Development of Nonparametric Path Model using Multivariate Adaptive Regression Spline (MARS) Method in Compliance Behavior of Paying Credit in Bank

ENDANG KRISNAWATI, ADJI ACHMAD RINALDO FERNANDES, SOLIMUN
Faculty of Math and Science
University of Brawijaya
St. Veteran, Ketawanggede, Lowokwaru, Malang, East Java 65145
INDONESIA

Abstract: - The purpose of this study is to develop a Non-parametric Path with the MARS (Multivariate Adaptive Regression Spline) approach which is applied to the behavior of paying credit compliance at Bank. prospective debtor by a Bank. The data used in this study is primary data using a research instrument in the form of a questionnaire. There are 7 variables, namely 5 exogenous variables in the form of 5C variables (Character (X1), Capacity (X2), Capital (X3), Collateral (X4), Condition of Economy (X5)), and two endogenous variables, namely Punctual Payment (Y1), Obedient Paying Behavior (Y2). Variable measurement technique is done by calculating the average score on the items. Sampling in this study used a purposive sampling technique with the criteria of respondents in the study were mortgage debtors (House Ownership Credit) at Bank X. Respondents obtained in this study were 100 respondents. The analysis used is nonparametric path with Multivariate Adaptive Regression Spline (MARS) approach. 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 4 variables, namely Character (X1), Capacity (X2), *Conditions of economy* (X5)*,* and On time pay (Y1) with a value of *generalized cross-validation* The smallest (GCV) obtained is 0.2496. The originality of this research is the development of a nonparametric path with the MARS approach that is able to capture interactions between existing variables and is also able to handle the limitations of the truncated spline to determine the position and number of knot points used when involving many predictor variables. There has been no previous research that has examined the development of a nonparametric path with the MARS approach.

Key-Words: Non-parametric Path, MARS, Credit-Paying Compliance Behavior, 5C Principle, Bank, Function Estimation

1 Introduction

Regression analysis is one of the statistical methods used to determine the relationship between predictor variables and response variables. There are three approaches in completing regression analysis, namely parametric regression, nonparametric regression, and semiparametric regression (a combination of parametric regression and non-parametric regression). Parametric regression is a regression in which the data pattern and regression curve are known, while nonparametric regression is a regression in which the data pattern and regression curve are unknown.

Analysis that involves more than one response variable and has a complex relationship requires a series of regression equations. One extension of regression analysis is path analysis. Path analysis is an analysis used to determine the relationship between response variables and predictor variables, where in path analysis there is the possibility of using intervening variables [1].

The limited information and the form of the function pattern of the relationship between the response variables and the predictor variables are considerations for using a nonparametric regression approach. The ability of nonparametric regression is to be able to find the shape of the unknown regression curve pattern by using parameters in each type of nonparametric regression which causes the curve in nonparametric regression to be flexible.

In Path analysis, problems often occur regarding the form of relationships and unknown data patterns, so the Parametric Path approach cannot be carried out. Therefore, other alternatives are needed that are able to handle this problem. 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 ([2], [3], [4])

There are many studies that discuss nonparametric regression with various developments based on the type of response variable, namely continuous response variables and categorical response variables. The continuous response variable has the nature of the data used as the response variable having an interval or ratio measurement scale. While the categorical response variable has the property that the data used as the response variable has a nominal or ordinal measurement scale.

One nonparametric regression approach is the Multivariate Adaptive Regression Spline (MARS). [5] introduced a nonparametric regression method, namely the Multivariate Adaptive Regression Spline (MARS). MARS is an approach to multivariate nonparametric regression analysis between the response variable and several predictor variables that is useful for overcoming the problem of high-dimensional data, namely data that has a predictor variable of $3 \leq r \leq 20$ and the data sample is $50 \leq n \leq 1000$. According to [5] the MARS method has the ability to capture additive and interaction effects on the predictor variables used, which is an advantage of MARS which is able to produce better estimates because it can not only capture additive effects. In the MARS method, the algorithm used is stepwise, which consists of forward stepwise and backward stepwise.

Statistics has an important role in various fields, one of which is Banking.Banks are financial institutions that provide services to channel funds or provide loans (credit) and collect funds from the wider community in the form of savings. Banks as financial institutions guarantee good finances so that debtors are assisted in financial matters. The bank program that is currently in demand by debtors is the Home Ownership Credit (KPR). This mortgage program is in demand by families who are less in terms of financing. Many banks offer a variety of mortgages, but of course the problem faced by decision makers is determining prospective debtors who are eligible for a mortgage [6]. For the sake of smooth credit activities between the bank and the debtor, the bank needs to assess and determine the debtor first before making a decision to accept or reject the applicant's mortgage request. This is done considering the risk of uncollectible credit is quite large. So that a debtor must meet the criteria that have been determined by the bank to be able to get credit. In this case the bank is required to be able to

make decisions. [7] in her research states that there are several principles that must be considered in making decisions to accept prospective debtors, including Collateral, Capacity, Capital, Character, and Condition of Economy.

