Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/120234

Development of Nonparametric Truncated Spline at Various Levels of Autocorrelation of Longitudinal Generating Data

Ananda Ajeng Puspitasari[*], *Adji Achmad Rinaldo Fernandes, Achmad Efendi, Suci Astutik and Eni Sumarminingsih*

Department of Statistics, Faculty of Mathematics and Natural Sciences, 65145 Indonesia, Brawijaya University, Indonesia

Received: 8 Jun. 2022, Revised: 20 Aug. 2022, Accepted: 20 Nov. 2022.
Published online: 1 May 2023.

Abstract: The purpose of this study is to obtain the best approach in longitudinal data using Path Nonparametric Truncated Spline with PLS and PWLS approach. The data used in this study is longitudinal generation data. There are three variables, five levels of autocorrelation, and the data does not meet the assumption of linearity. The degree of autocorrelation is divided into very low autocorrelation (|0.01-0.20|), low (|0.21-0.40|), moderate (|0.41-0.60|), high (|0.61-0.80|), and very high (|0.81-1.00|). The analysis used is path nonparametric truncated spline with PLS and PWLS approach. The result of this research is the estimation of nonparametric path function using PLS and PWLS approach on various interactions. Based on the Relative Efficiency (ER) value obtained, PWLS approach is a better approach than PLS approach at all levels of autocorrelation when the number of observations in the subject is five (T=5). On the number of observations in fifteen and twenty subjects (T=15, 20), PWLS approach was better than PLS approach when the autocorrelation of the data was more than 0.20. The originality of this research is the development of a nonparametric path with the PLS and PWLS approach that is able to obtain the best approach in limitations of autocorrelation between observations and in one subject are interconnected. There has been no previous research that has examined the best approach of a path nonparametric with PLS and PWLS approach.

Keywords: Longitudinal Data, Path Nonparametric Truncated Spline, PLS, PWLS.

1 Introduction

Path analysis is a development of regression analysis where the analysis can accommodate several structural equations [5]. Structural equations in path have characteristics that involve at least one exogenous variable, at least one intervening endogenous variable and there is one pure endogenous variable [7]. Exogenous variables are variables that can affect pure endogenous variables, and intervening variables are intermediary exogenous variables that affect pure endogenous variables.

Similar to regression analysis, path also has several assumptions, including linearity assumptions, residual normality assumptions and homoscedasticity assumptions [16]. The linearity assumption is an assumption that must first be met before the path is carried out. Path analysis assumes that the relationship between exogenous variables and endogenous variables can be explained through a known function and that function is a linear function. If this assumption is violated, it is impossible to estimate path [3].

Path Nonparametric analysis can be used if the linearity assumption is not met or the relationship pattern between exogenous variables and endogenous variables is unknown. A popular nonparametric path analysis is the nonparametric path truncated spline. Spline has the ability to generalize complex and complex statistical modeling very well. One of the splines that are of interest in applications is the truncated spline [1].

In this study we apply longitudinal data. Longitudinal Data is data collected where each subject has two or more specific observation time periods. According to [18], the nature of longitudinal data is independent between observations and dependent between observations within the subject. The nature of the dependence on observations in this subject gives rise to autocorrelation.

Autocorrelation in longitudinal data is a close relationship between one observation and another in one subject. In parametric analysis, autocorrelation in the data will cause the variance of the estimator to be no longer minimum even though it remains unbiased [8], this also applies to path because the variance of the estimated function is no longer

[*]Corresponding author e-mail: anandaajengps@student.ub.ac.id

minimum. Therefore, autocorrelation is a problem to be solved. The approach commonly used in estimating path is PLS (Penalized Least Square) but when there is autocorrelation in the data, it must be weighted using the PWLS (Penalized Weighted Least Square) [7].

This study examines the comparison of nonparametric path on longitudinal data. The data design in this study is longitudinal with three variables, five levels of autocorrelation, and does not meet the assumption of linearity. Generating data is the best alternative because of some of the constraints that have been mentioned above. By comparing the level of autocorrelation, it is hoped that the researcher will be able to obtain the optimum level of autocorrelation on the longitudinal data modeled by nonparametric path.

