Bootstrapping technique in structural equation modeling: a Monte Carlo study

Nor Iza Anuar Razak¹, Zamira Hasanah Zamzuri² and Nur Riza Mohd Suradi³
¹,²,³School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600, Bangi, Selangor, Malaysia

E-mail: aizanuar@gmail.com

Abstract. Structural Equation Modeling (SEM) is a powerful statistical technique that used to measure the causal relationships between variables. SEM is common among social science researchers, but not with clinical researchers as in the clinical field, data commonly available in smaller sample size. Hence, it is affecting the performance of SEM. This study was to propose the use of Double Bootstrap method on SEM (DBSEM) with smaller samples. DBSEM is an extension of Bootstrap method on SEM (BSEM) in which we resample residuals from original model SEM. With the estimated residual errors with sample size $n$ as a population, a bootstrap sample of $n$ persons with residual errors was drawn randomly with replacement. The DBSEM was expected to offer a practical and efficient performance compared to the original SEM. The double residual bootstrap method (resample with replacement) was used on SEM. A Monte Carlo simulation with the normal data distribution ($n = 30, 50, 75, \text{ and } 100$) was used to test the performance of models. Several point estimators such as Standard Error (SE), Mean Square Error (MSE) and Root Mean Square Error (RMSE) were used to measure models performance. The performance of DBSEM model is far well better compared to the original model, SEM. All point estimators for DBSEM showed a decreasing value compared to SEM (point estimators values are high). The result shows that for BSEM and DBSEM model, there are steep decreasing values in SE, MSE and RMSE. Since the point estimates values for DBSEM are relatively lesser compared to SEM, we can conclude that double bootstrap increase model’s accuracy and its reliability.

1. Introduction
Structural Equation Modeling (SEM) is famously known as multivariate models. Back to history dated to around 1918, a famous geneticist Sewall Wright has developed path analysis and graphical modeling simultaneously for analyzing cause-effect relationship and evaluate his biological system data. It spreads like a virus in statistical science since then. SEM basically consists of a dependent variable ($Y$) and independent variables ($x$) and has numerous advantage. For example, one can assess the reliability and validity of the model measures. Besides, SEM is a statistical model which can explain the relationship among multiple variables and combination of multiple types of analyses such as factor analyses, regressions and correlations. Furthermore, by using SEM one can evaluate whether the collected data fit a previously hypothesized model of relationships through simultaneous linear analyses. SEM models can be depicted as two regression equations:
\[ M = a_0 + aX + e_M \]
\[ Y = b_0 + bM + c'X + e_Y \]  

(1)

where \( a_0 \) and \( b_0 \) are intercepts, \( Y \), \( X \) and \( M \) represent the dependent variable, independent variable and mediation variable respectively. Meanwhile, \( a \), \( b \) and \( c \) are parameters estimated by using SEM or path analysis method, while \( e_M \) and \( e_Y \) indicate residuals or measurement errors.

In the SEM literature, [1] in their research have compared the performance of Partial Least Square (PLS) and Maximum Likelihood (ML) bootstrapping techniques in SEM by using a Monte Carlo study. The accuracy and efficiency of ML and PLS based bootstrapping in SEM were compared under several conditions such as sample sizes (\( n = 30, 50, 75, \) and 100) and distributional assumptions (normal, \( \chi^2 \) with \( df = 3 \), t-distributed with \( df = 5 \) and uniform). Results show that, PLS-based bootstrapping works best for smaller sample sizes. Contradict to that, ML-based bootstrapping is suggested for a larger sample size. Likewise, [2] have studied the concept of bootstrapping Structural Equation Models with smaller samples. With mealtime rituals in diabetes management as a case, bootstrapping method and the Bollen-Stine bootstrapped \( \chi^2 \) test were used in order to test the stability and appropriateness of the model. The model of maternal reports consists of the family social environment as exogenous variables, and mealt ime rituals coupled with child hemoglobin A1c levels as endogenous variables. In general, this research proposes a proper procedure of bootstrapping technique and also the Bollen-Stine approach used for testing models in SEM by using smaller samples.

