BAREKENG: Journal of Mathematics and Its Applications
June 2025 Volume 19 Issue 2 Page 1193-1202
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol19iss2pp1193-1202

# DEVELOPMENT OF CLUSTER INTEGRATION WITH VARIAN BASED STRUCTURAL EQUATION MODELING TO MANAGE HETEROGENEOUS DATA

**Hanifa Sepriadi [1*], Atiek Iriany[2], Solimun[3], Adji Achmad Rinaldo Fernandes[4]**

[1,2,3,4]*Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Brawijaya*
*Jln. Veteran Ketawanggede, Lowokwaru, Malang, 65145, Indonesia*

*Corresponding author's e-mail: ** ***Sepriadi1412@student.ub.ac.id***

*ABSTRACT*

***Article History:***

*Received: 28th September 2024*
*Revised: 3rd February 2025*
*Accepted: 5th March 2025*
*Published: 1th April 2025*

***Keywords:***

*Cashless Society;*
*Cluster;*
*Heterogeneous;*
*Structural Equation Modeling.*

*In the application of SEM to multivariate data, the individuals collected not only come from the same population but also from several groups (clusters). This data is heterogeneous. When SEM is applied to heterogeneous data, there will be a risk of bias in estimating equations in the measurement and structural models because there are differences between groups in the data. The purpose of this study is to overcome heterogeneous data in modeling cashless behavior with cluster using a dummy approach. This study used primary data from a survey in Bekasi City using a questionnaire with 100 respondents. Based on the study's results, it is known that using clustering in SEM can overcome heterogeneous data, which is indicated by the high coefficient of determination of 96.12%. Banks can use the results of this study to design products and services that are more in line with customer needs and preferences while encouraging financial inclusion in the digital era.*



***How to cite this article:***

H. Sepriadi, A. Iriany, Solimun and A. A. R. Fernandes., "DEVELOPMENT OF CLUSTER INTEGRATION WITH VARIAN BASED STRUCTURAL EQUATION MODELING TO MANAGE HETEROGENEOUS DATA," *BAREKENG: J. Math. & App.*, vol. 19, iss. 2, pp. 1193-1202, June, 2025.


Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

**Research Article · Open Access**

## 1. INTRODUCTION

Multivariate analysis is an analysis method that deals with relationships between variables simultaneously [1]. According to Solimun [2], one of the characteristics of multivariate analysis lies in the use of methods that relate to a large number of variables (multivariable) and are obtained simultaneously (not separately) from each research object. Multivariate techniques can be classified into 2: dependency and interdependence. One application of dependency techniques is the use of SEM analysis.

SEM analyzes the complex relationship of many factors to accommodate the form of reciprocal relationships [3]. SEM is a development of regression and path analysis, which simultaneously tests the relationship between variables and indicator models [2]. SEM is a multivariate analysis method used to determine the relationship between variables that cannot be measured directly (latent variables). SEM analysis is complex because it involves several exogenous and endogenous variables that are interconnected to form a model. SEM combines two models: the inner model (structural model) and the outer model (measurement model).

In the application of SEM to multivariate data, the individuals collected not only come from the same population but also from several groups (clusters). This data is heterogeneous. When SEM is applied to heterogeneous data, there will be a risk of bias in estimating equations in the measurement model and structural model because there are differences between groups in the data [4]. There are various approaches to overcoming the problem of heterogeneous data, such as cluster analysis, normalization, ensemble algorithms, and deep learning.

Cluster analysis is one of the multiple variables (multivariate) analyses included in the interdependence method; namely, the independent or explanatory variables are not distinguished from the dependent or response variable [5]. Cluster analysis aims to group objects that have the same properties into the same cluster, where between clusters have different properties [6]. In general, there are two methods in cluster analysis, namely hierarchical methods and non-hierarchical methods. In this study using non-hierarchical cluster analysis to determine the clusters to be formed as many as two clusters, the formation of these two clusters is based on an engagement strategy with active and passive consumer behavior in using mobile banking.

