Songklanakarin J. Sci. Technol.
42 (4), 897-909, Jul. - Aug. 2020

http://www.sjst.psu.ac.th

Original Article

Penalized spline estimator with multi smoothing parameters in bi-response multi-predictor nonparametric regression model for longitudinal data

Anna Islamiyati[1], Fatmawati[2], and Nur Chamidah[2*]

[1] *Department of Mathematics, Faculty of Mathematics and Natural Sciences, Hasanuddin University, Makassar, Sulawesi Selatan, 90245 Indonesia*

[2] *Department of Mathematics, Faculty of Sciences and Technology, Airlangga University, Surabaya, Jawa Timur, 60115 Indonesia*

Received: 29 October 2018; Revised: 14 May 2019; Accepted: 24 May 2019

Abstract

Penalized spline estimators that depend on a smoothing parameter is one type of estimator used in the estimation regression curve in nonparametric regression. The smoothing parameter is one of the most important components in the penalized spline estimator because it is related to the smoothness of the regression curve. In this paper, we determine the optimum number of smoothing parameters in a bi-response multi-predictor nonparametric regression model. Based on the result of the simulation study, we find that the optimum number of smoothing parameters corresponds to the number of predictor variables in each response. We also apply the estimated model to case of blood glucose levels in type 2 diabetes patients. The results of study show that there are different patterns of changes in blood glucose levels, both day and night, based on the length of care, the calorie diet, and the carbohydrate diet.

Keywords: penalized spline estimator, multi-smoothing parameters, longitudinal data, blood glucose levels, type 2 diabetes patients

1. Introduction

The smoothing spline estimators for estimating the regression curve of the nonparametric regression model have been introduced by Eubank (1999), Green and Silverman (1994), Wahba (1990), and Wang (1998). The smoothness of the curve is related to the smoothing parameter, symbolized by λ (lambda). However, for the purposes of real data analysis, we need not only the smoothness of the curve, but also a smooth curve that can be interpreted visually. Therefore, in this paper, we develop the use of the penalized spline estimators proposed by Claenskens, Kribovokova, and Opsomer (2009), Eilers and Marx (1996), and Ruppert and Carrol (1997, 2000) to the penalized spline estimator with multi-smoothing parameters for longitudinal data.

Some studies of the penalized spline estimator such as Durban, Harezlak, Wand, and Carrol (2005) have examined the individual curves of longitudinal data using penalized spline in the semi-parametric regression model. Yao and Lee (2006) proposed an iterative procedure of the penalized splines in principal component analysis. Lee and Oh (2007) used the penalized spline through M-type robust estimates to analyze the daily growth of ozone. Aydin and Yilmaz (2018) used modified spline estimators for nonparametric regression models with right-censored data, especially when the censored response observations are converted to synthetic data.

There are several previous researchers who have studied data containing two or more correlated response variables. For cross-sectioned data, Soo and Bates (1996) used multi-response spline regression. Chamidah and Lestari

*Corresponding author
Email address: nur-c@fst.unair.ac.id

(2016), Lestari, Anggraeni, and Saifudin (2017), Lestari, Budiantara, Sunaryo, and Mashuri (2010, 2012), Lestari, Fatmawati, and Budiantara (2017), Lestari, Fatmawati, Budiantara, and Chamidah (2018), Lestari, Anggraeni, and Saifudin (2018), and Lestari, Chamidah, and Saifudin (2019) used natural smoothing spline estimators to estimate the regression curve of multi-response nonparametric regression models. Chamidah, Budiantara, Sunaryo, and Ismaini (2012) and Chamidah and Saifudin (2013) used local polynomial and kernel estimators, respectively. Chamidah and Rifada (2016) have studied local linear estimators in the bi-response semiparametric regression model for estimating median growth charts of children in Surabaya, Indonesia. Lestari, Fatmawati, Budiantara, and Chamidah (2018) estimated regression curves by using spline and kernel estimators. Chamidah, Fadilah, Tjahjono, and Lestari (2018) used local linear estimators for designing a growth reference chart of children. For longitudinal data, Wang, Guo, and Brown (2000) analyzed hormonal balance data involving two responses based on measurement time using smoothing spline estimator. Durban *et al*. (2005) estimated the individual curves of longitudinal data. Budiantara *et al*. (2009) discussed weighted spline estimator for longitudinal data. Islamiyati, Fatmawati, and Chamidah (2017) estimated the function of goodness of fit in criteria penalized spline on bi-response nonparametric regression for longitudinal data. Islamiyati, Fatmawati, and Chamidah (2018) analyzed bi-response case based on measurement time using a single penalty. In real cases involving time, additional predictors and assumed to influence the response, for example, in the case of genes of Glioblastoma cancer and blood glucose levels of diabetes studied using the parametric approach. Lee, Du, Wei, Hayes, and Liu (2012) and Lee and Liu (2012) used a Gaussian model in examining several types of genes in Glioblastoma cancer that were measured repeatedly by considering neuron, axon and synaptic transmission. Sun *et al*. (2016) examined the effect of fasting time on the day and night blood glucose. Festa *et al*. (2017) used the Cox regression model in the analysis of moderate and severe hypoglycemia based on age, sex, and race. Hettiaratchi, Ekanayake, Welihinda, and Perera (2011) studied glycemic and insulinemic responses to breakfast and the succeeding second meal in type 2 diabetes. Besides that, we may use bi-response multi-predictor nonparametric regression model in longitudinal data. The model is used to resolve bi-response and multi-predictor cases that could not be accurately analyzed through a parametric approach.

The penalized spline estimation in the bi-response multi-predictor regression model consisting of the goodness of fit and the penalty function. The penalty function that we use in this paper is a quadratic penalty derived from the spline regression coefficient. We need to elaborate the penalty function of every response constructed from a set of predictors. In the additive of penalty, the number of smoothing parameters depends on the number of predictors involved in each response function. Also, the number of knots and order of the spline also play a role in determining the number of smoothing parameters that can be involved in the penalty function. For knots selection, we used the fixed selection method proposed by Ruppert (2002). Montoya, Ulloa, and Miller (2014) have tested the fixed selection method which gives the smallest GCV value. In simulation study, we examine the optimum number of smoothing parameters that can be involved in the penalty function of bi-response multi-predictor model. We show that the number of smoothing parameters in each response for the bi-response multi-predictor nonparametric regression model is adjusted for the number of predictors. We can possibly reduce the number of smoothing parameters with computational considerations, but need further research in other simulation studies.

Furthermore, the study of the theoretical development of model estimation, including the description of the penalty additive in the penalized spline criterion of the bi-response multi-predictor regression are described in the second section. In section 3, we present a simulation study of some functions of y_1 and y_2 with the number of samples and the correlation values varying. This simulation study shows the number of optimum smoothing parameters that should be involved in the penalty function. The optimization of the penalty additive in the penalized spline estimator is indicated by the minimum GCV value. Subsequently, the fourth section shows the application of multi-predictor bi-response model in type 2 diabetes data. These data were obtained from the hospital of Hasanuddin University South Sulawesi Indonesia. The factors involved in the model are daytime blood glucose (y_1), nighttime blood glucose (y_2), treatment time (t_1), the calorie diet (t_2), and the carbohydrate diet (t_3). The last section contains the conclusion of this paper to be used as material for further research development.

