

Nonparametric Smoothing Spline Approach in Examining Investor Interest Factors

Yossy Maynaldi Pratama¹, Adji Achmad Rinaldo Fernandes¹,
Ni Wayan Surya Wardhani¹, Rosita Hamdan²

¹Statistics Department, Brawijaya University, Indonesia

²Department of Development Economics, University of Malaysia Serawak, Malaysia

fernandes@ub.ac.id

ABSTRACT

Article History:

Received : 16-11-2023

Revised : 22-03-2024

Accepted : 25-03-2024

Online : 01-04-2024

Keywords:

Nonparametric;
Smoothing Spline;
Tourism;
Investment.



The nonparametric approach is an appropriate approach for patterns of relationships between predictor variables and response variables that are not or have not been known in form. In other words, there is no complete information about the pattern of relationships between variables. Curve estimation is determined based on relationship patterns in existing data. The nonparametric approach has great flexibility for estimating regression curves. This study aims to form a model on investor interest factors in improving tourism investment decisions with a nonparametric approach. The nonparametric method used is the smoothing spline regression method. The smoothing spline method is used because the modeling results from the smoothing spline approach can follow the relationship model between variables contained in the data. Thus, this method really helps researchers to model relationships between variables that are not linear and whose linear form is unknown. The results of the analysis showed that the nonparametric smoothing spline regression analysis method could model data by 94.63%, indicates that data variance can be explained by 94.63% with models, while other variance outside the study explain the remaining 5.37%. That is, investment motivation is one of the most important factors to improve investment decisions.



<https://doi.org/10.31764/jtam.v8iY2.20192>



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license

A. INTRODUCTION

One kind of statistics is regression analysis used to ascertain how predictor and response variables relate to one another (B. Xu et al., 2021). If the linearity assumption in regression analysis is satisfied and the data can be expressed in terms of specific functions, such as exponential, cubic, linear, quadratic, or cubic polynomial, a parametric approach can be used (Wadhvani & Shukla, 2019). If the shape of the curve is unknown, a nonparametric approach is used (Ibocache-Pulgar et al., 2023). The Ramsey Regression Equation Specification Error Test (RESET) is one tool for determining linearity. A nonparametric method can be applied if none of the assumptions are met. In estimating curve shapes, nonparametric approaches to regression analysis have high flexibility (Fernandes et al., 2014a).

Numerous nonparametric techniques exist truncated splines, wavelets, kernels, polynomials, MARS, and smoothing splines are a few examples (Matteson & James, 2014). As shown in these examples, smoothing splines have a unique quality that allows them to be

adjusted with good flexibility, resulting in a variety of curve shapes depending on different smoothing parameters (Wood et al., 2016). The spline curve estimator's smoothness or roughness is controlled by the smoothing parameter. Many researchers, including the following, have studied splines: Gu (2014) which discusses *smoothing spline* ANOVA for exponential distribution family epidemiology case study of Wisconsin diabetic retinopathy, Lestari et al. (2018) about the estimation of regression function nonparametric using smoothing spline and kernel estimators and Fernandes et al. (2014b) About spline *estimation* for double-response nonparametric regression of longitudinal data multi predictors. Nonparametric regression with *splines* has been applied in many fields. One of them is by Wylasmi (2016) which uses a *smoothing spline* approach in the health sector, namely on infant weight growth data in Malang City. With a *truncated spline* approach on poverty level data in East Java.

Before to employing regression analysis, several assumptions need to be taken into account, including the assumptions of linearity, residual normality, and residual homoscedasticity (Lai & Wang, 2013). One of the fundamental presumptions for figuring out the relationship between predictor and response variables is the linearity test (Fernandes et al., 2019). Regression analysis assumes that the relationship between predictor variables and response variables can be explained through a known function and that function is a linear. If this assumption is violated, estimating parametric path functions cannot be done (Nurcahayani et al., 2021).

The first assumption is the assumption of linearity. When the assumption of linearity in the data is not met, there are several solutions besides using a parametric approach, namely using a nonparametric approach. Nonparametric techniques come in a variety of forms, including wavelet, MARS, kernel, polynomial, truncated spline, and smoothing spline. One unique feature of the nonparametric smoothing spline approach is its high degree of adaptability, which allows for the creation of different curve shapes using a range of smoothing parameter values (Mielke, 2015). The spline curve estimator's smoothness or roughness is controlled by the smoothing parameter. Research on *splines* is mostly done by Hidayat et al. (2021) about regression curve estimation by using mixed smoothing spline and kernel (MsS-K) model and Budiantara et al. (2019) in comparison of smoothing and truncated spline estimators in estimating blood pressure models.

The nonparametric approach can be used when the pattern of relationships between predictor and response variables is nonlinear and has no known nonlinear form (quadratic, cubic, polynomial, and so on). Regression analysis defines the relationship between predictor variables and response variables with functions $f(x)$. Because splines can track the pattern of relationships between predictor and response variables, they are used in nonparametric regression analysis. flexibly because splines are one unique nonparametric regression technique that can effectively adapt to changes in data behavior (Mariati et al., 2021). The approach used in spline guessing *is the truncated spline approach and the smoothing spline approach*. In *truncated splines*, a knotted base is used, *while in smoothing splines*, the *estimation of the function is based on the criteria of model accuracy and the size of the smoothness of the curve which has been regulated by smoothing parameters* (Hidayati et al., 2019).

