Mathematics and Statistics 10(6): 1293-1303, 2022 http://www.hrpub.org
DOI: 10.13189/ms.2022.100615

Estimation of Nonparametric Path Fourier Series and Truncated Spline Ensemble Models

Atiek Iriany*, Adji Achmad Rinaldo Fernandes

Department of Statistics, University of Brawijaya, Malang, 65145, Indonesia

Received September 13, 2022; Revised October 28, 2022; Accepted November 15, 2022

Cite This Paper in the Following Citation Styles

(a): *[1] Atiek Iriany, Adji Achmad Rinaldo Fernandes , "Estimation of Nonparametric Path Fourier Series and Truncated Spline Ensemble Models," Mathematics and Statistics, Vol. 10, No. 6, pp. 1293 - 1303, 2022. DOI: 10.13189/ms.2022.100615.*

(b): *Atiek Iriany, Adji Achmad Rinaldo Fernandes (2022). Estimation of Nonparametric Path Fourier Series and Truncated Spline Ensemble Models. Mathematics and Statistics, 10(6), 1293 - 1303. DOI: 10.13189/ms.2022.100615.*

Abstract To ascertain whether there is a causal connection between exogenous and endogenous factors, one method is to perform path analysis. The linearity assumption is the one that has the power to alter the model. The model's shape is impacted by the linearity assumption. The path analysis is parametric if the linearity assumption is true, but non-parametric path analysis is used if the non-linear form is unknown and there is no knowledge of the data pattern. If the non-linear form is unknown and there is no knowledge of the data pattern, non-linear path analysis is used. This study's goal was to calculate the nonparametric route function using a combination of truncated spline and Fourier series methods. The findings demonstrated that nonparametric path analysis only in cases where the linearity presumption is violated can one employ the Fourier series and truncated spline. Then, using the Ordinary Least Square (OLS) approach, the estimator of Nonparametric Regression-Based Path Analysis was obtained, delivering an estimation result that is not unique because it makes use of a nonparametric approach. The contribution of this paper can be used as reference material, especially analysis in statistics. With this paper, it is hoped that it can be applied in various fields. Suggestions for further research can develop this research with other models.

Keywords Fourier Series, Nonparametric Path Analysis, Regression Analysis, Truncated Spline

1. Introduction

In 1934, Wright developed route analysis for the first time [1]. Path analysis is used to assess the model of the relationship between variables in the form of cause and effect [2]. A route analysis is one way determining if there is a causal relationship between exogenous and endogenous components [3]. In addition to establishing the direct influence of exogenous factors on endogenous variables, path analysis is used to determine if exogenous variables have an indirect effect on endogenous variables through mediating endogenous variables [4]. Mediation models are those in which the effect of an antecedent or independent variable (X) on the dependent variable (Y) is communicated via a third intervening or mediating variable (M) [5].

The regression technique approach known as nonparametric path analysis assumes that the regression function's curve's form is unknown. Nonparametric route analysis curves merely assume that the curve is smooth [6]. It is presumed that the function space contains the function curve [7]. The difference between parametric and nonparametric approaches is that the former seeks to force the data to follow a particular pattern, while the latter provides the data with the freedom to seek out its regression curve pattern, making the latter more flexible and unbiased [8]. The use of path analysis for parametric, non-linear and parametric methods can be seen in Figure 1.

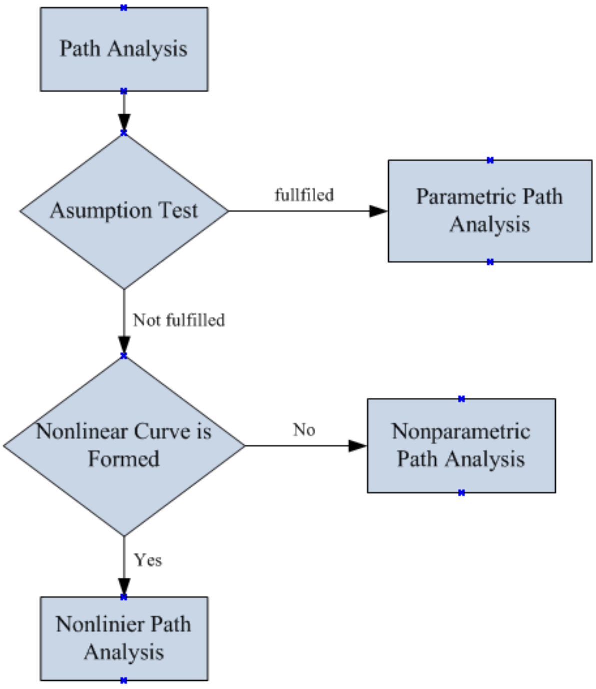

Figure 1. Tree diagrams use parametric path analysis, nonlinear path analysis, and nonparametric path analysis

In nonparametric path analysis, smoothing parameters are needed to determine the size of the smoothness or roughness of the curve in describing the data. The smoothing parameter of a very small value provides a very rough regression curve estimator. On the other hand, if the smoothing parameter value is very large, it will produce a very smooth nonparametric regression curve estimator [9]. As a result, in estimating the nonparametric regression function, it is necessary to choose the optimal smoothing parameter to obtain the most suitable estimator for the data. [10] used the Generalized Cross Validation (GCV) and Cross Validation (CV) methods as smoothing parameters in nonparametric regression analysis with Fourier series.

Nonparametric regression is the basis of nonparametric path analysis for the pattern of relationships between exogenous, endogenous dependent, and endogenous mediating variables [11]. Several approaches can be used in nonparametric path analysis, namely using Moving Averages, Fourier Series, Splines, Kernels, Local Polynomials, and Wavelets [12,13].

Since the Fourier series is a trigonometric polynomial with adjustable characteristics, the model can accommodate local data characteristics. Some studies have studied Fourier series nonparametric regression modeling. Hidayat et al. [7] states that splines are part of a polynomial that has continuous and segmented (truncated) properties. Amalia & Nur [13] conducted research by utilizing a nonparametric Fourier series regression approach and using the GCV and CV methods. Adrianingsih et al. [10] found that the model nonparametric regression with Fourier series estimator tends to follow the actual data pattern. Dani & Adrianingsih [8] conducted a study and showed that the nonparametric regression model with the Fourier series estimator tends to follow the actual data pattern. Both approaches have the advantage that they can be precise or close to the actual data pattern, so when the two approaches are combined it is hoped that they can make more flexible function estimates for data patterns. Based on the above background, in this study, we will predict a function estimator for a nonparametric path function using Fourier series and truncated spline approaches. The contribution of this paper can be used as reference material, especially analysis in statistics. With this paper, it is hoped that it can be applied in various fields.

2. Literature Review

2.1. Nonparametric Regression Analysis

Modeling data patterns using nonparametric regression is very versatile, minimizing the subjectivity of the researcher [14]. If the normality, non-multicollinearity, and homoscedasticity assumptions of parametric regression analysis were not met, nonparametric regression analysis was utilized [15]. This strategy works effectively for drawing conclusions when there is little to no prior knowledge about the regression curve or data pattern [16].

When parametric regression is used on unknown data, there is a risk that it will produce an unrepresentative regression model, which will lead to erroneous conclusions from hypothesis testing [17]. The pattern of the connection between the predictor variable (X) and the response variable (Y), whose curve form is not yet known, can be determined using a nonparametric regression model [18]:

$$Y_i = \hat{f}\left(x_i\right) + \varepsilon_i \qquad (1)$$

The study is carried out using parametric route analysis if the linearity presumption is true. However, the analysis uses non-linear and/or non-parametric pathways if the linearity supposition is violated. Where:

Y_i: the response variable's value
x_i: the predictor variable's value
$\hat{f}$: regression curve
i: 1, 2, ..., n.
N: the abundance of observations.
ε_i: erors in the i-th observation.

2.2. Nonparametric Path Analysis Based on Fourier Series

The abstract should concisely state the purpose of the investigation. Fourier Series Route Analysis is one method that can be applied to nonparametric path analysis.

