International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol.3 Issue 9, September – 2019, Pages: 1-8

Smoothing Spline Estimator in Multiresponse Nonparametric Regression for Predicting Blood Pressures and Heart Rate

Budi Lestari
Department of Mathematics, Faculty of Mathematics and Natural Sciences, The University of Jember
Jalan Kalimantan 37 Kampus Tegal Boto, Jember 68121 Indonesia
Doctoral Study Program, Faculty of Science and Technology, Airlangga University
Jalan Mulyorejo Kampus C UNAIR Surabaya 60115 Indonesia
lestari.statistician@gmail.com and budi.lestari-2016@fst.unair.ac.id

Fatmawati
Department of Mathematics, Faculty of Science and Technology, Airlangga University
Jalan Mulyorejo Kampus C UNAIR Surabaya 60115 Indonesia
fatma47unair@gmail.com and fatmawati@fst.unair.ac.id

I Nyoman Budiantara
Department of Mathematics, Faculty of Mathematics, Computing and Data Science, Sepuluh Nopember Institute of Technology
Jalan Arif Rahman Hakim, Kampus Sukolilo, Keputih, Surabaya 60111 Indonesia
nyomanbudiantara65@gmail.com

Abstract: *The basic idea of nonparametric regression is to let the data decide which regression function fits the best without imposing any specific form on it. Consequently, nonparametric methods are in general more flexible. They can uncover structure in the data that might otherwise be missed. In the real cases, we are frequently faced the problem in which two or more response variables are observed at several values of the predictor variables, and there are correlations among responses. For example, blood pressures and pulse are observed at several values of body mass index. Multiresponse nonparametric regression model provides powerful tools for modeling the functions which represent association of these variables. Estimating of regression function is the main problem in this model. Smoothing spline estimator has powerful and flexible properties for estimating the regression function. In this paper we discuss theoretically a method to estimate regression function and optimal smoothing parameter of blood pressures and heart rate models based on smoothing spline estimator in multiresponse nonparametric regression. The estimated regression function can be obtained by taking solution of penalized weighted least square optimization by using reproducing kernel Hilbert space approach. Next, we can get the optimal smoothing parameter by minimizing generalized cross validation function. In this research we obtained plots of predicted values of systolic and diastolic blood pressures and heart rate. The results show that patients who have high BMI, their systolic and diastolic blood pressures, and heart rate lead to hight. It means that patients who have overweight and obese categories lead to rentant suffering hypertension.*

Keywords — Blood Pressures, Multiresponse Nonparametric Regression, Heart Rate, Smoothing Parameter, Smoothing Spline Estimator.

1. INTRODUCTION

Statistical analysis often involves building mathematical models which examine the relationship between response and predictor variables. Spline smoothing is a general class of powerful and flexible modeling techniques. Research on smoothing spline models has attracted a great deal of attention in recent years, and the methodology has been widely used in many areas. Smoothing spline estimator with its powerful and flexible properties is one of the most popular estimators used for estimating regression curve of the nonparametric regression model. There are many researchers who have considered spline estimator for estimating regression function of the nonparametric regression model. Researchers [1-3] used original spline estimator to estimate regression curve of smooth data. M-type spline to overcome outliers in nonparametric regression has been proposed by [4-5]. Confidence interval for original spline model by using Bayesian approach has been constructed by [6]. Next, [7] compared between generalized cross validation (GCV) and generalized maximum likelihood (GML) for choosing the smoothing parameter in the generalized spline smoothing problem. Relaxed spline and quantile spline were introduced by [8-9]. Smoothing spline models with correlated random errors has been discussed by [10]. Some techniques for spline statistical model building by using reproducing kernel Hilbert spaces were introduced by [11]. A method that combines smoothing spline estimates of different smoothness to form a final improved estimate was proposed by [12]. Asymptotic property of smoothing splines estimators in functional linear regression with errors-in-variables has been studied by [13].

Next, [14] have studied smoothing spline estimation of variance functions. Further, [15] showed goodness of spline estimator rather than kernel estimator in estimating nonparametric regression model for gross national product data. Also, [16] have studied the determination of an optimum smoothing parameter for nonparametric regression using smoothing spline. All these researchers studied spline estimators in case of single response nonparametric regression models only.

In the real cases, we are frequently faced the problem in which two or more dependent variables are observed at several values of the independent variables, and there are correlations between responses. Multiresponse nonparametric regression model provides powerful tools to model the functions which represent association of these variables. There are many researchers who have considered nonparametric models for multiresponse data. Spline smoothing for estimating nonparametric functions from bivariate data with the same correlation of errors has been studied by [17]. Methods for estimating nonparametric regression model with spatially correlated errors were proposed by [18]. Next, [19] and [20] have studied spline estimators in multi-response nonparametric regression model with equal correlation of errors and unequal correlation of errors, respectively. Multiresponse nonparametric regression model approach to design child growth chart has been used by [21]. A mathematical statistics method for estimating regression curve of the multiresponse nonparametric regression model in case of heteroscedasticity of variance was proposed by [22]. Estimating regression function of the homoscedastic multiresponse nonparametric regression in which the number of observations were unbalance has been discussed by [23]. Next, [24-26] proposed smoothing spline estimator for estimating of the multiresponse nonparametric regression model by using reproducing kernel Hilbert space (RKHS). In addition, [27] discussed construction of covariance matrix in case of homoscedasticity of variances of errors. Also, [28] discussed estimating of both covariance matrix and optimal smoothing parameter. Further, [29] used spline to show ability of covariance matrix. But, these researchers have not discussed estimating of smoothing parameter in multi-response nonparametric regression model when the variances of errors are not the same for cross-section data. In addition, all these researchers have not discussed application of the estimated model on the real case data.

According to [30], risk of health has correlation with increasing of body mass index (BMI), and BMI more than or equals to 23 kg/m2 was categorized as overweight or obesity. Since level of overweight can be measured by BMI (body mass index), then increasing of BMI can also cause the increasing of systolic and diastolic blood pressures. Also, [31] have shown that the increasing of BMI of someone who were least than 60 years old caused the increasing of systolic and diastolic blood pressures. Next, [32] pointed out that increasing and decreasing of systole and diastole blood pressures were significantly caused by increasing and decreasing of BMI for all sex and all ages. Researchers [33] stated that BMI significantly influenced to systolic and diastolic blood pressures of Ethiopian, Vietnamese, and Indonesian. Also, [34] pointed out that BMI affects systolic and diastolic blood pressures of females and males. In addition, [35] have shown that there was positively correlation between BMI and both systole and diastole blood pressures of children 8-16 years old. Further, [36] pointed out that BMI and overweight or obesity can cause the increasing of blood pressures (systolic and diastolic). Therefore, in this paper, we discuss methods to estimate regression function and optimum smoothing parameter of the multiresponse nonparametric regression model if it is applied to blood pressures and heart rate affected by BMI.

