RANGE : Jurnal Pendidikan Matematika
e-ISSN :2685-2373

Volume 7, Nomor 2, Januari 2026, pp. 435-451

Estimation of Path Coefficients Based on Optimal RMSEA Index in Structural Equation Modeling

Justin Eduardo Simarmata[1]*, Ferdinandus Mone[2], Debora Chrisinta[3], Miko Purnomo[4], Alejandro Jr. V. Matute[5]
[1,2]Department of Mathematics Education, University of Timor, East Nusa Tenggara, Indonesia
[3]Department of Information Technology, University of Timor, East Nusa Tenggara, Indonesia
[4]Department of Mathematics, University of Timor, East Nusa Tenggara, Indonesia
[5]Department of Information Technology, Laguna State Polytechnic University, Philippines
*Correspondence: justinesimarmata@unimor.ac.id

Received: August 23rd, 2025. Accepted: January 29th, 2026. Published: January 31st, 2026

ABSTRACT

Structural Equation Modeling (SEM) is a statistical approach widely used to analyze causal relationships between latent and observed variables. A key issue in SEM lies in selecting an appropriate parameter estimation method, as it strongly affects the accuracy and interpretation of results. Among the most common estimation techniques are Maximum Likelihood (ML) and Weighted Least Squares (WLS). This study aims to compare the performance of ML and WLS in estimating path coefficients within SEM analysis. Using simulated data generated with the simulate Data function from a predefined structural model, three scenarios are examined with sample sizes of 500 and 1000. Data transformation procedures are applied to ensure consistency before model testing. Each SEM model is then estimated using both ML and WLS, and the results are evaluated through Root Mean Square Error of Approximation (RMSEA) values obtained from 100 replications. Findings indicate that WLS generally outperforms ML in terms of model fit and stability. In the first scenario with a sample size of 500, WLS achieves a lower average RMSEA (0.0141) compared to ML (0.0172). With a sample size of 1000 in the second scenario, both methods produce similar RMSEA values (0.009 for WLS and 0.0096 for ML), though WLS demonstrates lower variability. In the third scenario, also with a sample size of 1000, WLS records an average RMSEA of 0.0074 versus 0.0092 for ML. Overall, the results suggest that WLS is more effective and reliable than ML in providing accurate parameter estimates across different data conditions and sample sizes.

Keywords: ML, Parameter Estimation, RMSEA, SEM, WLS.

How to Cite: Simarmata, J.E., Mone, F. and Chrisinta, D., Purnomo, M., Matute, A. Jr. V. (2026). Estimation of Path Coefficient Parameter Based on The Best RMSEA Value in Structural Equation Modeling Weighted Least Square. *Range: Jurnal Pendidikan Matematika*, 7(2), 435-451.

Introduction

Structural Equation Modeling (SEM) Analysis is a complex statistical method used to test and measure causal relationships between observed and unobserved variables in a conceptual model (Hair Jr et al., 2021; Kline, 2023). SEM allows researchers to examine simultaneous relationships between latent variables (variables that cannot be directly measured) (Simarmata et al., 2024; Yu, 2025) and measured variables (indicators) (Aghaei et al., 2023). When analyzing the relationships between latent variables and indicators in Structural Equation Modeling (SEM), parameter estimation is performed. Parameter estimation is the process of determining the values of model parameters that describe the extent to which latent variables can be explained by measured indicator variables (Kong et al., 2022; Pratiwi et al., 2020).

The initial step involves formulating the structural equation of the SEM model by linking latent variables based on their indicators. Once the model is established, a parameter estimation method is chosen, such as Maximum Likelihood (ML) or least squares methods like Ordinary Least Squares (OLS) or Weighted Least Squares (WLS), depending on the nature of the data and the distribution encountered. Parameter estimation involves estimating the values of model parameters that best explain the covariance patterns between indicator variables and latent variables (Bakk & Kuha, 2021; Hayes, 2021; Surianti, 2021). This process involves finding parameter values that minimize the differences between the model predictions and the observed empirical data (Khairi et al., 2021). Initialization of initial parameter values is crucial, followed by using optimization algorithms to iteratively update these values (Chauhan & Yadav, 2024). The results of parameter estimation provide information about the strength and direction of relationships between latent variables and their indicators. Model evaluation is then performed using various model fit indices and statistical tests to ensure the generated model fits the observed data. If the evaluation results indicate a poor fit, model adjustments can be made, such as adding or removing paths or modifying the structural and measurement equations (Ardi, 2020; Wang & Cai, 2024).

The selection of parameter estimation methods in Structural Equation Modeling (SEM) analysis is crucial, as it can significantly affect the interpretation of results. Therefore, this study will compare the parameter estimation results of ML and WLS on path coefficients in SEM analysis. The choice of WLS is informed by the findings of Dany (2017), which demonstrated that WLS provided more accurate estimates than Ordinary Least Squares (OLS). To strengthen this evidence, the present study undertakes a systematic comparison between WLS and the widely applied ML method in SEM parameter estimation. Unlike previous studies, the novelty of this research lies in its integration of simulation-based data generation with comparative evaluation across different model types (saturated, independence, and hypothesized structural models), thereby providing a more rigorous and comprehensive assessment of estimation accuracy and model fit. This approach not only offers empirical validation of estimation techniques but also contributes methodological insights into the conditions under which WLS can serve as a robust alternative to ML in SEM applications.

Recent developments in Structural Equation Modeling (SEM) have highlighted increasing concerns regarding the suitability of parameter estimation methods under different data conditions, particularly when assumptions of multivariate normality and continuous measurement are violated. Although Maximum Likelihood (ML) estimation remains the most widely adopted approach due to its theoretical properties and computational efficiency, numerous recent studies have reported its sensitivity to non-normal data, discretized indicators, and model complexity. As a result, alternative estimation methods such as Weighted Least Squares (WLS) have gained growing attention, especially in simulation-based and applied SEM research.