Since the Fourier series is a flexible trigonometric polynomial, it may easily adjust to the local nature of the [19]. Data with a trigonometric distribution (sine and cosine) can be overcome by the Fourier series, which is an advantage [10]. According to the Nonparametric Regression model in equation (1), the Fourier Series approximates $f(x_i)$ $f(x_i)$as follows:

Minimize $\varepsilon_i{}^2$

$$Min\{\sum_{i=1}^{n}\varepsilon_i^2\} = Min\left\{\sum_{i=1}^{n}\left(y_i - f(x_i)\right)^2\right\} \quad (2)$$

Depending on how smooth the function f is, the following penalty is also applied in addition to minimizing equation (2):

$$\int_0^{\pi}\frac{2}{\pi}\left(f^{(2)}(x)\right)^2 dx \quad (3)$$

This means that by using Penalized Least Squares (PLS) to finish the optimization, the estimator for the regression curve f can be obtained.

$$Min\left\{n^{-1}\sum_{i=1}^{n}\left(y_i - f(x_i)\right)^2 + \lambda\int_0^{\pi}\frac{2}{\pi}\left(f^{(2)}(x)\right)^2 dx\right\} \quad (4)$$

To resolve equation (4), start by determining the value of P(a):

$$P(a) = \int_0^{\pi}\frac{2}{\pi}\left(f^{(2)}(x)\right)^2 dx$$

$$P(a) =$$

$$\frac{2}{\pi}\int_0^{\pi}\left(\sum_{k=1}^{K}(k^2 a_k coskx)^2 + 2\sum_{k<j}^{K}\sum_{k<j}^{K}(k^2 a_k coskx)\left(j^2 a_j cosjx\right)\right)dx$$

$$P(a) = \sum_{k=1}^{K}k^4 a_k^2 \quad (5)$$

Smoothing parameter can control between the goodness of fit and smoothness of the function. For a very large, a very smooth solution function will be obtained, while for a very small, a very coarse solution will be obtained. f can be approximated by the function x because it is a continuous function, with:

$$f(x) = bx + \frac{1}{2}a_0 + \sum_{k=1}^{K}a_k coskx \quad (6)$$

Using equation (6), it is possible to write

$$Min\left\{n^{-1}\sum_{i=1}^{n}\left(y_i - f(x_i)\right)^2 + \lambda\int_0^{p}\frac{2}{p}\left(f^{(2)}(x)\right)^2 dt\right\}$$

$$Min\left\{n^{-1}\sum_{i=1}^{n}\left[y_i - bx - \frac{1}{2}a_0 - \sum_{k=1}^{K}a_k coskx\right]^2 + \lambda\sum_{k=1}^{K}k^4 a_k^2\right\} \quad (7)$$

$$Min\{n^{-1}(y - \boldsymbol{Xa})'^{(y-\boldsymbol{Xa})} + \lambda\boldsymbol{a'Da}\}$$

$$= Min\{n^{-1}y'y - n^{-1}a'X'y - n^{-1}(a'X'y)' + a'(n^{-1}X'X + \lambda D)a\}$$

Where:

$$D = diag(0,0,1^4,2^4,\dots,K^4) \quad (8)$$

If equation (7) is known as Q(a), then we can obtain it by partially subtracting Q(a) from a and equating it to zero:

$$\frac{\partial Q(a)}{\partial a} = 0 - 2n^{-1}\boldsymbol{X}'y + 2\left(n^{-1}\boldsymbol{X}'\boldsymbol{X} + \lambda\boldsymbol{D}\right)\boldsymbol{a}s \quad (9)$$

$$\hat{a}(\lambda) = \left(n^{-1}X'X + \lambda D\right)^{-1} n^{-1}X'y$$

It can be expressed as a matrix based on the characteristics of the Fourier Series estimator in equation (9).

$$\underset{\sim}{f} = \mathbf{X}\underset{\sim}{a} + \underset{\sim}{e} \quad (10)$$

$$\hat{\underset{\sim}{f}} = \mathbf{X}\underset{\sim}{a} \quad (11)$$

Where:

$$\underset{\sim}{a} = \left(b, \frac{1}{2}a_0, a_1, \dots, a_K\right) \quad (12)$$

$$\mathbf{X} = \begin{pmatrix} x_1 & 1 & \cos x_1 & \cos 2x_1 & \cdots & \cos Kx_1 \\ x_2 & 1 & \cos x_2 & \cos 2x_2 & \cdots & \cos Kx_2 \\ x_3 & 1 & \cos x_3 & \cos 2x_3 & \cdots & \cos Kx_3 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ x_n & 1 & \cos x_n & \cos 2x_n & \cdots & \cos Kx_n \end{pmatrix} \quad (13)$$

If equation (11) is translated, it will look as follows:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_1 \\ \vdots \\ y_M \end{pmatrix} = \begin{pmatrix} x_1 & 1 & \cos x_1 & \cos 2x_1 & \cdots & \cos Kx_1 \\ x_2 & 1 & \cos x_2 & \cos 2x_2 & \cdots & \cos Kx_2 \\ x_3 & 1 & \cos x_3 & \cos 2x_3 & \cdots & \cos Kx_3 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ x_n & 1 & \cos x_n & \cos 2x_n & \cdots & \cos Kx_n \end{pmatrix}\begin{pmatrix} b \\ \frac{1}{2}a_0 \\ a_1 \\ \vdots \\ a_K \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ e_K \end{pmatrix}$$

Consequently, the following is how the estimator for the Fourier series' nonparametric path function is obtained [14]:

$$\hat{f}_\lambda(x_i) = \hat{b}(\lambda)x_i + \frac{1}{2}\hat{a}_0(\lambda) + \sum_{k=1}^{K}\hat{a}_k(\lambda)\cos kx_i \quad (14)$$

2.3. Truncated Spline-based Nonparametric Path Analysis

The spline is a part of regression analysis, more specifically from nonparametric regression and semiparametric regression. Research in the field of splines that is independent and has character, requires a comprehensive process, following stages and very long [20]. In modeling data patterns, nonparametric spline regression has advantages, namely: (a) Splines have very special and very good statistical interpretations. The spline model was obtained from the optimization of the PLS method. (b) The spline can handle smooth data/functions. (c) The spline is very good at handling data whose behavior changes at specific sub-intervals. (d) Spline is very good at generalizing intricate and complex statistical modeling [21] The Truncated Spline polynomial is one of the spline models that applications find interesting [22].

Splines are parts or pieces of a polynomial that have segmented and continuous (truncated) [10]. The advantage of Truncated Spline polynomial regression is that it tends to find its form of estimating the regression curve [23]. This can happen because the spline has a common fusion point that shows a pattern of data behavior called knot points. The knot point allows for effective adjustment of local characteristics so that the spline has high flexibility. In addition, the advantage of the Truncated Spline polynomial regression is that it is objective, so it can minimize the element of subjectivity. The optimization is done by using the least square method so that it has easy, simple, and good mathematical calculations to help with statistical calculations.

Simple equations are used in a nonparametric path analysis along with exogenous, mediating, and endogenous variables. Equation (15) is a straightforward path analysis model based on the undefined Truncated Spline approach [24]

$$\begin{aligned} y_{1i} &= f_{1.1}\left(x_{1i}\right)+\varepsilon_{1i} \\ y_{2i} &= f_{1.2}\left(x_{1i}\right)+f_{2.2}\left(x_{1i}\right)+\varepsilon_{1i} \end{aligned} \quad (15)$$

In equation (15) it can be shown in the form of a matrix as in equation (16):

$$\begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2n} \end{pmatrix} = \begin{pmatrix} f_{1.1}\left(x_{11}\right) \\ f_{1.1}\left(x_{12}\right) \\ \vdots \\ f_{1.1}\left(x_{1n}\right) \\ f_{1.2}\left(x_{11}\right)+f_{2.2}\left(y_{11}\right) \\ f_{1.2}\left(x_{12}\right)+f_{2.2}\left(y_{12}\right) \\ \vdots \\ f_{1.2}\left(x_{1n}\right)+f_{2.2}\left(y_{1n}\right) \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{1n} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \vdots \\ \varepsilon_{2n} \end{pmatrix} \quad (16)$$

The truncated spline has parameters determined by polynomial degrees and knots as smoothing parameters. Equation (17) illustrates how the X matrix is designed for the nonparametric path analysis:

$$X = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix}_{(2+p+K)\times 2n} \quad (17)$$

