Estimation of Covariance Matrix on Bi-Response Longitudinal Data Analysis with Penalized Spline Regression

by Fatmawati Fatmawati

Submission date: 07-Feb-2020 08:16PM (UTC+0800)

Submission ID: 1253119865

File name: 6_paper_2018_J._Phys.__Conf._Ser._979_012093.pdf (748.54K)

Word count: 3413

Character count: 16520

PAPER • OPEN ACCESS

Estimation of Covariance Matrix on Bi-Response Longitudinal Data Analysis with Penalized Spline Regression

To cite this article: A Islamiyati *et al* 2018 *J. Phys.: Conf. Ser.* **979** 012093

IOP Conf. Series: Journal of Physics: Conf. Series **979** (2018) 012093 doi:10.1088/1742-6596/979/1/012093

Estimation of Covariance Matrix on Bi-Response Longitudinal Data Analysis with Penalized Spline Regression

A Islamiyati[1], Fatmawati[2] and N Chamidah[2]

[1]Department of Mathematics, Faculty of Mathematics and Natural Sciences, Hasanuddin University, Indonesia
[2]Department of Mathematics, Faculty of Sciences and Technology, Airlangga University, Indonesia

E-mail: annaislamiyati_701@gmail.com

Abstract. The correlation assumption of the longitudinal data with bi-response occurs on the measurement between the subjects of observation and the response. It causes the auto-correlation of error, and this can be overcome by using a covariance matrix. In this article, we estimate the covariance matrix based on the penalized spline regression model. Penalized spline involves knot points and smoothing parameters simultaneously in controlling the smoothness of the curve. Based on our simulation study, the estimated regression model of the weighted penalized spline with covariance matrix gives a smaller error value compared to the error of the model without covariance matrix.

1. Introduction

The covariance matrix is usually used as a weighting in the model estimation in case of violation of assumptions such as the correlation between errors. Error correlation can occur if the data contains more than one correlated response. The covariance matrix can be assumed to be known [1] and estimated with the smoothing spline estimator [2]. However, if the covariance matrix is unknown, then it is necessary to be estimated. The covariance matrix can be estimated from the data and it may have an impact on the estimation results of the smoothing spline estimator [3]. The estimation of the covariance matrix in cross sectional data to overcome the correlation between the responses has been estimated by using the maximum likelihood method in spline regression [4]. Besides that, it also has been estimated by using the expected value of the error based on the local polynomial estimators [5], and the kernel estimators [6].

Our study analyses longitudinal data which is a combination of cross sectional and time series data. The analysis of cross sectional data that should be performed on longitudinal data will provide a different interpretation of a problem [7]. The longitudinal data consists of several subjects that are measured over time based on measurement time. Each subject is assumed to be independent, but the measurements in the subject are mutually dependent. The approach used for addressing the correlations that occur in longitudinal data can be done through principal component analysis [8], varying coefficient models [9] and mixed-effect penalized spline models [10].

The longitudinal data can consist of two responses that are assumed to be correlated called bi-response [11]. The correlation between the responses was overcome by using the covariance matrix and it is estimated through the smoothing spline estimator [12]. In this article, we use the covariance matrix as a weighted in the penalized spline criteria to solve correlations that occur in the data, both correlations between responses, or the correlation between observed data bi-response longitudinal data. The advantage of a penalized spline is the ability of the estimator to produce a smooth and fitted curve, and also the pattern change of curve regression can be seen visually. This is due to knot involvement and smoothing parameters simultaneously in controlling the smoothness of the curve. The penalized spline estimators consist goodness of fit and penalty function. The function of goodness of fit is truncated polynomials that have an optimal rate of convergence [13]. Knot points in the goodness of fit selected based on fixed methods [14]. We use the penalty function in the penalized spline criterion based on quadratic polynomials form of spline regression coefficient [15]. The covariance matrix is estimated based on the error of the nonparametric bi-response regression model based on the penalized spline estimator un-weighted. This is why the use of covariance matrix can produce to accurate estimation results. The advantages of using the weighted of the covariance matrix are shown through a longitudinal data simulation study on a correlation value of 0.9. The results show that the use of the covariance matrix in the penalized spline criteria in the case of longitudinal bi-response data provides a smaller error value.