Existing studies predominantly focus on comparing ML with other estimators in confirmatory factor analysis (CFA) or under limited structural conditions, often without explicitly examining the stability of path coefficient estimation across different model structures and sample sizes. Moreover, empirical evidence that systematically evaluates estimation accuracy using goodness-of-fit indicators—particularly RMSEA—across saturated, independence, and hypothesized structural models remains scarce. This gap indicates that the current state of the art has not fully addressed how estimation methods perform when model complexity and sample size vary simultaneously.

Therefore, there is an urgent need for a comprehensive simulation-based investigation that not only compares ML and WLS but also evaluates their robustness and stability across different structural scenarios. Addressing this need, the present study integrates simulated data generation with repeated estimation procedures to assess the performance of ML and WLS based on RMSEA behavior. By positioning RMSEA as a central evaluation metric and examining multiple model scenarios, this research contributes to the advancement of SEM methodology by providing clearer guidance on the optimal choice of estimation methods under varying data and model conditions.

Methods

To enhance methodological clarity and provide a clearer overview of the research procedure, this study adopts a structured simulation-based workflow. The overall methodological process, from data generation to model evaluation, follows a sequential set of stages illustrated in Figure 1 and described algorithmically in the following subsections.

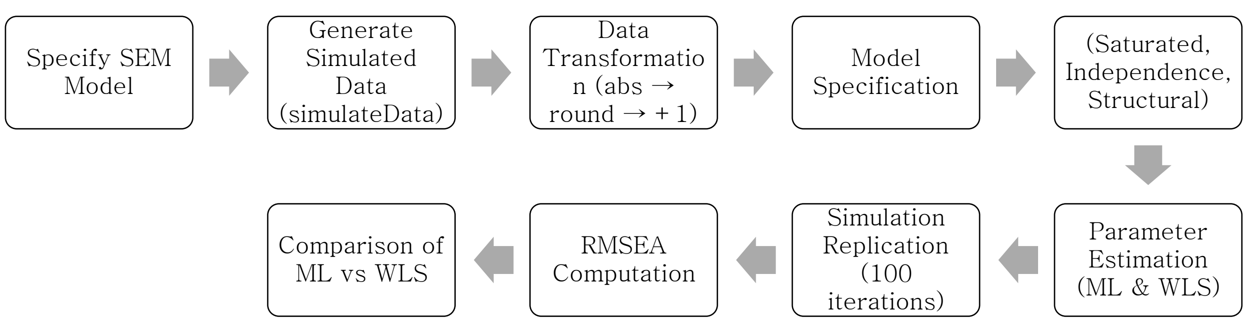

Figure 1. *Stages of the Proposed Simulation-Based SEM Estimation Procedure*

For completeness, the methodological procedure of this study can be summarized algorithmically as follows:

(1) Specify the SEM measurement and structural model.
(2) Generate simulated data using the simulateData() function for a given sample size.
(3) Transform the simulated data into positive integer values (range 1-5).
(4) Specify the saturated model, independence model, and hypothesized structural model.
(5) Estimate model parameters using Maximum Likelihood (ML) and Weighted Least Squares (WLS).
(6) Repeat the estimation process for 100 replications.
(7) Compute RMSEA values for each replication and estimation method.
(8) Compare estimation performance based on RMSEA stability and average values.

In this study, the function simulateData() is used to generate simulated data based on a predefined structural model (Grønneberg et al., 2022). This function allows for the creation of synthetic data that follows the structure and relationships between latent and observed variables according to the model being tested (Fonseca & Bacao, 2023). The model includes several latent variables measured by specific indicators and exhibits clear regression relationships (Van der Loo, 2012). In its implementation, simulateData() takes two main parameters: the predefined structural model and the sample size (sample.nobs), which determines the number of observations in the simulated dataset (Finch H., 2024).

After generating the simulated data using simulateData(), further transformation is performed on the data. The first step in this transformation is to take the absolute value of each element in the simulated dataset using the abs() function. This step ensures that all values in the dataset are non-negative. Next, these absolute values are rounded to the nearest integer using the round() function, which converts continuous values to discrete values more suitable for certain types of analysis. The final step is to add 1 to each element in the rounded dataset to ensure that there are no zero values and all minimum values are 1. This transformation process aims to produce a dataset consisting of positive integer values ranging from 1 to 5. These steps help avoid issues that may arise from negative or zero values in the analysis and make the data more suitable for specific types of analysis that require categorical values. Thus, the final result of this transformation is the dataset simData.n, which is ready for further analysis in this study.

As an example of implementation, a simple SEM model is defined and sample sizes of 500 and 1000 are specified (Rožman et al., 2020). The simulated data is generated using the simulateData() function and then processed into positive integer values. These steps ensure that the data used in the analysis is consistent with the predefined model and suitable for the intended analysis. The results of the transformation are displayed and verified to ensure data integrity before being used for further analysis.

This implementation example demonstrates how the use of the simulateData() function and data transformation can provide a synthetic dataset appropriate for research needs.

In the parameter estimation stage, both Maximum Likelihood (ML) and Weighted Least Squares (WLS) estimation methods are implemented using an iterative numerical optimization procedure. Specifically, parameter estimation is carried out using the Newton–Raphson algorithm, which updates parameter values iteratively based on the first and second derivatives of the objective function. At each iteration, the gradient vector and Hessian matrix are computed to adjust the parameter estimates until convergence is achieved according to a predefined tolerance criterion. This iterative estimation process is applied consistently across all simulation replications and scenarios, as illustrated in the methodological stages presented in Figure 1.

