LATEN VARIABLES PATH NONPARAMETRIC ANALYSIS: ESTIMATING FUNCTIONS AND TESTING HYPOTHESES

Fariq Hidayat[1], Adji Achmad Rinaldo Fernandes, Solimun

Abstract---*This study is to identify patterns of relationships between manifest variables and latent variables. Parametric path analysis is one of the proper analytical techniques used to form unknown curves. The phenomenon states it is difficult to get a curve shape or curve shape is unknown. Nonparametric path analysis is used to estimate the function with the truncated spline approach which contains parameters of polynomial degrees and knots. The case of post-mining revegetation technique is used as a case study because the pattern of relationships between variables is not linear. PCA is carried out to measure the indicators of each variable, so that the contribution of component variables and variables used are reflected by the strongest indicators. The results showed that the best model of path analysis with linear polynomial degrees and knot point 3 with GCV values of 1738,303.*

Key words---*Path Analysis, PCA, GCV, Truncated spline, polynomial, knots, Nonparametric, Post-mining Land*

I. INTRODUCTION

Nonparametric Regression Model that is widely used, one of which is the spline model. The spline model has excellent statistical and visual interpretation and has a high degree of flexibility (Eubank, 1999). Several types of spline functions that have been studied previously include smoothing spline (Eubank, 1999), B-Spline (Lyche and Morken, 2008), penalized spline (Griggs, 2013), thin plate spline (Wood, 2003), and so on. Budiantara (2005) developed a spline approach in nonparametric regression using the function base of the truncated spline family. Spline truncated function, is a polynomial function that is divided into several parts (regimes) at a knot point.

Knots point is a joint fusion point where the function is divided, or points that describe changes in data behavior at certain sub-intervals (Budiantara, 2009). Smoothing spline has high flexibility in data patterns and has no restrictions in each region, but must find the appropriate smoothing parameters based on the most optimal number of knot points on the model (Fernandes *et al*, 2014). Estimation of the curve or truncated spline nonparametric regression function using the Weighted Least Square (WLS) approach (Fernandes *et al*, 2015). The WLS approach is used for the case by considering the correlation between errors.

Path analysis is basically the development of parametric regression analysis, involving more than one function. However, the parametric path analysis model has not been able to accommodate if the curve shape is unknown (Fernandes *et al*, 2018). Therefore a path analysis model based on nonparametric regression was developed using the truncated spline approach. Hypothesis testing on nonparametric components needs to be done. Hypothesis testing is done by estimating parameters that are approximated by function estimates (Fernandes *et al*, 2019).

This study will discuss the combination of nonparametric truncated spline path analysis used in unknown curve shapes or relationships between variables that do not meet the linearity assumption. Because path analysis is a system modeling in

[1]*University of Brawijaya, Malang, Indonesia, Email: fernandes@ub.ac.id*

which endogenous variables are interrelated, WLS-based estimation will be used which considers the case of correlation between errors, and explains the use of latent variables in the analysis using PCA and simultaneously testing hypotheses on estimating functions.

II. LITERATURE REVIEW

Nonparametric Regression Analysis

Nonparametric regression analysis is a regression approach that aims to determine the pattern of the relationship between endogenous and exogenous, but the shape of the curve is unknown and no past information is found (Budiantara, 2000b). The following is the equation:

$$y = f(t_i) + \varepsilon, \quad t = 1,2,\ldots,i. \tag{2.3}$$

$f(t_i)$: t - regression function

ε : error or random error are normally distributed $E(\varepsilon) = 0$ and $V(\varepsilon) = \sigma_\varepsilon^2$

Multiendogen Nonparametric Regression Analysis

According to Rencher (2002) regression analysis can be distinguished based on the number of endogenous variables called multiendogenic nonparametric regression analysis, but endogenous variables are not interrelated. Following is the equation with the j-th endogenous:

$$y_j = f\left(t_{ji}\right) + \varepsilon_{ji}, \ \ j = 1,2,..,k, \quad x = 1,2,\ldots,i. \tag{2.4}$$

y_j : *j*-th endogenous variables l

$f\left(x_{ji}\right)$: *j*-th endogenous t-regression function

ε : random error are normally distributed $E(\varepsilon) = 0$ and $V(\varepsilon) = \sigma_\varepsilon^2$

Covariance matrix variant of random error [$\boldsymbol{\Sigma}^{-1}(\underset{\sim}{\sigma}^2)$] (Budiantara, et al. 2001*b*).

Parametric Path Analysis

In 1934 Wright developed parametric path analysis based on an extension of parametric regression analysis. Path analysis is a means of studying the direct and indirect effects of several variables in which some variables are seen as causes and other variables are seen as consequences (Dillon and Goldstein, 1984).

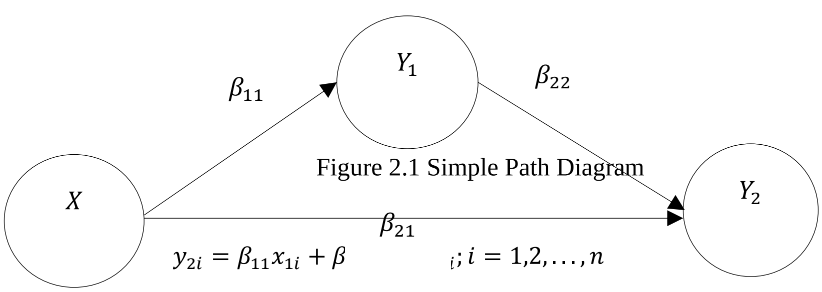

Figure 2.1 Simple Path Diagram

$$y_{1i} = \beta_{11}x_{1i} + \varepsilon_{1i}$$

$$y_{2i} = \beta_{11}x_{1i} + \beta \quad {}_i; i = 1,2,\ldots,n \tag{2.5}$$

Equation (2.5) can be written in matrix form, as follows:

$$\begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2n} \end{bmatrix} = \begin{bmatrix} x_{11} & 0 & 0 \\ x_{12} & 0 & 0 \\ \vdots & \vdots & \vdots \\ x_{1n} & 0 & 0 \\ 0 & x_{11} & y_{11} \\ 0 & x_{12} & y_{12} \\ \vdots & \vdots & \vdots \\ 0 & x_{1n} & y_{1n} \end{bmatrix} \begin{bmatrix} \beta_{1,11} \\ \beta_{1,21} \\ \beta_{2,21} \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{1n} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \vdots \\ \varepsilon_{2n} \end{bmatrix} \tag{2.6}$$

$$\underset{\sim}{y} = \mathbf{X}\underset{\sim}{\beta} + \underset{\sim}{\varepsilon} \tag{2.7}$$

di mana:

ε_{ki} : random error endogenous variable *k*-th observation *i*-th normally distributed

2n (two as many endogenous variables as observations), E($\underset{\sim}{\varepsilon}$)= $\underset{\sim}{0}$ (vectors measuring 2n) and V($\underset{\sim}{\varepsilon}$)= $\boldsymbol{\Sigma}_{2n\times 2n}$ variance covariance matrix $\boldsymbol{\Sigma}$ written as follows:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \Sigma_{11}^2 & 0 & \cdots & 0 & \Sigma_{(1,1)} & 0 & \cdots & 0 \\ 0 & \Sigma_{12}^2 & \cdots & 0 & 0 & \Sigma_{(2,2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma_{1n}^2 & 0 & 0 & \cdots & \Sigma_{(n,n)} \\ \Sigma_{(1,1)} & 0 & \cdots & 0 & \Sigma_{21}^2 & 0 & \cdots & 0 \\ 0 & \Sigma_{(2,2)} & \cdots & 0 & 0 & \Sigma_{22}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma_{(n,n)} & 0 & 0 & \cdots & \Sigma_{2n}^2 \end{bmatrix}_{(2n)\times(2n)} \tag{2.8}$$

Sub-matrix $\boldsymbol{\Sigma_{kk}}$ that is based on the matrix $\boldsymbol{\Sigma}_{2n\times 2n}$ can be summarized as a matrix $\boldsymbol{\Sigma}_{n\times n}$, as following:

$$\boldsymbol{\Sigma}_{kk} = \begin{bmatrix} \sigma_{k1}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{k2}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{kn}^2 \end{bmatrix}_{n\times n} \tag{2.9}$$

