Communications in Applied Analysis **20** (2016), 335-354

THE EFFECT OF CORRELATION BETWEEN RESPONSES IN BI-RESPONSE NONPARAMETRIC REGRESSION USING SMOOTHING SPLINE FOR LONGITUDINAL DATA

ADJI ACHMAD RINALDO FERNANDES[1] AND SOLIMUN[2]

[1,2]Departement of Statistics
Faculty of Natural Sciences
University of Brawijaya
Malang, INDONESIA

ABSTRACT: Studies on nonparametric regression model that developed at this time focus on single response model approach for longitudinal data or multi-responses model approach for cross section data. This study will be developed spline estimators in biresponses nonparametric regression for longitudinal data which accommodates correlation between observations of the same subject, and correlation between each responses. Initial study focused on the spline estimators form development and optimum smoothing parameters to estimate biresponses nonparametric regression curve in longitudinal data. In the final section, the result of theoretical study was applied on simulation data with situation of differences in the level of correlation. Biresponses nonparametric regression model which involved p predictors in longitudinal data, the estimators was expressed as follows:

$$\underset{\sim}{f}(x) \quad = \mathbf{T}^{*}\underset{\sim}{d}^{*} + \mathbf{V}^{*}\underset{\sim}{c}^{*}$$

While spline estimators which met the criteria of minimizing Penalized Weighted Least Square (PWLS):

$$\min_{f_{\ell ki}\in W_2^m[a_{\ell ki},b_{\ell ki}],k=1,2;\ \ell=1,2,\ldots,p;i=1,2,\ldots,N} \left\{ M^{-1}\left(\underset{\sim}{y}-\underset{\sim}{f}\right)^{\mathrm{T}}\mathbf{W}\left(\underset{\sim}{y}-\underset{\sim}{f}\right) + \sum_{k=1}^{2}\sum_{i=1}^{N}\lambda_{ki}\left[\sum_{\ell=1}^{p}\int_{a_{\ell ki}}^{b_{\ell ki}}\left[f_{\ell ki}^{(m)}\left(x_{\ell it}\right)\right]^{2}dx_{\ell it}\right]\right\},$$

Received: September 20, 2016

were $\underset{\sim}{\hat{f}}_{\underset{\sim}{\lambda}} = \mathbf{T}^* \underset{\sim}{\hat{d}}^* + \mathbf{V}^* \underset{\sim}{\hat{c}}^* = \mathbf{A}^*_{\underset{\sim}{\lambda}} \underset{\sim}{y}$, with

$$\mathbf{A}^*_{\underset{\sim}{\lambda}} = \mathbf{T}^* \left(\mathbf{T}^{*\mathrm{T}} \hat{\mathbf{U}}^{-1} \hat{\mathbf{W}} \mathbf{T}^*\right)^{-1} \mathbf{T}^{*\mathrm{T}} \hat{\mathbf{U}}^{-1} \hat{\mathbf{W}} + \mathbf{V}^* \hat{\mathbf{U}}^{-1} \hat{\mathbf{W}}^{-1} [\mathbf{I} - \mathbf{T}^* \left(\mathbf{T}^{*\mathrm{T}} \hat{\mathbf{U}}^{-1} \hat{\mathbf{W}}^{-1} \mathbf{T}^*\right)^{-1} \mathbf{T}^{*\mathrm{T}} \hat{\mathbf{U}}^{-1} \hat{\mathbf{W}}^{-1}].$$

Optimum smoothing parameters $\underset{\sim}{\lambda}_{opt} = (\lambda_{11(opt)}, \lambda_{12(opt)}, ..., \lambda_{2N(opt)})'$ for biresponses nonparametric regression spline estimators in longitudinal data was obtained by minimizing function of GCV. Application on simulation data showed that: 1) Characteristics of exponential and trigonometric function of experiment yielded the highest level of accuracy, 2) the application of spline estimator of biresponses give the less efective in correlation level $|\rho|$between 0.1-0.3. The use of variance-covariances matrix of random error which accomodate the correlation between each responses usefullness in condition of correlation level $|\rho|$>0.3. The higher level of correlation, the higher the R^2 level obtained.

AMS Subject Classification: 65D07

1. INTRODUCTION

Regression analysis is a method used to determine the relationship between the predictor variables with the response variables. One of the uses of regression analysis is in the analysis of longitudinal data, which is a combination of cross-section data and time-series, that is the observations which are made as many as r mutually independent subjects (cross-section) with each subject is repeatedly observed in n period of time (time-series) and between observations within the same subjects which are correlated [4].

In longitudinal data(x_{it}, y_{it}), the relationship between the predictor variables x_{it} with response variables y_{it} follows the regression model can be presented as follows:

$$y_{it} = f_i(x_{it}) + \varepsilon_{it}, \quad i = 1, 2, ..., r;\ t = 1, 2, ..., n. \tag{1}$$

In the applications in various fields, problems involving more than one correlated response variables are often encountered, so the regression model developed is multi-response (for two response variables called bi-responses) as follows:

$$y_{kit} = f_{ki}(x_{kit}) + \varepsilon_{kit}, \quad k = 1, 2;\ i = 1, 2, ..., r;\ t = 1, 2, ..., n. \tag{2}$$

Equations (1) and (2) are the regression models for longitudinal data with predictor variable x as the observation time (*design time points*) [11], and f is the regression curve relationship between the predictor variables with the response variable y for to-i subject. The curve f can be approached in three ways: parametric, nonparametric, or semi-parametric. The parametric regression approach is used when it is assumed that the shape of the curve f is known, while the nonparametric regression approach is used when the shape of the curve f is unknown. On the other hand, semi-parametric regression approach is used when it is assumed that the shape of the curve is partially known, and some others are unknown [5].

Regression models can be distinguished on the ground of the total responses involved, i.e. single and biresponse regression models (if involving two responses called biresponse). Single response model, as has been widely developed by previous researchers, comprised of one single response influenced by one or more predictors. On the other hand, the biresponse model is consisted of several models with the assumption that there is a correlation between responses. Several researchers have examined the biresponse nonparametric regression, including Wang, Guo & Brown (2000), Matias (2005), and Lestari, Budiantara, Sunaryo, & Mashuri (2010) with smoothing spline approach; as well as Chamidah, Budiantara, Sunaryo, and Zain (2012) with local polynomial approach. Basically, biresponse modeling purpose is to obtain a better model than a single response modeling, given such model does not only consider the predictor's influence on response, but also the relationship between responses. Representation of the relationship between responses is usually expressed in form of variance-covariance matrix, which is used as a weighting in the model parameter estimation. The smoothing spline approach in the biresponse regression model that has been conducted by the researchers above provides in-depth studies, be it in the selection process of smoothing parameter, regression curve estimation, and the study of correlation between responses in biresponse model. Therefore, deeper studies are required which capable of reviewing comparative degree of correlation between responses in biresponse nonparametric regression model with smoothing spline approach. The degree of correlation between responses in biresponse modeling (specifically for regression model, it involves only two responses) becomes the focus in the present research. It will later develop a simulation study on biresponse nonparametric regression model using smoothing spline approach at various degrees of correlations, which has the objective to identify the level of correlation between the appropriate responses to be used in biresponse model.