2 Materials and Methods

2.1 Data Longitudinal

According to [18], longitudinal data is data that is observed in certain time periods and carried out on several individuals. In longitudinal data between one individual in a group and another individual is considered independent but between observations in one individual is considered dependent [4].

There are three advantages of using longitudinal data compared to cross section data. First, longitudinal data more powerful than cross-sectional data because it can obtain the same statistical test power even though the number of subjects is smaller. Second, with the same number of subjects the error measurement results produce a more efficient estimator of the treatment effect. Third, longitudinal data can provide information about individual changes.

2.2 Path Nonparametric

Path Parametric analysis differ from the relationship pattern between exogenous variables and endogenous variables [1]. Exogenous variables are variables that can affect pure endogenous variables, and intervening variables are intermediary exogenous variables that affect pure endogenous variables. In linear patterns or known patterns of relationships between exogenous and endogenous variables, the analysis used is path, while in non-linear patterns or unknown relationship patterns the analysis used is nonparametric path. The relationship between endogenous and exogenous variables in path nonparametric f(x) where the function f(x) is a function whose form is not yet known [11]. One method to estimate the function f(x) on a nonparametric path is to use a truncated spline.

2.3 Model Path Nonparametric Truncated Spline Data Longitudinal

This study has the path nonparametric truncated spline

$$y_{1it} = f_1\left(x_{it}\right) + \varepsilon_{1it}; \tag{1}$$

$$y_{2it} = f_2\left(y_{1it}\right) + \varepsilon_{2it}; i = 1, 2, \ldots, N; t = 1, 2, \ldots, T \tag{2}$$

With:

x_{it} : is an exogenous variable, i is subject ,t is observations

y_{1it} : intervening variable, i is subject ,t is observations

y_{2it} : is a pure endogenous variable, i is subject ,t is observations

N : number of subjects

T : number of observations in each subject

ε_{1it} : random error in the truncated spline first

ε_{2it} : random error in the truncated spline second

According to [13], splines are parts or pieces of a polynomial that have segmented and continuous (truncated) properties.

[13] conducted a study with a regression curve f approximated by the spline function f knots K. Knot point (K) on the truncated spline is used to cut the function in order to get the right estimate. In matrix form it is stated as follows:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{nx1} = \begin{bmatrix} f(X_1) \\ f(X_2) \\ \vdots \\ f(X_n) \end{bmatrix}_{nx1} + \begin{bmatrix} \mathcal{E}_1 \\ \mathcal{E}_2 \\ \vdots \\ \mathcal{E}_n \end{bmatrix}_{nx1} \tag{3}$$

In matrix form, the spline regression model is presented as follows:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 & \cdots & X_1^p & (X_1 - K_1)_+^p & \cdots & (X_1 - K_k)_+^p \\ 1 & X_2 & \cdots & X_2^p & (X_2 - K_1)_+^p & \cdots & (X_2 - K_k)_+^p \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_n & \cdots & X_n^p & (X_n - K_1)_+^p & \cdots & (X_n - K_k)_+^p \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \\ \delta_1 \\ \vdots \\ \delta_k \end{bmatrix} + \begin{bmatrix} \mathcal{E}_1 \\ \mathcal{E}_2 \\ \vdots \\ \mathcal{E}_n \end{bmatrix} \tag{4}$$

It can also be written as:

$$\underset{\sim}{y} = \mathbf{X}\left[K_1, K_2, ..., K_k\right]\underset{\sim}{\beta} + \underset{\sim}{\varepsilon} \tag{5}$$

Furthermore, parameter estimation

$$\hat{\underset{\sim}{\beta}} = \left(\beta_0 \quad \beta_1 \quad \beta_2 \quad \mathrm{K} \quad \beta_p \quad \delta_1 \quad \mathrm{L} \quad \delta_k\right)^t \tag{6}$$

is obtained using the least square method, the optimization solution is as follows:

$$\underset{\underset{\sim}{\beta} \in R^{p+1+k}}{Min}\left\{\underset{\sim}{\varepsilon}^t \underset{\sim}{\varepsilon}\right\} = \underset{\underset{\sim}{\beta} \in R^{p+1+k}}{Min}\left\{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \mathrm{M} \\ \varepsilon_n \end{pmatrix}^t \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \mathrm{M} \\ \varepsilon_n \end{pmatrix}\right\} = \underset{\underset{\sim}{\beta} \in R^{p+1+k}}{Min}\left\{\left(\underset{\sim}{y} - \mathbf{X}\left[K_1, K_2, \mathrm{K}, K_k\right]\underset{\sim}{\beta}\right)^t \left(\underset{\sim}{y} - \mathbf{X}\left[K_1, K_2, \mathrm{K}, K_k\right]\underset{\sim}{\beta}\right)\right\} \tag{7}$$

