) percentage of infants aged 0-11 months who are breast milked 4-6 months, (x_3) labour force participation rate, (x_4) literacy numbers, (x_5) average school length.

3.2. Research Method

Modeling East Java life expectancy in 2013 using semiparametric regression estimation of Spline Truncated and Fourier Series, with the following steps:

- a. Explore data to find an overview of life expectancy data in East Java.
- b. Create a partial scatter plot between response variables with each predictor variable.
- c. Specifies the predictor variable approached by the Spline Truncated function and the Fourier Series function.
- d. Modeling life expectancy data in East Java using a mixed semiparametric approach of Spline Truncated and Fourier Series.
- e. Select the K knot point, smoother parameter and optimum k oscillation parameter using the GCV method and then set the best mix semiparametric regression.
- f. Calculate R^2 as part of the model's goodness criteria.
- g. Interpret the model with a semiparametric regression approach of Spline Truncated and Fourier series, and draw conclusions.

4. RESULTS AND DISCUSSIONS

4.1. Modeling Life Expectancy in East Java, Indonesia using Estimator Spline Truncated and Fourier Series in Semiparametric Regression

The initial stage before modeling the case of life expectancy in East Java Province is to know the form of relationship patterns between response variables and each predictor variable. Information on the form of relationship patterns between response variables and predictor variables can be obtained from the scatter plot and used to determine the appropriate type of regression curve in approaching the data pattern. Estimation of this model can be obtained using R software Faraway (2007), Rosenblad (2009) and (Fox & Weisberg, 2011). Modeling of life expectancy data in East Java has been tried using multiple linear methods, but in testing IIDN assumptions (Ident, Independent and normally distributed) indent assumptions are not met, so statistical inference cannot be done in the modeling. The life expectancy data has also been tried using Spline Truncated and produced R^2 of 99.77% with a mean square error of 0.0225. In addition, the Fourier series has been attempted and generated R^2 of 99.63% with a mean square error of 0.369. Although the Spline Truncated dan Fourier Series have a large R^2 and small mean square errors, they are not parsimoni. Therefore, modeling life expectancy data in East Java is modeled using semiparametric regression with a mixed estimator of Spline Truncated and Fourier Series, resulting in R^2 of 99.62% and mean square error of 0.0317. Modeling using this mixed estimator is more parsimoni compared to the three previous methods that have been tried.

Scatter plots of East Java Province life expectancy in 2013 for response variables to each predictor variable, presented in Figure 1, Figure 2, Figure 3, Figure 4, and Figure 5. Figure 1 shows that the variable relationship pattern form of life expectancy with infant mortality rate is liniarly negative. Figure 2 and Figure 3 show that the pattern form of the life expectancy relationship with the percentage of breastfeeding is unknown to the form of the relationship pattern and tends to experience repeated changes in data pattern behavior, so it can be modeled with the Fourier Series function. Figure 4 and Figure 5 show that the pattern shape of the relationship between the Life Expectancy Number variable and the Literate Number variable is unknown to the form of the relationship pattern and tends to experience behavioral changes at certain intervals, so it can be modeled with the Spline Truncated function. Figure 4 and Figure 5 show that the pattern shape of the relationship

between the Life Expectancy Number variable (y) and the Literate Number variable (t_1) is unknown to the form of the relationship pattern and tends to experience behavioral changes at certain intervals, so it can be modeled with the Spline Truncated function.

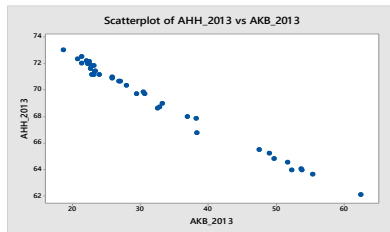


Figure 1. Scatter Plot between Life Expectancy and Infant Mortality

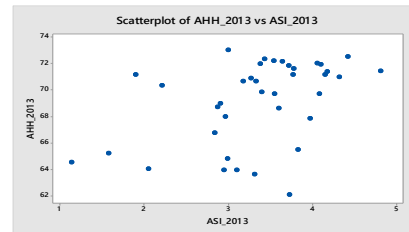


Figure 2. Scatter Plot between Life Expectancy and Percentage of Breastfeeding

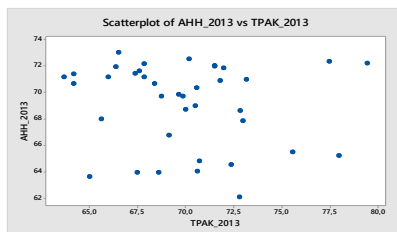


Figure 3. Scatter Plot between Life Expectancy and Labor Force Participation Rate