The Truncated Spline function is the foundation of the X matrix design. Equation (18) and (19) shows how the X matrix is constructed:

$$X = \begin{pmatrix} 1 & x_1 & \left(x_1 - t_{x1}\right) \\ 1 & x_2 & \left(x_2 - t_{x1}\right) \\ \vdots & \vdots & \vdots \\ 1 & x_n & \left(x_n - t_{x1}\right) \end{pmatrix} \quad (18)$$

$$X = \begin{pmatrix} 1 & x_1 & \left(x_1 - t_{x1}\right)_+ & y_{11} & \left(y_{11} - t_{y1}\right)_+ \\ 1 & x_2 & \left(x_2 - t_{x1}\right)_+ & y_{12} & \left(y_{12} - t_{y1}\right)_+ \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & \left(x_n - t_{x1}\right)_+ & y_{1n} & \left(y_{1n} - t_{y1}\right)_+ \end{pmatrix} \quad (19)$$

Equation (16) uses the variance-variance matrix denoted by equation (20) and a random variable whose error is assumed to be normally distributed to 2n variables:

$$\begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2n} \end{pmatrix} = \begin{pmatrix} f_{1.1}\left(x_{11}\right) \\ f_{1.1}\left(x_{12}\right) \\ \vdots \\ f_{1.1}\left(x_{1n}\right) \\ f_{1.2}\left(x_{11}\right)+f_{2.2}\left(y_{11}\right) \\ f_{1.2}\left(x_{12}\right)+f_{2.2}\left(y_{12}\right) \\ \vdots \\ f_{1.2}\left(x_{1n}\right)+f_{2.2}\left(y_{1n}\right) \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{1n} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \vdots \\ \varepsilon_{2n} \end{pmatrix} \quad (20)$$

3. Methods

This study was carried out to advance the theory of statistical modeling, specifically regression modeling. This study combined the truncated spline and Fourier series methods to derive the lemma theorem from nonparametric path analysis. Using the Ordinary Least Square (OLS) method, the nonparametric path analysis of the Fourier series function was estimated.

4. Result and Discussion

The Fourier series is a polynomial based on a cosine function that has flexibility so that it can adapt effectively to the local nature of the data. The Fourier series is good for describing periodic curves such as sine and cosine waves. The truncated spline is the sum of the polynomial functions with a truncated function. The truncated spline is used to show the function/data where there is a change in the behavior pattern of the curve that varies at different intervals.

Lemma 4.1. Forms of a Simple Fourier Series and Truncated Spline Nonparametric Path Analysis Model

If given paired data $(X_{1i}, X_{2i}, Y_{1i}, Y_{2i})$, the relationship between $(X_{1i}, X_{2i}, Y_{1i}, Y_{2i})$ them is modeled by additive nonparametric path analysis. Equation (21) shows a combined nonparametric path analysis model of the Fourier series and truncated spline.

$$y_{1i} = f(x_{1i}) + \hat{g}(x_{2i}) + \varepsilon_i$$
$$y_{2i} = f(x_{1i}, y_{1i}) + \hat{g}(x_{2i}, y_{1i}) + \varepsilon_i \quad (21)$$

Where y_{1i} and y_{2i} are response variables, $f(x_{1i})$, $\hat{g}(x_{2i})$, $f(x_{1i}, y_{1i})$, and $\hat{g}(x_{2i}, y_{1i})$ are path curves of unknown shape and ε_i is a random error, with a mean of zero and a variance σ^2 of assumed to have an independent normal distribution.

Let $f(x_{1i})$ be approximated by a Fourier series function. The model can be seen in equation (22).

$$\hat{f}_{1i} = \frac{1}{2}a_{01} + \sum_{j=1}^{p} b_{1j}X_{1i} + \sum_{j=1}^{p} b_{1j}X_{2i} + \sum_{k=1}^{K} \gamma_{1(2\times k)} cosKY_{1i}$$
$$\hat{f}_{2i} = \frac{1}{2}a_{02} + \sum_{j=1}^{p} b_{2j}X_{1i} + \sum_{j=1}^{p} b_{2j}X_{2i} + \sum_{k=1}^{K} \gamma_{1(2\times k)} cosKY_{1i} + \sum_{k=1}^{K} \gamma_{1(k)} cosKY_{1i} \quad (22)$$

Let $\hat{g}(x_{2i})$ be modeled by a short spline function. Equation (23) depicts the model in detail.

$$\hat{g}_{1i} = \sum_{v=1}^{m} \beta_{vj} x_{ji}^{v} + \sum_{u=1}^{r} \alpha_{uj} \left(x_{ji} - t_{uj} \right)_{+}^{m}$$
$$\hat{g}_{2i} = \sum_{v=1}^{m} \beta_{vj} x_{ji}^{v} + \sum_{u=1}^{r} \alpha_{uj} \left(x_{ji} - t_{uj} \right)_{+}^{m} + \alpha_{uj} \left(y_{ji} - t_{uj} \right)_{+}^{m} \quad (23)$$

with

$$\left(x_{ji} - t_{uj} \right)_{+}^{m} = \begin{cases} \left(x_j - t_{uj} \right); x_j \geq t_{uj} \\ 0 \quad ; x_j < t_{uj} \end{cases}$$

The combined nonparametric path analysis model of the Fourier series and truncated spline is shown in equation (24).

$$\hat{f}_{1i} = \frac{1}{2}a_{01} + \sum_{j=1}^{p} b_{1j} X_{1i} + \sum_{j=1}^{p} b_{1j} X_{2i} + \sum_{k=1}^{K} \gamma_{1(2\times k)} \cos KX_{1i} + \sum_{v=1}^{m} \beta_{vj} x_{ji}^{v} + \sum_{u=1}^{r} \alpha_{uj} \left(x_{ji} - t_{uj} \right)_{+}^{m}$$
$$\hat{f}_{2i} = \frac{1}{2}a_{02} + \sum_{j=1}^{p} b_{2j} X_{1i} + \sum_{j=1}^{p} b_{2j} X_{1i} + \sum_{k=1}^{K} \gamma_{1(2\times k)} \cos KX_{1i} + \sum_{v=1}^{m} \beta_{vj} x_{ji}^{v} + \sum_{u=1}^{r} \alpha_{uj} \left(x_{ji} - t_{uj} \right)_{+}^{m} + \alpha_{uj} \left(y_{ji} - t_{uj} \right)_{+}^{m} \quad (24)$$

When translated, equation (24) is shown in the equation below:

If K=2, m=1, and t=1

$$\hat{f}_{1i} = \frac{1}{2} a_{01} + b_{11j}X_{1i} + \gamma_{11} \cos X_{1i} + \gamma_{21} \cos 2\, X_{1i} + b_{21j}X_{2i} + \alpha_{11}(x_{2i} - t_{11})_{+} \quad (25)$$
$$\hat{f}_{2i} = \frac{1}{2} a_{02} + b_{12j}X_{1i} + \gamma_{12} \cos X_{1i} + \gamma_{22} \cos 2\, X_{1i} + b_{22j}X_{2i} + \alpha_{12}(x_{2i} - t_{12})_{+} + b_{32j}Y_{1i} + \alpha_{22}(Y_{1i} - t_{22})_{+}$$

If K=2, m=1, and t=2

$$\hat{f}_{1i} = \frac{1}{2}a_{01} + b_{11j}X_{1i} + \gamma_{11}\cos X_{1i} + \gamma_{21}\cos 2X_{1i} + b_{21j}X_{2i} + \alpha_{11}\left(x_{2i} - t_{11}\right)_{+} + \alpha_{21}\left(x_{2i} - t_{21}\right)_{+}$$
$$\hat{f}_{2i} = \frac{1}{2}a_{02} + b_{12j}X_{1i} + \gamma_{12}\cos X_{1i} + \gamma_{22}\cos 2X_{1i} + b_{22j}X_{2i} + \alpha_{12}\left(x_{2i} - t_{12}\right)_{+} + \alpha_{22}\left(x_{2i} - t_{22}\right)_{+} + b_{32j}Y_{1i} + \alpha_{32}\left(Y_{1i} - t_{32}\right)_{+} + \alpha_{42}\left(Y_{1i} - t_{42}\right)_{+} \quad (26)$$

