Advances in Social Science, Education and Humanities Research, volume 528
Proceedings of the 7th International Conference on Research, Implementation, and Education of Mathematics and Sciences (ICRIEMS 2020)

Smoothing Spline Estimator in Nonparametric Regression
(Application: Poverty in Papua Province)

Ni P. A. M. Mariati[1,2] I N. Budiantara[1,*] Vita Ratnasari[1]

[1] *Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia*
[2] *Universitas Mahasaraswati, Denpasar Bali, 80233, Indonesia*
[*]*Corresponding author. Email: nyomanbudiantara65@gmail.com*

ABSTRACT
Three estimates were obtained in estimating the regression curve, namely estimation of parametric regression, nonparametric regression and semiparametric regression. The most popular nonparametric regression option is smoothing spline. The advantage of smoothing spline is that it can use variable data at certain sub intervals, so this model needs to find its own data estimation. Smoothing Spline allows characters to function smoothly. In everyday life, data patterns are often found to change at certain sub-intervals, one of which is poverty data in Papua Province. Papua Province is ranked first in the percentage of poor people in Indonesia. The best nonparametric Smoothing Spline regression model for the poverty model in Papua Province with a generalized cross validation (GCV) value of 92.77 and R^2=99.99%.

Keywords: *Smoothing Spline, Poverty in Papua Province*.

1. INTRODUCTION

Nonparametric regression has high flexibility in estimating the regression curve. In the nonparametric regression view, the data are expected to find the estimation of the regression curve by themselves, without being influenced by the subjective factors of the study design [1]. Estimation methods that received a lot of attention and popular from several nonparametric regression researchers one of which is the Spline estimator. Spline has statistical properties that are useful for analyzing relationships in regression. Spline function has high flexibility and is able to handle data whose behavior changes in certain subintervals [2]. Spline is one form of estimator that is also often used in nonparametric regression because it has good visual interpretation, is flexible, and is able to handle smooth character functions [3-4]. The strength of Spline is that it can describe changes in behavior patterns of functions at certain sub intervals. Spline estimator approach is known by several methods such as Truncated Spline, Smoothing Spline, B-Spline, Exponential Spline and others. Among those methods that will be used to estimate the regression function is Smoothing Spline. Smoothing Spline estimates nonparametric regression functions that are assumed to be smooth in the sense that the function is contained within a particular function space and is often assumed to be in the sobolev space.

Smoothing Spline was first introduced by Whitaker in 1923 as a data pattern approach. Spline based on an optimization problem was developed by Reinsch in 1967 [5]. The Smoothing Spline estimator is obtained from a penalized least square (PLS) optimization [6]. Gurrin models Smoothing Spline in the case of biostatistics, in that case it is said that Smoothing Spline has a statistical interpretation and visual interpretation is very special and very good [7]. Besides that Smoothing Spline is able to handle data characters/functions that are smooth. Smoothing Spline also has a very good ability to handle data whose behavior changes at certain sub-intervals [8].

In everyday life data are often found to change at certain sub-intervals, one of which is poverty data. Poverty is a condition where a person or group of people cannot meet their economic needs to achieve prosperity and prosperity. Poverty is multidimensional in nature, which means that humans have many needs and vary so that poverty has various aspects, namely primary aspects consisting of assets, socio-political organizations, knowledge and skills [9]. Poverty Eradication is the goal of Sustainable development goals. One of the countries that has a problem with the percentage of poverty is Indonesia. Indonesia is a developing country with a very large population, with the fourth largest population in the world. There is a very close relationship between high unemployment, widespread poverty, education and uneven distribution of income. The level of open unemployment shows just the visible aspects of the employment problem in a developing country that is like the tip of an iceberg. If they do not work the consequence is that they cannot meet their needs properly, conditions like this have an impact on the creation and swelling of the amount of poverty [10]. Human development is the goal of development itself. Human development plays a role in shaping a country's ability to absorb modern technology and to develop its capacity to create sustainable growth and development [11].

Growth and poverty have a very strong correlation, because in the initial stages of the development process the poverty level tends to increase and as it approaches the final stage of development the number of poor people gradually decreases [12]. Research on increasing elementary school, junior high school, high school and university participation rates and economic growth has a negative and significant effect on poverty, meaning that increasing economic growth will reduce poverty. Likewise regarding education has a negative and significant effect on poverty, meaning that the higher the level of education will reduce poverty [13].

The unemployment rate has a significant effect on poverty levels, the government needs to improve the quality of human resources through improving the degree of public health and increasing access to education to remote areas [14]. In Rusdarti (2013) there is a table ranking the percentage of poor people by province in Indonesia, in that table, Papua Province was ranked first [15]. Research on macro mapping poverty. In this research based on the value of the poverty depth index because the depth index is a measure of the average expenditure gap of each poor population, in the study there was a ranking of the Poverty Depth Index in Indonesia, Papua Province was the highest depth index [16]. Papua Province is a province that has the highest poverty rate in Indonesia. So poverty programs and more attention from the government are still very much needed in Papua Province in overcoming the problems of poverty. Based on this, in this researches the percentage of poverty in Papua Province will be applied using the Smoothing Spline estimator in nonparametric regression.