2. Model and Estimates

2.1 The bi-response multi-predictor regression nonparametric model

The bi-response multi-predictor nonparametric regression model is a nonparametric regression model that contains two response variables. Given the data $\left(t_{ij1}, t_{ij2}, \ldots t_{ijp}, y_{1.ij}, y_{2.ij}\right)$, for $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots m_i$, as follow the bi-response multi-predictor nonparametric regression model for the longitudinal data:

$$y_{r.ij} = f_r\left(t_{ij1}, t_{ij2}, \ldots, t_{ijp}\right) + \varepsilon_{r.ij}, r = 1, 2\,. \tag{1}$$

Suppose that the model in (1) statisfies the additive properties of the predictor function, which in this case can be described as:

$$y_{r.ij} = f_r\left(t_{ij1}\right) + f_r\left(t_{ij2}\right) + \ldots + f_r\left(t_{ijp}\right) + \varepsilon_{r.ij}\ . \tag{2}$$

We assume random error $\varepsilon_{r.ij}, r = 1, 2$ as follows:

$$\begin{aligned} &E\left(\varepsilon_{1.ij}\right) = E\left(\varepsilon_{2.ij}\right) = 0, E\left(\varepsilon_{1.ij}^2\right) = \sigma_{1.i}^2, E\left(\varepsilon_{2.ij}^2\right) = \sigma_{2.i}^2, \\ &E\left(\varepsilon_{1.ij}, \varepsilon_{2.i^*j^*}\right) = E\left(\varepsilon_{2.ij}, \varepsilon_{1.i^*j}\right) = \begin{cases} \sigma_{12.i} & ; \quad i = i^* \text{and } j = j^* \\ 0 & ; \qquad \text{others.} \end{cases} \end{aligned} \tag{3}$$

2.2 The matrix of covariance

The model in (2) is assumed to have correlations between responses and between repeated measurements on the same subject. This leads to the estimation of the nonparametric regression model that uses the covariance matrix as a weight. We symbolize it by $\mathbf{\Omega}$ as follows:

$$\mathbf{\Omega} = \mathrm{Var}(\underset{\sim}{\varepsilon}) = E\left(\underset{\sim}{\varepsilon} - E(\underset{\sim}{\varepsilon})\right)^T \left(\underset{\sim}{\varepsilon} - E(\underset{\sim}{\varepsilon})\right) = E\left(\underset{\sim}{\varepsilon}^T \underset{\sim}{\varepsilon}\right)$$

$$\mathbf{\Omega} = E\left[\left(\underset{\sim}{\varepsilon}_{1.1}, \underset{\sim}{\varepsilon}_{1.2}, \ldots, \underset{\sim}{\varepsilon}_{1.n}, \underset{\sim}{\varepsilon}_{2.1}, \underset{\sim}{\varepsilon}_{2.2}, \ldots, \underset{\sim}{\varepsilon}_{2.n}\right)^T \left(\underset{\sim}{\varepsilon}_{1.1}, \underset{\sim}{\varepsilon}_{1.2}, \ldots, \underset{\sim}{\varepsilon}_{1.n}, \underset{\sim}{\varepsilon}_{2.1}, \underset{\sim}{\varepsilon}_{2.2}, \ldots, \underset{\sim}{\varepsilon}_{2.n}\right)\right],$$

where $\underset{\sim}{\varepsilon}_{r.1} = \left(\varepsilon_{r.11}, \varepsilon_{r.12}, \ldots, \varepsilon_{r.1m_1}\right)^T, \underset{\sim}{\varepsilon}_{r.2} = \left(\varepsilon_{r.21}, \varepsilon_{r.22}, \ldots, \varepsilon_{r.2m_2}\right)^T, \ldots, \underset{\sim}{\varepsilon}_{r.n} = \left(\varepsilon_{r.n1}, \varepsilon_{r.n2}, \ldots, \varepsilon_{r.nm_n}\right)^T$ for $r = 1,2$.

Based on the assumption in (3), we obtain the covariance matrix $\mathbf{\Omega}$ as follows:

$$\begin{pmatrix} \mathbf{\Sigma}_{1.1} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{\Sigma}_{12.1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{1.2} & & \mathbf{0} & \mathbf{0} & \mathbf{\Sigma}_{12.2} & & \mathbf{0} \\ \vdots & & \ddots & \vdots & & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & & \mathbf{\Sigma}_{1.n} & \mathbf{0} & \mathbf{0} & & \mathbf{\Sigma}_{12.n} \\ \mathbf{\Sigma}_{21.1} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{\Sigma}_{2.1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{21.2} & & \mathbf{0} & \mathbf{0} & \mathbf{\Sigma}_{2.2} & & \mathbf{0} \\ \vdots & & \ddots & \vdots & & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{\Sigma}_{21.n} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{\Sigma}_{2.n} \end{pmatrix}, \tag{4}$$

where $\mathbf{\Omega}$ is a variance-covariance matrix of the first and the second response. In general, the matrix $\mathbf{\Sigma}_{r.i}$ is the variance matrix in the *r*-response of the *i*-subject, $\mathbf{\Sigma}_{r.i} = diag\left(\sigma_i^2, \sigma_i^2, \ldots, \sigma_i^2\right)$ for $r = 1,2$, $i = 1,2,\ldots,n$, and the matrix $\mathbf{\Sigma}_{rs.i}$ is the covariance matrix in the *r*-response with the *s*-response of the *i*-subject, $\mathbf{\Sigma}_{rs.i} = diag\left(\sigma_{rs.i}, \sigma_{rs.i}, \ldots, \sigma_{rs.i}\right)$ for $r \neq s = 1,2$.

2.3 Estimation of the nonparametric regression function

The unknown functions $f_r\left(t_{ij1}\right), f_r\left(t_{ij2}\right), \ldots, f_r\left(t_{ijp}\right)$ in equation (2) are estimated using the penalized spline estimator and can be expressed as follows:

$$\left.\begin{aligned} f_r\left(t_{ij1}\right) &= \sum_{u_1=0}^{q_{r.1}} \beta_{r.u_1}\left(t_{ij1}\right)^{u_1} + \sum_{v_1=1}^{d_1} \beta_{r.(q_1+v_1)1}\left(t_{ij1} - K_{r.v_1}\right)_+^{q_{r.1}} \\ f_r\left(t_{ij2}\right) &= \sum_{u_2=0}^{q_{r.2}} \beta_{r.u_2}\left(t_{ij2}\right)^{u_2} + \sum_{v_2=1}^{d_2} \beta_{r.(q_2+v_2)2}\left(t_{ij2} - K_{r.v_2}\right)_+^{q_{r.2}} \\ &\vdots \\ f_r\left(t_{ijp}\right) &= \sum_{u_p=0}^{q_{r.p}} \beta_{r.u_p}\left(t_{ijp}\right)^{u_p} + \sum_{v_p=1}^{d_p} \beta_{r.(q_p+v_p)p}\left(t_{ijp} - K_{r.v_p}\right)_+^{q_{r.p}} \end{aligned}\right\} \tag{5}$$

Next, based on the equation (1), (2) and (5) we get:

$$f_r\left(t_{ij1}, t_{ij2}, \ldots, t_{ijp}\right) = \mathbf{X}_{r.1}\underset{\sim}{\beta}_{r.1} + \mathbf{X}_{r.2}\underset{\sim}{\beta}_{r.2} + \ldots + \mathbf{X}_{r.p}\underset{\sim}{\beta}_{r.p} \;. \tag{6}$$

Equation (6) can be expressed for two spline functions as follows:

$$\left.\begin{aligned} f_1\left(t_{ij1}, t_{ij2}, \ldots, t_{ijp}\right) &= \mathbf{X}_{1.1}\underset{\sim}{\beta}_{1.1} + \mathbf{X}_{1.2}\underset{\sim}{\beta}_{1.2} + \ldots + \mathbf{X}_{1.p}\underset{\sim}{\beta}_{1.p} = \mathbf{X}_1\underset{\sim}{\beta}_1 \\ f_2\left(t_{ij1}, t_{ij2}, \ldots, t_{ijp}\right) &= \mathbf{X}_{2.1}\underset{\sim}{\beta}_{2.1} + \mathbf{X}_{2.2}\underset{\sim}{\beta}_{2.2} + \ldots + \mathbf{X}_{2.p}\underset{\sim}{\beta}_{2.p} = \mathbf{X}_2\underset{\sim}{\beta}_2 \end{aligned}\right\}, \tag{7}$$

