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. 175-189

Development of Ramsey RESET to Identify the Polynomials Order of Smoothing Spline with Simulation Study

Muhammad Rafi Hasan Nurdin[1], Adji Achmad Rinaldo Fernandes[1], Eni Sumarminingsih[1], Muhammad Ohid Ullah[2]
[1]Departement of Statistics, Universitas Brawijaya, Indonesia
[2]Departement of Statistics, Shahjalal Univeristy of Science and Technology, Bangladesh
fernandes@ub.ac.id

Article History:
Received : 23-09-2024
Revised : 07-12-2024
Accepted : 07-12-2024
Online : 03-01-2025

Keywords:
Nonparametric Path;
Smoothing Spline;
Polynomial Order;
Simulation Study.

ABSTRACT

Path analysis is used to determine the effect of exogenous variables on endogenous variables. One of the assumptions in path analysis is the linearity assumption. The linearity assumption can be tested using Ramsey RESET. If the Ramsey RESET results show that all variables are non-linear then one of the alternative models that can be used is nonparametric smoothing spline. The smoothing spline method requires a smoothing spline polynomial order in estimating the nonparametric path analysis function. This polynomial order results in the smoothing spline method having good flexibility in data adjustment. The selection of the smoothing spline polynomial order becomes an obstacle because there is no test to determine the best order. Therefore, the purpose of this study is to find out how the value of V for order 3 and 4, develop Ramsey RESET to identify the best spline polynomial order, and evaluate the Ramsey RESET algorithm through simulation studies on various errors. The results of V values of order 3 and 4 can be obtained through the integral process and it is found that the higher the order, the value of V has a higher rank. Ramsey RESET development is done by modifying the second regression using nonparametric regression functions of order 2, 3, and 4. The simulation study results show that the classical Ramsey RESET can be used to detect linear shapes well because it is not affected by the value of the error variance. However, the classical Ramsey RESET has limitations in detecting non-linear forms other than quadratic and cubic forms so that other forms such as smoothing spline are needed. In testing non-linear models, the lowest p value is obtained in the form that matches the actual conditions, this can be interpreted that the modified Ramsey RESET can detect non-linear forms with spline polynomial orders well. The contribution of this research is to provide a test to identify the best smoothing spline polynomial order using Ramsey RESET modification

Crossref
https://doi.org/10.31764/jtam.v9i1.26785

—————————— ◆ ——————————

A. INTRODUCTION

Path analysis is a statistical method used to test the direct and indirect effects of exogenous variables on endogenous variables (Sandjojo, 2011). Path analysis has assumptions that must be met first, namely the assumption of linearity. Linearity is the nature of the relationship between variables is linear. The linearity assumption test uses the Ramsey Regression Specification Error Test (RESET) to determine whether the variable relationship is linear or non-linear (Solimun & Fernandes, 2023). The Ramsey test was first developed by Ramsey in 1969 (Ramsey, 1969). The linearity test using Ramsey RESET is done by testing two regression equations. The first equation is an equation of exogenous to endogenous variables, while the second equation is the first equation added (similar to the concept of auxiliary regression) with additional predictor variables in the form of quadratic response variables ($\hat{Y}_i^2$) and cubic response variables ($\hat{Y}_i^3$) from the first regression equation (Gujarati, 2004). Both regression equations obtained coefficient of determination values which were then compared to obtain the conclusion of Ramsey RESET. The relationship between the two variables is linear when the Ramsey test on the additional variables in the second equation is insignificant. This indicates that the addition of quadratic and cubic endogenous variables is not necessary. The Ramsey RESET results will determine whether the relationship between variables is linear or non-linear.

Ramsey RESET results that show the variable relationship is linear then parametric path analysis is used. If the test shows that the variables are non-linear, one of the alternative models that can be used is nonparametric path analysis. The approach that can be used in nonparametric models is spline (Wahyuningsih et al., 2019). Splines are pieces of polynomials, which polynomials have segmented properties. This segmented nature provides more flexibility than a regular polynomial, allowing it to conform more effectively to the characteristics of a function or data (Salam et al., 2022). Spline estimation has two methods, namely truncated spline and smoothing spline. Both methods have good flexibility in data adjustment so that various curve shapes can be obtained based on different smoothing coefficients (Takezawa, 2005). The smoothing spline method requires the order of the smoothing spline polynomial in the estimation of the nonparametric path analysis function (Purnama, 2020).

The estimation of the nonparametric smoothing spline path analysis function is obtained through $\mathbf{T}d + \mathbf{V}c$ (Fernandes et al., 2017). $\mathbf{T}d$ is the parametric path analysis equation and $\mathbf{V}c$ is the smoothing spline shape that is the penalty for each observation data. The polynomial order plays a role in smoothing each observation contained in **V**c. With this polynomial order, the smoothing spline method has good flexibility in data adjustment. The order of the spline polynomial affects how well the function captures the data distribution. The better the function, the higher the coefficient of determination, so it can be said that determining the order plays an important role. If wrong choice of order can result in the estimation of the function formed is not the best estimation (Eubank, 1999).

The choice of smoothing spline polynomial order becomes an obstacle because there is no test to determine the best order. Based on the problems that have been described, this study aims to find out how the value of **V** for order 3 and 4, develop Ramsey RESET to identify the best spline polynomial order, and evaluate the Ramsey RESET algorithm through simulation studies on various errors. Ramsey RESET development is done by modifying the second regression equation with a nonparametric smoothing spline regression function. The theoretical basis used in the development refers to the modification of Ramsey RESET by Solimun & Fernandes (2023) on quadratic or more than quadratic. The modification changes the first regression equation to a quadratic form and the second regression equation to a more than quadratic form. Based on the research of Solimun & Fernandes (2023), this study modifies Ramsey RESET with the first regression with a linear form and the second regression with a non-linear nonparametric smoothing spline form.

The modified Ramsey RESET algorithm is tested with simulation studies on various values of error variance. The simulation study was conducted using secondary data on cashless society and then error generation was performed to obtain y_{sim}. X on secondary data and y_{sim} were tested on the Ramsey RESET development algorithm. The results of this study can be a reference for the use of smoothing spline polynomial orders and testing the best order using Ramsey RESET. The results of the Ramsey RESET modification and algorithm simulation study are research contributions whose results can be used to test non-linear models and find out how well the algorithm detects linear and non-linear forms.

