Published in final edited form as:

Methodology (Gott). 2008 January 1; 4(4): 159–167. doi:10.1027/1614-2241.4.4.159.

Model error in covariance structure models: Some implications for power and Type I error

Donna L. Coffman

The Pennsylvania State University

Abstract

The present study investigated the degree to which violation of the parameter drift assumption affects the Type I error rate for the test of close fit and power analysis procedures proposed by MacCallum, Browne, and Sugawara (1996) for both the test of close fit and the test of exact fit. The parameter drift assumption states that as sample size increases both sampling error and model error (i.e. the degree to which the model is an approximation in the population) decrease. Model error was introduced using a procedure proposed by Cudeck and Browne (1992). The empirical power for both the test of close fit, in which the null hypothesis specifies that the Root Mean Square Error of Approximation (RMSEA) $\leq .05$, and the test of exact fit, in which the null hypothesis specifies that RMSEA = 0, is compared with the theoretical power computed using the MacCallum et al. (1996) procedure. The empirical power and theoretical power for both the test of close fit and the test of exact fit are nearly identical under violations of the assumption. The results also indicated that the test of close fit maintains the nominal Type I error rate under violations of the assumption.

When fitting a covariance structure model (CSM), the likelihood ratio test statistic (T) is asymptotically distributed as central chi-square if the model is correct in the population. However, all mathematical models are approximations, and in practice researchers wish to evaluate how well a model approximates the data. Mathematical models remain approximations even if they are fit in the population. Therefore, there is a degree of discrepancy between the ideal model which fits exactly in the population (the generating model) and the approximating model that researchers use to study observable phenomena (Cudeck & Henly, 1991, 2003; MacCallum, 2003; MacCallum & Tucker, 1991). When the mathematical model is not assumed to hold exactly in the population, the test statistic asymptotically follows a noncentral chi-square distribution as long as the degree of discrepancy in the population (F_0) is not too large (Steiger, Shapiro, & Browne, 1985). In order to derive the asymptotic distribution of the test statistic when models are approximations, it is necessary to invoke a parameter drift assumption. This assumption specifies that model error is proportional to sampling error and that as the sample size (N) increases, both sampling error and model error decrease. However, model error is not random but rather is a fixed constant. Although it is reasonable to assume that sampling error decreases as N increases, it is not reasonable to assume that as N increases, the approximating model becomes closer to the generating model. In other words, the parameter drift assumption never holds.

Address correspondence to: Donna L. Coffman, The Methodology Center, The Pennsylvania State University, 204 E. Calder Way, Ste. 400, State College, PA 16801, dlc30@psu.edu, phone: 814-863-9724, fax: 814-863-0000.

R programs for computing power and minimum sample size for a given level of power based on MacCallum et al. (1996) are available at www.personal.psu.edu/dlc30/research.htm. Web interactive versions (Preacher & Coffman, 2006) of these R programs are available at www.quantpsy.org

In practice, the implications of violation of the parameter drift assumption are that the asymptotic distribution of the test statistic is unknown, and therefore, there are consequences for any model evaluation procedure that makes use of the noncentral chi-square distribution of the test statistic under imperfect fit. In particular, violation of the parameter drift assumption has implications for the test of close fit (Browne & Cudeck, 1993). If the parameter drift assumption is violated, significance levels (e.g. p values) for this test may be invalid, leading to incorrect conclusions about model fit. In addition, violation of the parameter drift assumption has implications for power analysis procedures, because all such procedures rely on the test statistic following a noncentral distribution under the alternative hypothesis (see Lehmann, 1999). If the parameter drift assumption is violated, the distribution under the alternative hypothesis, H_1 , cannot be accurately specified, and results of the power analysis procedures may be invalid. Thus, the Satorra and Saris (1985) procedure for power analysis and the MacCallum et al. (1996) procedure for power analysis of the test of exact fit and the test of close fit may lead to incorrect power estimates and decision errors.

