

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

Rindu Hardianti¹, Solimun², Nurjannah³, Rosita Hamdan⁴

^{1,2,3}Departement of Statistics, University of Brawijaya, Indonesia

⁴Department of Development Economics, University Malaysia Serawak, Malaysia

rinduhardianti@student.ub.ac.id¹, solimun@ub.ac.id², nj_anna@ub.ac.id³, hrosita@unimas.my⁴

ABSTRACT

Article History:

Received : 31-08-2023

Revised : 09-12-2023

Accepted : 17-12-2023

Online : 19-01-2024

Keywords:

Nonlinear Principal Component Analysis;
Path Analysis;
Mixed Data;
Formative Indicator Models.

This research aims to obtain the main component score of the latent variable ability to pay, determine the strongest indicators forming the ability to pay on a mixed scale based on predetermined indicators, and model the ability to pay on time as mediated by fear of paying using path analysis. The data used is secondary data obtained through distributing questionnaires with a mixed data scale. The sampling technique used in the research was purposive sampling. The number of samples used in the research was 100 customers. The method used is nonlinear principal component analysis with path analysis modeling. The results of this research show that of the five indicators formed by the Principal Component, 74.8% of diversity or information is able to be stored, while 25.20% of diversity or other information is not stored (wasted). Credit term is the strongest indicator that forms the ability to pay variable. The variable ability to pay mortgage has a significant effect on payments by mediating the fear of being late in paying with a coefficient of determination of 73.63%.



<https://doi.org/10.31764/jtam.v8i1.17559>



This is an open access article under the **CC-BY-SA** license

A. 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 must be through indicators that reflect or structure them. Latent variables can be sorted into variables in the form of psychological attributes such as satisfaction in the form of conception variables, and can also be sorted into latent variables which are factual in nature such as the ability to pay mortgage variables which will be examined

in this research. In Cahyoningtyas's (2020) research, to obtain data on the ability to pay mortgages, it is measured through the indicators that make it up. An indicator model whose nature is to form or arrange variables is called a formative indicator model, where in this indicator model, the indicators that make it up are not required to have a common factor. According to David & Jacob (2014), to measure latent variable data with a formative indicator model, the principal component score method obtained through Principal Component Analysis is used.

Principal Component Analysis is a multivariate analysis introduced by Karl Pearson. Principal Component Analysis is a technique used to simplify data, by means of linear transformation to form a new coordinate system with maximum variation. Principal component analysis is used to reduce the dimensions of data without significantly reducing the characteristics of the data. Data reduction is carried out by means of linear transformation so that a new coordinate system is formed with maximum variation without significantly reducing the characteristics of the data (Gewers et al., 2021).

In principal component analysis there are two input matrices, namely the correlation matrix and the variance matrix (Yu et al., 2015). The variance matrix can be used when the variables to be analyzed have the same units. Meanwhile, the correlation matrix can be used for data that does not have the same units by transforming to standard normal form. In general, the correlation matrix used in principal component analysis is Pearson correlation. However, Pearson correlation has limitations, namely that it cannot calculate ordinal data scales (Chen et al., 2017). Therefore in this research uses nonlinear principal component analysis.

According to Mori et al. (2016), nonlinear principal component analysis is a development of principal components called Princals (Principal Component Analysis by Alternating Least Square). The purpose of nonlinear principal component analysis is to obtain principal component scores and obtain quantification of the categories of each variable. Nonlinear principal component analysis is used to analyze variables with nominal or ordinal data scales where these variables include "categorical" variables and is also used to resolve variables that have nonlinear relationships (Demir & Keskin, 2022).

The initial stage in nonlinear principal component analysis is the quantification process (Katayama et al., 2022). Quantification is assigning a numerical value to a category of a variable. The quantification technique used in this analysis is the optimal scaling technique. According to Heo & Lee (2019), data that has been quantified can be analyzed as in the linear principal component analysis model because all variables have a numerical scale. It can be concluded that nonlinear principal component analysis is the same as linear principal component analysis which goes through the process of quantifying categorical variables.

Jian & Yan (2015), also stated that in nonlinear principal component analysis, the correlation matrix is not calculated from the observed values of the variables, but from the values resulting from the quantification of the variables. Therefore, the correlation matrix in nonlinear principal component analysis can be said to be uncertain or unstable, depending on the quantification method used, which is called level analysis. Thus, the solution to nonlinear principal component analysis does not come from the correlation matrix, but from an "optimal scaling" process that depends on the analysis level.