B. METHODS

The statistical method used is nonparametric smoothing spline path analysis with polynomial orders 2 to 4. The purpose of this study is to estimate the nonparametric smoothing spline function and develop Ramsey RESET to identify the best smoothing spline polynomial order. This research uses secondary data on cashless whose variable relationships are non-linear to conduct a simulation study. Secondary data is obtained from lecturer research organized by BRIN RIIM Competition scheme. The data is the result of a questionnaire with a Likert scale of 1 for strongly disagree, 2 for disagree, 3 for undecided, 4 for agree, and 5 for strongly agree. The Likert scale on each statement item calculated the average score for each variable. The average score was used for function estimation in the simulation study. Estimating the nonparametric smoothing spline function using exogenous variable, one intervening endogenous variable, and one pure endogenous variable, the relationship between the three variables is shown in Figure 1 (Hamid et al., 2019).

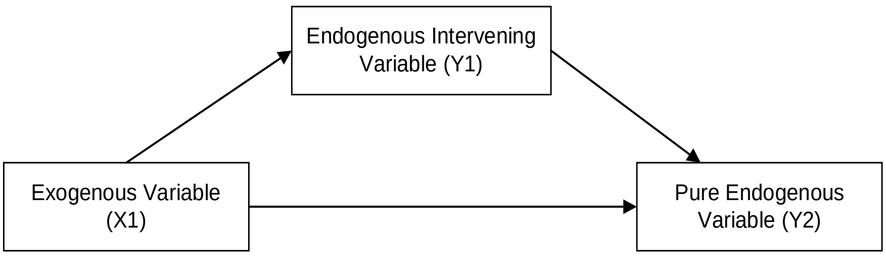

Figure 1. Simple Path Analysis Model

The basis function estimation for nonparametric smoothing spline path analysis is as follows (Pratama et al., 2024).

$$\underset{\sim}{f} = \mathbf{T}\underset{\sim}{\beta} + \mathbf{V}\underset{\sim}{\alpha} \tag{1}$$

with: **T** is a matrix of size $n \times m$ is $\underset{\sim}{\beta}$ a vector of size *m*.

$$\mathbf{T} = \begin{pmatrix} \langle \eta_1, \phi_1 \rangle & \langle \eta_1, \phi_2 \rangle & \cdots & \langle \eta_1, \phi_m \rangle \\ \langle \eta_2, \phi_1 \rangle & \langle \eta_2, \phi_2 \rangle & \cdots & \langle \eta_2, \phi_m \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \eta_n, \phi_1 \rangle & \langle \eta_n, \phi_2 \rangle & \cdots & \langle \eta_n, \phi_m \rangle \end{pmatrix}_{(n \times m)} \text{ and } \underset{\sim}{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{pmatrix}_{(m \times 1)} \tag{2}$$

The elemet a$\langle \eta_i, \phi_\ell \rangle$ is obatained from the formula as in equation (3).

$$\begin{aligned} \langle \eta_i, \phi_\ell \rangle &= \mathrm{L}_x \phi_\ell \\ &= \frac{x_i^{\ell-1}}{(\ell-1)!} \end{aligned} \tag{3}$$

Description: *i* is The *i*-th observation data (i = 1, 2, ..., n); ℓ is polynomial order a smoothing spline to-ℓ (ℓ = 1, 2, ..., m). **V** is a matrix of size $n\times n$ and $\underset{\sim}{\alpha}$ is a vector of size $n\times 1$.

$$\mathbf{V} = \begin{pmatrix} \langle \xi_1, \xi_1 \rangle & 0 & \cdots & 0 \\ 0 & \langle \xi_2, \xi_2 \rangle & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \langle \xi_n, \xi_n \rangle \end{pmatrix}_{(n\times n)} \text{ and } \underset{\sim}{\alpha} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}_{(n\times 1)} \tag{4}$$

with:

$$\begin{aligned} \langle \xi_i, \xi_i \rangle &= \mathrm{L}_x \xi_i \\ \langle \xi_i, \xi_i \rangle &= R_1(x_i, x_i) \\ \langle \xi_i, \xi_i \rangle &= \int_a^b \frac{(x_i-u)_+^{m-1}(x_i-u)_+^{m-1}}{\left((m-1)!\right)^2} du \\ \langle \xi_i, \xi_i \rangle &= \int_a^b \frac{(x_i-u)_+^{2(m-1)}}{\left((m-1)!\right)^2} du, i = 1,2,\dots,n \end{aligned} \tag{5}$$

If a $\mathrm{x}_i \in [0,1]$ in the spline polynomial order m=2, then equation (5) can be written as in equation (6).

$$\begin{aligned} \langle \xi_i, \xi_i \rangle &= \int_0^1 \frac{(x_i-u)_+^{2(2-1)}}{\left((2-1)!\right)^2} du \\ \langle \xi_i, \xi_i \rangle &= \int_0^1 \frac{(x_i-u)_+^{2}}{(1!)^2} du \\ \langle \xi_i, \xi_i \rangle &= \int_0^1 (x_i-u)^2 du \\ \langle \xi_i, \xi_i \rangle &= \left[u x_i^2 - u^2 x_i + \frac{u^3}{3} \right]_0^1 \\ \langle \xi_i, \xi_i \rangle &= x_i x_i - \frac{1}{2}(x_i + x_i) + \frac{1}{3} \\ \langle \xi_i, \xi_i \rangle &= x_i^2 - x_i + \frac{1}{3} \end{aligned} \tag{6}$$

Based on the equation that has been described, the function estimation for the simple path analysis model in Figure 1 is as follows (Fernandes et al., 2014).

X_1 toward Y_1

$$\begin{aligned} f_1(x_{1i}) &= \sum_{\ell 1=1}^{2} \beta_{1\ell 1-1} x_{1i}^{\ell 1-1} + \sum_{i=1}^{n} \alpha_{1i} \left(x_{1i}^2 - x_{1i} + \frac{1}{3} \right) \\ &= \beta_{10} + \beta_{11} x_{1i} + \alpha_{11} \left(x_{11}^2 - x_{11} + \frac{1}{3} \right) + \alpha_{12} \left(x_{12}^2 - x_{12} + \frac{1}{3} \right) + \ldots + \alpha_{1n} \left(x_{1n}^2 - x_{1n} + \frac{1}{3} \right) \end{aligned} \tag{7}$$

