5-1-2017

Limitations in the Systematic Analysis of Structural Equation Model Fit Indices

Sarah A. Rose
Wayne State University, rosey13579@gmail.com

Barry Markman
Wayne State University, barry.markman@wayne.edu

Shlomo Sawilowsky
Wayne State University, professorshlomo@gmail.com

Follow this and additional works at: http://digitalcommons.wayne.edu/jmasm

Part of the Applied Statistics Commons, Social and Behavioral Sciences Commons, and the Statistical Theory Commons

Recommended Citation

Limitations in the Systematic Analysis of Structural Equation Model Fit Indices

Sarah A. Rose
Wayne State University
Detroit, MI

Barry Markman
Wayne State University
Detroit, MI

Shlomo Sawilowsky
Wayne State University
Detroit, MI

The purpose of this study was to evaluate the sensitivity of selected fit index statistics in determining model fit in structural equation modeling (SEM). The results indicated a large dependency on correlation magnitude of the input correlation matrix, with mixed results when the correlation magnitudes were low and a primary indication of good model fit. This was due to the default SEM method of Maximum Likelihood that assumes unstandardized correlation values. However, this warning is not well-known, and is only obscurely mentioned in some textbooks. Many SEM computer software programs do not give appropriate error indications that the results are unsubstantiated when standardized correlation values are provided.

Keywords: Structural equation model, SEM, fit indices, RMSEA, SRMR, CFI, covariance matrices

Introduction

Wright (1918) presented the foundational theory of Structural Equation Modeling (SEM) for social and behavioral science research based on a path analysis used to model the bone size of rabbits. The novelty of the methodology was more generally accepted a half century later (Matsueda, 2011), coinciding with increasing use of computers, allowing for the more practical use of complicated matrix models. The development of more complicated analytical procedures was inevitable. Hoyle (1995) indicated, "with the increasing complexity and specificity of research questions in the social and behavioral sciences…has come increasing interest in SEM as a standard approach to testing research hypotheses" (p. 1).

Dr. Rose is an Adjunct Instructor of Educational Evaluation and Research. Email her at: ak1734@wayne.edu. Dr. Markman is a Professor of Educational Psychology and Educational Evaluation and Research. Email him at: barry.markman@wayne.edu. Dr. Sawilowsky is a Professor of Educational Evaluation and Research. Email him at: professorshlomo@gmail.com.

SEM is a powerful set of tools that can be used to explore data for the purpose of improving the understanding of the social, psychological, educational constructs and their interactions. It allows for a more complete and comprehensive analysis compared to other research methodologies, because it allows freedom in the evaluation of several model construct relationships simultaneously (Alavifar, Karimimalayer, & Anuar, 2012). The promise of this advantage should not be underestimated. The ability to take many variables and analyze them together using one test without the necessity for Bonferonni or similar corrections allows for considerable flexibility.

SEM models are developed by determining relationships between observed and/or latent variables to specify an initial model. The model is first analyzed to determine whether it is an appropriate approximation of the data construct. If the model is concluded to be an appropriate approximation, it is further analyzed to ascertain the magnitude and direction of relationships between the different variables.

As SEM was developed, it was designed primarily for the use of analysis of social and behavioral science data. Hence, the boundary conditions for performing SEM and determining model fit are steeped in the conditions typical of social and behavioral sciences, which includes multivariate normality (Gullen, 2000; Kline, 2011; Reinartz, Echambadi, & Chin, 2002; Tomarken & Waller, 2005). However, due to the capability of improving quality of life by analyzing data for complex research studies, SEM is increasingly being used in physical science research (e.g. Kelly, 2011; Ewing, Hamidi, Gallivan, Nelson, & Grace, 2014).