Furthermore, [3] used double bootstrap Data Envelopment Analysis (DEA) approach for assessing the efficiency of Greek hospitals. Double bootstrap procedure was used to investigate Technical Efficiencies (TE) and Scale Efficiency (SE) determinants. The double bootstrap method also was applied to avoid inconsistency problems. The outcome was, by using bootstrap methodology, TE scores are significantly lower compared to traditional DEA scores. Also, [4] in their study have discussed the advantages of Monte Carlo (MC) confidence intervals for indirect effects. Several methods such as delta method, the distribution of the product (DP) method, non-parametric percentile bootstrap and corrections, residual-based bootstrap and parametric bootstrap method were compared. Through simulation study, it has been concluded that in terms of enhancing precision, MC method give more advantages over other asymmetric CI methods because of smoothness of the sampling distribution. Contrasted with the other methods that oblige repeated model-fitting, the MC method will be exceptionally, and it is convenient.

In the year of 2011, [5] have modelled the performance of Australian hotels by using a Data Envelopment Analysis (DEA); and double bootstrap approach was used in their simulation study. The objective was to evaluate the technical efficiency of Australian hotels from the year of 2004 to 2007. A set of data consists of two inputs and six outputs were used in their analyses. In summary, empirical results proved that the DEA bootstrap method is better than traditional DEA by correcting the bias inherent in traditional DEA models. The normal specialized effectiveness of Australian lodgings progressed continuously from 76.17% in 2004 to reach its most elevated level of 80.84% in 2007. Prior to this, [6] have compared and rationale the performance of several methods in evaluating mediation effects. Three methods were used that are the normal approximation method, bootstrapping raw data method and bootstrapping error method. Several factors were taken into account such as a) sample size b) effect size c) distribution of residual errors d) coverage probability, e) power and f) confidence intervals. The simulation study shows that the bootstrapping error methods had the best coverage probability and the largest power when effect size was medium or large and the residual errors were independent and identically distributed (iid). Meanwhile, the bootstrapping raw data methods had the best coverage probability when the effect size was medium or large and the residual errors were non-iid. However, when the effect size was small, results seem like did not have any reliable conclusions for different methods. Lastly, if an only power of different methods were taken into consideration, the conclusions would be leading to the wrong idea.
The technique of bootstrapping confidence intervals for fit indexes in Structural Equation Modeling was explained by [7]. A simulation study was run to assess the achievement of two bootstrap methods; naive bootstrap and bootstrap method by [8]. Four major model misspecification conditions in the context of single-group confirmatory factor analysis (CFA) models were evaluated; true model (TM) conditions, correlated residuals (CR) conditions, cross-loading (CL) conditions and wrong model (WM) conditions. Within that, four indexes such as Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), Goodness-of-Fit Index (GFI), and Standardized Root Mean Square Residual (SRMR) were examined. Simulation results can be divided into several parts. For fit indexes based on noncentrality, such as CFI and RMSEA, the CIs under the YHY bootstrap had relatively good coverage rates for all conditions. Meanwhile, the CIs under the naive bootstrap had very low coverage rates when the fitted model had large degrees of freedom. In contrast, for GFI and SRMR, the CIs under both bootstrap methods had poor coverage rates in most conditions.

In spite of the usage of SEM, it is important to realize that not all researchers from all fields can fully employed this method. Researchers from clinical fields for example, are not completely used with this method. As identified by [2], there were exceptionally small percentage of empirical articles were published from the clinical fields like developmental/behavioural paediatrics and paediatric psychology field. Owing to the fact, sample size issue in SEM was identified as one of the factors contributed. Sample size is one of the elements influencing SEM’s performance. Because of the small sample size, maximum likelihood and generalized least squares estimators produce a little high of $\chi^2$ values, even when multivariate normality exists. As a consequence, this scenario can lead to reduce model’s reliability. Factual reality is, large sample size are not always on hand in a critical field such as clinical because of the rare case study being experimented. Even though there are several extensive literatures related with SEM models and bootstrap method, but relatively there is none attention has been given to double residual bootstrap method on SEM models. Hence, in this paper we propose double residual bootstrap method as to measure accuracy of statistical measure on SEM and to evaluate its performance on SEM.