2. Bi-response longitudinal data

Bi-response longitudinal data are longitudinal data involving two correlated responses. Longitudinal data are measured data from subjects, where each subject is measured repeatedly in a time interval. Longitudinal data are assumed to be independent among subjects, however, on the inter-observations in the same subjects, the data are mutually dependent [16]. Futhermore, if t_{ij} is the observation at the j^{th} time of the i^{th} subject, $y_{1.ij}$ is the 1st response variable at a time t_{ij} and $y_{2.ij}$ is the 2nd response variable at a time t_{ij}, then the longitudinal data bi-response was given by $(t_{ij}, y_{1.ij}, y_{2.ij})$, $j = 1,2,\ldots,m_i$, $i = 1,2,\ldots,n$, where m_i is the number of repeated measurement of the i^{th} individual. The relationship between t_{ij}, $y_{1.ij}$ and $y_{2.ij}$ can be expressed in the form of the regression model as follows.

$$y_{r\,ij} = f_r(t_{ij}) + \varepsilon_{r\,ij}, r = 1,2 \quad (1)$$

The longitudinal data structure of bi-response is shown in table 1.

Table 1. The structure of the longitudinal data bi-response to the predictor variable t_{ij}

r	i	j	t_{ij}	$y_{r\cdot ij}$	r	i	j	t_{ij}	$y_{r\cdot ij}$
1	1	1	t_{11}	$y_{1.11}$	2	1	1	t_{11}	$y_{2.11}$
		2	t_{12}	$y_{1.12}$			2	t_{12}	$y_{2.12}$
		⋮	⋮	⋮			⋮	⋮	⋮
		m_1	t_{1m_1}	$y_{1.1m_1}$			m_1	t_{1m_1}	$y_{2.1m_1}$
	⋮	⋮	⋮	⋮		⋮	⋮	⋮	⋮
	n	1	t_{n1}	$y_{1.n1}$		n	1	t_{n1}	$y_{2.n1}$
		2	t_{n2}	$y_{1.n2}$			2	t_{n2}	$y_{2.n2}$
		⋮	⋮	⋮			⋮	⋮	⋮
		m_n	t_{nm_n}	$y_{1.nm_n}$			m_n	t_{nm_n}	$y_{2.nm_n}$

Table 1 shows the structure of the longitudinal data containing two responses (r = 1,2). Each response consists of n subjects ($i = 1,2,\ldots,n$) which each subject is measured several times as much as m ($j = 1,2,\ldots,m_i$).

3. Penalized spline regression

The penalized spline is one of the estimators used in nonparametric regression in the estimation of the nonparametric regression function. Penalized spline estimator involves knot points and smoothing parameters simultaneously in controlling the smoothness of the curve. Penalty function in the penalized spline estimator takes the quadratic form of the truncated regression coefficient as recommended on the cross sectional data [17].

If the function $f_r(t_{ij})$ in equation (1) is assumed unknown and estimated through a penalized spline estimator, then it is expressed as follows.

$$f_r(t_{ij}) = \sum_{u=0}^{q} \beta_{r.u}(t_{ij})^u + \sum_{v=1}^{d} \beta_{r.(q+v)}(t_{ij} - K_v)_+^q \quad . \tag{2}$$

The equation (2) can be made in matrix form, as in (3).

$$\dot{f}(\vec{t}) = \mathbf{X}\vec{\beta}\,, \tag{3}$$

where $\vec{f}(\vec{t}) = \left[\vec{f}_1(\vec{t}), \vec{f}_2(\vec{t})\right]^T$ are an unknown form of regression function in the first and the second response.

Furthermore, the penalized spline regression model in equation (1) can be expressed to (4) based on the function at (3).

$$\vec{y} = \mathbf{X}\dot{\vec{\beta}} + \vec{\varepsilon}, \tag{4}$$

where $\dot{y}$ is a response vector containing the first and the second response. $\mathbf{X}$ is a predictor matrix in the first and the second response. $\underset{\sim}{\varepsilon}$ is an error vector in the first and the second response with $E(\vec{\varepsilon}) = \vec{0}$ and $\mathrm{Var}(\vec{\varepsilon}) = \mathbf{\Omega}$.