Problem Statement

The purpose of this study is to evaluate the sensitivity of selected fit index statistics in determining model fit. There are similarities between social and behavioral science and physical science data that make this transfer of methodologies apparently appropriate. Both data sets are parametric, can be assigned descriptive statistic values, can be formulated to provide frequency diagrams, and can be used with nonparametric tests. However, physical science data differ from the social behavioral science in several ways. In particular, physical science data typically have different distributions than that of social and behavioral science (e.g., Bradley, 1977, 1982; Ito, 1980; Micceri, 1989; Sawilowsky, Blair, & Micceri, 1990; Tan, 1982). Hence, the question arises: how well would SEM perform using non-normally distributed data commonly found in physical science data? However, an important preliminary step, the purpose of

this study, is to compare how various SEM fit indices work under standard normal conditions.

Model Fit

As the model is created, or specified, a foundational aspect of the SEM is to determine how well the model specified represents the data. It is imperative to specify the best model for the data to gain meaningful results. Model fit indices were developed to quantitatively and objectively assess the model fit. The matter of how to develop the fit statistics and which are the best to use has been a topic of great discussion. Kline (2011) indicated, "For at least 30 years the literature has carried an ongoing discussion about the best ways to test hypotheses and assess model fit" (p. 190).

There are dozens of fit indices measuring fit in a variety of ways. The plethora of indices presents two advantages: (1) They are useful for determining the performance of the model. SEM that is an improper fit to the data would provide inaccurate or erroneous results. (2) The complexity of variable matrices and sheer volume of analysis required point to a necessity for numerous fit index models. As the process is rigorous and complicated, so too the fit indices are difficult to simplify. It is therefore not surprising that currently no single fit index encompasses all the different indices in one comprehensive test (Gullen, 2000).

The complexity of analyzing the fit indices and the plethora of index tests from which to form a model fit assumption make it necessary to determine when models are truly a good fit to the data. Hooper, Coughlan, and Mullen (2008) indicated:

> Given the plethora of fit indices, it becomes a temptation to choose those fit indices that indicate the best fit…This should be avoided at all costs as it is essentially sweeping important information under the carpet. (p. 56)

Model fit indices have a short but rabid history. Initially, Chi-squared tests were used; however, the test was proved ineffectual due to the large sample sizes that are required for SEM analysis (Gullen, 2000). The Chi-squared test can be comparatively grossly underpowered for tiny data sets and fail to reach statistical significance. It can also be comparatively super-powered for huge data sets, reaching statistical significance in the presence of negligible differences (see, e.g., Kline, 2011, p. 201).

Various alternatives were therefore developed to supplement the model fit analysis (Bollen, 1989). Fit indices are classified into two categories: (1) Model Test Statistic, and (2) Approximate Fit Index (Kline, 2011).

Model Test Statistics and Chi-Squared

In the model test statistic, data are compared with a baseline model which is a covariance matrix of a sample from the data. If the covariance matrix of the overall data matches the covariance matrix of the sample, the model is considered a good fit. If the matrices differ, the discrepancies using the model need to be explained (Kline, 2011).

Model test statistics are typically developed as a "badness-of-fit" (Kline, 2011, p. 193) test. This means that failure to reject the null hypothesis indicates a good fit. Therefore, it is preferable for the resultant model test statistic to be as small as possible. The basic model test statistic is the model Chi-squared test. This test was developed by Karl Pearson (1900) and has withstood the test of time. It is probably the most well-known and accepted fit statistic. Its value lies in that it is nonparametric. The formula is (Neave & Worthington, 1988):

$$\chi^2 = \sum \frac{\left(\text{Observed} - \text{Expected}\right)^2}{\text{Expected}} \tag{1}$$

Therefore, the Chi-squared statistic is a percentage of the squared deviation from the expected over the expected score. A large Chi-squared statistic indicates a large deviation from the expected distribution. Indication of poor model fit occurs when the Chi-squared statistic value is greater than the critical value based on the nominal alpha.

Although the Chi-squared statistic in this context is apparently nonparametric, there are several factors that can adversely impact it such as large correlations among variables, unique variance, and large sample size (Kline, 2011). When observed variables are highly correlated, the Chi-squared value tends to increase. Unique variances among variables, being a product of score unreliability, result in a loss of statistical power. As the Chi-squared test is a badness-of-fit test, the loss of power reduces the probability of determining a poor model fit. As indicated above, the Chi-squared value tends to increase with sample size.

