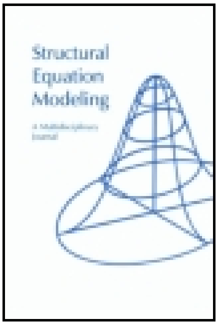

Structural Equation Modeling: A Multidisciplinary Journal
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hsem20

The Relation Among Fit Indexes, Power, and Sample Size in Structural Equation Modeling

Kevin H. Kim

Version of record first published: 19 Nov 2009

STRUCTURAL EQUATION MODELING, *12*(3), 368–390

The Relation Among Fit Indexes, Power, and Sample Size in Structural Equation Modeling

Kevin H. Kim
University of Pittsburgh

The relation among fit indexes, power, and sample size in structural equation modeling is examined. The noncentrality parameter is required to compute power. The 2 existing methods of computing power have estimated the noncentrality parameter by specifying an alternative hypothesis or alternative fit. These methods cannot be implemented easily and reliably. In this study, 4 fit indexes (RMSEA, CFI, McDonald's Fit Index, and Steiger's gamma) were used to compute the noncentrality parameter and sample size to achieve certain level of power. The resulting power and sample size varied as a function of (a) choice of fit index, (b) number of variables/degrees of freedom, (c) relation among the variables, and (d) value of the fit index. However, if the level of misspecification were held constant, then the resulting power and sample size would be identical.

The power of a study, and required sample size to achieve that power, are very important issues in any research. As structural equation modeling (SEM) becomes more widely used, the same beginning steps regarding sample size and power should be addressed.

There are four possible outcomes in any research, depending on whether or not the null hypothesis (H_0) is true or false and whether one rejects or does not reject the null hypothesis (see Table 1). There are two correct decisions and two errors (Type I and Type II). Type I error is rejecting a true H_0, whereas Type II error is not rejecting a false H_0. Looking at the rows of Table 1, a decision to either reject or not reject an H_0 can lead to a correct decision or an error. For instance, a decision to reject the H_0 can lead to a correct decision or a Type I error, whereas a decision to not reject the H_0 can lead to a correct decision or a Type II error. The correctness of these decisions depends on the probabilities of the Type I and Type II errors. The

Requests for reprints should be sent to Kevin H. Kim, Department of Psychology in Education, University of Pittsburgh, Pittsburgh, PA 15260. E-mail: khkim@pitt.edu

TABLE 1
Four Possible Outcomes in a Study (With Their Probabilities)

	True State	
Decision	*H_0 True*	*H_0 False*
Do not reject H_0	Correct decision ($1 - \alpha$)	Type II error (β)
Reject H_0	Type I error (α)	Correct decision ($1 - \beta$ = Power)

probability of a Type I error is known as α and the probability of a Type II error is known as β. The probability of not rejecting a true H_0 is $1 - \alpha$. The probability of rejecting a false H_0 is $1 - \beta$ or power. If β is large (i.e., low power) then the decision to not reject the H_0 may be incorrect. Likewise, if the α is large then the decision to reject the H_0 may be incorrect. In carrying out a research study, the values of α and β (consequently, power) should be set a priori to minimize errors. Typically, the α level is set at some arbitrary value (e.g., $\alpha = .05$). On the other hand, β depends on α and the noncentrality parameter, δ, which is an indication of sample size and the distance between null and alternative models. An increase in δ is associated with an increase in power (i.e., decrease in β).

In SEM, a not significant test statistic of overall fit is desired because the researcher typically does not want to reject an hypothesized model. However, this result can be due to lack of power. For example, a small sample size can guarantee low power. A not significant result in SEM will lead to a acceptance of a null hypothesis and may lead to publication. In other statistical methods a lack of power will result in demonstrating no effect and therefore it will not lead to publication. Therefore, power is an even more important issue in SEM than in other statistical methods.

This article examines the relation among fit indexes, power, and sample size in SEM. δ is required to compute power. The two main existing methods of computing power have estimated δ by specifying an alternative hypothesis or alternative fit. These methods cannot be implemented easily and reliably. Four fit indexes are used to compute δ and a sample size estimate to achieve a certain level of power. The different effects of these fit indexes on power and sample size estimates are discussed.

CURRENT METHODS

There are two main methods of computing power in SEM. The first method, introduced by Satorra and Saris (1985),[1] computes power by estimating the δ by $T_A = (N$

[1]The same idea was also presented by Matsueda and Bielby (1986). Original Satorra–Saris work was extended in Saris and Satorra (1993).

$- 1)F(\Sigma_A, \Sigma(\theta^*_A))$ where (a) θ^*_A is the vector minimizing $F(\Sigma_A, \Sigma(\theta))$; (b) $\Sigma_A = \Sigma(\theta_A)$ is an alternative covariance matrix where θ_A is a specific alternative parameter value that lies near the subspace formed by θ; (c) $F(\cdot)$ is the discrepancy fit function, for example normal theory likelihood function; and (d) $\Sigma(\theta)$ is a covariance matrix according to the null hypothesis (i.e., the model-implied covariance matrix). Power in SEM can be computed by $Pr(\chi^2_{obs} > \chi^2_{crit})$ under a noncentral χ^2 distribution where χ^2_{obs} is the observed chi-square and χ^2_{crit} is the critical chi-square value at some value of α. Notice that power is defined in the population; that is, Σ_A and $\Sigma(\theta)$ are population covariance matrices. The true population covariance matrix will be designated simply as Σ.

There are at least three limitations to the Satorra–Saris method. The first limitation is that a specific alternative model must be defined, which is not always easy. Second, not all possible alternative models can be tested due to the technical limitation that θ_A must lie near the subspace formed by the θ (i.e., an alternative model is nested within the model according to a null hypothesis). Third, it can only be used to compute power, and not the sample size needed to achieve a given power in a future study, herein called the *proposed sample size,* because it requires raw data.[2] One place where the Satorra–Saris method may be useful is in computing power for model modifications.

Recognizing this limitation, Mooijaart (2003) and Yuan and Hayashi (2003) generalized the Satorra–Saris method by utilizing a bootstrap approach. A proposed sample size can be computed by bootstrapping a sample until a desired power is achieved. However, like the Satorra–Saris method, a raw data set is required to be able to implement the bootstrap. Unlike the Satorra–Saris method, the Mooijaart and Yuan–Hayashi methods can be used with nonnormal[3] as well as missing data. Also, an empirical distribution of test statistics can be used for calculations, instead of a theoretical distribution, due to the bootstrap methodology.

Muthén and Muthén (2002) solved the problem of a raw data requirement by implementing the Satorra–Saris method through a Monte Carlo study. They simulated a raw data set based on known parameters about a sample. The Muthén–Muthén method, like the Yuan–Hayashi method, can be used with nonnormal and missing data. However, their estimate of proposed sample size is only as good as the parameter estimates used in its computation. If the parameter estimates are incorrect, the power and proposed sample sizes will be incorrect as well.

The second method was introduced by MacCallum, Browne, and Sugawara (1996). They computed power by redefining the H_0 in SEM in terms of the root

[2]The proposed sample size could be called a sample size estimate. However, under the assumptions, the proposed sample size is a precise population value not subject to sampling variability. Sample data do not need to be used to compute such values.

