

## The performance of unweighted least squares and regularized unweighted least squares in estimating factor loadings in structural equation modeling

Nurul Raudhah Zulkifli<sup>a\*</sup>, Nazim Aimran<sup>a</sup> and Sayang Mohd Deni<sup>a</sup>

<sup>a</sup>*School of Mathematical Sciences, College of Computing, Informatics and Media, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia*

### CHRONICLE

#### Article history:

Received: February 20, 2023

Received in revised format: April 12, 2023

Accepted: June 8, 2023

Available online: June 8, 2023

#### Keywords:

Monte Carlo simulation

Regularization

Unweighted least square

Regularized unweighted least square

### ABSTRACT

In a confirmatory study, researchers are expected to employ the covariance-based structural equation modeling (CB-SEM). One of the key presumptions when utilizing CB-SEM is that the data is multivariate normal. Nevertheless, a perfect normal distribution is rarely observed in real-life data. To resolve this, the unweighted least square (ULS) is designed to specifically deal with non-normal data in SEM. However, ULS often yields improper solutions like negative, or boundary estimates of unique variances since it considers measurement errors in observed variables. The disturbance in SEM is reflected in unique variance, which is random error due to unreliability or measurement error and reliable variation in the item that indicates unknown latent causes. Consequently, this can generate bias in indicator loadings estimates. As an action to disentangle this issue, the present study proposes the implementation of regularization parameters by adding small positive values to the variance-covariance matrix. The ratio of bias to variance in a model can be improved to obtain the best estimation performance. Pro-Active Monte Carlo simulation was used to produce multivariate non-normal data with designated sample sizes and population characteristics. The data were analyzed using R Programming Environment by employing “psych”, “MASS”, “foreign”, “mvrnorm”, “purr”, and “semTools” packages with 1000 replications to produce multivariate non-normal data. Next, the “lavaan” package was used for SEM and regularized SEM analyses. The outcome of this study proves the capability of regularized ULS to improve parameter estimation.

© 2023 by the authors; licensee Growing Science, Canada.

## 1. Introduction

Structural equation modeling (SEM) is a second-generation statistical analysis approach. It was developed to assess the inter-relationships among several variables in a model (Awang, 2023; Afthanorhan et al., 2021; Ainur et al., 2017; Aimran et al., 2017; Zulkifli et al., 2022). In SEM, the unweighted least squares (ULS) estimation method is designed to work with non-normal data (Mindrila, 2010). According to Jung and Takane (2008), ULS often yields inaccurate solutions like a negative, or boundary estimates of unique variances since it considers measurement errors in observed variables. Unique variance manifests as the disturbance in SEM that is, random error due to unreliability or measurement error and reliable variation in the item that signifies unknown latent causes. In addition, ULS leads to more bias and less precision results in estimating parameters (Forero et al., 2009). When the data is non-normal, the estimates of the loadings may be biased or less efficient, which can lead to inaccurate conclusions about the relationships between the latent variables and the observed variables. Therefore, a vast literature has been promoting the method of regularization to overcome this matter (Jacobucci et al., 2016; Yuan & Bentler, 2017; Yuan & Chan, 2016).

\* Corresponding author.

E-mail address: [nurulraudhahzulkifli@gmail.com](mailto:nurulraudhahzulkifli@gmail.com) (N. R. Zulkifli)

ISSN 2561-8156 (Online) - ISSN 2561-8148 (Print)

© 2023 by the authors; licensee Growing Science, Canada.

