JTAM (Jurnal Teori dan Aplikasi Matematika)
http://journal.ummat.ac.id/index.php/jtam
p-ISSN 2597-7512 | e-ISSN 2614-1175
Vol. 9, No. 1, January 2025, pp. 38-53

Comparison of Nonparametric Path Analysis and Biresponse Regression using Truncated Spline Approach

Laila Nur Azizah[1], Usriatur Rohma[1], Adji Achmad Rinaldo Fernandes[1], Ni Wayan Surya Wardhani[1], Suci Astutik[1]
[1]Department of Statistics, University of Brawijaya, Indonesia
usriaturrohma@student.ub.ac.id

Article History:
Received : 20-09-2024
Revised : 22-12-2024
Accepted : 23-12-2024
Online : 01-01-2025

Keywords:
Nonparametric Path Analysis;
Data Biresponse Regression;
Truncated Spline;
Linear.

ABSTRACT

Nonparametric path analysis and biresponse nonparametric regression are two flexible statistical approaches to analyze the relationship between variables without assuming a certain form of relationship. This study compares the performance of the two methods with the truncated spline approach, which has the advantage of determining the shape of the regression curve through optimal selection of knot points. This study aims to evaluate the best model based on linear and quadratic polynomial degree with 1, 2, and 3 knot points. The model is applied to data with 100 samples and simulated data of various sample levels. The results show that the best model in nonparametric path analysis is a quadratic model with three knots, while the best model in biresponse nonparametric regression is a quadratic model with two knots. Biresponse nonparametric regression has a coefficient of determination of 88.8% which is higher than the nonparametric path analysis of 70.9%. The best biresponse nonparametric regression model is the model with quadratic order and two knots.

https://doi.org/10.31764/jtam.v9i1.26739
This is an open access article under the CC–BY-SA license

A. INTRODUCTION

Path analysis was first developed by Wright in 1934 (Fernandes, 2016). Path analysis is used to test the relationship model between variables in the form of cause and effect (Solimun, 2002). Path analysis is a technique that can be used to determine whether there is a causal relationship between exogenous variables and endogenous variables. Path analysis is not only used to determine the direct effect of exogenous variables on endogenous variables, but also explains whether or not there is an indirect effect given from exogenous variables to endogenous variables through mediating endogenous variables.

There are six assumptions underlying path analysis, namely (1) the relationship between variables is linear and additive, (2) the residuals are normally distributed, (3) the relationship pattern between variables is recursive, (4) the minimum endogenous variable is on an interval measurement scale, (5) the research variables are measured without error and (6) the model being analyzed is specified based on relevant theories and concepts (Solimun, 2010). The assumption that can make the model change is the linearity assumption. The linearity assumption has an influence on the shape of the model. If the linearity assumption is met then the path analysis is parametric, but if the linearity assumption is not met there are 2 possibilities, nonlinear path analysis is used when the non-linear form is known, but if the nonlinear form is unknown and there is no information about the data pattern then use nonparametric path analysis. The relationship between variables can be determined using a linearity test, one of which is the Regression Specification Error Test (RESET) method.

Nonparametric regression is a regression method approach where the shape of the curve of the regression function is unknown. In nonparametric regression curves, the curve is simply assumed to be smooth. The function curve is assumed to be contained in a certain function space (Eubank, 1999). The difference between parametric and nonparametric is that in a parametric approach the data tends to be forced to follow a certain pattern, while the nonparametric approach is given the freedom to find its own regression curve pattern so that it is very flexible and objective (Hidayat et al., 2018). Nonparametric regression is the basis of nonparametric path analysis for relationship patterns between exogenous, pure endogenous, and mediated endogenous variables. There are several approaches that can be used in nonparametric path analysis, namely using Moving Average, Fourier Series, and Spline, Kernel, Local Polynomial and Wavelet (Prahutama, 2013).

Splines are used in nonparametric path analysis because they can follow the pattern of relationships between exogenous and endogenous variables and are very flexible (Eubank, 1999). According to (Hidayat et al., 2018), spline is part or pieces of polynomials that have segmented and continuous (truncated) properties. The advantage of truncated spline is that it tends to find its own form of estimating the regression curve. This can happen because the spline has a combination point that shows the pattern of data behavior called the knot point. This study uses a truncated spline by considering the existence of knot points in determining the most optimal points.

A relationship in regression analysis is not always between predictor variables and one response variable. Multi-response regression is a regression model when there is more than one response variable and one response variable has a relationship with another response variable. Multiresponse regression allows the relationship between variables to be seen through the variance matrix (Härdle & Liang, 2007). The form of path analysis is similar to birresponse regression. The similarity lies in the presence of two response variables (Y), but the relationship between response variables is different. In path analysis, the relationship between variables Y1 and Y2 is a causal relationship. Whereas in birresponse regression, there is no relationship between response variables, but between response variables are correlated, so the function estimation process uses weighted least square to accommodate the correlation between responses.

Previous research related to truncated spline nonparametric path analysis is research Efendi et al. (2021) entitled "Modelling of Path Nonparametric Truncated Spline Linear, Quadratic, and Cubic in Model on Time Paying Bank Credit". From the results of the study, it was found that the best model produced using truncated spline nonparametric path analysis was a model with a linear polynomial degree of 2 knots. The research can still be developed because it only compares 2 knot points at various polynomial degrees, besides that in hypothesis testing only uses Linear Function Parameter Hypothesis (HFLP) testing. One example of development is to compare various knot points at linear and quadratic polynomial degrees using jackknife resampling hypothesis testing.

This study aims to address the gaps in the literature by modelling data using a nonparametric path analysis with a truncated spline approach and comparing it with biresponse regression. By comparing various knot points at linear and quadratic polynomial degrees, it is expected to obtain the best truncated spline function. In addition, to see which analysis is better, a comparison of the results of nonparametric path analysis with biresponse nonparametric regression is carried out. The results of this comparison are expected to provide deeper insight into the effectiveness and accuracy of the two methods in estimating causal relationships and regression patterns in complex data. These insights will be crucial for researchers in choosing the appropriate analytical approach for their studies.

B. METHODS

1. Truncated Spline Nonparametric Regression Analysis

Nonparametric regression is used when the assumptions of parametric regression are not met, one of which is because the curve does not follow a linear, quadratic and polynomial shape. Truncated spline has the advantage of handling data patterns that show sharp changes, either in the form of increases or decreases, by using knot points, which are intersection points that show changes in data behavior patterns (Firpha & Achmad, 2022). The truncated spline nonparametric regression model is as follows.

