IOP Conference Series: Materials Science and Engineering

PAPER • OPEN ACCESS

The Estimation Function Approach Smoothing Spline Regression Analysis for Longitudinal Data

To cite this article: Risnawati *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **546** 052064

IOP Conf. Series: Materials Science and Engineering **546** (2019) 052064 doi:10.1088/1757-899X/546/5/052064

The Estimation Function Approach Smoothing Spline Regression Analysis for Longitudinal Data

Risnawati,[1] Adji Achmad Rinaldo Fernandes,[1] Nurjannah[1*]

[1)]Departement of Statistics, Faculty of Mathematics and Natural Sciences, Brawijaya University, Indonesia.

[*]Corresponding author: fernandes@ub.ac.id

Abstract. It is given the paired data following a nonparametric regression model for longitudinal data. The regression curve is approached by the smoothing spline function. The smoothing spline is a function that is able to map the data well and has a small error variance. This current study aims to obtained from completing PWLS (Penalized Weighted Least Square) optimaziton. Besides, the GCV method is used to select the smoothing parameter.
Keywords: nonparametric, longitudinal, smoothing spline, PWLS

1. Introduction

A method used in statistics in order to determine the pattern of the relationships between the two variables is regression analysis. If the form of the regression curve is known, then the approach uses the parametric regression model [1]. However, in real life, there is a pattern of relationships that are unknown is shape. The appropriate approach for the data pattern that is not known/unknown in the form of regression curve is nonparametric regression [2]. The nonparametric regression has the ability to find a form of a regression curve pattern that is not known yet. This ability is supported by the existence of parameters in each type of nonparametric regression method which estimates the regression curve to be more flexible. Therefore, the data is expected to find its own regression curve estimation form without being influenced by subjectivity of researcher [3]. Some of the models used by researchers are the kernel, spline, Fourier series and wavelet.

Spline has a very good ability in handling the data shoes behavior changes at sub-specified intervals [3]. The approach to estimating spline functions in nonparametric regression models is basically divided into two forms, namely penalized spline [4] and smoothing spline [5]. The uses of the regression form on the penalized spline approach requires accuracy in determining the number of knots and location of knots and location of knots, whereas smoothing spline is not required for the selection of knots, because the estimation of functions is based on the criteria of model accuracy and the size of smoothness of the curve set by smoothing parameters. This shows that the smoothing spline approach has better flexibility than the penalized spline approach.

The study on smoothing spline estimators actually has been carried out [6]. However, the estimators has several weaknesses in which the estimator is only able to handle cross-section data and is not able to handle longitudinal data so that it cannot be used to obtain a model for each subject. A type of data used in regression analysis is longitudinal data. The longitudinal data is a combination of cross section data and time series. In relation to the description above, this study will develop a smoothing spline estimator in nonparametric regression for longitudinal data. This model is expected to be able to deal with the weaknesses of previous research. This study will solve several objectives, namely, to obtain the form of smoothing spline estimator and to select the smoothing parameter selection in order to estimate the nonparametric regression curve in the longitudinal data.

2. Theoretical Framework

2.1. Longitudinal Data

On the other hand, the longitudinal data is also called as repeated measurement data, which means that the observations carried out on N subjects that are mutually independent with each subject being observed directly in a period of time. In longitudinal data, the observation in the same subject is dependent or correlated [7]. The longitudinal data has different characteristics when it is compared to time series data and cross section data. Moreover, in longitudinal data, the number of time series is relatively short because it allows only two or three different measurement times for each subject [8]. Besides, the observation between each subject are assumed to be independent of each other, but observation in the same subject are dependent [6].

2.2. Reproducing Kernel Hilbert Space (RKHS)

The reproducing kernel from a Hilbert $\mathcal{H}$ space is a R function defining on $[a,b]\times[a,b]$ so that for each particular point $x \in [a,b]$, applies for $R_x \in \mathcal{H}$, by formulated

$$R_x(x_s) = R(x_t, x_s) \text{ and } f(t) = \langle R_{x_t}, f \rangle, f \in \mathcal{H} \tag{1}$$

RKHS is a Hilbert space of real value functions on $[a,b]$ with the properties for each $x \in [a,b]$, functional $\mathrm{L}_x f = f(x)$ which is a limited linear functional, meaning that are real numbers such the following

$$|\mathrm{L}_x f| = |f(x)| \leq d\|f\| \tag{2}$$

The reproducing kernel from $\mathcal{H}$ is a R function defining on $[a,b]\times[a,b]$ so that for each point remains $x \in [a,b]$ which can be formulated:

$$R_x \in \mathcal{H},\ R_x(x_s) = R(x_s, x_t) \text{ and } \mathrm{L}_x f = \langle R_x, f \rangle = f(x), \quad f \in \mathrm{H} \tag{3}$$

If $\mathcal{H}$ of an RKHS, then it can be decomposed into $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$ with $\mathcal{H}_0 = \mathcal{H}_1^{\perp}$ and $\mathcal{H}_0, \mathcal{H}_1$ for each sub-space in $\mathcal{H}$ [6].

2.3. Nonparametric Spline Regression for Longitudinal Data

The nonparametric regression approach assuming the pattern of relations between response variables and predictors can be described in a particular function. However, in the application obtaining that function is precisely very difficult. The approach that should be used in this condition is a nonparametric regression approach [2]. It states the relationship between an one predictor with an one response for longitudinal data involving N subjects on the T observation of each subject, following the regression model:

$$y_{it} = f_i(x_{it}) + e_{it}; i = 1,2,\ldots,N; \mathrm{t} = 1,2,\ldots,\mathrm{T} \tag{4}$$

y_{it} is the response to the i-subject and the t-time observation, x_{it} is the i-subject predictor and the t-time observation, f_i is the regression curve for predictor relationship with the responses to i-subject, N is the number of subjects, T is the number of observation on each subject, moreover, e_{it} is the random error on i-subject and t-time observation.