Approximate Fit Indices

The second type of fit statistic is the approximate fit index. The difference between approximate fit indices and model test statistics is that fit statistics are based on continuous measures. There is not a dichotomous conclusion to either reject or accept a null hypothesis. The value of the fit statistic, as it compares to an ideal value in magnitude, provides a representation of the fit. For example, the ideal value for CFI fit index is 1.0. A model resulting in a CFI of 0.90 would be a better fit than a model resulting in a CFI value of 0.85. As the null hypothesis is not rejected at a decided alpha value, the magnitude of the value has meaning. Therefore, these fit indices can be considered as "rules-of-thumb" as opposed to "golden rules" (Kline, 2011, p. 197).

Approximate fit indices do not "distinguish between what may be sampling error and what may be real covariance evidence against the model" (Kline, 2011, p. 195). Thus, they do not provide information in regards to specification error. These tests are typically goodness-of-fit tests, which mean the ideal index statistic occurs at a value of a specified magnitude (e.g., 1.0 as opposed to zero). The most common of the approximate fit indices are RMSEA, SRMR and the CFI.

Root Mean Square Error Approximation (RMSEA)

The RMSEA is a parsimony-adjusted index. It is not a measure of central tendency but follows a non-central Chi-squared distribution. It has a high and a low value that are provided by most SEM software. The RMSEA is a badness-of-fit test. Therefore, a good fit indicator occurs when the RMSEA low value is less than 0.05 and the high value is less than 0.10. (Kline, 2011).

As a parsimony-adjusted index, the RMSEA adjusts for parsimonious characteristics. It is obtained by dividing by degrees of freedom of the SEM model (Kline, 2011):

$$\text{RMSEA} = \sqrt{\frac{\chi_M^2 - df_M}{df_M (N-1)}} \tag{2}$$

where df_M = degrees of freedom of the SEM, N = sample size, and χ_M^2 = Chi-squared statistic value.

A small Chi-squared value indicates a good model fit. A model with a large degree of freedom, or a parsimonious model, results in a small RMSEA value. In other words, parsimonious models that have small deviations would indicate a

good model fit per this index. The equation is further divided by the sample size. Therefore, the parsimonious effect of the equation increases as sample size increases.

The limitations of RMSEA are obvious. It contains inherent prejudices towards models that have large sample sizes and large degrees of freedom. A model with a moderate-to-large variation from the expected values, but with a large sample size, could pass the RMSEA criteria for model fit.

Standardized Root Mean Square Residual (SRMR)

Although the name is similar to the RMSEA, the two indices are quite different (Iacobucci, 2009). The SRMR is a measure of the standardized value of the square root of the mean absolute covariance squared residual. A good fit value would be close to zero. Hu and Bentler (1999) opined a maximum allowable value for a good fit is approximately 0.09.

The formula, as given by Iacobucci (2009) and Schermelleh-Engel, Moosbrugger, & Muller (2003), is

$$\text{SRMR} = \sqrt{\frac{\sum_{i=1}^{p}\sum_{j=1}^{i}\left[\frac{\left(s_{ij}-\hat{\sigma}_{ij}\right)}{\left(s_{ii}s_{jj}\right)}\right]^2}{\frac{k\left(k+1\right)}{2}}} \tag{3}$$

where k = observed endogenous variables + observed exogenous variables, s_{ij}, s_{ii}, and s_{jj} = values from the covariance matrix, and $\hat{\sigma}_{ij}$ = value from the expected matrix covariance.

Comparative Fit Index (CFI)

The CFI is an incremental fit index and a parsimony-adjusted index, where the data set is compared to the Chi-squared values of a baseline model. It performs well even with small sample sizes. It is a goodness-of-fit test where a value of 1 indicates the best fit. The CFI was developed with the assumption that latent variables are not correlated (Hooper, Coughlan, & Mullen, 2008). Therefore, models with highly correlated latent variables can result in an inaccurate assessment of model fit.