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

where: $\hat{f}(X_i)$ is Regression function for which the shape of the pattern to be estimated is unknown; X_i is The i-th predictor variable; *i* is 1,2,3,..., *n* where n is the number of observations; *j* is 1,2,3, ..., *p;p*≥1 with p is the order of the spline regression polynomials; k is 1,2,3, ..., K where K is the number of knot points. The function $\left(X_i - K_k \right)_+^p$ is a truncated function given by:

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

2. Truncated Spline Biresponse Nonparametric Regression Analysis

Multi-response regression is an approach model where one response variable has a relationship with more than one response variable. (Fernandes & Solimun, 2021) describe the nonparametric multi-response regression model as shown in equation (2).

$$y_{ki} = \sum_{\ell=1}^{p} f_{\ell k}(x_{\ell i}) + \varepsilon_{ki}; \quad k = 1,2; \quad i = 1,2,\ldots,\text{n}; \; l = 1,2,\ldots,p \tag{2}$$

with the truncated spline function in equation (3).

$$(x_{ki}-k_i)_+ = \begin{cases} (x_{ki}-k_i)\,; x_{ki} \geq k_i \\ \quad 0 \quad\;\; ; x_{ki} < k_i \end{cases} \tag{3}$$

In this case:

y_{ki} : The k-th response variable at the i-th observation
$x_{\ell i}$: The l-th predictor variable at the i-th observation
$f_{\ell k}$: The regression function linking the l-th predictor with the k-th response
n : The number of observations
p : The number of predictor variables
ε_{ki} : The error in the k-th response at the i-th observation

If equation (2) is applied to biresponse nonparametric regression, the regression model is obtained as shown in equation (3).

$$\begin{aligned} y_{1i} &= \sum_{\ell=1}^{p} f_{\ell 1}(x_{\ell i}) + \varepsilon_{1i} \\ y_{2i} &= \sum_{\ell=1}^{p} f_{\ell 2}(x_{\ell i}) + \varepsilon_{2i} \end{aligned} \tag{3}$$

The birespon nonparametric regression model for linear order with 1 knot point is as follows. $\hat{f}_{1i} = \hat{\beta}_{01} + \hat{\beta}_{11}x_{1i} + \hat{\beta}_{21}(x_{1i} - k_{11})_+ + \hat{\beta}_{31}x_{2i} + \hat{\beta}_{41}(x_{2i} - k_{21})_+$

$$\hat{f}_{2i} = \hat{\beta}_{02} + \hat{\beta}_{12}x_{1i} + \hat{\beta}_{22}(x_{1i} - k_{12})_+ + \hat{\beta}_{32}x_{2i} + \hat{\beta}_{42}(x_{2i} - k_{22})_+ \tag{4}$$

The birespon nonparametric regression model for linear order with 2 knots is as follows.

$$\begin{aligned} \hat{f}_{1i} &= \hat{\beta}_{01} + \hat{\beta}_{11}x_{1i} + \hat{\beta}_{21}(x_{1i} - k_{11})_+ + \hat{\beta}_{31}(x_{1i} - k_{12})_+ + \hat{\beta}_{41}x_{2i} + \hat{\beta}_{51}(x_{2i} - k_{21})_+ + \hat{\beta}_{61}(x_{2i} - k_{31})_+ \\ \hat{f}_{2i} &= \hat{\beta}_{02} + \hat{\beta}_{12}x_{1i} + \hat{\beta}_{22}(x_{1i} - k_{11})_+ + \hat{\beta}_{32}(x_{1i} - k_{12})_+ + \hat{\beta}_{42}x_{2i} + \hat{\beta}_{52}(x_{2i} - k_{21})_+ + \hat{\beta}_{62}(x_{2i} - k_{31})_+ \end{aligned} \tag{5}$$

The birespon nonparametric regression model for quadratic order with 1 knot point is as follows.

$$\begin{aligned} \hat{f}_{1i} &= \hat{\beta}_{01} + \hat{\beta}_{11}x_{1i} + \hat{\beta}_{21}x_{1i}^{\;2} + \hat{\beta}_{31}(x_{1i} - k_{11})^2_{\;+} + \hat{\beta}_{41}x_{2i} + \hat{\beta}_{51}x_{2i}^{\;2} + \hat{\beta}_{61}(x_{2i} - k_{21})^2_{\;+} \\ \hat{f}_{2i} &= \hat{\beta}_{02} + \hat{\beta}_{12}x_{1i} + \hat{\beta}_{22}x_{1i}^{\;2} + \hat{\beta}_{32}(x_{11i} - k_{11})^2_{\;+} + \hat{\beta}_{42}x_{2i} + \hat{\beta}_{52}x_{2i}^{\;2} \end{aligned} \tag{6}$$

The birespon nonparametric regression model for quadratic order with 2 knots is as follows.

$$\begin{aligned}\hat{f}_{1i} &= \hat{\beta}_{01} + \hat{\beta}_{11}x_{1i} + \hat{\beta}_{21}x_{1i}{}^2 + \hat{\beta}_{31}(x_{1i} - k_{11})_+{}^2 + \hat{\beta}_{41}(x_{1i} - k_{21})_+{}^2 + \hat{\beta}_{51}x_{2i} + \hat{\beta}_{61}x_{2i}{}^2 \\ &\quad + \hat{\beta}_{71}(x_{21i} - k_{12})_+{}^2 + \hat{\beta}_{81}(x_{2i} - k_{22})_+{}^2 \\ \hat{f}_{2i} &= \hat{\beta}_{02} + \hat{\beta}_{12}x_{1i} + \hat{\beta}_{22}x_{1i}{}^2 + \hat{\beta}_{32}(x_{1i} - k_{11})_+{}^2 + \hat{\beta}_{42}(x_{11i} - k_{21})_+{}^2 + \hat{\beta}_{52}x_{2i} + \hat{\beta}_{62}x_{2i}{}^2 \\ &\quad + \hat{\beta}_{72}(x_{2i} - k_{12})_+{}^2 + \hat{\beta}_{82}(x_{2i} - k_{22})_+{}^2\end{aligned} \tag{7}$$

3. Truncated Spline Nonparametric Path Analysis

Parametric path analysis cannot overcome when the regression curve is unknown and the linearity assumption is not met. Therefore, nonparametric path analysis was developed. The general equation of nonparametric path analysis with two exogenous variables and two endogenous variables can be written as follows:

$$\begin{aligned}Y_{1i} &= f_1(X_{1i}) + f_1(X_{2i}) + \varepsilon_{1i}; \\ &= f_1(X_{1i}, X_{2i}) + \varepsilon_{1i}; i = 1, 2, \ldots, n \\ Y_{2i} &= f_2(X_{1i}) + f_2(X_{2i}) + f_2(Y_{1i}) + \varepsilon_{2i}; \\ &= f_2(X_{1i}, X_{2i}, Y_{1i}) + \varepsilon_{2i}; i = 1, 2, \ldots, n\end{aligned} \tag{8}$$

where: $f_1(X_{1i}, X_{2i})$ is nonparametric path function between exogenous variables and endogenous variables endogenous variables (Y_{1i}); and $f_2(X_{1i}, X_{2i}, Y_{1i})$ is nonparametric path function between exogenous variable (X_{1i}, X_{2i}, Y_{1i}) and endogenous variable (Y_{2i}) The quadratic moment truncated spline nonparametric path model with 1 knot point for two exogenous variables and two endogenous variables is as follows:

$$\begin{aligned}&\hat{f}_{1i} = \hat{\beta}_{10} + \hat{\beta}_{11}X_{1i} + \hat{\beta}_{12}X_{1i}^2 + \hat{\beta}_{13}(X_{1i} - K_{11})_+^2 + \hat{\beta}_{14}X_{2i} + \hat{\beta}_{15}X_{2i}^2 + \hat{\beta}_{16}(X_{2i} - K_{21})_+^2 \\ &\hat{f}_{2i} = \hat{\beta}_{20} + \hat{\beta}_{21}X_{1i} + \hat{\beta}_{22}X_{1i}^2 + \hat{\beta}_{23}(X_{1i} - K_{31})_+^2 + \hat{\beta}_{24}X_{2i} + \hat{\beta}_{25}X_{2i}^2 + \hat{\beta}_{26}(X_{2i} - K_{41})_+^2 + \hat{\beta}_{27}\hat{f}_{1i} \\ &+ \hat{\beta}_{28}\hat{f}_{1i}^2 + \hat{\beta}_{29}\left(\hat{f}_{1i} - K_{51}\right)_+^2\end{aligned} \tag{9}$$

where the truncated function:

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

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

$$(X_{1i} - K_{31})_+^2 = \begin{cases}(X_{1i} - K_{31})^2 & ; X_{1i} \geq K_{31} \\ 0 & ; X_{1i} < K_{31}\end{cases}$$

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

$$\left(\hat{f}_{1i} - K_{51}\right)_+^2 = \begin{cases}\left(\hat{f}_{1i} - K_{51}\right)^2 & ; \hat{f}_{1i} \geq K_{51} \\ 0 & ; \hat{f}_{1i} < K_{51}\end{cases}$$

4. Optimal Knot Point Selection

Wu & Zhang (2006) stated that the method used to determine the optimal knot is the Generalized Cross Validation (GCV) method. If the optimal knot point is obtained, the best spline function is obtained. The GCV formula is as follows.

$$GCV(\mathbf{K}) = \frac{MSE(\mathbf{K})}{\left[n^{-1} trace(\mathbf{I} - \mathbf{A}(\mathbf{K}))\right]^2} \tag{10}$$

Where $MSE(\mathbf{K}) = n^{-1}\sum_{i=1}^{n}(Y_i - Y_i)^2$ and $\mathbf{K}$ are knot points and $\mathbf{A}(\mathbf{K})$ matrix is obtained from:

$$\hat{\mathbf{Y}} = \mathbf{A}(\mathbf{K})\mathbf{Y}$$

$$\mathbf{A}(\mathbf{K}) = \hat{\mathbf{Y}}^{-1}\mathbf{Y}$$

$$A[\mathbf{K}] = \mathbf{X}[\mathbf{K}]\left(\mathbf{X}[\mathbf{K}]^{\mathbf{T}}\mathbf{X}[\mathbf{K}]\right)^{-1}\mathbf{X}[\mathbf{K}]^{\mathbf{T}}$$

5. Model Fit Measures

The coefficient of determination is a measure of the contribution of predictor variables to the response variable. The coefficient of determination is used to determine how much diversity can be explained by the model formed. According to (Fernandes & Solimun, 2021), the coefficient of determination formula is as follows.

$$R^2 = 1 - \frac{\sum_{k=1}^{3}\sum_{i=1}^{n}(y_{ki} - \hat{f}_{ki})^2}{\sum_{k=1}^{3}\sum_{i=1}^{n}(y_{ki} - \bar{y}_k)^2}; 0 \le R^2 \le 1 \tag{11}$$

where: R^2 is Total coefficient of determination; y_{ki} is The i-th value of the endogenous variable; $\hat{f}_{ki}$ is The i-th function estimator for the endogenous variable; $\bar{y}_k$ is Average of endogenous variables; i is 1,2, …, n with n number of observations.

6. Jackknife Resampling

A simple resampling technique has been used long before the bootsrap method was invented, namely jackknife resampling. In 1949 the jackknife method was first discovered by Quenouille which is used to estimate the bias of an estimator by removing some sample observations. The jackknife method is known as a resampling method without returns, so there is an intertwined relationship in each resampling process. The jackknife method can be used to construct the variance of an estimator (Rodliyah, 2016). According to Aidi & Saufitra (2007) the jackknife method can be divided based on the amount of data removed into jackknife. In WarpPLS 6 software, the algorithm of the jackknife resampling process is called delete one, which is done by removing one sample and repeating it on each sample until the last. Suppose there is a sample $x = (x_1, x_2, \dots, x_n)$ and $\hat{\theta} = s(x)$ is an estimate for a parameter.

7. Hypothesis Testing (Resampling)

Hypothesis testing uses t test statistics, where parameter estimates and standard errors are obtained from jackknife resampling. Hypothesis testing with t test statistics as follows.

$$\text{Test statistic } t = \frac{\hat{\beta}_j}{SE_{\hat{\beta}_j}} \sim t_{n-1} \tag{12}$$

The hypothesis used for the test statistics in formula (6) is as follows. $H_0 : \beta_j = 0$ (there is no partial influence); and $H_1 : \beta_j \neq 0$ (there is a partial influence). The test criteria, namely if the test statistic $t > t_{\alpha/2(n-1)}$ then H_0 is rejected, which means that there is a significant influence between exogenous variables on endogenous variables.

8. Research Methods and Research Model

This study uses truncated spline nonparametric path analysis as the primary analytical method. In addition, nonparametric biresponse regression is also employed to provide a comparative perspective on the results. The path model was analyzed to determine the best nonparametric path function between linear and quadratic polynomial degrees with 1, 2, and 3 knot points and then tested the best model hypothesis with the t test at the jackknife resampling stage. The software used in this research is R Studio. The data used in this study is primary data from a research grant conducted by Fernandes in 2024 with a sample size of 100 samples. The variables used in this study consisted of two exogenous variables, mediating endogenous variables, and pure endogenous variables. For the truncated spline nonparametric path analysis, the variables consist of two exogenous variables, mediating endogenous variables, and pure endogenous variables. Meanwhile, in the nonparametric biresponse regression analysis, the variables include two predictor variables and two response variables. The research model can be seen in Figure 1 and Figure 2.

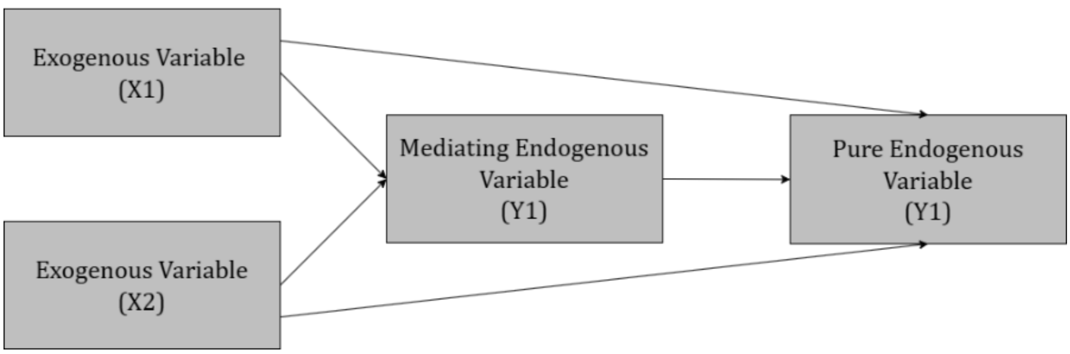

Figure 1. Nonparametric Path Analysis Research Model

If the mediation of endogenous variables does not exist, and there is a correlation between endogenous variables, then the research model looks like the following, namely using biresponse nonparametric regression.