The second assumption is the assumption of normality. The normality assumption is used to determine whether the residuals in the resulting model are normally distributed (Hainmueller et al., 2019). The third assumption is residual homoscedasticity. The

homoscedasticity assumption is used to determine whether the variety of residues produced by regression analysis has relatively the same value. Parameter estimation in a parametric approach using OLS produces an unbiased and consistent estimator that is considered less efficient because it does not minimize residual variety (Du & Bentler, 2022). An alternative is needed to produce an efficient and best estimator, namely using Weighted Least Square (WLS) (P. Xu et al., 2023). The Weighted Least Square (WLS) method is a development of the OLS method that will provide a more accurate solution than the OLS method when the data is identified outliers. It is possible that the model produced by the OLS method still contains outliers so that it will affect the diversity of the rest of the models, while the WLS method minimizes the outliers in the data (Kang et al., 2020). In contrast to the nonparametric approach, the handling of heteroscedasticity cases in the weighted spline nonparametric approach uses the Weighted Penalized Least Square (PWLS) method. Regarding PWLS, numerous studies have been carried out, including: Ouyang et al. (2011) examined the effect of penalties on PWLS reconstruction images for low-dose CBCT, and Mardianto et al. (2023) about the autocorrelation comparison on longitudinal data. The residual variety's degree of heterogeneity will vary depending on the data. The Mean Absolute Percentage Error (MAPE) is one statistic that assesses this. The accuracy of a method to form a model that corresponds to calculating the absolute percentage of average error is measured by MAPE, which has a value between 0 and 1 (Abatzoglou, 2013).

Explanations related to the methods that have been described can be used to review more deeply related to the tourism sector, especially the tourism investment sector in Indonesia. The potential for tourism in Indonesia is very large considering that Indonesia consists of tens of thousands of islands (Connell, 2013). They have abundant biodiversity and are surrounded by two continents causing cultural influences that add to the cultural richness of the archipelago in addition to the original culture of the population that is not influenced by outside cultures (Wahyuni et al., 2023). Tourism trends that are most in demand by tourists to Indonesia are the natural beauty, biodiversity and cultural diversity of Indonesia. The natural potential that exists in Indonesia includes unique flora and fauna such as Komodo dragons, corpse flowers, and other very rare types. Landscapes in an area can be sought after by tourists such as Raja Ampat, Bunaken, Badr Island and others. Indonesia with all its natural wealth and biodiversity in it is an attraction for investors who want to invest in Indonesia. Tourism is a very interesting and promising sector to be developed today (Higgins-Desbiolles, 2018). The tourism sector is an activity that never dies and becomes very important for a country (Higgins-Desbiolles, 2021). The existence of tourism, more specifically for local governments, tourist attractions will be income for the region itself. This study was conducted to model tourism investment decisions that are influenced by investment interests. The analysis used to determine the influence of investment interest on tourism investment decisions in Indonesia uses the nonparametric smoothing spline regression analysis method.

B. METHODS

1. Nonparametric Smoothing Spline Regression

Smoothing spline is one of the function estimation models in regression analysis that can be applied to path analysis. *Spline* has special characteristics and good variability in adjusting data behavior patterns (Fernandes et al., 2014b). The nonparametric smoothing spline regression model has the following function as in the following equation.

$$f = Td + Vc \tag{1}$$

T is a matrix of size $n \times m$ and d is an m -sized vector as presented in the following Equation.

$$T = \begin{bmatrix} T_1 & 0 \\ 0 & T_2 \end{bmatrix} \tag{2}$$

$$T_k = \begin{pmatrix} \langle \eta_{k1}, \phi_{k1} \rangle & \langle \eta_{k1}, \phi_{k2} \rangle & \cdots & \langle \eta_{k1}, \phi_{km} \rangle \\ \langle \eta_{k2}, \phi_{k1} \rangle & \langle \eta_{k2}, \phi_{k2} \rangle & \cdots & \langle \eta_{k2}, \phi_{km} \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \eta_{kn}, \phi_{k1} \rangle & \langle \eta_{kn}, \phi_{k2} \rangle & \cdots & \langle \eta_{kn}, \phi_{km} \rangle \end{pmatrix}, \text{ and } d = \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_m \end{pmatrix} \tag{3}$$

- k : Index for endogenous variables ($k=1,2,3,..N$)
- T_k : Submatrix T for endogenous variables
- m : Derailed polynomial *spline*

Element $\langle \eta_{ki}, \phi_{k\ell} \rangle$ in the matrix T is obtained from a formula like the one in the following equation.

$$\langle \eta_{ki}, \phi_{k\ell} \rangle = \frac{x_i^{\ell-1}}{(\ell-1)!}, \text{ with } i = 1, 2, \dots, n; \ell = 1, 2, \dots, m \tag{4}$$

V is a matrix of size $n \times n$ and c is a sized vector n as presented in the following Equation.