To summarize, when the model does not fit exactly in the population, a parameter drift assumption is invoked so that the asymptotic distribution of the test statistic may be derived. The asymptotic distribution is noncentral chi-square as long as the degree of discrepancy in the population is not too large (i.e. the parameter drift assumption holds). Because the parameter drift assumption is routinely violated in practice, it is of interest to know the severity of the consequences and the robustness of statistical methods and results in the face of violations. The purpose of this article is to examine effects of violation of the parameter drift assumption on the Type I error rates of the test of close fit and power analysis procedures for the test of close fit and the test of exact fit.

The Noncentral Chi-Square Distribution

As previously stated, the test statistic under Maximum Likelihood (ML) estimation is asymptotically distributed as central chi-square if the null hypothesis of exact fit is true (Steiger et al., 1985). If the null hypothesis is false, the test statistic under ML is asymptotically distributed as a noncentral chi-square with d degrees of freedom and noncentrality parameter, λ .

The noncentrality parameter is asymptotically related to the test statistic by

$$E(F_{ML}) = \frac{\lambda}{n} + \frac{d}{n} \quad (1)$$

where $n = N - 1$ and F_{ML} is the minimum of the ML discrepancy function, or

$$E(T) = nE(F_{ML}) = \lambda + d, \quad (2)$$

and thus,

$$\lambda = E(T) - d. \quad (3)$$

Therefore, λ is a function of n , d , and F_{ML} .

RMSEA

The population value of the Root Mean Square Error of Approximation (RMSEA), ε , is defined as

$$\varepsilon = \sqrt{F_0/d} \quad (4)$$

and is a measure of model discrepancy per degree of freedom in the population. To obtain a point estimate of ε , an estimate of the population discrepancy function value, F_0 , is needed. The sample discrepancy function value, $\hat{F}_{ML}$, is a biased estimator of F_0 . Browne and Cudeck (1993) presented a less biased estimator of F_0 ,

$$\hat{F}_0 = \max \left\{ \hat{F}_{ML} - d/n, 0 \right\}. \quad (5)$$

Thus a less biased estimate of the RMSEA can be obtained by substituting the estimate of F_0 from Equation (5) into Equation (4), resulting in

$$\hat{\varepsilon} = \sqrt{\hat{F}_0/d} = \sqrt{\max \left\{ \frac{\hat{F}_{ML}}{d} - \frac{1}{n}, 0 \right\}}. \quad (6)$$

Given Equation (1), ε may be written in terms of λ ,

$$\varepsilon = \sqrt{\frac{\lambda}{nd}}, \quad (7)$$

and thus $\hat{\varepsilon}$ may be written in terms of the point estimate of the noncentrality parameter, $\hat{\lambda}$, resulting in

$$\hat{\varepsilon} = \sqrt{\max \left\{ \frac{\hat{\lambda}}{nd}, 0 \right\}}. \quad (8)$$

If the test statistic does not follow a noncentral chi-square distribution, then an estimate of the noncentrality parameter no longer has the same meaning. Because the point estimate of the RMSEA involves an estimate of the noncentrality parameter (see Equation (8)), the point estimate of the RMSEA may be inaccurate if the test statistic does not follow a noncentral chi-square distribution.

Test of Close Fit

The noncentrality parameter can be defined in terms of ε by algebraically manipulating Equation (7),

$$\lambda = nd\varepsilon^2. \quad (9)$$

Given the relationship between λ and ε , the conventional chi-square test for exact fit used to test overall model fit in CSM can be reframed as a test of $H_0: \varepsilon_0 = 0$. In this case, $F_0 = 0$, and the test statistic follows a central chi-square distribution. However, a more useful test of close fit can be specified as a test of $H_0: \varepsilon_0 \leq .05$ (Browne & Cudeck, 1993). Under this null hypothesis, the test statistic follows a noncentral chi-square distribution (if the parameter drift assumption holds) with noncentrality parameter, $\lambda = nd(.05)^2$. In this case, we do not assume that $F_0 = 0$. The null hypothesis of $\varepsilon_0 \leq .05$ is typically chosen for the test of close fit because it has been suggested that RMSEA values less than .05 indicate close fit, values in the range of .08 to .10 indicate mediocre fit, and values greater than .10 indicate poor fit, although these values are intended as guidelines for interpretation and not as absolute thresholds (Browne & Cudeck, 1993).