In the estimation of the covariance matrix, we need estimates of the nonparametric bi-response regression model un-weighted. The estimation criteria is called penalized least square (PLS) which can be expressed in the form of vector as in (5).

$$\mathrm{PLS} = \left(\vec{y} - \mathbf{X}\vec{\beta}\right)^T \left(\vec{y} - \mathbf{X}\vec{\beta}\right) + \lambda\vec{\beta}^T \mathbf{D}\vec{\beta}. \tag{5}$$

4. Estimation of covariance matrix

The assumption for error random $\varepsilon_{r.ij}$ is related to the variance of the error.

$$E\left(\varepsilon_{r.ij}^2\right) = \sigma_{r.ij}^2, E\left(\varepsilon_{r\,ij}, \varepsilon_{s.i^*j}\right) = \begin{cases} \sigma_{r.ij}\sigma_{s.ij} & ; \; i = i^* \\ 0 & ; \; i \neq i^* \end{cases} \tag{6}$$

Equation (6) occurs at $r \neq s = 1,2$. The model of bi-response nonparametric regression on longitudinal data is assumed to have a correlation between responses and to have a correlation between measurement data on the same subject. This led to the estimation of bi-response nonparametric regression model on longitudinal data using a weighted covariance matrix from $\vec{\varepsilon}$ is $\mathbf{\Omega}$.

$$\mathbf{\Omega} = \mathrm{Var}(\vec{\varepsilon}) = E\left[\vec{\varepsilon} - E(\vec{\varepsilon})\right]\left[\vec{\varepsilon} - E(\vec{\varepsilon})\right]^T = E\left(\vec{\varepsilon}\vec{\varepsilon}^T\right)\,, \tag{7}$$

where $\vec{\varepsilon} = \left(\vec{\varepsilon}_1, \vec{\varepsilon}_2\right)^T, \vec{\varepsilon}_1 = \left(\vec{\varepsilon}_{1.1}, \vec{\varepsilon}_{1.2}, \ldots, \vec{\varepsilon}_{1.n}\right)^T, \vec{\varepsilon}_2 = \left(\vec{\varepsilon}_{2.1}, \vec{\varepsilon}_{2.2}, \ldots, \vec{\varepsilon}_{2.n}\right)^T$.

Based on the assumption in (6), we can make a form the matrix of covariance as follows.

$$\mathbf{\Omega} = \begin{bmatrix} \mathbf{\Sigma}_{11.i} & \mathbf{\Sigma}_{12.i} \\ \mathbf{\Sigma}_{21.i} & \mathbf{\Sigma}_{22.i} \end{bmatrix}. \tag{8}$$

Equation (8) can also be expressed as in (9).

$$\boldsymbol{\Omega}=\begin{bmatrix} \boldsymbol{\Sigma}_{11.11} & \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{\Sigma}_{12.11} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{11.22} & \cdots & \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_{12.22} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Sigma}_{11.nn} & \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Sigma}_{12.nn} \\ \boldsymbol{\Sigma}_{21.11} & \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{\Sigma}_{22.11} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{21.22} & \cdots & \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_{22.22} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Sigma}_{21.nn} & \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Sigma}_{22.nn} \end{bmatrix} \tag{9}$$

The matrix $\boldsymbol{\Sigma}_{rr.i}$ is the matrix of the variance in the same response, $r=1,2$ with the same subject, $i=1,2,\ldots,n$ that is $\boldsymbol{\Sigma}_{rr.i}=\text{diag}\left(\boldsymbol{\Sigma}_{rr.11},\boldsymbol{\Sigma}_{rr.22},\ldots,\boldsymbol{\Sigma}_{rr.nn}\right)$. The matrix $\boldsymbol{\Sigma}_{rs.ii}$ is the matrix of covariance in the different response, $r\neq s=1,2$ with the same subject, $i=1,2,\ldots,n$ that is $\boldsymbol{\Sigma}_{rs.i}=\text{diag}\left(\boldsymbol{\Sigma}_{rs.11},\boldsymbol{\Sigma}_{rs.22},\ldots,\boldsymbol{\Sigma}_{rs.nn}\right)$. The matrix of covariance $\boldsymbol{\Omega}$ in equation (9) is used in estimating bi-response nonparametric regression model in longitudinal data with a penalized spline estimator. The matrix of covariance is used to overcome the correlation between responses. The estimation covariance matrix is related to the estimation of error variance in the measurement data. The covariance matrix at (9) is assumed to be unknown, so it estimated as set forth in Theorem 1.

