J. Stat. Appl. Pro. **11**, No. 3, 781-794 (2022) 781

Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/110303

Truncated Spline Path Analysis Modeling on in Company X with the Government's Role as a Mediation Variable

*Farid Ubaidillah**, *Adji Achmad Rinaldo Fernandes, Atiek Iriany, Ni Wayan Surya Wardhani and Solimun Solimun*

Department of Statistic, Brawijaya University, Malang 65141, Indonesia

Received: 9 Jan. 2022, Revised: 13 Mar. 2022, Accepted: 16 May 2022
Published online: 1 Sep. 2022

Abstract: This study aims to model the path analysis of linear, quadratic, and cubic truncated splines at one and two knot points using Iteratively Reweighted Least Square (IRLS) and determine the best model in the case of Community Welfare with the Role of the Government as a mediating variable. The data used in this study is primary data. Data analysis used in this study is nonparametric path analysis with Iteratively Reweighted Least Square (IRLS) using Rstudio software. The estimation of the nonparametric path model function truncated spline of linear, quadratic, and cubic 1 and 2 point knots indicates that changes in data behavior patterns occur when the character / personality of the debtor is quite good and good at paying credit on time. The novelty of this research is to model the development of nonparametric path analysis, namely nonparametric truncated spline analysis of linear, quadratic, and cubic orders using Iterative Reweighted Least Square (IRLS) on community welfare data.

Keywords: Path Analysis, Nonparametric Path Analysis, Iteratively Reweighted Least Square (IRLS), Truncated Spline.

1 Introduction

Regression analysis is one of the methods in statistical science that aims to determine the pattern of the relationship between one or more predictor variables and the unknown response variable. In the process of modeling regression analysis, there is a regression curve that is used as an explanation of the functional relationship between predictor variables and response variables where the data pattern described is approached by parametric regression [1]. In practice, not all relationships between variables can follow a certain relationship pattern according to parametric regression modeling [2]. The relationship between variables that cannot follow a certain pattern is modeled using nonparametric regression. In nonparametric regression, the form of the regression function is assumed to be unknown and the linearity assumption is not met [3].

The most popular nonparametric regression approach model used is the Spline model. A spline is a section of a segmented polynomial connected through knots. The knot point itself is a combination of common points that explain how the behavior changes of the spline function at different intervals. The segmented nature causes the spline to have high flexibility compared to the ordinary polynomial model, thus making the spline regression adapt more effectively to the characteristics of a data [4].

In the current development of statistics and research, the relationship between variables is not only limited to one response variable with one or several response variables. Modeling for more than one response variable with one or more predictor variables is called path analysis modeling. Path analysis is an analysis that can be used to analyze causal relationships and direct or indirect effects of several observed variables [5]. Path analysis based on nonparametric regression is a path analysis that is in accordance with the pattern of the relationship between predictor variables and response variables whose curve shape is not known [6].

In the regression model, the method of least squares (OLS) is often used for parameter estimation. The least squares method (OLS) is considered the best method for estimating the regression parameters if the assumptions are met. However, if the data used does not meet these assumptions, then the parameter estimation results are invalid and biased. One of the violations of

*Corresponding author e-mail: farid.ub@student.ub.ac.id

these assumptions is the assumption of heteroscedasticity caused by an inhomogeneous variance [7]. Therefore, we need a method to deal with this problem, namely the Iterative Reweighted Least Square (IRLS) estimation [8].

Company X is one of the companies under the State-Owned Enterprises (BUMN) which is engaged in logistics and food in Indonesia. Basically the company has a goal to get a lot of profit. Another goal of the company is that many products can be purchased or known by the public. Therefore, the company must prioritize the wants and needs of the community [9].
Based on the description above, this study aims to model the best model using linear, quadratic, and cubic truncated spline path analysis at one and two knot points using Iteratively Reweighted Least Square (IRLS) on the 5C principle on the welfare of society at company X. with the government's role as a mediating variable and knowing the factors that affect the welfare of the community which is the originality of the research. In this study, nonparametric truncated spline path analysis function was used to estimate the Iterative Reweighted Least Square (IRLS) approach.

2 Literature Review

2.1 Nonparametric Regression

Regression analysis is one of the statistical methods commonly used to see the functional relationship between the response variables that are influenced by the predictor variables. According to [1], nonparametric regression is a statistical method used to determine the relationship between the response variable and predictor whose function is not known, it is only assumed that the function is smooth in the sense that it is contained in a certain function space, so that nonparametric regression has greater flexibility. high. Because the shape of the regression curve is unknown, nonparametric regression can form any function in estimating the shape of the regression curve, either linear or nonlinear.
According to [10], nonparametric regression models in general can be presented as follows:

$$y_i = f(x_i) + \varepsilon_i \text{ , for } i = 1,2,\dots,n \tag{1}$$

Where: y_i : the value of the response variable at the i-th observation
$f(x_i)$: observation
x_i : the value of the predictor variable at the i-th observation
ε_i : error i
n : many observations

Truncated Spline Nonparametric Regression

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. 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 Penalized Least Square (PLS) method. (b) Spline is able to handle smooth data/functions. (c) Spline has the ability to handle data whose behavior changes at certain sub-intervals very well. (d) Spline has the ability to generalize complex and complex statistical modeling very well. One of the spline models that are of interest in applications is the truncated spline polynomial [11].
According to [4], splines are parts or pieces of polynomials that have segmented and continuous (truncated) properties. The advantage of truncated spline polynomial regression is that it tends to find its own form of regression curve estimation. This can happen because the spline has a common fusion point that shows the occurrence of 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 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 statistical calculations.
[10] conducted a study with a regression curve f approximated by the spline function f and knots K. In matrix form it is stated as follows:

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

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

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 & \cdots & X_1^p & (X_1-K_1)_+^p & \cdots & (X_1-K_k)_+^p \\ 1 & X_2 & \cdots & X_2^p & (X_2-K_1)_+^p & \cdots & (X_2-K_k)_+^p \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_n & \cdots & X_n^p & (X_n-K_1)_+^p & \cdots & (X_n-K_k)_+^p \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \\ \delta_1 \\ \vdots \\ \delta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

It can also be written as:

$$\underset{\sim}{y} = \mathbf{X}\left[K_1, K_2, \ldots, K_k\right]\underset{\sim}{\beta} + \underset{\sim}{\varepsilon} \tag{2}$$

Next, the parameter estimation

$$\hat{\underset{\sim}{\beta}} = \begin{pmatrix} \beta_0 & \beta_1 & \beta_2 & \cdots & \beta_p & \delta_1 & \cdots & \delta_k \end{pmatrix}^t \tag{3}$$

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

$$\underset{\underset{\sim}{\beta} \in R^{p+1+k}}{Min}\left\{\underset{\sim}{\varepsilon}^t \underset{\sim}{\varepsilon}\right\} = \underset{\underset{\sim}{\beta} \in R^{p+1+k}}{Min}\left\{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}^t \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}\right\}$$

$$= \underset{\underset{\sim}{\beta} \in R^{p+1+k}}{Min}\left\{\left(\underset{\sim}{y} - \mathbf{X}\left[K_1, K_2, \cdots, K_k\right]\underset{\sim}{\beta}\right)^t \left(\underset{\sim}{y} - \mathbf{X}\left[K_1, K_2, \cdots, K_k\right]\underset{\sim}{\beta}\right)\right\} \tag{4}$$

The elaboration of the sum squared error matrix is presented as follows:

$$\begin{aligned} \sum_{i=1}^{n} \varepsilon_i^2 &= \underset{\sim}{\varepsilon}^t \underset{\sim}{\varepsilon} \\ &= \left(\underset{\sim}{y} - \mathbf{X}\left[K_1, K_2, \cdots, K_k\right]\underset{\sim}{\beta}\right)^t \left(\underset{\sim}{y} - \mathbf{X}\left[K_1, K_2, \cdots, K_k\right]\underset{\sim}{\beta}\right) \\ &= \left(\underset{\sim}{y} - \mathbf{X}[\mathbf{K}]\underset{\sim}{\beta}\right)^t \left(\underset{\sim}{y} - \mathbf{X}[\mathbf{K}]\underset{\sim}{\beta}\right) \\ &= \underset{\sim}{y}^t \underset{\sim}{y} - 2\underset{\sim}{\beta}^t \mathbf{X}[\mathbf{K}]^t \underset{\sim}{y} + \underset{\sim}{\beta}^t \mathbf{X}[\mathbf{K}]^t \mathbf{X}[\mathbf{K}]\underset{\sim}{\beta} \end{aligned} \tag{5}$$