Based on this foundation, the generated data are subsequently applied to three distinct types of models in order to evaluate model performance and overall fit under varying structural assumptions. In particular, the analyses are conducted using the saturated model, the independence (null) model, and the theoretically proposed structural equation model. This comparative framework enables the study to examine not only the adequacy of the hypothesized model but also its relative performance with respect to boundary conditions, thereby providing a more rigorous assessment of parameter estimation and model fit. The generated data will be applied to the following three types of models:

a) Scenario 1

n=500

Measurement model

$L1 =\sim x1 + x2 + x3$

$L2 =\sim y1 + y2 + y3 + y4$

$L3 =\sim y5 + y6 + y7 + y8$

Regressions

$L2 \sim L1$

$L3 \sim L1 + L2$

Scenario 1 is designed to evaluate the performance of Maximum Likelihood (ML) and Weighted Least Squares (WLS) estimation methods under a moderate sample size condition (n = 500). This scenario employs a relatively simple measurement and structural model to examine the stability and accuracy of path coefficient estimation when data availability is limited. The comparison focuses on how each estimation method performs in terms of model fit, as measured by RMSEA, across repeated simulations.

b) Scenario 2

n=1000

measurement model

$L1 = \sim x1 + x2 + x3$

$L2 = \sim x4 + x5 + x6$

$L3 = \sim y1 + y2 + y3 + y4$

$L4 = \sim y5 + y6 + y7 + y8$

$L5 = \sim z1 + z2 + z3$

regressions

$L3 \sim L1$

$L4 \sim L1 + L3$

$L5 \sim L4 + L2$

$L2 \sim L1$

Scenario 2 investigates the performance of ML and WLS estimation methods under a larger sample size condition (n = 1000) with an increased number of latent variables and structural relationships. This scenario is intended to assess whether increasing sample size improves estimation accuracy and model fit stability, as well as to examine potential differences between ML and WLS when applied to more complex SEM structures.

c) Scenario 3

n=1000

measurement model

$L1 = \sim x1 + x2 + x3$

$L2 = \sim x4 + x5 + x6$

$L3 = \sim y1 + y2 + y3 + y4$

$L4 = \sim y5 + y6 + y7 + y8$

$L5 = \sim js1 + js2 + js3$

$L6 = \sim wl1 + wl2 + wl3 + wl4$

regressions

$L3 \sim L1$

$L4 \sim L1 + L3$

$L5 \sim L4 + L2$

$L6 \sim L5 + L4$

$L2 \sim L1$

Scenario 3 further extends the analysis by applying ML and WLS estimation methods to a more complex structural model using a large sample size (n = 1000). This scenario emphasizes the robustness of each estimation method under higher model complexity, allowing for a comprehensive evaluation of their consistency and reliability in estimating path coefficients and achieving optimal model fit.

To evaluate model fit across estimation methods and simulation replications, this study employs the Root Mean Square Error of Approximation (RMSEA) as the primary goodness-of-fit index. RMSEA is a widely used fit measure in Structural Equation Modeling (SEM) that assesses the degree to which a model approximates the population covariance structure while accounting for model complexity. Unlike absolute fit indices, RMSEA penalizes overly complex models and provides an assessment of model misspecification per degree of freedom (Lai & Green, 2016).

Mathematically, RMSEA is computed based on the model chi-square statistic, degrees of freedom, and sample size, and is expressed as:

$$RMSEA = \sqrt{\max(\frac{\chi^2 - df}{df(N-1)}, 0)} \qquad (1)$$

where χ^2denotes the chi-square value of the fitted SEM model, dfrepresents the degrees of freedom, and $N$is the sample size. Lower RMSEA values indicate better model fit, with values below 0.05 generally interpreted as close fit and values between 0.05 and 0.08 indicating acceptable fit.

In the simulation framework of this study, RMSEA is computed at each iteration of the estimation process. For each replication, the SEM model is estimated using Maximum Likelihood (ML) and Weighted Least Squares (WLS), after which the corresponding chi-square statistic and degrees of freedom are obtained. These quantities are then used to calculate the RMSEA value for that iteration. This procedure is repeated for 100 iterations for each scenario and estimation method. The sequence of RMSEA values obtained across iterations forms the basis for evaluating estimation performance and stability. Accordingly, RMSEA values are plotted against the iteration index, where the X-axis represents the simulation iterations and the Y-axis represents the corresponding RMSEA values. This visualization allows for a direct comparison of the average RMSEA levels and variability between ML and WLS across different scenarios.

Results and Discussion

Structural Equation Modeling (SEM) is a statistical analysis technique that combines factor analysis and path analysis to model the relationships between latent variables and observed variables. In SEM, regression coefficients (path coefficients) can be estimated using various methods, including Maximum Likelihood (ML) and Weighted Least Squares (WLS). The Maximum Likelihood (ML) method aims to find parameter values that maximize the likelihood (fit) of the observed data.

$$L(\theta) = f(Y \mid \theta) \quad (2)$$

$$L(\theta) = \frac{1}{(2\pi)^{\frac{np}{2}} |\Sigma(\theta)|^{\frac{n}{2}}} \; exp\left(-\frac{1}{2} \sum_{i=1}^{n} \left(y_i - \mu(\theta)\right)^T \Sigma(\theta)^{-1} \left(yi - \mu(\theta)\right) \right)$$

$$\ell(\theta) = log L(\theta)$$

$$\ell(\theta) = -\frac{n}{2}\left(log|\Sigma(\theta)| + tr(S\Sigma(\theta)^{-1}) - p\, log(2\pi)\right)$$

$$\frac{\partial \ell(\theta)}{\partial \theta} = 0$$

$$\frac{\partial \ell(\theta)}{\partial \theta} = \frac{\partial}{\partial \theta}\left(-\frac{n}{2}\left(log|\Sigma(\theta)| + tr(S\Sigma(\theta)^{-1}) - p\, log(2\pi)\right)\right)$$