The puroses of this study are: (1) To obtain the function form of nonparametric bi-responses regression, (2) to obtain the spline estimator in estimating the nonparametric bi-responses regression curve, (3) to obtain the comparison of correlation degree between responses in biresponse nonparametric regression model with smoothing spline approach. The benefits of research are: (1) to handle the modeling in which the shape of regression curve is unknown, and there is no complete past information on the relationship pattern between the predictors of responses; (2) to address the most inter-correlated response problems in nonparametric regression model; (3) it can be used as one nonparametric regression model approach in numerous real cases, both in the economic field and exact science field.

2. MATERIALS AND METHODS

Bi-responses nonparametric regression model for longitudinal data which involves r subject on n observation in each subject is as follows:

$$y_{kit} = f_{ki}(x_{it}) + \varepsilon_{kit}, \quad k = 1, 2;\ i = 1, 2, ..., r;\ t = 1, 2, ..., n, \tag{3}$$

with:

y_{kit}: $k$response variable, of $i$subject, in $t$observation

f_{ki}: regression curve corresponding to k response variable, of $i$subject

ε_{kit}: *Error* random variable from estimation results in k response and i subject

ε_{kit} variable is an error random variable assumed to be normally distributed to N-variat ($N = 2rn$), with zero mean and variance-covariance matrix $\mathbf{W}^{-1}$ is follows:

$$\mathbf{W}^{-1} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \mathbf{0} & \cdots & \mathbf{0} & \Psi_{11} & \Psi_{12} & \cdots & \Psi_{1\mathrm{r}} \\ \mathbf{0} & \boldsymbol{\Sigma}_{12} & \cdots & \mathbf{0} & \Psi_{21} & \Psi_{22} & \cdots & \Psi_{2\mathrm{r}} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Sigma}_{1\mathrm{r}} & \Psi_{\mathrm{r}1} & \Psi_{\mathrm{r}2} & \cdots & \Psi_{\mathrm{rr}} \\ \Psi_{11} & \Psi_{21} & \cdots & \Psi_{\mathrm{r}1} & \boldsymbol{\Sigma}_{21} & \mathbf{0} & \cdots & \mathbf{0} \\ \Psi_{12} & \Psi_{22} & \cdots & \Psi_{\mathrm{r}1} & \mathbf{0} & \boldsymbol{\Sigma}_{22} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \Psi_{1\mathrm{r}} & \Psi_{21} & \cdots & \Psi_{\mathrm{rr}} & \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Sigma}_{2\mathrm{r}} \end{bmatrix}. \tag{4}$$

The spline approach generally defines f_{ki} in equation (3) in form of an unknown regression curve, but f_{ki} is only assumed as smooth, in a sense of being contained in a specified function space, especially Sobolev space or written as

$$f_{ki} \in W_2^m[a_{ki}, b_{ki}]; \quad k = 1, 2;\ i \quad = 1, 2, ..., r,$$

with:

$$W_2^m[a_{ki}, b_{ki}] = \left\{ f_{ki} : \int_{a_{ki}}^{b_{ki}} [f_{ki}^{(m)}(x)]^2 dx < \infty \right\}, \tag{5}$$

for a positive integer m. Optimization Penalized Weighted Least Square (PWLS) involves weighting in form of random *error* variance-covariance matrix as has been described in equation (8). To obtain the estimate of the regression curve f_{ki} using optimization PWLS that is the completion of optimization as follows (see [5]):

$$\min_{f_{ki}\in W_2^m[a_{ki},b_{ki}],k=1,2;i=1,2,\ldots,r}\left\{N^{-1}(\underset{\sim}{y}-\underset{\sim}{f})'\mathbf{W}(\underset{\sim}{y}-\underset{\sim}{f}) + \sum_{k=1}^{2}\sum_{i=1}^{r}\lambda_{ki}\int_{a_{ki}}^{b_{ki}}(f_{ki}^{(m)}(x_{ki}))^2dx_{ki}\right\}. \quad (6)$$

The PWLS optimization in equation (6) in addition to considering the weight, also considers the use of *2r* smoothing parameter λ_{ki} as a controller between the goodness of fit (the first segment) and the roughness penalty (second segment).

The data used in this study were simulation data on the various experimental functions (Exponential, Trigonometry, and Polynomial). The research was carried out in Statistics Laboratory, Department of Mathematics, Faculty of Mathematics and Natural Sciences, Brawijaya University. Research Stages: to compare the abilities of smoothing spline estimator in biresponse nonparametric regression in variance-covariance matrix considering the correlation between response variables with variance-covariance matrix at various levels of correlation ($|\rho|$ = 0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9), with the following steps; (a) establishing biresponse nonparametric regression model with a single predictor, and N =20, 50, 100 (sample size), setting the variation of smoothing spline order polynomial m = 4; (b) setting the value $x_i, i = 1, 2, ..., N$ with $x_i \in [0, 1]$; (c) generating random error $\underset{\sim}{\varepsilon}$ with the distribution N-variat (=*2N*) with E($\underset{\sim}{\varepsilon}$) = $\underset{\sim}{0}$ and Var($\underset{\sim}{\varepsilon}$) = $\mathbf{\Sigma}$ in two conditions as follows: Condition 1: the variance-covariance matrix $\mathbf{\Sigma}_{\text{MK}}$ considers the correlation between responses. Condition 2: variance-covariance matrix *random error* $\mathbf{\Sigma}_{\text{TMK}}$ does not consider the correlation between responses; (d) obtaining the regression curve f_k from exponential, polynomials, and trigonometry functions; (e) estimating the regression curve $\hat{f}_k$ using smoothing spline estimator that minimizes PWLS; (f) comparing the results of variance-covariance matrix that considers the correlation between the response variables and variance-covariance matrix at various degrees of correlation ($|\rho|$ = 0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9) based on RMSE and R^2 values.

3. RESULT AND DISCUSSION

3.1. FORM OF BI-RESPONSES NONPARAMETRIC REGRESSION FUNCTION FOR LONGITUDINAL DATA

Assume that the data follow the bi-responses nonparametric regression models for longitudinal data:

$$\underset{\sim t}{y} = L_x \underset{\sim}{f} + \underset{\sim t}{\varepsilon}, \tag{7}$$

with $\underset{\sim t}{y} = (y_{11t}, y_{12t}, ..., y_{1rt}, y_{21t}, y_{22t}, ...y_{2rt})'$ as the response variable, L_x limited linear function, and $\underset{\sim}{f} = (f_{11}, f_{12}, ..., f_{1r}, f_{21}, f_{22}, ..., f_{2r})'$ is the inknown function and is assumed *smooth* in a sense of contained in space H. In order to obtain the function (7), Lemma is presented in the following:

Lemma. (Form of bi-responses nonparametric regression function for longitudinal data) *If given the data pairs following the* $(x_{it}, y_{1it}, y_{2it})$*bi-responses nonparametric regression model for longitudinal data as given in the equation (3),then the form of the bi-responses nonparametric regression function for longitudinal data is:* $\underset{\sim}{f}(x) = \mathbf{T}\underset{\sim}{d} + \mathbf{V}\underset{\sim}{c}$.