Estimating Parametric Path Analysis with Ordinary Least Square

OLS method with the principle of minimizing ($\underset{\sim}{\varepsilon}^T\underset{\sim}{\varepsilon}$) with the following solutions:

$$\underset{\sim}{\varepsilon} = \underset{\sim}{y} - \mathbf{X}\underset{\sim}{\beta} \tag{2.10}$$

$$min\{Q\} = min\{\underset{\sim}{\varepsilon}^T\underset{\sim}{\varepsilon}\} = min\{(\underset{\sim}{y} - \boldsymbol{X}\underset{\sim}{\beta})^T(\underset{\sim}{y} - \boldsymbol{X}\underset{\sim}{\beta})\} \tag{2.11}$$

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

$$= \left(\underset{\sim}{y}^T\underset{\sim}{y} - 2\underset{\sim}{\beta}^T\boldsymbol{X}^T\underset{\sim}{y} + \underset{\sim}{\beta}^T\boldsymbol{X}^T\boldsymbol{X}\underset{\sim}{\beta}\right) \tag{2.12}$$

Optimization is carried out on Q and then derived partially with respect to $\underset{\sim}{\beta}$ and equated to zero.

$$\frac{\partial(Q)}{\partial(\underset{\sim}{\beta})} = \underset{\sim}{0}$$

$$-2\mathbf{X}^{\mathrm{T}}\underset{\sim}{y} + 2\mathbf{X}^{\mathrm{T}}\mathbf{X}\hat{\underset{\sim}{\beta}} = \underset{\sim}{0}$$

$$-\mathbf{X}^{\mathrm{T}}\underset{\sim}{y} + \mathbf{X}^{\mathrm{T}}\mathbf{X}\hat{\underset{\sim}{\beta}} = \underset{\sim}{0}$$

$$\mathbf{X}^{\mathrm{T}}\mathbf{X}\hat{\underset{\sim}{\beta}} = \mathbf{X}^{\mathrm{T}}\underset{\sim}{y}$$

$$\hat{\underset{\sim}{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}^T\underset{\sim}{y} \qquad (2.13)$$

Estimating Parametric Path Analysis with Weighted Least Square

Weighted Least Square (WLS) optimization is able to accommodate the correlation between errors using weighted (weighting) in the form of an inverse of the error covariance variance matrix (Hutahayan *et al*, 2019). Then the equation is given as follows:

$$min\{Q\} = min\{\underset{\sim}{\varepsilon}^T\boldsymbol{\Sigma}^{-1}\underset{\sim}{\varepsilon}\} = min\left\{(\underset{\sim}{y} - \boldsymbol{X}\underset{\sim}{\beta})^T\boldsymbol{\Sigma}^{-1}(\underset{\sim}{y} - \boldsymbol{X}\underset{\sim}{\beta})\right\} \qquad (2.14)$$

Partial derivatives are performed, as follows:

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

$$= \underset{\sim}{y}^T\boldsymbol{\Sigma}^{-1}\underset{\sim}{y} - 2\underset{\sim}{\beta}^T\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\underset{\sim}{y} + \underset{\sim}{\beta}^T\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x}\underset{\sim}{\beta}$$

After that it was demoted to $\underset{\sim}{\beta}$:

$$\frac{\partial Q(\underset{\sim}{\beta})}{\partial\underset{\sim}{\beta}} = \underset{\sim}{y}^T\boldsymbol{\Sigma}^{-1}\underset{\sim}{y} + -2\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\underset{\sim}{y} + 2\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\underset{\sim}{\beta}$$

$$\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\underset{\sim}{y} = \boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\hat{\underset{\sim}{\beta}}$$

Then the estimator will be obtained $\underset{\sim}{\beta}$ as follows:

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

Linear Assumption Test

Linearity test aims to determine the linear relationship between two or more variables. One of the linearity tests is the Regression Specification Error Test (RESET) method. The RESET approach utilizes OLS to minimize the number of errors squared from each observation (Gujarati, 2004).

Regression Equation 1

$$y_{ji} = \sum_{\ell=1}^{p}\beta_{1,j\,\ell}x_{\ell i} + \sum_{s=1}^{j-1}\beta_{2,js}y_{si} + \varepsilon_{ji} \qquad (2.19)$$

$j = 1,2,\dots,q; i = 1,2,\dots,n$

Regression Equation 2

$$y_{ji} = \sum_{\ell=1}^{p}\beta_{1,j\,\ell}x_{\ell i} + \sum_{s=1}^{k-1}\beta_{2,js}y_{si} + \beta_{3,j}\hat{y}_{ji}^2 + \beta_{4,j}\hat{y}_{ji}^3 + \varepsilon_{ji}; \qquad (2.20)$$

$j = 1,2,\dots,q; i = 1,2,\dots,n$

The hypothesis used is as follows:

$$H_0 : \beta_{3,1} = \beta_{3,2} = \ldots = \beta_{3,q} = \beta_{4,1} = \beta_{4,2} = \ldots = \beta_{4,q} = 0 \qquad \text{(linear model)}$$

H_0 : at least $\beta_j \neq 0; j = 3; l = 1, 2, ..., q$ (non linear model)

Estimating parameters using OLS and then counting R_1^2 from the equation (2.19), and R_2^2 from the equation (2.20).

The test statistics used are as follows:

$$F_{hit} = \frac{\frac{R_2^2 - R_1^2}{2}}{\frac{1-R_2^2}{nq-2q}} \sim F_{(2,nq-2q)} \tag{2.21}$$

Decision criteria: if $F_{hit} > F_{(2,nq-2q)}$, then decline H_0 in other words the linearity assumption is not fulfilled.

Nonparametric Path Analysis

Modeling the relationship between exogenous variables and endogenous variables with conditions not meeting linearity assumptions and nonlinear shape patterns of unknown relationships will be approximated by nonparametric regression analysis in cases is nonparametric path analysis (Fernandes, et al. 2017). Development of nonparametric path analysis as follows:

$y_{1i} = f_{1.1}(x_i) + \varepsilon_{1i};$

$y_{2i} = f_{2.1}(x_i) + f_{2.2}(y_{1i}) + \varepsilon_{2i}; i = 1,2,\ldots,n$

$y_{3i} = f_{3.1}(x_i) + f_{3.2}(y_{1i}) + f_{3.3}(y_{2i}) + \varepsilon_{2i}; i = 1,2,\ldots,n$

The matrix notation of the above equation is as follows:

$$y_{ji} = f_k(x_i) + \varepsilon_{ki}; \quad k = 1,2 \quad i = 1,2,\ldots,n$$

Where is:

x_i : *i*-th exogenous

y_{ji} : *i*-th endogenous of i-th observations

f_k : regression function in the *j*-th endogenous

ε_{ji} : the endogenous error *j*-th of *i*-th observations

Truncated Spline Estimates in Nonparametric Path Analysis

Schumaker (2007) describes the spline function based on the order polynomial form *m*. Estimation of the f regression curve can be used to identify the shape of the curve of regression analysis. Nonparametric regression approach is used when the f regression curve is not / not known shape, so that the form of estimating the f regression curve is determined based on the existing data (Eubank, 1999).

If the f regression curve approaches the spline truncated function with the knots points $k1, k2, \ldots, kk$, presented in the form of:

$$f(x_i) = \alpha_0 + \sum_{h=1}^{m} \alpha_h x_i^h + \sum_{k=1}^{K} \beta_k (x_i - k_k)_+^m \tag{2.22}$$

The linear spline is truncated with m degrees of polynomial and *k* = *k* or three points, namely knots $k1, k2, \ldots, kk$ then the function of the spline ($f(x_i)$ as follows:

Where is:

$$(x_i - k_k)_+^m \begin{cases} \beta_1 x_i, & x_1 < k_1 \\ \beta_1 x_i + \beta_2(x_i - k_1), & k_1 < x_i < k_2 \\ \beta_1 x_i + \beta_2(x_i - k_1) + \beta_3(x_i - k_2), & k_2 < x_i < k_3 \\ \vdots & \vdots \\ \beta_1 x_i + \beta_2(x_i - k_1) + \beta_3(x_i - k_2) + \cdots + \beta_k(x_i - k_k)^m, & x_i \geq k_k \\ 0, & x_i < k_k \end{cases}$$

the estimation of the regression curve $\underset{\sim}{f}$ obtained from completing optimization:

$$\underset{\underset{\sim}{\theta}^* \in R^{m+\gamma+1}}{\mathrm{Min}} ([\underset{\sim}{y} - B(x_i, k_1, k_2, \ldots, k_k)\underset{\sim}{\theta}^*]^{\mathrm{T}}[\underset{\sim}{y} - B(x_i, k_1, k_2, \ldots, k_k)\underset{\sim}{\theta}^*]) \quad (2.23)$$