Based on the description that has been explained, it is very important to do about the development of the Path model, especially in the case of unknown data patterns. So, in this study, we want to develop a Nonparametric Path model with the MARS method on various interactions that are applied to paying obedient behavior. 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. The originality of this research is the development of a nonparametric path with the MARS approach that is able to capture the interactions between existing variables.

2 Literature Review

2.1 Nonparametric Path Analysis
Nonparametric Path Analysis is a development of Nonparametric Regression Analysis. Nonparametric Path Analysis is able to overcome the weakness of Parametric Path analysis, namely the pattern of relationships between variables that are not known [8]. If given paired data (x_{1i}, y_{1i}, y_{2i}) $i = 1, 2, \ldots, n$; following a simple Nonparametric Path model:

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

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

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

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 ([2], [3], [4]). In the Nonparametric Path approach, the estimation form of the relationship pattern model is determined based on the existing data patterns [9].

The unknown pattern of the relationship between endogenous and endogenous can be estimated using the Spline function approach or the Fourier Series. The spline approach has high flexibility and is able to handle patterns of data relationships whose behavior changes at certain sub-intervals [2]. This has also been shown by [10] who compared the smoothing spline function with the kernel numerically.

2.2 Multivariate Adaptive Regression Spline (MARS)

MARS is an approach to non-parametric regression. MARS was first introduced by [5]. 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 [5] the MARS method can be used for

(X_v) as many predictor variables, where v there are

many predictor variables $3 \leq v \leq 20$ and the

sample size of data used is $50 \leq n \leq 1000$.

According to [11] Things that need to be considered in building the MARS model are:

a. 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.

b. 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.

c. Interaction is the result of cross multiplication between correlated variables. If $MI > 3$ The resulting model is increasingly complex and difficult to interpret.

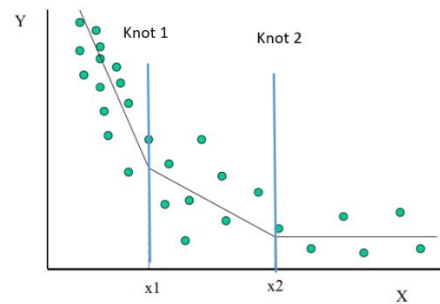The following is an illustration of the MARS model.

Fig. 1: MARS model illustration [12]

According to [13] based on the following nonparametric regression function,

$$y_i = f(x_i) + \varepsilon_i \text{ , for } i = 1,2,\dots,n \tag{3}$$

Then the MARS model is generally expressed in the following equation

$$y_i = a_0 + \sum_{m=1}^{M} a_m \prod_{k=1}^{Km} \left[S_{km}\left(x_{v(k,m)} - t_{km}\right)\right]_+ + \varepsilon_i \tag{4}$$

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

$$\hat{f}(x_i) = \hat{a}_0 + \sum_{m=1}^{M} \hat{a}_m \prod_{k=1}^{Km} \left[S_{km}\left(x_{v(k,m)} - t_{km}\right)\right]_+ \tag{5}$$

2.3 MARS Best Model Selection

According to [5], in determining the best MARS model, it is by looking at the smallest generalized cross-validation (GCV) value. The following is the GCV formula:

$$GCV(M) = \frac{\frac{1}{N} \sum_{i=1}^{N} [y_i - \hat{f}_M(x_i)]^2}{\left[1 - \frac{\hat{C}(M)}{N}\right]^2} \tag{6}$$

2.4 5C Prinsip Principle

For the sake of smooth credit activities between the bank and the debtor, the bank needs to assess and determine the debtor first before making a decision to accept or reject the applicant's mortgage request, this is done considering the risk of uncollectible credit is quite large. So that a debtor must meet the criteria that have been determined by the bank to be able to get credit. In this case the bank is required to be able to make decisions. [7] in her research states that there are several 5C principles that must be

considered in making decisions to accept prospective debtors, namely as follows:

a) *Character* This is the nature of the customer that is related to his willingness to pay credit installments later.
b) *Capital* is a capital analysis that aims to see the ability of the customer to bear the financing burden and the risk burden that may be experienced.
c) *Capacity* look at the customer's ability to pay installments and other needs, namely the customer's job, length of work, income.
d) *Collateral* which are goods given by customers as collateral.
e) *Condition of Economy* are political, socio-economic and cultural conditions and situations that play a role in determining economic conditions that affect the smooth running of a prospective debtor's business.