[3]Satorra (2003) generalized the Satorra–Saris method to nonnormal data.

mean square error of approximation (RMSEA) rather than Σ. The population RMSEA is defined as:

$$\varepsilon = \sqrt{\frac{F(\Sigma, \Sigma(\theta))}{df}}$$

where Σ is the true population covariance matrix and $\Sigma(\theta)$ is the population covariance matrix under the null hypothesis. The H_0 in SEM is $\Sigma = \Sigma(\theta)$ and the alternative hypothesis (H_a) is $\Sigma \neq \Sigma(\theta)$. Σ_A is not required in this method. MacCallum et al. created three different H_0 and H_a (not close fit, $H_0 : \varepsilon \geq .05$ and $H_a : \varepsilon < .05$; close fit, $H_0 : \varepsilon \leq .05$ and $H_a : \varepsilon > .05$; and exact fit, $H_0 : \varepsilon = 0$ & $H_a : \varepsilon > 0$). A researcher would choose one of these hypotheses depending on how close a model fit he or she wishes to test. However, these different hypotheses do not specify exact RMSEA values, which are needed to compute power under the MacCallum–Browne–Sugawara (MBS) method. They arbitrarily chose a pair of values for each hypothesis (not close fit, $\varepsilon_0 = .05$ and $\varepsilon_a = .01$; close fit, $\varepsilon_0 = .05$ and $\varepsilon_a = .08$; and exact fit, $\varepsilon_0 = 0$ and $\varepsilon_a = .05$; where ε_0 denotes value of the H_0 and ε_a denotes value of the H_a). Based on these values, they computed two noncentral χ^2 distributions: according to ε_0, which yields a noncentrality parameter $\delta_0 = (N-1)df\varepsilon^2{}_0$ where df is the degrees of freedom and N is the sample size, and according to ε_a which yields a noncentrality parameter $\delta_a = (N-1)df\varepsilon^2{}_a$. These are obtained from $\delta = (N-1)F(\Sigma, \Sigma(\theta))$ and the defining formula for ε. Power in the MBS method is defined as the area under the true distribution of the test statistic, beyond $\chi^2{}_{crit}$ in the appropriate direction ($\chi^2{}_{crit}$ in MBS is under their H_0; i.e., ε_0 noncentral χ^2 distribution, which will be denoted $\chi^2{}_{crit,\delta_0}$). For instance, in close fit, power = $Pr(\chi^2{}_{df,\delta_a} \geq \chi^2{}_{crit,\delta_0})$, and in not close fit, Power = $Pr(\chi^2{}_{df,\delta_a} \leq \chi^2{}_{crit,\delta_0})$.

MacCallum and Hong (1997) extended the MBS method to allow the use of the Goodness-of-Fit index (GFI) and the Adjusted Goodness-of-Fit Index (AGFI) instead of RMSEA. They stated that the population GFI is equal to Steiger's (1989) γ and the population AGFI is equal to $\gamma^* = 1 - \frac{p(p+1)}{2df}\gamma$ where p is number of variables and $\gamma = \frac{p}{p + 2\left(\frac{\delta}{N-1}\right)}$.

MacCallum and Hong chose to compute the power using only the close fit hypothesis (GFI, $\gamma_0 = .95$ and $\gamma_a = .90$; and AGFI, $\gamma^*{}_0 = .95$ and $\gamma^*{}_a = .90$; where γ_0 and $\gamma^*{}_0$ denote values of the H_0 and γ_a and $\gamma^*{}_a$ denote values of the H_a).

A major limitation to the MBS method is that only one goodness-of-fit statistic is used. Clearly, different indexes can lead to different conclusions. For example, MacCallum and Hong (1997) showed that the power computed using RMSEA and AGFI increased as degrees of freedom increased; however, the power computed using GFI had the opposite effect. They concluded that due to the differences in the results, power computed using GFI cannot be trusted. This conflicting result is explained in this article.

There has been much discussion on the issue of exact versus close fit in SEM (e.g., SEMNET, SEM listserv). The question of whether or not a model that does not fit the data perfectly (or exactly) is useful has supporters on both sides of the issue. Indeed, one of the reasons there are more than one set of H_0 and H_a for the MBS method was to address this issue. However, power has traditionally been computed for exact fit (or test) with other statistical methods. Power-related calculations in this article are limited to the traditional hypothesis of exact fit.

PROPOSED METHOD

In SEM, if the null hypothesis is true, that is, $\Sigma = \Sigma(\theta)$, then if all standard conditions (e.g., independent observations, large sample size, correct assumed distribution, etc.) are met, a model test statistic, $\chi^2_{obs} = (N-1)\,F(S, \Sigma(\hat{\theta}))$ for an estimator $\hat{\theta}$ based on the sample covariance matrix S and minimizing an appropriate discrepancy function $F(\cdot)$, has an asymptotic central chi-square distribution, χ^2_{df}, with $\frac{p(p+1)}{2} - q$ degrees of freedom where q is the number of parameters. The expected value of the χ^2_{df} will be its degrees of freedom ($E(\chi^2_{df}) = df$). However, if the null hypothesis is not true, that is, $\Sigma \neq \Sigma(\theta)$, then under appropriate regularity conditions,[4] a model test statistic has a noncentral chi-square distribution, $\chi^2_{df,\delta}$, with $\frac{p(p+1)}{2} - q$ degrees of freedom and noncentrality parameter δ. Its expected value will be its degrees of freedom plus the noncentrality parameter ($E(\chi^2_{df,\delta}) = df + \delta$). Evidently, δ is the amount of mean shift in the χ^2 variate under null and alternative hypotheses.

A model test statistic is significant if the χ^2_{obs} is greater than the χ^2_{crit} at some specified α level under a χ^2_{df} distribution ($Pr(\chi^2_{obs} > \chi^2_{crit} \mid \chi^2_{df}) = \alpha$). χ^2_{crit} is chosen by the investigator for the given df; usually it is a value for which $\alpha = .05$. Therefore, power is equal to the probability that χ^2_{obs} is greater than χ^2_{crit} under a $\chi^2_{df,\delta}$ distribution (power = $Pr(\chi^2_{obs} > \chi^2_{crit} \mid \chi^2_{df,\delta})$). Notice that the probability associated with power increases directly as δ increases, because increasing the value of δ shifts the expected value of the noncentral distribution further to the right (see Figure 1).

[4]For a description of the regularity conditions see Steiger, Shapiro, and Browne (1985). These include that the misspecification is not too extreme. If the misspecification is extreme, the noncentral χ^2 distribution may not well approximate the empirical distribution of $(N-1)F(S,\Sigma(\hat{\theta}))$ (see Curran, Bollen, Paxton, Kirby, & Chen, 2002).

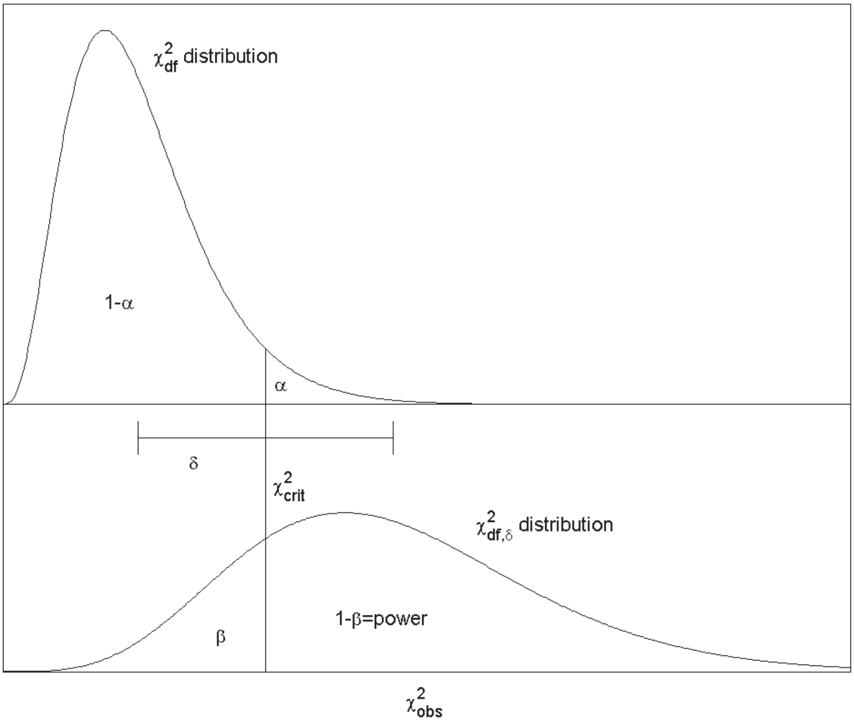

FIGURE 1 An illustration of the relations among α, β, δ and power in χ^2 distribution.

