CAUCHY: Jurnal Matematika Murni dan Aplikasi
Volume 10(1) (2025), Pages 82-95
p-ISSN: 2086-0382; e-ISSN: 2477-3344

Development of A Combined Truncated Spline And Kernel Semiparametric Path Model

Usriatur Rohma*, Adji Achmad Rinaldo Fernandes , Suci Astutik , Solimun

Department of Statistics, Faculty of Mathematics and Natural Sciences, Brawijaya University

Email: usriaturrohma@student.ub.ac.id

ABSTRACT

Semiparametric path analysis is a combination of parametric and nonparametric path analysis performed when the linearity assumption in some relationships is not met. In this study, the development of semiparametric path function estimation was carried out by combining two truncated spline and kernel approaches. In addition, the purpose of this study is to determine the significance of function estimation using t-test statistics at the jackknife resampling stage. This research was conducted in 135 Junrejo sub-districts of Batu district. The results showed that the development of a combined semiparametric path function estimation of truncated spline and kernel with weighted least square allows a more flexible and accurate estimation in modeling waste management behavior patterns. 2. The significance of the best truncated spline nonparametric path estimation in the model of the effect of Environmental Quality and the Use of Waste Banks on the Economic Benefits of Waste through the Use of the 3R Principles using t test statistics at the jackknife resampling stage shows that all exogenous variables have a significant effect on endogenous variables.

Keywords: Semiparametric Path Analysis; Truncated Spline; Kernel; Ramsey RESET; Economic Benefits of Waste

INTRODUCTION

Path analysis is a statistical method used to examine the direct and indirect effects between variables that act as causes and effects, with one of the assumptions in classical path analysis being the linearity assumption [1]. When the linearity assumption is not met, nonparametric path analysis becomes an alternative that does not rely on the assumption that the shape of the curve is known. Truncated spline is one type of nonparametric regression model that can handle changing patterns in certain sub-intervals [2]. The advantage of the truncated spline is that it is good at handling data with sharply rising or falling patterns using knot points. Not only spline, another nonparametric path approach is using kernel. The kernel nonparametric path can be used to estimate the conditional expected value of a random variable so as to find a nonlinear relationship between a pair of response and predictor random variables and obtain the

Submitted: November 16, 2024 Reviewed: December 28, 2024 Accepted: January 19, 2025
DOI: http://dx.doi.org/10.18860/ca.v10i1.29849

appropriate weights.

If in a path modeling the variable relationship is partly known curve while partly the shape is unknown, semiparametric path analysis is used. The semiparametric path approach is a combination of parametric and nonparametric approaches so that in one model, parameters and functions are estimated [3]. This method allows flexible and accurate modeling because it utilizes the strengths of both approaches. Thus, semiparametric path analysis can better capture the complexity of the relationship between variables.

Semiparametric path analysis can be applied to various fields, one of which is in the field of waste economics, so that this study will analyze the relationship between the influence of environmental quality variables and the use of waste banks on the economic benefits of waste mediated by the use of the 3R principle (Reduce, Reuse, Recycle). The topic is interesting to study because domestic waste basically has significant economic potential. Recycling, conversion into energy sources, and job creation are just a few of the benefits that can be obtained through efficient waste management. Therefore, changing people's view of waste into an economic value can be a strategic step in achieving self-sufficiency in waste management.

Waste management issues in Batu City, East Java, have become increasingly pressing due to rapid population growth and economic activities, which have significantly increased waste volumes [4]. Despite the local government's efforts with recycling programs and community education, waste management infrastructure is still inadequate, so the imbalance between waste production and management remains a challenge in some areas [5]. In addition, people's mindset towards waste, including their perspective, attitude, and behavior, plays an important role in solving this problem. The perception of people who see waste only as a useless object causes a lack of concern for waste. It is hoped that by analyzing the mindset of the community, including the community's views and understanding of the economic benefits of waste, awareness of the economic potential and positive role of waste in the environment can be increased, thus encouraging active participation in maintaining environmental sustainability in Batu City.

The development of a combined semiparametric path of truncated spline and kernel has not been done by many researchers. Research conducted by [6] who examined the combined nonparametric path of the fourier series and truncated spline which resulted in the conclusion that the form of the relationship between variables in the data of obedient behavior to pay at the bank can be described with only one approach, namely fourier series nonparametric path analysis only or truncated spline nonparametric path analysis only. This previous study used a combination of two nonparametric approaches instead of semiparametric.

Based on previous research, modeling will be carried out using semiparametric path analysis with a combined approach of truncated spline and kernel. By comparing 3 knot points on the degree of linear polynomial truncated spline and comparing 3 types of kernel, we want to get the best model in estimating the combined semiparametric path function of truncated spline and kernel. The urgency of this research is to conduct theoretical development regarding the combined semiparametric path of truncated spline and kernel. Where semiparametric modeling does not only capture one pattern but can capture more than one pattern. In addition, another urgency is to evaluate the effectiveness of policies, and develop more appropriate strategies to increase awareness and participation in sustainable waste management efforts.

METHODS

Linearity Test with Ramsey RESET

Ramsey RESET is used to test the form of the linearity relationship. Hypothesis testing used in the Ramsey RESET linearity test with the following steps [3].

1) Regress X_1 on Y_1 and calculate the estimated value of the response variable $\hat{Y}_i$. View $\hat{Y}_i$ as the endogenous variable in the model.