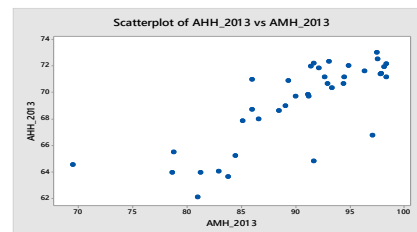


Figure 4. Scatter Plot Life Expectancy and Literacy Rate

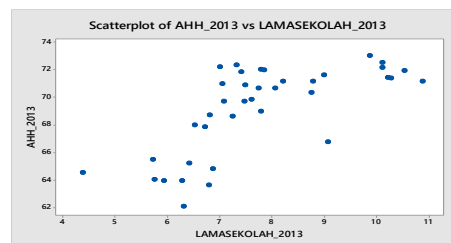


Figure 5. Scatter Plot between Life Expectancy and Average Length of School

In defining which variables are approached with Spline Truncated and Fourier Series in addition to using scatter plots can also be done by checking the GCV values of each predictor variable using equations (13).

$$GCV(K, \lambda, k) = \frac{MSE(K, \lambda, k)}{\left(n^{-1}tr(I - S(K, \lambda, k))\right)^2}, \quad (13)$$

with

$$MSE(K, \lambda, k) = n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2.$$

Based on the calculation of GCV obtained a minimum GCV value of 0.0516573. Thus the data of Life Expectancy In East Java in 2013 can be approached with a mixture of semiparametric regression spline truncated and fourier series, where there is one variable

approached with linear parametrics namely Infant Mortality Rate, two predictor variables approached with Spline Truncated which is the variable Literacy Rate and the Variable Participation Rate of the Workforce , as well as two predictor variables approached with the Fourier Series are the percentage variable of breast-fed infants and the average variable of the length of school. A full list of each predictor variable can be found in Table 1.

The semiparametric regression model of a mixture of Spline Truncated and Fourier Series with one parametric component, two Truncated Spline components and two Fourier Series components depends on knot points, k oscillations and smoothing parameters. Comparison of minimum GCV values between models, shown in Table 2. Table 2 shows that the regression model that has a minimum GCV is a model with three knot points and $k = 2$ of 0.0516263. However, because difference in GCV values at three knot points with oscillations $k = 1$, $k = 2$ and $k = 3$ is not much different and with the consideration of parsimoni model, the best model is selected with three point knots and oscillation $k = 1$. The estimated parameters of the semiparametric model of a mixture Spline Truncated and Fourier Series with three knot points and oscillations $k = 1$ are shown in Table 3. Based on the estimated results in Table 3, a plot can be made between y and \hat{y} for the 38 observations shown by Figure 6. Figure 6 shows that the graph \hat{y} is close to actual data graph. Based. Based on calculations, the semiparametric regression model of a mixture of Spline Truncated and Fourier Series with three knot points and $k = 1$ oscillations have an R^2 value of 99.62%, which means variable Infant Mortality Rate, percentage of infants aged 0-11 months who are breast-fed 4-6 months, Labor Force Participation Rate, Literacy Rate, and average length of school are able to explain the response variable Life Expectancy Figure in 2013 of 99.62%.

Table 1. Parametric and Nonparametric Components

| No | Predictor Variables | Components | Model |
|----|---------------------------------|---------------|---------------------------------|
| 1 | Infant Mortality Rate | Parametric | Linear Parametric |
| 2 | Literacy Rate | Nonparametric | Spline <i>Truncated</i> Funtion |
| 3 | Average Length of School | | Spline <i>Truncated</i> Funtion |
| 4 | Percentage of Babies breast-fed | | Fourier Series Function |
| 5 | Labour Force Participation Rate | | Fourier Series Function |

