Modeling of Blood Pressures Based on Stress Score using Least Square Spline Estimator in Bi-response Non-parametric Regression

Nur Chamidah[1]*, Budi Lestari[2], Toha Saifudin[3]

[1]Department of Mathematics, Faculty of Science and Technology, Airlangga University, Jalan Mulyorejo Kampus C-UNAIR, Surabaya 60115, Indonesia

[2]Department of Mathematics, University of Jember, Indonesia

[3]Department of Mathematics, Faculty of Science and Technology, Airlangga University, Jalan Mulyorejo Kampus C-UNAIR, Surabaya 60115, Indonesia

Email: nur-c@fst.unair.ac.id; lestari.statistician@gmail.com; tohasaifudin@fst.unair.ac.id

Abstract

The basic idea of non-parametric regression is to let the data decide which regression function fits best without imposing any specific form on it. Consequently, non-parametric regression methods are in general more flexible. They can uncover structure in the data that might otherwise be missed. Bi-response non-parametric regression model provides powerful tools for modeling the regression function which represents association between blood pressures and stress score. Spline estimator has powerful and flexible properties for estimating the regression function. In this paper we discuss methods to estimate blood pressure affected by a stress score using least squared spline estimator. The results show that the estimated regression function is linear in observation and biased estimator. Also, we obtain the minimum GCV value of 389.9907, and optimal smoothing parameter values of 0.5255788 and 2.544688.

Keywords: Blood Pressure and Stress Score, GCV, 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 function of the non-parametric regression model. There are many researchers who have considered spline estimator for estimating regression function of the non-parametric regression model. Kimeldorf & Wahba (1971), Craven & Wahba (1979), and Wahba (1990) used original spline estimator to estimate regression function of smooth data. Cox (1983), and Cox & O'Sullivan (1996) proposed M-type spline to overcome outliers in non-parametric regression. Wahba (1983) has constructed confidence intervals for original spline model by using Bayesian approach. Wahba (1985) compared between generalised cross validation (GCV) and generalised maximum likelihood (GML) for choosing the smoothing parameter in the generalised spline smoothing problem. Oehlert (1992), and Koenker et al. (1994) introduced relaxed spline and quantile spline, respectively. Wang (1998) discussed smoothing spline models with correlated random errors. Wahba (2000) introduced some techniques for spline statistical model building by using reproducing kernel Hilbert spaces. Lee (2004) proposed a method that combines smoothing spline estimates of different smoothness to form a final improved estimate. Cardot et al. (2007) gave asymptotic property of smoothing splines estimators in functional linear regression with errors-in-variables. Liu et al. (2007) studied smoothing spline estimation of variance functions. Aydin (2007) showed goodness of spline estimators rather than kernel estimators in estimating non-parametric regression models for gross national product data. Aydin et al. (2013) studied the determination of an optimal smoothing parameter for non-parametric regression using smoothing spline. Ramadan et al. (2019) used least squared spline estimator to determine wasting nutritional status. Chamidah et al. (2019b) and Murbarani et al. (2019) used spline estimator to estimate median growth charts of children and to estimate AIDS models, respectively. Also, Oktavitri et al. (2019) used spline estimators for predicting suspended and attached process behavior in anaerobic batch reactors. Chamidah et al. (2018), and Chamidah et al. (2019a) used a non-parametric regression approach to improve classification accuracy of cysts and tumors, and to design a children's growth chart. All these researchers studied spline estimators in the case of a single response non-parametric regression model only.

In the real cases, we frequently face the problem in which two or more dependent variables are observed at several values of the independent variables, and there are correlations between the responses. Multi-response non-parametric regression models provide powerful tools to model the functions which represent the association of these variables. There are many researchers who have considered non-parametric models for multi-response data. Wang et al. (2000) studied spline smoothing for estimating non-parametric functions from bivariate data with the same correlation of errors. Fernandez & Opsomer (2005) proposed methods of estimating non-parametric regression model with spatially correlated errors. Chamidah & Saifudin (2013) estimated children's growth by using the multi-response, non-parametric regression model approach. Chamidah & Lestari (2016) discussed estimating the regression curve of the homoscedastic multi-response non-parametric regression in which the number of observations were unbalance. Lestari *et al.* (2017) proposed smoothing spline estimator for estimating of the multi-response non-parametric regression model by using reproducing kernel Hilbert space (RKHS). Lestari et al. (2018) discussed the construction of a covariance matrix in the case of homoscedasticity of variances of errors. Islamiyati et al. (2018) used penalised spline regression to estimate the covariance matrix on bi-response non-parametric regression. Lestari et al. (2019) discussed estimating of both the covariance matrix and optimal smoothing parameter. But, these researchers have not discussed estimating the smoothing parameter in multi-response to the non-parametric regression model when the variances of errors are not the same for the cross-section data. In addition, Chamidah & Saifudin (2013), Chamidah & Lestari (2016), Islamiyati et al. (2018), and Lestari et al. (2017, 2018, and 2019) have not discussed theoretically the application of the estimated model on the real case data.