X_1 toward Y_2

$$\begin{aligned} f_{2.1}(x_{1i}) &= \sum_{\ell 2=1}^{2} \beta_{2\ell 2-1} x_{1i}^{\ell 2-1} + \sum_{i=1}^{n} \alpha_{2i} \left(x_{1i}^2 - x_{1i} + \frac{1}{3} \right) \\ &= \beta_{20} + \beta_{21} x_{1i} + \alpha_{21} \left(x_{11}^2 - x_{11} + \frac{1}{3} \right) + \alpha_{22} \left(x_{12}^2 - x_{12} + \frac{1}{3} \right) + \ldots + \alpha_{2n} \left(x_{1n}^2 - x_{1n} + \frac{1}{3} \right) \end{aligned} \tag{8}$$

Y_1 toward Y_2

$$\begin{aligned} f_{2.2}(\mathrm{y}_{1i}) &= \sum_{\ell 3=1}^{2} \beta_{3\ell 3-1} y_{1i}^{\ell 3-1} + \sum_{i=1}^{n} \alpha_{3i} \left(y_{1i}^2 - y_{1i} + \frac{1}{3} \right) \\ &= \beta_{30} + \beta_{31} x_{1i} + \alpha_{31} \left(y_{11}^2 - y_{11} + \frac{1}{3} \right) + \alpha_{32} \left(y_{12}^2 - y_{12} + \frac{1}{3} \right) + \ldots + \alpha_{3n} \left(y_{1n}^2 - y_{1n} + \frac{1}{3} \right) \end{aligned} \tag{9}$$

The estimation of the joint function of X_1 and Y_1 toward Y_2 is as follows.

$$\begin{aligned} f_2(x_{1i}, y_{1i}) &= \beta_{20} + \beta_{21} x_{1i} + \beta_{21} y_{1i} + \alpha_{21} \left(x_{1i}^2 - x_{1i} + y_{1i}^2 - y_{1i} + \frac{2}{3} \right) + \alpha_{22} \left(x_{1i}^2 - x_{1i} + y_{1i}^2 - y_{1i} + \frac{2}{3} \right) \\ &+ \ldots + \alpha_{2n} \left(x_{1i}^2 - x_{1i} + y_{1i}^2 - y_{1i} + \frac{2}{3} \right) \end{aligned} \tag{10}$$

Based on the estimation of the nonparametric smoothing spline function with order 2, the estimation of order 3 and 4 can be obtained. Estimation of nonparametric smoothing spline using the Penalized Weighted Least Square (PWLS) method (Fernandes et al., 2019). The selection of PWLS is based on research by Fernandes et al (2019) which says that the PWLS method is better than PLS in estimating nonparametric smoothing spline functions. The PWLS method requires an optimal smoothing coefficient through Generalized Cross Validation (GCV) (Zebua, 2021). The optimum smoothing coefficient is chosen from the minimum GCV value (Sayuti et al., 2013). Errors in determining the polynomial order can result in not obtaining the minimum GCV and optimal smoothing coefficient so that the function estimation cannot be estimated properly. The steps in the research are as follows.

1. Develop equation (5) to obtain the 3rd and 4th order values of **V**, where it is assumed that $x_i \in [0,1]$.
2. Perform Ramsey RESET modification on the second regression equation using equation (7) and the equation obtained in step 1.
3. The results of Ramsey RESET modification are used to identify the best smoothing spline polynomial order. The algorithm formed in the Ramsey modification is tested for consistency with simulation studies under different error variance conditions.
4. Conducting a simulation study using secondary data, where the variable relationship is non-linear so as to use an alternative nonparametric smoothing spline model. Simulation studies were conducted on the condition that the variable relationship was linear, non-linear with order 2, non-linear with order 3, and non-linear with order 4. The simulation study steps are:
   a. Estimating the nonparametric smoothing spline path function using equation (1) for a polynomial order of 2.
   b. In the estimation result, the function obtaining $\underset{\sim}{f} = \mathbf{T}\underset{\sim}{\beta} + \mathbf{V}\underset{\sim}{\alpha}$, to conduct a linear simulation study, the one used is $\mathbf{T}\underset{\sim}{\beta}$, then the equation $\underset{\sim}{y}_{sim} = \mathbf{T}\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ is used to obtain the value of $\underset{\sim}{y}_{sim}$ by generating normally distributed errors ($\varepsilon \sim N(0, \sigma_e^2)$ with error variances (ev) of 0.1 to 0.9. The various values of ev in this step are used to test the level of Ramsey RESET consistency. The error variance used was 0.1 to 0.9 because we wanted to know the effect of adding error variance on the results of detecting linear and non-linear properties through Ramsey RESET. The error variance used is the error variance that has been standardized so that the value is from 0 to 1, it can be interpreted that if using an error variance of 0.1 to 0.9 can reflect the condition of low to high error variance.
   c. The results in step (a) are used for non-linear simulation studies with an order of 2 (m=2). The general equation of the nonparametric model is a $\underset{\sim}{y}_{sim} = \underset{\sim}{f} + \underset{\sim}{\varepsilon}$ (Rosadi et al., 2022) error generation with various conditions as in step (b), this aims to test the level of consistency of Ramsey modification to identify the order of the smoothing spline polynomial.
   d. Estimating the nonparametric smoothing spline path function using orders 3 and 4.
   e. Generate errors as in step (c) for orders 3 and 4 respectively.
   f. Each non-linear condition is tested on the modified Ramsey RESET for order 2,3,4 to check whether the generation model formed is appropriate or not through the p value.
   g. Perform 1000 replications of generation for linear (step b), non-linear conditions of order 2 to 4 (steps c and e). The selection of 1000 replications is because it is assumed to be sufficient to illustrate the best and consistent results for each condition.
   h. Calculate the average of 1000 replications for each simulation condition.
   i. Interpretation of simulation study results.

C. RESULT AND DISCUSSION

1. Estimation of Smoothing Spline Nonparametric Path Analysis Function

The matrix **V** in equation (4) is the shape of the smoothing spline. If $x \in [0,1]$ and the spline polynomial order is 3 the equation (4) becomes as follows.

$$
\begin{aligned}
\langle \xi_i, \xi_i \rangle &= \int_0^1 \frac{(x_{1i}-u)^{2(2)}}{2!^2} du \\
&= \int_0^1 \frac{(x_{1i}^2 - 2x_{1i}u + u^2)^2}{4} du \\
&= \int_0^1 \left[\frac{x_{1i}^4 - 4x_{1i}^3 u + 6x_{1i}^2 u^2 - 4x_{1i}u^3 + u^4}{4} \right] du \\
&= \frac{1}{4}\left[x_{1i}^4 u - 2x_{1i}^3 u^2 + 2x_{1i}^2 u^3 - x_{1i}u^4 + \frac{u^5}{5} \right]_0^1 du \\
\langle \xi_i, \xi_i \rangle &= \frac{1}{4}\left[x_{1i}^4 - 2x_{1i}^3 + 2x_{1i}^2 - x_{1i} + \frac{1}{5} \right]
\end{aligned} \tag{11}
$$

Based on the calculation results in equation (11), the nonparametric smoothing spline path function of polynomial spline order 3 for simple path analysis as Figure 1 is as follows.