Table 2. Minimum GCV Value Comparison

| No | Model | GCV |
|----|--|------------|
| 1 | 1 point knots 97.2873 and 10.64. $k = 1$ oscillation | 0.0789276 |
| 2 | 1 point knots 97.2873 and 10.64. $k = 2$ oscillation | 0.0789381 |
| 3 | 1 point knots 97.2873 and 10.64. $k = 3$ oscillation | 0.0789485 |
| 4 | 2 point knots (90.1343;94.2671) and (9.0326;9.96143). $k = 1$ oscillation | 0.0603866 |
| 5 | 2 point knots (91.51191;94.2671) and (9.3423;9.96143). $k = 2$ oscillation | 0.0576090 |
| 6 | 2 point knots (90.1343;94.2671) and (9.03286;9.96143). $k = 3$ oscillation | 0.0577269 |
| 7 | 3 point knots (85.7431;87.5513). (94.9756;8.04625) and (8.4525;10.0775). $k = 1$ oscillation | 0.0546567 |
| 8 | 3 point knots (85.7431;87.5513). (94.7837; 8.04625) and (8.4525;10.0775). $k = 2$ oscillation | 0.0516263* |
| 9 | 3 point knots (85.7431;87.5513). (94.7837; 8.04625) and (8.4525; 10.0075). $k = 3$ oscillation | 0.0516573 |

Table 3. Estimated Parameters of Semiparamtric Model Mix of Spline Truncated and Fourier Series with 3 Knot Points and $k = 1$ Oscillations

| Variables | Parameters | Estimation |
|-----------|---------------|----------------------|
| x | β_0 | -0.00000000000568583 |
| | β_1 | -0.2500250 |
| t_1 | ϕ_{11} | -0.0315882 |
| | α_{11} | -0.1820790 |
| | α_{12} | 0.2772730 |
| | α_{13} | -0.5084910 |
| t_2 | ϕ_{21} | 0.1760080 |
| | α_{21} | -2.0568500 |
| | α_{22} | 3.3071400 |
| | α_{23} | -2.0945400 |
| z_1 | b_1 | 0.1621760 |
| | a_{01} | 74.4552000 |
| | a_{11} | 2.2383300 |
| z_2 | b_1 | 0.0278582 |
| | a_{02} | 74.4552000 |
| | a_{21} | -0.0020111 |

Based on Figure 6 it appears that the graph is close to the actual data graph. From the calculation, the semiparametric regression model of Spline Truncated and Fourier Series with 3 knot points and $k=1$ oscillation has an R^2 of 99.62%, which means the Variable Infant Mortality Rate, the percentage of infants aged 0-11 months who are breast-fed 4-6 months, Labor Force Participation Rate, Literacy Rate, and average length of school are able to explain the response variable Life Expectancy of 99.6%.

Here is written a semiparametric regression model of a mixture of Spline Truncated and Fourier Series with 3 knot points and $k = 1$ oscillations:

$$\hat{y}_i = 148.9104 - 0,250025x_i - 0,0315882t_{1i} - 0,182079(t_{1i} - 85,7431)_+ + 0,277273(t_{1i} - 87,5513)_+ - 0,508491(t_{1i} - 94,7837)_+ + 0,176008t_{2i} - 2,05685(t_{2i} - 8,04625)_+ + 3,30714(t_{2i} - 8,4525)_+ - 2,09454(t_{2i} - 10,0775)_+ + 0,162176z_{1i} + 2,23833\cos z_{1i} + 0,0278582z_{2i} - 0,00201112\cos z_{2i}$$

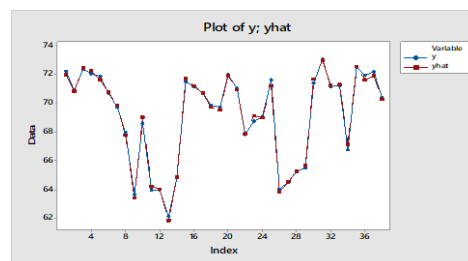


Figure 6. Plot data y and \hat{y}

The following is the model of each data group and its interpretation:

1. Model for Infant Mortality Rate variable

Assuming that other independent variables other than the Infant Mortality Rate are constant, the model is generally obtained :

$$\hat{y}_i = -0,250025x_i + c_1$$

with,

$$c_1 = 148.9104_i - 0,0315882t_{1i} - 0,182079(t_{1i} - 85,7431)_+ + 0,277273(t_{1i} - 87,5513)_+ - 0,508491(t_{1i} - 94,7837)_+ + \\ + 0,176008t_{2i} + -2,05685(t_{2i} - 8,04625)_+ + 3,30714(t_{2i} - 8,4525)_+ - 2,09454(t_{2i} - 10,0775)_+ + \quad \text{The} \\ + 0,162176z_{1i} + 2,23833\cos z_{1i} + 0,0278582z_{2i} - 0,00201112\cos z_{2i}$$

regression coefficient of Infant Mortality Rate is -0.250025, meaning that if other independent variables are constant and the Infant Mortality Rate variable increases by 1%, then the Life Expectancy Rate will decrease by 0.250025 year. There is negative relationship between the variable Infant Mortality Rate and Life Expectancy, the higher the Infant Mortality Rate, the lower the Life Expectancy.