Hypertension is often referred to as the silent killer because it takes the life of affected individuals without showing symptoms. However, the factors causing the disease (around 90%) are still unknown. The number of people living with hypertension is predicted to become 1.56 billion worldwide by the year 2025. The sickness is associated with cardiovascular diseases (CVD) risk factors, incidence, and mortality. It is also found to be prevalent among people of 35 years of age and above, currently smoking, and obese. The Seventh Report on the Joint National Committee on Prevention, Detection, Evaluation and Treatment of High Blood Pressure created a category called "pre-hypertension" which was defined as a systolic blood pressure (SBP) of 120-139 millimeters of mercury (mmHg) and a diastolic blood pressure (DBP) of 80-89 mmHg. Pre-hypertension, even in the low range (SBP: 120-130 mmHg or DBP: 80-85 mmHg), has been confirmed to have a higher risk of developing into hypertension. Hypertension has been associated with increased risk of coronary artery and cardiovascular and cerebrovascular diseases. A meta-analysis also reported that lower blood pressure could also lead to cardiovascular and chronic kidney diseases. This situation is critical in the Southeast Asian region with studies reporting HTN as an important risk factor for the attributable burden.

Several studies have found different risk factors for hypertension such as obesity, family history, stress levels, heart rate, and an unhealthy lifestyle (Brown *et al.*, 2000, and Roka *et al.*, 2015, Andriani & Chamidah, 2019, Lestari *et al.*, 2019). Furthermore, previous research showed the classification accuracy using binary logistic regression to be 72.5352% greater than the C4.5 algorithm which was 64.0845%. Therefore, it could be said that binary logistic regression is better than the C4.5 algorithm. However, the variables used were considered influential on hypertension through a regression curve without a pattern, therefore, in this paper we propose a bi-response non-parametric regression model approach. Moreover, a smoothing spline estimator used to estimate the regression function which describes the functional relationship between response variables and predictor variables. In this research, we discuss theoretical methods to estimate regression function and to select optimal smoothing parameters in bi-response non-parametric regression models that are a part of the multi-response non-parametric regression model. In addition, we give a numerical example to represent the application of this method on the real case data, i.e. an estimation of the regression function describing an association between blood pressure (systolic and diastolic blood pressure) and stress score.

2. Methods

Firstly, we consider the multi-response non-parametric regression model as given by Lestari et al. (2012, 2017, 2018, 2019) and by applying it to data of blood pressure and stress level. Next, the estimated regression function can be obtained by taking a solution of penalised weighted least square optimisation and by reproducing the kernel Hilbert space approach. Then, the optimal smoothing parameter can be obtained by minimising generalised cross validation (GCV) function.

In the numerical example, we used the secondary data obtained from the Heart Poly Surabaya Hajj Hospital, Surabaya, Indonesia. Data consists of systolic blood pressure, diastolic blood pressure, and stress level of 59 patients suffering from hypertension. Steps taken in analysing the data are: (1) descriptive statistical analysis was conducted on the predictor variables associated with the response variable; (2) blood pressure was modelled in Cardiac Poly outpatients in Surabaya Hajj Hospital with the bi-response non-parametric regression model approach based on smoothing estimators with the following steps: (a) making a scatterplot of systolic and diastolic blood pressure versus stress level to investigate their patterns; (b) testing the correlation between systolic blood pressure and diastolic blood pressure; (c) estimating the regression function by using the smoothing spline estimator; (d). Making a plot of the estimated systolic and diastolic blood pressures based on results obtained from step (c); (e) analysing and interpreting the estimated model of blood pressure associated with the stress score.