The derivative of the log-likelihood function consists of two components, the first is from $log|\Sigma(\theta)|$, which is:

$$\frac{\partial log|\Sigma(\theta)|}{\partial \theta} = tr\left(\frac{\Sigma(\theta)^{-1} \partial \Sigma(\theta)}{\partial \theta}\right)$$

and $tr(S\Sigma(\theta)^{-1})$

$$\frac{\partial tr(S\Sigma(\theta)^{-1})}{\partial \theta} = -tr\left(\frac{S\Sigma(\theta)^{-1} \partial \Sigma(\theta) \Sigma(\theta)^{-1}}{\partial \theta}\right)$$

Combining the two derivatives, the result is:

$$\frac{\partial \ell(\theta)}{\partial \theta} = -\frac{n}{2}\left(tr\left(\frac{\Sigma(\theta)^{-1} \partial \Sigma(\theta)}{\partial \theta}\right) - tr\left(\frac{S\Sigma(\theta)^{-1} \partial \Sigma(\theta) \Sigma(\theta)^{-1}}{\partial \theta}\right)\right)$$

$$-\frac{n}{2}\left(tr\left(\frac{\Sigma(\theta)^{-1} \partial \Sigma(\theta)}{\partial \theta}\right) - tr\left(\frac{S\Sigma(\theta)^{-1} \partial \Sigma(\theta) \Sigma(\theta)^{-1}}{\partial \theta}\right)\right) = 0$$

$$tr\left(\frac{\Sigma(\theta)^{-1} \partial \Sigma(\theta)}{\partial \theta}\right) - tr\left(\frac{S\Sigma(\theta)^{-1} \partial \Sigma(\theta) \Sigma(\theta)^{-1}}{\partial \theta}\right) = 0$$

$$tr\left(\frac{\Sigma(\theta)^{-1}\partial\Sigma(\theta)}{\partial\theta}\right) = tr\left(\frac{S\Sigma(\theta)^{-1}\partial\Sigma(\theta)\Sigma(\theta)^{-1}}{\partial\theta}\right)$$

To solve this equation, we need a system of equations that connects the model parameters θ with the data. In practice, this is done using numerical methods such as the Newton-Raphson method, which is an iterative technique that uses information from the gradient and Hessian to update parameter estimates. To apply the Newton-Raphson method, we also need the second derivative of the log-likelihood function with respect to the parameter θ, known as the Hessian matrix (the matrix of second derivatives of the log-likelihood function) (Willis et al., 2020):

$$H(\theta) = \frac{\partial^2 \ell(\theta)}{\partial\theta\partial\theta^T} \quad (3)$$

For the log-likelihood function in the context of the multivariate normal distribution, the first component of the Hessian originating from $log|\Sigma(\theta)|$ is:

$$\frac{\partial^2 log|\Sigma(\theta)|}{\partial\theta\partial\theta^T} = -tr\left(\frac{\Sigma(\theta)^{-1}\partial\Sigma(\theta)\Sigma(\theta)^{-1}\partial\Sigma(\theta)}{\partial\theta\partial\theta^T}\right) \quad (4)$$

The second component comes from $tr(S\Sigma(\theta)^{-1})$:

$$\frac{\partial^2 tr(S\Sigma(\theta)^{-1})}{\partial\theta\partial\theta^T} = tr\left(\frac{S\Sigma(\theta)^{-1}\partial\Sigma(\theta)\Sigma(\theta)^{-1}\partial\Sigma(\theta)\Sigma(\theta)^{-1}}{\partial\theta\partial\theta^T}\right) \quad (5)$$

Combining these components, we obtain the Hessian of the log-likelihood:

$$H(\theta) = -\frac{n}{2}\left(-tr\left(\frac{\Sigma(\theta)^{-1}\partial\Sigma(\theta)\Sigma(\theta)^{-1}\partial\Sigma(\theta)}{\partial\theta\partial\theta^T}\right) + tr\left(\frac{S\Sigma(\theta)^{-1}\partial\Sigma(\theta)\Sigma(\theta)^{-1}\partial\Sigma(\theta)\Sigma(\theta)^{-1}}{\partial\theta\partial\theta^T}\right)\right) \quad (6)$$

Equations (3)-(6) provide the mathematical foundation required for implementing the Newton-Raphson optimization method in parameter estimation. These equations define the gradient vector and the Hessian matrix of the log-likelihood function, which are essential for iteratively updating the parameter estimates. Based on these formulations, the Newton–Raphson method is applied through the following procedural steps. Steps of the Newton-Raphson method (Pho, 2022) are as follows:

1. Determine an initial value of the parameter vector, denoted as θ^0. This initial estimate may be obtained from prior information, simplified assumptions, or default starting values provided by the estimation software.

2. At the t-th iteration, compute the gradient vector $\nabla\ell(\theta^t)$, which consists of the first derivatives of the log-likelihood function, and compute the Hessian matrix $H(\theta^t)$, which contains the second derivatives.
3. Determine the inverse of the Hessian matrix $H(\theta^t)^{-1}$, which is required to adjust the parameter estimates.
4. Update the parameter vector using the Newton–Raphson update rule:

$$\theta^{t+1} = \theta^t - H(\theta^t)^{-1}\nabla\ell(\theta^t) \quad (7)$$

   where $\nabla\ell(\theta)$ denotes the gradient vector of the log-likelihood function.
5. The iterative process continues until convergence is achieved, that is, when the change in parameter estimates between two successive iterations becomes sufficiently small:

$$|\theta(t+1) - \theta(t)| < \epsilon \quad (8)$$

   where ϵ is a pre-set tolerance.

