Nexus of
Structural Equation Modeling
Volume 01, Issue 01, January 2026, pages 17-21

DOI: https://doi.org/xx.xxxxx/xxxxxxxxx-xx
E-ISSN : xxxx-xxxxx

Open Access - Research Article

Check for updates

Partial Least Square (PLS) Model Estimator with PATHMOX Approach Using Mediation Variables

Ika Puspita [1], Bambang Widjanarko Otok [1], and Sri Sulastri [1*].

[1,] *Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia.*

Article Info

Article history:

Received : February 28th, 2026
Revised : April 30th, 2026
Accepted : Mei 30th, 2026
Published : June 10th, 2026

Published by:

ARCSTICS PUBLISHING

ABSTRACT

One of the multivariate statistical analysis techniques that can explain complex relationship patterns involving many variables is *Structural Equation Modeling* (SEM). *Variance-Based* SEM (VB-SEM), also known as *Partial Least Squares* (PLS), is a more flexible method in SEM because it is able to overcome various limitations, such as the scale of different variable indicators, limited sample counts, and data distribution assumptions. PATHMOX is a special strategy in PLS-SEM that is used to segment structural model results to overcome the observed heterogeneity. This study aims to build a structural equation model (SEM-PLS) and segment the results of structural equation models using the PATHMOX approach. The PATHMOX algorithm is carried out by comparing structural models between segments through the *F-global test*. The path coefficient is estimated with OLS as in multiple linear regression analysis between latent variables Y_i and Y_j.

Keywords: *Mediation, OLS, PATHMOX, PLS, SEM,*

****Corresponding Author:*** *srisulatri2580@gmail.com*

1. Introduction

SEM is a multivariate statistical analysis method used to analyze complex relationship structures involving many variables. This method allows the simultaneous analysis of various relationships, explaining the causal relationships between latent variables, and accounting for measurement errors in model estimation [1]. The SEM model consists of two main components: (1) the measurement model, which connects observed variables with latent variables through confirmatory factor analysis, and (2) structural model, which explains the relationships between latent variables using simultaneous regression equations [2].

The SEM model was originally developed based on the covariance-based SEM (CB-SEM) approach by [3], [4]. This approach starts from a theoretical study that is explicitly formulated by the researcher based on a literature review. Therefore, CB-SEM is used to confirm models that have been developed based on theory with available empirical data [5]. However, the use of CB-SEM requires the fulfillment of several parametric assumptions, such as observation variables with a normal multivariate distribution, independence between observations, reflective indicators, and a large enough sample number to produce a stable estimate [6].

To overcome the limitations in fulfilling assumptions in covariance-based SEM (CB-SEM), a component-based SEM or variance approach was developed, known as Partial Least Squares Structural Equation Modeling (PLS-SEM). This method was first introduced [7]. PLS-SEM is a more flexible and powerful method than CB-SEM, especially because it does not require strict data distribution assumptions, can be used on small samples, and is able to handle various types of data scales, ranging from nominal (category), ordinal, interval, to ratio [6]

PLS-SEM has been applied in a variety of studies with different sample characteristics. In studies involving many variables and indicators from diverse populations such as gender, marital status, age, and education level, the main challenge that often arises is heterogeneity in the data. This heterogeneity can affect the results of the analysis and interpretation of the PLS-SEM model, so it is important to understand the types of heterogeneity that exist. In general, heterogeneity can be categorized into two main types, namely observed heterogeneity and unobserved heterogeneity. Observed heterogeneity occurs when the number and definition of segments are known in advance, while unobserved heterogeneity arises when the source of heterogeneity, both the number and the definition, cannot be clearly identified [8].

In studies where the variables that are the source of heterogeneity are not known beforehand, grouping observations into segments a priori is not possible. Empirically, several approaches such as FIMIX-PLS, REBUS-PLS, and PLS-POS have been used to address unobserved heterogeneity. [9] used FIMIX-PLS to detect heterogeneity in models containing latent variables, focusing on the grouping of elderly data in the city of Surabaya. [10] in her research on Modeling the Nutritional Status of Toddlers in Southeast Sulawesi Province applied PLS-SEM to identify valid indicators and significant latent variables on the nutritional status of toddlers, then used REBUS-PLS to detect heterogeneity by segmenting data. Meanwhile, a study conducted by [11] analyzed the indicators of ITS students' online shopping behavior using SEM-PLS and PLS-POS to identify student segments based on their online shopping behavior. However, when heterogeneity in data comes from observed heterogeneity, the PATHMOX approach can be applied. For example, a study by [12] used PATHMOX-PLS to analyze the clinical picture of HIV/AIDS patients, aiding in segmenting data based on different patient characteristics.

