The Effect of Human Development Index on Poverty Model in Indonesia using Penalized Basis Spline Nonparametric Regression.

N Hasanah*, S Bahri and N Fitriyani
Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Mataram, Jl. Majapahit 62 Mataram, Indonesia

*Corresponding author: nurlizahasana07@gmail.com

Abstract. Poverty is an issue of special concern to various countries in the world, including Indonesia. One factor that can affect poverty is the Human Development Index (HDI). The aim of this study is to model of the problem of poverty based on HDI. Using nonparametric spline regression was used with the penalized basis spline (PB-Spline) approach, which can overcome the problem of selecting many knot points and the location of knots in spline regression. This study was obtained three models, i.e. model with one knot point, model with two knot points, and model with three knot points. Based on indicators of the value of Generalized Cross Validation (GCV) and the value of Mean Square Error (MSE), the best model was a model with three knots, with smoothing parameter value of 1000, 11.26236 value of GCV and 11.08420 value of MSE.

1. Introduction
Poverty is an issue of particular concern to various countries in the world, including Indonesia. Indonesia is a country with a high poverty rate, with more than 10% of Indonesia's population is categorized as poor [1]. Based on several previous studies, we can see that the Human Development Index (HDI) significantly affects poverty levels [2]. The impact of the HDI on poverty can be determined by analyzing the relationship between the two variables. A nonparametric regression approach can be used to estimate the model. This method is not bound by assumptions and has high flexibility because estimating the regression curve adjusts the data without being influenced by the researcher's subjectivity factor [3]. Some of the nonparametric regression approaches that have been developed are spline, kernel, MARS, Fourier series, wavelet, and others.

The spline is a polynomial slice that has segmented properties that are joined by knot points and can explain the character of the data [3]. The advantages of spline regression in study are that the model will tend to seek its estimation wherever the data moves. It can overcome fluctuating data patterns with knot points and produces relatively smooth curves [4]-[7]. The spline model's weakness during high order and the selection of a large number of knots can be overcome by using the B-splines basis, namely by building a base based on the knot point. The problem in creating the B-spline's basic function is the knots' determination and placement, namely the place where the polynomial pieces are connected to the B-splines [2].

P-splines contains a function that takes into account smoothing parameters [2]. The P-splines consists of two components namely B-splines component and the distinction penalty. The B-splines is
developed into a P-splines that is finished using the B-splines with the same knot distance, called the PB-splines [8]. Optimizing the number and placement of knots is resolved by first determining the number of knots used on the B-splines. Furthermore, the knots’ order is carried out with the concept of equal space knots, which is to adjust the knots’ position so that the distance between the knots is the same. Then to increase the smoothness, a differentiating penalty is given [9].

This study aims to model the percentage of poverty based on human development index using PB-Spline nonparametric regression. So that, through this model the government is expected to make efforts to manage resources to improve the human development index in order to reduce the percentage of Poverty.

2. Method
The analysis steps were:

a. Preparing for the data obtained from central Bureau of Statistics about to the problem of poverty in Indonesia year 2018.

b. Determining the variables used with Human Development Index as a independent variabel and Poverty Percentage as a dependent variable and then input the dependent variable and the independent variable.

c. Determining a descriptive statistics of each variable and scatter the plot between the response variable and the predictor variable to find out the initial data pattern.

d. Determining knot point by defining a new predictor that contains the unique value of the predictor and then sort. Next determine the warts of the new predictor as the knot point by dividing the new predictor as much as the part.

e. Determining the base order of B-Spline function to find the base value of the B-Spline function on PB-Spline estimator by using the following Equation [10]

\[ B_{j,p}(x) = \frac{x-k_j}{k_{j+p-1}-k_j} B_{j,p-1}(x) + \frac{k_{j+p}-x}{k_{j+p}-k_{j+1}} B_{j+1,p-1}(x) \]  

(1)

Where \( \alpha_j \) is a coefficient regression, \( B_{j-p,p}(x) \) is a B-spline define over the knot, \( p \) is a degree of B-spline function and \( k_j \) is a knot points.

f. Modeling the dependent variable using linear PB-splines regression with knots points obtained from quantile samples

\[ f(x_i) = \sum_{j=1}^{p+1} \alpha_j (z_i; D) B(x_i)_{j-p,p} \]  

(2)

g. Calculating the GCV value for each knot in the PB-spline linear regression model using the following equations

\[ GCV = \frac{n^{-1}\sum (y_i - \hat{y}_i)^2}{n^{-1}tracel[I - S]} \]  

(3)
h. Selecting of knot points and smoothing parameter based on the minimum GCV value of the PB-spline linear regression model for each knot points.

i. Modeling using the optimal knot points obtained.

j. Calculating the estimated parameter value.

\[
\hat{\alpha} = (B^T B + \lambda D_d D_d)^{-1} B^T Y
\]

k. Determining the best model based on MSE using the following equation \cite{8}

\[
MSE = n^{-1} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2
\]

l. Interpreting the models and calculating coefficient determinant \((R^2)\) using the following equation

\[
R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2}
\]

3. Result and Discussion

Before further analysis or modeling using PB-Spline nonparametric regression, the first thing to do is investigate the pattern of the relationship between dependent variables and independent variables by creating graphs. The form pattern relationship between dependent variables and independent variables can be seen in Figure 1.