An alternative model does not need to be specified, although it is implied by δ. A nonfitting model will have a $\chi^2_{df,\delta}$ distribution. In practice, however, δ is unknown, and meaningful values of δ are hard to know a priori. Hence, useful operational procedures for computing δ are required. This article proposes that δ in SEM can be computed using a fit index. Unfortunately, not all fit indexes can be expressed in terms of δ. The fit indexes that can be expressed in terms of δ are the Comparative fit index (CFI; Bentler, 1990), RMSEA (Steiger & Lind, 1980), McDonald's (1989) fit index, and Steiger's (1989) γ.

Using the CFI,[5] a noncentrality parameter can be computed from

$$CFI = 1 - \frac{\delta}{\delta_B}$$

where δ_B is a noncentrality parameter for a base model. Therefore, $\delta = (1 - \text{CFI})\,\delta_B$ where $\delta_B = T_B - df_B$, T_B is the population value of the test statistic, and df_B is the de-

[5]The same as McDonald and Marsh's (1990) relative noncentrality index between values of 0 and 1.

grees of freedom for the base model (usually, the independence model). Letting $N = N_{\mathrm{CFI}}$ for clarity, after substituting,

$$\delta = (1 - \mathrm{CFI})(T_B - df_B)$$

where $T_B = (N_{\mathrm{CFI}} - 1)F_B$, $F_B = F(\Sigma(\theta_B),\Sigma)$, and θ_B is the parameter value of the base model:

$$\delta = (1 - \mathrm{CFI})[(N_{\mathrm{CFI}} - 1)F_B - df_B]. \tag{1}$$

F_B is required to compute δ and as shown later, power and proposed sample size using the CFI. However, this can be computed based on the model being tested and comparing it to an independence model. Consider the usual normal theory maximum likelihood (ML) function $log\,|\,\Sigma(\theta)\,| - log\,|\,S\,| + tr(S\Sigma^{-1}(\theta)) - p$ and replace S by Σ (in the population) and $\Sigma(\theta)$ by Σ_B, the population baseline model. Then $F_B = log\,|\,\Sigma_B\,| - log\,|\,\Sigma\,| + tr(\Sigma\Sigma_B^{-1}) - p$. Using a correlation matrix instead of a covariance matrix for scale invariant models,[6] this simplifies to

$$F_B = log\,|I| - log\,|\rho| + tr(\rho I) - p = -log\,|\rho| \tag{2}$$

where ρ is the correlation matrix based on the values of the model parameters and I is the identity matrix of $p \times p$.

Likewise for the RMSEA, using N_ε for the sample size to differentiate it from N_{CFI}, as defined previously, $\varepsilon = \sqrt{\frac{F(\Sigma,\Sigma(\theta))}{df}}$. Thus, with $\delta = (N_\varepsilon - 1)F(\Sigma,\Sigma(\theta))$, we have

$$\delta = (N_\varepsilon - 1)\varepsilon^2\, df. \tag{3}$$

Similarly, Steiger's $\gamma = \frac{p}{p + 2\left(\frac{\delta}{N_\gamma - 1}\right)}$, so that

$$\delta = (N_\gamma - 1)\left[\frac{p}{2}\left(\frac{1}{\gamma} - 1\right)\right] \tag{4}$$

and for McDonald's fit index, $\mathrm{Mc} = \exp\left[-\frac{1}{2}\left(\frac{\delta}{N_{\mathrm{Mc}} - 1}\right)\right]$,

$$\delta = (N_{\mathrm{Mc}} - 1)[-2log(\mathrm{Mc})]. \tag{5}$$

Each of the Equations 1 through 5 shows that the noncentrality parameter (consequently power) is a function of the sample size and a fit index. The fit index plays the role of an effect size. Obviously, the same δ can be obtained by an infinite combination of a sample size and a fit index. As expected, for a given index, increasing sample size always increases power. However, some additional information may be necessary, depending on which equation is being used (i.e., p, df, F_B, and df_B). A major purpose of this article is to evaluate how these various factors influence computations to determine proposed sample sizes to achieve a given level of power.

[6]For full set of conditions, see Shapiro and Browne (1990).

TABLE 2
List of Critical Noncentrality Parameters ($\delta_{1-\beta}$) by Degrees of Freedom and Power at α = .05

	Power			*Power*	
df	*.80*	*.90*	*df*	*.80*	*.90*
1	7.849	10.507	26	23.200	28.784
2	9.635	12.654	27	23.546	29.194
3	10.903	14.171	28	23.885	29.596
4	11.935	15.405	29	24.219	29.991
5	12.828	16.469	30	24.547	30.379
6	13.624	17.419	35	26.107	32.225
7	14.351	18.284	40	27.557	33.940
8	15.022	19.083	45	28.918	35.549
9	15.650	19.829	50	30.204	37.069
10	16.241	20.532	60	32.593	39.891
11	16.802	21.198	70	34.787	42.483
12	17.336	21.833	80	36.829	44.893
13	17.847	22.439	90	38.745	47.155
14	18.338	23.022	100	40.556	49.293
15	18.811	23.583	125	44.721	54.206
16	19.268	24.125	150	48.483	58.643
17	19.710	24.650	175	51.942	62.721
18	20.139	25.158	200	55.160	66.515
19	20.555	25.652	225	58.182	70.077
20	20.961	26.132	250	61.039	73.444
21	21.356	26.600	300	66.353	79.706
22	21.741	27.057	350	71.238	85.462
23	22.118	27.503	400	75.785	90.818
24	22.486	27.939	450	80.055	95.848
25	22.847	28.366	500	84.093	100.604

Let $\delta_{1-\beta}$ be a value of δ that achieves a power equal to $1 - \beta$ (in Figure 1, this is just δ). Given a certain level of power and degrees of freedom for a model, $\delta_{1-\beta}$ can be computed from power = 1 – CDF($\chi^2_{df,\delta}$,χ^2_{crit}) where CDF(·) is a cumulative density function of χ^2_{crit} under a $\chi^2_{df,\delta}$ (Table 2; see Appendix). The $\delta_{1-\beta}$ values increase as *df* increases. The values of $\delta_{1-\beta}$ for power = .90 are higher than for power = .80 (Figure 2). Only $\delta_{1-\beta}$ is required to ensure power. As noted, the same $\delta_{1-\beta}$ can be achieved by an infinite combination of a sample size and a fit index. However, not all values of a fit index are of interest, only good fit (e.g., CFI = .95). A minimum sample size necessary to achieve a certain level of power, the proposed sample size, can be computed given a value of the fit index by replacing δ from Equations 1 through 5 with $\delta_{1-\beta}$ and solving for *N*s,

$$N_{\text{CFI}} = \frac{\delta_{1-\beta} + df_B(1-\text{CFI})}{F_B(1-\text{CFI})} + 1 \tag{6}$$