3 Methodology

The data used in this study is primary data using a research instrument in the form of a questionnaire. There are 7 variables, namely with 5 exogenous variables, namely the 5C variable (Character (X1), Capacity (X2), Capital (X3), Collateral (X4), Condition of Economy (X5)), and two endogenous variables, namely Punctual Payment (Y1), Obedient Paying Behavior (Y2). The research instrument used has gone through the testing process, that is, it has been declared valid and reliable. Variable measurement technique is done by calculating the average score on the existing items. Sampling in this study used a purposive sampling technique, namely a sampling technique in which researchers determine certain criteria in sampling. The criteria for respondents in this study were mortgage debtors at Bank with the respondents obtained were 100 debtors. 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 Lemma and Theorem

Lemma 1

If given paired data $(x_{1i}, x_{2i}, \ldots, x_{pi}, y_{1i})$ and $(x_{1i}, x_{2i}, \ldots, x_{pi}, y_{2i})$ $i = 1, 2, \ldots, n$. If the relationship between the predictor and response is expressed in a regression function whose form is unknown and can be approximated by a nonparametric biresponse model, then the relationship can be expressed in the following form:

$$y = B(x, t)\alpha + \varepsilon$$

Where,

$$y = (y_1', y_2')' = (y_{11}, y_{12}, \ldots, y_{1n}, y_{21}, y_{22}, \ldots, y_{2n})'$$

$$B_{(x,t)} = \begin{pmatrix} B_{1(x,t)} & 0 \\ 0 & B_{2(x,t)} \end{pmatrix}$$

With:

$$B_{l(x,t)} = \begin{pmatrix} B_{1(x,t)} & 0 \\ 0 & B_{2(x,t)} \end{pmatrix}, l = 1,2$$

$$\alpha = (\alpha_1', \alpha_2')' = (\alpha_{11}, \alpha_{12}, \ldots, \alpha_{1M_1}, \alpha_{21}, \alpha_{22}, \ldots, \alpha_{2M_2})'$$

$$\varepsilon = (\varepsilon_1', \varepsilon_2')' = (\varepsilon_{11}, \varepsilon_{12}, \ldots, \varepsilon_{1M_1}, \varepsilon_{21}, \varepsilon_{22}, \ldots, \varepsilon_{2M_2})'$$

with $\varepsilon \sim N(0, W)$

Proof:
The function and is a biresponse nonparametric function which is assumed to be of unknown form. Because these functions are approximated by MARS functions such as, the biresponse MARS regression function is expressed as follows:

$$f_1 = a_{10} + \sum_{m_1=1}^{M_1} a_{1m_1} \prod_{k_1=1}^{Km_1} [S_{k_1 m_1}(x_{v(k_1, m_1)i} - t_{k_1 m_1})]_+ + \varepsilon_i$$

With

$$x_{v(k_1, m_1)} \in \{x_j\}_{j=1}^p, \ t_{k_1 m_1} \in \{x_{v(k_1, m_1)}\}_{i=1}^n, \ m_1 = 1, 2, \ldots, M_1$$

If   then

$$S_{k_1 m_1} = +1 + (x_{v(k_1 m_1)i} - t_{k_1 m_1})_+ = \begin{cases} x_{v(k_1 m_1)i} - t_{k_1 m_1} \text{ ,jika } x_{v(k_1 m_1)i} > t_{k_1 m_1} \\ 0, \text{ other} \end{cases}$$

If   then

$$S_{k_1 m_1} = -1 - (x_{v(k_1 m_1)i} - t_{k_1 m_1})_+ = \begin{cases} x_{v(k_1 m_1)i} - t_{k_1 m_1} \text{ ,jika } t_{k_1 m_1} > x_{v(k_1 m_1)i} \\ 0, \text{ selainya} \end{cases}$$

And the function is as follows f_2

$$f_2 = a_{20} + \sum_{m_2=1}^{M_2} a_{2m_2} \prod_{k_2=1}^{Km_2} [S_{k_2 m_2}(x_{v(k_2 m_2)i} - t_{k_2 m_2})]_+ + \varepsilon_i$$