X1 toward Y1

$$
\begin{aligned}
f_1(x_{1i}) &= \sum_{\ell 1=1}^{3} \beta_{1\ell 1-1} x_{1i}^{\ell 1-1} + \sum_{i=1}^{n} \alpha_{1i} \left(\frac{1}{4}\left(x_{1i}^4 - 2x_{1i}^3 + 2x_{1i}^2 - x_{1i} + \frac{1}{5} \right)\right) \\
&= \beta_{10} + \beta_{11}x_{1i} + \beta_{12}x_{1i}^2 + \alpha_{11}\left(\frac{1}{4}\left(x_{1i}^4 - 2x_{1i}^3 + 2x_{1i}^2 - x_{1i} + \frac{1}{5} \right)\right) \\
&\quad + \alpha_{12}\left(\frac{1}{4}\left(x_{1i}^4 - 2x_{1i}^3 + 2x_{1i}^2 - x_{1i} + \frac{1}{5} \right)\right) + \ldots + \alpha_{1n}\left(\frac{1}{4}\left(x_{1i}^4 - 2x_{1i}^3 + 2x_{1i}^2 - x_{1i} + \frac{1}{5} \right)\right)
\end{aligned} \tag{12}
$$

X1 toward Y2

$$
\begin{aligned}
f_{2.1}(x_{1i}) &= \sum_{\ell 2=1}^{3} \beta_{2\ell 2-1} x_{1i}^{\ell 2-1} + \sum_{i=1}^{n} \alpha_{2i} \left(\frac{1}{4}\left(x_{1i}^4 - 2x_{1i}^3 + 2x_{1i}^2 - x_{1i} + \frac{1}{5} \right)\right) \\
&= \beta_{20} + \beta_{21}x_{1i} + \beta_{22}x_{1i}^2 + \alpha_{21}\left(\frac{1}{4}\left(x_{1i}^4 - 2x_{1i}^3 + 2x_{1i}^2 - x_{1i} + \frac{1}{5} \right)\right) \\
&\quad + \alpha_{22}\left(\frac{1}{4}\left(x_{1i}^4 - 2x_{1i}^3 + 2x_{1i}^2 - x_{1i} + \frac{1}{5} \right)\right) + \ldots + \alpha_{2n}\left(\frac{1}{4}\left(x_{1i}^4 - 2x_{1i}^3 + 2x_{1i}^2 - x_{1i} + \frac{1}{5} \right)\right)
\end{aligned} \tag{13}
$$

Y_1 toward Y_2

$$\begin{aligned} f_{2.2}(y_{1i}) &= \sum_{\ell 3=1}^{3} \beta_{3\ell 3-1} y_{1i}^{\ell 3-1} + \sum_{i=1}^{n} \alpha_{3i} \left(\frac{1}{4}\left(y_{1i}^4 - 2y_{1i}^3 + 2y_{1i}^2 - y_{1i} + \frac{1}{5} \right) \right) \\ &= \beta_{30} + \beta_{31} y_{1i} + \beta_{32} y_{1i}^2 + \alpha_{31} \left(\frac{1}{4}\left(y_{1i}^4 - 2y_{1i}^3 + 2y_{1i}^2 - y_{1i} + \frac{1}{5} \right) \right) \\ &+ \alpha_{32} \left(\frac{1}{4}\left(y_{1i}^4 - 2y_{1i}^3 + 2y_{1i}^2 - y_{1i} + \frac{1}{5} \right) \right) + \ldots + \alpha_{3n} \left(\frac{1}{4}\left(y_{1i}^4 - 2y_{1i}^3 + 2y_{1i}^2 - y_{1i} + \frac{1}{5} \right) \right) \end{aligned} \quad (14)$$

with $i = 1,2,3, \ldots, n$, and $\ell = 1,2,3$. If $x \in [0,1]$ and spline polynomial order of 4, equation (4) becomes as follows.

$$\begin{aligned} \langle \xi_i, \xi_i \rangle &= \int_0^1 \frac{(x_{1i} - u)^{2(3)}}{(3!)^2} du \\ &= \int_0^1 \frac{(x_{1i}^2 - 2x_{1i}u + u^2)^3}{(6)^2} du \\ &= \int_0^1 \frac{\left(x_{1i}^6 - 6x_{1i}^5 u + 15x_{1i}^4 u^2 - 20x_{1i}^3 u^3 + 15x_{1i}^2 u^4 - 6x_{1i} u^5 + u^6 \right)}{36} du \\ &= \frac{1}{36}\left[x_{1i}^6 u - 3x_{1i}^5 u^2 + 5x_{1i}^4 u^3 - 5x_{1i}^3 u^4 + 3x_{1i}^2 u^5 - x_{1i} u^6 + \frac{u^7}{7} \right]_0^1 du \\ \langle \xi_i, \xi_i \rangle &= \frac{1}{36}\left[x_{1i}^6 - 3x_{1i}^5 + 5x_{1i}^4 - 5x_{1i}^3 + 3x_{1i}^2 - x_{1i} + \frac{1}{7} \right] \end{aligned} \quad (15)$$

Based on the calculation results of equation (15), the nonparametric smoothing spline polynomial order spline 4 path function for simple path analysis as Figure 1 is as follows.