Power Analysis Procedures for Tests of Model Fit

Given the distribution of T under H_0 and H_1 , an estimate of power can be obtained. As with any power analysis procedure, to define the distribution under H_1 , a value for λ must be specified. There are several different procedures that have been proposed for specifying λ when conducting power analysis in CSM. These procedures include those proposed by MacCallum et al. (1996), MacCallum and Hong (1997), and Satorra and Saris (1985) (see also Saris & Satorra, 1993).

MacCallum et al. (1996) developed a procedure for determining the power of overall tests of model fit in CSM. The development of this capability follows from work by Steiger and Lind (1980) and Browne and Cudeck (1993). MacCallum et al. (1996) defined effect size in terms of a pair of RMSEA values, ε_0 and ε_1 , from which one can define two overlapping chi-square distributions. For example, if the null hypothesis is that of exact fit, $H_0: \varepsilon_0 = 0$, and the alternative hypothesis is that model fit is mediocre, $H_1: \varepsilon_1 = .08$, then the test statistic under the null hypothesis follows a central chi-square distribution and the test statistic under the alternative hypothesis follows a noncentral chi-square distribution (if the parameter drift assumption holds) with noncentrality parameter $\lambda_1 = nd(.08)^2$. Similarly, if the null hypothesis is that of close fit, $H_0: \varepsilon_0 \leq .05$ and the alternative hypothesis is that of mediocre fit, $H_1: \varepsilon_1 = .08$, then the test statistic under the null hypothesis follows a noncentral chi-square distribution (if the parameter drift assumption holds) with noncentrality parameter $\lambda_0 = nd(.05)^2$ and the test statistic under the alternative hypothesis follows a noncentral chi-square distribution (if the parameter drift assumption holds) with noncentrality parameter $\lambda_1 = nd(.08)^2$. In general, one could determine power using any two values of the RMSEA. For example, $\varepsilon_0 \leq .05$ and $\varepsilon_1 = .10$ may also be reasonable values. In this case, one could determine the power for a test of the null hypothesis of close fit when the alternative hypothesis is that model fit is poor. In general, given the distribution of the test statistic under specified null and alternative hypotheses, power is then computed by standard methods.

The MacCallum et al. (1996) procedure, as described above, assumes that the test statistics under the alternative hypotheses follow a noncentral chi-square distribution. For the test of close fit, the procedure also assumes the test statistic under the null hypothesis follows a noncentral chi-square distribution. One then finds the critical value for the noncentral chi-square distribution under the null hypothesis for a given alpha level and computes the area above (to the right of) that value under the alternative noncentral chi-square distribution. Because the procedure relies on the assumption that the test statistics follow noncentral chi-square distributions when the model does not fit exactly in the population, it also relies on

the parameter drift assumption. If the parameter drift assumption is violated, then the test statistics under the null and alternative hypotheses do not necessarily follow a noncentral chi-square distribution. It follows that the critical value used by this procedure may be incorrect. If critical values are incorrect, then Type I error rates would deviate from nominal rates and empirical power would not match theoretical power.

Purpose of the Current Study

The primary purpose of the current study is to investigate the Type I error for the test of close fit and to evaluate the power properties for the test of close fit and the test of exact fit. This purpose will be accomplished through a simulation study in which the degree of model error in the population is specified and thus known. Second, the manner in which the alternative model is specified to evaluate power properties in the present study differs from the usual method of introducing model error. The method by which model error is introduced in the present study is described next.