The CFI is a function of the Chi-squared value and degrees of freedom of the model. The formula is (Kline, 2011):

$$CFI = 1 - \frac{\chi_M^2 - df_M}{\chi_B^2 - df_B} \qquad (4)$$

where df_X = degrees of freedom of the SEM/Baseline models, χ_X^2 = Chi-squared statistic value for the SEM/Baseline models, M = SEM model, and B = baseline model. This equation results in higher values for models with larger degrees of freedom, resulting in a more favorable fit statistic. Hu and Bentler (1999) opined a minimum CFI of 0.95 is necessary to indicate an acceptable fit.

Model Fit Indices Overview

Although multivariate normality is a baseline assumption of the model fit indices (Kline, 2011; Schermelleh-Engel et al., 2003), the formulas for calculating the model fit statistics are apparently nonparametric. It would therefore be reasonable to assume that the model fit index equations could be used to assess model fit for any distribution. However, the robustness of the formulas have not yet been assessed, and the capability of the indices to measure model fit for physical science data is of great interest.

Methodology

Monte Carlo simulation theory requires that baseline theories be tested prior to performance of Monte Carlo simulations on the problem statement. Therefore, it is required to verify model fit indices when normality is not violated as a prerequisite to any study on models that violate underlying assumptions.

Monte Carlo simulations using correlation matrices of randomly selected values of an incrementally increasing correlation range was conducted. The correlation matrices were of randomly selected values, of no model, and no relationship. Model fit indices should indicate a poor model fit for all simulations, meaning they should not exceed the Type-I error rate dictated by nominal α. Therefore, assessment of legitimacy of the model fit index results was based on the percentage of times the results indicated a poor model fit.

At first a Monte Carlo was performed using RStudio based on four variables and 10,000 repetitions of varying correlation matrices of randomly selected numbers between negative and positive 0.1. The results from this simulation

series were mixed in terms of model fit, indicating meaningless results. It was therefore a matter of interest to determine the minimum allowable correlation values under which the model fit indices would provide legitimate and meaningful results.

Monte Carlo simulations based on four variables and correlation matrices of randomly selected values of incrementally-increasing correlation ranges were performed. Each Monte Carlo simulation contained 1,000 repetitions and was performed for sample size of $n = 50$, 100, 150, 200, 300, and 500. The correlation range was a base value ± 0.015. Base values were incrementally increased at every hundredths place, beginning from 0.04, 0.05,…, 0.26, and 0.27. As such, 24 Monte Carlo Simulations were performed for six different sample sizes.

Results

Minimum Correlation Coefficient for SEM

The first Monte Carlo simulation included a correlation matrix of random values from a range of 0.04 ± 0.015. All model fit indices results included in the analyses (Chi-squared, RMSEA Lower, RMSEA Upper, SRMR, and CFI) were an indication of a poor model fit 0% of the time. Refer to Table 1 below.

As the correlation matrix values were increased in magnitude, the results of the model fit indices became meaningless. The percentages of greater than and less than critical values did not result in percentage numbers that added to 100%. The fit index results ceased to be meaningless as the correlation magnitudes were continuously increased, and instead were an indication of a poor model fit with increasing reliability. At a certain correlation magnitude (e.g. when correlation was equal to 0.08 ± 0.015 as in Table 2), the results of the model fit indices were an indication of a poor model fit for the conditions studied for all Monte Carlo repetitions. A summary of these results (select simulations) is provided in Table 3.

Each model fit index resulted in legitimate results at different correlation magnitudes. The best model fit index, which resulted in legitimate model fit estimation at the lowest correlation magnitude, was RMSEA Upper at a correlation of 0.08 for all sample sizes. The next best model fit index was CFI, with valid estimation of model fit at a minimum correlation value of 0.16. The next best model fit index was SRMR, with valid model fit estimation at a minimum correlation value of 0.17 for large sample sizes and 0.18 for sample size of 50. The next best model fit index following SRMR was Chi-squared, with valid model fit estimation at a minimum correlation value of 0.24. The model fit index

that performed the poorest was RMSEA Lower, with valid model fit estimation at a minimum correlation of 0.27. Refer to Table 4 below.

