CAUCHY: Jurnal Matematika Murni dan Aplikasi
Volume 10(1) (2025), Pages 201-212
p-ISSN: 2086-0382; e-ISSN: 2477-3344

Spearman Rank Correlation Principal Component Analysis for Mixed Indicators in Structural Equation Modeling

Lisa Asaliontin[1], Eni Sumarminingsih[1], Solimun[1], Mohammad Ohid Ullah[2,*]

[1]Department of Statistics, Brawijaya University, Indonesia
[2]Department of Statistics, Shahjalal Univeristy of Science and Technology, Bangladesh

Email: lisaasaliontin@student.ub.ac.id, ohid-sta@sust.edu*

ABSTRACT

Structural Equation Modeling (SEM) integrates variable relationships and indicator models simultaneously. Not all latent variables are metric, often mixing metric and non-metric scales, a topic seldom explored. This study aims to determine the performance of Spearman Rank Correlation Principal Component Analysis (PCA) in SEM with mixed-scale indicators in a mixed measurement model (formative and reflective). Applied to a case study on Fertilizer Repurchase Decisions with 250 research samples, Spearman Rank Correlation PCA was used for mixed-scale formative models and Confirmatory Factor Analysis (CFA) for reflective models. The results showed that Spearman Rank Correlation PCA have good performance, capturing 78.62% of the variance for mixed scale indicators and the SEM model confirmed significant relationships among variables with a coefficient of determination is 80%. The results demonstrate the flexibility of SEM in handling the complexity of mixed-scale data without sacrificing estimation accuracy and providing insight into customer behavior in making fertilizer repurchase decisions

Keywords: Farmer Demographics; Fertilizer Repurchase Decision; Mixed Scale Indicator; Spearman Rank Correlation PCA; Structural Equation Modeling

INTRODUCTION

Structural Equation Modeling (SEM) is statistical modeling that involves relationships between variables and also indicator models simultaneously [1]. SEM data analysis provides a comprehensive explanation of the study's variable relationships. For a set of distinct multiple regression equations calculated concurrently, SEM offers the most suitable and effective estimation method [2]. SEM is an analysis that uses latent variables, namely variables that cannot be measured directly (unobservable variables) so that they require indicators as measuring tools. In SEM, a set of linear relationships with multiple dependent variables can be estimated at the same time, and latent variable models can be specified to estimate the relationships between latent constructs and observed indicators [3]. Researchers can look at observations and theoretical ideas at the same time with SEM [4].The relationships between latent variables are described by the

Submitted: December 22, 2024 Reviewed: January 12, 2025 Accepted: March 10, 2025
DOI: http://dx.doi.org/10.18860/ca.v10i1.29976

structural model. On the other hand, the measurement model explains how latent variables and observed variables (indicators) relate to one another.

There are two main types of measurement models: formative and reflective. Reflective measurement, which is based on factor analysis, requires a common factor because it is an indicator model that reflects variables. However, a common factor is not necessary for formative measurement; that is, the indicators are not connected with one another [5]. In formative measurement, indicators collectively define the latent variable, so Principal Component Analysis (PCA) is used to combine indicators into one or more principal components so that a linear combination is formed that can effectively represent the latent variable.

Principal Component Analysis (PCA) is a technique for combining variables from multidimensional data into new variables, expressed as linear combinations of the original variables, while minimizing the loss of information from the original data set [6]. Conventional PCA, which uses Pearson correlation matrix in it, can only be used on metric-scale indicators (interval and ratio). However, in practical scenarios, indicators can be non-metric (nominal and ordinal) or a mixture of metric and nonmetric [7]. So that Spearman rank correlation is used instead of Pearson correlation in the correlation matrix to overcome the limitations of using Principal Component Analysis [8]. Previous research shows that the use of Spearman Rank Correlation successfully maintains the nonlinear correlation structure between variables and is computationally more efficient. However, the gap studies on PCA Rank Spearman Correlation are still very limited, and no application has been found in Structural Equation Modeling. Therefore, this study aims to evaluate the performance of Spearman Rank Correlation PCA in Structural Equation Modeling applied to Fertilizer Repurchase Decision modeling, which involves mixed scaled indicators (metric and non-metric) and mixed measurement models (formative and reflective) in it.