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} \quad (1)$$

Calculating the coefficient of determination (R^2) of regression, hereafter denoted as R_1^2 with the following equation:

$$R_1^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \quad (2)$$

2) Regressing X_1 and two additional predictor variables, namely $\hat{Y}_i^2$ and $\hat{Y}_i^3$ against $\hat{Y}_i^*$ (new regression equation) calculate the estimated value of the response variable $\hat{Y}_i^*$ then:

$$\hat{Y}_i^* = \hat{\beta}_0^* + \hat{\beta}_1^* X_{1i} + \hat{\beta}_2 \hat{Y}_i^2 + \hat{\beta}_3 \hat{Y}_i^3 \quad (3)$$

Calculating the coefficient of determination (R^2) of the regression, hereafter denoted as R_2^2 with the following equation:

$$R_2^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i^*)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \quad (4)$$

3) Perform linearity test between variables

H_0 : $\beta_2 = \beta_3 = 0$ vs

H_1 : at least one $\beta_k \neq 0; k = 2,3$

with the test statistic following the F distribution according to Equation (5)

$$F_{hitung} = \frac{(R_2^2 - R_1^2)/m}{(1 - R_2^2)/(n-k-1-m)} \sim F_{(m,n-k-1-m)} \quad (5)$$

$$Nilai - p = P\left(F_{(m,n-k-1-m)} \geq F_{hitung}\right) < \alpha$$

where:

m : The number of additional predictor variables.

k : The number of initial predictor variables.

n : The number of observations.

4) The test statistic approximates the F distribution with m and (n-k-1-m) degrees of freedom. Reject H_0 if the F test statistic > $F_{(m,n-k-1-m)}$ or p-value < α. Thus, it can be concluded that the relationship between variables is not linear. Conversely, if Accept H_0 then the relationship between variables is linear.

Truncated Spline Semiparametric Path Analysis

Semiparametric path analysis is a combination of parametric path and nonparametric regression. According to [7], the semiparametric path states that the regression curve is partly known and partly unknown in shape. The semiparametric path model if the relationship between X_1 and X_2 to Y_1 uses a parametric approach, while the rest uses a nonparametric truncated spline approach can be written in the following equation:

1) Linear polynomial degree truncated spline semiparametric path model with 1 knot

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

$$\hat{f}_{1i} = \hat{\beta}_{10} + \hat{\beta}_{11}X_{1i} + \hat{\beta}_{12}X_{2i}$$

$$\hat{f}_{2i} = \hat{\beta}_{20} + \hat{\beta}_{21}X_{1i} + \hat{\beta}_{22}(X_{1i} - K_{11})_+ + \hat{\beta}_{23}X_{2i} + \hat{\beta}_{24}(X_{2i} - K_{21})_+ + \hat{\beta}_{25}Y_{1i} + \hat{\beta}_{26}(Y_{1i} - K_{31})_+ \quad (6)$$

Where the function is truncated:

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

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

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

2) Linear polynomial degree truncated spline semiparametric path model with 2 knot points
The semiparametric truncated spline path model when linear (order p=1) with 2 knot points for two exogenous variables and two endogenous variables will be obtained as follows.

$$\hat{f}_{1i} = \hat{\beta}_{10} + \hat{\beta}_{11}X_{1i} + \hat{\beta}_{12}X_{2i}$$

$$\begin{aligned}\hat{f}_{2i} = {} & \hat{\beta}_{20} + \hat{\beta}_{21}X_{1i} + \hat{\beta}_{22}(X_{1i} - K_{11})_+ + \hat{\beta}_{23}(X_{1i} - K_{12})_+ + \hat{\beta}_{24}X_{2i} + \hat{\beta}_{25}(X_{2i} - K_{21})_+ \\ & + \hat{\beta}_{26}(X_{2i} - K_{22})_+ + \hat{\beta}_{27}Y_{1i} + \hat{\beta}_{28}(Y_{1i} - K_{31})_+ + \hat{\beta}_{29}(Y_{1i} - K_{32})_+ \end{aligned} \quad (7)$$

Where the function is truncated:

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

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

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

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

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

$$(Y_{1i} - K_{32})_+ = \begin{cases} (Y_{1i} - K_{32}) & ; Y_{1i} \geq K_{32} \\ 0 & ; Y_{1i} < K_{32} \end{cases}$$

3) Linear polynomial degree truncated spline semiparametric path model with 3 knot points
The semiparametric truncated spline path model when linear (order p=1) with 3 knot points for two exogenous variables and two endogenous variables will be obtained as follows.

$$\hat{f}_{1i} = \hat{\beta}_{10} + \hat{\beta}_{11}X_{1i} + \hat{\beta}_{12}X_{2i}$$

$$\begin{aligned}\hat{f}_{2i} = {} & \hat{\beta}_{20} + \hat{\beta}_{21}X_{1i} + \hat{\beta}_{22}(X_{1i} - K_{11})_+ + \hat{\beta}_{23}(X_{1i} - K_{12})_+ + \hat{\beta}_{24}(X_{1i} - K_{13})_+ + \hat{\beta}_{25}X_{2i} \\ & + \hat{\beta}_{26}(X_{2i} - K_{21})_+ + \hat{\beta}_{27}(X_{2i} - K_{22})_+ + \hat{\beta}_{28}(X_{2i} - K_{23})_+ + \hat{\beta}_{29}Y_{1i} + \hat{\beta}_{30}(Y_{1i} - K_{31})_+ \\ & + \hat{\beta}_{31}(Y_{1i} - K_{32})_+ + \hat{\beta}_{32}(Y_{1i} - K_{33})_+ \end{aligned} \quad (8)$$

Where the function is truncated:

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

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

$$(X_{1i}-K_{13})_+ = \begin{cases}(X_{1i}-K_{13}) & ; X_{1i} \geq K_{13} \\ 0 & ; X_{1i} < K_{13}\end{cases}$$