Comparison of PLS and PWLS

In the spline to estimate the regression curve is as follows [2].

$$\min\left\{M^{-1}\left(\underset{\sim}{y} - \underset{\sim}{f}\right)^T\left(\underset{\sim}{y} - \underset{\sim}{f}\right) + \sum_{i=1}^{N}\lambda_i \int_{a_i}^{b_i}\left(f_i^{(m)}(x_{it})\right)^2 dx_{it}\right\} \tag{8}$$

The function in equation 8 is aspline obtained by minimizing the Penalized Least Square (PLS). Goodness of fit is a function to measure the fit is on the first segment, while roughness penalty which is a function to measure the roughness of the curve is on the second segment. Estimating the regression curve $\hat{f}$ by minimizing PLS is equivalent to the PWLS approach without considering the presence of autocorrelation. According to [4], if there is paired data following $E\left(\underset{\sim}{\varepsilon}\right) = \underset{\sim}{0}$ and $Var\left(\underset{\sim}{\varepsilon}\right) = \mathbf{\Sigma}$ estimate spline that minimizes PWLS is as follows.

$$\min\left\{M^{-1}\left(\underset{\sim}{y} - \underset{\sim}{f}\right)^T \Sigma^{-1}\left(\underset{\sim}{y} - \underset{\sim}{f}\right) + \sum_{i=1}^{N}\lambda_i \int_{a_i}^{b_i}\left(f_i^{(m)}(x_{it})\right)^2 dx_{it}\right\} \tag{9}$$

It is known that $\hat{\underset{\sim}{f}}_\lambda = \mathbf{A}_\lambda \underset{\sim}{y}$ with:

$$\mathbf{A}_\lambda = \mathbf{T}(\mathbf{T}^T\hat{\mathbf{U}}^{-1}\hat{\mathbf{\Sigma}}^{-1})^{-1}\mathbf{T}^T\hat{\mathbf{U}}^{-1}\hat{\mathbf{\Sigma}}^{-1} + \mathbf{V}\hat{\mathbf{U}}^{-1}\hat{\mathbf{\Sigma}}^{-1}\left[\mathbf{I} - \mathbf{T}(\mathbf{T}^T\hat{\mathbf{U}}^{-1}\hat{\mathbf{\Sigma}}^{-1}\mathbf{T})^{-1}\mathbf{T}^T\hat{\mathbf{U}}^{-1}\hat{\mathbf{\Sigma}}^{-1}\right] \tag{10}$$

$\hat{\mathbf{U}} = \hat{\mathbf{\Sigma}}^{-1}\mathbf{V} + M\mathbf{\Lambda}$, M size 2NT. weights $\mathbf{\Sigma}^{-1}$ applied to the spline without considering autocorrelation are as follows.

$$\mathbf{\Sigma_{TM}} = \begin{bmatrix} \mathbf{\Sigma_{11}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma_{12}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{\Sigma_{1N}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{\Sigma_{21}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{\Sigma_{2N}} \end{bmatrix}_{2NT\times 2NT} \tag{11}$$

It $\mathbf{\Sigma}$ is stated as follows.

$$\mathbf{\Sigma}_{ki} = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix}_{T\times T} ; k = 1, 2 \tag{12}$$

Meanwhile, to estimate the regression curve for the case of autocorrelation, the weighting applied also takes into account autocorrelation such as matrix (13).

$$\mathbf{\Sigma_{DM}} = \begin{bmatrix} \mathbf{\Sigma_{11}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma_{12}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{\Sigma_{1N}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{\Sigma_{21}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{\Sigma_{2N}} \end{bmatrix}_{2NT\times 2NT} \tag{13}$$