In the PLS-SEM model, the relationships between latent variables can be differentiated into two types: direct relationships and indirect relationships. In direct relationships, exogenous variables directly affect endogenous variables without going through other variables. Meanwhile, in indirect relationships, the influence of exogenous variables on endogenous variables occurs through one or more mediating variables that are intermediates in the relationship. The use of mediation variables aims to see whether the influence of an exogenous variable on endogenous variables becomes stronger, weaker, or even insignificant after going through a mediation variable [13]. The PATHMOX approach in the PLS-SEM model, which allows exogenous variables to affect endogenous variables directly without going through other variables, has been applied by [12] The study showed that the model used met the criteria of validity and reliability, with an R-square value of 80% in the structural model, which indicates a good model suitability. In addition, the measurement model has also been proven to meet the necessary statistical criteria. However, the application of the PATHMOX approach by involving segmentation/indirect relationship variables in the PLS-SEM model has never been carried out. In fact, the use of segmentation variables has the potential to divide the population into segments that have structural differences in the relationships between variables, which can provide a deeper understanding of the heterogeneity that exists in the population.

2. Method Details

2.1. Structure Equation Modeling

Combining two statistical techniques factor analysis, which was established in psychology/psychometrics or sociology, and simultaneous equation models, which were developed in econometrics is known as structural equation modeling, or SEM. The two components of the SEM model are 1) the measurement component, which uses the confirmatory factor model to link observable and latent variables, and 2) the structural component, which uses simultaneous regression equations to link latent variables.

2.2. Partial Least Square (PLS)

Herman Wold created partial least squares, a soft modeling technique, in 1985 to address the drawbacks of parametric assumptions in covariance-based SEM (CB-SEM). Even in cases when there is multicollinearity among exogenous variables, this approach can be applied without assuming a multivariate normal distribution. Additionally, it works well with a variety of data scales and small sample sizes. PLS is also intended for use in scenarios when the measuring model is formative or the theoretical underpinnings are not well established. The following is one way to formulate the model (1):

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\,\boldsymbol{\xi} + \boldsymbol{\zeta} \tag{1}$$

The formula for exogenous latent variables with reflecting indicators is as follows (2):

$$\mathbf{x} = \boldsymbol{\Lambda}_{\mathbf{x}}\boldsymbol{\xi} + \boldsymbol{\delta} \tag{2}$$

And reflective indicators (3):

$$\mathbf{y} = \boldsymbol{\Lambda}_{\mathbf{y}}\boldsymbol{\eta} + \boldsymbol{\varepsilon} \tag{3}$$

2.3. PLS Path Modeling Segmentation Tree (PATHMOX)

With research data that contains numerous segmentation variables and no prior knowledge of the factors, PLS Path Modeling Segmentation Tree (PATHMOX) is a segmentation technique that uses decision trees. Gaston Sanchez presented the PATHMOX technique in 2009 to distinguish between various segments. 'MOX' in PATHMOX is said to be derived from the Nahuatl (Aztec) language, meaning to split into groups.

To handle the disparities in data groups seen in PLS-PM (Partial Least Squares-Path Modeling), this approach was created. The technique creates a segmentation tree by progressively dividing the data according to the binary segmentation principle. Based on the properties of the produced segments, each branch or node in this tree has a unique path model. PATHMOX's primary objective is to detect variations in path models, not to forecast results, in contrast to predictive techniques. This method helps identify and differentiate between different PLS-PM models in the data. PATHMOX will divide the data according to segmentation variables in the most efficient manner, resulting in groups with the most different model structures from one another.

2.4. PATHMOX Algorithm

The PATHMOX (Path Modeling Segmentation Trees) algorithm developed a novel segmentation method for path modeling. The way this algorithm operates is by building a path model tree with a structure akin to a decision tree, in which every node stands for a distinct model for various segments. Using the p-value threshold comparison for two parameters as the criteria and the F-test for path coefficients as the split criterion, the PATHMOX algorithm constructs a binary decision tree for PLS-PM models. Starting from the root node, the method estimates the path model for the entire population. The algorithm's steps are as follows, with the root node serving as the parent node:

1) Applying the structural model at the root node to estimate the path model for the entire population.