doi: 10.5267/j.ijds.2023.6.004

Arruda (2017) stated that regularization can be explained simply as the state of having been made regular. Also, the inclusion of information to solve poorly identified problems is typically described using the mathematical notion of regularization. These definitions are similar to the informal definition given by Bickel and Li (2006), which highlights the modification of a method for providing effective solutions in difficult conditions. The majority of the applications focus on multicollinearity, overfitting, or sparsity problems. Numerous regularization techniques address the parsimony principles such as smoothing, model selection, or methods to control model complexity. In a similar vein, regularization can be used to speed up lengthy computations or invert matrices. Recent research has demonstrated that regularized techniques for regression are beginning to be employed in covariance modelling. Some regularization methods specific to SEM methodology have recently been applied and studied (Jacobucci et al., 2016; Yuan & Bentler, 2017; Yuan & Chan, 2016). Methods like lasso and ridge regression, for instance, have been included into SEM (Jacobucci et al., 2016; Jung, 2013). Jacobucci et al. (2016) proposed regularized structural equation modelling, or regSEM, as a method of penalizing parameters to reduce model complexity and improve generalizability of models. Even though regularized ULS in RegSEM has been introduced for improving the effect of unique variance, however, the method is seen less efficient as the method regularized directly on the specific parameter matrix, that leads to over shrinkage of the estimation problem. To address the issue, this study will demonstrate the implementation of the regularization parameter to each element in the sample covariance matrix in the ULS estimator. This approach seeks the optimal tradeoff between bias and assessing variability by adding small positive values to the elements of the covariance matrix. By doing so, it can be proven that the ratio of bias to variance in a model can be optimized to yield the best estimation performance.

## 2. Methodology

### 2.1 Simulation research model

The existing method developed by Vale & Maurelli (1983) was used in this study to produce non-normally distributed data using Monte Carlo Markov Chain (MCMC) simulation techniques, with given skewness and kurtosis values of 2 and 7, respectively (Pavlov et al., 2020). Three population models with various true indicator loading values were generated. Each model consisted of four latent constructs, each with four items and correlates at 0.7. On the other hand, the homogeneous true indicator loading of 0.7 (for Model 1), 0.8 (for Model 2), and 0.9 (for Model 3), was predetermined respectively. The sample sizes chosen were 50, 100, 200, and 500. In path modelling, sample sizes of 100 to 200 are often used as the starting point (Awang, 2023; Henseler & Chin, 2010) with 50 samples indicating a small sample, and 500 demonstrating a large sample. Next, to ensure the consistency of the findings, 1000 replications of each sample size were performed resulting in the generation of  $3 \times 4 \times 1000 = 12,000$  datasets. ULS and regularized ULS were employed to estimate the indicator loading. For regularization, Arruda & Bentler (2017), Jacobucci et al. (2016), and Jung (2018) recommended that the optimal value of the regularization parameter,  $\lambda$  was chosen based on model performance through the smallest value of RMSEA. The simulation process and SEM analysis were carried out using the R statistical programming environment (R Core Team, 2018). The “psych” package, “MASS” package, “foreign” package, “mvnnonnorm” package, “purr” package and the “semTools” package were applied to produce multivariate non-normal data. Next, the “lavaan” package developed by Rosseel (2012) was used for SEM and regularized SEM analyses correspondingly. The three population models are as illustrated in Fig. 1, Fig. 2, and Fig. 3.

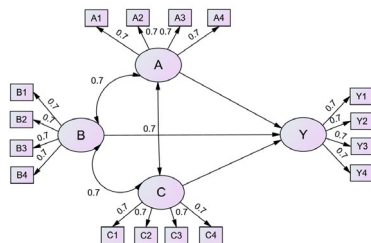


Fig. 1. Model 1

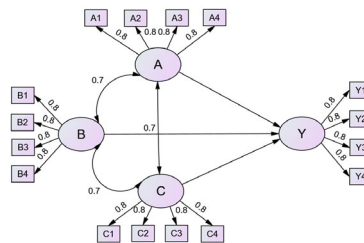


Fig. 2. Model 2

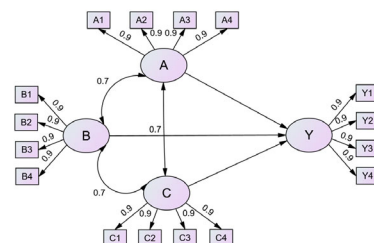


Fig. 3. Model 3

### 2.2 Estimation Methods

#### a. Unweighted Least Squares (ULS)

The Unweighted Least Squares (ULS) estimation method was used to analyze non-normal data to assess fit and coefficients in CB-SEM. McDonald and Bollen (1990) claimed that ULS minimizes the fit function by using derivatives:

$$F_{ULS} = \frac{1}{2} tr \{ [S - \Sigma(\theta)] \}^2 \quad (1)$$

where  $tr$  is the trace of the matrix,  $S$  is the sample covariance matrix,  $\Sigma(\theta)$  is the model-implied covariance matrix and  $\theta$  is the  $(t \times 1)$  vector of parameters. The sum of squares of each element in the residual matrix  $(S - \Sigma(\theta))$  is minimized using the

fit function  $F_{ULS}$ . Compared to ML and WLS, ULS has the advantage of producing a consistent estimator, but unlike ML, no distributional assumptions are required (Schermelleh-Engel, Moosbrugger & Muller, 2003).