The form of the matrix $\mathbf{\Sigma_{11}} = \mathbf{\Sigma_{11}} = \ldots = \mathbf{\Sigma_{kNT}}$ in the matrix (13) can be broken down into the following sub-matrix.

$$\mathbf{\Sigma}_{ki} = \begin{bmatrix} \sigma^2 & \rho_k\sigma^2 & \cdots & \rho_k^{T-1}\sigma^2 \\ \rho_k\sigma^2 & \sigma^2 & \cdots & \rho_k^{T-2}\sigma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \rho_k^{T-1}\sigma^2 & \rho_k^{T-2}\sigma^2 & \cdots & \sigma^2 \end{bmatrix}_{T\times T} ; k = 1, 2 \tag{14}$$

2.4 Relative Efficiency

In this research, relative efficiency is used to compare path nonparametric truncated spline based on PLS and PWLS. The comparison was carried out by ratifying the variance of the residuals for the PLS (weighting $\mathbf{\Sigma}_{TM}$) and PWLS (weighting $\mathbf{\Sigma}_{DM}$) estimators.

$$eff(\hat{f}_{DM}, \hat{f}_{TM}) = \frac{V(\hat{f}_{TM})}{V(\hat{f}_{DM})} \tag{15}$$

(Wackerly et al., 2014)
Description:
$V(\hat{f}_{DM})$: Variation of residuals in the PWLS function (Weighting $\mathbf{\Sigma}_{TM}$) $V(\hat{f}_{TM})$: Variance of residuals on the PLS function (Weighting $\mathbf{\Sigma}_{DM}$)

If the result of the calculation is more than one, it can be stated that the path nonparametric truncated spline based on PWLS (Weighter $\boldsymbol{\Sigma}_{DM}$) is better than path nonparametric truncated spline based $\boldsymbol{\Sigma}_{TM}$).

The statistical application should not be complicated and difficult, it should rather be simple and easy so that it is user-friendly [15].

2.5 Data Source

The data used in this study are simulation data that meet the following criteria:

1. The autocorrelation level is divided into five, namely |0.01–0.20|; |0.21-0.40|; |0.41-0.60|; |0.61-0.80|; and |0.81-1.00|.
2. Number of subjects N = 4.
3. Number of observations in subject T =5, 15, and 20.
4. Maximum iteration is 100 times and there are 30 repetitions (U).

Steps

The steps of this research are as follows:

1. The path analysis diagram used is as shown in Figure 1.

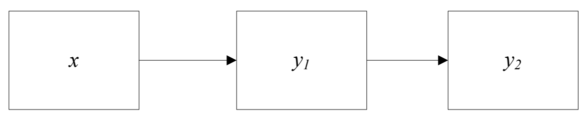

Fig.1: Path Analysis Diagram.

2. Setting the value $x_t, t = 1, 2, \ldots, T$ with $x_t \in [0,1]$ is the design time point with the formula $x_t = \dfrac{2t-1}{2T}$.
3. Generating autocorrelation.
4. Form a matrix $\boldsymbol{\Sigma}_{DM}$ and $\boldsymbol{\Sigma}_{TM}$ generate random error 2NT-variate where $E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$ and $\text{Var}(\underset{\sim}{\varepsilon}) = \boldsymbol{\Sigma}$ with two conditions, namely:
   The first condition without any weighting.

$$\boldsymbol{\Sigma}_{\mathbf{TM}} = \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{11}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\mathbf{12}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_{\mathbf{1N}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_{\mathbf{21}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_{\mathbf{2N}} \end{bmatrix}_{2NT \times 2NT}$$

   With $\boldsymbol{\Sigma}$ stated as follows.

$$\boldsymbol{\Sigma}_{ki} = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix}_{T \times T} ; k = 1, 2$$

   The second condition is a matrix with the following weighting.

$$\mathbf{\Sigma_{DM}} = \begin{bmatrix} \mathbf{\Sigma_{11}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma_{12}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{\Sigma_{1N}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{\Sigma_{21}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{\Sigma_{2N}} \end{bmatrix}_{2NT \times 2NT}$$

With $\mathbf{\Sigma}$ stated as follows.

$$\mathbf{\Sigma}_{ki} = \begin{bmatrix} \sigma^2 & \rho_k \sigma^2 & \cdots & \rho_k^{T-1}\sigma^2 \\ \rho_k \sigma^2 & \sigma^2 & \cdots & \rho_k^{T-2}\sigma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \rho_k^{T-1}\sigma^2 & \rho_k^{T-2}\sigma^2 & \cdots & \sigma^2 \end{bmatrix}_{T \times T} ; k = 1,2$$

In these two conditions, it is determined σ^2 that EV = 0.01.

