

Nonparametric Path Analysis on Consumer Satisfaction and Consumer Engagement in PT Pertamina

ADJI ACHMAD RINALDO FERNANDES, SOLIMUN, LAILIL MUFLIKHAH, AISYAH ALIFA,
ENDANG KRISNAWATI, NI MADE AYU ASTARI BADUNG,
ERLINDA CITRA LUCKI EFENDI
Brawijaya University
Jl. Veteran, Malang 65145 East Java
INDONESIA

Abstract: - The purpose of this research is to apply nonparametric path analysis on consumer satisfaction and consumer engagement of PT Pertamina. The results of the analysis are expected to be able to provide an estimate of the function in determining consumer satisfaction and consumer engagement of PT Pertamina. This study uses primary data involving five variables, namely Digitalization (X1), Consumer Needs (X2), Consumer Service (X3), Consumer Satisfaction (Y1), Consumer Engagement (Y3). Variable measurement technique is done by calculating the average score on the items. Sampling in this study used a purposive sampling technique with the respondent's criteria being company leaders. The result of this research is the estimation of nonparametric Path function using MARS approach on various interactions. The best estimate of the function of obedient behavior in paying credit is when it involves 3 variables, namely the digitization variable (X1), Consumer Needs (X2), Consumer Service (X3) with a value of generalized cross-validation The smallest (GCV) obtained is 0.2833. The originality of this research is that the variables used are the results of DNA analysis (Discourse Network Analysis), where the analysis extracts information from cyberspace which is then formed as the main issue and becomes a variable. In addition, there is no previous research that examines nonparametric path analysis on PT Pertamina's consumer satisfaction and engagement.

Key-Words: - Non-Paremetric Path, Consumer Satisfaction, Consumer Engagement, PT Pertamina, Function Estimation, Flexible Model

Received: March 18, 2021. Revised: November 14, 2021. Accepted: December 19, 2021. Published: January 9, 2022.

1 Introduction

Path analysis is an extension of multiple linear regression analysis that has more than one equation in the form of a system. In path analysis, the terms exogenous and endogenous variables are used. So that the model obtained is unbiased and can be used accurately, there are several assumptions that must be met. However, often the data found in the field do not meet the assumptions in the analysis to be used, namely the assumptions of normality and heteroscedasticity. Another thing that needs to be considered before starting the analysis is the form of the relationship between variables. There are times when the relationship between variables is linear, quadratic and even more than cubic. Therefore,

Path analysis based on nonparametric regression is a path/regression approach that is suitable for the pattern of the relationship between exogenous and endogenous which is not/not yet known [1]. The unknown pattern of the relationship between endogenous and exogenous can be estimated using the Spline function approach [2]. This study developed a nonparametric regression-based path

analysis using a spline path estimator. To get the regression curve estimation, Penalized Least Square (PLS) optimization or Penalized Weighted Least Square (PWLS) optimization is used [3].

Statistics has an important role in various fields, one of which is in the field of BUMN. PT Pertamina is a State-Owned Enterprise (BUMN) in charge of managing oil and gas mining in Indonesia. In general, PT Pertamina's market is divided into two, namely retail (Business to Consumer - B2C) and industrial (Business to Business - B2B). Until now, PT Pertamina continues to innovate to improve customer satisfaction, both B2C and B2B. The Industrial and Marine Business at PT Pertamina, for example, is currently developing a single point of contact with the concept of "Pertamina One Solution". This concept is expected to facilitate service to consumers, especially in the industrial market so that PT Pertamina can provide all the needs and desires of consumers.

Customer satisfaction is a response from consumers on the performance that has been given

in accordance with customer expectations. According to Wang in [4] customer satisfaction is a level where the needs, desires and expectations of customers can be met which will result in repeat purchases or continued loyalty. Customer satisfaction is the company's benchmark for how things are going in the future or even there are some things that must be changed because customers feel dissatisfied or disadvantaged which will affect customer engagement. Someone who returns to buy, and will tell others about his good experience with the product, it can be said that the customer is satisfied [5].

Based on the description that has been explained, it is very important to do research on Nonparametric Path analysis with a spline approach on PT Pertamina's consumer satisfaction and engagement. This research is very important because there are still weaknesses in the parametric path which requires the relationship between variables to be linear and the data pattern that must be known. So it is necessary to develop a nonparametric Path model that is able to estimate the function when the variable relationship is not linear and the data pattern is not known.