Proof. The function $\underset{\sim}{f} = (f_{11}, f_{12}, ..., f_{1r}, f_{21}, f_{22}, ..., f_{2r})'$ the unknown function and assumed *smooth* in the sense of being contained in the space H. Then the space H is decomposed into a direct sum of two spaces H_0 and H_1, that is$H = H_0 \oplus H_1$, with $H_0 = H_1^{\perp}$. Suppose the basis for the space H_0 is $\{\phi_{ki1}, \phi_{ki2}, \cdots, \phi_{kim}\}$ and the basis for the space H_1 is$\{\xi_{ki1}, \xi_{ki2}, \cdots, \xi_{kin}\}$, then for each function $f_{ki} \in H$can be presented individually as:

$$f_{ki} = g_{ki} + h_{ki}, \tag{8}$$

Furthermore, for each function $f_{ki} \in H$can be presented individually as:

$$f_{ki} = g_{ki} + h_{ki} = \sum_{j=1}^{m} d_{kij}\phi_{kij} + \sum_{t=1}^{n} c_{kit}\xi_{kit} = \underset{\sim ki}{\phi'}\ \underset{\sim ki}{d} + \underset{\sim ki}{\xi'}\ \underset{\sim ki}{c}. \tag{9}$$

L_x is a limited linear function in the space H and function $f_{ki} \in H$, obtained

$$L_{x_{it}} f_{ki} = L_{x_{it}}(g_{ki} + h_{ki}) = f_{ki}(x_{it}). \tag{10}$$

Based on the Riesz Representation Theorem and $L_{x_{it}}$ is a limited linear function in the space H, obtained a single value $\eta_{kit} \in H$ which is the representative of $L_{x_{it}}$, and completes the equation:

$$L_{x_{it}} f_{ki} = \langle \eta_{kit}, f_{ki} \rangle = f_{ki}(x_{it}), \quad f_{ki} \in H. \tag{11}$$

Based on the equation (17), then $f_{ki}(x_{it})$ in the equation (9) can be expressed as:

$$f_{ki}(x_{it}) = \left\langle \eta_{kit}, \underset{\sim ki}{\phi'} \underset{\sim ki}{d} \right\rangle + \left\langle \eta_{kit}, \underset{\sim ki}{\xi'} \underset{\sim ki}{c} \right\rangle. \quad (12)$$

The description of the equation (12) for $k = 1, i = 1$, obtained:

$$f_{11}(x_{1t}) = \left\langle \eta_{11t}, \underset{\sim 11}{\phi'} \underset{\sim 11}{d} \right\rangle + \left\langle \eta_{11t}, \underset{\sim 11}{\xi'} \underset{\sim 11}{c} \right\rangle, \quad t = 1, 2, \ldots, n,$$

$$\underset{\sim 11}{f}(x_1) = \begin{pmatrix} f_{11}(x_{11}) \\ f_{11}(x_{12}) \\ \vdots \\ f_{11}(x_{1n}) \end{pmatrix}$$

$$= \begin{pmatrix} d_{111}\langle \eta_{111}, \phi_{111}\rangle + \ldots + d_{11m}\langle \eta_{111}, \phi_{11m}\rangle + c_{111}\langle \eta_{111}, \xi_{111}\rangle + \ldots + c_{11n}\langle \eta_{111}, \xi_{11n}\rangle \\ d_{111}\langle \eta_{112}, \phi_{111}\rangle + \ldots + d_{11m}\langle \eta_{112}, \phi_{11m}\rangle + c_{111}\langle \eta_{112}, \xi_{111}\rangle + \ldots + c_{11n}\langle \eta_{112}, \xi_{11n}\rangle \\ \vdots \\ d_{111}\langle \eta_{11n}, \phi_{111}\rangle + \ldots + d_{11m}\langle \eta_{11n}, \phi_{11m}\rangle + c_{111}\langle \eta_{11n}, \xi_{111}\rangle + \ldots + c_{11n_1}\langle \eta_{11n}, \xi_{11n}\rangle \end{pmatrix}$$

$$= \begin{pmatrix} \langle \eta_{111}, \phi_{111}\rangle & \langle \eta_{111}, \phi_{112}\rangle & \cdots & \langle \eta_{111}, \phi_{11m}\rangle \\ \langle \eta_{112}, \phi_{111}\rangle & \langle \eta_{112}, \phi_{112}\rangle & \cdots & \langle \eta_{112}, \phi_{11m}\rangle \\ \vdots & \vdots & \vdots & \vdots \\ \langle \eta_{11n}, \phi_{111}\rangle & \langle \eta_{11n}, \phi_{112}\rangle & \cdots & \langle \eta_{11n}, \phi_{11m}\rangle \end{pmatrix} \begin{pmatrix} d_{111} \\ d_{112} \\ \vdots \\ d_{11m} \end{pmatrix}$$

$$+ \begin{pmatrix} \langle \eta_{111}, \xi_{111}\rangle & \langle \eta_{111}, \xi_{112}\rangle & \cdots & \langle \eta_{111}, \xi_{11n}\rangle \\ \langle \eta_{112}, \xi_{111}\rangle & \langle \eta_{112}, \xi_{112}\rangle & \cdots & \langle \eta_{112}, \xi_{11n}\rangle \\ \vdots & \vdots & \vdots & \vdots \\ \langle \eta_{11n}, \xi_{111}\rangle & \langle \eta_{11n}, \xi_{112}\rangle & \cdots & \langle \eta_{11n}, \xi_{11n}\rangle \end{pmatrix} \begin{pmatrix} c_{111} \\ c_{112} \\ \vdots \\ c_{11n} \end{pmatrix}$$

$$\underset{\sim 11}{f}(x_1) = \mathbf{T}_{11}\underset{\sim 11}{d} + \mathbf{V}_{11}\underset{\sim 11}{c}. \quad (13)$$

with:

$$\mathbf{T}_{11} = \begin{pmatrix} \langle \eta_{111}, \phi_{111}\rangle & \langle \eta_{111}, \phi_{112}\rangle & \cdots & \langle \eta_{111}, \phi_{11m}\rangle \\ \langle \eta_{112}, \phi_{111}\rangle & \langle \eta_{112}, \phi_{112}\rangle & \cdots & \langle \eta_{112}, \phi_{11m}\rangle \\ \vdots & \vdots & \vdots & \vdots \\ \langle \eta_{11n}, \phi_{111}\rangle & \langle \eta_{11n}, \phi_{112}\rangle & \cdots & \langle \eta_{11n}, \phi_{11m}\rangle \end{pmatrix}$$

and

$$\mathbf{V}_{11} = \begin{pmatrix} \langle \xi_{111}, \xi_{111}\rangle & \langle \xi_{111}, \xi_{112}\rangle & \cdots & \langle \xi_{111}, \xi_{11n}\rangle \\ \langle \xi_{112}, \xi_{111}\rangle & \langle \xi_{112}, \xi_{112}\rangle & \cdots & \langle \xi_{112}, \xi_{11n}\rangle \\ \vdots & \vdots & \vdots & \vdots \\ \langle \xi_{11n}, \xi_{111}\rangle & \langle \xi_{11n}, \xi_{112}\rangle & \cdots & \langle \xi_{11n}, \xi_{11n}\rangle \end{pmatrix}.$$

Then from the same way, it was obtained the result for $k = 1, 2$; $i = 2, 3, ..., r$,

$$\underset{\sim ki}{f}(x_i) = \mathbf{T}_{ki}\underset{\sim ki}{d} + \mathbf{V}_{ki}\underset{\sim ki}{c}. \tag{14}$$

