**International Journal of Academic and Applied Research (IJAAR)**
**ISSN: 2643-9603**
**Vol. 3 Issue 12, December – 2019, Pages: 75-78**

# Use of Penalized Spline Linear to Identify Change in Pattern of Blood Sugar based on the Weight of Diabetes Patients

**Anna Islamiyati[1], Raupong[2], Anisa[3]**
[1,2,3]Departement of Statistics, Hasanuddin University, Makassar, 90245, Indonesia
annaislamiyati@unhas.ac.id

***Abstract:*** Penalized linear spline regression contains first degree, knots and smoothing parameters which work simultaneously in the modeling. The flexible nature of the spline makes it possible to model data that has the possibility of variations of patterns in one model. That problem cannot be overcome by using linear regression in the parametric approach. Therefore, in this article, we show the use of the penalized spline linear regression models that contain several segments as a form of pattern in certain intervals. Furthermore, it is applied in the relationship of body weight with blood sugar in diabetic patients. Based on the model, we got 2 patterns of changes in blood sugar levels based on body weight. There is an upward pattern in body weight of less than 58.5 kg and a downward pattern in body weight from 58.5 kg upwards.

**Keywords—** cross section; diabetes; knot; penalized spline; smoothing parameter

## 1. INTRODUCTION

The diversity of data is increasingly found in the real world. The tendency of the increasingly large data at this time is a challenge for researchers to analyze big data into a reference decision. A common problem that cannot be avoided for big data is irregular patterns. Fluctuations go up and down, data spread unevenly, outliers get bigger, and there are still many data problems that will be found in real data. The parametric approach can only be used when the form of its function follows a parametric form causing limitation of that method. Therefore, developing a nonparametric regression approach that can be used in any data condition.

In the nonparametric regression approach, several estimators that have been developed include Budiantara et al. (2012) [1] with the weighted spline estimator. Chamidah et al. (2012) [2] with local polynomial, Chamidah and Saifuddin. (2013) [3] with kernel estimator, Aydin and Yilmaz (2018) [4], Lestari et al. (2019) [5] with smoothing spline, and Islamiyati (2018, 2019) [6, 7] uses truncated polynomial spline. Spline is an estimator that is widely used in real data because of the flexibility of the model through the knots. Montoya and Miller (2014) [8], Islamiyati et al. (2017) [9] develops penalized spline. The advantage of the penalized spline is the involvement of knots and smoothing parameters in smoothing the regression curve. Therefore, penalized spline is more easily interpreted visually and it is able to produce a smooth model estimation. This has been shown by Islamiyati et al. (2019) [10] through simulation studies, Islamiyati et al. (2018) [11] on bi-response longitudinal data, and Islamiyati et al. (2019) [12] on the multi-predictor regression model. However, this research is included in the longitudinal study.

One big data usually contain factors which can be measured repeatedly or called longitudinal data. In addition, there are also factors that can only be measured once, but have big data, such as body weight, height, and age. Therefore, this study emphasizes the measurement data only once or what is commonly called cross sectional. Islamiyati et al. (2019) [13] have modeled diabetes data, but only considered treatment time as longitudinal data. In this study, we examined body weight in relation to blood sugar from diabetic patients. We use a penalized linear spline in which the placement of knots is done by a trial and error system that takes into account the minimum GCV value. Data on blood sugar levels in diabetics have a tendency to fluctuate at each measurement time, so it is suitable to be modeled with the penalized spline approach. We examined variations in trends in blood sugar levels in different weight segments. Each segment produced by linear spline penalized spline regression can provide information about changes in blood sugar levels. That information is a variation or condition of certain body weight that should be of medical and patient concern in relation to blood sugar levels for diabetics.

## 2. ESTIMATION OF PENALIZED SPLINE LINEAR REGRESSION MODEL

The penalized spline estimator is formed from the truncated spline function in the Penalized Least Square (PLS) criteria. The penalized spline estimator uses knots and smoothing parameters simultaneously in estimating the nonparametric regression function. Ruppert (2002) [14] explains that the function is truncated with the order $q$ based on the knot point $a < K_1 < \ldots < K_d < b$, expressed by $f(t_i)$ as in equation (1). Functions $f(t_i)$ can be expressed in matrix form, namely:

$$\underset{\sim}{f} = \mathbf{X}\underset{\sim}{\beta}, \tag{1}$$

**International Journal of Academic and Applied Research (IJAAR)**
**ISSN: 2643-9603**
**Vol. 3 Issue 12, December – 2019, Pages: 75-78**

where $\mathbf{X} = \begin{bmatrix} 1 & x_1 & (x_1 - K_1)_+ & \dots & (x_1 - K_d)_+ \\ 1 & x_2 & (x_2 - K_1)_+ & & (x_2 - K_d)_+ \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & (x_n - K_1)_+ & \dots & (x_n - K_d)_+ \end{bmatrix}$, and $\underset{\sim}{\beta} = [\beta_0 \quad \beta_1 \quad \beta_2 \quad \dots \quad \beta_{1+d}]^T$.