Method

Introducing Model Error

Typically, in simulation studies error is introduced into models using specification errors. Specification errors may include omitted variables, incorrect functional form such as nonlinear relationships, and incorrectly specified structural or measurement relations. For example, in factor analysis models, model error may result from a large number of minor factors that are not of theoretical interest, or individual differences (MacCallum, 2003; MacCallum & Tucker, 1991). These errors occur because the approximating model differs from the generating model. However, most studies do not incorporate all of these various types of specification errors at once. Further, if one introduces nonlinear relationships, one must choose which specific relationships will be nonlinear, the form of the nonlinear relationships, and their size. If structural or measurement relations are specified incorrectly, one must choose which specific relationships will be incorrect, as well as their parameter values. However, in any empirical application, we do not know in what manner the model is incorrect. That is, we do not know which or how many relationships have been omitted or are nonlinear.

It is highly unlikely that a particular model is wrong because one or two specific relationships are missing, or because a particular relationship is quadratic. Rather, models are incorrect (to varying degrees) due to a variety of reasons. For example, one of the relationships is nonlinear, there are several relationships missing, and so on. The method that will be used to introduce model error in the present study is not specific; it potentially includes all types of possible errors. Further it is global model error rather than particular to the locale of the omitted relationship.

Generating Covariance Matrices with a Specified Minimum Discrepancy Function Value in the Population

Cudeck and Browne (1992) developed a procedure for constructing a covariance matrix with a specified minimum discrepancy function value in the population. One advantage of this method is that there is no need to designate the specific nature of the model error. The

method finds a positive definite covariance matrix, $\sum_0^*$ for p measured variables such that

$$\sum_0^* = \sum (\theta_0) + \mathbf{E} \quad (10)$$

where the $p \times p$ symmetric matrix $\mathbf{E}$ is chosen such that a class of discrepancy functions, $F(\sum_0^*, \sum(\theta_0))$, which includes ML and Ordinary Least Squares (OLS), has the specified parameter vector, θ_0 , as a minimizer and specified minimum, δ_0 . For the current study, $\delta_0 = F_0$. Essentially, the procedure treats CSMs as nonlinear regression models (Browne, 1982).

Because $\mathbf{S}$ is an unbiased estimator of $\sum_0^*$ in Equation (10), the expected value of

$$\mathbf{E} = \mathbf{S} - \sum(\theta_0) \quad (11)$$

is a null matrix.

Model Types

Because power is dependent on the type of model as well as degrees of freedom (Saris, Satorra, & Sörbom, 1987), I specified three model types; a restricted 2-factor model, a structural model, and a path analysis model. The parameter values were chosen based on actual empirical data rather than choosing random or arbitrary values for the parameters. Even though it is not possible to include every type of model configuration in the simulation study, it is hoped that by choosing three model types and parameters based on realistic and plausible values, the simulation results will be as generalizable as possible.

Model A: Confirmatory Factor Analysis (CFA) Model—A 2-factor model was specified as shown in Figure 1. This model was taken from Maydeu-Olivares and Coffman (2006) and their parameter estimates are used as parameter values here. For this model, $d = 13$.

Model B: Structural Equation Model (SEM)—The structural model shown in Figure 2 was specified as the population model with the parameter values given on the paths. This model was taken from Maydeu-Olivares and Lado (2003) and their parameter estimates are used as parameter values here. For this model, $d = 12$, which is similar to the d for the CFA model. Thus, the primary difference between the CFA model and the SEM model is that the latter includes structural parameters.

Model C: Path Analysis Model—The final model examined here was a path analysis model. This model is shown in Figure 3 and the parameter values are given on the paths. For this model, $d = 2$. Thus, the path model has fewer d than the SEM and CFA models.