2. MATERIAL AND METHODS

Firstly, consider multiresponse nonparametric regression model as given by [22, 24-25, 28]:

$$y_{ki} = f_k(t_{ki}) + \varepsilon_{ki}, \quad k = 1,2,...,p, \quad i = 1,2,...,n_k \quad (1)$$

where $Var(\varepsilon_{ki}) = \sigma_{ki}^2$. Next, suppose that we apply the model in (1) to data of blood pressures and pulse that are affected by BMI such that we have the blood pressures and pulse model as follows:

$$y_{ki} = f_k(t_{ki}) + \varepsilon_{ki}; \; k = 1,2,3 \; ; \; i = 1,2,...,n_k \quad (2)$$

where $Var(\varepsilon_{ki}) = \sigma_{ki}^2$, y_{1i}, y_{2i} and y_{3i} are response variables that represent the first response (i.e., systolic blood pressure), the second response (i.e., diastolic blood pressure), and the third response (i.e., heart rate), respectively; and $f_k(t_{ki})$ are unknown regression functions which represent function of predictor variable (i.e., BMI).

The estimated regression function can be obtained by taking solution of penalized weighted least square optimization by using reproducing kernel Hilbert space approach. Next, we can get the optimal smoothing parameter by minimizing generalized cross validation function. Finally, we apply the estimated model that we have obtained to the real case data, i.e., blood pressures and heart rate affected by BMI. In this case, we use the estimated estimate systolic and diastolic blood pressures, and heart rate.

3. RESULTS AND DISCUSSION

In this section, we give results and discussion about estimating regression function, estimating optimal smoothing parameter, and predicting blood pressures and heart rate by using smoothing spline estimator.

3.1 Estimation of Regression Function

Firstly, we consider a paired data set that follows the blood pressures and heart rate model as given in (2), i.e.,:

$$y_{ki} = f_k(t_{ki}) + \varepsilon_{ki} \; ; \; i = 1,2,...,n_k \; ; \; a_k \le t_k \le b_k \; ; \quad (3)$$

where k represents the number of response, $n_k = n$ for $k = 1,2,3$ and f_1, f_2, f_3 are unknown regression functions

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol.3 Issue 7, July – 2019, Pages: 1-1

assumed to be smooth in Sobolev space $W_2^m[a_k,b_k]$, and ε_{ki} are zero-mean independent random errors with variance σ_{ki}^2. The main objective of nonparametric regression analysis is estimate unknown regression functions $f_k \in W_2^m[a_k,b_k]$ in model (3). Next, suppose that $\underset{\sim}{y}=(\underset{\sim}{y}_1,\underset{\sim}{y}_2,\underset{\sim}{y}_3)'$, $\underset{\sim}{f}=(\underset{\sim}{f}_1,\underset{\sim}{f}_2,\underset{\sim}{f}_3)'$, $\underset{\sim}{\varepsilon}=(\underset{\sim}{\varepsilon}_1,\underset{\sim}{\varepsilon}_2,\underset{\sim}{\varepsilon}_3)'$, and $\underset{\sim}{t}=(\underset{\sim}{t}_1,\underset{\sim}{t}_2,\underset{\sim}{t}_3)'$ where $\underset{\sim}{y}_k=(y_{k1},...,y_{kn})'$, $\underset{\sim}{f}_k=(f_k(t_{k1}),...,f_k(t_{kn}))'$, $\underset{\sim}{\varepsilon}_k=(\varepsilon_{k1},\varepsilon_{k2},...,\varepsilon_{kn})'$, $\underset{\sim}{t}_k=(t_{k1},t_{k2},...,t_{kn})'$. Therefore, for $i=1,2,...,n$ and $k=1,2,3$ we can write equation (3) in the following equation:

$$\underset{\sim}{y}=\underset{\sim}{f}+\underset{\sim}{\varepsilon} \qquad (4)$$

where $E(\underset{\sim}{\varepsilon})=\underset{\sim}{0}$, and $Cov(\underset{\sim}{\varepsilon})=[W(\underset{\sim}{\sigma}^2)]^{-1}=diag(W_1(\underset{\sim}{\sigma}_1^2),W_2(\underset{\sim}{\sigma}_2^2),W_3(\underset{\sim}{\sigma}_3^2))$. Estimating of the functions $\underset{\sim}{f}$ in (4) by using smoothing spline estimator appears as a solution to the penalized weighted least-square (PWLS) minimization problem, i.e., determine $\hat{\underset{\sim}{f}}$ that can make the following PWLS minimum:

$$\underset{f_1,f_2,f_3\in W_2^m}{Min}\{(\sum_{k=1}^{3}n_k)^{-1}(\underset{\sim}{y}_1-\underset{\sim}{f}_1)'W_1(\underset{\sim}{y}_1-\underset{\sim}{f}_1)+...+(\underset{\sim}{y}_3-\underset{\sim}{f}_3)'W_3(\underset{\sim}{y}_3-\underset{\sim}{f}_3)+$$
$$\lambda_1\int_{a_1}^{b_1}(f_1^{(2)}(t))^2dt+...+\lambda_3\int_{a_3}^{b_3}(f_3^{(2)}(t))^2dt\} \qquad (5)$$

for pre-specified value $\underset{\sim}{\lambda}=(\lambda_1,\lambda_2,\lambda_3)'$. Note that, in equation (5), the first term represents the sum squares of errors and it penalizes the lack of fit. While, the second term which is weighted by $\underset{\sim}{\lambda}$ represents the roughness penalty and it imposes a penalty on roughness. It means that the curvature of $\underset{\sim}{f}$ is penalized by it. In equation (5), λ_k $(k=1,2,3)$ is called as the smoothing parameter. The solution will be vary from interpolation to a linear model, if λ_k varies from 0 to $+\infty$. So that, if $\lambda_k\to+\infty$, the roughness penalty will dominate in (5), and the smoothing spline estimate will be forced to be a constant. If $\lambda_k\to 0$, the roughness penalty will disappear in (5), and the spline estimate will interpolate the data. Thus, the trade-off between the goodness of fit given by:

$$(\sum_{k=1}^{3}n_k)^{-1}(\underset{\sim}{y}_1-\underset{\sim}{f}_1)'W_1(\underset{\sim}{y}_1-\underset{\sim}{f}_1)+...+(\underset{\sim}{y}_3-\underset{\sim}{f}_3)'W_3(\underset{\sim}{y}_3-\underset{\sim}{f}_3)$$

and smoothness of the estimate given by:

$$\lambda_1\int_{a_1}^{b_1}(f_1^{(2)}(t))^2dt+...+\lambda_p\int_{a_k}^{b_k}(f_p^{(2)}(t))^2dt$$

is controlled by the smoothing parameter λ_k. The solution for minimization problem in (5) is a smoothing spline estimator where its function basis is a "natural cubic spline" with $t_1,t_2,...,t_{n_k}$ $(k=1,2,...,p)$ as its knots. Based on this concept, a particular structured spline interpolation that depends on selection of the smoothing parameter λ_k value becomes a appropriate approach of the functions f_k $(k=1,2,3)$ in model (1). Let $\underset{\sim}{f}=(\underset{\sim}{f}_1,\underset{\sim}{f}_2,\underset{\sim}{f}_3)'$ where $\underset{\sim}{f}_k=(f_k(t_{k1}),f_k(t_{k2}),...,f_k(t_{kn}))'$, $k=1,2,3$, be the vector of values of function f_k $(k=1,2,3)$ at the knot points $t_1,t_2,...,t_{n_k}$ $(k=1,2,3)$. If we express the model of paired data set into a general smoothing spline regression model, we will get the following expression:

$$y_{ki}=L_{t_k}f_k+\varepsilon_{ki},\ i=1,2,...,n_k;\ k=1,2,3 \qquad (6)$$

where $f_k\in H_k$ (H_k represents Hilbert space) is an unknown smooth function, and $L_{t_k}\in H_k$ is a bounded linear functional.

Suppose H_k can be decomposed into two subspaces U_k and W_k as follows:

$$H_k = U_k \oplus W_k$$

where U_k is orthogonal to W_k, $k=1,2,3$. Suppose that $\{u_{k1},u_{k2},...,u_{km_k}\}$ and $\{\omega_{k1},\omega_{k2},...,\omega_{kn_k}\}$ are bases of spaces U_k and W_k, respectively. Then, we can express every function $f_k\in H_k$ ($k=1,2,3$) into the following expression:

$$f_k=g_k+h_k$$

where $g_k\in U_k$ and $h_k\in W_k$. Since $\{u_{k1},u_{k2},...,u_{km_k}\}$ is basis of space U_k and $\{\omega_{k1},\omega_{k2},...,\omega_{kn_k}\}$ is basis of space W_k, then for every $f_k\in H_k$ ($k=1,2,3$) follows:

$$f_k=\sum_{j=1}^{m_k}d_{kj}u_{kj}+\sum_{i=1}^{n_k}c_{ki}\omega_{ki}=\underset{\sim}{u}_k'\underset{\sim}{d}_k+\underset{\sim}{\omega}_k'\underset{\sim}{c}_k; \qquad (7)$$

where $k=1,2,...,p$; $d_{kj}\in\mathscr{R}$; $c_{ki}\in\mathscr{R}$; $\underset{\sim}{u}_k=(u_{k1},u_{k2},...,u_{km_k})'$, $\underset{\sim}{d}_k=(d_{k1},d_{k2},...,d_{km_k})'$, $\underset{\sim}{\omega}_k=(\omega_{k1},\omega_{k2},...,\omega_{kn_k})'$, and $\underset{\sim}{c}_k=(c_{k1},c_{k2},...,c_{kn_k})'$. Furthermore, since $L_{t_{ki}}$ is a function which is bounded and linear in H_k, and $f_k\in H_k$, $k=1,2,3$ then we have

$$L_{t_{ki}}f_k=L_{t_{ki}}(g_k+h_k)=g_k(t_{ki})+h_k(t_{ki})=f_k(t_{ki}). \qquad (8)$$

Based on model (3), and by applying the Riesz representation theorem [3], and because of $L_{t_{ki}}\in H_k$ is bounded linear functional, then according to Wang (2011) there is a representer $\xi_{ki}\in H_k$ of $L_{t_{ki}}$ which follows:

$$L_{t_{ki}}f_k=\langle\xi_{ki},f_k\rangle=f_k(t_{ki}),\ f_k\in H_k \qquad (9)$$

where $\langle \cdot, \cdot \rangle$ denotes an inner product. Based on (6) and by applying the properties of the inner product, we get:

$$f_k(t_{ki}) = \langle \xi_{ki}, \underset{\sim}{u}'_k \underset{\sim}{d}_k + \underset{\sim}{\omega}'_k \underset{\sim}{c}_k \rangle$$
$$= \langle \xi_{ki}, \underset{\sim}{u}'_k \underset{\sim}{d}_k \rangle + \langle \xi_{ki}, \underset{\sim}{\omega}'_k \underset{\sim}{c}_k \rangle . \quad (10)$$

Next, by applying equation (10), for $k = 1$ we have:

$f_1(t_{1i}) = \langle \xi_{1i}, \underset{\sim}{u}'_1 \underset{\sim}{d}_1 \rangle + \langle \xi_{1i}, \underset{\sim}{\omega}'_1 \underset{\sim}{c}_1 \rangle$, $i = 1, 2, ..., n_1$;

and for $i = 1, 2, 3, ..., n_1$ we have:

$$\underset{\sim}{f}_1(t_1) = (f_1(t_{11}), f_1(t_{12}), ..., f_1(t_{1n_1}))'$$
$$= K_1 \underset{\sim}{d}_1 + \Sigma_1 \underset{\sim}{c}_1 , \quad (11)$$

where:

$$K_1 = \begin{bmatrix} \langle \xi_{11}, u_{11} \rangle & \langle \xi_{11}, u_{12} \rangle & \cdots & \langle \xi_{11}, u_{1m_1} \rangle \\ \langle \xi_{12}, u_{11} \rangle & \langle \xi_{12}, u_{12} \rangle & \cdots & \langle \xi_{12}, u_{1m_1} \rangle \\ \vdots & \vdots & \vdots & \vdots \\ \langle \xi_{1n_1}, u_{11} \rangle & \langle \xi_{1n_1}, u_{12} \rangle & \cdots & \langle \xi_{1n_1}, u_{1m_1} \rangle \end{bmatrix},$$

$$\Sigma_1 = \begin{bmatrix} \langle \xi_{11}, \omega_{11} \rangle & \langle \xi_{11}, \omega_{12} \rangle & \cdots & \langle \xi_{11}, \omega_{1n_1} \rangle \\ \langle \xi_{12}, \omega_{11} \rangle & \langle \xi_{12}, \omega_{12} \rangle & \cdots & \langle \xi_{12}, \omega_{1n_1} \rangle \\ \vdots & \vdots & \vdots & \vdots \\ \langle \xi_{1n_1}, \omega_{11} \rangle & \langle \xi_{1n_1}, \omega_{12} \rangle & \cdots & \langle \xi_{1n_1}, \omega_{1n_1} \rangle \end{bmatrix},$$