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

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

$$(X_{2i}-K_{23})_+ = \begin{cases}(X_{2i}-K_{23}) & ; X_{2i} \geq K_{23} \\ 0 & ; X_{2i} < K_{23}\end{cases}$$

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

$$(Y_{1i}-K_{32})_+ = \begin{cases}(Y_{1i}-K_{32}) & ; Y_{1i} \geq K_{32} \\ 0 & ; Y_{1i} < K_{32}\end{cases}$$

$$(Y_{1i}-K_{33})_+ = \begin{cases}(Y_{1i}-K_{33}) & ; Y_{1i} \geq K_{33} \\ 0 & ; Y_{1i} < K_{33}\end{cases}$$

Optimal Knot Point Selection

The best spline estimator is obtained from the optimal knot point. If the optimal knot points are obtained, the best spline function will be obtained. The selection of the best knot point is more about parsimony or simplicity of the model. The smallest GCV value is the optimal knot point. The following equation is presented [8].

$$GCV(\underset{\sim}{\mathrm{K}}) = \frac{MSE(\underset{\sim}{\mathrm{K}})}{\left[n^{-1} trace(\mathbf{I}-\mathbf{A}(\underset{\sim}{K}))\right]^2} \tag{9}$$

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

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

$$\mathbf{A}[\underset{\sim}{K}] = \mathbf{X}[\underset{\sim}{K}]\left(\mathbf{X}[\underset{\sim}{K}]^{\mathrm{T}}\hat{\mathbf{\Sigma}}^{-1}\mathbf{X}[\underset{\sim}{K}]\right)^{-1}\mathbf{X}[\underset{\sim}{K}]^{\mathrm{T}}\hat{\mathbf{\Sigma}}^{-1}$$

Kernel Semiparametric Path Analysis

Semiparametric path analysis is a combination of parametric path and nonparametric regression. According to [9], semiparametric paths are used when the regression curve is partly known and partly unknown in shape. The equation model can be written in Equation (10).

$$\begin{aligned}\hat{f}_{1i} &= \hat{\beta}_{10} + \hat{\beta}_{11}X_{1i} + \hat{\beta}_{12}X_{2i} \\ \hat{f}_{2i} &= \hat{\beta}_{20} + \hat{\beta}_{21}X_{1i} + \hat{f}(X_{1i}) + \hat{\beta}_{22}X_{2i} + \hat{f}(X_{2i}) + \hat{\beta}_{23}Y_{1i} + \hat{f}(Y_{1i})\end{aligned} \tag{10}$$

Based on Equation (10), the semiparametric kernel path equation is formed as follows.

1) Polynomial degree kernel semiparametric path model p=1
   The kernel semiparametric path model when the polynomial degree p = 1 for two exogenous variables and two endogenous variables will be obtained as follows:

$$\begin{aligned}\hat{f}_{1i} &= \hat{\alpha}_{10} + \hat{\alpha}_{11}X_{1i} + \hat{\alpha}_{12}X_{2i} \\ \hat{f}_{2i} &= \hat{\alpha}_{20} + \hat{\alpha}_{21}(X_{1i}-X) + \hat{\alpha}_{22}(X_{2i}-X) + \hat{\alpha}_{23}(Y_{1i}-Y)\end{aligned} \tag{11}$$

2) Polynomial degree kernel semiparametric path model p=2
   The kernel semiparametric path model when the polynomial degree p = 2 for two exogenous variables and two endogenous variables will be obtained as follows:

$$\begin{aligned}\hat{f}_{1i} &= \hat{\alpha}_{10} + \hat{\alpha}_{11}X_{1i} + \hat{\alpha}_{12}X_{2i} \\ \hat{f}_{2i} &= \hat{\alpha}_{20} + \hat{\alpha}_{21}(X_{1i}-X) + \hat{\alpha}_{22}(X_{1i}-X)^2 + \hat{\alpha}_{23}(X_{2i}-X) + \hat{\alpha}_{24}(X_{2i}-X)^2 \\ &\quad + \hat{\alpha}_{25}(Y_{1i}-Y) + \hat{\alpha}_{26}(Y_{1i}-Y)^2\end{aligned} \tag{12}$$

3) Polynomial degree kernel semiparametric path model p=3
   The kernel semiparametric path model when the polynomial degree p = 3 for two exogenous variables and two endogenous variables will be obtained as follows:

$$\hat{f}_{1i} = \hat{\alpha}_{10} + \hat{\alpha}_{11} X_{1i} + \hat{\alpha}_{12} X_{2i}$$

$$\hat{f}_{2i} = \hat{\alpha}_{20} + \hat{\alpha}_{21}(X_{1i} - X) + \hat{\alpha}_{22}(X_{1i} - X)^2 + \hat{\alpha}_{23}(X_{1i} - X)^3 + \hat{\alpha}_{24}(X_{2i} - X) + \hat{\alpha}_{25}(X_{2i} - X)^2 + \hat{\alpha}_{26}(X_{2i} - X)^3 + \hat{\alpha}_{27}(Y_{1i} - Y) + \hat{\alpha}_{28}(Y_{1i} - Y)^2 + \hat{\alpha}_{29}(Y_{1i} - Y)^3 \quad (13)$$

Kernel Type

A good kernel selection criterion is based on the minimum kernel risk that can be obtained from the optimal kernel or the minimum variance. The kernel estimator for the density function $g_h(u)$ as in Equation (14).

$$g_h(u) = \frac{1}{n} \sum_{i=1}^{n} K_h(x_i - x) = \frac{1}{n_h} \sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right) \quad (14)$$

Equation (14) shows that $g_h(u)$ depends on the kernel type K_h and the smoothing parameter h. The shape of the kernel weights is determined by the kernel type K_h, while the size of the weights is determined by the smoothing parameter h called bandwidth. In this study using the three best kernel types as weights can be seen in Table 1.