$\mathbf{T}_{ki}$ is the matrix of order $n \times m$, $\underset{\sim ki}{d}$ is the vector of order m, $\mathbf{V}_{ki}$ is the matrix of order $n \times n$, $\underset{\sim ki}{c}$ is the vector of order n. Thus, the form of spline estimator $\underset{\sim}{f}(x)$is as follows:

$$\underset{\sim}{f}(x) = \mathbf{T}\underset{\sim}{d} + \mathbf{V}\underset{\sim}{c}. \tag{15}$$

$\mathbf{T}$ is the matrix of order $(2rn) \times (2rm)$ and $\mathbf{V}$ is the matrix of order $(2rn) \times (2rn)$ as follows:

$$\mathbf{T} = \begin{pmatrix} \mathbf{T}_{11} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_{12} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{T}_{2r} \end{pmatrix} \text{ and } \mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{12} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{V}_{2r} \end{pmatrix}.$$

Based on the above Lemma, it was used as the basis to obtain estimates of the spline as shown in the following theorem:

Theorem. *The spline estimator in the bi-responses nonparametric regression for longitudinal data. If given the data pairs following the bi-responses nonparametric regression models involving a single predictor variable on longitudinal data that meets the functional form of bi-responses nonparametric regression for longitudinal data as presented in the Lemma above, and with the assumption* $E(\underset{\sim t}{\varepsilon}) = \underset{\sim}{0}$ *and* $Var(\underset{\sim t}{\varepsilon}) = \mathbf{W}^{-1}$*(presented in the equation 4) so the spline estimator minimizing PWLS in the equation (6) is* $\underset{\sim\alpha}{\hat{f}} = \mathbf{A}\left(\underset{\sim}{\alpha}\right)\underset{\sim}{y}$ *with:*

$$\mathbf{A}\left(\underset{\sim}{\alpha}\right) = \mathbf{T}\left(\mathbf{T}'\mathbf{M}^{-1}\mathbf{W}\mathbf{T}\right)^{-1}\mathbf{T}'\mathbf{M}^{-1}\mathbf{W} + \mathbf{V}\mathbf{M}^{-1}\mathbf{W}[\mathbf{I} - \mathbf{T}\left(\mathbf{T}'\mathbf{M}^{-1}\mathbf{W}\mathbf{T}\right)^{-1}\mathbf{T}'\mathbf{M}^{-1}\mathbf{W}].$$

Proof. Given the equation (15), then regression curve estimator $\underset{\sim}{f}$ will be obtained. For the purposes of this estimation, *Reproducing Kernel Hilbert Space* (RKHS) approach will be used if we want to meet the estimation $\underset{\sim}{f}$ which completes the PWLS estimation:

$$\min_{\substack{f_{ki} \in H \\ k=1,2;\, i=1,2,..,r}} \left\{ \left\| \mathbf{W}\underset{\sim}{\varepsilon} \right\|^2 \right\}$$

$$= \min_{\substack{f_{ki} \in H \\ k=1,2; i=1,2,..,r}} \left\{ \left\| \mathbf{W}(\underset{\sim}{y} - \underset{\sim}{f}) \right\|^2 \right\}, \quad (16)$$

With constrain: $\| f_{ki} \|^2 < \gamma_{ki}, \gamma_{ki} \geq 0$

Then function $H = \mathrm{W}_2^m[a_{ki}, b_{ki}]$ used was *Sobolev* order-2 space which is defined as follows:

$$\mathrm{W}_2^m[a_{ki}, b_{ki}] = \left\{ f \; ; \int_{a_{ki}}^{b_{ki}} [f_{ki}^{(m)}(x_i)]^2 \, dx_i < \omega \right\},$$

with $a_{ki} \leq x_{ki} \leq b_{ki}$. Based on that space, *norm* for each function $f_{ki} \in \mathrm{W}_2^m[a_{ki}, b_{ki}]$, defined is $\| f_{ki} \|^2 = \int_{a_{ki}}^{b_{ki}} [f_{ki}^{(m)}(x_i)]^2 dx_i$.

Optimization with constrain on the equation (16) can be written as:

$$\min_{\substack{f_{ki} \in \mathrm{W}_2^m[a_{ki}, b_{ki}] \\ k=1,2; i=1,2,..,r}} \left\{ \left\| \mathbf{W} \underset{\sim}{\varepsilon} \right\|^2 \right\}$$

$$= \min_{\substack{f_{ki} \in \mathrm{W}_2^m[a_{ki}, b_{ki}] \\ k=1,2; i=1,2,..,r}} \left\{ \left\| \mathbf{W}(\underset{\sim}{y} - \underset{\sim}{f}) \right\|^2 \right\}, \quad (17)$$

with constrain

$$\int_{a_{ki}}^{b_{ki}} [f_{ki}^{(m)}(x_i)]^2 dx_i < \gamma_{ki}, \quad \gamma_{ki} \geq 0.$$

Weighed optimization with constrain (17) is equivalent to completing PWLS optimization with the equation (6). To complete the optimization, first *penalty* component must be described:

$$\int_{a_{11}}^{b_{11}} \left[f_{11}^{(m)}(x_1) \right]^2 dx_1 = \|P_1 f_{11}\|^2 = \langle P_1 f_{11}, P_1 f_{11} \rangle$$

$$= \left(\underset{\sim 11}{\xi'} \underset{\sim 11}{c} \right)' \left(\underset{\sim 11}{\xi'} \underset{\sim 11}{c} \right) = \underset{\sim 11}{c'} \left(\underset{\sim 11}{\xi} \underset{\sim 11}{\xi'} \right) \underset{\sim 11}{c} = \underset{\sim 11}{c'} \mathbf{V}_{11} \underset{\sim 11}{c}.$$

As a result

$$\lambda_{11} \int_{a_{11}}^{b_{11}} \left[f_{11}^{(m)}(x_{11}) \right]^2 dx_1 = \lambda_{11} \underset{\sim 11}{c'} \mathbf{V}_{11} \underset{\sim 11}{c}. \quad (18)$$

Using the same way, it was obtained:

$$\lambda_{ki} \int_{a_{12}}^{b_{12}} \left[f_{ki}^{(m)}(x_i) \right]^2 dx_i = \lambda_{ki} \underset{\sim ki}{c'} \mathbf{V}_{ki} \underset{\sim ki}{c}. \quad (19)$$

Based on the equation (19), *penalty* value gained:

$$\sum_{k=1}^{2}\sum_{i=1}^{r}\lambda_{ki}\int_{a_{ki}}^{b_{ki}}\left[f_{ki}^{(m)}(x_i)\right]^2 dx_i \;=\underset{\sim}{c}'\lambda\mathbf{V}\underset{\sim}{c}, \tag{20}$$

with $\lambda = diag(\lambda_{11}\mathbf{I}_{1n},\ \lambda_{12}\mathbf{I}_{1n},\ \cdots,\ \lambda_{2r}\mathbf{I}_{2n})$.