where $\underset{\sim}{\beta}_1 = \left(\underset{\sim}{\beta}_{1.1} \quad \underset{\sim}{\beta}_{1.2} \quad \cdots \quad \underset{\sim}{\beta}_{1.p}\right)^T$, $\underset{\sim}{\beta}_2 = \left(\underset{\sim}{\beta}_{2.1} \quad \underset{\sim}{\beta}_{2.2} \quad \cdots \quad \underset{\sim}{\beta}_{2.p}\right)^T$ is the regression coefficient in the first and second responses. Next, $\underset{\sim}{\beta}_{1.1}$ is the spline regression coefficient on the first response and predictor. In general, for $r=1,2$ and $h=1,2,\ldots,p$, we obtained $\underset{\sim}{\beta}_{r.h} = \left(\beta_{r.0h}, \beta_{r.1h}, \beta_{r.2h}, \ldots, \beta_{r.q_h h}, \beta_{r.(q_h+1)h}, \ldots, \beta_{r.(q_h+d_h)h}\right)^T$. Next, $\mathbf{X}_1 = \left(\mathbf{X}_{1.1} \quad \mathbf{X}_{1.2} \quad \ldots \quad \mathbf{X}_{1.p}\right)$ is the matrix X in the first response, $\mathbf{X}_2 = \left(\mathbf{X}_{2.1} \quad \mathbf{X}_{2.2} \quad \ldots \quad \mathbf{X}_{2.p}\right)$ is the matrix **X** in the second response. The matrix **X** on the response *r* and the predictor *h* is symbolized by $\mathbf{X}_{r.h}$ as follows:

$$
\mathbf{X}_{r.h} = \begin{pmatrix}
1 & t_{11h} & t_{11h}^2 & \cdots & t_{11h}^{q_{r.h}} & (t_{11h} - K_{r.1h})^{q_{r.h}} & \cdots & \left(t_{11h} - K_{r.d_{r.h}h}\right)^{q_{r.h}} \\
1 & t_{12h} & t_{12h}^2 & & t_{12h}^{q_{r.h}} & (t_{12h} - K_{r.1h})^{q_{r.h}} & & \left(t_{11h} - K_{r.d_{r.h}h}\right)^{q_{r.h}} \\
\vdots & & & & \vdots & & & \vdots \\
1 & t_{1m_1h} & t_{1m_1h}^2 & \cdots & t_{1m_1h}^{q_{r.h}} & \left(t_{1m_1h} - K_{r.1h}\right)^{q_{r.h}} & \cdots & \left(t_{1m_1h} - K_{r.d_{r.h}h}\right)^{q_{r.h}} \\
\vdots & & & & \vdots & & & \vdots \\
1 & t_{n1h} & t_{n1h}^2 & & t_{n1h}^{q_{r.h}} & (t_{n1h} - K_{r.1h})^{q_{r.h}} & & \left(t_{n1h} - K_{r.d_{r.h}h}\right)^{q_{r.h}} \\
1 & t_{n2h} & t_{n2h}^2 & & t_{n2h}^{q_{r.h}} & (t_{n2h} - K_{r.1h})^{q_{r.h}} & & \left(t_{n2h} - K_{r.d_{r.h}h}\right)^{q_{r.h}} \\
\vdots & & & & \vdots & & & \vdots \\
1 & t_{nm_nh} & t_{nm_nh}^2 & \cdots & t_{nm_nh}^{q_{1.1}} & \left(t_{nm_nh} - K_{r.1h}\right)^{q_{r.h}} & \cdots & \left(t_{nm_nh} - K_{r.d_{r.h}h}\right)^{q_{r.h}}
\end{pmatrix}.
$$

Based on (7), model (1) can be expressed in the form of the first response vector $\underset{\sim}{y}_1 = \left(y_{1.11}, y_{1.12}, \ldots, y_{1.1m_1}, \ldots, y_{1.n1}, y_{1.n1}, \ldots, y_{1.nm_n}\right)^T$ and the second response vector $\underset{\sim}{y}_2 = \left(y_{2.11}, y_{2.12}, \ldots, y_{2.1m_1}, \ldots, y_{2.n1}, y_{2.n1}, \ldots, y_{2.nm_n}\right)^T$ as follows:

$$
\left.\begin{aligned}
\underset{\sim}{y}_1 &= \mathbf{X}_1 \underset{\sim}{\beta}_1 + \underset{\sim}{\varepsilon}_1 \\
\underset{\sim}{y}_2 &= \mathbf{X}_2 \underset{\sim}{\beta}_2 + \underset{\sim}{\varepsilon}_2
\end{aligned}\right\}. \tag{8}
$$

Based on (8), the bi-response multi-predictor nonparametric regression model for the longitudinal data based on the penalized spline estimator as given in (1) can be expressed in the following matrix notation:

$$
\underset{\sim}{y} = \mathbf{X}\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}
$$

where $\underset{\sim}{y} = \left(\underset{\sim}{y}_1, \underset{\sim}{y}_2\right)^T$ is a vector of response variable that contains two response variables. The matrix $\mathbf{X}$ is expressed as the matrix **X** in the first and the second responses that is $\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{pmatrix}$. Next, $\underset{\sim}{\beta} = \left(\underset{\sim}{\beta}_1 \quad \underset{\sim}{\beta}_2\right)^T$ is the spline regression coefficient vector in the first and the second responses. Next, $\underset{\sim}{\varepsilon} = \left(\underset{\sim}{\varepsilon}_1, \underset{\sim}{\varepsilon}_2\right)^T$ is a vector of random error in the first and the second responses.

The estimator of bi-response multi-predictor nonparametric regression model can be obtained by carrying out the following penalized weighted least square (PWLS):

$$
\begin{aligned}
\text{PWLS} &= \underset{\sim}{\varepsilon}^T \mathbf{\Omega}^{-1} \underset{\sim}{\varepsilon} + \sum_{r=1}^{2} \lambda_r \mathrm{M}_r \\
&= \underset{\sim}{y}^T \mathbf{\Omega}^{-1} \underset{\sim}{y} - 2\underset{\sim}{\beta}^T \mathbf{X}^T \mathbf{\Omega}^{-1} \underset{\sim}{y} + \underset{\sim}{\beta}^T \mathbf{X}^T \mathbf{\Omega}^{-1} \mathbf{X} \underset{\sim}{\beta} + \underset{\sim}{\beta}^T \mathbf{D}_{\underset{\sim}{\lambda}} \underset{\sim}{\beta}.
\end{aligned}
$$

The elaboration of the penalty function $\sum_{r=1}^{2} \lambda_r \mathrm{M}_r$ is given in section (2.4). Next, we get the parameter estimate of $\underset{\sim}{\beta}$ as follows:

$$\hat{\underset{\sim}{\beta}} = \left(\mathbf{X}^T\mathbf{\Omega}^{-1}\mathbf{X} + \mathbf{D}_{\underset{\sim}{\lambda}}\right)^{-1}\mathbf{X}^T\mathbf{\Omega}^{-1}\underset{\sim}{y}. \tag{9}$$

Finally, we obtain estimator of bi-response multi-predictor nonparametric regression model:

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

2.4 Additive penalty

The penalty function in the bi-response multi-predictor nonparametric regression model is expressed as an additive function of the predictor as follows:

$$\sum_{r=1}^{2}\lambda_r \mathrm{M}_r = \lambda_1\mathrm{M}_1 + \lambda_2\mathrm{M}_2. \tag{10}$$

where $\mathrm{M}_1 = \int_{a_1}^{b_1}\left[g_1^{c_{1.1}}(t_{ij1})\right]^2 dt_{ij1} + \int_{a_2}^{b_2}\left[g_1^{c_{1.2}}(t_{ij2})\right]^2 dt_{ij2} + \ldots + \int_{a_p}^{b_p}\left[g_1^{c_{1.p}}(t_{ijp})\right]^2 dt_{ijp}$, $\mathrm{M}_2 = \int_{a_1}^{b_1}\left[g_2^{c_{2.1}}(t_{ij1})\right]^2 dt_{ij1} + \int_{a_2}^{b_2}\left[g_2^{c_{2.2}}(t_{ij2})\right]^2 dt_{ij2} + \ldots + \int_{a_p}^{b_p}\left[g_2^{c_{2.p}}(t_{ijp})\right]^2 dt_{ijp}$, and $g_r \in W_2$ is the function of the response contained in the Sobolev Space, $c_{r.h}$ is the derivative order in the predictor *h*, and the response *r*, that is $c_{r.h} = (q_{r.h}+1)$, $q_{r.h}$ is the spline order of the spline regression function of the predictor *h* and the response *r*.

