**Nonlinear Principal Component Analysis in Path Analysis with Mixed Data Formative Indicator Models**

**Rindu Hardianti1, Solimun2, Nurjannah3**

1,2,3Department of Statistics, Brawijaya University, Indonesia

rinduhardianti@student.ub.ac.id1, solimun@ub.ac.id2, nj\_anna@ub.ac.id3

|  |  |  |
| --- | --- | --- |
|  |  | **ABSTRACT** |
| **Article History:**Received : D-M-20XXRevised : D-M-20XXAccepted : D-M-20XXOnline : D-M-20XX |  | This study aims to obtain the main component score of the ability to pay latent variable, determine the strongest indicators forming the ability to pay on a mixed scale based on defined indicators, and model the ability to pay on time mediated by fear of paying using path analysis. The data used in this study is secondary data from mortgage-paying customers with a sample size of 100. The method used is nonlinear principal component analysis with path analysis modeling. The results of this study indicate that the eleven variables formed by PC1 or X1 are able to store diversity or information by 74.8%, while 25.20% of diversity or other information is not stored (wasted). The credit term is the strongest indicator that forms the ability to pay variable. The variable ability to pay mortgages has a significant effect on payments by mediating the fear of paying late with a coefficient of determination of 80.40%. |
| **Keyword*:***Nonlinear Principal Component Analysis;Path Analysis;Mixed Data Formative;Indicator Models. |
| E:\UMUM\OJSQ\JTAM\JTAM-Barcode copy.jpg |
| C:\Users\WINDOWS 7\Documents\Indeksi\New-Cros.jpg<https://doi.org/10.31764/jtam.vXiY.ZZZ> | C:\Users\WINDOWS 7\Documents\Indeksi\88x31.png**This is an open access article under the CC–BY-SA license** |

—————————— ◆ ——————————

1. **INTRODUCTION**

Multivariate analysis is one of statistical analysis that simultaneously analyze several variables in individuals or objects (Astutik et al., 2018). With multivariate analysis, the effect of several variables on other variables can be analyzed at once. Meanwhile, according to Solimun & Fernandes (2017) multivariate analysis can be said to be the use of statistical methods related to several variables where the measurements are carried out from each research object. The variable itself is a characteristic of the subject or object that is relevant to the problem being studied, where there are various variables viewed from various points of view. Based on the measurement process, variables are divided into manifest variables (observable) and latent variables (unobservable).

According to Solimun et al. (2017), in general, latent variables are defined as variables that cannot be measured directly, but the measurements must be carried out through indicators that reflect or compile them. Latent variables can be classified into variables in the form of psychological attributes such as satisfaction in the form of conception variables, and can also be classified into factual latent variables such as the ability to pay mortgages which will be examined in this study (Fernandes et al., 2014; Hang et al., 2016; James et al., 2013; O’Rourke et al., 2013). To get data on the ability to pay mortgages, can be measured through the indicators that compose it. The indicator model that forms or composes variables is called the formative indicator model, where in this indicator model, the indicators that compose it are not required to have a common factor. To measure latent variable data with formative indicator models, the main component score method is used which is obtained through Principal Component Analysis.

Principal Component Analysis is a multivariate analysis introduced by Karl Pearson. The purpose of Principal Component Analysis is to reduce the number of variable sets, and can also be used to obtain latent variable principal component scores (David & Jacobs, 2014; Solimun et al., 2017). However, in its application, not all indicators of latent variables are in the form of metric data. To analyze data with a mixed scale, consisting of non-metric scales, nonlinear principal component analysis with nonlinear transformations is carried out. The nonlinear transformation in question is principal component analysis with optimal scaling or optimal transformation from qualitative scale to quantitative values (Chen et al., 2017; Fernandes, Hutahayan, et al., 2019; Vegelius & Jin, 2021; Yu et al., 2015; Zhang & Pan, 2019).

Path analysis is a method used to determine causal relationships between endogenous and exogenous variables. Path analysis connects the existence of a mediating variable. Mediating variables are relationships between exogenous variables and endogenous variables with the result that exogenous variables cannot influence changes or the existence of endogenous variables directly (Efendi et al., 2021; Fernandes, Darmanto, et al., 2019; Fernandes & Solimun, 2017; Kock, 2016).