5. Generating residual data according to points 2 and 3.
6. Generating data on intervening variables and pure endogenous variables with the function $y_{1it} = f_1(x_{it}) + \varepsilon_{1it}$ and $y_{2it} = f_2(y_{1it}) + \varepsilon_{2it}$.
   a) For the value $y_{1it}, t = 1,2,\ldots,T$ obtained from the logit equation $y_{1it} = \dfrac{e^{0,5+0,7x_{it}}}{1+e^{0,5+0,7x_{it}}}$ and added to the remainder at point 4.
7. For the value $y_{2it}, t = 1,2,\ldots,T$ obtained from the logit equation $y_{2it} = \dfrac{e^{0,5+0,5y_{1it}}}{1+e^{0,5+0,5y_{1it}}}$ and added to the remainder at point 4.
8. Form a matrix $\mathbf{T}$ and form a matrix $\mathbf{V}$ with the equation as below.
9.

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

$$\mathbf{T_1} = \begin{bmatrix} \mathbf{T_{11}} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{T_{12}} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{T_{1N}} \end{bmatrix}_{NT \times Nm}$$

$$\mathbf{T_2} = \begin{bmatrix} \mathbf{T_{21}} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{T_{22}} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{T_{2N}} \end{bmatrix}_{NT \times Nm}$$

$$\mathbf{T}_{1i} = \mathbf{T}_{2i} = \begin{bmatrix} \langle \eta_{i1}, \phi_{i1} \rangle & \langle \eta_{i1}, \phi_{i2} \rangle & \cdots & \langle \eta_{i1}, \phi_{im} \rangle \\ \langle \eta_{i2}, \phi_{i1} \rangle & \langle \eta_{i2}, \phi_{i2} \rangle & \cdots & \langle \eta_{i2}, \phi_{im} \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \eta_{iT}, \phi_{i1} \rangle & \langle \eta_{iT}, \phi_{i2} \rangle & \cdots & \langle \eta_{iT}, \phi_{im} \rangle \end{bmatrix}_{T \times m} \quad \begin{aligned} \langle \eta_{iT}, \phi_{im} \rangle &= \frac{x_{it}^{j-1}}{(j-1)!}, t = 1,2,\ldots,T; j = 1,2,..,m \\ \langle \eta_{iT}, \phi_{im} \rangle &= \frac{y_{1it}^{j-1}}{(j-1)!}, t = 1,2,\ldots,T; j = 1,2,...,m \end{aligned}$$

While the matrix **V** is like the equation below.

$$\mathbf{V} = \begin{bmatrix} \mathbf{V_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{V_2} \end{bmatrix}_{2NT \times 2NT}$$

$$\mathbf{V_1} = \begin{bmatrix} \mathbf{V_{11}} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{V_{12}} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{V_{1N}} \end{bmatrix}_{NT \times NT}$$

$$\mathbf{V_2} = \begin{bmatrix} \mathbf{V_{21}} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{V_{22}} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{V_{2N}} \end{bmatrix}_{NT \times NT}$$

Where:

$$\langle \xi_{iT}, \xi_{iT} \rangle = x_{it} x_{is} - \frac{1}{2}(x_{it} + x_{is}) + \frac{1}{3} \quad \langle \xi_{iT}, \xi_{iT} \rangle = y_{1it} y_{1is} - \frac{1}{2}(y_{1it} + y_{1is}) + \frac{1}{3}$$

10. Selecting the optimal smoothing parameter (λ) using the GCV method.

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

11. Estimating function in PWLS (Weighter $\mathbf{\Sigma}_{DM}$) and PLS ($\mathbf{\Sigma}_{TM}$)

The PLS function is as follows.