$$V = \begin{bmatrix} V_1 & 0 \\ 0 & V_2 \end{bmatrix} \tag{5}$$

$$V_k = \begin{pmatrix} \langle \eta_{k1}, \xi_{k1} \rangle & 0 & \cdots & 0 \\ 0 & \langle \eta_{k2}, \xi_{k2} \rangle & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \langle \eta_{kn}, \xi_{kn} \rangle \end{pmatrix}, \text{ and } c = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} \tag{6}$$

Remember $\langle \eta_{kl}, \xi_{ki} \rangle = \langle \xi_{ki}, \xi_{ki} \rangle$ when $l = i$, then V can be written to:

$$V_k = \begin{pmatrix} \langle \xi_{k1}, \xi_{k1} \rangle & 0 & \dots & 0 \\ 0 & \langle \xi_{k2}, \xi_{k2} \rangle & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \langle \xi_{kn}, \xi_{kn} \rangle \end{pmatrix} \tag{7}$$

V_k is a submatrix of the matrix V for endogenous k-variables and element $\langle \xi_{ki}, \xi_{ks} \rangle$ obtained from the following equation.

$$\begin{aligned} \langle \xi_i, \xi_j \rangle &= L_x \xi_j = R_1(x_i, x_j) \\ &= \int_a^b \frac{(x_i - u)_+^{m-1} (x_j - u)_+^{m-1}}{((m-1)!)^2} du, i = 1, 2, \dots, n; j = 1, 2, \dots, n \end{aligned} \tag{8}$$

When $x \in [0, 1]$ and $m = 2$ (linear) The following results were obtained

$$\langle \xi_i, \xi_j \rangle = \int_a^b (x_i - u)(x_j - u) du \tag{9}$$

$$= \int_a^b [x_i x_j - x_i u - x_j u + u^2] du \tag{10}$$

$$= x_i x_j u - \frac{1}{2} x_i u^2 - \frac{1}{2} x_j u^2 + \frac{1}{3} u^3 \Big|_a^b \tag{11}$$

$$\langle \xi_i, \xi_j \rangle = x_i x_j - \frac{1}{2} (x_i + x_j) + \frac{1}{3} \tag{12}$$

The nonparametric path analysis model is a development of the model on nonparametric regression analysis. The nonparametric path analysis function formed based on simple parametric path analysis can be written as follows (Fernandes et al., 2019):

$$Y_{1i} = f_{1.1}(X_{1i}) + \varepsilon_{1i} \tag{13}$$

Equation (13) can be written in matrix form as presented equation (14)

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1n} \end{bmatrix} = \begin{bmatrix} f(X_{11}) \\ f(X_{12}) \\ \vdots \\ f(X_{1n}) \end{bmatrix} = \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{1n} \end{bmatrix} \tag{14}$$

Using spline's polynomial ordo $m = 2$, equation (13) can be replace like in to equation (14).

$$\tilde{f} = \mathbf{T}\tilde{d} + \mathbf{V}\tilde{c} \tag{15}$$

\mathbf{T} is the matrix with dimension $n \times m$. According to equation (6), \mathbf{T} can be written into equation (16).

$$\mathbf{T} = \begin{pmatrix} \mathbf{T}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

$$\mathbf{T} = \begin{pmatrix} 1 & X_{11} \\ 1 & X_{12} \\ \vdots & \vdots \\ 1 & X_{1n} \end{pmatrix} \tag{16}$$

Also matrix \mathbf{V} can be written into equation (17).

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

$$\mathbf{V}_1 = \begin{bmatrix} \langle \xi_{11}, \xi_{11} \rangle & 0 & \dots & 0 \\ 0 & \langle \xi_{22}, \xi_{22} \rangle & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & \langle \xi_{nn}, \xi_{nn} \rangle \end{bmatrix} \tag{17}$$

Vectors \tilde{d} and \tilde{c} can be written as equation (18) and (19)

$$\tilde{d} = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} \tag{18}$$

$$\tilde{c} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \tag{19}$$

Equation (15) can be solved as follows.

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1n} \end{bmatrix} = \begin{bmatrix} 1 & X_{11} \\ 1 & X_{12} \\ \vdots & \vdots \\ 1 & X_{1n} \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} + \begin{bmatrix} \langle \xi_{11}, \xi_{11} \rangle & 0 & \dots & 0 \\ 0 & \langle \xi_{22}, \xi_{22} \rangle & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & \langle \xi_{nn}, \xi_{nn} \rangle \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \tag{20}$$

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1n} \end{bmatrix} = \begin{bmatrix} d_1 + d_2 X_{11} \\ d_1 + d_2 X_{12} \\ \vdots \\ d_1 + d_2 X_{1n} \end{bmatrix} + \begin{bmatrix} c_1 \langle \xi_{11}, \xi_{11} \rangle \\ c_2 \langle \xi_{22}, \xi_{22} \rangle \\ \vdots \\ c_n \langle \xi_{nn}, \xi_{nn} \rangle \end{bmatrix} \tag{21}$$