Penalized spline estimators through PLS criteria formed from truncated functions are as follows:

$$\text{PLS} = \sum_{i=1}^{n} \left(y_i - f(t_i)\right)^2 + \lambda \int_a^b \left[ f^{(m)}(t) \right]^2 dt \ . \tag{2}$$

for each f in the Sobolev space $W_2^m[a,b] = \left\{ f \left| \int_a^b \left[ f^{(m)}(t) \right]^2 dt < \infty \right. \right\}$. If $W_2^m = P \oplus L$, then $f = g + h$, for each $g \in P$, and $h \in L$.

Equation (2) can also be stated as:

$$\text{PLS} = \left\{ \sum_{i=1}^{n} \left(y_i - f(t_i)\right)^2 + \lambda \sum_{v=1}^{d} \beta_{(1+v)}^2 \right\}. \tag{3}$$

Equation (3) can also be stated as follows:

$$\text{PLS} = \left(\underset{\sim}{y} - \mathbf{X}\underset{\sim}{\beta}\right)^T \left(\underset{\sim}{y} - \mathbf{X}\underset{\sim}{\beta}\right) + \lambda \underset{\sim}{\beta}^T \mathbf{D} \underset{\sim}{\beta} \ . \tag{4}$$

where $\lambda$ is smoothing parameter, $v$ is the number of knots $(v = 1, 2, \dots, d)$, $\underset{\sim}{\beta}$ is a spline regression parameter vector, and **D** is a diagonal matrix containing 0 and 1 knot points, or $\mathbf{D} = \text{diag}\left(\underset{\sim}{0}_{q+1}, \underset{\sim}{1}_d\right)$.

The estimation of parameter $\underset{\sim}{\beta}$ is obtained by reducing PLS in equation (4) to b $\underset{\sim}{\beta}$, namely:

$$\hat{\underset{\sim}{\beta}} = \left(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{D}\right)^{-1} \mathbf{X}^T \underset{\sim}{y} \tag{5}$$

Based on equation (5), the estimation of the penalized spline regression function is stated:

$$\hat{\underset{\sim}{f}}(t) = \mathbf{X}\hat{\underset{\sim}{\beta}} = \mathbf{X}\left(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{D}\right)^{-1} \mathbf{X}^T \underset{\sim}{y} \ ,$$

where $\lambda$ is smoothing parameter and $\mathbf{D} = \text{diag}\left(\underset{\sim}{0}_{q+1}, \underset{\sim}{1}_d\right)$.

Furthermore, the GCV criteria obtained in the penalized spline linear nonparametric regression model are as follows:

$$\text{GCV}(\underset{\sim}{\lambda}) = \frac{\left(\underset{\sim}{y} - \hat{\underset{\sim}{f}}(\underset{\sim}{t})\right)^T \left(\underset{\sim}{y} - \hat{\underset{\sim}{f}}(\underset{\sim}{t})\right)}{N^{-1}\left(\text{trace}\left(\mathbf{I} - \mathbf{A}(\underset{\sim}{\lambda})\right)\right)^2} \ , \tag{6}$$

where $\mathbf{A}(\underset{\sim}{\lambda}) = \mathbf{X}\left(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{D}\right)^{-1} \mathbf{X}^T$ is the hat matrix of the refinement parameter of size $n$ x $n$.

## 3. Patterns of Change of Blood Sugar based on Weight

Blood sugar levels in diabetics vary greatly. In this article, the data tested were 81 diabetic patients along with the weight of each patient.

### 3.1 Descriptive of Data

Data analyzed are shown in Table 1 which is related to data on blood sugar levels and patient body weight.

**Table 1**. Descriptive of patient's blood sugar and body weight

| Variable | Min | Max | Average | Stdev |
|---|---|---|---|---|
| Blood sugar | 126 | 549 | 292 | 107.64 |
| Weight | 32 | 85 | 57,76 | 11.38 |

Based on Table 1, the highest blood sugar in diabetic patients is around 540 mg/dl and the lowest is 126 mg/dl with an average of 292 mg/dl. This figure is certainly a high value of the normal blood sugar standard of around 130 mg/dl. Furthermore, diabetics have the highest body weight of 85 kg and the lowest of 32 kg with an average of 57.76 kg. Then the data is plotted as in Figure 1.