If equation (5) is derived from and then equated to zero, we get:

$$\frac{\partial\left(\underset{\sim}{\varepsilon}^t \underset{\sim}{\varepsilon}\right)}{\partial \underset{\sim}{\beta}} = \frac{\partial\left(\underset{\sim}{y}^t \underset{\sim}{y} - 2\underset{\sim}{\beta}^t \mathbf{X}[\mathbf{K}]^t \underset{\sim}{y} + \underset{\sim}{\beta}^t \mathbf{X}[\mathbf{K}]^t \mathbf{X}[\mathbf{K}]\underset{\sim}{\beta}\right)}{\partial \underset{\sim}{\beta}} = 0$$

$$\hat{\underset{\sim}{\beta}} = \left(\mathbf{X}[\mathbf{K}]^t \mathbf{X}[\mathbf{K}]\right)^{-1} \mathbf{X}[\mathbf{K}]^t \underset{\sim}{y} \tag{6}$$

With

$$\mathbf{X}[\mathbf{K}] = \mathbf{X}\left[K_1, K_2, \ldots, K_k\right]$$

$$= \begin{bmatrix} 1 & X_1 & \cdots & X_1^p & (X_1-K_1)_+^p & \cdots & (X_1-K_k)_+^p \\ 1 & X_2 & \cdots & X_2^p & (X_2-K_1)_+^p & \cdots & (X_2-K_k)_+^p \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_n & \cdots & X_n^p & (X_n-K_1)_+^p & \cdots & (X_n-K_k)_+^p \end{bmatrix}$$

Thus, the estimation of the K spline knot regression curve is obtained:

$$\begin{aligned} \hat{f}(X_i) &= \mathbf{X}[\mathbf{K}]\hat{\underset{\sim}{\beta}} \\ &= \mathbf{X}[\mathbf{K}]\left(\mathbf{X}[\mathbf{K}]^t \mathbf{X}[\mathbf{K}]\right)^{-1} \mathbf{X}[\mathbf{K}]^t \underset{\sim}{y} \end{aligned} \tag{7}$$

So that it is obtained:

$$\hat{f}(X_i) = A[\mathbf{K}]y \tag{8}$$

with $A[\mathbf{K}] = \mathbf{X}[\mathbf{K}]\left(\mathbf{X}[\mathbf{K}]^t\,\mathbf{X}[\mathbf{K}]\right)^{-1}\mathbf{X}[\mathbf{K}]^t$ is a function of the knot point and $\mathbf{K} = (K_1, K_2, ..., K_k)^t$ are the knot points.

In general, the regression model is similar to equation (1), f(X_i) is a regression curve that is approximated by a spline function of order p with knot points K_1, K_2,…,K_k (first knot point to k knot) derived from equation. The following function is obtained:

$$\hat{f}(X_i) = \hat{\beta}_0 + \sum_{j=1}^{p}\hat{\beta}_j X_i^j + \sum_{k=1}^{K}\delta_k\left(X_i - K_k\right)_+^p \tag{9}$$

If equation (9) is substituted into equation (1), the nonparametric spline regression equation is obtained as follows:

$$Y_i = \beta_0 + \sum_{j=1}^{p}\beta_j X_i^j + \sum_{k=1}^{K}\delta_k\left(X_i - K_k\right)_+^p + \varepsilon_i$$

where $i = 1,2,3, \dots, n$.

$j = 1,2,3, \dots, p\ ; p \geq 1$.

$k = 1,2,3, \dots, K$.

n : Number of observations.

p : Ordo Spline.

K : Number of knots.

The function$(X_i - K_k)_+^p$ is a truncated function given by:

$$(X_i - K_k)_+^p = \begin{cases}(X_i - K_k)^p & ;\ X_i \geq K_k \\ 0 & ;\ X_i < K_k\end{cases}$$

where p is the order of Spline and K is the selected knot [12].

Quadratic Polynomial Degree Truncated Spline Model (Order p=2) with 1 Knot Point The general form of the quadratic polynomial degree truncated spline regression model is as follows:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \delta_1 (X_i - K_1)_{+\,+}^2 + \varepsilon_{1i} \tag{10}$$

where function is truncated:

$$(X_i - K_k)_+^2 = \begin{cases}(X_i - K_k)^2 & ;\ X_i \geq K_k \\ 0 & ;\ X_i < K_k\end{cases}$$

The nonparametric truncated spline regression model on the dependent and independent variables single when quadratic (order p=2) with 1 knot will be obtained as follows:

$$\hat{f}_{1i} = \hat{\beta}_{10} + \hat{\beta}_{11} X_{1i} + \hat{\beta}_{12} X_{1i}^2 + \hat{\delta}_{11} (X_{1i} - K_{11})_+^2 + \tag{11}$$

where function truncated :

$$(X_{1i} - K_{11})_+^2 = \begin{cases}(X_{1i} - K_{11})^2 & ;\ X_{1i} \geq K_{11} \\ 0 & ;\ X_{1i} < K_{11}\end{cases}$$

2.2 Nonparametric Path Analysis

Path analysis is an analysis that studies the causal relationship between exogenous and endogenous variables, and looks for the most efficient path. Path analysis can be done if it meets the assumption of linearity and the shape of the regression curve is known. If the linearity assumption is not met and the shape of the regression curve is not known, then path analysis cannot be used, so an analysis that can overcome this problem is developed, namely nonparametric path analysis. The nonparametric path analysis function is presented in equation (12).

$$\begin{aligned} y_{1i} &= f_1(x_{1i}) + f_1(x_{2i}) + f_1(x_{3i}) + f_1(x_{4i}) + f_1(x_{5i}) + \varepsilon_{1i} \\ &= f_1(x_{1i}, x_{2i}, x_{3i}, x_{4i}, x_{5i}) \\ y_{2i} &= f_2(x_{1i}) + f_2(x_{2i}) + f_2(x_{3i}) + f_2(x_{4i}) + f_2(x_{5i}) + f_2(y_{1i}) + \varepsilon_{2i} \\ &= f_2(x_{1i}, x_{2i}, x_{3i}, x_{4i}, x_{5i}, y_{1i}) \\ i &= 1,2, \dots, n \end{aligned} \tag{12}$$

The nonparametric truncated spline path model when linear (order p=1) with 1 knot point for five exogenous variables and two endogenous variables will be obtained as follows.

$$\hat{f}_{1i} = \hat{\beta}_{10} + \hat{\beta}_{11}X_{1i} + \hat{\delta}_{11}(X_{1i} - K_{11})_+ + \hat{\beta}_{12}X_{2i} + \hat{\lambda}_{11}(X_{2i} - K_{21})_+ + \hat{\beta}_{13}X_{3i} + \hat{\gamma}_{11}(X_{3i} - K_{31})_+ + \hat{\beta}_{14}X_{4i} + \hat{\rho}_{11}(X_{4i} - K_{41})_+ + \hat{\beta}_{15}X_{5i} + \hat{\theta}_{11}(X_{5i} - K_{51})_+ \tag{13}$$

$$\hat{f}_{2i} = \hat{\beta}_{20} + \hat{\beta}_{21}X_{1i} + \hat{\delta}_{21}(X_{1i} - K_{11})_+ + \hat{\beta}_{22}X_{2i} + \hat{\lambda}_{21}(X_{2i} - K_{21})_+ + \hat{\beta}_{23}X_{3i} + \hat{\gamma}_{21}(X_{3i} - K_{31})_+ + \hat{\beta}_{24}X_{4i} + \hat{\rho}_{21}(X_{4i} - K_{41})_+ + \hat{\beta}_{25}X_{5i} + \hat{\theta}_{21}(X_{5i} - K_{51})_+ + \hat{\beta}_{26}f_{1i} + \hat{\alpha}_{21}(f_{1i} - K_{61})_+$$

The nonparametric truncated spline path model when linear (order p=1) with 2 knot points for five exogenous variables and two endogenous variables will be obtained as follows:

$$\hat{f}_{1i} = \hat{\beta}_{10} + \hat{\beta}_{11}X_{1i} + \hat{\delta}_{11}(X_{1i} - K_{11})_+ + \hat{\delta}_{12}(X_{1i} - K_{12})_+ + \hat{\beta}_{12}X_{2i} + \hat{\lambda}_{11}(X_{2i} - K_{21})_+ + \hat{\lambda}_{12}(X_{2i} - K_{22})_+ + \hat{\beta}_{13}X_{3i} + \hat{\gamma}_{11}(X_{3i} - K_{31})_+ + \hat{\gamma}_{12}(X_{3i} - K_{32})_+ + \hat{\beta}_{14}X_{4i} + \hat{\rho}_{11}(X_{4i} - K_{41})_+ + \hat{\rho}_{12}(X_{4i} - K_{42})_+ + \hat{\beta}_{15}X_{5i} + \hat{\theta}_{11}(X_{5i} - K_{51})_+ + \hat{\theta}_{12}(X_{5i} - K_{52})_+$$