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

The PWLS function is as follows.

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

12. Calculate the MSE on the function with the PWLS (weighting $\mathbf{\Sigma}_{DM}$) and PLS (weighting $\mathbf{\Sigma}_{TM}$) approaches according to the equation below.

$$MSE = \frac{1}{2NT} \left(\sum_{k=1}^{2} \sum_{i=1}^{N} \sum_{t=1}^{T} e_{kit}^{\ 2} \right)$$

13. Comparing the results of the path $\hat{f}$ with PWLS (weighting $\mathbf{\Sigma}_{DM}$) and PLS (weighting $\mathbf{\Sigma}_{TM}$) approaches with Relative Efficiency (ER) according to the equation below.

$$eff(\hat{f}_{DM}, \hat{f}_{TM}) = \frac{V(\hat{f}_{TM})}{V(\hat{f}_{DM})}$$

Where the MSE used is the average MSE for U repetition.

3 Results and Discussion

3.1 Comparison of Estimating Path Nonparametric Truncated Spline

Following is an example of a comparison of path nonparametric truncated spline at a moderate level of autocorrelation (|0.41–0.60|) in the first and second path of the 4th subject shown in figure 2 and figure 3:

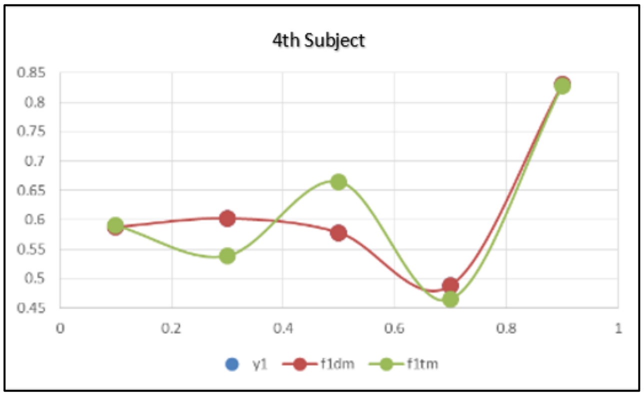

Fig. 2: Function Path on the 4th Subject Path First Moderate Autocorrelation Level.

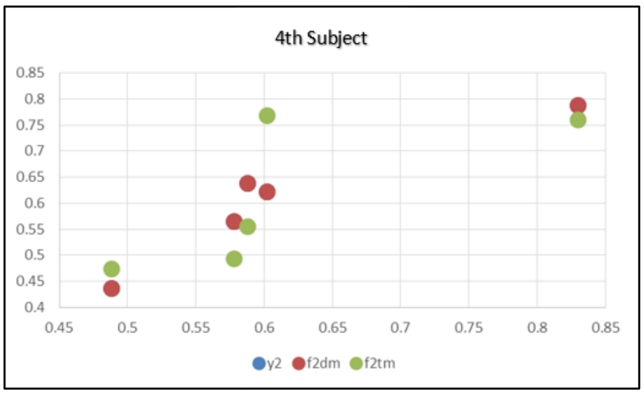

Fig.3: Function Path on the 4th Subject of the Second Path of Moderate Autocorrelation.

In Figures 2 and 3, the actual value denoted by the blue dot is not visible because it coincides with the predicted results of the PWLS (DM) function. The MSE value for the model that considers PWLS (DM) autocorrelation is 0.00004514 while for the model that does not consider PLS (TM) autocorrelation is 0.001569191.

3.2 Relative Efficiency of DM and TM Functions

Comparison of path nonparametric truncated spline with PWLS (DM Weighting) and PLS (TM Weighting) approaches was carried out with Relative Efficiency.

Table 1: The Relative Efficiency of the DM Estimator Function on TM.

Degree Autocorrelation	T=5	T=15	T=20
|0.01-0.20|	1.313838	1.007006	0.846229
|0.21-0.40|	2.305355	2.312333	1.676502
|0.41-0.60|	22.14056	5.812017	5.412136
|0.61-0.80|	10.55138	5.670869	7.491817

Table 1 provides information that at a very low level of autocorrelation with the number of observations in five subjects (T=5), the PWLS approach is better than PLS in estimating the function path (ER>1) . This also applies to other autocorrelation conditions because the Relative Efficiency value is more than one.