$\underset{\sim}{d}_1 = (d_{11}, d_{12}, ..., d_{1m_1})'$, and $\underset{\sim}{c}_1 = (c_{11}, c_{12}, ..., c_{1n_1})'$. In the similar process, we obtain: $\underset{\sim}{f}_2(t_2) = K_2 \underset{\sim}{d}_2 + \Sigma_2 \underset{\sim}{c}_2$; $\underset{\sim}{f}_3(t_3) = K_3 \underset{\sim}{d}_3 + \Sigma_3 \underset{\sim}{c}_3$. Therefore, the regression function $\underset{\sim}{f}(t)$ can be expressed as:

$$\underset{\sim}{f}(t) = (\underset{\sim}{f}_1(t_1), \underset{\sim}{f}_2(t), \underset{\sim}{f}_3(t))'$$
$$= (K_1 \underset{\sim}{d}_1, K_2 \underset{\sim}{d}_2, K_3 \underset{\sim}{d}_3)' + (\Sigma_1 \underset{\sim}{c}_1, \Sigma_2 \underset{\sim}{c}_2, \Sigma_3 \underset{\sim}{c}_3)'$$
$$= diag(K_1, K_2, K_3)(\underset{\sim}{d}_1, \underset{\sim}{d}_2, \underset{\sim}{d}_3)' +$$
$$diag(\Sigma_1, \Sigma_2, \Sigma_3)(\underset{\sim}{c}_1, \underset{\sim}{c}_2, \underset{\sim}{c}_3)'$$
$$= K\underset{\sim}{d} + \Sigma \underset{\sim}{c} . \quad (12)$$

In equation (12), K is a $(N \times M)$-matrix and $\underset{\sim}{d}$ is a vector of parameters with dimension $(M \times 1)$ (where $N = \sum_{k=1}^{3} n_k = 3n$, $M = \sum_{k=1}^{3} m_k = 3m$) that are expressed as: $K = diag(K_1, K_2, K_3)$, and $\underset{\sim}{d} = (\underset{\sim}{d}'_1, \underset{\sim}{d}'_2, \underset{\sim}{d}'_3)'$, respectively. Also, Σ is a $(N \times N)$-matrix, and $\underset{\sim}{c}$ is a $(N \times 1)$-vector of parameters which are expressed as:

$\Sigma = diag(\Sigma_1, \Sigma_2, \Sigma_3)$, and $\underset{\sim}{c} = (\underset{\sim}{c}'_1, \underset{\sim}{c}'_2, \underset{\sim}{c}'_3)'$, respectively.

Therefore, we can write model in (3) as follows:

$$\underset{\sim}{y} = K\underset{\sim}{d} + \Sigma \underset{\sim}{c} + \underset{\sim}{\varepsilon} .$$

We use the RKHS method to obtain the estimation of $\underset{\sim}{f}$, by solving the following optimization:

$$\underset{\substack{f_k \in \mathcal{H}_k \\ k=1,2,3}}{Min} \left\{ \left\| W^{\frac{1}{2}}(\underset{\sim}{\sigma}^2) \underset{\sim}{\varepsilon} \right\|^2 \right\} = \underset{\substack{f_k \in \mathcal{H}_k \\ k=1,2,3}}{Min} \left\{ \left\| W^{\frac{1}{2}}(\underset{\sim}{\sigma}^2)(\underset{\sim}{y} - \underset{\sim}{f}) \right\|^2 \right\}, \quad (13)$$

with constraint:

$$\int_{a_k}^{b_k} [f_k^{(m)}(t_k)]^2 dt_k < \gamma_k \ , \ \gamma_k \geq 0 . \quad (14)$$

To solve the optimization (13) with constraint (14) is equivalent to solve the optimization PWLS:

$$\underset{\substack{f_k \in W_2^m[a_k, b_k] \\ k=1,2,3}}{Min} \left\{ N^{-1}(\underset{\sim}{y} - \underset{\sim}{f})' W(\underset{\sim}{\sigma}^2)(\underset{\sim}{y} - \underset{\sim}{f}) + \sum_{k=1}^{3} \lambda_k \int_{a_k}^{b_k} [f_k^{(m)}(t_k)]^2 dt_k \right\}, \quad (15)$$

where λ_k, $k = 1, 2, 3$ are smoothing parameters that control trade-off between goodness of fit represented by: $N^{-1}(\underset{\sim}{y} - \underset{\sim}{f})' W(\underset{\sim}{\sigma}^2)(\underset{\sim}{y} - \underset{\sim}{f})$

and the roughness penalty measured by:

$$\lambda_1 \int_{a_1}^{b_1} [f_1^{(m)}(t_1)]^2 dt_1 + ... + \lambda_3 \int_{a_3}^{b_3} [f_3^{(m)}(t_3)]^2 dt_3 .$$

To get the solution to (15), we first decompose the roughness penalty as follows:

$$\int_{a_1}^{b_1} [f_1^{(m)}(t_1)]^2 dt_1 = \left\| P f_1 \right\|^2 = \langle P f_1, P f_1 \rangle$$
$$= \langle \underset{\sim}{\omega}'_1 \underset{\sim}{c}_1, \underset{\sim}{\omega}'_1 \underset{\sim}{c}_1 \rangle = \underset{\sim}{c}'_1 (\underset{\sim}{\omega}_1 \underset{\sim}{\omega}'_1) \underset{\sim}{c}_1 = \underset{\sim}{c}'_1 \Sigma_1 \underset{\sim}{c}_1$$

It implies:

$$\lambda_1 \int_{a_1}^{b_1} [f_1^{(m)}(t_1)]^2 dt_1 = \lambda_1 \underset{\sim}{c}'_1 \Sigma_1 \underset{\sim}{c}_1 . \quad (16)$$