If K=2, m=2, and t=1

$$\hat{f}_{1i} = \frac{1}{2}a_{01} + b_{11j}X_{1i} + \gamma_{11}\cos X_{1i} + \gamma_{21}\cos 2X_{1i} + b_{21j}X_{2i} + b_{31j}X_{2i}^2 + \alpha_{11}\left(x_{2i} - t_{11}\right)_{+}^2$$
$$\hat{f}_{2i} = \frac{1}{2}a_{02} + b_{12j}X_{1i} + \gamma_{12}\cos X_{1i} + \gamma_{22}\cos 2X_{1i} + b_{22j}X_{2i} + b_{32j}X_{2i}^2 + \alpha_{12}\left(x_{2i} - t_{12}\right)_{+}^2 + b_{42j}Y_{1i} + b_{52j}Y_{1i}^2 + \alpha_{22}\left(Y_{1i} - t_{22}\right)_{+}^2 \quad (27)$$

If K=2, m=2, and t=2

$$\hat{f}_{1i} = \frac{1}{2}a_{01} + b_{11j}X_{1i} + \gamma_{11}\cos X_{1i} + \gamma_{21}\cos 2X_{1i} + b_{21j}X_{2i} + b_{31j}X_{2i}^2 + \alpha_{11}\left(x_{2i} - t_{11}\right)_{+}^2 + \alpha_{21}\left(x_{2i} - t_{21}\right)_{+}^2$$
$$\hat{f}_{2i} = \frac{1}{2}a_{02} + b_{12j}X_{1i} + \gamma_{12}\cos X_{1i} + \gamma_{22}\cos 2X_{1i} + b_{22j}X_{2i} + b_{32j}X_{2i}^2 + \alpha_{12}\left(x_{2i} - t_{12}\right)_{+}^2 + \alpha_{22}\left(x_{2i} - t_{22}\right)_{+}^2 + b_{42j}Y_{1i} + b_{52j}Y_{1i}^2 + \alpha_{32}\left(Y_{1i} - t_{32}\right)_{+}^2 \quad (28)$$

If K=3, m=1, and t=1

$$\hat{f}_{1i} = \frac{1}{2}a_{01} + b_{11j}X_{1i} + \gamma_{11}\cos X_{1i} + \gamma_{21}\cos 2X_{1i} + \gamma_{31}\cos 3X_{1i} + b_{21j}X_{2i} + \alpha_{11}\left(x_{2i} - t_{11}\right)_{+}$$
$$\hat{f}_{2i} = \frac{1}{2}a_{02} + b_{12j}X_{1i} + \gamma_{12}\cos X_{1i} + \gamma_{22}\cos 2X_{1i} + \gamma_{32}\cos 3X_{1i} + b_{22j}X_{2i} + \alpha_{12}\left(x_{2i} - t_{12}\right)_{+} + b_{32j}Y_{1i} + \alpha_{22}\left(Y_{1i} - t_{22}\right)_{+} \quad (29)$$

If K=3, m=1, and t=2

$$\hat{f}_{1i} = \frac{1}{2}a_{01} + b_{11j}X_{1i} + \gamma_{11}\cos X_{1i} + \gamma_{21}\cos 2X_{1i} + \gamma_{31}\cos 3X_{1i} + b_{21j}X_{2i} + \alpha_{11}\left(x_{2i} - t_{11}\right)_{+} + \alpha_{21}\left(x_{2i} - t_{21}\right)_{+}$$
$$\hat{f}_{2i} = \frac{1}{2}a_{02} + b_{12j}X_{1i} + \gamma_{12}\cos X_{1i} + \gamma_{22}\cos 2X_{1i} + \gamma_{32}\cos 3X_{1i} + b_{22j}X_{2i} + \alpha_{12}\left(x_{2i} - t_{12}\right)_{+} + \alpha_{22}\left(x_{2i} - t_{22}\right)_{+} + b_{32j}Y_{1i} + \alpha_{32}\left(Y_{1i} - t_{32}\right)_{+} + \alpha_{42}\left(Y_{1i} - t_{42}\right)_{+} \quad (30)$$

If K=3, m=2, and t=1

$$\begin{aligned}\hat{f}_{1i} &= \frac{1}{2}a_{01} + b_{11j}X_{1i} + \gamma_{11}\cos X_{1i} + \gamma_{21}\cos 2X_{1i} + \gamma_{31}\cos 3X_{1i} \\ &+ b_{21j}X_{2i} + b_{31j}X_{2i}^2 + \alpha_{11}\left(x_{2i} - t_{11}\right)_+^2 \\ \hat{f}_{2i} &= \frac{1}{2}a_{02} + b_{12j}X_{1i} + \gamma_{12}\cos X_{1i} + \gamma_{22}\cos 2X_{1i} + \gamma_{32}\cos 3X_{1i} \\ &+ b_{22j}X_{2i} + b_{32j}X_{2i}^2 + \alpha_{12}\left(x_{2i} - t_{12}\right)_+^2 + b_{42j}Y_{1i} + b_{52j}Y_{1i}^2 \\ &+ \alpha_{22}\left(Y_{1i} - t_{22}\right)_+^2\end{aligned} \quad (31)$$

If K=3, m=2, and t=2

$$\begin{aligned}\hat{f}_{1i} &= \frac{1}{2}a_{01} + b_{11j}X_{1i} + \gamma_{11}\cos X_{1i} + \gamma_{21}\cos 2X_{1i} + \gamma_{31}\cos 3X_{1i} \\ &+ b_{21j}X_{2i} + b_{31j}X_{2i}^2 + \alpha_{11}\left(x_{2i} - t_{11}\right)_+^2 + \alpha_{21}\left(x_{2i} - t_{21}\right)_+^2 \\ \hat{f}_{2i} &= \frac{1}{2}a_{02} + b_{12j}X_{1i} + \gamma_{12}\cos X_{1i} + \gamma_{22}\cos 2X_{1i} + \gamma_{32}\cos 3X_{1i} \\ &+ b_{22j}X_{2i} + b_{32j}X_{2i}^2 + \alpha_{12}\left(x_{2i} - t_{12}\right)_+^2 + \alpha_{22}\left(x_{2i} - t_{22}\right)_+^2 + b_{42j}Y_{1i} \\ &+ b_{52j}Y_{1i}^2 + \alpha_{32}\left(Y_{1i} - t_{32}\right)_+^2\end{aligned} \quad (32)$$

If K=4, m=1, and t=1

$$\begin{aligned}\hat{f}_{1i} &= \frac{1}{2}a_{01} + b_{11j}X_{1i} + \gamma_{11}\cos X_{1i} + \gamma_{21}\cos 2X_{1i} + \gamma_{31}\cos 3X_{1i} + \gamma_{41}\cos 4X_{1i} \\ &+ b_{21j}X_{2i} + \alpha_{11}\left(x_{2i} - t_{11}\right)_+ \\ \hat{f}_{2i} &= \frac{1}{2}a_{02} + b_{12j}X_{1i} + \gamma_{12}\cos X_{1i} + \gamma_{22}\cos 2X_{1i} + \gamma_{32}\cos 3X_{1i} + \gamma_{42}\cos 4X_{1i} \\ &+ b_{22j}X_{2i} + \alpha_{12}\left(x_{2i} - t_{12}\right)_+ + b_{32j}Y_{1i} + \alpha_{22}\left(Y_{1i} - t_{22}\right)_+\end{aligned} \quad (33)$$

If K=4, m=1, and t=2

$$\begin{aligned}\hat{f}_{1i} &= \frac{1}{2}a_{01} + b_{11j}X_{1i} + \gamma_{11}\cos X_{1i} + \gamma_{21}\cos 2X_{1i} + \gamma_{31}\cos 3X_{1i} + \gamma_{41}\cos 4X_{1i} \\ &+ b_{21j}X_{2i} + \alpha_{11}\left(x_{2i} - t_{11}\right)_+ + \alpha_{21}\left(x_{2i} - t_{21}\right)_+ \\ \hat{f}_{2i} &= \frac{1}{2}a_{02} + b_{12j}X_{1i} + \gamma_{12}\cos X_{1i} + \gamma_{22}\cos 2X_{1i} + \gamma_{32}\cos 3X_{1i} + \gamma_{42}\cos 4X_{1i} \\ &+ b_{22j}X_{2i} + \alpha_{12}\left(x_{2i} - t_{12}\right)_+ + \alpha_{22}\left(x_{2i} - t_{22}\right)_+ + b_{32j}Y_{1i} + \alpha_{32}\left(Y_{1i} - t_{32}\right)_+ \\ &+ \alpha_{42}\left(Y_{1i} - t_{42}\right)_+\end{aligned} \quad (34)$$