For each of the three models, I computed Σ_0 and then added error using the Cudeck and Browne (1992) method to produce Σ with a specified lack of fit in the population. For each model type, the specified lack of fit in the population was varied by specifying the population discrepancy function value such that ε was 0.025, 0.05, 0.075, 0.10, and 0.15. These values were chosen to represent a range of lack of fit. A value of 0.025 represents a model that fits very well, whereas a value of 0.10 represents a model that fits poorly. A value of 0.15 represents a model that certainly would not be interpreted in practice and was thus chosen to represent an extreme.

The population covariance matrices with a given discrepancy function value were obtained using a FORTRAN 77 program (run on a Fortran 90 compiler) written by Robert Cudeck and entitled Covariance Matrices with Specified Structure (COMSS). This program was modified so that a given degree of model error in the population may be specified for models that include a structural component. Before proceeding with the simulation I fit each

model to the appropriate population covariance matrix $\mathbf{I}$ obtained from COMSS as a check that the discrepancy function equaled the value $\mathbf{I}$ specified and that the parameter values were equal to those $\mathbf{I}$ specified.

Sampling and Replications

Using 5 different values for N (200, 300, 500, 1000, and 5000), I drew $r = 1000$ random sample covariance matrices from each $\mathbf{\Sigma}$ for each value of N and each value of ε and fit the appropriate model to the sample covariance matrices. The sample covariance matrices were generated randomly and fit to the model using Matlab. All models were fit using ML estimation.

The sample covariance matrices were randomly drawn from a Wishart distribution with parameters p and $\mathbf{\Sigma}$ using a procedure described by Monahan (2001). First, a Cholesky decomposition of the population covariance matrix was performed. Second, $p(p - 1)/2$ random normal variates were generated and loaded into the upper triangle of a symmetric matrix, $\mathbf{A}$. Third, p random chi-square variates with the appropriate d were generated and the square root of these were loaded into the diagonal of $\mathbf{A}$. The random sample covariance matrix is then given as

$$\mathbf{S} = N^{-1}[\mathbf{T}'(\mathbf{A}'\mathbf{A})\mathbf{T}] \quad (12)$$

where $\mathbf{T}$ is the upper triangular matrix obtained from the Cholesky decomposition.

Summary

In summary, for each of three models and five levels of misfit, I constructed a population covariance matrix with a specified discrepancy function value in the population. That is, for each model type I generated $\mathbf{\Sigma}$ with a known amount of misspecification in the population. Random sample covariance matrices were then drawn from each of these population covariance matrices for various sample sizes. I examined three model types, five levels of ε , and five levels of N , resulting in 75 conditions. For each condition, there were 1000 replications.

Results

The primary outcome examined is the power analysis estimates for the test of exact fit and the test of close fit, and the Type I error rates for the test of close fit. I obtained $\hat{F}_{ML}$ for each sample and computed T , the RMSEA, and the test of close fit. I compared the empirical power to the theoretical power obtained from the MacCallum et al. (1996) method and I compared the empirical rejection rates to the nominal level. I retained replications that resulted in nonconverged or out of bounds parameter estimates (Heywood cases).

Power Analysis

The proportion of times the null hypothesis is rejected out of the 1000 random samples is the empirically observed power at that particular N . This proportion was computed for each combination of ε , N , and model type. I compared the empirical power estimates to the theoretical power computed using the MacCallum et al. (1996) procedure and observed how this correspondence was affected by model misspecification. I conducted power analysis for the test of exact fit in which $H_0: \varepsilon_0 = 0$ and the test of close fit in which $H_0: \varepsilon_0 \leq .05$.

Power for the Test of Exact Fit—Counting the number of rejections for the test of exact fit provided the empirical power estimates. The null hypothesis is that the model fits the population covariance matrix exactly, $H_0: \varepsilon_0 = 0$, and the alternative hypothesis is that the model does not fit the population covariance matrix exactly, $H_1: \varepsilon_1 > 0$. Therefore, the alternative hypothesis is true and the null hypothesis should be rejected for a correct decision.