X_1 toward Y_1

$$\begin{aligned} f_1(x_{1i}) &= \sum_{\ell 1=1}^{4} \beta_{1\ell 1-1} x_{1i}^{\ell 1-1} + \sum_{i=1}^{n} \alpha_{1i} \left(\frac{1}{36}\left(x_{1i}^6 - 3x_{1i}^5 + 5x_{1i}^4 - 5x_{1i}^3 + 3x_{1i}^2 - x_{1i} + \frac{1}{7} \right) \right) \\ &= \beta_{10} + \beta_{11} x_{1i} + \beta_{12} x_{1i}^2 + \beta_{13} x_{1i}^3 + \alpha_{11} \left(\frac{1}{36}\left(x_{1i}^6 - 3x_{1i}^5 + 5x_{1i}^4 - 5x_{1i}^3 + 3x_{1i}^2 - x_{1i} + \frac{1}{7} \right) \right) + \\ & \alpha_{12} \left(\frac{1}{36}\left(x_{1i}^6 - 3x_{1i}^5 + 5x_{1i}^4 - 5x_{1i}^3 + 3x_{1i}^2 - x_{1i} + \frac{1}{7} \right) \right) + \ldots + \\ & \alpha_{1n} \left(\frac{1}{36}\left(x_{1i}^6 - 3x_{1i}^5 + 5x_{1i}^4 - 5x_{1i}^3 + 3x_{1i}^2 - x_{1i} + \frac{1}{7} \right) \right) \end{aligned} \quad (16)$$

X_1 toward Y_2

$$\begin{aligned} f_{2.1}(x_{1i}) &= \sum_{\ell 2=1}^{4} \beta_{2\ell 2-1} x_{1i}^{\ell 2-1} + \sum_{i=1}^{n} \alpha_{2i} \left(\frac{1}{36} \left(x_{1i}^6 - 3x_{1i}^5 + 5x_{1i}^4 - 5x_{1i}^3 + 3x_{1i}^2 - x_{1i} + \frac{1}{7} \right) \right) \\ &= \beta_{20} + \beta_{21} x_{1i} + \beta_{22} x_{1i}^2 + \beta_{23} x_{1i}^3 + \alpha_{21} \left(\frac{1}{36} \left(x_{1i}^6 - 3x_{1i}^5 + 5x_{1i}^4 - 5x_{1i}^3 + 3x_{1i}^2 - x_{1i} + \frac{1}{7} \right) \right) + \\ &\quad \alpha_{22} \left(\frac{1}{36} \left(x_{1i}^6 - 3x_{1i}^5 + 5x_{1i}^4 - 5x_{1i}^3 + 3x_{1i}^2 - x_{1i} + \frac{1}{7} \right) \right) + \ldots + \\ &\quad \alpha_{2n} \left(\frac{1}{36} \left(x_{1i}^6 - 3x_{1i}^5 + 5x_{1i}^4 - 5x_{1i}^3 + 3x_{1i}^2 - x_{1i} + \frac{1}{7} \right) \right) \end{aligned} \tag{17}$$

Y_1 toward Y_2

$$\begin{aligned} f_{2.2}(\mathrm{y}_{1i}) &= \sum_{\ell 3=1}^{4} \beta_{3\ell 3-1} y_{1i}^{\ell 3-1} + \sum_{i=1}^{n} \alpha_{3i} \left(\frac{1}{36} \left(y_{1i}^6 - 3y_{1i}^5 + 5y_{1i}^4 - 5y_{1i}^3 + 3y_{1i}^2 - y_{1i} + \frac{1}{7} \right) \right) \\ &= \beta_{30} + \beta_{31} y_{1i} + \beta_{32} y_{1i}^2 + \beta_{33} y_{1i}^3 + \alpha_{31} \left(\frac{1}{36} \left(y_{1i}^6 - 3y_{1i}^5 + 5y_{1i}^4 - 5y_{1i}^3 + 3y_{1i}^2 - y_{1i} + \frac{1}{7} \right) \right) + \\ &\quad \alpha_{32} \left(\frac{1}{36} \left(y_{1i}^6 - 3y_{1i}^5 + 5y_{1i}^4 - 5y_{1i}^3 + 3y_{1i}^2 - y_{1i} + \frac{1}{7} \right) \right) + \ldots + \\ &\quad \alpha_{3n} \left(\frac{1}{36} \left(y_{1i}^6 - 3y_{1i}^5 + 5y_{1i}^4 - 5y_{1i}^3 + 3y_{1i}^2 - y_{1i} + \frac{1}{7} \right) \right) \end{aligned} \tag{18}$$

with $i = 1,2,3, \ldots, n$, and $\ell = 1,2,3,4$

2. Deveploment of Ramsey RESET

The steps of developing Ramsey RESET to identify the order of the smoothing spline polynomial are as follows.

a. Calculate the estimated values of the first regression parameter (equation 19) using the Ordinary Least Square (OLS) method.

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i \tag{19}$$

b. Calculating the coefficient of determination of the first regression.

$$R^2_{first} = 1 - \frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2} \tag{20}$$

c. Calculate the second regression estimated value using the following three equations in turn.
   1) Equation (7) for smoothing spline polynomial order of 2.
   2) Equation (12) for smoothing spline polynomial order of 3.
   3) Equation (15) for a polynomial smoothing spline order of 4.

d. Calculate each coefficient of determination in the third step.

$$R^2_{second} = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{f})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \tag{21}$$

e. The three coefficients of determination obtained are then compared in turn with the coefficient of determination of the first regression to obtain the value of the F test statistic.

$$F_{count} = \frac{(R^2_{second} - R^2_{first})/\mathrm{m}}{(1 - R^2_{second})/(n-k)} \sim F_{m,n-k} \tag{22}$$

Description: R^2_{first} is coefficient of determination of the first regression at step 1; R^2_{second} is coefficient of determination of the second regression at step 3; Y_i is exogenous variable; $\bar{Y}$ is average of endogenous variables; $\widehat{Y}_i$ is estimated value of endogenous variables as step 3; $\hat{f}$ is function estimation as step 3; n is number of observations; I is value at the i-th observation; i = 1,2,3,...n; m is the number of additional exogenous; variables in the second regression; and k is number of parameters in the second regression.

3. Simulation Study

Simulations were conducted 1000 times for each Ramsey and modify Ramsey algorithm model, namely linear, smoothing with polynomial spline order 2, smoothing with polynomial spline order 3, and smoothing with polynomial spline order 4. The simulation results are shown in Table 1.

