International Research Journal of Multidisciplinary Scope (IRJMS), 2024; 5(4):392-401

Original Article | ISSN (O): 2582-631X DOI: 10.47857/irjms.2024.v05i04.01411

The Effect of Regularized Unweighted Least Squares on CFI, TLI, RMSEA and TLI in Structural Equation Modeling

Nurul Raudhah Zulkifli*, Nazim Aimran, Sayang Mohd Deni

School of Mathematical Sciences, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia. *Corresponding Author's Email: nurulraudhahzulkifli@gmail.com

Abstract

In a confirmatory investigation, researchers are mandated to employ covariance-based structural equation modeling (CB-SEM). A crucial assumption inherent in CB-SEM is the multivariate normality of the data. However, real-world data rarely conforms to a perfectly normal distribution. To address this, unweighted least squares (ULS) is specifically tailored for handling non-normally distributed data in SEM. Nonetheless, ULS often yields unsatisfactory outcomes, such as negative or boundary estimates of unique variances, as it accounts for measurement errors in observed variables. In the realm of SEM, unique variance manifests as disturbance, arising from unreliability or measurement error and reliable variation in items indicating latent causes that are not explicitly known. One common cause of improper solutions in SEM is non-convergence, wherein the estimation fails to reach a minimum fit function. To address this challenge, the present study proposes the regularization of the ULS estimator to rectify inadequacies in model fit. Multivariate non-normally distributed data, with predetermined population parameters and sample sizes, were generated through Pro-Active Monte Carlo simulation and subsequently analyzed using the R Programming Environment. The results reveal the effectiveness of the regularized ULS in enhancing model fit indices such as the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR).

Keywords: Fitness Indexes, Monte Carlo, Regularization, SEM, ULS.

Introduction

The second-generation statistical analysis method, known as structural equation modeling (SEM), is proficient in examining the complex interconnections among multiple variables within a model (1–4). Past study has demonstrated that SEM outperforms ordinary least squares (OLS) in achieving the best-fit model for estimation (4). SEM is capable of handling both normal and non-normal data, with the unweighted least square (ULS) estimator specifically designed for non-normal data (5). Given that real-life data often deviate from normality (6, 7), this paper presents comparable findings of various estimation methods in covariance-based SEM (CB-SEM), including maximum likelihood (ML), ULS, and regularized ULS, for analyzing the inter-relationships of variables. ULS frequently leads to improper solutions, such as negative or boundary estimates of unique variances, due to its consideration of measurement errors in observed variables (8). In SEM, unique variance is depicted as disturbance, involving random error (arising from unreliability or measurement error) and reliable variation in the item (indicating unknown latent causes). Improper solutions in SEM are partially attributed to non-convergence, where the estimation fails to reach a minimum fit function. Hence, regularization techniques were introduced to overcome this issue. Regularization is defined as the condition of having been rendered regular in layman's terms. It is frequently used in the area of mathematics to describe the inclusion of information to resolve an improper problem. Methods of regularization are frequently used in fields including statistics and machine learning. Multicollinearity and overfitting are prevalent in most applications. There has been a vast amount of work in statistics dealing with regularization in a wide spectrum of problems (9–15). According to scholars, regularization is the class of techniques needed to modify maximum likelihood to give reasonable answers in unstable situations such as negative or boundary estimates of unique variances (11).

(Received 24th June 2024; Accepted 22nd October 2024; Published 30th October 2024)

The negative or boundary estimates of unique variances means that during estimation, the computed value for the unique variances (variances of individual variables) might either be negative or reach the boundary constraints of the estimation procedure. Hence, previous study extended the use of regularization to structural equation modeling, namely, regularized structural equation modeling (RegSEM) (14). RegSEM penalizes certain parameters, such as factor loadings and path coefficients. The aim is to develop models that are simpler and easier to comprehend.

Through the inclusion of penalties on particular parameters within a SEM model, researchers gain significant flexibility in simplifying model complexity, addressing poor fitting models, and improving loading estimation accuracy. Several regularization studies have been emphasized and utilized in SEM (16–19). Arruda and Bentler introduced regularized Generalized Least Squares (GLS), which employs a maximum posteriori (MAP) covariance matrix for the regularized weight matrix (9). The objective of the study is to evaluate the performance of the ML, GLS, RGLS, and rGLS test statistics in a confirmatory factor model across different sample sizes. The results showed that at all sample sizes below 500, regularized GLS outperformed ML and GLS. However, in larger samples, their performance was equivalent. As a result, this study suggests that regularization using alternative estimators in SEM to nonnormal data should be used more widely. This study only focused on regularizing only the weight matrix in GLS. Recent study introduced regularized PLSc, a ridge-type regularization to be included into consistent partial least squares (PLSc). By including the regularization parameter in the estimation, PLSc and regularization parameter were integrated with the goal of addressing potential multicollinearity issues. When there is significant multicollinearity, regularized PLSc should be considered (20). There is general agreement that regularized parameter estimates are, on average, more accurate than their nonregularized least-squares counterparts. Additionally, at lower sample sizes, the regularization impact is more noticeable.