Based on the equation (18), the *Goodness of fit* on the PWLS optimization (16 can be written as:

$$N^{-1}\left(\underset{\sim}{y}-\underset{\sim}{f}\right)'\mathbf{W}\left(\underset{\sim}{y}-\underset{\sim}{f}\right)=N^{-1}\left(\underset{\sim}{y}-\mathbf{T}\underset{\sim}{d}-\mathbf{V}\underset{\sim}{c}\right)^{T}\mathbf{W}\left(\underset{\sim}{y}-\mathbf{T}\underset{\sim}{d}-\mathbf{V}\underset{\sim}{c}\right). \tag{21}$$

By combining the *goodness of fit* (18) and the *penalty* (20), the PWLS optimization can be presented in the form of:

$$\min_{\substack{\underset{\sim}{c}\in\Re^{2rn}\\ \underset{\sim}{d}\in\Re^{2rm}}}\left\{N^{-1}\left(\underset{\sim}{y}-\mathbf{T}\underset{\sim}{d}-\mathbf{V}\underset{\sim}{c}\right)'\mathbf{W}\left(\underset{\sim}{y}-\mathbf{T}\underset{\sim}{d}-\mathbf{V}\underset{\sim}{c}\right)+\underset{\sim}{c}'\lambda\mathbf{V}\underset{\sim}{c}\right\}$$

$$=\min_{\substack{\underset{\sim}{c}\in\Re^{2rn}\\ \underset{\sim}{d}\in\Re^{2rm}}}\left\{(\underset{\sim}{y}'\mathbf{W}\underset{\sim}{y}-2\underset{\sim}{d}'\mathbf{T}'\mathbf{W}\underset{\sim}{y}-2\underset{\sim}{c}'\mathbf{V}'\mathbf{W}\underset{\sim}{y}+\underset{\sim}{d}'\mathbf{T}'\mathbf{W}\mathbf{T}\underset{\sim}{d}+\underset{\sim}{d}'\mathbf{T}'\mathbf{W}\mathbf{V}\underset{\sim}{c}\right.$$

$$\left.+\underset{\sim}{c}'V'WT\underset{\sim}{d}+\underset{\sim}{c}'\ (\mathbf{V}'\mathbf{W}\mathbf{V}+N\lambda\mathbf{V})\,\underset{\sim}{c}\right\}$$

$$=\min_{\substack{\underset{\sim}{c}\in\Re^{2rn}\\ \underset{\sim}{d}\in\Re^{2rm}}}\left\{\ Q(\underset{\sim}{c},\underset{\sim}{d})\ \right\}. \tag{22}$$

The completion of the optimization (22), was obtained by partially derivating $Q(\underset{\sim}{c},\underset{\sim}{d})$ against $\underset{\sim}{c}$ then the result was equated to zero, and gave the result:

$$\begin{aligned}-2\mathbf{V}'\mathbf{W}\underset{\sim}{y}+2\mathbf{V}'\mathbf{W}\mathbf{T}\underset{\sim}{d}+2\left(\mathbf{V}'\mathbf{W}\mathbf{V}+N\lambda\mathbf{V}\right)\underset{\sim}{c}&=0,\\ -\mathbf{W}\underset{\sim}{y}+\mathbf{W}\mathbf{T}\underset{\sim}{d}+[\mathbf{W}\mathbf{V}+N\lambda\mathbf{I}]\underset{\sim}{c}&=0.\end{aligned} \tag{23}$$

Suppose given matrix $\mathbf{M}=\mathbf{W}\mathbf{V}+N\lambda\mathbf{I}$

Then the equation (23) can be written as:

$$-\mathbf{W}\underset{\sim}{y}+\mathbf{W}\mathbf{T}\underset{\sim}{d}+\mathbf{M}\underset{\sim}{c}\ =0$$

$$\mathbf{M}\underset{\sim}{c}=\mathbf{W}(\underset{\sim}{y}-\mathbf{T}\underset{\sim}{d}). \tag{24}$$

The equation (24) was multiply with $\mathbf{M}^{-1}$ it was obtained the equation:

$$\hat{\underset{\sim}{c}} = \mathbf{M}^{-1}\mathbf{W}(\underset{\sim}{y} - \mathbf{T}\underset{\sim}{d}) \tag{25}$$

Then partial derivatif against $\underset{\sim}{d}$ then the result was equated to zero, it gave the result of:

$$-\mathbf{T}'\mathbf{W}\underset{\sim}{y} + \mathbf{T}'\mathbf{W}\mathbf{T}\underset{\sim}{d} + \mathbf{T}'\mathbf{W}\mathbf{V}\underset{\sim}{c} = \underset{\sim}{0}.$$

Because the equation (25), it was obtained the equation:

$$-\mathbf{T}'\mathbf{W}\underset{\sim}{y} + \mathbf{T}'\mathbf{W}\mathbf{T}\underset{\sim}{d} + \mathbf{T}'[\mathbf{W}\mathbf{V}\mathbf{M}^{-1}]\mathbf{W}(\underset{\sim}{y} - \mathbf{T}\underset{\sim}{d}) = \underset{\sim}{0}. \tag{26}$$

Because $\mathbf{M} = \mathbf{W}\mathbf{V} + N\lambda\mathbf{I}$ so $\mathbf{V} = [\mathbf{W}]^{-1}(\mathbf{M} - N\lambda\mathbf{I})$.

As a result, it was obtained the equation:

$$\mathbf{V}\mathbf{M}^{-1} = \mathbf{W}^{-1}(\mathbf{M} - N\lambda\mathbf{I})\mathbf{M}^{-1} = \mathbf{W}^{-1}(\mathbf{I} - N\lambda\mathbf{M}^{-1})$$

By doubling the equation above with $\mathbf{W}$ it was obtained:

$$\mathbf{W}\mathbf{V}\mathbf{M}^{-1} = \mathbf{I} - N\lambda\mathbf{M}^{-1}$$

This equation was substituted in the equation (26) it was obtained the equation:

$$-\mathbf{T}'\mathbf{W}\underset{\sim}{y} + \mathbf{T}'\mathbf{W}\mathbf{T}\underset{\sim}{d} + \mathbf{T}'[\mathbf{I} - N\lambda\mathbf{M}^{-1}]\mathbf{W}(\underset{\sim}{y} - \mathbf{T}\underset{\sim}{d}) = \underset{\sim}{0}.$$

If the equation above was describedit was gained the equation:

$$\hat{\underset{\sim}{d}} = \left(\mathbf{T}'\mathbf{M}^{-1}\mathbf{W}\mathbf{T}\right)^{-1}\mathbf{T}'\mathbf{M}^{-1}\mathbf{W}. \tag{27}$$

The equation (26) was substituted into the equation (24) it was gained:

$$\hat{\underset{\sim}{c}} = \mathbf{M}^{-1}\mathbf{W}[\mathbf{I} - \mathbf{T}\left(\mathbf{T}'\mathbf{M}^{-1}\mathbf{W}\mathbf{T}\right)^{-1}\mathbf{T}'\mathbf{M}^{-1}\mathbf{W}]\underset{\sim}{y}. \tag{28}$$