2. Method
Bootstrap method was first introduced by Efron in the year of 1979 as a computer-based method [9, 10]. Statistically, bootstrapping means any test or metric that relies on random sampling with replacement. The definition of bootstrapping will be given by a statistical method which it produces bootstrap data sets or some called it as phantom data sets for estimating the standard error of $\hat{\theta}$ and create confidence interval [11]. The bootstrap is a technique to estimate the uncertainty of a statistic without making distributional assumptions such as normality. There are several types of bootstrap such as jackknife, wild bootstrap, block bootstrap, smooth bootstrap, residual bootstrap and raw data bootstrap.

This study will be focusing on bootstrapping residual method; which is one of the resampling method. The main concept of bootstrapping is to estimate the empirical sampling distribution of a parameter by resampling from a sample with replacement, which all goes back to bootstrap analogy, which is “the population is to the sample as the sample is to the bootstrap samples.” In spite of the fact that each re-sample has the same number of elements as the original sample, the replacement method guarantees that each of these re-samples are likely to be slightly and randomly diverse than the original sample [1]. There are two essential points need to be taken into consideration in bootstrapping error method. As explained by [6], the design matrix ($X$) does not turn over the bootstrap samples, which means ($X$) remains the same through the entire bootstrapping process. Lastly, the residual errors have to be independently and identically distributed (iid) in the bootstrapping error method. Double residual bootstrap method is constructed by the following steps:

**Step 1:**
Let say $M_i = M_{1},\ldots, M_{n}$ and $Y_i = Y_{1},\ldots, Y_{n}$ are sample data with $\sim N(0,1)$, $i = 1,\ldots, n$. Data are independent and identically distributed (iid). Parameters $a_0,b_0,b$ and $c$ are then estimated by using Maximum Likelihood Estimation (MLE).
3rd International Conference on Mathematical Sciences and Statistics

IOP Conf. Series: Journal of Physics: Conf. Series 1132 (2018) 012072
doi:10.1088/1742-6596/1132/1/012072

Step 2:
Fitted SEM model is constructed based on the parameter values $\bar{a}_0, \bar{b}_0, \bar{b}$ and $\bar{c}$; and presented by $M_i = \bar{a}_0 + \bar{a}X_i + \bar{e}_{M_i}$ and $Y_i = \bar{b}_0 + \bar{b}M_i + \bar{c}X_i + \bar{e}_{Y_i}$. Next, compute the residual errors $\hat{e}_{M_i}$ and $\hat{e}_{Y_i}$. Residual errors can be depicted as $\hat{e}_{M_i} = M_i - \bar{M}_i$ and $\hat{e}_{Y_i} = Y_i - \bar{Y}_i$ with sample data $i = 1, \ldots, N$.

Step 3:
Next step is to implement the residual bootstrap technique. By using the estimated residual errors, considering $\hat{e}_{M_i}$ and $\hat{e}_{Y_i}$ as a population, randomly draw with replacement the paired residual errors $(\hat{e}_{M_i}, \hat{e}_{Y_i})$ in order to obtain bootstrap residual values.

Step 4:
New bootstrap sample $\bar{Y}_i^b, \bar{M}_i^b$ and $X$ are obtained; which are derived from $\bar{M}_i^b = \bar{a}_0 + \bar{a}X_i + \bar{e}_{M_i}^b$ and $\bar{Y}_i^b = \bar{b}_0 + \bar{b}ar{M}_i^b + \bar{c}X_i + \bar{e}_{Y_i}^b$.