At a very low level of autocorrelation with the number of observations in fifteen subjects (T=15), the PWLS approach was as good as PLS (ER=1). However, in other autocorrelation conditions, the estimation of the path PWLS approach is better than PLS because the Relative Efficiency value is already greater than one.

At a very low level of autocorrelation with the number of observations in twenty subjects (T=20), the PLS approach was better than the PWLS approach in estimating the path (ER<1). However, in other autocorrelation conditions, the estimation of the path function of the PWLS approach is better than PLS because the Relative Efficiency value is already greater than one.

Based on the Relative Efficiency (ER) value obtained for the number of observations in five subjects (T=5), it can be concluded that the estimation of path using the PWLS approach is better than PLS for all levels of autocorrelation from the lowest to the highest.

In the number of observations in fifteen subjects (T=15), path estimation using the PWLS approach was better than PLS for all levels of autocorrelation except for very low autocorrelation levels because at very low autocorrelation levels PWLS and PLS were equally good (ER was close to one).

In the number of observations in twenty subjects (T=20) with a very low autocorrelation level, the estimation of path using the PLS approach is better than PWLS. On the other level of autocorrelation, path estimation using PWLS approach is better than PLS.

Based on the Relative Efficiency value at a very low level of autocorrelation and several conditions on the number of observations in the subject, it can be concluded that the estimation of path using the PLS approach will be better than PWLS when the number of observations in the subject increases.

4 Conclusion and Suggestions

Based on the Relative Efficiency (RE) value obtained for the number of observations in five subjects (T=5), it can be concluded that the estimation of path using the PWLS approach is better than PLS for all levels of autocorrelation from the lowest to the highest. In the number of observations in fifteen subjects (T=15), path estimation using the PWLS approach was better than PLS for all levels of autocorrelation except for very low autocorrelation levels because at very low autocorrelation levels PWLS and PLS were equally good (ER was close to one).

In the number of observations in twenty subjects (T=20) with a very low autocorrelation level, the estimation of path using the PLS approach is better than PWLS. On the other level of autocorrelation, path estimation using PWLS approach is better than PLS. Based on the Relative Efficiency value at a very low level of autocorrelation and several conditions on the number of observations in the subject, it can be concluded that the estimation of path using the PLS approach will be better than PWLS when the number of observations in the subject increases. If there is data that has conditions like in this study, the data can be analyzed by path.

Acknowledgement
Thank you to all parties for their support and input in the preparation of this research. All the support and input given is very useful for the perfection of this research.

References

[1] Al-Musadieq, M., Nurjannah, N., Raharjo, K., Solimun, S., and Fernandes, A. A. R., *The Mediating Effect of Work Motivation on The Influence of Job Design and Organizational Culture Against HR Performance*. Journal of Management Development., **37(6)**,452-469, (2018).

[2] Budiantara, I.N., *Optimum Journal of Basic Sciences*. Model spline with knots. University of Jember, Indonesia., (2011).

[3] Eubank, R.L., *Nonpatametric Regression and Spline Smoothing*. Marcel Dekker, New York., (1999).

[4] Fernandes, A.A.R., Budiantara, I.N., Otok, B.W., and Suhartono., *Spline Estimator for Bi-Responses and Multi-Predictors Nonparamtric Regression Model in Case of Longitudinal Data*. Journal of Mathematics and Statistics.,**11(2)**, 61-69, (2015).

[5] Fernandes, A. A. R., Budiantara, I. N. I., Otok, B. W., and Suhartono., *Reproducing Kernel Hilbert Space for Penalized Regression Multi-Predictors: Case in Longitudinal Data. International Journal of Mathematical Analysis*., **8(40)**, 1951-1961, (2014).