Based on the Equation (27) and Equation (28), it was obtained the estimator for bi-responses nonparametric regression curve for longitudinal data involving a single predictor variable as follows:

$$\begin{aligned}\hat{\underset{\sim}{f}}_{\lambda} &= \mathbf{T}\hat{\underset{\sim}{d}} + \mathbf{V}\hat{\underset{\sim}{c}} \\ &= \{\mathbf{T}\left(\mathbf{T}'\mathbf{M}^{-1}\mathbf{W}\mathbf{T}\right)^{-1}\mathbf{T}'\mathbf{M}^{-1}\mathbf{W} \\ &\quad + \mathbf{V}\mathbf{M}^{-1}\mathbf{W}[\mathbf{I} - \mathbf{T}\left(\mathbf{T}'\mathbf{M}^{-1}\mathbf{W}\mathbf{T}\right)^{-1}\mathbf{T}'\mathbf{M}^{-1}\mathbf{W}]\underset{\sim}{y} \\ &= \mathbf{A}\left(\underset{\sim}{\lambda}\right)\}\underset{\sim}{y}.\end{aligned} \tag{29}$$

with:

$$\mathbf{A}\left(\underset{\sim}{\lambda}\right) = \mathbf{T}\left(\mathbf{T}'\mathbf{M}^{-1}\mathbf{W}\mathbf{T}\right)^{-1}\mathbf{T}'\mathbf{M}^{-1}\mathbf{W} + \mathbf{V}\mathbf{M}^{-1}\mathbf{W}[\mathbf{I} - \mathbf{T}\left(\mathbf{T}'\mathbf{M}^{-1}\mathbf{W}\mathbf{T}\right)^{-1}\mathbf{T}'\mathbf{M}^{-1}\mathbf{W}].$$

3.2. SIMULATION

The spline estimator on the equation (29) was applied in the data simulation. In this study, exponential function used was

$$f_{ki}(x) = b_1(e^{-b_2x} - 4e^{-b_3x} + 3e^{-b_4x}).$$

The Plot between predictor variable x_{ki} and response variable y_{ki} to be given to Appendix 1. From the simulation results, no form of a particular pattern (the pattern was less clear form) between the predictor variables x with response variable y and i subject (y_{ki}). The next stage is to selection of the value of m based on the highest value of coefficient of determiation (R^2), the lowest value of Generalized Cross Validation (GCV) and Minimum Square Error (MSE) as follows:

m	2	3	4
R^2	85.80%	92.94%	94.63%
MSE	1.2220	0.8442	0.7366
GCV	181.37	110.34	108.011

Table 1: The results of m value selection

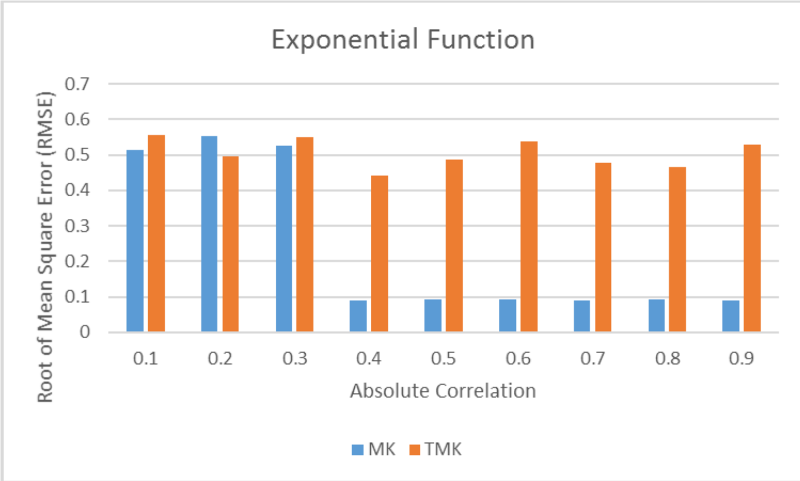
From the results of the selection value of m, it showed that the value of m =4 (cubic spline) gave the best results (the highest R^2, and the lowest GCV-MSE). The next stage is to choose the smoothing parameter based on the value of the minimum GCV. Appendix 2 presents the results of the partial smoothing parameter (λ_{ki}) based on the minimum value of GCV, by conditioning the other parameters which were constant. The optimizations results showed that the optimization of the minimum value of GCV is 108.011 for each parameter value as follows:

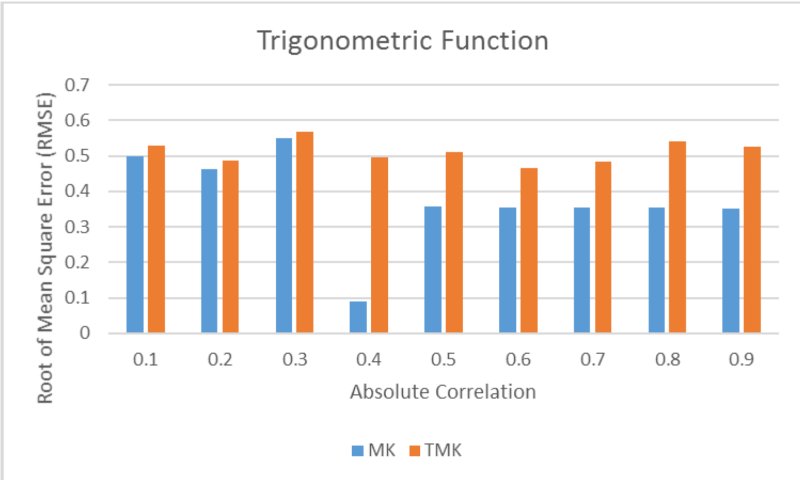
$$\lambda_{11} = 4.3; \quad \lambda_{12} = 3.0; \quad \lambda_{13} = 4.7; \quad \lambda_{21} = 4.1; \quad \lambda_{22} = 2.8; \quad \lambda_{23} = 4.7.$$

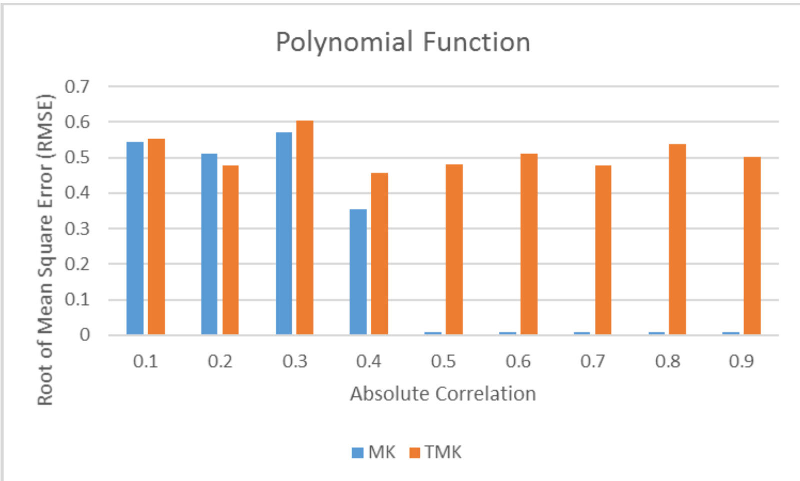
Appendix 3 shows the results of bi-responses spline estimator for longitudinal data that provide the minimum value of GCV. bi-responses spline estimator for longitudinal data on the simulated data was highly dependent on optimal smoothing parameter. Of the optimal solution, the curve can describe 94.63% of the variability of the original data.

Figure 1: Experimental Function Variation

Representation of Simulation Data. This research employed simulation data to implement biresponse nonparametric regression model with smoothing spline approach.