(14)

$$\hat{f}_{2i} = \hat{\beta}_{20} + \hat{\beta}_{21}X_{1i} + \hat{\delta}_{21}(X_{1i} - K_{11})_+ + \hat{\delta}_{22}(X_{1i} - K_{12})_+ + \hat{\beta}_{22}X_{2i} + \hat{\lambda}_{21}(X_{2i} - K_{21})_+ + \hat{\lambda}_{22}(X_{2i} - K_{22})_+ + \hat{\beta}_{23}X_{3i} + \hat{\gamma}_{21}(X_{3i} - K_{31})_+ + \hat{\gamma}_{22}(X_{3i} - K_{32})_+ + \hat{\beta}_{24}X_{4i} + \hat{\rho}_{21}(X_{4i} - K_{41})_+ + \hat{\rho}_{22}(X_{4i} - K_{42})_+ + \hat{\beta}_{25}X_{5i} + \hat{\theta}_{21}(X_{5i} - K_{51})_+ + \hat{\theta}_{22}(X_{5i} - K_{52})_+ + \hat{\beta}_{26}f_{1i} + \hat{\alpha}_{21}(f_{1i} - K_{61})_+ + \hat{\alpha}_{22}(f_{1i} - K_{62})_+$$

$$(X_{1i} - K_{11})_+ = \begin{cases} (X_{1i} - K_{11}) & ; X_{1i} \geq K_{11} \\ 0 & ; X_{1i} < K_{11} \end{cases}$$
$$(X_{1i} - K_{12})_+ = \begin{cases} (X_{1i} - K_{12}) & ; X_{1i} \geq K_{12} \\ 0 & ; X_{1i} < K_{12} \end{cases}$$
$$(X_{2i} - K_{21})_+ = \begin{cases} (X_{2i} - K_{21}) & ; X_{2i} \geq K_{21} \\ 0 & ; X_{2i} < K_{21} \end{cases}$$
$$(X_{2i} - K_{22})_+ = \begin{cases} (X_{2i} - K_{22}) & ; X_{2i} \geq K_{22} \\ 0 & ; X_{2i} < K_{22} \end{cases}$$
$$(X_{3i} - K_{31})_+ = \begin{cases} (X_{3i} - K_{31}) & ; X_{3i} \geq K_{31} \\ 0 & ; X_{3i} < K_{31} \end{cases}$$
$$(X_{3i} - K_{32})_+ = \begin{cases} (X_{3i} - K_{32}) & ; X_{3i} \geq K_{32} \\ 0 & ; X_{3i} < K_{32} \end{cases}$$
$$(X_{4i} - K_{41})_+ = \begin{cases} (X_{4i} - K_{41}) & ; X_{4i} \geq K_{41} \\ 0 & ; X_{4i} < K_{41} \end{cases}$$
$$(X_{4i} - K_{42})_+ = \begin{cases} (X_{4i} - K_{42}) & ; X_{4i} \geq K_{42} \\ 0 & ; X_{4i} < K_{42} \end{cases}$$
$$(X_{5i} - K_{51})_+ = \begin{cases} (X_{5i} - K_{51}) & ; X_{5i} \geq K_{51} \\ 0 & ; X_{5i} < K_{51} \end{cases}$$
$$(X_{5i} - K_{52})_+ = \begin{cases} (X_{5i} - K_{52}) & ; X_{5i} \geq K_{52} \\ 0 & ; X_{5i} < K_{52} \end{cases}$$
$$(f_{1i} - K_{61})_+ = \begin{cases} (f_{1i} - K_{61}) & ; X_{6i} \geq K_{61} \\ 0 & ; X_{6i} < K_{61} \end{cases}$$
$$(f_{1i} - K_{62})_+ = \begin{cases} (f_{1i} - K_{62}) & ; X_{6i} \geq K_{62} \\ 0 & ; X_{6i} < K_{62} \end{cases}$$

The nonparametric truncated spline path model at quadratic (order p=2) with 1 knot point for five exogenous variables and two endogenous variables will be obtained as follows.

$$\hat{f}_{1i} = \hat{\beta}_{10} + \hat{\beta}_{11}X_{1i} + \hat{\beta}_{12}X_{1i}^2 + \hat{\delta}_{11}(X_{1i} - K_{11})_+^2 + \hat{\beta}_{13}X_{2i} + \hat{\beta}_{14}X_{2i}^2 + \hat{\lambda}_{11}(X_{2i} - K_{21})_+^2 + \hat{\beta}_{15}X_{3i} + \hat{\beta}_{16}X_{3i}^2 + \hat{\gamma}_{11}(X_{3i} - K_{31})_+^2 + \hat{\beta}_{17}X_{4i} + \hat{\beta}_{18}X_{4i}^2 + \hat{\rho}_{11}(X_{4i} - K_{41})_+^2 + \hat{\beta}_{19}X_{5i} + \hat{\beta}_{110}X_{4i}^2 + \hat{\theta}_{11}(X_{5i} - K_{51})_+^2 + \hat{\beta}_{212}f_{1i}^2 + \alpha_{21}(f_{1i} - K_{61})_+^2$$

$$\hat{f}_{2i} = \hat{\beta}_{20} + \hat{\beta}_{21}X_{1i} + \hat{\beta}_{22}X_{1i}^2 + \hat{\delta}_{21}(X_{1i} - K_{11})_+^2 + \hat{\beta}_{23}X_{2i} + \hat{\beta}_{24}X_{2i}^2 + \hat{\lambda}_{21}(X_{2i} - K_{21})_+^2 + \hat{\beta}_{25}X_{3i} + \hat{\beta}_{26}X_{3i}^2 + \hat{\gamma}_{21}(X_{3i} - K_{31})_+^2 + \hat{\beta}_{27}X_{4i} + \hat{\beta}_{28}X_{4i}^2 + \hat{\rho}_{21}(X_{4i} - K_{41})_+^2 + \hat{\beta}_{29}X_{5i} + \hat{\beta}_{210}X_{5i}^2 + \hat{\theta}_{21}(X_{5i} - K_{51})_+^2 + \hat{\beta}_{211}f_{1i} \quad (15)$$

The nonparametric truncated spline path model at quadratic time (order p=2) with 2 knot points for five exogenous variables and two endogenous variables will be obtained as follows.

$$\hat{f}_{1i} = \hat{\beta}_{10} + \hat{\beta}_{11}X_{1i} + \hat{\beta}_{12}X_{1i}^2 + \hat{\delta}_{11}(X_{1i} - K_{11})_+^2 + \hat{\delta}_{12}(X_{1i} - K_{12})_+^2 + \hat{\beta}_{13}X_{2i} + \hat{\beta}_{14}X_{2i}^2 + \hat{\lambda}_{11}(X_{2i} - K_{21})_+^2 + \hat{\lambda}_{12}(X_{2i} - K_{22})_+^2 + \hat{\beta}_{15}X_{3i} + \hat{\beta}_{16}X_{3i}^2 + \hat{\gamma}_{11}(X_{3i} - K_{31})_+^2 + \hat{\gamma}_{12}(X_{3i} - K_{32})_+^2 + \hat{\beta}_{17}X_{4i} + \hat{\beta}_{18}X_{4i}^2 + \hat{\rho}_{11}(X_{4i} - K_{41})_+^2 + \hat{\rho}_{12}(X_{4i} - K_{42})_+^2 + \hat{\beta}_{19}X_{5i} + \hat{\beta}_{110}X_{5i}^2 + \hat{\theta}_{11}(X_{5i} - K_{51})_+^2 + \hat{\theta}_{12}(X_{5i} - K_{51})_+^2$$