Table 1. Monte Carlo simulation percentage of model fit indices (indication of poor model fit); correlation matrix magnitudes range of 0.04 ± 0.015

Model Fit Index	Sample Size					
	50	100	150	200	300	500
Chi-squared	0%	0%	0%	0%	0%	0%
RMSEA Lower	0%	0%	0%	0%	0%	0%
RMSEA Upper	0%	0%	0%	0%	0%	0%
SRMR	0%	0%	0%	0%	0%	0%
CFI	0%	0%	0%	0%	0%	0%

Table 2. Monte Carlo simulation percentage of model fit indices (indication of poor model fit); correlation matrix magnitudes range of 0.08 ± 0.015

Model Fit Index	Sample Size					
	50	100	150	200	300	500
Chi-squared	0%	0%	0%	0%	N/A	N/A
RMSEA Lower	0%	0%	0%	0%	0%	0%
RMSEA Upper	100%	100%	100%	100%	100%	100%
SRMR	0%	0%	0%	0%	0%	0%
CFI	0%	0%	N/A	N/A	N/A	N/A

Table 3. Monte Carlo simulation percentage of model fit indices (indication of poor model fit); correlation matrix magnitudes range ± 0.015

Model Fit Index	Correlation	Sample Size					
		50	100	150	200	300	500
Chi-squared	0.04	0%	0%	0%	0%	0%	0%
RMSEA Lower		0%	0%	0%	0%	0%	0%
RMSEA Upper		0%	0%	0%	0%	0%	0%
SRMR		0%	0%	0%	0%	0%	0%
CFI		0%	0%	0%	0%	0%	0%
Chi-squared	0.06	0%	0%	0%	0%	0%	N/A
RMSEA Lower		0%	0%	0%	0%	0%	0%
RMSEA Upper		35%	N/A	N/A	N/A	N/A	N/A
SRMR		0%	0%	0%	0%	0%	0%
CFI		0%	0%	0%	N/A	N/A	N/A
Chi-squared	0.08	0%	0%	0%	0%	N/A	N/A
RMSEA Lower		0%	0%	0%	0%	0%	0%
RMSEA Upper		100%	100%	100%	100%	100%	100%
SRMR		0%	0%	0%	0%	0%	0%
CFI		0%	0%	N/A	N/A	N/A	N/A

Table 3, continued.

Model Fit Index	Correlation	Sample Size					
		50	100	150	200	300	500
Chi-squared	0.16	0%	N/A	N/A	N/A	N/A	N/A
RMSEA Lower		0%	0%	N/A	N/A	N/A	N/A
SRMR		54%	N/A	N/A	N/A	N/A	N/A
CFI		100%	100%	100%	100%	100%	100%
Chi-squared	0.18	0%	N/A	N/A	N/A	N/A	N/A
RMSEA Lower		0%	N/A	N/A	N/A	N/A	N/A
SRMR		100%	100%	100%	100%	100%	100%
Chi-squared	0.24	100%	100%	100%	100%	100%	100%
RMSEA Lower		0%	N/A	N/A	N/A	N/A	N/A
Chi-squared	0.27	100%	100%	100%	100%	100%	100%
RMSEA Lower		100%	100%	100%	100%	100%	100%
RMSEA Upper		100%	100%	100%	100%	100%	100%
SRMR		100%	100%	100%	100%	100%	100%
CFI		100%	100%	100%	100%	100%	100%

Table 4. Minimum correlation values for valid model fit index measurement

Model Fit Index	Sample Size					
	50	100	150	200	300	500
Chi-squared	0.24	0.24	0.24	0.24	0.24	0.24
RMSEA Lower	0.27	0.27	0.27	0.27	0.27	0.27
RMSEA Upper	0.08	0.08	0.08	0.08	0.08	0.08
SRMR	0.18	0.17	0.17	0.17	0.17	0.17
CFI	0.16	0.16	0.16	0.16	0.16	0.16

Conclusion

Originally, a Monte Carlo simulation with randomly selected correlation values between - 0.1 and + 0.1 was performed. The results were meaningless, with mixed results in terms of fit. The output of the latest repetition of the Monte Carlo simulation was extracted and compared with the output from Amos Graphics to ensure that a programming error did not occur. The results were the same within rounding error.