Recently, a study by Robitzsch A explores different aspects of implementing regularized single-group and multiple-group structural equation modeling (SEM) (19). The finding indicates that, in certain cases, using a fixed regularization parameter is better than choosing an optimal one based on the Bayesian information criterion for estimating structural parameters. Additionally, it compares widely used penalty functions in regularized SEM across various R packages with a recently proposed penalty function in Mplus software. The study offers practical guidance for implementing regularized SEM in future software applications. However, past methods place their focus heavily on the application of RegSEM method to different estimators such as ML using different model and data conditions. Therefore, our objective is to assess the performance of regularized ULS, as well as ML and ULS, across various SEM fit indices (e.g., CFI, TLI, RMSEA, and SRMR).

Methodology

Simulation Research Framework

In this study, we employed Monte Carlo Markov Chain (MCMC) simulation procedures to generate non-normally distributed data, adhering to the standard approach with specified skewness and kurtosis value of 2 and 7, respectively (21). Three distinct population models were constructed, each featuring specific specifications of true indicator loadings. These models consisted of four latent constructs, with uniform true indicator loadings of 0.7, 0.8, and 0.9, along with a correlation of 0.7 for each pair (22). To mitigate underestimation concerns associated with small sample sizes, high indicator loadings were established. The relationships between constructs in the population were consistently characterized as homogeneous. Sample sizes of 50, 100, 200, and 500 were generated for the purpose of evaluating consistency, aligning with the common sample sizes employed in path modeling (22, 23). Following that, the nonnormal data were generated based on specified population models across four sample sizes (n= 50, 100, 200 and 500) to gain insights into the robustness, efficiency, and generalizability of the methods under different conditions. It allows for a more comprehensive understanding of how changes in sample size impact the validity and reliability of simulation results. To obtain consistent outcomes for the analyses, 1,000 replications of each sample size is conducted, resulting in the generation of 3 models x 4 sample sizes x 1,000 replications = 12,000 datasets. The estimation of indicator loadings was performed using Maximum Likelihood (ML), Unweighted Least Squares (ULS), and regularized ULS methods. For regularization, we chose the optimal value of the regularization parameter, λ using cross-validation (24). The simulation process and subsequent data analysis for structural equation modeling (SEM) were carried out using the R statistical programming environment. Various packages, including “psych”, “MASS”, “foreign”, “mvrnonnorm”, and “semTools” were employed to generate multivariate non-normal data. Following this, SEM and regularized SEM analyses were conducted using the “lavaan” and “regsem” packages, respectively. Figure 1, Figure 2, and Figure 3 visually present the population models that underwent testing with maximum likelihood (ML), unweighted least square (ULS), and regularized ULS adapted from (22). For the estimation purposes, IBM SPSS AMOS software was utilized, incorporating both maximum likelihood (ML) and unweighted least square (ULS) estimators. In the context of regularized unweighted least squares estimation, penalties were applied to the coefficients derived from ULS.