$$\hat{f}_{2i} = \hat{\beta}_{20} + \hat{\beta}_{21}X_{1i} + \hat{\beta}_{22}X_{1i}^2 + \hat{\delta}_{21}(X_{1i} - K_{11})_+^2 + \hat{\delta}_{22}(X_{1i} - K_{12})_+^2 + \hat{\beta}_{23}X_{2i} + \hat{\beta}_{24}X_{2i}^2 + \hat{\lambda}_{21}(X_{2i} - K_{21})_+^2 + \hat{\lambda}_{22}(X_{2i} - K_{22})_+^2 + \hat{\beta}_{25}X_{3i} + \hat{\beta}_{26}X_{3i}^2 + \hat{\gamma}_{21}(X_{3i} - K_{31})_+^2 + \hat{\gamma}_{22}(X_{3i} - K_{32})_+^2 + \hat{\beta}_{27}X_{4i} + \hat{\beta}_{28}X_{4i}^2 + \hat{\rho}_{21}(X_{4i} - K_{41})_+^2 + \hat{\rho}_{22}(X_{4i} - K_{42})_+^2 + \hat{\beta}_{29}X_{5i} + \hat{\beta}_{210}X_{5i}^2 + \hat{\theta}_{21}(X_{5i} - K_{51})_+^2 + \hat{\theta}_{22}(X_{5i} - K_{52})_+^2 + \hat{\beta}_{211}f_{1i} + \hat{\beta}_{212}f_{1i}^2 + \hat{\alpha}_{21}(f_{1i} - K_{61})_+^2 + \hat{\alpha}_{22}(f_{1i} - K_{62})_+^2 \quad (16)$$

where function truncated :

$$(X_{1i} - K_{11})_+^2 = \begin{cases} (X_{1i} - K_{11})^2 & ; X_{1i} \geq K_{11} \\ 0 & ; X_{1i} < K_{11} \end{cases}$$

$$(X_{1i} - K_{12})_+^2 = \begin{cases} (X_{1i} - K_{12})^2 & ; X_{1i} \geq K_{12} \\ 0 & ; X_{1i} < K_{12} \end{cases}$$

$$(X_{2i} - K_{21})_+^2 = \begin{cases} (X_{2i} - K_{21})^2 & ; X_{2i} \geq K_{21} \\ 0 & ; X_{2i} < K_{21} \end{cases}$$

$$(X_{2i} - K_{22})_+^2 = \begin{cases} (X_{2i} - K_{22})^2 & ; X_{2i} \geq K_{22} \\ 0 & ; X_{2i} < K_{22} \end{cases}$$

$$(X_{3i} - K_{31})_+^2 = \begin{cases} (X_{3i} - K_{31})^2 & ; X_{3i} \geq K_{31} \\ 0 & ; X_{3i} < K_{31} \end{cases}$$

$$(X_{3i} - K_{32})_+^2 = \begin{cases} (X_{3i} - K_{32})^2 & ; X_{3i} \geq K_{32} \\ 0 & ; X_{3i} < K_{32} \end{cases}$$

$$(X_{4i} - K_{41})_+^2 = \begin{cases} (X_{4i} - K_{41})^2 & ; X_{4i} \geq K_{41} \\ 0 & ; X_{4i} < K_{41} \end{cases}$$

$$(X_{4i} - K_{42})_+^2 = \begin{cases} (X_{4i} - K_{42})^2 & ; X_{4i} \geq K_{42} \\ 0 & ; X_{4i} < K_{42} \end{cases}$$

$$(X_{5i} - K_{51})_+^2 = \begin{cases} (X_{5i} - K_{51})^2 & ; X_{5i} \geq K_{51} \\ 0 & ; X_{5i} < K_{51} \end{cases}$$

$$(X_{5i} - K_{52})_+^2 = \begin{cases} (X_{5i} - K_{52})^2 & ; X_{5i} \geq K_{52} \\ 0 & ; X_{5i} < K_{52} \end{cases}$$

$$(f_{1i} - K_{61})_+^2 = \begin{cases} (f_{1i} - K_{61})^2 & ; Y_{1i} \geq K_{61} \\ 0 & ; Y_{1i} < K_{61} \end{cases}$$

$$(f_{1i} - K_{62})_+^2 = \begin{cases} (f_{1i} - K_{62})^2 & ; Y_{1i} \geq K_{62} \\ 0 & ; Y_{1i} < K_{62} \end{cases}$$

The nonparametric truncated spline path model at cubic time (order p=3) with 1 knot point for five exogenous variables and two endogenous variables will be obtained as follows.

$$\hat{f}_{1i} = \hat{\beta}_{10} + \hat{\beta}_{11}X_{1i} + \hat{\beta}_{12}X_{1i}^2 + \hat{\beta}_{13}X_{1i}^3 + \hat{\delta}_{11}(X_{1i} - K_{11})_+^3 + \hat{\beta}_{14}X_{2i} + \hat{\beta}_{15}X_{2i}^2 + \hat{\beta}_{16}X_{2i}^3 + \hat{\lambda}_{11}(X_{2i} - K_{21})_+^3 + \hat{\beta}_{17}X_{3i} + \hat{\beta}_{18}X_{3i}^2 + \hat{\beta}_{19}X_{3i}^3 + \hat{\gamma}_{11}(X_{3i} - K_{31})_+^3 + \hat{\beta}_{110}X_{4i} + \hat{\beta}_{111}X_{4i}^2 + \hat{\beta}_{112}X_{4i}^3 + \hat{\rho}_{11}(X_{4i} - K_{41})_+^3 + \hat{\beta}_{113}X_{5i} + \hat{\beta}_{114}X_{5i}^2 + \hat{\beta}_{115}X_{5i}^3 + \hat{\theta}_{11}(X_{5i} - K_{51})_+^2$$

$$\hat{f}_{2i} = \hat{\beta}_{20} + \hat{\beta}_{21}X_{1i} + \hat{\beta}_{22}X_{1i}^2 + \hat{\beta}_{23}X_{1i}^3 + \hat{\delta}_{21}(X_{1i} - K_{11})_+^3 + \hat{\beta}_{24}X_{2i} + \hat{\beta}_{25}X_{2i}^2 + \hat{\beta}_{26}X_{2i}^3 + \hat{\lambda}_{21}(X_{2i} - K_{21})_+^3 + \hat{\beta}_{27}X_{3i} + \hat{\beta}_{28}X_{3i}^2 + \hat{\beta}_{29}X_{3i}^3 + \hat{\gamma}_{21}(X_{3i} - K_{31})_+^3 + \hat{\beta}_{210}X_{4i} + \hat{\hat{\beta}}_{211}X_{4i}^2 + \hat{\beta}_{212}X_{4i}^3 + \hat{\rho}_{21}(X_{4i} - K_{41})_+^3 + \hat{\beta}_{213}X_{5i} + \hat{\beta}_{214}X_{5i}^2 + \hat{\beta}_{215}X_{5i}^3 + \hat{\theta}_{21}(X_{5i} - K_{51})_+^3 + \hat{\beta}_{216}f_{1i} + \hat{\beta}_{217}f_{1i}^2 + \hat{\beta}_{218}f_{1i}^3 + \hat{\alpha}_{21}(f_{1i} - K_{61})_+^3 \quad (17)$$

The nonparametric truncated spline path model at cubic time (order p=3) with 2 knot points for five exogenous variables and two endogenous variables will be obtained as follows.

$$\hat{f}_{1i} = \hat{\beta}_{10} + \hat{\beta}_{11}X_{1i} + \hat{\beta}_{12}X_{1i}^2 + \hat{\beta}_{13}X_{1i}^3 + \hat{\delta}_{11}(X_{1i} - K_{11})_+^3 + \hat{\delta}_{12}(X_{1i} - K_{12})_+^3 + \hat{\beta}_{14}X_{2i} + \hat{\beta}_{15}X_{2i}^2 + \hat{\beta}_{16}X_{2i}^3 + \hat{\lambda}_{11}(X_{2i} - K_{21})_+^3 + \hat{\lambda}_{12}(X_{2i} - K_{22})_+^3 + \hat{\beta}_{17}X_{3i} + \hat{\beta}_{18}X_{3i}^2 + \hat{\beta}_{19}X_{3i}^3 + \hat{\gamma}_{11}(X_{3i} - K_{31})_+^3 + \hat{\gamma}_{12}(X_{3i} - K_{32})_+^3 + \hat{\beta}_{110}X_{4i} + \hat{\beta}_{111}X_{4i}^2 + \hat{\beta}_{112}X_{4i}^3 + \hat{\rho}_{11}(X_{4i} - K_{41})_+^3 + \hat{\rho}_{12}(X_{4i} - K_{42})_+^3 + \hat{\beta}_{113}X_{5i} + \hat{\beta}_{114}X_{5i}^2 + \hat{\beta}_{115}X_{5i}^3 + \hat{\theta}_{11}(X_{5i} - K_{51})_+^3 + \hat{\theta}_{12}(X_{5i} - K_{51})_+^3$$