### b. Regularized Unweighted Least Squares

In order to produce an accurate estimate, particularly for non-ideal conditions and non-normal data, regularization method is applied to the estimation of the variance-covariance matrix in the ULS estimator. The ULS estimator is a popular method for estimating SEM parameters for non-normal data. However, it can suffer from issues such as biased estimates, unstable solutions, and improper solutions (Forero et al., 2009). Improper solutions like negative or boundary estimates of unique variances can occur when there are errors in the measurement of the observed variables. It represents the minimum amount of variance that cannot be explained by the model. Regularization techniques can improve the stability of the covariance matrix estimates, which can also help to reduce the risk of negative error variances by ensuring enough variance for the latent variables. The values in the sample variance-covariance matrix play a crucial role in the estimation of model parameters in SEM (Kline, 2016). The quality and reliability of parameter estimations, such as factor loading, can be significantly impacted by the accuracy and stability of the sample variance-covariance matrix. Stabilizing the variance-covariance matrix requires applying some forms of regularization to the sample covariance matrix. Regularization achieves this by adding a regularization parameter to the sample variance-covariance matrix, which can control the amount of regularization applied. Consequently, in this study, improved estimators ( $\hat{S}$ ) is used instead of  $S$  in ULS estimators through the addition of the regularization parameter,  $\lambda$ . This technique seeks the ideal trade-off between bias and assessing variability (Jacobucci et al., 2019) by balancing the fit of the model to the data and the complexity of the model including controlling the amount of shrinkage applied to the sample covariance matrix. This helps to improve the stability of the covariance matrix, especially when the sample size is small, or the number of variables is large. The improved sample variance-covariance matrix ( $\hat{S}$ ), which incorporates the addition of regularization parameter,  $\lambda$  to each element of the sample covariance matrix is demonstrated below:

$$\hat{S} = S + \lambda \quad (2)$$

where  $S$  is the sample covariance matrix and  $\lambda$  is the regularization parameter. The value of lambda plays a crucial role in determining how much weight to be assigned to the variance-covariance matrix (Arruda & Bentler, 2017). Several lambda values ( $\lambda > 0$ ) are tested and an optimal  $\lambda$  is chosen depending on the performance of the model.  $\lambda$  with the smallest RMSEA is chosen for each model. RMSEA is a fitness index that measures the discrepancy between the observed and predicted covariance matrices, adjusted for the complexity of the model. The goal of using the RMSEA criterion to select the regularization parameter is to find a balance between model fit and complexity that results in the most accurate and stable estimates of the model parameters. This study varied  $\lambda$  across multiple values, ranging from 0 to 1 in equal increments. Hence, with a regularization parameter that has been considered in the sample covariance matrix of ULS, sufficient covariance between the measured variables is provided to ensure enough variance for the latent variables. Accordingly, the improved regularized ULS technique is as follows:

$$F_{regULS} = \frac{1}{2} tr \{[\hat{S} - \Sigma(\theta)]\}^2 \quad (3)$$

where  $tr$  is the trace of the matrix,  $\hat{S}$  is the regularized sample covariance matrix,  $\Sigma(\theta)$  is the model-implied covariance matrix and  $\theta$  is the ( $t \times 1$ ) vector of parameters.

### 2.3 Comparative Bias Index (CBI)

The population data was generated using different sample sizes and specified criteria, as had been previously mentioned. The actual values for model parameters like true indicator loading were found to be the population value. This value is required to generate simulated data. To examine the bias of parameter estimates produced from simulation data, the CBI developed by Aimran et al. (2017) was calculated for comparison.

$$CBI = 1 - \frac{|\hat{\theta} - \theta|}{\theta} \quad (4)$$

where  $\hat{\theta}$  represents an estimate of the model parameter while  $\theta$  denotes the parameter's actual value. A CBI value of  $\geq 0.8$  denotes an estimate with acceptable bias, whereas a CBI value of  $> 0.9$  denotes an estimate that is unbiased or has low bias. Otherwise, the bias estimate is unacceptable.