Next, by similar way, we get:

$$\lambda_2 \int_{a_2}^{b_2} [f_2^{(m)}(t_2)]^2 dt_2 = \lambda_2 \underset{\sim}{c}'_2 \Sigma_2 \underset{\sim}{c}_2 , \ \dots ,$$

$$\lambda_3 \int_{a_3}^{b_3} [f_3^{(m)}(t_3)]^2 dt_3 = \lambda_3 \underset{\sim}{c}'_3 \Sigma_3 \underset{\sim}{c}_3 . \quad (17)$$

Based on (16) and (17), we have penalty:

$$\sum_{k=1}^{3} \lambda_k \int_{a_k}^{b_k} [f_k^{(m)}(t_k)]^2 dt_k \} = \underset{\sim}{c}' \lambda \Sigma \underset{\sim}{c} \quad (18)$$

where $\lambda = diag(\lambda_1 I_{n_1}, \lambda_2 I_{n_2}, \lambda_3 I_{n_3})$. We can express the goodness of fit in (15) as follows:

$$N^{-1}(\underset{\sim}{y} - \underset{\sim}{f})' W(\underset{\sim}{\sigma}^2)(\underset{\sim}{y} - \underset{\sim}{f}) =$$
$$N^{-1}(\underset{\sim}{y} - K\underset{\sim}{d} - \Sigma\underset{\sim}{c})' W(\underset{\sim}{\sigma}^2)(\underset{\sim}{y} - K\underset{\sim}{d} - \Sigma\underset{\sim}{c}) .$$

If we combine the goodness of fit and the roughness penalty, we will have optimization PWLS:

$$\underset{\substack{\underset{\sim}{c} \in R^{3n} \\ \underset{\sim}{d} \in R^{3m}}}{Min} \left\{ (\underset{\sim}{y} - K\underset{\sim}{d} - \Sigma\underset{\sim}{c})' W(\underset{\sim}{\sigma}^2)(\underset{\sim}{y} - K\underset{\sim}{d} - \Sigma\underset{\sim}{c}) + \underset{\sim}{c}' N \lambda \Sigma \underset{\sim}{c} \right\}$$
$$= \underset{\substack{\underset{\sim}{c} \in R^{3n} \\ \underset{\sim}{d} \in R^{3m}}}{Min} \left\{ Q(\underset{\sim}{c}, \underset{\sim}{d}) \right\} . \quad (19)$$

To get the solution to (19), firstly we must take the partially differential of $Q(\underset{\sim}{c},\underset{\sim}{d})$ and then their results are equaled to zeros as follows:

$$\frac{\partial Q(\underset{\sim}{c},\underset{\sim}{d})}{\partial \underset{\sim}{c}}=\underset{\sim}{0} \Leftrightarrow \hat{\underset{\sim}{c}}=M^{-1}W(\underset{\sim}{\sigma}^2)(\underset{\sim}{y}-K\underset{\sim}{d}). \quad (20)$$

$$\frac{\partial Q(\underset{\sim}{c},\underset{\sim}{d})}{\partial \underset{\sim}{d}}=\underset{\sim}{0} \Leftrightarrow$$

$$\hat{\underset{\sim}{d}}=[K'M^{-1}W(\underset{\sim}{\sigma}^2)K]^{-1}K'M^{-1}W(\underset{\sim}{\sigma}^2)\underset{\sim}{y}. \quad (21)$$

Next, if we substitute (21) into (20) , we obtain:

$$\hat{\underset{\sim}{c}}=M^{-1}W(\underset{\sim}{\sigma}^2)[I-K(K'M^{-1}W(\underset{\sim}{\sigma}^2)K)^{-1}K'M^{-1}W(\underset{\sim}{\sigma}^2)]\underset{\sim}{y}. \quad (22)$$

Finally, based on (12), (21) and (22), we get the smoothing spline estimator which can be expressed as follows:

$$\hat{\underset{\sim}{f}}_{\lambda}(\underset{\sim}{t})=\begin{pmatrix}\hat{\underset{\sim}{f}}_{1,\lambda_1}(\underset{\sim}{t}_1)\\ \hat{\underset{\sim}{f}}_{2,\lambda_2}(\underset{\sim}{t}_2)\\ \vdots \\ \hat{\underset{\sim}{f}}_{p,\lambda_p}(\underset{\sim}{t}_p)\end{pmatrix}=K\hat{\underset{\sim}{d}}+\Sigma\hat{\underset{\sim}{c}}=H(\underset{\sim}{\lambda})\underset{\sim}{y} \quad (23)$$

where

$$H(\underset{\sim}{\lambda})=K[K'M^{-1}W(\underset{\sim}{\sigma}^2)K]^{-1}K'M^{-1}W(\underset{\sim}{\sigma}^2)+\Sigma M^{-1}W(\underset{\sim}{\sigma}^2)$$
$$[I-K(K'M^{-1}W(\underset{\sim}{\sigma}^2)K)^{-1}K'M^{-1}W(\underset{\sim}{\sigma}^2)],$$

and $\hat{\underset{\sim}{f}}_{\lambda}(\underset{\sim}{t})$ is smoothing spline with a natural cubic spline as a basis function with knots at $t_1,t_2,...,t_{n_k}$ $(k=1,2,3)$, for a fixed smoothing parameter $\underset{\sim}{\lambda}>\underset{\sim}{0}$. $H(\underset{\sim}{\lambda})$ is a positive-definite (symmetrical) smoother matrix that depends on smoothing parameter $\underset{\sim}{\lambda}$ and the knot points $t_1,t_2,...,t_{n_k}$ $(k=1,2,3)$. Yet, it does not depend on $\underset{\sim}{y}$.

Based on estimated model we have in (23), we conclude that the estimated model is a linear function in observation. In addition, by taking expectation of equation (23), i.e., $E(\hat{\underset{\sim}{f}}_{\lambda}(t))$ we obtain that the estimated regression function in (23) is a biased estimator Further discussion about this estimator can be obtained on [3], [37-41].

3.2. Estimation of Optimal Smoothing Parameter

Researcher [3] has shown that in uniresponse spline nonparametric regression, if smoothing parameter (λ) value is very small $(\lambda \to 0)$ then it will give a very rought estimator of nonparametric regression function. In contrary, if smoothing parameter (λ) value is very large $(\lambda \to \infty)$ then it will give a very smooth estimator of nonparametric regression function. Therefore, we need to select optimum smoothing parameter (λ) in order to obtain estimator that is suitable with data. For this need, some researchers have proposed some selection methods, for instance [2] proposed cross validation (CV) method, [10] proposed unbiased risk (UBR) method, and [3] proposed generalized cross validation (GCV) method. Not only does uniresponse spline nonparametric regression, but also multiresponse spline nonparametric regression depends on smoothing parameter λ_k, $k=1,2,3$.