The empirical and theoretical power curves for the test of exact fit for the CFA, SEM, and path models are given in Figures 4 – 6, respectively. As shown in the figures, the theoretical power closely matched the empirical power. For all three models, as the sample size increases, the power increases as expected. Also, as ε increases, power increases as expected. As can be seen from Figures 4 and 5, the empirical power curves for the SEM and CFA models are nearly identical. However, the empirical power curves for the path model are very different from the SEM and CFA models. The empirical power is much lower in nearly all conditions than the corresponding conditions of the SEM and CFA models except for the highest sample size and ε conditions. Recall that the CFA and SEM models have $d = 13$ and $d = 12$, respectively, whereas the path model has $d = 2$.

It is important to note that, for the SEM and CFA models when $\varepsilon = .025$, the power to reject the null hypothesis of exact fit is very low unless $N = 1000$ or 5000 . Although power is substantially higher when $\varepsilon = .05$, it is still not at the recommended level of .80 unless $N = 1000$ or 5000 . The situation is even worse for the path model. For $N = 1000$, power to reject the null hypothesis of exact fit does not exceed the recommended level of .80 until $\varepsilon = .075$. For $N = 500$, power does not reach .80 until $\varepsilon = .10$ and for $N = 300$, power does not reach .80 until $\varepsilon = .15$. For $N = 200$, the power does not reach .80 for any of the values of ε considered here.

Power for the Test of Close Fit—Counting the number of rejections for the test of close fit for the models in which the specified population discrepancy is .075, .10, and .15 provided the empirical power estimates for the test of close fit. The null hypothesis is that the fit of the model to the population covariance matrix is close, $H_0: \varepsilon_0 \leq .05$, and the alternative hypothesis is that it is not, $H_1: \varepsilon_1 > .05$. Therefore, the alternative hypothesis is true and the null hypothesis should be rejected for a correct decision.

I compared the empirical power with the power computed using the MacCallum et al. (1996) procedure (theoretical power). Both the empirical and theoretical power for the test of close fit for all models are given in Table 1.

The theoretical power closely matches the empirical power for all sample size conditions and ε conditions examined here for all three models. In most cases, the theoretical power and empirical power differ only in the third decimal place or not at all. In some conditions, the two values differ in the second decimal place. The two values did not differ in the first decimal place for any of the conditions. In summary, the MacCallum et al. (1996) procedure closely matches the empirical power for the conditions examined in the present simulation.

Although the theoretical power closely matched the empirical power, it should be noted that the power to reject the null hypothesis when $\varepsilon = .075$ is very low especially for the path model. For the SEM and CFA models, power does not reach the recommended level of .80 until $N = 1000$ and for the path model, power does not reach the recommended level of .80 until $N = 5000$.

Type I Error

Counting the number of rejections for the test of close fit for the models in which the specified population discrepancy is .025 or .05 provided the empirical Type I error rate. Because the null hypothesis for the test of close fit is $H_0: \varepsilon_0 \leq .05$, the null hypothesis is true and rejecting the null hypothesis is a Type I error. The nominal Type I error rate was set to .05. Therefore, the empirical Type I error rate should approximately equal .05 when the specified population discrepancy is .05.

I computed the empirical Type I error rates for the test of close fit for those conditions with $\varepsilon = .025$ and .05. These values are presented in Table 2 for all models. Comparing these values with the nominal Type I error rate of .05, the SEM and CFA models maintain Type I error well. On the other hand, the path model does not maintain the nominal Type I error as well (see Table 2). For $N = 300$ and $N = 1000$, the empirical Type I error matches the nominal level. For $N = 200$ and $N = 500$, the empirical Type I error is greater than the nominal level and for $N = 5000$, the empirical Type I error is smaller than the nominal level. Thus, there is not a clear pattern other than that the path model does not maintain nominal Type I error rates as well as the SEM and CFA models.