Theorem 1
If the nonparametric regression model with two responses based on the un-weighted penalized spline in each sample $t_{r.ij}$ is

$$\dot{y}_{r.i}=\mathbf{X}_{r.t_{ij}}\vec{\beta}_r^{(0)}+\vec{\varepsilon}_{r.i}^{(0)}, r=1,2; i=1,2,\ldots,n ,$$

then we get

1. $\hat{\sigma}_{r.ij}^2=\dfrac{\vec{y}_{r.i}^T\mathbf{P}_{rr}\vec{y}_{r.i}-\vec{y}_{r.i}^T\mathbf{A}_r^T\mathbf{P}_{rr}\mathbf{A}_r\vec{y}_{r.i}}{\text{tr}\left(\mathbf{P}_{rr}\right)}, \mathbf{P}_{rr}=\left[\mathbf{I}-\mathbf{A}_r\right]^T\left[\mathbf{I}-\mathbf{A}_r\right]$
2. $\hat{\sigma}_{rs.ij}=\dfrac{\vec{y}_{r.i}^T\mathbf{P}_{rs}\vec{y}_{s.i^*}-\vec{y}_{r.i}^T\mathbf{A}_r^T\mathbf{P}_{rs}\mathbf{A}_s\vec{y}_{s.i^*}}{\text{tr}\left(\mathbf{P}_{rs}\right)}, \ \mathbf{P}_{rs}=\left[\mathbf{I}-\mathbf{A}_r\right]^T\left[\mathbf{I}-\mathbf{A}_s\right]$

Proof
The coefficient of regression $\vec{\beta}_r^{(0)}$ is a vector of the regression coefficient on the un-weighted penalized spline. The covariance matrix is $\boldsymbol{\Omega}$ in the criteria of PLS. The PLS criterion in (5) can be described in the matrix form.

$$\text{PLS}=\dot{y}_{r.i}^T\dot{y}_{r.i}-2\vec{\beta}_r^{(0)T}\mathbf{X}_{r.t_{ij}}^T\dot{y}_{r.i}+\vec{\beta}_r^{(0)T}\mathbf{X}_{r.t_{ij}}\mathbf{X}_{r.t_{ij}}\vec{\beta}_r^{(0)}+\vec{\lambda}_r^T\vec{\beta}_r^{(0)T}\mathbf{D}_r\vec{\beta}_r^{(0)} \tag{10}$$

If equation (10) is differentiable to $\vec{\beta}_r^{(0)}$, then obtained

$$\hat{\vec{\beta}}^{(0)}=\left(\mathbf{X}_{r.t_{ij}}^T\mathbf{X}_{r.t_{ij}}^T+\vec{\lambda}_r^T\mathbf{D}_r\right)^{-1}\mathbf{X}_{r.t_{ij}}^T\underset{\sim}{\dot{y}}_{r.i} .$$

Furthermore, the estimation of nonparametric bi-response regression function in longitudinal data through un-weighted penalized spline estimator can be expressed as follows.

$$\hat{\vec{f}}_r\left(t_{ij}\right)=\mathbf{X}_{r.t_{ij}}\hat{\vec{\beta}}^{(0)}=\mathbf{X}_{r.t_{ij}}\left(\mathbf{X}_{r.t_{ij}}^T\mathbf{X}_{r.t_{ij}}^T+\vec{\lambda}_r^T\mathbf{D}_r\right)^{-1}\mathbf{X}_{r.t_{ij}}^T\vec{y}_{r.i} . \tag{11}$$

If the matrix of the smoothing parameter is $\mathbf{A}_r=\mathbf{X}_{r.t_{ij}}\left(\mathbf{X}_{r.t_{ij}}^T\mathbf{X}_{r.t_{ij}}^T+\vec{\lambda}_r^T\mathbf{D}_r\right)^{-1}\mathbf{X}_{r.t_{ij}}^T$, then the equation (11) can be written into

$$\hat{\vec{f}}_r\left(t_{ij}\right)=\mathbf{A}_r\vec{y}_{r.i} .$$

Next, $E\left(\varepsilon_{r.ij}^{2(0)}\right)$ is obtained on the assumption of $\varepsilon_{r.ij}$ on (7).

$$E\left(\varepsilon_{r.ij}^{(0)}\varepsilon_{r.ij}^{(0)}\right)=E\left\{\dot{y}_{r.i}^{T}\left[\mathbf{I}-\mathbf{A}_{r}\right]^{T}\left[\mathbf{I}-\mathbf{A}_{r}\right]\dot{y}_{r.i}\right\}=E\left\{\dot{y}_{r.i}^{T}\mathbf{P}_{rr}\dot{y}_{r.i}\right\}, \tag{12}$$

where $\mathbf{P}_{rr}=\left[\mathbf{I}-\mathbf{A}_{r}\right]^{T}\left[\mathbf{I}-\mathbf{A}_{r}\right]$.