In the process of the penalty function, we use the definition of the delta dirac function in forming the truncated element of the spline function in each predictor with the number of points of knots as many as $v = 1,2,\ldots d_{r.h}$. The definition of delta dirac function in this paper is given as follows:

$$\delta\left(t_{ijh} - K_{r.vh}\right) = \begin{cases}\left(t_{ijh} - K_{r.vh}\right) & ; \quad t_{ijh} \geq K_{r.vh} \\ 0 & ; \quad t_{ijh} < K_{r.vh}\end{cases}. \tag{11}$$

The delta dirac function shift causes the formation of new functions, that is $g_{r.h}$ where truncated element is $\left(t_{ijh} - K_{r.1h}\right)_+^{q_{r.h}+1}, \left(t_{ijh} - K_{r.2h}\right)_+^{q_{r.h}+1}, \ldots, \left(t_{ijh} - K_{r.d_{r.h}h}\right)_+^{q_{r.h}+1}$. Next, we get function $g_{r.h}$ as follows:

$$g_{r.h} = \mathrm{B}_0 + \ldots + \mathrm{B}_{q_{r.h}} + \mathrm{B}_{(q_{r.h}+1)} + \ldots + \mathrm{B}_{(q_{r.h}+d_{r.h})}. \tag{12}$$

where $\mathrm{B}_0 = \beta_{r.0h}$, $\mathrm{B}_{q_{r.h}} = \beta_{r.q_{r.h}h}\left(t_{ijh}\right)^{q_{r.h}}$, $\mathrm{B}_{(q_{r.h}+1)} = \beta_{r.(q_{r.h}+1)h}(t_{ijh} - K_{r.1h})_+^{q_{r.h}+1}$, and $\mathrm{B}_{(q_{r.h}+d_{r.h})} = \beta_{r.(q_{r.h}+d_{r.h})h}(t_{ijh} - K_{r.d_{r.h}h})_+^{q_{r.h}+1}$. Subsequently function $g_{r.h}$ in equation (12) is derivated with respect to $t_{ij1}, t_{ij2}, \ldots, t_{ijp}$ with the higher order derivated $c_{r.h} = q_{r.h}+1$, the result is as follows:

$$\frac{\partial^{c_{r.h}} g_{r.h}}{\partial t_{ijh}} = \frac{\partial^{q_{r.h}+1} g_{r.h}}{\partial t_{ijh}} = (q_{r.h}+1)!\left(\beta_{r.(q_{r.h}+1)h} + \ldots + \beta_{r.(q_{r.h}+d_{r.h})h}\right) = (q_{r.h}+1)!\sum_{v_h=1}^{d_{r.h}}\beta_{r.(q_{r.h}+v_h)h}. \tag{13}$$

If $(q_{r.h}+1)!\sum_{v_h=1}^{d_{r.h}}\beta_{r.(q_{r.h}+v_h)h} = \Delta_{r.h}$ is the derivative function $g_{r.h}$ with respect to t_{ijh} in the *r*-response and *h*-predictor of function with respect to t_{ijh}, then $\Delta_{r.h}$ for $h = 1,2,\ldots,p$ is as follow:

$$\frac{\partial^{q_1+1} g_{r.1}}{\partial t_{ij1}} + \frac{\partial^{q_2+1} g_{r.2}}{\partial t_{ij2}} + \ldots + \frac{\partial^{q_p+1} g_{r.p}}{\partial t_{ijp}} = \Delta_{r.1} + \Delta_{r.2} + \ldots + \Delta_{r.p}. \tag{14}$$

where $\Delta_{r.1}=(q_1+1)!\sum_{v_1=1}^{d_{r.1}}\beta_{r.(q_{r.1}+v_1)1},\Delta_{r.2}=(q_2+1)!\sum_{v_2=1}^{d_{r.2}}\beta_{r.(q_{r.2}+v_2)2},\ldots,\Delta_{r.p}=(q_p+1)!\sum_{v_p=1}^{d_{r.p}}\beta_{r.(q_{r.p}+v_p)p}$.

Next, by taking integral over $[a_h,b_h]$ of squared (14), we get M1 and M2 as follows:

$$\mathrm{M}_1=\sum_{v_1=1}^{d_{1.1}}\beta^2_{1.(q_{1.1}+v_1)1}(C_{1.1})+\sum_{v_2=1}^{d_{1.2}}\beta^2_{1.(q_{1.2}+v_2)2}(C_{1.2})+\ldots+\sum_{v_p=1}^{d_{1.p}}\beta^2_{1.(q_{1.p}+v_p)p}(C_{1.p}) \;, \tag{15}$$

$$\mathrm{M}_2=\sum_{v_1=1}^{d_{2.1}}\beta^2_{2.(q_{2.1}+v_1)1}(C_{2.1})+\sum_{v_2=1}^{d_{2.2}}\beta^2_{2.(q_{2.2}+v_2)2}(C_{2.2})+\ldots+\sum_{v_p=1}^{d_{2.p}}\beta^2_{2.(q_{2.p}+v_p)p}(C_{2.p}) \;, \tag{16}$$

where $C_{1.1},C_{1.2},\ldots,C_{1.p}$ is a constant in the first response where $C_{1.1}>0,C_{1.2}>0,\ldots,C_{1.p}>0$ and $C_{2.1},C_{2.2},\ldots,C_{2.p}$ is a constant in the first response where $C_{2.1}>0,C_{2.2}>0,\ldots,C_{2.p}>0$. In general, the value of $C_{r.p}$ for $r=1,2$ and $h=1,2,\ldots,p$ is $C_{r.p}=\{c_{r.p}!\}^2(b_p-a_p)$. If we elaborate the components in (15) and (16), for r = 1,2 then we get:

$$\begin{aligned}\lambda_1\mathrm{M}_1&=\lambda_1\left(\underset{\sim}{\beta}_{1.1}^T\mathbf{D}_{1.1}\underset{\sim}{\beta}_{1.1}(C_{1.1})+\underset{\sim}{\beta}_{1.2}^T\mathbf{D}_{1.2}\underset{\sim}{\beta}_{1.2}(C_{1.2})+\ldots+\underset{\sim}{\beta}_{1.p}^T\mathbf{D}_{1.p}\underset{\sim}{\beta}_{1.p}(C_{1.p})\right)\\&=\underset{\sim}{\beta}_{1.1}^T\lambda_{1.1}\mathbf{D}_{1.1}\underset{\sim}{\beta}_{1.1}+\underset{\sim}{\beta}_{1.2}^T\lambda_{1.2}\mathbf{D}_{1.2}\underset{\sim}{\beta}_{1.2}+\ldots+\underset{\sim}{\beta}_{1.p}^T\lambda_{1.p}\mathbf{D}_{1.p}\underset{\sim}{\beta}_{1.p}\\&=\underset{\sim}{\beta}_{1.1}^T\mathbf{D}_{\lambda_{1.1}}\underset{\sim}{\beta}_{1.1}+\underset{\sim}{\beta}_{1.2}^T\mathbf{D}_{\lambda_{1.2}}\underset{\sim}{\beta}_{1.2}+\ldots+\underset{\sim}{\beta}_{1.p}^T\mathbf{D}_{\lambda_{1.p}}\underset{\sim}{\beta}_{1.p}\\&=\underset{\sim}{\beta}_{1}^T\mathbf{D}_{\lambda_1}\underset{\sim}{\beta}_{1}\end{aligned} \tag{17}$$

where $\lambda_1C_{1.1}=\lambda_{1.1}$, $\lambda_1C_{1.2}=\lambda_{1.2},\ldots,$ $\lambda_1C_{1.p}=\lambda_{1.p}$, $\mathbf{D}_{\lambda_1}=Diag\left(\mathbf{D}_{\lambda_{1.1}},\mathbf{D}_{\lambda_{1.2}},\ldots,\mathbf{D}_{\lambda_{1.p}}\right)$, $\lambda_{1.1}\mathbf{D}_{1.1}=\mathbf{D}_{\lambda_{1.1}},\lambda_{1.2}\mathbf{D}_{1.2}=\mathbf{D}_{\lambda_{1.2}},\ldots,\lambda_{1.p}\mathbf{D}_{1.p}=\mathbf{D}_{\lambda_{1.p}}$ is matrix **D** in the first response and predictor 1 to *p*. In general, for matrix **D** in the first response is $\mathbf{D}_{\lambda_{1.h}}=\mathrm{diag}\left(\lambda_{1.h}a_{1.0h},\lambda_{1.h}a_{1.1h},\ldots,\lambda_{1.h}a_{1.q_{1.h}h},\lambda_{1.h}a_{1.(q_{1.h}+1)h},\lambda_{1.h}a_{2.(q_{1.h}+2)h},\ldots,\lambda_{1.h}a_{2.(q_{1.h}+d_{1.h})h}\right),h=1,2,\ldots,p,$ $a_{1.0h},a_{1.1h},\ldots,a_{1.q_{1.h}h}=0$ and $a_{1.(q_{1.h}+1)h},a_{1.(q_{1.h}+2)h},\ldots,a_{1.(q_{1.h}+d_{1.h})h}=1$. Furthermore, the penalty in the second response in equation (18) is obtained in the same way as in the first response.