2 Literature Review

2.1 Nonparametric Path Analysis

Nonparametric path is an appropriate approach for the pattern of relationships between exogenous and endogenous, as well as fellow endogens whose form is unknown, or there is no complete past information about the pattern of relationships [1]. In the Nonparametric Path approach, the estimation form of the relationship pattern model is determined based on the existing data patterns. The unknown pattern of the relationship between endogenous and endogenous can be estimated using the Spline function approach or the Fourier Series [6]. The spline approach has high flexibility and is able to handle patterns of data relationships whose behavior changes at certain sub-intervals [1][7]. This has also been shown by [8] who compared the smoothing spline function with the kernel numerically, and [9] who compared the spline smoothing function with a numerically better kernel spline function. If given paired data following a simple Nonparametric Path model i.e:

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

$$y_{2i} = f_{1.2}(x_{1i}) + f_{2.2}(y_{2i}) + \varepsilon_{2i}; i = 1, 2, \dots, n$$

2.2 Spline Regression

Spline regression is a nonparametric regression method that aims to minimize diversity and estimate the behavior of data that tends to differ. The spline approach has the ability to overcome data patterns that show a sharp rise or fall with the help of knot points, and the resulting curve is relatively smooth. Knot points are joint points that indicate changes in data behavior patterns [1]. Spline is a piecewise polynomial of order q and has a continuous derivative with knots of order $(q-1)$ [10][11]. In a univariate spline with K knots, the base function is as follows:

In general, the spline approach states the relationship between p predictors and a single response with the model shown as follows:

$$y_i = f_1(x_{1i}) + f_2(x_{2i}) + \dots + f_p(x_{pi}) + \varepsilon; i = 1, 2, \dots, n$$

The equation can be written in the following form:

$$y_i = \sum_{j=1}^p f_j(x_{ji}) + \varepsilon_i, i = 1, 2, \dots, n$$

If f_j approached with a truncated spline function which has q order of polynomials and r knots then the equation can be written as:

$$y_i = \theta_0 + \sum_{j=1}^p \left(\sum_{l=1}^q \theta_{jl} x_{ji}^l + \sum_{h=1}^r y_{jh} (x_{ji} - t_{jh})^q \right) + \varepsilon_i;$$

$$i = 1, 2, \dots, n$$

$$(x_{ji} - t_{jh})^q = \begin{cases} (x_{ji} - t_{jh})^q, & x_{ji} \geq t_{jh} \\ 0, & x_{ji} < t_{jh} \end{cases}$$

Information:

y_i : response to observation to- i

x_{ji} : predictor to- j on the observation i

θ_0 : intercept

θ_{jl} : polynomial coefficient on predictor th j and the order of l

y_{jh} : truncated coefficient on the j th predictor and knots to- h

t_{jh} : value of knots on predictor j and knots to- h

r : number of knots

q : order of polynomials *truncated spline*

p : number of predictors

n : many observations

ε_i : random error on the i th observation

2.3 Multivariate Adaptive Regression Spline (MARS)

MARS is an approach to non-parametric regression. MARS was first introduced by [10]. The MARS model is focused on overcoming the problem of high dimensions and discontinuities in the data. MARS is able to estimate the contribution of the basis function to the response variable, by capturing not only the adaptive effect but also the interaction effect between predictors. MARS is an extension of the approach *Recursive Partition Regression* (RPR) which produces a discontinuous model at knots. According to [10] the MARS method can be used for as many predictor variables.

According to [12] things that need to be considered in building the MARS model are:

- Knot, is the value of the predictor variable when *slope* a regression line undergoes a change that can be defined as the end of one segment as well as the beginning of another segment. At each knot point, it is expected that there will be continuity of the basis function between *oneregion* with *region* other. The maximum observations between knots (MO) are 0, 1,2 and 3 observations [13].
- The basis function (B) is the interval between successive knots. In general, the selected basis function is a polynomial with a continuous derivative at every knot point. The maximum allowable basis function is 2 to 4 times as many predictor variables. There is a limitation of the base function used so that the resulting model is not too complex.
- Interaction is the result of cross multiplication between correlated variables. The maximum

number of interactions (MI) allowed is 1, 2 or 3. If $MI > 3$ the resulting model is increasingly complex and difficult to interpret.