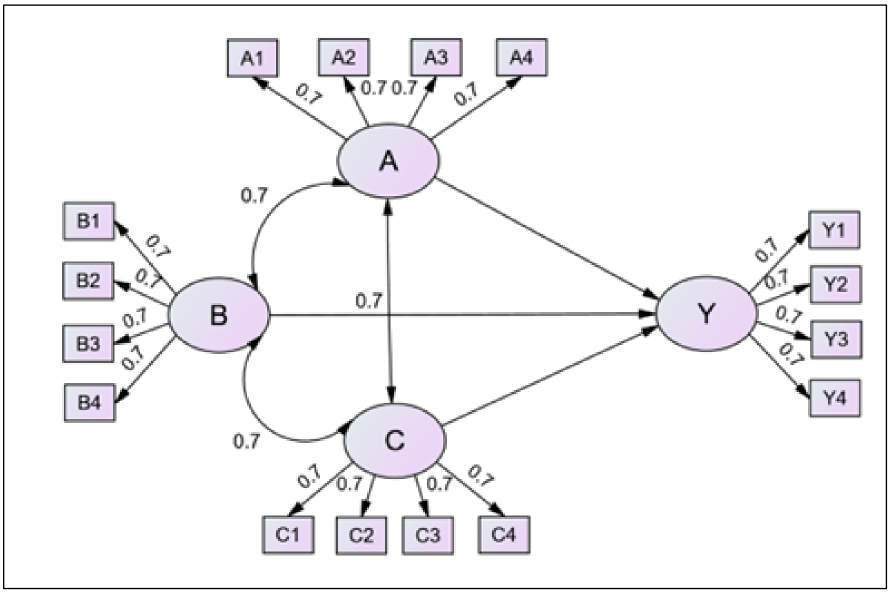

Figure 1: Reflective Measurement Model 1

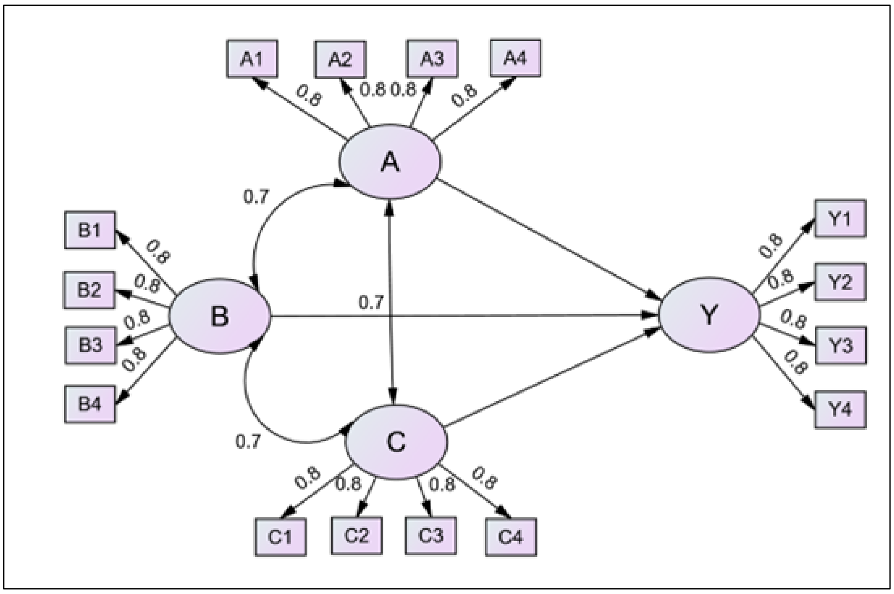

Figure 2: Reflective Measurement Model 2

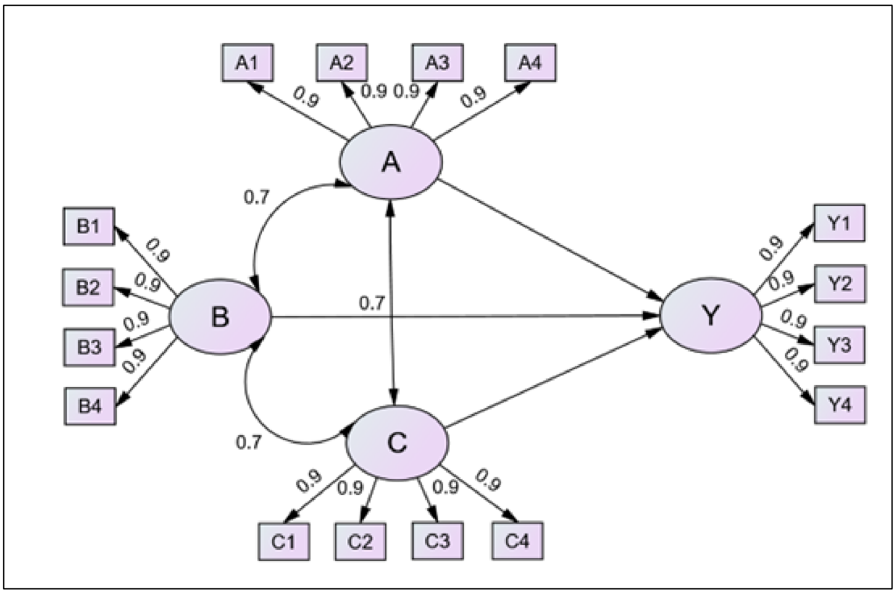

Figure 3: Reflective Measurement Model 3

Estimation Methods

Maximum Likelihood (ML)