Table 1. Simulation Study Result

Variable Relationship	Error Varians	Average P-value Classic	Average P-value for Modify Ramsey		
			m=2	m=3	m=4
Linier	0.1	0.499446	-	-	-
	0.2	0.506781	-	-	-
	0.3	0.501517	-	-	-
	0.4	0.495723	-	-	-
	0.5	0.516068	-	-	-
	0.6	0.493368	-	-	-
	0.7	0.502344	-	-	-
	0.8	0.491643	-	-	-
	0.9	0.486540	-	-	-

Variable Relationship	Error Varians	Average P-value Classic	Average P-value for Modify Ramsey		
			m=2	m=3	m=4
Non-Linear (m=2)	0.1	0.000215*	**7.73×10^{-20}***	8.98×10^{-20}*	9.53×10^{-20}*
	0.2	0.009755*	**2.45×10^{-8}***	2.91×10^{-8}*	7.42×10^{-7}*
	0.3	0.044469*	**0.000078***	0.000689*	0.002208*
	0.4	0.068109	**0.000086***	0.000191*	0.005227*
	0.5	0.100687	**0.000782***	0.002707*	0.011911*
	0.6	0.133111	**0.000782***	0.003400*	0.013545*
	0.7	0.170226	**0.002147***	0.007025*	0.028452*
	0.8	0.193173	**0.003596***	0.010228*	0.032083*
	0.9	0.223640	**0.003984***	0.006450*	0.028149*
Non-Linear (m=3)	0.1	0.000119*	3.94×10^{-20}*	**1.14×10^{-21}***	4.13×10^{-20}*
	0.2	0.009197*	4.50×10^{-8}*	**3.13×10^{-10}***	6.32×10^{-7}*
	0.3	0.032617*	0.000058*	**0.000024***	0.000857*
	0.4	0.062233	0.001218*	**0.000831***	0.004444*
	0.5	0.081514	0.001027*	**0.000268***	0.005739*
	0.6	0.119540	0.003702*	**0.001035***	0.016586*
	0.7	0.139922	0.004234*	**0.000931***	0.019482*
	0.8	0.165355	0.005638*	**0.001995***	0.023563*
	0.9	0.195661	0.006959*	**0.001254***	0.027445*
Non-Linear (m=4)	0.1	0.000110*	2.75×10^{-23}*	1.66×10^{-23}*	**8.93×10^{-24}***
	0.2	0.004150*	2.74×10^{-9}*	1.23×10^{-10}*	**1.29×10^{-12}***
	0.3	0.020556*	0.007487*	0.000284*	**0.000002***
	0.4	0.058143	0.017101*	0.002400*	**0.000056***
	0.5	0.071014	0.042029*	0.005177*	**0.000083***
	0.6	0.088201	0.061564	0.007970*	**0.000811***
	0.7	0.122423	0.087707	0.010371*	**0.000572***
	0.8	0.139451	0.117832	0.014866*	**0.001037***
	0.9	0.179629	0.138990	0.028291*	**0.000977***

Description: (a) m denotes the order of the spline polynomials (m=2,3,4); (b) Classic is the Ramsey RESET developed by Ramsey in 1969; (c) The p-value shown is the average p-value of 1000 replications in each condition; (d) (*) shows significant (reject H_0 → non-linier); (e) Bold indicates the best condition because it has the smallest p-value. If it is not linear then one alternative that can be used is nonparametric path analysis; dan (f) Yellow highlights indicate the relationship is linear. If it is linear, parametric path analysis is used. Based on table 1, the average p value of Ramsey RESET and modified Ramsey RESET on the generation data (ysim) and X secondary data with linear conditions, non-linear with 2nd order, non-linear with 3rd order, and non-linear with 4th order. The results in the linear condition are consistent in various conditions of the value of ev and show the results that the variable relationship is linear. This shows that the classic Ramsey RESET algorithm developed by Ramsey in 1969 has no effect on the error variance value so it is very good for testing whether the variable relationship is linear or non-linear (quadratic or cubic). A visualization of the p-value movement in the linear condition with classical Ramsey RESET is shown in Figure 2.

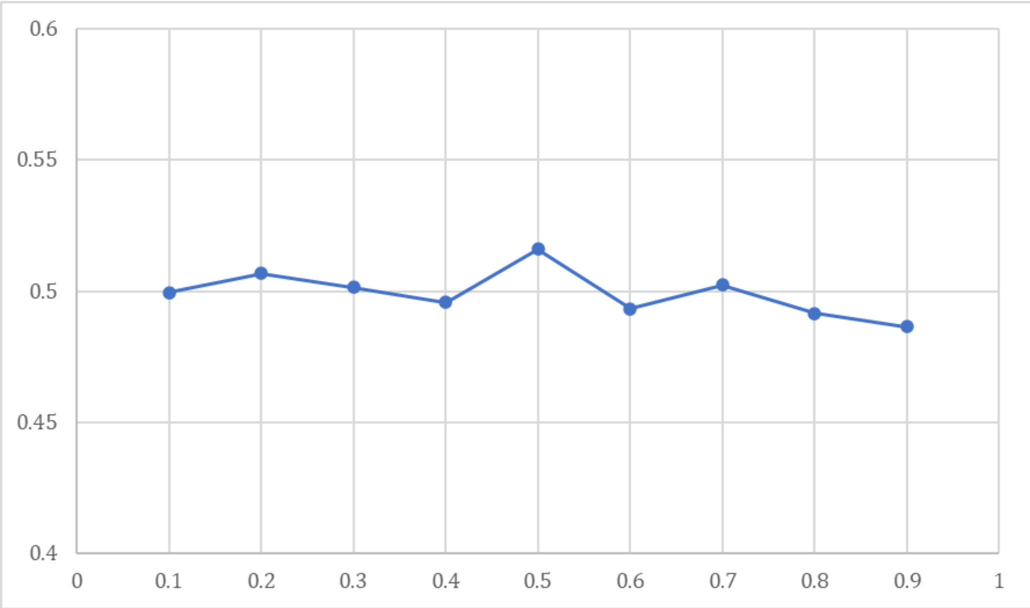

Figure 2. P-Value Movement in Linear Condition Simulation Study

The results on the non-linear variable relationship model with polynomial smoothing spline order 2 (m = 2) show good results because the model built with the smallest p value results in order 2. The smaller the p value indicates that the more appropriate the use of the model formed (this lowest value is given a bold mark). The non-linear model with order 2 is very sensitive to ev, where the higher ev, the greater the p value. When ev is 0.4, the classical Ramsey RESET results show that the variables are linear but when tested using 2nd, 3rd, and 4th order results are nonlinear. This shows that classical Ramsey has shortcomings in testing non-linear models because classical Ramsey only tests quadratic and cubic models. The visualization of p-value movement in the non-linear condition with order 2 in the Ramsey RESET modification is shown in Figure 3.