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1n} \end{bmatrix} = \begin{bmatrix} d_1 + d_2 X_{11} + c_1 \langle \xi_{11}, \xi_{11} \rangle \\ d_1 + d_2 X_{12} + c_2 \langle \xi_{22}, \xi_{22} \rangle \\ \vdots \\ d_1 + d_2 X_{1n} + c_n \langle \xi_{nn}, \xi_{nn} \rangle \end{bmatrix} \tag{22}$$

Based on calculations equations (20), (21), and (22), smoothing spline regression nonparametric model for spline polynomials ordo $m=2$ (linier) can be written in equation (23).

$$f(x_{1i}) = \sum_{\ell=1}^2 d_{\ell} x_i^{1\ell-1} + c_{1i} \left(x_i^2 - x_i + \frac{1}{3} \right), i = 1, 2, 3, \dots, n; \ell = 1, 2 \tag{23}$$

2. Penalized Weighted Least Square (PWLS)

Nonparametric regression curves have no known shape, however f_k assumed smooth to obtain regression curve estimation f_k . In nonparametric path models, a single predictor that satisfies the assumed form $E(\varepsilon) = 0$ and $Var(\varepsilon) = \Sigma$ can be solved by the PWLS method (Fernandes et al., 2014b) that is:

$$\text{Min} \left\{ M^{-1} (\underline{y} - \underline{f})^T (\underline{y} - \underline{f}) + \sum_{k=1}^2 \lambda_k \int_{a_k}^{b_k} (f_i^{(m)}(x))^2 dx \right\} \tag{24}$$

Value of M In the equation above represents the value $2n$. Optimization in the Equation above considers: 2 refiner parameters λ_k as a controller between *goodness of fit (first segment)* and *roughness penalty (second segment)* where λ_k is a smoothing parameter that simultaneously controls these two things. The *goodness of fit* field in PWLS optimization is:

$$R(\underline{f}) = M^{-1} (\underline{y} - \underline{f})^T (\underline{y} - \underline{f}) \tag{25}$$

The *roughness penalty* is:

$$J(\underline{f}) = \sum_{k=1}^2 \lambda_k \int_{a_k}^{b_k} (f_i^{(m)}(x))^2 dx \tag{26}$$

3. Optimum Smoothing Parameter (λ)

Finer parameters λ is a balance controller between the conformity of the curve to the data and the smoothness of the curve. If λ has large value, the estimation of the functions obtained will be smoother, whereas if λ has small value, the estimation of the function obtained will be rougher (Lai & Wang, 2013). The method for determining the optimal smoothing parameter is the mean square error (MSE) and *generalized cross validation* (GCV).

4. Mean Square Error (MSE)

Mean square error (MSE) is the presumptive value of the residual variety. MSE is also defined as the square value of the difference between the estimator and the population parameter (Fernandes et al., 2019). The best model is a model with a minimum *MSE*, which means that the presumptive value of the model is close to the actual value. The *MSE* formula can be written as follows (Nurcahayani et al., 2021).

$$MSE = n^{-1} \sum_{i=1}^n (Y_i - f(x_i))^2 \tag{27}$$

$$= n^{-1} (Y - f(x))'(Y - f(x)) \tag{28}$$

$$= n^{-1} ((I - G)Y)'((I - G)Y) \tag{29}$$

where n: Sample size; Y_i : Response variables; x_i : Predictor variables; $f(x_i)$: Regression Function; $G: X(X'X)^{-1}X'$

5. Generalized Cross Validation (GCV)

GCV is a *cross-validation modification* (CV). CV is a method of selecting a model based on the predictive ability of the model. The formula of CV is as follows.

$$CV = n^{-1} \sum_{i=1}^n \left(\frac{y_i - f(x_i)}{1 - g_{ii}} \right)^2 \tag{30}$$

with g_{ii} is the i^{th} diagonal element of the matrix G . The *GCV equation* is obtained by substituting g_{ii} with $\sum_{i=1}^n g_{ii} = n^{-1}Tr(I - G)$. Value of $Tr(I - G)$ is the sum of the diagonal elements of the matrix $(I - G)$. GCV functions are defined as:

$$GCV = n^{-1} \sum_{i=1}^n \left(\frac{y_i - f(x_i)}{n^{-1}Tr(I - G)} \right)^2 \tag{31}$$

The selection of optimal parameters is one of the important things, because it is related to the acquisition of an optimal *spline* estimator (Sifriyani et al., 2017). When the *GCV* has a minimum value, the spline regression model can be bound to have a small error as well.