Once convergence is reached, θ is the parameter estimate that maximizes the log-likelihood function. The results are evaluated by examining the model fit using RMSEA. Various metrics can be used to check the fit, but this study uses RMSEA because it is currently one of the most popular measures of goodness-of-model fit within SEM (Lai & Green, 2016).

Next, the Weighted Least Squares (WLS) method is used to address heteroscedasticity or correlation among measurement errors. In the context of SEM, to minimize the difference between observed parameters and those predicted by the model, considering the appropriate weight matrix. The WLS objective function is expressed as follows:

$$F(\theta) = (s - \sigma(\theta))'W(s - \sigma(\theta)) \quad (9)$$

Parameter estimation is done by minimizing the WLS objective function:

$$\hat{\theta} = arg\min_{\theta} F(\theta) \quad (10)$$

The minimization process involves iterations until we reach the parameter value that minimizes $F(\theta)$. This process involves a vector s consisting of observed sample covariances or correlations, which are the empirical data from the sample, while the model parameter vector $\sigma(\theta)$ includes predicted covariances or correlations dependent on parameter θ. The weight matrix W, chosen as the inverse of the variance-covariance matrix of s, provides efficient estimation in the context of heteroscedasticity:

$$W = Var(s)^{-1} \quad (11)$$

The minimization process involves calculating the gradient and Hessian of $F(\theta)$. The gradient of the objective function $F(\theta)$ is the vector of first derivatives of $F(\theta)$ with respect to θ:

$$\nabla F(\theta) = -2W\big(s - \sigma(\theta)\big)\frac{(\partial\sigma(\theta))}{\partial\theta} \quad (12)$$

The Hessian of the objective function $F(\theta)$ is the matrix of second derivatives of $F(\theta)$ with respect to θ:

$$H(\theta) = 2\left(\frac{\partial\sigma(\theta)}{\partial\theta}\right)' W \left(\frac{\partial\sigma(\theta)}{\partial\theta}\right) \quad (13)$$

Using iterative methods like Newton-Raphson, the parameter update formula is obtained as:

$$\theta^{t+1} = \theta^t - H(\theta^t)^{-1}\nabla F(\theta^t) \quad (14)$$

Equation (14) describes the Newton–Raphson update rule for parameter estimation. Here, θ^t is the parameter vector at the *t*-th iteration, $\nabla F(\theta^t$ denotes the gradient of the objective function with respect to the parameter vector, and $H(\theta^t)^{-1}$ is the inverse of the Hessian matrix of second derivatives of F evaluated at θ^t. This update iteratively adjusts the parameter estimates by moving in the direction determined by the inverse Hessian and the gradient until convergence is achieved, leading to the optimal solution. This study includes three main simulations evaluating the performance of each method under different data conditions. Each simulation tests the strengths and weaknesses of WLS and ML in providing accurate and reliable estimates. Results and discussions are detailed for each simulation, showing RMSEA values for each method over 100 repetitions.