One of the credit facilities provided by banks in Indonesia is the Home Ownership Credit (KPR). According to Budiono (2016), KPR is a credit facility provided by a certain bank to individual customers, both for those who want to buy a house and those who want to repair their house. Before banks provide mortgages to customers, it is important for the bank to have an assessment of potential customers (Hayati, 2016; Sumardi & Fernandes, 2018). So that congestion in the process of returning customer mortgages to the bank can be minimized. To overcome the risk of bad credit, customers must pay mortgages on time. By looking at the influence of the ability to pay mortgages and the fear of paying late on timely mortgage payments, the bank can decide if the customer is able to pay mortgages on time or not.

Based on the explanation above, the researcher wants to find out more about how to get the main component score from the latent variable ability to pay mortgages with its constituent indicators, namely guarantee documents, education, collectability status, work experience, joint income, RPA, form of business entity, credit term, number of family responsibilities, savings ownership, and loan to value. In addition, this research also determines how the influence of the ability to pay mortgages on paying mortgages on time is mediated by the fear of paying mortgages.

1. **METHODS**

The data used is secondary data regarding the ability of customers to pay bank mortgages, fear of paying late, and paying mortgages on time. This data was obtained from research conducted at Bank X in 2022. The subjects studied were the mortgage debtors customers. This study uses data consisting of 100 samples. The data scale used is a mixed data scale. This study uses two analytical methods with a mixed method approach, namely *Principal Component Analysis* (PCA) Nonlinear with path analysis. This research is designed to answer the problems that have been formulated, as well as to achieve research objectives by involving hypothesis testing to determine the effect between research variables.



**Figure 1.** Research Model

The following is an indicator of the variable ability to pay mortgages.

**Table 1.** Variable Indicators of Ability to Pay Mortgages

|  |  |  |
| --- | --- | --- |
| **Indicator** | **Data Scale** | **Category** |
| Guarantee Document (X1.1) | Ordianal | Other | 1 |
| Parent SHM / Main SHGB | 2 |
| SHGB | 3 |
| SHM | 4 |
| Education (X1.2) | Ordinal | SD, SMP, SMA | 1 |
| Diploma | 2 |
| S1,S2,S3 | 3 |
| Collectability Status (X1.3) | Ordinal | In Special Attention | 1 |
| Fluent | 2 |
| Work Experience (X1.4) | Ordinal | $\leq 3$Year | 1 |
| $>3 - \leq 6$Year | 2 |
| $>6 - \leq 12$Year | 3 |
| $>12 - \leq 18$Year | 4 |
| $>18 - \leq 20$Year | 5 |
| $>20$Year | 6 |
| Joint Income (X1.5) | Ordinal | Non-joint income | 1 |
| Have joint income | 2 |
| RPA (Instalment Income Ratio) (X1.6) | Ordinal | $\leq 1,5$Million | 1 |
| $>1,5 - \leq 2$Million | 2 |
| $>2 - \leq 2,5$Million | 3 |
| $>2,5 - \leq 3$Million | 4 |
| $>3$Million | 5 |
| Form of Business Entity (X1.7) | Ordinal | Other | 1 |
| PERUM, PERSERO | 2 |
| PT Non Tbk | 3 |
| Credit Term (X1.8) | Ordinal | $>240$Month | 1 |
| $>180 - \leq 240$Month | 2 |
| $>120 - \leq 180$Month | 3 |
| $>60 - \leq 120$Month | 4 |
| $>48 - \leq 60$Month | 5 |
| $>36 - \leq 48$Month | 6 |
| $\leq 36$Month | 7 |
| Number of Family Responsibilities (X1.9) | Ordinal | $>2$Person | 1 |
| 2 persons | 2 |
| 1 person | 3 |
| 0 People | 4 |
| Savings Ownership (X1.10) | Ordinal | Do not have | 1 |
| Have Savings in Other Banks | 2 |
| Have Savings | 3 |
| Loan to Value (X1.11) | Ordinal | $>95\%$  | 1 |
| $>90\% - \leq 95\%$  | 2 |
| $>80\% - \leq 90\%$  | 3 |
| $>70\% - \leq 80\%$  | 4 |
| $>60\% - \leq 70\%$  | 5 |
| $>50\% - \leq 60\%$  | 6 |
| $\leq 50\%$  | 7 |

1. **RESULT AND DISCUSSION**
2. **Principal Component Analysis**

There are nine indicators that form the bank's ability to pay mortgages variable. These indicators are analyzed by nonlinear principal components to obtain principal component scores for the variables. Eigen value is used to select the principal component. The selected component must have 𝑒𝑖𝑔𝑒𝑛 𝑣𝑎𝑙𝑢𝑒 > 1. The eigen value results from the principal component analysis of the ability to pay mortgages are presented in Table 1.