Fit index results should be consistent regardless of whether or not a meaningful model is produced. Examination of the model fit results should indicate a good or a poor model fit when a reasonable model is assessed. However, examination of the results should never indicate a good model fit on a poorly-defined model. In this case, the correlation values between variables were small and the paths were not significant. Therefore the model, having no relationships,

should result in an indication of poor model fit when assessed using model fit index tests. This indication of poor model fit should occur uniformly for all model fit index tests and for all sample sizes, or at least within the Type I error rate set by nominal alpha.

These findings were discussed with colleagues. One believed that, with caution (presumably ignoring fit results in the absence of a good model), there were some insights that could be garnered based on the results. This viewpoint was amplified by another colleague, who replicated the results via Mplus, and hence urged extreme caution, because of SEMs ability to produce a well-fitted model that is nevertheless bereft of significant covariances.

As a beginning to approaching the model fit assessment with caution, additional research was conducted to determine what SEM conditions caused the model fit index results to be meaningless. The Monte Carlo simulation models were assessed to discover common characteristics. A consistent attribute was the low correlation values between the variables. It appeared when the correlation values between variables were low, the results of the model fit indices were meaningless. Additional research was therefore conducted to determine what constituted a low correlation, and whether there was a minimum allowable correlation value between variables that is a prerequisite for a SEM to be meaningful.

Additional Monte Carlo simulations were conducted, with 1,000 repetitions and varying magnitudes of correlation matrices. The magnitudes of the correlation values were randomly selected from a base value ± 0.015. Twenty-four Monte Carlo simulations were performed, with the base value increasing from 0.04 to 0.27 at every hundredths place value (i.e. 0.04, 0.05 ,0.06, etc.). The model fit indices would be legitimized by the percentage of times a poor model fit was indicated, as the variables had no relationship and correlation values were randomly selected.

As the correlation matrix values were increased in magnitude, the results of the model fit indices became first illogical and then finally logical with an indication of a poor model fit occurring with increasing reliability. At a certain correlation magnitude range (e.g. when correlation was equal to 0.08 ± 0.015 as in Table 2), the results of the model fit indices were an indication of a poor model fit for all sample sizes studied for all Monte Carlo repetitions.

Table 5. Minimum correlation values

Rank	Model Fit Index	Minimum Correlation Value
1	RMSEA Upper	0.08
2	CFI	0.16
3	SRMR	0.18
4	Chi-squared	0.24
5	RMSEA Lower	0.27

Table 6. Correlation matrix

Variables	z	X_1	X_2	X_3
z	1.000	0.104	0.098	0.115
X_1	0.104	1.000	0.100	0.088
X_2	0.098	0.100	1.000	0.109
X_3	0.115	0.088	0.109	1.000

Each model fit index resulted in legitimate results at different correlation magnitudes; refer to Table 3 above. Model fit indices can be ranked from best to worst based on the minimum correlation values required before legitimate results were acquired. The model fit indices, from best to worst, are listed in Table 5 above with their respective minimum correlation values and ranks.

The results from the last repetition of the Monte Carlo simulation with correlation range of 0.1 ± 0.015 and sample size of 500 were extracted (refer to Table 6 above and the Lavaan output below) to better understand the results of the Monte Carlo simulations and to verify the conclusions determined above. The results of the model fit index tests were mixed. The p-value for the Chi-squared test was 0.003, an indication of a poor model fit. The RMSEA Upper value was 0.133, an indication of a poor model fit. The RMSEA Lower value was 0.044, an indication of a good model fit. The CFI value was 0.505, an indication of a poor model fit. The SRMR value was 0.055, an indication of a good model fit.