The MARS model is generally expressed in the following equation:

$$y_i = a_0 + \sum_{m=1}^M a_m \prod_{k=1}^{K_m} \left[S_{km} \left(x_{v(k,m)} - t_{km} \right) \right] + \varepsilon_i$$

The estimation of the MARS model is written in the following equation:

$$\tilde{f}(x_i) = \hat{a}_0 + \sum_{m=1}^M \hat{a}_m \prod_{k=1}^{K_m} \left[S_{km} \left(x_{v(k,m)} - t_{km} \right) \right]$$

with,

$\tilde{f}(x_i)$: base function

a_0 : coefficient of the base function β_0

a_m : coefficient of the m th function basis

M : number of base functions

K_m : number of interactions on the m th basis function

S_{km} : is +1 if the knot is to the right of the subregion, and -1 if the knot is to the left of the subregion

$x_{v(k,m)}$: predictor variable to- v , select to k and subregion to m

t_{km} : knot value of predictor variable $X_{vi(k,m)}$

2.4 MARS Best Model Selection

According to [10], in determining the best MARS model, it is by looking at the smallest generalized cross-validation (GCV) value. The following is the GCV formula:

$$GCV(M) = \frac{\frac{1}{N} \sum_{i=1}^N [y_i - \hat{f}_M(x_i)]^2}{\left[1 - \frac{\hat{c}(M)}{N} \right]^2}$$

3 Methodology

This study uses primary data involving five variables, namely Digitalization (X1), Consumer Needs (X2), Consumer Service (X3), Consumer Satisfaction (Y1), Consumer Engagement (Y3). The research instrument used in primary data is a questionnaire with a Likert scale. Variable measurement technique is done by calculating the

average score on the items. Sampling in this study used a purposive sampling technique with the respondent's criteria being company leaders. The total respondents in this study were 200 respondents. This study uses nonparametric path analysis with the MARS method approach. The combination of the base functions used in this study is 2 to 4 times the exogenous variable with a minimum observation (MO) of 0.1,2,3 and a maximum interaction of 1,2,3.

4 Result

4.1 Linearity Check

Information about the pattern of relationships between variables is very necessary in statistical modeling because the information is used to determine the method used is a parametric regression approach or a nonparametric regression approach. In nonparametric regression modeling. This study uses data exploration using a scatter plot to check linearity.

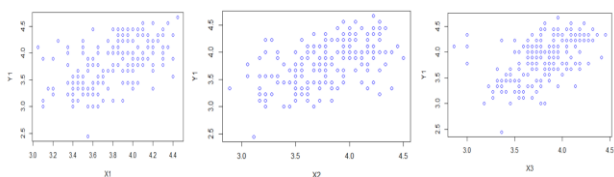


Fig. 1: Scatter Plot Exogenous variable with variable Y1

Figure 1 shows that the exogenous variable to the Y1 variable has unknown patterns of relationships. This indicates that the use of nonparametric paths with the MARS approach for model estimation is appropriate.

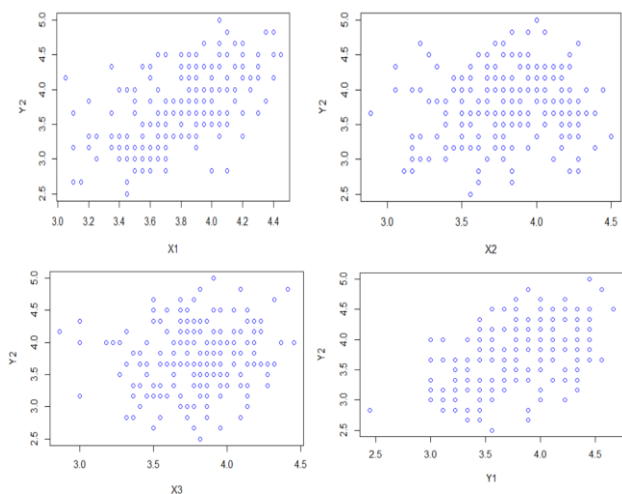


Fig. 2: Scatter Plot Exogenous variable with variable Y1

Figure 2 also shows that the exogenous variable to the Y2 variable has unknown patterns of relationships. This indicates that the use of nonparametric paths with the MARS approach for model estimation is appropriate.