2. RESULT AND DISCUSSION

The results and discussion will discuss the smoothing spline estimator in nonparametric regression and application of poverty data in Papua province. Which will be explained in subsection 2.1 and 2.2.

2.1. Smoothing Spline Estimator in Nonparametric Regression

The data provided is in pairs $\left(x_{1i}, x_{2i}, \ldots, x_{pi}, y_i\right)$ which is assumed that the predictor variables were $\left(x_{1i}, x_{2i}, \ldots, x_{pi}\right)$ and the respon variables y_i the relationship between the two variables follows nonparametric regression multivariable model:

$$y_i = f\left(x_{1i}, x_{2i}, \ldots, x_{pi}\right) + \varepsilon_i,\ i = 1, 2, \ldots, n.$$

Assume the multivariable nonparametric regression model is additive, so the regression model is obtained as follows:

$$y_i = \sum_{j=1}^{p} g_j\left(x_{ji}\right) + \varepsilon_i,\ i = 1, 2, \ldots, n.$$

Component regression curve $g_j(x_{ji})$ is assumed to be in Sobolev space $W_2^m\left[a_j, b_j\right],\ g_j \in W_2^m\left[a_j, b_j\right],\ \mathrm{j} = 1, 2, \ldots, \mathrm{p}$ with:

$$W_2^m\left[a_j, b_j\right] = \left\{ \int_{a_j}^{b_j} \left(g_j^{(m)}(x_{ji})\right)^2 dx_{ji} < \infty \right\} [17].$$

Component regression curve $\underset{\sim}{g}$ involves one predictor, then $\underset{\sim}{g}$ it can be written as follows: $\underset{\sim}{g} = (\underset{\sim}{g}_1^T, \underset{\sim}{g}_2^T, \ldots, \underset{\sim}{g}_n^T)^T$. It will first be explained for one predictor as follows: The component of a regression curve is a regression curve of unknown shape and assumed to be smooth in the sense of contained in space $W = W_2^m\left[a_j, b_j\right]$. Then H space can be decomposed into direct sum of two perpendicular W_0 and W_1 spaces, namely: $W = W_0 \oplus W_1$, with $W_0 \perp W_1$.

If the basis on W_0 space is $\left\{\theta_{j1}, \theta_{j2}, \ldots, \theta_{jm}\right\}$ where m is the polynomial spline order and the base on W_1 space is $\left\{\psi_{j1}, \psi_{j2}, \ldots, \psi_{jn}\right\}$ where n ad where n is the number of observations, then for each function $u_j \in W_0$ it can be written as follows:

$$u_j = c_{j1}\theta_{j1} + c_{j2}\theta_{j2} + \ldots + c_{jm}\theta_{jm} = \sum_{k=1}^{m} c_{jk}\theta_{jk} = \underset{\sim}{\theta}_j' \underset{\sim}{c}_j,$$

while for each function $v_j \in W_1$ as follows:

$$\begin{aligned} v_j &= \varsigma_{j1}\psi_{j1} + \varsigma_{j2}\psi_{j2} + \ldots + \varsigma_{jn}\psi_{jn} = \left(d_1\tau_j\right)\psi_{j1} + \\ &\quad + \left(d_2\tau_j\right)\psi_{j2} + \ldots + \left(d_n\tau_j\right)\psi_{jn} \\ &= \sum_{i=1}^{n} \left(d_i\tau_j\right)\psi_{ji} = \tau_j \underset{\sim}{\psi}_j' \underset{\sim}{d} \end{aligned}$$

where c_j and d_j a constant. So for each function $g_j \in W$ can be described as follows:

$$g_j = u_j + v_j = \sum_{k=1}^{m} c_{jk}\theta_{jk} + \sum_{i=1}^{n}\left(d_i\tau_j\right)\psi_{ji}$$
$$= \underset{\sim}{\theta}'_j \underset{\sim}{c}_j + \tau_j \underset{\sim}{\psi}'_j \underset{\sim}{d},\ j = 1, 2, \ldots, \text{p} \quad (1)$$

where

$\underset{\sim}{\theta}_j = \left\{\theta_{j1}, \theta_{j2}, ..., \theta_{jm}\right\}'$, $\underset{\sim}{\psi}_j = \left\{\psi_{j1}, \psi_{j2}, ..., \psi_{jn}\right\}'$

$\underset{\sim}{c}_j = \left\{c_{j1}, c_{j2}, ..., c_{jm}\right\}'$ and $\underset{\sim}{d} = \left(d_1, d_2, \ldots, d_n\right)'$.