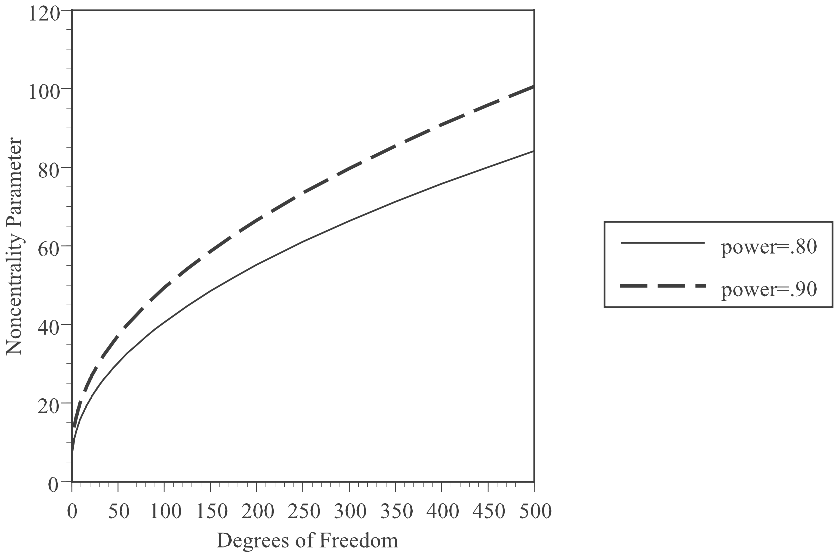

FIGURE 2 Plot of noncentrality parameter by degrees of freedom for power of .80 and .90

$$N_{\varepsilon} = \frac{\delta_{1-\beta}}{\varepsilon^2 df} + 1 \tag{7}$$

$$N_{\gamma} = \frac{2\gamma\delta_{1-\beta}}{p(1-\gamma)} + 1 \tag{8}$$

$$\text{and } N_{\mathrm{Mc}} = -\frac{1}{2}\left(\frac{\delta_{1-\beta}}{\log(\mathrm{Mc})}\right) + 1 \tag{9}$$

These equations provide a minimum sample size required to detect a difference between a model with perfect fit (e.g., CFI = 1.00 and $\varepsilon = 0.00$) and models with a less than perfect fit (e.g., CFI = .95 and $\varepsilon = .05$). The relation among the fit indexes, sample size, and power, Equations 6 through 9, may also depend on other variables (i.e., p, df, F_B, and df_B). To gauge a general pattern of relation, these additional variables must be known.

PROPOSED SAMPLE SIZE

CFI

The proposed minimum sample size to achieve a given level of power using the CFI is a function of $\delta_{1-\beta}$, df_B, F_B, and CFI (see Equation 6). An increase in the sample size is associated with an increase in the $\delta_{1-\beta}$ and CFI. However, an increase in the sample size is also associated with a decrease in F_B. F_B is a function of the correlations among the variables, ρ_{ij}, and the number of variables, p. An increase in ρ_{ij} and p is associated with an increase in F_B, which is associated with a decrease in the proposed sample size. An increase in p is also associated with an increase in df_B (see Figure 3). To further understand the combination of these effects, the sample size estimates in several typical confirmatory factor analysis (CFA) models were computed (see Tables 3–5). The CFA model chosen was a simple two- (or three- or five-) factor model with three indicators, each having a simple cluster structure. In these tables, λ_x is the common factor loading used for all nonzero loadings, and ϕ is the common value of the factor intercorrelation among pairs of factors. Contrary to the figure, an increase in the number of variables was not associated with an increase in the proposed sample size. This occurred because an increase in the number of variables for a fixed level of intercorrelation is offset by an increase in F_B.

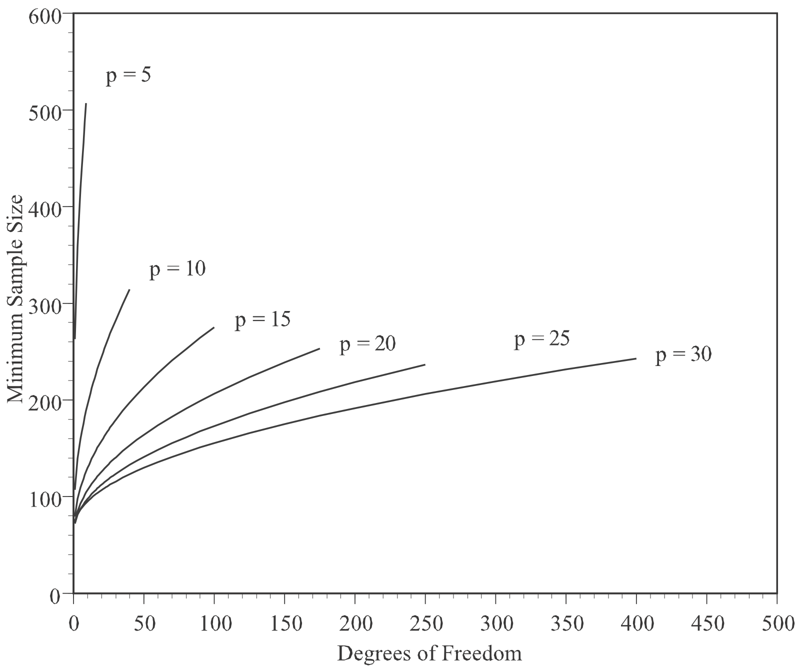

FIGURE 3 The proposed minimum sample sizes by degree of freedom using CFI = .95 and $\rho_{ij} = .30$ at various number of variables. *Note*: Only models with $df < df_B$ are graphed. $p = 5$ with $\rho_{ij} = .30$ generates $F_B = .6382$; $p = 10$, $F_B = 1.9017$; $p = 15$, $F_B = 3.448$; $p = 20$, $F_B = 4.8747$; $p = 25$, $F_B = 6.4561$; $p = 30$, $F_B = 8.0714$.

TABLE 3
Proposed Minimum Sample Sizes Using CFI for $p = 6$, $df = 8$,
Two-Factor Confirmatory Factor Analysis Model With $\phi = .30$

Power	*CFI*	N_{CFI}	
.80	.90	225	67
	.95	429	127
	.99	2,061	607
.90	.90	280	83
	.95	540	159
	.99	2,612	769
		$\lambda_x = .60$	$\lambda_x = .80$
		$F_B = .7366$	$F_B = 2.5042$

Note. CFI = Comparative fit index.

TABLE 4
Proposed Minimum Sample Sizes Using CFI for $p = 9$, $df = 24$,
Three-Factor Confirmatory Factor Analysis Model With $\phi = .30$

Power	*CFI*	N_{CFI}	
.80	.90	228	69
	.95	424	128
	.99	1,990	597
.90	.90	276	83
	.95	519	156
	.99	2,465	740
		$\lambda_x = .60$	$\lambda_x = .80$
		$F_B = 1.1485$	$F_B = 3.8308$

Note. CFI = Comparative fit index.

TABLE 5
Proposed Minimum Sample Sizes Estimates Using CFI for $p = 15$, $df = 80$,
Five-Factor Confirmatory Factor Analysis Model With $\phi = .30$

Power	*CFI*	N_{CFI}	
.80	.90	235	73
	.95	417	129
	.99	1,872	578
.90	.90	275	85
	.95	496	154
	.99	2,270	701
		$\lambda_x = .60$	$\lambda_x = .80$
		$F_B = 2.0245$	$F_B = 6.5620$

Note. CFI = Comparative fit index.