In this section we discuss selection method for selecting the optimum smoothing parameter in multiresponse nonparametric regression model for data of blood pressures and pulse. Regression function estimator of multiresponse nonparametric regression model for data of blood pressures and pulse as given in equation (23) can be expressed as follows:

$$\hat{\underset{\sim}{f}}_{\lambda}(t)=H(\lambda_1,\lambda_2,\lambda_3;\underset{\sim}{\sigma}^2)\underset{\sim}{y} \quad (24)$$

where $\underset{\sim}{\sigma}^2=(\sigma_1^2,\sigma_2^2,\sigma_3^2)'$. MSE (Mean Square Error) of (24) can be determined as follows:

$$MSE(\lambda_1,\lambda_2,\lambda_3;\underset{\sim}{\sigma}^2)=\frac{(\underset{\sim}{y}-\hat{\underset{\sim}{f}}_{\lambda}(t))'W(\underset{\sim}{\sigma}^2)(\underset{\sim}{y}-\hat{\underset{\sim}{f}}_{\lambda}(t))}{\sum_{k=1}^{3}n_k}$$
$$=\frac{\left(\underset{\sim}{y}-H(\lambda_1,\lambda_2,\lambda_3;\underset{\sim}{\sigma}^2)\underset{\sim}{y}\right)'W(\underset{\sim}{\sigma}^2)\left(\underset{\sim}{y}-H(\lambda_1,\lambda_2,\lambda_3;\underset{\sim}{\sigma}^2)\underset{\sim}{y}\right)}{\sum_{k=1}^{3}n_k}$$
$$=\frac{\left[\left(I_N-H(\lambda_1,\lambda_2,\lambda_3;\underset{\sim}{\sigma}^2)\right)\underset{\sim}{y}\right]'W(\underset{\sim}{\sigma}^2)\left[\left(I_N-H(\lambda_1,\lambda_2,\lambda_3;\underset{\sim}{\sigma}^2)\right)\underset{\sim}{y}\right]}{N}$$
$$=\frac{\left\|\left(W(\underset{\sim}{\sigma}^2)\right)^{\frac{1}{2}}\left(I_N-H(\lambda_1,\lambda_2,\lambda_3;\underset{\sim}{\sigma}^2)\right)\underset{\sim}{y}\right\|^2}{N}.$$

where $\left(W(\underset{\sim}{\sigma}^2)\right)^{\frac{1}{2}}$ is a diagonal matrix, and $N=\sum_{k=1}^{3}n_k$.

Next, we define a quantity (further it is called as GCV function) as follows:

$$G(\lambda_1,\lambda_2,\lambda_3;\underset{\sim}{\sigma}^2)=\frac{N^{-1}\left\|\left(W(\underset{\sim}{\sigma}^2)\right)^{\frac{1}{2}}\left(I_N-H(\lambda_1,\lambda_2,\lambda_3;\underset{\sim}{\sigma}^2)\right)\underset{\sim}{y}\right\|^2}{\left[N^{-1}trace\left(I_N-H(\lambda_1,\lambda_2,\lambda_3;\underset{\sim}{\sigma}^2)\right)\right]^2}$$

The optimum smoothing parameter $\underset{\sim}{\lambda}_{opt}$ is obtained by taking the solution of the following optimization:

$$G_{opt}(\lambda_{1(opt)},\lambda_{2(opt)},\lambda_{3(opt)};\underset{\sim}{\sigma}^2)$$
$$=\underset{\lambda_1\in R^+,\lambda_2\in R^+,\lambda_3\in R^+}{Min}\left\{\frac{N^{-1}\left\|\left(W(\underset{\sim}{\sigma}^2)\right)^{\frac{1}{2}}\left(I_N-H(\lambda_1,\lambda_2,\lambda_3;\underset{\sim}{\sigma}^2)\right)\underset{\sim}{y}\right\|^2}{\left[N^{-1}trace\left(I_N-H(\lambda_1,\lambda_2,\lambda_3;\underset{\sim}{\sigma}^2)\right)\right]^2}\right\},$$

where $\|\underset{\sim}{v}\| = \sqrt{v_1^2 + \ldots + v_p^2}$ for $\underset{\sim}{v} = (v_1, v_2, v_3)'$, and $\underset{\sim}{\lambda}_{opt} = (\lambda_{1(opt)}, \lambda_{2(opt)}, \lambda_{3(opt)})'$.

3.3. Estimation of Blood Pressures and Heart Rate

In this section, we give an example application of smoothing spline estimator on real case data set, i.e., association between blood pressures, heart rate, and body mass index (BMI). The data was recorded from a District General Hospital of Trenggalek city, East Java Province, Indonesia in 2018. The data draws the association between blood pressures (systolic and diastolic), heart rate and BMI of patients. Data consists of three response variables (in this case, $k = 1, 2, 3$) and a predictor variable. The first response variable (y_1) is systolic blood pressure, the second response variable (y_2) is diastolic blood pressure, and the third response variable (y_3) is heart rate. While the predictor variabel (t) is body mass index (BMI).

The estimation results give plot of estimated systolic blood pressure versus BMI, estimated diastolic blood pressure versus BMI, and estimated heart rate versus BMI as given in Figure 1, Figure 2, and Figure 3, respectively.

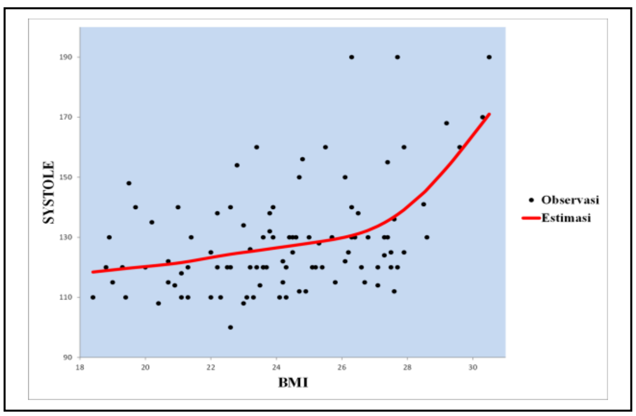

Figure 1. Plot of systolic blood pressure versus BMI.