3. Results and Discussion

In this section, we give the results and discussion about methods of estimation of regression function and selection optimal smoothing spline in the bi-response non-parametric regression model by using the least squared spline estimator. Also, we give a numerical example of application of the method on the real case data.

3.1. Estimation of Regression Function Using Bi-response Least Square Spline Estimator

A regression analysis including two response variables shows there is significant correlation between response variables not only logically but also mathematically which is called a bi-response regression analysis. Bi-response non-parametric regression approach is used when it's regression function form is unknown. Generally, the bi-response non-parametric regression model can be expressed as follows:

$$\underset{\sim}{y}_i = \underset{\sim}{f}(t_i) + \underset{\sim}{\varepsilon}_i \qquad (1)$$

where $\underset{\sim}{y}_i = \left(y_i^{(1)} \quad y_i^{(2)}\right)^T$, $\underset{\sim}{f}(t_i) = \left(f^{(1)}(t_i) \quad f^{(2)}(t_i)\right)^T$, and $\underset{\sim}{\varepsilon}_i = \left(\varepsilon_i^{(1)} \quad \varepsilon_i^{(2)}\right)$ is zero mean random error with variance Σ_i. Function f is an unknown regression curve that can be approached by truncated spline function as follows:

$$f^{(r)}(t_i) = \beta_0^{(r)} + \beta_1^{(r)} t_i + \ldots + \beta_p^{(r)} t_i^{\,p} + \sum_{k=1}^{K} \beta_{p+k}^{(r)} (t_i - \delta_k^{\,(r)})_+^p \qquad (2)$$

where $\delta_1, \delta_2, \ldots, \delta_K$ are knots and $(t_i - \delta_k)^p$ is a truncated function which is expressed as follows:

$$(t_i - \delta_k)_+^p = \begin{cases} (t_i - \delta_k)^p, & t_i \geq \delta_k \\ 0 & , t_i < \delta_k \end{cases} \qquad (3)$$

The parameter $\underset{\sim}{\beta}$ in (2) can be estimated by using weighted least square (WLS) method that minimiszes the weighted sum of squared errors as follows:

$$\begin{aligned}\underset{\sim}{\varepsilon}^T \mathbf{W} \underset{\sim}{\varepsilon} &= (\underset{\sim}{y} - \mathbf{X}\underset{\sim}{\beta})^T \mathbf{W} (\underset{\sim}{y} - \mathbf{X}\underset{\sim}{\beta}) \\ &= (\underset{\sim}{y}^T - \underset{\sim}{\beta}^T \mathbf{X}^T)(\mathbf{W}\underset{\sim}{y} - \mathbf{W}\mathbf{X}\underset{\sim}{\beta}) \\ &= \underset{\sim}{y}^T \mathbf{W}\underset{\sim}{y} - \underset{\sim}{y}^T \mathbf{W}\mathbf{X}\underset{\sim}{\beta} - \underset{\sim}{\beta}^T \mathbf{X}^T \mathbf{W}\underset{\sim}{y} + \underset{\sim}{\beta}^T \mathbf{X}^T \mathbf{W}\mathbf{X}\underset{\sim}{\beta} \\ &= \underset{\sim}{y}^T \mathbf{W}\underset{\sim}{y} - 2\underset{\sim}{\beta}^T \mathbf{X}^T \mathbf{W}\underset{\sim}{y} + \underset{\sim}{\beta}^T \mathbf{X}^T \mathbf{W}\mathbf{X}\underset{\sim}{\beta}\end{aligned}$$

Next, we have $\dfrac{\partial \underset{\sim}{\varepsilon}^T \mathbf{W} \underset{\sim}{\varepsilon}}{\partial \underset{\sim}{\beta}^T} = -2\mathbf{X}^T \mathbf{W}\underset{\sim}{y} + 2\mathbf{X}^T \mathbf{W}\mathbf{X}\underset{\sim}{\beta} = 0$

$$\Leftrightarrow 2\mathbf{X}^T \mathbf{W}\mathbf{X}\underset{\sim}{\beta} = 2\mathbf{X}^T \mathbf{W}\underset{\sim}{y}$$