$$\begin{aligned}\lambda_2\mathrm{M}_2&=\lambda_2\left(\underset{\sim}{\beta}_{2.1}^T\mathbf{D}_{2.1}\underset{\sim}{\beta}_{2.1}(C_{2.1})+\underset{\sim}{\beta}_{2.2}^T\mathbf{D}_{2.2}\underset{\sim}{\beta}_{2.2}(C_{2.2})+\ldots+\underset{\sim}{\beta}_{2.p}^T\mathbf{D}_{2.p}\underset{\sim}{\beta}_{2.p}(C_{2.p})\right)\\&=\underset{\sim}{\beta}_{2}^T\mathbf{D}_{\lambda_2}\underset{\sim}{\beta}_{2}\end{aligned} \tag{18}$$

where $\mathbf{D}_{\lambda_2}=Diag\left(\mathbf{D}_{\lambda_{21}},\mathbf{D}_{\lambda_{22}},\ldots,\mathbf{D}_{\lambda_{2p}}\right)$, $\lambda_{2.1}\mathbf{D}_{2.1}=\mathbf{D}_{\lambda_{21}},\lambda_{2.2}\mathbf{D}_{2.2}=\mathbf{D}_{\lambda_{22}},\ldots,\lambda_{2.p}\mathbf{D}_{2.p}=\mathbf{D}_{\lambda_{2p}}$ is matrix **D** in the second response and predictor 1 to *p*. In general, matrix **D** in the second response is

$\mathbf{D}_{\lambda_{2h}}=\mathrm{diag}\left(\lambda_{2.h}a_{2.0h},\lambda_{2.h}a_{2.1h},\ldots,\lambda_{2.h}a_{2.q_{2.h}h},\lambda_{2.h}a_{2.(q_{2.h}+1)h},\lambda_{2.h}a_{2.(q_{2.h}+2)h},\ldots,\lambda_{2.h}a_{2.(q_{2.h}+d_{2.h})h}\right)$ where $a_{2.0h},a_{2.1h},\ldots,a_{2.q_{2.h}h}=0$ and $a_{2.(q_{2.h}+1)h},a_{2.(q_{2.h}+2)h},\ldots,a_{2.(q_{2.h}+d_{2.h})h}=1$.

By substituting (17) and (18) into (10) we get:

$$\sum_{r=1}^{2}\lambda_r\mathrm{M}_r=\underset{\sim}{\beta}_1^T\mathbf{D}_{\lambda_1}\underset{\sim}{\beta}_1+\underset{\sim}{\beta}_2^T\mathbf{D}_{\lambda_2}\underset{\sim}{\beta}_2=\underset{\sim}{\beta}^T\mathbf{D}_{\lambda}\underset{\sim}{\beta} \tag{19}$$

where $\mathbf{D}_{\lambda}=Diag\left(\mathbf{D}_{\lambda_1},\mathbf{D}_{\lambda_2}\right)$

Based on the penalty function in the first and second responses, it shows that the number of smoothing parameters in the first response is as many as predictors, so also in the second response. So, we have to consider in the bi-response multi-predictor spline regression, that the smoothing parameters are involved in the model are as many as the predictors. We show the involvement of smoothing parameters in each function of some predictors through a simulation study in the third section.

3. Simulation Study

In this section, we simulate a longitudinal data set consisting of two responses and three predictors using R-code. The purpose of this simulation study is to demonstrate the ability of the smoothing parameters in the penalty additive function in the bi-response multi-predictor nonparametric regression model. In general, the bi-response regression model is expressed as follows:

$$\underset{\sim}{y} = \underset{\sim}{f} + \underset{\sim}{\varepsilon}, \quad r = 1,2,$$

where $\underset{\sim}{\varepsilon}$ is assumed to be normally distributed with $E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$ and $\mathrm{var}(\underset{\sim}{\varepsilon}) = \mathbf{\Omega}$. Matrix $\mathbf{\Omega}$ is a covariance matrix corresponding to (5) in which variance, $\sigma_{ij}^2 = (t_{ij} \times \mathrm{sd})^2$, where $\mathrm{sd} = 0.1, 10, 30, 50$ in all cases in this simulation. We conduct a simulation study by defining the form of a predictor function for every response function, the number of observations, and the response correlation coefficient. The function used in this simulation for the first and the second responses are given in equations (20) and (21), respectively.

$$\left.\begin{aligned} f_{1.1} &= 0.5 + 2t_1 - 1t_1^2 + 2(t_1 - 2)_+^2 - 2(t_1 - 4)_+^2 \\ f_{1.2} &= \sin(2\pi t_2 + 3) \\ f_{1.3} &= 7 + 3\sin(3\pi t_3) \end{aligned}\right\}. \qquad (20)$$

$$\left.\begin{aligned} f_{2.1} &= 0.5 + 2t_1 + 1.5t_1^2 - (t_1 - 2)_+^2 - 4(t_1 - 4)_+^2 \\ f_{2.2} &= \sin(3\pi t_2 + 3) \\ f_{2.3} &= 9 + 3\sin(2\pi t_3) \end{aligned}\right\}. \qquad (21)$$

In this simulation study, we use functions (20) and (21) for different correlation values between the y_1 and y_2 responses, which are $\rho = \pm 0.9$, $\rho = 0.8$, $\rho = 0.7$, and $\rho = \pm 0.6$ for 50 subjects, and each subject measured 3 to 10 times. We analyze the data on each correlation value by using three (according to the number of predictors) and one smoothing parameter based on the GCV criterion. The optimization of the number of smoothing parameters involved in the additive penalty function is shown through the boxplot. GCV values on three smoothing parameters $(\lambda_1, \lambda_2, \lambda_3)$ are symbolized GCV_3, and the GCV value on one parameter smoothing (λ) is symbolized GCV_1. We can see in the boxplot (Figure 1) that a minimum GCV value is obtained with the use of three smoothing parameters for each different response function in the case of two mutually correlated responses. The results of this simulation are briefly given in Table 1.

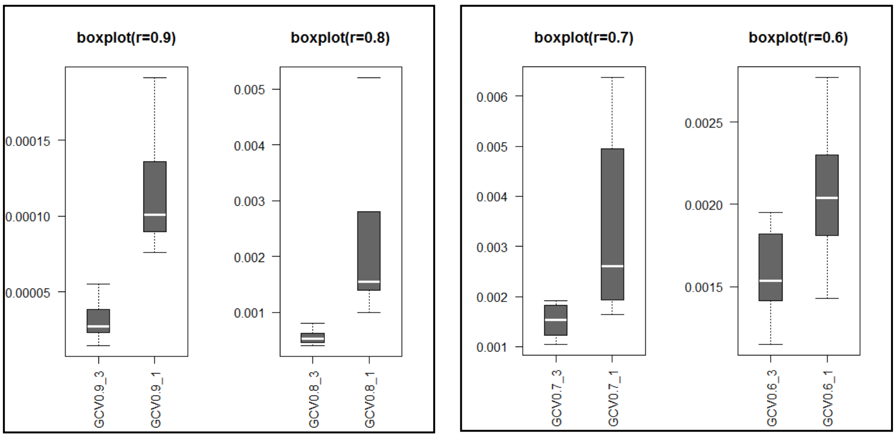

Figure 1. (Different spline function form) Boxplot of GCV values of estimation with (GCV_3) and λ (GCV_1) on every response with correlation between responses ($r = 0.6, 0.7, 0.8, 0.9$).