In the nonparametric regression model for the longitudinal data, random error $\underset{\sim}{e} = (e_{11}, e_{12}, ..., e_{NT})$, is assumed to be normally distributed NT- variat with the mean $E(\underset{\sim}{e}) = 0$ (vector sized NT) and the matrix of variance-covariance $Var(\underset{\sim}{e}) = \Sigma$ (matrix sized $NT \times NT$) as follow:

$$\Sigma = \begin{bmatrix} \Sigma_1 & 0 & \cdots & 0 \\ 0 & \Sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma_N \end{bmatrix}_{(NT)\times(NT)} \tag{5}$$

The matrix form Σ can be simplified into a sub-matrix can be presented as follows:

$$\Sigma_{11} = \begin{bmatrix} \sigma_{11}^2 & \rho\sigma_{11}^2 & \rho 2\sigma_{11}^2 & \cdots & \rho^{T-1}\sigma_{11}^2 \\ \rho\sigma_{11}^2 & \sigma_{12}^2 & \rho\sigma_{11}^2 & \cdots & \rho^{T-2}\sigma_{11}^2 \\ \rho^2\sigma_{11}^2 & \rho\sigma_{11}^2 & \sigma_{13}^2 & \cdots & \rho^{T-3}\sigma_{11}^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1}\sigma_{11}^2 & \rho^{T-2}\sigma_{11}^2 & \rho^{T-3}\sigma_{11}^2 & \cdots & \sigma_{1T}^2 \end{bmatrix}_{T\times T}$$

The homogeny assumption $(\sigma_{11}^2 = \sigma_{12}^2 = \sigma_{13}^2 ... = \sigma_{1T}^2 = \sigma^2)$, the structure of the covariance matrix is denoted by Autoregressive (AR) structure which has two parameters to define all variances and co-variances, namely the variant parameter σ^2 and correlation parameter ρ.

The coefficient autocorrelation of i- subject on the v- time lag is symbolized $r_{y_{i(v)}}$ which is presented as follows [10].

$$r_{y_{i(v)}} = \frac{\sum_{t=1}^{T-v} (y_{it} - \bar{y}_i)(y_{i(t+v)} - \bar{y}_i)}{\sum_{t=1}^{T} (y_{it} - \bar{y}_i)^2}$$

The description of v is the index of time lag

$$\Sigma_{ii} = \begin{bmatrix} \sigma_i^2 & \rho\sigma_i^2 & \rho^2\sigma_i^2 & \cdots & \rho^{T-1}\sigma_i^2 \\ \rho\sigma_i^2 & \sigma_i^2 & \rho\sigma_i^2 & \cdots & \rho^{T-2}\sigma_i^2 \\ \rho^2\sigma_i^2 & \rho\sigma_i^2 & \sigma_i^2 & \cdots & \rho^{T-3}\sigma_i^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1}\sigma_i^2 & \rho^{T-2}\sigma_i^2 & \rho^{T-3}\sigma_i^2 & \cdots & \sigma_i^2 \end{bmatrix}_{T\times T}, i = 1, 2, ...N$$

In the structure of AR (1), residual variance σ^2 is constantly assumed and the co-variance residual is assumed $\sigma^2\rho, \sigma^2\rho^2$ etc [7]. The sub-matrix element Σ is the diagonal matrix value of observed variance in the response, and between observation in the response which are mutually independent, marked with a non-diagonal element in the matrix value of 0 [8].

2.4. The Optimization Completion of Penalized Weighted Least Square (PWLS)

The estimation curve completion f_i of the longitudinal data for the formula (4) uses penalized weighted least square (PWLS) which involves the weights in the form of the inverse covariance matrix random error symbolized Σ. To obtain an estimate of the regression curve f_i, it uses PWLS which means that the optimization completion as follows [9]:

$$\text{Min}\left\{ M^{-1}(\underset{\sim}{y} - \underset{\sim}{f})^{\mathrm{T}}\Sigma^{-1}(\underset{\sim}{y} - \underset{\sim}{f}) + \sum_{i=1}^{N}\lambda_i \int_{a_i}^{b_i} (f_i^{(m)}(x_{it}))^2 dx_{it} \right\} \tag{6}$$

with $m = NT, \underset{\sim}{y} = (y_{11}, y_{12}, ..., y_{NT})^T$ and $\underset{\sim}{f} = (f_1(x_{11}), f_2(x_{12}), ..., f_N(x_{NT}))^T$ the optimization of PWLS on the formula (6) is also considering the use of smoothing parameter l_i as the controller between the goodness of fit and roughness penalty. The optimization reducing of PWLS above becomes goodness of fit $R(f)$ and roughness penalty $j(f)$ [12]:

$$R(f) = M^{-1}(\underset{\sim}{y} - \underset{\sim}{f})^{\mathrm{T}} \Sigma^{-1} (\underset{\sim}{y} - \underset{\sim}{f}) \tag{7}$$

$$J(f) = \sum_{i=1}^{N} \lambda_{ji} \int_{a_{ji}}^{b_{ji}} (f_{ji}^{(m)}(x_{jit}))^2 dx_{jit} \tag{8}$$

2.5. The Smoothing Parameter in the Nonparametric Regression

In order to obtain the optimal spline estimator and smoothness of the curve, it is very dependent on the selection of refining parameters l . Choosing the smoothing parameter l optimal in nonparametric spline regression in longitudinal data is based on the generalized cross validation (GCV) method. The optimal parameter estimation is obtained by minimizing the GCV function so that the regression curve estimation obtained. The choice of optimal l by the GCV method is defined as follows with $M = NT$ [3]. After getting the optimal parameter estimation by minimizing the GCV function, the regression curve estimation will be obtained.

3. Research Methodology

The step conducted in the research related to obtaining the estimation of smoothing spline for the longitudinal data are illustrated as follows:

1) The paired data are given (x_{it}, y_{it}).
2) The regression of the nonparametric model for longitudinal data is provided.
3) Approaching $f_i(x_{it})$ by having smoothing spline degree m with the smoothing coefficient l_i
4) Deciding the random error $\underset{\sim}{e}_{it}$ normal distributed M-variat with $(M = NT), \mathrm{E}(\underset{\sim}{e}) = 0$ and $Var(\underset{\sim}{e}) = 0$.
5) Obtaining the form of the regression function f and determining the matrix design of $\mathbf{T}$ and $\mathbf{V}$ by having RKHS approach.
6) To estimate the regression function $\hat{f}$ by minimizing PWLS.
7) To estimate the variance-covariance matrix S thus, it is obtained the matrix estimation of $\hat{\mathrm{S}}$
8) Accomplishing to choose the optimal smoothing parameter.