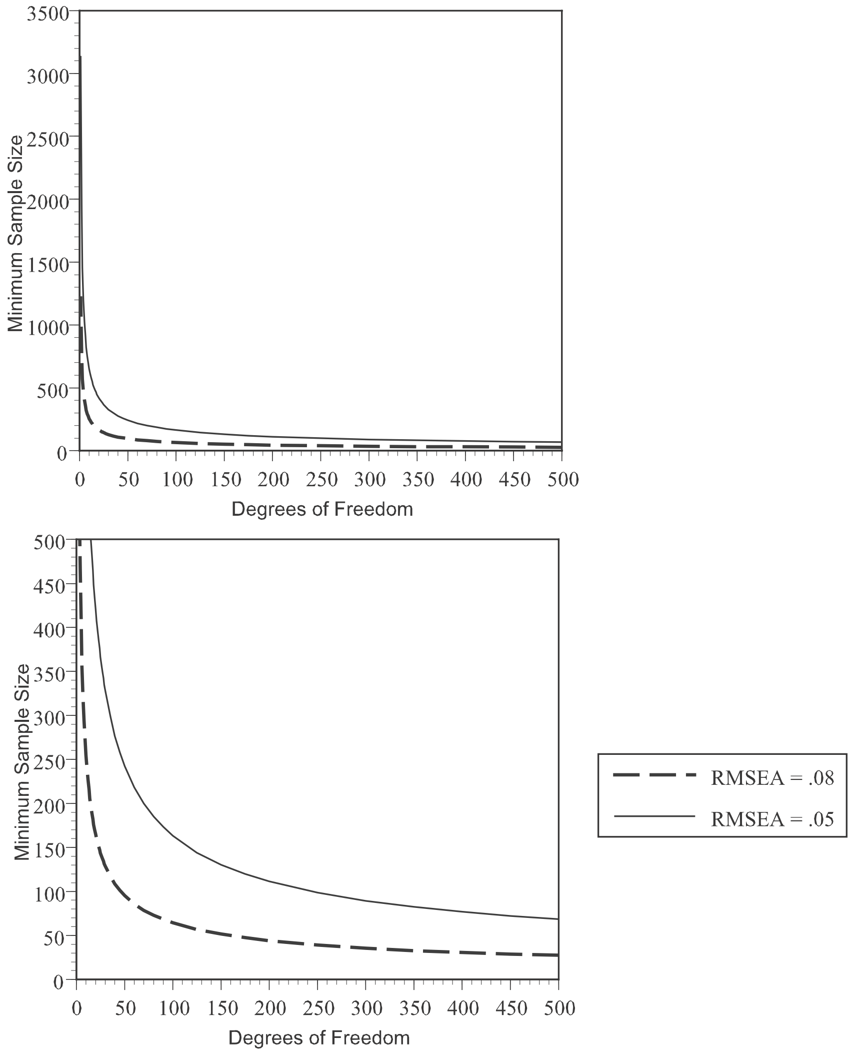

FIGURE 4 Proposed minimum sample sizes by degrees of freedom using ε.

RMSEA

The proposed minimum sample size using ε is a function of $\delta_{1-\beta}$, *df,* and ε (see Equation 7). The proposed sample size increases as the fit index decreases. However, the sample size decreases as the degrees of freedom increases (see Figure 4). It is not affected by ρ_{ij}. Several typical CFA model proposed sample sizes using ε are shown in Tables 6 through 8.

TABLE 6
Proposed Minimum Sample Sizes Using for McDonald's Fit Index, Steiger's γ, and Root Mean Squared Error of Approximation for $p = 6$ and $df = 8$

Power	*Mc*	N_{Mc}	γ	N_γ	ε	N_ε
.80	.90	72	.90	46	.08	294
	.95	147	.95	96	.05	752
	.99	748	.99	497	.01	18,779
.90	.90	92	.90	58	.08	374
	.95	187	.95	122	.05	955
	.99	950	.99	631	.01	23,855

TABLE 7
Proposed Minimum Sample Sizes Using McDonald's Fit Index, Steiger's γ, and Root Mean Squared Error of Approximation for $p = 9$ and $df = 24$

Power	*Mc*	N_{Mc}	γ	N_γ	ε	N_ε
.80	.90	108	.90	46	.08	147
	.95	220	.95	96	.05	376
	.99	1,120	.99	496	.01	9,370
.90	.90	134	.90	57	.08	183
	.95	273	.95	119	.05	467
	.99	1,391	.99	616	.01	11,642

TABLE 8
Proposed Minimum Sample Sizes Using McDonald's Fit Index, Steiger's γ, and Root Mean Squared Error of Approximation for $p = 15$ and $df = 80$

Power	*Mc*	N_{Mc}	γ	N_γ	ε	N_ε
.80	.90	176	.90	45	.08	73
	.95	360	.95	94	.05	185
	.99	1,833	.99	487	.01	4,605
.90	.90	214	.90	55	.08	89
	.95	439	.95	115	.05	225
	.99	2,234	.99	594	.01	5,613

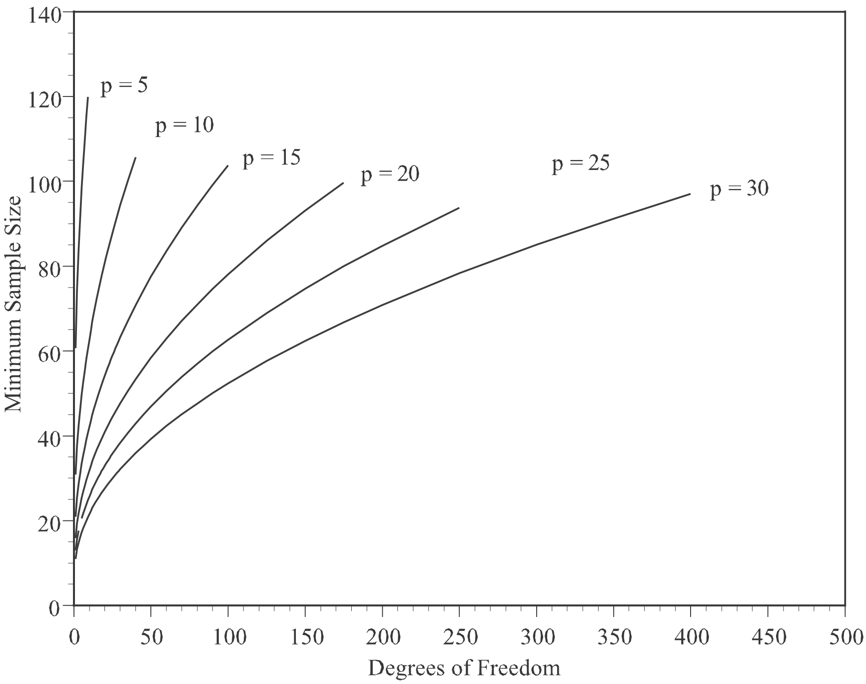

FIGURE 5 Proposed minimum sample sizes by degrees of freedom using Steiger's $\gamma = .95$ at various number of variables. *Note*: Only models with $df < df_B$ are graphed.

Steiger's γ

The proposed minimum sample size using γ is a function of $\delta_{1-\beta}$, p, and γ (see Equation 8). The proposed sample size increases as the fit index increases. However, the sample size is not affected by an increase in p. An increase in the number of variables is offset by an increase in $\delta_{1-\beta}$. The sample sizes computed using γ are not affected by ρ_{ij} (see Figure 5). Several typical CFA model proposed sample sizes using γ are shown in Tables 6 through 8.