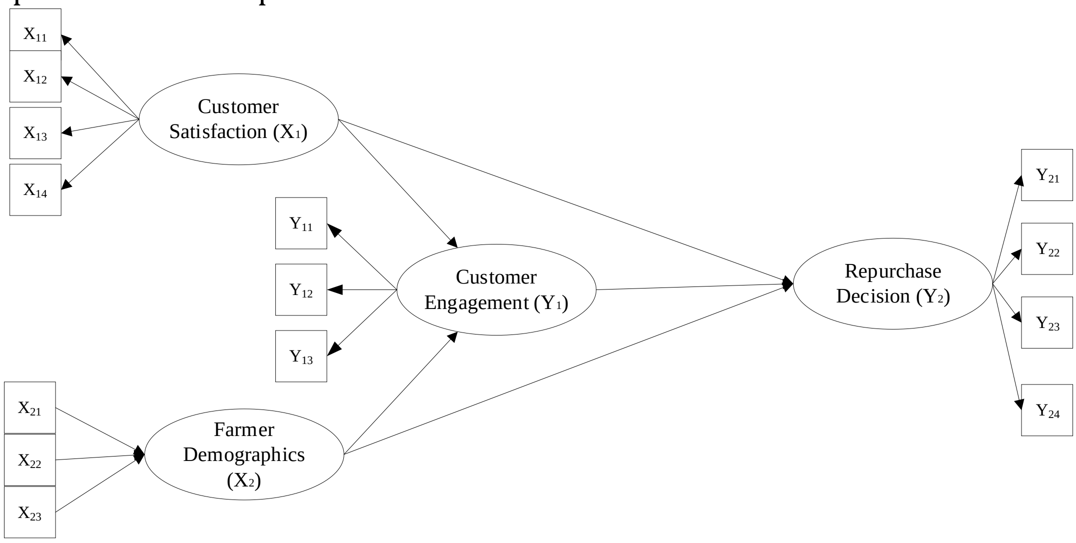
In a business, retaining customers is essential to ensure they continue to purchase products in the future. This also applies to PT. X, a company engaged in selling fertilizers, it is important to pay attention to its customers (farmers) in making Fertilizer repurchase decisions. This encourages PT. X to understand the characteristics of farmers in both external and internal aspects. External aspects include Farmer Demographics, while internal aspects include Customer Satisfaction and Engagement. Customer satisfaction acts as a key measure of the quality delivered to customers through the product or service and the accompanying support provided. [9]. Customer Engagement refers to emotional attachment that a customer experience during the repeated and ongoing interactions [10]. With these dynamics, the interaction between Farmer Demographics, Customer Satisfaction, and Customer Engagement is considered an important determining factor in understanding Fertilizer Repurchase Decisions, making it interesting to study.

This research was conducted to evaluate the performance of Spearman Rank Correlation PCA on Structural Equation Modeling in Fertilizer Repurchase Decision Case. This study focuses on a mixed-scale indicator model, it emphasizes a more flexible approach that integrates both metric and non-metric data, which is often found in real-world scenarios. By evaluating the performance of Spearman Rank Correlation PCA, particularly within a formative measurement model, this study not only addresses a gap in the literature but also provides new insights into the practical application of SEM with mixed-scale data.

METHODS

Data

This study is a quantitative research that utilizes secondary data collected through questionnaires. The research focuses on farmers as the customers of PT. X fertilizer, with a sample size of 250 respondents. The research model is as follows:



Gambar 1. Research Model

Tables 1 and 2 display the variables and indicators that were employed.

Tabel 1. Research Variables and Indicators

Variables	Measurement Model	Indicator	Scale
Customer satisfaction (X_1)	Reflective	General Satisfaction (X_{11}) Customer Complaints Handling (X_{12}) Satisfaction of Service Received (X_{13}) Customer Rating (X_{14})	Interval
Farmer Demographics (X_2)	Formative	Farmer Age (X_{21}) Farmer's Last Education (X_{22}) Family Income Level (X_{23})	Ratio Ordinal Ordinal
Customer Engagement (Y_1)	Reflective	Conscious Attention (Y_{11}) Focused Participation (Y_{12}) Social Connection (Y_{13})	Interval
Repurchase Decision (Y_2)	Formative	Purchase Incentives (Y_{21}) Time from Purchase to Goods Received (Y_{22}) Seller Response Speed (Y_{23}) Transaction System (Y_{24})	Interval

The Farmer Demographic Variable (X_2) is a mixed data scale variable because there are 2 data scales at once in one variable, namely the ratio and ordinal scales which can be detailed in Table 2.