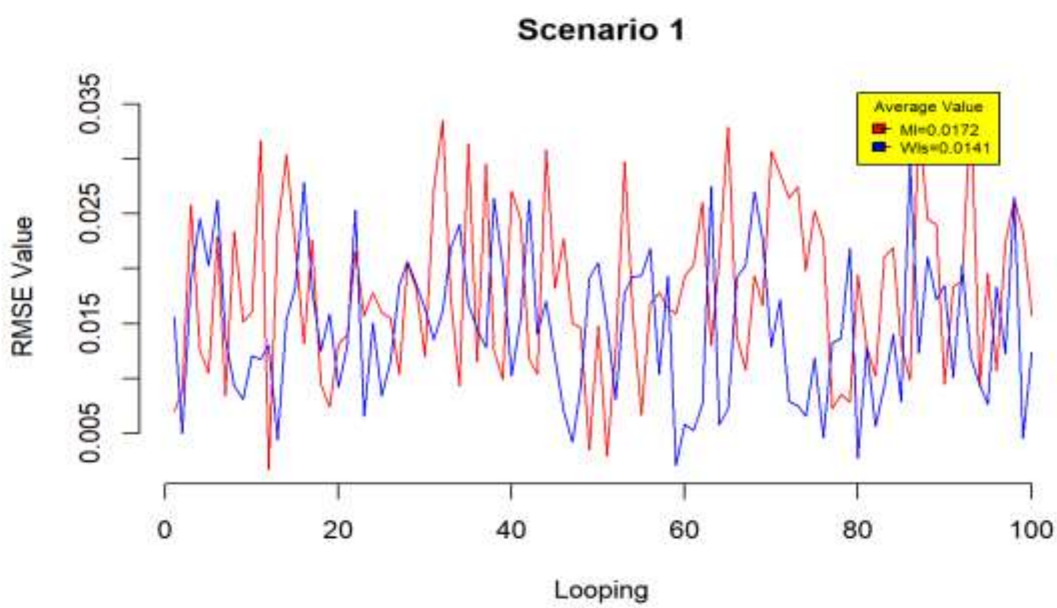

***Figure 2**. Comparison of RMSEA Values for the First Scenario*

The measurement model used in this study involves three main latent variables (Latent Variables). Variable $L1$ is measured using indicators $x1, x2$, and $x3$; $L2$ is measured by indicators $y1, y2, y3$, and $y4$; and $L3$ is measured by indicators $y5, y6, y7$, and $y8$. Regression analysis shows that $L2$ is predicted by $L1$, while $L3$ is predicted by $L1$ and $L2$. Simulation results are shown in a plot depicting RMSEA values over 100 iterations. The X-axis indicates the iterations, while the Y-axis shows the RMSEA values. In this plot, there are two lines representing two different methods or models. The red line (ML) shows RMSEA fluctuations with an average of 0.0172, while the blue line (WLS) shows different patterns with an average RMSEA of 0.0141. The RMSEA fluctuation analysis shows that both methods exhibit RMSEA variation over iterations, but the blue line tends to have lower and more stable RMSEA values compared to the red line. The lower average RMSEA value of the blue method (0.0141) indicates better model fit compared to the red method (0.0172). Therefore, based on these simulation results, the model or method represented by the blue line (WLS) shows better fit in terms of RMSEA compared to the model represented by the red line (ML). The overall conclusion of this study is that considering model fit based on RMSEA, the method represented by the blue line is a more optimal choice for application in this context.

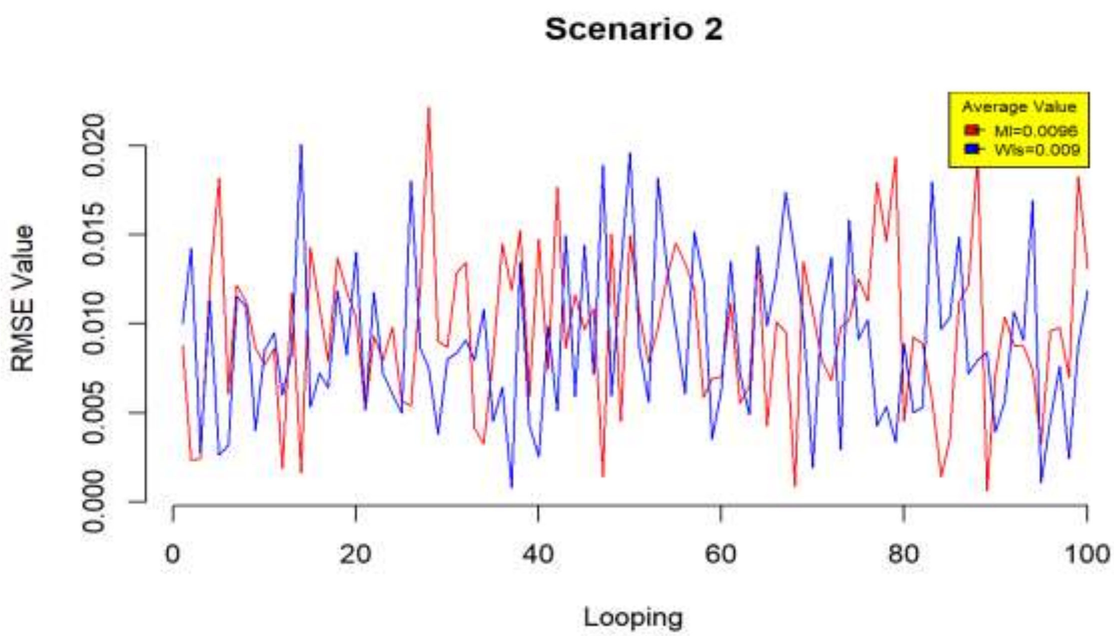

Figure 3. *Comparison of RMSEA Values for the Second Scenario*

In the second scenario, the measurement model used in this study defines the relationship between latent variables ($L1, L2, L3, L4, L5$) and observed variables:

$$(x1, x2, x3, x4, x5, x6, y1, y2, y3, y4, y5, y6, y7, y8, z1, z2, z3)$$

Latent variable $L1$ is associated with $x1, x2$, and $x3$; $L2$ is associated with $x4, x5$, and $x6$; $L3$ is associated with $y1, y2, y3$, and $y4$; $L4$ is associated with $y5, y6, y7$, and $y8$; and $L5$ is associated with $z1, z2$, and

$z3$. In the tested regression model, latent variable $L3$ is predicted by $L1, L4$ is predicted by $L1$ and $L3, L5$ is predicted by $L4$ and $L2$, and $L2$ is predicted by $L1$. These relationships show how these latent variables influence each other in a complex structural model. The results shown in the graph indicate that both scenarios have very similar average RMSEA values, with ML slightly higher (0.0096 compared to 0.009 for WLS). However, the graph also shows significant variation in RMSEA values during the looping process, possibly indicating fluctuations in model performance during iterations.