As the use of cashless payments increases, new challenges and opportunities arise in the banking sector. Banks must adapt to these changes in consumer behavior by developing services that better suit user needs. A deep understanding of the factors influencing cashless behavior is essential to formulating the right strategy. Cashless behavior is becoming an increasingly relevant topic, especially after the COVID-19 pandemic. The COVID-19 pandemic has driven significant changes in people's habits, including how they transact. The need to reduce physical contact and minimize the spread of the virus has made people switch from cash payments to non-cash payments, either through cards, digital wallets, or other payment applications [7]. This change is not only seen in big cities but also in areas that previously relied more on cash transactions.

This research aims to discover and develop Cluster-Structural Equation Modeling to overcome heterogeneous data on cashless behavior. Banks can use the results of this study to design products and services that are more in line with customer needs and preferences while encouraging financial inclusion in the digital era.

## 2. RESEARCH METHODS

### 2.1 Data

The data used in this study is secondary data from Solimun's research grant [8]. The sample unit in this study is BNI customers in Jakarta who use the mobile banking application, so the population of this study is all BNI Bank customers in Jakarta. The population in this study was 1500 BNI customers in Jakarta. The sampling technique used was nonprobability sampling because the survey was conducted directly. Therefore, this study uses quota sampling to obtain heterogeneous data.

The sample size in this study was set at 100 BNI customers who used Mobile Banking. The research instrument used in this research is a questionnaire. The research questionnaire was distributed through a community survey using a Google form with a Likert scale. The variables measured are Economy (**$X_1$**), Adoption (**X₂**), Financial Technology (**Y₁**), and Cashless Behavior (**Y**₂). Researchers checked All variables for validity and reliability, and the results showed that all questionnaire items were valid and reliable so that they could be used for further analysis.

## 2.2 Research Model and Research Step

The research model used can be seen in **Figure 1**.

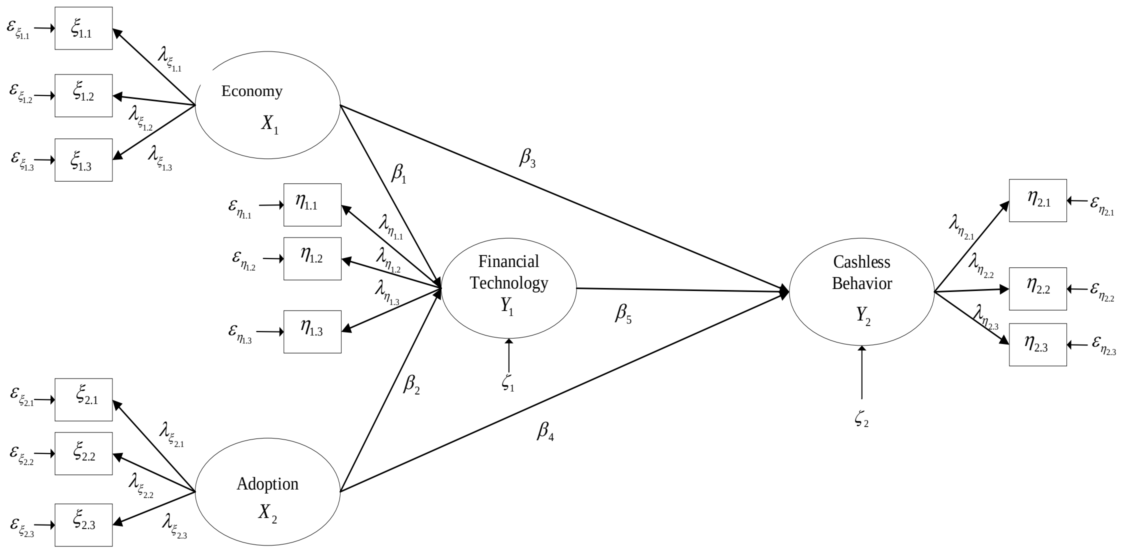


**Figure 1. Research Model**

The model used in **Figure 1** will use Cluster Analysis and Structural Equation Modeling. This study uses the help of R Studio Software. The steps in this study are as follows:

a. Prepare secondary data.
b. Forming dummy variables from the clusters formed to distinguish between clusters. After the dummy variable is formed, it is multiplied by each variable.
c. From the cluster results, a linearity test is carried out using Ramsey RESET.
d. Perform confirmatory factor analysis on reflective indicators of each variable and form a measurement model.
e. Perform path analysis to form a structural model.
f. Conducting hypothesis testing of the SEM function of the best model with two-way t-test statistics using replication values and standard errors generated from the jackknife resampling process.
g. Calculating the direct effect, indirect effect, and total effect.
h. Interpreting results.