McDonald's Fit Index

The proposed minimum sample size using Mc is a function of $\delta_{1-\beta}$ and Mc (see Equation 9). The sample size increases as the value of the fit index and the number of variables increase (see Figure 6). Even though p is not in the equation, it is still affected by p because $\delta_{1-\beta}$ is affected by df. The proposed sample size using Mc is not affected by ρ_{ij}. Several typical CFA model sample sizes using Mc are shown in Tables 6 through 8.

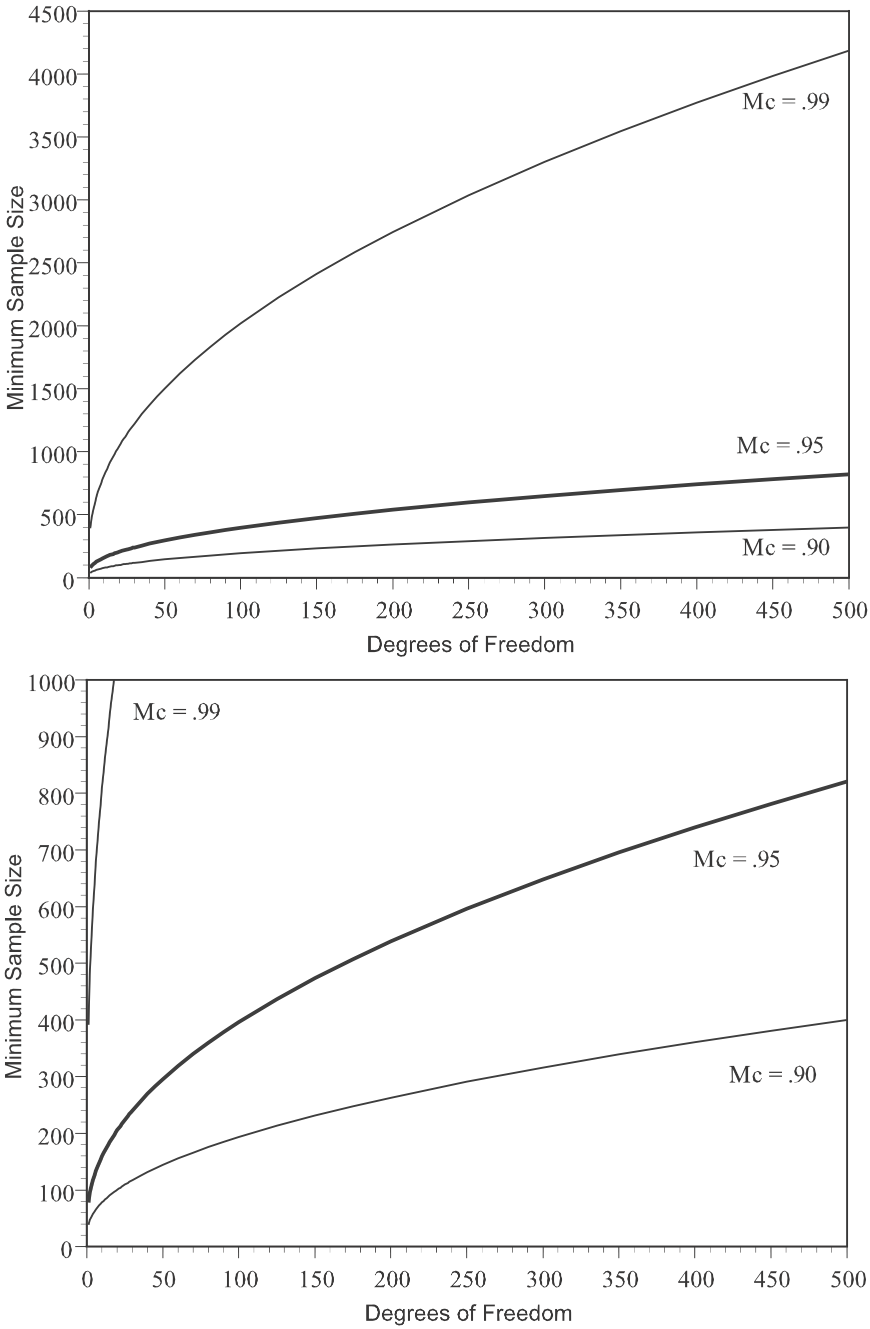

FIGURE 6 Proposed minimum sample sizes by degrees of freedom using Mc.

Relation Among Proposed Sample Sizes and Fit Indexes

In computing proposed sample sizes, ε produced a very different pattern of results than the other methods. However, on closer inspection, the pattern of proposed sample sizes looks very similar to results produced by CFI and Steiger's γ at certain values of p and ρ. A simple example produced using CFI by generating different df from different numbers of p is shown in Figure 7. The plot is very similar to one generated by ε; see Figure 4. Thus it can be said that the pattern of results generated by ε is a special case of that given by CFI and Steiger's γ.

The different magnitude of proposed sample sizes by the four fit indexes is not contradictory information. Rather, different proposed sample sizes were computed because different values of fit index were measuring different levels of misspecification in a model. For example, a value of CFI = .95 may or may not represent the same degree of misspecification as ε = .05 in a given model. Naturally, if they do not represent the same misspecification, they will not produce the same proposed sample sizes. What level of fit indexes will produce the same misspecification? This can be computed from Equations 1 through 5. All four equa-

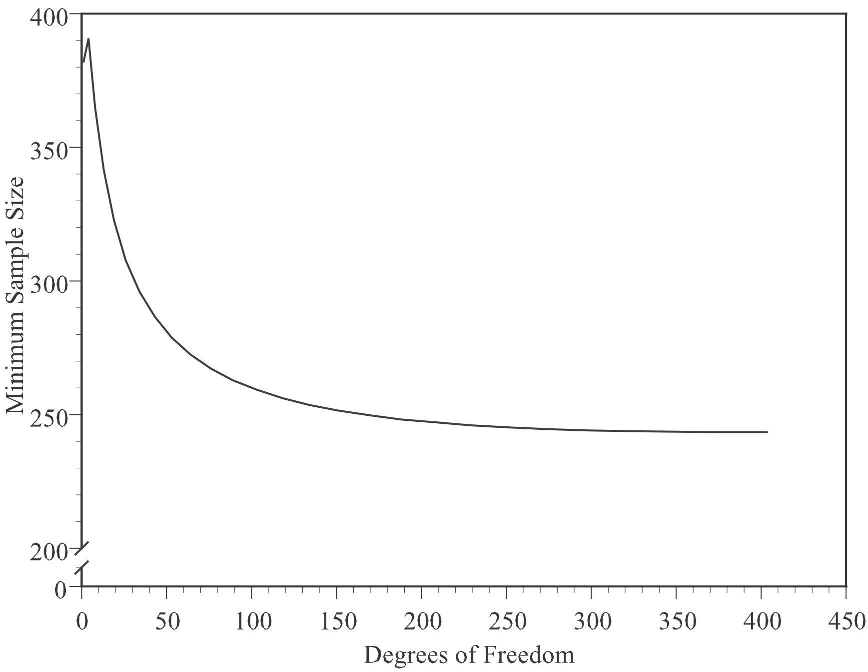

FIGURE 7 Proposed minimum sample sizes by degrees of freedom for one-factor CFA model using CFI = .95, different p, and ρ = .30.