Tabel 1. Research Variables for Principal Component Analysis

Variables	Indicator	Observation Data (Score)
Farmer Demographics (X_2)	Farmer Age (X_{21})	Year No School (1)

	Farmer's Last Education (X_{22})	Elementary School (2) Junior High School (3) High School (4) Lecture (5)
	Family Income Level (X_{23})	$< Rp500{,}000$per month(1) $Rp500{,}000 - Rp1{,}000{,}000$ per month (2) $Rp1{,}000{,}001 - Rp3{,}000{,}000$ per month (3) $Rp3{,}000{,}001 - Rp10{,}000{,}000$ per month (4) $Rp10{,}000{,}000 - Rp30{,}000{,}001$ per month (5)

Spearman Rank Correlation PCA

Principal Component Analysis (PCA) is a statistical method that applies an orthogonal transformation to convert a dataset of potentially correlated variables into a new set of linearly uncorrelated variables known as principal components [11]. PCA reduces the dimensions of a dataset while preserving as much of its statistical variability as possible [12]. This is achieved by computing eigenvectors and eigenvalues from the covariance or correlation matrix of the input data. Through this process, PCA performs a linear transformation, converting the high-dimensional input vector into a lower-dimensional representation with uncorrelated components [13]. Spearman rank correlation PCA is a PCA by using a Spearman rank correlation matrix instead of the conventional Pearson correlation matrix as its input. Like PCA in general, Spearman rank correlation PCA is used to acquire data representing latent variables, namely in the form of principal component scores. Spearman rank correlation coefficient using equation (1).

$$r_{sj} = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{N^3 - N} \tag{1}$$

The following is the correlation matrix which can be seen in equation (2)

$$\boldsymbol{R} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1m} \\ r_{21} & 1 & \cdots & r_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & 1 \end{bmatrix} \tag{2}$$

information:

r_{sj} : correlation value between indicators in the same variable
$\boldsymbol{R}$: sample correlation matrix

Confirmatory Factor Analysis (CFA)

Confirmatory Factor Analysis (CFA) is a statistical method used to assess the measurement of latent variables by modeling their relationships with observed indicators believed to represent them [14]. CFA assesses whether the empirical evidence supports an imagined link between manifest indicators and latent components [15]. Latent variables, are those that are not directly measurable but can be constructed and shaped by factor loadings on the observable indicators [2]. Reflective models, also known as confirmatory factor models, use confirmatory factor analysis to extract latent variable data in the form of factor scores.

Ramsey's RESET Test

The relationship between independent and dependent variables must be linear, according to the linear regression assumption of linear relationships [16]. Ramsey first proposed the Regression Specification Error verify, or RESET, in 1969 as a way to verify the linearity assumption. The steps to apply RESET are as follows [17]:

Regress X_1 on Y_i. The equation of $\hat{Y}_i$ as an endogenous variable in the model is presented in the form of equation (3).

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} \tag{3}$$

Calculating the coefficient of determination according to equation (3) as R_1^2 which is presented in the equation (4).

$$R_1^2 = 1 - \frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \tag{4}$$

Regress X_1 on Y_i nd two additional predictor variables $\hat{Y}_i^2$ and $\hat{Y}_i^3$. Furthermore Y_i^* as the response variable is presented in equation (5).

$$Y_i^* = \beta_0^* + \beta_1^* X_{1i} + \beta_2 \hat{Y}_i^2 + \beta_3 \hat{Y}_i^3 + \varepsilon_i^* \tag{5}$$

Then, calculate $\hat{Y}_i^*$ as per the model in equation (6).

$$\hat{Y}_i^* = \hat{\beta}_0^* + \hat{\beta}_1^* X_{1i} + \hat{\beta}_2 \hat{Y}_i^2 + \hat{\beta}_3 \hat{Y}_i^3 \tag{6}$$

Calculate the coefficient of determination according to equation (6) as R_2^2 presented in equation (7).

$$R_2^2 = 1 - \frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i^*\right)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \tag{7}$$

Testing the linearity between predictor variables and the response with the following hypothesis.

$H_0 : \beta_2 = \beta_3 = 0$

$H_1 : there\ is\ at\ least\ one\ \beta_j \neq 0; j = 2,3$

with test statistics following the F distribution according to equation (8).

$$F_{value} = \frac{(R_2^2 - R_1^2)/m}{(1 - R_2^2)/(n - k - 1 - m)} \sim F_{m,n-k-1-m} \tag{8}$$

Information:
n : number of observations
k : the number of initial predictor variables
m : number of additional predictor variables
Drawing conclusions based on the results obtained with the criteria if the test statistic $F_{value} > F_{(\alpha,m,n-k-1-m)}$ or $p - value > \alpha$ then Accept H_0 means that the relationship between variables is linear.

Analysis Steps

The analysis methods used are Spearman Rank Correlation PCA and CFA in SEM. Data analysis in this study was carried out using the help of R Studio software. The analysis steps are as follows.