2) Finding the optimal split: examining the inner model's coefficient equality. The model is computed for each partition, and all potential binary splits for the latent variable are produced.

3) Using PLS-PM to compute the inner model for every segment and determine the associated path coefficients.

4) Using PLS-PM to compute the inner model for every segment and determine the associated path coefficients.

5) Comparing the inner models using the p-value computation and the test of equivalence for path coefficients.

6) Out of all binary splits across all segmentation variables, choose the best split (with the smallest p-value).

7) A new path model is calculated for every element in the child nodes if the p-value is higher than the chosen threshold. A p-value is deemed statistically significant if it is less than 0.05.

3. Results and Discussion

The PATHMOX algorithm is a decision tree-based hierarchical segmentation method used to identify heterogeneous structures in the PLS-SEM model.

The structural model used in this study consists of three endogenous latent constructs, namely, , $\eta_1\eta_2$and η_3 with two exogenous constructs ξ_1 and . ξ_2The forms of structural equations used are:

$$\eta_1 = \gamma_{11}\xi_1 + \zeta_1$$

$$\eta_2 = \gamma_{21}\xi_1 + \zeta_2$$

$$\eta_3 = \gamma_{31}\xi_1 + \gamma_{32}\xi_2 + \beta_{31}\eta_1 + \beta_{32}\eta_2 + \zeta_3$$

With ζ_i is *an error* that is assumed to be distributed normally $E(\zeta) = 0$ and var.$(\zeta) = \sigma^2 I$

The matrices that can be obtained are as follows:

$$\eta = [\eta_1, \eta_2, \eta_3]^T$$

$$\xi = [\xi_1, \xi_2]^T$$

B is the vector of the path coefficient:

$$\mathrm{B} = [\gamma_{11}, \gamma_{21}, \gamma_{31}, \gamma_{32}, \beta_{31}, \beta_{32}]^T$$

The structural similarities that are formed are:

$\eta = f(\xi, \mathrm{B}) + \zeta$

It is assumed that *the parent node* is divided into two *child nodes* or segments A and B, so that the structural similarity of each segment is:

Segment A: $\eta^A = f(\xi^A B^A) + \zeta^A$

Segment B: $\eta^B = f(\xi^B B^B) + \zeta^B$

with and $\zeta^A \sim N(0, \sigma^2 I)\zeta^B \sim N(0, \sigma^2 I)$

Hypothesis testing for both segments can be said that the initial hypothesis shows that the equations for segment A and segment B are the same, while the counterhypothesis is that the equations are different. This is shown with the following matrix shape:

Zero hypothesis: The structural model in segment A and segment B is the same H_0

$$H_0: \begin{bmatrix} \eta^A \\ \eta^B \end{bmatrix} = \begin{bmatrix} \xi^A \\ \xi^B \end{bmatrix} [\mathrm{B}] + \begin{bmatrix} \zeta^A \\ \zeta^B \end{bmatrix}$$

Alternative hypothesis: The structural models in segment A and segment B are differentH_1

$$H_1: \begin{bmatrix} \eta^A \\ \eta^B \end{bmatrix} = \begin{bmatrix} \xi^A & 0 \\ 0 & \xi^B \end{bmatrix} \begin{bmatrix} B^A \\ B^B \end{bmatrix} + \begin{bmatrix} \zeta^A \\ \zeta^B \end{bmatrix}$$

The following matrix is determined:

$$X_0 = \begin{bmatrix} \xi^A \\ \xi^B \end{bmatrix}_{(n,p)}, X = \begin{bmatrix} \xi^A & 0 \\ 0 & \xi^B \end{bmatrix}_{(n,2p)} \text{ and } A = \begin{bmatrix} I_p \\ I_p \end{bmatrix}_{(2p,p)}$$

where p is the number of exogenous latent variables for and is an identity matrix ordered $\eta I_p p$. Based on the lemma from

Lebart (1979) in (Lamberti et al., 2017) testing the statistical test F with p and free degrees, namely:$(n-2p)$.

$$F_{Block} = \frac{(SS_{H_0} - SS_{H_1})/p}{SS_{H_0}/(n-2p)}$$

After the parent node is divided into two *child nodes* based on the best segmentation variable, the structural parameters (path coefficient) for each segment (child node) are estimated separately, using the PLS-SEM algorithm.