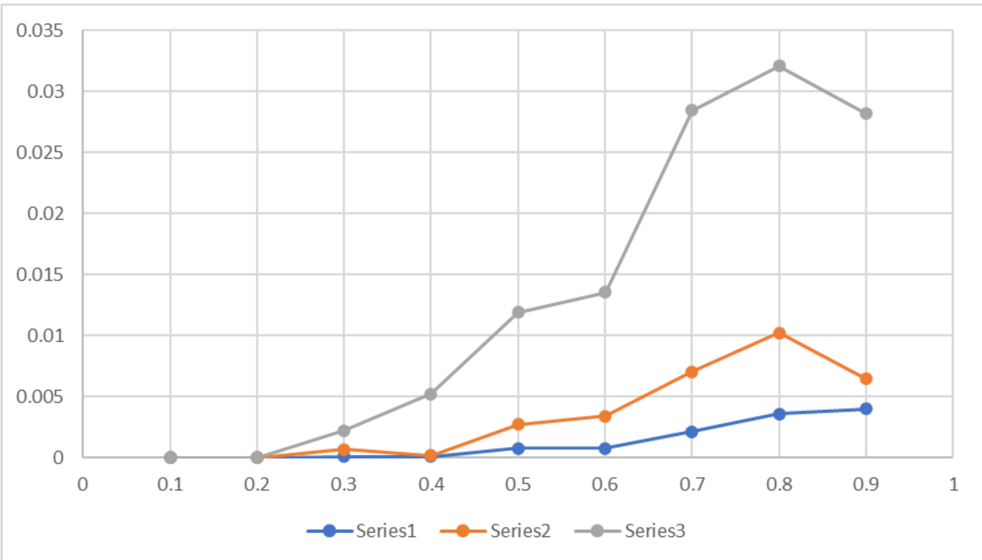

Figure 3. P-Value Movement in Simulation Study of Nonlinear Condition (m=2)

Based on Figure 3, it can be seen that the p-value tends to increase when the error variance increases, there is only a decrease in the p-value when it is 0.9. The results in non-linear conditions with orders 3 and 4 are not much different from non-linear conditions with order 2. The movement of the p-value at orders 3 and 4 is shown in Figures 4 and 5.

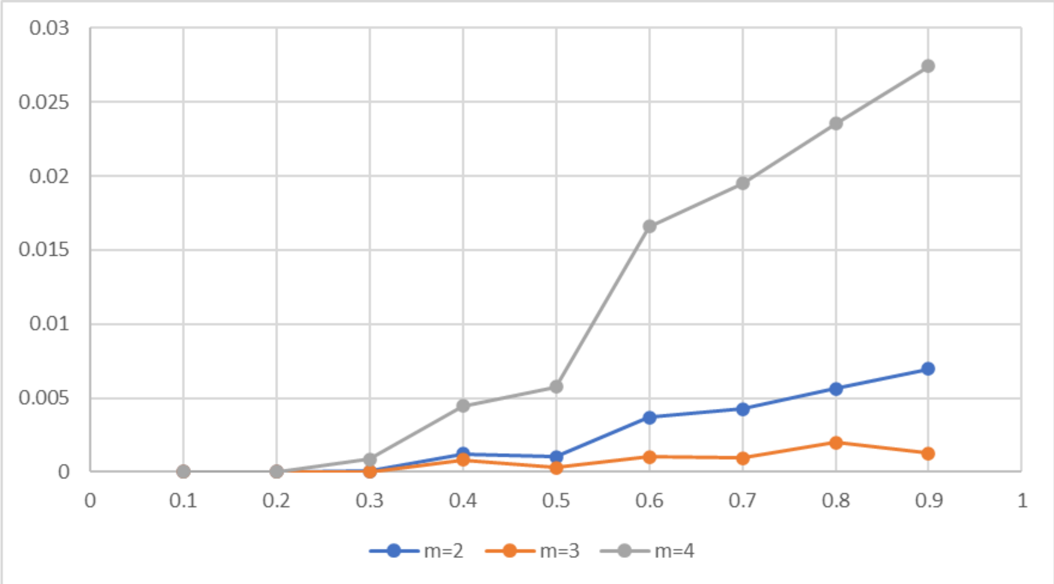

Figure 4. P-Value Movement in Simulation Study of Nonlinear Condition (m=3)

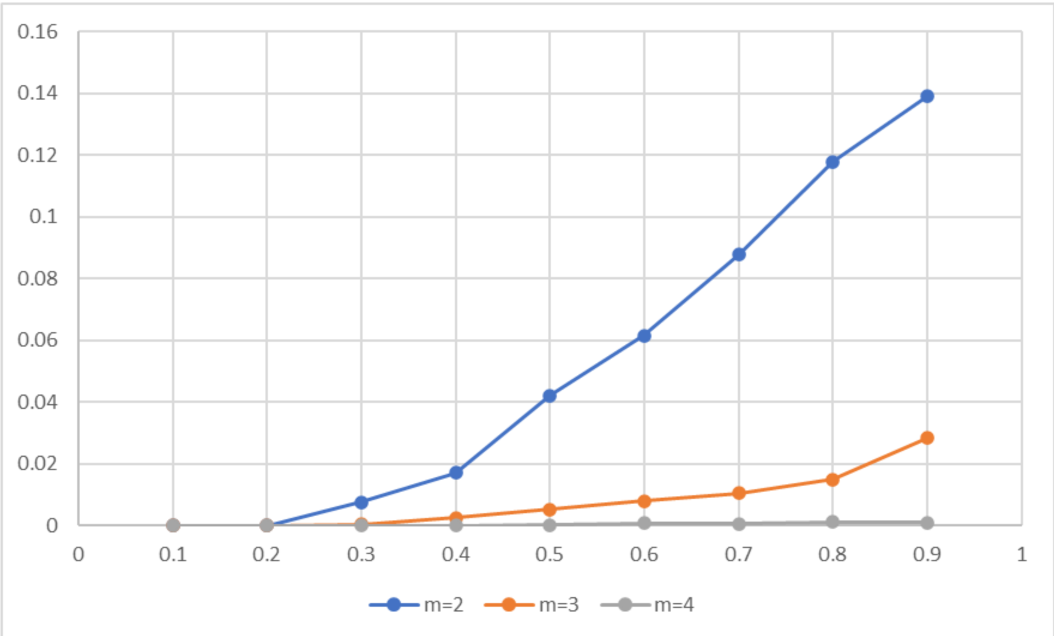

Figure 5. P-Value Movement in Simulation Study of Nonlinear Condition (m=4)

The results on the non-linear variable relationship model with smoothing spline polynomial orders 3 and 4 (m=3,4) obtained consistent results between the model generated and the modified Ramsey RESET test. This is measured by the p value between linear and non-linear conditions (m=2,3,4). The assessment criteria are the same as order 2, where the smaller it is, the more suitable it is for estimating the model function formed. The modified Ramsey RESET algorithm is sensitive to ev, where the greater ev, the greater the resulting p value which can be interpreted that the fit of the model used is decreasing. When ev is 0.4, the classic Ramsey RESET results show that the variables are linear but when tested using orders 2, 3, and 4, the results are not linear. This shows that classical Ramsey RESET has shortcomings in testing non-linear models because classical Ramsey RESET only tests quadratic and cubic models. These results are the same as the second-order conditions so that the classic Ramsey is still lacking in estimating non-linearity and testing other models such as smoothing spline is needed.