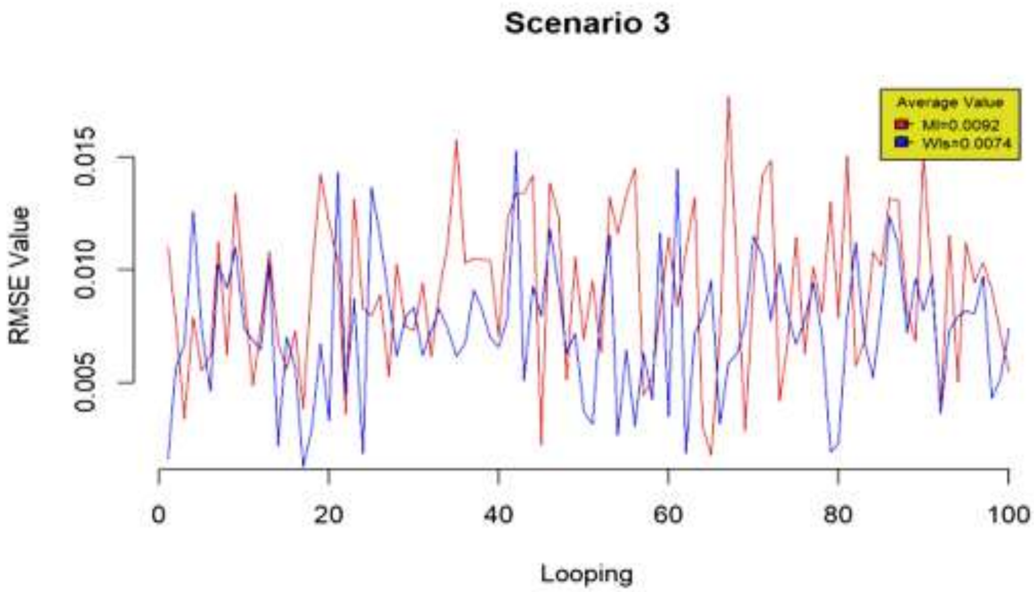

***Figure 4**. Comparison of RMSEA Values for the Third Scenario*

In the third scenario, the measurement model used in this study defines the relationship between latent variables ($L1, L2, L3, L4, L5, L6$) and observed variables:

$$(x1, x2, x3, x4, x5, x6, y1, y2, y3, y4, y5, y6, y7, y8, js1, js2, js3, wl1, wl2, wl3, wl4)$$

Latent variable $L1$ is measured by indicators $x1, x2$, and $x3$; $L2$ by $x4, x5$, and $x6$; $L3$ by $y1, y2, y3$, and $y4$; $L4$ by $y5, y6, y7$, and $y8$; $L5$ by $js1, js2$, and $js3$; and $L6$ by $wl1, wl2, wl3$, and $wl4$. In the tested regression model, latent variable $L3$ is predicted by $L1, L4$ is predicted by $L1$ and $L3, L5$ is predicted by $L4$ and $L2, L6$ is predicted by $L5$ and $L4$, and $L2$ is predicted by $L1$. These relationships illustrate how these latent variables influence each other in a complex structural model. The results shown in the graph indicate that both scenarios have relatively low average RMSEA values, with WLS having a slightly lower average (0.0074) compared to ML (0.0092). The graph also shows significant variation in RMSEA values during the looping process, possibly indicating fluctuations in model performance during iterations.

The measurement and regression models used to test relationships among latent variables show that the Weighted Least Squares (WLS) method tends to have lower and more stable RMSE values compared to the Maximum Likelihood (ML) method. In scenario 1, WLS has an average RMSEA of 0.0141, lower than ML's average of 0.0172. Scenario 2 shows very similar average RMSE values between the two methods, with ML slightly higher (0.0096) compared to WLS (0.009). Scenario 3 shows that WLS again performs better with an average RMSE of 0.0074 compared to ML's 0.0092. Overall, the simulation results from the three scenarios confirm that the WLS method provides better model fit and is more optimal compared to ML in the context of the structural models tested.

Based on the three scenarios considering sample size, the Weighted Least Squares (WLS) method tends to provide better model fit compared to the Maximum Likelihood (ML) method, especially as the sample size increases. Consider the result study (Beauducel & Herzberg, 2009) and (Schweizer et al., 2023) WLS better fit results under MLE (maximum likelihood estimation). In all scenarios, WLS consistently shows lower and more stable average RMSE values compared to ML, indicating that WLS is a more optimal estimation method for application in the context of the tested structural models, both for smaller sample sizes ($n = 500$) and larger ones ($n = 1000$).

Conclusion

This study compared the Weighted Least Squares (WLS) and Maximum Likelihood (ML) methods in path analysis using Structural Equation Modeling (SEM). The goal was to evaluate the effectiveness of both methods under different data conditions through threemain simulations. The first simulation results showed that the WLS method had an average RMSEA of 0.0141, lower than ML's average of 0.0172, indicating that WLS provided better model fit compared to ML with a sample size of 500. In the second simulation, although both methods showed very similar average RMSEA values (0.009 for WLS and 0.0096 for ML), WLS was still superior with lower fluctuation and slightly lower average RMSEA compared to ML. This simulation was conducted with a larger sample size of 1000, showing that WLS still maintained better performance compared to ML. The third simulation again showed that WLS had a lower average RMSEA (0.0074) compared to ML (0.0092). This simulation, also with a sample size of 1000, demonstrated the stability and superiority of WLS in providing better model fit. Overall, the results from the three scenarios confirm that the WLS method is superior in providing accurate and reliable parameter estimates in SEM analysis. The WLS method consistently shows lower and more stable average RMSEA values compared to ML, both for smaller and larger sample sizes. Therefore, WLS is a more optimal estimation method for application in the context of the structural models tested in this study. Future studies may extend this work by comparing WLS and ML with other SEM estimation methods, such as robust or Bayesian approaches, under different data distribution conditions and higher model complexity. Additionally, incorporating alternative goodness-of-fit indices alongside RMSEA and applying the proposed framework to empirical datasets may further enhance the generalizability of the findings.

Acknowledgement

The research team extends its highest appreciation to all parties who have provided support and contributed to the completion of this research. The team is particularly grateful to the Institute for Research and Community Service (LPPM) of Universitas Timor for funding this research through Research Implementation Contract Number: 129/UN60.6/PP/2024. Furthermore, sincere appreciation is also conveyed to the partner institutions, SMA Negeri Taekas and SMK Katolik Kefamenanu, for their cooperation and readiness in providing the necessary information and data throughout the research process.

References

Aghaei, S., Shahbazi, Y., Pirbabaei, M., & Beyti, H. (2023). A hybrid SEM-neural network method for modeling the academic satisfaction factors of architecture students. *Computers and Education: Artificial Intelligence*, *4*(100122). https://doi.org/10.1016/j.caeai.2023.100122

Ardi, N. (2020). Structural Equation Modelling-Partial Least Square to Determine the Correlation of Factors Affecting Poverty in Indonesian Provinces. *IOP Conference Series: Materials Science and Engineering (Vol. 846, No. 1, p. 012054)* . https://doi.org/10.1088/1757-899X/846/1/012054.