### 2.4 Root Mean Square Error (RMSEA)

As shown below, the RMSEA computes the difference due to approximation per degree of freedom:

$$RMSEA = \sqrt{\frac{F_0}{df}} \quad (5)$$

where  $F_0$  denotes the discrepancy between the method used to generate the data and the model that was fitted. The acceptable

cut-off values for RMSEA are  $\leq 0.08$  (Awang, 2023).

### 3. Result

The results of the CBI and RMSEA values for all indicator loadings and models are reviewed in Table 1, Table 2 and Table 3, respectively.

**Table 1**  
The CBI, RMSEA and optimal regularization values for Model 1

Sample size	Items	s		RMSEA		Optimal regularization parameter ( $\lambda$ ) for regularized ULS
		ULS	Regularized ULS	ULS	Regularized ULS	
50	A1	0.879	0.937	0.157	0.044	0.75
	A2	0.997	0.874			
	A3	0.919	0.800			
	A4	0.940	0.891			
	B1	<b>0.746</b>	0.939			
	B2	0.984	0.877			
	B3	<b>0.790</b>	0.891			
	B4	0.879	0.901			
	C1	0.833	0.863			
	C2	0.951	0.800			
	C3	0.800	0.860			
	C4	0.944	0.857			
	Y1	0.947	0.880			
	Y2	0.850	0.961			
Y3	0.994	0.974				
Y4	0.933	0.904				
100	A1	0.823	0.887	0.116	0.026	0.10
	A2	0.963	0.844			
	A3	0.894	0.947			
	A4	0.899	0.804			
	B1	0.803	0.894			
	B2	0.959	<b>0.789</b>			
	B3	0.947	0.833			
	B4	<b>0.701</b>	<b>0.777</b>			
	C1	0.941	0.989			
	C2	0.951	0.927			
	C3	0.906	0.854			
	C4	0.986	0.830			
	Y1	0.934	0.961			
	Y2	0.836	0.849			
Y3	0.851	0.859				
Y4	0.863	0.931				
200	A1	0.900	0.921	0.134	0.040	0.20
	A2	0.881	0.909			
	A3	0.993	0.993			
	A4	0.870	0.874			
	B1	<b>0.781</b>	0.927			
	B2	0.904	0.867			
	B3	<b>0.789</b>	0.894			
	B4	<b>0.741</b>	0.851			
	C1	0.989	0.964			
	C2	0.817	0.991			
	C3	0.811	0.920			
	C4	0.907	0.904			
	Y1	0.991	0.997			
	Y2	0.873	0.890			
Y3	0.820	0.893				
Y4	0.953	0.967				
500	A1	0.846	0.980	0.117	0.035	0.40
	A2	<b>0.766</b>	0.993			
	A3	0.874	0.903			
	A4	<b>0.717</b>	0.981			
	B1	0.856	0.983			
	B2	<b>0.690</b>	0.986			
	B3	<b>0.757</b>	0.991			
	B4	<b>0.686</b>	0.966			
	C1	0.983	0.923			
	C2	0.849	0.913			
	C3	<b>0.773</b>	0.979			
	C4	<b>0.740</b>	0.984			
	Y1	0.920	0.949			
	Y2	0.801	0.956			
Y3	0.820	0.950				
Y4	0.889	0.976				