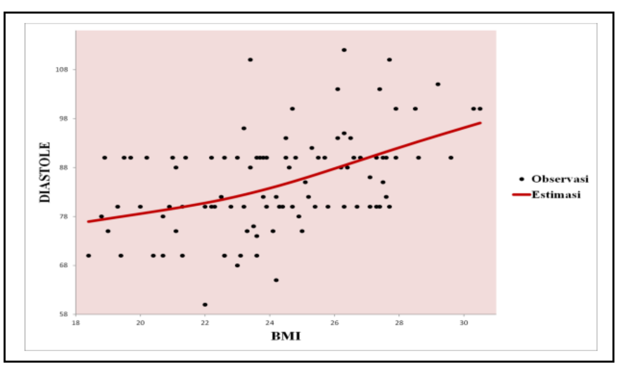

Figure 2. Plot of diastolic blood pressure versus BMI.

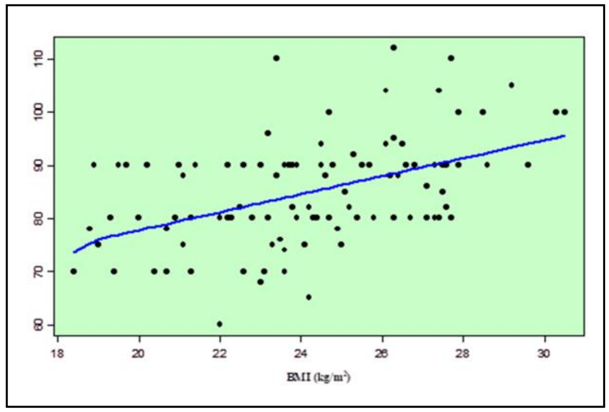

Figure 3. Plot of heart rate versus BMI.

According to [42], body weight clasification can be determined by BMI, i.e., underweight (BMI < 18.5), normal (18.5 ≤ BMI ≤ 22.9), and overweight (BMI ≥ 23) in which the body weight clasification is determined by BMI in certain interval. We can show patterns of underweight, normal, and overweight effects on systolic and diastolic blood pressures, and heart rate by using smoothing spline estimator that includes all observations as knots of spline.

Figure 1 shows that if patients have BMI least than 19 kg/cm^2 or have BMI from 19 kg/cm^2 to 28.5 kg/cm^2, their systolic blood pressures go up slowly along with increasing their BMI. Furthermore, if patients have BMI greater than 28.5 kg/cm^2, their systolic blood pressures go up sharply along with increasing their BMI. It means that if patients have BMI with underweight category or between normal and overweight category, according to [42], then their systolic blood pressures go up slowly along with increasing their BMI. This increasing systolic blood pressures will go up sharply if patients have BMI in obese category. Next, Figure 2 shows that the increasing BMI will raise diastolic blood pressure slowly not only for patient who has underweight and overweight categories but also for patient who has obese category along with increasing their BMI. Finaly, Figure 3 shows that heart rate of patients who have BMI less than 19 kg/m^2 have not shown sharply increasing, however, heart rate of patients who have BMI more than 19 kg/m^2 raise sharply along with increasing their BMI. In addition, patients who have high BMI, their systolic and diastolic blood pressures, and heart rate lead to hight. Furthermore, since BMI is determined by body weight devided by square of hight then patients who have overweight and obese categories lead to rentant suffering hypertension.

4. CONCLUSION

In estimating of the regression function of the multiresponse nonparametric regression model based on smoothing spline estimator, we use all observation points as knots. In this case, the results of analysis show that patients who have overweight and obese categories lead to rentant suffering hypertension.

5. ACKNOWLEDGMENT

Many thanks to Director of the Directorate of Research and Public Service, the Directorate General of Reinforcing of Research and Development, the Ministry of Research, Technology, and Higher Education of the Republic of Indonesia for funding this research via the Doctoral Dissertation Research Grant (*Hibah Penelitian Disertasi Doktor*) in the fiscal year 2019.

6. REFERENCE

[1] Kimeldorf, G., and Wahba, G. (1971). Some result on tchebycheffian spline functions. J. of Math. Anal. and Appl., Vol. 33, pp. 82-95.

[2] Craven, P., and Wahba, G. (1979). Smoothing noisy data with spline function: Estimating the correct degree of smoothing by the method of generalized cross validation. Numer. Math., Vol. 31, pp. 377-403.

[3] Wahba, G. (1990). Spline Models for Observational Data. SIAM Philadelphia. Pennsylvania.

[4] Cox, D. D. (1983). Asymptotic for M-type smoothing spline. Annals. Statistics. Vol. 11, pp. 530-551.

[5] Cox, D. D., and O'Sullivan, F. (1996). Penalized likelihood type estimators for generalized nonparametric regression. J. Mult. Anal., Vol. 56, pp. 185-206.

[6] Wahba, G. (1983). Bayesian confidence intervals for the cross-validated smoothing spline. J. Royal Stat. Society, Series B., Vol. 45, pp. 133-150.

[7] Wahba, G. (1985). A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem. Annals Stat., Vol. 13, pp. 1378-1402.

[8] Oehlert, G. W. (1992). Relaxed boundary smoothing spline. Annals. Statistics, Vol. 20, pp. 146-160.

[9] Koenker, R., Pin, N. G., and Portnoy, S. (1994). Quantile smoothing splines. Biometrics, Vol. 81, pp. 673-680.

[10]Wang, Y. (1998). Smoothing spline models with correlated random errors. J. Amer. Stat. Assoc., Vol. 93, pp. 341-348.

[11] Wahba, G. (2000). An introduction to model building with reproducing kernel hilbert spaces. Available at: http://www.stat.wisc.edu/~wahba/ftp1/interf/rootpar1r.pdf

[12] Lee, T., C., M. (2004). Improved smoothing spline regression by combines estimates of different smoothness. Stat. & Prob. Letters, Vol. 67, pp. 133-140.

[13] Cardot, Crambes, H.C., Kneip, A., and Sarda, P. (2007). Smoothing splines estimators in functional linear regression with errors-in-variables. Computational Stat.Data Anal., Vol. 51, pp. 4832-4848.

[14] Liu, A., Tong, T., and Wang, Y. (2007). Smoothing Spline Estimation of Variance Functions. J. Comput. Graphical Stat., Vol. 16, pp. 312-329.

[15] Aydin, D. (2007). A comparison of the nonparametric regression models using smoothing spline and kernel regression. World Acad. Sci. Eng. Technol., Vol. 26, pp. 730-734.

[16] Aydin, D., Memmedhi, M., and Omay, R.E. (2013). Smoothing parameter selection for nonparametric regression using smoothing spline. European J. of Pure & Appl. Math., Vol. 6, pp. 222-238.