The coefficient of the path is then estimated with OLS as in multiple linear regression analysis between latent variables Y_i and is as follows: Y_j

$$\mathbf{Y_i} = \boldsymbol{\beta}_\mathbf{i}\mathbf{Y_j} + \mathbf{u_i}$$

$$\mathbf{u_i} = \mathbf{Y_i} - \boldsymbol{\beta}_\mathbf{i}\mathbf{Y_j}$$

Then calculate as follows:$\mathbf{u_i^T u_i}$

$$\mathbf{u_i^T u_i} = (\mathbf{Y_i} - \boldsymbol{\beta}_\mathbf{i}\mathbf{Y_j})^\mathbf{T}(\mathbf{Y_i} - \boldsymbol{\beta}_\mathbf{i}\mathbf{Y_j})$$

$$= (\mathbf{Y_i}^\mathbf{T} - \boldsymbol{\beta}_\mathbf{i}{}^\mathbf{T}\mathbf{Y_j}^\mathbf{T})(\mathbf{Y_i} - \boldsymbol{\beta}_\mathbf{i}\mathbf{Y_j})$$

$$= \mathbf{Y_i}^\mathbf{T}\mathbf{Y_i} - \mathbf{Y_i}^\mathbf{T}\boldsymbol{\beta}_\mathbf{i}\mathbf{Y_j} - \boldsymbol{\beta}_\mathbf{i}{}^\mathbf{T}\mathbf{Y_j}^\mathbf{T}\mathbf{Y_i} + \boldsymbol{\beta}_\mathbf{i}{}^\mathbf{T}\mathbf{Y_j}^\mathbf{T}\mathbf{Y_j}\boldsymbol{\beta}_\mathbf{i}$$

$$= \mathbf{Y_i}^\mathbf{T}\mathbf{Y_i} - 2\boldsymbol{\beta}_\mathbf{i}{}^\mathbf{T}\mathbf{Y_j}^\mathbf{T}\mathbf{Y_i} + \boldsymbol{\beta}_\mathbf{i}{}^\mathbf{T}\mathbf{Y_j}^\mathbf{T}\mathbf{Y_j}\boldsymbol{\beta}_\mathbf{i}$$

$$\frac{\partial(\mathbf{u_i^T u_i})}{\partial\boldsymbol{\beta}_\mathbf{i}} = 0$$

$$\frac{\partial}{\partial\mathbf{w_j}}(\mathbf{Y_i}^\mathbf{T}\mathbf{Y_i} - 2\boldsymbol{\beta}_\mathbf{i}{}^\mathbf{T}\mathbf{Y_j}^\mathbf{T}\mathbf{Y_i} + \boldsymbol{\beta}_\mathbf{i}{}^\mathbf{T}\mathbf{Y_j}^\mathbf{T}\mathbf{Y_j}\boldsymbol{\beta}_\mathbf{i}) = 0$$

$$-2\mathbf{Y_j}^\mathbf{T}\mathbf{Y_i} + 2\mathbf{Y_j}^\mathbf{T}\mathbf{Y_j}\boldsymbol{\beta}_\mathbf{i} = 0$$

$$\mathbf{Y_j}^\mathbf{T}\mathbf{Y_j}\widehat{\boldsymbol{\beta}}_\mathbf{i} = \mathbf{Y_j}^\mathbf{T}\mathbf{Y_i}$$

$$(\mathbf{Y_j}^\mathbf{T}\mathbf{Y_j})^{-1}(\mathbf{Y_j}^\mathbf{T}\mathbf{Y_j})\widehat{\boldsymbol{\beta}}_\mathbf{i} = (\mathbf{Y_j}^\mathbf{T}\mathbf{Y_j})^{-1}(\mathbf{Y_j}^\mathbf{T}\mathbf{Y_i})$$

So the form of estimation is written as follows: $\widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\beta}}_\mathbf{i} = (\mathbf{Y_j}^\mathbf{T}\mathbf{Y_j})^{-1}(\mathbf{Y_j}^\mathbf{T}\mathbf{Y_i})$

After the internal model estimation was carried out for each segment of the PATHMOX separation result, a comparison of the quality of the structural model (inner model) was carried out based on a quality *test*. This comparison includes the significance of the path *coefficient*, as well as the pattern of relationships between latent constructs at each node compared to the global model.