Such that we obtain:

$$\hat{\underset{\sim}{\beta}} = (\mathbf{X}^T \mathbf{W}\mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}\underset{\sim}{y} \tag{4}$$

where $\mathbf{W} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix}^{-1}$ is a weighted matrix which is inverse of the covariance matrix of response-1 and response-2 errors, $\mathbf{\Sigma}_{rr}$ ($r = 1,2$) is diagonal matrix of covariance of r^{th} response error, and $\mathbf{\Sigma}_{12}$, $\mathbf{\Sigma}_{21}$ are diagonal matrix of covariance between response-1 error and response-2 error. Based on (4) we get the estimated model of biresponsebi-response nonparametricnon-parametric regression as follows:

$$\hat{\underset{\sim}{y}} = \mathbf{X}\hat{\underset{\sim}{\beta}} = \mathbf{X}(\mathbf{X}^T \mathbf{W}\mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}\underset{\sim}{y} = \mathbf{A}\underset{\sim}{y} \tag{5}$$

Therefore we have mean squared error (MSE) as follows:

$$MSE(\lambda) = \frac{1}{2n}(\underset{\sim}{y} - \hat{\underset{\sim}{y}})^T (\underset{\sim}{y} - \hat{\underset{\sim}{y}}) = \frac{1}{2n}(\underset{\sim}{y} - \mathbf{A}\underset{\sim}{y})^T (\underset{\sim}{y} - \mathbf{A}\underset{\sim}{y}) \tag{6}$$

Next, we use the obtained MSE value to calculate generalised cross validation (GCV). Then, we use the minimum GCV to determine optimum knots. The GCV value in bi-response non-parametric regression model is expressed by:

$$GCV(\lambda)=\frac{MSE(\lambda)}{\left[\frac{1}{2n}tr(\mathbf{I}-\mathbf{A})\right]^2} \quad (7)$$

where $\mathbf{A}=\mathbf{X}(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}$, and λ is a smoothing parameter expressed as $\lambda=\left(p,K,\left(\delta_1,\delta_2,...,\delta_K\right)\right)$ with order p, number of knot K, knots $(\delta_1,\delta_2,...,\delta_K)$.

3.2 Estimation of an Optimal Smoothing Parameter

Wahba (1990) has shown that in uni-response spline nonparametricnon-parametric regression, if smoothing parameter (λ) value is very small $(\lambda \to 0)$ then it will give a very rought estimator of nonparametricnon-parametric regression function. In contrary, if the smoothing parameter (λ) value is very large $(\lambda \to \infty)$ then it will give a very smooth estimator of nonparametricnon-parametric regression function. Therefore, we need to select the optimum smoothing parameter (λ) in order to obtain an estimator that is suitable with the data. For this need, some researchers have proposed some selection methods, for instance, Craven & Wahba (1979) proposed a cross validation (CV) method, Wang (1998) proposed an unbiased risk (UBR) method, and Wahba (1990) proposed a generaliszed cross validation (GCV) method. Not only does uniresponse spline nonparametricnon-parametric regression, but also multiresponsemulti-response spline nonparametricnon-parametric regression depends on the smoothing parameter λ_k, $k=1,2,3$.

In this section we discuss the selection method for selecting the optimal smoothing parameter in multiresponsethe multi-response nonparametricnon-parametric regression model for data of blood pressures and pulse. The Rregression function estimator of multiresponsemulti-response nonparametricnon-parametric regression model for data of blood pressures and stress score as given by Lestari et al. (2018b and 2019b) can be expressed as follows:

$$\hat{\underset{\sim}{f}}_\lambda(t)=H(\lambda_1,\lambda_2;\underset{\sim}{\sigma}^2)\underset{\sim}{y} \quad (8)$$

where $\underset{\sim}{\sigma}^2 = (\sigma_1^2, \sigma_2^2)'$. MSE (Mean Square Error) of (8) can be determined as follows:

$$MSE(\lambda_1, \lambda_2; \underset{\sim}{\sigma}^2) = (\sum_{k=1}^{2} n_k)^{-1} (\underset{\sim}{y} - \hat{\underset{\sim}{f}}_\lambda(t))' W(\underset{\sim}{\sigma}^2)(\underset{\sim}{y} - \hat{\underset{\sim}{f}}_\lambda(t))$$

$$= (\sum_{k=1}^{2} n_k)^{-1} \left(\underset{\sim}{y} - H(\lambda_1, \lambda_2; \underset{\sim}{\sigma}^2)\underset{\sim}{y}\right)' W(\underset{\sim}{\sigma}^2)\left(\underset{\sim}{y} - H(\lambda_1, \lambda_2; \underset{\sim}{\sigma}^2)\underset{\sim}{y}\right)$$