**Table 1.** Eigen Value of the Ability to Pay Mortgages Variables

|  |  |
| --- | --- |
|  | **Eigen Value** |
| **PC1** | 2,076 |
| **PC2** | 0,251 |
| **PC3** | 0,212 |
| **PC4** | 0,143 |
| **PC5** | 0,094 |
| **PC6** | 0,519 |
| **PC7** | 0,846 |
| **PC8** | 0,322 |
| **PC9** | 0,244 |
| **PC10** | 0,119 |
| **PC11** | 0,658 |

Based on table 1, there is only one component that have 𝑒𝑖𝑔𝑒𝑛 𝑣𝑎𝑙𝑢𝑒 > 1, then the selected component is the first component. The proportion of cumulative variance in the first component is 0.784, meaning that the first main component contains 74.8% of the data from the variable Ability to Pay KPR. The weight of the first component in the Mortgage Paying Variable

**Table 2.** The Weight of the Principal Components of the Ability to Pay Mortgage Variables

|  |  |
| --- | --- |
| **Indicator** | **Weight** |
| Guarantee Document (X1.1) | 0.361 |
| Education (X1.2) | 0.386 |
| Collectability Status (X1.3) | 0.438 |
| Work Experience (X1.4) | 0.507 |
| Joint Income (X1.5) | 0.522 |
| RPA (Instalment Income Ratio) (X1.6) | 0.476 |
| Form of Business Entity (X1.7) | 0.446 |
| Credit Term (X1.8) | 0.538 |
| Number of Family Responsibilities (X1.9) | 0.361 |
| Savings Ownership (X1.10) | 0.392 |
| Loan to Value (X1.11) | 0.345 |

Based on the weight of the principal components, the indicator that has the greatest weight is the X18 indicator, namely the Credit Period. This means that the X18 indicator is able to characterize the Mortgage Paying Ability variable. Based on table 2, the principal component linear combination equation can be formed to get the component score which is the value of the Mortgage Paying Ability variable (X1).

$$X\_{1}=0,361X\_{1.1}+0,386X\_{1.2}+0,438X\_{1.3}+0,507X\_{1.4}+0,522X\_{1.5}+0,476X\_{1.6}+0,446X\_{1.7}+0,538X\_{1.8}+0,361X\_{1.9}+0,392X\_{1.10}+0,345X\_{1.11}$$

1. **Path Analysis**
2. A Path Analysis assumptions

The results of testing the assumptions of path analysis in this study are as follows.

1. The relationship between variables is linear and additive

The linearity test was carried out using the RESET method with Rstudio software with the output results shown in table 3.

**Table 3.** Linearity Test Results

|  |  |  |
| --- | --- | --- |
| **Variable Relations** | **P-Value** | **Connection** |
| X1 against Y1 | 0.4377 | linear |
| X1 against Y2 | 0.5698 | linear |
| Y1 against Y2 | 0.6714 | linear |

Based on Table 3. it can be seen that there is a relationship between exogenous variables and endogenous variables$p-value>0,05$ means accept $H\_{0}$so that it can be said that the assumption of linearity has been met.

1. Minimal endogenous variable in interval measurement scale

The data used is secondary data in the form of a Likert scale, where the score produced on the Likert scale is data that is close to the interval scale. Then the score that has been obtained is carried out by a scaling process using the Summated Rating Scale (SRS) method. Therefore, the assumption of a minimum endogenous variable measuring interval scale is fulfilled.

1. Normality Assumption

The normality assumption test is used to identify residuals in normally distributed research or not. The regression model can be said to be good if the residuals are normally distributed. Based on the results of the analysis obtained value $p-value (0,2)>α (0,05)$, it can be concluded to reject $H\_{0}$. So it can be concluded that with a significant level of 5%, the residuals are normally distributed.

1. Models are Recursive

The model in path analysis is said to be recursive if it has a one-way relationship pattern. In Figure 2 it can be seen that each exogenous variable has one-way causality and there is no two-way (reciprocal) relationship so that it can be said that the model is recursive.

1. The analyzed model is correctly specified based on relevant theories and concepts
2. Parameter Estimation and Hypothesis Testing

Parameter estimation in the path analysis is carried out to estimate the path coefficient. This is used to determine the relationship between exogenous variables and endogenous variables, while hypothesis testing is used to test the significance of the path coefficient partially. The hypothesis used is as follows.