Conclusions

The results indicated that the test of close fit reasonably maintains Type I error rates and the MacCallum et al. (1996) procedure for the power of the test of close fit and for the test of exact fit worked well. As would be expected, power increased as the sample size increased and as ε increased for all three models. Increasing values of ε represent increasing values of λ and therefore an increase in effect size. For both the test of close fit and the test of exact fit, power was lower when d was lower as indicated by the difference in the results for the path model compared to those for the SEM and CFA models.

The MacCallum et al. (1996) procedure is valid at least for the conditions examined here, which included a value of $\varepsilon = .15$. This value is generally considered to represent substantial model misspecification. Yet, the power calculations based on the MacCallum et al. (1996) procedure matched the empirical power extremely well.

One advantage of the present study is the method by which model misspecification was introduced. Previous simulation studies on power have constructed the alternative model by adding or deleting paths. Saris et al. (1987) have found that the likelihood ratio test has unequal power for the same size of misspecification introduced into different model parameters. Further, power depends on the size of the misspecified parameter (Kaplan, 1995). Therefore, different power estimates may be obtained depending on the particular method used to specify of the alternative model. More generally, in the case of simulation studies, the method of introducing misspecification into the model may lead to different results (Cudeck & Browne, 1992). A limitation of the present study is that the data were generated from a multivariate normal distribution. The results presented here may not hold under violations of multivariate normality.

Although the empirical power and theoretical power matched closely, there was low power for the test of exact fit to reject H_0 , especially for the path model. The implication of this finding is that the researcher may conclude that the model fits well when in fact it does not. The power for the test of close fit was also low when $\varepsilon = .075$, especially for the path model. Based on these results, applied researchers should keep in mind that the probability of rejecting the model when in fact the model is incorrect is very low when evaluating models with low degrees of freedom. The practical implication of low power for models with few degrees of freedom is that researchers may conclude that the model fits well even when it

does not. Given the excellent performance of the MacCallum et al. (1996) power procedure, it is recommended that applied researchers utilize this procedure for conducting power analysis.

Acknowledgments

This article is based on the author's dissertation; advisor, Robert C. MacCallum. The author would like to thank Robert C. MacCallum, Linda Collins, and Kristopher J. Preacher for helpful feedback on this manuscript.

Preparation of this article was supported by NIDA Center Grant P50 DA100075 and NIDA Training Grant T32 DA017629-01A1.

References

- Browne, MW. Covariance structures. In: Hawkins, DM., editor. Topics in applied multivariate analysis. New York: Cambridge University Press; 1982.
- Browne, MW.; Cudeck, R. Alternative ways of assessing model fit. In: Bollen, KA.; Long, JS., editors. Testing structural equation models. Newbury Park, CA: Sage; 1993.
- Cudeck R, Browne MW. Constructing a covariance matrix that yields a specified minimizer and a specified minimum discrepancy function value. *Psychometrika* 1992;57(3):357–369.
- Cudeck R, Henly SJ. Model selection in covariance structures analysis and the “problem” of sample size: A clarification. *Psychological Bulletin* 1991;109:512–519. [PubMed: 2062982]
- Cudeck R, Henly SJ. A realistic perspective on pattern representation in growth data: Comment on Bauer and Curran (2003). *Psychological Methods* 2003;8:378–383. [PubMed: 14596498]
- Kaplan, D. Statistical power in structural equation modeling. In: Hoyle, RH., editor. Structural equation modeling: Concepts, issues, and applications. Thousand Oaks, CA: Sage; 1995.
- Lehmann, EL. Elements of large sample theory. New York: Springer; 1999.
- MacCallum RC. Working with imperfect models. *Multivariate Behavioral Research* 2003;38(1):113–139.
- MacCallum RC, Browne MW, Sugawara HM. Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods* 1996;1(2):130–149.
- MacCallum RC, Hong S. Power analysis in covariance structure modeling using GFI and AGFI. *Multivariate Behavioral Research* 1997;32(2):193–210.
- MacCallum RC, Tucker LR. Representing sources of error in the common-factor model: Implications for theory and practice. *Psychological Bulletin* 1991;109(3):502–511.
- Maydeu-Olivares A, Coffman DL. Random intercept item factor analysis. *Psychological Methods* 2006;11:344–362. [PubMed: 17154751]
- Maydeu-Olivares A, Lado N. Marketing orientation and business economic performance. *International Journal of Service Industry Management* 2003;14:284–309.
- Monahan, JF. Numerical methods of statistics. New York: Cambridge University Press; 2001.
- Preacher, KJ.; Coffman, DL. Computing power and minimum sample size for RMSEA [computer software]. 2006 May. Available from www.quantpsy.org
- Saris, WE.; Satorra, A. Power evaluations in structural equation models. In: Bollen, KA.; Long, JS., editors. Testing structural equation models. Newbury Park, CA: Sage; 1993.
- Saris WE, Satorra A, Sörbom D. The detection and correction of specification errors in structural equation models. *Sociological Methodology* 1987;17:105–129.
- Satorra A, Saris WE. Power of the likelihood ratio test in covariance structure analysis. *Psychometrika* 1985;50(1):83–90.
- Steiger, JH.; Lind, JC. Statistically based tests for the number of common factors. Paper presented at the annual meeting of the Psychometric Society; Iowa City, IA. 1980 Jun.
- Steiger JH, Shapiro A, Browne MW. On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika* 1985;50(3):253–264.