Table 1. Kernel types used as weights

Kernel Type	Kernel Function
Gaussian Kernel	$K_h(u) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{u^2}{2}),\ -\infty < u < \infty$
Kernel Triangle	$K_h(u) = (1 - \lvert u \rvert), (\lvert u \rvert \le 1)$
Uniform Kernel	$K_h(u) = \frac{1}{2},\ -1 \le u \le 1$

Where:

u_i : $\frac{x_i - x}{h}, i = 1,2,\dots,n$

x : predictor variable

x_1 : the i-th value of the predictor variable and h is the bandwidth

Optimal Bandwidth Selection

The selection of the optimal bandwidth is the most important problem in estimating the kernel density. Bandwidth (h) is a smoothing parameter to control the smoothness of the curve. Generalized Cross Validation (GCV) criterion is one of the methods to obtain the optimal bandwidth.

$$GCV(\underset{\sim}{h}) = \frac{MSE(\underset{\sim}{h})}{\left(n^{-1} tr\left[\mathbf{I} - (\mathbf{A}(\underset{\sim}{h}))\right]\right)^2} \quad (15)$$

where h is the bandwidth and the weight matrix $\mathbf{A}(\underset{\sim}{h})$ is obtained from:

$$\hat{\underset{\sim}{Y}} = \mathbf{A}(\underset{\sim}{h})\underset{\sim}{Y}$$

$$\mathbf{A}(\underset{\sim}{h}) = \hat{\underset{\sim}{Y}}^{-1}\underset{\sim}{Y}$$

$$\mathbf{A}(\underset{\sim}{h}) = \mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}$$

Hypothesis Testing

According to [10], resampling is the process of resampling from existing samples so that new samples are obtained. Resampling is used in hypothesis testing, however, resampling can be used to bring up all possible combinations. This is certainly quite time consuming, so computation is required [11]. The jackknife method can be used to construct the variance of an estimator [12]. According to [13] the jackknife method can be divided based on the amount of data removed into jackknife.

$$\text{Test statistics } t = \frac{\bar{\hat{\beta}}^*_{kl}}{SE_{\hat{\beta}^*_{kl}}} \sim t_{n-1} \tag{16}$$

Where:

$\bar{\hat{\beta}}^*_{kl}$: Average jackknife process parameter estimates

$SE_{\hat{\beta}^*_{kl}}$: Standard error jackknife

Data Source

The data used in this study are secondary data obtained from research grants conducted by Fernandes et al. (2024) regarding community mindset towards the economic benefits of waste in Batu City. The population in this study were all people of Junrejo Subdistrict. The sample in this study was the people of Junrejo District as many as 135 respondents. The research was conducted using latent variables derived from the Likert measurement scale. The research instrument lattice as shown in Table 2.

Table 2. Research instrument grids

Variables	Indicator
Environmental Quality (X_1)	Environmental Awareness ($X_{1.1}$)
	Environmental Maintenance ($X_{1.2}$)
	Community Attitude towards the Environment ($X_{1.3}$)
Use of Waste Bank (X_2)	The concept of waste banks in people's homes ($X_{2.1}$)
	Effectiveness of Waste Bank ($X_{2.2}$)
	Waste Bank Operational Efficiency ($X_{2.3}$)
Use of the 3R Principle (Y_1)	Reduce Effectiveness ($Y_{1.1}$)
	Reuse Effectiveness ($Y_{1.2}$)
	Recycle Effectiveness ($Y_{1.3}$)
	Waste Management Efficiency ($Y_{1.4}$)
Economic Benefits of Waste (Y_2)	Economic Waste Management Efficiency ($Y_{2.1}$)
	Utilization of Waste as an Economic Resource ($Y_{2.2}$)

Research Model

This study uses the variables of Environmental Quality (X_1), Use of Waste Bank (X_2), Use of the 3R Principle (Y_1)Principles, and Economic Benefits of Waste (Y_2). The research model can be seen in Figure 1.

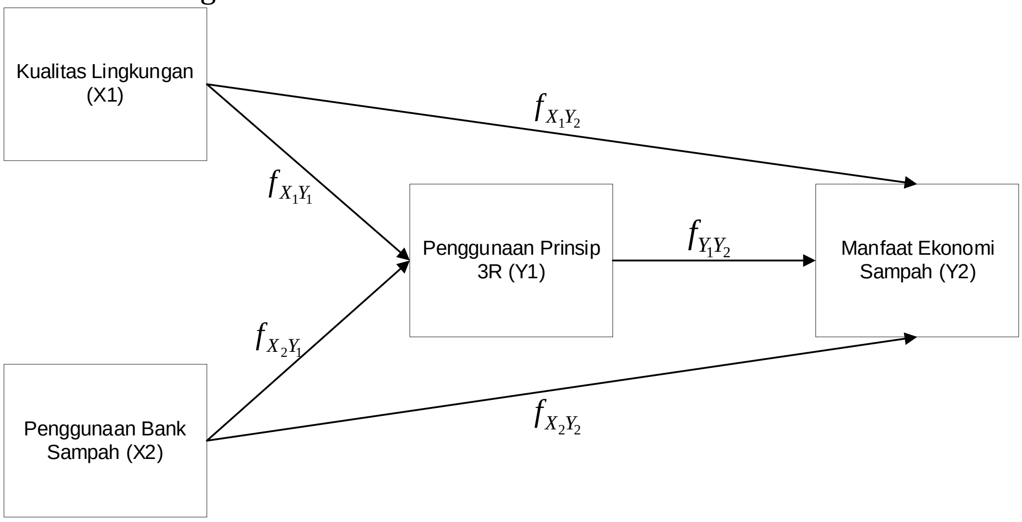

Figure 1. Research Model

Research Stages

The steps in this study are as follows:

1) Create a path diagram based on exogenous and endogenous variables in the study
2) Assign data based on the path diagram model that has been formed for exogenous and endogenous variables.
3) Testing the assumption of linearity of the relationship between variables using Ramsey's Regression Specification Error Test (RESET) with an Fcount value.
4) Estimating the combined semiparametric path function of truncated spline and kernel.
5) Determine the best model based on the largest coefficient of determination R^2.
6) Conduct hypothesis testing of the path function of the best model with the t test statistic at the jackknife resampling stage and the standard error value resulting from the resampling process

RESULTS AND DISCUSSION

Linearity Assumption Test Results

Testing the assumption of linearity is done using the Regression Specification Error Test (RESET) developed by Ramsey. The results of the linearity assumption test can be seen in Table 3.

Table 3. Ramsey Linearity Test Results RESET

Variable Relationship	F Test Statistics	p-value	Description
Environmental Quality (X_1)→ Use of 3R Principles (Y_1)	1.5234	0.2167	Linear
Use of Waste Bank (X_2)→ Use of 3R Principles (Y_1)	3.3156	0.0374	Nonlinear
Environmental Quality (X_1)→ Economic Benefits of Waste (Y_2)	3.1735	0.0430	Nonlinear

Variable Relationship	F Test Statistics	p-value	Description
Use of Waste Bank (X_2)→ Economic Benefits of Waste (Y_2)	3.5078	0.0311	Nonlinear
Use of 3R Principles (Y_1)→ Economic Benefits of Waste (Y^2)	0.8722	0.4149	Linear

Based on Table 3, it can be seen that the results of linearity testing with Ramsey RESET show that there are two relationships that have a p-value of more than 0.05, which means that the relationship is linear, namely the relationship between the variables of Environmental Quality (X_1) to the Use of 3R Principles (Y_1) and the Use of 3R Principles (Y_1) to the Economic Benefits of Waste (Y_2). The relationship between the Use of Waste Bank (X_2) to the Use of 3R Principles (Y_1), Environmental Quality (X_1) to the Economic Benefits of Waste (Y_2), and the Use of Waste Bank (X_2) to the Economic Benefits of Waste (Y_2) has a p-value of less than 0.05 so that the test decision taken is to reject H_0 which means the relationship is not linear. Because in this study there is a linear and non-linear relationship, path analysis is used with a semiparametric approach.

Best Model Selection

The criterion used to determine the best model is a small GCV value. The best model is a model that has optimal knot points and bandwidth. The selection of the smallest GCV to obtain the optimal knot point can be done using a graph that can be seen in Figure 2 and Figure 3. While the optimal bandwidth is obtained from the determination of the smallest GCV which is done using a graph can be seen in Figure 4, Figure 5.

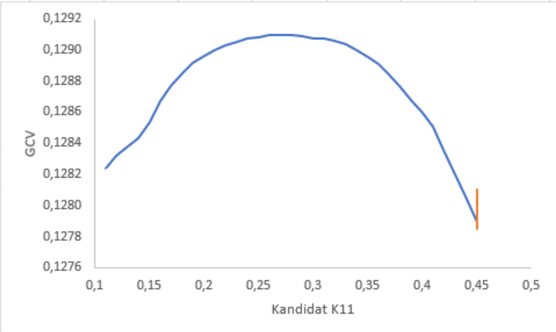

Figure 2. Candidate *knot* points K_{11} with GCV

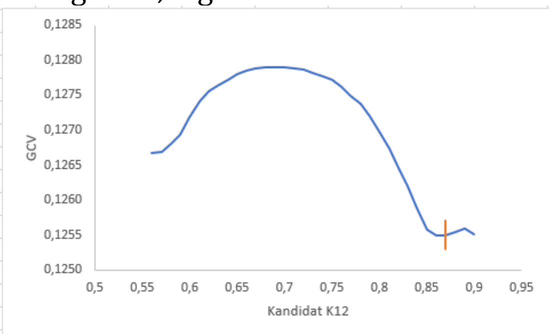

Figure 3. Candidate *knot* points K_{12} with GCV

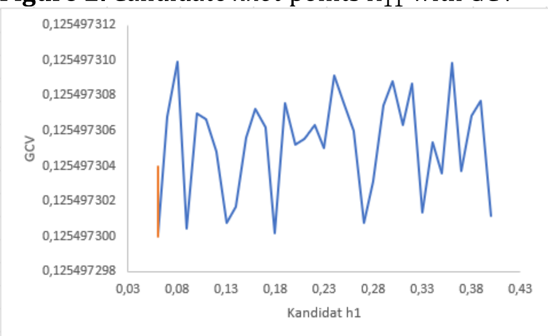

Figure 4. Candidate *bandwidth* h_1 with GCV

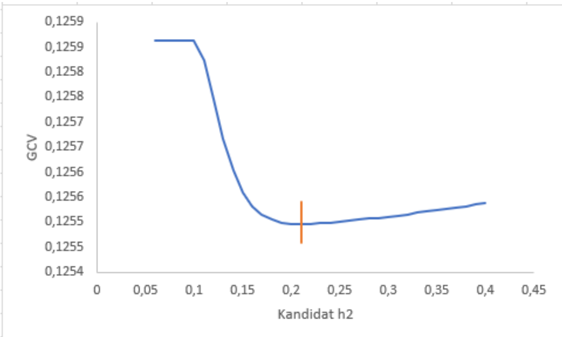

Figure 5. Candidate *bandwidth* h_2 with GCV

The selection of knot points and optimal bandwidth in each model using GCV can be seen in Table 4.