* $H\_{0}: ρ\_{xy}=0$(There is no significant effect of exogenous variables on endogenous variables) vs
* $H\_{1}: ρ\_{xy}\ne 0$(There is a significant effect of exogenous variables on endogenous variables)

**Table 3.** Results of Parameter Estimation and Hypothesis Testing

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable Relations** | **Path Coefficient** | **p-values** | **Decision** |
| X1 against Y1 | 0.4181 | 0.0045 | Reject $H\_{0}$ |
| X1 against Y2 | 0.4452 | 0.0127 | Reject $H\_{0}$ |
| Y1 against Y2 | 0.3939 | 0.0012 | Reject $H\_{0}$ |

Based on Table 3. It can be seen that the decision Reject $H\_{0}$, which means there is a significant influence of exogenous variables on endogenous variables, with the results of estimating the parameters of the path analysis can be formed as follows:

$Z\_{Y1}=0,4452Z\_{X1}$

$Z\_{Y2}=0,4452Z\_{X1}+0,3939Z\_{Y1}$

With diagrams and path coefficients as follows:



**Figure 2.** Path Diagram and Coefficient

1. Path Analysis Model Validity

The validity of the model in path analysis can be known by looking at the total determination coefficient value. The total coefficient of determination is used to explain the diversity of data that can be explained by the model. Calculation of the total determination coefficient obtained from each model that is formed in the path analysis. The coefficient of determination for each model is obtained using the RStudio software. The results of the coefficient of determination can be seen in Table 4.

**Table 4.** Coefficient of Determination

|  |  |
| --- | --- |
| **Model** | **Coefficient of Determination (R 2 )** |
| 1 | 0.6123 |
| 2 | 0.5537 |

Based on Table 4, the total determination coefficient is calculated as follows.

1. Model 1

$R\_{1}^{2}=0,6123$

$P\_{e1}=\sqrt{1-R\_{1}^{2}}=0,6226$

1. Model 2

$R\_{1}^{2}=0,5537$

$P\_{e2}=\sqrt{1-R\_{2}^{2}}=0,6680$

The total determination coefficient of model 1 and model 2 is as follows.

$R\_{t}^{2}=1-P\_{e1}^{2}P\_{e2}^{2}$

$R\_{t}^{2}=1-(\left(0,6226\right)^{2}×\left(0,6680\right)^{2})$

$R\_{t}^{2}=0,8040$

The total determination coefficient value of 0.8040 explains that 80.40% of the data diversity can be explained by the research model, while 19.6% of the data diversity is explained by other variables outside the model.

1. **Discussion**

The results showed that the ability to pay mortgages variable, the credit term indicator is the indicator which best characterizes the ability to pay mortgages variable compared to other indicators. This means that the credit term indicator is the most capable indicator to represent the variable of the ability to pay mortgages. The longer the credit period, the lower the monthly installments that must be paid by the customer. This can help ease the burden of monthly payments and give customers more time to repay loans. The longer the credit term, the higher the amount of interest to be paid, so that the overall cost of the loan will increase. Therefore, before taking credit for a certain period of time, it is important to consider the customer's financial capabilities so that they can pay installments consistently and on time.

Based on the research results, the variable ability to pay mortgages has a significant effect on timely payments, which is mediated by the fear of paying late. The ability to pay shows the extent to which the customer has the financial capacity to pay installments according to the agreement. If the customer has a good paying ability, that is, has sufficient income or cash flow to cover credit payments on schedule, this helps ensure that customers can pay debts consistently and don't miss the payments. On the other hand, if customers have problems paying on time, for example, they often miss payments or experience repeated delays, this could have a negative impact on their financial credibility. In some cases, this may lead to additional penalties or fees, and may even result in default, which may negatively affect the ability to obtain credit in the future. Therefore, it is important for customers to have good paying ability and be responsible for maintaining discipline in paying credit installments on time.

1. **CONCLUSSION**

Based on the results of the analysis it can be concluded that:

1. Nonlinear principal component analysis was used to obtain latent variables with indicators on a non-metric scale and it was found that the credit term indicator was the indicator that best characterized the ability to pay mortgages compared to other indicators. The results showed that the eleven variables formed by PC1 or X1 were able to store diversity or information by 74.8%, while 25.20% of diversity or other information was not stored (wasted).
2. Based on the results of the path analysis, it can be concluded that the ability to pay mortgages has a significant effect on timely payments, which is mediated by fear of paying. With a total coefficient of determination of 0.8040, it explains that 80.40% of the data diversity can be explained by the research model, while 19.6% of the data diversity is explained by other variables outside the model. And the path analysis model is obtained as follows:

$Z\_{Y1}=0,4452Z\_{X1}$

$Z\_{Y2}=0,4452Z\_{X1}+0,3939Z\_{Y1}$

**ACKNOWLEDGEMENT**

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors would like to thank the editors and anonymous reviewers for their comments that help improve the quality of this work.