Bakk, Z., & Kuha, J. (2021). Relating latent class membership to external variables: An overview. *British Journal of Mathematical and Statistical Psychology*, *74*(2), 340–362. https://doi.org/10.1111/bmsp.12227.

Beauducel, A., & Herzberg, P. Y. (2009). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling*, *13*(2), 186–203. https://doi.org/10.1207/s15328007sem1302_2

Chauhan, D., & Yadav, A. (2024). JawaAn archive-based self-adaptive artificial electric field algorithm with orthogonal initialization for real-parameter optimization problems. *Applied Soft Computing*, *150*(111109). https://doi.org/10.1016/j.asoc.2023.111109

Dany, R. (2017). *Perbandingan Pendugaan Parameter Berbasis OLS (Ordinary Least Square) dan WLS (Weighted Least Square) Pada Analisis Path* [Doctoral dissertation, Universitas Brawijaya]. https://repository.ub.ac.id/id/eprint/8711/

Finch H. (2024). A Comparison of Methods for Synthesizing Results from Previous Research to Obtain Priors for Bayesian Structural Equation Modeling. *Psych.*, *6*(1), 45–88. https://doi.org/10.3390/psych6010004

Fonseca, J., & Bacao, F. (2023). Tabular and latent space synthetic data generation: a literature review. *J Big Data*, *10*(115). https://doi.org/10.1186/s40537-023-00792-7

Grønneberg, S., Foldnes, N., & Marcoulides, K. M. (2022). covsim: An R Package for Simulating Non-Normal Data for Structural Equation Models Using Copulas. *Journal of Statistical Software*, *102*(3), 1–45. https://doi.org/10.18637/jss.v102.i03

Hair Jr, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., Danks, N. P., & Ray, S. (2021). *An introduction to structural equation modeling. Partial least squares structural equation modeling (PLS-SEM) using R: a workbook*. Springer International Publishing.

Hayes, T. (2021). R-squared change in structural equation models with latent variables and missing data. *Behavior Research Methods*, *53*(5), 2127–2157. https://doi.org/10.3758/s13428-020-01532-y.

Khairi, M. I., Susanti, D., & Sukono, S. (2021). Study on Structural Equation Modeling for Analyzing Data. *International Journal of Ethno-Sciences and Education Research*, *1*(3), 52–60. https://doi.org/10.46336/ijeer.v1i3.295.

Kline, R. B. (2023). Principles and practice of structural equation modeling. In *BMC Public Health* (Issue 1093). Guilford publications.

Kong, X., Jiang, X., Zhang, B., Yuan, J., & Ge, Z. (2022). Latent variable models in the era of industrial big data: Extension and beyond. *Annual Reviews in Control*, *54*, 167–199. https://doi.org/10.1016/j.arcontrol.2022.09.005

Lai, K., & Green, S. B. (2016). The problem with having two watches: Assessment of fit when RMSEA and CFI disagree. *Multivariate Behavioral Research*, *51*(2–3), 220–239. https://doi.org/10.1080/00273171.2015.1134306

Pho, K. H. (2022). Improvements of the Newton–Raphson method. *Journal of Computational and Applied Mathematics*, *408*(114106). https://doi.org/10.1016/j.cam.2022.114106

Pratiwi, I., Yanuar, F., & Yozza, H. (2020). Pendugaan Parameter MIU Dari Distribusi Log-Normal Dengan Menggunakan Metode Maximum Likelihood Estimation (MLE) dan Metode Bayes. *Jurnal Matematika UNAND* , *9*(2), 84–92. https://doi.org/10.25077/jmu.9.2.84-92.2020

Rožman, M., Tominc, P., & Milfelner, B. (2020). A Comparative Study Using Two SEM Techniques on Different Samples Sizes for Determining Factors of Older Employee's Motivation and Satisfaction. *Sustainability*, *12*(6), 2189. https://doi.org/10.3390/su12062189

Schweizer, K., DiStefano, C., & French, B. (2023). A maximum likelihood approach for asymmetric non-normal data using a transformational measurement model. *Frontiers in Applied Mathematics and Statistics*, *9*(1095769.). https://doi.org/10.3389/fams.2023.1095769

Simarmata, J. E., Mone, F., & Chrisinta, D. (2024). Structural Equation Modeling: The Influence of School Environment on Students' Interest in Selecting State University. *JTAM (Jurnal Teori Dan Aplikasi Matematika)*, *8*(2), 378–389. https://doi.org/10.31764/jtam.v8i2.19921

Surianti, H. (2021). Estimasi Parameter Model Regresi Linier dengan Pendekatan Metode Bayesian. *Journal of Mathematics: Theory and Applications*, *3*(2), 57–61. https://doi.org/10.31605/jomta.v3i2.1364.

Van der Loo, M. P. (2012). Learning RStudio for R Statistical Computing. In *Ecotrophic* (Issue 2). Packt Publishing Ltd.

Wang, Q., & Cai, X. (2024). Active-learning class activities and shiny applications for teaching support vector classifiers. *Journal of Statistics and Data Science Education*, *32*(2), 202–216. https://doi.org/10.1080/26939169.2023.2231065

Yu, X. (2025). A method for evaluating the learning effectiveness of MOOC English online education based on fuzzy clustering decision tree. *International Journal of Continuing Engineering Education and Life Long Learning*, *35*(1–2), 46–61. https://doi.org/10.1504/IJCEELL.2025.143798