$$= (\sum_{k=1}^{2} n_k)^{-1} \left\| \left(W(\underset{\sim}{\sigma}^2)\right)^{\frac{1}{2}} \left(I_{\sum_{k=1}^{2} n_k} - H(\lambda_1, \lambda_2; \underset{\sim}{\sigma}^2) \right) \underset{\sim}{y} \right\|^2 .$$

where $\left(W(\underset{\sim}{\sigma}^2)\right)^{\frac{1}{2}}$ is a diagonal matrix.

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

$$G(\lambda_1, \lambda_2; \underset{\sim}{\sigma}^2) = \frac{\left(\sum_{k=1}^{2} n_k\right)^{-1} \left\| \left(W(\underset{\sim}{\sigma}^2)\right)^{\frac{1}{2}} \left(I_{\sum_{k=1}^{2} n_k} - H(\lambda_1, \lambda_2; \underset{\sim}{\sigma}^2) \right) \underset{\sim}{y} \right\|^2}{\left[\left(\sum_{k=1}^{2} n_k\right)^{-1} trace\left(I_{\sum_{k=1}^{2} n_k} - H(\lambda_1, \lambda_2; \underset{\sim}{\sigma}^2) \right) \right]^2} .$$

The optimal smoothing parameter $\underset{\sim}{\lambda}_{opt} = (\lambda_{1(opt)}, \lambda_{2(opt)})'$ is obtained by taking the solution of the following optimiszation:

$$G_{opt}(\lambda_{1(opt)}, \lambda_{2(opt)}; \underset{\sim}{\sigma}^2) = \underset{\lambda_1 \in R^+, \lambda_2 \in R^+}{Min} \left\{ G(\lambda_1, \lambda_2; \underset{\sim}{\sigma}^2) \right\}$$

$$= \underset{\lambda_1 \in R^+, \lambda_2 \in R^+}{Min} \left\{ \frac{\left(\sum_{k=1}^{2} n_k\right)^{-1} \left\| \left(W(\underset{\sim}{\sigma}^2)\right)^{\frac{1}{2}} \left(I_{\sum_{k=1}^{2} n_k} - H(\lambda_1, \lambda_2; \underset{\sim}{\sigma}^2) \right) \underset{\sim}{y} \right\|^2}{\left[\left(\sum_{k=1}^{2} n_k\right)^{-1} trace\left(I_{\sum_{k=1}^{2} n_k} - H(\lambda_1, \lambda_2; \underset{\sim}{\sigma}^2) \right) \right]^2} \right\},$$

where the norm $\|\underset{\sim}{v}\| = \sqrt{v_1^2 + v_2^2}$ for vector $\underset{\sim}{v} = (v_1, v_2)'$.

3.3 Numerical Example

The secondary data used consists of systolic blood pressure, diastolic blood pressure, and stress levels of 59 patients suffering from hypertension. Based on the correlation test result we conclude that there is correlation between systolic and diastolic blood pressures with a correlation value is 0.581. Also, results of the scatter plot of systolic and diastolic blood pressures versus stress level showed that there was no certain pattern of parametric regression patterns. Based on these results, we conclude that the bi-response non-parametric regression model approach is applicable to analyse the data. Estimating results of regression function based on least square spline estimator give a minimum GCV value of 389.9907, and optimal smoothing parameters of 0.5255788 and 2.544688. Furthermore, plots of the estimated systolic and diastolic blood pressures are given in Fig. 1 and Fig. 2, respectively.