Note: Bold values indicate unacceptable bias estimates

For Model 1, the loading of each item under the corresponding constructs was set to 0.7. The findings for Model 1 are depicted in Table 1. The finding revealed that ULS outperforms regularized ULS with no bias estimates ( $\geq 0.8$ ) for a sample size of 50. Next, for 100 samples, a number of unfavorable bias estimates ( $< 0.8$ ) was observed for both estimation methods. The biased

estimates generated by regularized ULS are due to overestimation of indicator loadings. Thus, this finding suggests that the proposed method generates comparable results to ULS for a sample size of 100. However, the regularized technique produces a compelling result without any indicators exhibiting unacceptable bias estimates when involving a large sample ( $n \geq 200$ ). Conversely, there are more unacceptable bias estimates for ULS at large sample size ( $n=500$ ). Moreover, the RMSEA values improved and achieved an acceptable threshold of 0.08 when employing regularized ULS. For the selection of the optimal regularization parameter, a small sample model ( $n=50$ ) requires an extremely large regularization parameter value,  $\lambda=0.75$  to obtain a better fit model with minimal RMSEA. Nevertheless, the optimal regularization parameters for sample sizes of 100, 200 and 500 are 0.10, 0.20 and 0.40, respectively, denoting a non-apparent difference.

**Table 2**  
The CBI, RMSEA and optimal regularization values for Model 2

Sample size	Items	CBI		RMSEA		Regularization parameter ( $\lambda$ ) for regularized ULS
		ULS	Regularized ULS	ULS	Regularized ULS	
50	A1	<b>0.768</b>	0.839	0.123	0.077	0.80
	A2	<b>0.740</b>	0.970			
	A3	<b>0.613</b>	0.934			
	A4	<b>0.518</b>	0.916			
	B1	0.981	0.835			
	B2	<b>0.740</b>	0.909			
	B3	0.873	0.926			
	B4	<b>0.765</b>	0.959			
	C1	0.964	0.855			
	C2	<b>0.713</b>	0.946			
	C3	<b>0.588</b>	0.946			
	C4	0.848	0.920			
	Y1	<b>0.698</b>	0.966			
	Y2	0.983	0.983			
	Y3	0.898	0.994			
	Y4	0.960	0.978			
100	A1	0.893	0.986	0.141	0.031	0.10
	A2	<b>0.650</b>	<b>0.794</b>			
	A3	<b>0.611</b>	0.805			
	A4	<b>0.575</b>	0.853			
	B1	0.813	0.996			
	B2	<b>0.696</b>	0.809			
	B3	<b>0.700</b>	0.843			
	B4	<b>0.643</b>	<b>0.776</b>			
	C1	0.823	0.958			
	C2	<b>0.743</b>	<b>0.791</b>			
	C3	<b>0.565</b>	<b>0.790</b>			
	C4	0.835	0.801			
	Y1	<b>0.750</b>	0.915			
	Y2	0.913	0.896			
	Y3	0.953	0.893			
	Y4	<b>0.788</b>	0.880			
200	A1	0.959	0.984	0.137	0.036	0.20
	A2	<b>0.779</b>	0.845			
	A3	<b>0.669</b>	0.865			
	A4	<b>0.724</b>	0.904			
	B1	0.948	0.975			
	B2	<b>0.718</b>	0.875			
	B3	<b>0.750</b>	0.894			
	B4	<b>0.650</b>	0.835			
	C1	0.866	0.994			
	C2	<b>0.648</b>	0.850			
	C3	<b>0.718</b>	0.850			
	C4	<b>0.741</b>	0.866			
	Y1	<b>0.781</b>	0.939			
	Y2	0.923	0.920			
	Y3	0.909	0.914			
	Y4	0.901	0.911			
500	A1	0.951	0.915	0.121	0.022	0.45
	A2	<b>0.731</b>	0.941			
	A3	<b>0.731</b>	0.973			
	A4	<b>0.788</b>	0.999			
	B1	0.964	0.910			
	B2	<b>0.724</b>	0.993			
	B3	<b>0.780</b>	0.989			
	B4	<b>0.701</b>	0.944			
	C1	0.914	0.934			
	C2	<b>0.720</b>	0.958			
	C3	<b>0.716</b>	0.958			
	C4	<b>0.720</b>	0.981			
	Y1	0.886	0.988			
	Y2	0.868	0.969			
	Y3	0.868	0.960			
	Y4	0.843	0.969			