## 2.3 K-Means Cluster

The K-means algorithm is a non-hierarchical cluster analysis method that will allocate objects into k groups based on the nearest mean. In addition, grouping within the same cluster is based on similar characteristics so that objects with different characteristics will be grouped into other clusters [9]. Thus, the goal of clustering is to minimize variation within a cluster and maximize variation between clusters.

K-means cluster analysis starts with randomly selecting initial cluster centers from a set of objects. Then, each object is tested and assigned to one of the predefined cluster centers depending on the minimum distance between objects and each cluster. This study uses Manhattan distance because Manhattan distance has advantages in clustering heterogeneous data, which lies in its ability to handle scale differences between variables with more robustness to outliers than Euclidean distance and is more effective in high-dimensional space. After all, it only considers absolute differences in each dimension. The distance used is the Manhattan that wrote in **Equation (1)**:

$$d(X_i, X_j) = \sum_{r=1}^{p} |X_{ri} - X_{rj}| \tag{1}$$

Explanation:

$d(X_i, X_j)$ : distance between $i$ and $j$

$X_{ri}$ : the value of variable *r* in the *i*-th observation

$X_{rj}$ : the value of variable *r* in the *j*-th observation

$p$ : number of data variables

The cluster this study uses dummy variables to present the qualitative nature of the data as quantitative. This study uses two clusters, and we obtained the dummy function in **Equation (2)**:

$$D = \begin{cases} 1, & \text{cluster 1} \\ 0, & \text{other} \end{cases} \tag{2}$$

### 2.4 Structural Equation Modeling (SEM)

SEM is a multivariate statistical technique that examines the relationship between variables in a model, both between indicators and latent variables, and between latent variables [10]. SEM is a combination of structural models and measurement models simultaneously [11]. In this illustration, variable X is an exogenous variable, and variable Yi is an endogenous variable. Exogenous variables are variables whose values are determined outside the model, while endogenous variables are variables whose values are determined in the model.

The integration of k-means cluster with SEM can be seen in the **Equation (3)**:

$$\begin{aligned} Y_{1i} &= \beta_{01} + \beta_{11}X_{1i} + \beta_{21}X_{2i} + \beta_{31}D + \beta_{41}D_iX_{1i} + \beta_{51}D_iX_{2i} + \varepsilon_{1i} \\ Y_{2i} &= \beta_{02} + \beta_{12}X_{1i} + \beta_{22}X_{2i} + \beta_{32}Y_{1i} + \beta_{42}D_i + \beta_{52}D_iX_{1i} + \beta_{62}D_iX_{2i} + \beta_{72}D_iY_{1i} + \varepsilon_{2i} \end{aligned} \tag{3}$$

Explanation:

$Y_i$ : endogenous latent variable at the $i$ observation

$\beta$ : coefficient of influence of latent variables

$x$ : exogenous latent variable

$D$ : dummy variable

$\varepsilon$ : error

### 2.5 Resampling

Resampling is the process of re-sampling from existing samples to obtain new samples. The new samples are obtained from the original samples taken randomly, either with or without replacement. Applying the resampling method allows the validity of data free from assumptions or, in other words, does not require the assumption of normality. The jackknife method is a resampling method without returns. Therefore, there is an intertwined relationship in each resampling process. Suppose there is an initial sample $x = (x_1, x_2, \ldots, x_n)$ and $\hat{\theta} = s(x)$ is the estimate of a parameter. The steps for estimating the standard error of the jackknife are as follows [12].

a. Resampling by deleting 1 row of data in each jackknife sample.

$$x_{(i)} = x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$$

b. Calculating the corresponding jackknife replications for each jackknife sample.

$$\hat{\theta}_{(i)} = s(x_{(i)}); i = 1, 2, \ldots, n \tag{4}$$

c. Estimating the standard error using the standard deviation for the jackknife replicated n times.

$$\widehat{SE}_{jack} = \left[ \frac{n-1}{n} \sum_{i=1}^{n} \left( \hat{\theta}_{(i)} - \hat{\theta}_{(.)} \right)^2 \right]^{1/2} \tag{5}$$

### 2.6 Hypothesis Testing

Hypothesis testing using test statistics, where parameter estimates and standard errors from jackknife resampling results. Hypothesis testing with test statistics is done using the following formula:

$$t\text{-test statistics} = \frac{\hat{\beta}_j}{SE_{\hat{\beta}_j}} \sim t_{n-1} \tag{6}$$

The hypothesis used for the test statistics in **Equation (6)** is as follows.