If K=4, m=2, and t=1

$$\begin{aligned}\hat{f}_{1i} &= \frac{1}{2}a_{01} + b_{11j}X_{1i} + \gamma_{11}\cos X_{1i} + \gamma_{21}\cos 2X_{1i} + \gamma_{31}\cos 3X_{1i} + \gamma_{41}\cos 4X_{1i} \\ &+ b_{21j}X_{2i} + b_{31j}X_{2i}^2 + \alpha_{11}\left(x_{2i} - t_{11}\right)_+^2 \\ \hat{f}_{2i} &= \frac{1}{2}a_{02} + b_{12j}X_{1i} + \gamma_{12}\cos X_{1i} + \gamma_{22}\cos 2X_{1i} + \gamma_{32}\cos 3X_{1i} + \gamma_{42}\cos 4X_{1i} \\ &+ b_{22j}X_{2i} + b_{32j}X_{2i}^2 + \alpha_{12}\left(x_{2i} - t_{12}\right)_+^2 + b_{42j}Y_{1i} + b_{52j}Y_{1i}^2 \\ &+ \alpha_{22}\left(Y_{1i} - t_{22}\right)_+^2\end{aligned} \quad (35)$$

If K=4, m=2, and t=2

$$\begin{aligned}\hat{f}_{1i} &= \frac{1}{2}a_{01} + b_{11j}X_{1i} + \gamma_{11}\cos X_{1i} + \gamma_{21}\cos 2X_{1i} + \gamma_{31}\cos 3X_{1i} + \gamma_{41}\cos 4X_{1i} \\ &+ b_{21j}X_{2i} + b_{31j}X_{2i}^2 + \alpha_{11}\left(x_{2i} - t_{11}\right)_+^2 + \alpha_{21}\left(x_{2i} - t_{21}\right)_+^2 \\ \hat{f}_{2i} &= \frac{1}{2}a_{02} + b_{12j}X_{1i} + \gamma_{12}\cos X_{1i} + \gamma_{22}\cos 2X_{1i} + \gamma_{32}\cos 3X_{1i} + \gamma_{42}\cos 4X_{1i} \\ &+ b_{22j}X_{2i} + b_{32j}X_{2i}^2 + \alpha_{12}\left(x_{2i} - t_{12}\right)_+^2 + \alpha_{22}\left(x_{2i} - t_{22}\right)_+^2 + b_{42j}Y_{1i} \\ &+ b_{52j}Y_{1i}^2 + \alpha_{32}\left(Y_{1i} - t_{32}\right)_+^2\end{aligned} \quad (36)$$

If K=5, m=1, and t=1

$$\begin{aligned}\hat{f}_{1i} &= \frac{1}{2}a_{01} + b_{11j}X_{1i} + \gamma_{11}\cos X_{1i} + \gamma_{21}\cos 2X_{1i} + \gamma_{31}\cos 3X_{1i} + \gamma_{41}\cos 4X_{1i} \\ &+ \gamma_{51}\cos 5X_{1i} + b_{21j}X_{2i} + \alpha_{11}\left(x_{2i} - t_{11}\right)_+ \\ \hat{f}_{2i} &= \frac{1}{2}a_{02} + b_{12j}X_{1i} + \gamma_{12}\cos X_{1i} + \gamma_{22}\cos 2X_{1i} + \gamma_{32}\cos 3X_{1i} + \gamma_{42}\cos 4X_{1i} \\ &+ \gamma_{52}\cos 5X_{1i} + b_{22j}X_{2i} + \alpha_{12}\left(x_{2i} - t_{12}\right)_+ + b_{32j}Y_{1i} + \alpha_{22}\left(Y_{1i} - t_{22}\right)_+\end{aligned} \quad (37)$$

If K=5, m=1, and t=2

$$\begin{aligned}\hat{f}_{1i} &= \frac{1}{2}a_{01} + b_{11j}X_{1i} + \gamma_{11}\cos X_{1i} + \gamma_{21}\cos 2X_{1i} + \gamma_{31}\cos 3X_{1i} + \gamma_{41}\cos 4X_{1i} \\ &+ \gamma_{51}\cos 5X_{1i} + b_{21j}X_{2i} + \alpha_{11}\left(x_{2i} - t_{11}\right)_+ + \alpha_{21}\left(x_{2i} - t_{21}\right)_+ \\ \hat{f}_{2i} &= \frac{1}{2}a_{02} + b_{12j}X_{1i} + \gamma_{12}\cos X_{1i} + \gamma_{22}\cos 2X_{1i} + \gamma_{32}\cos 3X_{1i} + \gamma_{42}\cos 4X_{1i} \\ &+ \gamma_{52}\cos 5X_{1i} + b_{22j}X_{2i} + \alpha_{12}\left(x_{2i} - t_{12}\right)_+ + \alpha_{22}\left(x_{2i} - t_{22}\right)_+ + b_{32j}Y_{1i} \\ &+ \alpha_{32}\left(Y_{1i} - t_{32}\right)_+ + \alpha_{42}\left(Y_{1i} - t_{42}\right)_+\end{aligned} \quad (38)$$

If K=5, m=2, and t=1

$$\begin{aligned}\hat{f}_{1i} &= \frac{1}{2}a_{01} + b_{11j}X_{1i} + \gamma_{11}\cos X_{1i} + \gamma_{21}\cos 2X_{1i} + \gamma_{31}\cos 3X_{1i} + \gamma_{41}\cos 4X_{1i} \\ &+ \gamma_{51}\cos 5X_{1i} + b_{21j}X_{2i} + b_{31j}X_{2i}^2 + \alpha_{11}\left(x_{2i} - t_{11}\right)_+^2 \\ \hat{f}_{2i} &= \frac{1}{2}a_{02} + b_{12j}X_{1i} + \gamma_{12}\cos X_{1i} + \gamma_{22}\cos 2X_{1i} + \gamma_{32}\cos 3X_{1i} + \gamma_{42}\cos 4X_{1i} \\ &+ \gamma_{52}\cos 5X_{1i} + b_{22j}X_{2i} + b_{32j}X_{2i}^2 + \alpha_{12}\left(x_{2i} - t_{12}\right)_+^2 + b_{42j}Y_{1i} + b_{52j}Y_{1i}^2 \\ &+ \alpha_{22}\left(Y_{1i} - t_{22}\right)_+^2\end{aligned} \quad (39)$$

If K=5, m=2, and t=2

$$\begin{aligned}\hat{f}_{1i} &= \frac{1}{2}a_{01} + b_{11j}X_{1i} + \gamma_{11}\cos X_{1i} + \gamma_{21}\cos 2X_{1i} + \gamma_{31}\cos 3X_{1i} + \gamma_{41}\cos 4X_{1i} \\ &+ \gamma_{51}\cos 5X_{1i} + b_{21j}X_{2i} + b_{31j}X_{2i}^2 + \alpha_{11}\left(x_{2i} - t_{11}\right)_+^2 + \alpha_{21}\left(x_{2i} - t_{21}\right)_+^2 \\ \hat{f}_{2i} &= \frac{1}{2}a_{02} + b_{12j}X_{1i} + \gamma_{12}\cos X_{1i} + \gamma_{22}\cos 2X_{1i} + \gamma_{32}\cos 3X_{1i} + \gamma_{42}\cos 4X_{1i} \\ &+ \gamma_{52}\cos 5X_{1i} + b_{22j}X_{2i} + b_{32j}X_{2i}^2 + \alpha_{12}\left(x_{2i} - t_{12}\right)_+^2 + \alpha_{22}\left(x_{2i} - t_{22}\right)_+^2 + b_{42j}Y_{1i} \\ &+ b_{52j}Y_{1i}^2 + \alpha_{32}\left(Y_{1i} - t_{32}\right)_+^2\end{aligned} \quad (40)$$

If the path analysis model is presented in the form of a matrix, in general, the matrix formed is as follows:

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ \vdots \\ y_{1n} \\ y_{21} \\ y_{22} \\ y_{23} \\ \vdots \\ y_{2n} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & x_{11} & \cos x_{11} & \cdots & \cos Kx_{11} & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \frac{1}{2} & x_{12} & \cos x_{12} & \cdots & \cos Kx_{12} & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \frac{1}{2} & x_{13} & \cos x_{12} & \cdots & \cos Kx_{13} & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2} & x_{1n} & \cos x_{1n} & \cdots & \cos Kx_{1n} & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & x_{21} & x_{21}^2 & \cdots & x_{21}^m & \left(x_{21} - t_{21}\right)_+^m & \left(x_{21} - t_{22}\right)_+^m & \cdots & \left(x_{21} - t_{2r}\right)_+^m \\ 0 & 0 & 0 & \cdots & 0 & x_{22} & x_{22}^2 & \cdots & x_{22}^m & \left(x_{22} - t_{21}\right)_+^m & \left(x_{22} - t_{22}\right)_+^m & \cdots & \left(x_{22} - t_{2r}\right)_+^m \\ 0 & 0 & 0 & \cdots & 0 & x_{23} & x_{23}^2 & \cdots & x_{23}^m & \left(x_{23} - t_{21}\right)_+^m & \left(x_{23} - t_{22}\right)_+^m & \cdots & \left(x_{23} - t_{2r}\right)_+^m \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & x_{2n} & x_{2n}^2 & \cdots & x_{2n}^m & \left(x_{2n} - t_{21}\right)_+^m & \left(x_{2n} - t_{22}\right)_+^m & \cdots & \left(x_{2n} - t_{2r}\right)_+^m \end{bmatrix} \begin{bmatrix} b \\ \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_K \\ \beta_{21} \\ \beta_{22} \\ \vdots \\ \beta_{m2} \\ \alpha_{21} \\ \alpha_{22} \\ \vdots \\ \alpha_{2r} \end{bmatrix}$$

Then the above model can be written as follows:

$$y = Wa + X\beta + \varepsilon \tag{41}$$

Where:

Y_i : the response variable's value for observation i

x_i : the predictor variable's value for observation i

$\hat{f}$: regression curve

i : 1, 2, ..., *n*.

t: knot point

K: Oscillation rate

p: Spline Ordo

ε_i : remainder on i-th observation

Proof

In equation (21), the following equations can be generated and after that they will be used as a matrix.

First Part

Simple linear regression model:

$$Y_i = \beta_0 + \beta_i X_i + \varepsilon_i \tag{42}$$

Where:

Y_i: the response variable's value for observation i

β_0: intercept parameters

β_1: parameter slope

X_i: the predictor variable's value for observation i

ε_i: remainder in i-th observation

Multiple linear regression analysis is used when there are multiple predictor variables. The equation can be used to express the multiple linear regression model (43).

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \tag{43}$$

The equation can be used to represent the general linear regression model (44)

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{i(p-1)} + \varepsilon_i \tag{44}$$

The matrix on multiple linear regression analysis can be used to estimate the parameters. Furthermore, the multiple linear regression equation is shown in equation (45) below.

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \cdots + \beta_p X_{1p-1} + \varepsilon_1 \\ Y_2 &= \beta_0 + \beta_1 X_{21} + \beta_2 X_{22} + \cdots + \beta_p X_{2p-1} + \varepsilon_2 \\ Y_3 &= \beta_0 + \beta_1 X_{31} + \beta_2 X_{32} + \cdots + \beta_p X_{3p-1} + \varepsilon_3 \\ &\vdots \\ Y_n &= \beta_0 + \beta_1 X_{n1} + \beta_2 X_{n2} + \cdots + \beta_p X_{np-1} + \varepsilon_n \end{aligned} \tag{45}$$

The matrix form is

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1p-1} \\ 1 & X_{21} & X_{22} & \cdots & X_{2p-1} \\ 1 & X_{31} & X_{32} & \cdots & X_{3p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{np-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{pmatrix} \tag{46}$$

Second Part

The simple path analysis model is shown in equation (48)

$$\begin{aligned} Y_{1i} &= f_1(X_{1i}, X_{2i}) + \varepsilon_{1i} \\ Y_{2i} &= f_2(X_{1i}, X_{2i}, Y_{1i}) + \varepsilon_{2i} \end{aligned} \tag{47}$$

$$\begin{aligned} Y_{1i} &= \beta_{10} + \beta_{11} X_1 + \beta_{12} X_2 + \varepsilon_{1i} \\ Y_{2i} &= \beta_{20} + \beta_{21} X_1 + \beta_{22} X_2 + \beta_{23} Y_1 + \varepsilon_{2i} \end{aligned} \tag{48}$$

With matrix from:

$$\underset{\sim}{Y}_{2nx1} = \boldsymbol{X}_{2nx7} \underset{\sim}{\beta}_{7x1} + \underset{\sim}{\varepsilon}_{2nx1} \tag{49}$$

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1n} \\ Y_{21} \\ Y_{22} \\ \vdots \\ Y_{2n} \end{bmatrix} = \begin{bmatrix} X_X & \underset{\sim}{0}_{nx4} \\ \underset{\sim}{0}_{nx3} & X_{XY} \end{bmatrix} \begin{bmatrix} \beta_{10} \\ \beta_{11} \\ \beta_{12} \\ \beta_{20} \\ \beta_{21} \\ \beta_{22} \\ \beta_{23} \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{1n} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \vdots \\ \varepsilon_{2n} \end{bmatrix}$$

where

$$X_X = \begin{bmatrix} 1 & X_{11} & X_{21} \\ 1 & X_{12} & X_{22} \\ & \vdots & \\ 1 & X_{1n} & X_{2n} \end{bmatrix}; \; X_{XY} = \begin{bmatrix} 1 & X_{11} & X_{21} & Y_{11} \\ 1 & X_{12} & X_{22} & Y_{12} \\ & & \vdots & \\ 1 & X_{1n} & X_{2n} & Y_{1n} \end{bmatrix}$$

With:

Y_{hi}: endogenous variable h-th, observation i-th

X_i: exogenous variable observation i-th

β : parameters for predictor variables

ε_{hi}: random error endogenous variable h-th, observation i-th

Third Part

Equations (50) and (51) present nonparametric regression models that can be created after understanding the equations and multiple linear regression models (51).

$$Y_{1i} = f_1(X_{1i}, X_{2i}) + \varepsilon_{1i} \tag{50}$$

$$\hat{f}_\lambda(x_i) = \hat{b}(\lambda)x_i + \frac{1}{2}\hat{a}_0(\lambda) + \sum_{k=1}^{2} \hat{a}_k(\lambda)\cos kx_i \tag{51}$$

With the equation and matrix form like the following equation:

$$\underset{\sim}{f}_{2nx1} = \boldsymbol{X}_{2nx7}\alpha_{7x1} \tag{52}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} & 1 & cosx_{11} & cos2x_{11} & x_{21} & cosx_{21} & cos2x_{21} \\ x_{12} & 1 & cosx_{12} & cos2x_{12} & x_{22} & cosx_{22} & cos2x_{22} \\ x_{13} & 1 & cosx_{13} & cos2x_{13} & x_{23} & cosx_{23} & cos2x_{23} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{1n} & 1 & cosx_{1n} & cos2x_{1n} & x_{2n} & cosx_{2n} & cos2x_{2n} \end{pmatrix} \begin{pmatrix} b \\ \frac{1}{2}a_0 \\ a_1 \\ \vdots \\ a_7 \end{pmatrix} \tag{53}$$

The equations shown in the previous equations, namely the regression, path, and nonparametric regression models can be made into functions and matrices as follows.

$$\underset{\sim}{f}_{2nx1} = \boldsymbol{X}_{2nx17}\alpha_{17x1} \tag{54}$$

$$\mathbf{X}=\begin{pmatrix} \frac{1}{2} & x_{11} & \cos x_{11} & \cdots & \cos 2x_{21} & 0 & 0 & 0 & \cdots & 0 \\ \frac{1}{2} & x_{12} & \cos x_{12} & \cdots & \cos 2x_{22} & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2} & x_{1n} & \cos x_{1n} & \cdots & \cos 2x_{2n} & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & \frac{1}{2} & x_{11} & \cos x_{11} & \cdots & \cos 2y_{11} \\ 0 & 0 & 0 & \cdots & 0 & \frac{1}{2} & x_{12} & \cos x_{12} & \cdots & \cos 2y_{12} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \frac{1}{2} & x_{1n} & \cos x_{1n} & \cdots & \cos 2y_{1n} \end{pmatrix} \tag{55}$$