Note: Bold values indicate unacceptable bias estimates

Next, Table 2 displays the indicator loadings computed by the CBI for Model 2. For Model 2, each item loading was set to 0.8. The research revealed that the regularized ULS consistently outperformed ULS across all sample sizes (n= 50, 100, 200, 500) with the lowest RMSEA value. On the other hand, ULS yields numerous low bias indicator estimates (CBI <0.8). A thorough review of CBI values reveals that regularized ULS produces an estimate that is superior to ULS (CBI value close to 1). Also, for the selection of the optimal regularization parameter, a small sample model (n=50) requires an extremely large regularization parameter value,  $\lambda=0.80$  to obtain a better fit model with the smallest RMSEA. Meanwhile, the optimal regularization parameters needed for sample sizes of 100, 200 and 500 are 0.10, 0.20 and 0.45, correspondingly, denoting a non-apparent difference. From this finding, a similar pattern is observed from Model 1 in selecting the optimal value of  $\lambda$  for the regularized ULS method. Small sample (n=50) requires a very large  $\lambda$  to generate a better fit model, hence yielding a more unbiased indicator loading estimate.

**Table 3**  
The CBI, RMSEA and optimal regularization values for Model 3

Sample size	Items	CBI		RMSEA		Regularization parameter ( $\lambda$ ) for regularized ULS
		ULS	Regularized ULS	ULS	Regularized ULS	
50	A1	<b>0.768</b>	0.810	0.107	0.059	0.65
	A2	<b>0.740</b>	0.943			
	A3	<b>0.613</b>	0.984			
	A4	<b>0.518</b>	0.961			
	B1	0.981	0.809			
	B2	<b>0.740</b>	0.992			
	B3	0.873	0.981			
	B4	<b>0.765</b>	0.976			
	C1	0.964	<b>0.781</b>			
	C2	<b>0.713</b>	0.970			
	C3	<b>0.588</b>	0.956			
	C4	0.848	0.994			
	Y1	<b>0.698</b>	0.984			
	Y2	0.983	0.950			
	Y3	0.898	0.940			
	Y4	0.960	0.939			
100	A1	0.893	0.834	0.118	0.059	0.55
	A2	<b>0.650</b>	0.922			
	A3	<b>0.611</b>	0.963			
	A4	<b>0.575</b>	0.941			
	B1	0.813	0.832			
	B2	<b>0.696</b>	0.973			
	B3	<b>0.700</b>	0.964			
	B4	<b>0.643</b>	0.958			
	C1	0.823	0.802			
	C2	<b>0.743</b>	0.950			
	C3	<b>0.565</b>	0.936			
	C4	0.835	0.976			
	Y1	<b>0.750</b>	0.977			
	Y2	0.913	0.941			
	Y3	0.953	0.931			
	Y4	<b>0.788</b>	0.928			
200	A1	0.959	0.860	0.080	0.041	0.45
	A2	<b>0.779</b>	0.899			
	A3	<b>0.669</b>	0.939			
	A4	<b>0.724</b>	0.919			
	B1	0.948	0.857			
	B2	<b>0.718</b>	0.951			
	B3	<b>0.750</b>	0.944			
	B4	<b>0.650</b>	0.938			
	C1	0.866	0.824			
	C2	<b>0.648</b>	0.928			
	C3	<b>0.718</b>	0.911			
	C4	<b>0.741</b>	0.952			
	Y1	<b>0.781</b>	0.969			
	Y2	0.923	0.931			
	Y3	0.909	0.921			
	Y4	0.901	0.914			
500	A1	0.951	0.888	0.081	0.029	0.35
	A2	<b>0.731</b>	0.870			
	A3	<b>0.731</b>	0.911			
	A4	<b>0.788</b>	0.893			
	B1	0.964	0.884			
	B2	<b>0.724</b>	0.926			
	B3	<b>0.780</b>	0.922			
	B4	<b>0.701</b>	0.916			
	C1	0.914	0.849			
	C2	<b>0.720</b>	0.901			
	C3	<b>0.716</b>	0.884			
	C4	<b>0.720</b>	0.927			
	Y1	0.886	0.960			
	Y2	0.868	0.920			
	Y3	0.868	0.910			
	Y4	0.843	0.900			

Note: Bold values indicate unacceptable bias estimates

Next, every item loading under each of the corresponding constructs for Model 3 was set to 0.9. The outcome is tabulated in Table 3. Consistent with the findings for Model 2, the regularized ULS method obviously appears to be better than ULS for all sample sizes ( $n=50, 100, 200, 500$ ) with the lowest value of RMSEA. On this ground, regularized ULS is indeed able to improve the performance of the existing method; ULS, in estimating the parameters for non-normal data. However, this method requires a large sample ( $n \geq 500$ ) to yield more precise loading estimates where the CBI value is  $> 0.9$  or close to 1. Besides, a small sample model ( $n=50$ ) requires an extremely large regularization parameter value,  $\lambda=0.65$  to obtain a better fit model with the smallest RMSEA. However, there are no discernible differences in the optimal regularization parameter across sample sizes.