The knots are substituted:

$$\underset{\sim}{y} = \mathbf{T}(\underset{\sim}{k})\underset{\sim}{\gamma} + \underset{\sim}{\varepsilon} \quad (2.24)$$

Estimation of the truncated spline regression model obtained using the OLS method, based on the following equation:

$$\underset{\sim}{\varepsilon} = \underset{\sim}{y} - \mathbf{T}(\underset{\sim}{k})\underset{\sim}{\gamma}$$

$$\sum_{i=1}^{n} {\varepsilon_i}^2 = \underset{\sim}{\varepsilon}^T \underset{\sim}{\varepsilon} \quad (2.25)$$

it can be written as

$$\underset{\sim}{\varepsilon}^T \underset{\sim}{\varepsilon} = (\underset{\sim}{y} - \mathbf{T}(\underset{\sim}{k})\underset{\sim}{\gamma})^T(\underset{\sim}{y} - \mathbf{T}(\underset{\sim}{k})\underset{\sim}{\gamma}) \quad (2.25)$$

So the nonparametric regression curve estimator is obtained:

$$\hat{\underset{\sim}{\gamma}} = \left(\mathbf{T}'(\underset{\sim}{k})\mathbf{T}(\underset{\sim}{k})\right)^{-1}\mathbf{T}'(\underset{\sim}{k})\underset{\sim}{y} \quad (2.32)$$

$$\hat{\underset{\sim}{\gamma}} = \mathbf{A}((\underset{\sim}{k})\underset{\sim}{y}) \quad (2.33)$$

Optimal Knots Point Selection

To get the optimum spline estimator the Least Square optimization approach is used by using a family function that contains knots (Budiantara et al. 2006).

$$GCV(\underset{\sim}{k}) = min\left[\frac{n^{-1}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{[n^{-1}tr(\boldsymbol{I} - \boldsymbol{H}(\underset{\sim}{k}))]^2}\right]$$

$n^{-1}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ is *mean square error* and $\mathbf{I}$ is identity matrix, $\mathbf{A} = \left(\mathbf{T}'(\underset{\sim}{k})\mathbf{T}(\underset{\sim}{k})\right)^{-1}\mathbf{T}'(\underset{\sim}{k})\underset{\sim}{y}$ obtained from $\hat{\underset{\sim}{\gamma}} = \mathbf{A}((\underset{\sim}{k})\underset{\sim}{y})$.

Hypothesis Test Formulation

The **X** matrix is an exogenous matrix containing m polynomials as many knots as k points, or written as follows:

$$\mathbf{X} = \begin{bmatrix} 1 & {x_{11}}^m & (x_{11} - k_{11})_+^m & (x_{11} - k_{21})_+^m & \cdots & (x_{11} - k_{k1})_+^m \\ 1 & {x_{21}}^m & (x_{21} - k_{11})_+^m & (x_{21} - k_{21})_+^m & \cdots & (x_{21} - k_{k1})_+^m \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & {x_{n1}}^m & (x_{n1} - k_{11})_+^m & (x_{n1} - k_{21})_+^m & \cdots & (x_{21} - k_{k1})_+^m \end{bmatrix}_{k\times(m+r)p}$$

$$\underset{\sim}{\gamma} = \begin{bmatrix} \beta_0 \\ \beta_{11} \\ \beta_{m1} \\ \beta_{(m+1)1} \\ \vdots \\ \beta_{(m+k)1} \end{bmatrix}_{(m+k)\times 1} ; \qquad \underset{\sim}{\varepsilon} = \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{21} \\ \varepsilon_{31} \\ \varepsilon_{41} \\ \vdots \\ \varepsilon_{n1} \end{bmatrix}_{(n\times 1)}$$

Hypothesis formulation is used to test the significance of the function of the truncated spline nonparametric regression model simultaneously as follows:

H_0 : $\underset{\sim}{\gamma} = 0$, vs

H_1 : at least $\underset{\sim}{\gamma} \neq 0$ (2.38)

Determine the parameter space in the $H(\Omega)$ space as follows:

$$\Omega = \{\gamma = (\beta_0, \beta_{m1}, \beta_{(m+1)1}, \dots, \beta_{(m+k)1}), \sigma_\Omega^2\} \tag{2.39}$$

Determine the parameter space in the $H(\omega)$ space as follows::

$$\omega = \left\{\gamma = (\beta_0, \beta_{m1}, \beta_{(m+1)1}, \dots, \beta_{(m+k)1}), \sigma_\omega^2 | \underset{\sim}{\gamma}_\omega = 0\right\} \tag{2.40}$$

Decision Criteria

Simultaneous hypothesis testing on the model with the truncated spline approach will be examined, if given the equation:

$$j = 1, 2, 3$$

$$f_{2i} = \beta_{20} + \beta_{21}x_{1i} + \delta_{21}(x_{1i} - k_{11})_+ + \delta_{22}(x_{1i} - k_{12})_+ + \delta_{23}(x_{1i} - k_{13})_+ + \beta_{22}y_{1i}$$
$$+\gamma_{21}(y_{1i} - k_{21})_+ + \gamma_{22}(y_{1i} - k_{22})_+ + \gamma_{23}(y_{1i} - k_{23})_+$$

$$f_{3i} = \beta_{30} + \beta_{31}x_{1i} + \delta_{31}(x_{1i} - k_{11})_+ + \delta_{32}(x_{1i} - k_{12})_+ + \delta_{33}(x_{1i} - k_{13})_+ + \beta_{32}y_{1i}$$
$$+\gamma_{31}(y_{1i} - k_{21})_+ + \gamma_{32}(y_{1i} - k_{22})_+ + \gamma_{33}(y_{1i} - k_{23})_+ + \beta_{33}y_{2i} + \lambda_{31}(y_{2i} - k_{31})_+$$
$$+\lambda_{32}(y_{2i} - k_{32})_+ + \lambda_{33}(y_{2i} - k_{33})_+$$

$$f_{ji} = \beta_{j0} + \sum_{h=1}^{m} \beta_h x_i^h + \sum_{k=1}^{K} \delta_k (y_i - k_k)_+^m + \sum_{l=1}^{L} \lambda_l (y_i - k_k)_+^m \tag{2.41}$$

With the following hypothesis:

H_0 : $\beta_0 = \cdots = \beta_h = \beta_{m1} = \delta_{(m+1)1} = \cdots = \delta_{(m+k)1} = \lambda_{(m+1)1} = \cdots = \lambda_{(m+l)1} = 0$, vs

H_1 : at least $\beta_0 \neq 0$ or $\beta_h \neq 0$ or $\delta_{m1} \neq 0$ or $\delta_{(m+k)1} \neq 0$ or $\lambda_{m1} \neq 0$ or $\lambda_{(m+l)1} \neq 0$

where are exogenous variables $h = 1,2, \dots, p$ and knots $k = 1,2, \dots, K$

Furthermore, F test statistics are obtained as follows

$$F_{hitung} = \frac{\dfrac{SSR}{1 + (m + K)h}}{\dfrac{SSE}{n - (1 + (m + K) \times h)}}$$

The area of rejection of the hypothesis based on the F test statistic is H_0 rejected if the F test statistic is obtained as follows $F_{hitung} > F_{\alpha,(1+(m+K)\times h),(n-(1+(m+K)\times h)}$ which states that there is one parameter that is not equal to zero or at least one exogenous variable that has a significant effect on endogenous variables.

Principle Component Analysis

According to Solimun (2010) PCA is used to get data from latent variables (unobservable variables). Formative indicators that are appropriate for use in PCAs or models with fomative indicators require common factors, but consistently correlations between indicators are not required.

PCA has the term total diversity meaning that the component provides information.

$$
\begin{aligned}
PC_1 &= a_{11}X_1 + a_{21}X_2 + \cdots + a_{p1}X_p + \varepsilon_i \\
PC_2 &= a_{12}X_1 + a_{22}X_2 + \cdots + a_{p2}X_p + \varepsilon_i \\
&\vdots \\
PC_p &= a_{12}X_1 + a_{22}X_2 + \cdots + a_{pk}X_p + \varepsilon_i
\end{aligned} \tag{2.42}
$$

Path Diagram in the case of Land Rehabilitation Technique in South Kalimantan

Damage due to mining has increased especially since the development of unconventional mining. The impact of coal mining activities, both conventional and unconventional mining on the physical environment in the form of an increase in critical land due to reduced forest, damage to agricultural land and gardens. Through the study of several variables based on previous research, the following is the relationship between the variables supporting the post-mining land revegetation technique.