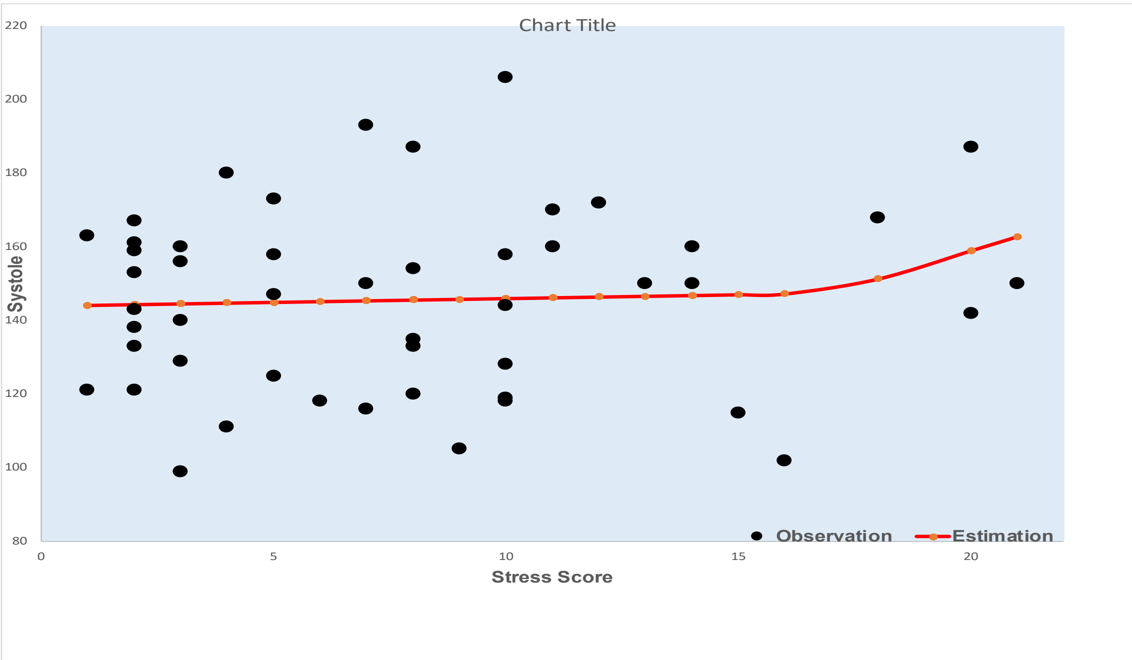

Figure 1. Plots of the estimated systolic blood pressure versus stress score

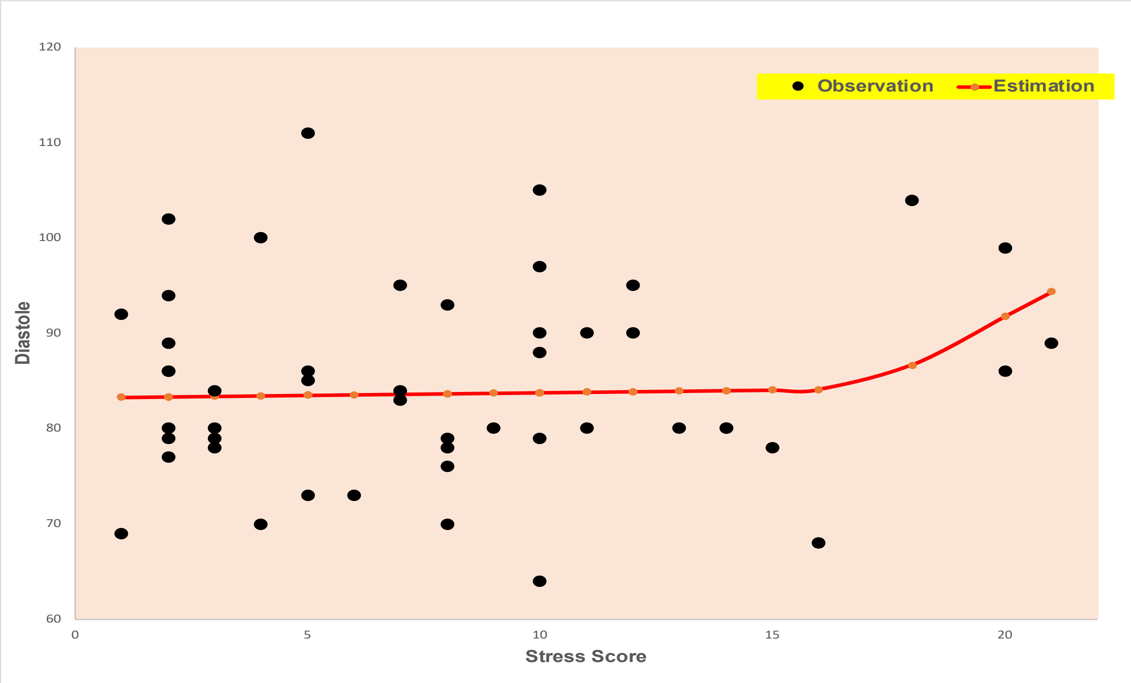

Figure 2. Plots of the estimated diastolic blood pressure versus stress score

Fig. 1 shows that systolic blood pressure slowly increases in a row with increasing stress score. Also, Fig. 2 shows that systolic blood pressure slowly increases in a row with increasing stress score. However, the increasing of systolic blood pressure is sharper than that of diastolic blood pressure. It means that stress level can cause the increasing of both systolic and diastolic blood pressure.