#### 4. Discussion

The performance of the regularized ULS and the traditional ULS methods in terms of CBI was investigated in this current study. Monte Carlo Markov Chain method was used to generate data with different sample sizes using a simple model that fits certain conditions (e.g., non-normal, complete data). Thus, some conclusions were derived from the results. As previously mentioned, the three models' actual item loadings are uniformly set between 0.7 and 0.9. For Model 1, when  $n=50$ , ULS outperforms regularized ULS with no bias estimates ( $\geq 0.8$ ), indicating that ULS is favorable for small sample sizes. Meanwhile, for  $n=100$ , the finding discloses that a number of unfavorable bias estimates ( $< 0.8$ ) was observed for both estimation methods. This finding suggests that the proposed method (i.e., the regularized ULS) generates comparable results to those of ULS when the sample sizes are 50 and 100. However, the regularized ULS method outperforms the ULS when the sample size is sufficiently large (i.e.,  $n \geq 200$ ). Moreover, the RMSEA values improved and achieved an acceptable threshold of 0.08 when employing regularized ULS. Conversely, the number of undesirable bias estimates for ULS increases as the sample size increases ( $n \geq 200$ ). Thus, it is suggested that the study is carried out using a regularized ULS estimator when data with an actual loading of 0.7 is to be simulated.

For Model 2 and Model 3, ULS yields several low bias indicator estimates (CBI  $< 0.8$ ). However, the regularized ULS is capable of yielding unbiased estimates across all sample sizes ( $n=50, 100, 200$  and  $500$ ). In light of this finding, when the actual indicator loading is high, regularized ULS is suitable to be employed for small to large samples. Therefore, it can be inferred from the results that when true indicator loadings are high, the loadings are precisely estimated and hence lie within the permissible bias range. Considering greater sample sizes ( $n=200$ ), it can hence be deduced that regularized ULS can be utilized to generate more precise parameter estimates in simulation studies with population indicator loadings greater than 0.7. The biases of indicator loading estimates in ULS for all sample sizes might be due to its underestimation. Thus, the new regularized approach has been proven to overcome this matter.

In addition, the choice of regularization parameter,  $\lambda$  is crucial in yielding better estimation. A small sample ( $n=50$ ) requires very large  $\lambda$  to generate a good fit model, hence resulting in more unbiased indicator loading estimate. In addition, the differences between optimal regularization parameters across large sample sizes ( $n \geq 100$ ) are not noticeable. Still, the selection of tuning parameters should depend on the model characteristics and sample size. By incorporating a positive constant value into the sample variance-covariance matrix, the stability of the covariance matrix is enhanced, thereby mitigating the risk of negative error variances. The values within the sample variance-covariance matrix play a crucial role in determining the model parameters. The accuracy and stability of the sample variance-covariance matrix can significantly impact the estimated parameter values. Thus, selecting the optimal value of the regularization parameter is crucial for generating the best estimate values (Arruda & Bentler, 2017). A greater  $\lambda$  value is necessary to achieve more stable and precise estimates of certain constructs especially for small sample size. Therefore, it is vital to evaluate the model through the smallest value of RMSEA and test various choices of the regularization parameter accordingly (Jacobucci et al, 2016).

#### 5. Conclusion

In summary, it is suggested that regularized ULS is a great estimator if the true indicator loadings are high (e.g.,  $\geq 0.7$ ). It has been discovered in this present study that the addition of regularization parameter to each element in the sample covariance matrix of ULS estimator has been proven to enhance the loading estimates particularly when using non-normal data as well as when the indicator loading for the simulated data is large (e.g., 0.7). Since real-life data are usually non-normal, the findings of this study can be used by the policy makers and researchers to obtain a more accurate estimation in analyzing the inter-relationships between variables.