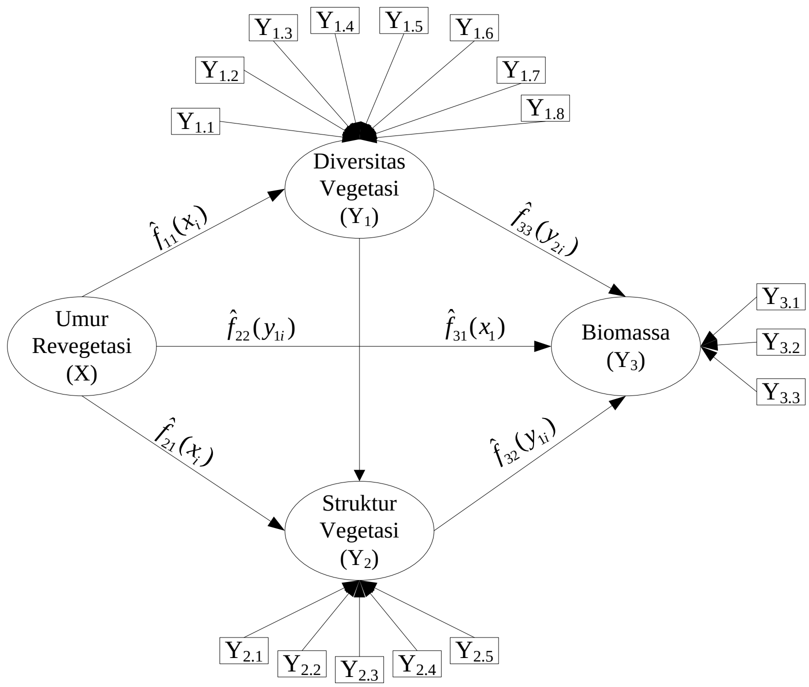

Figure 2.2. Research Model on the Case of Land Reforming Techniques in South Kalimantan.

The mathematical model formed by the truncated spline function approach at polynomial degree *m* and the *k*-knot point, then the function estimator form is obtained as follows:

$$f_{11}(y_{1i}) = \alpha_{01} + \alpha_{11}x_i + \alpha_{21}x_{21}{}^2 + \alpha_{31}x_i{}^3 + \cdots + \alpha_{p1}x_i{}^m + \beta_{11}(x_i - k_{11})_+^m + \beta_{21}(x_i - k_{21})_+^m + \cdots + \beta_{m1}(x_i - k_{ki})_+^m \tag{2.43}$$

$$f_{21}(y_{2i}) = \alpha_{02} + \alpha_{12}x_i + \alpha_{23}x_{22}{}^2 + \alpha_{32}x_i{}^3 + \cdots + \alpha_{p2}x_i{}^m + \beta_{12}(x_i - k_{12})_+^m + \beta_{22}(x_i - k_{22})_+^m + \cdots + \beta_{m2}(x_i - k_{k2})_+^m \tag{2.44}$$

$$f_{31}(y_{3i}) = \alpha_{03} + \alpha_{13}x_i + \alpha_{23}x_2{}^2 + \alpha_{33}x_i{}^3 + \cdots + \alpha_{p3}x_i{}^m + \beta_{13}(x_i - k_{13})_+^m + \beta_{23}(x_i - k_{23})_+^m + \cdots + \beta_{m3}(x_i - k_{k3})_+^m \tag{2.45}$$

$$f_{22}(y_{2i}) = \alpha_{04} + \alpha_{12}y_{1i} + \alpha_{21}y_{1i}{}^2 + \alpha_{31}y_{1i}{}^3 + \cdots + \alpha_{p1}y_{1i}{}^m + \beta_{12}(x_i - k_{12})_+^m + \beta_{22}(y_{1i} - k_{22})_+^m + \cdots + \beta_{m2}(y_{1i} - k_{k2})_+^m \tag{2.46}$$

$$f_{32}(y_{3i}) = \alpha_{05} + \alpha_{23}y_{1i} + \alpha_{21}y_{11}{}^{2} + \alpha_{31}x_i{}^{3} + \cdots + \alpha_{p1}x_i{}^{m} + \beta_{12}((x_{1i} - k_{12})_+^m + \beta_{22}((x_{1i} - k_{22})_+^m + \cdots + \beta_{m2}(x_{1i} - k_{k2})_+^m \quad (2.47)$$

$$f_{33}(y_{3i}) = \alpha_{06} + \alpha_{1}y_{2i} + \alpha_{21}y_{21}{}^{2} + \alpha_{31}y_{2i}{}^{3} + \cdots + \alpha_{p1}y_{2i}{}^{m} + \beta_{13}(y_{2i} - k_{13})_+^m + \beta_{23}(y_{2i} - k_{23})_+^m + \cdots + \beta_{m3}(y_{2i} - k_{k3})_+^m \quad (2.48)$$

III. METHODOLOGY

Data Sources

The data used in this study are secondary data which are primary data from Anshari, et al. (2018) on evaluating the success of several revegetation techniques that aim to improve biomass and soil quality on post-mining land in South Kalimantan. The study was conducted in October 2016 to December 2017 in the Laboratory of Animal Ecology and Diversity, Department of Biology, Faculty of Mathematics and Sciences, Brawijaya University, Indonesia.

Stages of Analysis

The stages of the research carried out in this study are:

1. Optimize equations containing functions $\underset{\sim}{f}$ that meet OLS optimization requirements. It is known that the spline function is truncated

$$\underset{f_\ell \in W_2{}^m[a_\ell, b_\ell]}{\text{Min}} \left[N^{-1} \sum_{i=1}^{N} \left(y_i - \sum_{\ell=1}^{p} f_\ell(x_{\ell i}) \right)^2 \right]$$

2. Estimates $\underset{\sim}{f}$, $\underset{\sim}{\hat{\Sigma}}$ and knots and place knots simultaneously or concurrently. The optimal knot point will meet the GCV criteria..
3. Formulate a hypothesis test for parameters in the truncated spline function equation:

   $H_0: \beta_0 = \beta_{m1} = \beta_{(m+1)1} = \ldots = \beta_{(m+k)1} = 0$, vs

   H_1: at least $\beta_{(m+k)1} \neq 0$
4. Get the estimation of the regression curve

$$\underset{\underset{\sim}{\theta}^* \in R^{m+\gamma+1}}{\text{Min}} \left(\left[\underset{\sim}{y} - B(x_i, k_1, k_2, \ldots, k_k)\underset{\sim}{\theta}^*\right]^{\mathrm{T}} \left[\underset{\sim}{y} - B(x_i, k_1, k_2, \ldots, k_k)\underset{\sim}{\theta}^*\right] \right)$$

5. Testing the significance of parameters simultaneously or simultaneously. Modeling a nonparametric path analysis using the truncated spline approach in the case of evaluating the success of revegetation techniques in increasing biomass and soil quality in post-mining coal fields in South Kalimantan.

IV. Results and Discussion

Model of nonparametric truncated spline path of linear polynomial degree.

The results of processing the linear model with the optimum knots point for each equation with GCV coefficient values are presented in Table 1.

Table 1. Optimum knot points for linear models

Linear	**1 *Knot***				
	k_1	**R^2**	**GCV**	**R^2_{total}**	**GCV_{total}**
	1	0,797	1825,753	0,801	1800,957
	1,3	0,799	1814,134		

	110			0,801	1800,957		
	2 Knot						
	k_1	**k_2**		**R^2**	**GCV**	**R^2_{total}**	**GCV_{total}**
	1	3,7		0,807	1769,301	0,810	1747,563
	1,84	2,32		0,810	1750,206		
	123	210		0,810	1747,563		
	3 *Knot*						
	k_1	**k_2**	**k_3**	**R^2**	**GCV**	**R^2_{total}**	**GCV_{total}**
	1	2,2	3,5	0,809	1751,664	0,811	**1738,303**
	0,95	1,73	2,32	0,810	1744,270		
	100	198	357	0,811	1738,303		

Nonparametric path model of truncated spline degrees of quadratic polynomials.

The results of the quadratic model processing with the optimum knot points for each equation with GCV coefficient values are presented in Table 2.