Where:

$\underset{\sim}{f}\left(X_{ij}\right)$: nonparametric regression function vector

$\boldsymbol{X}_{ij}$: exogenous variable

$\underset{\sim}{\alpha}_{ij}$: The j-th observation's parameter vector for the i-th exogenous variable

Theorem 4.1. Ordinary Least Square

Based on Lemma 4.1, the nonparametric path model can be approximated using the least squares approach. Thus, the estimation that can minimize the Fourier and truncated spline is shown as follows.

$$\underset{\beta\epsilon R}{Min}\left(\underset{\sim}{\varepsilon}'\underset{\sim}{\varepsilon}\right)=\left\{n^{-1}\sum_{i=1}^{n}\left(y_i-f_1(x_{1i})\right)^2+\lambda\int_0^{\pi}\frac{2}{\pi}\left(f_1^{(2)}(x_{1i})\right)^2\partial x_{1i}\right\}$$

$$Min_{\beta\epsilon R}\left(\underset{\sim}{\varepsilon}'\underset{\sim}{\varepsilon}\right)=\{n^{-1}(y-\boldsymbol{Wa}-\boldsymbol{X}\beta)'(y-\boldsymbol{Wa}-\boldsymbol{X}\beta)+\lambda\boldsymbol{a}'\boldsymbol{Da}\} \tag{56}$$

$$\underset{\beta\epsilon R}{Min}\left(\underset{\sim}{\varepsilon}'\underset{\sim}{\varepsilon}\right)=\{Q(\boldsymbol{a},\beta)\}$$

Equation 56 needs to be described to obtain an estimate of its function. The decomposition of equation (56) is as follows:

$$Q(\boldsymbol{a},\beta)=n^{-1}(y-\boldsymbol{Wa}-\boldsymbol{X}\beta)'(y-\boldsymbol{Wa}-\boldsymbol{X}\beta)+\lambda\boldsymbol{a}'\boldsymbol{Da}=n^{-1}(y'-\boldsymbol{a}'\boldsymbol{W}'-\beta'\boldsymbol{X}')(y-\boldsymbol{Wa}-\boldsymbol{X}\beta)+\lambda\boldsymbol{a}'\boldsymbol{Da}$$

$$=n^{-1}\begin{pmatrix}y'y-\boldsymbol{a}'\boldsymbol{W}'y-\beta'\boldsymbol{X}'y-y'\boldsymbol{Wa}+\boldsymbol{a}'\boldsymbol{W}'\boldsymbol{Wa}+\\ \beta'\boldsymbol{X}'\boldsymbol{Wa}-y'\boldsymbol{X}\beta+\boldsymbol{a}'\boldsymbol{W}'\boldsymbol{X}\beta+\beta'\boldsymbol{X}'\boldsymbol{X}\beta\end{pmatrix}+\lambda\boldsymbol{a}'\boldsymbol{Da}=n^{-1}y'y-2n^{-1}\boldsymbol{a}'\boldsymbol{W}'y-n^{-1}\beta'\boldsymbol{X}'y+2n^{-1}\boldsymbol{a}'\boldsymbol{W}'\boldsymbol{X}\beta$$

$$-n^{-1}y'\boldsymbol{X}\beta+n^{-1}\beta'\boldsymbol{X}'\boldsymbol{X}\beta+\boldsymbol{a}'(n^{-1}\boldsymbol{W}'\boldsymbol{W}+\lambda\boldsymbol{D})\boldsymbol{a} \tag{57}$$

Estimates of $\boldsymbol{a}$ and β can be obtained using the ordinary least squares method, by reducing the total squares of error as follows:

$$\frac{\partial Q(a,\beta)}{\partial a}=\left(\frac{1}{\partial a}\right)\partial\begin{pmatrix}n^{-1}y'y\text{-}2n^{-1}a'W'y\text{-}n^{-1}\beta'X'y+2n^{-1}a'W'X\beta\\ \text{-}n^{-1}y'X\beta+n^{-1}\beta'X'X\beta+a'^{\left(n^{-1}W'W+\lambda D\right)}a\end{pmatrix}\frac{\partial Q(\boldsymbol{a},\beta)}{\partial\boldsymbol{a}}=$$

$$\left(\frac{1}{\partial\boldsymbol{a}}\right)\boldsymbol{\partial}\begin{pmatrix}n^{-1}y'y-2n^{-1}\boldsymbol{a}'\boldsymbol{W}'y-n^{-1}\beta'\boldsymbol{X}'y+2n^{-1}\boldsymbol{a}'\boldsymbol{W}'\boldsymbol{X}\beta\\ -n^{-1}y'\boldsymbol{X}\beta+n^{-1}\beta'\boldsymbol{X}'\boldsymbol{X}\beta+\boldsymbol{a}'^{(n^{-1}\boldsymbol{W}'\boldsymbol{W}+\lambda\boldsymbol{D})}\boldsymbol{a}\end{pmatrix}=$$

$$-2n^{-1}\boldsymbol{W}'y+2n^{-1}\boldsymbol{W}'\boldsymbol{X}\beta+2(n^{-1}\boldsymbol{W}'\boldsymbol{W}+\lambda\boldsymbol{D})\boldsymbol{a}=$$

$$2(-n^{-1}\boldsymbol{W}'y+n^{-1}\boldsymbol{W}'\boldsymbol{X}\beta+n^{-1}\boldsymbol{W}'\boldsymbol{W}+\lambda\boldsymbol{D})\boldsymbol{a}$$

$$0=\left(-n^{-1}\boldsymbol{W}'y+n^{-1}\boldsymbol{W}'\boldsymbol{X}\hat{\beta}\right)(n^{-1}\boldsymbol{W}'\boldsymbol{W}+\lambda\boldsymbol{D})\hat{\boldsymbol{a}}\left(n^{-1}\boldsymbol{W}'y-n^{-1}\boldsymbol{W}'\boldsymbol{X}\hat{\beta}\right)=(n^{-1}\boldsymbol{W}'\boldsymbol{W}+\lambda\boldsymbol{D})\hat{\boldsymbol{a}}$$

$$\hat{\boldsymbol{a}}=\left(n^{-1}\boldsymbol{W}'\boldsymbol{W}+\lambda\boldsymbol{D}\right)^{-1}\left(n^{-1}\boldsymbol{W}'y-n^{-1}\boldsymbol{W}'\boldsymbol{X}\hat{\beta}\right)$$

$$\hat{\boldsymbol{a}} = \left(\boldsymbol{W}'\boldsymbol{W} + n\lambda\boldsymbol{D}\right)^{-1}\boldsymbol{W}'\left(y - \boldsymbol{X}\hat{\beta}\right)$$

$$\hat{\boldsymbol{a}} = S(K,\lambda)\boldsymbol{W}'\left(y - \boldsymbol{X}\hat{\beta}\right) \quad (58)$$

Equation (58) can be minimized by partially deriving relating to and equaling zero.

$$\frac{\partial Q(\boldsymbol{a},\beta)}{\partial\beta} = -2n^{-1}\boldsymbol{X'y} + 2n^{-1}\boldsymbol{X'Wa} + 2n^{-1}\boldsymbol{X'X}\beta\,\frac{\partial Q(a,\beta)}{\partial\beta} = -2n^{-1}X'y + 2n^{-1}X'Wa + 2n^{-1}X'X\beta = 2n^{-1}(-\boldsymbol{X'y} + \boldsymbol{X'Wa} + \boldsymbol{X'X}\beta)$$

$$0 = 2n^{-1}\left(-\boldsymbol{X'y} + \boldsymbol{X'W}\hat{\boldsymbol{a}} + \boldsymbol{X'X}\hat{\beta}\right) \quad (59)$$

$$\boldsymbol{X}'\boldsymbol{X}\hat{\beta} = \boldsymbol{X}'\boldsymbol{y} - \boldsymbol{X}'\boldsymbol{W}\hat{\boldsymbol{a}}$$

$$\hat{\beta} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{y} - \boldsymbol{X}'\boldsymbol{W}\hat{\boldsymbol{a}}\right)$$

$$\hat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{W}\hat{\boldsymbol{a}})$$