4.2 Nonparametric Path Function Estimation with MARS Approach

Based on linearity examination, it was found that the relationship between variables was unknown as well as random data patterns so that the nonparametric path approach using the MARS method was appropriate. In the estimation of the Nonparametric Path function with the MARS approach, it is formed by a combination of the Basis Function, Maximum Interaction (MI) and Maximum Observation (MO) which is shown in Table 4.2. The basis function used is 2 to 4 times the number of predictor variables. There is a limitation on the basis function used, so that the resulting model is not complex. Because this study uses 3 exogenous variables, the basis functions used are 6, 9, 12. The maximum interactions used are 1,2 and 3. The maximum observations used are 0.1,2 and 3. Table 1.

Table 1. MARS Nonparametric Path GCV Values

Inter action	Function	Base Function	MO	GCV	Total GCV
1	Function 1	9	0	0.0963	0.2889
	Function 2	9	0	0.1925	
2	Function 1	12	0	0.1064	0.2835
	Function 2	12	0	0.1771	
3	Function 1	12	1	0.1010	0.2833
	Function 2	12	1	0.1823	

Based on Table 1, the best estimated function with 3 interactions with the smallest GCV value in the first function is 0.1010 and the second function has the smallest GCV value of 0.1823. The best estimates of the first and second functions are obtained when involving three exogenous variables, namely Digitization (X1), Consumer Needs (X2), Consumer Services (X3). Best first function estimation in model development Path nonparametric with MARS approach is shown in equations (1) and (2).

$$\hat{f}_{1i} = 3.176758 + 0.9711612 * \max(0, X1 - 3.45) - 0.5074453 * \max(0, X1 - 3.85) + 2.151026 * \max(0, 3.318182 - X3) + 0.7680472 * \max(0, X3 - 3.318182) - 0.9059963 * \max(0, X1 - 3.45) * \max(0, 3.5 - X3) - 1.012747 * \max(0, X1 - 3.45) * \max(0, X2 - 3.944444) * \max(0, X3 - 3.5) - 4.10493 * \max(0, X1 - 3.45) * \max(0, 3.944444 - X2) * \max(0, X3 - 3.5)$$

(1)

$$\hat{f}_{2i} = 3.285635 + 1.878941 * \max(0, X1 - 3.45) - 1.058779 * \max(0, X1 - 3.85) + 2.827457 * \max(0, 3.318182 - X3) - 0.09334885 * \max(0, X3 - 3.318182) - 7.504701 * \max(0, X1 - 3.45) * \max(0, 3.5 - X3) - 3.312039 * \max(0, X1 - 3.45) * \max(0, X2 - 3.944444) * \max(0, X3 - 3.5) - 1.016443 * \max(0, X1 - 3.45) * \max(0, 3.944444 - X2) * \max(0, X3 - 3.5)$$

(2)

In nonparametric path function estimation with the MARS approach, there are several statistics used to measure the magnitude of the error from the obtained function estimation. Figure 4.1 shows the magnitude of the generalization R^2 (GRSq) and R^2 (RSq) when a number of exogenous variables are used. The greater the value of generalized R^2 , the better the estimated function obtained.

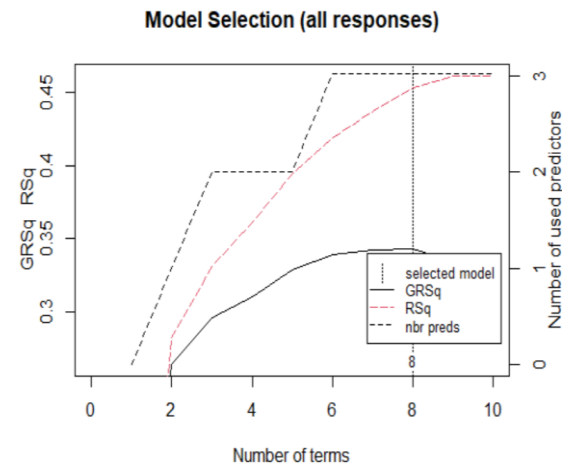


Fig. 3: MARS Nonparametric Path Model Selection Graph with 3 interactions

Based on Figure 3, it can be seen that the selected model is a model that uses 3 exogenous variables, namely the generalization R^2 is 0.3426 and R^2 is 0.4531. Figure 3 shows if the predictor variables used are more than three, then there is a decrease in generalization R^2 and if the predictor variables used are less than three then the values of generalization R^2 and R^2 are getting smaller.