4. Conclusion

Based on the estimated model that we have obtained, we conclude that the estimated model is a linear function in observation. In addition, by taking expectation of the estimated model, i.e., $E(\hat{\underset{\sim}{f}}_{\lambda}(t))$ we obtain that the estimated regression function is a biased estimator. For this real case data, we obtain the minimum GCV value of 389.9907, and optimal smoothing parameter values of 0.5255788 and 2.544688. Further, either systolic blood pressure or diastolic blood pressure tends to slowly increase in a row with increasing of stress score. However, the increasing of systolic blood pressure is sharper than that of diastolic blood pressure.

Acknowledgment

Many thanks to the 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 Fundamental Research Grant (*Hibah Penelitian Dasar*) in the fiscal year 2019 with Contract Number: 726/UN3.14/LT/2019.

References

Andriani, P. and Chamidah, N. (2019). Modelling of Hypertension Risk Factors Using Logistic Regression to Prevent Hypertension in Indonesia, *IOP Conf. Series: Journal of Physics: Conf. Series* **1306** 012027. Doi:10.1088/1742-6596/1306/1/012027.

Aydin, D. (2007). A comparison of the non-parametric regression models using smoothing spline and kernel regression. *World Acad. Sci. Eng. Technol.*, **26**, 730-734.

Aydin, D., Memmedhi, M., and Omay, R.E. (2013). Smoothing parameter selection for non-parametric regression using smoothing spline. *European J. of Pure & Appl. Math.*, **6**, 222-238.

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*, **8**(9), 605-619.

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.*, **51,** 4832-4848.

Chamidah, N., and Saifudin, T. (2013). Estimation of Children Growth Based on Kernel Smoothing in Multi-response Non-parametric Regression. *Applied Mathematical Sciences*, **7**(37), 1839-1847.

Chamidah, N., and Lestari, B. (2016). Spline Estimator in Homoscedastic Multi-response Non-parametric Regression Model in Case of Unbalanced Number of Observations. *Far East Journal of Mathematical Sciences (FJMS)*, **100**, 1433-1453.

Chamidah, N., Gusti, K. H., Tjahjono, E., and Lestari, B. (2019a). Improving of Classification Accuracy of Cyst and Tumor Using Local Polynomial Estimator. *TELKOMNIKA*, **17**(3), 1492-1500.

Chamidah, N., Tjahjono, E., Fadilah, A. R., Lestari, B. (2018). Standard Growth Charts for Weight of Children in East Java Using Local Linear Estimator, *Journal of Physics: Conference Series*, **1097** 012092.

Chamidah, N., Zaman, B., Muniroh, L., Lestari, B. (2019b). Estimation of Median Growth Charts for Height of Children in East Java Province of Indonesia Using Penalized Spline Estimator. *International Conference Proceedings GCEAS*, **5**, 68-78.

Cox, D. D. (1983). Asymptotic for M-type smoothing spline. *Annals. Statistics*. **11**, 530-551.

Cox, D. D., and O'Sullivan, F. (1996). Penalized likelihood type estimators for generalized non-parametric regression. *J. Mult. Anal.*, **56**, 185-206.

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.*, **31**, 377-403.

Eubank, R. L. (1999). *Non-parametric Regression and Smoothing Spline*. Marcel Deker, New York.

Fernandez, F. M., and Opsomer, J. D. (2005). Smoothing parameter selection methods for non-parametric regression with spatially correlated errors. *Canadian J. Statistics*, **33**, 279-295.

Green, P. J., and Silverman, B. W. (1994). *Non-parametric Regression and Generalized Linear Models*. Chapman Hall, New York.

Hardle, W. (1991). *Applied Non-parametric Regression*. Cambridge University Press. Cambridge.

Islamiyati, A., Fatmawati, and Chamidah, N. (2018). Estimation of Covariance Matrix on Bi-response Longitudinal Data Analysis with Penalized Spline Regression. *Journal of Physics: Conference Series*, **979**, 012093.