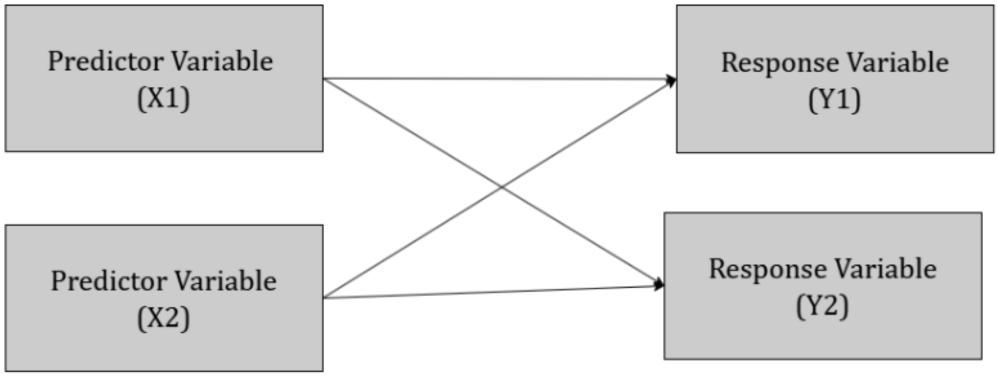

Figure 3. Nonparametric Biresponse Regression Analysis Research Model

The following is a flow chart showing the research method presented explicitly in Figure 3.

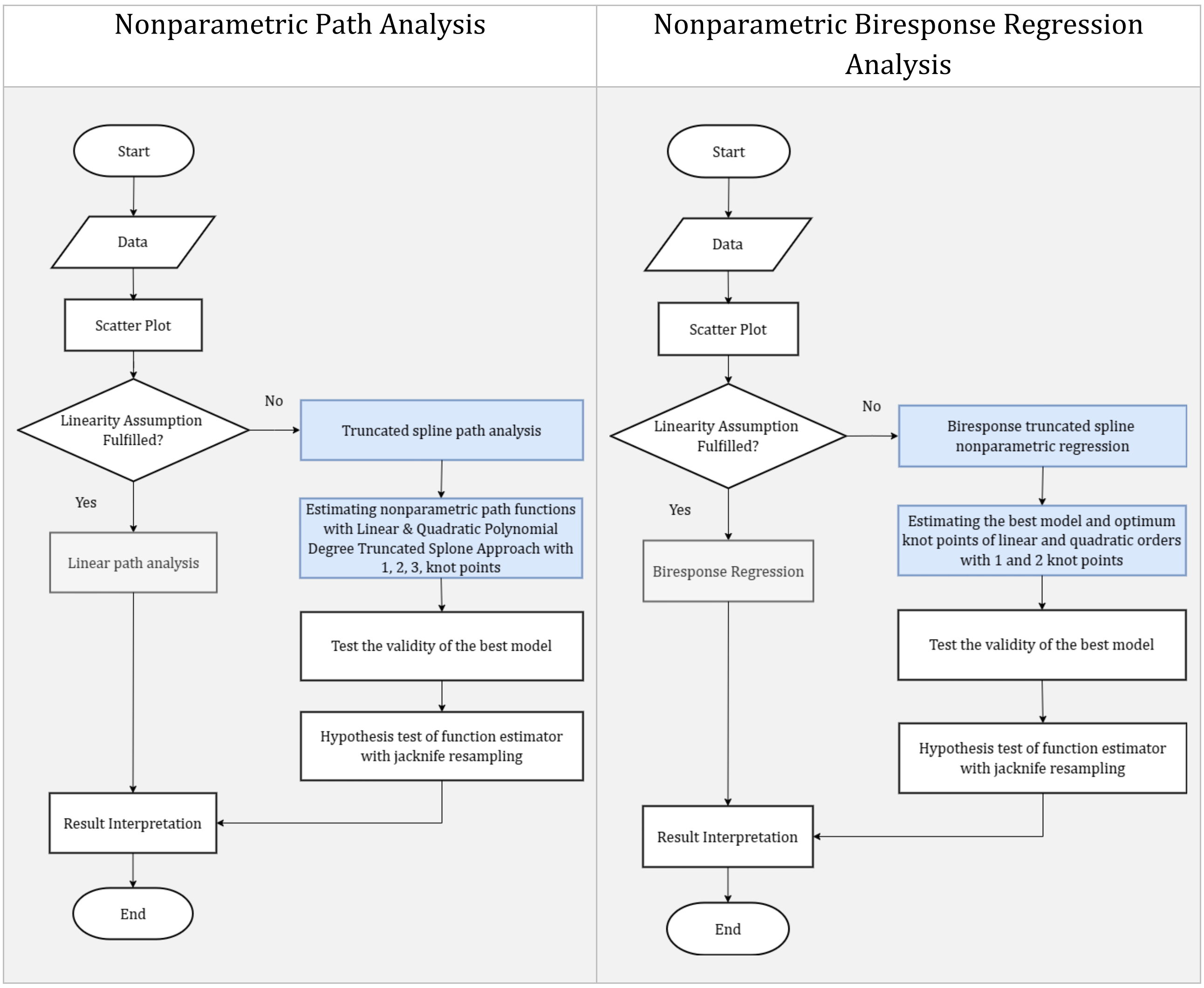

Figure 3. Flow Chart

C. RESULT AND DISCUSSION

1. Linearity Test

The linearity test aims to determine the existence of a linear relationship between more than two variables. The results of the linearity test with Ramsey's RESET are presented in Table 1.

Table 1. Linearity Assumption Test Results

Variables	P-value	Relationship
$(X_1) \rightarrow (Y_1)$	<0.001	Nonlinear
$(X_2) \rightarrow (Y_1)$	<0.001	Nonlinear
$(X_1) \rightarrow (Y_2)$	0.04	Nonlinear
$(X_2) \rightarrow (Y_2)$	0.04	Nonlinear
$(Y_1) \rightarrow (Y_2)$	0.03	Nonlinear

Based on Table 1 it can be seen that the test results with Ramsey's RESET show that the relationship between variables has a p-value <0.05 so H_0 is rejected. With a real level of 5%, it is known that the relationship between variables is not linear and the form of nonlinearity has not been found.

2. Best Model Selection

The criterion used to determine the best model is a small GCV value. The best model is the model that has the optimal knot points.

Table 2. Comparison of GCV and R^2 of each Model

Analysis	Orde	Knot Point	GCV Model	R^2
Nonparametric Truncated Spline Path Analysis	Linear	1	6.5778	0.3487
		2	6.3714	0.4308
		3	6.0898	0.5419
	Quadratic	1	6.7152	0.3826
		2	5.7995	0.5828
		3	5.2492	0.6933
Biresponse Truncated Spline Regression	Linear	1	0.3861	0.8412
		2	0.3645	0.8743
	Quadratic	1	0.3642	0.8751
		2	0.3561	0.8852

Based on Table 3, it can be seen that the nonparametric truncated spline path analysis lies in the GCV value of 5.2492 with the highest coefficient of determination of 69.33%. While in biresponse truncated spline regression, the lowest GCV value is 0.3561 with a coefficient of determination of 88.52%. This means that in nonparametric truncated spline analysis, the best model is the quadratic order model with three knot points. Meanwhile, in biresponse truncated spline regression, the best model is the quadratic model with two knot points.