$$\hat{f}_{2i} = \hat{\beta}_{20} + \hat{\beta}_{21}X_{1i} + \hat{\beta}_{22}X_{1i}^2 + \hat{\beta}_{23}X_{1i}^3 + \hat{\delta}_{21}(X_{1i} - K_{11})_+^3 + \hat{\delta}_{22}(X_{1i} - K_{12})_+^3 + \hat{\beta}_{24}X_{2i} + \hat{\beta}_{25}X_{2i}^2 + \hat{\beta}_{26}X_{2i}^3 + \hat{\lambda}_{21}(X_{2i} - K_{21})_+^3 + \hat{\lambda}_{22}(X_{2i} - K_{22})_+^3 + \hat{\beta}_{27}X_{3i} + \hat{\beta}_{28}X_{3i}^2 + \hat{\beta}_{29}X_{3i}^3 + \hat{\gamma}_{21}(X_{3i} - K_{31})_+^3 + \hat{\gamma}_{22}(X_{3i} - K_{32})_+^3 + \hat{\beta}_{210}X_{4i} + \hat{\beta}_{211}X_{4i}^2 + \hat{\beta}_{212}X_{4i}^3 + \hat{\rho}_{21}(X_{4i} - K_{41})_+^3 + \hat{\rho}_{22}(X_{4i} - K_{42})_+^3 + \hat{\beta}_{213}X_{5i} + \hat{\beta}_{214}X_{5i}^2 + \hat{\beta}_{215}X_{5i}^3 + \hat{\theta}_{21}(X_{5i} - K_{51})_+^3 + \hat{\theta}_{22}(X_{5i} - K_{52})_+^3 + \hat{\beta}_{216}f_{1i} + \hat{\beta}_{217}f_{1i}^2 + \hat{\beta}_{218}f_{1i}^3 + \hat{\alpha}_{21}(f_{1i} - K_{61})_+^3 + \hat{\alpha}_{22}(f_{1i} - K_{62})_+^3 \quad (18)$$

where function truncated :

$$(X_{1i} - K_{11})_+^3 = \begin{cases} (X_{1i} - K_{11})^3 & ; X_{1i} \geq K_{11} \\ 0 & ; X_{1i} < K_{11} \end{cases}$$

$$(X_{1i} - K_{12})_+^3 = \begin{cases} (X_{1i} - K_{12})^3 & ; X_{1i} \geq K_{12} \\ 0 & ; X_{1i} < K_{12} \end{cases}$$

$$(X_{2i} - K_{21})_+^3 = \begin{cases} (X_{2i} - K_{21})^3 & ; X_{2i} \geq K_{21} \\ 0 & ; X_{2i} < K_{21} \end{cases}$$

$$(X_{2i} - K_{22})_+^3 = \begin{cases} (X_{2i} - K_{22})^3 & ; X_{2i} \geq K_{22} \\ 0 & ; X_{2i} < K_{22} \end{cases}$$

$$(X_{3i} - K_{31})_+^3 = \begin{cases} (X_{3i} - K_{31})^3 & ; X_{3i} \geq K_{31} \\ 0 & ; X_{3i} < K_{31} \end{cases}$$

$$(X_{3i} - K_{32})_+^3 = \begin{cases} (X_{3i} - K_{32})^3 & ; X_{3i} \geq K_{32} \\ 0 & ; X_{3i} < K_{32} \end{cases}$$

$$(X_{4i} - K_{41})_+^3 = \begin{cases} (X_{4i} - K_{41})^3 & ; X_{4i} \geq K_{41} \\ 0 & ; X_{4i} < K_{41} \end{cases}$$

$$(X_{4i} - K_{42})_+^3 = \begin{cases} (X_{4i} - K_{42})^3 & ; X_{4i} \geq K_{42} \\ 0 & ; X_{4i} < K_{42} \end{cases}$$

$$(X_{5i} - K_{51})_+^3 = \begin{cases} (X_{5i} - K_{51})^3 & ; X_{5i} \geq K_{51} \\ 0 & ; X_{5i} < K_{51} \end{cases}$$

$$(X_{5i}-K_{52})_+^3=\begin{cases}(X_{5i}-K_{52})^3 & ;\ X_{5i}\geq K_{52}\\ 0 & ;\ X_{5i}<K_{52}\end{cases}$$

$$(f_{1i}-K_{61})_+^3=\begin{cases}(f_{1i}-K_{61})^3 & ;\ f_{1i}\geq K_{61}\\ 0 & ;\ f_{1i}<K_{61}\end{cases}$$

$$(f_{1i}-K_{62})_+^3=\begin{cases}(f_{1i}-K_{62})^3 & ;\ f_{1i}\geq K_{62}\\ 0 & ;\ f_{1i}<K_{62}\end{cases}$$

Nonparametric Truncated Spline Path Analysis Function Estimation Using WLS

In nonparametric path analysis, parameter estimation can be done using the weighted least square (WLS) method, which minimizes the following equation.

$$\min\left\{\underset{\sim}{\varepsilon}^T\boldsymbol{\Sigma}^{-1}\underset{\sim}{\varepsilon}\right\}=\min\left\{\left(\underset{\sim}{y}-\boldsymbol{X}\underset{\sim}{\beta}\right)^T\boldsymbol{\Sigma}^{-1}\left(\underset{\sim}{y}-\boldsymbol{X}\underset{\sim}{\beta}\right)\right\} \tag{21}$$

Then do the partial derivatives, and get the following equation:.

$$\begin{aligned}Q\left(\underset{\sim}{\beta}\right) &= \left(\underset{\sim}{y}-\boldsymbol{X}\underset{\sim}{\beta}\right)^T\boldsymbol{\Sigma}^{-1}\left(\underset{\sim}{y}-\boldsymbol{X}\underset{\sim}{\beta}\right)\\ &= \left(\underset{\sim}{y}^T-\underset{\sim}{\beta}^T\boldsymbol{X}^T\right)\boldsymbol{\Sigma}^{-1}\left(\underset{\sim}{y}-\boldsymbol{X}\underset{\sim}{\beta}\right)\\ &= \underset{\sim}{y}^T\boldsymbol{\Sigma}^{-1}\underset{\sim}{y}-\underset{\sim}{\beta}^T\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\underset{\sim}{y}-\underset{\sim}{y}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\underset{\sim}{\beta}+\underset{\sim}{\beta}^T\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\underset{\sim}{\beta}\\ &= \underset{\sim}{y}^T\boldsymbol{\Sigma}^{-1}\underset{\sim}{y}-2\underset{\sim}{\beta}^T\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\underset{\sim}{y}+\underset{\sim}{\beta}^T\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\underset{\sim}{\beta}\end{aligned} \tag{20}$$

Next, equation (2.20) is derived with respect to ~ and equated to zero.

$$\begin{aligned}\frac{\partial Q\left(\underset{\sim}{\beta}\right)}{\partial\underset{\sim}{\beta}} &= -2\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\underset{\sim}{y}+2\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\underset{\sim}{\beta}\\ 0 &= -2\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\underset{\sim}{y}+2\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\underset{\sim}{\beta}\\ 2\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\underset{\sim}{\beta} &= 2\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\underset{\sim}{y}\\ \boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\underset{\sim}{\beta} &= \boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\underset{\sim}{y}\\ \underset{\sim}{\hat{\beta}} &= (\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\underset{\sim}{y}\end{aligned} \tag{21}$$

The next step is the equation (21) formed into $\underset{\sim}{\hat{f}}=\boldsymbol{X}\underset{\sim}{\hat{\beta}}$, and the estimator of the spline function is obtained as follows:.

$$\underset{\sim}{\hat{f}}=\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\underset{\sim}{y} \tag{22}$$

with $\underset{\sim}{\hat{f}}=\left(\hat{f}_1,\hat{f}_2\right)^T$. Equation (22) can be simplified in the form $\underset{\sim}{\hat{f}}=\boldsymbol{H}(\underset{\sim}{\boldsymbol{K}})\underset{\sim}{y}$, di mana $\boldsymbol{H}(\underset{\sim}{\boldsymbol{K}})=\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}$. $\boldsymbol{H}(\underset{\sim}{\boldsymbol{K}})$ is a function of the knot point, while$\boldsymbol{K}=(k_{11},\dots,k_{1K},k_{21},\dots,k_{2K})$ are knot points.