Based on Riesz's Representation Theorem (Wahba, 1990) by describing L_x as linear functional is limited to space W and $g_j \in W$ then the equation can be presented as follows::

$$\mathrm{L}_x g_j = \mathrm{L}_x\left(u_j + v_j\right) = \mathrm{L}_x u_j + \mathrm{L}_x v_j = u_j(x_{ji}) + v_j(x_{ji})$$
$$= g_j(x_{ji})$$

L_x is a linear functional limited to space W, so a single value is obtained $\eta_i \in W$ which is a representation of L_x and satisfy the equation:

$$\mathrm{L}_x g_j = \left\langle \eta_{ji}, g_j \right\rangle \quad (2)$$
$$= g_j(x_{ji}),\ g_j \in W$$

with $\langle .,. \rangle$ is inner product. Based on the nature of the inner product, equation (2) can be written as:

$$g_j(x_i) = \left\langle \eta_{ji}, g_j \right\rangle = \left\langle \eta_{ji}, \underset{\sim}{\theta}'_j \underset{\sim}{c}_j + \tau_j \underset{\sim}{\psi}'_j \underset{\sim}{d} \right\rangle$$
$$= \left\langle \eta_{ji}, \underset{\sim}{\theta}'_j \underset{\sim}{c}_j \right\rangle + \left\langle \eta_{ji}, \tau_j \underset{\sim}{\psi}'_j \underset{\sim}{d} \right\rangle \quad (3)$$

Based on equation (3) for i = 1, it can be stated as follows:

$$g_j(x_{j1}) = \left\langle \eta_{ji}, \underset{\sim}{\theta}'_j \underset{\sim}{c}_j \right\rangle + \left\langle \eta_{ji}, \tau_j \underset{\sim}{\psi}'_j \underset{\sim}{d} \right\rangle$$
$$= \left\langle \eta_{j1}, \begin{pmatrix} \theta_{j1} & \theta_{j2} & \cdots & \theta_{jm} \end{pmatrix} \begin{pmatrix} c_{j1} \\ c_{j2} \\ \vdots \\ c_{jm} \end{pmatrix} \right\rangle +$$
$$+ \left\langle \eta_{j1}, \begin{pmatrix} \tau_j\psi_{j1} & \tau_j\psi_{j2} & \cdots & \tau_j\psi_{jn} \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{pmatrix} \right\rangle$$
$$= c_{j1}\left\langle \eta_{j1}, \theta_{j1} \right\rangle + c_{j2}\left\langle \eta_{j1}, \theta_{j2} \right\rangle + ... + c_{jm}\left\langle \eta_{j1}, \theta_{jm} \right\rangle +$$
$$d_1\tau_j\left\langle \eta_{j1}, \psi_{j1} \right\rangle + d_2\tau_j\left\langle \eta_{j1}, \psi_{j2} \right\rangle + ... + d_2\tau_j\left\langle \eta_{j1}, \psi_{jn} \right\rangle$$

If the process is continued in the same way, then for i = n obtained:

$$g_j(x_{jn}) = c_{j1}\left\langle \eta_{jn}, \theta_{j1} \right\rangle + c_{j2}\left\langle \eta_{jn}, \theta_{j2} \right\rangle + ... +$$
$$+ c_{jm}\left\langle \eta_{jn}, \theta_{jm} \right\rangle + d_1\tau_j\left\langle \eta_{jn}, \psi_{j1} \right\rangle + \quad (4)$$
$$+ d_2\tau_j\left\langle \eta_{jn}, \psi_{j2} \right\rangle + ... + d_2\tau_j\left\langle \eta_{jn}, \psi_{jn} \right\rangle$$

Based on equation (4) vectors $g_j(x_j)$ can be expressed in the form of:

$$\underset{\sim}{g}_j(x_j) = \begin{pmatrix} \underset{\sim}{g}_j(x_{j1}) \\ \underset{\sim}{g}_j(x_{j2}) \\ \vdots \\ \underset{\sim}{g}_j(x_{jn}) \end{pmatrix}$$
$$= \begin{pmatrix} \langle \eta_{j1}, \theta_{j1} \rangle & \langle \eta_{j1}, \theta_{j2} \rangle & \cdots & \langle \eta_{j1}, \theta_{jm} \rangle \\ \langle \eta_{j2}, \theta_{j1} \rangle & \langle \eta_{j2}, \theta_{j2} \rangle & \cdots & \langle \eta_{j2}, \theta_{jm} \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \eta_{jn}, \theta_{j1} \rangle & \langle \eta_{jn}, \theta_{j2} \rangle & \cdots & \langle \eta_{jn}, \theta_{jm} \rangle \end{pmatrix} \begin{pmatrix} c_{j1} \\ c_{j2} \\ \vdots \\ c_{jm} \end{pmatrix} +$$
$$\begin{pmatrix} \tau_j\langle \eta_{j1}, \psi_{j1} \rangle & \tau_j\langle \eta_{j1}, \psi_{j2} \rangle & \ldots & \tau_j\langle \eta_{j1}, \psi_{jn} \rangle \\ \tau_j\langle \eta_{j2}, \psi_{j1} \rangle & \tau_j\langle \eta_{j2}, \psi_{j2} \rangle & \ldots & \tau_j\langle \eta_{j2}, \psi_{jn} \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \tau_j\langle \eta_{jn}, \psi_{j1} \rangle & \tau_j\langle \eta_{jn}, \psi_{j2} \rangle & \ldots & \tau_j\langle \eta_{jn}, \psi_{jn} \rangle \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{pmatrix}$$
$$\underset{\sim}{g}_j = \mathbf{U}_j \underset{\sim}{c}_j + \tau_j \mathbf{V}_j \underset{\sim}{d} \quad (5)$$