Table 2 Optimal knot points for quadratic models

Quadratic	**1 *Knot***						
	k_1			**R^2**	**GCV**	**R^2_{total}**	**GCV_{total}**
	1,6			0,796	1866,144	0,799	1837,515
	1,96			0,796	1866,071		
	141			0,799	1837,515		
	2 Knot						
	k_1	**k_2**		**R^2**	**GCV**	**R^2_{total}**	**GCV_{total}**
	1,1	3		0,800	1838,184	0,802	1831,839
	1,01	1,68		0,801	1834,807		
	134	343		0,802	1831,839		
	3 *Knot*						
	k_1	**k_2**	**k_3**	**R^2**	**GCV**	**R^2_{total}**	**GCV_{total}**
	1,1	2,2	3,1	0,803	1835,118	0,782	2007,732
	0,95	2,06	2,51	0,803	1833,746		
	110	234	332	0,782	2007,732		

Model of nonparametric path truncated spline degree cubic polynomial.

The results of processing cubic models with optimum knots for each equation with GCV coefficient values are presented in Table 3.

Table 3. Optimum knots for cubic models

Cubic	**1 *Knot***				
	k_1	**R^2**	**GCV**	**R^2_{total}**	**GCV_{total}**
	1,4	0,768	209,018	0,780	2011,084

	1,38			0,769	2093,269		
	160			0,780	2011,084		
	2 Knot						
	k1	**k2**		R^2	**GCV**	R^2_{total}	**GCV$_{total}$**
	1,6	3,9		0,720	2427,670	0,781	2010,080
	1,68	2,26		0,721	2425,902		
	192	357		0,781	2010,080		
	3 *Knot*						
	k1	**k2**	**k3**	R^2	**GCV**	R^2_{total}	**GCV$_{total}$**
	1,7	2,6	3,9	0,749	2242,982	0,760	2168,574
	0,95	1,6	2,26	0,750	2241,402		
	100	192	278	0,760	2168,574		

The minimum GCV value is in the linear polynomial path truncated spline analysis model with three knots that is equal to 1738,303.

Best Model.

Based on the estimation of the function using the WLS method, the best nonparametric path model that has a minimum GCV value is the degree of linear polynomial with three knots that have an R^2 value of 81,1 percent.

$$\hat{f}_{11} = 1,488 + 0,453x_{1i} - 0,542(x_{1i} - 1)_+ + 0,219(x_{1i} - 2,2)_+ - 0,296(x_{1i} - 3,5)_+$$

$$\hat{f}_{21} = 430,930 - 207,733x_{1i} + 268,879(x_{1i} - 1)_+ + 5,171(x_{1i} - 2,2)_+ - 139,704(x_{1i} - 3,5)_+$$

$$-80,059y_{1i} + 33,143(y_{1i} - 0,95)_+ + 204,356(y_{1i} - 1,73)_+ - 297,281(y_{1i} - 2,32)_+$$

$$\hat{f}_{31} = 1041,166 - 469,992x_{1i} + 491,529(x_{1i} - 1)_+ + 106,438(x_{1i} - 2,2)_+ - 149,318(x_{1i} - 3,5)_+$$

$$-235,179y_{1i} - 88,969(y_{1i} - 0,95)_+ + 1156,890(y_{1i} - 1,73)_+ - 1166,301(y_{1i} - 2,32)_+$$

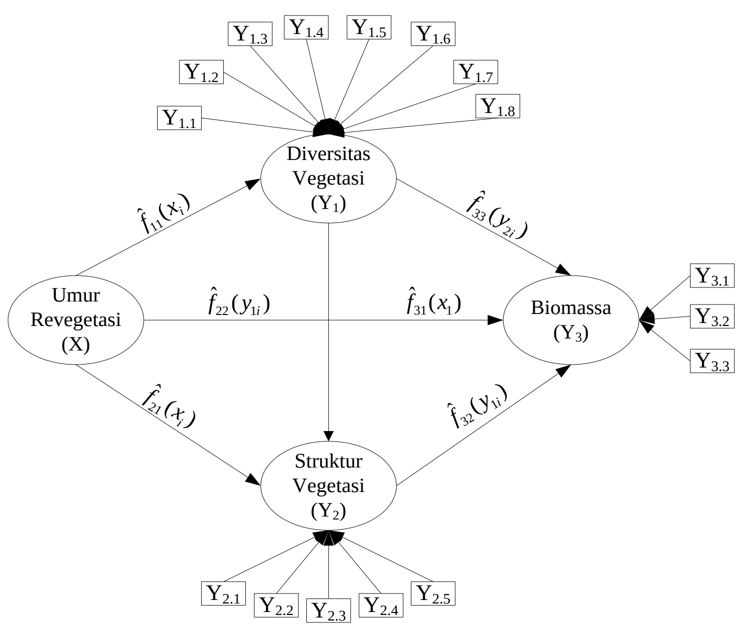

Figure 4.1 Path analysis diagram

Hypotesis Testing

Testing the parameter linear hypothesis in the best model of path truncated spline analysis is used to determine whether there is an influence of exogenous variables on endogenous variables.

1. Hypothesis Testing X1 against Y1

$$f_{1i} = \beta_{10} + \beta_{11}x_{1i} + \delta_{11}(x_{1i}-1)_+ + \delta_{12}(x_{1i}-2,2)_+ + \delta_{13}(x_{1i}-3,5)_+$$
$$= \beta_{10} + \beta_{11}x_{1i} + \delta_{11}x_{1i} - \delta_{11} + \delta_{12}x_{1i} - 2,2\delta_{12} + \delta_{13}x_{1i} - 3,5\delta_{13}$$
$$= (\beta_{10} - \delta_{11} - 2,2\delta_{12} - 3,5\delta_{13}) + (\beta_{11} + \delta_{11} + \delta_{12} + \delta_{13})x_{1i}$$

Simultaneous hypothesis testing

$$\beta_{11} = \delta_{11} = \delta_{12} = \delta_{13} = 0$$

$$\begin{bmatrix} 0 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \beta_{10} \\ \beta_{11} \\ \delta_{11} \\ \delta_{12} \\ \delta_{13} \end{bmatrix} = 0$$

2. Hypothesis Testing X1 against Y2

$$f_{2i} = \beta_{20} + \beta_{21}x_{1i} + \delta_{21}(x_{1i}-1)_+ + \delta_{22}(x_{1i}-2,2)_+ + \delta_{23}(x_{1i}-3,5)_+ + \beta_{22}y_{1i}$$
$$+\gamma_{21}(y_{1i}-0,95)_+ + \gamma_{22}(y_{1i}-1,73)_+ + \gamma_{23}(y_{1i}-2,32)_+$$
$$f_{2i} = \beta_{20} + \beta_{21}x_{1i} + \delta_{21}x_{1i} - \delta_{21} + \delta_{22}x_{1i} - \delta_{22}2,2 + \delta_{23}x_{1i} - \delta_{23}3,5$$
$$+\beta_{22}y_{1i} + \lambda_{21}y_{1i} - 0,95\lambda_{21} + \lambda_{22}y_{1i} - 1,73\lambda_{22} + \lambda_{23}y_{1i} - 2,32\lambda_{23}$$
$$f_{2i} = (\beta_{20} - \delta_{21} - 2,2\delta_{22} - 3,5\delta_{23} - 0,95\gamma_{21} - 1,73\gamma_{22} - 2,32\gamma_{23})$$
$$+(\beta_{21} + \delta_{21} + \delta_{22} + \delta_{23})x_{1i} + (\beta_{22} + \gamma_{21} + \gamma_{22} + \gamma_{23})y_{1i}$$

Simultaneous hypothesis testing

$$H_0 : \beta_{21} = \delta_{21} = \delta_{22} = \delta_{23} = 0$$

$$\begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_{20} \\ \beta_{21} \\ \beta_{22} \\ \delta_{21} \\ \delta_{22} \\ \delta_{23} \\ \gamma_{21} \\ \gamma_{22} \\ \gamma_{23} \end{bmatrix} = [0]$$