2. Model for Literacy Rate variable

Assuming that other independent variables other than the Literacy Rate are constant, the model is generally obtained :

$$\hat{y}_i = -0,0315882t_{1i} - 0,182079(t_{1i} - 85,7431)_+ + 0,277273(t_{1i} - 87,5513)_+ - 0,508491(t_{1i} - 94,7837)_+ + c_2$$

where,

$$c_2 = 148.9104 - 0,250025x_i + 0,176008t_{2i} - 2,05685(t_{2i} - 8,04625)_+ + 3,30714(t_{2i} - 8,4525)_+ \\ - 2,09454(t_{2i} - 10,0775)_+ + 0,162176z_{1i} + 2,23833\cos z_{1i} + 0,0278582z_{2i} - 0,00201112\cos z_{2i}$$

The model can be interpreted by using the following functions:

$$\hat{y}_i = \begin{cases} -0,0315882t_{1i} + c_2, & t_{1i} \leq 85,7431 \\ 15,612017 - 0,2136672t_{1i} + c_2, & 85,7431 < t_{1i} \leq 87,5513 \\ -8,663583 + 0,0636058t_{1i} + c_2, & 87,5513 < t_{1i} \leq 94,7837 \\ 39,53317 - 0,4448852t_{1i} + c_2, & t_{1i} > 94,7837, \end{cases}$$

- East Java Province Literacy Rate which is less than 85.7431% include Jember, Bondowoso, Situbondo, Probolinggo, Bojonegoro, Tuban, Bangkalan, Sampang, Pamekasan and Sumenep Regency, when literacy rate is increased by 1%, then the Life Expectancy in these areas will decrease by 0.0315882 years.
- Literacy rate of East Java Province in the range of values of 85.7431% to 87.5513%, covering Lumajang and Ngawi Regency. If the Literacy Rate increases by 1%, then the Life Expectancy in these regions will decrease by 0.2136672 year.
- Literacy Rate of East Java Province in the range of 87.5513% to 94.7837% include Pacitan, Ponorogo, Trenggalek, Blitar, Kediri, Malang, Banyuwangi, Pasuruan, Mojokerto, Jombang, Nganjuk, Madiun, Magetan and Lamongan Regency, as well as Probolinggo and Batu City. If literacy rate increase by 1%, then life expectancy in these regions in these regions will increase by 0.0636 year.
- Literacy Rate of East Java Province greater than 94.7837% include Tulungagung, Sidoarjo and Gresik Regency, as well as Kediri, Blitar, Malang, Pasuruan, Mojokerto, and Surabaya City. If literacy increases by 1%, the Life Expectancy in these regions will decrease by 0.4448 year.

3. Model for Average Length of School Variable

Assuming that other independent variables other than the average length of school are constant, the general model is obtained:

$$\hat{y}_i = 0,176008t_{2i} - 2,05685(t_{2i} - 8,04625)_+ + 3,30714(t_{2i} - 8,4525)_+ - 2,09454(t_{2i} - 10,0775)_+ + c_3$$

where,

$$c_3 = 148.9104 - 0,250025x_i - 0,0315882t_{i_1} - 0,182079(t_{i_1} - 85,7431)_+ + 0,277273(t_{i_1} - 87,5513)_+ +$$

$$- 0,508491(t_{i_1} - 94,7837)_+ + 0,162176z_{i_1} + 2,23833\cos z_{i_1} + 0,0278582z_{2i} - 0,00201112\cos z_{2i}$$

The model can be interpreted by using the following functions:

$$\hat{y}_i = \begin{cases} 0,176008t_{2i} + c_3, & t_{i_1} \leq 8,04625 \\ 16,5499 - 1,880842t_{2i} + c_3, & 8,04625 < t_{i_1} \leq 8,4525 \\ -11,4037 + 1,42629t_{2i} + c_3, & 8,4525 < t_{i_1} \leq 10,0775 \\ 9,704 - 0,668242t_{2i} + c_3, & t_{i_1} > 10,0775 \end{cases}$$