The regression coefficients for the exogenous variables were 0.088 for X_1, 0.079 for X_2, and 0.098 for X_3. Although these values were low, the coefficients for X_1 and X_3 were statistically significant. This is illogical, as the correlation magnitudes in the correlation matrix were all low. Statistically significant paths between variables are therefore a contradictory conclusion. These results solidified the conclusion above that a SEM with a correlation matrix of low values would result in illogical outcomes.

Lavaan Output for Sample Size of 500 and Four Variables, Repetitions = 1,000

```
Number of observations            500
Estimator                         ML
Minimum Function Test Statistic   14.059
Degrees of freedom                3
P-value (Chi-square)              0.003

User model versus baseline model:
 Comparative Fit Index (CFI)      0.505
 Tucker-Lewis Index (TLI)         0.010
 Number of free parameters        7
 RMSEA                            0.086
rmsea.ci.lower                    0.044
rmsea.ci.upper                    0.133
 90 Percent Confidence Interval  0.044 0.133
 P-value RMSEA <= 0.05            0.075
 SRMR                            0.055

Parameter estimates:
 Information           Expected
 Standard Errors       Standard

Regressions:
z ~    Estimate     Std.err     Z-value     P(>|z|)
x1     0.088        0.044       1.990       0.047
x2     0.079        0.044       1.786       0.074
x3     0.098        0.044       2.232       0.026

Covariances:
 x1 ~~x2      0.000
 x3           0.000
 x2 ~~x3      0.000

Variances:
  z      0.970 0.061
 x1      0.998 0.063
 x2      0.998 0.063
 x3      0.998 0.063
```

SEM is a collection of procedures that are assessed based on a plethora of fit or lack of fit statistics that could be subjectively chosen or ignored to support or eliminate a model. Dozens of caveats (such as those listed in Kline, 2011, e.g., at its core it relates to non-experimental data and hence there can never be causation (p. 8), a poor model can be preserved by modifying the hypotheses on which it is based (p. 8), alternative models may not be ruled out (p. 8), it is a large sample technique (p. 11), it eschews hypothesis testing and hence is veiled behind subjectivity (p. 13), the statistical significance of estimated parameters are dependent on the algorithm adopted (p. 13), a maximum likelihood estimate cannot tolerate even a single missing datum (p. 48), a nonpositive definite matrix cannot be analyzed (p. 49), ill-scaled covariance matrices cannot be handled (p. 67)) severely limit SEM outside of textbook examples.

Moreover, Kline (2011) noted,

> It may be problematic to submit for analysis just a correlation matrix without standard deviations or specify that all standard deviations are 1.0, which standardizes everything. This is because the default method of ML estimation (and most other methods, too) assumes that the variables are unstandardized. This means that if a correlation matrix without standard deviations is analyzed, the results may not be correct…Some SEM computer programs give warning message or terminate the run if the researcher requests the analysis of a correlation matrix only with standard ML estimation. By the same token, it would also be problematic to convert raw scores to z scores and then submit for analysis the data file of standardized scores. (p. 49)

These cautions from Kline (2011) appear to explain why a systematic Monte Carlo study conducted by inputting an incrementally increasing correlation matrices, such as was attempted in this study, cannot be successful. The standard procedure of starting the study with a null zero order correlation matrix to show the relevant fit indices reject, or fail to reject as appropriate to the index, is not possible, precluding a presentation of the power spectrum of the competitors based on systematically increasing (or decreasing based on the type of fit index) the matrix. The restrictions indicated by Kline (2011) were mentioned in an obscure section of the textbook, and were omitted by most other textbook authors. Thus, this limitation and the egregious results from the non-compliance are not well-publicized.

It appears it is necessary to start with a good model in order for the model fit indices to provide a proper assessment. This is circuitous, for how can a good model be assessed if the baseline condition for meaningful results is a good model? Analysts must consider this paradox, and decide if SEM outside of textbook examples is truly meaningful.