$H_0 : \beta_j = 0$

$H_1 : \beta_j \neq 0$

Furthermore, the results of the t-test statistics are compared with the t-table. The test criteria, namely if the test statistic $t > t_{\frac{\alpha}{2}(n-1)}$ then $H_0$ is rejected, which means that there is a significant influence between exogenous variables on endogenous variables.

### 2.7 Model Validity

One way to test the fit of the model is to calculate the coefficient of total determination **[13]**. The SEM model is evaluated by looking at the goodness of fit model value. The goodness of fit model in question is an index and measure of the goodness of the relationship between latent variables. One way to determine the value of the goodness of fit model is to look at the presentation of explained variance through the coefficient of determination ($R^2$) for endogenous latent constructs. Q-square predictive relevance ($Q^2$) is used for structural models that can measure how well the observed value is produced by the model. The magnitude of $Q^2$ has a value range of $0 < Q^2 < 1$, where the closer to 1 means the better the model. The magnitude of $Q^2$ is equivalent to the coefficient of total determination. The calculation of $Q^2$ can be seen as in **Equation (7)**.

$$Q^2 = 1 - \left(1 - R^2_{1,adj}\right)\left(1 - R^2_{2,adj}\right)\cdots\left(1 - R^2_{p,adj}\right) \tag{7}$$

With the coefficient of determination ($R^2_{adj}$) can be seen as in **Equation (8)**.

$$R^2_{p,adj} = 1 - \left( \frac{\sum_{i=1}^{n} (y_{pi} - \hat{y}_{pi})^2 / (n-k-1)}{\sum_{i=1}^{n} (y_{pi} - \bar{y}_p)^2 / (n-1)} \right) \tag{8}$$

Explanation:

$R^2_{p,adj}$: corrected coefficient of determination in the *p*-th inner model equation

$y_{pi}$ : *i*-th value of the *p*-th endogenous variable

$\hat{y}_{pi}$ : *i*-th function estimator for the *p*-th endogenous variable

$\bar{y}_p$ : average of endogenous variables *n* : numbers of observation

*k* : number of exogenous variables in the model

### 2.8 Research Variable

The research uses latent variables from the Likert measurement scale. The variables measured are Economy $(\boldsymbol{X_1})$, Adoption $(\boldsymbol{X_2})$, Financial Technology $(\boldsymbol{Y_1})$, and Cashless Behavior $(\boldsymbol{Y_2})$. The variables used will be described into several indicators as in **Table 1**.

**Table 1. Research Variables**

| Variable | Indicators |
|---|---|
| Economy $(X_1)$ | Digitalization $(X_{1.1})$<br>Transaction $(X_{1.2})$<br>System $(X_{1.3})$ |
| Adoption $(X_2)$ | Acceptance $(X_{2.1})$<br>Perception $(X_{2.2})$<br>Security $(X_{2.3})$ |
| Financial Technology $(Y_1)$ | Ease of Use $(Y_{1.1})$<br>Accessibility of Technology $(Y_{1.2})$<br>User Satisfaction $(Y_{1.3})$ |
| Cashless Behavior $(Y_2)$ | Frequency of Use $(Y_{2.1})$<br>Contextual Factors $(Y_{2.2})$<br>Habit $(Y_{2.3})$ |

## 3. RESULTS AND DISCUSSION

### 3.1 K-Means Cluster

Cluster analysis aims to classify and organize objects with various characteristics based on specific characteristics. Therefore, using non-hierarchical cluster analysis, cluster analysis will be carried out to classify customers who use mobile banking using the K-Means method with the Manhattan distance measure. Manhattan distance measure can be obtained using **Equation (1)** and then grouped with K-Means. The results of the centroid calculation for each cluster are shown in **Table 2**.