with:

$$\mathbf{U}_j = \begin{pmatrix} \langle \eta_{j1}, \theta_{j1} \rangle & \langle \eta_{j1}, \theta_{j2} \rangle & \cdots & \langle \eta_{j1}, \theta_{jm} \rangle \\ \langle \eta_{j2}, \theta_{j1} \rangle & \langle \eta_{j2}, \theta_{j2} \rangle & \cdots & \langle \eta_{j2}, \theta_{jm} \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \eta_{jn}, \theta_{j1} \rangle & \langle \eta_{jn}, \theta_{j2} \rangle & \cdots & \langle \eta_{jn}, \theta_{jm} \rangle \end{pmatrix}, \underset{\sim}{c}_j = \begin{pmatrix} c_{j1} \\ c_{j2} \\ \vdots \\ c_{jm} \end{pmatrix}$$

$$\mathbf{V}_j = \begin{pmatrix} \tau_j\langle \eta_{j1}, \psi_{j1} \rangle & \tau_j\langle \eta_{j1}, \psi_{j2} \rangle & \ldots & \tau_j\langle \eta_{j1}, \psi_{jn} \rangle \\ \tau_j\langle \eta_{j2}, \psi_{j1} \rangle & \tau_j\langle \eta_{j2}, \psi_{j2} \rangle & \ldots & \tau_j\langle \eta_{j2}, \psi_{jn} \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \tau_j\langle \eta_{jn}, \psi_{j1} \rangle & \tau_j\langle \eta_{jn}, \psi_{j2} \rangle & \ldots & \tau_j\langle \eta_{jn}, \psi_{jn} \rangle \end{pmatrix}, \underset{\sim}{d} = \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{pmatrix}$$

If we taken $W = W_2^m\left[a_j, b_j\right],\ j = 1, 2, \ldots, p$ then:

$$\left\langle \eta_{ji}, \theta_{jk} \right\rangle = \mathrm{L}_{xj}\theta_{jk}$$
$$= \frac{x_{ji}^{k-1}}{(k-1)!}, i = 1, 2, \ldots, n; k = 1, 2, \ldots, m$$

Because,

$$\left\langle \eta_{ji}, \psi_{ji} \right\rangle = \left\langle \theta_{jk} + \psi_{ji}, \psi_{ji} \right\rangle$$
$$= \left\langle \theta_{jk} + \psi_{jt} \right\rangle + \left\langle \psi_{ji}, \psi_{jt} \right\rangle$$
$$\left\langle \eta_{ji}, \psi_{ji} \right\rangle = \left\langle \psi_{ji}, \psi_{jt} \right\rangle$$

then the matrix $\mathbf{V}_j$ can be written as:

$$\mathbf{V}_j = \begin{pmatrix} \langle \psi_{j1}, \psi_{j1} \rangle & \langle \psi_{j1}, \psi_{j2} \rangle & \cdots & \langle \psi_{j1}, \psi_{jn} \rangle \\ \langle \psi_{j2}, \psi_{j1} \rangle & \langle \psi_{j2}, \psi_{j2} \rangle & \cdots & \langle \psi_{j2}, \psi_{jn} \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \psi_{jn}, \psi_{j1} \rangle & \langle \psi_{jn}, \psi_{j2} \rangle & \cdots & \langle \psi_{jn}, \psi_{jn} \rangle \end{pmatrix}$$

so the spline estimator form can be stated in the following form:

$$\begin{aligned} \underset{\sim}{g}(x_1, x_2, \ldots, x_p) &= \sum_{j=1}^{p} g_j(x_j) \\ &= \sum_{j=1}^{p} \left(\mathbf{U}_j \underset{\sim}{c}_j + \tau_j \mathbf{V}_j \underset{\sim}{d} \right) \\ &= \mathbf{U}\underset{\sim}{c} + \mathbf{V}_{\underset{\sim}{\tau}} \underset{\sim}{d} \end{aligned} \tag{6}$$

where

$$\mathbf{U} = \begin{pmatrix} \mathbf{U}_1 & \mathbf{U}_1 & \ldots & \mathbf{U}_p \end{pmatrix}, \ \underset{\sim}{c} = \begin{pmatrix} \underset{\sim}{c}_1 & \underset{\sim}{c}_2 & \ldots & \underset{\sim}{c}_p \end{pmatrix}',$$

$$\mathbf{V}_\tau = \begin{pmatrix} \tau_1 \mathbf{V}_1 & \tau_2 \mathbf{V}_2 & \ldots & \tau_p \mathbf{V}_p \end{pmatrix}, \ \underset{\sim}{d} = \begin{pmatrix} \underset{\sim}{d}_1 & \underset{\sim}{d}_2 & \ldots & \underset{\sim}{d}_n \end{pmatrix}'.$$