[17] Wang, Y., Guo, W., and Brown, W. B. (2000). Spline smoothing for bivariate data with applications to association between hormones. Statistica Sinica, Vol. 10, pp. 377-397.

[18] Fernandez, F. M., and Opsomer, J. D. (2005). Smoothing parameter selection methods for nonparametric regression with spatially correlated errors. Canadian J. Statistics, Vol. 33, pp, 279-295.

[19] Lestari, B., Budiantara, I. N., Sunaryo, S., and Mashuri, M. (2009). Spline estimator in homoscedastic multi-response nonparametric regression model. Proc. Indonesian Math. Soc. Int. Conf. on Math. & Its Appl., Oct. 12-13, Yogyakarta-Indonesia. (Available at: http://digilib.its.ac.id/ITS-Article-40312000750/18579).

[20] Lestari, B., Budiantara, I. N., Sunaryo, S., and Mashuri, M. (2010). Spline estimator in multi-response nonparametric regression model with unequal correlation of errors. J. Math. Stat., Vol. 6, pp. 327-332.

[21] Chamidah, N., Budiantara, I. N., Sunaryo, S., and Zain, I. (2012). Designing of child growth chart based on multi-response local polynomial modeling. J. Math. Stat., Vol. 8, pp. 342-347.

[22] Lestari, B., Budiantara, I. N., Sunaryo, S., and Mashuri, M. (2012). Spline smoothing for multiresponse nonparametric regression model in case of heteroscedasticity of variance. J. Math. and Stat., Vol. 8(3), pp. 377-384.

[23] Chamidah, N., and Lestari, B. (2016). Spline Estimator in Homoscedastic Multi-Response Nonparametric Regression Model in Case of Unbalanced Number of Observations. Far East Journal of Mathematical Sciences (FJMS), Vol. 100, pp. 1433-1453.

[24] Lestari, B., Fatmawati, and Budiantara, I. N. (2017). Estimasi Fungsi Regresi Nonparametrik Multirespon Menggunakan Reproducing Kernel Hilbert Space Berdasarkan Estimator Smoothing Spline. Proceed.ing of National Seminar on Mathematics and Its Applications (SNMA) 2017, Airlangga University, Surabaya, pp. 243-250.

[25] Lestari, B., Fatmawati, Budiantara, I. N., & Chamidah, N. (2018). Estimation of regression function in multiresponse nonparametric regression model using smoothing spline and kernel estimators. Journal of Physics: Conference Series, 1097, 012091. DOI: 10.1088/1742-6596/1097/1/012091.

[26] Lestari, B., Chamidah, N., & Saifudin, T. (2019). Estimasi fungsi regresi dalam model regresi nonparametrik birespon menggunakan estimator smoothing spline dan estimator kernel. Jurnal Matematika, Statistika & Komputasi (JMSK), Vol. 15(2), pp. 20-24. DOI: 10.20956/jmsk.v15i2.5565.

[27] Lestari, B., Anggraeni, D., and Saifudin, T. (2018). Estimation of Covariance Matrix based on Spline estimator in Homoscedastic Multiresponses Nonparametric Regression Model in Case of Unbalance Number of Observations. Far East Journal of Mathematical Sciences, Vol. 108(2), pp. 341-355.

[28] Lestari, B., Fatmawati, and Budiantara, I. N. (2019). Spline Estimator and Its Asymptotic Properties in Multiresponse Nonparametric Regression Model. Songklanakarin Journal of Science and Technology, In Press. (http://rdo.psu.ac.th/sjstweb/ArticleInPress.php).

[29] Islamiyati, A., Fatmawati, Chamidah, N. (2019). Ability of Covariance Matrix in Bi-Response Multi-Predictor Penalized Spline Model Through Longitudinal Data Simulation. International Journal of Academic and Applied Research (IJAAR), Vol. 3(3), pp. 8-11.

[30] IOTF-WHO (2000). The Asia-Pasific Perspective: Redefining Obesity and Its Treatment. Health Communications Australia Pty Limited, Sydney.

[31] Brown, C. D., Higgins, M., Donato, K. A., Rohde, F. C., Garrison, R. , Obarzanek, E., Ernst, N. D., and Horan, M. (2000). Body Mass Index and the Prevalence of Hypertension and Dyslipidemia. Obesity Research, Vol. 8(9), pp. 605-619.

[32] Droyvold, W. B., Midthjell, K. , Nilsen, T. I. L., and Holmen, J. (2005). Change in Body Mass Index and Its Impact on Blood Pressure: A Prospective Population Study. International Journal of Obesity, Vol. 29, pp. 650-655.

[33] Tesfaye, F., Nawi, NG., Van Minh, H., Byass, P., Berhane, Y., Bonita, R., and Wall, S. (2007), Association Between Body Mass Index and Blood Pressure Across Three Populations in Africa and Asia. J. Human Hypertension, Vol. 21, pp. 28-37.

[34] Kumar, T. A., Sudhir, U., Srinivasan, G., and Punith, K. (2008). Association of Body Mass Index with Blood Pressure in the Elderly. J. Indian Academy of Clinical Medicine, Vol. 9(4), pp. 274-276.

[35] Nanaware, N. L., Gavkare, A. M., and Surdi, A. D. (2011). Study of Correlation of Body Mass Index (BMI) with Blood Pressure in School Going Children and Adolescents. International Journal of Recent Trends in Science and Technology, Vol. 1(1), pp. 20-26.

[36] Roka, R., Michimi, A., dan Macy, G. (2015). Associations Between Hypertension and Body Mass Index and Waist Circumference in U.S. Adults: A Comparative Analysis by Gender. High Blood Pressure & Cardiovascular Prevention, Vol. 22(3), pp. 265-273.

[37] Green, P. J., and Silverman, B. W. (1994). Nonparametric Regression and Generalized Linear Models. Chapman Hall, New York.

[38] Eubank, R. L. (1999). Nonparametric Regression and Smoothing Spline. Marcel Deker, New York.

[39] Hardle, W. (1991). Applied Nonparametric Regression. Cambridge University Press. Cambridge.

[40] Schimek, M. G., (2000). Smoothing and Regression, John Wiley & Sons. New York.

[41] Watson, G. S. (1964). Smooth Regression Analysis. Sankhya Series A., Vol. 26, pp. 359-372.

[42] IOTF-WHO. (2000). *The Asia-Pasific Perspective: Redefining Obesity and Its Treatment*. Health Communications Australia Pty Limited, Sydney.