Table 4. Ramsey Linearity Test Results RESET

Variable Relationship	Optimal Knot	Optimal Bandwidth	GCV	R^2_{adj}
Use of Waste Bank (X_2)→ Use of 3R Principles (Y_1)	$K_{11} = 0.45$	-	0.1278	0,9477
	$K_{12} = 0.87$	-	0.1255	0,9522
Environmental Quality (X_1)→ Economic Benefits of Waste (Y_2)	-	$h_1 = 0.06$	0.1255	0,9522
Use of Waste Bank (X_2)→ Economic Benefits of Waste (Y_2)	-	$h_3 = 0.21$	0.1255	0,9522

Based on Table 4, the determination of the optimal *knot* point K_{11} on the relationship between the Use of Waste Bank (X_2) to the Use of 3R Principles (Y_1) is located at a value of 0.45 with a GCV value of 0.1278 and a value of 0.9477. R^2_{adj} of 0.9477. This indicates that 94.7% of the data diversity can be explained by a linear model with two *knot* points. While K_{12} is located at a value of 0.87 with a GCV value of 0.1255 and a value of R^2_{adj} of 0.9522. Determination of the optimal *bandwidth* in the relationship between Environmental Quality (X_1) to the Economic Benefits of Waste (Y_2) is located at a value of 0.06 with a GCV value of 0.1255 and a value of 0.9522. R^2_{adj} of 0.9522. While the optimal *bandwidth* in the relationship between Waste Bank Usage (X_2) to Waste Economic Benefits (Y_2) is located at a value of 0.21 with a GCV value of 0.1255 and a value of R^2_{adj} 0,9522.

Best Truncated Spline and Kernel Combined Semiparametric Path Modeling Results

Estimation of the combined truncated spline and kernel semiparametric path function is carried out after obtaining the optimal knot points and bandwidth. The results of estimating the combined truncated spline and kernel semiparametric path function are as follows.

$$\hat{f}_{1i} = 0,1283 + 0,1785X_{1i} + 0,7817X_{2i} - 0,9298(X_{2i} - 0,45)_+ + 1,5604(X_{2i} - 0,87)_+$$

$$\begin{aligned}\hat{f}_{2i} &= 1,7355 + 0,8618(X_{1i} - X) + 0,9996(X_{1i} - X)^2 + 0,3917(X_{2i} - X) \\ &+ 0,4927(X_{2i} - X)^2 + 0,5581Y_{1i}\end{aligned} \tag{17}$$

Where the function is truncated:

$$(X_{2i} - 0,45)_+ = \begin{cases} (X_{2i} - 0,45) & ; X_{2i} \geq 0,45 \\ 0 & ; X_{2i} < 0,45 \end{cases}$$

$$(X_{2i} - 0,87)_+ = \begin{cases} (X_{2i} - 0,87) & ; X_{2i} \geq 0,87 \\ 0 & ; X_{2i} < 0,87 \end{cases}$$

Hypothesis Testing Results

Hypothesis testing was conducted on the combined truncated spline and kernel semiparametric path model as follows.

$H_{01} : \beta_{11X_1Y_1} = 0$ vs $H_{11} : \beta_{11X_1Y_1} \neq 0$

$H_{02} : \beta_{12X_2Y_1} = 0$ vs $H_{12} : \beta_{12X_2Y_1} \neq 0$

$H_{03} : \beta_{13X_2Y_1} = 0$ vs $H_{13} : \beta_{13X_2Y_1} \neq 0$

$H_{04}:\beta_{14X_2Y_1}=0$ vs $H_{14}:\beta_{14X_2Y_1}\neq 0$

$H_{05}:\alpha_{21X_1Y_2}=0$ vs $H_{15}:\alpha_{21X_1Y_2}\neq 0$

$H_{06}:\alpha_{22X_1Y_2}=0$ vs $H_{16}:\alpha_{22X_1Y_2}\neq 0$

$H_{07}:\alpha_{23X_2Y_2}=0$ vs $H_{17}:\alpha_{23X_2Y_2}\neq 0$

$H_{08}:\alpha_{24X_2Y_2}=0$ vs $H_{18}:\alpha_{24X_2Y_2}\neq 0$

$H_{09}:\beta_{21Y_1Y_2}=0$ vs $H_{19}:\beta_{21Y_1Y_2}\neq 0$

Based on the results of the combined semiparametric path modeling of truncated spline and kernel, it is obtained that the form of the relationship between $X_1 \rightarrow Y_1$ and $Y_1 \rightarrow Y_2$ is linear, $X_2 \rightarrow Y_1$ has the form of a linear truncated spline relationship with 2 knot points, $X_1 \rightarrow Y_2$ and $X_2 \rightarrow Y_2$ has the form of a gaussian kernel relationship. Based on equation (4.21), further analysis of Semiparametric Path of direct influence can be done with secondary data written in the form of functions. Data patterns approached by linear parametric and nonparametric truncated spline can be interpreted, while data patterns approached by nonparametric kernel cannot be interpreted. The following are the direct effects on secondary data.