Furthermore, the description of the penalty components $\sum_{j=1}^{p} \lambda_j \int_{a_j}^{b_j} \left[g_j^{(m)}(x_j) \right]^2 dx_j$ obtained by the decomposition as follows:

$$\begin{aligned} \sum_{j=1}^{p} \lambda_j \int_{a_j}^{b_j} \left[g_j^{(m)}(x_j) \right]^2 dx_j &= \sum_{j=1}^{p} \frac{\lambda}{\tau_j} \int_{a_j}^{b_j} \left[g_j^{(m)}(x_j) \right]^2 dx_j \\ &= \lambda \sum_{j=1}^{p} \tau_j^{-1} \int_{a_j}^{b_j} \left[g_j^{(m)}(x_j) \right]^2 dx_j \end{aligned}$$

where $\int_{a_j}^{b_j} \left[g_j^{(m)}(x_j) \right]^2 dx_j = \left\| P_1 g_j \right\|^2$, then

$$\lambda \sum_{j=1}^{p} \tau_j^{-1} \int_{a_j}^{b_j} \left[g_j^{(m)}(x_j) \right]^2 dx_j = \lambda \sum_{j=1}^{p} \left\| P_1 g_j \right\|^2 .$$

Next,

$$\begin{aligned} \left\| P_1 g_j \right\|^2 &= \left\langle P_1 g_j, P_1 g_j \right\rangle = \left\langle P_1 \left(\underset{\sim}{\theta}_j' \underset{\sim}{c}_j + \tau_j \underset{\sim}{\psi}_j' \underset{\sim}{d} \right), P_1 \left(\underset{\sim}{\theta}_j' \underset{\sim}{c}_j + \tau_j \underset{\sim}{\psi}_j' \underset{\sim}{d} \right) \right\rangle \\ &= \underset{\sim}{d}' \begin{pmatrix} (\tau_j \psi_{j1}), (\tau_j \psi_{j1}) & (\tau_j \psi_{j1}), (\tau_j \psi_{j2}) & \cdots & (\tau_j \psi_{j1}), (\tau_j \psi_{jn}) \\ (\tau_j \psi_{j2}), (\tau_j \psi_{j1}) & (\tau_j \psi_{j2}), (\tau_j \psi_{j2}) & \cdots & (\tau_j \psi_{j2}), (\tau_j \psi_{jn}) \\ \vdots & \vdots & \ddots & \vdots \\ (\tau_j \psi_{jn}), (\tau_j \psi_{j1}) & (\tau_j \psi_{jn}), (\tau_j \psi_{j2}) & \cdots & (\tau_j \psi_{jn}), (\tau_j \psi_{jn}) \end{pmatrix} \underset{\sim}{d} \\ &= \underset{\sim}{d}' \tau_j \mathbf{V}_j \underset{\sim}{d} \end{aligned} \tag{7}$$

So that the quantity of components is obtained by substituting equation (7) into the penalty component, it can be obtained:

$$\begin{aligned} \sum_{j=1}^{p} \lambda_j \int_{a_j}^{b_j} \left[g_j^{(m)}(x_{ji}) \right]^2 dx_{ji} &= \lambda \sum_{j=1}^{p} \left\| P_1 g_j \right\|^2 \\ &= \lambda \sum_{j=1}^{p} \underset{\sim}{d}' \tau_j \mathbf{V}_j \underset{\sim}{d} \\ &= \lambda \underset{\sim}{d}' \mathbf{V}_{\underset{\sim}{\tau}} \underset{\sim}{d} \end{aligned} \tag{8}$$

where $\lambda_j = \dfrac{\lambda}{\tau_j}, \ j = 1, 2, \ldots, p,$

$\underset{\sim}{d}' = \begin{pmatrix} d_1 & d_2 & \ldots & d_n \end{pmatrix}$ and

$\mathbf{V}_{\underset{\sim}{\tau}} = \tau_1 \mathbf{V}_1 + \tau_2 \mathbf{V}_2 + \ldots + \tau_p \mathbf{V}_p$

The component Goodness of fit is:

$$n^{-1} \left(\underset{\sim}{y} - \underset{\sim}{g} \right)' \left(\underset{\sim}{y} - \underset{\sim}{g} \right) = n^{-1} \left(\underset{\sim}{y} - \mathbf{U}\underset{\sim}{c} - \mathbf{V}\underset{\sim}{d} \right)' \left(\underset{\sim}{y} - \mathbf{U}\underset{\sim}{c} - \mathbf{V}\underset{\sim}{d} \right) \tag{9}$$