Equation (12) described by using the properties of the matrix and the result it is as follows.

$$\begin{aligned}E\left(\vec{y}_{r.i}^{T}\mathbf{P}_{rr}\vec{y}_{r.i}\right)&=\mathrm{tr}\left(\mathbf{P}_{rr}\sigma_{r.ij}^{2}\mathbf{I}\right)+\left(\mathbf{A}_{r}\vec{y}_{r.i}\right)^{T}\mathbf{P}_{rr}\left(\mathbf{A}_{r}\vec{y}_{r.i}\right)\\&=\sigma_{r.ij}^{2}\mathrm{tr}\left(\mathbf{P}_{rr}\right)+\dot{y}_{r.i}^{T}\mathbf{A}_{r}^{T}\left[\mathbf{I}-\mathbf{A}_{r}\right]^{T}\left[\mathbf{I}-\mathbf{A}_{r}\right]\mathbf{A}_{r}\vec{y}_{r.i}\\&=\sigma_{r.ij}^{2}\mathrm{tr}\left(\mathbf{P}_{rr}\right)+\vec{y}_{r.i}^{T}\mathbf{A}_{r}^{T}\mathbf{P}_{rr}\mathbf{A}_{r}\vec{y}_{r.i}\end{aligned} \tag{13}$$

Based on (13), the estimated variance of the error is as follows.

$$\hat{\sigma}_{r.ij}^{2}=\frac{\vec{y}_{r.i}^{T}\mathbf{P}_{rr}\vec{y}_{r.i}-\vec{y}_{r.i}^{T}\mathbf{A}_{r}^{T}\mathbf{P}_{rr}\mathbf{A}_{r}\vec{y}_{r.i}}{\mathrm{tr}\left(\mathbf{P}_{rr}\right)},\ \mathbf{P}_{rr}=\left[\mathbf{I}-\mathbf{A}_{r}\right]^{T}\left[\mathbf{I}-\mathbf{A}_{r}\right].\ \blacksquare$$

The covariance assumption of error is expressed by $E\left(\varepsilon_{r.ij}^{(0)},\varepsilon_{s.i^{*}j}^{(0)}\right)$, so $Cov\left(\varepsilon_{r.ij}^{(0)}\varepsilon_{s.i^{*}j}^{(0)}\right)=\sigma_{r.ij}\sigma_{s.ij}$ to $r=s=1,2$ and $i=i^{*}$, as for $Cov\left(\varepsilon_{r.ij}^{(0)}\varepsilon_{s.i^{*}j}^{(0)}\right)=0$ to $i\neq i^{*}$. Next, $E\left(\varepsilon_{r.ij}^{(0)},\varepsilon_{s.i^{*}j}^{(0)}\right)$ is described as follows.

$$E\left(\varepsilon_{r.ij}^{(0)},\varepsilon_{s.i^{*}j}^{(0)}\right)=E\left\{\vec{y}_{r.i}^{T}\left[\mathbf{I}-\mathbf{A}_{r}\right]^{T}\left[\mathbf{I}-\mathbf{A}_{s}\right]\vec{y}_{s.i^{*}}\right\}=E\left\{\vec{y}_{r.i}^{T}\mathbf{P}_{rs}\vec{y}_{s.i^{*}}\right\}, \tag{14}$$

where $\mathbf{P}_{rs}=\left[\mathbf{I}-\mathbf{A}_{r}\right]^{T}\left[\mathbf{I}-\mathbf{A}_{s}\right]$.

Furthermore, the equation (14) is described and the result is obtained at (15).

$$\begin{aligned}E\left(\vec{y}_{r.i}^{T}\mathbf{P}_{rs}\vec{y}_{s.i^{*}}\right)&=\mathrm{tr}\left(\mathbf{P}_{rs}\sigma_{r.ij}\sigma_{s.i^{*}j}\mathbf{I}\right)+\left(\mathbf{A}_{r}\vec{y}_{r.i}\right)^{T}\mathbf{P}_{rs}\left(\mathbf{A}_{s}\vec{y}_{s.i^{*}}\right)\\&=\sigma_{r.ij}\sigma_{s.i^{*}j}\mathrm{tr}\left(\mathbf{P}_{rs}\right)+\dot{y}_{r.i}^{T}\mathbf{A}_{r}^{T}\left[\mathbf{I}-\mathbf{A}_{r}\right]^{T}\left[\mathbf{I}-\mathbf{A}_{s}\right]\mathbf{A}_{s}\dot{y}_{s.i^{*}}\\&=\sigma_{r.ij}\sigma_{s.i^{*}j}\mathrm{tr}\left(\mathbf{P}_{rs}\right)+\vec{y}_{r.i}^{T}\mathbf{A}_{r}^{T}\mathbf{P}_{rs}\mathbf{A}_{s}\vec{y}_{s.i^{*}}\end{aligned} \tag{15}$$