4. Result and Discussion

4.1. The Estimator Spline to Estimate Nonparametric Regression Curves in Longitudinal Data

If the paired data are given $(x_{it}, y_{it}); i = 1,2,...,N; t = 1,2,...,T$ following the model nonparametric regression for longitudinal data as given in the formula [9] as follows:

$$y_{it} = f_i(x_{it}) + e_{it}; i = 1,2,...,N; \mathrm{t} = 1,2,...,\mathrm{T} \tag{9}$$

Then, the form of nonparametric function for longitudinal data is described as:

$$\underset{\sim}{f} = \mathbf{T}\underset{\sim}{d} + \mathbf{V}\underset{\sim}{c}$$

Proof:

Suppose that in the formula [9], it is rewritten as $y_{it} = \mathrm{L}_x f_i + e_{it}$

The function $\underset{\sim}{f} = (f_1(x_{11}), f_2(x_{12}), ..., f_N(x_{NT}))^{\mathrm{T}}$ is unknown regression curve and assumed to be smooth in the sense contained in the $\mathcal{H}$ space. Then $\mathcal{H}$ is decomposed to be
$\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$ with $\mathcal{H}_0 = \mathcal{H}_1^{\perp}$
For example, basis $\mathcal{H}_0$ and $\mathcal{H}_1$ for each is follows:
Basis $\mathcal{H}_0 = \{\phi_{i1}, \phi_{i2}, ..., \phi_{im}\}$ with m is the order of polynomial spline
Basis $\mathcal{H}_1 = \{x_{i1}, x_{i2}, ..., x_{iT}\}$ and T the number of the time of the observation
Then the function of $f_i \in \mathcal{H}$ to be $f_i = g_i + h_i$, having the function $g_i \in \mathcal{H}_0$ and $h_i \in \mathcal{H}_1$ for each is as follows:

$$g_i = \sum_{j=1}^{m} d_{ij}\phi_{ij} = \underset{\sim}{\phi}_i^{\mathrm{T}} \underset{\sim}{d}_i \text{ and } h_i = \sum_{t=1}^{T} c_{it}\xi_{it} = \underset{\sim}{\xi}_i^{\mathrm{T}} \underset{\sim}{c}_i \text{ in which } d_{ij} \;\&\; c_{it} \text{ it is constant.}$$

Therefore, $f_i = g_i + h_i = \underset{\sim}{\phi}_i^{\mathrm{T}} \underset{\sim}{d}_i + \underset{\sim}{x}_i^{\mathrm{T}} \underset{\sim}{c}_i$ (10)

By describing L_x linear functional limited to $\mathcal{H}$ space and the function of $f_i \in \mathcal{H}$ and $h_{it} \in \mathcal{H}$, a, then the formula (10) can be presented as follows:

$$\mathrm{L}_x f_i = \langle h_{it}, f_i \rangle = f_i(x_{it}),\; f_i \in \mathcal{H} \tag{11}$$

Based on the formula (10) and (11), then $f_i(x_{it})$ it can be described as:

$$\mathrm{L}_x f_i = \langle h_{it}, f_i \rangle = \langle h_{it}, \underset{\sim}{\phi}_i^{\mathrm{T}} \underset{\sim}{d}_i + \underset{\sim}{x}_i^{\mathrm{T}} \underset{\sim}{c}_i \rangle = \langle h_{it}, \underset{\sim}{\phi}_i^{\mathrm{T}} \underset{\sim}{d}_i \rangle + \langle h_{it}, \underset{\sim}{x}_i^{\mathrm{T}} \underset{\sim}{c}_i \rangle \tag{12}$$

Next, the formula (9), for $i = 1, t = 1$ then it is obtained:

$$\begin{aligned} f_1(x_{11}) &= \langle \eta_{11}, \underset{\sim}{\phi}_1^T \underset{\sim}{d}_1 \rangle + \langle \eta_{11}, \underset{\sim}{\xi}_1^T \underset{\sim}{c}_1 \rangle = \langle \eta_{11}, \phi_{11} d_{11} \rangle + \langle \eta_{11}, \phi_{12} d_{12} \rangle + ... + \langle \eta_{11}, (\xi_{11} c_{1T}) \rangle \\ &= d_{11} \langle \eta_{1T}, \phi_{11} \rangle + d_{12} \langle \eta_{11}, \phi_{12} \rangle + ... + c_{1T} \langle \eta_{11}, \xi_{1T} \rangle \end{aligned}$$

If the process is conducted in a similar way, for $t = T$, then it is obtained:

$$f_1(x_{1T}) = \langle \eta_{1T}, \underset{\sim}{\phi}_1^T \underset{\sim}{d}_1 \rangle + \langle \eta_{1N}, \underset{\sim}{\xi}_1^T \underset{\sim}{c}_1 \rangle$$

as a result, the vector from the function of f_1 in the form of:

$$f_1 = \begin{pmatrix} f_1(x_{11}) \\ f_1(x_{12}) \\ \vdots \\ f_1(x_{1T}) \end{pmatrix} = \begin{pmatrix} d_{11}\langle \eta_{11}, \phi_{11} \rangle + ... + d_{1m}\langle \eta_{11}, \phi_{1m} \rangle + c_{11}\langle \eta_{11}, \xi_{11} \rangle + ... + c_{1T}\langle \eta_{11}, (\xi_{1T}) \rangle \\ d_{11}\langle \eta_{12}, \phi_{11} \rangle + ... + d_{1m}\langle \eta_{12}, \phi_{1m} \rangle + c_{11}\langle \eta_{12}, \xi_{11} \rangle + ... + c_{1T}\langle \eta_{12}, (\xi_{1T}) \rangle \\ \vdots \\ d_{11}\langle \eta_{iT}, \phi_{11} \rangle + ... + d_{1m}\langle \eta_{1T}, \phi_{1m} \rangle + c_{11}\langle \eta_{1T}, \xi_{11} \rangle + ... + c_{1T}\langle \eta_{1T}, (\xi_{1T}) \rangle \end{pmatrix}$$

$$= \begin{pmatrix} \langle \eta_{11}, \phi_{11} \rangle & \langle \eta_{11}, \phi_{12} \rangle & \cdots & \langle \eta_{11}, \phi_{1m} \rangle \\ \langle \eta_{12}, \phi_{11} \rangle & \langle \eta_{12}, \phi_{12} \rangle & \cdots & \langle \eta_{12}, \phi_{1m} \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \eta_{1T}, \phi_{11} \rangle & \langle \eta_{1T}, \phi_{12} \rangle & \cdots & \langle \eta_{1T}, \phi_{1m} \rangle \end{pmatrix} \begin{pmatrix} d_{11} \\ d_{12} \\ \vdots \\ d_{1m} \end{pmatrix} + \begin{pmatrix} \langle \eta_{11}, \xi_{11} \rangle & \langle \eta_{11}, \xi_{12} \rangle & \cdots & \langle \eta_{11}, \xi_{1T} \rangle \\ \langle \eta_{12}, \xi_{11} \rangle & \langle \eta_{12}, \xi_{12} \rangle & \cdots & \langle \eta_{12}, \xi_{1T} \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \eta_{1T}, \xi_{11} \rangle & \langle \eta_{1T}, \xi_{12} \rangle & \cdots & \langle \eta_{1T}, \xi_{1T} \rangle \end{pmatrix} \begin{pmatrix} c_{11} \\ c_{12} \\ \vdots \\ c_{1T} \end{pmatrix}$$