- The average length of school in East Java which is less than 9.03286 years includes Pacitan, Ponorogo, Trenggalek, Tulungagung, Blitar, Kediri, Malang, Lumajang, Jember, Banyuwangi, Bondowoso, Situbondo, Probolinggo, Pasuruan, Nganjuk, Madiun, Magetan, Ngawi, Bojonegoro, Tuban, Lamongan, Bangkalan, Sampang, Pamekasan, and Sumenep Regency. If the average length of school increases by 1 year, then the Life Expectancy in these areas will increase by 0.176008 year.
- The average length of school in East Java in the range of 8.04625 years to 8.4525 years are Mojokerto and Jombang Regency. If the average length of school increases by 1 year, then the Life Expectancy in these areas will decrease by 1,8808 years.
- The average length of school in East Java in the range of 8.4525 years to 10,0775 years include Gresik Regency as well as Blitar, Probolinggo, Pasuruan and Batu City. If the average length of school increases by 1 year, then the Life Expectancy in these areas will increase by 1.42629 years.

The average length of school in East Java which is more than 10.0775 years includes Sidoarjo Regency, as well as Kediri, Blitar, Malang, Probolinggo, Pasuruan, Mojokerto, Madiun, Surabaya, and Batu City. If the average length of school increases by 1%, then life expectancy will decrease by 0.6682 years. Based on the results of these interpretations, the model interpretation for Literacy Figure data is less than 85.7431, the model for Literacy Figure data in the range of 85.7431 – 87.5513% and the model for Literacy Figure data greater than 94.7837% is not in accordance with the theory, where when the Literacy Rate increases, the Life Expectancy Rate decreases. As well as the interpretation of the model for the average school length data in the range of 8.04625 – 8.4525 years and the model for the average school length data greater than 10.0775 years is also not in accordance with the theory, where when the average length of school increases, the Life Expectancy rate decreases.

Literacy Rate and average length of school are components of education. Where education is one of the factors that contribute greatly to the development of social and economic life of a region. Literacy rate less than 85.7431% is the horseshoe area. The regions that fall into the category of second poulitice areas are those that have low Literacy Rates compared to other regions in East Java. However, when the Literacy Rate increases, the Life Expectancy decreases. The theory is not possible because the livelihoods in these areas are about 60% generally farmers and fishermen. Where the profession does not require high educational specifications.

The Literacy Rate of East Java Province, which is in the range of 85.7431% to 87.5513%, covers Lumajang and Ngawi Regency, as well as the Literacy Rate which is in the range greater than 94.7837% covering Tulungagung, Sidoarjo and Gresik Regency, Kediri, Blitar, Malang, Pasuruan, Mojokerto, and Surabaya City. The range of literacy rates is quite good. In addition, these areas are regions that have a fairly high Literacy Rate. However, when the Literacy Rate increased in the region, the life expectancy rate decreased.

This is not in accordance with the theory, possibly due to the unequal literacy rate in these regions.

The average length of schooling in East Java Province, which ranges from 8.4625 to 8.4525 years, covers Mojokerto and Jombang districts. The majority of livelihoods in Jombang Regency are private labor/employees and trade and services. Meanwhile, in Mojokerto Regency, the majority work in the industrial sector. The workforce in these two regions is dominated by graduates with a junior high school education level. The theory mismatch is possible because the workforce in the two regions is dominated by workers with junior high school education specifications, so that the contributions made by workers from these two regions have not been able to meet the needs of the industrial sector and cause social and economic disparities in the region. Based on the regencies/cities whose average length of schooling is greater than 10.0775 years, there are several regencies/cities in this range, namely Sidoarjo Regency as well as Kediri, Blitar, Malang, Probolinggo, Pasuruan, Mojokerto, Madiun, Surabaya, and Batu City. It is possible that there is a theoretical mismatch between the average length of schooling and the life expectancy in these areas, because these areas have a high but uneven average length of schooling.

4. CONCLUSION

Based on the analysis carried out, can be drawn conclusions as follows: Modeling life expectancy data in East Java using a mixed estimator of Spline Truncated and Fourier Series produces a value of R^2 of 99.62% which means that predictor variables are able to explain the life expectancy response variable of 99.62%. From the modeling results obtained that the higher the Infant Mortality Rate, the lower the Life Expectancy. This is in accordance with the life expectancy theory which states that the Life Expectancy is inversely proportional to the Infant Mortality Rate. As for the Variable Literacy Figures and the average variable of the length of school, there are interpretations that are not in accordance with the theory. Where there are several regencies/cities in East Java Province, when literacy rates and the average length of school has increased, the life expectancy has decreased. The discrepancy with the theory occurs because of the different characteristics of each regency / city in East Java Province. In addition, it is possible by the uneven literacy rate and the average length of school in East Java Province. This research is expected to be a consideration to the government of East Java Province in taking policy on population issues especially regarding improving the welfare of the community.