Path analysis is a multivariate analysis method used to test models of relationships between variables in the form of cause and effect (Kock, 2016). Path analysis was first developed by Wright in 1934 and was used to determine the direct and indirect effects between variables (Vegelius & Jin, 2021). 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 (Fernandes & Solimun, 2017). Path analysis can be applied in various fields, one of which is banking.

Home Ownership Credit (mortgage) is one of the most common credit facilities provided by banks in Indonesia and in many other countries. According to Budiono (2016), mortgage 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 (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.

Previous research related to timeliness of payments was conducted by Efendi et al. (2021). In this research it can be concluded that there is a significant relationship between lifestyle which is a representation of the variable ability to pay mortgages and the timeliness of paying mortgages. However, there has not been much research regarding the calculation of latent variables, namely the ability to pay mortgages using the constituent indicators on an ordinal scale. Therefore, in this research, a non-linear principal component analysis was carried out to measure the variable of ability to pay mortgages and determine the most significant indicators in forming the variable of ability to pay mortgages. After conducting a nonlinear principal component analysis, a path analysis was then carried out to find out the relationship between the variable ability to pay mortgages and the timeliness of paying mortgages.

B. METHODS

The data in the research is secondary data obtained through distributing questionnaires with a mixed data scale, namely a Likert scale for the variables of fear of being late in paying and paying mortgages on time. Meanwhile, the variable ability to pay mortgages uses an ordinal data scale. The sampling technique used was purposive sampling. Purposive sampling is a sampling technique that is based on certain characteristics or conditions that are the same as the characteristics of the population. The number of samples used in the research was 100 customers. 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. The research model used is shown in Figure 1.

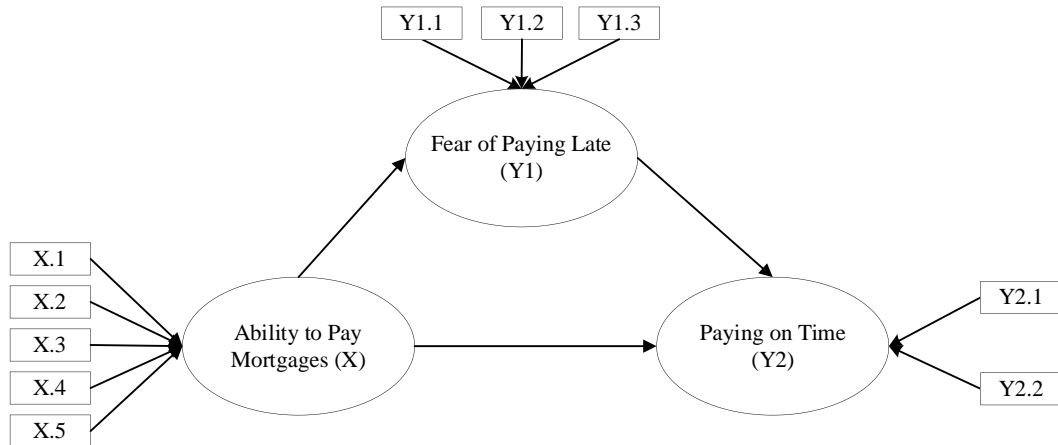


Figure 1. Research Model

The following is an indicator of the variable ability to pay mortgages, as shown in Table 1.

Table 1. Variable Indicators of Ability to Pay Mortgages

Indicator	Data Scale	Category	
Work Experience (X.1)	Ordinal	≤ 3Year	1
		> 3 – ≤ 6Year	2
		> 6 – ≤ 12Year	3
		> 12 – ≤ 18Year	4
		> 18 – ≤ 20Year	5
		> 20Year	6
RPA (Instalment Income Ratio) (X.2)	Ordinal	≤ 1.5Million	1
		> 1.5 – ≤ 2Million	2
		> 2 – ≤ 2.5Million	3
		> 2.5 – ≤ 3Million	4
		> 3Million	5
Form of Business Entity (X.3)	Ordinal	Other	1
		PERUM. PERSERO	2
		PT Non Tbk	3
Credit Term (X.4)	Ordinal	> 240Month	1
		> 180 – ≤ 240Month	2
		> 120 – ≤ 180Month	3
		> 60 – ≤ 120Month	4
		> 48 – ≤ 60Month	5
		> 36 – ≤ 48Month	6
		≤ 36Month	7
Loan to Value (X.5)	Ordinal	> 95%	1
		> 90% – ≤ 95%	2
		> 80% – ≤ 90%	3
		> 70% – ≤ 80%	4
		> 60% – ≤ 70%	5
		> 50% – ≤ 60%	6
		≤ 50%	7