1. Identifying data based on the nature of the indicators of the latent variables, namely formative for the Farmer Demographics variable (X_2) and reflective for

Customer Satisfaction (X_1), Customer Engagement (Y_1), and Repurchase Decisions (Y_2)
2. Conducting Spearman Rank PCA on Variable Farmer Demographics (X_2) and CFA on Variable Customer Satisfaction (X_1), Customer Engagement (Y_1), and Repurchase Decisions (Y_2)
3. Creating a Structural Equation Modeling path diagram.
4. Convert path diagrams into systems of equations to form structural and measurement models.
5. Checking the linearity assumption of the structural model of SEM analysis with Ramsey's RESET Test
6. Estimating parameters using the PLS (Partial Least Square) method
7. Conducting hypothesis testing
8. Checking validity on structural models and measurement models
9. Interpreting the results

RESULTS AND DISCUSSION

Spearman Rank Correlation PCA

Variables that have indicators with mixed scales are analyzed using Spearman rank correlation PCA, is used to generate eigenvalues and component weights. Table 3 displays the resultant eigenvalues.

Table 3. Eigenvalues Spearman Rank Correlation PCA

Variables	Dimensions/PC	Eigenvalues	Diversity Proportion	Cumulative Percent Diversity
Farmer Demographics (X_2)	1	2,4004	0.7862	0.7862
	2	0.5245	0.1718	0.9580
	3	0.1307	0.0428	1

The first component, also known as PC1, which represents the latent variable under analysis, is the primary component that has the highest cumulative value of percent diversity on Table 3. In the X_2 variable, the main component produced is able to store 78.62% of the diversity of all indicators, while the remaining 21.38% is not stored (wasted).

From the eigenvalues, the component weights will be obtained. Table 3 shows the results of the component weights of each indicator. The component weights show the relative contribution of each indicator to the formation of each main component. The resulting component weights are presented in Table 4.

Table 4. Weights Spearman Rank Correlation PCA

Variables	Indicator	Component weight
Farmer Demographics (X_2)	Farmer Age (X_{21})	0.6354
	Farmer's Last Education (X_{22})	0.3482
	Family Income Level (X_{23})	0.6893

Based on the weight of the main components of Spearman rank correlation, the indicator variable Demographic Farmers (X_2) which has the greatest weight is the indicator of Farmer Income Level (X_{23}) which means that the indicator of Farmer Income Level (X_{23}) is the dominant indicator in shaping the Demographic Farmer Variable (X_2).

Table 4 allows for the creation of a linear combination equation of the primary components, which yields a component score that serves as the Farmer Demographics variable's value (X_2).

$$X_2 = 0.6354X_{2.1} + 0.3482X_{2.2} + 0.6893X_{2.3}$$

Confirmatory Factor Analysis

Variables that have indicators with interval scales and are reflective are analyzed using Confirmatory Factor Analysis. Table 5 shows the results of the loading values of each indicator.

Table 5. Confirmatory Factor Analysis Component Loadings

Variables	Indicator	Loading	P-Value
Customer satisfaction (X_1)	General Satisfaction (X_{11})	0.635	<0.001
	Customer Complaints Handling (X_{12})	0.630	<0.001
	Satisfaction of Service Received (X_{13})	0.674	<0.001
	Customer Rating (X_{14})	0.888	<0.001
Customer Engagement (Y_1)	Conscious Attention (Y_{11})	0.553	<0.001
	Focused Participation (Y_{12})	0.721	<0.001
	Social Connection (Y_{13})	0.418	<0.001
Repurchase Decision (Y_2)	Purchase Incentives (Y_{21})	0.726	<0.001
	Time from Purchase to Goods Received (Y_{22})	0.927	<0.001
	Seller Response Speed (Y_{23})	0.568	<0.001
	Transaction System (Y_{24})	0.139 ns	0.608

It is evident from Table 5 that in the Customer Satisfaction (X_1), and Customer Engagement (Y_1) variables, all indicators have a p-value smaller than 0.05 so that these indicators can reflect the Customer Satisfaction (X_1), Customer Engagement (Y_1) variables. Meanwhile, in the Repurchase Decision Variable (Y_2) there is a Transaction System indicator (Y_{24}) with a p-value> 0.05, which can be concluded that the Transaction System indicator does not reflect the Repurchase Decision variable so that the indicator does not need to be used in the analysis. Indicators that reflect the Repurchase Decision Variable (Y_2) are Item Purchase Incentives (Y_{21}), Purchase Time until Goods Received (Y_{22}) and Seller Response Speed (Y_{23}).