4. Conclusion

The *Partial Least Squares Structural Equation Modeling* (PLS-SEM) study with the PATHMOX approach was conducted to identify structural heterogeneity in the data through segmentation based on non-indicator variables.

The best split selection is done to compare between segment A and segment B. This comparison is done using the F test, with the equation of the F test for both segments. Calculate *the inner model* for each segment, including the estimation of each path coefficient using the PLS-SEM method which is estimated through the *Ordinary Least Squares* (OLS) approach. Compare the inner model between segments based on *the test of quality*, which includes the significance of the path and the pattern of relationships between latent constructs. Determine the best split based on the smallest p-value as the primary segment for each subsequent segmentation variable.

Author Contributions Statement

Author 1: Conceptualization, Methodology
Author 2: Formal analysis, or investigation
Author 3: Preparation of the original draft or writing of the review and editing.

Conflict of Interest Statement

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests.

Reference

[1] J. Hair, W. Black, B. Babin, and R. Anderson, "Multivariate Data Analysis: A Global Perspective," in *Multivariate Data Analysis: A Global Perspective*, vol. 7th, 2010.

[2] S. Yamin and H. Kurniawan, "SEM: Belajar Lebih Mudah Teknik Analisis Data Kuesioner dengan LISREL-PLS," 2009, *Buku*.

[3] Karl G. Joreskog, "A General Method for Estimating A Linear Structural Equation System," *ETS Research Bulletin Series*, vol. 2, 1970.

[4] Y. Ezure, M. Chatfield, D. L. Paterson, and L. Hall, "Applications and reporting of causal inference modelling in infectious disease studies: A systematic review," *Infect. Dis. Model.*, vol. 11, no. 1, 2026, doi: 10.1016/j.idm.2025.09.006.

[5] S. H. dan P. Wardoyo, *Structural Equation Modeling: Untuk Penelitian Manajemen Menggunakan AMOS 18.00*. 2013.

[6] D. Novitasari and M. Asbari, "Urgensi Kepemimpinan dan Mentalitas Siap Berubah Terhadap Kinerja Pegawai di Musim Pandemi Covid-19," *Jurnal Riset Ekonomi*

Manajemen (REKOMEN), vol. 4, no. 1, 2020, doi: 10.31002/rn.v4i1.2712.

[7] W. W. Chin, "The partial least squares approach to structural equation modelling. In Marcoulides G. A. (Ed.)," in *Modern Methods for Business Research*, vol. 295, no. 2, 1998.

[8] G. Lamberti, T. B. Aluja, and G. Sanchez, "The Pathmox approach for PLS path modeling segmentation," *Appl. Stoch. Models Bus. Ind.*, vol. 32, no. 4, 2016, doi: 10.1002/asmb.2168.

[9] J. F. Hair, M. Sarstedt, L. M. Matthews, and C. M. Ringle, "Identifying and treating unobserved heterogeneity with FIMIX-PLS: part I – method," 2016. doi: 10.1108/EBR-09-2015-0094.

[10] B. W. Otok, Purhadi, R. Sriningsih, and D. S. Dila, "Segmentation of toddler nutritional status using REBUS and FIMIX partial least square in Southeast Sulawesi," *MethodsX*, vol. 12, 2024, doi: 10.1016/j.mex.2023.102515.

[11] W. B. H. Pratama, "Analisis Indikator Perilaku Belanja Online Mahasiswa Its Dengan Structural Equation Modelling-Partial Least Square (Sem-Pls) Dan Pls Prediction-Oriented Segmentation (Plspos)," *Institut Teknologi Sepuluh Nopember*, vol. 22, no. 3, 2018.

[12] A. H. Hiariey, "Analisis Path Modeling Segmentation Partial Least Square (PATHMOX-PLS) pada Gambaran Klinis Pasien HIV/AIDS," *Institut Teknologi Sepuluh November*, 2018.

[13] I. K. G. ANTARA, I. P. E. N. KENCANA, and K. JAYANEGARA, "VARIABEL LATEN SEBAGAI MODERATOR DAN MEDIATOR DALAM HUBUNGAN KAUSAL," *E-Jurnal Matematika*, vol. 2, no. 4, 2013, doi: 10.24843/mtk.2013.v02.i04.p056.