6. Goodness of Fit

Goodness of fit can be seen from several criteria, one of them is determination coefficient score (R^2). According to Gujarati (2003), determination coefficient (R^2) is an overall measure of the goodness of the function estimator model. R^2 score described in the proportion form or percentage from total variant on endogenous variable which is described by exogenous variables. R^2 score can be obtain from equation (32).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{f}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \tag{32}$$

7. Hypothesis Testing

According to (Liu & Wang, 2004) there are different treatments to test the hypothesis on a nonparametric smoothing spline approach. Testing is performed using t-SKL test (Symmetrized Kullback-Leibler). Let \hat{f}_0 be the regression function under the null model. Then $\hat{f}_0 = Hy$, where $\hat{f}_0 = (\hat{f}_0(t_1), \dots, \hat{f}_0(t_n))'$ and $H = T(T'T)^{-1}T'$. Note that H is an idempotent hat matrix and $A(\lambda)H = HA(\lambda) = H$. Define

$$S_1 = \sum_{i=1}^n (\hat{f}_\lambda(t_i) - \hat{f}_0(t_i))^2 \tag{33}$$

$$S_2 = \sum_{i=1}^n (y_i - \hat{f}_\lambda(t_i))^2 \tag{34}$$

$$S_3 = \sum_{i=1}^n (y_i - \hat{f}_0(t_i))^2 \tag{35}$$

where S_1 measures the difference between \hat{f}_0 and \hat{f}_λ , S_2 is the residual sum of squares under H_1 , and S_3 is the residual sum of square under H_0 . Where the hypothesis is determined as follows.

$$H_0 : f = constant$$

$$H_1 : f \neq constant$$

For non-Gaussian data, Xiang & Wahba (1995) proposed the SKL test based on the SKL distance between \hat{f}_0 and \hat{f}_λ :

$$t_{SKL} = \frac{1}{n} \left[E_{\hat{f}_0} \left(\log \left(\frac{\hat{f}_0}{\hat{f}_\lambda} \right) \right) + E_{\hat{f}_\lambda} \left(\log \left(\frac{\hat{f}_\lambda}{\hat{f}_0} \right) \right) \right] \tag{36}$$

For Gaussian data, it reduces to

$$t_{SKL} = \frac{1}{n\sigma^2} \left\| \hat{f}_\lambda - \hat{f}_0 \right\|^2 = \frac{1}{n\sigma^2} S_1 \tag{37}$$

The test criteria, if $t_{SKL} \geq t_{\frac{\alpha}{2}(n-1)}$ then H_0 rejected. Which mean there is significant effect between exogenous variable towards endogenous variable.

8. Research Variables

The data used in this study are secondary data. Secondary data was obtained from a research grant regarding regional investment decisions in the 2023 National Competitive Basic Research. This study uses variables of investment interest (X), and investment decision (Y). The amount of secondary data used was as many as 500 respondents. Research instruments used in secondary data, namely questionnaires with Likert scales and indicators measuring variables of investment interest and investment decisions are presented in the following Table 1. In this study, validity and reliability checks have been carried out and provide results that all items are valid and reliable.

Table 1. Research Variables

Research Variables	Indicators
Investment Interest (X)	(1). Benefit
	(2). Capital
	(3). Motivation
	(4). Profit
	(5). Education
Investment Decisions (Y)	(1). Profitability
	(2). <i>Market Orientation</i>
	(3). Tax Protection
	(4). Company Growth
	(5). Kind of Investment

9. Investment Interest

Understanding investment from an economic point of view is a commitment to sacrifice funds with a definite amount in the present to get profits in the future (Pajar & Pustikaningsih, 2017). So investment interest is the desire to find information about the type of investment starting from advantages, weaknesses, investment performance, and so on. The characteristics of people who are interested in investing can be known by how much effort potential investors make in finding out about a type of investment, studying and then practicing it. Another characteristic that can be seen is that potential investors will try to take the time to learn more about the investment or immediately try to invest in this type of investment, even adding to the portion of existing investments (Kusumawati, 2020).

10. Investment Decisions

Investment decision is one of the functions of financial management that involves allocating funds, both funds sourced from inside and outside the company in various forms of investment decisions with the aim of obtaining greater profits from the cost of funds in the future. Forms of investment include short-term investments consisting of investments into cash, short-term securities, receivables, and inventories. Long-term investments in the form of land, buildings, vehicles, machinery, production equipment and more. In addition to getting profits in the future, another goal is to maximize the value of the company.

C. RESULT AND DISCUSSION

1. Linearity Test

The linearity test is used to determine the pattern of relationships between variables. The Ramsey Reset Test was used to test linearity in this study. Data on latent variables in testing the linearity assumption were obtained using the average score. The following are the results of the linearity test in the following Table 2.

Table 2. Linearity Test Results

Relationship	F- Value	p-Value	Information
X towards Y	10.049	0.002	Non-Linear

Based on Table 2, it can be explained that the relationship between the predictor variable (X) and the response variable (Y) is not linear because it has a p-value ($0.002 < 0.05$). The parametric approach cannot be used in this case because the data does not satisfy the assumption of linearity. Another solution that can be used is to use a nonparametric approach. The nonparametric approach can be used because the relationship between the predictor (X) and response (Y) variables does not satisfy the assumption of linearity. A visualization of the relationship pattern between the investment interest variable as a predictor variable (X) and the investment decision variable as a response variable (Y) is shown as follows.