In projecting the fit and coefficients within CB-SEM, we opted for the elegant approach of maximum likelihood (ML) estimation. ML leverages derivatives with finesse to minimize the following fit function:

$$F_{ML} = \log|\Sigma(\theta)| + tr\left(S\Sigma^{-1}(\theta)\right) - \log|S| - (p+q)\,. \quad [1]$$

Where the covariance matrix of the theoretical model is denoted as Σ, and the sample covariance matrix is defined as S. For a square matrix$\mathrm{B}$, $|\mathrm{B}|$ implies the determinant of B; $tr(\mathrm{B})$ defines the sum of the diagonal elements of B; and $p + q$ is the total numbers of manifest variables indicators. The fitting function in ML is derived under the assumption that the observed variables exhibit a normal distribution.

Unweighted Least Squares (ULS)

In light of the non-normal distribution in the data used for this study, we appraised the performance of the unweighted least square (ULS). As outlined by Shi and Maydeu-Olivares (25), the ULS fit function can be formulated as:

$$F_{ULS} = \varepsilon_r' \varepsilon_r\,. \quad [2]$$

Here, $\varepsilon_r = \kappa - \kappa_0$ represents the disparity between the population threshold and polychoric correlations.

Regularized Unweighted Least Squares

The ULS estimation method in SEM has certain disadvantages, particularly related to conditioning problems. To address these issues, several solutions involving regularization have been proposed in the statistical literature. Regularization techniques aim to improve the stability and accuracy of parameter estimation. One regularization approach is shrinkage, which dates back to the work of Stein C (26). Shrinkage methods have been implemented in SEM, as seen in a study (27). This method involves penalizing specific parameters in the model to achieve regularization. Regularized structural equation modeling (RegSEM), which applies regularization techniques to SEM was introduced previously (14). RegSEM penalizes certain parameters in the model to encourage stability and improve estimation accuracy. In general, regularization techniques aim to bring stability and conformity to problematic situations. This help to provide reasonable answers and address instability issues in the estimation of SEM parameters. The goal of RegSEM is to improve the accuracy and stability of SEM estimates, particularly in situations where the sample size is small, or the model is complex (28). The distinction between RegSEM and other forms of regularization comes in terms of the specification of which parameters to penalize. RegSEM allows the regularization of parameters from general SEM models. RegSEM involves adding a penalty term to the likelihood function that encourages the model to have smaller coefficients and a simpler structure. This helps to improve the generalization of the model. One advantage of RegSEM in estimating parameters is that regularization can help to prevent the occurrence of negative variances or other implausible parameter estimates that can arise in the presence of Heywood cases (18). By adding a penalty term to the likelihood function, regularization can constrain the parameter estimates to reasonable ranges and avoid extreme values that can result from overfitting or other issues in the model. Moreover, regularization can help to improve the estimation precision of the parameters in the presence of Heywood cases by reducing the impact of small or noisy data points on the estimates. This can lead to more stable and accurate estimates, even when the data is complex or challenging to model. Overall, RegSEM can be a valuable tool in dealing with Heywood cases and other challenges in structural equation modeling and can help to improve the reliability and interpretability of the parameter estimates. A regularized extension into ULS, which incorporates ridge-type regularization by minimizing:

$$F_{regULS} = F_{ULS} + \lambda \sum_{j=1}^{p} \beta_j^2 . \qquad [3]$$

In this context, λ denotes the regularization parameter (or tuning parameter), *p* is the number of covariables used in the model, β is the coefficient for x_j. The initial summation signifies the ULS fit function that we aim to minimize, while the second summation illustrates the penalty imposed on the coefficients. A certain degree of shrinkage is identified where the advantages of reducing variance surpass the trade-off of increased bias, ultimately yielding more accurate estimates. The tuning parameter plays a crucial role in regulating the impact on bias and variance. With an increase in the tuning parameter, the effect diminishes, resulting in a reduction of variance. However, beyond a certain point, the model starts losing essential properties, leading to an introduction of bias and causing the model to underfit. Therefore, the choice of the tuning parameter should be made judiciously, taking into account the delicate balance between bias and variance. To select the optimal value of λ, a range of values, typically ranging from 20 to 100, is considered, and the model is run for each penalty value (23). The initial penalty value is set to zero, and then it is progressively increased. This approach is adopted due to the likelihood of encountering estimation issues in SEMs involving latent variables. If estimation problems arise, the testing process can be stopped. The RegSEM method is implemented as a user-friendly package called 'regsem' in the R programming language (28, 29). This package simplifies the process of fitting a model using the 'lavaan' package and subsequently applying regularization using 'regsem'.