Table 1. The GCV value of the simulated data in the case of different functions with $\sigma_{ij}^2 = t_{ij} \times sd$

sd	*r*	λ	*n* = 20	50	100	300
50	-0.6	$\lambda_1,\lambda_2,\lambda_3$	0.0207*	0.00609*	0.000512*	0.0000267
		λ	0.0237	0.01012	0.001034	0.0000652
0.1	-0.9	$\lambda_1,\lambda_2,\lambda_3$	0.000171*	0.0000514*	0.00000817*	0.00000127
		λ	0.00452	0.000182	0.0000116	0.00000143
50	0.6	$\lambda_1,\lambda_2,\lambda_3$	0.0139*	0.00279*	0.00272*	0.0000464
		λ	0.0162	0.00629	0.00494	0.0000978
30	0.7	$\lambda_1,\lambda_2,\lambda_3$	0.0516*	0.000822*	0.000408*	0.0000218
		λ	0.0768	0.00126	0.00168	0.0000532
10	0.8	$\lambda_1,\lambda_2,\lambda_3$	0.00947*	0.000668*	0.000166*	0.0000111
		λ	0.0159	0.00226	0.000767	0.0000656
0.1	0.9	$\lambda_1,\lambda_2,\lambda_3$	0.00037*	0.0000189*	0.0000105*	0.00000101
		λ	0.0077	0.000179	0.0000276	0.00000108

Furthermore, our simulations on several different subjects, i.e., n = 20, 50, 100, and 300. Every subject is measured 3 to 10 times. Boxplots of GCV values are shown in in Figure 2. We make boxplots of GCV values based on GCV values obtained through iterations involving three and one smoothing parameters. The boxplot of GCV values shows that the value in GCV_3 is smaller than GCV_1. The difference of value of the GCV_3 and GCV_1 is very large for n = 20, 30, 50, and 100. For the number of subjects n = 300 (± 3,000 observed data), the value of GCV_3 is smaller than the value of GCV_1 and the difference is small. Therefore, the penalized spline estimator in bi-response multi-predictor nonparametric regression model must involve as many smoothing parameters as there are predictors in the function of each response in penalty additive functions.

Based on the simulation study, we find that the number of smoothing parameters involved in the penalty function of the bi-response multi-predictor model is as many as the number of its predictors. We can use a smoothing parameter to reduce computing costs only when the sample size is large.

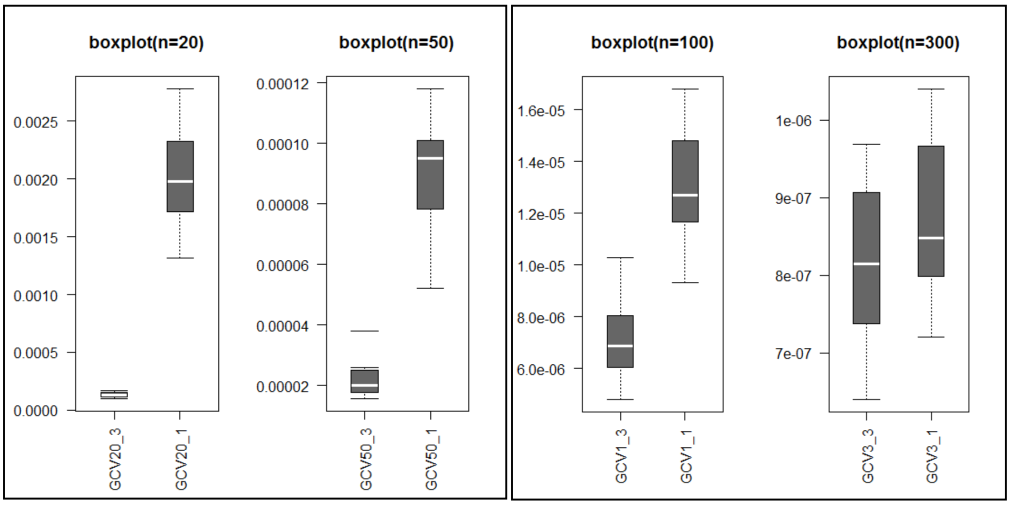

Figure 2. (Different spline function form) Boxplots of GCV values of estimation with $\lambda_1,\lambda_2,\lambda_3$ (GCV_3) and λ (GCV_1) in every response with the number of subjects n = 20, 50, 100, 300.

4. Application of Blood Glucose Level Data of Type 2 Diabetes Patients

The pattern of change in blood glucose in type 2 diabetes patients is analyzed through a bi-response multi-predictor nonparametric regression model using a penalized spline estimator. Data of type 2 diabetes patients are obtained from the Hospital of Hasanuddin University, South Sulawesi, Indonesia. Blood glucose levels of patients that serve as response variables are daytime random blood glucose (y_1) and nighttime random blood glucose (y_2). Changes in blood glucose levels are measured by treatment time (t_1), the total diet of calorie (t_2), and the total diet of carbohydrate (t_3). We have patients with varying treatment times, so that total observations were 418 measurements.

The correlation coefficient between y_1 and y_2 is 0.6. It means that there is a correlation between daytime blood glucose and nighttime blood glucose for type 2 diabetes patients. In the first analysis, we use a partial analysis of each patient. The partial analysis performs on the time factor of treatment, the diet of calorie, and the diet of carbohydrate. The patterns of the daytime and the nightime blood glucose of type 2 diabetes patients are shown in Figure 3.

The result of the partial model estimates given in Figure 3 shows the different curve shapes of every predictor variable. The number of knots, the spline order and the smoothing parameters are chosen based on a minimum value of the GCV which is shown in Table 2. The obtained optimal order equals 2 in every predictor, whereas the number of knots and lambda values are different. The optimal lambda (λ) values are also shown in Table 2.

Further, type 2 diabetes data are analyzed using bi-response nonparametric regression by using the penalized spline estimator. Based on Table 2, we obtain the minimum GCV value of order spline optimum, i.e., 2, and the number of smoothing parameters, i.e., 3 for every response. The number of knots is different in every predictor. The covariance matrix is involved in the estimation of the bi-response regression model as shown in Figure 4 and Figure 5. The covariance matrix is the result of the variance estimate of the errors of the first and the second responses. The GCV values of the regression model with the covariance matrix give smaller

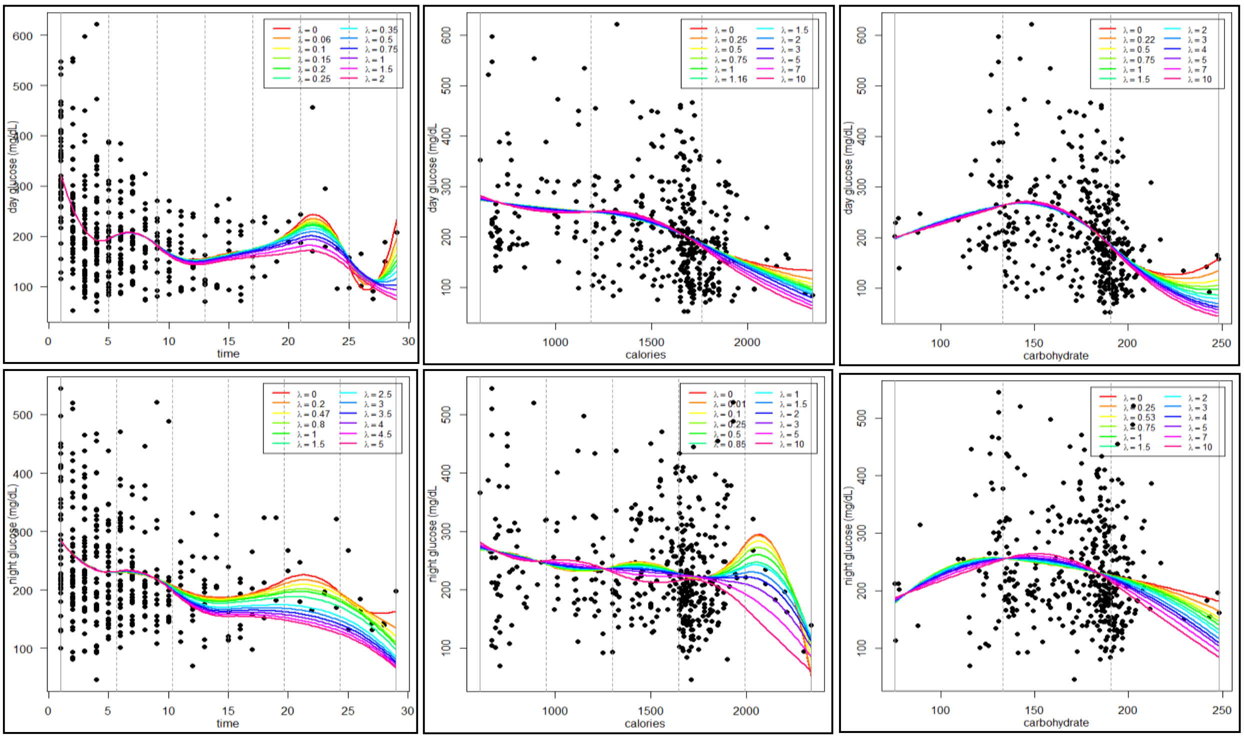

Figure 3. Estimation of the regression curve of the first response (above) and the second response (below).