$$\underset{\sim}{f}_1 = \mathbf{T}_1 \underset{\sim}{d}_1 + \mathbf{V}_1 \underset{\sim}{c}_1 \tag{13}$$

because and its data is the longitudinal data, then based on formula [8] then the t-observation is affected by the observation of $(t-1), (t-2), \text{etc}$. Therefore, the $\mathbf{V}_1$ matrix in the form of triangle matrix is presented as follows:

$$\mathbf{V}_1 = \begin{bmatrix} \langle \xi_{11}, \xi_{i1} \rangle & 0 & \cdots & 0 \\ \langle \xi_{12}, \xi_{11} \rangle & \langle \xi_{12}, \xi_{12} \rangle & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \langle \xi_{1T}, \xi_{11} \rangle & \langle \xi_{1T}, \xi_{12} \rangle & \cdots & \langle \xi_{1T}, \xi_{1T} \rangle \end{bmatrix}_{T \times T}$$

with the description $\mathbf{T}_1$ is a sized matrix $T \times m$, $\underset{\sim}{d}_1$ and is a vector sized m, $\mathbf{V}_1$ and is sized matrix $T \times T$, $\underset{\sim}{c}_1$ and is vector sized T

Moreover, by taking the similar way for $i = 1, 2, \ldots, N$ is describe as follows:

$$f_i = \begin{pmatrix} f_i(x_{i1}) \\ f_1(x_{i2}) \\ \vdots \\ f_i(x_{iT}) \end{pmatrix} = \begin{pmatrix} d_{i1} \langle \eta_{i1}, \phi_{i1} \rangle + \ldots + d_{im} \langle \eta_{i1}, \phi_{im} \rangle + c_{i1} \langle \eta_{11}, \xi_{11} \rangle + \ldots + c_{iT} \langle \eta_{11}, (\xi_{iT}) \rangle \\ d_{i1} \langle \eta_{i2}, \phi_{i1} \rangle + \ldots + d_{im} \langle \eta_{i2}, \phi_{im} \rangle + c_{i1} \langle \eta_{12}, \xi_{11} \rangle + \ldots + c_{iT} \langle \eta_{12}, (\xi_{iT}) \rangle \\ \vdots \\ d_{i1} \langle \eta_{iT}, \phi_{i1} \rangle + \ldots + d_{im} \langle \eta_{iT}, \phi_{im} \rangle + c_{i1} \langle \eta_{1T}, \xi_{11} \rangle + \ldots + c_{iT} \langle \eta_{1T}, (\xi_{iT}) \rangle \end{pmatrix}$$

$$\underset{\sim}{f}_i = \mathbf{T}_i \underset{\sim}{d}_i + \mathbf{V}_i \underset{\sim}{c}_i \tag{14}$$

$\mathbf{T}_i$ is sized matrix $T \times m$, $\underset{\sim}{d}_i$ is sized matrix $T \times T$ and $\underset{\sim}{c}_i$ is sized vector T.

Therefore, the estimator spline form of $\underset{\sim}{f}$ can be illustrated as follows:

$$\underset{\sim}{f} = \begin{pmatrix} \mathbf{T}_1 \underset{\sim}{d}_1 + \mathbf{V}_1 c_1 \\ \mathbf{T}_2 \underset{\sim}{d}_2 + \mathbf{V}_2 c_2 \\ \vdots \\ \mathbf{T}_N \underset{\sim}{d}_N + \mathbf{V}_N c_N \end{pmatrix} = \begin{bmatrix} \mathbf{T}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{T}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{T}_N \end{bmatrix} \begin{pmatrix} \underset{\sim}{d}_1 \\ \underset{\sim}{d}_2 \\ \vdots \\ \underset{\sim}{d}_N \end{pmatrix} + \begin{bmatrix} \mathbf{V}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{V}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{V}_N \end{bmatrix} \begin{pmatrix} \underset{\sim}{c}_1 \\ \underset{\sim}{c}_2 \\ \vdots \\ \underset{\sim}{c}_N \end{pmatrix}$$

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

❖ **Proven**

Furthermore, to obtain the estimation of the regression curve f by using penalized weighted least square optimization, it is presented as follows:

$$\text{Min}\left\{ M^{-1}(\underset{\sim}{y} - \underset{\sim}{f})^{\mathrm{T}} \Sigma^{-1} (\underset{\sim}{y} - \underset{\sim}{f}) + \sum_{i=1}^{N} \lambda_i \int_{a_i}^{b_i} (f_i^{(m)}(x_{it}))^2 \, dx_{it} \right\}$$

Assumption $\mathrm{E}(\underset{\sim}{\varepsilon}) = 0$ and $\mathrm{Var}(\underset{\sim}{\varepsilon}) = \Sigma$, then

$$\hat{\underset{\sim}{f}}_\lambda = \mathbf{A}_\lambda \underset{\sim}{y}$$

With the description $\mathbf{A}_\lambda \underset{\sim}{y} = \mathbf{T}\left(\mathbf{T}^{\mathrm{T}}\hat{\mathbf{U}}^{-1}\hat{\mathbf{\Sigma}}^{-1}\mathbf{T}\right)^{-1} \mathbf{T}^{\mathrm{T}}\hat{\mathbf{U}}^{-1}\hat{\mathbf{\Sigma}}^{-1}\underset{\sim}{y} + \mathbf{V}\hat{\mathbf{U}}^{-1}\hat{\mathbf{\Sigma}}^{-1}\left[\mathbf{I} - \mathbf{T}\left(\mathbf{T}^{\mathrm{T}}\hat{\mathbf{U}}^{-1}\hat{\mathbf{\Sigma}}^{-1}\mathbf{T}\right)^{-1}\mathbf{T}^{\mathrm{T}}\hat{\mathbf{U}}^{-1}\hat{\mathbf{\Sigma}}^{-1}\right]$

$$\hat{\mathbf{U}} = \hat{\mathbf{\Sigma}}^{-1}\mathbf{V} + \mathbf{MA}$$

Proof:

Because the formula (15) function is $\underset{\sim}{f} = \mathbf{T}\underset{\sim}{d} + \mathbf{V}\underset{\sim}{c}$, then the nonparametric regression model for longitudinal data in formula (6) can be described as :

$$\underset{\sim}{y} = \underset{\sim}{f} + \underset{\sim}{\varepsilon} = \mathbf{T}\underset{\sim}{d} + \mathbf{V}\underset{\sim}{c} + \underset{\sim}{\varepsilon}$$

Then it is used $\mathcal{H} = \mathbf{W}_2^m[a_i, b_i]$ which means *Sobolev* space of order-2 defined as follows:

$\mathbf{W}_2^m[a_i, b_i] = f_i : \left\{ \int_{a_i}^{b_i} \left[f_i^{(m)}(x_{it}) \right]^2 dx_{it} < \infty \right\}$ with $a_i \le x_{it} \le b_i$ dan $i = 1, 2 \ldots, N$. based on that space

By having the RKHS approach, so that it is obtained the estimation of $\underset{\sim}{f}$ that fulfill the optimization of PWLS:

$$\min\left\{\left\|\Sigma^{-\frac{1}{2}}\underset{\sim}{\varepsilon}\right\|^2\right\}=\min\left\{\left\|\Sigma^{-\frac{1}{2}}(\underset{\sim}{y}-\underset{\sim}{f}\right\|^2\right\} \text{ with the obstacles } \int_{a_i}^{b_i}\left[f_i^{(m)}(x_{it})\right]^2 dx_{it}<\infty \tag{16}$$

Then it is used $\mathcal{H}=\mathbf{W}_2^m[a_i,b_i]$ which means *Sobolev* space of order-2 defined as follows: $\mathbf{W}_2^m[a_i,b_i]=f_i:\left\{\int_{a_i}^{b_i}\left[f_i^{(m)}(x_{it})\right]^2 dx_{it}<\infty\right\}$ with $a_i\le x_{it}\le b_i$ dan $i=1,2...,N.$ based on that space, it is

The optimization of the weighted with constraints on the formula (16) is equivalent to completing optimization of PWLS:

$$\text{Min}\left\{M^{-1}(\underset{\sim}{y}-\underset{\sim}{f})^{\mathrm{T}}\Sigma^{-1}(\underset{\sim}{y}-\underset{\sim}{f})+\sum_{i=1}^{N}\lambda_i\int_{a_i}^{b_i}(f_i^{(m)}(x_{it}))^2dx_{it}\right\} \tag{17}$$

Both $M=2NT$ and λ_i are the smoothing parameter that controls between goodness of fit, $M^{-1}(\underset{\sim}{y}-\underset{\sim}{f})^{\mathrm{T}}\Sigma^{-1}(\underset{\sim}{y}-\underset{\sim}{f})$ *and penalty*, $\sum_{i=1}^{N}\lambda_i\int_{a_i}^{b_i}(f_i^{(m)}(x_{it}))^2dx_{it}$

Then, it does the decomposition of the penalty component first as follows:

$$\int_{a_1}^{b_1}\left[(f_1^{(m)}(x_{11}))\right]^2dx_{11}=\|P_1f_1\|^2=\langle P_1f_1,P_1f_1\rangle$$

where P_1 is an orthogonal projection f_1 ke $\mathcal{H}_1$ in $\mathbf{W}_2^m[a_i,b_i]$.

$$\begin{aligned}\int_{a_1}^{b_1}\left[(f_1^{(m)}(x_{11}))\right]^2dx_{11}&=\left\langle P_1\left(\underset{\sim}{\phi}_1^{\mathrm{T}}\underset{\sim}{d}_1+\underset{\sim}{\xi}_1^{\mathrm{T}}\underset{\sim}{c}_1\right),P_1\left(\underset{\sim}{\phi}_1^{\mathrm{T}}\underset{\sim}{d}_1+\underset{\sim}{\xi}_1^{\mathrm{T}}\underset{\sim}{c}_1\right)\right\rangle=\left\langle\underset{\sim}{\xi}_1^{\mathrm{T}}\underset{\sim}{c}_1,\underset{\sim}{\xi}_1^{\mathrm{T}}\underset{\sim}{c}_1\right\rangle=\left(\underset{\sim}{\xi}_1^{\mathrm{T}}\underset{\sim}{c}_1\right)^{\mathrm{T}}\left(\underset{\sim}{\xi}_1^{\mathrm{T}}\underset{\sim}{c}_1\right)=\underset{\sim}{c}_1\left(\underset{\sim}{\xi}_1^{\mathrm{T}}\underset{\sim}{\xi}_1^{\mathrm{T}}\right)\underset{\sim}{c}_1\\&=\underset{\sim}{c}_{11}^{T}\mathbf{V}_1\underset{\sim}{c}_1\end{aligned}$$

as a result, $\int_{a_1}^{b_1}\left[(f_1^{(m)}(x_{11}))\right]^2dx_{11}=\lambda_1\underset{\sim}{c}_1^T\mathbf{V}_1\underset{\sim}{c}_1$. In a similar way, it is generally obtained:

$\int_{a_i}^{b_i}\left[(f_i^{(m)}(x_{it}))\right]^2dx_{it}=\lambda_i\underset{\sim}{c}_i^T\mathbf{V}_i\underset{\sim}{c}_i$, therefore, it is obtained the penalty value:

$$\sum_{i=1}^{N}\lambda_i\int_{a_i}^{b_i}(f_i^{(m)}(x_{it}))^2dx_{it}=\lambda_1\underset{\sim}{c}_1^T\mathbf{V}_1\underset{\sim}{c}_1+\lambda_2\underset{\sim}{c}_2^T\mathbf{V}_2\underset{\sim}{c}_1+...+\lambda_N\underset{\sim}{c}_N^T\mathbf{V}_N\underset{\sim}{c}_N$$

$$=\begin{pmatrix}\underset{\sim}{c}_1^{\mathrm{T}} & \underset{\sim}{c}_2^{\mathrm{T}} & \cdots & \underset{\sim}{c}_N^{\mathrm{T}}\end{pmatrix}\begin{pmatrix}\lambda_1\mathbf{V}_1 & 0 & 0 & 0\\ 0 & \lambda_2\mathbf{V}_2 & 0 & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \lambda_N\mathbf{V}_N\end{pmatrix}\begin{pmatrix}\underset{\sim}{c}_1\\ \underset{\sim}{c}_2\\ \vdots\\ \underset{\sim}{c}_N\end{pmatrix}=\underset{\sim}{c}^{\mathrm{T}}\mathbf{\Lambda}\mathbf{V}\underset{\sim}{c},\;\;\mathbf{\Lambda}=\begin{bmatrix}\lambda_1\mathbf{I}_N & \mathbf{0}_N & \cdots & \mathbf{0}_N\\ \mathbf{0}_N & \lambda_2\mathbf{I}_N & \cdots & \mathbf{0}_N\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0}_N & \mathbf{0}_N & \cdots & \lambda_N\mathbf{I}_N\end{bmatrix}$$

Based on the result of the formula (12), then it is obtained *goodness of fit* on the optimization of PWLS of the formula (15) that can be drawn as follows:

$$M^{-1}\left(\underset{\sim}{y}-\underset{\sim}{f}\right)^{\mathrm{T}}\mathbf{\Sigma}^{-1}\left(\underset{\sim}{y}-\underset{\sim}{f}\right)=M^{-1}\left(\underset{\sim}{y}-\mathbf{T}\underset{\sim}{d}-\mathbf{V}\underset{\sim}{c}\right)^{\mathrm{T}}\mathbf{\Sigma}^{-1}\left(\underset{\sim}{y}-\mathbf{T}\underset{\sim}{d}-\mathbf{V}\underset{\sim}{c}\right)$$