3. Hypothesis Testing X1 against Y3

$$f_{3i} = \beta_{30} + \beta_{31}x_{1i} + \delta_{31}(x_{1i}-1)_+ + \delta_{32}(x_{1i}-2,2)_+ + \delta_{33}(x_{1i}-3,5)_+ + \beta_{32}y_{1i}$$
$$+\gamma_{31}(y_{1i}-0,95)_+ + \gamma_{32}(y_{1i}-1,73)_+ + \gamma_{33}(y_{1i}-2,32)_+ + \beta_{33}y_{2i}$$
$$+\lambda_{31}(y_{2i}-100)_+ + \lambda_{32}(y_{2i}-198)_+ + \lambda_{33}(y_{2i}-357)_+$$

$$f_{3i} = \beta_{30} + \beta_{31}x_{1i} + \delta_{31}x_{1i} - \delta_{31} + \delta_{32}x_{1i} - 2,2\delta_{32} + \delta_{33}x_{1i} - 3,5\delta_{33}$$
$$+\beta_{32}y_{1i} + \gamma_{31}y_{1i} - 0,95\gamma_{31} + \gamma_{32}y_{1i} - 1,73\gamma_{32} + \gamma_{33}y_{1i} - 2,32\gamma_{33}$$
$$+\beta_{33}y_{2i} + \lambda_{31}y_{2i} - 100\lambda_{31} + \lambda_{32}y_{2i} - 198\lambda_{32} + \lambda_{33}y_{2i} - 357\lambda_{33}$$

$$f_{3i} = (\beta_{30} - \delta_{31} - 2,2\delta_{32} - 3,5\delta_{33} - 0,95\gamma_{31} - 1,73\gamma_{32} - 2,32\gamma_{33}$$
$$-100\lambda_{31} - 198\lambda_{32} - 357\lambda_{33}) + (\beta_{31} + \delta_{31} + \delta_{32} + \delta_{33})x_{1i}$$
$$+(\beta_{32} + \gamma_{31} + \gamma_{32} + \gamma_{33})y_{1i} + (\beta_{33} + \lambda_{31} + \lambda_{32} + \lambda_{33})y_{2i}$$

Simultaneous hypothesis testing

$$H_0 : \beta_{31} = \delta_{31} = \delta_{32} = \delta_{33} = 0$$

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_{30} \\ \beta_{31} \\ \beta_{32} \\ \beta_{33} \\ \delta_{31} \\ \delta_{32} \\ \delta_{33} \\ \gamma_{31} \\ \gamma_{32} \\ \gamma_{33} \\ \lambda_{31} \\ \lambda_{32} \\ \lambda_{33} \end{bmatrix} = 0$$

4. Hypothesis Testing Y1 against Y2

$$f_{2i} = \beta_{20} + \beta_{21}x_{1i} + \delta_{21}(x_{1i}-1)_+ + \delta_{22}(x_{1i}-2,2)_+ + \delta_{23}(x_{1i}-3,5)_+ + \beta_{22}y_{1i}$$
$$+\gamma_{21}(y_{1i}-0,95)_+ + \gamma_{22}(y_{1i}-1,73)_+ + \gamma_{23}(y_{1i}-2,32)_+$$

$$f_{2i} = \beta_{20} + \beta_{21}x_{1i} + \delta_{21}x_{1i} - \delta_{21} + \delta_{22}x_{1i} - \delta_{22}2,2 + \delta_{23}x_{1i} - \delta_{23}3,5$$
$$+\beta_{22}y_{1i} + \lambda_{21}y_{1i} - 0,95\lambda_{21} + \lambda_{22}y_{1i} - 1,73\lambda_{22} + \lambda_{23}y_{1i} - 2,32\lambda_{23}$$

$$f_{2i} = (\beta_{20} - \delta_{21} - 2,2\delta_{22} - 3,5\delta_{23} - 0,95\gamma_{21} - 1,73\gamma_{22} - 2,32\gamma_{23})$$
$$+(\beta_{21} + \delta_{21} + \delta_{22} + \delta_{23})x_{1i} + (\beta_{22} + \gamma_{21} + \gamma_{22} + \gamma_{23})y_{1i}$$

Simultaneous hypothesis testing

$$H_0 : \beta_{22} = \gamma_{21} = \gamma_{22} = \gamma_{23} = 0$$

$$\begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \beta_{20} \\ \beta_{21} \\ \beta_{22} \\ \delta_{21} \\ \delta_{22} \\ \delta_{23} \\ \gamma_{21} \\ \gamma_{22} \\ \gamma_{23} \end{bmatrix} = [0]$$

5. Hypothesis Testing Y1 against Y3

$$\begin{aligned} f_{3i} = {} & \beta_{30} + \beta_{31} x_{1i} + \delta_{31}(x_{1i} - 1)_+ + \delta_{32}(x_{1i} - 2,2)_+ + \delta_{33}(x_{1i} - 3,5)_+ + \beta_{32} y_{1i} \\ & + \gamma_{31}(y_{1i} - 0,95)_+ + \gamma_{32}(y_{1i} - 1,73)_+ + \gamma_{33}(y_{1i} - 2,32)_+ + \beta_{33} y_{2i} \\ & + \lambda_{31}(y_{2i} - 100)_+ + \lambda_{32}(y_{2i} - 198)_+ + \lambda_{33}(y_{2i} - 357)_+ \end{aligned}$$

$$\begin{aligned} f_{3i} = {} & \beta_{30} + \beta_{31} x_{1i} + \delta_{31} x_{1i} - \delta_{31} + \delta_{32} x_{1i} - 2,2\delta_{32} + \delta_{33} x_{1i} - 3,5\delta_{33} \\ & + \beta_{32} y_{1i} + \gamma_{31} y_{1i} - 0,95\gamma_{31} + \gamma_{32} y_{1i} - 1,73\gamma_{32} + \gamma_{33} y_{1i} - 2,32\gamma_{33} \\ & + \beta_{33} y_{2i} + \lambda_{31} y_{2i} - 100\lambda_{31} + \lambda_{32} y_{2i} - 198\lambda_{32} + \lambda_{33} y_{2i} - 357\lambda_{33} \end{aligned}$$

$$\begin{aligned} f_{3i} = {} & (\beta_{30} - \delta_{31} - 2,2\delta_{32} - 3,5\delta_{33} - 0,95\gamma_{31} - 1,73\gamma_{32} - 2,32\gamma_{33} \\ & - 100\lambda_{31} - 198\lambda_{32} - 357\lambda_{33}) + (\beta_{31} + \delta_{31} + \delta_{32} + \delta_{33}) x_{1i} \\ & + (\beta_{32} + \gamma_{31} + \gamma_{32} + \gamma_{33}) y_{1i} + (\beta_{33} + \lambda_{31} + \lambda_{32} + \lambda_{33}) y_{2i} \end{aligned}$$

Simultaneous hypothesis testing

$$H_0 : \beta_{32} = \gamma_{31} = \gamma_{32} = \gamma_{33} = 0$$

$$\begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_{30} \\ \beta_{31} \\ \beta_{32} \\ \beta_{33} \\ \delta_{31} \\ \delta_{32} \\ \delta_{33} \\ \gamma_{31} \\ \gamma_{32} \\ \gamma_{33} \\ \lambda_{31} \\ \lambda_{32} \\ \lambda_{33} \end{bmatrix} = 0$$

6. Hypothesis testing Y2 against Y3

$$\begin{aligned} f_{3i} &= \beta_{30} + \beta_{31}x_{1i} + \delta_{31}(x_{1i} - 1)_+ + \delta_{32}(x_{1i} - 2,2)_+ + \delta_{33}(x_{1i} - 3,5)_+ + \beta_{32}y_{1i} \\ &\quad + \gamma_{31}(y_{1i} - 0,95)_+ + \gamma_{32}(y_{1i} - 1,73)_+ + \gamma_{33}(y_{1i} - 2,32)_+ + \beta_{33}y_{2i} \\ &\quad + \lambda_{31}(y_{2i} - 100)_+ + \lambda_{32}(y_{2i} - 198)_+ + \lambda_{33}(y_{2i} - 357)_+ \end{aligned}$$

$$\begin{aligned} f_{3i} &= \beta_{30} + \beta_{31}x_{1i} + \delta_{31}x_{1i} - \delta_{31} + \delta_{32}x_{1i} - 2,2\delta_{32} + \delta_{33}x_{1i} - 3,5\delta_{33} \\ &\quad + \beta_{32}y_{1i} + \gamma_{31}y_{1i} - 0,95\gamma_{31} + \gamma_{32}y_{1i} - 1,73\gamma_{32} + \gamma_{33}y_{1i} - 2,32\gamma_{33} \\ &\quad + \beta_{33}y_{2i} + \lambda_{31}y_{2i} - 100\lambda_{31} + \lambda_{32}y_{2i} - 198\lambda_{32} + \lambda_{33}y_{2i} - 357\lambda_{33} \end{aligned}$$

$$\begin{aligned} f_{3i} &= (\beta_{30} - \delta_{31} - 2,2\delta_{32} - 3,5\delta_{33} - 0,95\gamma_{31} - 1,73\gamma_{32} - 2,32\gamma_{33} \\ &\quad - 100\lambda_{31} - 198\lambda_{32} - 357\lambda_{33}) + (\beta_{31} + \delta_{31} + \delta_{32} + \delta_{33})x_{1i} \\ &\quad + (\beta_{32} + \gamma_{31} + \gamma_{32} + \gamma_{33})y_{1i} + (\beta_{33} + \lambda_{31} + \lambda_{32} + \lambda_{33})y_{2i} \end{aligned}$$