With

$$x_{v(k_2 m_2)} \in \{x_j\}_{j=1}^p \text{ ,} t_{k_2 m_2} \in \{x_{v(k_2 m_2)}\}_{i=1}^n, m_2 = 1,2,...,M_2,$$

If   then

$$S_{k_2 m_2} = +1 + (x_{v(k_2 m_2)i} - t_{k_2 m_2})_+ = \begin{cases} x_{v(k_2 m_2)i} - t_{k_2 m_2} \text{ ,jika } x_{v(k_2 m_2)i} > t_{k_2 m_2} \\ 0, \text{ selainya} \end{cases}$$

If   then

$$S_{k_2 m_2} = -1 - (x_{v(k_2 m_2)i} - t_{k_2 m_2})_+ = \begin{cases} x_{v(k_2 m_2)i} - t_{k_2 m_2} \text{ ,jika } t_{k_2 m_2} > x_{v(k_2 m_2)i} \\ 0, \text{ selainya} \end{cases}$$

The biresponse MARS function above can be simplified to the function f_1 and f_2 as follows:

$$f_1 = a_{10} + \sum_{m_1=1}^{M_1} a_{1m_1} B_{m_{1i}}(x, t)$$

$$f_2 = a_{20} + \sum_{m_2=1}^{M_2} a_{2m_2} B_{m_{2i}}(x, t)$$

With $B_{m_{1i}}(x, t)$ and is the base function for $B_{m_{2i}}(x, t) f_1$ and in the MARS model which is expressed as follows:

$$B_{m_{1i}}(x, t) = \prod_{k_1=1}^{Km_1} [S_{k_1 m_1}(x_{v(k_1 m_1)i} - t_{k_1 m_1})]_+$$

$$B_{m_{2i}}(x, t) = \prod_{k_2=1}^{Km_2} [S_{k_2 m_2}(x_{v(k_2 m_2)i} - t_{k_2 m_2})]_+$$

If it is described based on the number of observations (i=1,2,3, …, n), the bi-response MARS function is described as follows:

$$f_1(x_{11}, x_{21}, ..., x_{p1}) = a_{10} + \sum_{m_1=1}^{M_1} a_{1m_1} \prod_{k_1=1}^{Km_1} [S_{k_1 m_1}(x_{v(k_1 m_1)1} - t_{k_1 m_1})]_+$$

$$f_1(x_{12}, x_{22}, ..., x_{p2}) = a_{10} + \sum_{m_1=1}^{M_1} a_{1m_1} \prod_{k_1=1}^{Km_1} [S_{k_1 m_1}(x_{v(k_1 m_1)2} - t_{k_1 m_1})]_+$$

$$\vdots$$

$$f_1(x_{1n}, x_{2n}, ..., x_{pn}) = a_{10} + \sum_{m_1=1}^{M_1} a_{1m_1} \prod_{k_1=1}^{Km_1} [S_{k_1 m_1}(x_{v(k_1 m_1)n} - t_{k_1 m_1})]_+$$

and function f_2 is as follows:

$$f_2(x_{11}, x_{21}, ..., x_{p1}) = a_{20} + \sum_{m_2=1}^{M_2} a_{2m_2} \prod_{k_2=1}^{Km_2} [S_{k_2 m_2}(x_{v(k_2 m_2)1} - t_{k_2 m_2})]_+$$

$$f_2(x_{12}, x_{22}, ..., x_{p2}) = a_{20} + \sum_{m_2=1}^{M_2} a_{2m_2} \prod_{k_2=1}^{Km_2} [S_{k_2 m_2}(x_{v(k_2 m_2)2} - t_{k_2 m_2})]_+$$

$$\vdots$$

$$f_2(x_{1n}, x_{2n}, \ldots, x_{pn}) = a_{20} + \sum_{m_2=1}^{M_2} a_{2m_2} \prod_{k_2=1}^{Km_2} [S_{k_2m_2}(x_{v(k_2m_2)n} - t_{k_2m_2})]_+$$

Based on the base function f_1 ($m_1 = 1,2,3, \ldots, M_1$ and then the function is as follows: $m_2 = 1,2,3, \ldots, M_2$)

$$f_1(x_{11}, x_{21}, \ldots, x_{p1}) = a_{10} + a_{11} \prod_{k_1=1}^{K_1} [S_{k_11}(x_{v(k_1m_1)1} - t_{k_1m_1})]_+ + \cdots + a_{1M_1} \prod_{k_1=1}^{K_1} [S_{k_1M_1}(x_{v(k_1m_1)1} - t_{k_1m_1})]_+$$

$$f_1(x_{11}, x_{21}, \ldots, x_{p1}) = a_{10} + a_{11} \prod_{k_1=1}^{K_1} [S_{k_11}(x_{v(k_1m_1)2} - t_{k_1m_1})]_+ + \cdots + a_{1M_1} \prod_{k_1=1}^{K_1} [S_{k_1M_1}(x_{v(k_1m_1)2} - t_{k_1m_1})]_+$$