3. Best Truncated Spline Nonparametric Path Model

Based on the explanation in sub chapter C.2, the best nonparametric truncated spline path model is obtained, namely the quadratic polynomial degree truncated spline nonparametric path model (order p = 2) with 3 knot points and biresponse truncated spline regression (order p = 2) with 2 knot points. Here are the optimal knot points and the best model goodness test.

Table 3. Optimal Knot Points of Truncated Spline Nonparametric Path Model

Variable Relationship	Nonparametric Truncated Spline Path Analysis		Biresponse Truncated Spline Regression	
	Knot Point Optimal	**Final GCV and R^2**	**Knot Point Optimal**	**Final GCV and R^2**
$X_1 \to Y_1$	K_{11} = 1.12	GCV = 5.2442 $R^2 = 0.7096$	K_{11} = 2.81	GCV = 0.353 $R^2 = 0.888$
	K_{12} = 1.54		K_{12} = 3.19	
	K_{13} = 1.84			
$X_2 \to Y_1$	K_{21} = 2.02		K_{21} = 2.12	
	K_{22} = 2.82		K_{22} = 3.11	
	K_{23} = 5			
$X_1 \to Y_2$	K_{31} = 1.24		K_{31} = 2.81	
	K_{32} = 1.58		K_{32} = 3.19	
	K_{33} = 2.93			
$X_2 \to Y_2$	K_{41} = 1.79		K_{41} = 2.08	
	K_{42} = 3.31		K_{42} = 3.08	
	K_{43} = 5			
$Y_1 \to Y_2$	K_{51} = 3.54			
	K_{52} = 4.51			
	K_{53} = 6.28			

Based on Table 3, for Nonparametric Truncated Spline Path Analysis, it can be seen that the GCV value is 5.2442 and the R^2 value is 0.7096. It can be interpreted that the model formed can explain the response variable by 70.9% and the rest is explained by other factors that cannot be known in the model by 29.1%. The following are the results of the best model estimator.

$$\hat{f}_{1i} = -21,84 + 63,91X_{1i} - 34,08X_{1i}^2 + 54,16(X_{1i} - 1,12)_+^2 - 29,72(X_{1i} - 1,54)_+^2 + 10,61(X_{1i} - 1,84)_+^2$$
$$+0,54X_{2i} - 0,67X_{2i}^2 + 2,98(X_{2i} - 2,02)_+^2 - 2,53(X_{2i} - 2,82)_+^2 - 0,39(X_{2i} - 5)_+^2$$
$$\hat{f}_{2i} = -183,09 + 218,23X_{1i} - 108,25X_{1i}^2 + 215,95(X_{1i} - 1,24)_+^2 - 121,63(X_{1i} - 1,58)_+^2 + 29,16(X_{1i} - 2,93)_+^2$$
$$+5,01X_{2i} - 4,72X_{2i}^2 + 9,38(X_{2i} - 1,79)_+^2 - 4,77(X_{2i} - 3,31)_+^2 - 3,16(X_{2i} - 5)_+^2 + 74,65\hat{f}_{1i} - 12,14\hat{f}_{1i}^2$$
$$+32,99\left(\hat{f}_{1i} - 3,54\right)_+^2 - 20,42\left(\hat{f}_{1i} - 4,51\right)_+^2 - 6,95\left(\hat{f}_{1i} - 6,28\right)_+^2$$

From the best model estimation results above, then form a relationship pattern between exogenous variables to endogenous variables in the research model using a quadratic polynomial degree truncated spline nonparametric path function with 3 knot points as follows.

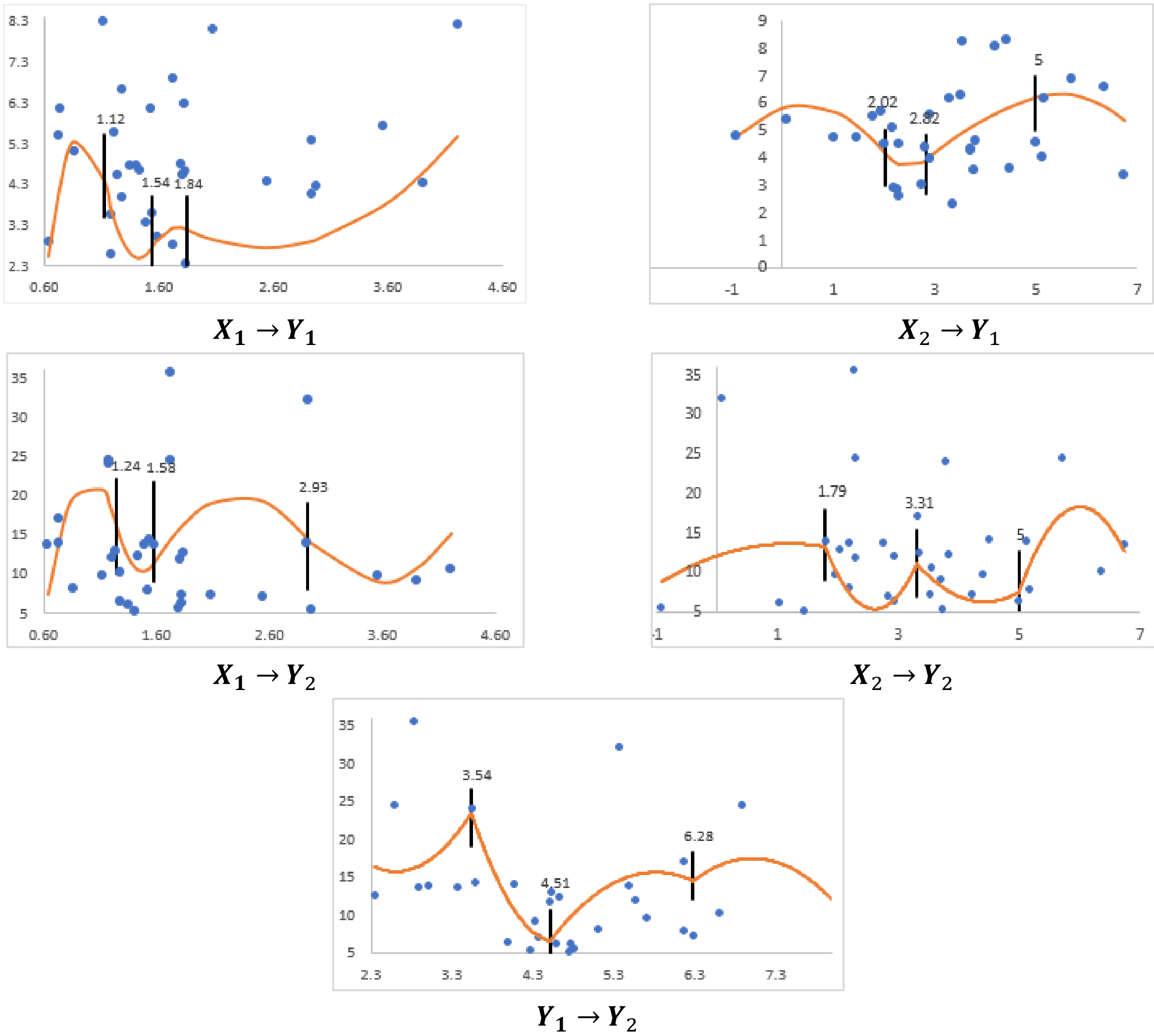

Figure 4. Relationship Pattern Curve of the Best Combined Truncated Spline Path Model Analysis