Simultaneous hypothesis testing

$$H_0 : \beta_{33} = \lambda_{31} = \lambda_{32} = \lambda_{33} = 0$$

$$\begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \beta_{30} \\ \beta_{31} \\ \beta_{32} \\ \beta_{33} \\ \delta_{31} \\ \delta_{32} \\ \delta_{33} \\ \gamma_{31} \\ \gamma_{32} \\ \gamma_{33} \\ \lambda_{31} \\ \lambda_{32} \\ \lambda_{33} \end{bmatrix} = 0$$

The test statistics used in testing the parameter linear hypothesis are as follows:

$$F = \frac{Q/p(k')}{Y' P_x{}^{\perp} Y / P(P_x{}^{\perp})}$$

$$= \frac{Q/p(k')}{JK\ sisa/(n-p)}$$

$$= \frac{Q}{S^2} \qquad (4.4)$$

Following in summary the results of hypothesis testing for the truncated spline path model of three knots linear polynomial degrees are presented in Table 4:

Table 4. Results of testing linear parameter hypothesis

Function	**F**	**F-table**	***P-value***	**Decisiom**	**Conclusion**
$X_1 \rightarrow Y_1$	4.906	2,402	**0,001**	Rejected H_0	**Significant**
$X_1 \rightarrow Y_2$	0.0002	2,402	0,999	Accepted H_0	Not Significant

$Y_1 \rightarrow Y_2$	0.0001	2,402	0,999	Accepted H_0	Not Significant
$X_1 \rightarrow Y_3$	1.51E-05	2,402	0,999	Accepted H_0	Not Significant
$Y_1 \rightarrow Y_3$	8.98E-06	2,402	0,999	Accepted H_0	Not Significant
$Y_2 \rightarrow Y_3$	0.187	2,402	0,945	Accepted H_0	Not Significant

Based on Table 4., the results of testing the hypothesis on the function estimator are significant only when the relationship X1 to Y1, while testing the hypothesis on the estimation of the relationship function that is not yet significant.

Interpretation of the Truncated Spline Nonparametric Path Model

1. Interpretation of vegetation age (X) on vegetation diversity (Y1)

Here is the second nonparametric path model when vegetation diversity is in the regime:

regime **first** $x<1$

$$\hat{f}_{1i} = 1,488 + 0,453x_i$$

regime **second** 1<x<2,2

$$\hat{f}_{1i} = 2,301 - 0,089x_i$$

regime **third** 2,2<x<3,5

$$\hat{f}_{1i} = 1,754 + 0,130x_i$$

regime **fourth** $x>3,5$

$$\hat{f}_{1i} = 2,789 - 0,166x_i$$

The graph of the relationship between vegetation age variables and vegetation diversity is presented in the form of a curve as follows:

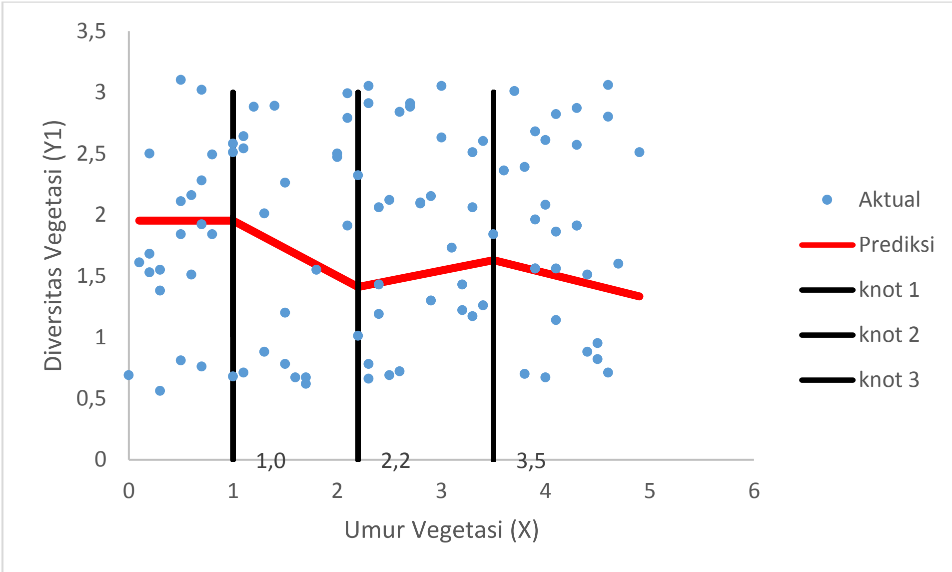

Figure 4.2 Plot relationship between vegetation age (X) and vegetation diversity (Y1)

2. Interpretation of vegetation age (X) to vegetation structure (Y2) assuming the other variables are constant.

Here is the second nonparametric path model when the vegetation structure is in the regime:

***regime* first** $x<1$

$$\hat{f}_{2i} = 430,930 - 207,733x_i$$

***regime* second** 1<x<2,2

$$\hat{f}_{2i} = 27,611 + 61,146x_{1i}$$

***regime* third** $2,2<x_1<3,5$

$$\hat{f}_{2i} = 14,684 + 66,317x_{1i}$$

***regime* fourth** $x_1>3,5$

$$\hat{f}_{2i} = 503,648 - 73,387x_{1i}$$

The graph of the relationship between vegetation age variables and vegetation structure is presented in the form of a curve as follows:

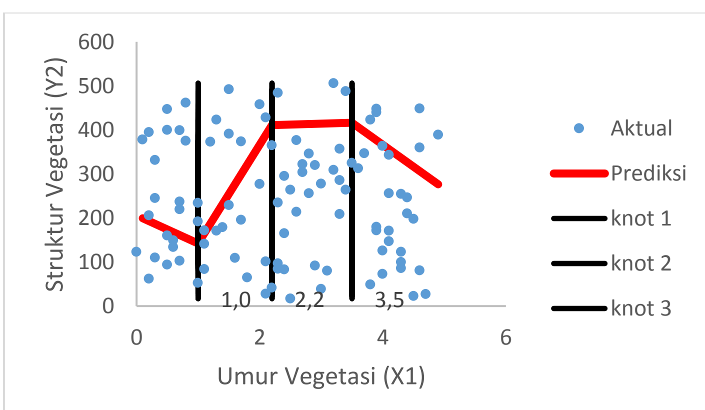

Figure 4.6 Plot relationship between vegetation age (X1) and vegetation structure (Y2)

3. Interpretation of vegetation age (X1) on post-mining land biomass (Y3) assuming the other variables are constant

Here is the second nonparametric path model when the Vegetation Structure is in the regime:

***regime* first** $x_1 < 1$

$$\hat{f}_{3i} = 1041{,}166 - 469{,}992 x_{1i}$$

***regime* second** $1 < x_1 < 2{,}2$

$$\hat{f}_{3i} = 549{,}637 + 21{,}537 x_{1i}$$

***regime* third** $2{,}2 < x_1 < 3{,}5$

$$\hat{f}_{3i} = 283{,}542 + 127{,}975 x_{1i}$$

***regime* fourth** $x_1 > 3{,}5$

$$\hat{f}_{3i} = 806{,}155 - 21{,}343 x_{1i}$$

The graph of the relationship between vegetation age variables and post-mining land biomass is presented in the form of a curve as follows:

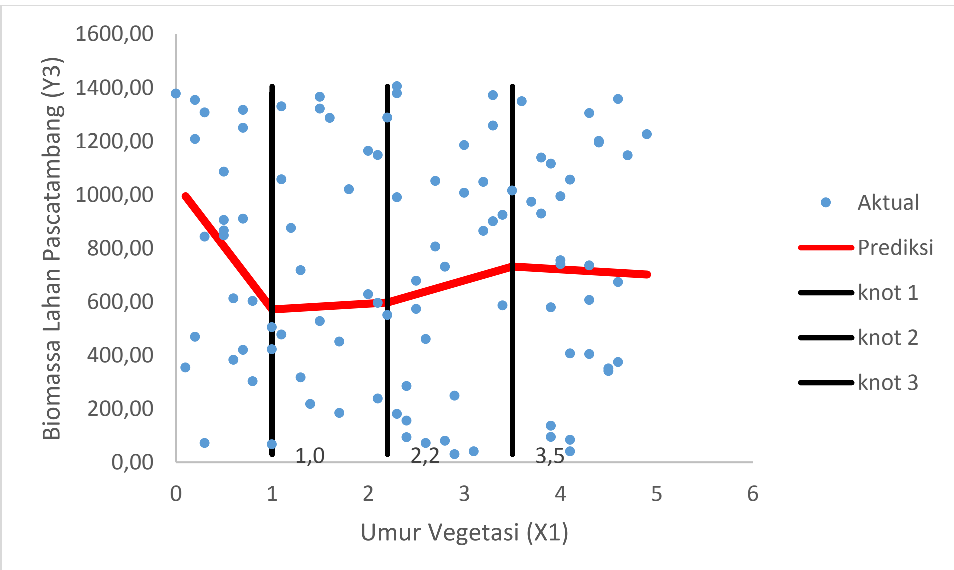

Figure 4.7 Plot relationship between vegetation age (X1) and post-latrine land biomass (Y3)