**Table 2. K-Means Cluster Analysis Centroid**

| Variable | Cluster 1 | Cluster 2 |
|---|---|---|
| **Economy $(X_1)$** | 2.858 | 2.202 |
| **Adoption $(X_2)$** | 2.909 | 2.291 |
| **Financial Technology $(Y_1)$** | 2.897 | 2.204 |
| **Cashless Behavior $(Y_2)$** | 2.867 | 2.125 |

K-Means method from three cluster presented in **Figure 2**.

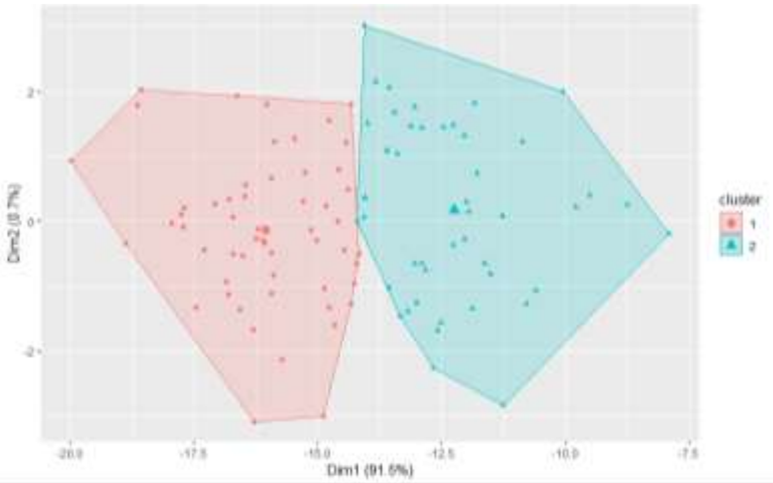


**Figure 2. Result K-Means Cluster**

**Figure 2** shows that customers are divided into two clusters using the K-Means method, and the distance used is Manhattan Distance. Cluster 1 is shown in red; Cluster 2 is shown in blue. Details of the members in each cluster can be seen in **Table 3**.

**Table 3. Total Member from Two K-Means Cluster**

| Cluster | Costumer Group Code | Total Member Cluster |
|---|---|---|
| **Cluster 1** | 1, 8, 10, 14, 15, 16, 17, 18, 20, 21, 22, 24, 28, 29, 30, 31, 32, 33, 34, 35, 38, 39, 40, 41, 43, 44, 47, 48, 49, 50, 52, 53, 54, 55, 56, 57, 60, 61, 67, 69, 70, 76, 78, 81, 82, 83, 84, 85, 86, 87, 88, 92, 95, 96, 100 | 55 |
| **Cluster 2** | 2, 3, 4, 5, 6, 7, 9, 11, 12, 13, 19, 23, 25, 26, 27, 36, 37, 42, 45, 46, 51, 58, 59, 62, 63, 64, 65, 66, 68, 71, 72, 73, 74, 75, 77, 79, 80, 89, 90, 91, 93, 94, 97, 98, 99 | 45 |

### 3.2 Ramsey RESET

The assumption test that must be met in SEM is linearity between the variables used. If the assumptions are met, then the approach used is parametric. Still, if the assumptions are not met, modeling will be carried out using a semiparametric or non-parametric approach. One method for testing the linearity assumption is the Regression Specification Error Test or RESET, which Ramsey introduced in 1969 **[14]**.

RESET linearity testing uses the following hypothesis.

$H_0$: $\beta_{p+1} = \beta_{p+2} = \cdots = \beta_{p+m}$

$H_1$: There is at least one difference, $\beta_{p+j} j = 1,2, \ldots, m$.

**Table 4. Ramsey Reset Result**

| Relationships | P-value | | | Results |
|---|---|---|---|---|
| | Overall | Group 1 | Group 2 | |
| Economy → Financial Technology | 0.408 | 0.619 | 0.519 | Linear Relationships |
| Adoption → Financial Technology | 0.534 | 0.275 | 0.099 | Linear Relationships |
| Economy → Cashless Behavior | 0.784 | 0.332 | 0.473 | Linear Relationships |
| Adoption → Cashless Behavior | 0.663 | 0.522 | 0.606 | Linear Relationships |
| Financial Technology → Cashless Behavior | 0.550 | 0.299 | 0.316 | Linear Relationships |

Based on the linearity test using RAMSEY RESET, it can be seen that all direct relationships between exogenous latent variables and endogenous variables show a linear relationship in each cluster; both Cluster 1 and Cluster 2 show linear relationship results.