The completing optimization of PWLS both goodness of fit and penalty are presented as follows:

$$\min\left\{M^{-1}\left(\underset{\sim}{y}-\underset{\sim}{f}\right)^{\mathrm{T}}\Sigma^{-1}\left(\underset{\sim}{y}-\underset{\sim}{f}\right)+\sum_{i=1}^{N}\lambda_i\int_{a_i}^{b_i}\left(f_i^{(m)}(x_{it})\right)^2 dx_{it}\right\}=\min\left\{M^{-1}(\underset{\sim}{y}-\mathrm{T}\underset{\sim}{d}-\mathrm{V}\underset{\sim}{c})^{T}\Sigma^{-1}(\underset{\sim}{y}-\mathrm{T}\underset{\sim}{d}-\mathrm{V}\underset{\sim}{c})+\underset{\sim}{c}^{\mathrm{T}}\Lambda\mathrm{V}\underset{\sim}{c}\right\}$$

$$=\min\left\{\left[\underset{\sim}{y}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\underset{\sim}{y}-2\underset{\sim}{d}^{\mathrm{T}}\mathbf{T}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\underset{\sim}{y}-2\underset{\sim}{c}^{\mathrm{T}}\mathbf{V}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\underset{\sim}{y}+\underset{\sim}{d}^{\mathrm{T}}\mathbf{T}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\mathbf{T}\underset{\sim}{d}+\underset{\sim}{d}^{\mathrm{T}}\mathbf{T}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\mathbf{V}\underset{\sim}{c}+\underset{\sim}{c}^{\mathrm{T}}\mathbf{V}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\mathbf{T}\underset{\sim}{d}+\underset{\sim}{c}^{\mathrm{T}}\left(\mathbf{V}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\mathbf{V}+M\mathbf{\Lambda}\mathbf{V}\right)\underset{\sim}{c}\right]M^{-1}\right\}$$

$$=\min\left\{\mathrm{Q}\left(\underset{\sim}{c},\underset{\sim}{d}\right)\right\} \tag{18}$$

The completing optimization on the formula (16) is obtained through the partial derivatives $\mathrm{Q}\left(\underset{\sim}{c},\underset{\sim}{d}\right)$ successively against $\underset{\sim}{c}$ and $\underset{\sim}{d}$, then it is equated with 0. The following is a partial derivative:

$\dfrac{\partial\mathrm{Q}\left(\underset{\sim}{c},\underset{\sim}{d}\right)}{\partial\underset{\sim}{c}}=\underset{\sim}{0}$, then it is obtained: $-\mathbf{\Sigma}^{-1}\underset{\sim}{y}+\mathbf{\Sigma}^{-1}\mathbf{T}\underset{\sim}{d}+\left[\mathbf{\Sigma}^{-1}\mathbf{V}+M\mathbf{\Lambda}\mathbf{I}\right]\hat{\underset{\sim}{c}}=\underset{\sim}{0}$

For example, it is given a matrix $\mathbf{U}=\mathbf{\Sigma}^{-1}\mathbf{V}+M\mathbf{\Lambda}$

Then, it can be re-written: $\hat{\underset{\sim}{c}}=\mathbf{U}^{-1}\mathbf{\Sigma}^{-1}\left(\underset{\sim}{y}-\mathbf{T}\underset{\sim}{d}\right)$ (19)

Then, the partial derivative $\dfrac{\partial\mathrm{Q}\left(\underset{\sim}{c},\underset{\sim}{d}\right)}{\partial\underset{\sim}{d}}=\underset{\sim}{0}$ is obtained:

$$-\mathbf{T}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\underset{\sim}{y}+\mathbf{T}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\mathbf{T}\hat{\underset{\sim}{d}}+\mathbf{T}^{\mathrm{T}}\left[\mathbf{\Sigma}^{-1}\mathbf{V}\mathbf{U}^{-1}\right]\mathbf{\Sigma}^{-1}\left(\underset{\sim}{y}-\mathbf{T}\hat{\underset{\sim}{d}}\right)=\underset{\sim}{0} \tag{20}$$

Because $\mathbf{U}=\mathbf{\Sigma}^{-1}\mathbf{V}+M\mathbf{\Lambda}$, then $\mathbf{V}=\mathbf{\Sigma}\left(\mathbf{U}-M\mathbf{\Lambda}\mathbf{I}\right)$ it will be obtained as follows:

$$\mathbf{V}\mathbf{U}^{-1}=\mathbf{\Sigma}\left(\mathbf{U}-M\mathbf{\Lambda}\mathbf{I}\right)\mathbf{U}^{-1}=\mathbf{\Sigma}\left(\mathbf{I}-M\mathbf{\Lambda}\mathbf{U}^{-1}\right)$$

$$\mathbf{\Sigma}^{-1}\mathbf{V}\mathbf{U}^{-1}=\mathbf{I}-M\mathbf{\Lambda}\mathbf{U}^{-1} \tag{21}$$

The formula (21) is substituted into the formula (20), so that it is obtained:

$$-\mathbf{T}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\underset{\sim}{y}+\mathbf{T}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\mathbf{T}\hat{\underset{\sim}{d}}+\mathbf{T}^{\mathrm{T}}\left[\mathbf{I}-M\mathbf{\Lambda}\mathbf{U}^{-1}\right]\mathbf{\Sigma}^{-1}\left(\underset{\sim}{y}-\mathbf{T}\hat{\underset{\sim}{d}}\right)=\underset{\sim}{0}$$

If the formulation above is described, then:

$$-M\mathbf{\Lambda}\mathbf{T}^{\mathrm{T}}\mathbf{U}^{-1}\mathbf{\Sigma}^{-1}\underset{\sim}{y}+M\mathbf{\Lambda}\mathbf{T}^{\mathrm{T}}\mathbf{U}^{-1}\mathbf{\Sigma}^{-1}\mathbf{T}\underset{\sim}{d}=\underset{\sim}{0}$$

$$\hat{\underset{\sim}{d}}=\left(\mathbf{T}^{\mathrm{T}}\mathbf{U}^{-1}\mathbf{\Sigma}^{-1}\mathbf{T}\right)^{-1}\mathbf{T}^{\mathrm{T}}\mathbf{U}^{-1}\mathbf{\Sigma}^{-1}\underset{\sim}{y} \tag{22}$$