**REFERENCES**

Astutik, S., Solimun, & Darmanto. (2018). *Multivariate Analysis: Theory and Application with SAS*. Brawijaya Press University.

Budiono, D. (2016). *The behavior of corporate taxpayers in fulfilling tax obligations: Humanistic theory perspective*. FEBSOS.

Chen, B., Yang, J., Jeon, B., & Zhang, X. (2017). Kernel quaternion principal component analysis and its application in RGB-D object recognition. *Neurocomputing*, *266*, 293–303. https://doi.org/10.1016/j.neucom.2017.05.047

David, C. C., & Jacobs, D. J. (2014). Principal component analysis: a method for determining the essential dynamics of proteins. *Protein Dynamics: Methods and Protocols*, 193–226. https://doi.org/10.1007/978-1-62703-658-0\_11

Efendi, E. C. L., Fernandes, A. A. R., & Mitakda, M. B. T. (2021). Modeling of Path Nonparametric Truncated Spline Linear, Quadratic, and Cubic in Model on Time Paying Bank Credit. *WSEAS Transactions on International Journal of Electrical Engineering and Computer Science,* *3*, 52–60.

Fernandes, A. A. R., Budiantara, I. N., Otok, B. W., & Suhartono. (2014). Reproducing Kernel Hilbert space for penalized regression multi-predictors: Case in longitudinal data. *International Journal of Mathematical Analysis*, *8*(40), 1951–1961.

Fernandes, A. A. R., Darmanto, Astuti, A. B., Solimun, Amaliana, L., Nurjannah, Yanti, I., Arisoesilaningsih, E., & Isaskar, R. (2019). Smoothing Spline Nonparametric Path: Application for Green Product and Green Marketing Strategy towards Green Product Purchasing Intention. *IOP Conference Series: Earth and Environmental Science*, *239*(1). https://doi.org/10.1088/1755-1315/239/1/012018

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., & Solimun. (2017). Moderating effects orientation and innovation strategy on the effect of uncertainty on the performance of the business environment. *International Journal of Law and Management*, *59*(6), 1211–1219.

Hang, J., Zhang, J., & Chen, M. (2016). Application of multi-class fuzzy support vector machine classifier for fault diagnosis of wind turbine. *Fuzzy Sets and Systems*, *297*, 128–140. https://doi.org/https://doi.org/10.1016/j.fss.2015.07.005

Hayati, I. K. (2016). Analysis of Quality of Worklife (Qwl) Implementation Against Job Satisfaction and Commitment of the Employees. *The Management Journal of Binaniaga*, *1*(01), 9–18.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning*. Springer.

Kock, N. (2016). Advantages of nonlinear over segmentation analyses in path models. *International Journal of E-Collaboration (Ijec)*, *12*(4), 1–6.

O’Rourke, N., Psych, R., & Hatcher, L. (2013). *A step-by-step approach to using SAS for factor analysis and structural equation modeling*. Sas Institute.

Solimun, & Fernandes, A. A. R. (2017). *Investigation of the mediating variable: What is necessary? (case study in management research)*. *59*(6), 1059–1067. https://doi.org/https://doi.org/10.1108/IJLMA-09-20 16 -0077

Solimun, Fernandes, A. A. R., & Nurjannah. (2017). *Multivariate Statistical Method Structural Equation Modeling (SEM) WarpPLS Approach*. UB Press.

Sumardi, S., & Fernandes, A. A. R. (2018). The mediating effect of service quality and organizational commitment on the effect of management process alignment on higher education performance in Makassar, Indonesia. *Journal of Organizational Change Management*, *31*(2), 410–425.

Vegelius, J., & Jin, S. (2021). A semiparametric approach for structural equation modeling with ordinal data. Structural Equation Modeling. *A Multidisciplinary Journal*, *28*(4), 497–505.

Yu, H., Khan, F., & Ikram, G. (2015). An alternative formulation of PCA for process monitoring using distance correlation. *Industrial & Engineering Chemistry Research*, *55*(3), 656–669.

Zhang, Z., & Pan, P. (2019). A hybrid intrusion detection method based on improved fuzzy C-means and support vector machine. *A Hybrid Intrusion Detection Method Based on Improved Fuzzy C-Means and Support Vector Machine.*