### 3.3 Structural Equation Modeling (SEM)

In SEM there are two models, namely the outer model and the inner model **[15]**, the following are the results of the outer model.

**Table 5. Outer Model Result**

| Variables | Indicators | Loading Factor | | | | P-value | Composite Reliability |
|---|---|---|---|---|---|---|---|
| | | Group 1 | | Group 2 | | | |
| Economy | Digitalization | 0.767 | * | 0.647 | * | 0.020 | 0.607 |
| | Transaction | 0.731 | * | 0.633 | * | 0.043 | 0.557 |
| | System | 0.690 | * | 0.658 | * | 0.032 | 0.588 |
| Adoption | Security | 0.615 | * | 0.617 | * | 0.025 | 0.487 |
| | Acceptance | 0.715 | * | 0.695 | * | 0.037 | 0.435 |
| | Perception | 0.522 | * | 0.732 | * | 0.040 | 0.420 |

| Variables | Indicators | Loading Factor | | | | P-value | Composite Reliability |
|---|---|---|---|---|---|---|---|
| | | Group 1 | | Group 2 | | | |
| Financial Technology | Ease of Use | 0.726 | * | 0.710 | * | 0.046 | 0.401 |
| | Accessibility of Technology | 0.717 | * | 0.689 | * | 0.042 | 0.516 |
| | User Satisfaction | 0.774 | * | 0.761 | * | 0.035 | 0.615 |
| Cashless Behavior | Frequency of Use | 0.728 | * | 0.788 | * | 0.030 | 0.586 |
| | Contextual Factors | 0.741 | * | 0.558 | * | 0.036 | 0.525 |
| | Habit | 0.749 | * | 0.761 | * | 0.041 | 0.449 |

Table 5 shows the indicators with each variable's most significant factor loading values. For economic variables, the digitalization indicator has a loading factor of 0.767 (Group 1) and 0.647 (Group 2), highlighting the critical role of digitalization in the cashless economy, with a significant p-value of 0.020. The transaction and system indicators also exhibit strong contributions, with loading factors of 0.731 and 0.690 (Group 1), and 0.633 and 0.658 (Group 2), with all the p-values below 0.05.

Regarding the adoption variable, the security and acceptance indicators have significant loading factors, with values of 0.615 and 0.715 (Group 1), and 0.617 and 0.695 (Group 2), and significant p-values of 0.025 and 0.037, respectively. The perception indicator shows a lower contribution in Group 1 (0.522) but a stronger influence in Group 2 (0.732), with a significant p-value of 0.040, indicating a difference in the perception of technology adoption between groups.

For the financial technology variable, the ease of use and accessibility of technology indicators exhibit strong contributions with loading factors of 0.726 and 0.717 (Group 1), and 0.710 and 0.689 (Group 2), with significant p-values of 0.046 and 0.042. User satisfaction has the strongest impact, with loading factors of 0.774 (Group 1) and 0.761 (Group 2), showing that user satisfaction is a key factor in using cashless financial technology.

Lastly, in the cashless behavior variable, frequency of use has a loading factor of 0.728 (Group 1) and 0.788 (Group 2), indicating that frequency of use is an essential indicator of cashless behavior, with a p-value of 0.030. The contextual factors indicator plays a stronger role in Group 1 (0.741) but a weaker one in Group 2 (0.558), showing variation in the influence of contextual factors. The habit indicator demonstrates strong contributions in both groups, with loading factors above 0.7 and a significant p-value of 0.041.

Inner models are specifications of the relationships between latent variables. Latent variables and indicators, often called manifest variables, can be standardized without losing their general characteristics. In this study, the inner model hypothesis was tested using the t-test with a real level of 5% for direct and indirect effects.