Based on Figure 4, it can be seen that the relationship graph between the X_1 and Y_1 in the quadratic polynomial degree truncated spline nonparametric path model with 3 knot points is divided into four regimes. The first regime is shown when (X_{1i}) < 1.12 where X_1 is in the very low category, the second regime when 1.12 ≤ (X_{1i}) < 1.54 or X_1 is in the low category, the third regime when 1.54 ≤ (X_{1i}) < 1 where $X_1$is in the high category, the fourth regime whe(X_{1i}) ≥1.84 or X_1 is in the very high category. The relationship between X_2 and $Y_1$is a nonparametric truncated spline of quadratic polynomial degree with 3 knots divided into four regimes. The first regime is shown when (X_{2i})< 2.02 where X_2 is in the very low category, the second regime when 2.02 ≤ (X_{2i}) < 2.82 or X_2 is in the low category, the third regime when 2.82 ≤ (X_{2i})< 5 where X_2 is in the high category, and the fourth regime when (X_{2i})≥ 5% or X_2 is in the very high category. The increase in the estimation line in regime 3 means that the higher the X_2 will increase Y_1. However, this is contradicted in regime 4, where there is a decrease in the estimation line of the Y_1 even though the X_2 is higher than in regime 3.

The relationship between X_1 and Y_2 is a nonparametric truncated spline of quadratic polynomial degree with 3 knots divided into four regimes. The first regime is shown when(X_{1i})< 1.24 where X_1 is in the very low category, the second regime when 1.24 ≤ (X_{1i}) <

1.58 or X_1 is in the low category, the third regime when 1.58 ≤ (X_{1i}) < 2.93, where $X_1$is in the high category, and the fourth regime when (X_{1i}) ≥ 2.93 or $X_1$is in the very high category. The increase in the estimation line in regime 4 which previously experienced a decrease means that the higher the X_1will increase the Y_2. However, this is contradicted in regime 1, where there is an increase in the estimation line of Y_2 even though X_1 is lower than regime 4.

The relationship between the variable X_2 and Y_2 is a nonparametric truncated spline of quadratic polynomial degree with 3 knots divided into four regimes. The first regime is shown when (X_{2i}) < 1.79 where $X_2$is in the very low category, the second regime when 1.79 ≤ (X_{2i}) < 3.31 or $X_2$is in the low category, the third regime when 3.31 ≤ (X_{2i}) < 5 where X_2 is in the high category, and the fourth regime when (X_{2i}) ≥ 5 or $X_2$is in the very high category. The increase in the estimation line of Y_2 in regime 4 means that the higher the X_2 will increase the Y_2. However, this is contradicted in regime 1, where there is an increase in the estimation line of Y_2 even though the X_2 is lower than regime 4.

The relationship between the variable Y_1 and Y_2 is a nonparametric truncated spline degree quadratic polynomial 3 knots which is divided into four regimes. The first regime is shown when (Y_{1i}) < 3.54 where the Y_1 is in the very low category, the second regime when 3.54 ≤ (Y_{1i}) < 4.51 or the Y_1 is in the low category, the third regime when 4.51% ≤ (Y_{1i}) < 6.28 where the Y_1 is in the high category, and the fourth regime when (Y_{1i}) ≥ 6,28 or the Y_1 is in the very high category. The increase in the estimation line in regime 4, then the decrease in the estimation line of Y_2 means that the higher the Y_1 has not been able to increase the Y_2 optimally. Other evidence can be seen in regime 1 where when the Y_1 is lower than in regime 4, it is more optimal in increasing the Y_2. Based on Table 4, for BiresponseTruncated Spline Regression Analysis, it can be seen that the GCV value is 0.353 and the R^2 value is 0.888. It can be interpreted that the model formed can explain the response variable by 88.9% and the rest is explained by other factors that cannot be known in the model by 11.1%. The following are the results of the best model estimator.

$$\hat{f}_{1i} = -5,034 - 2,472x_{1i} + 0,651x_{1i}^2 - 2,109(x_{1i} - 2,814)_+^2 + 3,094(x_{11i} - 3,191)_+^2 + 2,045x_{2i} - 0,214x_{2i}^2$$
$$-0,726(x_{2i} - 2,118)_+^2 + 14,435(x_{2i} - 3,118)_+^2$$
$$\hat{f}_{2i} = 7,332 - 0,013x_{1i} + 0,098x_{1i}^2 - 0,284(x_{1i} - 2,814)_+^2 - 4,022(x_{11i} - 3,191)_+^2 + 0,211x_{2i} - 0,014x_{2i}^2$$
$$-0,189(x_{2i} - 2,118)_+^2 - 20,694(x_{2i} - 3,118)_+^2$$

from the best model estimation results above, then form a relationship pattern between predictor variables to research variables in the research model using a quadratic polynomial degree truncated spline regression function with 2 knot points as follows.