⋮

$$f_1(x_{11}, x_{21}, \ldots, x_{p1}) = a_{10} + a_{11} \prod_{k_1=1}^{K_1} [S_{k_11}(x_{v(k_1m_1)2} - t_{k_1m_1})]_+ + \cdots + a_{1M_1} \prod_{k_1=1}^{K_1} [S_{k_1M_1}(x_{v(k_1m_1)2} - t_{k_1m_1})]_+$$

and regression function f_2 as follows:

$$f_2(x_{11}, x_{21}, \ldots, x_{p1}) = a_{20} + a_{21} \prod_{k_2=1}^{K_2} [S_{k_21}(x_{v(k_2m_2)1} - t_{k_2m_2})]_+ + \cdots + a_2M_2 \prod_{k_2=1}^{KM_2} [S_{k_21}(x_{v(k_2m_2)1} - t_{k_2m_2})]_+$$

$$f_2(x_{12}, x_{22}, \ldots, x_{p2}) = a_{20} + a_{21} \prod_{k_2=1}^{K_2} [S_{k_21}(x_{v(k_2m_2)1} - t_{k_2m_2})]_+ + \cdots + a_2M_2 \prod_{k_2=1}^{KM_2} [S_{k_21}(x_{v(k_2m_2)1} - t_{k_2m_2})]_+$$

⋮

$$f_2(x_{1n}, x_{2n}, \ldots, x_{pn}) = a_{20} + a_{21} \prod_{k_2=1}^{K_2} [S_{k_21}(x_{v(k_2m_2)1} - t_{k_2m_2})]_+ + \cdots + a_2M_2 \prod_{k_2=1}^{KM_2} [S_{k_21}(x_{v(k_2m_2)1} - t_{k_2m_2})]_+$$

The nonparametric MARS biresponse function in the above equation can be written as follows:

$$f(x) = B(x, t)\alpha$$

Where is the biresponse MARS function containing and which is expressed in the following vector form:

$$f(x) f_1(x) f_2(x)$$

$$f(x) = \left(\frac{f_1(x)}{f_2(x)}\right) = \begin{pmatrix} f_1(x_{11}, x_{21}, \ldots, x_{p1}) \\ f_1(x_{11}, x_{21}, \ldots, x_{p1}) \\ \vdots \\ f_1(x_{11}, x_{21}, \ldots, x_{p1}) \\ \hline f_2(x_{11}, x_{21}, \ldots, x_{p1}) \\ f_2(x_{12}, x_{22}, \ldots, x_{p2}) \\ \vdots \\ f_2(x_{1n}, x_{2n}, \ldots, x_{pn}) \end{pmatrix}$$

$B(x, t)$ is a diagonal matrix containing the basis function for $B_1(x, t)$ and base function for $B_2(x, t)$ as follows:

$$f_1(x) B_2(x, t) f_2(x)$$

$$B_{l(x,t)} = \begin{pmatrix} B_{1(x,t)} & 0 \\ 0 & B_{2(x,t)} \end{pmatrix}, l = 1,2$$

With,

$$B_{1(x,t)} = \begin{pmatrix} 1 & \prod_{k_1=1}^{K_1}[S_{k_11}(x_{v(k_11)1} - t_{k_11})]_+ & \cdots & \prod_{k_1=1}^{KM_1}[S_{k_1M_1}(x_{v(k_1M_1)1} - t_{k_1M_1})]_+ \\ 1 & \prod_{k_1=1}^{K_1}[S_{k_11}(x_{v(k_11)2} - t_{k_11})]_+ & \cdots & \prod_{k_1=1}^{KM_1}[S_{k_1M_1}(x_{v(k_1M_1)2} - t_{k_1M_1})]_+ \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \prod_{k_1=1}^{K_1}[S_{k_11}(x_{v(k_11)n} - t_{k_11})]_+ & \cdots & \prod_{k_1=1}^{KM_1}[S_{k_1M_1}(x_{v(k_1M_1)n} - t_{k_1M_1})]_+ \end{pmatrix}$$