Step 5:
Again, estimate the parameters of $\bar{a}, \bar{b}$ and $\bar{c}$ by using Step 1.

Step 6:
Repeat Steps 3-5 for a total of $B$ (Bootstrap sample size) times.

Step 7:
Models performance are evaluated by using point estimates and 95% normal and $t$-distribution confidence intervals were constructed.

Step 8:
For double residual bootstrapping in SEM, the whole steps (Step 1-7) are repeated by using the estimated residual errors of fitted BSEM $(\hat{e}_{M_i}^b, \hat{e}_{Y_i}^b)$ as a population. Finally, the performance of DBSEM will be compared with BSEM and SEM.

3. Simulation study
A detailed Monte Carlo simulation study was executed to test the performance of double residual bootstrap method on SEM. Initial parameter values of $a$, $b$ and $c$ each is sets as 0.3, 0.3 and 0.2. Normal data distributions were generated with mean = 0 and variance = 1. Gaussian distribution has been chosen because it is the most popular in empirical applications [12]. Several sample sizes $n = 30, 50, 75,$ and $100$ were taken into account. For each of sample sizes, 1000 bootstrap replication ($B = 1000$) were replicated. Models performance were evaluated based on several point estimates and interval estimates.

The point estimates included were Standard Error (SE), Mean Square Error (MSE) and Root Mean Square Error (RMSE). Meanwhile, for interval estimates, 95% normal and $t$-distribution of confidence intervals (CI) were constructed. The simulation study was run by using the R statistical programming environment. For the purpose of comparison, summary simulation results of point estimates and interval estimates for all three models (SEM, BSEM and DBSEM) each is summarized in Table 1 and Table 2.

4. Results and Discussion
Referring to the Table 1, we can conclude when the residual bootstrap method was applied to the SEM model, there were decreasing in value for SE, MSE and RMSE. For example, with $n = 30$, SE values for original SEM to BSEM was decreasing from 0.2240244 to 0.2197523. The same pattern goes to MSE and RMSE value which each is decreasing from 1.3678061 to 1.3497809 and 1.1695324 to 1.1480601 respectively. It has been shown that, this pattern (decreasing in point estimate values) are constants for other sample sizes with $n = 50, 75$ and $100$. In the same way, when the double residual bootstrap method was implemented on SEM model, point estimate values (SE, MSE and RMSE) also shows a decrescent pattern. For instance, with $n = 50$, SE values for SEM to DBSEM was decreasing from 0.1565819 to 0.1504670. In the same way, the MSE (1.028454 to 1.005193) and RMSE (1.0141271 to 0.9991439) values for DBSEM also decreasing compared to SEM.
Results from Table 1, it proved that residual bootstrap and double residual bootstrap method gave less variation, thus lead to more accurate and reliable model. Other than that, notably from Table 1, we can see a consistent decreasing pattern in SE values as the sample size increase. For BSEM, decreasing values are from 0.2197523 → 0.1545118 → 0.1177495 → 0.10231414, meanwhile for DBSEM the changes are from 0.1644386 → 0.1504670 → 0.1114908 → 0.09568257.

| $n$  | Model | SE     | MSE    | RMSE  |
|------|-------|--------|--------|-------|
| 30   | SEM   | 0.2240244 | 1.3678061 | 1.1695324 |
|      | BSEM  | 0.2197523 | 1.3497809 | 1.1480601 |
|      | DBSEM | 0.1644386 | 0.7967854 | 0.8890069 |
| 50   | SEM   | 0.1565819 | 1.028454 | 1.0141271 |
|      | BSEM  | 0.1545118 | 1.017774 | 1.0036998 |
|      | DBSEM | 0.1504670 | 1.005193 | 0.9991439 |
| 75   | SEM   | 0.1192995 | 0.9410450 | 0.9700747 |
|      | BSEM  | 0.1177495 | 0.9906773 | 0.9926078 |
|      | DBSEM | 0.1114908 | 0.8203586 | 0.9033375 |
| 100  | SEM   | 0.10278125 | 0.8803299 | 0.9364797 |
|      | BSEM  | 0.10231414 | 0.8803299 | 0.9364797 |
|      | DBSEM | 0.09568257 | 0.8803299 | 0.9364797 |