The simulation function employed is as follows: each simulation function uses sample size variation (N =20, 50, 100), the degree of ($|\rho| =$ 0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9), and weighting matrix variation ($\mathbf{\Sigma}_{\text{MK}}$ and $\mathbf{\Sigma}_{\text{TMK}}$). Below is the representation of simulation results using the exponential function with a sample size of N =20 and the degree of correlation of 0.5. Table 1 presents the selection of optimal smoothing parameter based on the minimum value of GCV, with order polynomial m=4. The biresponse spline estimator result plot is given in Appendix 3. The graph shows that good predictive abilities can be produced by spline estimator in which nearly all the points can be approached by the prediction curve.

The following presents the comparison abilities of spline estimator in biresponse nonparametric regression on two types of weighting matrix ($\mathbf{\Sigma}_{\text{MK}}$ dan $\mathbf{\Sigma}_{\text{TMK}}$), and at various degree of correlation ($|\rho|$ = 0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9) by considering the variation of simulation functions (exponential, trigonometric, and polynomials), as well as the sample size (N = 20, 50, 100). Meanwhile, the comparison criteria use the highest R^2 value and the lowest RMSE. The comparison result is presented in Table 3 (experimental function variation), and Table 4 (sample size variation). The application of spline estimator of biresponses give the less efective in correlation level $|\rho|$between 0.1-0.3. The use of variance-covariances matrix of random error which accomodate the correlation between each responses usefullness in condition of correlation level $|\rho|$>0.3. The higher level of correlation, the higher the R^2 level obtained.

4. CONCLUSION AND RECCOMENDATION

Based on the results of the study presented on the previous part, several things can be concluded as follows:

1. Bi-responses nonparametric regression mdel on longitudinal data on the equation $y_{kit} = f_{ki}(x_{it}) + \varepsilon_{kit}$has a function form $\underset{\sim}{f}(x) = \mathbf{T}\underset{\sim}{d} + \mathbf{V}\underset{\sim}{c}$, with

$$\mathbf{T} = \begin{pmatrix} \mathbf{T}_{11} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_{12} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{T}_{2r} \end{pmatrix},$$

$$\mathbf{T}_{ki} = \begin{pmatrix} \langle \eta_{ki1}, \phi_{ki1} \rangle & \langle \eta_{ki1}, \phi_{ki2} \rangle & \cdots & \langle \eta_{ki1}, \phi_{kim} \rangle \\ \langle \eta_{ki2}, \phi_{ki1} \rangle & \langle \eta_{ki2}, \phi_{ki2} \rangle & \cdots & \langle \eta_{ki2}, \phi_{kim} \rangle \\ \vdots & \vdots & \vdots & \vdots \\ \langle \eta_{kin}, \phi_{ki1} \rangle & \langle \eta_{kin}, \phi_{ki2} \rangle & \cdots & \langle \eta_{kin}, \phi_{kim} \rangle \end{pmatrix},$$

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{12} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{V}_{2r} \end{pmatrix},$$

$$\mathbf{V}_{ki} = \begin{pmatrix} \langle \xi_{ki1}, \xi_{ki1} \rangle & \langle \xi_{ki1}, \xi_{ki2} \rangle & \cdots & \langle \xi_{ki1}, \xi_{kin} \rangle \\ \langle \xi_{ki2}, \xi_{ki1} \rangle & \langle \xi_{ki2}, \xi_{ki2} \rangle & \cdots & \langle \xi_{ki2}, \xi_{kin} \rangle \\ \vdots & \vdots & \vdots & \vdots \\ \langle \xi_{kin}, \xi_{ki1} \rangle & \langle \xi_{kin}, \xi_{ki2} \rangle & \cdots & \langle \xi_{kin}, \xi_{kin} \rangle \end{pmatrix}.$$

2. Bi-responses nonparametric regression of the spline estimator on longitudinal data which meet the criteria of minimizing PWLS

$$\min_{f_{ki} \in W_2^m [a_{ki}, b_{ki}], k=1,2;\ i=1,2,\ldots,r} \left\{ N^{-1} \left(\underset{\sim}{y} - \underset{\sim}{f} \right)' \mathbf{W} \left(\underset{\sim}{y} - \underset{\sim}{f} \right) + \sum_{k=1}^{2} \sum_{i=1}^{r} \lambda_{ki} \int_{a_{ki}}^{b_{ki}} \left[f_{ki}^{(m)} (x_i) \right]^2 dx_i \right\}$$

is $\underset{\sim\alpha}{\hat{f}} = \mathbf{A}^* \left(\underset{\sim}{\lambda} \right) \underset{\sim}{y}$, with

$$\mathbf{A}^* \left(\underset{\sim}{\lambda} \right) = \mathbf{T}^* \left(\mathbf{T}^{*'} \mathbf{M}^{-1} \mathbf{W} \mathbf{T}^* \right)^{-1} \mathbf{T}^{*'} \mathbf{M}^{-1} \mathbf{W} + \mathbf{V}^* \mathbf{M}^{-1} \mathbf{W} [\mathbf{I} - \mathbf{T}^* \left(\mathbf{T}^{*'} \mathbf{M}^{-1} \mathbf{W} \mathbf{T}^* \right)^{-1} \mathbf{T}^{*'} \mathbf{M}^{-1} \mathbf{W}].$$

3. The simulation results show that the spline estimator can be applied to the generation of data with $m = 4$ (cubic spline) which gives the value of R^2 of 94.63%. The application of spline estimator of biresponses give the less efective in correlation level $|\rho|$between 0.1-0.3. The use of variance-covariances matrix of random error which accomodate the correlation between each responses usefullness in condition of correlation level $|\rho|>0.3$. The higher level of correlation, the higher the R^2 level obtained.

Based on the conclusions obtained, several matters suggested are as follows (1) in this research, the theory developed in biresponse nonparametric regression function is based on spline estimator using two responses. It is suggested for future researchers to develop multiple responses nonparametric regression model estimation involving more than two responses; (2) this research sets random error variance as homogeneous, hence future researchers are suggested to continue in similar case by considering heterogeneity of random error variances; (3) this research uses two responses or biresponse. In future studies, it is advisable to use more than two responses or multiple responses.

ACKNOWLEDGEMENTS

Many thanks to University of Brawijaya for financial support, and Dr. Iin Nur Chozin for provision of data.

REFERENCES

[1] A. Antoniadis, T. Sapatinas, Estimation and inference in functional mixed-effects models, *Computational Statistics and Data*, **51** (2007), 4793-4813.

[2] I.N. Budiantara, Subanar, and Z. Soejoeti, Weighted Spline Estimator, In: *Proc 51-st Session of the International Statistical Institute*, Istanbul (1997), 333-334.

[3] H. Chen and Y. Wang, A penalized spline approach to functional mixed effects model analysis, *Biometrics*, **67** (2011), 861-870.

[4] P.J. Diggle, Y.K. Liang, S.L. Zeger, *Analysis of Longitudinal Data*, Second Edition, Oxford Statistical Science Series, **13**, New York, 2006.

[5] R.L. Eubank, *Nonparametric Regression and Spline Smoothing*, Second Edition, New York, Marcel Dekker, 1999.

[6] W. Hardle, *Applied Nonparametric Regression*, New York, Cambridge University Press, 1990.

[7] B. Lestari, I.N. Budiantara, S. Sunaryo, and M. Mashuri, Spline estimator in multiresponse nonparametric regression model with unequal correlation of errors, *Mathematics and Statistics*, **6**, no. 3 (2010), 327-332.