The formula (22) is substituted into the formula (19), and then it is obtained:

$$\hat{\underset{\sim}{c}}=\mathbf{U}^{-1}\mathbf{\Sigma}^{-1}\left[\mathbf{I}-\mathbf{T}\left(\mathbf{T}^{\mathrm{T}}\mathbf{U}^{-1}\mathbf{\Sigma}^{-1}\mathbf{T}\right)^{-1}\mathbf{T}^{\mathrm{T}}\mathbf{U}^{-1}\mathbf{\Sigma}^{-1}\underset{\sim}{y}\right] \tag{23}$$

Based on the formula (22) and (23), it is obtained the estimator for longitudinal data of nonparametric regression curves as follows:

$$\hat{\underset{\sim}{f}}=\mathbf{T}\hat{\underset{\sim}{d}}+\mathbf{V}\hat{\underset{\sim}{c}} \tag{24}$$

$$\hat{\underset{\sim}{f}}=\left\{\mathbf{T}\left(\mathbf{T}^{\mathrm{T}}\mathbf{U}^{-1}\mathbf{\Sigma}^{-1}\mathbf{T}\right)^{-1}\mathbf{T}^{\mathrm{T}}\mathbf{U}^{-1}\mathbf{\Sigma}^{-1}\underset{\sim}{y}+\mathbf{V}\mathbf{U}^{-1}\mathbf{\Sigma}^{-1}\left[\mathbf{I}-\mathbf{T}\left(\mathbf{T}^{\mathrm{T}}\mathbf{U}^{-1}\mathbf{\Sigma}^{-1}\mathbf{T}\right)^{-1}\mathbf{T}^{\mathrm{T}}\mathbf{U}^{-1}\mathbf{\Sigma}^{-1}\right]\right\}\underset{\sim}{y}$$

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

$$\hat{\underset{\sim}{f}}=\mathbf{A}_{\lambda}\underset{\sim}{y} \tag{25}$$

❖ **Proven**

4.2. Matrix Estimation of Error Variance-Covariance

By using Maximum Likelihood Estimator (MLE) method, ε_{it} the random error is obtained from the results of the estimation on i-subject of t-observation is is assumed by distributed NT-variat and the mean $E(\underset{\sim}{\varepsilon})=0$ (vector sizedNT) and variance-covariance matrix $\mathrm{Var}(\underset{\sim}{\varepsilon})=0$ (matrix sized $NT\times NT$) as follows :

$$\hat{\Sigma} = \begin{bmatrix} \hat{\Sigma}_{11} & 0 & \cdots & 0 \\ 0 & \hat{\Sigma}_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{\Sigma}_{NN} \end{bmatrix}_{NT\times NT}$$

To obtain the estimation $\hat{\Sigma}_{ii}; i = 1,2,\dots,N$ with $\underset{\sim}{y}_i = \left(\underset{\sim}{y}_1, \underset{\sim}{y}_2, \dots, \underset{\sim}{y}_n\right)$ then, the function likelihood y_i as follows:

$$L\left(f_i, \Sigma_i \mid \underset{\sim}{y}_i\right) = \frac{\mathrm{TT}}{2}\ln(2\pi) - \frac{\mathrm{T}}{2}\ln|\Sigma| - \frac{1}{2}\operatorname{tr}\left[\boldsymbol{\Sigma}_{ii}^{-1}\sum_{t=1}^{T}(\underset{\sim}{y}_{it} - \underset{\sim}{f}_{it})(\underset{\sim}{y}_{it} - \underset{\sim}{f}_{it})^{\mathrm{T}}\right]$$

for example $\mathbf{W}_{ii} = \frac{\sum_{t=1}^{T}(\underset{\sim}{y}_{it} - \underset{\sim}{f}_{it})(\underset{\sim}{y}_{it} - \underset{\sim}{f}_{it})^{\mathrm{T}}}{T}$,Then $\ln\left[L(\underset{\sim}{f}_i, \boldsymbol{\Sigma}_{ii} \left|\underset{\sim}{y}_i\right.)\right] = -\frac{TT}{2}\ln(2\pi) - \frac{T}{2}\ln|\boldsymbol{\Sigma}_{ii}| - \frac{T}{2}\operatorname{tr}\left[\boldsymbol{\Sigma}_{ii}^{-1}\mathbf{W}_{ii}\right]$ the estimator for variance-covariance matrix $\hat{\boldsymbol{\Sigma}}_{ii}$ is obtained by maximizing the function $\ln\left[L(\underset{\sim}{f}_i, \boldsymbol{\Sigma}_{ii} \left|\underset{\sim}{y}_i\right.)\right]$ by formulating $\frac{\partial \ln\left[L(\underset{\sim}{f}_i, \boldsymbol{\Sigma}_{ii} \left|\underset{\sim}{y}_i\right.)\right]}{\partial \boldsymbol{\Sigma}_{ii}} = \mathbf{0}$.

$$\frac{\partial \ln\left[L(\underset{\sim}{f}_i, \boldsymbol{\Sigma}_{ii} \left|\underset{\sim}{y}_i\right.)\right]}{\partial \boldsymbol{\Sigma}_{ii}} = \frac{\partial\left[-\frac{TT}{2}\ln(2\pi) - \frac{T}{2}\ln|\boldsymbol{\Sigma}_{ii}| - \frac{T}{2}\operatorname{tr}\left[\boldsymbol{\Sigma}_{ii}^{-1}\mathbf{W}_{ii}\right]\right]}{\partial \boldsymbol{\Sigma}_{ii}} = -\frac{T}{2}\boldsymbol{\Sigma}_{ii}^{-1}\boldsymbol{\Sigma}_{ii}\boldsymbol{\Sigma}_{i}^{-1} + \frac{T}{2}\boldsymbol{\Sigma}_{ii}^{-1}\mathbf{W}_{ii}\boldsymbol{\Sigma}_{ii}^{-1}$$

$= -\frac{T}{2}\boldsymbol{\Sigma}_{ii}^{-1}(\boldsymbol{\Sigma}_{ii} - \mathbf{W}_{ii})\boldsymbol{\Sigma}_{ii}^{-1} = 0$ So that $\hat{\boldsymbol{\Sigma}}_{ii} = \mathbf{W}_{ii}$

Therefore, the estimator of variance-covariance matrix *error* $\hat{\boldsymbol{\Sigma}}_{ii}$ for nonparametric regression models for single predictors and single responses to longitudinal data which is presented as follows:

$$\hat{\boldsymbol{\Sigma}}_{ii} = \frac{(\underset{\sim}{y}_i - \hat{\underset{\sim}{f}}_i)(\underset{\sim}{y}_i - \hat{\underset{\sim}{f}}_i)^{\mathrm{T}}}{T}; i = 1,2,\dots,N$$