Note. SEM: Structural Equation Modeling; BSEM: Bootstrap Structural Equation Modeling; DBSEM: Double Bootstrap Structural Equation Modeling; SE: Standard error, MSE: Mean Square Error; RMSE: Root Mean Square Error; $n$: sample size

The outcomes of study also visualized in Graph 1, Graph 2 and Graph 3 below

Graph 1. Standard Error (SE)
Besides, in this simulation study, 95% normal and $t$-distribution confidence intervals (CI) were constructed and were reported in Table 2. The outcomes were, for 95% normal CIs, both BSEM and DBSEM models gave narrower CIs compared to the original SEM. As can be seen from Table 2, with $n = 30$, BSEM-N and DBSEM-N both yielded narrower normal CIs with 0.7229207 and 0.5409549 respectively, way narrower than the original SEM-N, that is 0.7369748. Equally important to notice, the same trend (narrowed normal CIs) is consistent for all BSEM-N and DBSEM-N models, regarding all sample sizes. In addition, as well as 95% $t$-distribution confidence intervals, residual bootstrap method yielded narrowed CIs to BSEM. As an example, with $n = 50$, CI for BSEM-$t$ is 0.5180937, meanwhile, CI for SEM-$t$ is 0.5250351. Subsequently, the CI of DBSEM-$t$ for the same sample size was narrowed down to 0.5045312. Then again, the same trend (narrowed $t$-distribution CIs) is constant for all BSEM-$t$ and DBSEM-$t$ models, for all sample sizes.

The confidence interval assesses the accuracy and precision of the point estimate. Assuming that a CI method yields CIs with acceptable coverage, narrow confidence intervals are generally preferred to wider ones (Preacher & Selig, 2012). Hence, referring to CIs result, we can sum up that bootstrap method does increase model’s reliability.
Table 2. 95% normal and t-distribution confidence intervals

| n  | Model   | Lower       | Upper       | CI width   |
|----|---------|-------------|-------------|------------|
| 30 | SEM-N   | -0.3553151  | 0.3816596   | 0.7369748  |
|    | BSEM-N  | -0.3193102  | 0.4036105   | 0.7229207  |
|    | DBSEM-N | -0.1823648  | 0.35859     | 0.5409549  |
|    | SEM-t   | -0.3674737  | 0.3938182   | 0.7612919  |
|    | BSEM-t  | -0.3312369  | 0.4155373   | 0.7467741  |
|    | DBSEM-t | -0.1912895  | 0.3675147   | 0.5588042  |
| 50 | SEM-N   | -0.146347   | 0.3687616   | 0.5151087  |
|    | BSEM-N  | -0.1307682  | 0.3775304   | 0.5082985  |
|    | DBSEM-N | -0.1611012  | 0.3338912   | 0.4949924  |
|    | SEM-t   | -0.1513102  | 0.3737248   | 0.5250351  |
|    | BSEM-t  | -0.1356658  | 0.382428    | 0.5180937  |
|    | DBSEM-t | -0.1658706  | 0.3386606   | 0.5045312  |
| 75 | SEM-N   | -0.2070066  | 0.1854538   | 0.3924605  |
|    | BSEM-N  | -0.2152162  | 0.1721452   | 0.3873612  |
|    | DBSEM-N | -0.1565082  | 0.2102638   | 0.3667721  |
|    | SEM-t   | -0.2094944  | 0.1879416   | 0.3974361  |
|    | BSEM-t  | -0.2176716  | 0.1746007   | 0.3922722  |
|    | DBSEM-t | -0.1588332  | 0.2125888   | 0.3714220  |
| 100| SEM-N   | -0.1672922  | 0.170828    | 0.3381202  |
|    | BSEM-N  | -0.1654041  | 0.1711795   | 0.3365836  |
|    | DBSEM-N | -0.3502803  | -0.03551263 | 0.3147677  |
|    | SEM-t   | -0.1688892  | 0.172425    | 0.3413142  |
|    | BSEM-t  | -0.1669938  | 0.1727692   | 0.3397630  |
|    | DBSEM-t | -0.3517669  | -0.03402596 | 0.3177410  |