D. CONCLUSION AND SUGGESTIONS

The development of nonparametric smoothing spline function estimation with order 3 and 4 results in a longer form of function estimation because the higher the order, the more penalty is given. The use of an appropriate order is very important and necessary because the selection of the wrong order will result in the estimation of a function with a low coefficient of determination. The higher the order required, the more unformed the distribution of data used because the higher the order, the more the penalty for each observation can accommodate the distribution of data. Therefore, if there is an indication of nonlinearity when tested using the classic Ramsey RESET, further tests can be performed using modified Ramsey RESET to identify the best spline polynomial order. The modified Ramsey RESET algorithm can be applied to the research data well because it has been proven from the results of the simulation study that the appropriate variable relationship model is formed with the smallest p value in the model formed.

In reality, the distribution of data is so diverse that there are several conditions that may require a spline polynomial order of more than 4. However, future researchers need to consider the order to be developed because of the high computational level in estimating the nonparametric smoothing spline path function with a large order. In addition, the simulation study in this research is only limited to one sample size of 200. In future research, it can be recommended to use various sample sizes during simulation studies. The hope of using various sample sizes is to know the impact of sample size on p-value. Although this study still has some limitations such as one sample size, the use of the modified Ramsey RESET algorithm can be used because it has been tested to have a high level of accuracy at various stages of the simulation study.

REFERENCES

Gujarati. (2004). Basic Econometrics, Fourth Edition. In *The Economic Journal* (Vol. 82, Issue 326). https://doi.org/10.2307/2230043

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

Fernandes, A. A. R., Budiantara, I. N., Otok, B. W., & Suhartono. (2014). Spline estimator for bi-responses nonparametric regression model for longitudinal data. *Applied Mathematical Sciences*, *8*,issue? 5653–5665. https://doi.org/10.12988/ams.2014.47566

Fernandes, A. A. R., Solimun, & Arisoesilaningsih, E. (2017). *Estimation of spline function in nonparametric path analysis based on penalized weighted least square (PWLS)*. 020030. https://doi.org/10.1063/1.5016664

Hamid, M., Sufi, I., Konadi, W., & Yusrizal, A. (2019). Analisis Jalur Dan Aplikasi Spss Versi 25 Edisi Pertama. In *Aceh. Kopelma Darussalam*.

Lestari, B. (2018). Estimasi Fungsi Regresi Dalam Model Regresi Nonparametrik Birespon Menggunakan Estimator Smoothing Spline dan Estimator Kernel. *Jurnal Matematika Statistika Dan Komputasi*, *15*(2), 20. https://doi.org/10.20956/jmsk.v15i2.5710

Pratama, Y. M., Fernandes, A. A. R., Wardhani, N. W. S., & Hamdan, R. (2024). Nonparametric Smoothing Spline Approach in Examining Investor Interest Factors. *JTAM (Jurnal Teori Dan Aplikasi Matematika)*, *8*(2), 425. https://doi.org/10.31764/jtam.v8i2.20192

Purnama, D. I. (2020). A Comparison between Nonparametric Approach: Smoothing Spline and B-Spline to Analyze The Total of Train Passangers in Sumatra Island. *EKSAKTA: Journal of Sciences and Data Analysis*, *1*(1), 73–80. https://doi.org/10.20885/eksakta.vol1.iss1.art11

Ramsey, J. B. (1969). Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *31*(2), 350–371. https://doi.org/10.1111/j.2517-6161.1969.tb00796.x

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*, *546*(5), 052024. https://doi.org/10.1088/1757-899X/546/5/052024

Rosadi, S., Rinaldi, A., & Gunawan, W. (2022). Implementasi Metode Regresi Nonparametrik Spline untuk Menganalisis Keuntungan Produksi Batu-Bata. *Jurnal Ilmiah Matematika Dan Terapan*, *19*(2), 215–226. https://doi.org/10.22487/2540766x.2022.v19.i2.16150

Salam, N., Sukmawaty, Y., & Halida, A. (2022). Estimasi Model Regresi Nonparametrik Dengan Metode B-Spline. *Jurnal Sistem Media Bina Ilmiah*, *16*(10), 7631–7638.

Sandjojo, N. (2011). *Metode Analisis Jalur (Path Analysis) dan Aplikasinya*. Pustaka Sinar Harapan.

Sayuti, A., Kusnandar, D., & Mara, M. N. (2013). Generalized Cross Validation Dalam Regresi Smoothing Spline. *Buletin Ilmiah Mat. Stat. Dan Terapannya (Bimaster)*, *02*(3), 191–196. https://jurnal.untan.ac.id/index.php/jbmstr/article/view/3862/3869

Solimun, & Fernandes, A. A. R. (2023). Innovation-Based Research Using Structural Flexibility and Acceptance Model (SFAM). *Cogent Business and Management*, *10*(1). Page?https://doi.org/10.1080/23311975.2022.2128255

Solimun., Fernandes, A. A. R., & Nurjannah. (2017). *Metode Statistika Multivariat Pemodelan Persamaan Struktural (SEM) Pendekatan WarpPLS*. UB Press.

Takezawa, K. (2005). *Introduction to Nonparametric Regression*. Wiley. https://doi.org/10.1002/0471771457

Wahyuningsih, T. D., Handajani, S. S., & Indriati, D. (2019). Penerapan Generalized Cross Validation dalam Model Regresi Smoothing Spline pada Produksi Ubi Jalar di Jawa Tengah. *Indonesian Journal of Applied Statistics*, *1*(2), 117. https://doi.org/10.13057/ijas.v1i2.26250

Zebua, H. I. (2021). Pemodelan Kemiskinan di Sumatera Utara Menggunakan Regresi Nonparametrik Kernel dan Splines. *Seminar Nasional Official Statistics*, *2021*(1), 899–907. https://doi.org/10.34123/semnasoffstat.v2021i1.1087