**Table 6. Inner Model Result**

| Relationships | Coefficient | | P value | |
|---|---|---|---|---|
| | Group 1 | Group 2 | Group 1 | Group 2 |
| Economy->Financial Technology | 0.279 | 0.253 | 0.005 | 0.011 |
| Adoption->Financial Technology | 0.368 | 0.321 | 0.007 | 0.027 |
| Economy ->Cashless Behavior | 0.217 | 0.276 | 0.001 | 0.006 |
| Adoption -> Cashless Behavior | 0.257 | 0.312 | 0.010 | 0.008 |
| Financial Technology -> Cashless Behavior | 0.280 | 0.331 | 0.005 | 0.001 |

Group 1:

$$\begin{aligned} Y_{1i} &= 0.279X_{1i} + 0.368X_{2i} \\ Y_{2i} &= 0.217X_{1i} + 0.257X_{2i} + 0.280Y_{1i} \end{aligned} \quad (9)$$

Group 2:

$$\begin{aligned} Y_{1i} &= 0.253X_{1i} + 0.321X_{2i} \\ Y_{2i} &= 0.276X_{1i} + 0.321X_{2i} + 0.331Y_{1i} \end{aligned} \quad (10)$$

In Group 1, the relationship between economy and financial technology has a coefficient of 0.279, with a highly significant p-value of 0.005. This indicates that economic factors have a moderate but statistically significant positive influence on financial technology adoption in cashless behavior. The relationship between adoption and financial technology is even stronger, with a coefficient of 0.368 and a significant p-value of 0.007. This suggests that technology adoption plays a vital role in facilitating the use of financial technology in Group 1.

When looking at the direct influence of the economy on cashless behavior, the coefficient is 0.217 with a very significant p-value of 0.001. The significance implies that economic factors have a moderately positive and significant effect on promoting cashless behavior in this group. The impact of adoption on cashless behavior has a coefficient of 0.257 and a p-value of 0.010, indicating that adoption of technology significantly positively influences encouraging cashless behavior, though slightly less than its effect on financial technology. Finally, the relationship between financial technology and cashless behavior is also strong, with a coefficient of 0.280 and a highly significant p-value of 0.005. The significance shows that financial technology has a meaningful and direct impact on driving cashless behavior in Group 1.

In Group 2, the relationship between economy and financial technology has a coefficient of 0.253, with a significant p-value of 0.011. This significance shows that economic factors still positively influence financial technology adoption, although slightly less than in Group 1. The effect of adoption on financial technology in Group 2 is also positive, with a coefficient of 0.321 and a p-value of 0.027. While this relationship is significant, it is slightly weaker than in Group 1, suggesting that adoption plays a minor role in driving financial technology use.

The relationship between economy and cashless behavior is more substantial in Group 2, with a coefficient of 0.276 and a p-value of 0.006, indicating a moderate and significant positive impact of economic factors on cashless behavior, even more so than in Group 1. Adoption also has a strong positive effect on cashless behavior in Group 2, with a coefficient of 0.312 and a p-value of 0.008, showing that technology adoption significantly encourages cashless behavior, with a more significant influence than Group 1. Lastly, the relationship between financial technology and cashless behavior is the strongest in Group 2, with a coefficient of 0.331 and an exceptionally significant p-value of 0.001, highlighting that financial technology has a very substantial and direct impact on promoting cashless behavior, more so than Group 1.

### 3.4 Model Validity

Model validity is measured using the coefficient of determination using **Equation 7**; model validity is measured for groups and without groups. The results of the coefficient of determination can be seen in **Table 7**.

**Table 7. Validity Model Result**

| Model Validity | Non-Group | Group |
|---|---|---|
| Q-square predictive relevance ($Q^2$) | 80.35% | 96.12% |

**Table 7** presents the results of model validity measurement using the coefficient of determination for non-grouped (Nongroup) and grouped (Group) data. In the non-grouped model (Non-Group), the Q-square predictive relevance ($Q^2$) value is 80.35%, which indicates that the model can explain 80.35% of the data variability, demonstrating good validity in predicting the observed variables. On the other hand, in the grouped model (Group), the $Q^2$ value increases to 96.12%, which means that when the data is grouped, the model can explain 96.12% of the data variability, indicating a significantly higher level of validity in predicting the outcomes compared to the non-grouped model. Overall, these results suggest that data grouping substantially enhances the model's predictive ability and validity.