V. Conclusion

Based on the research objectives and the results and discussions that have been carried out, the conclusions can be drawn, namely:

1. Estimating the function of truncated spline nonparametric path analysis using the WLS approach obtained the best model, namely a model with linear polynomial degrees with 3 knots based on the lowest GCV value of 1738,303 with the following model:

$$\hat{f}_{11} = 1{,}488 + 0{,}453 x_{1i} - 0{,}542 (x_{1i} - 1)_+ + 0{,}219 (x_{1i} - 2{,}2)_+ - 0{,}296 (x_{1i} - 3{,}5)_+$$

$$\hat{f}_{21} = 430,930 - 207,733x_{1i} + 268,879(x_{1i} - 1)_+ + 5,171(x_{1i} - 2,2)_+ - 139,704(x_{1i} - 3,5)_+$$

$$-80,059y_{1i} + 33,143(y_{1i} - 0,95)_+ + 204,356(y_{1i} - 1,73)_+ - 297,281(y_{1i} - 2,32)_+$$

$$\hat{f}_{31} = 1041,166 - 469,992x_{1i} + 491,529(x_{1i} - 1)_+ + 106,438(x_{1i} - 2,2)_+ - 149,318(x_{1i} - 3,5)_+$$

$$-235,179y_{1i} - 88,969(y_{1i} - 0,95)_+ + 1156,890(y_{1i} - 1,73)_+ - 1166,301(y_{1i} - 2,32)_+$$

$$3,928y_{2i} - 5,145(y_{2i} - 100)_+ + 0,671(y_{2i} - 198)_+ + 1,279(y_{2i} - 357)_+$$

2. The results of testing the hypothesis on the estimator of a significant function only when the relationship X1 to Y1, while testing the hypothesis on the estimation of the function of the relationship that has not been significant.

REFERENCES

[1] Anshari, M.F., Boedianto, Fernandes, A.A.R., dan Ariesoesilaningsih, E. (2018). Hydroseeding application using pioneer local plant seeds for coal postmining soil in Tanah Laut Regency, South Kalimantan. *Journal of Degraded and Mining Lands Management*. Hal. 1335-1345.

[2] Budiantara, I.N. (2009). Spline Dalam Regresi Nonparametrik dan Semiparametrik.Sebuah Permodelan Statistika Masa Kini dan Masa Mendatang. *Pidato Pengukuhan Untuk Jabatan Guru Besar*. Institut Teknologi Sepuluh November, Surabaya.

[3] Budiantara, I.N., Suryadi, F., Otok, B. dan Guritno, S. (2006). "Pemodelan B-Spline dan MARS pada Nilai Ujian Masuk Terhadap IPK Mahasiswa Jurusan Disain Komunikasi UK Petra. Surabaya". *Jurnal Teknik Industri*, 8, 1-13.

[4] Budiantara, I.N. (2005). Regresi Spline Linier. *Makalah Pembicara Utama pada Seminar Nasional Matematika.* FMIPA. Universitas Diponogoro, Semarang.

[5] Budiantara, I.N. (2001*b*). "Estimasi Parametrik dan Nonparametrik untuk Pendekatan Kurva Regresi". *Makalah Pembicara Utama pada Seminar Nasional Statistika V*. Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh Nopember (ITS). Surabaya.

[6] Budiantara, I.N. (2000*b*)." Optimasi dan Proyeksi Dalam Regresi Nonparametrik Spline*". Majalah Berkala Matematika dan Ilmu Pengetahuan Alam (BMIPA). Universitas Gadjah Mada*, 10, 35-44.

[7] Budiantara, I.N. (2000*a*). Metode U, GML, CV dan GCV dalam Regresi Nonparametrik Spline. *Majalah Ilmiah Himpunan Matematika Indonesia* (MIHMI), 6, 285-290.

[8] Budiantara, I.N., Subanar, dan Soejoeti, Z. (1997). Weighted Spline Estimator. *Bulletin of the International Statistical Institute*, Vol. 51 No.1, hal. 333-334.

[9] Dillon, W.R. dan Goldstein, M. (1984). *Multivariate Analysis Methods and Applications*. John Willey and Sons. New Jersey.

[10] Eubank, R.L. (1999). *Nonparametrik Regression and Spline Smoothing*, Second Edition. Marcel Dekker Inc. New York.

[11] Fernandes, A.A.R.,Nyoman Budiantara, I.,Otok, B.W.,Suhartono, (2014), "Spline estimator for bi-responses nonparametric regression model for longitudinal data ", *Applied Mathematical Sciences*, Vol 8 No 114, pp 5653-5665.

[12] Fernandes, A.A.R, Solimun, dan Arisoesilaningsih, E. (2017). Estimation of spline function in nonparametrik *path* analysis based on penalized weighted least square (PWLS). *AIP Proceeding, ICWOMAA*. 1913, 020037.

[13] Fernandes, A.A.R, Budiantara, I.N., Otok, B.W. dan Suhartono. (2015). Spline Estimator for Bi-Endogenses and Multi-Prediktiors Nonparametrik Regression Model in Case of Longitudinal Data. *Journal of Mathematics and Statistics*, Vol 11, No 2, 2015, hal. 61-69.

[14] Fernandes, A.A.R, Jansen, P., Sa'adah, U, Solimun, Nurdjannah, Amaliana, L., and Efendi, A. (2018). "Comparison of Spline Estimator at Various Levels of Autocorrelation in Smoothing Spline Nonparametric Regression For Longitudinal Data", *Communications in Statistics – Theory and Methods*, Vol 46, No 24, pp 12401-12424.

[15] Fernandes, A.A.R., Hutahayan, B., Solimun, Arisoesilaningsih, E., Yanti, I., Astuti, A.B., Nurjannah, & Amaliana, L, (2019). "Comparison of Curve Estimation of the Smoothing Spline Nonparametric Function Path Based on PLS and PWLS In Various Levels of Heteroscedasticity", *IOP Conference Series: Materials Science and Engineering*, Forthcoming Issue.

[16] Gujarati, D.N. (2004). *Basic Econometric*, (4th Ed.). The McGraw-Hill Companies. New York.

[17] Hair, J.F., Black, W.C., Babin, B.J. dan Anderson, R.E. (2014). *Multivariate Data Analysis*. Prentice Hall. New Jersey.

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

[19] Hutahayan, B., Solimun, Fernandes, A.A.R., Arisoesilaningsih, E., Yanti, I., Astuti, A.B., Nurjannah, & Amaliana, L, (2019), "Mixed Second Order Indicator Model: The First Order Using Principal Component Analysis and The Second Order Using Factor Analysis", *IOP Conference Series: Materials Science and Engineering*, Forthcoming Issue.
[20] Howell, J.R. (2007). *Analysis Using Smoothing Splines As Implemented In LME In R*. Thesis. Brigham Young University.
[21] Rencher, A. dan Schaalje, G. (2007), *Linear Models in Statistics*, 2nd Edition, John Willey and Sons Inc., New Jersey.
[22] Rencher, A. (2002). *Methods of Multivariate Analysis.Second Edition*. John Wiley dan Sons.,Inc. New York.
[23] Schumaker, L. (2007). *Basic Theory: Spline Function*. Cambridge University Press. New York.
[24] Soesilaningsih, E., Indriyani, S., Retnowati, R., dan Fernandes, A.A.R. (2009). *Pemodelan Pertumbuhan Vegetatif dan Produksi Umbi Porang pada Beberapa Umur Tanaman, Kondisi Vegetasi, Tanah serta Iklim Agroforestri*. Staff Research Grant I-MHERE. Malang.
[25] Solimun. (2010). *Analisis Multivariat Pemodelan Struktural Metode Partial Least Square-PLS*. CV. Citra. Malang.
[26] Wahba, G. (1990). *Spline Models for Observational Data*. SIAM. Pensylvania.