Solution for Regression Coefficients Using IRLS

To minimize the regression equation, the first partial derivative of the equation with respect to the parameter, must be equal to 0. So that will produce a necessary condition for the minimum that results in the equation of the derivative function:

$$\frac{\partial f(x)}{\partial\beta_j}=0$$

$$\sum_{i=1}^{n}X_{ij}\psi\left(\frac{y_i-\sum_{j=0}^{k}x_{ij}\beta_j}{s}\right)=0,\qquad j=0,1,\dots,k \tag{23}$$

with $\psi=\rho'$ dan x_{ij} is the i-th observation on the regressor ke-j and $x_{i0}=1$. Define a weight function fungsi

$$w(u_i) = \frac{\psi\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{s}\right)}{y_i - \sum_{j=0}^{k} x_{ij}\beta_j} \tag{24}$$

and $w_i = w(u_i)$. Then the equation of the derivative function above can be written as

$$\sum_{i=1}^{n} X_{ij} \mathrm{w_i}\left(y_i - \sum_{j=0}^{k} x_{ij}\beta_j\right) = 0, \qquad j = 0,1,\dots,k \tag{25}$$

n general, the function is not linear and the derivative equation above must be solved by the iteration method.
Regression coefficient estimation with M-estimation is performed with least squares estimation with iterative weighting. This estimation procedure requires an iteration process where w_i will change in each iteration so that we get _0,β _1,…,β _k. The procedure is called Iteratively Reweighted Least Squares (IRLS). To use IRLS, assume that an initial estimate of _0 exists and s is a scale estimate. For parameters where p is the number of parameters to be estimated, then

$$\sum_{i=1}^{n} X_{ij} w_i^0\left(y_i - \sum_{j=0}^{k} x_{ij}\hat{\beta}_j^0\right) = 0, \qquad j = 0,1,\dots,k \tag{26}$$

$$w_i^0 = \begin{cases} \dfrac{\psi\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\hat{\beta}_j^0}{s}\right)}{y_i - \sum_{j=0}^{k} x_{ij}\hat{\beta}_j^0}, & \text{jika } y_i \neq \sum_{j=0}^{k} x_{ij}\hat{\beta}_j^0 \\ 1, & \text{jika } y_i = \sum_{j=0}^{k} x_{ij}\hat{\beta}_j^0 \end{cases}$$

For the case of multiple regression, the parameter calculation can be obtained from the matrix equation

$$\mathbf{X'W^0X}\hat{\beta} = \mathbf{X'W^0Y} \tag{27}$$

$\mathbf{W^0}$ is a diagonal matrix of size (n × n) of weights with diagonal elements$w_1^0, w_2^0, \dots, w_n^0$. Therefore, the one-step estimator

$$\text{is } \hat{\beta} = (\mathbf{X'W^0X})^{-1}\mathbf{X'W^0Y} \tag{28}$$

In the next step, we recalculate the weight from w_i=w(u_i) but use ^1 instead of ^0, and so on. This iteration calculation is stopped when changes occur in the regression coefficient, namely the difference between ^(t+1) and ^t is less than 0.1% with t = 0.1,.... Estimated regression coefficient with estimation -M IRLS can be written by the formula:

(β)^^(t+1)=(X^' W^t X)^(-1) X^' W^t Y (29)

The least squares estimate can be used as a starting value, ^0. Furthermore, for ^2 can be written as follows:

$$\hat{\beta}^2 = \left(\mathbf{X'W^1X}\right)^{-1}\mathbf{X'W^1Y} \tag{30}$$

Generalized Cross Validation (GCV)

The GCV method is a method that is very often used to select the most optimal knot point. The best spline is characterized by the optimal knot point obtained. One method of selecting the optimal knot point is GCV. The best spline model with optimal knot points is obtained from the smallest GCV value. As previously explained, the GCV method is a modified form of the CV method. Where the CV equation is as follows.:

$$\mathrm{CV}(\tilde{k}) = n^{-1}\sum_{i=1}^{n}\left(\frac{y_i - \hat{\mu}_{k(j)}}{1 - \mathbf{A}(k_1, k_2, \dots, k_j)}\right)^2 \tag{31}$$

While the GCV equation is given by the following equation:

$$\mathrm{GCV}(\tilde{k}) = n^{-1}\sum_{i=1}^{n}\left(y_i - \hat{\mu}_{k(j)}\right)^2 \mathrm{Q} \tag{32}$$

where:

$$Q = \left\{\frac{1 - \mathbf{A}(k_1, k_2, \dots, k_j)}{n^{-1}\mathrm{trace}\left[\mathbf{I} - \mathbf{A}(k_1, k_2, \dots, k_j)\right]}\right\}^2$$

The GCV equation can also be written with the following equation:

$$GCV(\tilde{k}) = n^{-1} \frac{\sum_{i=1}^{n}(y_i - \hat{f}(x))^2}{(1 - n^{-1}\text{trace}[\mathbf{A}(\tilde{k})])^2} \quad (33)$$

where : $\hat{f}(x) = \mathbf{A}(\tilde{k})\tilde{Y} = [\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t]\tilde{Y}$

$\mathbf{A}(\tilde{k}) = [\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t]$

$MSE = n^{-1}\sum_{i=1}^{n}(y_i - \hat{f}(x))^2$

I = Matriks Identitas

n = Many Observations

3 Research Method

3.1Research data

The data used in this study is primary data. Data was collected through a questionnaire with a Likert scale. The sampling technique used is purposive sampling. Purposive sampling is a sampling technique based on certain characteristics or conditions that are the same as the characteristics of the population. The sample used is 100 customers of Company X. This study uses several variables, namely Company Performance (X_1), Product Quality (X_2), Government Role (Y_1), and Community Welfare (Y_2),

Research Model

The variables used consisted of three exogenous variables, one endogenous variable, and one mediating variable. Exogenous variables include Company Performance (X1) and Product Quality (X2) variables. The mediating variable is the role of the government (Y1). While the endogenous variable is community welfare (Y2). The research path diagram is shown in Figure 1.

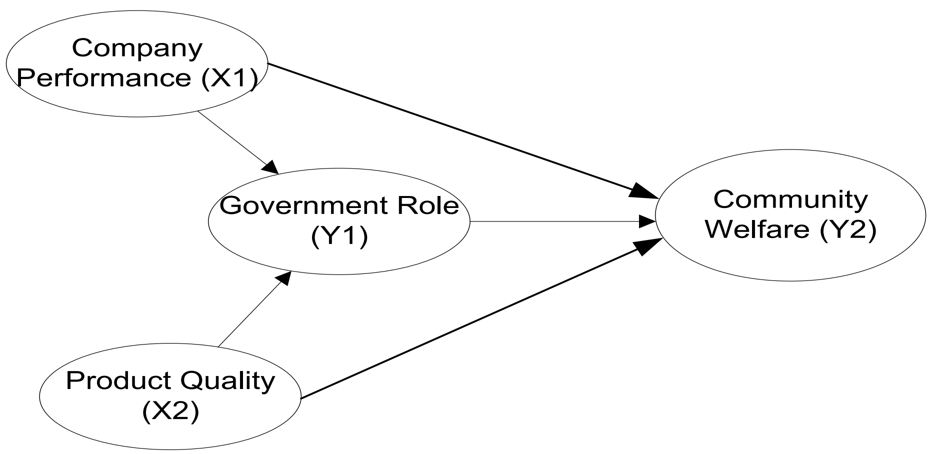

Fig.1: Research Path Diagram.

Research procedure

The steps taken are as follows:

1. Create a path diagram.
2. Test the linearity assumption with Ramsey's Regression Specification Error Test (RESET).
3. Estimating the path function with linear, quadratic, and cubic truncated splines with 1 knot point and 2 knot points using the IRLS method.
4. Get the results of estimating the function of each polynomial degree, namely linear,quadratic, and cubic with the number of knots, namely 1 knot point, and 2 knot point
5. Selection of optimum model and optimum knot point based on the smallest GCV coefficient.
6. Interpreting the results of the best nonparametric truncated spline path function estimation results and drawing conclusions from the results of the analysis.

4 Results of Analysis and Discussion

Linearity Test

In statistical modeling, information about the pattern of relationships between variables is needed to determine whether a method used is a parametric or nonparametric approach. To use a nonparametric approach, a nonlinear relationship is required. The results of linearity testing with the Ramsey RESET test are presented in Table 4.1.

Table 4.1: Ramsey RESET Linearity Test Results.

Variabel	**P-Value**	**Relationship**
X_1 to Y_1	0,6417	Linier
X_2 to Y_1	0,0007	Non Linier
X_1 to Y_2	0,9428	Linier
X_2 to Y_2	0,1257	Linier
Y_1 to Y_2	0,0107	Non Linier

Based on the results of linearity testing in Table 4.1, the test results show that there are five linear relationships between variables, but there are two nonlinear relationships between variables, so a nonparametric approach can be used.

Truncated Spline Nonparametric Path Modeling

Nonparametric truncated spline path modeling was performed on linear, quadratic, and cubic polynomial degrees with 1 and 2 knot points. The best nonparametric truncated spline path model is obtained when the knot point is optimal. Meanwhile, to obtain the optimal knot point, it is necessary to find the smallest GCV value. In addition, the selection of the best polynomial degree can also be done by looking at the coefficient of the largest determinant. The results of the calculation of the GCV value and the determinant coefficient for each model are shown in Table 4.2.