The direct effect between variable X_1 on Y_1 by assuming the relationship between other variables is constant, denoted by c, can be seen in Equation (18).

$$\hat{f}_{1i}=0{,}1785X_{1i}+c_1 \tag{18}$$

The direct effect between variable X_2 on Y_1 by assuming the relationship between other variables is constant, denoted
by *c,* can be seen in Equation (19).

$$\hat{f}_{1i}=0{,}7817X_{2i}-0{,}9298(X_{2i}-0{,}45)_{+}+1{,}5604(X_{2i}-0{,}87)_{+}+c_2 \tag{19}$$

With the following *truncated spline* conditions:
When $X_2 < 0.45$

$$\hat{f}_{1i}=0{,}7817X_{2i}+c_{21} \tag{20}$$

When $0.45 \leq X_2 < 0.87$

$$\hat{f}_{1i}=0{,}4184-0{,}1481X_{2i}+c_{22} \tag{21}$$

When $X_2 \geq 0.87$

$$\hat{f}_{1i}=-0{,}9391+1{,}4123X_{2i}+c_{23} \tag{22}$$

The direct effect between the variable Y_1 on Y_2 by assuming the relationship between other variables is constant, denoted by *c,* can be seen in Equation (23).

$$\hat{f}_{2i}=0{,}5581Y_{1i}+c_3 \tag{23}$$

The direct effect on the combined truncated spline and kernel semiparametric path calculated in the previous equation can be presented in tabular form in Table 5.

Table 5. Direct Effect Test Results

Variable Relationship	Coefficient	Estimation	Critical Ratio	p-value
Environmental Quality (X_1) → Use of 3R Principles (Y_1)	$\hat{\beta}_{11}X_{1i}$	0.1785	1.8305	0.0752**
Use of Waste Bank (X_2) → Use of 3R Principles (Y_1)	$\hat{\beta}_{12}X_{2i}$ ($X_2 < K_{11}$)	0.7817	8.0148	<0.0001*
	$\hat{\beta}_{12}X_{2i}+\hat{\beta}_{13}(X_{2i}-K_{11})$ ($K_{11} \leq X_2 < K_{12}$)	-0.1481	-9.5337	<0.0001*

Variable Relationship	Coefficient	Estimation	Critical Ratio	p-value
	$\hat{\beta}_{12}X_{2i}+\hat{\beta}_{13}(X_{2i}-K_{11})$ $+\hat{\beta}_{14}(X_{2i}-K_{12})$ $(X_2 \geq K_{12})$	1.4123	15.9988	<0.0001*
Use of 3R Principles (Y_1)→ Economic Benefits of Waste (Y^2)	$\hat{\beta}_{21}Y_{1i}$	0,5581	5.7221	<0.0001*

Notes: *Significant with $\alpha = 0{,}05$ and ** Significant with $\alpha = 0{,}1$

Based on Table 5. testing the semiparametric path model for significant effects as follows.

1) Environmental Quality (X_1) on the Use of 3R Principles (Y_1)

Testing the direct effect between Environmental Quality on the Use of the 3R Principles (Y_1) obtained a coefficient of 0.1785 with a *p-value* of 0.0752 which means significant. Thus, there is a significant direct influence between Environmental Quality on the Use of the 3R Principles (Y_1). The path coefficient is positive, indicating that the relationship between the two is positive. This means that the better the Environmental Quality (X_1), the higher the tendency of individuals or communities to apply the 3R Principles (Reduce, Reuse, Recycle) in their daily activities. This condition illustrates that improving environmental quality can motivate more environmentally friendly behavior, so that people are more encouraged to apply sustainable waste management principles.

2) The Use of Waste Bank (X_2) on the Use of 3R Principles (Y_1)

Testing the direct effect between the Use of Waste Bank (X_2) on the Use of 3R Principles (Y_1) is divided into three regions (regimes) with different conditions. The first regime is when the variable Waste Bank Usage (X_2) is less than the knot point of 0.45, the coefficient is 0.7817 with a p-value <0.0001, which means significant. This result indicates that in the first regime, when the Use of Waste Bank is below the value of 0.45, there is a significant and positive direct effect on the Use of 3R Principles (Y_1). The positive coefficient indicates that when the level of Waste Bank usage (X_2) increases within the limit below 0.45, there will be a significant increase in the Use of 3R Principles. This condition illustrates that at the initial level of waste bank usage, any increase in contribution to the waste bank substantially encourages people to more actively apply the 3R principles in their waste management. This may be due to the increased awareness of the benefits of better waste management, which encourages environmentally friendly behavior at an early stage.

In the second regime, when the Use of Waste Bank (X_2) is more than 0.45 but less than 0.87, the coefficient is -0.1481 with a p-value <0.0001, which shows a significant result. This result indicates that in the second regime, there is a significant but negative direct effect between the Use of Waste Bank (X_2) and the Use of 3R Principles (Y_1). The negative coefficient indicates that in this range, an increase in the use of waste banks actually tends to decrease the application of the 3R principles. This means that when the use of waste banks is in the range between 0.45 and 0.87, there is a tendency for people to rely on waste banks for waste management, thus reducing individual initiative in applying the 3R principles, such as reducing or recycling waste themselves. This condition may occur because people begin to feel that the contribution to the waste bank is sufficient in managing waste, so people are less encouraged to carry out additional waste management through the 3R principles. This phenomenon shows the importance of maintaining a balance between the use of waste bank facilities and increasing individual awareness in independent waste management.

In the third regime, namely when the Use of Waste Bank (X_2) is more than the knot point of 0.87, the coefficient is 1.4123 with a p-value <0.0001 which shows a significant result. This result indicates that in the third regime, there is a significant direct effect between the Use of Waste Bank (X_2) and the Use of 3R Principles (Y_1). The positive coefficient indicates that when waste bank usage exceeds the value of 0.87, there is a substantial increase in the application of the 3R principles. In other words, at high levels of waste bank usage, people are more motivated to be active in independent waste management using the 3R principles. This could be due to the higher awareness and involvement of the community in sustainable waste management, so that participation in waste banks encourages people to adopt other environmentally friendly behaviors. This condition illustrates that in the advanced stage of using waste banks, the community's contribution does not only stop at the use of waste banks, but also extends to the application of the 3R principles in daily life. This shows a synergistic positive effect between the use of waste management facilities and personal waste management practices.