After getting the goodness of fit component in equation (9) and the penalty component in equation (8), then complete the optimization of Penalized Least Squares (PLS) as follows:

$$\begin{aligned} & \underset{g_j \in W_2^m[a_j, b_j]}{Min} \left\{ n^{-1} \left(\underset{\sim}{y} - \mathbf{U}\underset{\sim}{c} - \mathbf{V}_{\underset{\sim}{\tau}} \underset{\sim}{d} \right)' \left(\underset{\sim}{y} - \mathbf{U}\underset{\sim}{c} - \mathbf{V}_{\underset{\sim}{\tau}} \underset{\sim}{d} \right) + \lambda \underset{\sim}{d}' \mathbf{V}_{\underset{\sim}{\tau}} \underset{\sim}{d} \right\} \\ &= \underset{g_j \in W_2^m[a_j, b_j]}{Min} \left\{ \square (\underset{\sim}{c}, \underset{\sim}{d}) \right\} \end{aligned} \tag{10}$$

$$\frac{\square\, \partial(\underset{\sim}{c}, \underset{\sim}{d})}{\partial \underset{\sim}{d}} = -2\mathbf{V}_{\underset{\sim}{\tau}}' \underset{\sim}{y} + 2\mathbf{V}_{\underset{\sim}{\tau}}' \mathbf{U}\underset{\sim}{c} + 2\mathbf{V}_{\underset{\sim}{\tau}}' \mathbf{V}_{\underset{\sim}{\tau}} \hat{\underset{\sim}{d}} + 2n\lambda \mathbf{V}_{\underset{\sim}{\tau}}' \hat{\underset{\sim}{d}} = 0$$

$$-\underset{\sim}{y} + \mathbf{U}\underset{\sim}{c} + \left[\mathbf{V}_\tau + n\lambda \mathbf{I} \right] \hat{\underset{\sim}{d}} = 0 \tag{11}$$

For example $M = \mathbf{V}_\tau + n\lambda \mathbf{I}$, so equation (11) becomes:

$$\hat{\underset{\sim}{d}} = \mathbf{M}^{-1} \left(\underset{\sim}{y} - \mathbf{U}\underset{\sim}{c} \right) \tag{12}$$

$$\frac{\partial \square\, (\underset{\sim}{c}, \underset{\sim}{d})}{\partial \underset{\sim}{c}} = -2\mathbf{U}'\underset{\sim}{y} + 2\mathbf{U}'\mathbf{U}\underset{\sim}{c} + 2\mathbf{U}'\mathbf{V}_{\underset{\sim}{\tau}} \underset{\sim}{d} = 0 \tag{13}$$

Substitution equation (12) to equation (13), so we get:

$$-2\mathbf{U}'\underset{\sim}{y} + 2\mathbf{U}'\mathbf{U}\hat{\underset{\sim}{c}} + 2\mathbf{U}'\mathbf{V}_{\underset{\sim}{\tau}} \mathbf{M}^{-1} \left(\underset{\sim}{y} - \mathbf{U}\underset{\sim}{c} \right) = 0 \tag{14}$$

Given, $M = \mathbf{V}_{\underset{\sim}{\tau}} + n\lambda \mathbf{I}$, then $\mathbf{V}_{\underset{\sim}{\tau}} = \mathbf{M} - n\lambda \mathbf{I}$. As a result, the following equation is obtained:

$$\begin{aligned} \mathbf{V}_{\underset{\sim}{\tau}} \mathbf{M}^{-1} &= (\mathbf{M} - n\lambda \mathbf{I}) \mathbf{M}^{-1} \\ &= \left(\mathbf{I} - n\lambda \mathbf{M}^{-1} \right) \end{aligned} \tag{15}$$

By substituting equation (15) for equation (14) then solving it is obtained:

$$\hat{\underset{\sim}{c}} = \left(\mathbf{U}'\mathbf{M}^{-1}\mathbf{U} \right)^{-1} \mathbf{U}'\mathbf{M}^{-1} \underset{\sim}{y} \tag{16}$$

then substitute equation (16) for equation (12), so that:

$$\hat{\underset{\sim}{d}} = \mathbf{M}^{-1} \left(\mathbf{I} - \mathbf{U} \left(\mathbf{U}'\mathbf{M}^{-1}\mathbf{U} \right)^{-1} \mathbf{U}'\mathbf{M}^{-1} \right) \underset{\sim}{y} \tag{17}$$

Thus, the Smoothing Spline model becomes:

$$\begin{aligned} \hat{g} &= \mathbf{U}\hat{\underset{\sim}{c}} + \mathbf{V}_\tau \hat{\underset{\sim}{d}} \\ &= \Big[\mathbf{U} \left(\mathbf{U}'\mathbf{M}^{-1}\mathbf{U} \right)^{-1} \mathbf{U}'\mathbf{M}^{-1} + \\ &\qquad \mathbf{V}_\tau \mathbf{M}^{-1} \left(\mathbf{I} - \mathbf{U} \left(\mathbf{U}'\mathbf{M}^{-1}\mathbf{U} \right)^{-1} \mathbf{U}'\mathbf{M}^{-1} \right) \Big] \underset{\sim}{y} \end{aligned} \tag{18}$$

2.2. Application Data Poverty in Papua Province use Smoothing Spline Estimator in Nonparametric Regression

In the section smoothing spline estimator in the nonparametric regression using poverty data in Papua Province. The variables of interests were taken from the 2019 poverty data. Namely data on poverty in the province of Papua The response variable is the percentage of Poverty (y) while successive predictor variables Open Unemployment Rate (x_1), Labor Force Participation Rate (x_2), Human Development Index (x_3), and income per capita (x_4).