Fitness Indexes

Comparative Fit Index (CFI)

The Comparative Fit Index (CFI), introduced previously (30), evaluates the enhancement in the fit of the hypothesized model in comparison to a baseline model. The population CFI is articulated as follows:

$$CFI = 1 - \frac{F_0}{F_B} . \qquad [4]$$

Where F_0 and F_B signify the fit function for the postulated model (i.e., the hypothesized model) and the standard model (i.e., the baseline model) in which all observed variables are considered to be uncorrelated. Acceptable cut-off values for CFI typically fall above 0.90, signifying a good fit.

Tucker-Lewis Index (TLI)

The Tucker-Lewis Index (TLI) functions as an incremental fit index. The Non-Normed Fit Index (NNFI), also known as TLI, was introduced to mitigate the influence of sample size on the Normed Fit Index's drawback. TLI is calculated using the formula presented below (31):

$$TLI = \frac{\left(F_i / df_i\right) - \left(F_t / df_t\right)}{\left(F_i / df_i\right) - 1/(N-1)} . \qquad [5]$$

Where, $df_t = s - t$ is the number of degree of freedom for the target model. The acceptable cut-off values for TLI is TLI > 0.90.

Root Mean Square Error of Approximation (RMSEA)

The population Root Mean Square Error of Approximation (RMSEA) gauges the discrepancy attributed to approximation per degree of freedom and is calculated as follows:

$$RMSEA = \sqrt{\frac{F_0}{df}}. \quad [6]$$

where F_0 implies the difference between the data-generating process and the fitted model. df refers to the degrees of freedom of the proposed model. The established threshold for an acceptable RMSEA is set at <0.08.

Standardized Root Mean Square Residual (SRMR)

The Standardized Root Mean Square Residual (SRMR) serves as an indicator of approximate fit. The population SRMR is defined as:

$$SRMR = \sqrt{\frac{\varepsilon_s^{'} \varepsilon_s}{t}}. \quad [7]$$

In this context, ε_s represents the vector of the population standardized residual covariances and $t = p(p+1)$ denotes the number of unique components in the residual covariance (correlation) matrix. The SRMR serves as a standardized effect size measure that reflects model misfit, approximately representing the average standardized residual covariance (32). According to the recommendation in a previous study (33), an SRMR value of ≤ 0.08 is considered an appropriate threshold for assessing model fit.

Results

Table 1, Table 2 and Table 3 summarize the performance of ML, ULS, and regularized ULS based on fitness indexes: CFI, TLI, RMSEA, and SRMR across the three pre-specified models.

Table 1: Model 1 Fit Indexes (True Loadings Set at 0.7)

Sample size	Fitness Indexes	Estimation methods		
		ML	ULS	Regularized ULS
50	CFI	.806	.987	.974
	TLI	.762	.984	.968
	RMSEA	.123	.087	.124
	SRMR	.086	.084	.086
100	CFI	.931	.993	.992
	TLI	.915	.991	.990
	RMSEA	.068	.059	.069
	SRMR	.062	.064	.063
200	CFI	.944	.991	.992
	TLI	.931	.989	.990
	RMSEA	.054	.057	.054
	SRMR	.051	.050	.052
500	CFI	.975	.993	.995
	TLI	.969	.992	.994
	RMSEA	.030	.036	.030
	SRMR	.034	.033	.035

Note: Values in bold denote unacceptable fit – calculation was based on IBM-SPSS AMOS and R Programming software. The threshold for CFI ≥ 0.9, TLI ≥ 0.9, RMSEA < 0.08, SRMR < 0.08.

The assessment of model fit, gauged through CFI, TLI, RMSEA, and SRMR measures for ML, ULS, and regularized ULS in Model 1, is encapsulated in Table 1. Notably, at a small sample size (n = 50), the CFI and TLI measures for Model 1 were deemed unfit when compared to ULS and regularized ULS. Nevertheless, with an upswing in the sample size, ML's CFI and TLI measures reached commendable levels, hinting at the sample size's impact on these metrics when ML was applied. Conversely, both RMSEA and SRMR measures for all estimators were deemed unsatisfactory at a smaller sample size (n = 50). In summary, the outcomes from Model 1 highlight that, when dealing with non-normal data, ULS and regularized ULS consistently surpass ML, showcasing superior model fit, as indicated by CFI and TLI measures across varying sample sizes (50, 100, 200, and 500).