Kimeldorf, G., andWahba, G. (1971). Some result on tchebycheffian spline functions. *J. of Math. Anal. and Appl.*, **33**, 82-95.

Koenker, R., Pin, N. G., and Portnoy, S. (1994). Quantile smoothing splines. *Biometrics*. **81**, 673-680.

Lee, T., C., M. (2004). Improved smoothing spline regression by combines estimates of different smoothness. *Stat. and Prob. Letters*, **67**, 133-140.

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*, *Faculty of Sciences and Technology, Airlangga University, Surabaya*, 243-250.

Lestari, B., Anggraeni, D., and Saifudin, T. (2018a). Estimation of Covariance Matrix based on Spline estimator in Homoscedastic Multi-Responses Non-parametric Regression Model in Case of Unbalance Number of Observations. *Far East Journal of Mathematical Sciences (FJMS)*, **108**(2), 341-355.

Lestari, B., Fatmawati, Budiantara, I. N., & Chamidah, N. (2018b). Estimation of Regression Function in Multi-response Non-parametric Regression Model Using Smoothing Spline and Kernel Estimators. *Journal of Physics: Conference Series*, **1097**, 012091. DOI: 10.1088/1742-6596/1097/1/012091.

Lestari, B., Chamidah, N., & Saifudin, T. (2019a). Estimasi Fungsi Regresi dalam Model Regresi Nonparametrik Birespon Menggunakan Estimator Smoothing Spline dan Estimator Kernel. *Jurnal Matematika, Statistika dan Komputasi (JMSK)*, **15**(2), 20-24. DOI: 10.20956/jmsk.v15i2.5565.

Lestari, B., Fatmawati, and Budiantara, I. N. (2019b). Spline Estimator and Its Asymptotic Properties in Multi-response Non-parametric Regression Model. *Songklanakarin Journal of Science and Technology*, In Press.

Lestari, B., Fatmawati, and Budiantara, I. N. (2019c). Smoothing Spline Estimator in Multi-response Non-parametric Regression for Predicting Blood Pressures and Heart Rate, *International Journal of Academic and Applied Research* (IJAAR), **3**(9), 1-8.

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

Murbarani, N., Swastika, Y., Dwi, A., Aris, B., Chamidah, N. (2019). Modeling of the Percentage of AIDS Sufferers in East Java Province using Non-parametric Regression Approach Based on Truncated Spline Estimator. *Indonesian Journal of Statistics and Its Applications,* **3(**2), 139-147.

Oehlert, G. W. (1992). Relaxed boundary smoothing spline. *Annals. Statistics*, **20**, 146-160.

Oktavitri, N. I., Kuncoro, E. P., Purnobasuki, H., and Chamidah, N. (2019). Prediction of Suspended and Attached Process Behavior in Anaerobic Batch Reactor Using Non-parametric Regression Model Approach Based on Spline Estimator. *Eco. Env. & Cons.* **25** (April Suppl. Issue) , S96-S100.

Ramadan, W., Chamidah, N., Zaman, B., Muniroh, L. and Lestari, B. (2019). Standard Growth Chart of Weight for Height to Determine Wasting Nutritional Status in East Java Based on Semiparametric Least Square Spline Estimator. *IOP Conf. Series: Materials Science and Engineering,* **546** 052063. DOI: 10.1088/1757-899X/546/5/052063.

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*, **22**(3), 265-273.

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

Wahba, G. (1978). Improper Prior, Spline Smoothing and the Problem of Guarding Against Model Errors in Regression. *Journal of the Royal Statistical Society Series B.*, **40**, 364-372.

Wahba, G. (1983). Bayesian confidence intervals for the cross-validated smoothing spline. *Journal of the Royal Statistical Society Series B*, **45**, 133-150.

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

Wahba, G. (1990). *Spline Models for Observational Data*. SIAM Philadelphia. Pennsylvania.

Wahba, G. (1992). *Multivariate Functional and Operator Estimation, Based on Smoothing Spline and Reproducing Kernels*. (http://www.stat.wisc.edu/~wahba/ftp1/oldie/santafe1992.pdf).

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).

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

Watson, G. S. (1964). Smooth Regression Analysis. *Sankhya Series A*., **26,** 359-372.