According to the factor loading values, it can be seen that the most dominant indicator in reflecting the variable is the indicator with the highest outer loading value. The indicator that best reflects the Customer Satisfaction variable (X_1) is Customer Assessment (X_{14}) with a loading of 0.888. In the Customer Engagement variable (Y_1), the most dominant indicator in reflecting the Customer Engagement Variable is Focused Participation (Y_1) with a loading of 0.721. Meanwhile, the indicator of Purchase Time until Goods Received (Y_{22}) is the most dominant indicator in reflecting the Repurchase Decision variable (Y_2) with a loading of 0.927.

SEM Model Path Diagram

The path diagram resulting from the design of the SEM Model can be seen in Figure 3.

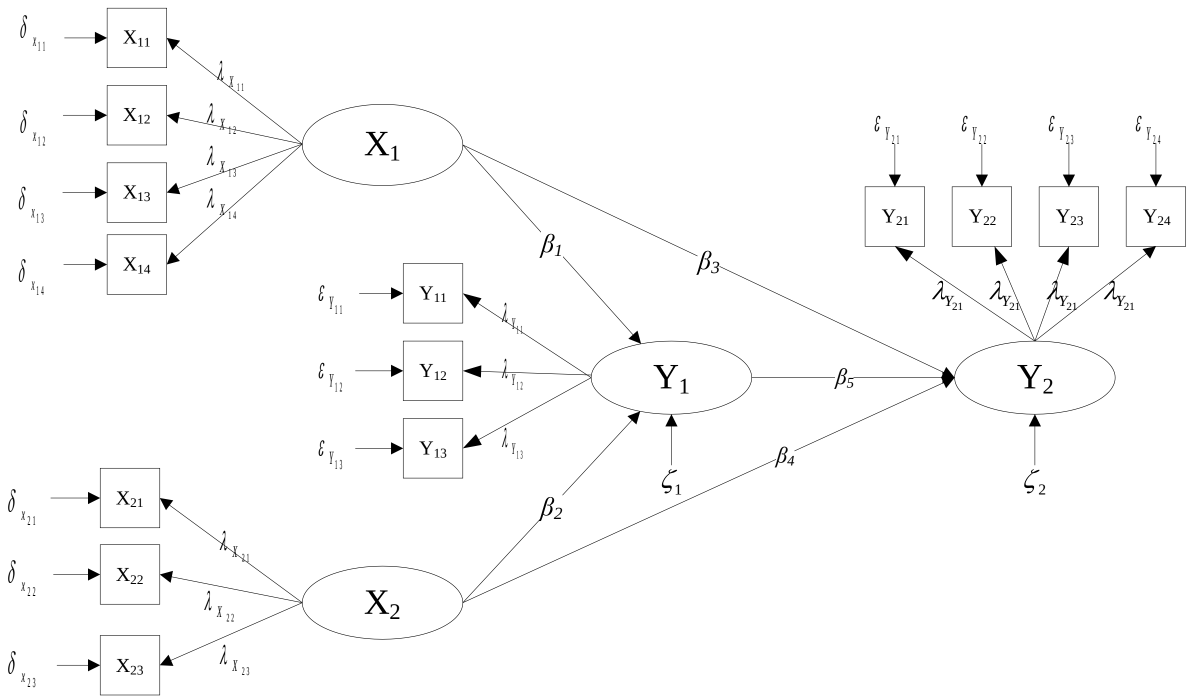


Figure 2. SEM Path Diagram

Information:

X_i	: exogenous latent variable to-i
Y_g	: endogenous latent variable to-g
X_{ij}	: exogenous variable to-i indicato r to-j
Y_{gk}	: endogenous variable to-g indicator to-k
λ_{xij}	: coefficient loading exogenous variable to-i indicator to-j
λ_{ygk}	: coefficient loading endogenous variable to-g indicator to-k
β	: coefficient of influence of latent variables
δ_{X_i}	: error measurement on exogenous latent variables
ε_{Y_g}	: error measurement on endogenous latent variables
ζ_g	: error g-model

Based on the SEM model, an inner model or structural model is formed as follows:

$$Y_{1i} = \beta_{01} + \beta_1 X_{1i} + \beta_1 X_{1i} + \zeta_{1i} \tag{9}$$

$$Y_{2i} = \beta_{02} + \beta_3 X_{1i} + \beta_4 X_{2i} + \beta_5 Y_{1i} + \zeta_{2i} \tag{10}$$

Which can be described in matrix form as follows:

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1n} \\ Y_{21} \\ Y_{22} \\ \vdots \\ Y_{2n} \end{bmatrix}_{2n\times 1} = \begin{bmatrix} 1 & X_{11} & X_{21} & 0 & 0 & 0 & 0 \\ 1 & X_{12} & X_{22} & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{1n} & X_{2n} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & X_{11} & X_{21} & Y_{11} \\ 0 & 0 & 0 & 1 & X_{12} & X_{22} & Y_{12} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 1 & X_{1n} & X_{2n} & Y_{1n} \end{bmatrix}_{2n\times 7} \begin{bmatrix} \beta_{01} \\ \beta_{1} \\ \beta_{2} \\ \beta_{02} \\ \beta_{3} \\ \beta_{4} \\ \beta_{5} \end{bmatrix}_{7\times 1} + \begin{bmatrix} \zeta_{11} \\ \zeta_{12} \\ \vdots \\ \zeta_{1n} \\ \zeta_{21} \\ \zeta_{22} \\ \vdots \\ \zeta_{2n} \end{bmatrix}_{2n\times 1}$$

Or it can be written in matrix notation, as in equation (11):

$$\boldsymbol{Y}_{2n\times1} = \boldsymbol{X}_{2n\times7}\boldsymbol{\beta}_{7\times1} + \boldsymbol{\zeta}_{7n\times1} \tag{11}$$

Results of Structural Model Linearity Assumption Test

The assumptions in SEM analysis are only related to structural modeling, where the relationship between latent variables in the structural model is linear. Table 6 displays the outcomes of the linearity assumption test using Ramsey's RESET Test.

Table 6. Linearity Test Results

Variables	P-value	Connection
X1 to Y1	0.6326	Linear
X1 to Y2	0.9954	Linear
X2 to Y1	0.7224	Linear
X2 to Y2	0.8063	Linear
Y1 to Y2	0.8932	Linear

Table 6 shows that a p-value > α (5%) is produced by the link between latent variables, so it is decided that H_0 is accepted and it is concluded that all relationships between latent variables in the fertilizer repurchase decision data are linear.

Parameter Estimation and Hypothesis Testing of Structural Models

Hypothesis testing of the structural model using the t-test. The results of the structural model hypothesis testing can be seen in Table 7. and Figure 3. below.

Table 7. Results of Structural Model Hypothesis Testing

Relationship Between Variables	Path Coefficient	p-value	Information
X1 to Y1	0.0625	0.043	Significant
X2 toY1	0.2662	<0.001	Significant
X1 to Y2	0.6879	<0.001	Significant
X2 to Y2	0.0815	<0.001	Significant
Y1 to Y2	0.1333	<0.001	Significant

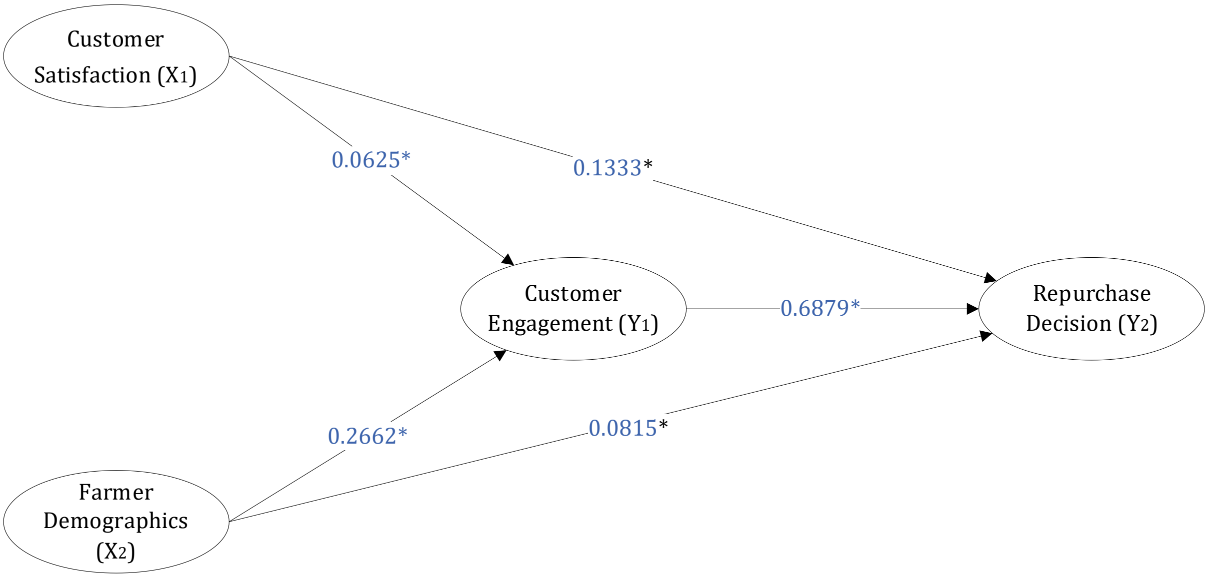


Figure 3. Structural Model Hypothesis Testing

Based on the inner model path coefficients, the function estimate is obtained through:

$$Y_{1i} = \beta_{01} + \beta_1 X_{1i} + \beta_1 X_{1i} + \zeta_{1i}$$

$$Y_{2i} = \beta_{02} + \beta_3 X_{1i} + \beta_4 X_{2i} + \beta_5 Y_{1i} + \zeta_{2i}$$