References

Alavifar, A., Karimimalayer, M., & Anuar, M. K. (2012). Structural equation modeling VS multiple regression. *Engineering Science and Technology: An International Journal, 2*(2), 326-329. Retrieved from http://www.estij.org/papers/vol2no22012/25vol2no2.pdf

Bollen, K. A. (1989). A new incremental fit Index for general structural equation models. *Sociological Methods & Research, 17*(3), 303-316. doi: 10.1177/0049124189017003004

Bradley, J. V. (1977). A common situation conducive to bizarre distribution shapes. *The American Statistician, 31*(4), 147-150. doi: 10.2307/2683535

Bradley, J. V. (1982). The insidious L-shaped distribution. *Bulletin of the Psychometric Society, 20*(2), 85-88. doi: 10.3758/BF03330089

Ewing, R., Hamidi, S., Gallivan, F., Nelson, A. C., & Grace, J. B. (2014). Structural equation models of VMT growth in US urbanised areas. *Urban Studies, 51*(14), 3079-3096. doi: 10.1177/0042098013516521

Gullen, J. A. (2000). *Goodness of fit as a single factor structural equation model* (Unpublished doctoral dissertation). Wayne State University, Detroit, MI.

Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods, 6*(1), 53-60. Retrieved from http://www.ejbrm.com/volume6/issue1

Hoyle, R. H. (Ed.). (1995). *Structural equation modeling: Concepts, issues, and applications*. Thousand Oaks, CA: Sage Publications.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. doi: 10.1080/10705519909540118

Iacobucci, D. (2009). Structural equations modeling: Fit indices, sample size, and advanced topics. *Journal of Consumer Psychology, 20*(1), 90-98. doi: 10.1016/j.jcps.2009.09.003

Ito, P. K. (1980). Robustness of ANOVA and MANOVA test procedures. In P. R. Krishnaiah (Ed.), *Handbook of Statistics* (Vol. 1, pp. 199-236). Amsterdam, Netherlands: North-Holland. doi: 10.1016/S0169-7161(80)01009-7

Kelly, S. (2011). Do homes that are more energy efficient consume less energy?: A structural equation model of the English residential sector. *Energy, 36*(9), 5610-5620. doi: 10.1016/j.energy.2011.07.009

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: The Guilford Press.

Matsueda, R. L. (2011). *Key advances in the history of structural equation modeling* (Working paper no. 114). Seattle, WA: University of Washington Center for Statistics and the Social Sciences. Retrieved from https://www.csss.washington.edu/Papers/2012/wp114.pdf

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105*(1), 156-166. doi: 10.1037/0033-2909.105.1.156

Neave, H. R., & Worthington, P. L. (1988). *Distribution free tests*. Boston, MA: Unwin Hyman Inc.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series 5, 50*(302), 157-175. doi: 10.1080/14786440009463897

Reinartz, W. J., Echambadi, R., & Chin, W. W. (2002). Generating non-normal data for simulation of structural equation models using Mattson's method. *Multivariate Behavioral Research, 37*(2), 227-244. doi: 10.1207/S15327906MBR3702_03

Sawilowsky, S. S., Blair, R. C., & Micceri, T. (1990). REALPOPS.LIB: a PC Fortran library of eight real distributions in psychology and education. *Psychometrika, 55*(4), 729.

Schermelleh-Engel, K., Moosbrugger, H., & Muller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research, 8*(2), 23-74. Retrieved from https://www.dgps.de/fachgruppen/methoden/mpr-online/issue20/art2/mpr130_13.pdf

Tan, W. Y. (1982). Sampling distributions and robustness of t, F, and variance-ratio in two samples and ANOVA models with respect to departure from normality. *Communications in Statistics – Theory and Methods, 11*(21), 2485-2511.

Tomarken, A. J., & Waller, N. G. (2005). Structural equation modeling: Strengths, limitations, and misconceptions. *Annual Review of Clinical Psychology, 1*(1), 31-65. doi: 10.1146/annurev.clinpsy.1.102803.144239

Wright, S. (1918). On the nature of size factors. *Genetics, 3*(4), 367-374. Retrieved from http://www.genetics.org/content/3/4/367