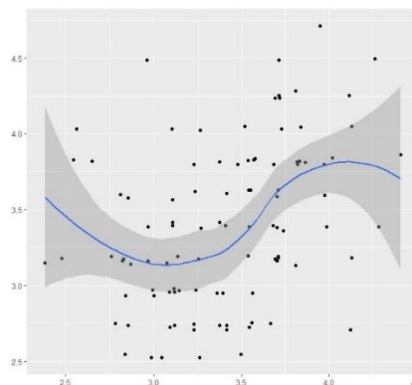


Figure 1. Linearity Plot Visualization

In the figure above, it can be seen that the pattern of the relationship between the investment interest variable as a predictor variable (X) to the investment decision variable as a response variable (Y) has a pattern that is not linear and does not form quadratic and cubic patterns so that it cannot use a parametric approach to estimate the regression coefficient

parameters. So that the nonparametric approach is used to estimate the regression coefficient parameters because the nonparametric approach can estimate regression coefficient parameters in data that have a pattern of relationships between predictor variables (X) and response variables (Y) that are nonlinear and the pattern is unknown. This is because the nonparametric approach has flexibility in data patterns so that the results of estimating regression coefficient parameters can follow the pattern of relationships between predictor variables (X) and response variables (Y) of unknown form. The predictor variable (X) in this study is the investment interest variable and the response variable (Y) in this study is the investment decision variable.

2. Nonparametric Smoothing Spline Estimation

To obtain a nonparametric model with the best smoothing parameters, the GCV value must be minimum. The iteration to obtain the minimum GCV performed using ‘nlminb’ package to solve the optimization with R software yields the lambda and GCV pairs described by the following Table 3.

Table 3. Lambda and GCV values.

Lambda	GCV
0.001	9.0271513218
0.00116224	9.0271513100
0.00126458	9.0271513041
0.00182164	9.0271512836
0.00233956	9.0271512733
0.00314073	9.0271512640
0.00413704	9.0271512575
0.00549611	9.0271512525
0.00728268	9.0271512487
0.00966435	9.0271512458
0.01283050	9.0271512437
0.01703570	9.0271512420
0.02271229	9.0271512408

The main principle of the *Generalized Cross Validation* (GCV) method is the process of selecting the best model by entering λ values and taking λ which produces the lowest GCV value. To obtain the minimum GCV, which is characteristic for determining the model with the best λ , a three-stage iteration is carried out. In the end, a minimum GCV of 9.0271 was obtained when the λ value was 0.02271. Estimating curves with a nonparametric *smoothing spline* approach requires two vector coefficients, vector d and vector c . Table 4 below presents the coefficients of estimating non-parametric regression functions in estimators with the PWLS approach.

Table 4. Smoothing Spline Estimation Function.

Symbol	Coefficient
d_{11}	1.2040
d_{12}	-1.8227
$c_{1,1}$	2.7121
...	...
$c_{1,500}$	4.2754

Based on Table 4, the estimator equation for the nonparametric *smoothing spline* regression function can be written as follows.

$$f(x_i) = 1.2040 - 1.8227x_i + 2.7121\left(x_1x_i - \frac{1}{2}(x_1 + x_i) + \frac{1}{3}\right) + \dots + 4.2754\left(x_{500}x_i - \frac{1}{2}(x_{500} + x_i) + \frac{1}{3}\right)$$

The values of d_{11} and d_{12} can be interpreted like slope and intercept values in simple linear regression analysis. While the values $c_{1.1}$ to $c_{1.500}$ are penalty values in each data. Where this penalty value provides the accuracy of the resulting modelling on the nonparametric smoothing spline approach. So that in each of the first data up to the 500th data, a penalty is applied. This is the main characteristic of the nonparametric smoothing spline approach where the modelling obtained produces a model that can closely follow the pattern of the data owned.

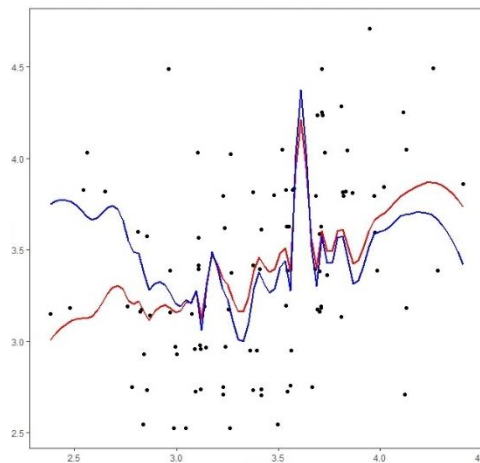


Figure 2. Smoothing Spline Modelling Visualization

The measure of goodness of the model (R2 test statistics) obtained is 0.94627. This means that 94.63% of the diversity of data has been able to be explained by the nonparametric *smoothing spline function estimator model*, and the rest (5.37%) is explained by other variables outside the model. *One of the* advantages of estimating the smoothing spline function is that the pattern of relationships between variables is more sensitive to the original data. This is due to the penalty element in the model that can produce function estimators with high flexibility. The penalty in the model can be seen in the equation above where each observation has a mathematical element that moves based on the observation, for example in the first observation (X_1) penalized $2.7121\left(x_1x_i - \frac{1}{2}(x_1 + x_i) + \frac{1}{3}\right)$, and so on.