The optimal basis functions of Path MARS equations (1) and (2) are taken using a backward process based on the minimum GCV value for each response partially. In selecting this optimal basis function, the minimum GCV value for response 1 is 0.1010 and for response 2 is 0.1823. The selection of the optimal basis function is important, because when it involves too many basic functions, the model obtained will be more complex (not parsimony).

5 Conclusion

Based on the results and discussions conducted in the study, it can be concluded that:

1. The estimation of the nonparametric path model on the compliance behavior data is when it involves 3 interactions. The smallest GCV value for the estimation of the first function is 0.1010 while the smallest GCV value for the second estimation function is 0.1823. In addition, the generalized value obtained R^2 as big as 0.3426 and R^2 of 0.4531 with a total GCV value of 0.2833. Model estimation with minimum GCV value is obtained when involving three exogenous variables for the first and second functions.

2. Variables that have a dominant influence on consumer satisfaction and engagement are the digitization variable (X1), Consumer Needs (X2), and Consumer Service (X3). This can be used as information for PT Pertamina in increasing consumer satisfaction and engagement. Consumer satisfaction and engagement will increase if Digitization, Consumer Needs, and Consumer Service.

References:

- [1] Budiantara, IN. Subanar and Z. Soejoeti. *Bulls. int. stats. Inst*, 51, 333-334. 1997
- [2] Wahba, G. Convergence rates of "thin plate" smoothing splines when the data are noisy. In *Smoothing Techniques for Curve Estimation* (pp. 233-245). Springer, Berlin, Heidelberg. 1979.
- [3] Fernandes, AAR, & Solimun, S. The Effect Of Correlation Between Responses In Bi-Response Nonparametric Regression Using Smoothing Spline For Longitudinal Data. *Communications In Applied Analysis*, 20(3). 2016.
- [4] Wang, Y., Wu, L., & Engel, B. Prediction of sewage treatment cost in rural regions with multivariate adaptive regression splines. *Waters*, 11(2), 195. 2019.
- [5] Wedarini, NMS. The effect of product quality on customer satisfaction and loyalty telkom flexi. *E-Journal of Management*, 2(5). 2012.
- [6] Fernandes, A. A. R., & Cahyoningtyas, R. A. Structural equation modelling on Latent Variables to identify farmers satisfaction in East Java using Mixed-Scale Data. (2021, May). In *Journal of Physics: Conference Series* (Vol. 1872, No. 1, p. 012022). IOP Publishing.
- [7] Fernandes, A. A. R., Solimun, F. U., Aryandani, A., Chairunissa, A., Alifa, A., Krisnawati, E., ... & Rasyidah12, F. L. N. Comparison Of Cluster Validity Index Using Integrated Cluster Analysis With Structural Equation Modelingthe War-Pls Approach. *Journal of Theoretical and Applied Information Technology*, 99(18). 2021.
- [8] Liang, L., Yang, F., Cook, WD, & Zhu, J. DEA models for supply chain efficiency evaluation. *Annals of operations research*, 145(1), 35-49. 2006.
- [9] Aydin, D. A comparison of the nonparametric regression models using smoothing spline and kernel regression. *World Academy of Science, Engineering and Technology*, 36, 253-257. 2007.
- [10] Friedman, JH. *Multivariate Adaptive Regression Splines*. The *Annals of Statistics*. Vol.19. Number 1. pp 1-14. 1991.
- [11] Fernandes, A.A.R., Solimun. Multi-responses model in patients suffering from decubitus wound using generalized penalized Spline. *International Journal of PharmTech Research*. 9(9), pp. 488-497. 2016.
- [12] Agwil, W., Rahmi, I., & Yozza, H.. Prediction of Forest Fire Area Based on Meteorological Data Using Multivariate Adaptive Regression Splines (MARS) Approach. *Journal of Mathematics*, 77-84. 2012.
- [13] Setiyawati, A. Customer Satisfaction Study To Achieving Customer Loyalty (Case Study on Consumers at Bangun Rejeki Semarang Stores) (Doctoral dissertation, Universitas Diponegoro). 2009.

**Creative Commons Attribution License 4.0
(Attribution 4.0 International, CC BY 4.0)**

This article is published under the terms of the Creative Commons Attribution License 4.0

https://creativecommons.org/licenses/by/4.0/deed.en_US