Table 2: Model 2 Fit Indexes (True Loadings Set at 0.8)

Sample size	**Fitness Indexes**	**Estimation methods**		
		ML	ULS	Regularized ULS
50	CFI	.976	1.000	.995
	TLI	.971	1.041	.993
	RMSEA	.037	.000	.037
	SRMR	.075	.074	.076
100	CFI	.963	1.000	.989
	TLI	.954	1.030	.986
	RMSEA	.043	.000	.043
	SRMR	.061	.059	.060
200	CFI	.988	1.000	.997
	TLI	.986	1.011	.996
	RMSEA	.024	.000	.025
	SRMR	.042	.040	.042
500	CFI	.978	1.000	.995
	TLI	.973	1.000	.994
	RMSEA	.034	.004	.031
	SRMR	.031	.031	.032

Note: Values in bold denote unacceptable fit – calculation was based on IBM-SPSS AMOS and R Programming software. The threshold for CFI ≥ 0.9, TLI ≥ 0.9, RMSEA < 0.08, SRMR < 0.08.

Table 2 outlines the performances of ML, ULS, and regularized ULS for Model 2. The results reveal that, in non-normal conditions, the ULS estimator yielded an overfit model across all sample sizes (50, 100, 200, and 500) based on CFI, TLI, and RMSEA measures. This implies that regularized ULS demonstrated a superior fit compared to both ULS and ML for Model 2, where every item loading underlying the respective constructs was set at 0.8. Intriguingly, ML also exhibited the ability to generate a better fit model than ULS when non-normal data were utilized. When comparing ML and regularized ULS using CFI and TLI measures, regularized ULS exhibited superior performance than ML. However, in terms of the SRMR measure, all models displayed comparable fit values. Consequently, we deduce that when analyzing non-normal data with the Model 2 specification, regularized ULS emerges as the method producing the most optimal fit model.

Table 3: Model 3 Fit Indexes (True Loadings Set at 0.9)

Sample Size	**Fitness Indexes**	**Estimation Methods**		
		ML	**ULS**	**Regularized ULS**
50	CFI	.848	1.000	.966
	TLI	.814	1.015	.958
	RMSEA	.116	.000	.117
	SRMR	.089	.081	.086
100	CFI	.949	1.000	.985
	TLI	.938	1.021	.981
	RMSEA	.061	.000	.062
	SRMR	.063	.058	.062
200	CFI	.974	1.000	.994
	TLI	.968	1.007	.992
	RMSEA	.046	.000	.046
	SRMR	.043	.039	.043

Sample Size	Fitness Indexes	Estimation Methods		
		ML	**ULS**	**Regularized ULS**
500	CFI	.989	1.000	.998
	TLI	.986	1.002	.997
	RMSEA	.030	.000	.030
	SRMR	.026	.025	.026

Note: Values in bold denote unacceptable fit – calculation was based on IBM-SPSS AMOS and R Programming software. The threshold for CFI ≥ 0.9, TLI ≥ 0.9, RMSEA < 0.08, SRMR < 0.08.

The findings for Model 3 are depicted in Table 3. Among the three estimators, generalized ULS emerged as the optimal estimator, consistently achieving model fitness within acceptable threshold values across all sample sizes (50, 100, 200, and 500). Similar to Model 1, Model 3 was deemed unfit when ML was applied to a small sample size (n = 50). However, there was an improvement in fitness indexes with an increase in sample size under ML. Similar to Model 2, the ULS estimator resulted in an overfit model, evident from CFI, TLI, and RMSEA measures across all sample sizes (50, 100, 200, and 500) under non-normal conditions. This underscores the effectiveness of regularized ULS in enhancing model fit, particularly when dealing with non-normal data and extremely large indicator loadings (0.9).