C. RESULT AND DISCUSSION

1. Principal Component Analysis

There are five 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. According to Jolliffe & Cadima (2016) in determining how many principal components (PC) should be taken, three methods are used, one of which is by using eigen values. The selected component must have an eigen value of more than one. The eigen value results from the principal component analysis of the ability to pay mortgages are presented in Table 2.

Table 2. Eigen Value of the Ability to Pay Mortgages Variables

Principal Component	Eigen Value	Variance (%)	Cummulative Variance (%)
1	2.076	0.748	0.748
2	0.251	0.140	0.888
3	0.212	0.096	0.984
4	0.143	0.014	0.998
5	0.094	0.002	1.000

Based on Table 2, there is only one component that have eigen value of more than one, 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 Mortgages. The weight of the first component in the Mortgage Paying Variable, as shown in Table 3.

Table 3. The Weight of the Principal Components of the Ability to Pay Mortgage Variables

Indicator	Weight
Work Experience (X.1)	0.361
RPA (Instalment Income Ratio) (X.2)	0.386
Form of Business Entity (X.3)	0.438
Credit Term (X.4)	0.522
Loan to Value (X.5)	0.507

Based on Table 3, the weight of the principal components, the indicator that has the greatest weight is the X.4 indicator, namely the Credit Period. This means that the Credit Period 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 (X).

$$X = 0.361X_1 + 0.386X_2 + 0.438X_3 + 0.522X_4 + 0.507X_5 \quad (1)$$

2. Path Analysis

a. 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 4.

Table 4. Linearity Test Results

Variable Relations	P-Value	Connection
X against Y1	0.2455	linear
X against Y2	0.3960	linear
Y1 against Y2	0.5750	linear

Based on Table 4 it can be seen that there is a relationship between exogenous variables and endogenous variables $p - \text{value} > 0.05$ means accept H_0 so that it can be said that the assumption of linearity has been met.

2) 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.

3) 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 $p - \text{value} (0.2) > \alpha (0.05)$, 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.

4) 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.

5) The analyzed model is correctly specified based on relevant theories and concepts.

b. 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. $H_0: \rho_{xy} = 0$ (There is no significant effect of exogenous variables on endogenous variables) vs; $H_1: \rho_{xy} \neq 0$ (There is a significant effect of exogenous variables on endogenous variables), as shown in Table 5.

Table 5. Results of Parameter Estimation and Hypothesis Testing

Variable Relations	Path Coefficient	p-values	Decision
X against Y1	0.4282	0.0052	Reject H ₀
X against Y2	0.4356	0.0276	Reject H ₀
Y1 against Y2	0.3452	0.0090	Reject H ₀

Based on Table 5 it can be seen that the decision Reject H₀, 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:

$$\begin{aligned} Z_{Y1} &= 0.4282Z_X \\ Z_{Y2} &= 0.4356Z_X + 0.3452Z_{Y1} \end{aligned} \tag{2}$$

With diagrams and path coefficients as shown in Figure 2.