![Figure 1. Scatter Plot between Human Development Index and Poverty Percentage.](Image)

The following scatter plot describes how data scatter patterns where it is known that data points have different patterns at certain intervals. Then the analysis is done using one of the regression analysis spline namely PB-Spline. Since \(n = 34\) in this case and only trial value of many knots point \((K)\) less than \(n - p - 1\) are used. And this study we used one, two and three knots for modeling poverty. The model is obtained by selecting the optimal finer parameters and knot points. Because in the selection of optimal knots and optimal finer parameters must meet the minimum GCV criteria. Figure 2 are three prediction models obtained by model using the optimal knots obtained.
Figure 2. Prediction Models with One Knot Points, Two Knot Points and Three Knot Points.

To measure the accuracy of the model obtained on the estimation used MSE. Table 1 is a comparison of MSE values for the optimal one, two and three knot models obtained.

| Many Knot | Knot Points | MSE  |
|-----------|-------------|------|
| 1         | $k_1 = 73.26647$ | 12.94277 |
| 2         | $k_1 = 64.26206$, $k_2 = 64.86235$ | 11.90754 |
| 3         | $k_1 = 64.26206$, $k_2 = 64.86235$, $k_3 = 75.06735$ | 11.08420 |

4. Conclusion

The study conducted three models is a model with a one-knot point, a model with two-knot points, and a model with three-knot points. Based on indicators of the value of Generalized Cross-Validation (GCV) and the amount of Mean Square Error (MSE), the best model was a model with three knots, with a 1000 value of smoothing parameter, 11.26236 value of GCV, 11.08420 value of MSE, and 56.1% value of the coefficient of determination ($R^2$). The best model obtained is listed in Equation (7).

$$\hat{y} = \hat{\alpha}_1 B_{-1,2} + \hat{\alpha}_2 B_{0,2} + \hat{\alpha}_3 B_{1,2} + \hat{\alpha}_4 B_{2,2} + \hat{\alpha}_5 B_{3,2}$$

$$= 25.702932 B_{-1,2} + 19.592885 B_{0,2} + 13.484852 B_{1,2} + 7.385023 B_{2,2} + 1.295194 B_{3,2}$$  (7)

On a simplified base and formed the following function
\[
\hat{y} = \begin{cases} 
474.9746852599 - 6.110047x & ,60.06 < x \leq 64.26206 \\
404.2761960647 - 6.108033x & ,64.26206 < x \leq 64.86235 \\
533.2621581981 - 6.099829x & ,64.86235 < x \leq 75.06735 \\
497.0460194941 - 6.089829x & ,75.06735 < x \leq 80.47 \\
0.60029 & ,x = 80.47 \\
10.205 & ,x = 5.40265 \\
\end{cases}
\] 

(8)

Acknowledgments

This work was supported by the Statistical Research Areas, Universitas Mataram.

References

[1] Nansadiqa L, Masbar R and Majid M S A 2019 Does economic growth matter for poverty reduction in Indonesia East African Scholars Journal of Economics, Business and Management 2 (2) 46-52

[2] Novianto, S and Sudarsono H 2018 Analysis Of Poverty Level in Districts/Cities Of Central Java Jurnal Ekonomi Pembangunan 16 (1) 1-12

[3] Fahrika A I, Salam H and Buhasyim M A 2020 Effect of Human Development Index (HDI), Unemployment, and Investment Realization toward Poverty in South Sulawesi-Indonesia The International Journal of Social Sciences World 2 (2) 110-116

[4] Massaid A, Hanif M, Febrianti D and Chamidah N 2019 Modelling of Poverty Percentage of Non-Food Per Capita Expenditures in Indonesia Using Least Square Spline Estimator IOP Conference Series: Materials Science and Engineering 546 (5) 052044

[5] Sriliana I, Sunandi E and Rafflesia U 2018 The Poverty Modeling Using Small Area Estimation with Semiparametric P-spline Asian Journal of Applied Sciences 6 (4). 11-19

[6] Mariati N P A M, Budiantara I N and Ratnasari V 2019 Modeling Poverty Percentages in the Papua Islands using Fourier Series in Nonparametric Regression Multivariable Journal of Physics: Conference Series 1397 (1) 012071

[7] Dewanti P, Budiantara I N and Rumiati A T 2020 Modelling of SDG’s Achievement in East Java Using Bi-responses Nonparametric Regression with Mixed Estimator Spline Truncated and Kernel. In Journal of Physics: Conference Series 1562 (1) 012016

[8] Iorio C, Frasso G, D’Ambrosio A and Siciliano R 2018 A P-spline based clustering approach for portfolio selection Expert Systems with Applications 95 88-103

[9] Wagner J, Münnich R, Hill J, Stoffers J and Udelhoven T 2017 Non-parametric small area models using shape-constrained penalized B-splines Journal of the Royal Statistical Society: Series A 180 (4) 1089-1109

[10] Zamani P, Amirabadi-Farahani M, Aliarabi H and Malecky M 2016 Comparison of different Legendre and B-Spline random regression models of lamb's estimate variance components for average birth weight per lambing in Mehraban sheep Iranian Journal of Animal Science 46 (4) 407-415