By performing standardization, the following equation is produced:

$$Y_{Y1} = \beta_1 Z_{X1} + \beta_2 Z_{X2} + \zeta_1$$

$$Y_{Y2} = \beta_3 Z_{X1} + \beta_4 Z_{X2} + \beta_5 Z_{Y1} + \zeta_2$$

So the estimated function of the Fertilizer Repurchase Decision is:

$$Z_{Y_1} = 0.0625\ Z_{X_1} + 0.2662 Z_{X_2}$$

$$Z_{Y_2} = 0.6879 Z_{X_1} + 0.0815 Z_{X_2} +\ 0.1333 Z_{Y1}$$

The relationship between Customer Satisfaction (X_1) and Customer Engagement (Y_1) is significant at the 5% significance level with a coefficient value of 0.1333. This indicates that the higher the level of customer satisfaction, which includes general satisfaction, customer complaint handling, service satisfaction, and customer assessment, the stronger the attachment felt by customers (farmers). That is, when customers are satisfied, they tend to have a stronger engagement to the product or service they use. The relationship between Customer Satisfaction (X_1) and Repurchase Decision (Y_2) is significant, with a coefficient of 0.0625. This shows that the level of customer satisfaction affects their decision to repurchase. This indicates that the higher the level of customer satisfaction which includes general satisfaction, customer complaint handling, service satisfaction, and customer assessment, the higher the likelihood of farmers to repurchase the fertilizer. Furthermore, the relationship between Farmer Demographics (X_2) and Customer Engagement (Y_1) is significant with a coefficient of 0.2662. This indicates that the demographic characteristics of farmers play an important role in increasing customer engagement, meaning that the demographic characteristics of farmers, namely the age, the latest education and the level of family income of farmers have an important role in shaping customer interest and attachment to fertilizer products. In addition, the relationship between Farmer Demographics (X_2) and Repurchase Decision (Y_2) is significant, with a coefficient of 0.0815. This confirms that farmers' demographic characteristics, such as age, latest education level, and family income, play an important role in determining fertilizer repurchase decisions. Furthermore, the relationship between Customer Engagement (Y_1) and Repurchase Decision (Y_2) is significant with a coefficient value of 0.6879. This means that the higher the engagement felt by farmers, which includes conscious Attention, focused participation, and social connection, the higher the likelihood of farmers to make repeat purchases. Customer attachment plays an important role in shaping repurchase decisions, where when customers feel emotionally or functionally attached to a product or service, they are more likely to consider using it again in the future. This confirms that efforts to increase customer attachment can be an effective strategy in retaining customers to buy fertilizer products again.

Structural Model Evaluation

The R-Square value or coefficient of determination of each endogenous variable in each model can be used to evaluate the structural model. Table 8 displays the endogenous variables' R-Square values.

Table 8. R-Square Structural Model

Coefficient of Determination Y_1	Coefficient of Determination Y_2	Total Determination Coefficient
0.3222	0.6291	1 – (0.3222)(0.6291) = 0.7973≈ 0.8

Based on Table 8, the total determination coefficient value is 0.8, which means that the model can explain 80% of the data diversity or varians, while the remaining 20% is explained by other variables outside the research model.

CONCLUSIONS

Based on the results and discussion, it can be concluded that the Spearman Rank Correlation PCA showed good performance, capturing 78.62% of the variance for mixed scale indicators, namely in the Farmer Demographic Variable (X_2). In addition, the SEM model confirmed significant relationships among variables with a coefficient of determination of 80%, which provides insight into customer (farmer) behavior in making fertilizer purchase decisions. The result showed that Spearman Rank Correlation PCA was successfully applied to Structural Equation Modeling (SEM) to handle data with mixed scaled indicators (metric and non-metric) and was able to be integrated with CFA to handle data with mixed measurement models (reflective and formative). This demonstrates the flexibility of SEM in handling data complexity well, while maintaining estimation accuracy, and provides new opportunities to properly understand the relationships between latent variables in more complex models, especially in studies with mixed-scale data.