3. Hypothesis Testing

Hypothesis testing in this study does not use t-test as commonly used, but uses SKL (Symmetrized Kullback-Leibler) t-test. The t-SKL test is used because in nonparametric smoothing spline modelling special treatment is required where the t-SKL test statistics are obtained by comparing the value of the function when it is 0 and when it is the same value as the smoothing parameter ("lambda"). The test criterion t SKL (Symmetrized Kullback-Leibler)

is that when the test statistics $t_{SKL} \geq t_{\frac{\alpha}{2}(n-1)}$ then H_0 is rejected. The rejection of H_0 means that there is a significant influence between the investment interest variable as a predictor variable (X) and the investment decision variable as a response variable (Y). Based on the results of hypothesis testing using the t-SKL test, the following results were obtained.

Table 1. Results of Hypothesis Testing using t-SKL

Relationship	t- score	p-value	Information
X towards Y	7.369	0.002	Significant

Based on the table above, it can be seen that the results of hypothesis testing using the SKL (Symmetrized Kullback-Leibler) t-test produce a significance value of < 0.05 . The significance value or commonly referred to as a p-value of less than 0.05 means reject H_0 , so it can be concluded that there is a significant influence between predictor variables or in this study is an investment interest variable (X) on response variables, namely investment decision variables (Y).

D. CONCLUSION AND SUGGESTIONS

Based on the analysis that has been done, it was found that a nonparametric smoothing spline approach is suitable to be used to determine how the influence of investment motivation on investment decisions is related. The high degree of flexibility of the nonparametric smoothing spline approach makes this approach most likely to be used in various disciplines. Data challenges do not become a serious problem if a nonparametric smoothing spline approach is used to model or conjecture a function in a particular problem. The advice that researchers can give is to develop it on more complex models such as using more than one predictor variable. In addition, from the modelling that has been done, in the next research can be done on data that has outliers. The flexibility of the nonparametric smoothing spline approach allows the resulting model to model well data that has outliers.

REFERENCES

- Abatzoglou, J. T. (2013). Development of gridded surface meteorological data for ecological applications and modelling. *International Journal of Climatology*, 33(1), 121–131. <https://doi.org/10.1002/joc.3413>
- Connell, J. (2013). Contemporary medical tourism: Conceptualisation, culture and commodification. *Tourism Management*, 34, 1–13. <https://doi.org/10.1016/j.tourman.2012.05.009>
- Du, H., & Bentler, P. M. (2022). Distributionally weighted least squares in structural equation modeling. *Psychological Methods*, 27(4), 519–540. <https://doi.org/10.1037/met0000388>
- 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). <https://doi.org/10.1088/1757-899X/546/5/052024>
- Fernandes, A. A. R., Nyoman Budiantara, I., Otok, B. W., & Suhartono. (2014a). Spline estimator for bi-responses nonparametric regression model for longitudinal data. *Applied Mathematical Sciences*, 8(113–116), 5653–5665. <https://doi.org/10.12988/ams.2014.47566>
- Fernandes, A. A. R., Nyoman Budiantara, I., Otok, B. W., & Suhartono. (2014b). Spline estimator for bi-responses nonparametric regression model for longitudinal data. *Applied Mathematical Sciences*, 8(113–116), 5653–5665. <https://doi.org/10.12988/ams.2014.47566>