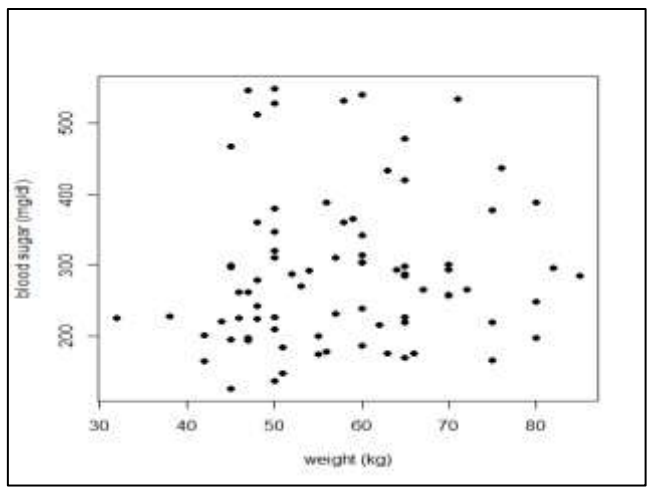


**Fig 1**. Scatter plot of data

Based on Figure 1, the plot between blood sugar and body weight does not show a well-known parametric pattern, for example linear, quadratic or cubic. Therefore, the approach used to model the data is penalized spline linear nonparametric regression.

**International Journal of Academic and Applied Research (IJAAR)**
**ISSN: 2643-9603**
**Vol. 3 Issue 12, December – 2019, Pages: 75-78**

### 3.2 Data Analysis with Penalized Spline Linear Regression

Data analyzed by the penalized linear spline regression considers knot points and smoothing parameters. The knots are tried from 1 to 3 knots with a smoothing parameter of 0.1. The complete results are shown in Figure 2.

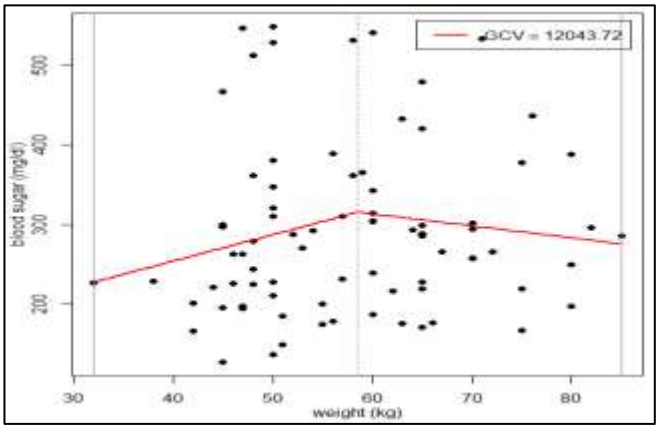


**Fig 2**. Estimation of the penalized spline linear regression curve at 1 knot point, $K = 58.5$.

Figure 2 shows 2 patterns of changes in blood sugar with the selected knot point being 58.5. The GCV value for the lambda smoothing parameter = 0.1 is 12043.72.

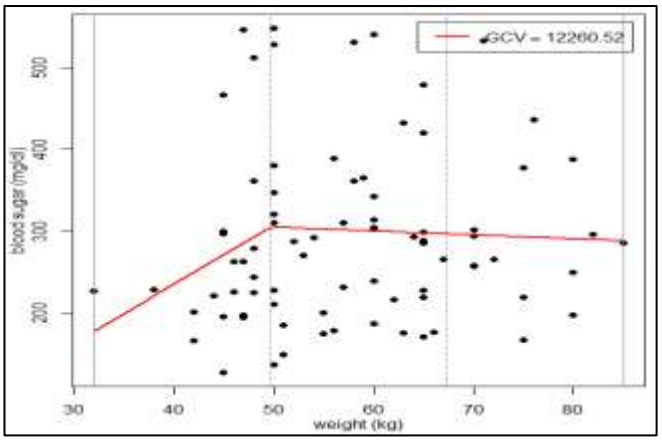


**Fig 3**. Estimation of the penalized spline linear regression curves at 2 knots, namely $K_1 = 49.7$ and $K_2 = 67.3$.

Figure 3 is a penalized linear spline regression model for 2 knots, namely 49.7 and 67.3. There are 3 patterns of change that occur based on these knots.

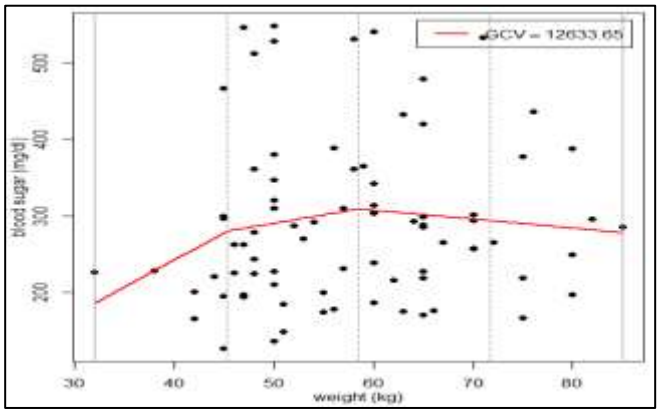


**Fig 4**. Estimation of the penalized spline linear regression curve at 3 knots, $K_1 = 45.25$, $K_2 = 58.50$, $K_3 = 71.75$.