Figure 1.
Model A: Confirmatory Factor Analysis Model.

Figure 2.
Model B: Structural Equation Model.

Figure 3.
Model C: Path Analysis Model.

Figure 4.
Empirical and Theoretical Power Curves for Test of Exact Fit: CFA.

Figure 5.
Empirical and Theoretical Power Curves for Test of Exact Fit: SEM.

Figure 6.
Empirical and Theoretical Power Curves for Test of Exact Fit: Path Model.

Table 1

Power for Test of Close Fit

CFA		.075		.10		.15	
N	Empirical	Theoretical	Empirical	Theoretical	Empirical	Theoretical	
200	0.269	0.269	0.681	0.690	0.998	0.997	
300	0.371	0.379	0.865	0.865	1.000	1.000	
500	0.543	0.569	0.981	0.980	1.000	1.000	
1000	0.858	0.853	1.000	1.000	1.000	1.000	
5000	1.000	1.000	1.000	1.000	1.000	1.000	
SEM		.075		.10		.15	
N	Empirical	Theoretical	Empirical	Theoretical	Empirical	Theoretical	
200	0.250	0.257	0.665	0.661	0.994	0.996	
300	0.374	0.360	0.848	0.841	0.999	1.000	
500	0.569	0.542	0.971	0.973	1.000	1.000	
1000	0.821	0.828	0.999	1.000	1.000	1.000	
5000	1.000	1.000	1.000	1.000	1.000	1.000	
path model		.075		.10		.15	
N	Empirical	Theoretical	Empirical	Theoretical	Empirical	Theoretical	
200	0.106	0.115	0.243	0.232	0.597	0.590	
300	0.150	0.140	0.310	0.310	0.751	0.756	
500	0.171	0.186	0.446	0.450	0.925	0.923	
1000	0.279	0.289	0.707	0.706	0.999	0.997	
5000	0.797	0.799	1.000	0.999	1.000	1.000	

Table 2

Type I Error Rates for Test of Close Fit.

N	CFA			SEM			path model		
	0.025	0.05	0.025	0.025	0.05	0.025	0.025	0.05	0.05
200	0.005	0.047	0.008	0.008	0.050	0.022	0.022	0.063	0.063
300	0.001	0.051	0.003	0.003	0.053	0.015	0.015	0.049	0.049
500	0.001	0.052	0.001	0.001	0.055	0.011	0.011	0.063	0.063
1000	0.000	0.051	0.000	0.000	0.050	0.003	0.003	0.051	0.051
5000	0.000	0.052	0.000	0.000	0.054	0.000	0.000	0.039	0.039