- Gu, C. (2014). *Journal of Statistical Software Smoothing Spline ANOVA Models: R Package gss*. <http://www.jstatsoft.org/>
- Gujarati, D. N. (2003). *Basic Econometrics* (L. Sutton & A. Bright, Eds.; Fourth Edition). Gary Burke.
- Hainmueller, J., Mummolo, J., & Xu, Y. (2019). How Much Should We Trust Estimates from Multiplicative Interaction Models? Simple Tools to Improve Empirical Practice. *Political Analysis*, 27(2), 163–192. <https://doi.org/10.1017/pan.2018.46>
- Hidayat, R., Budiantara, I. N., Otok, B. W., & Ratnasari, V. (2021). The regression curve estimation by using mixed smoothing spline and kernel (MsS-K) model. *Communications in Statistics - Theory and Methods*, 50(17), 3942–3953. <https://doi.org/10.1080/03610926.2019.1710201>
- Hidayati, L., Chamidah, N., & Nyoman Budiantara, I. (2019). Spline Truncated Estimator in Multiresponse Semiparametric Regression Model for Computer based National Exam in West Nusa Tenggara. *IOP Conference Series: Materials Science and Engineering*, 546(5), 052029. <https://doi.org/10.1088/1757-899X/546/5/052029>
- Higgins-Desbiolles, F. (2018). Sustainable tourism: Sustaining tourism or something more? *Tourism Management Perspectives*, 25, 157–160. <https://doi.org/10.1016/j.tmp.2017.11.017>
- Higgins-Desbiolles, F. (2021). The “war over tourism”: challenges to sustainable tourism in the tourism academy after COVID-19. *Journal of Sustainable Tourism*, 29(4), 551–569. <https://doi.org/10.1080/09669582.2020.1803334>
- Ibacache-Pulgar, G., Villegas, C., López-Gonzales, J. L., & Moraga, M. (2023). Influence measures in nonparametric regression model with symmetric random errors. In *Statistical Methods and Applications* (Vol. 32, Issue 1, pp. 1–25). Institute for Ionics. <https://doi.org/10.1007/s10260-022-00648-z>
- Kang, S., Kim, T., & Chung, W. (2020). Hybrid RSS/AOA Localization using Approximated Weighted Least Square in Wireless Sensor Networks. *Sensors*, 20(4), 1159. <https://doi.org/10.3390/s20041159>
- Kusumawati, Y. T. (2020). The Effect Of Financial Performance And Growth Opportunity On Investment Decisions In Food And Beverage Companies In Indonesia Stock Exchange. *Jurnal Ekonomi Dan Manajemen*, 14(1), 144–157. <https://doi.org/10.30650/jem.v14i1.253>
- Lai, M. J., & Wang, L. (2013). Bivariate penalized splines for regression. *Statistica Sinica*, 23(3), 1399–1417. <https://doi.org/10.5705/ss.2010.278>
- Lestari, B., Fatmawati, Budiantara, I. N., & Chamidah, N. (2018). Estimation of Regression Function in Multi-Response Nonparametric Regression Model Using Smoothing Spline and Kernel Estimators. *Journal of Physics: Conference Series*, 1097(1). <https://doi.org/10.1088/1742-6596/1097/1/012091>
- Liu, A., & Wang, Y. (2004). Hypothesis testing in smoothing spline models. *Journal of Statistical Computation and Simulation*, 74(8), 581–597. <https://doi.org/10.1080/00949650310001623416>
- Mardianto, M. F. F., Gunardi, & Utami, H. (2023). *Interval estimation for nonparametric regression using Fourier series estimator in longitudinal data*. 030010. <https://doi.org/10.1063/5.0103799>
- Mariati, N. P. A. M., Budiantara, I. N., & Ratnasari, V. (2021). The Application of Mixed Smoothing Spline and Fourier Series Model in Nonparametric Regression. *Symmetry*, 13(11), 2094. <https://doi.org/10.3390/sym13112094>
- Matteson, D. S., & James, N. A. (2014). A Nonparametric Approach for Multiple Change Point Analysis of Multivariate Data. *Journal of the American Statistical Association*, 109(505), 334–345. <https://doi.org/10.1080/01621459.2013.849605>
- Mielke, J. (2015). An ultrasound study of Canadian French rhotic vowels with polar smoothing spline comparisons. *The Journal of the Acoustical Society of America*, 137(5), 2858–2869. <https://doi.org/10.1121/1.4919346>
- Nurchayani, H., Budiantara, N., & Zain, I. (2021). *Mathematics The Curve Estimation of Combined Truncated Spline and Fourier Series Estimators for Multiresponse Nonparametric Regression*. <https://doi.org/10.3390/math>
- Nyoman Budiantara, I., & Lestari, B. (2019). *Comparison of Smoothing and Truncated Spline Estimators in Estimating Blood Pressure Models*. www.ijicc.net
- Ouyang, L., Solberg, T., & Wang, J. (2011). Effects of the penalty on the penalized weighted least-squares image reconstruction for low-dose CBCT. *Physics in Medicine and Biology*, 56(17), 5535–5552. <https://doi.org/10.1088/0031-9155/56/17/006>

- Pajar, R. C., & Pustikaningsih, A. (2017). *Pengaruh Motivasi Investasi dan Pengetahuan Investasi terhadap Minat Investasi di Pasar Modal Pada Mahasiswa FE UNY*.
- Sifriyani, Haryatmi, Budiantara, I. N., & Gunardi. (2017). Geographically weighted regression with spline approach. *Far East Journal of Mathematical Sciences*, 101(6), 1183–1196. <https://doi.org/10.17654/MS101061183>
- Wadhvani, R., & Shukla, S. (2019). Analysis of parametric and non-parametric regression techniques to model the wind turbine power curve. *Wind Engineering*, 43(3), 225–232. <https://doi.org/10.1177/0309524X18780398>
- Wahyuni, D., Sari, I. K. P., Prasetio, A., & Berliando, M. (2023). *Environmental Sustainability Analysis in the Tourism Sector: A Perspective on Resilient and Sustainable Tourism Development in Indonesia*. <https://www.researchgate.net/publication/372405393>
- Wood, S. N., Pya, N., & Säfken, B. (2016). Smoothing Parameter and Model Selection for General Smooth Models. *Journal of the American Statistical Association*, 111(516), 1548–1563. <https://doi.org/10.1080/01621459.2016.1180986>
- Wylasmi, V. (2016). *Regresi Nonparametrik Dengan Pendekatan Smoothing Spline Pada Data Longitudinal (Aplikasi Pada Pertumbuhan Berat Badan Bayi Di Kota Malang)*.
- Xiang, D., & Wahba, G. (1995). *Testing the Generalized Linear Model Null Hypothesis versus "Smooth" Alternatives*.
- Xu, B., Luo, Y., Xu, R., & Chen, J. (2021). Exploring the driving forces of distributed energy resources in China: Using a semiparametric regression model. *Energy*, 236. <https://doi.org/10.1016/j.energy.2021.121452>
- Xu, P., Liu, J., & Shi, Y. (2023). Almost unbiased weighted least squares location estimation. *Journal of Geodesy*, 97(7), 68. <https://doi.org/10.1007/s00190-023-01742-0>