Based on Figure 2-4 shows that the regression model that gives a minimum GCV value is at 1 knot point. Therefore, the model used to identify variations in changes in blood sugar levels based on body weight is a penalized spline linear regression model with 1 knot point. This shows there are 2 patterns of changes in diabetic blood sugar levels of patients with diabetes based on body weight. The first pattern occurs when blood sugar increases as low body weight up to 58.5 kg. That means there is a tendency for blood sugar patterns to rise sharply with increasing body weight to 58.5 kg. This can occur because the body weight is still in the normal category, so that someone is not strictly in the diabetic diet program. That causes blood sugar to rise rapidly. The different conditions we see in the second segment are body weight above 58.5 kg. It looks blood sugar decreases with increasing body weight. In this condition shows that weight does not have a positive effect on increasing blood sugar, it makes the patient aware of going on a diet so that his blood sugar drops.

The results of this article show that it is not always that if you gain weight, your blood sugar will also increase. Through penalized spline, another tendency is shown that excess weight actually reduces blood sugar in people with diabetes. That happened because patients were careful in consuming their daily food.

## 4 References

[1] Budiantara, I.N., Ratna, M., Zain, I. and Wibowo, W. (2012). Modeling the percentage of poor people in Indonesia using spline nonparametric regression approach. International Journal of Basic & Applied Sciences, 12 : 119-124.

[2] Chamidah, N., Budiantara, I.N., Sunaryo, S. and Ismaini, Z. (2012). Designing of child growth chart based on multi response local polynomial modeling. Journal of Mathematics and Statistics. 8, 342-347.

[3] Chamidah, N. and Saifudin, T (2013). Estimation of children growth based on kernel smoothing in multi response nonparametric regression. Applied Mathematical Sciences, 7 (37).

[4] Aydin, D. and Yilmaz, E. (2018). Modified spline regression based on randomly right-censored data: A comparative study. Communications in Statistics-Simulation and Computation. 47 : Issue 9.

[5] Lestari, B., Fatmawati and Budiantara, I.N. (2019). Smoothing spline estimator in multiresponse nonparametric regression for predicting blood pressure and heart rate. International Journal of Academic and Applied Research (IJAAR), 3 (9) : 1-8.

[6] Islamiyati, A. (2019). Regresi spline polynomial truncated biprediktor untuk identifikasi perubahan jumlah trombosit pasien demam berdarah dengue. Al khwarizmi, 7 (2) : 97-110.

[7] Islamiyati, A. Spline polynomial truncated dalam regresi nonparametrik. Jurnal Matematika, Statistika Dan Komputasi, 14 (1) : 54–60.

[8] Montoya, L.E., Ulloa, N. and Miller, V. (2014). A simulation study comparing knot selection methods with equally spaced knot in a penalized regression spline. International Journal of Statistic and Probability, 3 (3), 96-110.

[9] Islamiyati, A., Fatmawati and Chamidah, N. (2017). Fungsi goodness of fit dalam kriteria penalized spline pada estimasi regresi nonparametrik birespon untuk data longitudinal. Proseding Seminar Nasional Matematika dan Aplikasinya. UNAIR Surabaya.

[10] Islamiyati, A., Fatmawati and Chamidah, N. (2019) Ability of covariance matrix in bi-response multi-prredictor penalized spline model through longitudinal data simulation. International Journal of Academic and Applied Research (IJAAR), 3(3) : 8-11.

[11] Islamiyati, A., Fatmawati and Chamidah, N. (2018). Estimation of covariance matrix on bi-response longitudinal data analysis with penalized spline regression. J. Phys.: Conf. Ser. **979** 012093.

[12] Islamiyati, A., Fatmawati and Chamidah, I.N. (2019). Penalized spline estimator with multi smoothing parameters in biresponse multipredictor regression model for longitudinal data. Songklanakarin Journal of Science and Technology, In Press SJST-2018-0423.R2.

[13] Islamiyati, A., Fatmawati and Chamidah, I.N. (2019). Changes in Blood Glucosa 2 Hours After Meals in Type 2 Diabetes Patients Based On length of treatment at Hasanuddin University Hospital, Indonesia. Rawal Medical Journal, Accepted.

[14] Ruppert, D. (2002). Selecting the number of knot for penalized spline. Journal of Computational and Graphical Statistics, 11 (4), 735-757.