REFERENCES

[1] Solimun, Fernandes A. A. R., Nurjannah, "*Metode Statistika Multivariat Pemodelan Persamaan Struktural (SEM)*", UB Press, 2017. https://opac.perpusnas.go.id/DetailOpac.aspx?id=1141341
[2] Hair J.F., Black W. C., Babin B. J., and Anderson R. E., "*Multivariate Data Analysis (Eighth Edition)*", vol. 19, no. 3, 2019. DOI: 10.5117/2006.019.003.007.
[3] Ryu E., "Model fit evaluation in multilevel structural equation models," *Front.*

Psychol., vol. 5, no. FEB, pp. 1–9, 2014. DOI: 10.3389/fpsyg.2014.00081.

[4] Albahri A. S., Alnoor, A., Zaidan, A. A., Albahri, O. S., Hameed, H., Zaidan, B. B., Peh, S. S., Zain, A. B., Siraj, S. B., Alamoodi, A. H., & Yass, A. A.., "Based on the multi-assessment model: Towards a new context of combining the artificial neural network and structural equation modelling: A review," *Chaos, Solitons and Fractals*, vol. 153, p. 111445, 2021. DOI: 10.1016/j.chaos.2021.111445.

[5] Chen H., and Wang Y., "A Penalized Spline Approach to Functional Mixed Effects Model Analysis," *Biometrics*, vol. 67, no. 3, pp. 861–870, 2011. DOI: 10.1111/j.1541-0420.2010.01524.x.

[6] Watanabe K., Naito K., and Koch I., "Robustness of Principal Component Analysis with Spearman's Rank Matrix," *J. Stat. Theory Pract.*, vol. 18, no. 1, pp. 1–29, 2024. DOI: 10.1007/s42519-023-00358-z.

[7] Solimun, Fernandes A. A. R., Nurjannah, Erwinda E. G., Hardianti R., and Arini L. H. Y., "*Metodologi Penelitian Variabel Mining berbasis Big Data dalam Pemodelan Sistem untuk mengungkap Research Novelty*", UB Press, 2023. https://books.google.co.id/books/about/Metodologi_Penelitian.html?id=GfXaEAAAQBAJ&redir_esc=y

[8] Yu H., Khan F., and Garaniya V., "A Sparse PCA for Nonlinear Fault Diagnosis and Robust Feature Discovery of Industrial Processes," *AIChE J.*, vol. 62, no. 5, pp. 215–228, 2016. DOI: 10.1002/aic.

[9] Ganguli S., and Roy S. K., "Generic technology-based service quality dimensions in banking: Impact on customer satisfaction and loyalty," *Int. J. Bank Mark.*, vol. 29, no. 2, pp. 168–189, 2011. DOI: 10.1108/02652321111107648.

[10] Bansal R., and Chaudhary K., "Customer Engagement- A Literature Review," *Glob. Int. Res. Thoughts*, vol. 2, no. 1, pp. 15–20, 2016. [Online]. Available: https://www.researchgate.net/publication/318224473

[11] Karamizadeh S., Abdullah S. M., Manaf A. A., Zamani M., and Hooman A., "An Overview of Principal Component Analysis," *J. Signal Inf. Process.*, vol. 04, no. 03, pp. 173–175, 2013. DOI: 10.4236/jsip.2013.43b031.

[12] Jollife I. T., and Cadima J., "Principal component analysis: A review and recent developments," *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 374, no. 2065, 2016. DOI: 10.1098/rsta.2015.0202.

[13] Kong X., Hu C., and Duan Z., "*Principal component analysis networks and algorithms*", 2017. DOI: 10.1007/978-981-10-2915-8.

[14] Roos J. M., and Bauldry S., "*Confirmatory Factor Analysis*. United States of America", SAGE Publications, 2022. https://books.google.co.id/books/about/Confirmatory_Factor_Analysis.html?id=sT1KEAAAQBAJ&redir_esc=y

[15] Goretzko D., Siemund K., and Sterner P., "Evaluating Model Fit of Measurement Models in Confirmatory Factor Analysis," *Educ. Psychol. Meas.*, vol. 84, no. 1, pp. 123–144, 2024. DOI: 10.1177/00131644231163813.

[16] Schreiber-Gregory D., and Bader K., "Logistic and Linear Regression Assumptions: Violation Recognition and Control," *SESUG Pap. 247*, no. May, pp. 1–22, 2018, [Online]. Available: https://www.researchgate.net/publication/341354759

[17] Fernandes A. A. R., Solimun, Nurjannah, and Amaliana L., "*Analisis Regresi Dalam Pendekatan Fleksibel (Ilustrasi dengan Paket Program R)*", 2021. https://books.google.co.id/books/about/Analisis_Regresi_dalam_Pendekatan_Fleksi.html?id=SKFgEAAAQBAJ&redir_esc=y