REFERENCES

- Asrini, L. J., & Budiantara, I. N. (2014). Fourier Series Semiparametric Regression Models (Case study: The Production of Lowland Rice Irrigation in Central Java). *ARPN Journal of Engineering and Applied Sciences*, 9(9), 1501–1506.
- Budiantara, I. N., Ratnasari, V., Ratna, M., & Zain, I. (2015). The Combination of Spline and Kernel Estimator for Nonparametric Regression and its Properties. *Applied Mathematical Sciences*, 9(122), 6083–6094. <https://doi.org/10.12988/ams.2015.58517>
- Cleries, R., Martínez, J. M., Valls, J., Pareja, L., Esteban, L., Gispert, R., ... Borràs, J. M. (2009). Life Expectancy and Age-Period-Cohort Effects: Analysis and Projections of Mortality in Spain Between 1977 and 2016. *Public Health*, 123(2), 156–162.

<https://doi.org/10.1016/j.puhe.2008.10.026>

- Ellis, D. M., Draper, N. P., & Smith, H. (1968). Applied Regression Analysis. *Applied Statistics*, 17(1), 83. <https://doi.org/10.2307/2985274>
- Faraway, J. J. (2007). Extending the Linear Model With R: Generalized Linear, Mixed Effects and Nonparametric Regression Models. *Journal of the American Statistical Association*, 102, 1477–1477.
- Fox, J., & Weisberg, S. (2011). Appendix - Nonparametric Regression in R. In *An R Companion to Applied Regression* (pp. 1–17).
- Halicioglu, F. (2011). Modeling Life Expectancy in Turkey. *Economic Modelling*, 28(5), 2075–2082. <https://doi.org/10.1016/j.econmod.2011.05.002>
- Nisa, K., Budiantara, I. N., & Rumiati, A. T. (2016). Modeling Multivariable Semiparametric Regression using Combined Estimators of Fourier Series and Truncated Spline. In *IPTEK Journal of Proceedings Series*.
- Nisa, K., Budiantara, I. N., & Rumiati, A. T. (2017). Multivariable Semiparametric Regression Model with Combined Estimator of Fourier Series and Kernel. In *IOP Conference Series: Earth and Environmental Science* (Vol. 58). <https://doi.org/10.1088/1755-1315/58/1/012028>
- Pane, R., Nyoman Budiantara, I., Zain, I., & Otok, B. W. (2014). Parametric and Nonparametric Estimators in Fourier Series Semiparametric Regression and Their Characteristics. *Applied Mathematical Sciences*, 8(101–104), 5053–5064. <https://doi.org/10.12988/ams.2014.46472>
- Rismal, Budiantara, I. N., & Prastyo, D. D. (2016). Mixture Model of Spline Truncated and Kernel in Multivariable Nonparametric Regression. In *AIP Conference Proceedings* (Vol. 1739). <https://doi.org/10.1063/1.4952565>
- Rosenblad, A. (2009). J.J. Faraway: Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models. *Computational Statistics*, 24(2), 369–370. <https://doi.org/10.1007/s00180-009-0152-1>
- Sugiantari. (2013). Analisis Faktor-Faktor yang Mempengaruhi Angka Harapan Hidup di Jawa Timur Menggunakan Regresi Semiparametrik Spline. *Jurnal Sains Dan Seni Pomits*, 2(1), D37–D41.
- Wayan Sudiarsa, I., Nyoman Budiantara, I., Suhartono, S., & Purnami, S. W. (2015). Combined Estimator Fourier Series and Spline Truncated in Multivariable Nonparametric Regression. *Applied Mathematical Sciences*, 9(97–100), 4997–5010. <https://doi.org/10.12988/ams.2015.55394>
- Wibowo, W., Haryatmi, S., & Budiantara, I. N. (2013). Modeling of Regional Banking Activities using Spline Multiresponse Semiparametric Regression. *International Journal of Applied Mathematics and Statistics*, 44(14), 391–398.