4.3. Choosing the Optimal Smoothing Parameters Methods

The method of selecting the optimal smoothing parameters in the weighted spline estimator in nonparametric regression on single predictors and single responses for longitudinal data uses the Generalized Cross Validation (GCV) method. It is known in the formula (23) weighted spline estimator as follows $\hat{\underset{\sim}{f}} = \mathbf{A}_\lambda \underset{\sim}{y}$, so that goodness of fit from spline estimator is described as follows:

$$R(f) = M^{-1}(\underset{\sim}{y} - \hat{\underset{\sim}{f}})^{\mathrm{T}}\hat{\Sigma}^{-1}(\underset{\sim}{y} - \hat{\underset{\sim}{f}}) = M^{-1}(\underset{\sim}{y} - \mathbf{A}_\lambda \underset{\sim}{y})^{\mathrm{T}}\hat{\Sigma}^{-1}(\underset{\sim}{y} - \mathbf{A}_\lambda \underset{\sim}{y}) = M^{-1}\left(\left[\underset{\sim}{y}^{\mathrm{T}}(\mathbf{I} - \mathbf{A}_\lambda)^{\mathrm{T}}\right]\left[\hat{\Sigma}^{-1}\right]^{\frac{1}{2}^{\mathrm{T}}}\right)\left(\left[\hat{\Sigma}^{-1}\right]^{\frac{1}{2}}\left[(\mathbf{I} - \mathbf{A}_\lambda)\underset{\sim}{y}\right]\right)$$

Therefore, GCV is defined as follows:

$$GCV = \frac{M^{-1}\underset{\sim}{y}^{\mathrm{T}}(\mathbf{I}\text{-}\mathbf{A}_\lambda)^{\mathrm{T}}(\mathbf{I}\text{-}\mathbf{A}_\lambda)\underset{\sim}{y}}{\left(M^{-1}trace(\mathbf{I}\text{-}\mathbf{A}_\lambda)\right)^2}$$

5. Conclusion

In relation to the results and the discussion above, it can be concluded that the nonparametric regression model involving a single predictor and a single response is defined as $(x_{it}y_{it})$ which means that the formula of the longitudinal data is $y_{it} = f_i(x_{it}) + \varepsilon_{it}$ that has the function

$$\underset{\sim}{f} = \mathbf{T}\underset{\sim}{d} + \mathbf{V}\underset{\sim}{c}.$$

Moreover, the spline estimator satisfying the minimum criteria of PWLS which is presented as follows:

$$\text{Min}\left\{ M^{-1}(\underset{\sim}{y}-\underset{\sim}{f})^{\mathrm{T}}\Sigma^{-1}(\underset{\sim}{y}-\underset{\sim}{f})+\sum_{i=1}^{N}\lambda_i\int_{a_i}^{b_i}(f_i^{(m)}(x_{it}))^2\,dx_{it}\right\} \text{ is } \hat{\underset{\sim}{f}}=\mathbf{T}\hat{\underset{\sim}{d}}+\mathbf{V}\hat{\underset{\sim}{c}}=\mathbf{A}_{\underset{\sim}{\lambda}}\underset{\sim}{y},$$

The estimation of the matrix of the variance covariance random error is presented as follows:

$$\hat{\mathbf{\Sigma}}_{ii}=\mathbf{W}_{ii}=\frac{(\underset{\sim}{y}_i-\hat{\underset{\sim}{f}}_i)(\underset{\sim}{y}_i-\hat{\underset{\sim}{f}}_i)^{\mathrm{T}}}{T},\quad i=1,2,\ldots,N.$$

The smoothing parameter is presented as follows:

$$GCV=\frac{M^{-1}\underset{\sim}{y}^{\mathrm{T}}\left(\mathbf{I}-\mathbf{A}_{\lambda}\right)^{\mathrm{T}}\left(\mathbf{I}-\mathbf{A}_{\lambda}\right)\underset{\sim}{y}}{\left(M^{-1}trace\left(\mathbf{I}-\mathbf{A}_{\lambda}\right)\right)^2}$$

References

[1] Hardle, W. (1990). Applied Nonparametric Regression. New York : Cambridge University Press.

[2] Budiantara, I.N. (2001) . Estimasi Parametrik dan Non parametrik untuk Pendekatan Kurva Regresi Makalah Pembicara Utama Pada Seminar Nasional Statistika V, Jurusan Statistika, FMIPA, Institut Teknologi Sepuluh November (ITS), Surabaya.

[3] Eubank, R.L. (1988), *Spline* Smoothing and Nonparametric Regression, New York: Marcel Deker.

[4] Heckman, N., Lockhart R., and Nielsen J.D. (2009). *Penalized Regression, Mixed Effects Models and Appropriate Modelling*. Retrieved, December, 12, 2012. Website: http://www.stat.ubc.ca/~nancy/pubs/lmetechreport.pdf.

[5] Budiantara, I.N., Lestari, B., dan Islamiyati, A. (2009). *Estimator Spline Terbobot Spline Parsial Terbobot dalam Regresi Nonparametrik dan Semi parametrik Heteroskedastik untuk Data* Longitudinal. Laporan Akhir Program Hibah Kompetensi Tahun 1. Lembaga Penelitian dan Pengabdian Kepada Masyarakat. Institut Teknologi Sepuluh Nopember. Surabaya.

[6] Wu, H., and Zhang, J.T. (2006). *Nonparametric Regression Methods for* Longitudinal *Data Analysis*. New Jersey: John Wiley and Sons, Inc.

[7] Diggle, P.J., Liang, Y.K. and Zeger, S.L., (2006), *Analysis of* Longitudinal *Data.* Second Edition. New York: Oxford.

[8] Verbekke, G., and Molenberghs, G. (2000). *Linear Mixed Model for* Longitudinal *Data*. Springer Series in statistics. New York: Springer Verlag.

[9] Fernandes, A.A.R. (2016). Estimator Spline dalam Regresi Nonparametrik Birespon Untuk Data Longitudinal (Studi Kasus Pada Pasien Penderita TB Paru Malang). Institut Teknologi Sepuluh Nopember. Surabaya.

[10] Gujarati, D. (2004). Ekonometrika Dasar. Terjemahan : Zain, S. Jakarta : Erlangga.

[11] Weiss, R.E. (2005). *Modelling* Longitudinal *Data*. Springer Texts in Statistic New York. Retrieved, January, 31, 2012. Website: http://www.biostat.ucla. edu/books/mld.

[12] Lestari, B., Budiantara, I.N., Sunaryo, S., & Mashuri, M. (2012). Spline Smoothing for Multi-response Nonparametric Regresiion Model in Case of heteroscedascity of Variance. Journal of Mathematics and Statistics, 8(3), 377-384.