Discussion

This study assessed the effect of regularization to SEM on fitness indexes such as CFI, TLI, RMSEA and SRMR. The application of regularization to SEM demonstrated that regularization could improve the fit of the model. In this study, we considered different types of models to generate non-normal and complete data using simulations with various sample sizes. As stated, the true loadings of indicators for the three models are homogenous between 0.7 to 0.9. In summary, the outcomes from Model 1 indicate that, across all sample sizes (50, 100, 200, and 500), both ULS and regularized ULS consistently produced a superior fit compared to ML, as measured by CFI and TLI. For Model 2, where all item loadings were set at 0.8, regularized ULS demonstrated improved fit over ULS and ML. Notably, ULS resulted in overfit model estimations across all sample sizes based on CFI, TLI, and RMSEA measures. Similarly, ULS yielded overfit models across various sample sizes. The findings suggest that regularized ULS effectively enhances model fit, particularly when dealing with non-normal data and substantial indicator loadings (e.g., 0.8 and 0.9). Conversely, ML rendered the model unfit, particularly at a small sample size (n = 50). This underscores the significant influence of sample size on the application of regularization in SEM, as emphasized in previous research (14, 18). This study offers a proficient solution to counter the issue of inadequate fit performance observed in ULS (8). Furthermore, it provides valuable insights into the wider application of regularization in Structural Equation Modeling (SEM), particularly in situations where regularization is acknowledged to outperform the traditional estimator in the context of CB-SEM. In situations where the data is non-normal and the true loading is large (e.g., 0.8 and 0.9), researchers should be aware that the model fit indexes can change significantly with the use of the regularization method. Our findings underscore the potential advantages of integrating regularization with alternative estimators in Structural Equation Modeling (SEM), moving beyond the common reliance on Maximum Likelihood (ML) estimation. This is particularly relevant when dealing with non-normally distributed data, as indicated by the insights in past studies (18, 22). We infer that when simulating data with true loadings set at 0.8 and 0.9, conducting the study using regularized ULS as an alternative fitting function in CB-SEM could offer a viable and beneficial approach. We propose the optimal value for the regularization parameter, λ, as 0.01, determined through cross-validation, effectively addressing the overfitting concern in the context of this study. The regularization method consistently demonstrated improved model fit compared to the non-regularized ULS. The choice of lambda plays a crucial role in determining the extent of penalization for parameter estimation, thereby enhancing the fit function (22, 23, 28). It is important to note that the comparison of ML, ULS, and regularized ULS applications across the three predefined models may not be universally applicable to all models. The conclusions drawn are specific to the models examined within the scope of this study. For a more comprehensive understanding, it is advisable to apply the proposed method to a diverse set of applications involving more complex models.

In conclusion, we investigated how the inclusion of regularization in the structural equation modeling (SEM) estimator affects different fitness indexes in this simulation study. Our results demonstrate that applying regularization to the SEM estimator leads to notable improvements in model fit, particularly in situations involving non-normal data and higher indicator loadings (e.g., 0.8 and 0.9) for simulated data (34). Given the prevalent non-normality in real-world data, the implications of this study offer valuable insights for policymakers and researchers aiming to enhance the accuracy of estimations when analyzing the inter-relationship of variables.

Abbreviation

Nil.

Acknowledgement

Nil.

Author Contributions

Dr. Nurul Raudhah Zulkifli: Study design, Data, Analysis and interpretation, Drafting of the manuscript, Revision of the manuscript, Dr. Ahmad Nazim Aimran: Study design, Revision of the manuscript, Assoc. Prof. Dr. Sayang Mohd Deni: Study design.

Conflict of Interest

The authors affirm the absence of any identified conflicts of interest associated with this publication and emphasize that there has been no considerable financial support for this work that might have affected its outcomes.

Ethics Approval

Not applicable.

Funding

No financial support received for the study.