[8] X. Lin, N. Wang, A.H. Welsh and R. Carrol, Equivalent kernels of smoothing splines in nonparametric regression for clustered/longitudinal data, *Biometrika*, **91**, no. 1 (2006), 177-193.

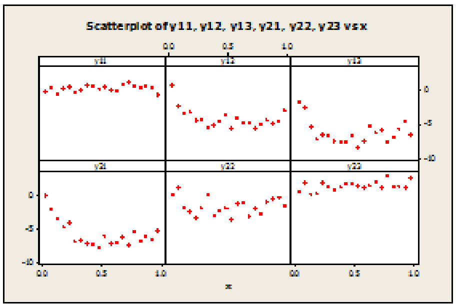
[9] G. Wahba, *Spline Models for Observational Data*, Pensylvania, SIAM, 1990.

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

[11] H. Wu and J.T. Zhang, *Nonparametric Regression Methods for Longitu-dinal Data Analysis*, New Jersey, A John-Wiley and Sons, 2006, Inc Publication.

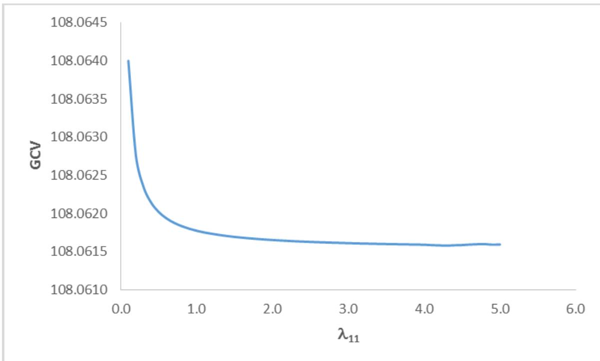
APPENDIX 1

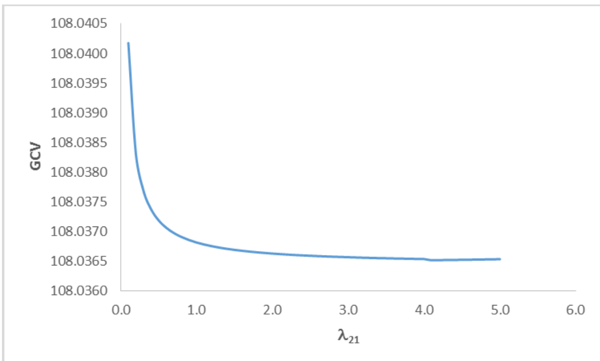
Plot of simulation data.

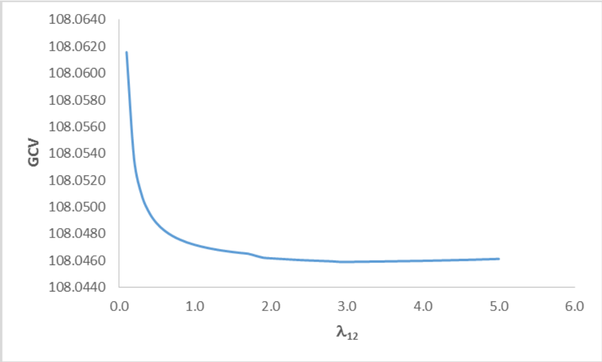


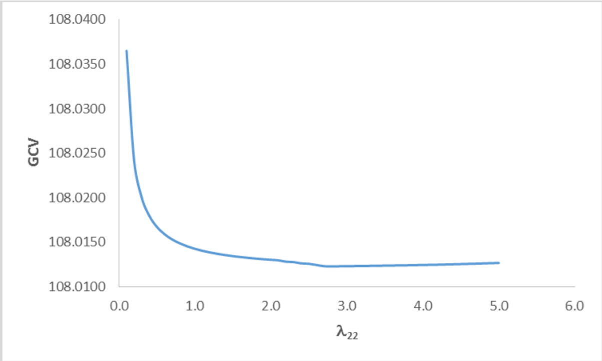
APPENDIX 2

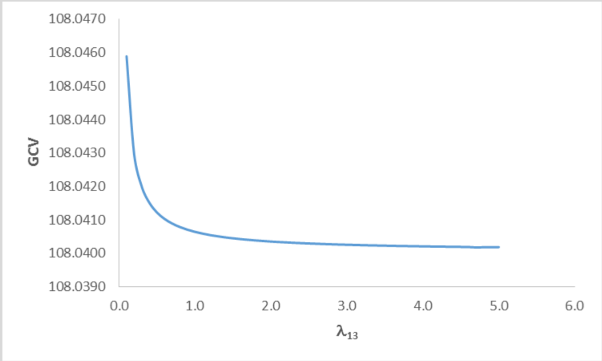
Smoothing parameter.

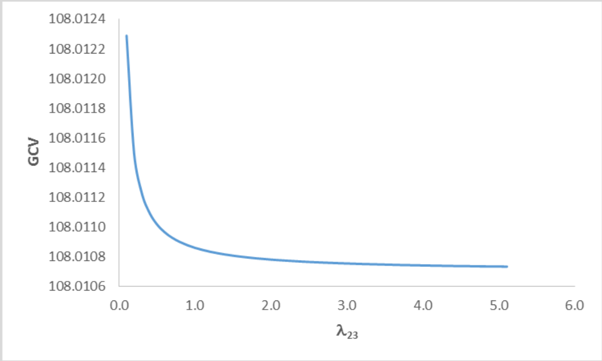






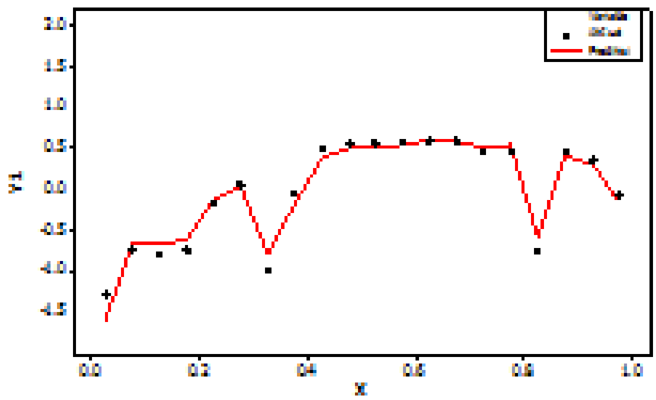


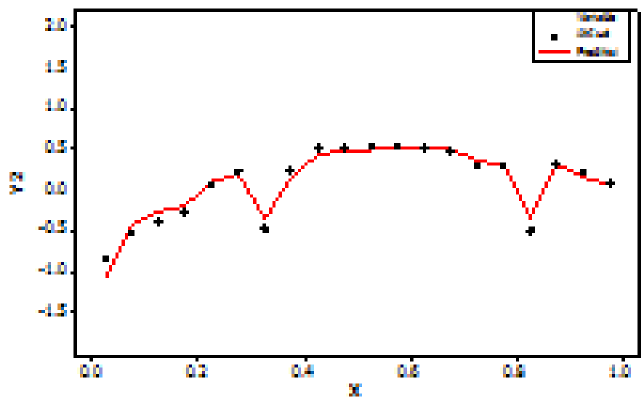




APPENDIX 3

Spline estimator of bi-responses longitudinal data.


Y1
X
0.0
0.2
0.4
0.6
0.8
1.0


Y2
X
0.0
0.2
0.4
0.6
0.8
1.0