Note. Lower, upper and CI width are percentile measure for 95% normal and t-distribution confidence intervals; SEM-N and SEM- t: Normal and t-distribution confidence intervals for SEM; BSEM-N and BSEM- t: Normal and t-distribution confidence intervals for Bootstrap SEM; DBSEM-N and DBSEM- t: Normal and t-distribution confidence intervals for Double Bootstrap SEM; n is sample size.

5. Conclusion
To summarize, this paper has practically demonstrated the ability of residual bootstrap method as an aid on SEM with smaller sample sizes. The objective of the study was to evaluate the performance of the double residual bootstrap method on SEM. It has been shown that, point estimates values (SE, MSE and RMSE) for BSEM and DBSEM were greatly decreasing compared to the original SEM. As BSEM and DBSEM gave less variation, this has led to more accurate and reliable models. Another essential point is, for interval estimates, the CIs width of BSEM and DBSEM models were way narrowed than the original SEM, regarding all sample sizes. Considering the shorter intervals are always preferred, it can be concluded that the performance of DBSEM and BSEM were slightly improved compared to SEM.

Future research may address the performance of other bootstrapping methods such as jacknife, wild bootstrap and block bootstrap. Besides, another statistical software such as SAS and AMOS may want to be considered in the next study. Another recommendation is, researchers may consider to running more complex SEM models with more variables involved.

Acknowledgements
Author would like to thank Faculty of Science and Technology, (FST) UKM for the grant FRGS/1/2015/ST06/UKM/02/1.
References

[1] Sharma P N and Kim K H 2013 A comparison of PLS and ML bootstrapping techniques in SEM: a Monte Carlo study New perspectives in partial least squares and related methods (New York: Springer) pp 201-208

[2] Ievers-Landis C E, Burant C J and Hazen R 2011 The concept of bootstrapping of structural equation models with smaller samples: an illustration using mealtime rituals in diabetes management J. Dev. Behav. Pediatr. 32 619-626

[3] Kounetas K and Papathanassopoulos F 2013 How efficient are Greek hospitals? a case study using a double bootstrap DEA approach Eur. J. Health Econ. 14 979-994

[4] Preacher K J and Selig J P 2012 Advantages of monte carlo confidence intervals for indirect effects Commun. Methods Meas. 6 77-98

[5] Assaf A G and Agbola F W 2011 Modelling the performance of Australian hotels: a DEA double bootstrap approach Tourism Econ. 17 73-89

[6] Zhang Z and Wang L 2008 Methods for evaluating mediation effects: rationale and comparison New trends in psy. 595-604

[7] Zhang X and Savalei V 2016 Bootstrapping confidence intervals for fit indexes in structural equation modeling Struct. Equ. Modeling 23 392-408

[8] Yuan K H, Hayashi K and Yanagihara H 2007 A class of population covariance matrices in the bootstrap approach to covariance structure analysis Multivariate Behav. Res. 42 261–281

[9] Efron B 1979 The 1977 RIETZ lecture Ann. Stat. 7 1-26

[10] Singh K and Xie M 2008 Bootstrap: a statistical method unpublished manuscript URL http://www.stat.rutgers.edu/home/mxie/RCPapers/bootstrap.pdf

[11] Efron B and Tibshirani R J 1993 An Introduction to the Bootstrap (London: Chapman and Hall)

[12] Pascual L, Romo J and Ruiz E 2006 Bootstrap prediction for returns and volatilities in GARCH models Comput. Stat. Data Anal. 50 2293-2312