References

1. Awang Z, Afthanorhan W, Lim SH, Zainudin NFS. SEM Made Simple 2.0 A Gentle Approach of Structural Equation Modelling. Gong Badak: Penerbit Unisza. 2023.
2. Ainur AK, Sayang MD, Jannoo Z, Yap BW. Sample size and non-normality effects on goodness of fit measures in structural equation models. Pertanika J Sci Technol. 2017;25(2):575–86.
3. Aimran AN, Ahmad S, Afthanorhan A, Awang Z. The assessment of the performance of covariance-based structural equation modeling and partial least square path modeling. AIP Conf Proc. 2017;1842 (1):030001.
4. Nazim A, Ahmad S. A comparison between Ordinary Least Square (OLS) And Structural Equation Modeling (SEM) methods in estimating the influencial factors of 8th grades student's mathematics achievement in Malaysia. Int J Sci Eng Res. 2013;4(7):717–22.
5. Mîndrilă D. 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 for Digital Society. 2010;1(1):60–6.
6. Gao S, Mokhtarian PL, Johnston RA. Nonnormality of data in structural equation models. Transp Res Rec. 2008;2082(1):116–24.
7. Hair, Hult GT, Ringle CM, Sarstedt M, Thiele KO. Mirror, mirror on the wall: a comparative evaluation of composite-based structural equation modeling methods. J Acad Mark Sci. 2017;45(5):616–32.
8. Jung S, Takane Y. Regularized common factor analysis. New Trends in Psychometrics. 2007;1(1):1–10.
9. Arruda EH, Bentler PM. A Regularized GLS for Structural Equation Modeling. Structural Equation Modeling. 2017;24(5):657–65.
10. Bauer F, Pereverzev S, Rosasco L. On regularization algorithms in learning theory. J Complex. 2007;23(1):52–72.
11. Bickel PJ, Li B, Tsybakov AB, van de Geer SA, Yu B, Valdés T, *et al*. Regularization in statistics. Test. 2006;15(2):271–344.
12. Bickel PJ, Levina E. Regularized estimation of large covariance matrices. Ann Stat. 2008;36(1):199–227.
13. Hoang HS, Talagrand O. On Regularization Approach to Parameter Estimation and its Application to Design of Stable Filters. IFAC Proceedings Volumes. 1993;26(2):291–6.
14. Jacobucci R, Grimm KJ, McArdle JJ. Regularized Structural Equation Modeling. Structural Equation Modeling. 2016;23(4):555–66.
15. Pourahmadi M. Covariance estimation: The GLM and regularization perspectives. Statistical Science. 2011;26(3):369–87.
16. Eminita V, Notodiputro KA, Sartono B. Variable that influence achievement of indonesian students in the program international student assessment (PISA) 2015 using structural equation modelling (SEM). J Phys Conf Ser. 2020;1521(4):042041.
17. Finch WH, Miller JE. A Comparison of Regularized Maximum-Likelihood, Regularized 2-Stage Least Squares, and Maximum-Likelihood Estimation with Misspecified Models, Small Samples, and Weak Factor Structure. Multivariate Behav Res. 2021;56(4):608–26.
18. Jacobucci R, Brandmaier AM, Kievit RA. A Practical Guide to Variable Selection in Structural Equation Modeling by Using Regularized Multiple-Indicators, Multiple-Causes Models. Adv Methods Pract Psychol Sci. 2019;2(1):55–76.

19. Robitzsch A. Implementation Aspects in Regularized Structural Equation Models. Algorithms. 2023;16(9):446.
20. Jung S, Park JH. Consistent partial least squares path modeling via regularization. Frontiers in Psychology. 2018;9:174.
21. Vale CD, Maurelli VA. Simulating multivariate nonnormal distributions. Psychometrika. 1983;48(3):465–71.
22. Zulkifli NR, Aimran N, Deni S. The performance of unweighted least squares and regularized unweighted least squares in estimating factor loadings in structural equation modeling. International Journal of Data and Network Science. 2023;7(3):1017–24.
23. Zulkifli NR, Aimran N, Deni S, Sapri A. A comparative study of the performance of unweighted least squares and regularized unweighted least squares in structural equation modeling. AIP Conf Proc. 2024;3150(1):040011.
24. Jacobucci R, Grimm KJ, Brandmaier AM, Serang S, Kievit RA, Scharf F, *et al*. Package 'regsem'. 2023. https://github.com/Rjacobucci/regsem/
25. Shi D, Maydeu-Olivares A. The Effect of Estimation Methods on SEM Fit Indices. Educ Psychol Meas. 2020;80(3):421–45.
26. Stein C. Estimation of a covariance matrix. Rietz Lecture. 1975;(1):1–6.
27. Yuan KH, Chan W. Structural equation modeling with near singular covariance matrices. Comput Stat Data Anal. 2008;52(10):4842–58.
28. Li X, Jacobucci R. Regularized structural equation modeling with stability selection. Psychol Methods. 2022;27(4):497–518.
29. Li X, Jacobucci R, Ammerman BA. Tutorial on the use of the regsem package in R. Psych. 2021;3(4):579-92.
30. Bentler PM, Yuan KH. Structural equation modeling with small samples: Test statistics. Multivariate Behav Res. 1999;34(2):181–97.
31. Schermelleh-Engel K, Moosbrugger H, Müller H. Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. Methods of psychological research online. 2003;8(2):23-74.
32. Shi D, Maydeu-Olivares A, DiStefano C. The relationship between the standardized root mean square residual and model misspecification in factor analysis models. Multivariate Behav Res. 2018;53(5):676–94.
33. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equ Modeling. 1999;6(1):1–55.
34. Zulkifli R, Aimran N, Deni S, Badarisam F. 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. 2022;6(2):391–400.