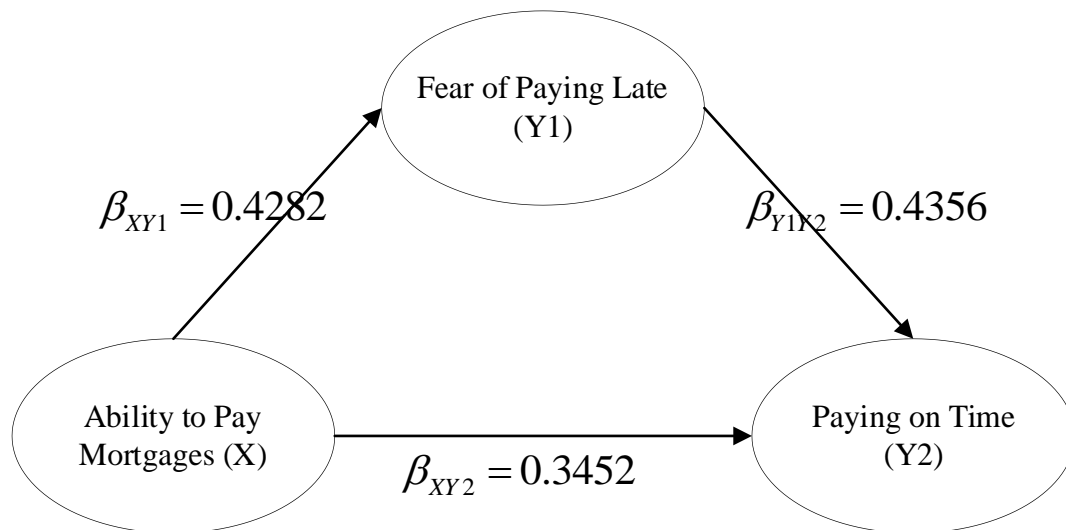


Figure 2. Path Diagram and Coefficient

c. 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 6.

Table 6. Coefficient of Determination

Model	Coefficient of Determination (R ²)
1	0.6250
2	0.5672

Based on Table 6, the total determination coefficient is calculated as follows.

1) Model 1

$$R_1^2 = 0.6250$$

$$P_{e1} = \sqrt{1 - R_1^2} = 0.7806$$

2) Model 2

$$R_2^2 = 0.5672$$

$$P_{e2} = \sqrt{1 - R_2^2} = 0.6578$$

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

$$R_t^2 = 1 - P_{e1}^2 P_{e2}^2$$

$$R_t^2 = 1 - ((0.7806)^2 \times (0.6578)^2)$$

$$R_t^2 = 0.7363$$

The total determination coefficient value of 0.7363 explains that 73.63% of the data diversity can be explained by the research model, while 26.37% of the data diversity is explained by other variables outside the model.

3. Discussion

The research results show that nonlinear principal component analysis can measure the variable ability to pay mortgages with indicators arranged on an ordinal scale. This statement is supported by the results of research conducted by Cahyoningtyas et al. (2020), which revealed that latent variables with an ordinal scale can be measured using nonlinear principal component analysis. Demir & Keskin (2022) stated that nonlinear principals component analysis can be used in variables that use ordinal scales. Based on the analysis, it was found that of the 5 indicators forming the ability to pay mortgages variable, credit term is the indicator that has the strongest value to form the ability to pay mortgage variable with a value of 0.522. Meanwhile, the weakest indicator in forming the ability to pay mortgage variable is the work experience indicator with a value of 0.361.

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.

The model formed is good for use, this can be seen through the coefficient of determination of the model. Where in model 1 the coefficient of determination is 0.6250, which means that

the variable ability to pay mortgages is able to explain the fear of paying late by 62.50% and the remaining 37.50% is explained by other variables outside the study. On the other hand, model 2 has a coefficient of determination of 0.5672 which means that the ability to pay mortgages variable and the fear of paying late variable are able to explain the variable of paying of time by 56.72% and the remaining 43.28% is explained by other variables outside this study. And the total determination coefficient value of 0.7363 explains that 73.63% of the data diversity can be explained by the research model, while 26.37% of the data diversity is explained by other variables outside the model.

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. This is in line with research conducted by Efendi et al. (2021), 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.

D. CONCLUSION

Based on the results of the analysis it can be concluded that, 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 principal component were able to store diversity or information by 74.8%, while 25.20% of diversity or other information was not stored (wasted). Empirically, 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 0.7363, it explains that 73.63% of the data diversity can be explained by the research model, while 26.37% of the data diversity is explained by other variables outside the model. And the path analysis model is obtained as follows:

$$\begin{aligned} Z_{Y1} &= 0.4282Z_X \\ Z_{Y2} &= 0.4356Z_X + 0.3452Z_{Y1} \end{aligned}$$