Let $\sigma_{r.ij}\sigma_{s.i^{*}j}=\sigma_{rs.ij}$, then the estimated variance of the error is as follows.

$$\hat{\sigma}_{rs.ij}=\frac{\vec{y}_{r.i}^{T}\mathbf{P}_{rs}\vec{y}_{s.i^{*}}-\vec{y}_{r.i}^{T}\mathbf{A}_{r}^{T}\mathbf{P}_{rs}\mathbf{A}_{s}\vec{y}_{s.i^{*}}}{\mathrm{tr}\left(\mathbf{P}_{rs}\right)},\ \mathbf{P}_{rs}=\left[\mathbf{I}-\mathbf{A}_{r}\right]^{T}\left[\mathbf{I}-\mathbf{A}_{s}\right].\ \blacksquare$$

As a result of the theorem, estimation of covariance matrices $\mathbf{\Omega}$ on (9) in the nonparametric bi-response regression model on the longitudinal data based on the penalized spline estimator, it is as follows.

$$\hat{\mathbf{\Omega}}=\begin{bmatrix}\hat{\mathbf{\Sigma}}_{11.11} & \mathbf{0} & \cdots & \mathbf{0} & \hat{\mathbf{\Sigma}}_{12.11} & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \hat{\mathbf{\Sigma}}_{11.22} & \cdots & \mathbf{0} & \mathbf{0} & \hat{\mathbf{\Sigma}}_{12.22} & \cdots & \mathbf{0}\\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & \mathbf{0} & \cdots & \hat{\mathbf{\Sigma}}_{11.nn} & \mathbf{0} & \mathbf{0} & \cdots & \hat{\mathbf{\Sigma}}_{12.nn}\\ \hat{\mathbf{\Sigma}}_{21.11} & \mathbf{0} & \cdots & \mathbf{0} & \hat{\mathbf{\Sigma}}_{22.11} & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \hat{\mathbf{\Sigma}}_{21.22} & \cdots & \mathbf{0} & \mathbf{0} & \hat{\mathbf{\Sigma}}_{22.22} & \cdots & \mathbf{0}\\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & \mathbf{0} & \cdots & \hat{\mathbf{\Sigma}}_{21.nn} & \mathbf{0} & \mathbf{0} & \cdots & \hat{\mathbf{\Sigma}}_{22.nn}\end{bmatrix} \tag{16}$$

with $\hat{\mathbf{\Sigma}}_{rr.ii}=\mathrm{diag}\left(\hat{\sigma}_{r.i}^{2},\hat{\sigma}_{r.i}^{2},\ldots,\hat{\sigma}_{r.i}^{2}\right)$ and $\hat{\mathbf{\Sigma}}_{rs.ii}=\mathrm{diag}\left(\hat{\sigma}_{r.i}\hat{\sigma}_{s.i},\hat{\sigma}_{r.i}\hat{\sigma}_{s.i},\ldots,\hat{\sigma}_{r.i}\hat{\sigma}_{s.i}\right)$.

5. Simulation Study

The ability of covariance matrix as a weighting in the nonparametric bi-response regression model of the longitudinal data is shown through a simulation study. We use a covariance matrix corresponding to equation (19). Simulated studies were performed on the number of subjects 5, 10 and 30, while each subject was measured from 3 to 5 times the measurement. The correlation between the first and the second response is 0.9 and $\varepsilon_r \sim N(0,1)$. The model of each response to the number of knot points of 2 and the lambda is 0.5 is shown in equation (17).

$$\begin{aligned} y_{1.ij} &= f_1(t_{ij}) + \varepsilon_{1.ij},\ \varepsilon_{1.ij} \sim N(0,1), \\ y_{2.ij} &= f_2(t_{ij}) + \varepsilon_{2.ij},\ \varepsilon_{2.ij} \sim N(0,1), \end{aligned} \tag{17}$$

where $f_1 = 3.5 + 2t_{ij} - t_{ij}^2 + 2(t_{ij} - 2)_+^2 - 2(t_{ij} - 4)_+^2$ and $f_2 = 6.5 - t_{ij} + 2t_{ij}^2 + 1.5(t_{ij} - 2)_+^2 + 1.5(t_{ij} - 4)_+^2$.