Table 4.2: GCV Value of Truncated Spline Nonparametric Path Model.

Polynomial Degrees	**Many Knots**	**GCV Value**	**R^2**
Linier	1 *Knot*	4,6378	0,9820
	2 *Knot*	4,4523	0,9834
Kuadratik	1 *Knot*	4,2687	0,9840
	2 *Knot*	4,2223	0,9851
Kubik	1 *Knot*	4,4536	0,9834
	2 *Knot*	4,2597	0,9848

Based on Table 4.2, it is known that the smallest GCV value of 4.2223 and the largest R2 value of 0.9851 resulted in a nonparametric truncated spline path model with quadratic polynomial degrees (order p=2) with 2 knot points. The model has an R2 value of 98.51% which indicates that the former model explains the diversity of endogenous variables by 98.51% and the rest is explained by other factors that cannot be known in the model by 1.49%. Nonparametric path model truncated spline quadratic polynomial degree (order p=2) with 2 knot points for three exogenous variables and two endogenous variables, namely as follows.

$$\begin{aligned}\hat{f}_{1i} = \;& 28,9448 - 1,0334X_{1i} + 0,2032X_{1i}^2 + 0,0414(X_{1i}-3)_+^2 - 1,1395(X_{1i}-3,8)_+^2 \\ & -9,3725X_{2i} + 1,7252X_{2i}^2 - 2,0502(X_{2i}-3)_+^2 - 2,0791(X_{2i}-4)_+^2 - 2,8469X_{3i} \\ & +0,5793X_{3i}^2 - 0,5806(X_{3i}-2,7)_+^2 - 0,0168(X_{3i}-3,2)_+^2 - 0,4091X_{4i} + 0,0838X_{4i}^2 \\ & -0,0343(X_{4i}-3)_+^2 + 0,9227(X_{4i}-3,8)_+^2 - 7,4427X_{5i} + 1,5503X_{5i}^2 \\ & -3,1335(X_{5i}-2,8)_+^2 + 1,8673(X_{5i}-3,3)_+^2\end{aligned}$$

$$\begin{aligned}\hat{f}_{2i} = \;& 10,0672 + 2,7210X_{1i} - 0,370X_{1i}^2 + 1,3003(X_{1i}-3)_+^2 - 2,5005(X_{1i}-3,8)_+^2 - 2,4852X_{2i} \\ & +0,4703X_{2i}^2 - 0,6088(X_{2i}-3)_+^2 - 0,0784(X_{2i}-4)_+^2 + 1,4792X_{3i} - 0,3793X_{3i}^2 \\ & +1,0663(X_{3i}-2,7)_+^2 - 0,8481(X_{3i}-3,2)_+^2 + 3,7425X_{4i} - 0,7437X_{4i}^2 \\ & +1,6570(X_{4i}-3)_+^2 - 1,4620(X_{4i}-3,8)_+^2 - 4,7747X_{5i} + 0,9509X_{5i}^2 \\ & -1,5293(X_{5i}-2,8)_+^2 + 0,4344(X_{5i}-3,3)_+^2 - 6,5017f_{1i} + 1,4050f_{1i}^2 \\ & -2,4445(f_{1i}-2,7)_+^2 + 1,3564(f_{1i}-3,2)_+^2\end{aligned}$$

From the truncated spline function that has been obtained previously, then the values of f1 and f2 are obtained. In this study, the first step to estimate the nonparametric path smoothing spline function is to do iterative weighting. The first step in using iterative weighted (IRLS) is to estimate the function using PLS (Penalixed Least Square) which is then iterated using PWLS (Penalized Weighted Least Square). The weight used in iterative weighted is the square of the error formed in the previous equation and the weighting is always updated in each iteration. By using R software, the optimal lambda estimation results are obtained at each iteration. The balance controller between the fit of the curve to the data and the smoothness of the curve is indicated by the smoothing parameter. The smoothing parameter is 0.006 for model 1 and 0.142 for model 2. Based on the resulting output, the coefficient estimator value is as follows.

Table 4.3: Coefficient Value of Smoothing Spline Nonparametric Path Model.

Koefisien	f_1	f_2
$\hat{d}_{11}$	0.3018	0.9605
$\hat{d}_{12}$	0.8249	0.1045
$\hat{c}_{11}$	0.1610	0.5662
⋮	⋮	⋮
$\hat{c}_{1499}$	0.4667	0.9148
$\hat{c}_{1500}$	0.1599	0.5728

The function of each endogenous variable can be seen in the equation below and substitute the value of x to obtain an estimate of the endogenous variable.
Fungsi 1:

$$\hat{f}_{11}=\hat{d}_{11}+\hat{d}_{12}x_{11}+\hat{c}_{11}\left[x_{11}x_{11}-\frac{1}{2}\left(x_{11}+x_{11}\right)+\frac{1}{3}\right]+\hat{c}_{12}\left[x_{11}x_{12}-\frac{1}{2}\left(x_{11}+x_{12}\right)+\frac{1}{3}\right]$$
$$+\ldots+\hat{c}_{1500}\left[x_{51}x_{5100}-\frac{1}{2}\left(x_{51}+x_{5100}\right)+\frac{1}{3}\right]$$

$$\hat{f}_{11}=0.3018+0.8249x_{11}+0.1610\left[x_{11}x_{11}-\frac{1}{2}\left(x_{11}+x_{11}\right)+\frac{1}{3}\right]+0.9603\left[x_{11}x_{12}-\frac{1}{2}\left(x_{11}+x_{12}\right)+\frac{1}{3}\right]$$
$$+\ldots+0.1599\left[x_{51}x_{5100}-\frac{1}{2}\left(x_{51}+x_{5100}\right)+\frac{1}{3}\right]$$

Fungsi 2:

$$\hat{f}_{21}=\hat{d}_{21}+\hat{d}_{22}\left(x_{11}+y_{2i}\right)+\hat{c}_{21}\left[x_{11}x_{11}-\frac{1}{2}\left(x_{11}+x_{11}\right)+\frac{1}{3}+y_{11}y_{11}-\frac{1}{2}\left(y_{11}+y_{11}\right)+\frac{1}{3}\right]$$
$$+\hat{c}_{22}\left[x_{11}x_{12}-\frac{1}{2}\left(x_{11}+x_{12}\right)+\frac{1}{3}+y_{11}y_{12}-\frac{1}{2}\left(y_{11}+y_{12}\right)+\frac{1}{3}\right]$$
$$+\ldots+\hat{c}_{2500}\left[x_{51}x_{5100}-\frac{1}{2}\left(x_{51}+x_{5100}\right)+\frac{1}{3}+y_{11}y_{1100}-\frac{1}{2}\left(y_{11}+y_{1100}\right)+\frac{1}{3}\right]$$

$$\hat{f}_{21} = 0.9605 + 0.1045\left(x_{11} + y_{2i}\right) + 0.5662\left[x_{11}x_{11} - \frac{1}{2}\left(x_{11} + x_{11}\right) + \frac{1}{3} + y_{11}y_{11} - \frac{1}{2}\left(y_{11} + y_{11}\right) + \frac{1}{3}\right]$$

$$+0.0443\left[x_{11}x_{12} - \frac{1}{2}\left(x_{11} + x_{12}\right) + \frac{1}{3} + y_{11}y_{12} - \frac{1}{2}\left(y_{11} + y_{12}\right) + \frac{1}{3}\right]$$

$$+\ldots + 0.5728\left[x_{51}x_{5100} - \frac{1}{2}\left(x_{51} + x_{5100}\right) + \frac{1}{3} + y_{11}y_{1100} - \frac{1}{2}\left(y_{11} + y_{1100}\right) + \frac{1}{3}\right]$$

The Best Truncated Spline Nonparametric Path Model Relationship Pattern

The nonparametric path model truncated spline quadratic polynomial degree (order p=2) with 2 knot points has several relationships, namely the relationship between the variables of Company Performance, Product Quality, and Company Performance (X1) with the role of the government (Y1) and Community welfare (Y2) . In the form of an image, only the relationship between one exogenous variable and one endogenous variable is presented to facilitate discussion, where the influence of other exogenous variables on endogenous variables is considered constant. If presented in the form of a curve, the relationship between variables (X1) and (Y1) is as follows.