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. https://books.google.co.id/books/about/Analisis_Multivariat.html?id=BvhqDwAAQBAJ&redir_esc=y
- Budiono, D. (2016). *The behavior of corporate taxpayers in fulfilling tax obligations: Humanistic theory perspective*. FEBSOS. <http://jonuns.com/index.php/journal/article/view/542>
- Cahyoningtyas, R. A., Solimun, & Fernandes, A. A. R. (2020). The implementation of first order and second order with mixed measurement to identify farmers satisfaction. *Mathematics and Statistics*, 8(6), 671–682. <https://doi.org/https://doi.org/10.13189/ms.2020.080607>
- 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/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*, 1084, 193–226. https://doi.org/https://doi.org/10.1007/978-1-62703-658-0_11
- Demir, C., & Keskin, S. (2022). Introduction of Nonlinear Principal Component Analysis with an Application in Health Science Data. *Eastern Journal of Medicine*, 27(3), 394–402. <https://doi.org/10.5505/ejm.2022.09068>
- 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. <https://doi.org/https://doi.org/10.13189/ms.2021.090611>
- Fernandes, A. A. R., & Solimun. (2017). Moderating effects orientation and innovation strategy on the effect of uncertainty on the performance of business environment. *International Journal of Law and Management*, 59(6), 1211–1219. <https://doi.org/10.1108/IJLMA-10-2016-0088>
- Gewers, F., Ferreira, G. R., Arruda, H. F. de, & Silva, F. N. (2021). Principal Component Analysis: A Natural Approach to Data Exploration. *ACM Computing Surveys (CSUR)*, 54(4), 1–34. <https://doi.org/https://doi.org/10.1145/3447755>
- Heo, S., & Lee, J. H. (2019). Parallel neural networks for improved nonlinear principal component analysis. *Computers & Chemical Engineering*, 127, 1–10. <https://doi.org/https://doi.org/10.1016/j.compchemeng.2019.05.011>
- Jiang, Q., & Yan, X. (2015). Nonlinear plant-wide process monitoring using MI-spectral clustering and Bayesian inference-based multiblock KPCA. *Journal of Process Control*, 32, 38–50. <https://doi.org/https://doi.org/10.1016/j.jprocont.2015.04.014>
- Jolliffe, I. T., & Cadima, J. (2016). Principal Component Analysis: A Review and Recent Developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202. <https://doi.org/https://doi.org/10.1098/rsta.2015.0202>
- Katayama, H., Mori, Y., & Kuroda, M. (2022). Variable Selection in Nonlinear Principal Component Analysis. In *Advances in Principal Component Analysis* (pp. 79–89). IntechOpen. <https://doi.org/http://dx.doi.org/10.5772/intechopen.103758>
- Kock, N. (2016). Advantages of nonlinear over segmentation analyses in path models. *International Journal of E-Collaboration (Ijec)*, 12(4), 1–6. <https://doi.org/https://doi.org/10.4018/ijec.2016100101>
- Mori, Y., Kuroda, M., & Makino, N. (2016). *Nonlinear Principal Component Analysis and Its Applications*. Springer Singapore. <https://doi.org/https://doi.org/10.1007/978-981-10-0159-8>
- Solimun, S; Fernandes, A. A. R. & N. (2017). *Metode Statistika Multivariat. Pemodelan Persamaan Struktural (SEM)*. UB Press. [https://books.google.co.id/books?hl=id&lr=&id=GrRVDwAAQBAJ&oi=fnd&pg=PR5&dq=Solimun,+S%3B+Fernandes,+A.+A.+R.+%26+N.+%282017%29.+Metode+Statistika+Multivariat.+Pemodelan+Persamaan+Struktural+\(SEM\).+UB+Press.&ots=nva-f-b8cH&sig=f-mwQtLFiSsdF1YW4NNdyqGj-fw&redir_esc=y#v=onepage&q&f=false](https://books.google.co.id/books?hl=id&lr=&id=GrRVDwAAQBAJ&oi=fnd&pg=PR5&dq=Solimun,+S%3B+Fernandes,+A.+A.+R.+%26+N.+%282017%29.+Metode+Statistika+Multivariat.+Pemodelan+Persamaan+Struktural+(SEM).+UB+Press.&ots=nva-f-b8cH&sig=f-mwQtLFiSsdF1YW4NNdyqGj-fw&redir_esc=y#v=onepage&q&f=false)
- 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-09-2016-0077>

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.

<https://doi.org/https://doi.org/10.1108/JOCM-11-2016-0247>

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.

<https://doi.org/https://doi.org/10.1080/10705511.2020.1848431>

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.

<https://doi.org/https://doi.org/10.1021/acs.iecr.5b03397>