Table 2. Order, knots, and lambda values based on GCV values.

Order	Response	Predictor	Number of knots	Lambda (λ)	GCV
1	y_1	t_1	6	0.06	0.04423
		t_2	2	1.16	
		t_3	2	0.22	
	y_2	t_1	5	0.47	
		t_2	4	0.10	
		t_3	2	0.53	
2	y_1	t_1	6	0.06	0.04017
		t_2	2	1.16	
		t_3	2	0.22	
	y_2	t_1	5	0.47	
		t_2	4	0.10	
		t_3	2	0.53	
3	y_1	t_1	6	0.06	0.04079
		t_2	2	1.16	
		t_3	2	0.22	
	y_2	t_1	5	0.47	
		t_2	4	0.10	
		t_3	2	0.53	

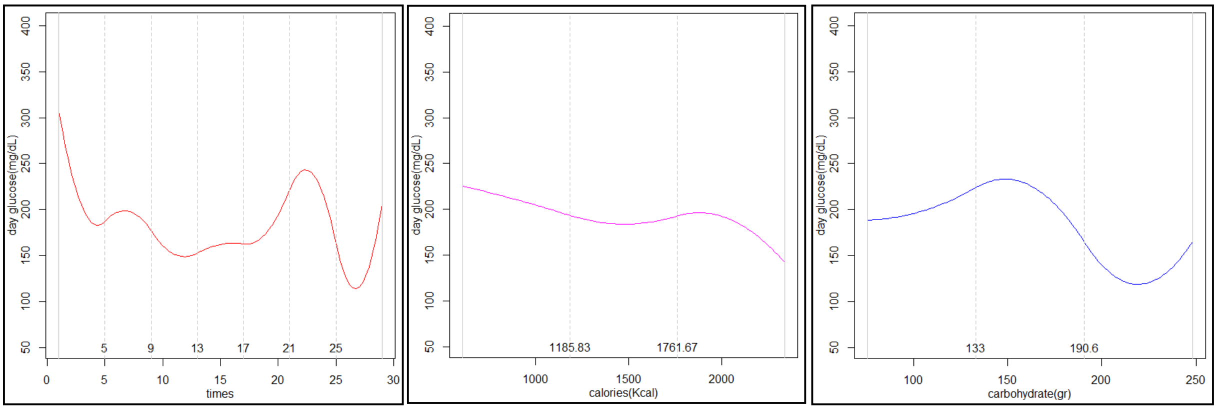

Figure 4. Curve estimation of bi-response multi-predictor regression based on penalized spline of the first response

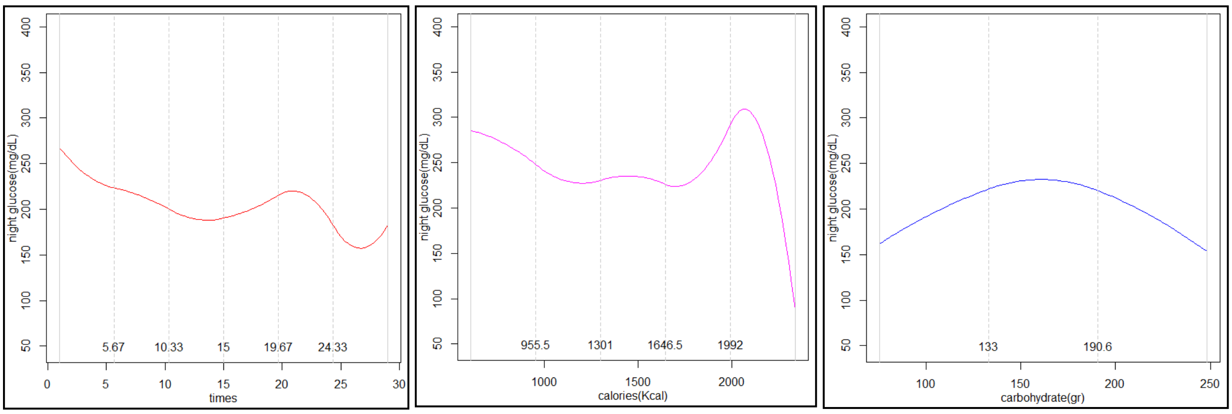

Figure 5. Curve estimation of bi-response multi-predictor regression based on penalized spline of the second response

error values. The obtained GCV value is 0.04017. Based on the analysis, we also find that the use of the covariance matrix together with the smoothing parameters is able to produce the smoothest estimator.

Figure 4 and Figure 5 show changes in blood sugar levels day and night through a quadratic penalized spline model. Based on the time of treatment, we found seven patterns of changes in day blood sugar levels and six patterns of changes in night blood sugar levels. It tends to go up at a certain time, but goes down at other times. However, it increases at the end of treatment. The estimation of bi-response multi-predictor nonparametric regression model are given as follows:

$$\begin{aligned}\hat{y}_1 = {} & 506.22 - 294.25t_1 - 239.26t_1^2 - 315.59(t_1 - 5)_+^2 - 287.14(t_1 - 9)_+^2 - 295.68(t_1 - 13)_+^2 \\ & - 169.76(t_1 - 17)_+^2 - 408.61(t_1 - 21)_+^2 - 87.61(t_1 - 25)_+^2 - 27.42t_2 - 64.23t_2^2 \\ & - 29.08(t_2 - 1{,}185.83)_+^2 - 165.38(t_2 - 1{,}761.67)_+^2 + 2.48t_3 + 70.61t_3^2 - 115.23(t_3 - 133)_+^2 \\ & + 70.75(t_3 - 190.6)_+^2\end{aligned}$$

$$\begin{aligned}\hat{y}_2 = {} & 271.94 - 75.60t_1 - 87.78t_1^2 - 121.76(t_1 - 5.67)_+^2 - 107.96(t_1 - 10.33)_+^2 - 71.44(t_1 - 15)_+^2 \\ & - 172.76(t_1 - 19.67)_+^2 - 72.76(t_1 - 24.33)_+^2 - 17.49t_2 - 75.18t_2^2 - 52.34(t_2 - 955.5)_+^2 \\ & - 83.52(t_2 - 1{,}301)_+^2 + 77.55(t_2 - 1{,}646.5)_+^2 - 490.75(t_2 - 1{,}992)_+^2 + 76.22t_3 + 119.65t_3^2 \\ & + 73.04(t_3 - 133)_+^2 - 14.15(t_3 - 190.6)_+^2\end{aligned}$$

Furthermore, type 2 diabetes patients in the hospital get a calorie diet from the hospital nutritionist. It is intended to decrease blood glucose levels to normal. The standard of calories has been determined by the nutritionists based on the condition of patients. Figure 4 shows that the effect of calorie diet on changes in patients blood glucose levels in the daytime tends to decrease quadratically. There are three patterns of changes, the first is blood glucose levels decreased, then slightly increased in the middle of treatment, and the last decreased after a strict the higher calorie diet. At night, there are five patterns of changes in blood glucose levels due to the calorie diet, and a high calorie diet can make the night blood glucose levels are decreased.