By substituting Equation (58) into equation (59), the beta is found as follows:

$$\hat{\beta} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\left(\boldsymbol{y} - \boldsymbol{W}\left(S(K,\lambda)\boldsymbol{W}'^{(y-\boldsymbol{X}\hat{\beta})}\right)\right) = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\left(\boldsymbol{y} - \boldsymbol{W}S(K,\lambda)\boldsymbol{W}'^{(y-\boldsymbol{X}\hat{\beta})}\right) =$$

$$(\boldsymbol{X'X})^{-1}\boldsymbol{X'y} - (\boldsymbol{X'X})^{-1}\boldsymbol{X'W}S(K,\lambda)\boldsymbol{W'}y + (\boldsymbol{X'X})^{-1}\boldsymbol{X'W}S(K,\lambda)\boldsymbol{W'X}\hat{\beta} \quad (60)$$

To obtain the result of the function estimator, it is necessary to subtract both sides $\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{W}S(K,\lambda)\mathbf{W}'\mathbf{X}\hat{\beta}$ in equation (61) as follows:

$$\hat{\beta} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}S(K,\lambda)\boldsymbol{W}'\boldsymbol{X}\hat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}S(K,\lambda)\boldsymbol{W}'y(\boldsymbol{I} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}S(K,\lambda)\boldsymbol{W}'\boldsymbol{X})\hat{\beta} =$$

$$(\boldsymbol{X'X})^{-1}\boldsymbol{X'}(\boldsymbol{y} - \boldsymbol{W}S(K,\lambda)\boldsymbol{W'}y)(I-(X'X)^{-1}X'WS(K,\lambda)W'X)\hat{\beta} = (X'X)^{-1}X'(I-WS(K,\lambda)W')y(\boldsymbol{I} - (\boldsymbol{X'X})^{-1}\boldsymbol{X'W}S(K,\lambda)\boldsymbol{W'X})\hat{\beta} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\left(\boldsymbol{I} - \boldsymbol{W}S(K,\lambda)\boldsymbol{W}'\right)y\hat{\beta} =$$

$$\left((\boldsymbol{I} - (\boldsymbol{X'X})^{-1}\boldsymbol{X'W}S(K,\lambda)\boldsymbol{W'X})\right)^{-1}(\boldsymbol{X'X})^{-1}\boldsymbol{X'}(\boldsymbol{I} - \boldsymbol{W}S(K,\lambda)\boldsymbol{W'})y \quad (61)$$

5. Conclusions

The use of nonparametric path analysis leads to the following conclusion: only in cases where the linearity presumption is violated can one use the Fourier series and truncated spline. Then, because it employs a nonparametric methodology, the estimator of the Nonparametric Regression-Based Path Analysis, which combined Fourier series and truncated spline using the OLS approach, provides an estimation result that is not singular. Different combinations of lambdas, oscillations, orders, and knots will produce various outcomes. However, this non-unique outcome will offer a graphical representation that is more resemblant to the initial data distribution. Adrianingsih et al. [10] conducted research using nonparametric regression analysis because the data did not form a certain pattern, the results showed nonparametric regression analysis using the Fourier series and using three oscillations resulted in a coefficient of determination of 84.08%. The contribution of this paper can be used as reference material, especially analysis in statistics. With this paper, it is hoped that it can be applied in various fields. Suggestions for further research can develop this research with other models.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors would like to thank the editor and anonymous reviewers for their comments that help improve the quality of this work.

REFERENCES

[1] Fernandes Fernandes, A. A. R. 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.

[2] Solimun, M. S. Multivariate Analysis Structural Equation Modelling (SEM) Lisrel dan Amos, Malang: Fakultas MIPA Universitas Brawijaya, 2002.

[3] Fernandes, A. A. R., & 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, 2018,

[4] Solimun & 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

[5] Eubank, R. L. Nonparametric regression and spline smoothing, CRC press, 1999.

[6] Sumardi, S., & 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.

[7] Hidayat, R., Yuliani, Y., & Sam, M. Model Regresi Nonparametrik Dengan Pendekatan Spline Truncated, Prosiding, 3(1), 2018. https://journal.uncp.ac.id/index.php/proceding/article/view/840/725

[8] Dani, A. T. R., Arianingsih, N. Y. & Ainurrochmah, A. Pengujian Hipotesis Simultan Model Regresi Nonparametrik Spline Truncated dalam Pemodelan Kasus Ekonomi, JAMBURA Journal of Probability and Statistics, 1(2), pp.98-106, 2020.

[9] Fernandes, A. A. R., Hutahayan, B., Arisoesilaningsih, E., Yanti, I., Astuti, A. B., & 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 (Vol. 546, No. 5, p. 052024). IOP Publishing, 2019.

[10] Prahutama, A. Model Regresi Nonparametrik dengan Pendekatan Deret Fourier pada Kasus Tingkat Pengangguran Terbuka di Jawa Timur, In Prosiding Seminar Nasional Statistika UNDIP, 2013.

[11] Fernandes, S., & Rinaldo, A. A. R. A. A. 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, 2018.

[12] Takezawa, K. Introduction to nonparametric regression (Vol. 606), John Wiley & Sons, 2006.

[13] Amalia, S. H. & Nur, I. M. Pemodelan Regresi Nonparametrik Deret Fourier pada Kasus Tingkat Kemiskinan di Provinsi Sumatera Utara, Prosiding Seminar Nasional Mahasiswa Unimus. Vol. 2, 2019.

[14] Adrianingsih, N. Y., Andrea, T. R. D., & Alifta, A. Pemodelan Dengan Pendekatan Deret Fourier Pada Kasus Tingkat Pengangguran Terbuka Di Nusa Tenggara Timur, Edusaintek, 4, 2020.

[15] Raharjo, K., Nurjannah, N., Solimun, S., & Fernandes, A. A. R The influence of organizational culture and job design on job commitment and human resource performance. Journal of Organizational Change Management, 2018.

[16] Dani, A. T. R., & Adrianingsih, N. Y. Pemodelan Regresi Nonparametrik dengan Estimator Spline Truncated vs Deret Fourier. Jambura Journal of Mathematics, 3(1), 26-36, 2021.

[17] 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, 2017.

[18] Sholiha, A., Kuzairi, K., & Madianto, M. F. F. Estimator Deret Fourier Dalam Regresi Nonparametrik dengan Pembobot Untuk Perencanaan Penjualan Camilan Khas Madura, Zeta-Math Journal, 4(1), 18-23, 2018.

[19] Wisisono, I.R.N, Nurwahidah, A.I., & Andriyana, Y. Regresi Nonparametrik dengan Pendekatan Deret Fourier pada Data Debit Air Sungai Citarum, Jurnal Matematika "MANTIK&quot, 4(2), 75-82, 2018.

[20] Purbawangsa, I. B. A., Solimun, S., Fernandes, A. A. R., & 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.

[21] Benny Hutahayan, A. A. R. F. S. N.. Comparison of use of Linkage in Integrated Cluster with Discriminal Analysis Approach. International Journal of Advanced Science and Technology, 29(3), 5654 – 5668, 2020.

[22] Budiantara, I. Nyoman. Penelitian Bidang Regresi Spline Menuju Terwujudnya Penelitian Statistika yang Mandiri dan Berkarakter, Prosiding Seminar Nasional FMIPA Undiksha, 9–28, 2011. https://ejournal.undiksha.ac.id/index.php/semnasmipa/article/view/2736/2316

[23] Fernandes, A. A. R., Budiantara, I. N. I., Otok, B. W., & Suhartono. Reproducing Kernel Hilbert space for penalized regression multi-predictors: Case in longitudinal data, International Journal of Mathematical Analysis, 8(40), 1951-1961, 2014.

[24] Hidayat, M. F., Achmad, R. F. A., & Solimun. Estimation of Truncated Spline Function in Non-parametric Path Analysis Based on Weighted Least Square (WLS). IOP Conference Series: Materials Science and Engineering, 546(5), 052027, 2019. https://doi.org/10.1088/1757-899X/546/5/052027