3) Use of the 3R Principles (Y_1) to the Economic Benefits of Waste (Y_2)

Testing the direct effect between the Use of the 3R Principles (Y_1) on the Economic Benefits of Waste (Y_2) obtained a coefficient of 0.5581 with a p-value of <0.0001 which means significant. So, there is a significant direct influence between the Use of the 3R Principles (Y_1) on the Economic Benefits of Waste (Y_2). The path coefficient is positive, indicating that the relationship between the two is positive. In other words, the higher the application of the 3R Principles (Reduce, Reuse, Recycle), the greater the economic benefits that can be obtained from waste management. This indicates that an increase in the use of the 3R principles in waste management will have a better economic impact, such as increased resource use efficiency, reduced disposal costs, and potential additional income from recycling or reprocessing waste.

CONCLUSIONS

Based on the results of the analysis and discussion that has been carried out, it is concluded that the development of a combined truncated spline and kernel semiparametric path function estimation with weighted least square allows for a more flexible and accurate estimation in modeling waste management behavior patterns. The best model used is the model with the minimum GCV value with a component of two knot points and two bandwidths as follows.

$$\hat{f}_{1i} = 0,1283 + 0,1785X_{1i} + 0,7817X_{2i} - 0,9298(X_{2i} - 0,45)_+ + 1,5604(X_{2i} - 0,87)_+$$

$$\hat{f}_{2i} = 1,7355 + 0,8618(X_{1i} - X) + 0,9996(X_{1i} - X)^2 + 0,3917(X_{2i} - X)$$
$$+ 0,4927(X_{2i} - X)^2 + 0,5581Y_{1i}$$

The semiparametric path model above has a coefficient of determination R^2 of 0.9521 Thus, the combined truncated spline and kernel semiparametric path model can be used to model and explain the relationship between Environmental Quality and the Use of Waste Banks on the Economic Benefits of Waste with the Use of the 3R Principles as a mediating variable of 0.9521.

The significance of the best truncated spline nonparametric path estimation in the model of the effect of Environmental Quality and the Use of Waste Banks on the Economic Benefits of Waste through the Use of the 3R Principles using the t test statistic at the jackknife resampling stage shows that all exogenous variables have a significant effect on endogenous variables. Based on the results of the analysis in this study, suggestions that can be given are that further research can use other resampling methods such as bootstrap and blindfold.

REFERENCES

[1] Solimun, *Analisis Multivariat Pemodelan Struktural*. Malang: CV Citra Malang, 2010.
[2] I. N. Budiantara, B. Lestari, and A. Islamiyati, "Weighted Partial Spline Estimators in Heteroscedastic Nonparametric and Semi-parametric Regression for Longitudinal Data," 2009.
[3] A. A. R. Fernandes and Solimun, *Analisis Regresi dalam Pendekatan Fleksibel*. Malang: Universitas Brawijaya Press, 2021.
[4] P. D. Apriliyanti, "Waste Management Based on Community Empowerment in the Brantas River Watershed of Batu City," 2024.
[5] D. Werdiono, "Seriously Tackling Waste at Tlekung Landfill in Batu City".
[6] F. L. N. Rasyidah, "Estimasi Fungsi Jalur Nonparametrik Deret Fourier Dan Truncated Spline Pada Data Simulasi," *Brawijaya*, 2022.
[7] R. L. Eubank, *Nonparametric Regression and Spline Smoothing*. CRC Press, 1999. doi: 10.1201/9781482273144.
[8] I. D. A. M. I. Wulandari and I. N. Budiantara, "Analysis of Factors Affecting the Percentage of Poor Population and Food Per Capita Expenditure in East Java using Birespon Spline Nonparametric Regression," *ITS Journal of Science and Arts*, vol. 3, no. 1, pp. 30–35, 2014.
[9] H. Nurcahayani, I. N. Budiantara, and I. Zain, "The Curve Estimation of Combined Truncated Spline and Fourier Series Estimators for Multiresponse Nonparametric Regression," *Mathematics*, vol. 9, no. 10, p. 1141, May 2021, doi: 10.3390/math9101141.
[10] A. A. R. Fernandes, I. N. Budiantara, B. W. Otok, and Suhartono, "Spline Estimator for Bi-Responses and Multi-Predictors Nonparametric Regression Model in Case of Longitudinal Data," *J Math Stat*, vol. 11, no. 2, pp. 61–69, Feb. 2015, doi: 10.3844/jmssp.2015.61.69.
[11] C. Ho. Yu, "Resampling methods: concepts, applications, and justification," *Practical Assessment, Research & Evaluation*, vol. 8, no. 19, pp. 1–23, 2003.
[12] I. Rodliyah, "Perbandingan Metode Bootstrap dan Jackknife dalam Mengestimasi Parameter Regresi Linier Berganda.," *JMPM: Jurnal Matematika dan Pendidikan Matematika*, vol. I, no. 1, pp. 76–86, 2016.
[13] M. N. Aidi, "Perbaikan Metode Kriging Biasa (Ordinary Kriging) Melalui Pemecahan Matriks C Menjadi Beberapa Anak Matriks Non Overlap Untuk Mewakili Drift Pada Peubah Spasial," *Forum Statistika dan Komputasi*, vol. 12, no. 1, 2007.