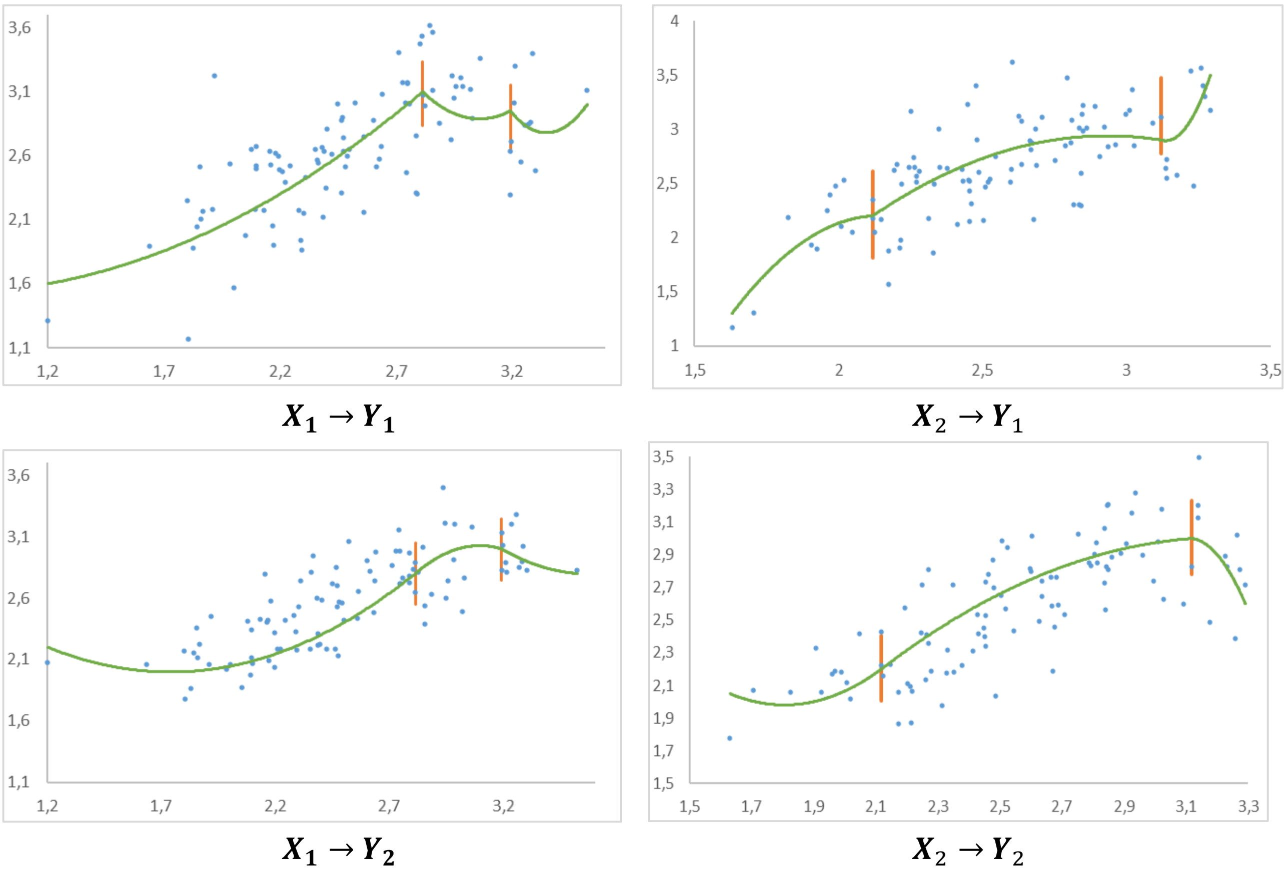

Figure 5. Relationship Pattern Curve of the Best Combined Birespone Truncated Spline Regression Model Analysis

Based on Figure 5, it can be seen that the relationship model between the predictor variables and the response variable can be flexible to follow the data pattern. Each relationship is divided into three regimes. When X_1 is less than 2.814 (<74%) or X_1 is good enough, it causes a significant increase in Y_1 and Y_2. However, when X_1 is more than 2.814 and less than 3.191 (74% - 89%) or X_1 is good, the value of Y_1 tends to decrease while Y_2 tends to increase. el response can be flexible to follow the data pattern. Conversely, when X_1 is worth more than 3,191 (> 89%) or X_1 is very good, the value of Y_1 tends to increase while Y_2 tends to decrease. This is in accordance with the results of the hypothesis test which states that there is a significant influence between X_1 on Y_1 and Y_2 as indicated by significant changes in response values.

The relationship between X_2 and Y_1 and Y_2 in the nonparametric birespon truncated spline regression model of quadratic polynomial degree (order p=2) with 2 knots is divided into three regimes. The first regime is shown when X_{2i} < 2.118. The second regime is when 2.118 < X_{2i} < 3.118 and the third regime is when X_{2i} > 3.118. When X_2 is less than 2.118 (< 12%) or X_2 is good enough, it causes an increase in Y_1 and Y_2. However, when X_1 is more than 2.118 and less than 3.118 (12% - 90%) or x_2 is good, there is a significant increase in Y_1 and Y_2. Meanwhile, when X_2 is worth more than 3.118 (> 90%) or X_2 is very good, Y_1 tends to increase while Y_2 tends to decrease. This is in accordance with the results of the hypothesis test which states that the significant effect of X_2 on Y_1 and Y_2 only occurs when the marketing relationship is of high value. Based on Table 4, it can be seen that birresponse nonparametric regression analysis has a higher coefficient of determination than truncated spline nonparametric path analysis. However, to determine the consistency of the coefficient of determination of the two analyses, a simulation study was conducted based on the sample size with the following analysis results, as shown in Table 5.

Table 5. Simulation Study Results

Sample Size	Average of Coefficient Determination	
	Nonparametric Truncated Spline Path Analysis	**Biresponse Truncated Spline Regression**
100	0.7143	0.8762
300	0.7235	0.8852
500	0.7595	0.8974

The analysis shows that the average coefficient of determination for Nonparametric Truncated Spline Path Analysis and Biresponse Truncated Spline Regression method increases as the sample size increases. At a sample size of 100, the Nonparametric Truncated Spline Path Analysis method was able to explain about 71.43% of the variation in the data, while the Biresponse Truncated Spline Regression method performed better with a coefficient of determination of 87.62%. When the sample size increased to 300, the coefficient of determination of the path analysis method increased to 72.35%, but remained lower than the biresponse method, which recorded a coefficient of determination of 88.52%. At the largest sample size of 500, the performance of both methods improved further. The path analysis method recorded a coefficient of determination of 75.95%, while the biresponse method reached 89.74%, close to 90%. Overall, the Biresponse Truncated Spline Regression method consistently performed better than Nonparametric Truncated Spline Path Analysis, especially at larger sample sizes, indicating a better ability to explain data variation.

4. Hypothesis Testing of the Best Model

Hypothesis testing using the t test is carried out on the best model, namely the quadratic polynomial degree truncated spline nonparametric path model (order p = 2) with 3 knot points through the jackknife method which begins with resampling jackknife. Jackknife resampling in this study was carried out by removing two random observations at each resampling stage. Resampling is done 1000 times. The hypothesis used is as follows. $H_0 : \beta_j = 0$ (there is no partial influence); and $H_1 : \beta_j \neq 0$ (there is a partial influence), as shown in Table 6.

Table 6. Best Model Hypothesis Testing Results

Relationship	Function Estimator	Test Statistic t	p-value	Decision
$\boldsymbol{X_1 \rightarrow Y_1}$	$\beta_{11}X_1$	14.62	<0.001	Significant
	$\beta_{12}X_1^2$	-15.15	<0.001	Significant
	$\beta_{13}(X_1 - K_{11})^2$	15.77	<0.001	Significant
	$\beta_{14}(X_1 - K_{12})^2$	-15.48	<0.001	Significant
	$\beta_{15}(X_1 - K_{13})^2$	13.51	<0.001	Significant
$\boldsymbol{X_2 \rightarrow Y_1}$	$\beta_{16}X_2$	8.84	<0.001	Significant
	$\beta_{17}X_2^2$	-23.65	<0.001	Significant
	$\beta_{18}(X_2 - K_{21})^2$	31.12	<0.001	Significant