tions are expressed in terms of δ. Therefore, they can be equated and be used to solve for values of fit indexes. For example, to compute what value of CFI is equal to ε for a given model, Equations 1 through 3 can be equated:

$$(1 - \mathrm{CFI})[(N - 1)F_B - df_B] = (N - 1)\varepsilon^2\, df \tag{10}$$

N_{CFI} and N_ε have been replaced by a common N. A CFI = .95 will measure the same misspecification as ε = .05 for only certain combinations of additional information: F_B, df_B, df, and N. Not just any combination of additional information will produce the same misspecification in a model for certain values of CFI and ε. For example, with the two-factor CFA model in Table 3 and CFI = .95, a sample size of 127 is required to achieve a power of .80. This sample size estimate implies that ε = .122. For this example model, a value of CFI = .95 measures the same level of misspecification as ε = .122. A CFI = .95 would measure the same misspecification as ε = .05 in this example if the F_B = .519. Likewise, the relation between CFI and the other two fit indexes can be computed:

$$\left(1 - \mathrm{CFI}\right)\left[\left(N - 1\right)F_B - df_B\right] = \left(N - 1\right)\left[\frac{p}{2}\left(\frac{1}{\gamma} - 1\right)\right] \tag{11}$$

$$(1 - \mathrm{CFI})[(N - 1)F_B - df_B] = (N - 1)[-2\,log(\mathrm{Mc})] \tag{12}$$

N_γ and N_{Mc} have also been replaced by a common N. Using the same two-factor CFA model, a CFI = .95 is equal to γ = .962 and Mc = .942. A similar type of relation can be computed by equating any combination of the fit indexes. Graphs show that there are monotonic relations among the fit indexes given a model (Figure 8), and also that not all combinations of values are possible (e.g., CFI = .95 and ε = .05).

STEP-BY-STEP INSTRUCTION WITH AN EXAMPLE

Here are the steps to compute a proposed sample size for each fit index. The first three steps are the same for all four fit indexes.

1. Come up with a model.
   Let us consider a three-factor, nine-variable CFA model (p = 9).
2. Compute df for the model.
   In the three-factor CFA model, df = 24.
3. Compute $\delta_{1-\beta}$.
   According to Table 2, $\delta_{1-\beta}$ = 22.486 for power = .80. For df (or power) not in the table, use the algorithms in the Appendix.

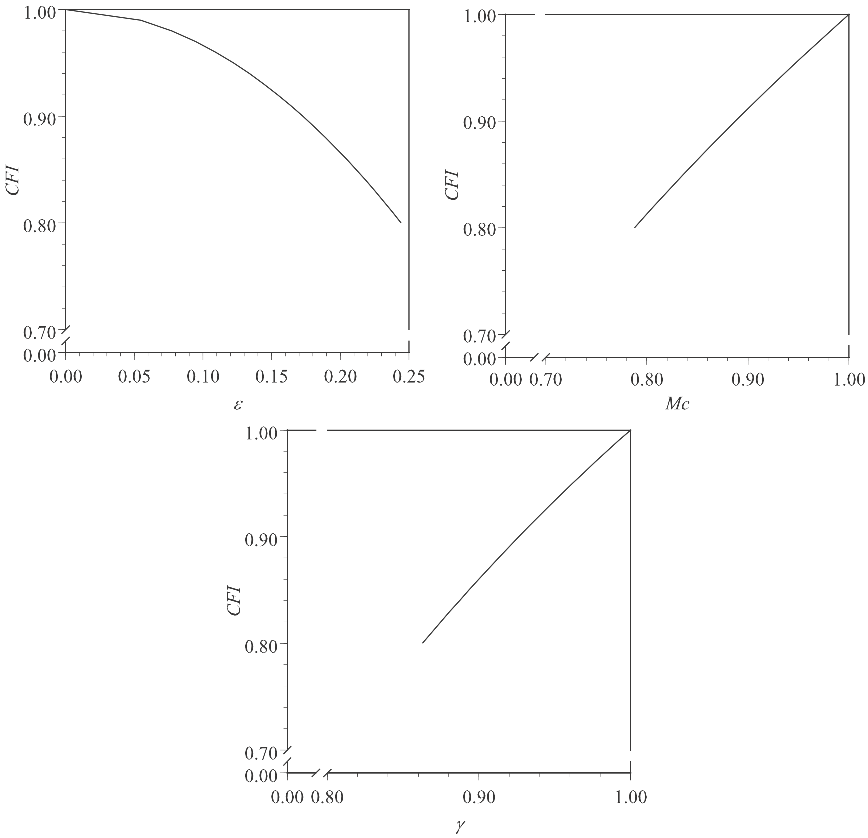

FIGURE 8 The relations among CFI, ε, Mc, and γ for two-factor CFA model with F_B = 2.5042, df_B = 15, df = 8, and p = 6.

CFI

- Choose a value of CFI.

$$\text{CFI} = .95$$

- Compute F_B using Equation 2.
  To compute F_B, a ρ matrix needs to be estimated first. This can be achieved by first estimating the standardized parameters of the model. Assume factor loadings of .60 ($\lambda_x = .60$) and correlation among the factors to be .30 (ϕ = .30). Compute the model correlation matrix based on the parameters (the same as computing the model-implied covariance matrix except using standardized parameters). Then using Equation 2, compute F_B.

$$F_B = 1.1485$$

- Compute df_B.

$$df_B = 36$$

- Compute a proposed sample size using Equation 6.

$$N_{\mathrm{CFI}} = \frac{\delta_{1-\beta} + df_B\left(1 - \mathrm{CFI}\right)}{F_B\left(1 - \mathrm{CFI}\right)} + 1 = \frac{22.486 + 36\left(1 - .95\right)}{1.1485\left(1 - .95\right)} + 1 = 424$$

RMSEA

- Choose a value of ε.

$$\varepsilon = .05$$

- Compute a proposed sample size using Equation 7.

$$N_{\varepsilon} = \frac{\delta_{1-\beta}}{\varepsilon^2 df} + 1 = \frac{22.486}{.05^2\left(24\right)} + 1 = 376$$

Steiger's γ

- Choose a value of γ.

$$\gamma = .95$$

- Compute a proposed sample size using Equation 8.

$$N_{\gamma} = \frac{2\gamma\delta_{1-\beta}}{p(1-\gamma)} + 1 = \frac{2\left(.95\right)\left(22.486\right)}{9\left(1 - .95\right)} + 1 = 96$$

Mc

- Choose a value of Mc.

$$\mathrm{Mc} = .95$$

- Compute a proposed sample size using Equation 9.

$$N_{\mathrm{Mc}} = -\frac{1}{2}\left(\frac{\delta_{1-\beta}}{\log(\mathrm{Mc})}\right) + 1 = -\frac{1}{2}\left(\frac{22.486}{\log(.95)}\right) + 1 = 220$$

DISCUSSION

Power and minimum sample size required to achieve a given power varied as a function of (a) choice of fit index, (b) number of variables and degrees of freedom, (c) ρ_{ij}, and (d) value of the fit index. The different proposed sample sizes are due to different levels of misspecification in a model, not different characteristics of the fit indexes. They will produce the same proposed sample size given the same level of misspecification in a model. It is hard to know what values of fit indexes measure the same level of misspecification without computing them using Equations 1, 3, 4, and 5. Some additional information about the data may be required.

ε and Mc use the smallest amount of information to compute proposed minimum sample size. They only require a value of fit index in addition to the degree of freedom of a model. CFI requires the most information. It requires not only a value of fit index along with the degree of freedom of a model, but the degree of freedom of an independent model and F_B, which depends on the relations among the variables. Steiger's γ requires the number of variables in addition to a value of a fit index and degrees of freedom. Each of these fit indexes produces different proposed minimum sample sizes due to different measure of fit. For example, a value of .95 for CFI will not always indicate the same model misspecification as a value of .05 for ε.