[6] Fernandes, A. A. R. and Solimun., *Moderating Effects Orientation and Innovation Strategy on The Effect of Uncertainty on The Performance of Business Environment*. International Journal of Law and Management., **59(6)**, 1211-1219, (2017).
[7] Fernandes, A. A. R., Hutahayan, B., Arisoesilaningsih, E., Yanti, I., Astuti, A. B., and Amaliana, L., *Comparison of Curve Estimation of The Smoothing Spline Nonparametric Function Path Based on PLS and PWLS In Various Levels of Heteroscedasticity.* In IOP Conference Series: Materials Science and Engineering, IOP Publishing., **546(5)**, 052024, (2019, June).
[8] Fernandes, A.A.R., Solimun, and Arisoesilaningsih, E., *Estimation of Spline in Nonparametric Analysis Based On Penalized Weighted Least Square (PWLS).* International Conference and Workshop on Mathematical Analysis and Its Application., **020030**, 1-9, (2017).
[9] Fernandes, A. A. R., and Taba, I. M., *Welding Technology as The Moderation Variable in The Relationships Between Government Policy and Quality of Human Resources and Workforce Competitiveness*. Journal of Science and Technology Policy Management., **10(1)**, 58-72, (2018).
[10] Gujarati, D.N. and Dawn C.P., *Basic Econometrics*, 5th Ed. New York: McGraw Hill, (2008).
[11] Hakim, W., and Fernandes, A.A.R., *Moderation Effect of Organizational Citizenship Behavior on The Performance of Lecturers. Journal of Organizational Change Management*., **30(7)**, 1136-1148, (2017).
[12] Hidayat, M. F. and Achmad, R. F. A., *Estimation of Truncated Spline Function In Non-Parametric Path Analysis Based on Weighted Least Square (WLS).* In IOP Conference Series: Materials Science and Engineering, IOP Publishing., **546(5)**, 052027, (2019, June).
[13] Hidayat, R., Yuliani, Y., and Sam, M., *Nonparametric Regression Model with Spline Truncated*. Proceedings., **3(1)**, 203-210, (2017).
[14] Kirono, I., Armanu, A., Hadiwidjojo, D., and Solimun, S., *Logistics Performance Collaboration Strategy and Information Sharing with Logistics Capability as Mediator Variable (Study in Gafeksi East Java Indonesia).* International Journal of Quality & Reliability Management., **36(8)**, 1301-1317, (2019).
[15] Purbawangsa, I. B. A., Solimun, S., Fernandes, A. A. R., and Rahayu, S. M., *Corporate Governance, Corporate Profitability Toward Corporate Social Responsibility Disclosure and Corporate Value (Comparative Study in Indonesia, China and India Stock Exchange in 2013-2016).* Social Responsibility Journal., **16(7)**, 983-999, (2019).
[16] Raharjo, K., Nurjannah, N., Solimun, S., and Fernandes, A. A. R., *The Influence of Organizational Culture and Job Design on Job Commitment and Human Resource Performance.* Journal of Organizational Change Management., **31(7)**, 1346-1367, (2018).
[17] Solimun and Fernandes, A.A.R., *Investigation the Mediating Variable: What is Necessary? (Case Study in Management Research).* International Journal of Law and Management., **59(6)**, 1059-1067, (2017).
[18] Solimun and Fernandes, A.A.R., *Investigation of Instrument Validity: Investigate the Consistency Between Criterion and Unidimensional In Instrument Validity (Case Study In Management Research).* International Journal of Law and Management., **59(6)**, 1203-1210, (2017).
[19] Sri, M. and Solimun, S., *The Relationship Between Audit Quality and Risk Taking Toward Value Creation In Indonesia.* Journal of Accounting in Emerging Economies., **9(2)**, 251-267, (2019).
[20] Sumardi, S. and Fernandes, A.A.R., *The Mediating Effect of Service Quality and Organizational Commitment on The Effect of Management Process Alignment on Higher Education Performance in Makassar, Indonesia*. Journal of Organizational Change Management., **31(2)**, 410-425, (2018).
[21] Sumardi, S., and Fernandes, A. A. R., *The Influence of Quality Management on Organization Performance: Service Quality and Product Characteristics as A Medium.* Property Management., **38(3)**, 383-403, (2020).
[22] Wackerly, D., Mendenhall, W., and Scheaffer, R.L., *Mathematical Statistics with Applications, Cengage Learning*., (2014).
[23] Wu, H. and Zhang, J.T., *Nonparametric Regression Methods for Longitudinal Data Analysis,* John Wiley and Sons, New Jersey., (2006).