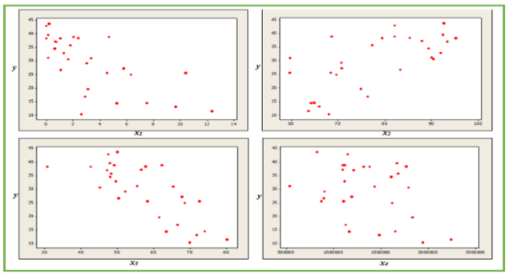

Figure 1 Scatterplot between the response variable with each predictor variable

Figure 1 shows that the data tend to change on a particular subinterval such as pattern smoothing spline. Fitting the poverty data using the smoothing spline regression model yield a GCV of 92.77, *MSE* is 0.001 and R^2=99.99%. The smoothing parameters obtained respectively for $x_1 = 1.0$; $x_2 = 0.2$; $x_3 = 0.1$; $x_4 = 0.4$. The R^2 value indicates that this model can explain poverty as much as 99,99%.

TABLE 1. y and $\hat{y}$

No	y	$\hat{y}$	No	y	$\hat{y}$
1	10.35	10.35089	16	14.41	14.41147
2	38.33	38.32784	17	16.83	16.82931
3	13.13	13.13149	18	30.95	30.94896
4	24.81	24.80922	19	38.79	38.78915
5	27.13	27.12984	20	29.13	29.1321
6	25.5	25.4999	21	38.24	37.1104
7	37.16	37.15978	22	39.52	39.51892
8	35.71	35.71093	23	36.93	36.93024
9	14.54	14.54035	24	34.52	34.52023
10	19.66	19.66033	25	38.24	38.24026
11	25.5	25.49976	26	31.12	31.11967
12	26.6	26.60023	27	42.92	42.91941
13	38.82	38.81876	28	43.65	43.64971
14	30.51	30.51094	29	11.49	11.49082
15	32.9	32.89949			

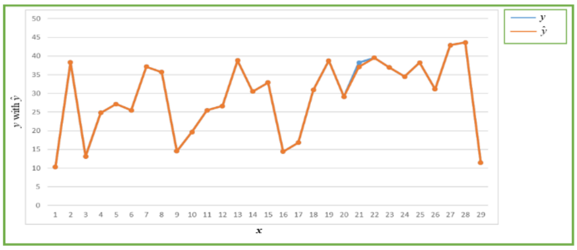

Figure 2 Line plot between the y with $\hat{y}$

Table 1 and Figure 2 shows that the smoothing spline regression model is good because the estimated value of y and $\hat{y}$ is relatively small. Which means that the estimation $\hat{y}$ results are close to y. Table 1 and Figure 2, the fitted values are closed to the observations. This indicates that the model can predict well. Figures 2, so that the estimation formula in equation (18) can be used appropriately to estimate the data poverty in Papua Province.

3. CONSCLUSION

Smoothing Spline Estimator in nonparametric regression is as follows:

$$\hat{g} = \left[\mathbf{U}\left(\mathbf{U}'\mathbf{M}^{-1}\mathbf{U}\right)^{-1}\mathbf{U}'\mathbf{M}^{-1} + \mathbf{V}_\tau\mathbf{M}^{-1}\left(\mathbf{I} - \mathbf{U}\left(\mathbf{U}'\mathbf{M}^{-1}\mathbf{U}\right)^{-1}\mathbf{U}'\mathbf{M}^{-1}\right)\right]\underset{\sim}{y}$$

with

$$\mathbf{U} = \left(\mathbf{U}_1 \quad \mathbf{U}_1 \quad \ldots \quad \mathbf{U}_p\right), \quad \mathbf{V}_\tau = \left(\tau_1\mathbf{V}_1 \quad \tau_2\mathbf{V}_2 \quad \ldots \quad \tau_p\mathbf{V}_p\right)$$

and $M = \mathbf{V}_\tau + n\lambda\mathbf{I}$. In the application of poverty data in the Province of Papua, the smoothing spline regression model has a GCV of 92.77, *MSE* be obtained 0.001 and R^2=99.99%. This means that this model can explain poverty as much as 99.99%.