The proposed method relies on the use of fit indexes. Therefore, a knowledge of the distribution of fit indexes would be advantageous. The distribution of ε was introduced by Steiger and Lind (1980) and has been known for sometime, whereas the distributions of other fit indexes have only recently been studied. Ogasawara (2001) derived an approximation to the distributions of several fit indexes (e.g., GFI, AGFI, CFI) with the assumption of multivariate normality using a delta method. How the distributions of these fit indexes impact the measure of noncentrality parameter and proposed sample size require further study. It would be valuable to consider an application of Ogasawara's findings to the computation of power and sample size (e.g., distribution of noncentrality parameter, instead of just a point estimate, to allow computation of a confidence interval).

According to recent Monte Carlo studies by Hu and Bentler (1998, 1999), goodness-of-fit statistics behave differently. They showed that different fit indexes were affected differently by sample size, estimation method, and model misspecification. For instance, CFI was found to be moderately sensitive to a measurement model misspecification but very sensitive to structural model misspecification. CFI was also found to be preferable when sample size is small. However, Mc, Steiger's γ, and ε behaved differently depending on the estimation method. The estimation methods had an affect on these fit indexes at small sample sizes but not at large sample sizes. Hence a test of fit should not be based solely on one goodness-of-fit statistic. Instead, Hu and Bentler recommended use of at least two different classes of goodness-of-fit statistics. If the evaluation of a model will be performed using two fit indexes, to ensure power for both values of the fit indexes, the maximum proposed sample size from different fit indexes must be used. For example, for a three-factor CFA model in Table 4 with power of .80 and CFI = .95 and $\lambda_x = .60$, the sample size is 417. The same model with $\varepsilon = .05$ needs 376 cases (Table 7). Therefore, at least 417 cases are needed to ensure a power of .80 for both fit indexes.

More recently, a new type of criticism of fit indexes was given by Browne, MacCallum, Kim, Andersen, and Glaser (2002). They showed that different fit values can be achieved with the same residuals under varying sizes of unique variances. How such weaknesses of fit indexes will affect power and proposed sample size computation has not been studied. The only methods for computing power and sample size that do not use fit indexes are the Satorra–Saris, bootstrap, and Muthén–Muthén methods. However, these methods have other restrictions that prevent them from being implemented in every scenario. The proposed and MBS methods use fit indexes. Fit indexes show consensus most of the time and are often used to judge a model. Until there is a better standard to judge a model, fit indexes will continue be used.

When a model is accepted with a certain fit (e.g., CFI = .95), a model is often judged to be good; that is, it reproduces the relation among variables correctly (see, e.g., Hu & Bentler, 1998, 1999). However, whether or not there was sufficient power to detect the difference between a CFI of .95 and a CFI of 1.00 (perfect fit) is not necessarily known. If a model is accepted simply because it has CFI = .95, this should not mean it has a good fit unless there was enough power to detect a difference between .95 and 1. For instance, if there is not enough power to detect a difference between a model with CFI of .90 and and a model with CFI of 1, then is CFI = .95 really a good fit? Or is it simply that the fit of the model is not different from that of a model with CFI = .90?

ACKNOWLEDGMENTS

An earlier version of this article was submitted as a dissertation for partial satisfaction of the requirement for a doctoral degree at the University of California, Los Angeles. I would like to thank Peter M. Bentler, Stephen G. West, Steven P. Reise, and James Sidanius for their help.

REFERENCES

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107,* 238–246.

Browne, M. W., MacCallum, R. C., Kim, C.-T., Andersen, B. L., & Glaser, R. (2002). When fit indices and residuals are incompatible. *Psychological Methods, 7,* 403–421.

Curran, P. J., Bollen, K. A., Paxton, P., Kirby, J., & Chen, F. (2002). The noncentral chi-square distribution in misspecified structural equation models: Finite sample results from a Monte Carlo simulation. *Multivariate Behavioral Research, 37,* 1–36.

Hu, L.-T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3,* 424–453.

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternative. *Structural Equation Modeling, 6,* 1–55.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1,* 130–149.

MacCallum, R. C., & Hong, S. (1997). Power analysis in covariance structure modeling using GFI and AGFI. *Multivariate Behavioral Research, 32,* 193–210.

Matsueda, R. L., & Bielby, W. T. (1986). Statistical power in covariance structure models. *Sociological Methods, 16,* 120–158.

McDonald, R. P. (1989). An index of goodness-of-fit based on noncentrality. *Journal of Classification, 6,* 97–103.

McDonald, R. P., & Marsh, H. W. (1990). Choosing a multivariate model: Noncentrality and goodness of fit. *Psychological Bulletin, 107,* 247–255.

Mooijaart, A. (2003). Estimating the statistical power in small samples by empirical distributions. In H. Yanai, A. Okada, K. Shigemasu, Y. Kano, & J. J. Meulman (Eds.), *New development in psychometrics* (pp. 149–156). Tokyo: Springer-Verlag.

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling, 9,* 599–620.

Ogasawara, H. (2001). Approximations to the distributions of fit indexes for misspecified structural equation models. *Structural Equation Modeling, 8,* 556–574.

Saris, W. E., & Satorra, A. (1993). Power evaluations in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 181–204). Thousand Oaks, CA: Sage.

Satorra, A. (2003). Power of chi-square goodness-of-fit in structural equation models: The case of non-normal data. In H. Yanai, A. Okada, K. Shigemasu, Y. Kano, & J. J. Meulman (Eds.), *New development in psychometrics* (pp. 57–68). Tokyo: Springer-Verlag.

Satorra, A., & Saris, W. E. (1985). Power of the likelihood ratio test in covariance structure analysis. *Psychometrika, 50,* 83– 90.

Shapiro, A., & Browne, M. W. (1990). On the treatment of correlation structures as covariance structures. *Linear Algebra and Its Applications, 127,* 567–587.

Steiger, J. H. (1989). *EzPATH: A supplementary module for SYSTAT and SYGRAPH.* Evanston, IL: SYSTAT.

Steiger, J. H., & Lind, J. C. (1980, May). *Statistically based tests for the number of common factors.* Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.

Steiger, J. H., Shapiro, A., & Browne, M. W. (1985). On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika, 50,* 253–263.

Yuan, K. H., & Hayashi, K. (2003). Bootstrap approach to interference and power analysis based on three test statistics for covariance structure models. *British Journal of Mathematical and Statistical Psychology, 56,* 93–110.

APPENDIX

SPSS Syntax for Calculating $\delta_{1-\beta}$

```
comment compute noncentrality parameter delta.
comment create two variables in the data editor
```

```
comment df and power first.

set mxloop = 10000.
compute #alpha = .05.
compute #df = df.
compute #power = power.
compute #crit = idf.chisq(1-#alpha, #df).
compute delta = rnd(#crit - #df).
compute #times = 1.
compute #direc = 1.
compute #amount = 10.
loop.
+ compute delta = delta + #direc*#amount.
+ compute #pow = 1-ncdf.chisq(#crit,#df,delta).
+ do if (#direc*(#power - #pow) < 0).
+ compute #times = #times + 1.
+ compute #direc = -1*#direc.
+ compute #amount = #amount/10.
+ end if.
end loop if (#times = 8).
execute.
```

SAS Program for Calculating $\delta_{1-\beta}$

```
data noncent;
alpha = .05;
df = 500;
power = .80;
crit = cinv(1-alpha,df);
delta = round(crit - df);
times = 1;
direc = 1;
amount = 10;
do until (times = 8);
delta = delta + direc*amount;
pow = 1 - probchi(crit, df, delta);
if (direc*(power-pow) < 0) then do;
times = times + 1;
direc = -1 * direc;
amount = amount/10;
end;
end;
proc print data = noncent;
var df alpha power delta;
run;
```