$$B_{2(x,t)} = \begin{pmatrix} 1 & \prod_{k_2=1}^{K_1}[S_{k_21}(x_{v(k_21)1} - t_{k_21})]_+ & \cdots & \prod_{k_2=1}^{KM_2}[S_{k_2M_2}(x_{v(k_2M_2)1} - t_{k_2M_2})]_+ \\ 1 & \prod_{k_2=1}^{K_1}[S_{k_21}(x_{v(k_21)2} - t_{k_21})]_+ & \cdots & \prod_{k_2=1}^{KM_2}[S_{k_2M_2}(x_{v(k_2M_2)n} - t_{k_2M_2})]_+ \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \prod_{k_2=1}^{K_1}[S_{k_21}(x_{v(k_21)n} - t_{k_21})]_+ & \cdots & \prod_{k_2=1}^{KM_2}[S_{k_2M_2}(x_{v(k_2M_2)n} - t_{k_2M_2})]_+ \end{pmatrix}$$

While α is a biresponse MARS parameter containing α_1 for $B_{1(x,t)}$ and parameters α_2 $B_{2(x,t)}$ which is expressed in vector form as follows:

$$\underset{\sim}{\alpha} = \begin{pmatrix} \underline{\underline{\alpha_1}} \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} \alpha_{10} \\ \alpha_{11} \\ \vdots \\ \underline{\underline{\alpha_{1M_1}}} \\ \alpha_{20} \\ \alpha_{21} \\ \vdots \\ \alpha_{2M_2} \end{pmatrix}$$

The form of the birespond MARS equation is then obtained as follows:

$$\underset{\sim}{y} = f(x) + \underset{\sim}{\varepsilon}$$

$$\underset{\sim}{y} = B(\underset{\sim}{x}, t)\underset{\sim}{\alpha} + \underset{\sim}{\varepsilon}$$

Random error $\underset{\sim}{\varepsilon}$ in the MARS equation, the

biresponse is normally distributed with mean $\mathbf{0}$ and the error covariance matrix W. The W matrix serves as a weight matrix that accommodates the correlation between responses when estimating the model parameters. To get an estimate of the model parameters, the matrix W is assumed to have a known value (fixed variable). The estimation results of the continuous biresponse MARS regression function are described in Theorem 1.

Theorem 1
If the equation of the biresponsive MARS Nonparametric Path function is given by Lemma 1

assuming $\varepsilon \sim N(0, W)$, where the matrix W (matrix

of order $2n \times 2n$) is assumed to be known, then the estimation of the biresponsive MARS

Nonparametric Path function $f(x)$ is as follows:

$$\hat{f}(x) = H(.)\underset{\sim}{y}$$

Where:

$$H(.) = B(\underset{\sim}{x}, t)(B'(\underset{\sim}{x}, t)W^{-1}B(\underset{\sim}{x}, t))^{-1}B(\underset{\sim}{x}, t)W^{-1}$$

4.2 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.
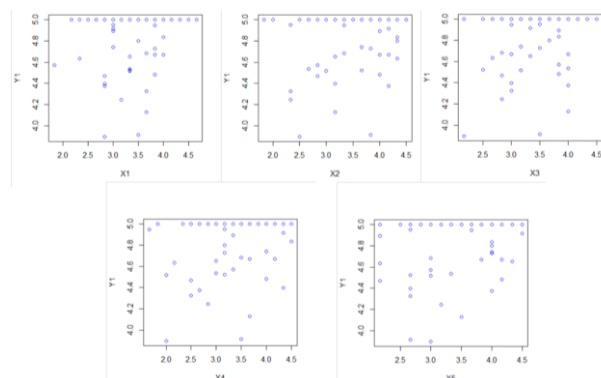

Fig. 2: Scatter Plot Exogenous variable with variable Y1

Figure 2 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.
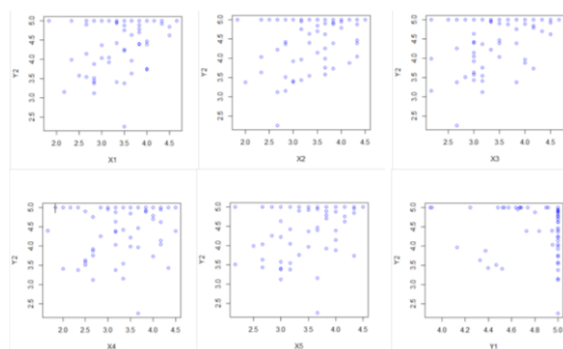

Fig. 3: Scatter Plot of Exogenous variable with variable Y2 variable

Figure 3 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.