We compare the error value of the simulated data with the error value after weight through the covariance matrix. It is shown through Box Plot in figure 1.

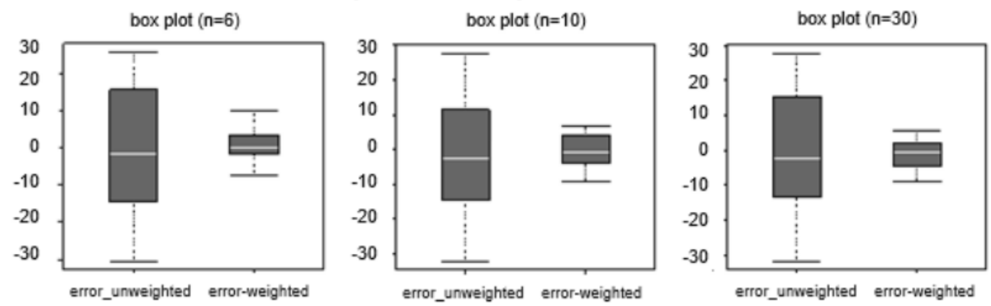

Figure 1. Plots of error of simulation data with error after weight through a matrix of covariance based on penalized spline estimators.

Figure 1 consists of three box plots for each number of subjects 5, 10 and 30. Each box plot shows an error from the model of simulated data without weighted (left) and error after weight with the covariance matrix (right). The results of the analysis show that the error of the simulated data without weighted is greater than the estimated error of the nonparametric bi-response regression model with the covariance matrix. The average error of the model of the covariance matrix rate close to zero with smaller error range. This result shows the superiority of the covariance matrix of error in overcoming the correlation that occurs in the longitudinal data of bi-response through the penalized spline estimator. The simultaneous process between the covariance matrix, knots and smoothing parameters are capable of produce estimation errors of smaller nonparametric bi-response regression models.

6. Conclusion

In this study, we described the covariance matrix estimation of the error in the bi-response longitudinal case through a penalized spline estimator. In addition, through simulation studies, we also demonstrated the ability of the covariance matrix as weight in correlated bi-response data. The result of the covariance matrix estimation through the penalized spline estimator is outlined in Theorem 1, and then it used in the simulation study. Our results have shown that the use of the covariance matrix as weight as set forth in Theorem 1 is capable of providing a smaller error value. The error value is shown through the box plot on several simulations of a number of different subjects.

Acknowledgments
Many thanks to the Directorate General of Higher Education, the Ministry of Research, Technology and Higher Education of the Republic of Indonesia passed through BPPDN grant and the University of Hasanuddin for financial support.

References

[1] Fessler J A 1991 Nonparametric fixed-interval smoothing with vector splines *IEEE Trans. Signal Processing* **39** 852-59
[2] Wahba G 1992 *Multivariate functional and operator estimation, based on smoothing splines and reproducing kernels Nonlinear Modelling and Forecasting SFI Studies in the Science of Complexity* ed by M. Casdagli and S. Eubank Vol XII (Addison Wesley Reading) pp 95-112
[3] Wang Y 1998 Smoothing spline models with correlated random errors *J. Amer. Statist. Assoc.* **93** 341-48
[4] Soo Y W and Bates D M 1996 Multi-response Spline Regression *Computational Statistics & Dana Analysis* **22** 619-31
[5] Chamidah N, Budiantara I N, Sunaryo S and Ismaini Z 2012 Designing of Child Growth Chart Based on Multi Response Local Polynomial Modeling *Journal of Mathematics and Statistics* **8** 342-347
[6] Chamidah N and Saifudin T 2013 Estimation of Children Growth Based on Kernel Smoothing in Multi response Nonparametric Regression *Applied Mathematical Sciences* **7** 1839-47
[7] Diggle P J, Liang Y K and Zeger Z L 2006 *Analysis of Longitudinal Data second edition* (New York: Oxford)
[8] Yao and Lee 2006 Penalized Spline Models for Functional Principal Component Analysis *Journal Royal Statistical Society Series B* **68** 3 – 25
[9] Liang H and Xiao Y 2006 Penalized for Longitudinal Data with an Application in AIDS Studies *Journal of Modern Applied Statistical Methods* **73** 13 – 22
[10] Chen H and Wang Y 2011 A Penalized Spline Approach to Functional Mixed Effects Model Analysis *Biometrics* **67** 861-70
[11] Fernandes A A R, Budiantara IN, Otok B W and Suhartono 2014 Spline Estimators for Bi-Responses Nonparametric Regression Model for Longitudinal Data *Applied Mathematical Sciences* **8** 5653 – 65
[12] Wang Y, Guo W and Brown M B 2000 Spline Smoothing for Bivariate Data with Application to Association Between Hormones *Statistica Sinica* **10** 377-397
[13] Claeskens G, Kribovokova T and Opsomer J D 2009 Asymptotic Properties of Penalized Spline Estimators *Biometrik* **96** 529-544
[14] Montoya L E, Ulloa N and Miller V 2014 A Simulation Study Comparing Knot Selection Methods with Equally Spaced Knot in a Penalized Regression Spline *International Journal of Statistic and Probability* **3** 96-110
[15] Ruppert D and Carrol R J 2000 Spatially-Adaptive Penalties for Spline Fitting *Australian and New Zealand Journal of Statistics* **42** 205-223
[16] Wu H and Zhang J T 2002 Local Polynomial Mixed Effects Models for Longitudinal Data *Journal of American Statistical Association* **97** 883-897
[17] Ruppert D 2002 Selecting the Number of Knot for Penalized Spline *Journal of Computational and Graphical Statistics* **11** 735-757