AUTHORS' CONTRIBUTIONS

Conceptualization, N.P.A.M.M., I.N.B., and V.R.; methodology, N.P.A.M.M., and I.N.B.; software, N.P.A.M.M., and V.R.; validation, N.P.A.M.M., I.N.B., and V.R.; formal analysis, N.P.A.M.M., and I.N.B.; investigation, N.P.A.M.M.; data curation, N.P.A.M.M.; writing original draft preparation, N.P.A.M.M.; writing review and editing, I.N.B., and V.R.; visualization, V.R.; supervision, I.N.B., and V.R.; project administration, I.N.B., and V.R.; All authors have read and agreed to the published version of the manuscript.

ACKNOWLEDGMENTS

Ni Putu Ayu Mirah Mariati thanks the (DIKTI), *Kemendikbud*, the Republic of Indonesia, which has given the BPPDN scholarship. The authors thank *Direktorat Jenderal Penguatan Riset dan Pengembangan (DIKTI), Kemendikbud*, the Republic of Indonesia, which funded this study via grant *Penelitian Disertasi Doktor (PDD)* in 2020.

REFERENCES

[1] G. Wahba, Bayesian Confidence Intervals for The Cross-validated Smoothing Spline, J.R. Statist. Ser.B., 1983, 45(1), pp.133-150, DOI: https://doi.org/10.1111/j.2517-6161.1983.tb01239.x.

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

[3] C. Gu and G. Wahba, Smoothing Spline ANOVA with Component-Wise Bayesian Confidence Intervals, Journal of Computational and Graphical Statistics, 1993, 2(1), pp. 97-117, DOI: https://doi.org/10.1080/10618600.1993.10474601.

[4] I. N. Budiantara, V. Ratnasari, M. Ratna, and Z. Ismaini, The Combination of Spline and Kernel Estimator for Nonparametric Regression and its Properties, Apllied Mathematical Sciences, 2015 9(122), pp.6083-6094, DOI: http://dx.doi.org/10.12988/ams.2015.58517.

[5] G. Wahba, Spline Models For Observasion Data Pensylvania, SIAM, 1990.

[6] S.N. Wood, On Confidence Intervals for Generalized Additive Models Based on Penalized Regression Spline, Australian & New Zealand Journal Of Statistics., 2006, 48, pp.445-464, DOI:http://dx.doi.org/10.1111/j.1467-842X.2006.00450.x

[7] C.L. Gurrin, K.J. Scurrah and M.L.Hazelton, Tutorial in Bio Statistics:Spline Smoothing with Linear Mixed Models, Statistics in Medicine, 2005, 24, pp.3361-3381, DOI:https://doi.org/10.1002/sim.2193..

[8] R. Hidayat, I.N. Budiantara, B.W. Otok and V. Ratnasari, The Regression Curve Estimation by Using Mixed Smoothing Spline and Kernel (Mss-K) Model, Communications in Statistics-Theory and Methods, 2020, pp. 1–12, DOI: https://doi.org/10.1080/03610926.2019.1710201.

[9] World Bank, Era Baru dalam Pengentasan Kemiskinan di Indonesia, Indonesia, Bank Dunia, 2007.

[10] L. Arsyad, Pengantar Perencanaan dan Pembangunan Ekonomi Daerah Yogyakarta, BPFE, 1999.

[11] M.P. Todaro and C.S Stephen, Pembangunan Ekonomi, Jakarta, Erlangga, 2006.

[12] S. Sukirno, Makro Ekonomi Teori Pengantar, Jakarta, Raja Grafindo Persada, 2012.

[13] B. Teerawichitchainan and J. Knodel, Long-Term Care Needs in the Context of Poverty and Population Aging: The Case of Older Persons in Myanmar, Journal of Cross-Cultural Gerontolog, 2018, 33(2), pp.143–162, DOI: https://doi.org/10.1007/s10823-017-9336-2.

[14] I. P. T. Suepa, I. Gani and Muliati, Pengaruh Pertumbuhan Ekonomi dan Pengangguran Serta Inflasi Terhadap Kemiskinan di Kalimantan Timur, Jurnal Ekonomi Mulawarman, 2018, 3(3), DOI: http://dx.doi.org/10.29264/jiem.v3i3.3753.

[15] Rusdarti and Karolina, Faktor-faktor yang Mempengaruhi Tingkat Kemiskinan di Provinsi Jawa Tengah, Jurnal Economia, 2013, 9(1), DOI: https://doi.org/10.21831/economia.v9i1.1371.

[16] A. E. S. Saputro and A. P. Utomo, Faktor-faktor yang Mempengaruhi Kemiskinan secara Makro di Lima Belas Provinsi Tahun 2007, Jurnal dan Organisasi Manajemen, 2010, 6(2), pp. 89-100, DOI:http://www.jurnal.ut.ac.id/index.php/jom/article/view/285.

[17] N. P. A. M Mariati, I. N. Budiantara and V. Ratnasari, Combination Estimation of Smoothing Spline and Fourier Series in Nonparametric Regression, Journal of Mathematics, 2020, pp. 1-10, DOI: https://doi.org/10.1155/2020/4712531.