4.3 Nonparametric Path Function Estimation with MARS Approach
In linearity testing, it was found that the relationship between the Condition of economy variable (X5) and Punctuality to pay (Y1) and the data pattern was unknown 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 4.1. 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 5 exogenous variables, the basis functions used are 10, 15, 20. The maximum interactions used are 1,2 and 3. The maximum observations used are 0.1,2 and 3.

Table 1. MARS Nonparametric Path GCV Values

Interaction	Function	Base Function	MO	GCV
1	Function 1	20	0	0.3005
	Function 2	20	3	0.2610
2	Function 1	20	0	0.2526
	Function 2	20	0	0.2562
3	Function 1	20	0	0.2526
	Function 2	20	3	0.2496

Based on Table 1, the best estimated function with 3 interactions with the smallest GCV value in the first function is 0.2526 and the second function has the smallest GCV value is 0.2496. The best estimate of the first function is obtained when involving three exogenous variables, namely *Capacity* (X2), *Collateral* (X4), and Conditions of economy (X5). While the second best function estimation is when it involvesfour variables namely *character* (X1), *Capacity* (X2), *Conditions of economy* (X5), and On time pay (Y1). Best first function estimation in model development *Path* nonparametric with MARS approach is shown in equations (4.1) and (4.2).

$$\hat{f}_{1i} = 3.4554 - 1.2462 * \max(0, 3.8333 - X_{2i}) - 0.9376 * \max(0, X_{2i} - 3.8333) - 0.5703 * \max(0, 4 - X_{4i}) + 2.2989 * \max(0, X_{5i} - 3) - 2.3416 * \max(0, X_{5i} - 3.5) + 0.9424 * \max(0, 3.3333 - X_{2i}) * \max(4 - X_{4i}) - 12.2209 * \max(0, 3.1667 - X_{2i}) * \max(X_{5i} - 3.5) + 0.6765 * \max(0, 3.6667 - X_{2i}) * \max(4 - X_{5i}) - 0.9861 * \max(0, X_{4i} - 3.3333) * \max(X_{5i} - 3.5)$$

$$(7)$$

$$\hat{f}_{2i} = 4.0975 - 0.8006 * \max(0, 4 - Y_{1i}) - 0.7851 * \max(0, 4.5 - X_3) * \max(0, 3 - X_{5i}) - 0.4519 * \max(0, 4 - X_{1i}) * \max(0, 4.5 - X_{3i}) * \max(0, X_{5i} - 3) + 0.7771 * \max(0, 3.3333 - X_{1i}) * \max(0, 3.5 - X_3) * \max(0, 4 - Y_1)$$

$$(8)$$

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 4.2 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.
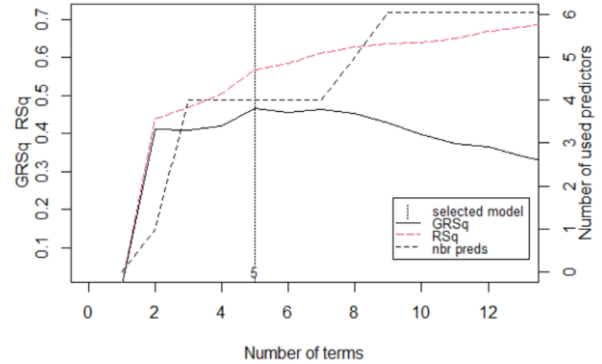

Fig. 4: Nonparametric Path MARS Model Selection Graph with 3 interactions

Based on Figure 4, it can be seen that the selected model is a model that uses 4 exogenous variables, namely the generalization R^2 value of 0.4652 and R^2 of 0.5678. Figure 4.2 shows that if the predictor variables used are more than four, then there is a decrease in generalization R^2 and if the predictor variables used are less than four then the values of generalization R^2 and R^2 are getting smaller.

The best estimation visualization on the nonparametric MARS path is shown in Figure 5 as follows.

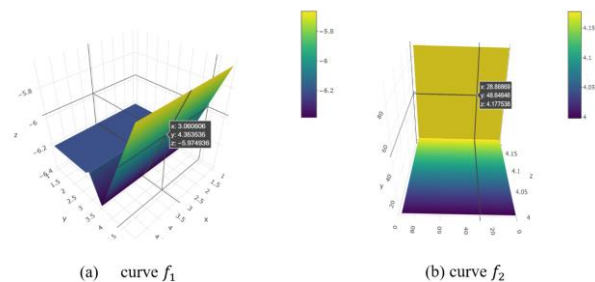(a) curve f_1 (b) curve f_2

Fig. 5: MARS Path Function Estimation Graph

The optimal basis functions of Path MARS equations (7) and (8) 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 0.2526 and for response 2 is 0.2496. 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). Complex models will cause overfitting cases, namely good models applied to training data but not well applied to testing data.