Estimation of Covariance Matrix on Bi-Response Longitudinal Data Analysis with Penalized Spline Regression

ORIGINALITY REPORT

11%	8%	11%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	A N R Chytrasari, Sri Haryatmi Kartiko, Danardono. "The extended nested-error regression model with penalized spline function for estimation under informative sampling", Journal of Physics: Conference Series, 2019 Publication	2%
2	www.m-hikari.com Internet Source	2%
3	thescipub.com Internet Source	1%
4	repository.uin-malang.ac.id Internet Source	1%
5	repository.unair.ac.id Internet Source	1%
6	Ayub Parlin Ampulembang, Bambang Widjanarko Otok, Agnes Tuti Rumiati, Budiasih. "Bi-responses nonparametric regression model using MARS and its properties", Applied	1%

Mathematical Sciences, 2015
Publication

7	Yilmaz, Hilal, and Neil D. Pearson. "Maximum likelihood estimation of covariance matrices with constraints on the efficient frontier", International Journal of Computational Economics and Econometrics, 2016. Publication	<1%
8	repository.unhas.ac.id Internet Source	<1%
9	Lestari. "Spline Estimator in Multi-Response Nonparametric Regression Model with Unequal Correlation of Errors", Journal of Mathematics and Statistics, 2010 Publication	<1%
10	guzlop-editoras.com Internet Source	<1%
11	Xiaoyan Lei. "Contact friction analysis with a simple interface element", Computer Methods in Applied Mechanics and Engineering, 2001 Publication	<1%
12	Evgenii L. Bashkirov. "Linear groups of degree four containing the subgroup SL2(k)⊗SL2(k)", Journal of Pure and Applied Algebra, 2011 Publication	<1%
13	Dursun Aydin, Ersin Yilmaz. "Modified spline	

regression based on randomly right-censored data: A comparative study", Communications in Statistics - Simulation and Computation, 2017
Publication | <1%

14 Zhaowen Shao, Lingtao Yu, Jun Li. "A Method for Self-Service Rehabilitation Training of Human Lower Limbs", 2019 IEEE 15th International Conference on Automation Science and Engineering (CASE), 2019
Publication | <1%

15 Hua Liang. "Modeling Antitumor Activity in Xenograft Tumor Treatment", Biometrical Journal, 06/2005
Publication | <1%

16 escholarship.org
Internet Source | <1%

17 A. E. Anwar, N. Chamidah. "Glaucoma Identification on Fundus Retinal Images Using Statistical Modelling Approach", IOP Conference Series: Materials Science and Engineering, 2019
Publication | <1%

18 pertambangan.fst.uinjkt.ac.id
Internet Source | <1%

19 Saif Benjaafar, Mohsen ElHafsi, Francis de Véricourt. "Demand Allocation in Multiple-Product, Multiple-Facility, Make-to-Stock | <1%

Systems", Management Science, 2004
Publication

20	"The Work of Raymond J. Carroll", Springer Nature, 2014 Publication	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		

Estimation of Covariance Matrix on Bi-Response Longitudinal Data Analysis with Penalized Spline Regression

GRADEMARK REPORT

FINAL GRADE

/0

GENERAL COMMENTS

Instructor

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8