Relationship	Function Estimator	Test Statistic t	p-value	Decision
	$\beta_{19}(X_2 - K_{22})^2$	-26.24	<0.001	Significant
	$\beta_{110}(X_2 - K_{23})^2$	-3.07	<0.001	Significant
$X_1 \to Y_2$	$\beta_{21}X_1$	-21.55	<0.001	Significant
	$\beta_{22}X_1^2$	-27.76	<0.001	Significant
	$\beta_{23}(X_1 - K_{31})^2$	25.69	<0.001	Significant
	$\beta_{24}(X_1 - K_{32})^2$	-22.88	<0.001	Significant
	$\beta_{25}(X_1 - K_{33})^2$	17.84	<0.001	Significant
$X_2 \to Y_2$	$\beta_{26}X_2$	2.32	<0.001	Significant
	$\beta_{27}X_2^2$	-7.00	<0.001	Significant
	$\beta_{28}(X_2 - K_{41})^2$	9.41	<0.001	Significant
	$\beta_{29}(X_2 - K_{42})^2$	-7.38	<0.001	Significant
	$\beta_{30}(X_2 - K_{43})^2$	-5.90	<0.001	Significant
$Y_1 \to Y_2$	$\beta_{31}Y_1$	20.38	<0.001	Significant
	$\beta_{32}Y_1^2$	-22.83	<0.001	Significant
	$\beta_{33}(Y_1 - K_{51})^2$	26.08	<0.001	Significant
	$\beta_{34}(Y_1 - K_{52})^2$	-18.92	<0.001	Significant
	$\beta_{35}(Y_1 - K_{53})^2$	-10.13	<0.001	Significant

In Table 6 it can be seen that the p-value has a value less than (0.05), so it can be decided to reject H_0 in testing the hypothesis of each variable.

D. CONCLUSION AND SUGGESTIONS

The analysis shows that the best nonparametric path model is the equation model with quadratic order and 3 knots. While the best birresponse nonparametric regression model is a model with quadratic order and 2 knots. Among the two analyses, the birresponse nonparametric regression has a higher coefficient of determination than the nonparametric path analysis, even at various sample sizes. The larger the sample size, the higher the coefficient of determination. However, the relationship in birresponse regression only applies to the relationship between predictor variables and response variables, not facilitating the presence of mediating variables such as path analysis. Future research is recommended to detect the consistency of the GCV value and the coefficient of determination in various scenarios of linear and quadratic models with various levels of knots to determine the best model scenario from both analyses.

ACKNOWLEDGEMENT

Praise be to Allah SWT who has bestowed His grace so that the writing of this journal can be completed. The author would like to thank all lecturers of the Department of Statistics, Universitas Brawijaya, parents, and friends who have provided prayers, suggestions, and input in the process of writing this journal.

REFERENCES

Aidi, M. N., & Saufitra, I. (2007). Perbaikan Metode Kriging Biasa (Ordinary Kriging) Melalui Pemecahan Matriks C Menjadi Beberapa Anak Matriks Non Overlap Untuk Mewakili Drift Pada Peubah Spasial. *In Forum Statistika dan Komputasi*, *12*(1).

Efendi, E. C. L., Achmad, A., Fernandes, R., Bernadetha, M., & Mitakda, T. (2021). *Modeling of Path Nonparametric Truncated Spline Linear, Quadratic, and Cubic in Model on Time Paying Bank Credit*.DOI/URL?

Eubank, R. L. (1999). *Nonparametric Regression and Spline Smoothing*. CRC Press. https://doi.org/10.1201/9781482273144

Fernandes, A. A. R. (2016). Estimator Spline Dalam Regresi Nonparametrik Birespon Untuk Data Longitudinal Studi Kasus Pada Pasien Penderita Tb Paru Di Malang. *Institut Teknologi Sepuluh Nopember Surbaya.* , *9*. DOI/URL?

Fernandes, A. A. R., Budiantara, I. N., Otok, B. W., & Suhartono. (2014). Reproducing kernel Hilbert space for penalized regression multi-predictors: case in longitudinal data. *International Journal of Mathematical Analysis*, *8*,issue? 1951–1961. https://doi.org/10.12988/ijma.2014.47212

Fernandes, A. A. R., Budiantara, I. N., Otok, B. W., & Suhartono. (2015). Spline Estimator for Bi-Responses and Multi-Predictors Nonparametric Regression Model in Case of Longitudinal Data. *Journal of Mathematics and Statistics*, *11*(2), 61–69. https://doi.org/10.3844/jmssp.2015.61.69

Fernandes, A. A. R., & Solimun. (2021). *Analisis Regresi dalam Pendekatan Fleksibel*. Universitas Brawijaya Press.

Firpha, R. M. P., & Achmad, A. I. (2022). Regresi Nonparametrik Spline Truncated untuk Pemodelan Persentase Penduduk Miskin di Jawa Barat Pada Tahun 2021. *Bandung Conference Series: Statistics*, *2*(2), 454–458. https://doi.org/10.29313/bcss.v2i2.4720

Härdle, W. (1990). *Applied nonparametric regression* . Cambridge university press. DOI/URL?

Hidayat, M. F., Adji Achmad, R. F., & Solimun. (2018). Estimation of Truncated Spline Function in Non-parametric Path Analysis Based on Weighted Least Square (WLS). *IOP Conference Series: Materials Science and Engineering*, *546*(5), 052027. https://doi.org/10.1088/1757-899X/546/5/052027

Iriany, A., & Fernandes, A. A. R. (2023). Hybrid Fourier series and smoothing spline path non-parametrics estimation model. *Frontiers in Applied Mathematics and Statistics*, *8*. https://doi.org/10.3389/fams.2022.1045098

Krämer, A., Green, J., Pollard, J., & Tugendreich, S. (2014). Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics*, *30*(4), 523–530. https://doi.org/10.1093/bioinformatics/btt703

Rodliyah, I. (2016). Perbandingan Metode Bootstrap Dan Jackknife Dalam Mengestimasi Parameter Regresi Linier Berganda. *JMPM: Jurnal Matematika dan Pendidikan Matematika*, *1*(1), 76. https://doi.org/10.26594/jmpm.v1i1.516

Santos, P., & Mattos, E. (2018). *Correcting the population of Brazilian municipalities using the Jackknife model **. DOI/URL?

Solimun. (2010). *Analisis Multivariat Pemodelan Struktural*. CV Citra Malang. DOI/URL?

Solimun, S., Fernandes, A. A. R., Rahmawati, I., Muflikhah, L., Ubaidillah, F., & Rohimi, A. Al. (2021). Research on Structural Flexibility and Acceptance Model (SFAM) Reconstruction Based on Disruption Innovation in the Social Humanities and Education Sector. *WSEAS TRANSACTIONS ON MATHEMATICS*, *20*, 657–675. https://doi.org/10.37394/23206.2021.20.70

Prahutama, A. (2013). Prosiding Seminar Nasional Statistika Model Regresi Nonparametrik Dengan Pendekatan Deret Fourier Pada Kasus Tingkat Pengangguran Terbuka Di Jawa Timur. *Prosiding Seminar Nasional Statistika UNDIP*, 69–76.

Ubaidillah, F., Fernandes, A. A. R., Iriany, A., Wardhani, N. W. S., & Solimun. (2022). Truncated Spline Path Analysis Modeling on in Company X with the Governments Role as a Mediation Variable. *Journal of Statistics Applications & Probability*, *11*(3), 781–794. https://doi.org/10.18576/jsap/110303

Wu, H., & Zhang, J. T. (2006). *Nonparametric regression methods for longitudinal data analysis: mixed-effects modeling approaches*. John Wiley & Sons. DOI/URL?