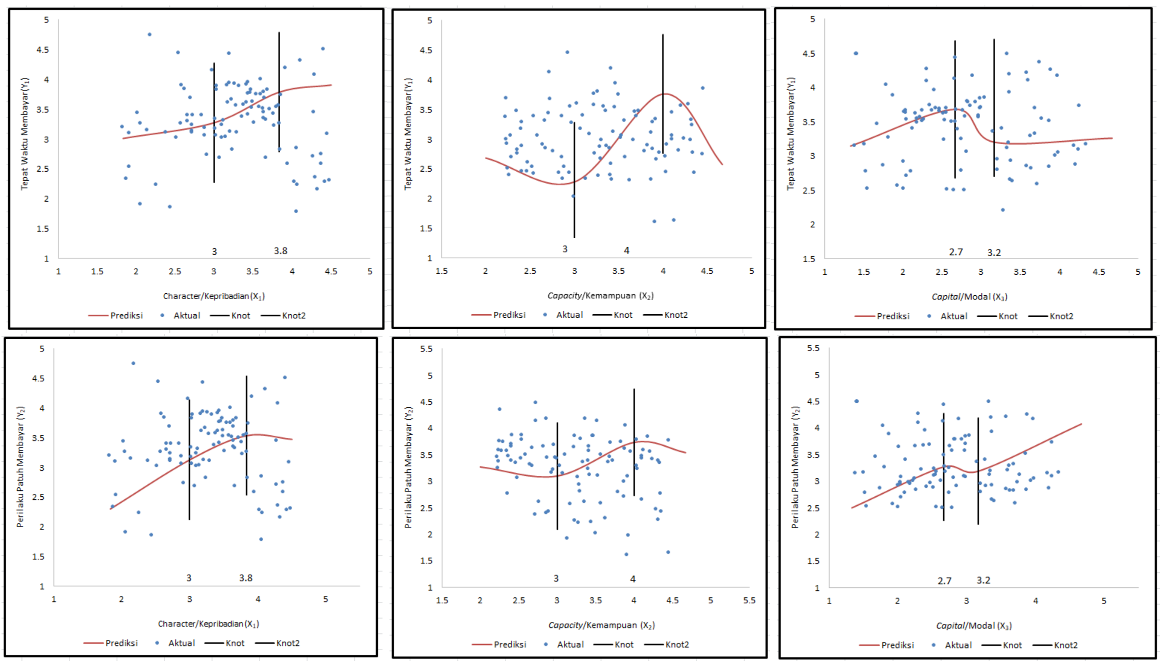

Fig. 4.1. 2 Point Quadratic Truncated Spline Nonparametric Path Model Functions Knot.
a) X1 to Y1, b) X2 to Y1, c) X3 to Y1, d) X4 to Y1, e) X5 to Y1,
f) X1 to Y2, g) X2 to Y2, h) X3 to Y2, i) X4 to Y2, j) X5 to Y2.

Based on Figure 4.1, the relationship between each endogenous variable and exogenous variables in the nonparametric path model truncated spline quadratic polynomial degree (order p=2) with 2 knot points is divided into three regimes. Changes in data behavior patterns occur during economic conditions during this pandemic. The pattern of data distribution of each variable shows a difference in the first regime, second regime, and third regime. One of them is the relationship between variables (X1) and (Y1). The first regime is shown when X1i < 3, the second regime when 3<X1i< 3.8, and the third regime when X〗 _1i > 3.8. The above results can be used in different directions, i.e. information analytics using artificial intelligence [13-16].

5 Conclusions

Based on the results of the analysis and discussion that has been carried out, it can be concluded that:

1. Estimation of nonparametric path model functions truncated spline orders of linear, quadratic, and cubic 1 and 2 knot points on endogenous variables that affect exogenous variables. The best model is obtained on the

nonparametric truncated spline¬ quadratic path model of 2 knots. This model has the smallest GCV value compared to other models, which is 4.2223 and the R2 value is 98.51%.

2. The estimation of the best nonparametric smoothing spline path model is obtained from the quadratic order truncated spline nonparametric path model function with 2 knot points where the model shows changes in data behavior patterns that occur when the company's performance is quite good.
3. Estimation of path analysis coefficients using the iterative weighted method (IRLS) can be applied to the wishes and needs of the community with the moderating variable of the role of the government

Suggestions

Some suggestions for further similar research are as follows:

1. This study only estimates nonparametric truncated spline fungi paths of linear, quadratic, and cubic orders. For further research, it is possible to test hypotheses and significance.
2. This study only uses one and two knot points. For further research, more than two knot points can be used.
3. This study does not interpret the path analysis which has a direct influence, an indirect effect, or a total effect. For further research, it is possible to carry out a more detailed interpretation using simulation data.
4. Reduce the tolerance limit on iterative weighted so that a more optimal lambda is obtained.

Conflicts of Interest: The authors declare that there is no conflict of interest regarding the publication of this article.

References

[1] Eubank, R. (1988). Spline Smoothing and Nonparametric Regression. Marcel Dekker, New York.

[2] Solimun & Fernandes, A. A. R. (2017). Investigation the mediating variable: What is necessary?(case study in management research). International Journal of Law and Management Vol. 59 No. 6, pp. 1059-1067 DOI 10.1108/IJLMA-09-2016-0077

[3] Fernandes, A. A. R., Hutahayan, B., Arisoesilaningsih, E., Yanti, I., Astuti, A. B., & Amaliana, L. (2019). 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**, 052024). IOP Publishing.

[4] Hidayat, M. F., & Achmad, R. F. A. (2019, June). Estimation of Truncated Spline Function in Non-parametric Path Analysis Based on Weighted Least Square (WLS). In IOP Conference Series: Materials Science and Engineering ., Vol. 546 No. 5, 052027.

[5] Achmad Rinaldo Fernandes, A., Janssen, P., Sa'adah, U., Solimun, Effendi, A., Nurjannah, & Amaliana, L. (2018). Comparison of spline estimator at various levels of autocorrelation in smoothing spline non parametric regression for longitudinal data. Communications in Statistics-Theory and Methods., Vol. **47 No. 21**, 5265-5285.

[6] Eubank, R.L. (1999). Spline Smoothing and Nonparametric Regression, Second Edition. New York : Marcel Dekker, Inc.

[7] Fernandes, A. A. R., Budiantara, I. N. I., Otok, B. W., & Suhartono. (2014). Reproducing Kernel Hilbert space for penalized regression multi-predictors: Case in longitudinal data. International Journal of Mathematical Analysis., Vol. **8 No. 40**, 1951-1961.

[8] Rasheed, B. A., Adnan, R., Saffari, S. E., & dano Pati, K. (2014). Robust weighted least squares estimation of regression parameter in the presence of outliers and heteroscedastic errors. Jurnal Teknologi., Vol. **71** No. 1 DOI: 10.11113/jt.v71.3609

[9] Purbawangsa, I. B. A., Solimun, S., Fernandes, A. A. R., & Rahayu, S. M. (2019). 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. vol. 16, No. 7, 983-999

[10] Fernandes, A. A. R., Solimun, & Arisoesilaningsih, E. (2017). Estimation of spline function in nonparametric path analysis based on penalized weighted least square (PWLS). In AIP Conference Proceedings (Vol. 1913, No. 1, p. 020030). AIP Publishing LLC

[11] Budiantara, I. N. (2011). Penelitian Bidang Regresi Spline Menuju Terwujudnya Penelitian Statistika yang Mandiri dan Berkarakter. In Prosiding Seminar Nasional MIPA.

[12] Bintariningrum, M. F., & Budiantara, I. N. (2014). Pemodelan regresi nonparametrik truncated splinedan aplikasinya pada Angka Kelahiran Kasar di Surabaya. Jurnal Sains dan Seni ITS., **3(1)**, D7-D12.

[13] Hamed Fazlollahtabar, (2021) Online Business Quality of Service Information Analytics using Artificial Intelligence, Inf. Sci. Lett. Vol. 10, No. 1, 93-100 doi:10.18576/isl/100111

[14] Ammar Al-Sammarraee and Nadya Alshareeda, (2021) The Role of Artificial Intelligence by Using Automatic Accounting Information System in Supporting the Quality of Financial Statement, Inf. Sci. Lett. Vol. 10, No. 2, 223-254, doi:10.18576/isl/100208

[15] Eman Wadie Abdel Halim, (2021) The Influence of University Students Inclination and Their Intentions towards Self-Employment Decision: Applied to Students of Entrepreneurship Clubs, Inf. Sci. Lett. Vol. 10, No. 2, PP:341-376, doi:10.18576/isl/100215

[16] Rajab I. Ismail, A. Al-Sammarraee, N. ALshareeda, M. Adel Al Araifi, (2021) The Impact of Monetary Policy Instruments on GDP and Inflation in the Kingdom of Bahrain Aligned with 2030 Vision, Inf. Sci. Lett. Vol. 10, No. 2, 277-321 doi:10.18576/isl/100212