#### References

- Afthanorhan, A., Awang, Z., Aimran, N., & Arifin, J. (2021). An Extensive Comparison Between CBSEM and Consistent PLS-SEM On Producing the Estimates of Construct Correlation in Applied Research. In *Journal of Physics: Conference Series* (Vol. 1874, No. 1, p. 012083). IOP Publishing.
- Aimran, A. N., Ahmad, S., Afthanorhan, A., & Awang, Z. (2017). The development of comparative bias index. *AIP Conference Proceedings*, 1870.
- Ainur, A. K., Sayang, M. D., Jannoo, Z., & Yap, B. W. (2017). Sample Size and Non-Normality Effects on Goodness of Fit

- Measures in Structural Equation Models. *Pertanika Journal of Science & Technology*, 25(2).
- Arruda, E. H., & Bentler, P. M. (2017). *A Regularized GLS for Structural Equation Modeling*. *Structural Equation Modeling*, 24(5), 657–665.
- Arruda, E. H. (2017). *Applications of Regularization to SEM: Shrinking Eigenvalues to Improve Stability of Covariance Matrices*. University of California, Los Angeles.
- Awang, Z. (2023). *SEM Made Simple 2.0: A Gentle Approach to Learning Structural Equation Modeling*. MPWS Rich Publication, Bangi.
- Bickel, P. J., & Li, B. (2006). Regularization in statistics. *TEST: Sociedad de Estadística e Investigación Operativa*, 15, 271–344.
- Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling*, 16(4), 625–641.
- Henseler, J., & Chin, W. W. (2010). A comparison of approaches for the analysis of interaction effects between latent variables using partial least squares path modeling. *Structural equation modeling*, 17(1), 82–109.
- Jacobucci, R., Brandmaier, A. M., & Kievit, R. A. (2019). A Practical Guide to Variable Selection in Structural Equation Modeling by Using Regularized Multiple-Indicators, Multiple-Causes Models. *Advances in Methods and Practices in Psychological Science*, 2(1), 55–76.
- Jacobucci, R., Grimm, K. J., & McArdle, J. J. (2016). Regularized Structural Equation Modeling. *Structural Equation Modeling*, 23(4), 555–566.
- Jung, S. (2013). Structural equation modeling with small sample sizes using two-stage ridge least-squares estimation. *Behavior Research Methods*, 45(1), 75–81.
- Jung, S., & Takane, Y. (2008). Regularized common factor analysis. *New trends in psychometrics*, 141–149.
- Kline, R. B. (2016). *Principles and practice of structural equation modeling*. Guilford publications.
- McDonald, R. P., & Bollen, K. A. (1990). Structural Equations with Latent Variables. *In Journal of the American Statistical Association*, 85(412).
- Mindrila, D. (2010). Maximum likelihood (ML) and diagonally weighted least squares (DWLS) estimation procedures: A comparison of estimation bias with ordinal and multivariate non-normal data. *International Journal of Digital Society*, 1(1), 60–66.
- Pavlov, G., Shi, D., & Maydeu-Olivares, A. (2020). Chi-square Difference Tests for Comparing Nested Models: An Evaluation with Non-normal Data. *Structural Equation Modeling*, 27(6), 908–917.
- R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48, 1–36.
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of psychological research online*, 8(2), 23–74.
- Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, 48(3), 465–471.
- Yuan, K. H., & Bentler, P. M. (2017). Improving the convergence rate and speed of Fisher-scoring algorithm: Ridge and anti-ridge methods in structural equation modeling. *Annals of the Institute of Statistical Mathematics*, 69, 1–27.
- Yuan, K. H., & Chan, W. (2016). Structural equation modeling with unknown population distributions: Ridge Generalized Least Squares. *Structural Equation Modeling: A Multidisciplinary Journal*, 23, 163–179.
- Zulkifli, R., Aimran, N., Deni, S. M., & Badarisam, F. N. (2022). A comparative study on the performance of maximum likelihood, generalized least square, scale-free least square, partial least square and consistent partial least square estimators in structural equation modeling. *International Journal of Data and Network Science*, 6(2), 391–400.