5 Conclusion

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

1. The estimation of the MARS nonparametric path model on the compliance behavior data is when it involves 3 interactions. The smallest GCV value for the first function estimate is 0.2526, while the smallest GCV value for the second function estimate is 0.2496. In addition, the generalized value obtained R^2 as big as 0.4652 and R2 of 0.5678 with a total GCV value of. Model estimation with minimum GCV value is obtained when involving three exogenous variables for the first function and 4 variables for the second function.

2. Variables that have a dominant influence on obedient behavior to pay are variable character (X1), Capacity (X2), *Conditions of economy*(X5)*,* and On time pay (Y1). This can be used as information for the Bank in assessing and determining the debtor first before making a decision to accept or reject the applicant's KPR request for smooth credit activities between the bank and the debtor. Prospective debtors will be more obedient in paying if (1) their character or personality is getting better, (2) they are able to manage their business well, (3) they are in political, socio-economic and cultural conditions and situations that affect the smooth running of the prospective debtor's business, (4) there is a guarantee provided by the prospective debtor.

3. This research will be useful for banks to make decisions when accepting prospective creditors, so that credit risk can be minimized and increase knowledge about the process of modeling obedient paying behavior to customers. Furthermore this research Can be used as a reference for further research regarding the development of modeling behavior in paying credit.

References:

[1] Kadir, T. Lung cancer prediction using machine learning and advanced imaging techniques. *Translational lung cancer research*, Vol.7, No.3, 2018, pp.304.

[2] Eubank, RL. *Nonparametric Regression and Spline Smoothing 2nd edition*, Marcel Dekker, Inc. New York, 1999.

[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*, Vol.20, No.3, 2016.

[4] Lestari B, Budiantara IN, Chamidah N. Smoothing parameter selection method for multiresponse nonparametric regression model using smoothing spline and Kernel estimators approaches. *InJournal of Physics: Conference Series*, Vol. 1397, No. 1, 2019, pp12-64.

[5] Friedmen, JH. Multivariate Adaptive Regression Splines. *The Annals of Statistics*. Vol.19, Number 1, 1991, pp 1-14.

[6] Al-Hafiz. Decision Support System for Determining Home Ownership Loans Implementing Multi-Objective Optimization On The Basis Of Ratio Analysis (Moora). *KOMIK (National Conference on Information and Computer Technology)*, Vol.1, No.1, 2017.

[7] Nuraeni, F. *Implementation of Algorithm C 4.5 and Principle 6C for Classification of Credit Quality in Financing Companies*. In Technopex 2018.

[8] Fernandes AA, Astuti AB, Amaliana L, Yanti I, Arisoesilaningsih E, Isaskar R. Smoothing Spline Nonparametric Path: Application for Green Product and Green Marketing Strategy towards Green Product Purchasing Intention. InIOP Conference Series: Earth and Environmental Science 2019 Feb 1 (Vol. 239, No. 1, p. 01 2018). *IOP Publishing*.

[9] Fernandes AA, Solimun M, Aryandani A, Chairunissa A, Krisnawati E, Al Rohimi A, Rasyidah FL. Development of Path Analysis Based on Nonparametric Regression on Vegetative Growth and Porang's Production Modeling. Journal of Hunan University Natural Sciences, Vol.48, No.4, 2021.

[10] Liang, L., Yang, F., Cook, WD, & Zhu, J. DEA models for supply chain efficiency evaluation. *Annals of operations research*, Vol. 145, No.1, 2006, pp. 35-49.

[11] Agwil, W., HG, I. R., & Yozza, H. Prediksi Luas Area Kebakaran Hutan Berdasarkan Data Meteorologi dengan Menggunakan Pendekatan Multivariate Adaptive Regression Splines (MARS). *Jurnal Matematika UNAND*, Vol.1, No.1, 2012, pp. 77-84.

Endang Krisnawati,
Adji Achmad Rinaldo Fernandes, Solimun

[12] Briand, LC, Freimut, B., & Vollei, F. Using multiple adaptive regression splines to support decision making in code inspections. *Journal of Systems and Software*, Vol.73, No.2, 2014, pp.205-217.

[13] Kishartini K, Safitri D, Ispriyanti D. Multivariate Adaptive Regression Splines (MARS) Untuk Klasifikasi Status Kerja Di Kabupaten Demak. *Jurnal Gaussian*. Vol.3, No.4, 2014, pp.711.