Furthermore, the pattern of the daytime blood glucose levels based on the carbohydrate diet tends to fall, but in the last time at the tendency to rise when the carbohydrate diet is raised. At night, the carbohydrate diet is set by the hospital nutritionist has been able to decrease the blood glucose levels quadratically. Next, the regression curve of every predictor based on the estimation of the bi-response multi-predictor nonparametric regression model through the penalized spline estimator with the covariance matrix, is shown in Figure 4 and Figure 5.

The approximate accuracy of the bi-response multi-predictor regression model by using the penalized spline estimator on blood glucose data of type 2 diabetes patients through the error values is shown in the box-plot in Figure 6. The box-plot shows that the error of the estimated bi-response multi-predictor model has the middle value close to zero.

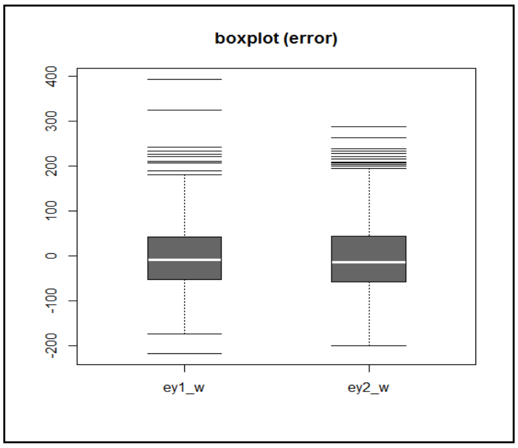

Figure 6. Boxplot of error values of estimation bi-response regression model using a penalized spline, (left) the first response, and (right) the second response

5. Conclusions

The results show that the smoothing parameters working simultaneously with the knots and weighted matrix are tools in generating an efficient regression model. Also, we find that the number of smoothing parameters involved in additive penalties is equal to the number of predictors. The performance of one smoothing parameter is able to give a minimum GCV value on estimation of bi-response multi-predictor regression model when the sample size is large. In the data analysis of blood glucose level of type 2 diabetes patients, we show the pattern of changes in blood glucose levels of patients that can provide an overview of changes in blood glucose level fluctuations of patients with type 2 diabetes.

References

Aydin, D., & Yilmaz, E. (2018). Modified spline regression based on randomly right-censored data: A comparative study. *Communications in Statistics-Simulation and Computation, 47*(9), 2587-2611.

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

Chamidah, N., & 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, 100*(9), 1433-1453.

Chamidah, N., & Rifada, M. (2016). Local linear estimation in bi-response semiparametric regression model for estimating median growth charts of children. *Far East Journal of Mathematical Sciences, 99*(8), 1233-1244.

Chamidah, N., & Saifudin, T. (2013). Estimation of children growth based on kernel smoothing in multi response nonparametric regression. *Applied Mathematical Sciences, 7*(37), 1839-1847.

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*, 012 092.

Claeskens, G., Kribovokova, T., & Opsomer, J. D. (2009). Asymptotic properties of penalized spline estimators. *Biometrika, 96*, 529-544.

Durban, M., Harezlak, J., Wand, M. P., & Carrol, R. J. (2005). Simple fitting of subject-specific curves for longitudinal data. *Statistica in Medicine, 24*, 1153-1167.

Eilers, P. H. C., & Marx, B. D. (1996). Flexible smoothing with b–splines and penalties. *Statistical Science, 11*, 89–121.

Eubank, R. L. (1999). *Nonparametric regression and spline smoothing (2nd ed.).* New York, NY: Marcel Dekke.

Festa, A., Heller, S. R., Seaquist, E., Duan, R., Hadjiyianni, I., & Fu, H. (2017). Association between mild and severe hypoglycemia in people with type 2 diabetes initiating insulin. *Journal of Diabetes and Its Complications, 31*, 1047-1052.

Green, P. J., & Silverman, B. W. (1994). *Nonparametric regression and generalized linear models (a roughness penalty approach).* New York, NY: Chapman and Hall.

Hettiaratchi, U. P. K., Ekanayake, S., Welihinda, J., & Perera, M. S. A. (2011). Glycemic and insulinemic responses to breakfast and succeeding second meal in type 2 diabetes. *International Journal of Diabetes in Developing Countries, 31* (4), 199-206.

Islamiyati, A., Fatmawati, & Chamidah, N. (2017). Fungsi goodness of fit dalam kriteria penalized spline pada estimasi regresi nonparametrik birespon untuk data longitudinal (in Indonesian). *Proceeding of National Seminar on Mathematics and Its Applications*, 216-221.

Islamiyati, A., Fatmawati, & Chamidah, N. (2018). Estimation of covariance matrix on bi-response longitudinal data analysis with penalized spline regression. *Journal of Physics: Conference Series, 979*. doi:10. 1088/1742-6596/979/1/012093

Lai, M. J., & Wang, L. (2013). Bivariate penalized spline for regression. *Statistica Sinica, 23*, 1399-1417.

Lee, M. C. T., & Oh, H. S. (2007). Robust penalized regression spline fitting with application to additively mixed modeling. *Computational Statistics, 22,* 159-171.

Lee, W., Du, Y., Wei, S., Hayes, D. N., & Liu, Y. (2012). Multiple response regression for Gaussian mixture model with known labels. *Statistical Analysis and Data Mining, 5*, 493-508.

Lee, W., & Liu, Y. (2012). Simultaneous multiple response regression and inverse covariance matrix estimation via penalized Gaussian maximum likelihood. *Journal of Multivariate Analysis, 111*, 241-255.

Lestari, B., Anggraeni, D., & Saifuddin,T. (2017). Konstruksi dan estimasi matriks kovariansi dalam model regresi nonparametric multirespon berdasarkan estimator smoothing spline untuk beberapa kasus ukuran sampel (in Indonesian). *Proceeding of National Seminar on Mathematics and Its Application*, 238-242.

Lestari, B., Anggraeni, D., & Saifuddin, T. (2018). Estimation of covariance matrix based on spline estimator in homoscedastic multi-response nonparametric regression model in case of unbalance number of observation. *Far East Journal of Mathematical Science, 108*(2), 341-355.

Lestari, B., Budiantara, I. N., Sunaryo, S., & Mashuri, M. (2010). Spline estimator in multi-response nonparametric regression model. *Journal of Mathemathics and Statistics, 6*(3), 327-332.

Lestari, B., Chamidah, N., & Saifudin, T., (2019). Estimasi Fungsi Regresi Dalam Model Regresi Nonparametrik Birespon Menggunakan Estimator Smoothing Spline dan Estimator Kernel (in Indonesian). *Jurnal Matematika, Statistika, dan Komputasi, 15*(2), 20-24.

Lestari, B., Budiantara, I. N., Sunaryo, S., & Mashuri, M. (2012). Spline smoothing for multi-response nonparametric regression model in case of heteroscedasticity of variance. *Journal of Mathemathics and Statistics, 8*(3), 337-384.

Lestari, B., Fatmawati, & Budiantara, I.N. (2017). Estimasi fungsi regresi nonparametric multirespon menggunakan reproducing kernel Hilbert space berdasarkan estimator smoothing spline. *Proceeding of National Seminar on Mathematics and Its Application*, 243-250.

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

Montoya, L. E., Ulloa, N., & 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.

Ruppert, D. (1997). Penalized spline. *Australian and New Zealand Journal of Statistics, 42*(2), 205-223.

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

Ruppert, D., & Carrol, R. J. (2000). Spatially-adaptive penalties for spline fitting. *Australian and New Zealand Journal of Statistics, 42*(2), 205-223.

Soo, Y. W., & Bates, D. M. (1996). Multi-response spline regression. *Computational Statistics and Dana Analysis, 22*(6), 619-631.

Sun, C., Li, X., Liu, L., Canet, M. J., Guan, Y., Fan, Y., & Zhou, Y. (2016). Effect of fasting time on measuring mouse blood glucose level. *International Journal of Clinical and Experimental Medicine, 9*(2), 4186-4189.

Wahba, G. (2000). *An introduction to model building with reproducing kernel Hilbert spaces*. The university of Wisconsin, Madison, WI.

Wang, Y. (1998). Smoothing spline models with correlated random errors. *Journal of The American Statistical Association, 93*, 341-348.

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

Yao, F., & Lee, T. C. M. (2006). Penalized spline models for functional principal component analysis. *Journal Royal Statistical Society - Series B, 68*(1), 3-25.