## 4. CONCLUSIONS

Based on the results of the analysis and discussion that has been carried out, the conclusions obtained are as follows:

a. SEM results show that adoption and financial technology influenced Group 1 and Group 2 cashless behavior. This shows that adoption and financial technology must be considered if you want to increase cashless behavior.

b. The coefficient of determination ($Q^2$) indicates that the model's predictive power improves when clustering is applied, as evidenced by a higher $Q^2$ value in the grouped data compared to the ungrouped data. This finding suggests that clustering enhances model predictability by capturing structural patterns within heterogeneous data rather than directly addressing estimation bias in SEM.

## REFERENCES


[1] Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *MULTIVARIATE DATA ANALYSIS* (8 ed.). Boston: Pearson.

[2] Solimun, Fernandes, A. A. R., dan Nurjannah. (2017). METODE STATISTIKA MULTIVARIAT PEMODELAN PERSAMAAN STRUKTURAL (SEM) PENDEKATAN WarpPLS. Malang: UB Press.

[3] Bollen. (1989). Structural Equations with Latent Variables. New York: A Wiley-Interscience Publication.

[4] Takasawa, I., Tanioka, K., & Yadohisa, H. (2023). CLUSTERED SPARSE STRUCTURAL EQUATION MODELING FOR HETEROGENEOUS DATA. *Journal of Classification*, 40(3), 588-613.

[5] Hu, H., Liu, J., Zhang, X., & Fang, M. (2023). AN EFFECTIVE AND ADAPTABLE K-MEANS ALGORITHM FOR BIG DATA CLUSTER ANALYSIS. PATTERN RECOGNITION, 139, 109404.

[6] Aggarwal, C. C., & Reddy, C. K. (2013). DATA CLUSTERING. ALGORITHMS AND APPLICATIONS. Chapman&Hall/CRC Data mining and Knowledge Discovery series, Londra.

[7] Chen, Y. H., Glymour, M., Riley, A., Balmes, J., Duchowny, K., Harrison, R., ... & Bibbins-Domingo, K. (2021). EXCESS MORTALITY ASSOCIATED WITH THE COVID-19 PANDEMIC AMONG CALIFORNIANS 18–65 YEARS OF AGE, BY OCCUPATIONAL SECTOR AND OCCUPATION: March through November 2020. PloS one, 16(6), e0252454.

[8] Solimun & Fernandes, A. A. R., "FAKTOR DETERMINAN PERKEMBANGAN CASHLESS SOCIETY UNTUK MENDUKUNG PEMULIHAN EKONOMI SAAT PANDEMI DENGAN DIMODERASI OLEH FINANCIAL TECHNOLOGY BERBASIS STATISTIKA DAN SAINS DATA". 2023.

[9] Aggarwal, C. C., & Reddy, C. K. (2013). DATA CLUSTERING. ALGORITHMS AND APPLICATIONS. Chapman&Hall/CRC Data mining and Knowledge Discovery series, Londra.

[10] 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. https://doi.org/http://dx.doi.org/10.12988/ijma.2014.47212

[11] Schumacker, R. E., & Lomax, R. G. (2016). *A BEGINNER'S GUIDE TO STRUCTURAL EQUATION MODELING* (4th ed.). New York: Routledge.

[12] S. Maunah, "ESTIMASI SKEWNESS (KEMIRINGAN) DENGAN MENGGUNAKAN METODE BOOTSTRAP DAN METODE JACKKNIFE," *Unnes Journal of Mathematics*, vol. 6, pp. 143–152, 2017.

[13] Solimun, Fernandes, A. A. R., dan Nurjannah. (2017). *METODE STATISTIKA MULTIVARIAT PEMODELAN PERSAMAAN STRUKTURAL (SEM) PENDEKATAN* WarpPLS. Malang: UB Press.

[14] Fernandes, A. A. R., Solimun. (2021). *ANALISIS REGRESI DALAM PENDEKATAN FLEKSIBEL*. UB Press.

[15] Astutik, S., Solimun, & Darmanto. (2018). *ANALISIS MULTIVARIAT TEORI DAN APLIKASINYA DENGAN SAS.* Malang: UB Press.